The Choice Set Matters: Some Unpleasant Automobile Simulations

Joe D. Campbell*

February 28, 2026

Abstract

Discrete choice models are used extensively for demand estimation. An important step in demand estimation using these models is the addition of the "outside option", the choice to not purchase any of the specified products. I demonstrate numerically that estimated elasticities are especially sensitive to the choice set definition using data taken directly from Berry et al. (1995) automobile dataset. Moreover, misspecified choice sets continue to impact estimated elasticities even when the model is perfectly specified, meaning that the model perfectly explains observed shares under the correctly specified choice set. The inclusion of market fixed effects ameliorates but does not eliminate this sensitivity to the choice set. These results demonstrate that specification of the choice set, more often art than science, has important implications for estimated elasticities.

1 Introduction

Since the publication of Berry et al. (1995), discrete choice models have become a workhorse for demand estimation in industrial organization and applied microeconomics more generally.

*joe.campbell.pro@gmail.com

Such models take the choice set as given and estimate the parameters of the model using the observed choices. However, in practice, the choice set is often not well-defined. In antitrust litigation, for instance, the choice set is often highly contested (Goodman et al., 2026). In this paper I demonstrate numerically that estimated elasticities are sensitive to the choice set definition using data taken directly from Berry et al. (1995) automobile dataset. I consider two choice set definitions: the full market as defined in the original paper, and a restricted market that excludes luxury vehicles. I estimate both the simpler logit model and the more complex BLP random coefficients model. I find that the estimated elasticities for mass market vehicles are sensitive to the choice set definition for both logit and BLP models. I then consider an alternative dataset with shares modified so that the model is "perfectly" specified, meaning that the model perfectly explains observed shares under the correctly specified choice set. Elasticities from the logit model are not sensitive to the choice set definition, but those from the BLP model remain sensitive to the choice set. These results demonstrate that definition of the choice set, often relegated to the appendix, has important implications for estimated elasticities. Put more forcefully, reported elasticities should be treated as correct only conditional on correct specification of the choice set, even if the underlying model is correctly specified.

2 Models and Estimation

I estimate demand using two models: logit and the BLP random coefficients model. This section sets up utility, gives elasticity formulas, and describes estimation.

2.1 Setup

Consider market $t = 1, \dots, T$ with J_t products indexed by j. Consumer i chooses among inside goods and an outside option (good 0). The outside option captures non-purchase or purchase outside the observed set.

Consumer i's indirect utility from product j in market t is:

$$u_{ijt} = \delta_{jt} + \mu_{ijt} + \varepsilon_{ijt} \tag{1}$$

where δ_{jt} is mean utility, μ_{ijt} captures observed heterogeneity, and ε_{ijt} is an idiosyncratic Type I extreme value shock. I normalize $u_{i0t} = \varepsilon_{i0t}$.

The mean utility has the parametric form:

$$\delta_{jt} = x'_{jt}\beta - \alpha p_{jt} + \xi_{jt} \tag{2}$$

where x_{jt} is observed product characteristics, p_{jt} is price, β and α are parameters, and ξ_{jt} is unobserved product quality.

2.2 Logit Model

In standard multinomial logit, $\mu_{ijt} = 0$ for all consumers. Under the Type I extreme value assumption, the probability that consumer i chooses product j is:

$$s_{jt}(\delta) = \frac{\exp(\delta_{jt})}{1 + \sum_{k=1}^{J_t} \exp(\delta_{kt})} \tag{3}$$

with the outside good share:

$$s_{0t}(\delta) = \frac{1}{1 + \sum_{k=1}^{J_t} \exp(\delta_{kt})} \tag{4}$$

Taking logs and differencing from the outside option gives:

$$\ln(s_{jt}) - \ln(s_{0t}) = \delta_{jt} = x'_{jt}\beta - \alpha p_{jt} + \xi_{jt} \tag{5}$$

2.2.1 Logit Elasticities

Logit has closed-form elasticities. The own-price elasticity for product j is:

$$\eta_{jj} = -\alpha p_{jt}(1 - s_{jt}) \tag{6}$$

The cross-price elasticity of product j with respect to the price of product k ($k \neq j$) is:

$$\eta_{jk} = \alpha p_{kt} s_{kt} \tag{7}$$

These expressions show IIA. The ratio of any two choice probabilities depends only on those two products. So η_{jk} depends on k's price and share, not on j's characteristics.

2.3 BLP Random Coefficients Model

The BLP model relaxes IIA by allowing preferences to vary across consumers:

$$\mu_{ijt} = \sum_{r=1}^{R} x_{jt}^{(r)} \sigma_r \nu_{ir} - p_{jt}\sigma_p \nu_{ip} \tag{8}$$

where $x_{jt}^{(r)}$ is the r-th characteristic with a random coefficient, ν_{ir} and ν_{ip} are taste draws, and σ_r and σ_p govern heterogeneity. Price has a random coefficient, so price sensitivity differs across consumers.

Consumer i chooses product j if $u_{ijt} > u_{ikt}$ for all $k \neq j$. Market share is the integral of individual probabilities over $\nu_i = (\nu_{i1}, \ldots, \nu_{iR}, \nu_{ip})$:

$$s_{jt}(\delta, \theta) = \int \frac{\exp(\delta_{jt} + \mu_{ijt}(\theta))}{1 + \sum_{k=1}^{J_t} \exp(\delta_{kt} + \mu_{ikt}(\theta))} dF(\nu_i) \tag{9}$$

where $\theta = (\sigma_1, \ldots, \sigma_R, \sigma_p)$ collects nonlinear parameters and $F(\cdot)$ is the taste distribution.

2.3.1 BLP Elasticities

BLP elasticities account for heterogeneous responses. The own-price elasticity for product j is:

$$\eta_{jj} = -\frac{p_{jt}}{s_{jt}} \int \frac{\partial s_{ijt}}{\partial p_{jt}} dF(\nu_i) = -\frac{p_{jt}}{s_{jt}} \int (\alpha + \sigma_p \nu_{ip}) s_{ijt}(1 - s_{ijt}) dF(\nu_i) \tag{10}$$

where $s_{ijt} = s_{ijt}(\delta, \theta)$ is individual i's choice probability.

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

$$\eta_{jk} = \frac{p_{kt}}{s_{jt}} \int \frac{\partial s_{ijt}}{\partial p_{kt}} dF(\nu_i) = -\frac{p_{kt}}{s_{jt}} \int (\alpha + \sigma_p \nu_{ip}) s_{ijt} s_{ikt} dF(\nu_i) \tag{11}$$

2.4 Estimation

I estimate both logit and BLP using two-step GMM with standard demand-side instruments based on characteristics of other products in the same market. When I include market fixed effects, I interact instruments with market dummies so they are not absorbed. For BLP, mean utilities are recovered from the share inversion at each candidate value of the nonlinear parameters before minimizing the GMM objective. Implementation uses `PyBLP`, following Conlon and Gortmaker (2020).

2.5 Model Specifications

I start from two baseline demand specifications. The first is logit without market fixed effects,

$$\ln(s_{jt}) - \ln(s_{0t}) = x'_{jt}\beta - \alpha p_{jt} + \xi_{jt}, \tag{12}$$

estimated by two-step GMM with BLP-style demand instruments. The second is the corresponding BLP specification,

$$\delta_{jt} = x'_{jt}\beta - \alpha p_{jt} + \xi_{jt}, \tag{13}$$

where δ_{jt} is recovered from inversion and nonlinear parameters are chosen by GMM. Following Berry et al. (1995), random coefficients are on price, horsepower, size, and air conditioning:

$$\theta = (\sigma_p, \sigma_{\text{hp}}, \sigma_{\text{size}}, \sigma_{\text{air}}).$$

As in the original paper, the marginal utility of price varies with inverse income, so the effective price coefficient α_i is scaled by consumer-specific income.

After baseline models, I estimate versions with market fixed effects. For logit:

$$\ln(s_{jt}) - \ln(s_{0t}) = x'_{jt}\beta - \alpha p_{jt} + \gamma_t + \tilde{\xi}_{jt}, \tag{14}$$

and for BLP:

$$\delta_{jt} = x'_{jt}\beta - \alpha p_{jt} + \gamma_t + \tilde{\xi}_{jt}. \tag{15}$$

In practice, market dummies enter the linear stage and instruments are interacted with market indicators so they are not absorbed.

2.6 Summary

Table 1 reports four specifications: baseline logit, baseline BLP, and each with market fixed effects. I estimate each on two choice sets (full and mass-market only) and two data types (original and ideal), for $4 \times 2 \times 2 = 16$ estimates.

Table 1: Model Specifications

Specification	Model	Market FE
1	Logit	No
2	Logit	Yes
3	BLP Random Coefficients	No
4	BLP Random Coefficients	Yes

3 Data and Empirical Design

I use the Berry et al. (1995) automobile data from 1971 to 1990. Each observation is a model-year product, and each year is a market. The data include prices, market shares, product characteristics, and firm identifiers.

I compare two choice sets. The first is the full market with all products. The second excludes luxury brands and keeps mass-market products only. Excluded luxury products are absorbed into the outside option.

To separate structural effects from econometric effects, I use an ideal-data exercise. For each specification, I first estimate parameters on the full market and recover unobserved quality. I then set unobserved quality to zero and recompute model-implied shares, holding observed characteristics and prices fixed.

This represents the best-case scenario for each model. If, under ideal data, differences across choice sets vanish, it indicates the model is capable of fully accounting for choice set restrictions—any observed differences in the original data are due to misspecification. If differences persist even with ideal data, this reflects an inherent sensitivity of the model to the choice set definition.

Combining two choice sets, two data types, and four specifications yields 16 estimates. I compare elasticities across these different configurations.

4 Theory

This section formalizes why misspecifying the choice set can be absorbed in simple logit (under correct specification), but not in random coefficients logit.

4.1 A Gumbel Aggregation Fact

Let $\{\varepsilon_{ijt}\}_{j=1}^{J_t}$ be i.i.d. Type I extreme value with CDF

$$F(\varepsilon) = \exp\{-\exp(-\varepsilon)\}.$$

Define the inclusive value in market t:

$$I_t(\delta_t) \equiv \log\left(\sum_{k=1}^{J_t} \exp(\delta_{kt})\right).$$

Proposition 1 (Maximum of shifted Type I EV shocks)**.** *If* $M_t \equiv \max_{1\le k\le J_t}\{\delta_{kt} + \varepsilon_{ikt}\}$*, then*

$$\Pr(M_t \le m) = \exp\{-\exp[-(m - I_t(\delta_t))]\},$$

so M_t *is also Type I extreme value with location parameter* $I_t(\delta_t)$*.*

Proof. Using independence,

$$\Pr(M_t \le m) = \prod_{k=1}^{J_t} \Pr(\delta_{kt} + \varepsilon_{ikt} \le m) = \prod_{k=1}^{J_t} \exp\{-\exp[-(m - \delta_{kt})]\}.$$

Therefore

$$\Pr(M_t \le m) = \exp\left\{-\sum_{k=1}^{J_t} \exp[-(m - \delta_{kt})]\right\} = \exp\left\{-\exp\left[-\left(m - \log\sum_{k=1}^{J_t} e^{\delta_{kt}}\right)\right]\right\}.$$

This is Type I extreme value with location $I_t(\delta_t)$. □

So the key closure property is about the *maximum* of shifted Type I EV shocks (equivalently, the log-sum-exp inclusive value), not the literal arithmetic sum of shocks.

4.2 Implication for Logit with Choice-Set Restriction

Let F denote the full set of inside goods and $M \subset F$ the observed subset after dropping some alternatives (e.g., luxury cars). For $j \in M$, the full-model logit probability is

$$s_{jt}^F = \frac{\exp(\delta_{jt})}{1 + \sum_{k \in F} \exp(\delta_{kt})}.$$

Conditioning on the event that a consumer chooses some product in M or the outside option, the restricted logit probability is

$$s_{jt}^M = \frac{\exp(\delta_{jt})}{1 + \sum_{k \in M} \exp(\delta_{kt})} = \frac{\exp(\delta_{jt} - \Delta_t)}{1 + \sum_{k \in F} \exp(\delta_{kt})},$$

where

$$\Delta_t \equiv \log \left(\frac{1 + \sum_{k \in F} \exp(\delta_{kt})}{1 + \sum_{k \in M} \exp(\delta_{kt})} \right).$$

Hence dropping products shifts all surviving inside-good indices by the same market-level constant $-\Delta_t$. With market fixed effects this shift is absorbed, leaving slope parameters unchanged under correct specification.

Proposition 2 (Choice-set invariance in correctly specified logit with market FE)**.** *Suppose data are generated by standard logit with common slope parameters across markets and products. If estimation includes market fixed effects, then estimating on F or on any subset $M \subset F$ yields the same (β, α) for products in M.*

Proof. From the previous display, for each $j \in M$:

$$\log s_{jt}^M - \log s_{0t}^M = \log s_{jt}^F - \log s_{0t}^F - \Delta_t,$$

where Δ_t is constant within market t. In the linear index representation

$$\log s_{jt} - \log s_{0t} = x_{jt}'\beta - \alpha p_{jt} + \gamma_t,$$

this only changes γ_t to $\gamma_t - \Delta_t$. Therefore within-market variation, and thus identified slopes (β, α), are unchanged. □

4.3 Why Random Coefficients (BLP) Is Different

In random coefficients logit,

$$u_{ijt} = \delta_{jt} + \mu_{ijt}(\nu_i) + \varepsilon_{ijt},$$

with $\mu_{ijt}(\nu_i)$ containing heterogeneous tastes, including random price sensitivity. For a fixed draw ν_i, dropping products still induces a log-sum-exp correction, but now it is

$$\Delta_{it}(\nu_i) = \log\left(\frac{1 + \sum_{k\in F}\exp(\delta_{kt} + \mu_{ikt}(\nu_i))}{1 + \sum_{k\in M}\exp(\delta_{kt} + \mu_{ikt}(\nu_i))}\right),$$

which varies across consumers through ν_i.

Proposition 3 (No general invariance in random coefficients logit)**.** *With non-degenerate random coefficients, restricting from F to M does not in general induce a common market-level additive shift in mean utility. Therefore market fixed effects cannot, in general, absorb the effect of omitted alternatives, and estimated substitution patterns can change even under correct specification.*

Proof. The correction term is $\Delta_{it}(\nu_i)$, not Δ_t. Since Δ_{it} depends on ν_i, there is no single constant that maps all individual probabilities under F to probabilities under M. Aggregating over $F(\nu)$ then changes the moments that identify the random-coefficient distribution. In particular, omitting high-price/high-quality products removes information about consumers with low price sensitivity and high taste for quality; this changes identified heterogeneity and therefore elasticities among remaining products. □

Intuitively, in simple logit, omitted products are absorbed into a market intercept via the inclusive value. In BLP, omitted products alter which parts of the heterogeneity distribution are identified, so the effect is not a single intercept shift.

4.4 Elasticities Depend on Other Prices Even with Fixed Parameters and Shares

The BLP cross- and own-price formulas can be written as

$$\eta_{jk,t} = -\frac{p_{kt}}{s_{jt}} \int (\alpha_i)\, s_{ij,t}\, s_{ik,t}\, dF(\nu_i), \qquad \alpha_i \equiv \alpha + \sigma_p \nu_{ip}.$$

Holding (α, σ_p) and observed aggregate shares s_{jt} fixed does *not* generally pin down $\eta_{jk,t}$ because individual probabilities $s_{ij,t}$ and $s_{ik,t}$ depend on the full denominator

$$1 + \sum_{\ell \in \mathcal{J}_t} \exp\bigl(\delta_{\ell t} + \mu_{i\ell t}\bigr),$$

which moves when the price of any product ℓ changes.

Therefore, even if two specifications match the same aggregate shares and use the same estimated parameters, elasticities can differ if the surrounding price vector (or available alternatives) differs. Elasticities are equilibrium objects of the entire choice set, not functions of each product's own (p_{jt}, s_{jt}) alone.

5 Empirical Results

I now present empirical results. For each specification, I estimate elasticities on the full market and the mass-market-only sample. I do this for original and ideal data, for 16 total estimates.

5.1 Original Data

On original data, logit without fixed effects moves away from the 45-degree line when the choice set is restricted (Figure 1). With fixed effects, the differences become even larger—a result of misspecification, not a contradiction of the earlier proposition, which holds only

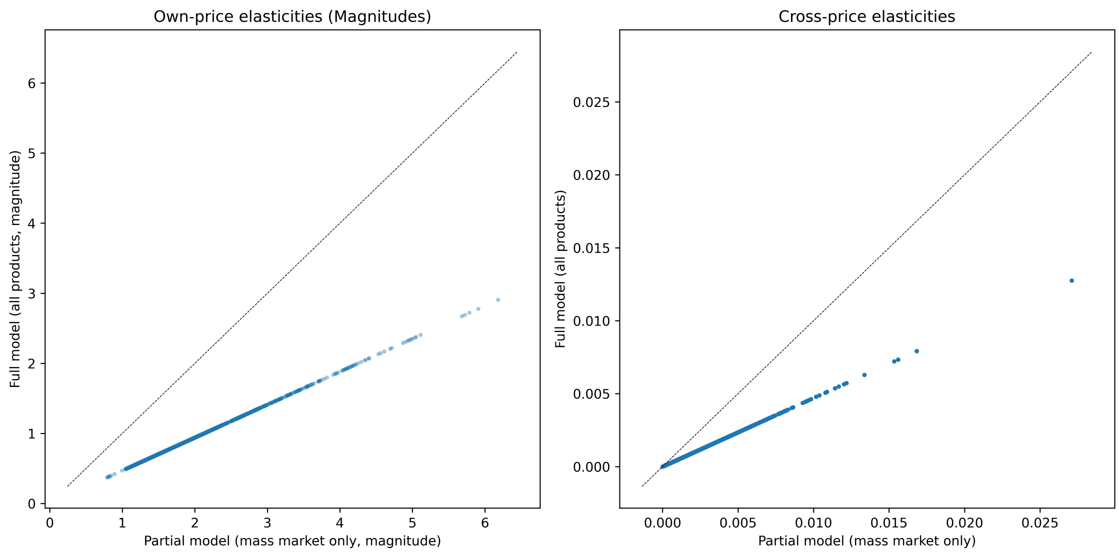

Figure 1: Elasticity Comparison: Logit without Market Fixed Effects (Original Data). Left panel shows own-price elasticities; right panel shows cross-price elasticities. Points represent elasticities estimated on the full market (x-axis) versus mass-market only (y-axis). The 45-degree line indicates perfect agreement.

under correct specification.

BLP behaves differently. In Figures 3 and 4, the partial (restricted) model actually yields *less* elastic (smaller-magnitude) own-price responses for products near the omitted luxury segment, relative to the full model. Cross-price elasticities among remaining products are not generally more extreme either, and market fixed effects make little difference to this pattern.

5.2 Perfect Specification

With perfect specification, the results for logit demonstrate that restricting the choice set has little effect on estimated elasticities, both with and without fixed effects (see Figures 5 and 6). The estimates from the mass-market-only sample closely match those from the full market in both cases.

For BLP, the gap remains significant under perfect specification (Figures 7 and 8), though the magnitude is smaller than with original data. Without fixed effects, the BLP estimates

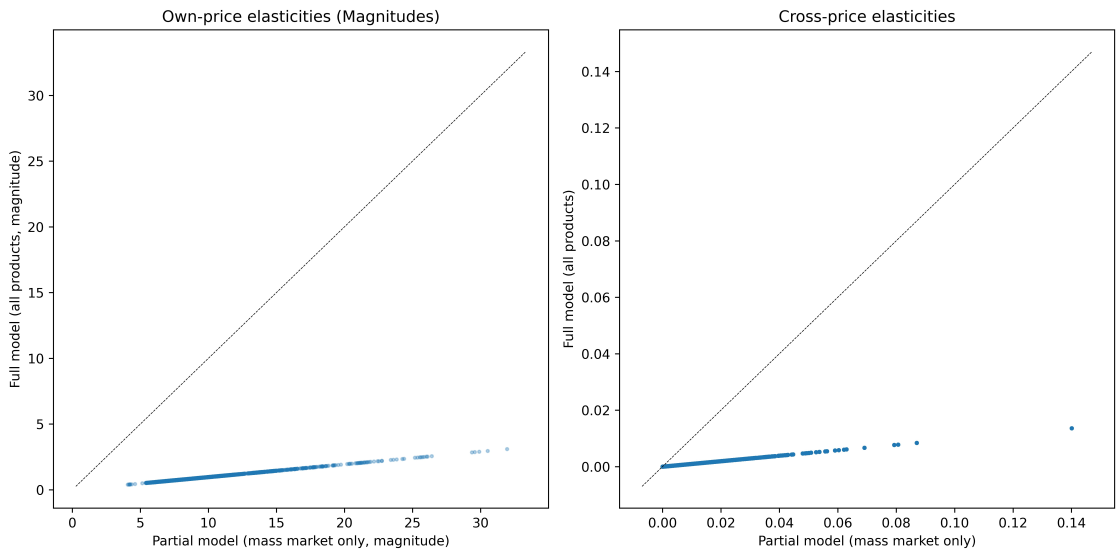

Figure 2: Elasticity Comparison: Logit with Market Fixed Effects (Original Data). In the original-data estimates, differences between full-market and mass-market elasticities remain visible, and can be larger than in the no-fixed-effects case.

from the mass-market-only sample are less elastic (closer to zero) than those from the full market. With fixed effects, however, the direction of the difference is less clear.

These changes are economically meaningful. Own-price elasticities move, and cross-price elasticities often move more. Changes are larger for products closer to the excluded segment in characteristic space. In many markets, own-price differences are on the order of 20–30% in absolute value, and cross-price differences can be larger. So the gap is not cosmetic; it is large enough to materially change substitution predictions. To isolate the mechanism, the next section varies prices of omitted products while holding observed-product prices and estimated preference parameters fixed.

6 Price Sensitivity and Unobserved-Good Prices

This section provides a direct mechanism check using original shares. I vary prices of omitted luxury goods by factors of 0.25, 1, and 4 and then recompute own-price elasticity magnitudes for non-luxury products. Figure 9 compares these magnitudes across the three counterfac-

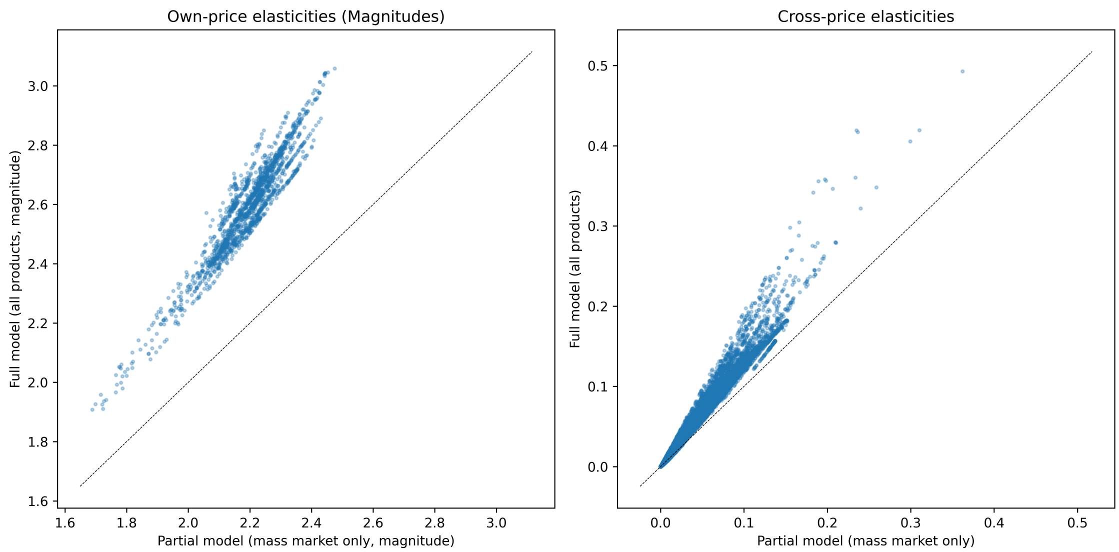

Figure 3: Elasticity Comparison: BLP without Market Fixed Effects (Original Data). The BLP model shows greater sensitivity to choice set restrictions than logit, with substantial scatter around the 45-degree line.

tuals.

The pattern is clear: when omitted goods are made more expensive, observed non-luxury products become closer substitutes for those omitted goods, and own-price elasticity magnitudes increase. Consumers therefore appear more price sensitive among observed products.

When omitted goods are made cheaper, substitution toward those omitted goods absorbs demand reallocation that would otherwise occur within the observed set, so own-price elasticity magnitudes for observed products decrease. Consumers therefore appear less price sensitive among observed products.

This is exactly the channel emphasized by the choice-set argument: elasticities for observed products depend on prices of alternatives that may be unobserved to the researcher.

7 Conclusion

This paper demonstrates how incomplete choice sets affect demand elasticities in the BLP automobile data. By comparing logit and BLP on original and ideal data, I eliminate

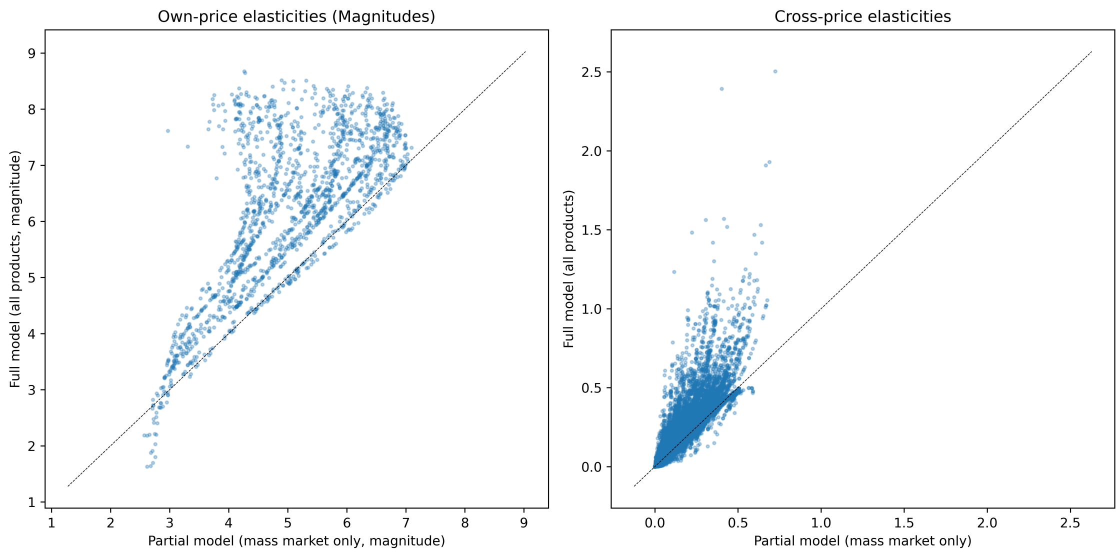

Figure 4: Elasticity Comparison: BLP with Market Fixed Effects (Original Data). Unlike logit, market fixed effects do not eliminate BLP's sensitivity to choice set definition.

misspecification as a source of sensitivity.

The main finding is an asymmetry across model classes. Logit with market fixed effects is close to invariant under correct specification. BLP remains sensitive to choice-set definition even with fixed effects and even on ideal data.

These findings relate to recent theoretical work on random utility models with incomplete choice sets by Kono et al. (2023).

The practical implications are important. BLP-style models are used in merger and antitrust work where substitution patterns matter directly. Even modest choice-set truncation can lead to meaningful errors in elasticities. The definition of the choice set is a key step in identifying elasticities, rather than a minor detail to be relegated to the appendix.

A natural next step is to consider partial identification of elasticities when the choice set is only imperfectly observed.

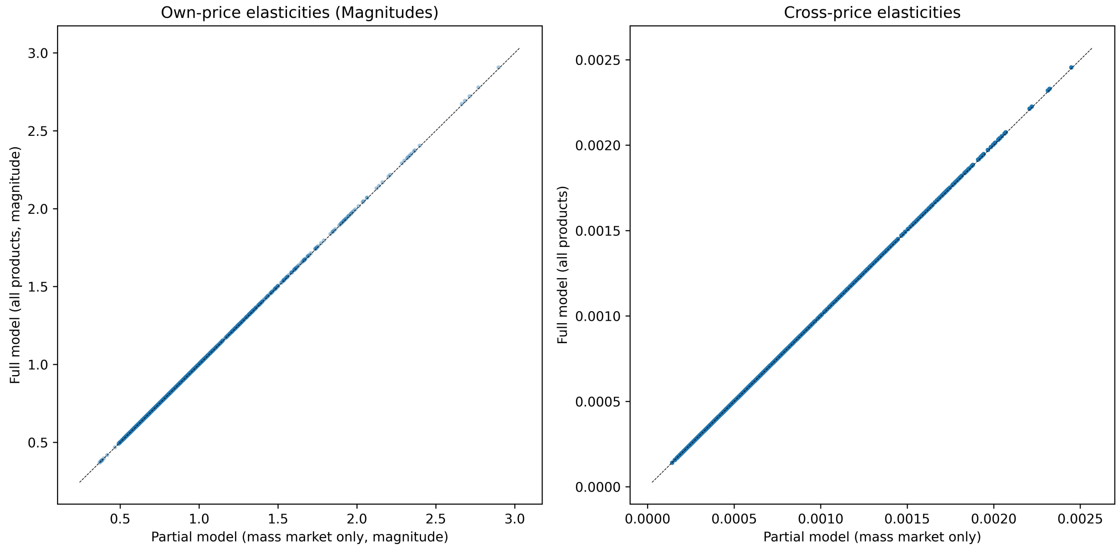

Figure 5: Elasticity Comparison: Logit without Market Fixed Effects (Perfect Specification). With $\xi_{jt} = 0$ by construction, econometric issues are eliminated, revealing the structural sensitivity to choice set definition.

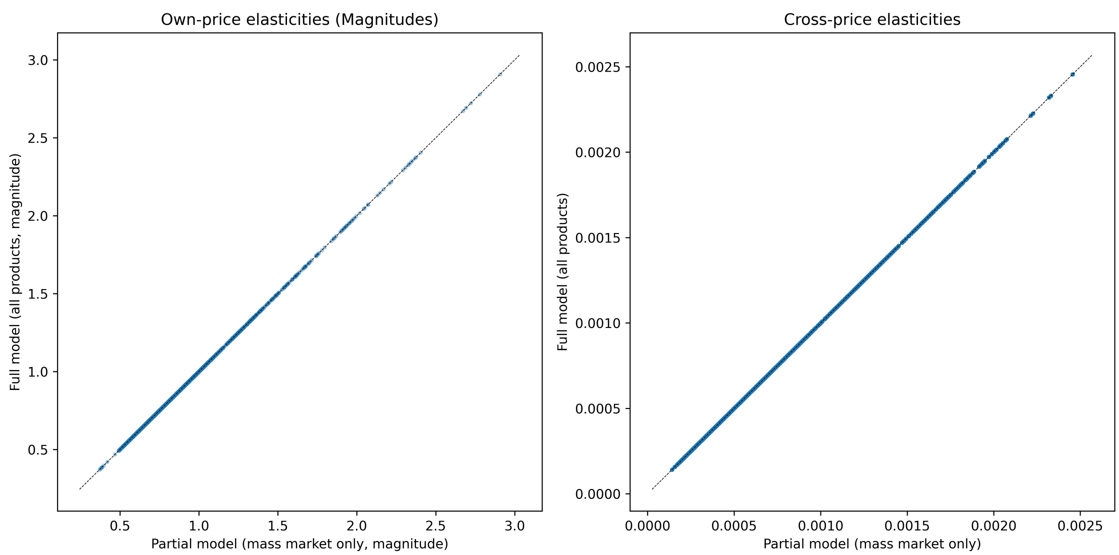

Figure 6: Elasticity Comparison: Logit with Market Fixed Effects (Perfect Specification). Near-perfect clustering on the 45-degree line confirms that logit with market fixed effects is structurally robust to choice set restrictions under correct specification.

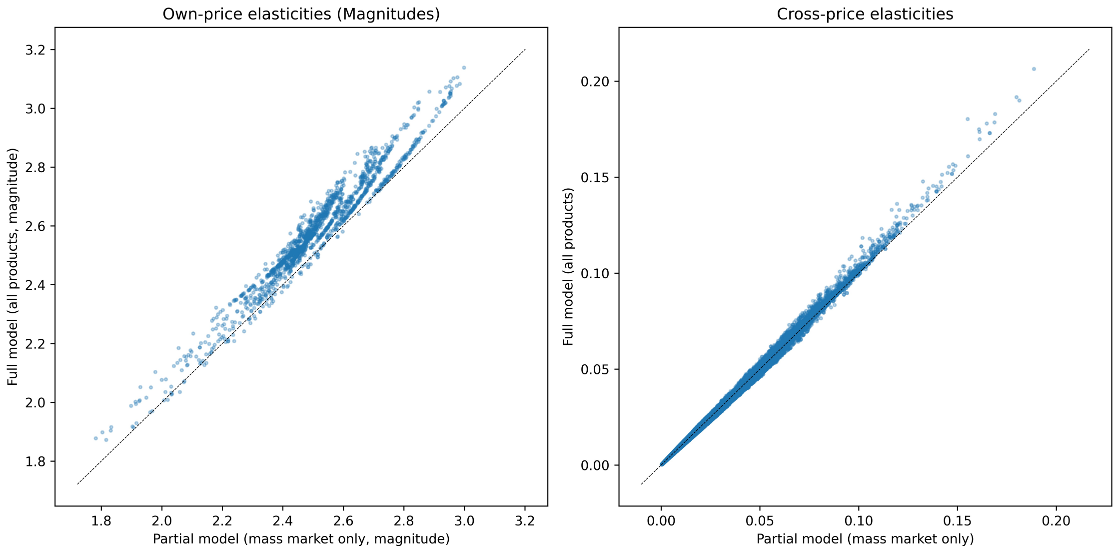

Figure 7: Elasticity Comparison: BLP without Market Fixed Effects (Perfect Specification). Even with perfect specification ($\xi_{jt} = 0$), BLP remains sensitive to choice set restrictions, demonstrating that the vulnerability is structural rather than econometric.

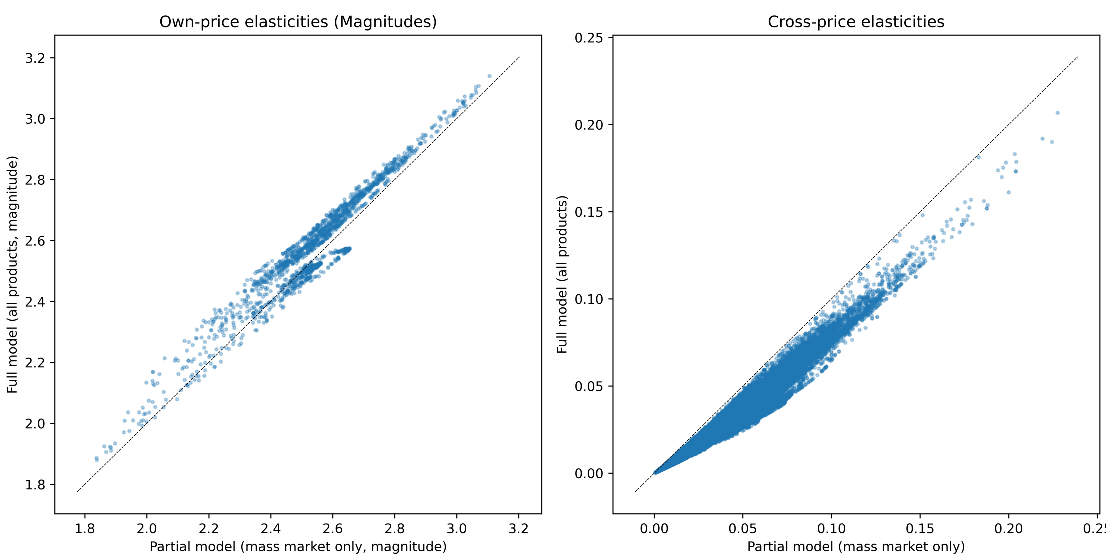

Figure 8: Elasticity Comparison: BLP with Market Fixed Effects (Perfect Specification). Market fixed effects do not resolve BLP's structural sensitivity to choice set definition. The persistent scatter demonstrates that excluding luxury brands fundamentally affects preference distribution identification.

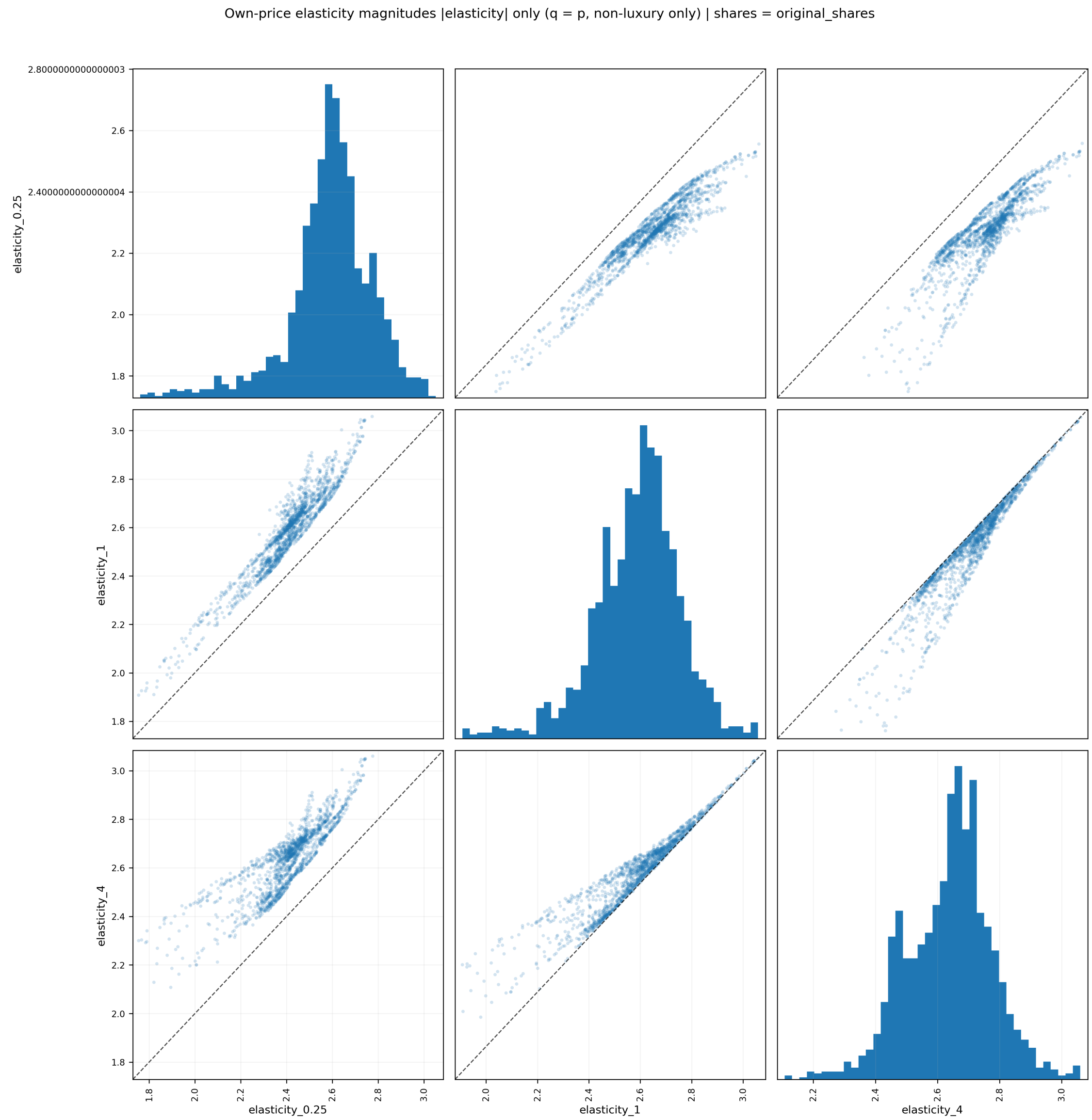

Figure 9: Own-price elasticity magnitudes under counterfactual omitted-good prices (Original Shares). Each panel compares non-luxury own-price elasticity magnitudes across omitted-luxury price factors 0.25, 1, and 4. Points above the 45-degree line indicate larger elasticity magnitudes under the y-axis counterfactual.

A Additional Figures: Own-Price Elasticities vs. Prices

This appendix presents scatter plots of own-price elasticities against product prices for all model specifications and data types. Each figure compares estimates from the full market (blue) and mass-market only (red) choice sets. These plots reveal how the relationship between prices and own-price elasticities changes with choice set restrictions.

A.1 Original Data

A.2 Ideal Data

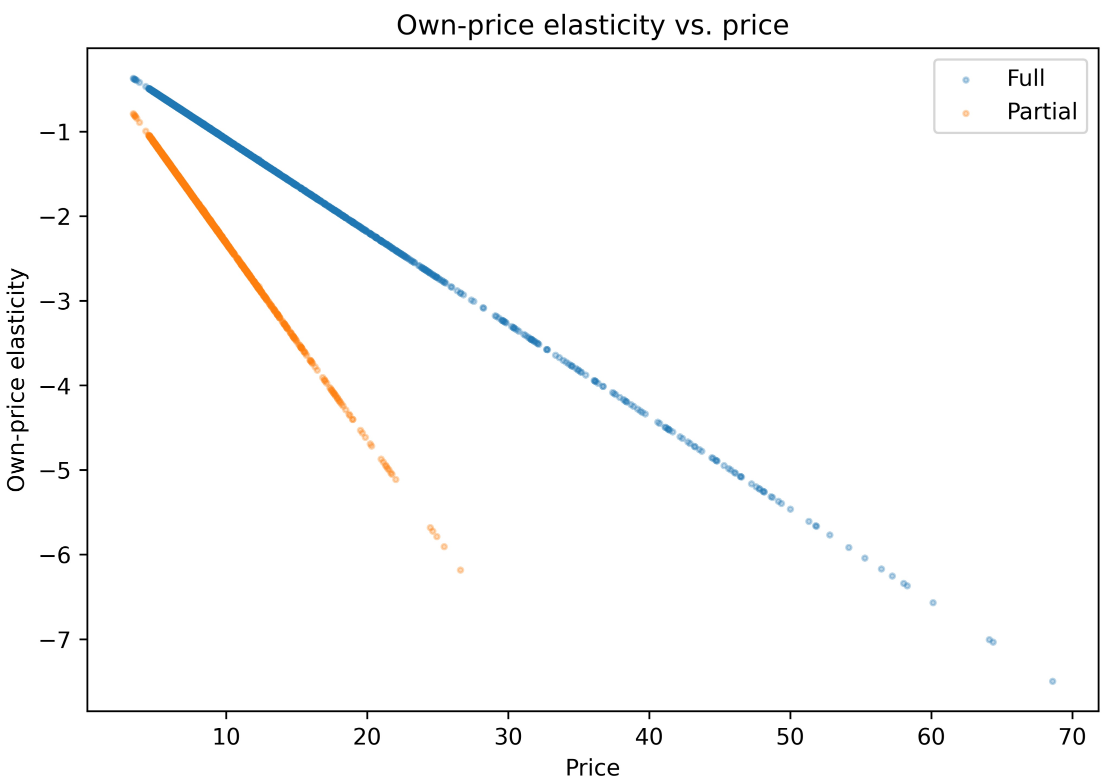

Figure 10: Own-Price Elasticities vs. Prices: Logit without Market Fixed Effects (Original Data)

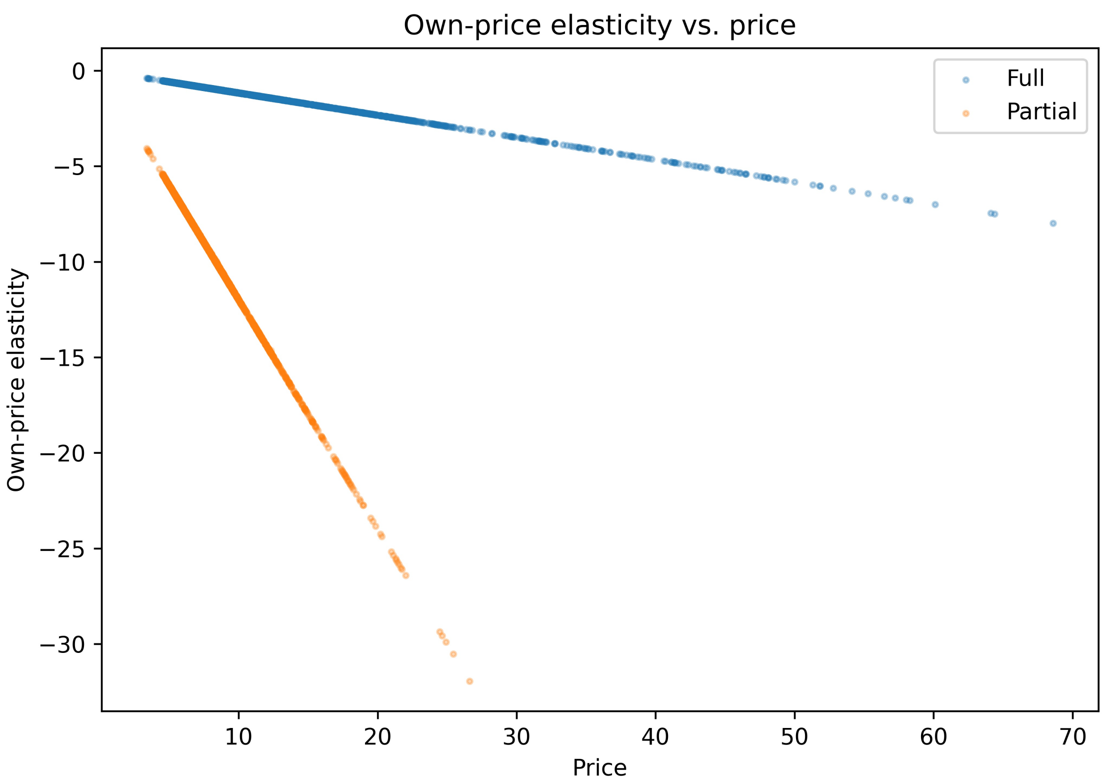

Figure 11: Own-Price Elasticities vs. Prices: Logit with Market Fixed Effects (Original Data)

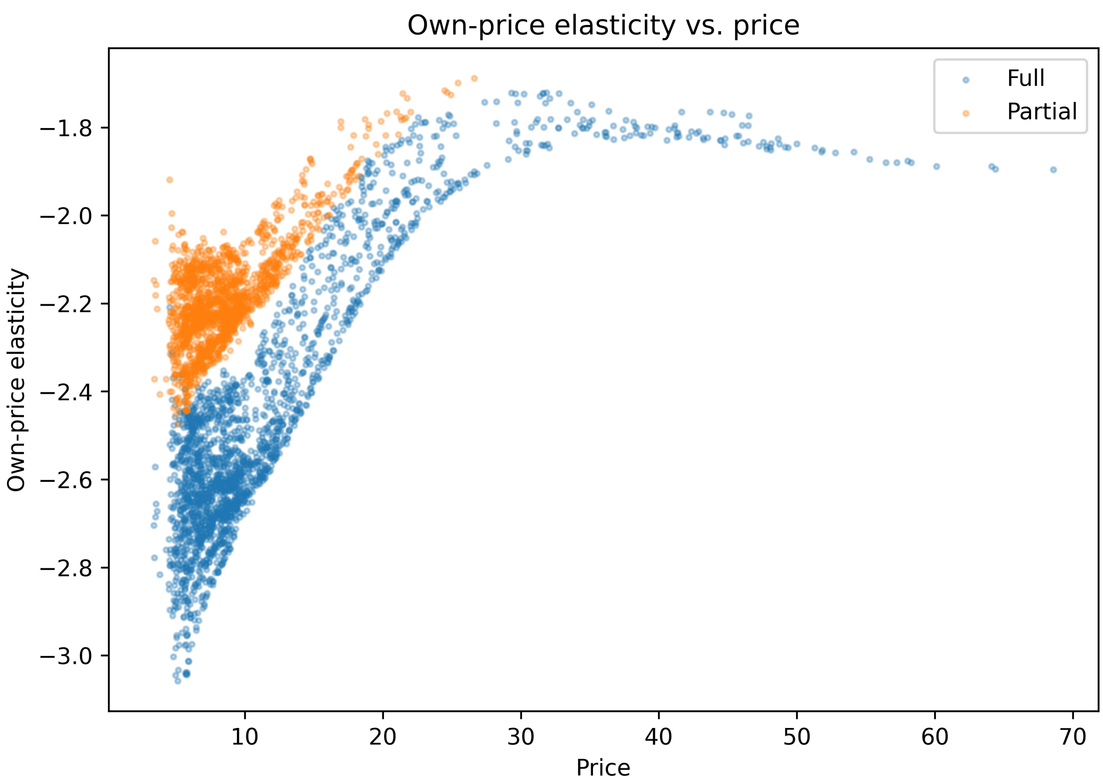

Figure 12: Own-Price Elasticities vs. Prices: BLP without Market Fixed Effects (Original Data)

Figure 13: Own-Price Elasticities vs. Prices: BLP with Market Fixed Effects (Original Data)

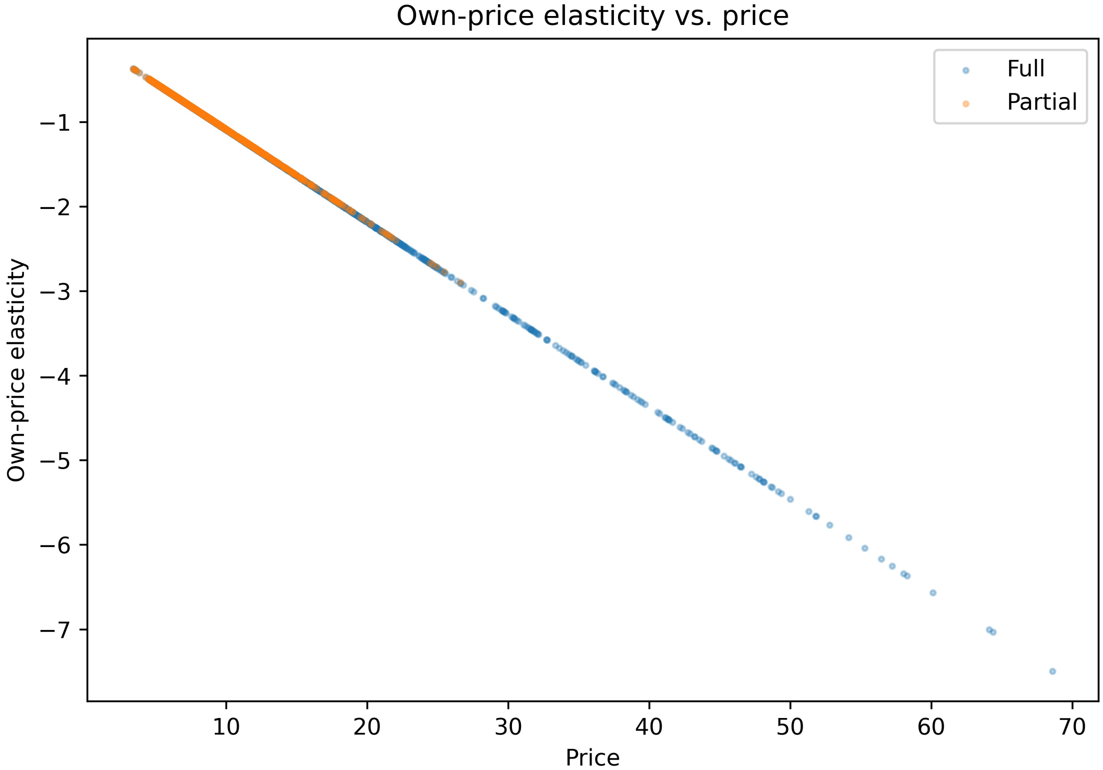

Figure 14: Own-Price Elasticities vs. Prices: Logit without Market Fixed Effects (Ideal Data)

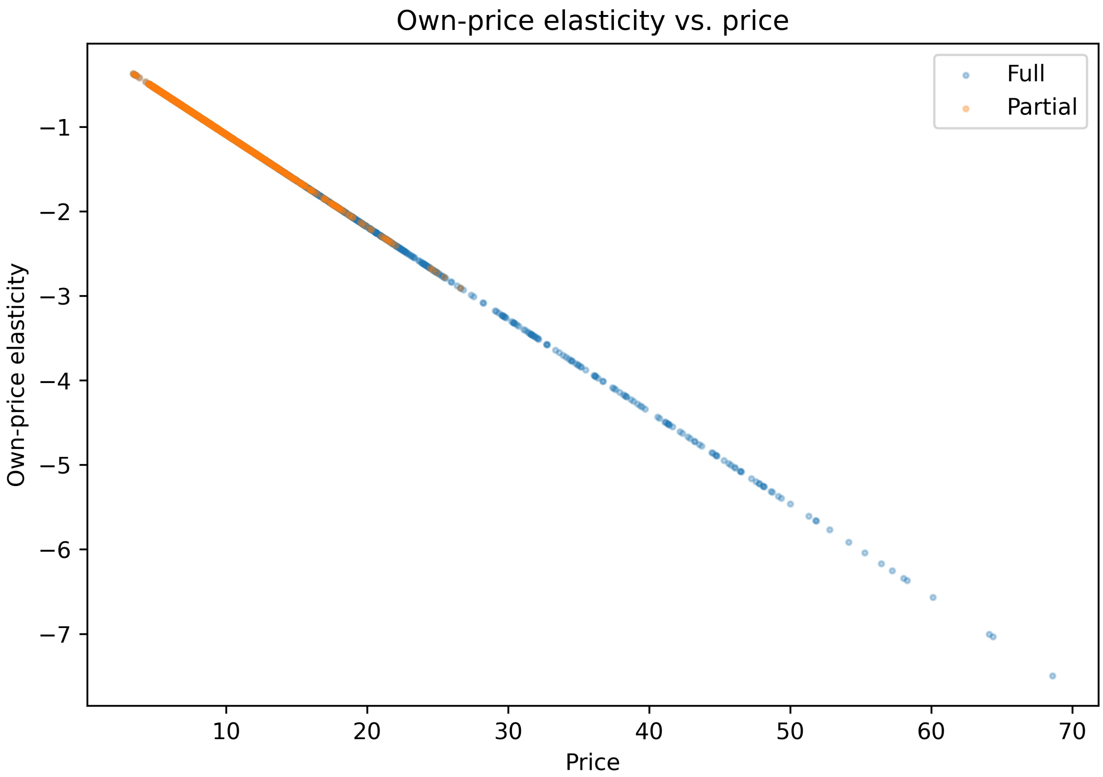

Figure 15: Own-Price Elasticities vs. Prices: Logit with Market Fixed Effects (Ideal Data)

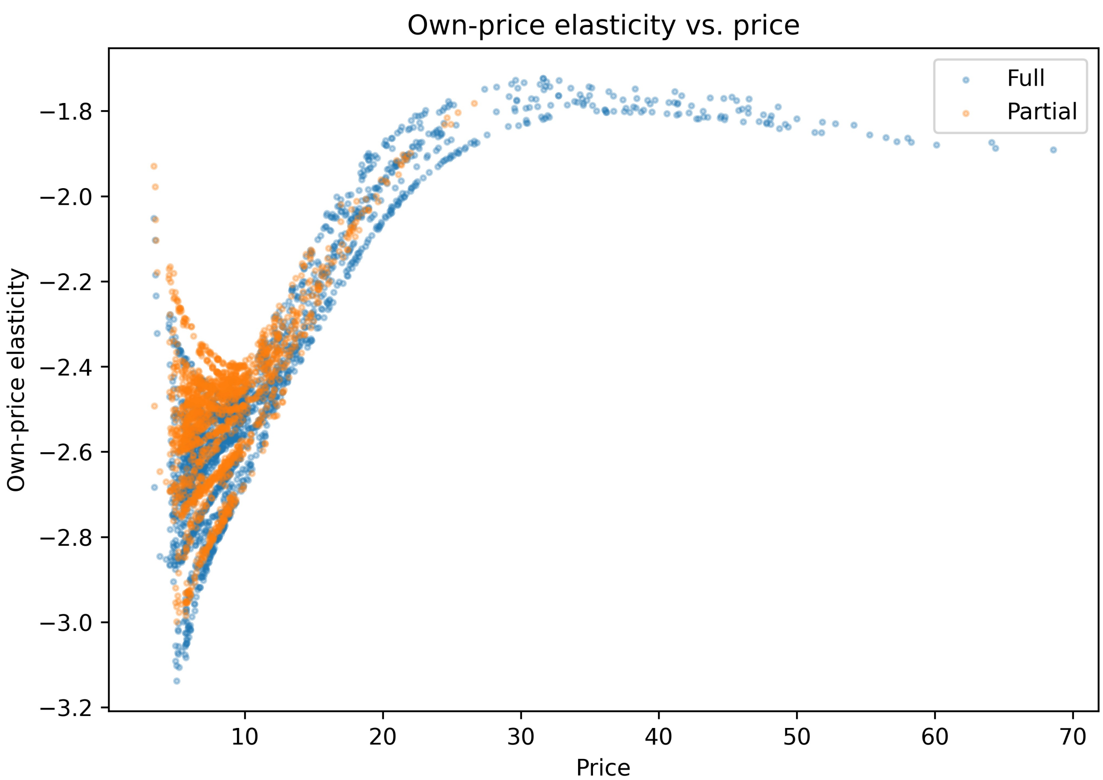

Figure 16: Own-Price Elasticities vs. Prices: BLP without Market Fixed Effects (Ideal Data)

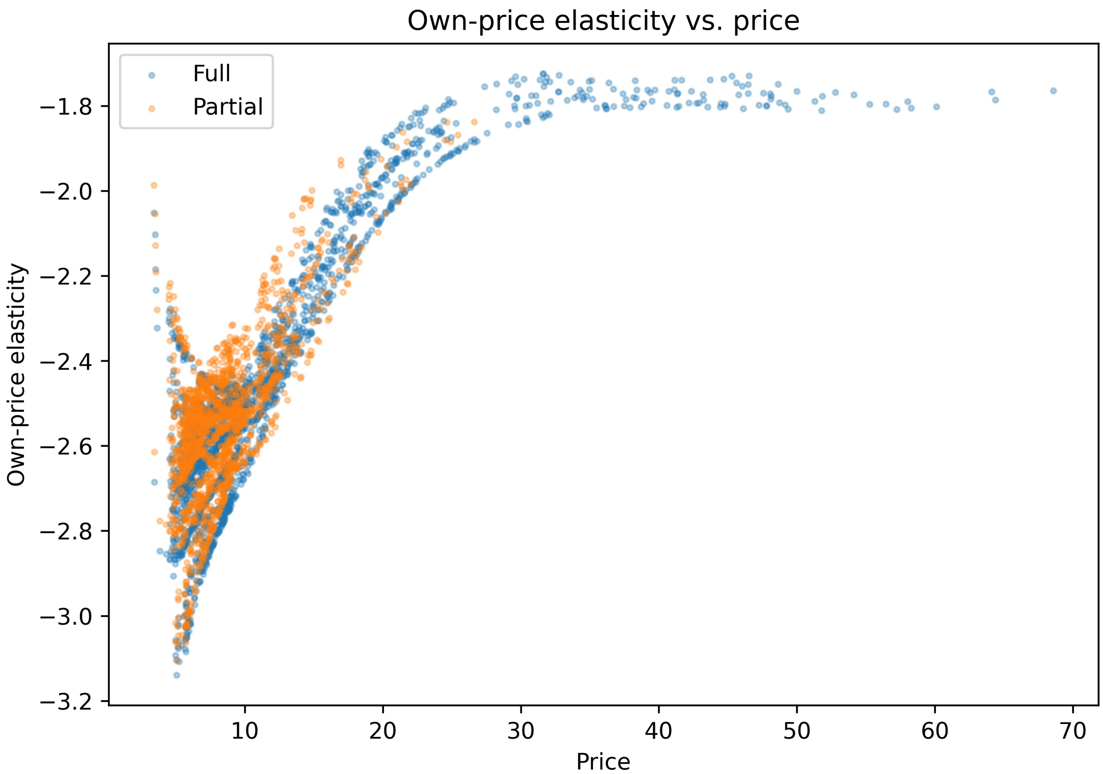

Figure 17: Own-Price Elasticities vs. Prices: BLP with Market Fixed Effects (Ideal Data)

References

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

Goodman, J., L. H. de Frahan, J. E. Holz, J. A. List, N. MacMenamin, E. C. McKay, M. Mogstad, S. Sadoff, and H. Sider (2026): "Consumer Demand and Market Competition with Time-Intensive Goods," NBER Working Paper No. 34743.

Kono, H., K. Saito, and A. Sandroni (2023): "Random Utility with Unobservable Alternatives," arXiv preprint arXiv:2302.03913.

Conlon, C., and J. Gortmaker (2020): "Best Practices for Differentiated Products Demand Estimation with `PyBLP`," *RAND Journal of Economics*, 51(4), 1108–1161.

Nevo, A. (2000): "A Practitioner's Guide to Estimation of Random-Coefficients Logit Models of Demand," *Journal of Economics & Management Strategy*, 9(4), 513–548.