cemmap

centre for microdata methods and practice

Estimating demand with recentered instruments

Kirill Borusyak
Mauricio Caceres Bravo
Peter Hull

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP10/25

Economic and Social Research Council

Estimating Demand with Recentered Instruments

Kirill Borusyak
UC Berkeley and NBER

Mauricio Caceres Bravo
Brown

Peter Hull
Brown and NBER*

April 2025

Abstract

We develop a new approach to estimating flexible demand models with exogenous supply-side shocks. Our approach avoids conventional assumptions of exogenous product characteristics, putting no restrictions on product entry, despite using instrumental variables that incorporate characteristic variation. The proposed instruments are model-predicted responses of endogenous variables to the exogenous shocks, recentered to avoid bias from endogenous characteristics. We illustrate the approach in a series of Monte Carlo simulations.

*Contact: k.borusyak@berkeley.edu, mauricio_caceres_bravo@brown.edu, peter_hull@brown.edu. We are grateful to Dan Ackerberg, Dmitry Arkhangelsky, Matthew Backus, Jean-Francois Houde, Jesse Shapiro, and numerous seminar participants for helpful comments. Adamson Bryant and Jacob Lefler provided excellent research assistance.

1 Introduction

Many economic analyses depend on accurate estimates of the demand for differentiated products. Prominent examples from the field of industrial organization (IO) include measuring welfare effects of mergers or new products and testing models of firm conduct; trade economists similarly use demand estimates to measure welfare effects of new tariffs and gains from trade or internal migration. Often, these analyses leverage structural models of demand that allow for rich and realistic substitution patterns—such as the mixed multinomial logit model popularized by Berry et al. (1995). To estimate these models with market-level data, researchers need multiple instrumental variables (IVs) that address the endogeneity of prices and other terms capturing the substitution patterns.

This paper develops a new approach to constructing powerful instruments for popular demand models by leveraging a set of exogenous supply-side shocks (for brevity, "cost shocks") like input price changes, new taxes or subsidies, markup regulations, or certain productivity and ownership shocks. Such shocks are increasingly found in empirical applications, where they serve as a natural instrument for price.[1] However, even when plausibly exogenous cost shocks are available, researchers typically identify the parameters governing substitution patterns (so-called "nonlinear" parameters) using other IVs constructed from observed characteristics of competing products. Prominent examples include nest size instruments in nested logit models, "BLP instruments" which average or sum the characteristics of a product's competitors, and more sophisticated versions like the efficient IVs of Berry et al. (1999) and the differentiation IVs of Gandhi and Houde (2020). Such instruments rely on the econometric exogeneity of product characteristics: a strong assumption that is often inconsistent with natural models of product entry (Ackerberg and Crawford, 2009; Petrin et al., 2022). Characteristic-based IVs can also be weak, and lack across-market variation when all markets have the same products (Reynaert and Verboven, 2014; Nevo, 2001).[2]

We propose instruments that combine product characteristics and cost shocks in a particular way: to predict the response of the demand model's endogenous variables to the exogenous shocks. These IVs are motivated by thinking of the model as structuring "spillover effects," of exogenous changes to product prices on the market shares of other products. For example, the nested logit model structures spillovers with a parameter σ that governs whether, when prices exogenously rise, consumers substitute primarily to local competitors in a product's nest or more evenly to all unaffected products in the market. To distinguish between those cases and thus identify σ, we propose instruments which predict how a product's within-nest market share changes in response to a set of cost shocks. A simple IV in this spirit is the the cost shock of a product less the average shock in its nest. We show how such IVs can be generally constructed from first-order approximations to model-implied responses, yielding shift-share instruments with cost shocks as the exogenous "shifts." We also propose a novel instrument construction from exact model-based predictions. We build intuition for these constructions in mixed logit models by considering in-

[1] Examples include Berry et al. (1999) and Goldberg and Verboven (2001) (exchange rate shocks), Li (2016) (subsidies), Miller and Weinberg (2017) (merger shocks), and Nakamura and Zerom (2010) (productivity shocks).

[2] BLP instruments are sometimes also used to estimate the price sensitivity parameter, when cost shocks are not available. Armstrong (2016) studies the weak instrument problem that arises in that context.

struments constructed to predict the impact of shocks via small "nonlinear" parameters (i.e., in a "local to logit" approximation, similar to Salanie and Wolak (2022)).

Instruments constructed this way are generally complex formulas of the exogenous cost shocks and the likely endogenous product characteristics. To avoid bias from the latter, we follow Borusyak and Hull (2023) in recentering the IVs: i.e., subtracting their expectation over the data-generating process of exogenous shocks. For example, a researcher may simulate this process by permuting observed shocks across comparable products; she could then recenter any formula IV by subtracting from each product's instrument value the average value across these counterfactual shocks, holding the characteristics of all products fixed. When the instruments are constructed from first-order approximations (i.e. as shift-share IVs) recentering is simpler as it only requires specifying and adjusting for the conditional mean of the shocks. In general, recentering ensures our IVs derive their validity only from the exogeneity of cost shocks—even though they also derive power from product characteristics.

We formalize this approach in a broad class of demand models, which includes both conventional mixed and nested logit from IO as well as analogous constant elasticity of substitution (CES) models from trade. We focus on estimating the parameters that govern own- and cross-price elasticities, which are central to many important policy counterfactuals. Consistency of recentered IV estimates follows when there are either many uncorrelated markets or many uncorrelated shocks that can affect multiple markets jointly while inducing sufficient across-product variation in the instruments. Unlike conventional characteristic-based IVs, our instruments can yield consistent estimates when all markets have the same products and product fixed effects are included. Asymptotic normality of our estimators follows with many market clusters; we further extend results in Adão et al. (2019) and Borusyak et al. (2022b) to show how many-shock asymptotic inference can be conducted with shift-share instruments. We characterize the asymptotically efficient recentered IVs, building on Chamberlain (1987, 1992), Newey and McFadden (1994), and Borusyak and Hull (2025). While our baseline analysis shows how powerful recentered IVs can be derived for a given demand model, we also consider non-parametric identification by building on Berry and Haile (2014). Notably, our approach applies even when the same products are sold in all markets and product fixed effects are included, as in Nevo (2001); characteristic-based IVs have no variation in those settings.

We compare this approach to conventional ones in a series of Monte Carlo simulations based on the data-generating process in Gandhi and Houde (2020). When characteristics are exogenous, the power of recentered instruments—whether derived from first-order approximations or exact model-based predictions—is comparable to that of Differentiation IVs and much better than that of BLP instruments. Expectedly, recentered IVs have less power with a lower variance of cost shocks while the power of characteristics-based IVs is lower with less variation in choice sets across markets. A simple model of strategic product entry introduces significant bias in characteristic-based IV estimates, while recentered IV estimates remain accurate.

This paper contributes to two main literatures. First, we contribute to an IO literature studying demand estimation without exogenous characteristics. The potential bias from endogenous characteristics has been noted as far back as Berry et al. (1995). Existing solutions to this concern broadly fall into two categories: some papers put additional structure on the model unobservables (i.e., the unobserved taste shifters) by assuming characteristic endogeneity is captured by controls (e.g., product fixed effects in Nevo (2001)) or imposing a particular statistical process for the unobservables (e.g., Sweeting (2013) and Moon et al. (2018)). Other papers explicitly models characteristic choice or product entry (e.g., Crawford et al. (2019) and Petrin et al. (2022)).[3] In contrast, our approach places no restrictions on how the model unobservables relate to observed characteristics and does not require a model of entry, relying only on cost shock exogeneity. This solution relates to an idea in Ackerberg and Crawford (2009) of searching for "orthogonal instruments" to identify own- and cross-price elasticities while leaving the relationship between characteristics and taste shifters unidentified. We propose a concrete way to achieve this goal, via recentered functions of exogenous cost shocks and endogenous characteristics.[4]

Second, we contribute to a recent econometrics literature on identification and estimation with shift-share IVs and other "formula" instruments combining exogenous shocks with other potentially endogenous data (Borusyak et al., 2022b, 2024; Adão et al. 2019; Borusyak and Hull, 2023, 2025). While this literature studies linear causal or structural models, we focus on nonlinear demand estimation. In this sense our work is also related to Adão et al. (2023) who identify parameters of a quantitative spatial model by the responses of endogenous variables to exogenous shocks; Borusyak et al. (2022a) follow a similar approach with a migration model. Notably, both of these papers work with linear approximations of their models—introducing inaccuracies when shocks are large—while we work directly with nonlinear demand models.[5]

Recent complementary work by Andrews et al. (2025) shows that recentered instruments are more robust to demand model misspecification than characteristic-based IVs, in the sense of recovering more interpretable causal estimands in a non-parametric potential outcomes framework. We demonstrate a different advantage of recentered IVs: that they can be used to relax the assumption of exogenous characteristics in a given demand specification. Moreover, we propose specific constructions of powerful recentered IVs tailored to a class of demand models.

Finally, our analysis relates to empirical studies using weighted sums or other transformations of shocks as instruments to estimate demand models. Recent examples include Costinot et al. (2016), Adão et al. (2017), Couture and Handbury (2020), Fajgelbaum et al. (2020), Adão et al. (2022), Barahona et al. (2023), and Adão et al. (2024); see also Fujiy et al. (2024) who estimate input demand by firms. Typically these instrument constructions arise from intuitive arguments instead of being derived as model-implied responses to the shocks, limiting their power. Moreover, while

[3] See also Fan (2013), who builds BLP-type instruments for a firm's endogenous characteristics from particular characteristics of the firm's competitors: namely, consumer characteristics in markets where competitors operate.

[4] A larger literature improves mixed logit demand estimation in other ways, maintaining the assumption of exogenous characteristics. See, e.g., Berry et al. (1999), Reynaert and Verboven (2014), and Gandhi and Houde (2020) on IV power, Salanie and Wolak (2022) and Lu et al. (2023) on alternative estimation methods, and Wang (2023) on allowing for non-parametric distributions of random coefficients.

[5] Adão et al. (2024) develop a similar approach to specification testing that applies to nonlinear models, but they do not propose an estimation procedure.

the constructions are sometimes simple enough to not need recentering, this consideration is also not typically part of the formal analysis. We show in a general setting how powerful instruments can be constructed by leveraging the structure of the model, and how instrument validity may be made more credible and transparent via explicit recentering.

The rest of this paper is structured as follows. Section 2 builds intuition for our approach in a simple nested logit demand model with randomized cost shocks. Section 3 develops our general approach and discusses asymptotic properties. Section 4 illustrates the approach with simulations. Section 5 concludes. All proofs are collected in the appendix.

2 Motivating Example: Nested Logit with Random Cost Shocks

We start with a simple example that illustrates the main logic of our approach as well as its advantages over conventional methods. Here we keep the presentation informal and intuitive, leaving formal assumptions and results for the more general analysis in Section 3.

2.1 Setting and Conventional Instruments

Consider a set of markets m, each with a set of differentiated products $j \in \mathcal{J}_m$ and an outside good $j = 0$. The products are grouped into "nests" $n(j)$; let $d_{jn} = \mathbf{1}[n(j) = n]$ denote mutually exclusive nest indicators. A researcher observes the nest allocation, along with the price p_{jm} and quantity share s_{jm} of each product. Finally, the researcher observes a set of shocks g_{jm} which increase products' marginal costs (e.g., via input prices) but do not directly affect demand.

The researcher correctly assumes that market shares arise from a nested logit demand model: a mass of consumers i in each market choose a single product or the outside good to maximize their utility $u_{ijm} = \alpha p_{jm} + \xi_{jm} + \varepsilon_{ijm}$, where ξ_{jm} is a common taste shifter and ε_{ijm} is an idiosyncratic taste shock that can be correlated across products in a nest. The outside good has a zero price and taste shifter, such that utility from it is $u_{i0m} = \varepsilon_{i0m}$. Conditional on the prices and taste shifters, the idiosyncratic shocks $(\varepsilon_{ijm})_{j\in\mathcal{J}_m\cup\{0\}}$ are distributed across consumers in such a way that the market shares satisfy:

$$\log\left(s_{jm}/s_{0m}\right) = \alpha p_{jm} + \sigma \log\left(s_{jm}/s_{n(j)m}\right) + \xi_{jm}, \tag{1}$$

where s_{0m} is the market share of the outside good in market m and s_{nm} is the total market share of products in nest n and market m.[6] From this equation the researcher is interested in estimating $\alpha < 0$, which determines the own-price sensitivity of demand, and $\sigma \in [0, 1)$ which captures the extent of within-nest correlation in the taste shocks that governs substitution patterns. These two parameters determine the matrix of cross-price elasticities, which is a key input to a variety of policy counterfactuals (e.g., merger analyses) and consumer welfare calculations.

Estimating α and σ generally requires finding two instruments which are uncorrelated with

[6] Appendix A.1 gives the formula for nested logit market shares and derives this expression from it. The nested CES model, commonly used in international trade and spatial economics, implies similar expressions; see Section 3.5.

the unobserved taste shifters ξ_{jm} but correlated with the two endogenous variables in equation (1): price p_{jm} and the log within-nest market share $\log(s_{jm}/s_{n(j)m})$. Here price endogeneity likely arises because more popular products (with higher ξ_{jt}) are likely of higher quality and therefore more expensive to produce; firms may moreover optimally charge higher markups for them. Endogeneity of $\log(s_{jm}/s_{n(j)m})$ further arises from the market shares' direct dependence on the taste shifters: i.e., more popular products will have larger within-nest market shares.

A natural instrument for the own-price sensitivity parameter α is the excluded cost shock g_{jm}, since higher costs are predicted to at least partially pass through to higher prices.[7] To justify this choice simply, suppose the g_{jm} are drawn in a randomized trial after product entry. Randomization and the natural exclusion restriction that the cost shocks do not directly affect demand ensure that g_{jm} is a valid instrument for equation (1), i.e. that $\operatorname{Cov}[g_{jm}, \xi_{jm}] = 0$.[8]

It is more difficult to find an instrument for the substitution parameter σ. One popular strategy is to construct instruments from the observed characteristics of other products, such as the nest indicators d_{jn}. In particular, in nested logit models, it is common to use the number of products in j's nest, $N_{n(j)m} = \sum_{k \in \mathcal{J}_m} d_{kn(j)}$ (or, equivalently, the number of other products excluding j; e.g., Goldberg and Verboven (2001), Town and Liu (2003), Miller and Weinberg (2017)). This instrument is expected to predict $\log\left(s_{jm}/s_{n(j)m}\right)$ because a product with a larger number of "local" competitors in its nest should have, on average, a smaller within-nest share.

Such characteristic-based instruments have at least two drawbacks. First, their validity hinges on the econometric exogeneity of the characteristics: a strong assumption that can be at odds with natural models of product entry. For example, the nest size instrument will be invalid (with $\operatorname{Cov}\left[N_{n(j)m}, \xi_{jm}\right] > 0$) when firms introduce more products in nests for which consumers have a higher preference in a particular market—a natural tendency of profit-optimizing firms with at least partial information on consumer tastes (see, e.g., Aguiar and Waldfogel (2018, p.506)). Even if the instrument is based on the product entry by firms other than the producer of j, it is not econometrically exogenous. Second, such instruments have no useful variation in some common empirical contexts. Specifically, $N_{n(j)m}$ and similar instruments cannot be used if all products are sold in all markets and product fixed effects are included (as is commonly done since Nevo (2001)). Given these issues, we look for different instruments.

2.2 Proposed Instruments

Consider an IV that measures the relative cost shock of product j vs. the average shock in its nest:

$$z_{jm} = g_{jm} - \frac{1}{N_{n(j)m}} \sum_{k \in \mathcal{J}_m} d_{kn(j)} g_{km}. \tag{2}$$

[7] We attribute an instrument to a particular endogenous variable informally; technically both instruments jointly identify both parameters when they are valid.

[8] Throughout, we call instruments "valid" when they are uncorrelated with the model error (i.e., the unobserved taste shifter). We consider an instrument's relevance, i.e. its correlation with endogenous variables, separately.

We next explain why z_{jm} is both relevant and valid, and how it exemplifies our general approach to constructing instruments as certain combinations of exogenous cost shocks and potentially endogenous characteristics (here, nest indicators) of all the products in the market.

First, this instrument is likely relevant (i.e., correlated with $\log\left(s_{jm}/s_{n(j)m}\right)$): if product j's local competitors in its nest have relatively low cost shocks, its within-nest market share is expected to be lower. This intuitive argument has a formal backing: z_{jm} approximates the model-predicted response of the endogenous variable $\log\left(s_{jm}/s_{n(j)m}\right)$ to the exogenous shocks. Specifically, consider a hypothetical scenario in which all products have equal prices $\check{p}_{km} = \check{\pi}_0$ and unobserved taste shifters $\check{\xi}_{km} = 0$, and thus all products within the nest have the same market shares. To this scenario we introduce an exogenous component of price variation, $\hat{p}_{km} = \check{p}_{km} + \check{\pi} g_{km}$ for some auxiliary constant $\check{\pi} \neq 0$ that aims to capture the pass-through of cost shocks into prices in a simple way. Nested logit demand characterizes the share response to this set of price changes; Appendix A.1 shows that, in a first-order approximation that is precise for small shocks, the resulting change in $\log\left(s_{jm}/s_{n(j)m}\right)$ is equal to z_{jm}, up to a constant scaling factor. While cost shocks is only one of the reasons why $\log\left(s_{jm}/s_{n(j)m}\right)$ deviates from equal shares, z_{jm} captures the component of variation in the within-nest shares due to the cost shocks and is thus likely relevant.

Second, z_{jm} is a valid instrument (i.e., $\mathbb{E}\left[z_{jm}\xi_{jm}\right] = 0$) when the cost shocks are exogenous in the sense we previously assumed to justify their use as an instrument for α. This claim is nontrivial because the formula for z_{jm} incorporates not only the cost shocks but also the nest dummies, which are likely econometrically endogenous. Nevertheless, z_{jm} is constructed in such a way that its validity stems only from the exogeneity of the shocks only. Intuitively, when cost shocks are random, it is also random whether the shock for a particular product is higher or lower than the average in its nest. Formally, z_{jm} is a *recentered* instrument, meaning its expectation over draws of the shocks is zero conditional on the other variation (Borusyak and Hull, 2023): $\mathbb{E}\left[z_{jm} \mid (d_{kn})_{k\in\mathcal{J}_m,n}\right] = \mathbb{E}\left[g_{jm}\right] - \frac{1}{N_{n(j)m}}\sum_{k\in\mathcal{J}_m} d_{kn(j)}\mathbb{E}\left[g_{km}\right] = 0$ when cost shocks are drawn randomly. Thus, z_{jm} is guaranteed to be uncorrelated with any function of the nest allocation which could create endogeneity problems. Other formula instruments can be adjusted via a recentering procedure that removes the component correlated with the characteristics—as we soon illustrate.

To recap, our proposal is to construct instruments as recentered model-based predictions of how relevant endogenous variables respond to the cost shocks. This general approach suggests three ways of improving the simple z_{jm} instrument, in the sense of likely power gains, by forming better predictions of the endogenous variable. First, consider an exact model-based prediction of $\log\left(s_{jm}/s_{n(j)m}\right)$ from the price predictions $\left(\hat{p}_{km}\right)_{k\in\mathcal{J}_m}$ in place of the first-order approximation in z_{jm}. Appendix A.1 shows this prediction can be written:

$$\widehat{\log}\left(s_{jm}/s_{n(j)m}\right) = \frac{\check{\alpha}}{1-\check{\sigma}}\check{\pi} g_{jm} - \log\left(\sum_{k\in\mathcal{J}_m} d_{kn(j)} \exp\left(\frac{\check{\alpha}}{1-\check{\sigma}}\check{\pi} g_{km}\right)\right), \tag{3}$$

where $\check{\alpha}$ and $\check{\sigma}$ are some preliminary estimates of α and σ and $\check{\pi}$, as before, is an estimate of the pass-through of cost shocks into prices.[9] Unlike z_{jm}, exogeneity of the cost shocks does not make this prediction a valid instrument; instead, like the conventional nest size instrument $N_{n(j)m}$, its validity also hinges on the econometric exogeneity of the nest allocation. To see this simply, note that $\widehat{\log}\left(s_{jm}/s_{n(j)m}\right)$ varies over products even in the absence of cost shocks ($g_{km} = \mu_g$ for all k). In fact, this variation is driven exactly by nest size: plugging in $g_{km} = \mu_g$ for all k yields $\widehat{\log}\left(s_{jm}/s_{n(j)m}\right) = -\log N_{n(j)m}$, showing that this prediction suffers from exactly the same endogeneity concerns as the conventional nest size instrument.

Following Borusyak and Hull (2023), we propose obtaining valid instruments by recentering model-based predictions like (3), using knowledge of how the exogenous shocks are drawn. Recentering is straightforward when the shocks are drawn randomly in an experiment: the researcher can re-draw many sets of counterfactual shocks from the experimental protocol, recompute the prediction under each set, average across shock counterfactuals to measure the expected prediction ($\mu_{jm} \equiv \mathbb{E}\left[\widehat{\log}\left(s_{jm}/s_{n(j)m}\right) \mid (d_{kn})_{k\in\mathcal{J}_m,n}\right]$ in the case of (3)), and subtract this expectation from the actual prediction.[10] Like z_{jm}, the recentered prediction

$$z_{jm}^{\text{exact}} = \widehat{\log}\left(s_{jm}/s_{n(j)m}\right) - \mu_{jm} \tag{4}$$

is a valid instrument regardless of any econometric endogeneity of the nest allocation. Below we discuss other ways to recenter predictions in observational data, where the shock data-generating process is unknown. Although recentering removes some variation in the prediction, which was not necessary with z_{jm}, starting from a better prediction still improves the first-stage (Borusyak and Hull, 2025). Note that inaccuracy of the initial $(\check{\alpha}, \check{\sigma})$ estimates in equation (3) is not an issue for the validity of z_{jm}^{exact}, given recentering, though it will likely affect power.

A second type of improvement comes from using additional data to better predict the endogenous variables' responses to exogenous shocks. One particularly useful input is the market shares s_{jm}^{pre} and prices p_{jm}^{pre} for the same products and the same market in an earlier period, before the shocks g_{km} were drawn. To the extent that prices and taste shifters are serially correlated, this yields better predictions of prices and market shares in the period of interest.[11] The instrument that results from incorporating this information is also very intuitive: Appendix A.1 shows that the first-order approximation of the model-predicted response of $\log\left(s_{jm}/s_{n(j)m}\right)$ to the exogenous shocks around the pre-period shares (rather than equal shares) yields

$$z_{jm}^{\text{weighted}} = g_{jm} - \frac{\sum_{k\in\mathcal{J}_m} d_{kn(j)} s_{km}^{\text{pre}} g_{km}}{\sum_{k\in\mathcal{J}_m} d_{kn(j)} s_{km}^{\text{pre}}}. \tag{5}$$

Like z_{jm}, this prediction is mean-zero over draws of random cost shocks, $\mathbb{E}\left[z_{jm}^{\text{weighted}} \mid (d_{kn})_{k\in\mathcal{J}_m,n}\right] =$

[9] Preliminary estimates $\check{\alpha}, \check{\sigma}$ can be obtained using simpler instruments, such as (2), in a first step, while $\check{\pi}$ can be obtained from a regression of prices and the own-product shocks. In Section 3 we show how a continuously updating estimator can bypass the need for $\check{\alpha}, \check{\sigma}$.

[10] We note that μ_{jm} is related but not equal to the value of z_{jm} with no shocks, $-\log N_{n(j)m}$. The difference arises because z_{jm} is a nonlinear function of the shocks, and thus taking the expectation of z_{jm} across the shock distribution is not the same as plugging in expectation of the shocks.

[11] A second complementary use of such data is that equation (1) can be estimated in time-differences, yielding more precise estimates from the same instruments when ξ_{jm} is serially correlated. We discuss this approach in Section 3.4.

0, making it a recentered instrument without further adjustment.[12] The appendix further shows that the two improvements can be combined: a researcher can incorporate the lagged share information to improve the recentered exact prediction (4), too. In either case, the recentered instrument will again be valid just by virtue of the exogenous shocks—even though it now draws power from variation in lagged market shares (as well as the nest allocation) and lagged market shares by themselves need not be exogenous.[13]

Finally, consider an instrument which uses a more realistic prediction of how prices respond to the full set of exogenous shocks. A researcher might, for example, specify an auxiliary pricing model which captures not only the pass-through of product j's cost shock to its own price but also how p_{jm} responds to competitor cost shocks g_{km} for $k \neq j$ (depending, for instance, on whether j and k are offered by the same firm). Substituting this model's price predictions into any of the above instrument constructions, in place of the simple $\hat{p}_{jm} = \breve{\pi} g_{jm}$ prediction, yields an instrument which is likely more powerful when such cost shock spillovers are important and can be estimated. The better price prediction can also be recentered and used in place of g_{jm} as an instrument for identifying α. Note that as with the initial $(\breve{\alpha}, \breve{\sigma})$ estimates in z_{jm}^{exact}, the validity of these instruments does not hinge on the accuracy of the pricing model.

3 General Approach

We now consider a broader class of demand models and formalize our general approach. Section 3.1 develops the baseline mixed logit model and the introduces key shock exogeneity assumption. Section 3.2 defines recentered IVs and establishes identification with them. Section 3.3 develops our proposal for constructing powerful recentered IVs from the structure of the model, while Section 3.4 establishes consistency and asymptotic inference. Section 3.5 discusses several extensions to the baseline model.

3.1 Setting

We consider a class of random utility models—canonical mixed logit demand—with market-level data as in Berry et al. (1995); see Berry and Haile (2021) and Gandhi and Nevo (2021) for more recent treatments. A researcher observes a set of markets m (which might correspond to regions, periods, or both) with differentiated products $j \in \mathcal{J}_m$, prices p_{jm}, and quantity shares s_{jm}.[14] Each product also has a vector of observed characteristics $x_{jm} \in \mathbb{R}^L$ (which includes an intercept), as

[12] z_{jm} and z_{jm}^{weighted} are examples of shift-share instruments, which average the exogenous shocks with a set of weights capturing differential shock exposure (Borusyak et al., 2022b). As discussed below, such instruments are often easier to recenter or require no recentering at all because they are linear in the shocks.

[13] We note that, unlike the nest size IV, z_{jm} and z_{jm}^{weighted} have cross-market variation even if all products are sold in all markets—so long as cost shocks vary across markets. Moreover, z_{jm}^{weighted} can vary across markets even if cost shocks do not, if the pre-period shares vary due to any unobserved cost or taste differences.

[14] Here we do not restrict whether the data consist of many markets or just a single one, whether the markets are randomly sampled, or whether the number of products per market is large. We return to these issues in Section 3.4.

well as an unobserved scalar taste shifter ξ_{jm}. All variables are normalized such that the outside good in each market, $j = 0$, has $p_{0m} = \xi_{0m} = 0$ and $x_{0m} = 0$.

Consumers i choose among all products and the outside good by maximizing their utility:

$$\max_{j \in \mathcal{J}_m \cup \{0\}} \delta_{jm} + \eta_{0i} p_{jm} + \sum_{\ell=1}^{L_1} \eta_{i\ell} x_{jm\ell} + \varepsilon_{ijm}. \tag{6}$$

Here product j's mean utility δ_{jm} is determined by its price, characteristics, and the taste shifter:

$$\delta_{jm} = \alpha p_{jm} + \beta' x_{jm} + \xi_{jm} \tag{7}$$

with $\alpha < 0$ and $\beta \in \mathbb{R}^L$. A subvector of characteristics $x_{jm}^{(1)} = (x_{jm1}, \ldots, x_{jmL_1})$, as well as (potentially) price, also enter utility with mean-zero "random coefficients" $\eta_i = (\eta_{i\ell})_{\ell=0}^{L_1}$ that capture heterogeneous consumer preferences. This η_i is *iid* across consumers and follows distribution $\mathcal{P}(\cdot;\sigma)$ that is known up to a vector of "nonlinear" parameters σ (typically Gaussian with independent components and standard deviations $\sigma_0, \ldots, \sigma_{L_1}$). Finally, ε_{ijm} is an extreme-value shock, *iid* across consumers and products including the outside good. Integrating out these shocks implies market shares satisfy

$$s_{jm} = \mathcal{S}_j(\boldsymbol{\delta}_m; \sigma, \boldsymbol{x}_m^{(1)}, \boldsymbol{p}_m) \equiv \int \frac{\exp\left(\delta_{jm} + \eta_{0i} p_{jm} + \sum_{\ell=1}^{L_1} \eta_{i\ell} x_{jm\ell}\right)}{1 + \sum_{k \in \mathcal{J}_m} \exp\left(\delta_{km} + \eta_{0i} p_{km} + \sum_{\ell=1}^{L_1} \eta_{i\ell} x_{km\ell}\right)} d\mathcal{P}(\eta_i; \sigma), \tag{8}$$

where bold symbols denote a collection of variables for all products in the market: $\boldsymbol{v}_m = (v_{jm})_{j \in \mathcal{J}_m}$ for any variable v_{jm}. The distribution $\mathcal{P}$ determines the patterns of product substitutability: for instance, the pure multinomial logit model corresponds to no variation in random coefficients (typically captured by $\sigma = 0$). The nested logit model considered in Section 2 is also a special case, with nest dummies as characteristics and with a particular choice of $\mathcal{P}(\cdot;\sigma)$ (McFadden, 1978). Berry (1994) and Berry et al. (1995) famously show that the share function $\mathcal{S}_j(\boldsymbol{\delta}_m; \sigma, \boldsymbol{x}_m^{(1)}, \boldsymbol{p}_m)$ is invertible, such that mean utilities can be derived from σ and observed data:

$$\delta_{jm} = \mathcal{D}_j\left(\boldsymbol{s}_m; \sigma, \boldsymbol{x}_m^{(1)}, \boldsymbol{p}_m\right), \tag{9}$$

for functions $\mathcal{D}_j(\cdot)$ that generally do not have a closed-form but can be computed numerically.

We focus on estimating parameters $\theta = (\alpha, \sigma')'$, which are central to a number of important policy counterfactuals that do not involve changing product characteristics. Indeed, the model implies that own- and cross-price elasticities can be characterized in terms of θ and observed data:

$$\frac{ds_{jm}}{dp_{km}} = \int (\alpha + \eta_{0i}) s_{ji} \left(\mathbf{1}[j = k] - s_{ki}\right) d\mathcal{P}(\eta_i; \sigma) \tag{10}$$

where

$$s_{ji} = \frac{\exp\left(\mathcal{D}_j\left(\boldsymbol{s}_m; \sigma, \boldsymbol{x}_m^{(1)}, \boldsymbol{p}_m\right) + \eta_{0i} p_{jm} + \sum_{\ell=1}^{L_1} \eta_{i\ell} x_{jm\ell}\right)}{1 + \sum_{k \in \mathcal{J}_m} \exp\left(\mathcal{D}_k\left(\boldsymbol{s}_m; \sigma, \boldsymbol{x}_m^{(1)}, \boldsymbol{p}_m\right) + \eta_{0i} p_{km} + \sum_{\ell=1}^{L_1} \eta_{i\ell} x_{km\ell}\right)}.$$

In contrast, a consistent estimate of the causal effect of characteristics on mean utility, β, is not needed when analyzing a merger (Ackerberg and Crawford (2009)). Similarly, the welfare gains from a new product can be computed without β.[15]

To identify θ we leverage share inversion, which implies a structural equation that is additive in the unobserved taste shifter:

$$\mathcal{D}_j\left(\boldsymbol{s}_m;\sigma,\boldsymbol{x}_m^{(1)},\boldsymbol{p}_m\right)=\alpha p_{jm}+\beta' x_{jm}+\xi_{jm}.$$

This representation permits estimation of θ from moment conditions of the form

$$\mathbb{E}\left[Z_{jm}\cdot\left(\mathcal{D}_j\left(\boldsymbol{s}_m;\sigma,\boldsymbol{x}_m^{(1)},\boldsymbol{p}_m\right)-\alpha p_{jm}-\beta' x_{jm}\right)\right]=0 \tag{11}$$

for some vector of instruments Z_{jm} that are uncorrelated with ξ_{jm}. If the instruments are also uncorrelated with x_{jm} (as ours will be), equation (11) will hold for any value of β.

To build instruments, we assume the researcher observes a set of supply-side shocks g_{jm} that vary by product and market. Shocks to input prices is a natural source of supply-side shocks commonly used in the IO literature (e.g., Villas-Boas (2007), Backus et al. (2021), Barahona et al. (2023)); other studies have used exchange rate shocks (e.g., Berry et al. (1999); Goldberg and Verboven (2001)), shocks from weather (as a productivity shifter; e.g., Nakamura and Zerom (2010)), product-specific subsidies (e.g., Li, 2016), taxes (e.g., Dearing, 2022), and markup shocks due to mergers (e.g., Miller and Weinberg, 2017).[16] We generically call the g_{jm} "cost shocks" and assume that they are exogenous in the following sense:

$$\mathbb{E}\left[\xi_{jm}\mid \boldsymbol{g}_m,\boldsymbol{x}_m\right]=\mathbb{E}\left[\xi_{jm}\mid \boldsymbol{x}_m\right]\qquad \forall m, j\in\mathcal{J}_m. \tag{12}$$

That is, we assume the taste shifter of each product j is mean-independent of cost shocks (both for j and its competitors), conditionally on the observed characteristics of all products.

To understand the economic content of the assumption, note that the unobserved taste shifter captures both objective characteristics of the product chosen by firms and the subjective preferences of consumers. Thus, several conditions have to be met for equation (12) to hold. First, cost shocks should not affect product entry decisions or firm choices of unobserved characteristics. A sufficient condition is that cost shocks are realized after those decisions have been made (but before prices are set—a condition required for relevance of cost shocks as instruments). Second, cost shocks should not influence consumer preferences. Such influence could be possible when cost shocks influence firms' advertising decisions or stem from the prices of inputs that affect consumer earnings.

In addition to these exclusion restrictions, cost shocks may not be *correlated with* variables that affect product entry or consumer preferences. This can be viewed as an independence condition, in the sense of Imbens and Angrist (1994), that is automatically satisfied in randomized experiments. Examples from the literature show how it can also hold in observational data. Several IO papers use

[15] While such gains require a prediction of the new product's mean utility, naturally based on its characteristics, this constitutes a prediction problem and not a causal problem which would require the structural parameter β. We are not aware of previous work making this point.

[16] We do not require the shocks to be independent across products and markets. For instance, all cars produced in the same country are assigned the same exchange rate shock, and input price shocks affect all products using this input, to different extents. For now we assume that the researcher can assign each product to the corresponding country of production or shares of different inputs but we relax this assumption in Section 3.5.

the exchange rate in the country of production as a cost shock for automobiles (e.g., Berry et al., 1999; Goldberg and Verboven, 2001; Grieco et al., 2024) and other industries (e.g., Nakamura and Zerom, 2010). While these studies use the *level* of the exchange rate, their *changes* over time may be particularly attractive as the g_{jm} because exchange rates are known to roughly follow a random walk (e.g., Kilian and Taylor (2003)). The same argument applies to many commodity price changes, for goods using these commodities as inputs (e.g., coffee in Nakamura and Zerom (2010)). For inputs traded on futures markets, Ackerberg and Crawford (2009) point out that the difference between the realized input price at the time when downstream firms set prices and the price of a futures contract at the earlier moment when product entry has been decided is guaranteed to be unrelated to the characteristics.

While the assumption in (12) is restrictive, it is important to highlight what it does *not* entail: it allows unobserved taste shifters ξ_{jm} to be arbitrarily correlated with the observed characteristics of both product j and its competitors.[17] This is in contrast to the prevalent approach in the literature (e.g., Berry et al., 1995, 1999; Gandhi and Houde, 2020) which imposes a stronger assumption:

$$\mathbb{E}\left[\xi_{jm} \mid \boldsymbol{x}_m, \boldsymbol{g}_m\right] = 0, \tag{13}$$

equivalent to imposing $\mathbb{E}\left[\xi_{jm} \mid \boldsymbol{x}_m\right] = 0$ in addition to equation (12). Under this stronger condition, instruments constructed as functions of own and competing products' characteristics are valid. This includes BLP instruments, computed as sums or averages of competitor characteristics, efficient instruments from Berry et al. (1999), as well as the differentiation IVs of Gandhi and Houde (2020) which capture the average distance between x_{jm} and characteristics of competitors or the number of products in the market with characteristics sufficiently close to x_{jm}.

As recognized as far back as Berry et al. (1995), however, the econometric exogeneity of observed characteristics is an unappealing restriction on product entry. It is often arbitrary which objective characteristics are observed or unobserved by the econometrician, such that there is no reason why the two groups should be uncorrelated with each other.[18] Moreover, in natural models of product entry, ξ_{jm} can be related to characteristics of competing products. This can happen, for instance, when all firms observe some information about the cost or demand conditions common to the market when making product entry decisions. Similar to the discussion in Section 2, in a market where consumers like small and fuel-efficient cars (i.e. their ξ_{jm} is predicted to be higher), we expect all firms to pivot towards these characteristics, whether in ways observed or unobserved to the econometrician. In this situation, BLP and differentiation IVs need not be valid.

In the rest of our analysis we maintain a modified version of (12) that allows shock exogeneity to be conditional on some other observed data $\boldsymbol{q}_m$ (as well as product characteristics):

Assumption 1 (Exogenous cost shocks)**.** $\mathbb{E}\left[\xi_{jm} \mid \boldsymbol{g}_m, \boldsymbol{x}_m, \boldsymbol{q}_m\right] = \mathbb{E}\left[\xi_{jm} \mid \boldsymbol{x}_m, \boldsymbol{q}_m\right], \forall m, j \in \mathcal{J}_m$.

[17]An additional feature of (12) is that it allows the cost shocks to be mutually correlated. Mutual correlations of taste shifters are also allowed, as when unobserved market-level demand and cost conditions affect the choice of unobserved quality of all firms in the market.

[18]Note, however, that it is generally necessary for characteristics with random coefficients to be observed. An interesting exception is provided by Adão et al. (2017) who show identification in a model with a random coefficient on unobserved mean utility; our approach applies to that model as well.

Both condition (12) and Assumption 1 hold when the shocks are unconditionally as-if randomly assigned; the modified assumption will then be helpful to construct more powerful instruments that use information in $\boldsymbol{q}_m$ (e.g., lagged market shares and prices as in Section 2). In other settings, conditioning on potential confounders in $\boldsymbol{q}_m$ may help make shock exogeneity more plausible, such as when the g_{jm} are systematically correlated with some observed q_{jm}.

3.2 Recentered Instruments

We say Z_{jm} is a vector of *formula* instruments when it can be written as $Z_{jm} = f_{jm}(\boldsymbol{g}_m, \boldsymbol{x}_m, \boldsymbol{q}_m)$ for some non-stochastic and known vector-valued functions $(f_{jm})_{m,j\in\mathcal{J}_m}$. We further say that Z_{jm} consists of *recentered* formula instruments (or just recentered IVs) when the formulas satisfy:

$$\mathbb{E}\left[f_{jm}(\boldsymbol{g}_m, \boldsymbol{x}_m, \boldsymbol{q}_m) \mid \boldsymbol{x}_m, \boldsymbol{q}_m\right] = 0, \quad \forall m, j \in \mathcal{J}_m. \tag{14}$$

Our first result shows this property characterizes the complete set of valid instruments in our setup:

Lemma 1. *Assumption 1 implies* $\mathbb{E}\left[Z_{jm}\xi_{jm}\right] = 0$ *if and only if* Z_{jm} *consists of recentered IVs.*

This result follows immediately from Proposition 1 in Borusyak and Hull (2025).[19] It shows that when researchers are only willing to assume that cost shocks are exogenous, in the sense of Assumption 1, the only justified instruments are recentered IVs. More positively, it shows that any candidate formula instrument $h_{jm}(\boldsymbol{g}_m, \boldsymbol{x}_m, \boldsymbol{q}_m)$ for some functions $(h_{jm})_{m,j\in\mathcal{J}_m}$ can be made a valid IV by recentering: i.e., by subtracting off its conditional mean $\mathbb{E}\left[h_{jm}(\boldsymbol{g}_m, \boldsymbol{x}_m, \boldsymbol{q}_m) \mid \boldsymbol{x}_m, \boldsymbol{q}_m\right]$.

Borusyak and Hull (2023) and Borusyak et al. (2024) discuss several strategies for recentering formula instruments. One general approach follows when it is possible to generate a set of counterfactual shock vectors $\boldsymbol{g}_m^{(c)}$, for $c = 1, \ldots, C$, which are either drawn from the same distribution as $\boldsymbol{g}_m$ (conditional on $\boldsymbol{x}_m$ and $\boldsymbol{q}_m$) or otherwise as likely to have been realized. For example, the $\boldsymbol{g}_m^{(c)}$ could be generated by redrawing from the same randomization protocol that generated $\boldsymbol{g}_m$ in a randomized trial, as in the Section 2 motivating example. In observational data, the counterfactual shocks can instead be generated by reshuffling the set of observed g_{jm} across comparable products, markets, or both. Given such $\boldsymbol{g}_m^{(c)}$, any formula instrument $h_{jm}(\boldsymbol{g}_m, \boldsymbol{x}_m, \boldsymbol{q}_m)$ can be recentered by recomputing instrument values under each counterfactual (holding fixed $\boldsymbol{x}_m$ and $\boldsymbol{q}_m$) and subtracting the average across these values, $\frac{1}{C}\sum_c h_{jm}(\boldsymbol{g}_m^{(c)}, \boldsymbol{x}_m, \boldsymbol{q}_m)$, for each j and m.[20] A second general approach follows when the formula instrument is linear in the shocks: i.e., when $h_{jm}(\boldsymbol{g}_m, \boldsymbol{x}_m, \boldsymbol{q}_m) = \sum_{k\in\mathcal{J}_m} w_{jkm} g_{km}$ for some exposure weights (or "shares") w_{jkm} that are functions of $(\boldsymbol{x}_m, \boldsymbol{q}_m)$. Recentering such shift-share IVs only requires de-meaning the cost shocks (or "shifts") by their conditional expectations, $\mathbb{E}\left[g_{km} \mid \boldsymbol{x}_m, \boldsymbol{q}_m\right]$.

Lemma 1 implies that a recentered IV vector Z_{jm} locally identifies model parameters θ under

[19] Like there, the "only if" part of Lemma 1 should be understood as follows: unless Z_{jm} consists of recentered instruments, it is possible to find a conditional distribution of ξ_{jm} such that Assumption 1 holds but $\mathbb{E}\left[Z_{jm}\xi_{jm}\right] \neq 0$.

[20] The number of counterfactuals C does not matter for recentered IV consistency, though it generally affects the asymptotic variance. See footnotes 19 and 21 in Borusyak and Hull (2023).

a rank condition (Rothenberg, 1971): that the matrix $\mathbb{E}\left[Z_{jm}\nabla'_{jm}\right]$ is full column rank, where

$$\nabla_{jm} = \frac{\partial}{\partial\theta}\left(\mathcal{D}_j\left(\boldsymbol{s}_m;\sigma,\boldsymbol{x}_m^{(1)},\boldsymbol{p}_m\right) - \alpha p_{jm} - \beta' x_{jm}\right) = \begin{pmatrix} -p_{jm} \\ \nabla_{jm}^{\sigma} \end{pmatrix}, \tag{15}$$

for $\nabla_{jm}^{\sigma} = \partial\mathcal{D}_j\left(\boldsymbol{s}_m;\sigma,\boldsymbol{x}_m^{(1)},\boldsymbol{p}_m\right)/\partial\sigma$ with the derivative evaluated at true parameter values.[21] Note again that θ is identified while β is not, since recentering makes Z_{jm} uncorrelated with x_{jm}. But this is unimportant since β does not directly enter price elasticities or important policy counterfactuals. We next discuss how likely powerful IVs can be constructed.

3.3 Constructing Model-Based IVs

We propose constructing recentered IVs that predict how the vector of model's residual derivatives ∇_{jm} responds to the exogenous cost shocks. Specifically, we consider an instrument vector of length $\dim(\theta)$ approximating:

$$\tilde{Z}_{jm} = \mathbb{E}\left[\nabla_{jm} \mid \boldsymbol{g}_m, \boldsymbol{x}_m, \boldsymbol{q}_m\right] - \mathbb{E}\left[\nabla_{jm} \mid \boldsymbol{x}_m, \boldsymbol{q}_m\right]. \tag{16}$$

The first term of $\tilde{Z}_{jm}$ is the best predictor of the residual derivatives given the cost shocks, product characteristics, and other data in $\boldsymbol{q}_m$, where the expectation is taken over the conditional distribution of $(\boldsymbol{p}_m, \boldsymbol{s}_m)$ that corresponds to different realizations of unobserved demand and cost shocks. The second term recenters this best predictor by its expectation over the shocks, $\mathbb{E}\left[\mathbb{E}\left[\nabla_{jm} \mid \boldsymbol{g}_m, \boldsymbol{x}_m, \boldsymbol{q}_m\right] \mid \boldsymbol{x}_m, \boldsymbol{q}_m\right] = \mathbb{E}\left[\nabla_{jm} \mid \boldsymbol{x}_m, \boldsymbol{q}_m\right]$. Hence $\tilde{Z}_{jm}$ captures how the model residual's derivative is affected by the specific draw of shocks. Note that this $\tilde{Z}_{jm}$ is guaranteed to satisfy the rank condition when cost shocks are relevant (i.e. when $\tilde{Z}_{jm} \neq 0$) since then $\mathbb{E}\left[\tilde{Z}_{jm}\nabla'_{jm}\right] = \mathbb{E}\left[\tilde{Z}_{jm}\tilde{Z}'_{jm}\right]$ which is generally fully rank.[22] Below we show $\tilde{Z}_{jm}$ is closely related to the asymptotically efficient recentered IV vector.

The overall logic of our approximations to $\tilde{Z}_{jm}$ is as follows. Instead of integrating over the unobserved shocks, we predict ∇_{jm} in a single "no-shock" scenario that would prevail in the absence of unexpected $\boldsymbol{g}_m$ shocks, i.e. when $\boldsymbol{g}_m = \mathbb{E}\left[\boldsymbol{g}_m \mid \boldsymbol{x}_m, \boldsymbol{q}_m\right]$. This scenario is constructed using the data in $(\boldsymbol{x}_m, \boldsymbol{q}_m)$ only. We then predict how prices and shares would deviate because of the realized shocks. Specifically: cost shocks affect prices and the researcher can construct an unexpected component of prices using an auxiliary model of cost shock pass-through, as in Section 2. In turn, prices affect market shares; using preliminary values of demand parameters, the researcher can then measure the impact of unexpected price changes on shares. Finally, $\partial\mathcal{D}_j/\partial\sigma$ is a function of the shares so it can be predicted for the scenario with shocks, as a function of $(\boldsymbol{g}_m, \boldsymbol{x}_m, \boldsymbol{q}_m)$. To approximate $\tilde{Z}_{jm}$, it then remains to recenter this prediction as in Section 3.2.

[21] As Newey and McFadden (1994) note, establishing global identification with nonlinear moment conditions is generally challenging; the objective function for mixed logit estimation is known to not be globally convex (Conlon and Gortmaker, 2020).

[22] Formally, $\mathbb{E}\left[\tilde{Z}_{jm}\nabla'_{jm}\right] = \mathbb{E}\left[\tilde{Z}_{jm}\mathbb{E}\left[\nabla_{jm} \mid \boldsymbol{g}_m, \boldsymbol{x}_m, \boldsymbol{q}_m\right]'\right] = \mathbb{E}\left[\tilde{Z}_{jm}\tilde{Z}'_{jm}\right]$ by the law of iterated expectations and the fact that $\mathbb{E}\left[\tilde{Z}_{jm}\mathbb{E}\left[\nabla_{jm} \mid \boldsymbol{x}_m, \boldsymbol{q}_m\right]'\right] = 0$ by virtue of recentering.

The IV construction proceeds in four steps. First, the researcher picks some preliminary values of the parameters $\breve{\alpha} < 0$ and $\breve{\sigma}$. For now, we view these as non-stochastic though it is without loss to allow them to be functions of $(\boldsymbol{x}_m, \boldsymbol{q}_m)$. We discuss in-sample estimation in Section 3.4.

Second, the researcher constructs the no-shock scenario which comprises of a prediction of prices and shares $(\breve{\boldsymbol{p}}_m, \breve{\boldsymbol{s}}_m)$ based on the information in $(\boldsymbol{x}_m, \boldsymbol{q}_m)$ only. With panel data and persistent cost and demand shocks, a natural choice is the prices and shares in a period prior to the realization of the $\boldsymbol{g}_m$ shocks (collected in $\boldsymbol{q}_m$). In a single cross-section, predicted prices and mean utilities may be the fitted values from regressing prices and $\mathcal{D}_j\left(\boldsymbol{s}_m; \breve{\sigma}, \boldsymbol{x}_m^{(1)}, \boldsymbol{p}_m\right)$ on characteristics, respectively, while predicted shares may follow from the model (i.e., equation (8)) at the parameters $\breve{\sigma}$ and implied mean utilities.[23]

Third, the researcher forms price predictions as deviations from $\breve{\boldsymbol{p}}_m$ due to the exogenous cost-shocks. For clarity here we work with the simplest predictions:

$$\hat{p}_{jm} = \breve{p}_{jm} + \breve{\pi}\tilde{g}_{jm}, \tag{17}$$

where $\tilde{g}_{jm} = g_{jm} - \mathbb{E}\left[g_{jm} \mid \boldsymbol{x}_m, \boldsymbol{q}_m\right]$ is product j's recentered cost shock and $\breve{\pi} \neq 0$ is a pass-through coefficient that is again assumed non-stochastic or measurable with respect to $(\boldsymbol{x}_m, \boldsymbol{q}_m)$, for now. These predictions need not be correct for the resulting instruments to be valid, as in Section 2. We consider extensions with more elaborate shock pass-through models below.

The price prediction immediately suggests the first instrument (recall equation (15)): $-\breve{\pi}\tilde{g}_{jm}$, or equivalently the recentered shock $\tilde{g}_{jm}$. The price prediction also implies a prediction for mean utilities that will shortly prove helpful:

$$\hat{\delta}_{jm} = \breve{\delta}_{jm} + \breve{\alpha}\breve{\pi}\tilde{g}_{jm} \quad \text{for } \breve{\delta}_{jm} = \mathcal{D}_j\left(\breve{\boldsymbol{s}}_m; \breve{\sigma}, \boldsymbol{x}_m^{(1)}, \breve{\boldsymbol{p}}_m\right).$$

Finally, the researcher generates predictions for the market shares and ultimately ∇_{jm}^{σ}. We propose two versions: a first-order approximation which yields a recentered shift-share IV, and an exact prediction that may yield a more powerful instrument but generally requires further recentering. Using the first-order approximation, one predicts market shares as

$$\begin{aligned}
\hat{s}_{jm} &= \breve{s}_{jm} + \sum_{k \in \mathcal{J}_m} \frac{\partial}{\partial \delta_{km}} \mathcal{S}_j(\breve{\boldsymbol{\delta}}_m; \breve{\sigma}, \boldsymbol{x}_m^{(1)}, \breve{\boldsymbol{p}}_m)\left(\hat{\delta}_{km} - \breve{\delta}_{km}\right) \\
&= \breve{s}_{jm} + \sum_{k \in \mathcal{J}_m} \frac{\partial}{\partial \delta_{km}} \mathcal{S}_j(\breve{\boldsymbol{\delta}}_m; \breve{\sigma}, \boldsymbol{x}_m^{(1)}, \breve{\boldsymbol{p}}_m)\breve{\alpha}\breve{\pi}\tilde{g}_{km}
\end{aligned}$$

[23] One could also include recentered shocks in these regressions to improve predictive power but take fitted values corresponding to the characteristics only. Nonlinear predictions, such as with machine learning algorithms, may improve precision, too.

and the model residual's derivative with respect to the vector σ as

$$\begin{aligned}\hat{\nabla}^{\sigma}_{jm} &= \check{\nabla}^{\sigma}_{jm} + \sum_{k\in\mathcal{J}_m} \frac{\partial^2}{\partial p_{km}\partial\sigma}\mathcal{D}_j\left(\check{\boldsymbol{s}}_m;\check{\sigma},\boldsymbol{x}^{(1)}_m,\check{\boldsymbol{p}}_m\right)(\hat{p}_{km}-\check{p}_{km}) \\ &\quad + \sum_{k\in\mathcal{J}_m} \frac{\partial^2}{\partial s_{km}\partial\sigma}\mathcal{D}_j\left(\check{\boldsymbol{s}}_m;\check{\sigma},\boldsymbol{x}^{(1)}_m,\check{\boldsymbol{p}}_m\right)(\hat{s}_{km}-\check{s}_{km}) \\ &= \check{\nabla}^{\sigma}_{jm} + \sum_{k\in\mathcal{J}_m} w_{jkm}\tilde{g}_{km}. \end{aligned} \tag{18}$$

Here $\check{\nabla}^{\sigma}_{jm} = \frac{\partial}{\partial\sigma}\mathcal{D}_j\left(\check{\boldsymbol{s}}_m;\check{\sigma},\boldsymbol{x}^{(1)}_m,\check{\boldsymbol{p}}_m\right)$ predicts the residual's derivative in the absence of shocks and

$$\begin{aligned} w_{jkm} &= \check{\pi}\frac{\partial^2}{\partial p_{km}\partial\sigma}\mathcal{D}_j\left(\check{\boldsymbol{s}}_m;\check{\sigma},\boldsymbol{x}^{(1)}_m,\check{\boldsymbol{p}}_m\right) \\ &\quad + \check{\alpha}\check{\pi}\sum_{k'\in\mathcal{J}_m}\frac{\partial^2}{\partial s_{k'm}\partial\sigma}\mathcal{D}_j\left(\check{\boldsymbol{s}}_m;\check{\sigma},\boldsymbol{x}^{(1)}_m,\check{\boldsymbol{p}}_m\right)\cdot\frac{\partial}{\partial\delta_{km}}\mathcal{S}_{k'}(\check{\boldsymbol{\delta}}_m;\check{\sigma},\boldsymbol{x}^{(1)}_m,\check{\boldsymbol{p}}_m) \end{aligned} \tag{19}$$

is the predicted first-order effect of $\tilde{g}_{km}$ on that derivative. Recentering the prediction in (18) eliminates the first term, resulting in the $\dim(\theta)\times 1$ vector of shift-share instruments

$$Z^{SSIV}_{jm} = \begin{pmatrix} -\check{\pi}\tilde{g}_{jm} \\ \sum_{k\in\mathcal{J}_m} w_{jkm}\tilde{g}_{km} \end{pmatrix}.$$

Again, an advantage of Z^{SSIV}_{jm} is that it only requires specification of the conditional shock means $\mathbb{E}\left[\boldsymbol{g}_m \mid \boldsymbol{x}_m, \boldsymbol{q}_m\right]$ for recentering. Also conveniently, the choice of $\check{\pi}$ is immaterial with this approach, as it only rescales the instruments. Without a random coefficient in price the first term in equation (19) drops out, making $\check{\alpha}$ immaterial too.

Alternatively, the researcher can obtain the exact prediction of how changes in mean utilities due to the cost shocks affect shares:

$$\hat{s}_{jm} = \mathcal{S}_j(\hat{\boldsymbol{\delta}}_m;\check{\sigma},\boldsymbol{x}^{(1)}_m,\hat{\boldsymbol{p}}_m),$$

and how those changes affect ∇^{σ}_{jm}:

$$\hat{\nabla}^{\sigma}_{jm} = \frac{\partial}{\partial\sigma}\mathcal{D}_j\left(\hat{\boldsymbol{s}}_m;\check{\sigma},\boldsymbol{x}^{(1)}_m,\hat{\boldsymbol{p}}_m\right).$$

This $\hat{\nabla}^{\sigma}_{jm}$ is a nonlinear function of the shocks which needs to be recentered, generally by specifying shock counterfactuals as described in Section 3.2.[24] The recentered formula IV vector is then:

$$Z^{FIV}_{jm} = \begin{pmatrix} -\check{\pi}\tilde{g}_{jm} \\ \hat{\nabla}^{\sigma}_{jm} - \mathbb{E}\left[\hat{\nabla}^{\sigma}_{jm}\mid \boldsymbol{x}_m,\boldsymbol{q}_m\right] \end{pmatrix}.$$

[24]This $\hat{\nabla}^{\sigma}_{jm}$ relates to the efficient IV construction of Berry et al. (1999): the two would be equivalent if the price predictions $\hat{p}_{jm}$ were taken from an equilibrium pricing model and mean utilities were set to $\hat{\delta}_{jm} = \check{\alpha}\hat{p}_{jm} + \check{\beta}'x_{jm}$ for an initial value $\check{\beta}$ of β. Our instrument differs in three ways: it is based on a shock pass-through model that need not be correct, it uses additional information (in particular, when lagged prices and shares are available), and it is recentered to avoid bias from endogenous characteristics. Conlon and Gortmaker (2020) propose an improvement on the Berry et al. (1999) instrument that integrates over an empirical distribution of ξ_{jm} rather than setting unobserved demand shocks to zero. Our approach could be similarly extended.

Appendix Proposition A1 builds intuition for these instruments by considering the case where the nonlinear parameters σ are the standard deviations of random coefficients and $\check{\sigma}$ is close to zero, which would correspond to a pure multinomial logit model.[25] In this "local-to-logit" approximation, the shift-share IV corresponding to the standard deviation σ_ℓ of a non-price characteristic $x_{jm\ell}$ can be written, up to a scaling factor, as:

$$z_{jm\ell}^{SSIV} \approx x_{jm\ell} \cdot \sum_{k \in \mathcal{J}_m} \check{s}_{km} \left(x_{km\ell} - \bar{x}_{m\ell} \right) \tilde{g}_{km} \quad \text{for } \bar{x}_{m\ell} = \sum_{k \in \mathcal{J}_m} \check{s}_{km} x_{km\ell}.$$

This IV an interaction between product j's own characteristic $x_{jm\ell}$ and a market-specific aggregate of the shocks: the share-weighted covariance across the products in the market between $x_{km\ell}$ and the recentered cost shock (including the outside good with a shock set to zero). The covariance is positive when the cost shocks unexpectedly make products with high $x_{km\ell}$ more expensive relative to other products in the market. Thus, $z_{jm\ell}^{SSIV}$ is performing an analysis similar to difference-in-differences: it compares changes in market shares for high-$x_{jm\ell}$ vs. low-$x_{jm\ell}$ products in markets where high-$x_{km\ell}$ vs. low-$x_{km\ell}$ products became less competitive because of the exogenous cost shocks. Identification with this instrument is therefore based on the core property of mixed logit models: that, after a cost shock, market shares are reallocated towards products with similar characteristics when σ_ℓ is large, but to all products evenly (in proportion of their market shares) when $\sigma = 0$. The instrument for the random coefficient in price has additional terms related to how cost shocks affect prices directly; see Appendix A.2.

3.4 Estimation and Asymptotics

We use a generalized method of moments (GMM) procedure to estimate θ. Specifically, we generalize the moment condition (11) to write:

$$\mathbb{E}\left[Z_{jm}\left(\check{\theta}, \check{\pi}\right) \cdot \left(\mathcal{D}_j\left(\boldsymbol{s}_m; \sigma, \boldsymbol{x}_m^{(1}, \boldsymbol{p}_m \right) - \alpha p_{jm} - \mathcal{B}_j(\boldsymbol{x}_m, \boldsymbol{q}_m; \gamma, \check{\theta}) \right) \right] = 0, \tag{20}$$

where $Z_{jm}\left(\check{\theta}, \check{\pi}\right)$ is a vector of recentered instruments of the same dimensionality as θ, constructed as above, now with the dependence on preliminary parameter values $(\check{\theta}, \check{\pi})$ made explicit. The $\mathcal{B}_j(\boldsymbol{x}_m, \boldsymbol{q}_m; \gamma, \check{\theta})$ term is included to reduce residual variation in the error $\Xi_{jm} = \beta' x_{jm} + \xi_{jm}$. Since recentered IVs are uncorrelated with any function of $(\boldsymbol{x}_m, \boldsymbol{q}_m)$, the moment conditions (20) hold at the true parameter values regardless of $\mathcal{B}_j(\cdot)$ and for any γ.

Two examples of $\mathcal{B}_j(\cdot)$ are illuminating. First, it can be set to $\gamma' x_{jm}$ with γ estimated as a projection coefficient (i.e., from regressing the estimate of Ξ_{jm} on x_{jm}). While γ need not coincide with the causal effect of characteristics on demand, β, this is not a problem for many important policy counterfactuals (as discussed in Section 3.1). Second, in panel data, one can set $\mathcal{B}_j$ to be the lagged value of Ξ_{jm} obtained from the lagged prices and shares using the initial parameter values, $\check{\theta}$ (and with no additional parameters, $\gamma = \emptyset$). This corresponds to estimating the model "in

[25] Salanie and Wolak (2022) use an approximation of mixed logit around $\sigma = 0$ to simplify share inversion and estimation. Our proof of Appendix Proposition A1 offers a new derivation that yields additional intuition, discussed in Appendix A.2.

differences," as commonly done in linear models and sometimes also for nonlinear demand models (e.g., Adão et al. (2017)). Such differencing helps reduce residual variation when the unobserved demand shifters are strongly serially correlated.

Our baseline estimator is the $(\hat{\theta}, \hat{\gamma})$ that solves the sample analog of condition (20), averaging across product-market pairs, along with the sample analog of a moment condition for γ:

$$\mathbb{E}\left[\frac{\partial \mathcal{B}_j(\boldsymbol{x}_m, \boldsymbol{q}_m; \gamma, \check{\theta})}{\partial \gamma} \cdot \left(\mathcal{D}_j\left(\boldsymbol{s}_m; \sigma, \boldsymbol{x}_m^{(1)}, \boldsymbol{p}_m\right) - \alpha p_{jm} - \mathcal{B}_j(\boldsymbol{x}_m, \boldsymbol{q}_m; \gamma, \check{\theta})\right)\right] = 0. \tag{21}$$

This condition defines γ as the coefficient giving the least-squares fit of Ξ_{jm} on $\mathcal{B}_j(\boldsymbol{x}_m, \boldsymbol{q}_m; \gamma, \check{\theta})$ (e.g. a projection on x_{jm} in the above linear case). It can be understood in the same way as how, in regression analyses of randomized control trials, predetermined controls are included to soak up residual outcome variation with the coefficients on them not interpreted causally. Here we assume the researcher has obtained initial values of θ by, say, an initial GMM procedure with $\tilde{g}_{jm}$ and conventional characteristic-based IVs as instruments. The researcher has also obtained an initial pass-through constant $\check{\pi}$, for example from least-squares estimation of $p_{jm} = \pi_0 + \pi \tilde{g}_{jm} + \epsilon_{jm}$.

We also consider a "continuously updating" estimator which, unlike the baseline estimator, does not require initial values of θ. This estimator replaces $\check{\theta}$ in the moment conditions (20)–(21) with θ; that is, the instruments Z_{jm} (and, if applicable, the $\mathcal{B}_j$ term) are updated when searching for the parameter estimate.[26] We use this estimator in our simulations, below. Another alternative is to use a two-step or iterative GMM procedure.

Consistency and asymptotic normality of these estimators follow from standard GMM theory (e.g. Newey and McFadden (1994, Theorems 2.6 and 3.1)) when there are many *iid* markets m (or, more generally, many *iid* market clusters: e.g., in a panel with many regions and a small number of time periods). In other cases, such as when there are only a few markets or when across-market linkages create dependences in the instruments and GMM residuals, asymptotic properties can be established from many *iid* shocks g_{jm} (or, more generally, many *iid* shock clusters) following Adão et al. (2019), Borusyak et al. (2022b), and Borusyak and Hull (2023). This strategy is helpful, for instance, when the markets represent regions and the shocks arise from exchange rate fluctuations, which affect all regions at once. Other shocks can affect the demand for similar products across multiple regions, too. We develop this approach in a setting where $\gamma = \emptyset$ (e.g. the differencing case discussed above), Z_{jm} consists of shift-share instruments, and the preliminary values $(\check{\theta}, \check{\pi})$ are

[26] We use the term "continuously updating" differently to the standard continuously updating estimator, where it refers to the choice of the GMM weighting matrix. That choice is not relevant to our just-identified setting.

non-stochastic.[27] The estimator $\hat{\theta}$ then solves:

$$\begin{aligned}0 &= \sum_m \sum_{j\in\mathcal{J}_m} \left(\sum_{k\in\mathcal{J}_m} w_{jkm}\tilde{g}_{km}\right)\left(\mathcal{D}_j\left(\boldsymbol{s}_m;\hat{\sigma},\boldsymbol{x}_m^{(1)},\boldsymbol{p}_m\right) - \hat{\alpha}p_{jm} - \mathcal{B}_j(\boldsymbol{x}_m,\boldsymbol{q}_m,\check{\theta})\right)\\ &= \sum_m \sum_{k\in\mathcal{J}_m} \tilde{g}_{km}\mathcal{R}_{km}(\hat{\theta}) \end{aligned} \tag{22}$$

where $\mathcal{R}_{km}(\hat{\theta}) = \sum_{j\in\mathcal{J}_m} w_{jkm}\left(\mathcal{D}_j\left(\boldsymbol{s}_m;\hat{\sigma},\boldsymbol{x}_m^{(1)},\boldsymbol{p}_m\right) - \hat{\alpha}p_{jm} - \mathcal{B}_j(\boldsymbol{x}_m,\boldsymbol{q}_m,\check{\theta})\right)$ is an "aggregated" shock-level residual in the sense of Adão et al. (2019) and Borusyak et al. (2022b).[28] Equation (22) represents $\hat{\theta}$ as the solution of a "shock-level" GMM procedure, with an estimable variance of $\sum_m \sum_{k\in\mathcal{J}_m} \tilde{g}_{km}\mathcal{R}_{km}(\theta)$ given many *iid* shocks or many shock clusters that allow for shock correlations across markets. Standard GMM expressions can then be applied, as before, regardless of the correlation structure in the residual $\mathcal{D}_j\left(\boldsymbol{s}_m;\sigma,\boldsymbol{x}_m^{(1)},\boldsymbol{p}_m\right) - \alpha p_{jm} - \mathcal{B}_j(\boldsymbol{x}_m,\boldsymbol{q}_m,\check{\theta})$ across both products and markets. Convergence of $\hat{\theta}$ only requires the shocks to induce sufficient variation across product-market pairs in the IVs.

Results on the asymptotic efficiency of recentered IVs for linear structural equations, developed by Borusyak and Hull (2025), can also be extended to characterize the optimal IV matrix, $Z^* = \left(Z_{jm}^{*\prime}\right)_{m,j\in\mathcal{J}_m}$ without assuming many independent markets. Under appropriate regularity conditions, it takes the form:

$$Z^* = \mathbb{E}\left[\xi\xi' \mid \boldsymbol{x},\boldsymbol{q}\right]^{-1}\left(\mathbb{E}\left[\nabla \mid \boldsymbol{g},\boldsymbol{x},\boldsymbol{q}\right] - \mathbb{E}\left[\nabla \mid \boldsymbol{x},\boldsymbol{q}\right]\right),$$

where $\xi = (\xi_{jm})_{m,j\in\mathcal{J}_m}$, $\nabla = \left(\nabla_{jm}'\right)_{m,j\in\mathcal{J}_m}$, and $\boldsymbol{v} = (\boldsymbol{v}_m)_m$ for any variable $\boldsymbol{v}$. The inner term in parentheses stacks the recentered best predictors of the model's residual derivatives, $\tilde{Z}_{jm}$. The recentered predictor is then adjusted by $\mathbb{E}\left[\xi\xi' \mid \boldsymbol{x},\boldsymbol{q}\right]^{-1}$, which can be understood as combining a partial residualization of $\tilde{Z}_{jm}$ on $\mathbb{E}\left[\xi \mid \boldsymbol{x},\boldsymbol{q}\right]$ and a reweighting by $\mathrm{Var}\left[\xi \mid \boldsymbol{x},\boldsymbol{q}\right]^{-1}$ (see Proposition 3 in Borusyak and Hull (2025)). This characterization provides further motivation for our focus on approximating $\tilde{Z}_{jm}$ as well as for the adjustment for $\mathcal{B}_j(\boldsymbol{x}_m,\boldsymbol{q}_m,\check{\theta})$ in (20), as a proxy for $\mathbb{E}\left[\xi_{jm} \mid \boldsymbol{x},\boldsymbol{q}\right]$, in estimation. Weighting by an estimate of the residual's inverse variance, as in feasible generalized least squares, is less popular in practice and not pursued here.

3.5 Extensions

We now develop several extensions of the baseline model. We consider them one by one to avoid notational clutter, but in practice they can be combined.

Observed Consumer Characteristics. In some applications, the researcher observes the distribution of consumer characteristics in each market and allows these consumer characteristics

[27] For recentered instruments that do not have a shift-share structure, Borusyak and Hull (2023) provide sufficient conditions for consistency in linear IV settings. Adapting them to our current setting is left to future work. No general asymptotic inference results are currently known for such instruments, even for linear IV settings.

[28] Note that we include the "price instrument" $-\check{\pi}\tilde{g}_{jm}$ as a shift-share IV here, with $w_{jkm} = -\check{\pi}\mathbf{1}[j=k]$.

correlate with tastes for product characteristics in $\boldsymbol{x}_m^{(1)}$. Our results extend immediately to that case. Specifically, the consumer with characteristics $c_i = (c_{ir})_{r=1}^R$ solves:

$$\max_{j \in \mathcal{J}_m \cup \{0\}} \delta_{jm} + \left(\sum_{r=1}^{R} \gamma_{r0} c_{ir} + \eta_{i0} \right) p_{jm} + \sum_{\ell=1}^{L_1} \left(\sum_{r=1}^{R} \gamma_{r\ell} c_{ir} + \eta_{i\ell} \right) x_{jm\ell} + \varepsilon_{ijm},$$

where $\gamma = (\gamma_{r\ell})$ serve as additional nonlinear parameters and extreme-value shocks ε_{ijm} are independent from (c_i, η_i). The distribution $\mathcal{P}_m(\cdot; \sigma)$ of (c_i, η_i) is known: typically c_i is assumed independent of η_i with the market-specific distribution taken from the data. Again, the model is invertible (see, e.g., Gandhi and Nevo (2021)), and the rest of the analysis goes through without change.

Using Lagged Prices and Shares with Product Entry and Exit. When lagged prices and shares are available, our baseline recommendation is to use them as $(\check{\boldsymbol{p}}_m, \check{\boldsymbol{s}}_m)$ when constructing the instruments. This approach requires a modification when some products have recently entered the market and their lagged information is not available. Moreover, if many products have exited, lagged shares may be a poor prediction of the current period's share in the absence of cost shocks.

In such cases, our proposal is to predict prices $\check{p}_{jm}$ and mean utilities $\check{\delta}_{jm}$ for all products in the current period and construct shares from them, as $\check{s}_{jm} = \mathcal{S}_j(\check{\boldsymbol{\delta}}_m; \check{\sigma}, \boldsymbol{x}_m^{(1)}, \check{\boldsymbol{p}}_m)$. For continuing products, lagged price can serve as $\check{p}_{jm}$, while mean utility can be obtained from the inversion of lagged shares, given $\check{\sigma}$. For new products, one may proceed as in a single cross-section, taking fitted values from regressions of realized price and implied mean utility $\mathcal{D}_j\left(\boldsymbol{s}_m; \check{\sigma}, \boldsymbol{x}_m^{(1)}, \boldsymbol{p}_m\right)$ on characteristics.

Identification of β via Instruments for Product Entry. Our baseline analysis assumes that the $\boldsymbol{g}_m$ shocks do not affect the characteristics of available products, which leaves the causal effect of characteristics on mean utility, β, unidentified. This is in contrast to the price coefficient α which is identified because price is affected by the shocks and the researcher is able to construct a relevant recentered instrument, $\check{\pi}\tilde{g}_{jm}$. If the researcher has access to shocks that affect some characteristics in a predictable way, those characteristics can be treated in the same way as the baseline model treats price, and the corresponding components of β become identified.

Incorporating a Pricing Model. Our baseline analysis focuses solely on demand estimation, leaving the supply side flexible. Although we require an auxiliary model of cost shock pass-through, it need not be correct and is indeed very simple in our baseline proposal. This modular approach has the advantage that demand can be estimated with fewer assumptions. However, some counterfactuals, such as a merger between two firms, require predicting how prices would change, and thus taking a stand on how firms set prices. In that case, the researcher may consider leveraging the pricing model to obtain more powerful—albeit less robust—demand estimates, too, as in Berry et al. (1995). While we leave the technical presentation to future drafts, this extension should be straightforward.

Estimated Mapping from Inputs to Products. Our baseline analysis assumes that the cost shocks are product-specific. When the shocks originate from input prices or exchange rate fluctuations, this requires (at least partial) knowledge of the mapping from inputs to products: e.g., the percentages of wheat and corn among the ingredients of ready-to-eat cereals (Barahona et al., 2023) or the country of assembly for each car model (Grieco et al., 2024). However, in some settings where input prices are observed, the mapping to products is not available. Villas-Boas (2007) addresses this problem by using interactions between market-specific input prices and product dummies as instruments. The intuition is that, when the same product is observed in sufficiently many markets (e.g., time periods), the sensitivity of each product's price to all inputs prices is revealed.

This insight can be adapted to our setting, in a two-step approach. First, the exposure of each product to the set of inputs is estimated by a product-specific (e.g., time-series) regression of price on recentered input price shocks. This regression should have sufficiently many observations per product, such that the parameters converge to some pseudo-true values, which need not reflect the true production function. Second, a product-specific cost shock is constructed as a shift-share aggregate of input price shocks as shifts with estimated exposures as shares and used in the rest of the analysis. We leave the precise asymptotic analysis of this approach to future drafts.

Alternative Demand Models. While we have focused on nested and mixed logit, our approach extends to other popular parametric demand systems. Most directly, nested and mixed constant elasticity of substitution (CES) models are closely related, with prices replaced with log prices and quantity shares replaced with expenditure shares. Our instrument construction then goes through. Similarly, our analysis extends directly to variations on the mixed logit model used in IO, such as the Hotelling model of spatial product differentiation (e.g. Houde, 2012) and the "principles of differentiation" model of Bresnahan et al. (1997) which combines multiple nest groupings.

Non-Parametric Demand. We follow Berry and Haile (2014) in considering non-parametric identification of demand under Assumption 1 instead of the conventional stronger assumption (our equation (13)) that they impose. As they show, identification requires an index restriction: that at least one component of $(p_{jm}, x'_{jm})'$ enters demand without a random coefficient. Moreover, exogenous cost shocks are only sufficient for non-parametric identification if price satisfies this property. We show that this result extends to our weaker assumption, too:

Proposition 1. *Consider the non-parametric inverse demand model with an index restriction on price:*

$$p_{jm} + \xi_{jm} = \mathcal{D}_j(\boldsymbol{s}_m, \boldsymbol{x}_m) \tag{23}$$

for an unknown set of functions $\mathcal{D}_j$. Suppose Assumption 1 holds with $\boldsymbol{q}_m = \emptyset$ and the cost shocks satisfy a completeness property: for any function $h(\boldsymbol{s}_m, \boldsymbol{x}_m)$ with finite expectation, $\mathbb{E}\left[h(\boldsymbol{s}_m, \boldsymbol{x}_m) \mid \boldsymbol{g}_m, \boldsymbol{x}_m\right] = 0$ a.s. implies $h(\boldsymbol{s}_m, \boldsymbol{x}_m) = 0$ a.s. Then $\mathcal{D}_j(\cdot)$ and the unobserved demand shifter ξ_{jm} are identified up to an additive term $\beta_j(\boldsymbol{x}_m)$. Moreover, cross-price elasticities are point-identified.

4 Monte Carlo Simulations

We now analyze the bias and variance properties of the recentered IV approach, relative to conventional alternatives, in a Monte Carlo simulation that largely follows Gandhi and Houde (2020). Section 4.1 describes the baseline data-generating process, where both conventional and recentered IVs are valid, and shows what data features drive the variance of estimates in the two approaches. Section 4.2 then introduces product characteristic endogeneity, demonstrating that our proposed IVs remain accurate while conventional characteristic-based IVs are significantly biased. Details of the algorithms used for estimation are reported in Appendix B.

4.1 Mixed Logit with Exogenous Characteristics

Simulation Design. We simulate a set of regions $r = 1, \ldots, 100$ in two periods $t \in \{1, 2\}$; hence $m = (r, t)$. In each period, consumers choose between products $j \in \mathcal{J}_m = \{1, \ldots, 15\}$ and the outside good to maximize their utility, according to equation (6). We consider $L_1 = 2$ observed time-invariant characteristics $x_{jr\ell} \overset{iid}{\sim} N(0, 1)$, in addition to the intercept $x_{jr0} = 1$. Random coefficients are placed on both characteristics, $x^{(1)}_{jm} = (x_{jm1}, x_{jm2})$ but not on price. The random coefficients $\eta_{i\ell} \overset{iid}{\sim} N(0, \sigma^2_\ell)$ have true standard deviations of $\sigma_\ell = 4$ for $\ell = 1, 2$. Product j's mean utility δ_{jm} is determined each period according to equation (7) with persistent unobserved demand shifters: $\xi_{jr1} \overset{iid}{\sim} N(0, 1)$ and $\xi_{jr2} = 0.9\xi_{jm1} + \sqrt{1 - 0.9^2} \cdot e_{jm}$ for $e_{jm} \overset{iid}{\sim} N(0, 1)$. We set $\beta_0 = 35$, $\beta_1 = \beta_2 = 2$, and $\alpha = -0.2 - 4\exp(0.5)$. Market shares are simulated with 1,000 independent draws:

$$s_{jm} = \frac{1}{1000} \sum_{i=1}^{1000} \frac{\exp\left(\delta_{jm} + \eta_i' x^{(1)}_{jm}\right)}{1 + \sum_{k \in \mathcal{J}_m} \exp\left(\delta_{km} + \eta_i' x^{(1)}_{km}\right)}. \tag{24}$$

Prices are set by a simultaneous Bertrand-Nash game where each product is produced by a single firm. In each period the price vector for each region $\boldsymbol{p}_{mt}$ is the solution to the following system of equations derived from the firms' first-order conditions:

$$\boldsymbol{p}_m = \boldsymbol{c}_m - \left[\frac{d\boldsymbol{S}(\boldsymbol{\delta}_m; \sigma, \boldsymbol{x}^{(1)}_m, \boldsymbol{p}_m)}{d\boldsymbol{p}'_m}\right]^{-1} \cdot \boldsymbol{S}(\boldsymbol{\delta}_m; \sigma, \boldsymbol{x}^{(1)}_m, \boldsymbol{p}_m). \tag{25}$$

where $c_{jm} = \gamma' x_{jm} + \omega_{jm} + g_{jm}$ is firm j's marginal cost (and the derivative with respect to price includes the effect through $\boldsymbol{\delta}_m$). We set $\gamma_0 = 5$ and $\gamma_1 = \gamma_2 = 1$ and generate persistent unobserved cost shocks: $\omega_{jr1} \overset{iid}{\sim} N(0, 1)$ and $\omega_{jr2} = 0.9\omega_{jr1} + \sqrt{1 - 0.9^2} \cdot w_{jr}$ for $w_{jr} \overset{iid}{\sim} N(0, 1)$. The observed cost shocks only happen in the second period, such that $g_{jr1} = 0$ and $g_{jr2} \overset{iid}{\sim} N(0, 0.2^2)$.[29] Note that the variance of g_{jr2} is only 4% of the variance of unobserved cost shocks, consistent with the limited exogenous shock variation expected in typical applications.

[29] The parameters of our simulation were picked to follow the simulation in Gandhi and Houde (2020) as much as possible. The deviations arise for three reasons: we have two periods, we distinguish between observed and unobserved costs shocks, and we do not have a random coefficient on price. Our value for α is picked as the average price coefficient: they set the linear coefficient on price to -0.2 and have (the negative of) log-normal random coefficients with the mean $-4e^{0.5}$. We also set $L_1 = 2$ instead of 4 and $\beta_0 = 35$ instead of 50.

We estimate this model for two alternative sets of moment conditions. For the conventional characteristic-based IVs, let Z^C_{jr2} be a vector collecting g_{jr2}, x_{jr}, and a set of two instruments for $\sigma = (\sigma_1, \sigma_2)$: either BLP (sum of competitor characteristics) instruments or the local or quadratic differentiation IVs for proposed by Gandhi and Houde (2020). These instruments are given by:

$$
\begin{aligned}
\text{BLP Sum of Characteristics IV}: \quad & z_{jr2\ell} = \sum_{k \in \mathcal{J}_r, k \neq j} x_{kr\ell} \\
\text{GH Quadratic Differentiation IV}: \quad & z_{jr2\ell} = \sum_{k \in \mathcal{J}_r, k \neq j} \left(x_{jr\ell} - x_{kr\ell}\right)^2 \\
\text{GH Local Differentiation IV}: \quad & z_{jr2\ell} = \sum_{k \in \mathcal{J}_r, k \neq j} \mathbf{1}\left[|x_{jr\ell} - x_{jr\ell}| < \kappa_\ell\right]
\end{aligned}
$$

with a proximity threshold κ_ℓ; we follow Gandhi and Houde (2020) and use the standard deviation of $x_{jm\ell}$. We then estimate (α, β, σ) via GMM using data from the second period only (when the cost shock is available) and the moment condition:

$$\mathbb{E}\left[Z^C_{jr2}\xi_{jr2}\right] = 0.$$

For the recentered instruments, let Z^R_{jr2} be a vector collecting g_{jr2} and either the shift-share or recentered exact prediction IVs proposed in Section 3. We recenter by permuting the cost-shocks 20 times across both products and markets. We use the continuously updating procedure proposed in Section 3.4 to bypass the need for initial estimates (iterative GMM yields very similar results). The moment condition is:

$$\mathbb{E}\left[Z^R_{jr2}\Delta\xi_{jr}\right] = 0$$

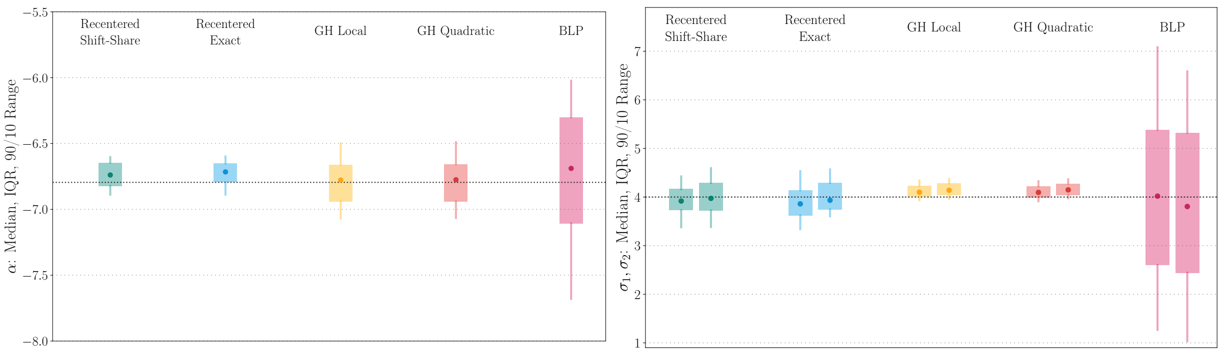
for $\Delta\xi_{jr} = \xi_{jr2} - \xi_{jr1}$. Differencing corresponds to setting $\mathcal{B}_j$ defined in Section 3.4 to the lagged values $\beta' x_{jr} + \xi_{jr1}$ (since characteristics are time-invariant they drop from the estimation) which helps reduce residual variation since the demand shifters are serially correlated. Since we simulate a panel with two periods, it is natural to set the prediction for prices and shares that we will use to construct Z^R_{jr2} to their pre-cost shock (i.e. period one) values, $(\check{p}_{jr2}, \check{s}_{jr2}) = (p_{jr1}, s_{jr1})$. See Appendix B for additional details on our estimation.

Baseline Results. Figure 1 shows the results of our estimation for 100 Monte Carlo simulations. As expected, each set of instruments yields approximately unbiased estimates for each of the parameters. The recentered IVs tend to estimate the price coefficient somewhat more precisely than the differentiation IVs, with a tighter distribution of estimates, while the reverse is true for the nonlinear parameters. BLP instrument estimates are considerably noisier for all parameters; we drop them going forward to focus on the leading characteristic-based IVs.

Sensitivity. We next study how the precision of the recentered IV and differentiation IV estimates varies with two key features of the data-generating process. Figure 2 shows that recentered IVs have less power to estimate the nonlinear parameters σ with a lower variance of cost shocks (both IV approaches benefit from more variable shocks for estimating the price coefficient). In turn,

Figure 1: Baseline Simulation Results

(a) Price Coefficient α (b) Non-Linear Parameters σ_1, σ_2



Notes. The two panels show the simulated distributions for the GMM estimates of α and (σ_1, σ_2) across 100 simulations of the data-generating process described in Section 4.1. The "Recentered Shift-Share" estimates use the shift-share IV described in Section 3.3; "Recentered Exact" estimates use the exact prediction IV (recentered around the average of 20 permutations of the cost shock); "GH Local," "GH Quadratic," and "BLP" correspond to the characteristic IVs described in Section 4.1. For each set of estimates, we plot the median, a box delineating the 25th and 75th percentiles, lines denoting the 10th and 90th percentiles, and a horizontal dashed line denoting the true value of the parameters.

Figure 3 shows that differentiation IVs have lower power for σ with less variation in choice sets across markets. While in our baseline simulation each market has an independent draw of product characteristics, here we make a subset of products the same across all markets, as when sold nationally. In the extreme case where all products are common across markets, differentiation IVs only have variation because product fixed effects are not included in our estimation procedure.

4.2 Endogenous Product Characteristics

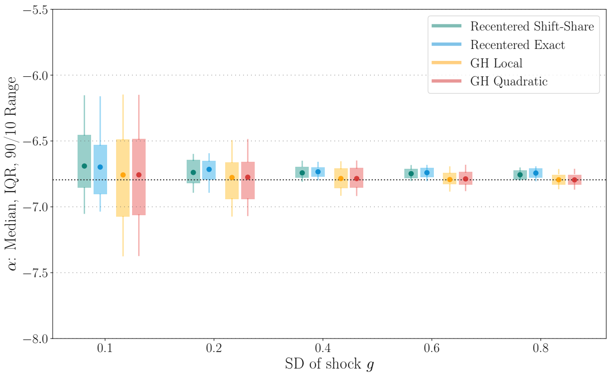
We now show that our proposed IVs continue to accurately estimate the price elasticity parameters even when product characteristics are endogenous, while the differentiation IVs do not. We consider a simple model of characteristic endogeneity that assumes each region has a time-invariant "bliss point" B_r for the first product characteristic.[30] Consumers dislike products far from the bliss point, which we model by subtracting $3(x_{jr1} - B_r)^2$ from ξ_{jrt}. Realizing this, firms introduce more products near the bliss point, which we model by centering the distribution of x_{jr1} around B_r: $x_{jm1} \overset{iid}{\sim} N(B_m, 1)$. Here the differentiation IVs of Gandhi and Houde (2020) are invalid because popular products are in the dense part of the distribution of product characteristics. By contrast, our proposed IVs only require exogeneity of the cost shocks.

Figure 4 summarizes the results: there is little bias when using our proposed IV and a moderate decrease in power relative to Figure 1. However, using either of the two differentiation IV strategies yields substantially biased estimates for σ_1: the estimates are equal to zero in all simulation draws. The reason is that, under exogenous entry, mixed logit predicts a negative correlation between market shares and the degree of local competition as proxied by the differentiated IVs. This

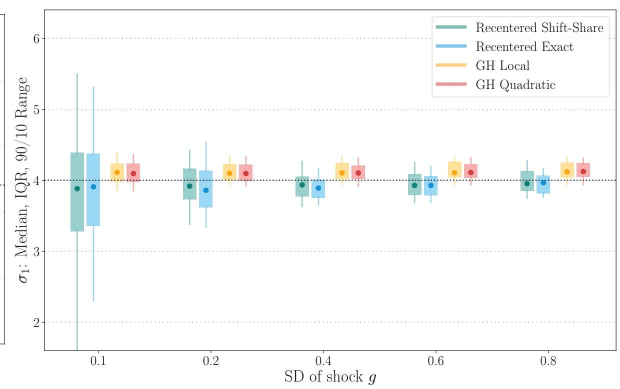
[30]This simulation is in spirit to Gandhi and Houde (2020), Section 4.4.

Figure 2: Role of Cost Shock Variation

(a) Price Coefficient α



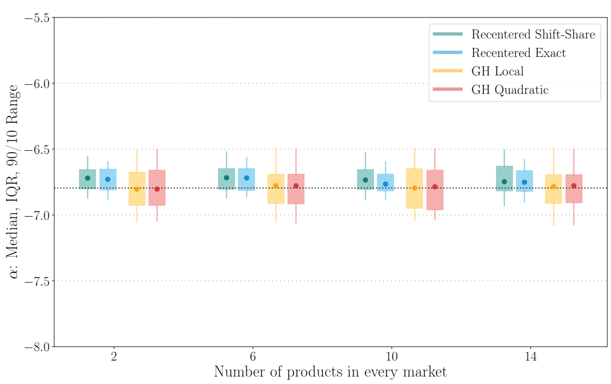
(b) Non-Linear Parameter σ_1



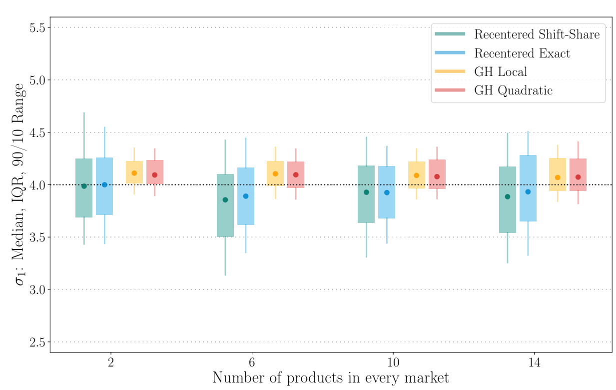
Notes. The two panels show the distributions of the indicated parameter estimates for different values of the standard deviation of the cost shock g_{jr2}. The data-generarating process is otherwise unchanged; see notes to Figure 1.

Figure 3: Role of Cross-Market Characteristic Variation

(a) Price Coefficient α

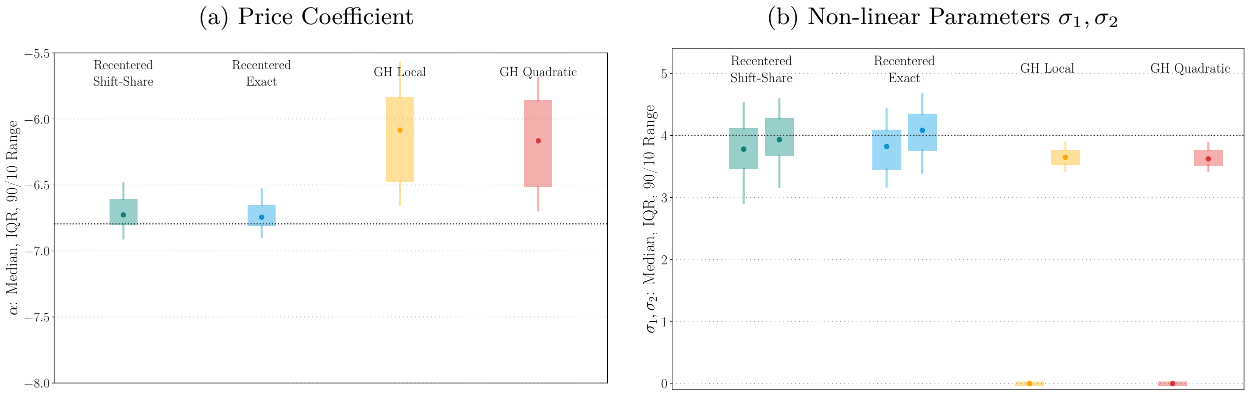


(b) Non-Linear Parameter σ_1



Notes. The two panels show the distributions of the indicated parameter estimates as we vary the number of common products across markets. In each simulation we set $x_{jr\ell} = x_{j1\ell}$ for indicated number of common products $j = 1, \ldots, C$. The data-generarating process is otherwise unchanged; see notes to Figure 1.

Figure 4: Endogenous Characteristics



Notes. The two panels show the distributions of the indicated parameter estimates as we introduce a bliss point B_r for the first characteristic in each region. The data-generarating process is otherwise unchanged; see notes to Figure 1.

negative correlation is especially strong when the variance of random coefficients is high. However, with endogenous entry, as in our simulations, market shares tend to correlate positively with being in the dense part of the characteristic space (because that means being close to the bliss point). This generates a strong negative bias in σ_1 when using the differentiation IVs.[31] Notably, this bias seems to also affect differentiation IV estimation of the price coefficient α and the other nonlinear parameter σ_2; recentered IV estimates remain unbiased for these parameters as well.

5 Conclusion

Modern demand models give a flexible yet tractable structure for substitution across goods. We develop new tools for bringing this structure to data by leveraging its predictions of how key endogenous variables respond to a set of exogenous supply-side shocks. Our recentered IV approach avoids the widespread but often implausible assumption of exogenous product characteristics, letting us "reuse" the exogenous shocks to construct multiple powerful instruments targeted at each of the nonlinear parameters of the model. Simulations suggest recentered IVs can have comparable power to leading characteristic-based IVs while avoiding severe bias from characteristic endogeneity. Future drafts will illustrate this approach in a real setting.

Several open paths remain in this agenda. First, we have only considered here demand estimation with market-level data; the role of recentered IV with individual choice-level data is an interesting question for future research. Second, while we have characterized non-parametric identification of demand with recentered IVs, flexible estimation (using, e.g., modern machine learning tools) is worth further study. Finally, we expect recentered IVs to be useful for identifying structural models beyond demand—such as for dynamic choice or strategic entry in IO, or other phenomena in macroeconomics, international trade, and spatial economics. Developing these extensions may

[31] We thank Jean-Francois Houde for pointing this out to us.

yield practical new ways to improve the credibility and transparency of structural estimation.

References

Ackerberg, D. A., and G. S. Crawford (2009): "Estimating Price Elasticities in Differentiated Product Demand Models with Endogenous Characteristics," *Unpublished*.

Adão, R., C. Arkolakis, and F. Esposito (2023): "General Equilibrium Effects in Space: Theory and Measurement," *Working Paper*.

Adão, R., P. Carrillo, A. Costinot, D. Donaldson, and D. Pomeranz (2022): "Imports, exports, and earnings inequality: Measures of exposure and estimates of incidence," *The Quarterly Journal of Economics*, 137, 1553–1614.

Adão, R., A. Costinot, and D. Donaldson (2017): "Nonparametric counterfactual predictions in neoclassical models of international trade," *American Economic Review*, 107, 633–689.

——— (2024): "Putting Quantitative Models to the Test: An Application to the U.S.-China Trade War," *The Quarterly Journal of Economics*, qjae041.

Adão, R., M. Kolesár, and E. Morales (2019): "Shift-share designs: Theory and inference," *The Quarterly Journal of Economics*, 134, 1949–2010.

Aguiar, L., and J. Waldfogel (2018): "Quality predictability and the welfare benefits from new products: Evidence from the digitization of recorded music," *Journal of Political Economy*, 126, 492–524.

Andrews, I., N. Barahona, M. Gentzkow, A. Rambachan, and J. M. Shapiro (2025): "Causal Interpretation of Structural IV Estimands," *Forthcoming Quarterly Journal of Economics*, 1–41.

Armstrong, T. B. (2016): "Large Market Asymptotics for Differentiated Product Demand Estimators With Economic Models of Supply," *Econometrica*, 84, 1961–1980.

Backus, M., C. Conlon, and M. Sinkinson (2021): "Common Ownership and Competition in the Ready-To-Eat Cereal Industry," *Working Paper*.

Barahona, N., C. Otero, and S. Otero (2023): "Equilibrium effects of food labeling policies," *Econometrica*, 91, 839–868.

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

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," *American Economic Review*, 89, 400–430.

Berry, S. T. (1994): "Estimating discrete-choice models of product differentiation," *The RAND Journal of Economics*, 242–262.

Berry, S. T., and P. A. Haile (2021): "Foundations of demand estimation," in *Handbook of industrial organization* Volume 4: Elsevier, 1–62.

Borusyak, K., R. Dix-Carneiro, and B. Kovak (2022a): "Understanding migration responses to local shocks," *Working Paper*.

Borusyak, K., and P. Hull (2023): "Non-Random Exposure to Exogenous Shocks," *Econometrica*, 91, 2155–2185.

——— (2025): "Optimal Formula Instruments," *Working Paper*.

Borusyak, K., P. Hull, and X. Jaravel (2022b): "Quasi-Experimental Shift-Share Research Designs," *Review of Economic Studies*, 89, 181–213.

——— (2024): "Design-Based Identification with Formula Instruments: a Review," *The Econometrics Journal*, 28, 83–108, 10.1093/ectj/utae003.

Bresnahan, T. F., S. Stern, and M. Trajtenberg (1997): "Market segmentation and the sources of rents from innovation: Personal computers in the late 1980s," *The Rand Journal of Economics*, 28, S17.

Chamberlain, G. (1987): "Asymptotic efficiency in estimation with conditional moment restrictions," *Journal of Econometrics*, 34, 305–334.

——— (1992): "Efficiency bounds for semiparametric regression," *Econometrica: Journal of the Econometric Society*, 567–596.

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

Costinot, A., D. Donaldson, and C. Smith (2016): "Evolving comparative advantage and the impact of climate change in agricultural markets: Evidence from 1.7 million fields around the world," *Journal of Political Economy*, 124, 205–248.

Couture, V., and J. Handbury (2020): "Urban revival in America," *Journal of Urban Economics*, 119, 103267.

Crawford, G. S., O. Shcherbakov, and M. Shum (2019): "Quality overprovision in cable television markets," *American Economic Review*, 109, 956–995.

Dearing, A. (2022): "Estimating structural demand and supply models using tax rates as instruments," *Journal of Public Economics*, 205, 104561.

Dekle, R., J. Eaton, and S. Kortum (2008): "Global rebalancing with gravity: Measuring the burden of adjustment,"Technical report, IMF Staff Papers.

Fajgelbaum, P. D., P. K. Goldberg, P. J. Kennedy, and A. K. Khandelwal (2020): "The return to protectionism," *The Quarterly Journal of Economics*, 135, 1–55.

Fan, Y. (2013): "Ownership consolidation and product characteristics: A study of the US daily newspaper market," *American Economic Review*, 103, 1598–1628.

Fujiy, B. C., D. Ghose, and G. Khanna (2024): "Production Networks and Firm-Level Elasticities of Substitution," *Working Paper*.

Gandhi, A., and J. F. Houde (2020): "Measuring Substitution Patterns in Differentiated Products Industries: The Missing Instruments," *Working Paper*.

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

Goldberg, P. K., and F. Verboven (2001): "The evolution of price dispersion in the european car market," *Review of Economic Studies*, 68, 811–848.

Grieco, P. L., C. Murry, and A. Yurukoglu (2024): "The Evolution of Market Power in the US Auto Industry," *Quarterly Journal of Economics*, 139, 1201–1253.

Halchenko, V., R. Trembovetska, V. Tychkov, and A. Storchak (2020): "The construction of effective multidimensional computer designs of experiments based on a quasi-random additive recursive Rd–sequence," *Applied Computer Systems*, 25, 70–76.

Houde, J.-F. (2012): "Spatial Differentiation and Vertical Mergers in Retail Markets for Gasoline," *American Economic Review*, 102, 2147–2182.

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

Kilian, L., and M. P. Taylor (2003): "Why is it so difficult to beat the random walk forecast of exchange rates?" *Journal of International Economics*, 60, 85–107.

Kocis, L., and W. J. Whiten (1997): "Computational investigations of low-discrepancy sequences," *ACM Transactions on Mathematical Software (TOMS)*, 23, 266–294.

Li, J. (2016): "Compatibility and Investment in the U.S. Electric Vehicle Market," *Working Paper*.

Lu, Z., X. Shi, and J. Tao (2023): "Semi-nonparametric estimation of random coefficients logit model for aggregate demand," *Journal of Econometrics*, 235, 2245–2265.

McFadden, D. (1978): "Modelling the Choice of Residential Location," in *Spatial Interaction Theory and Panning Models*: North Holland: Amsterdam, 75–96.

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

Moon, H. R., M. Shum, and M. Weidner (2018): "Estimation of random coefficients logit demand models with interactive fixed effects," *Journal of Econometrics*, 206, 613–644.

Nakamura, E., and D. Zerom (2010): "Accounting for incomplete pass-through," *The review of economic studies*, 77, 1192–1230.

Nevo, A. (2001): "Measuring market power in the ready-to-eat cereal industry," *Econometrica*, 69, 307–342.

Newey, W. K., and D. McFadden (1994): "Large sample estimation and hypothesis testing," *Handbook of econometrics*, 4, 2111–2245.

Newey, W. K., and J. L. Powell (2003): "Instrumental Variable Estimation of Nonparametric Models," *Econometrica*, 71, 1565–1578.

Petrin, A., M. Ponder, and B. Seo (2022): "Identification and estimation of discrete choice demand models when observed and unobserved characteristics are correlated,"Technical report, Working Paper.

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

Rothenberg, T. J. (1971): "Identification in parametric models," *Econometrica: Journal of the Econometric Society*, 577–591.

Salanie, B., and F. A. Wolak (2022): "Fast, Detail-free, and Approximately Correct: Estimating Mixed Demand Systems," *Working Paper*.

Sweeting, A. (2013): "Dynamic product positioning in differentiated product markets: The effect of fees for musical performance rights on the commercial radio industry," *Econometrica*, 81, 1763–1803.

Town, R., and S. Liu (2003): "The welfare impact of Medicare HMOs," *RAND Journal of Economics*, 719–736.

Varadhan, R., and C. Roland (2008): "Simple and Globally Convergent Methods for Accelerating the Convergence of Any EM Algorithm," *Scandinavian Journal of Statistics*, 35, 335–353.

Villas-Boas, S. B. (2007): "Vertical relationships between manufacturers and retailers: Inference with limited data," *Review of Economic Studies*, 74, 625–652.

Wang, A. (2023): "Sieve BLP: A semi-nonparametric model of demand for differentiated products," *Journal of Econometrics*, 235, 325–351.

Online Appendix

A Theoretical Appendix

A.1 Derivations for Section 2

Market Shares with Nested Logit Demand. Let $\delta_{jm} = \alpha p_{jm} + \xi_{jm}$ be product j's mean utility and $D_{nm} = \sum_{j \in \mathcal{J}_m} d_{jn} \exp\left(\delta_{jm}/(1-\sigma)\right)$. Then, as is well known (e.g, Berry (1994)), nested logit market shares satisfy

$$\frac{s_{jm}}{s_{n(j)m}} = \frac{\exp\left(\delta_{jm}/(1-\sigma)\right)}{D_{n(j)m}}, \tag{A1}$$

$$s_{nm} = \frac{D_{nm}^{1-\sigma}}{1 + \sum_{n'} D_{n'm}^{1-\sigma}}, \tag{A2}$$

$$s_{0m} = \frac{1}{1 + \sum_{n'} D_{n'm}^{1-\sigma}}. \tag{A3}$$

Equation (1) follows from simple manipulation of these terms.

Exact Prediction of Within-Nest Share. Let $\hat{p}_{jm} = \breve{\pi}_0 + \breve{\pi} g_{jm}$ where $\breve{\pi}_0$ is an intercept suppressed in Section 2. Plugging $\delta_{jm} = \breve{\alpha}\hat{p}_{jm}$ into the within-nest share expression (A1) yields (3):

$$\begin{aligned}
\widehat{\log} \frac{s_{jm}}{s_{n(j)m}} &= \frac{\breve{\alpha}}{1-\breve{\sigma}} \left(\breve{\pi}_0 + \breve{\pi} g_{jm}\right) - \log \sum_{k \in \mathcal{J}_m} d_{kn(j)} \exp\left(\frac{\breve{\alpha}}{1-\breve{\sigma}} \left(\breve{\pi}_0 + \breve{\pi} g_{km}\right)\right) \\
&= \frac{\breve{\alpha}}{1-\breve{\sigma}} \breve{\pi} g_{jm} - \log \sum_{k \in \mathcal{J}_m} d_{kn(j)} \exp\left(\frac{\breve{\alpha}}{1-\breve{\sigma}} \breve{\pi} g_{km}\right).
\end{aligned}$$

First-Order Approximation. We now take a first-order approximation of (3) around $g_{km} = \mu_g$ for all k:

$$\begin{aligned}
\widehat{\log} \frac{s_{jm}}{s_{n(j)m}} \approx{}& \frac{\breve{\alpha}}{1-\breve{\sigma}} \breve{\pi} \mu_g - \log \sum_{k \in \mathcal{J}_m} d_{kn} \exp\left(\frac{\breve{\alpha}}{1-\breve{\sigma}} \breve{\pi} \mu_g\right) \\
&+ \frac{\breve{\alpha}}{1-\breve{\sigma}} \breve{\pi} \left(g_{jm} - \mu_g\right) - \frac{\sum_{k \in \mathcal{J}_m} d_{kn} \exp\left(\frac{\breve{\alpha}}{1-\breve{\sigma}} \breve{\pi} \mu_g\right) \frac{\breve{\alpha}}{1-\breve{\sigma}} \breve{\pi} \left(g_{km} - \mu_g\right)}{\sum_{k \in \mathcal{J}_m} d_{kn} \exp\left(\frac{\breve{\alpha}}{1-\breve{\sigma}} \breve{\pi} \mu_g\right)} \\
={}& -\log N_{n(j)m} + \frac{\breve{\alpha}}{1-\breve{\sigma}} \breve{\pi} \left(g_{jm} - \frac{1}{N_{n(j)m}} \sum_{k \in \mathcal{J}_m} d_{kn} g_{km}\right).
\end{aligned}$$

We interpret the second term as the response of $\log(s_{jm}/s_{n(j)m})$ to the set of cost shocks (while the first term does not depend on the shocks and is eliminated by recentering). This response is therefore equal to z_{jm} in (2), up to a non-zero scaling factor that does not affect IV estimation.

Exact Prediction Using Lagged Shares. Suppose the choice set $\mathcal{J}_m$ has not changed since the pre-period and g_{jm} are shocks to price changes, such that $p_{jm} = p_{jm}^{\text{pre}} + \pi g_{jm} + \omega_{jm}$. Let ξ_{jm}^{pre} be the unobserved taste shifter in the pre-period[32] such that $\delta_{jm}^{\text{pre}} = \alpha p_{jm}^{\text{pre}} + \xi_{jm}^{\text{pre}}$ and, according to (A1),

$$\frac{s_{jm}^{\text{pre}}}{s_{n(j)m}^{\text{pre}}} = \frac{\exp\left(\delta_{jm}^{\text{pre}}/(1-\sigma)\right)}{\sum_{k\in\mathcal{J}_m} d_{kn(j)} \exp\left(\delta_{jm}^{\text{pre}}/(1-\sigma)\right)}.$$

We predict prices in the period of interest as $\hat{p}_{jm} = p_{jm}^{\text{pre}} + \check{\pi} g_{jm}$ and predict mean utilities by using the predicted price and the pre-period taste shifter:

$$\hat{\delta}_{jm} = \check{\alpha}\hat{p}_{jm} + \xi_{jm}^{\text{pre}} = \delta_{jm}^{\text{pre}} + \check{\alpha}\check{\pi} g_{jm}.$$

Substituting these into (A1) analogously to the exact hat algebra technique of Dekle et al. (2008) yields the exact prediction:

$$\begin{aligned}
\widehat{\log}\frac{s_{jm}}{s_{n(j)m}} &= \log \frac{\exp\left(\hat{\delta}_{jm}/(1-\check{\sigma})\right)}{\sum_{k\in\mathcal{J}_m} d_{kn(j)} \exp\left(\hat{\delta}_{km}/(1-\check{\sigma})\right)} \\
&= \log \frac{\exp\left(\delta_{jm}^{\text{pre}}/(1-\sigma)\right)\cdot \exp\left(\check{\alpha}\check{\pi} g_{jm}/(1-\check{\sigma})\right)}{\sum_{k\in\mathcal{J}_m} d_{kn(j)} \exp\left(\delta_{km}^{\text{pre}}/(1-\check{\sigma})\right)\cdot \exp\left(\check{\alpha}\check{\pi} g_{km}/(1-\check{\sigma})\right)} \\
&= \log \frac{\exp\left(\delta_{jm}^{\text{pre}}/(1-\check{\sigma})\right)}{\sum_{k\in\mathcal{J}_m} d_{kn(j)} \exp\left(\delta_{km}^{\text{pre}}/(1-\check{\sigma})\right)} + \log \frac{\exp\left(\check{\alpha}\check{\pi} g_{jm}/(1-\check{\sigma})\right)}{\sum_{k\in\mathcal{J}_m} d_{kn(j)} \frac{s_{km}^{\text{pre}}}{s_{n(j)m}^{\text{pre}}} \exp\left(\check{\alpha}\check{\pi} g_{km}/(1-\check{\sigma})\right)} \\
&= \log \frac{s_{jm}^{\text{pre}}}{s_{n(j)m}^{\text{pre}}} + \frac{\check{\alpha}\check{\pi}}{1-\check{\sigma}} g_{jm} - \log \sum_{k\in\mathcal{J}_m} d_{kn(j)} \frac{s_{km}^{\text{pre}}}{s_{n(j)m}^{\text{pre}}} \exp\left(\frac{\check{\alpha}\check{\pi}}{1-\check{\sigma}} g_{km}\right).
\end{aligned} \tag{A4}$$

As with exact hat algebra, lagged shares serve as sufficient statistics in this prediction, while lagged prices and taste shifters need not be observed or computed.

Recentering this prediction over the distribution of shocks yields a recentered exact prediction

$$\begin{aligned}
&\frac{\check{\alpha}\check{\pi}}{1-\check{\sigma}}\left(g_{jm}-\mu_g\right) - \log \sum_{k\in\mathcal{J}_m} d_{kn(j)} \frac{s_{km}^{\text{pre}}}{s_{n(j)m}^{\text{pre}}} \exp\left(\frac{\check{\alpha}\check{\pi}}{1-\check{\sigma}} g_{km}\right) \\
&\qquad + \mathbb{E}\left[\log \sum_{k\in\mathcal{J}_m} d_{kn(j)} \frac{s_{km}^{\text{pre}}}{s_{n(j)m}^{\text{pre}}} \exp\left(\frac{\check{\alpha}\check{\pi}}{1-\check{\sigma}} g_{km}\right) \mid \left(d_{kn}, s_{km}^{\text{pre}}\right)_{k\in\mathcal{J}_m,n}\right].
\end{aligned}$$

[32]Given parameters, ξ_{jm}^{pre} can be inverted from the pre-period prices and shares; however, this is not necessary.

First-Order Approximation Using Lagged Shares. Linearizing (A4) around $g_{km} = \mu_g$ for all k yields

$$\widehat{\log}\frac{s_{jm}}{s_{n(j)m}} \approx \log\frac{s_{jm}^{\text{pre}}}{s_{n(j)m}^{\text{pre}}} + \frac{\check{\alpha}\check{\pi}}{1-\check{\sigma}} g_{jm} - \frac{\check{\alpha}\check{\pi}}{1-\check{\sigma}}\mu_g - \sum_{k\in\mathcal{J}_m} d_{kn(j)} \frac{s_{km}^{\text{pre}}}{s_{n(j)m}^{\text{pre}}} \frac{\check{\alpha}\check{\pi}}{1-\check{\sigma}}\left(g_{km}-\mu_g\right)$$
$$= \log\frac{s_{km}^{\text{pre}}}{s_{n(j)m}^{\text{pre}}} + \frac{\check{\alpha}\check{\pi}}{1-\check{\sigma}}\left(g_{jm} - \sum_{k\in\mathcal{J}_m} d_{kn(j)}\frac{s_{km}^{\text{pre}}}{s_{n(j)m}^{\text{pre}}} g_{km}\right).$$

Recentering this expression eliminates the first term, while the second term is equal to z_{jm}^* from (5) up to a non-zero scaling factor.

A.2 Local-to-Logit Approximation

In this section we consider a model in which the random coefficients on $x_{jm}^{(1)}$, and potentially also a random coefficient on price, are equal to $\eta_{i\ell} = \sigma_\ell \nu_{i\ell}$ for mutually uncorrelated $\nu_{i\ell}$ with $\mathbb{E}\left[\nu_{i\ell}\right] = 0$ and $\mathbb{E}\left[\nu_{i\ell}^2\right] = 1$ (but any marginal distributions). We suppress the dependence of the $\mathcal{S}_j$ and $\mathcal{D}_j$ on $\boldsymbol{x}_m^{(1)}$ and $\boldsymbol{p}_j$ to simplify notation. We first state a lemma characterizing market shares and the share inversion function in the "local-to-logit" approximation to the mixed logit model, which provides a new intuition for the mechanics of mixed logit and for the approximations derived by Salanie and Wolak (2022, Theorem 2). We then characterize Z_{jm}^{SSIV} from Section 3.3 when the preliminary parameter values are local-to-logit, regardless of the true parameters.

Lemma A1. *In the above model, without a random coefficient on price, the following Taylor expansions in $\sigma = (\sigma_\ell)_{\ell=1}^{L_1}$ hold around $\sigma = 0$:*

$$\mathcal{S}_j(\boldsymbol{\delta}_m;\sigma) = \mathcal{S}_j(\boldsymbol{\delta}_m;0)\cdot\left[1+\sum_{\ell=1}^{L_1}\frac{\sigma_\ell^2}{2}\left(\left(x_{jm\ell}-\bar{x}_{m\ell}\right)^2 - \sum_{k\in\mathcal{J}_m\cup\{0\}} s_{km}\left(x_{km\ell}-\bar{x}_{m\ell}\right)^2\right)\right] + O(\sigma^3), \quad \text{(A5)}$$

$$\log\frac{\mathcal{S}_j(\boldsymbol{\delta}_m;\sigma)}{\mathcal{S}_0(\boldsymbol{\delta}_m;\sigma)} = \delta_{jm} + \sum_{\ell=1}^{L_1} a_{jm\ell}\frac{\sigma_\ell^2}{2} + O(\sigma^3), \quad \text{(A6)}$$

$$\mathcal{D}_j\left(\boldsymbol{s}_m;\sigma\right) = \log\frac{s_{jm}}{s_{0m}} - \sum_{\ell=1}^{L_1} a_{jm\ell}\frac{\sigma_\ell^2}{2} + O(\sigma^3), \quad \text{(A7)}$$

where $O(\sigma^q)$ indicates qth- and higher-order terms and

$$a_{jm\ell} = \left(x_{jm\ell} - \sum_{k\in\mathcal{J}_m} s_{km}x_{km\ell}\right)^2 - \left(0 - \sum_{k\in\mathcal{J}_m} s_{km}x_{km\ell}\right)^2, \quad \text{(A8)}$$

with s_{km} understood as $\mathcal{S}_k\left(\boldsymbol{\delta}_m;0\right)$ in equations (A5) and (A6). With a random coefficient on price, the same expressions apply with price viewed as another characteristic: i.e., with $x_{jm0} \equiv p_{jm}$ and with the summations including $\ell = 0$.

Intuitively, equation (A5) shows that—in the vicinity of simple multinomial logit—increasing σ_ℓ raises the share of good j if and only if the ℓth characteristic of this good is relatively unusual in its market, in the sense that its further away from the market average than the typical product is (where by "further" we mean averaging square distances with market share weights, counting the outside good as one of the products).

We now characterize the shift-share construction from Section 3.3 in this approximation:

Proposition A1. *In the model of this section, the ℓth shift-share instrument from (18) corresponding to the random coefficient on a non-price characteristic satisfies:*

$$\sum_{k\in\mathcal{J}_m} w_{jkm\ell}\tilde{g}_{km} = (-2\check{\alpha}\check{\pi}\check{\sigma}_\ell)\cdot x_{jm\ell}\cdot\sum_{k\in\mathcal{J}_m}\check{s}_{km}\left(x_{km\ell}-\bar{x}_{m\ell}\right)\tilde{g}_{km}+O(\check{\sigma}^2). \tag{A9}$$

The instrument for the random coefficient on price (if included in the model) is

$$\begin{aligned}\sum_{k\in\mathcal{J}_m} w_{jkm0}\tilde{g}_{km} =& 2\check{\sigma}_0\check{\pi}\left(\check{p}_{jm}-\bar{p}_m\right)\tilde{g}_{jm}-2\check{\sigma}_0\check{\pi}\check{p}_{jm}\sum_{k\in\mathcal{J}_m}\check{s}_{km}\tilde{g}_{km}\\ &-2\check{\alpha}\check{\pi}\check{\sigma}_0\cdot p_{jm}\cdot\sum_{k\in\mathcal{J}_m}\check{s}_{km}\left(\check{p}_{jm}-\bar{p}_m\right)\tilde{g}_{km}+O(\check{\sigma}^2),\end{aligned} \tag{A10}$$

where $\bar{p}_m=\sum_{k\in\mathcal{J}_m}\check{s}_{km}\check{p}_{km}$.

Intuition for these results follow from the above discussion of equation (A5). Focus first on non-price characteristics. While cost shocks cannot affect product entry or characteristics under our assumptions, they can still make certain products more or less unusual in the market by reallocating market shares and thus shifting the share-weighted average of characteristics $\bar{x}_{m\ell}$. Under our simple model of cost shock pass-through, $\bar{x}_{m\ell}$ increases whenever products k with higher $x_{km\ell}$ receive a lower shock—as captured by the covariance term in Proposition A1. When $\bar{x}_{m\ell}$ moves up, high-$x_{jm\ell}$ products become less unusual and lose market shares when σ_ℓ is higher while low-$x_{jm\ell}$ products become more unusual and gain market power in that case. Our instrument identifies σ_ℓ by tracing such differential responses, provided there is enough variation in the ℓth market-level aggregate shock. When there is a random coefficient on price, there are additional effects captured by the first two terms in (A10): cost shocks move the price of good j and prices of its competitors. For instance, the first term reflects that if j is priced higher than the market average, a high cost shock that increases its price makes the product vertically more unusual and raises the market share if σ_0 is larger.

We finally specialize equation (A9) to the common case where a random coefficient is included for the intercept $x_{jm\ell}=1$. Given the normalization $x_{0m}=0$, this captures heterogeneous preferences for all inside goods vs. the outside good. In that case,

$$\sum_{k\in\mathcal{J}_m} w_{jkm\ell}\tilde{g}_{km} = (-2\check{\alpha}\check{\pi}\check{\sigma}_\ell)\cdot\check{s}_{0m}\left(1-\check{s}_{0m}\right)\cdot\frac{\sum_{k\in\mathcal{J}_m}\check{s}_{km}\tilde{g}_{km}}{\sum_{k\in\mathcal{J}_m}\check{s}_{km}}+O(\check{\sigma}^2).$$

That is, the instrument is based on the share-weighted average shock to all inside products in the market, which move prices of all inside goods relative to the outside good. The average shock is scaled to place a higher weight on markets with the share of the outside good closer to 0.5.

Proof of Lemma A1. For this lemma, random coefficients on price can be handled simply by viewing price as another characteristic. We therefore omit random coefficients on price without loss of generality.

For consumer i in market m, let

$$s_{ji} = \frac{\exp\left(\delta_{jm} + \sum_{\ell=1}^{L_1} \sigma_\ell \nu_{i\ell} x_{jm\ell}\right)}{1 + \sum_{k\in\mathcal{J}_m} \exp\left(\delta_{km} + \sum_{\ell=1}^{L_1} \sigma_\ell \nu_{i\ell} x_{km\ell}\right)}$$

denote the probability of choosing product $j \in \mathcal{J}_m \cup \{0\}$ given her $\nu_i = (\nu_{i\ell})_{\ell=1}^{L_1}$. Then $s_{jm} = \mathbb{E}_\nu[s_{ji}]$ where $\mathbb{E}_\nu[\cdot]$ denotes the expectation with respect to the distribution of ν_i. Also, denote $\bar{x}_{i\ell} = \sum_{k\in\mathcal{J}_m} s_{ki} x_{km\ell}$ and $\bar{x}_{m\ell} = \sum_{k\in\mathcal{J}_m} s_{km} x_{km\ell}$. Then we have:

$$\frac{\partial \mathcal{S}_j}{\partial \sigma_\ell} = \mathbb{E}_\nu\left[\frac{\partial s_{ji}}{\partial \sigma_\ell}\right] = \mathbb{E}_\nu\left[\nu_{i\ell} s_{ji}\left(x_{jm\ell} - \bar{x}_{i\ell}\right)\right]$$

and

$$\frac{\partial^2 \mathcal{S}_j}{\partial \sigma_\ell \sigma_{\ell'}} = \mathbb{E}_\nu\left[\frac{\partial^2 s_{ji}}{\partial \sigma_\ell \partial \sigma_{\ell'}}\right] = \mathbb{E}_\nu\left[\nu_{i\ell} s_{ji}\left(\nu_{i\ell'}\left(x_{jm\ell} - \bar{x}_{i\ell}\right)\left(x_{jm\ell'} - \bar{x}_{i\ell'}\right) - \sum_{k\in\mathcal{J}_m} x_{km\ell} s_{ki} \nu_{i\ell'}\left(x_{km\ell'} - \bar{x}_{i\ell'}\right)\right)\right].$$

At $\sigma = 0$, all consumers have the same conditional choice probabilities, $s_{ji} = s_{jm}$, and thus

$$\begin{aligned}
\frac{\partial \mathcal{S}_j}{\partial \sigma_\ell}\,|_{\sigma=0} &= \mathbb{E}_\nu[\nu_{i\ell}] \cdot s_{jm}\left(x_{jm\ell} - \bar{x}_{m\ell}\right) \\
&= 0, \\
\frac{\partial^2 \mathcal{S}_j}{\partial \sigma_\ell \partial \sigma_{\ell'}}\,|_{\sigma=0} &\propto \mathbb{E}_\nu[\nu_{i\ell}\nu_{i\ell'}] = 0 \qquad \text{for } \ell' \neq \ell \\
\frac{\partial^2 \mathcal{S}_j}{(\partial \sigma_\ell)^2}\,|_{\sigma=0} &= \mathbb{E}_\nu\left[\nu_{i\ell}^2\right] s_{jm}\left[\left(x_{jm\ell} - \bar{x}_{m\ell}\right)^2 - \sum_{k\in\mathcal{J}_m} s_{km} x_{km\ell}\left(x_{km\ell} - \bar{x}_{m\ell}\right)\right] \\
&= s_{jm}\left[\left(x_{jm\ell} - \bar{x}_{m\ell}\right)^2 - \sum_{k\in\mathcal{J}_m\cup\{0\}} s_{km}\left(x_{km\ell} - \bar{x}_{m\ell}\right)^2\right].
\end{aligned}$$

By a second-order Taylor expansion,

$$\mathcal{S}_j(\boldsymbol{\delta}_m;\sigma) = \mathcal{S}_j(\boldsymbol{\delta}_m;0) \cdot \left[1 + \sum_\ell \frac{\sigma_\ell^2}{2}\left(\left(x_{jm\ell} - \bar{x}_{m\ell}\right)^2 - \sum_{k\in\mathcal{J}_m\cup\{0\}} s_{km}\left(x_{km\ell'} - \bar{x}_{m\ell'}\right)^2\right)\right] + O(\sigma^3),$$

establishing (A5). The Taylor approximation in logs then implies

$$\log \mathcal{S}_j(\boldsymbol{\delta}_m;\sigma) = \log \mathcal{S}_j(\boldsymbol{\delta}_m;0) + \sum_\ell \frac{\sigma_\ell^2}{2}\left(\left(x_{jm\ell} - \bar{x}_{m\ell}\right)^2 - \sum_{k\in\mathcal{J}_m\cup\{0\}} s_{km}\left(x_{km\ell'} - \bar{x}_{m\ell'}\right)^2\right) + O(\sigma^3).$$

Subtracting an analogous expression that holds for the outside good,

$$\log \frac{\mathcal{S}_j(\boldsymbol{\delta}_m;\sigma)}{\mathcal{S}_0(\boldsymbol{\delta}_m;\sigma)} = \log \frac{\mathcal{S}_j(\boldsymbol{\delta}_m;0)}{\mathcal{S}_0(\boldsymbol{\delta}_m;0)} + \sum_\ell \frac{\sigma_\ell^2}{2}\left((x_{jm\ell}-\bar{x}_{m\ell})^2 - (0-\bar{x}_{m\ell})^2\right) + O(\sigma^3)$$
$$= \delta_{jm} + \sum_\ell \frac{\sigma_\ell^2}{2}\left((x_{jm\ell}-\bar{x}_{m\ell})^2 - (0-\bar{x}_{m\ell})^2\right) + O(\sigma^3),$$

where the second line uses the standard result on share inversion with simple multinomial logit, yielding equation (A6).

Finally, plugging in $\mathcal{S}_j(\boldsymbol{\delta}_m;0) = \mathcal{S}_j(\boldsymbol{\delta}_m;\sigma) + O(\sigma)$ into equation (A6) yields (A7), as $\mathcal{D}_j$ is the mapping from the shares to δ_{jm}.

Proof of Proposition A1. We apply the steps of the shift-share IV construction in Section 3.3 to this setting. Because the above derivation shows that $\frac{\partial \mathcal{S}_j}{\partial \sigma_\ell}\mid_{\sigma=0}= 0$,

$$\frac{\partial \mathcal{S}_j(\check{\boldsymbol{\delta}}_m;\sigma)}{\partial \delta_{km}} = \frac{\partial \mathcal{S}_j(\check{\boldsymbol{\delta}}_m;0)}{\partial \delta_{km}} + O(\sigma^2) = s_{jm}\left(\mathbf{1}\left[j=k\right] - s_{km}\right) + O(\sigma^2)$$

for $s_{km} = \mathcal{S}_j\left(\check{\boldsymbol{\delta}}_m;0\right)$. Since $\mathcal{S}_j\left(\check{\boldsymbol{\delta}}_m;0\right) = \mathcal{S}_j\left(\check{\boldsymbol{\delta}}_m;\check{\sigma}\right) + O(\sigma^2)$ by equation (A5), we also have

$$\frac{\partial \mathcal{S}_j(\check{\boldsymbol{\delta}}_m;\sigma)}{\partial \delta_{km}} = \check{s}_{jm}\left(\mathbf{1}\left[j=k\right] - \check{s}_{km}\right) + O(\sigma^2),$$

and thus

$$\hat{s}_{jm} - \check{s}_{jm} = \check{\alpha}\check{\pi}\sum_{k\in\mathcal{J}_m} \frac{\partial \mathcal{S}_j(\check{\boldsymbol{\delta}}_m;\check{\sigma})}{\partial \delta_{km}}\tilde{g}_{km}$$
$$= \check{\alpha}\check{\pi}s_{jm}\left(\tilde{g}_{jm} - \sum_{k\in\mathcal{J}_m} s_{km}\tilde{g}_{km}\right) + O(\sigma^2).$$

Next, by equation (A7) and writing equation (A8) as

$$a_{jm\ell} = x_{jm\ell}^2 - 2x_{jm\ell}\cdot \sum_{k\in\mathcal{J}_m} s_{km}x_{km\ell}, \tag{A11}$$

we have

$$\frac{\partial^2 \mathcal{D}_j\left(\check{\boldsymbol{s}}_m;\sigma\right)}{\partial s_{km}\partial\sigma_\ell} = \sigma_\ell \frac{\partial a_{jm\ell}}{\partial s_{km}} + O(\sigma^2) = -2\sigma_\ell x_{jm\ell}x_{km\ell} + O(\sigma^2).$$

Also, recalling that $p_{jm} \equiv x_{jm0}$, equation (A5) implies that $\frac{\partial^2 \mathcal{D}_j(\check{\boldsymbol{s}}_m;\sigma)}{\partial p_{km}\partial\sigma_\ell} = O(\sigma^2)$ for $\ell \neq 0$ while for $\ell = 0$ equations (A7) and (A11) yield

$$\frac{\partial^2 \mathcal{D}_j\left(\check{\boldsymbol{s}}_m;\sigma\right)}{\partial p_{km}\partial\sigma_0} = \sigma_0 \frac{\partial a_{jm0}}{\partial p_{km}} + O(\sigma^2) = 2\sigma_0\left[\left(\check{p}_{jm} - \bar{p}_m\right)\mathbf{1}\left[j=k\right] - \check{p}_{jm}\check{s}_{km}\right] + O(\sigma^2).$$

By (18), the instrument for σ_ℓ, which is the recentered ℓth component of the predicted residual derivative evaluated at $\check{\boldsymbol{s}}_m$, $\check{\boldsymbol{p}}_m$, and $\check{\sigma}$, equals:

$$\hat{\nabla}^\sigma_{jm\ell} - \check{\nabla}^\sigma_{jm\ell} = \sum_{k\in\mathcal{J}_m} \frac{\partial^2 \mathcal{D}_j\left(\check{\boldsymbol{s}}_m;\check{\sigma}\right)}{\partial p_{km}\partial\sigma_\ell}\left(\hat{p}_{km} - \check{p}_{km}\right) + \sum_{k\in\mathcal{J}_m} \frac{\partial^2 \mathcal{D}_j\left(\check{\boldsymbol{s}}_m;\check{\sigma}\right)}{\partial s_{km}\partial\sigma_\ell}\left(\hat{s}_{km} - \check{s}_{km}\right).$$

For $\ell \neq 0$, the first term is equal to $O(\check{\sigma}^2)$ and thus

$$\begin{aligned}
\hat{\nabla}^{\sigma}_{jm\ell} - \check{\nabla}^{\sigma}_{jm\ell} &= -2\check{\sigma}_\ell \check{\alpha}\check{\pi} x_{jm\ell} \sum_{k \in \mathcal{J}_m} \check{s}_{km} x_{km\ell} \left(\tilde{g}_{km} - \sum_{k' \in \mathcal{J}_m} \check{s}_{k'm} \tilde{g}_{k'm} \right) + O(\check{\sigma}^2) \\
&= -2\check{\sigma}_\ell \check{\alpha}\check{\pi} x_{jm\ell} \sum_{k \in \mathcal{J}_m} \check{s}_{km} \left(x_{km\ell} - \bar{x}_{m\ell} \right) \tilde{g}_{km} + O(\check{\sigma}^2),
\end{aligned}$$

where the second line used the properties of covariances. In turn, for $\ell = 0$,

$$\begin{aligned}
\hat{\nabla}^{\sigma}_{jm0} - \check{\nabla}^{\sigma}_{jm0} &= 2\check{\sigma}_0 \check{\pi} \sum_{k \in \mathcal{J}_m} \left[\left(\check{p}_{jm} - \bar{p}_m \right) \mathbf{1}\left[j = k \right] - \check{p}_{jm} \check{s}_{km} \right] \tilde{g}_{km} \\
&\quad - 2\check{\sigma}_0 \check{\alpha}\check{\pi} \check{p}_{jm} \sum_{k \in \mathcal{J}_m} \check{s}_{km} \check{p}_{km} \left(\tilde{g}_{km} - \sum_{k' \in \mathcal{J}_m} \check{s}_{k'm} \tilde{g}_{k'm} \right) + O(\check{\sigma}^2) \\
&= 2\check{\sigma}_0 \check{\pi} \left[\left(\check{p}_{jm} - \bar{p}_m \right) \tilde{g}_{jm} - \check{p}_{jm} \sum_{k \in \mathcal{J}_m} \check{s}_{km} \tilde{g}_{km} \right] \\
&\quad - 2\check{\sigma}_0 \check{\alpha}\check{\pi} \check{p}_{jm} \sum_{k \in \mathcal{J}_m} \check{s}_{km} \left(\check{p}_{km} - \bar{p}_m \right) \tilde{g}_{km} + O(\check{\sigma}^2),
\end{aligned}$$

establishing the claims of the Proposition.

A.3 Proof of Proposition 1

Write $\beta_j(\boldsymbol{x}_m) = \mathbb{E}\left[\xi_m \mid \boldsymbol{x}_m \right]$, $\tilde{\xi}_{jm} = \xi_{jm} - \beta_j(\boldsymbol{x}_m)$, and $\tilde{\mathcal{D}}_j(\boldsymbol{s}_m, \boldsymbol{x}_m) = \mathcal{D}_j(\boldsymbol{s}_m, \boldsymbol{x}_m) - \beta_j(\boldsymbol{x}_m)$. Then we can rewrite equation (23) as

$$p_{jm} = \tilde{\mathcal{D}}_j(\boldsymbol{s}_m, \boldsymbol{x}_m) - \tilde{\xi}_{jm}$$

with

$$\mathbb{E}\left[\tilde{\xi}_{jm} \mid \boldsymbol{g}_m, \boldsymbol{x}_m \right] = \mathbb{E}\left[\xi_{jm} \mid \boldsymbol{g}_m, \boldsymbol{x}_m \right] - \beta_j(\boldsymbol{x}_m) = 0,$$

where the last equality uses Assumption 1. This is a standard non-parametric IV problem of Newey and Powell (2003). The completeness assumption implies that $\tilde{\mathcal{D}}_j(\boldsymbol{s}_m, \boldsymbol{x}_m)$ and $\tilde{\xi}_{jm}$ are identified, meaning that $\mathcal{D}_j(\boldsymbol{s}_m, \boldsymbol{x}_m)$ and ξ_{jm} are identified up to an additive term $\beta_j(\boldsymbol{x}_m)$.

Inverting $\boldsymbol{\mathcal{D}} = \left(\tilde{\mathcal{D}}_j \right)_{j \in \mathcal{J}_m}$ yields

$$\boldsymbol{s}_m = \tilde{\boldsymbol{\mathcal{D}}}^{-1} \left(\boldsymbol{p}_m + \boldsymbol{\xi}_m - \boldsymbol{\beta}(\boldsymbol{x}_m), \boldsymbol{x}_m \right).$$

Thus, cross-price elasticities are given by

$$\frac{d\boldsymbol{s}_m}{d\boldsymbol{p}'_m} = \frac{\partial \tilde{\boldsymbol{\mathcal{D}}}^{-1}}{\partial \boldsymbol{p}'_m} \left(\tilde{\boldsymbol{\mathcal{D}}}(\boldsymbol{s}_m, \boldsymbol{x}_m), \boldsymbol{x}_m \right),$$

where all terms are point-identified.

B Monte Carlo Simulation Details

B.1 Baseline Data-Generating Process

For each of 100 simulations, we generate a set of regions $r = 1, \ldots, 100$ in time periods $t = 1, 2$. Each market $m = (r, t)$ has $j = 1, \ldots, 15$ products and an outside good, $j = 0$. Each product has $L_1 = 2$ time-invariant characteristics with random coefficients and an intercept (with no random coefficient). The data-generating process is as follows:

- Random coefficients: $\eta_{i\ell} \overset{iid}{\sim} N(0, \sigma_\ell^2)$ for all $\ell = 1, \ldots, L_1$.
- Observed characteristics: $x_{jr\ell} \overset{iid}{\sim} N(0, 1)$ for all $\ell = 1, \ldots, L_1$; $x_{jr0} = 1$.
- Idiosyncratic preference shocks: $\varepsilon_{ijrt} \overset{iid}{\sim} T1EV(0, 1)$.
- Unobserved taste shifters: $\xi_{jr1} \overset{iid}{\sim} N(0, 1)$, $\xi_{jr2} = 0.9\xi_{jr1} + \sqrt{1 - 0.9^2} \cdot e_{jr2}$ with $e_{jr2} \overset{iid}{\sim} N(0, 1)$.
- Unobserved cost shifters: $\omega_{jr1} \overset{iid}{\sim} N(0, 1)$, $\omega_{jr2} = 0.9\omega_{jr1} + \sqrt{1 - 0.9^2} \cdot w_{jr2}$ with $w_{jr2} \overset{iid}{\sim} N(0, 1)$.
- Observed cost shocks: $g_{jr1} = 0$, $g_{jr2} \overset{iid}{\sim} N(0, 0.2^2)$.
- Marginal costs: $c_{jm} = \gamma' x_{jm} + \omega_{jm} + g_{jm}$.
- Prices: p_{jm} solve equation (25). To characterize this solution, first note that shares do not directly depend on prices here:

  $$\boldsymbol{S}(\boldsymbol{\delta}_m; \sigma, \boldsymbol{x}_m^{(1)}) = \int s_{jmi} d\mathcal{P}(\eta_i; \sigma)$$

  for

  $$s_{jmi} = \frac{\exp\left(\delta_{jm} + \sum_{\ell=1}^{L_1} \eta_{i\ell} x_{jm\ell}\right)}{1 + \sum_{k \in \mathcal{J}_m} \exp\left(\delta_{km} + \sum_{\ell=1}^{L_1} \eta_{i\ell} x_{km\ell}\right)}.$$

  Now write the derivative of $\boldsymbol{S}(\boldsymbol{\delta}_m; \sigma, \boldsymbol{x}_m^{(1)})$ in terms of own and cross price effects. Noting $\dfrac{d}{dp_{jm}} \delta_{jm} = \alpha$, we have

  $$\frac{d\boldsymbol{S}(\boldsymbol{\delta}_m; \sigma, \boldsymbol{x}_m^{(1)})}{d\boldsymbol{p}_m'} = \frac{d\boldsymbol{S}(\boldsymbol{\delta}_m; \sigma, \boldsymbol{x}_m^{(1)})}{d\boldsymbol{\delta}_m} \frac{d\boldsymbol{\delta}_m}{d\boldsymbol{p}_m} = \Lambda(\boldsymbol{\delta}_m; \sigma, \boldsymbol{x}_m^{(1)}) - \Gamma(\boldsymbol{\delta}_m; \sigma, \boldsymbol{x}_m^{(1)})$$

  where $\Lambda_{jk} = 0$ for $j \neq k$ and

  $$\Lambda_{jj}(\boldsymbol{\delta}_m; \sigma, \boldsymbol{x}_m^{(1)}) = \int \alpha s_{jmi} d\mathcal{P}(\eta_i; \sigma),$$
  $$\Gamma_{jk}(\boldsymbol{\delta}_m; \sigma, \boldsymbol{x}_m^{(1)}) = \int \alpha s_{jmi} s_{kmi} d\mathcal{P}(\eta_i; \sigma).$$

  Conlon and Gortmaker, 2020 note the fixed point of the following mapping gives the same solution:

  $$\boldsymbol{p}_m \hookleftarrow \boldsymbol{c}_m + \Lambda(\boldsymbol{\delta}_m; \sigma, \boldsymbol{x}_m^{(1)})\Gamma(\boldsymbol{\delta}_m; \sigma, \boldsymbol{x}_m^{(1)})(\boldsymbol{p}_m - \boldsymbol{c}_m) - \Lambda(\boldsymbol{\delta}_m; \sigma, \boldsymbol{x}_m^{(1)})^{-1} \boldsymbol{S}(\boldsymbol{\delta}_m; \sigma, \boldsymbol{x}_m^{(1)}).$$

We find the fixed point of this mapping for each market m and check that it satisfies Equation (25).[33]

- Parameters: $\sigma_\ell = 4$ for all $\ell = 1, \ldots, L_1$, $\alpha = -0.2 - 4\exp(0.5)$, $\beta_0 = 35, \beta_1 = \beta_2 = 2, \gamma_0 = 5, \gamma_1 = \gamma_2 = 1$.

We compute market shares of each product using equation (24) based on 1,000 draws of the random coefficient vectors that are the same across markets.

B.2 Computing and Inverting Market Shares

We approximate $\mathcal{S}_j(\boldsymbol{\delta}_{rt}; \sigma, \boldsymbol{x}_r^{(1)})$ using 250 elements of a 2-dimensional Halton sequence $\widetilde{h}_i = (\widetilde{h}_i^1, \widetilde{h}_i^2)$. Let $(h_i^1, h_i^2) = (\Phi^{-1}(\widetilde{h}_i^1), \Phi^{-1}(\widetilde{h}_i^2))$, with Φ^{-1} denoting the inverse of the standard normal CDF, and

$$\mathcal{S}_j\left(\boldsymbol{\delta}_{rt}; \sigma, \boldsymbol{x}_r^{(1)}\right) = \frac{1}{250}\sum_{i=1}^{250} s_{rtji}, \quad s_{rtji} = \frac{\exp\left(\delta_{jr2} + \sum_{\ell=1}^{2} \sigma_\ell h_i^\ell x_{jr}\right)}{1 + \sum_{k \in \mathcal{J}_r} \exp\left(\delta_{kr2} + \sum_{\ell=1}^{2} \sigma_\ell h_i^\ell x_{kr}^\ell\right)}.$$

To generate $\widetilde{h}_i^\ell$, we use the reverse-radix scrambling algorithm in Kocis and Whiten (1997) and skip the first 1,000 draws. Various derivatives of $\mathcal{S}_j$ that we will use below (e.g., $\partial \mathcal{S}_j / \partial \sigma_\ell$) are approximated in the same way, with s_{rtji} replaced by its derivatives.

We compute $\delta_{jrt} = \mathcal{D}_j(\boldsymbol{s}_{rt}; \sigma, \boldsymbol{x}_r^{(1)})$ as the fixed point of the contraction mapping

$$\delta_{jrt}^{(\iota+1)} \leftarrow \delta_{jrt}^{(\iota)} + \log s_{jrt} - \log \mathcal{S}_j(\boldsymbol{\delta}_{rt}^{(\iota)}; \sigma, \boldsymbol{x}_r^{(1)}),$$

with the starting point $\delta_{jrt}^{(0)} = \log(s_{jrt}/s_{0rt})$. To compute the fixed point, we use the "SQUAREM" method of Varadhan and Roland (2008) with a maximum of 10,000 iterations and a tolerance of $\epsilon^{5/6}$, where ϵ is the double-precision machine epsilon.

B.3 Computing Instruments

We focus here on the recentered IVs since the characteristic-based IVs are given in Section 4. To implement the recentered shift-share instrument from equation (18), we first note that $\mathcal{S}_j(\cdot)$ and $\mathcal{D}_j(\cdot)$ do not explicitly depend on price in this simulation. We further assume the researcher knows $\mathbb{E}\left[g_{jm} \mid \boldsymbol{x}_m, \boldsymbol{q}_m\right] = 0$, such that $\tilde{g}_{jm} = g_{jm}$. We obtain the pass-through coefficient $\breve{\pi}$ from a simple linear regression of Δp_{jr} on $\widetilde{g}_{jr2}$, and leave the discussion of other preliminary parameter values $(\breve{\alpha}, \breve{\sigma})$ to below. Thus we have for $\ell = 1, \ldots, L_1$:

$$Z_{jr2\ell}^{\text{SSIV}} = \breve{\alpha}\breve{\pi} \sum_{k, k' \in \mathcal{J}_r} \frac{\partial^2}{\partial s_{k'r} \partial \sigma_\ell} \mathcal{D}_j\left(\boldsymbol{s}_{r1}; \breve{\sigma}, \boldsymbol{x}_r^{(1)}\right) \cdot \frac{\partial}{\partial \delta_{kr}} \mathcal{S}_{k'}(\boldsymbol{\delta}_{r1}; \breve{\sigma}, \boldsymbol{x}_r^{(1)}) \cdot g_{kr2}. \tag{A12}$$

To compute the derivatives, we first apply the implicit function theorem to equation (8) and

[33] If it does not then we solve for a fixed point of equation (25) directly. In a small number of cases (less than 0.5 markets per simulation), we do not find a solution and drop the market from the sample.

find

$$\partial_{s'_{r1}} \mathcal{D}\left(\boldsymbol{s}_{r1}; \check{\sigma}, \boldsymbol{x}_r^{(1)}\right) = \left[\partial_{\delta'} \mathcal{S}(\boldsymbol{\delta}_{r1}; \check{\sigma}, \boldsymbol{x}_r^{(1)})\right]^{-1},$$
$$\partial_{\sigma_\ell} \mathcal{D}\left(\boldsymbol{s}_{r1}; \check{\sigma}, \boldsymbol{x}_r^{(1)}\right) = -\left[\partial_{\delta'} \mathcal{S}(\boldsymbol{\delta}_{r1}; \check{\sigma}, \boldsymbol{x}_r^{(1)})\right]^{-1} \partial_{\sigma_\ell} \mathcal{S}(\boldsymbol{\delta}_{r1}; \check{\sigma}, \boldsymbol{x}_r^{(1)})$$

for $\boldsymbol{\delta}_{r1} = \mathcal{D}\left(\boldsymbol{s}_{r1}; \check{\sigma}, \boldsymbol{x}_r^{(1)}\right)$. Next, we differentiate $\partial_{s'_{r1}} \mathcal{D}\left(\boldsymbol{s}_{r1}; \check{\sigma}, \boldsymbol{x}_r^{(1)}\right)$ with respect to σ_ℓ to obtain

$$\partial_{\sigma_\ell} \partial_{s'_{r1}} \mathcal{D}\left(\boldsymbol{s}_{r1}; \check{\sigma}, \boldsymbol{x}_r^{(1)}\right) = -\left[\partial_{\delta'} \mathcal{S}(\boldsymbol{\delta}_{r1}; \check{\sigma}, \boldsymbol{x}_r^{(1)})\right]^{-1} \left[\frac{d}{d\sigma_\ell} \partial_{\delta'} \mathcal{S}\left(\mathcal{D}(\boldsymbol{s}_{r1}; \check{\sigma}, \boldsymbol{x}_r^{(1)}); \check{\sigma}, \boldsymbol{x}_r^{(1)}\right)\right] \left[\partial_{\delta'} \mathcal{S}(\boldsymbol{\delta}_{r1}; \check{\sigma}, \boldsymbol{x}_r^{(1)})\right]^{-1}$$

where the total derivative with respect to σ_ℓ is:

$$\frac{d}{d\sigma_\ell} \partial_{\delta'} \mathcal{S}\left(\mathcal{D}(\boldsymbol{s}_{r1}; \check{\sigma}, \boldsymbol{x}_r^{(1)}); \check{\sigma}, \boldsymbol{x}_r^{(1)}\right) = \sum_{k \in \mathcal{J}_r} \partial_{\delta_k} \partial_{\delta'} \mathcal{S}(\boldsymbol{\delta}_{r1}; \check{\sigma}, \boldsymbol{x}_r^{(1)}) \cdot \partial_{\sigma_\ell} \mathcal{D}_k\left(\boldsymbol{s}_{r1}; \check{\sigma}, \boldsymbol{x}_r^{(1)}\right) + \partial_{\sigma_\ell} \partial_{\delta'} \mathcal{S}(\boldsymbol{\delta}_{r1}; \check{\sigma}, \boldsymbol{x}_r^{(1)}).$$

The recentered exact formula instrument is given by

$$Z_{jr2\ell}^{FIV} = \frac{\partial}{\partial \sigma_\ell} \mathcal{D}_j\left(\mathcal{S}(\boldsymbol{\delta}_{r1} + \check{\alpha}\check{\pi}\boldsymbol{g}_{r2}; \check{\sigma}, \boldsymbol{x}_r^{(1)}); \check{\sigma}, \boldsymbol{x}_r^{(1)}\right) - \frac{1}{20} \sum_{c=1}^{20} \frac{\partial}{\partial \sigma_\ell} \mathcal{D}_j\left(\mathcal{S}(\boldsymbol{\delta}_{r1} + \check{\alpha}\check{\pi}\boldsymbol{g}_{r2}^{(c)}; \check{\sigma}, \boldsymbol{x}_r^{(1)}); \check{\sigma}, \boldsymbol{x}_r^{(1)}\right), \tag{A13}$$

where $\boldsymbol{g}^{(c)}$ is obtained by randomly permuting the shocks $\boldsymbol{g}$ across products and markets.

B.4 Estimation Procedure

Given any moment condition $\mathbb{E}[h(\varphi)] = 0$ with the sample analog $h_N(\varphi) = \frac{1}{N} \sum_{j,r} h_{jr}(\varphi)$ where N is the number of product-market pairs, we estimate φ via non-linear GMM as

$$\widehat{\varphi} = \arg\min_{\varphi} Q_N(\varphi) = \frac{1}{2} h_N(\varphi)' W_N h_N(\varphi), \tag{A14}$$

where N is the number of product-market pairs and W_N is a positive-definite weighting matrix. Appendix B.4.1 details the estimation using characteristic-based IVs; Appendices B.4.2 and B.4.3 detail the estimation using recentered IVs.

B.4.1 Estimation with Characteristic-Based IVs

Let Z_{jr2} be a vector collecting g_{jr2}, x_{jr}, and two characteristic-based IVs for $\sigma = (\sigma_1, \sigma_2)$: either BLP (sum of competitor characteristics) IVs or the Gandhi and Houde (2020) local or quadratic differentiation IVs. We use the second time period to estimate the parameters $\varphi = (\sigma, \alpha, \beta)$, with

$$h_{jr}(\varphi) = \xi_{jr2} \cdot Z_{jr2}$$

for $\xi_{jr2} = \mathcal{D}_j(\boldsymbol{s}_{r2}; \sigma, \boldsymbol{x}_r^{(1)}) - \alpha p_{jr2} - \beta' x_{jr}$ and $\mathcal{D}_j(\boldsymbol{s}_{rt}; \sigma, \boldsymbol{x}_r^{(1)})$ from Appendix B.2.

We rewrite the problem (A14) as a numerical optimization over σ, concentrating out α, β. To do so, note that the minimization over α, β is a linear IV-GMM problem with the solution

$$(\alpha(\sigma), \beta(\sigma)')' = \left(X' Z W_N Z' X\right)^{-1} X' Z W_N Z' \mathcal{D}(\sigma), \tag{A15}$$

where X collects (p_{jr2}, x'_{jr}), Z collects $\left(Z'_{jr2}\right)$, and $\mathcal{D}(\sigma)$ collects $\mathcal{D}_j(\boldsymbol{s}_{r2}; \sigma, \boldsymbol{x}_r^{(1)})$ across jr pairs.

To solve $\min_\sigma Q_N(\sigma) \equiv Q_N(\sigma, \alpha(\sigma), \beta(\sigma))$, we use the following procedure:

1. Take 50 points for σ from $[0, 10]^2$; we use the deterministic "R_d" sequence in Halchenko et al. (2020). Choose $\sigma^{(0)}$ as the point that minimizes $Q_N(\sigma)$.

2. Optimize over σ via the Gauss-Newton regression algorithm, following the suggestion in Gandhi and Houde (2020):[34]

   - At each iteration, set $\sigma^{(\iota+1)} = \sigma^{(\iota)} + b^{(\iota)}$, with

   $$b^{(\iota)} = -\left(H_N(\sigma^{(\iota)})' W_N H_N(\sigma^{(\iota)})\right)^{-1} H_N(\sigma^{(\iota)})' W_N h_N(\sigma^{(\iota)})$$

   where

   $$h_N(\sigma) \equiv h_N(\sigma, \alpha(\sigma), \beta(\sigma)) = \frac{1}{N}\mathcal{D}(\sigma)' \left(Z - ZW_N Z'X(X'ZW_N Z'X)^{-1}X'Z\right)$$

   and

   $$H_N(\sigma)' \equiv \partial_\sigma h_N(\sigma)' = \frac{1}{N}\left(\partial_\sigma \mathcal{D}(\sigma)'\right)\left(Z - ZW_N Z'X(X'ZW_N Z'X)^{-1}X'Z\right).$$

   - Continue updating $\sigma^{(\iota)}$ while $\|b^{(\iota)}\| > \epsilon^{1/2}$ and $\iota < 100$.

3. If $\iota \geq 100$ or $\left\|\frac{d}{d\sigma}Q_N(\widehat{\sigma})\right\| > \epsilon^{1/3}$, discard step 2 and minimize $Q_N(\sigma)$ via a BFGS-based algorithm (as suggested in Conlon and Gortmaker, 2020) with the same starting parameter $\sigma^{(0)}$. We use MATLAB's implementation via `fmincon` with a lower bound of 0 for σ.

We note that while the system of moment conditions is just-identified and thus the choice of the positive definite matrix W_N should be irrelevant, in practice we set $W_N = \left(\frac{1}{N}\sum_{j,r} Z_{jr2}Z'_{jr2}\right)^{-1}$.

B.4.2 Continuously Updating Estimation with Recentered IVs

Let Z_{jr2} be a vector collecting g_{jr2} and either the shift-share or exact prediction IVs in Section 3. We use the sample analog of the moment condition in first-differences, with

$$h_{jr}(\theta) = \Delta\xi_{jr} \cdot Z_{jr2},$$

where $\Delta\xi_{jr} = \Delta\mathcal{D}_j(\boldsymbol{s}_{rt}; \sigma, \boldsymbol{x}_r^{(1)}) - \alpha\Delta p_{jr}$, and $\mathcal{D}_j(\boldsymbol{s}_{rt}; \sigma, \boldsymbol{x}_r^{(1)})$ is computed as in Appendix B.2. Note that we do not estimate β in this analysis. To estimate θ we follow steps similar to Appendix B.4.1:

1. Take 50 points from $[0, 10]^2$ in the same way. Define $\alpha(\sigma)$ as the regression slope of $\Delta\mathcal{D}_j(\boldsymbol{s}_{rt}; \sigma, \boldsymbol{x}_r^{(1)})$ on Δp_{jr} (and an intercept) instrumented with g_{jr2}. Choose $\sigma^{(0)}$ that minimizes $Q_N(\sigma, \alpha(\sigma))$

[34]To be more precise, we force our σ to be strictly positive by using the softplus transformation $\sigma \equiv \log(1+\exp(\tilde{\sigma}))$ and conducting the search over $\tilde{\sigma}$ (adjusting the derivatives by $\exp(\tilde{\sigma})/(1+\exp(\tilde{\sigma}))$ for the change of variables). Also, note that the Gauss-Newton regression can be derived using the first-order approximation $h_N(\sigma) \approx h_N(\sigma^{(\iota)}) + H_N(\sigma^{(\iota)})(\sigma - \sigma^{(\iota)})$; the algorithm solves for the step size $b^{(\iota)} \equiv \sigma - \sigma^{(\iota)}$ via linear GMM. Isolating b from the first-order condition $0 = \partial_b(h_N(\sigma^{(\iota)}) + H_N(\sigma^{(\iota)})b)' W_N (h_N(\sigma^{(\iota)}) + H_N(\sigma^{(\iota)})b)$ gives the expression.

among the 50 points while recomputing Z_{jr2} at each $(\sigma, \alpha(\sigma))$ via equation (A12) or equation (A13). Set $\alpha^{(0)} = \alpha(\sigma^{(0)})$.[35]

2. Estimate θ using Gauss-Newton regression, searching over both σ and α:[36]

   - At each step, set $\theta^{(\iota+1)} = \theta^{(\iota)} + b^{(\iota)}$, with

   $$b^{(\iota)} = -\left(H_N(\theta^{(\iota)})' W_N H(\theta^{(\iota)})\right)^{-1} H_N(\theta^{(\iota)})' W_N h_N(\theta^{(\iota)})$$

     and $H_N(\theta)$ computed numerically using finite-differences (to account for the fact the instrument also changes with the parameters).

   - Continue updating $\theta^{(\iota)}$ while $\|b^{(\iota)}\| > \epsilon^{1/2}$ and $\iota < 100$.

3. If $\iota \geq 100$ or $\|\partial_\theta Q_N(\widehat{\theta})\| > \epsilon^{1/3}$, discard step 2 and minimize $Q_N(\theta)$ using a BFGS-based algorithm with the same starting parameter $\theta^{(0)}$.

Like in Appendix B.4.1, the system of moment conditions here is just-identified; we set $W_N = I$ so that estimation will not require recomputing the weighting matrix at each iteration.

B.4.3 Iterative Estimation with Recentered IVs

Iterative estimation is similar to Appendix B.4.2, but the instrument at each step depends on the parameter estimates from the previous step. This allows us to concentrate out α and use analytical derivatives. We use the following procedure:

1. Choose $\sigma^{(0)}$ and $\alpha^{(0)}$ as in Step 1 in Appendix B.4.2.

2. Compute $Z^{(\iota)}_{jr2}$ via equation (A12) or equation (A13) using $\widehat{\theta}^{(\iota)} = (\widehat{\sigma}^{(\iota)}, \widehat{\alpha}^{(\iota)})$.

3. Obtain estimates $\widehat{\theta}^{(\iota+1)}$ as follows:

   - Concentrate out α as

   $$\alpha^{(\iota)}(\sigma) = \left(X' Z^{(\iota)} W_N Z^{(\iota)\prime} X\right)^{-1} X' Z^{(\iota)} W_N Z^{(\iota)\prime} \Delta\mathcal{D}\left(\sigma\right),$$

     where X collects $\Delta p_{jr} - \overline{\Delta p}$, $Z^{(\iota)}$ collects $Z^{(\iota)}_{jr2} - \overline{Z^{(\iota)}}$, and $\Delta\mathcal{D}(\sigma)$ collects $\Delta\mathcal{D}_j(\boldsymbol{s}_{rt}; \sigma, \boldsymbol{x}^{(1)}_r) - \overline{\Delta\mathcal{D}(\sigma)}$ across jr pairs (with bars denoting sample averages). Note that concentrating α out is possible because the instruments $Z^{(\iota)}_{jr2}$ are fixed based on $\widehat{\theta}^{(\iota)}$.

   - Starting from $\hat{\sigma}^{(\iota)}$, search over σ to minimize $Q^{(\iota)}_N(\sigma, \alpha^{(\iota)}(\sigma))$, which depends on ι via the instruments $Z^{(\iota)}_{jr2}$. Use Gauss-Newton regression iterations as in Step 2 of Appendix B.4.1. If failed, discard the estimates and use a BFGS-based algorithm as in Step 3 of of Appendix B.4.1. Save the result as $\hat{\sigma}^{(\iota+1)}$.

[35]Note that, unlike Appendix B.4.1, $\alpha(\sigma) \neq \arg\min_\alpha Q_N(\sigma, \alpha)$. Here the instrument depends on α, such that it is not possible to concentrate out α. Rather, $\alpha(\sigma)$ is a reasonable starting point that sets one of the moments to zero: $\mathbb{E}\left[\Delta\xi_{jr} \cdot g_{jr2}\right] = 0$. We do not use $\alpha(\sigma)$ in Step 2.

[36]Like in Appendix B.4.1, we use the softplus transformation for σ. In addition, we force α to be negative by setting $\alpha \equiv -\log(1 + \exp(\widetilde{\alpha}))$ and searching over $\tilde{\alpha}$ (adjusting the derivatives accordingly).

- Set $\hat{\alpha}^{(\iota+1)} = \alpha^{(\iota)}\left(\hat{\sigma}^{(\iota+1)}\right)$.

4. Continue updating $\widehat{\theta}^{(\iota)}$ while $\left\|\widehat{\sigma}^{(\iota+1)} - \widehat{\sigma}^{(\iota)}\right\| > \epsilon^{1/2}$ and $\iota < 100$.

As in Appendix B.4.2, we set $W_N = I$.