Fast, Detail-free, and Approximately Correct: Estimating Mixed Demand Systems*

Bernard Salanié[†] Frank A. Wolak[‡]

April 8, 2022

Abstract

Many econometric models used in applied work integrate over unobserved heterogeneity. We show that a class of these models that includes many random coefficients demand systems can be approximated by a "small-σ" expansion that yields a linear two-stage least squares estimator. While our estimator is only approximately correct, it is extremely fast and easy to implement. It is also detail-free: its implementation does not rely on the higher moments of the distribution of the random coefficients. We test our approach on the models of product shares and prices popular in empirical IO, with or without micromoments and with or without specifiying supply. Monte Carlo simulations suggest that our approximate estimator performs surprisingly well: its asymptotic bias is usually small, and it works well in finite samples. A simple Newton-Raphson correction further improves the estimates at minimal cost.

*We are grateful to Dan Ackerberg, Roy Allen, John Asker, Steve Berry, James Brand, Xiaohong Chen, Pierre Dubois, Jeremy Fox, Han Hong, Guy Laroque, Simon Lee, Arthur Lewbel, Thierry Magnac, Lars Nesheim, Ariel Pakes, Mathias Reynaert, Tobias Salz, Richard Smith, Pedro Souza, Frank Verboven, Martin Weidner, Ken Wolpin, and especially to Chris Conlon for their useful comments, as well as to seminar audiences at NYU, Rice, UCL, and the Stanford Institute for Theoretical Economics (SITE). We also thank Zeyu Wang for excellent research assistance.

[†]Department of Economics, Columbia University, 1022 International Affairs Building, 420 West 118th Street, New York, NY 10027, bsalanie@columbia.edu.

[‡]Department of Economics and Program on Energy and Sustainable Development, Stanford University, Stanford CA 94305-6072, wolak@zia.stanford.edu.

Introduction

Many econometric models are estimated from conditional moment conditions that express the mean independence of random unobservable terms η and instruments Z:

$$E(\eta|Z) = 0.$$

In structural models, the unobservable term is usually obtained by solving a set of equations—often a set of first-order conditions—that define the observed endogenous variables as functions of the observed exogenous variables and unobservables. That is, we start from

$$G(Y, \eta, \theta_0) = 0 \tag{1}$$

where Y is the vector of all observed random variables and θ_0 is the true value of the vector of unknown parameters. The parametric function G is assumed to be known and can depend on a vector of observed exogenous variables. If the solution exists and is unique, we invert this system into

$$\eta = F(Y, \theta_0)$$

and we seek an estimator of θ_0 by minimizing an empirical analog of a norm

$$\|E(F(Y, \theta)m(Z))\|$$

where $m(Z)$ is a vector of measurable functions of Z. We will assume throughout that the moment conditions identify θ.

Unless $F(Y, \theta)$ exists in closed form, inversion often is a step fraught with difficulties. Even when a simple inversion algorithm exists, it is still costly and must be done with a high degree of numerical precision, as errors may jeopardize the "outer" minimization problem. One alternative is to minimize an empirical analog of the norm

$$\|E(\eta\, m(Z))\|$$

subject to the structural constraints (1). This "MPEC approach" has met with some success in dynamic programming and empirical industrial organization (Su and Judd 2012, Dubé et al 2012). It still requires solving a nonlinearly constrained, nonlinear objective function minimization problem; convergence to a solution can be a challenging task in the absence of very good initial values. This is especially galling when the model has be estimated many times, as with Nash-in-Nash models.

We propose an alternative method that derives a linear model from a very simple series expansion. To fix ideas, suppose that θ can be decomposed into a pair (β, σ), where σ is a scalar whose true value is likely to be small. We rewrite (1) as

$$G(Y, F(Y, \beta_0, \sigma_0), \beta_0, \sigma_0) = 0.$$

Expanding $\sigma \to F(Y, \beta_0, \sigma)$ in a Taylor series at $\sigma = 0$ suggests a family of "approximate estimators" that minimize the empirical analogs of the norms:

$$\left\| E\left(\left(F(Y, \beta, 0) + \ldots + F_{\sigma\ldots\sigma}(Y, \beta, 0) \frac{\sigma^q}{q!} \right) m(Z) \right) \right\| \tag{2}$$

If the true value σ_0 is not too large, one may hope to obtain a satisfactory estimator for a small value of q. In general, this still requires solving a nonlinear minimization problem when $q > 0$. For $q = 1$, the first-order conditions of the problem are the usual normal equations. However, as we will see $F_\sigma(Y, \beta, 0) = 0$ in many interesting cases, so that we must go at least to the $q = 2$ expansion to identify σ.

The resulting estimators of β_0 and Σ_0 are only approximately correct, in the sense that they consistently estimate an approximation of the original model. On the other hand, they can be estimated very simply and fast by two-stage least-squares. As this is a linear problem, the optimal[1] instruments associated with the second-order conditional moment restrictions can be estimated directly from the data using nonparametric regressions[2]. Moreover, since our approximate estimators only rely on limited features of the data generating process, they are "detail-free" in ways that we will explore later.

As we will show, under very weak conditions the Berry, Levinsohn, and Pakes (1995) model (hereafter "macro-BLP") that is the workhorse of empirical IO belongs to the QLRC family. So do count models with unobserved heterogeneity, which are often used in insurance applications for instance[3] Moreover, our method may

[1]In the sense of Amemiya (1975).

[2]Alternatively, we can include flexible functions of the columns of Z in the instruments used to compute the 2SLS estimates.

[3]See Section 1.1 and Appendix B.2.

remain useful beyond this class of quasi-linear models, at the cost of requiring (simple) numerical optimization[4]. Another attractive feature of FRAC in this context is that it scales up nicely. Brand (2021a) has used the FRAC estimator to allow the distribution of the price sensitivity of retail consumers to vary at the three-digit ZIP code level; this would be completely infeasible with the standard GMM approach[5].

To test our method, we run two Monte-Carlo simulations on a macro-BLP model. In Section 5, we show that the asymptotic bias inherent in our method is usually small—and certainly much smaller than the sampling variation in many applications. Section 6 turns to a finite sample simulation modeled after Dubé et al (2012). We find that in regard to the mean values of the random coefficients, our estimation procedure performs as well as their recommended MPEC estimator across all parameter configurations considered. For a number of parameter configurations, our bias-corrected estimator in fact produces superior estimates of variance of the random coefficients relative to the MPEC estimator. We also demonstrate the usefulness of our procedure in testing whether the coefficient of a product characteristic is non-random, or whether it is zero. For the parameter configurations that we consider, the finite-sample size of each of these tests is well-approximated by the asymptotic size; and they are powerful enough to detect economically meaningful deviations from the null hypothesis with a high probability.

Our approach builds on "small-σ" approximations to the mapping F. Kadane (1971) pioneered the "small-σ" method. He applied it to a linear, normal simultaneous equation system and studied the properties of k-class estimators[6] when the number of observations n is fixed and σ goes to zero. He showed that when the number of observations is large, under these "small-σ asymptotics" the k-class estimators have biases in σ^2, and that their mean-squared errors differ by terms of order σ^4. Kadane argued that small σ, fixed n asymptotics are often a good approximation to finite-sample distributions when the estimation sample is large enough.

The small-σ approach was used by Chesher (1991) in models with measurement error. Most directly related to us, Chesher and Santos-Silva (2002) used a second-order approximation argument to reduce a mixed multinomial logit model to a "heterogene-

[4] We will illustrate this on a mixed nested logit in Appendix B.1.

[5] He has also published a Julia implementation of FRAC (Brand 2021b).

[6] Which include OLS and 2SLS.

ity adjusted" unmixed multinomial logit model in which mean utilities have additional terms[7]. They suggested estimating the unmixed logit and using a score statistic based on these additional covariates to test for the null of no random variation in preferences. Like them, we introduce additional covariates. Unlike them, we develop a method to estimate jointly the mean preference coefficients and parameters characterizing their random variation; and we only use linear instrumental variables estimators. To some degree, our method is also related to that of Harding and Hausman 2007, who use a Laplace approximation of the integral over the random coefficients in a mixed logit model without choice-specific random effects. Unlike them, we allow for endogeneous prices; our approach is also much simpler to implement.

An alternative approach developed by Lu, Shi and Tao (2021) applies semi-nonparametric techniques to the macro-BLP model in order to estimate the dependence of market shares on the covariates whose coefficients are random. Our estimator can be seen as a second-order truncation of theirs. Lu et al's estimator, unlike ours, is consistent as the number of products goes to infinity. Like ours, it does not require specifying the distribution of random coefficients. Our two-stage least-squares technique is of course simpler to implement than their partially linear model.

Section 1 introduces the class of random coefficient models to which our method applies. Section 2 presents the model popularized by Berry-Levinsohn-Pakes (1995) and discusses some of the difficulties that practitioners have encountered when taking it to data. We give a detailed description of our algorithm in Section 3. Readers not interested in the derivation of our formulæ can jump directly to our Monte Carlo simulations in Sections 5 and 6. Section 4 of the paper derives and discusses the properties of our method; it also proposes a simple corrected estimator. The proofs of some of our results are in Appendices, along with a more focused discussion of the mixed binary choice model (Appendix A) and extensions to a nested logit macro-BLP model (Appendix B.1) and to a count data model with unobserved heterogeneity (Appendix B.2).

[7] Ketz (2018) builds on a quadratic expansion in $\sigma_0 = 0$ to derive asymptotic distributions when the true σ_0 is on the boundary.

1 Quasi-linear Random Coefficients Models

Our method applies to random coefficient models that have a specific quasi-linear structure. Their defining characteristic is that the error term $\boldsymbol{\eta}$ and the mean coefficients $\boldsymbol{\beta}$ only enter the reduced form of (1) via a linear combination $\boldsymbol{\eta} + \boldsymbol{f}_1(\boldsymbol{Y})\boldsymbol{\beta}$:

$$\boldsymbol{G}(\boldsymbol{Y}, \boldsymbol{\eta}, \boldsymbol{\beta}, \boldsymbol{\Sigma}) \equiv \boldsymbol{G}^* \left(\boldsymbol{Y}, E_{\boldsymbol{\varepsilon}} \boldsymbol{A}^*(\boldsymbol{Y}, \boldsymbol{\eta} + \boldsymbol{f}_1(\boldsymbol{Y})\boldsymbol{\beta}, \boldsymbol{\varepsilon})\right). \tag{3}$$

The unobserved random vector $\boldsymbol{\varepsilon}$ is distributed independently of $\boldsymbol{Y}$ and $\boldsymbol{\eta}$. It is location-normalized by $E_{\boldsymbol{\varepsilon}}\boldsymbol{\varepsilon} = \boldsymbol{0}$ and it has a finite variance-covariance matrix $\boldsymbol{\Sigma}$. We will denote J the number of equations in the reduced form (the dimension of $\boldsymbol{G}$). We assume that the dimensions of $\boldsymbol{A}^*$ and $\boldsymbol{\eta}$ also equal J.

For simplicity, we will assume that $\boldsymbol{\Sigma}$ is positive definite; we will explain later how to incorporate linear constraints on its elements. In this specification, the functions $\boldsymbol{f}_1$, $\boldsymbol{G}^*$ and $\boldsymbol{A}^*$ are assumed to be known; our goal is to get approximate estimates of $\boldsymbol{\theta}_0 = (\boldsymbol{\beta}_0, \boldsymbol{\Sigma}_0)$. We assume that instruments $\boldsymbol{Z}$ are available and that the unknown parameters $\boldsymbol{\theta}_0$ are identified by the moment conditions $E_0\left(\boldsymbol{\eta}|\boldsymbol{Z}\right) = \boldsymbol{0}$.

1.1 Count Data

Our leading example in this paper will be the "macro-BLP" model of modern empirical industrial organization; Section 2 will describe it and show that it is indeed a QLRC model. Here we discuss another interesting QLRC instance—count data with unobserved heterogeneity.

The most basic data-generating process for count data is the Poisson model: for a subpopulation with observed characteristics X and unobserved η, we define $\lambda = \eta + X\beta$ and the average number K of events in this subpopulation follows a Poisson $\mathcal{P}(\lambda)$:

$$\Pr(K = k|\boldsymbol{X}) = p_k(\lambda(\boldsymbol{X})),$$

where $p_k(\lambda) \equiv \frac{\lambda^k \exp(-\lambda)}{k!}$.

This has well-known problems fitting individual count data. A common solution is to add unobserved heterogeneity at the individual level, e.g.

$$\lambda(\boldsymbol{X}_i, \epsilon_i) = \boldsymbol{X}_i\boldsymbol{\beta} + \epsilon_i$$

so that

$$\Pr(K=k|\boldsymbol{X}) = E_\epsilon p_k\left(\lambda(\boldsymbol{X},\epsilon)\right).$$

A very popular choice has u follow a Gamma distribution, independently of $\boldsymbol{X}$, in which case the count variable K has a negative binomial distribution. Let us go beyond this functional form and write

$$\Pr(K=k|\boldsymbol{X}) = E_{\boldsymbol{\varepsilon}} q_k(\eta_k + \boldsymbol{X}_k\boldsymbol{\beta}, \boldsymbol{\varepsilon}) \tag{4}$$

where the q_k are known non-negative functions, the η_k are unknown fixed effects, and the unobserved random vector $\boldsymbol{\varepsilon}$ has an unknown distribution with mean $\mathbf{0}$ and unknown variance-covariance matrix $\boldsymbol{\Sigma}$. Suppose we have consistent estimators $\hat{y}_k(X)$ of the left-hand side of (4) for J values $\{k_1,\ldots,k_J\}$, and assume that $E(\eta_k|\boldsymbol{Z})=0$ for each of these values. This describes a QLRC model with $\boldsymbol{Y}=(\hat{y}_{k_1}(X),\ldots,\hat{y}_{k_J}(X))$; $\boldsymbol{f}_1(\boldsymbol{Y})=\boldsymbol{X}$; $\boldsymbol{A}^*_j(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c})=q_{k_j}(\boldsymbol{b},\boldsymbol{c})$; and $\boldsymbol{G}^*_j(\boldsymbol{a},\boldsymbol{b})=\boldsymbol{a}_j-\boldsymbol{b}_j$.

1.2 Approximating QLRC Models

The quasi-linear structure of (3) yields straightforward expansions in this class of models. Remember that we define $\boldsymbol{F}$ as the inverse of $\boldsymbol{G}$ in the $\boldsymbol{\eta}$ dimension. Without further restrictions, $\boldsymbol{F}$ may well be empty- or many-valued, and it may not have the degree of smoothness that we need. We will focus on *regular* QLRC models, which combine invertibility, smoothness, and distributions of random coefficients with enough moments.

Definition 1 (Regular QLRC Models)**.** *A QLRC model is regular if and only if:*

1. *All moments of order 4 or less of $\boldsymbol{\varepsilon}$ are finite*

2. *$\boldsymbol{G}^*$ is twice differentiable with respect to its second argument*

3. *$\boldsymbol{A}^*$ is twice differentiable with respect to its last two arguments*

4. *the matrices*

$$\boldsymbol{G}^*_2 \equiv \left[\frac{\partial G^*_j}{\partial A^*_k}\left(\boldsymbol{Y}, \boldsymbol{A}^*(\boldsymbol{Y}, \boldsymbol{\eta}+\boldsymbol{f}_1(\boldsymbol{Y})\boldsymbol{\beta}, \mathbf{0})\right)\right]_{j,k=1,\ldots,J}$$

and

$$\boldsymbol{A}_2^* \equiv \left[\frac{\partial A_j^*}{\partial \eta_k}(\boldsymbol{Y}, \boldsymbol{\eta} + \boldsymbol{f}_1(\boldsymbol{Y})\boldsymbol{\beta}, \boldsymbol{0})\right]_{j,k=1,\dots,J}$$

are invertible for all $(\boldsymbol{Y}, \boldsymbol{\eta}, \boldsymbol{\beta})$.

As we will see in Section 2, macro-BLP models satisfy parts 2, 3, and 4 of Definition 1; so does the count data model with heterogeneity of Section 1.1 if the q_k functions are twice differentiable and invertible in their first argument. Note that part 1 of Definition 1 encompasses any distribution of random coefficients $\tilde{\boldsymbol{\beta}} = \boldsymbol{\beta} + \boldsymbol{\varepsilon}$ whose first four moments are finite. Any heteroskedasticity of $\boldsymbol{\varepsilon}$ can easily be accommodated via the first argument of the function $\boldsymbol{A}^*$.

We now state our main theorem, using the notation $\boldsymbol{h}_j$ to denote the j-th partial derivative of a function $\boldsymbol{h}$, and $\boldsymbol{h}_{jk}$ to be its second derivative in the j-th and k-th arguments.

Theorem 1 (Expansions for regular quasi-linear random coefficients models)**.** *Any regular QLRC model admits an inverse whose second-order expansion is*

$$F(\boldsymbol{Y}, \boldsymbol{\beta}, \boldsymbol{\Sigma}) \simeq \boldsymbol{f}_0(\boldsymbol{Y}) - \boldsymbol{f}_1(\boldsymbol{Y})\boldsymbol{\beta} - \boldsymbol{f}_2(\boldsymbol{Y})\boldsymbol{\Sigma} \tag{5}$$

where

- *the* J *variables* $\boldsymbol{f}_0(\boldsymbol{Y})$ *are uniquely defined by the system of equations*

$$\boldsymbol{G}^*(\boldsymbol{Y}, \boldsymbol{A}^*(\boldsymbol{Y}, \boldsymbol{f}_0(\boldsymbol{Y}), \boldsymbol{0})) = \boldsymbol{0} \tag{6}$$

- *and the linear operator* $\boldsymbol{f}_2(\boldsymbol{Y})$ *is defined by*

$$(\boldsymbol{f}_2(\boldsymbol{Y})\boldsymbol{\Sigma})_j = \frac{1}{2}\boldsymbol{D}_2(\boldsymbol{Y})\mathbf{Tr}\left(\frac{\partial^2 A_j^*}{\partial \varepsilon \partial \varepsilon'}(\boldsymbol{Y}, \boldsymbol{f}_0(\boldsymbol{Y}), \boldsymbol{0})\boldsymbol{\Sigma}\right) \tag{7}$$

for $j = 1, \dots, J$*, where*

$$\boldsymbol{D}_2(\boldsymbol{Y}) = \left(\boldsymbol{A}_2^*(\boldsymbol{Y}, \boldsymbol{f}_0(\boldsymbol{Y}), \boldsymbol{0})\right)^{-1}.$$

Equivalently,

$$\boldsymbol{f}_2(\boldsymbol{Y})\boldsymbol{\Sigma} = \sum_{l,n=1}^{M} \boldsymbol{K}^{ln}(\boldsymbol{Y})\Sigma_{ln}$$

where each $\boldsymbol{K}^{ln}$ *is a vector in* $\mathbb{R}^J$ *that solves the linear system*

$$\boldsymbol{A}_2^*(\boldsymbol{Y}, \boldsymbol{f}_0(\boldsymbol{Y}), \boldsymbol{0})\, \boldsymbol{K}^{ln}(\boldsymbol{Y}) = \frac{1}{2}\frac{\partial^2 \boldsymbol{A}^*}{\partial \varepsilon_l \partial \varepsilon_n}(\boldsymbol{Y}, \boldsymbol{f}_0(\boldsymbol{Y}), \boldsymbol{0}). \tag{8}$$

We will call these vectors the artificial regressors*.*

Proof. A full proof is Appendix C.1; we describe its main elements here. Since $\boldsymbol{\Sigma}$ is a positive definite matrix, it admits a unique Cholesky decomposition $\boldsymbol{\Sigma} = \boldsymbol{L}\boldsymbol{L}'$. Choose any non-zero coefficient L_{ij} and define $\sigma = |L_{ij}|$. This allows us to define $\boldsymbol{B} = \boldsymbol{L}/\sigma$ and $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{B}\boldsymbol{B}'$. Similarly, we denote $\boldsymbol{v} = \boldsymbol{L}^{-1}\boldsymbol{\varepsilon}$: it is a random vector with mean zero and a unit variance-covariance matrix, and $\varepsilon = \sigma \boldsymbol{B}\boldsymbol{v}$. We will expand $\boldsymbol{\eta}$ as a function of σ and $\boldsymbol{B}$ then recast our results in terms of $\boldsymbol{\Sigma}$.

With this new notation, we define $\mathcal{F}$ by

$$\begin{aligned}\boldsymbol{0} &= \boldsymbol{G}(\boldsymbol{Y}, \boldsymbol{F}(\boldsymbol{Y}, \boldsymbol{\beta}, \boldsymbol{\Sigma}), \boldsymbol{\beta}, \boldsymbol{\Sigma}) \\ &= \boldsymbol{G}^* \left(\boldsymbol{Y}, E_{\boldsymbol{v}} \boldsymbol{A}^*(\boldsymbol{Y}, \mathcal{F}(\boldsymbol{Y}, \boldsymbol{\beta}, \sigma, \boldsymbol{B}) + \boldsymbol{f}_1(\boldsymbol{Y})\boldsymbol{\beta}, \sigma \boldsymbol{B}\boldsymbol{v})\right).\end{aligned} \tag{9}$$

We first prove that the function $\mathcal{F}$ of (9) is well-defined and that it satisfies three very useful properties:

C1: $\mathcal{F}_3(\boldsymbol{Y}, \boldsymbol{\beta}, 0, \boldsymbol{B}) \equiv \boldsymbol{0}$

C2: $\mathcal{F}(\boldsymbol{Y}, \boldsymbol{\beta}, 0, \boldsymbol{B})$ is independent of $\boldsymbol{B}$ and affine in $\boldsymbol{\beta}$.

C3: the second derivative $\mathcal{F}_{33}(\boldsymbol{Y}, \boldsymbol{\beta}, 0, \boldsymbol{B})$ does not depend on $\boldsymbol{\beta}$.

Definition 1 implies that there exists a unique function $\boldsymbol{g}$ such that $\boldsymbol{g}(\boldsymbol{Y}) = E_{\boldsymbol{v}} \boldsymbol{A}^*(\boldsymbol{Y}, \mathcal{F}(\boldsymbol{Y}, \boldsymbol{\beta}, \sigma, \boldsymbol{B}) + \boldsymbol{f}_1(\boldsymbol{Y})\boldsymbol{\beta}, \sigma \boldsymbol{B}\boldsymbol{v})$. For $\sigma = 0$, part 3 of Definition 1 in turn gives us a unique $\boldsymbol{f}_0$ such that $\boldsymbol{g}(\boldsymbol{Y}) = \boldsymbol{A}^*(\boldsymbol{Y}, \boldsymbol{f}_0(\boldsymbol{Y}), \boldsymbol{0})$, which is (6). It translates into $\mathcal{F}(\boldsymbol{Y}, \boldsymbol{\beta}, 0, \boldsymbol{B}) = \boldsymbol{f}_0(\boldsymbol{Y}) - \boldsymbol{f}_1(\boldsymbol{Y})\boldsymbol{\beta}$, which implies property **C2**. With some linear algebra, we obtain **C1**, **C3**, and (7) from the identity

$$\begin{aligned}\boldsymbol{A}^*(\boldsymbol{Y}, \boldsymbol{f}_0(\boldsymbol{Y}), \boldsymbol{0}) &= \boldsymbol{g}(\boldsymbol{Y}) \\ &= E_{\boldsymbol{v}} \boldsymbol{A}^*(\boldsymbol{Y}, \mathcal{F}(\boldsymbol{Y}, \boldsymbol{\beta}, \sigma, \boldsymbol{B}) + \boldsymbol{f}_1(\boldsymbol{Y})\boldsymbol{\beta}, \sigma \boldsymbol{B}\boldsymbol{v}).\end{aligned}$$

□

Note that we did not use *any* distributional assumption on the random coefficients, beyond finite moments. Moreover, the method is detail-free in its implementation: the same formulæ can be applied to any QLRC model. The values taken by the terms in the expansions of course do depend on $\boldsymbol{F}_1$, $\boldsymbol{A}^*$ and $\boldsymbol{G}^*$. We give an illustration for a one-covariate mixed binary choice model without any distributional assumption in Appendix A.2.

1.3 Estimating QLRC Models

The expansion (5) suggests minimizing

$$\left\| E\left(\left(\boldsymbol{f}_0(\boldsymbol{Y}) - \boldsymbol{f}_1(\boldsymbol{Y})\boldsymbol{\beta} - \sum_{l,n=1}^{M} \boldsymbol{K}^{ln}(\boldsymbol{Y})\Sigma_{ln}\right)\boldsymbol{m}(\boldsymbol{Z})\right)\right\|.$$

Taking the parameters of interest to be $(\boldsymbol{\beta}, \boldsymbol{\Sigma})$, this is simply a two-stage least squares regression of $\boldsymbol{f}_0(\boldsymbol{Y})$ on $\boldsymbol{f}_1(\boldsymbol{Y})$ and $\boldsymbol{K}(\boldsymbol{Y})$ with instruments $\boldsymbol{m}(\boldsymbol{Z})$. The *artificial regressors* $\boldsymbol{K}(\boldsymbol{Y})$ can be computed directly from the data, using (8); and their estimated coefficients will be our approximate estimator of $\boldsymbol{\Sigma}$. More precisely, suppose we observe data $\boldsymbol{Y}_i$ for $i = 1, \ldots, N$ generated by a regular QLRC model. Our estimation algorithm is as follows:

Algorithm 1. ***Fast, Detail-free, and Approximately Correct (FRAC)[8] Estimation of Regular Quasi-linear Random Coefficient Models***

1. *For every observation $i = 1, \ldots, N$, take $\boldsymbol{f}_1(\boldsymbol{Y}_i)$ from the definition of the model in (3) and*
   - *invert (6) to compute $\boldsymbol{f}_0(\boldsymbol{Y}_i)$*
   - *use (8) to compute the artificial regressors $\boldsymbol{K}(\boldsymbol{Y}_i)$.*
2. *Run a two-stage least squares regression of $\boldsymbol{f}_0(\boldsymbol{Y})$ on $\boldsymbol{f}_1(\boldsymbol{Y})$ and $\boldsymbol{K}(\boldsymbol{Y})$, taking as instruments a flexible set of functions of the columns of $\boldsymbol{Z}$.*

[8] In a previous version of the paper, we called our method "robust", hence the R in "FRAC". "Detail-free" is a more accurate characterization; we decided to keep the FRAC acronym, which is more euphonic than FDFAC.

3. *Define $\hat{\boldsymbol{\beta}}_N$ to be the estimated coefficients associated with $\boldsymbol{f}_1$ and $\hat{\boldsymbol{\Sigma}}_N$ the estimated coefficients associated with $\boldsymbol{K}$.*

Since we only used a second-order expansion, our estimators do not converge to the true $\boldsymbol{\theta}_0$ as N goes to infinity; they have a probability limit

$$\boldsymbol{\theta}_2 = \text{plim}(\hat{\boldsymbol{\theta}}_{\boldsymbol{N}}).$$

and they are asymptotically normal around $\boldsymbol{\theta}_2$. Consistent estimates of the covariance matrix of the asymptotic distribution of $\sqrt{N}(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}_2)$ can be obtained from the expressions for the heteroskedasticity consistent covariance matrix for the 2SLS estimator given in White (1982).

In random coefficient models the matrix $\boldsymbol{\Sigma}$ is often taken to be diagonal; and some of its diagonal elements may be fixed at zero. Our algorithm easily adapts to these and other linear constraints of the form $\boldsymbol{\Sigma} = \boldsymbol{C}\boldsymbol{S}$ for functional independent parameters $\boldsymbol{S}$: we only need to redefine the artificial regressors as the product of $\boldsymbol{K}$ and $\boldsymbol{C}$. Imposing that the matrix $\boldsymbol{\Sigma}$ be positive definite would bring in nonlinear constraints and we have not attempted to do so.

The rest of this paper can be seen as an application of Algorithm 1 to the macro-BLP model of empirical industrial organization. For completeness, we describe its implementation to the count data model with heterogeneity of Section 1.1 in Appendix B.2.

2 The macro-BLP model

Much work in empirical IO is based on market share and price data. It has followed Berry et al (1995—hereafter BLP) in specifying a mixed multinomial logit model with product-level random effects. To deal with the endogeneity of prices implied by these product-level random effects, BLP use a Generalized Method Moments (GMM) estimator that relies on the mean independence of the product-level random effects and a set of instruments.

To fix ideas, we define "the standard model" as follows. Let J products be available on each of T markets. Each market contains an infinity of consumers who choose one of J products. Consumer i in market t derives a conditional indirect utility from consuming product j equal to

$$\boldsymbol{X}'_{jt}\tilde{\boldsymbol{\beta}}_i + \xi_{jt} + u_{ijt}.$$

There is also a good 0, the "outside good", whose utility for consumer i is typically normalized to equal u_{i0t}. The random variables $\tilde{\boldsymbol{\beta}}$ represent individual variation in tastes for observed product characteristics, while the $\boldsymbol{u}$ stand for unobserved heterogeneity observed by the individual, but not by the econometrician. The vectors $\tilde{\boldsymbol{\beta}}$ and $\boldsymbol{u}$ are independent of each other, and of the covariates $\boldsymbol{X}$ and product random effects $\boldsymbol{\xi}$.

The simplest such specification assumes that the elements of the vector $\boldsymbol{u}_{it} = (u_{i0t}, u_{i1t}, \ldots, u_{iJt})$ are independently and identically distributed (iid) as standard type-I Extreme Value (EV) variables; the product effects ξ_{jt} are unknown mean zero random variables conditional on a set of instruments; and the random variation in preferences $\tilde{\boldsymbol{\beta}}_i$ has a distribution which is known up to its mean $\bar{\boldsymbol{\beta}}_0$ and its variance-covariance matrix $\boldsymbol{V}_0$. This distribution is often modeled as independent, identically distributed $N(\bar{\boldsymbol{\beta}}_0, \boldsymbol{V}_0)$ random vectors with a diagonal variance-covariance $\boldsymbol{V}_0$; we won't need such a distributional assumption.

Like the original BLP paper, we allow for a more general structure:

$$\tilde{\boldsymbol{\beta}}_i = \boldsymbol{\Pi}_0 \boldsymbol{D}_i + \boldsymbol{\varepsilon}_i$$

where

- $\boldsymbol{\Pi}_0$ are unknown coefficients of a random vector $\boldsymbol{D}_i$ whose distribution is known and typically depends on the market t;

- $\boldsymbol{\varepsilon}_i$ has a mean zero distribution with an unknown, finite variance-covariance matrix $\boldsymbol{V}_0$;

- $\boldsymbol{\varepsilon}_i$ and $\boldsymbol{D}_i$ are distributed independently of each other.

In the literature, the $\boldsymbol{D}$ variables are often called "micromoments" or "demographics". Berry et al (1995) used one such variable to represent the distribution of income within each market. Nevo (2001) added age and children.

We break down D_i into its mean on market t and its within-market variation: $\boldsymbol{D}_i = \bar{\boldsymbol{D}}_t + \tilde{\boldsymbol{D}}_i$. This allows us to rewrite

$$\boldsymbol{X}'_{jt}\tilde{\boldsymbol{\beta}}_i = \bar{\boldsymbol{X}}'_{jt}\bar{\boldsymbol{\Pi}} + \boldsymbol{X}'_{jt}\boldsymbol{\Pi}\tilde{\boldsymbol{D}}_i + \boldsymbol{X}'_{jt}\boldsymbol{\varepsilon}_i$$

where $\bar{\boldsymbol{X}}_{jt} \equiv \boldsymbol{X}_{jt} \otimes \bar{\boldsymbol{D}}_t$ has nonrandom coefficients; and $\bar{\boldsymbol{\Pi}} \equiv \text{vec}(\boldsymbol{\Pi})$. We will use the notation $\boldsymbol{\nu}_i$ for the term $\boldsymbol{\Pi}\tilde{\boldsymbol{D}}_i + \boldsymbol{\varepsilon}_i$. so that $\boldsymbol{X}'_{jt}\tilde{\boldsymbol{\beta}}_i = \bar{\boldsymbol{X}}'_{jt}\bar{\boldsymbol{\Pi}} + \boldsymbol{X}'_{jt}\boldsymbol{\nu}_i$. Denote $\boldsymbol{\Omega}_t$ the (known) variance of D_i on market t. Then the variance-covariance matrix of $\boldsymbol{\nu}$ on this market is

$$\boldsymbol{\Sigma}_t \equiv V\boldsymbol{\nu}_i = \boldsymbol{\Pi}\boldsymbol{\Omega}_t\boldsymbol{\Pi}' + \boldsymbol{V}.$$

Models without micromoments are the special case when $\boldsymbol{D}_i$ is the constant 1. Then $\bar{D}_t \equiv 1$ and $\tilde{D}_i \equiv 0$, so that $\bar{\boldsymbol{X}}_{jt} \equiv \boldsymbol{X}_{jt}$ and $\boldsymbol{\nu}_i \equiv \boldsymbol{\varepsilon}_i$, with variance-covariance matrix $\boldsymbol{\Sigma}_t \equiv \boldsymbol{V}$.

Some of the covariates in $\boldsymbol{X}_{jt}$ may be correlated with the product-specific random effects. The usual example is a model of imperfect price competition where the prices firms set in market t depend on the value of the vector of unobservable product characteristics, $\boldsymbol{\xi}_t$, some of which the firms observe.

The parameters to be estimated are the mean coefficients $\boldsymbol{\Pi}_0$ and the variance-covariance matrix of the random coefficients $\boldsymbol{V}_0$. We collect them in $\boldsymbol{\theta}_0 = (\boldsymbol{\Pi}_0, \boldsymbol{V}_0)$. The data available consists of the market shares $(s_{1t}, \ldots, s_{Jt})$ and prices $(p_{1t}, \ldots, p_{Jt})'$ of the J varieties of the good, of the covariates $\boldsymbol{X}_t$, and of additional instruments $\boldsymbol{Z}_t$, all for market t. Note that the market shares do not include information on the proportion S_{0t} of consumers who choose to buy good 0. Typically the analyst estimates this from other sources. Let us assume that this is done, so that we can deal with the augmented vector of market shares $(S_{0t}, S_{1t}, \ldots, S_{Jt})$, with $S_{jt} = (1 - S_{0t})s_{jt}$ for $j \in \mathcal{J} = \{1, \ldots, J\}$.

The market shares for market t are obtained by integration over the variation in preferences, which comes from both $\tilde{\boldsymbol{D}}$ and $\boldsymbol{\varepsilon}$: for good $j \in \mathcal{J}$,

$$S_{jt} = E_{\tilde{\boldsymbol{D}},\boldsymbol{\varepsilon}} \left[\frac{\exp\left(\bar{\boldsymbol{X}}'_{jt}\bar{\boldsymbol{\Pi}} + \boldsymbol{X}'_{jt}\boldsymbol{\Pi}\tilde{\boldsymbol{D}} + \boldsymbol{X}'_{jt}\boldsymbol{\varepsilon} + \xi_{jt}\right)}{1 + \Sigma_{k=1}^J \exp\left(\bar{\boldsymbol{X}}'_{kt}\bar{\boldsymbol{\Pi}} + \boldsymbol{X}'_{kt}\boldsymbol{\Pi}\tilde{\boldsymbol{D}} + \boldsymbol{X}'_{kt}\boldsymbol{\varepsilon} + \xi_{kt}\right)} \right] \tag{10}$$

and $S_{0t} = 1 - \sum_{j=1}^J S_{jt}$.

Berry et al. (1995) assume that

$$E\left(\xi_{jt}|\boldsymbol{Z}_{jt}\right)=\mathbf{0}$$

for all $j \in \mathcal{J}$ and t. The instruments $\boldsymbol{Z}_{jt}$ may for instance be the characteristics of competing products, or cost-side variables. The procedure is operationalized by showing that for given values of $\boldsymbol{\theta}$, the system (10) defines an invertible mapping[9] in $\mathbb{R}^J$. Call $\Xi(\boldsymbol{S}_t,\boldsymbol{\theta})$ its inverse; a GMM estimator obtains by choosing functions $\boldsymbol{Z}^*_{jt}$ of the instruments and minimizing a well-chosen quadratic norm of the sample analogue of:

$$E\left(\Xi(\boldsymbol{S}_t,\boldsymbol{\theta})\boldsymbol{Z}^*_{jt}\right)$$

over $\boldsymbol{\theta}$.

These models have proved very popular; but their implementation has faced a number of issues. Some recent literature has focused on the sensitivity of the estimates to the instruments used in GMM estimation of the mixed multinomial logit model. Reynaert–Verboven (2014) showed that using linear combinations of the instruments can lead to unreliable estimates of the parameters of interest. They recommend using the optimal instruments given by the Amemiya (1975) formula:

$$\boldsymbol{Z}^*_{jt}=E\left(\frac{\partial \boldsymbol{\Xi}}{\partial \boldsymbol{\theta}}(\boldsymbol{S}_t,\boldsymbol{\theta}_0)|\boldsymbol{Z}_{jt}\right).$$

As implementing the Amemiya formula relies on a consistent first-step estimate of $\boldsymbol{\theta}_0$, this is still problematic. Gandhi and Houde (2020) propose "differentiation IVs" to approximate the optimal instruments for the parameters $\boldsymbol{V}$ of the distribution of the random preferences $\boldsymbol{\varepsilon}$. They also suggest a simple regression to detect weak instruments. An alternative is to use the Continuously Updating Estimator to build up the optimal instruments as minimization progresses. Armstrong (2016) points out that instruments based on the characteristics of competing products achieve identification through correlation with markups. But when the number of products is large, many models of the cost-side of the market yield markups that just do not have enough variation, relative to sampling error. This can give inconsistent or just uninformative estimates[10].

[9]See Berry (1994).

[10]Instruments that affect marginal cost directly (if available) do not require variation in the markup to shift prices, and therefore do not suffer from these issues. Variation in the number of products per market may also be used to restore identification, data permitting.

Computation has also been a serious issue. The original BLP approach used a "nested fixed point" (NFP) approach: every time the objective function to be minimized was evaluated for the current parameter values, a contraction mapping/fixed-point algorithm must be employed to compute the implied product effects $\boldsymbol{\xi}_t$ from the observed market shares $\boldsymbol{S}_t$ and for the current value of $\boldsymbol{\theta}$. This was both very costly and prone to numerical errors that propagate from the nested fixed point algorithm to the minimization algorithm. Dubé et al (2012) proposed a nonlinearly-constrained, nonlinear optimization problem to estimate $\boldsymbol{\theta}$. Their simulations suggest that this "MPEC" approach often outperforms the NFP method, sometimes by a large factor. Lee and Seo (2015) proposed an "approximate BLP" method that inverts a linearized approximation of the mapping from $\boldsymbol{\xi}_t$ to $\boldsymbol{S}_t$. They argue that this can be even faster than the MPEC approach to estimation. Petrin and Train (2010) have proposed a maximum likelihood estimator that replaces endogeneous regressors with a control function. This circumvents the need to compute the implied value of $\boldsymbol{\xi}$ for each value of $\boldsymbol{\theta}$, but still requires solving a nonlinear optimization problem to compute an estimate of $\boldsymbol{\theta}_0$. Solving a nonlinear optimization problem for a potentially large set of parameters is time-consuming. It typically requires starting values in the neighborhood of the optimal solution; closed-form gradients; and careful monitoring of the optimization algorithm by the analyst, as the objective function is not globally concave.

Conlon and Gortmaker (2020) cover all of these issues in great detail; and their Python module `pyblp` incorporates what they found to be the best practices (some of which they contributed.) Their conclusion is measured: "it is possible to obtain good performance even in small samples and without exogeneous cost-shifters, particularly when "optimal instruments" are employed along with supply-side restrictions." It is quite easy to incorporate such supply-side restrictions in our approach; we show it in Section 3.3.3.

The method we propose in this paper completely circumvents the need to solve a nonlinear optimization problem. It also avoids the computational burden of generating a large set of random draws from a multidimensional distribution. Our estimator relies on an approximate model that is exactly valid when there is no random variation in preferences, and becomes a coarser approximation as the amplitude of the random variation in elements of $\tilde{\boldsymbol{\beta}}_i$ grows. As such, our estimator is *not* a consistent estimator of the parameters of the BLP model. On the other hand, it has some very real advantages that may tip the scale in its favor. First, it requires a single linear 2SLS regression that can be computed in microseconds with off-the-shelf software[11]. Second, our estimator needs to assume very little about the form of the distribution of the random variation in preferences $\boldsymbol{\nu}$ (beyond its small scale), justifying the "detail-free" in our title.

Some readers may find the "approximate correctness" of our estimator unsatisfying. It at least yields "nearly consistent" starting values for the classical nested-fixed point and MPEC nonlinear optimization procedures at a minimal cost. This can address a major challenge associated with successfully implementing the MPEC estimation procedure. It also provides useful diagnoses about how well different parameters can be identified with a particular model and dataset; and a simple way to select between models, as we will explain.

3 2SLS Estimation in the Standard BLP Model

For the reader primarily interested in applying our method to empirical industrial organization, this section provides a step-by-step guide to implementing the estimator in the standard macro-BLP model. For simplicity, we concentrate on the most basic model.

3.1 Expansions and the Artificial Regressors

Let us first establish that the macro-BLP model belongs to the class of QLRC models we introduced in Section 1. To see this, consider a single market and define $\boldsymbol{Y} = (\boldsymbol{S}, \boldsymbol{X})$; $\boldsymbol{f}_1(\boldsymbol{Y}) = \bar{\boldsymbol{X}}$; and $\boldsymbol{\eta} = \boldsymbol{\xi}$.

Now let $\boldsymbol{G}_j^*(\boldsymbol{Y}, \boldsymbol{a}) = S_j - a_j$ and

$$A_j^* = \Pr\left(j = \arg \max_{J=0,1,\ldots,J} \left(\bar{\boldsymbol{X}}_k' \bar{\boldsymbol{\Pi}} + \boldsymbol{X}_k' \boldsymbol{\nu} + \xi_k + u_k \right) | \boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{\nu} \right)$$

[11] Fox et al. (2011) discretize the distribution of the random coefficients on a grid and estimate the corresponding probability masses. This also results in a least-squares estimator; theirs is constrained by linear inequalities and may be sensitive to the choice of the grid points. Nevo, Turner and Williams (2016) report a positive experience with a very large grid.

so that, denoting $b_j = \bar{\boldsymbol{X}}_j' \bar{\boldsymbol{\Pi}} + \xi_j$, we can rewrite

$$A_j^*(\boldsymbol{X}, \boldsymbol{b}, \boldsymbol{\nu}) \equiv \frac{\exp\left(\boldsymbol{b}_j + \boldsymbol{X}_j' \boldsymbol{\nu}\right)}{1 + \sum_{k=1}^J \exp\left(\boldsymbol{b}_k + \boldsymbol{X}_k' \boldsymbol{\nu}\right)}. \tag{11}$$

This recasts the macro-BLP model as a QLRC model, which is obviously regular. Applying Theorem 1 shows that $\boldsymbol{f}_0(\boldsymbol{y})$ has a very simple expression:

$$\boldsymbol{f}_0(\boldsymbol{Y}) = \log \frac{\boldsymbol{S}}{S_0}$$

where S_0 is the market share of good 0. This simply reflects the well-known fact that when the coefficients are not random, the model can be estimated by regressing the log-odds ratios of market shares on the covariates. To compute the artificial regressors, we use (8). We do this in Appendix C.2. To state our results, we introduce some useful notation:

Definition 2 (Market share weighting). *For any J-dimensional vector $\boldsymbol{T}$ of J components, we define the scalar*

$$e_{\boldsymbol{S}} \boldsymbol{T} = \sum_{k=1}^J S_k T_k.$$

By extension, if $\boldsymbol{m}$ is a $(J \times J)$ matrix with J columns $(\boldsymbol{m}_1, \ldots, \boldsymbol{m}_J)$, we define the vector

$$e_{\boldsymbol{S}} \boldsymbol{m} = \sum_{k=1}^J S_k \boldsymbol{m}_k.$$

It is important to note here that the operator $e_{\boldsymbol{S}}$ does *not* conserve mass, as the weights $(S_1, \ldots, S_J)$ only sum to $(1 - S_0)$.

Given these definitions, we summarize our results in the following theorem:

Theorem 2 (Artificial Regressors for Macro-BLP without micromoments). *In the macro-BLP model without micromoments ($\boldsymbol{D}_i \equiv 1$), the artificial regressors are given by*

$$K_j^{ln} = \frac{1}{2} \left(X_{jl} e_{\boldsymbol{S}} \boldsymbol{X}_n + X_{jn} e_{\boldsymbol{S}} \boldsymbol{X}_l - X_{jl} X_{jn} \right). \tag{12}$$

If the matrix $\boldsymbol{\Sigma}$ is restricted by $\boldsymbol{\Sigma}_{ln} = \sum_{p=1}^P C_{ln}^p S_p$, then the artificial regressors associated with the parameter S_p are

$$\sum_{l,n=1}^M C_{ln}^p K_j^{ln}.$$

In particular, if the matrix $\boldsymbol{\Sigma}$ *is diagonal, then the artificial regressors are given by*

$$K_j^{ll} = X_{jl}\left(e_{\boldsymbol{S}}\boldsymbol{X}_l - \frac{X_{jl}}{2}\right). \tag{13}$$

Models with micromoments require a bit more care, because of the quadratic term $\boldsymbol{\Pi}\boldsymbol{\Omega}_t\boldsymbol{\Pi}'$ in the variance of $\boldsymbol{\nu}$. As a consequence, the second-order expansion has an additional term:

$$F_j(\boldsymbol{Y}_t, \boldsymbol{\Pi}, \boldsymbol{V}) \simeq \log\frac{S_{jt}}{S_{0t}} - \bar{\boldsymbol{X}}_{jt}\bar{\boldsymbol{\Pi}} - \sum_{l,n=1}^{M} K_{jt}^{ln}V^{ln} - \sum_{l,n=1}^{M} K_{jt}^{ln}\left(\boldsymbol{\Pi}\boldsymbol{\Omega}_t\boldsymbol{\Pi}'\right)_{ln}. \tag{14}$$

Moreover, this term is a quadratic form of the same parameters that appear as coefficients of $\bar{\boldsymbol{X}}_{jt}$. We explain how to deal with this in Section 3.3.2.

3.2 Intuition

To understand the formula for the artificial regressors, first consider the model without micromoments ($\boldsymbol{D} \equiv 1$). For simplicity and like much of the literature, assume that the variance-covariance matrix $\boldsymbol{\Sigma} = \boldsymbol{V}$ is diagonal. Formula (13) shows that when $\boldsymbol{\Sigma}$ is relatively small, so that our approximate model is a reasonable one, its elements are identified from a simple quadratic form of the corresponding observed characteristics, weighted by the market shares. One attraction of this approximation is that the artificial regressors can be computed straightforwardly, and their variations examined, before resorting to any estimation. The presence of quadratic terms is not surprising, since the model multiplies $\boldsymbol{X}$ by $\boldsymbol{\varepsilon}$. The very simple form of the artificial regressors was less predictable. To understand it better, we turn to the $J = 1$ subcase—that is, a mixed logit model.

When $J = 1$, we have $e_{\boldsymbol{S}}\boldsymbol{T} = S_1T_1$ for any variable $\boldsymbol{T}$. The artificial regressors on market t are simply

$$\left(S_{1t} - \frac{1}{2}\right)X_{1tl}^2$$

for each covariate l that has a random coefficient. The focal role of the one-half market share is a consequence of the symmetric shape of the logistic distribution. With $J = 1$, the model is

$$S_{1t} = E_{\boldsymbol{\varepsilon}}L\left(\boldsymbol{X}(\bar{\boldsymbol{\Pi}} + \boldsymbol{\varepsilon})\right),$$

where $L(t) = 1/(1 + \exp(-t))$ is the cdf of the logistic. Our second order expansions bring in the second derivative of L, which is

$$L''(t) = L(t)(1 - L(t))(1 - 2L(t)).$$

As L has an inflection point at $t = 0$, where it equals 1/2, it is locally flat and a first-order certainty equivalence prevails: if the argument of L does not vary much around $L^{-1}(1/2) = 0$, then the model is second-order equivalent to a model with non-random coefficients. As a consequence, it is very hard to identify the variance of $\boldsymbol{\varepsilon}$ when $J = 1$ and the market share stays close to 1/2. Away from this region, the variance in the characteristics of the product and in its market share identifies the variance of the random coefficients.

With more products ($J > 1$), the term $e_{\boldsymbol{S}}\boldsymbol{X}$ introduces variations in the characteristics of other products into the artificial regressors. This gives more identifying power to the approximate estimator.

3.3 Estimating the Approximate macro-BLP Model

For notational simplicity, we assume that we use all $J \times T$ conditional moment restrictions:

$$E\left(\xi_{jt}|\boldsymbol{Z}_{jt}\right) = 0.$$

Adapting our procedure to subsets of moment restrictions is straightforward.

3.3.1 Without Micromoments

Our procedure runs as follows for the model without micromoments:

Algorithm 2. *FRAC estimation of the standard BLP model*

1. *For every market t, augment the market shares from $(s_{1t}, \ldots, s_{Jt})$ to $(S_{0t}, S_{1t}, \ldots, S_{Jt})$*
2. *For every product-market pair $(j \in \mathcal{J}, t)$:*
    (a) *compute the market-share weighted covariate vector $\boldsymbol{e}_t = \sum_{k=1}^{J} S_{kt}\boldsymbol{X}_{kt}$;*
    (b) *for every (m, n) for which Σ_{mn} is not set at zero, compute the "artificial regressor" K_{mn}^{jt} as*

- *if* $n = m$*:* $K^{jt}_{mm} = \left(\frac{X_{jtm}}{2} - e_{tm}\right) X_{jtm}$*;*
- *if* $n < m$*:* $K^{jt}_{mn} = X_{jtm}X_{jtn} - e_{tm}X_{jtn} - e_{tn}X_{jtm}$.

(c) for every $j = 1, \dots, J$*, define* $y_{jt} = \log(S_{jt}/S_{0t})$

3. *Run a two-stage least squares regression of* $\boldsymbol{y}$ *on* $\bar{\boldsymbol{X}}$ *and* $\boldsymbol{K}$*, taking as instruments a flexible set of functions of the columns of* $\boldsymbol{Z}$*. Define* $\hat{\boldsymbol{\Pi}}$ *to be the estimated coefficients associated with* $\boldsymbol{X}$ *and (the nonzero part of)* $\hat{\boldsymbol{\Sigma}}$ *to be the estimated coefficients associated with* $\boldsymbol{K}$*.*

4. *(optional[12]) Run a three-stage least squares (3SLS) regression across the* T *markets stacking the* J *equations for each product with a weighting matrix equal to the inverse of the sample variance of the residuals from step 3.*

Consistent estimates of the covariance matrix of the asymptotic distribution of $\sqrt{TJ}(\hat{\boldsymbol{\theta}} - \text{plim}(\hat{\boldsymbol{\theta}}))$ can be obtained from the expressions for the heteroskedasticity consistent covariance matrix for the 2SLS estimator given in White (1982).

Ideally, the "flexible set of functions of the columns of $\boldsymbol{Z}$" in step 3 should be able to span the space of the instruments $E(\boldsymbol{X}|\boldsymbol{Z})$ and $E(\boldsymbol{K}|\boldsymbol{Z})$ that are optimal for our approximate model. Alternatively, these instruments can be estimated by a nonparametric regressions of each column of $\boldsymbol{X}$ on the columns of $\boldsymbol{Z}$.

As is well-known, misspecification of one equation of the model can lead to inconsistency in 3SLS parameter estimates of all equations of the model. It is therefore not clear that Step 4 is worth the additional effort.

It is important to reiterate here that $\boldsymbol{e}$ is *not* a simple weighted average, as the weights do not sum to one, but only to $(1 - S_{0t})$. To illustrate, if $X_{jtm} \equiv 1$ is the constant, then e_{tm} is $(1 - S_{0t})$ and the artificial regressor that identifies the corresponding variance parameter is

$$K^{jt}_{mm} = S_{0t} - \frac{1}{2}.$$

More generally, if $X_{jtn} = \mathbb{1}(j \in \mathcal{J}_0)$ is a dummy that reflects whether variety j belongs to group $\mathcal{J}_0 \subset \mathcal{J}$, then it is easy to see that the corresponding variance parameter is

[12]This step should only be considered when T is large relative to J.

the coefficient of the artificial regressor

$$K^{jt}_{nn} = \mathbb{1}(j \in \mathcal{J}_0)\left(\frac{1}{2} - S_{\mathcal{J}_0 t}\right)$$

where $S_{\mathcal{J}_0 t}$ is the market share of group $\mathcal{J}_0$ on market t.

3.3.2 Adding Micromoments

In the presence of micromoments, we propose two complementary approaches that only require two-stage least-squares estimation:

- if the variance $\boldsymbol{\Omega}_t$ of the micromoments does not vary much across markets, Algorithm 2 gives estimates of $\boldsymbol{\Pi}$ and $\boldsymbol{\Sigma}$; and the variance of $\boldsymbol{\varepsilon}$ can be recovered as $\boldsymbol{V} = \boldsymbol{\Sigma} - \boldsymbol{\Pi}\boldsymbol{\Omega}\boldsymbol{\Pi}'$, where $\boldsymbol{\Omega}$ is an average of the $\boldsymbol{\Omega}_t$.

- this can be refined by a simple iterative procedure if the differences in the variances cannot be neglected. Start with the approach in the previous bullet point to get estimates $\boldsymbol{\Pi}^{(0)}, \boldsymbol{V}^{(0)}$. At each iterative step, given estimates $\boldsymbol{\Pi}^{(s)}$, subtract the term
$$\sum_{l,n=1}^{M} K^{ln}_{jt}\left(\boldsymbol{\Pi}^{(s)}\boldsymbol{\Omega}_t\left(\boldsymbol{\Pi}^{(s)}\right)'\right)$$
from $\log y_{jt}$ and apply step 3 of Algorithm 2 to obtain new estimates $\boldsymbol{\Pi}^{(s+1)}, \boldsymbol{V}^{(s+1)}$. Stop when the estimates stabilize.

3.3.3 Adding the Supply Side

Modeling supply jointly with demand has two advantages in the macro-BLP model: it adds identfying information, and it allows to run counterfactuals. Our approach easily accommodates a supply block.

To see this, suppose that a firm f produces a set of varieties V_f at constant marginal cost c^f_{kt} on market t for each $k \in V_f$. It chooses prices $(p_{kt})_{k\in V_f}$ to maximize

$$\sum_{k\in V_f}(p_{kt} - c^f_{kt})S_{kt}$$

where the market share of f depends on its prices, on competitor's prices, and on the characteristics of consumers' demand. In Nash equilibrium, all firms sell at the same prices on any given market.

The first-order conditions of this problem are

$$p_{jt} + \sum_{k \in V_f} (p_{kt} - c^f_{kt}) \frac{\partial S_{kt}}{\partial p_{jt}} = 0 \text{ for each } j \in V_f.$$

We can rewrite them as

$$c^f_{jt} = p_{jt}(1 + \mu_{jt})$$

where the markup μ_{jt} can be evaluated once the market demand functions are known.

Suppose for simplicity that on each market t, there are J firms, each of which produces only one variety: $V_j = \{j\}$ for $j = 1, \ldots, J$. Then Lerner's formula gives

$$\frac{1}{\mu_{jt}} = -\frac{\partial \log S_{jt}}{\partial \log p_{jt}}.$$

Since marginal costs must be positive, we specify

$$\log c^j_{jt} = \boldsymbol{W}_{jt}\boldsymbol{\gamma} + \omega_{jt}$$

where ω_{jt} is orthogonal to some functions $m_S(\boldsymbol{Z}_{jt})$ of the instruments. This gives a set of moment conditions

$$E\left(\log p_{jt} + \log(1 + \mu_{jt}) - \boldsymbol{W}_{jt}\boldsymbol{\gamma}\right) m_S(\boldsymbol{Z}_{jt}) = \boldsymbol{0}. \tag{15}$$

The markups ν_{jt} are complicated nonlinear functions of the parameters $\boldsymbol{\Pi}$ and $\boldsymbol{\Sigma}_t$ of the demand system. However, they are easy to evaluate once these parameters are estimated from the demand system. Replacing $\boldsymbol{\mu}$ with the estimated $\hat{\boldsymbol{\mu}}$ in (15) gives an estimating equation that is linear in the parameters $\boldsymbol{\gamma}$ and can again be estimated by two-stage least-squares.

This recursive approach provides us with approximate estimates of both consumer preferences and cost functions, using only two-stage least-squares estimation. Joint estimation would allow us to improve the estimates of demand parameters by using the fit of the supply equations; unfortunately, it cannot be done without breaking the appealing linearity of our recursive approach.

4 Pros and Cons of the 2SLS Estimation Approach

Our method has two obvious drawbacks. The first one is minor: since the elements of the variance-covariance matrix $\mathbf{\Sigma}$ are estimated as the coefficients of the corresponding artificial regressors $\boldsymbol{K}$, the resulting matrix $\hat{\mathbf{\Sigma}}$ may not be semi-definite positive.

The second drawback is more substantial: because this is only an approximate model, the resulting estimator $\hat{\boldsymbol{\theta}}$ will not converge to $\boldsymbol{\theta}_0$ as the number of markets T goes to infinity. We discuss this in much more detail in Section 4.1. For now, let us note that this drawback is tempered by several considerations. First, the number of markets available in empirical IO is typically small; finite-sample performance of the estimator is what matters, and we will examine that in Section 6. More importantly, our estimator has several useful features. Let us list six of them:

1. Because the estimator employs linear 2SLS, computing it is extremely fast and can be done in microseconds with any of-the-shelf software.

2. We do not have to assume any distributional form for the random variation in preferences $\boldsymbol{\varepsilon}$. This is a notable advantage over other method, which yield inconsistent estimates if the distribution of $\boldsymbol{\varepsilon}$ is misspecified.

3. Computing the optimal instruments does not require any first-step estimate because the estimating equation is linear. We can just use a flexible set of functions of the columns of $\boldsymbol{Z}$ that span the space of the optimal instruments $E(\boldsymbol{X}|\boldsymbol{Z})$ and $E(\boldsymbol{K}|\boldsymbol{Z})$.

4. Even if the econometrician decides to go for a different estimation method, our proposed 2SLS estimates obtained should provide a set of very good initial parameter values for a nonlinear optimization algorithm.

5. The confidence regions on the estimates will give useful diagnoses about the strength of identification of the parameters, both mean coefficients $\mathbf{\Pi}$ and their random variation $\mathbf{\Sigma}$. This would be very hard to obtain otherwise, except by trying different specifications.

6. There has been much interest in systematic specification searches in recent years; see e.g. Horowitz-Nesheim 2019 for a Lasso-based selection approach in

discrete choice models. With our method any number of variants can be tried in seconds, and model selection is drastically simplified.

4.1 The Quality of the Approximation

Ideally, we would be able to bound the approximation error in the expansion of ξ_j, and use this bound to majorize the error in our estimator in the manner described in Kristensen and Salanié (2017). While we have not gone that far, we can justify the local-to-zero validity of the expansion in the usual way. We are taking a mapping

$$\boldsymbol{S} = G\left(\boldsymbol{\xi}, \boldsymbol{X}, \sigma\right)$$

that is differentiable in both $\boldsymbol{\xi}$ and σ; inverting it to $\boldsymbol{\xi} = \Xi\left(\boldsymbol{S}, \boldsymbol{X}, \sigma\right)$; and taking an expansion to the right of $\sigma = 0$ for fixed market shares $\boldsymbol{S}$ and covariates $\boldsymbol{X}$. The validity of the expansion for small σ and fixed $(\boldsymbol{X}, \boldsymbol{S})$ depends on the invertibility of the Jacobian $G_{\boldsymbol{\xi}}$.

First consider the standard model. It follows from Berry 1994 that $G_{\boldsymbol{\xi}}$ is invertible if no observed market share hits zero or one. Applying the Implicit Function Theorem repeatedly shows that in fact the Taylor series of $\boldsymbol{\xi}$ converges over some interval $[0, \bar{\sigma}]$ if all moments of $\boldsymbol{\varepsilon}$ are finite; and that the expansion is valid at order L if the moments of $\boldsymbol{\varepsilon}$ are bounded to order $(L+1)$.

Characterizing this range of validity is trickier. Figure 1 uses formulæ derived in Appendix A to plot the first four coefficients of the expansion in $\Sigma_{11} X_1^2$ for the standard Gaussian binary model (that is, the Gaussian mixed logit) with one covariate X_1:

$$\xi_1 = \log \frac{S_1}{S_0} - \beta X_1 + \sum_{k=1}^{4} t_k(S_1) \left(\Sigma_{11} X_1^2\right)^k + O(\sigma^{10}).$$

Each curve plots the function t_k as market shares vary between zero and one. The visual impression is clear: the coefficients damp quickly. Beyond the first term (which corresponds to our 2SLS method), the coefficients are always smaller than 0.05 in absolute value. Of course, the approximation error also depends on the values taken by the covariates.

While this simple example can only be illustrative, we find the figure encouraging as to the practical range of validity of the approximation. To go further, in Section 5

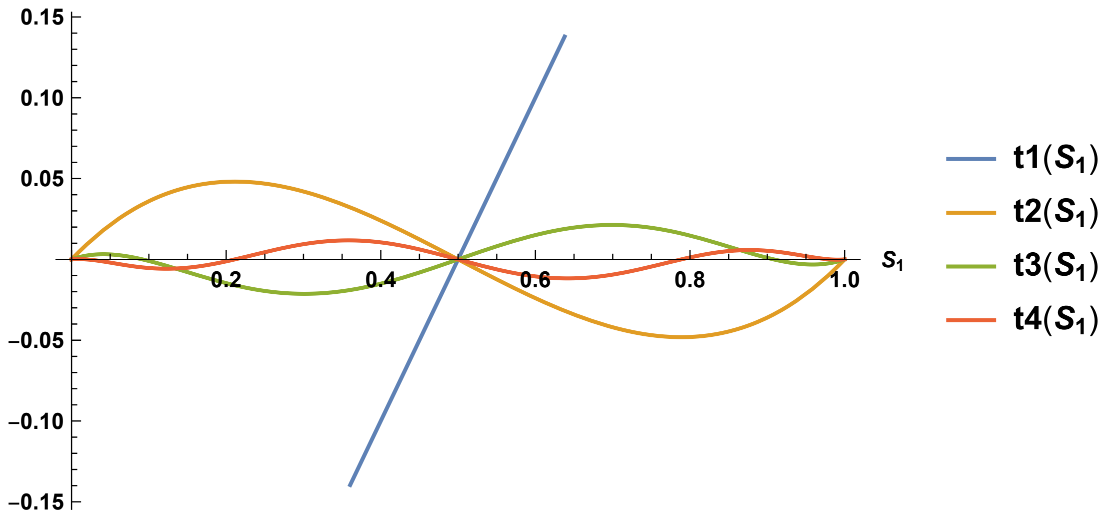


Figure 1: Coefficients $t_{1,2,3,4}(S_1)$

we simulate a simple multinomial logit model and we explore the properties of the estimated parameters and elasticities as the number of markets becomes very large. This allows us to quantify the asymptotic bias of our approximate estimators, anong other things.

4.2 Invariance to Higher-order Moments

Our expansions only rely on the properties of the derivatives of the logistic cdf $L(t) = \frac{1}{1+\exp(-t)}$ and on the first two moments of $\boldsymbol{\varepsilon}$. This has a distinct advantage over competing methods: the lower-order moments of $\boldsymbol{\varepsilon}$ can be estimated by 2SLS, and nothing more needs to be known about its distribution.

Suppose for instance that the analyst does not want to assume that $\boldsymbol{\varepsilon}$ has a symmetric distribution. Then the artificial regressors are unchanged. In the absence of symmetry, the approximate model may or may not be be a worse approximation; in any case, running Algorithm 2 should still provide useful estimators of the elements of $\boldsymbol{\Sigma}_0$.

We follow with a modification of our multinomial logit random coefficients modeling framework to account for the third and fourth moments of $\boldsymbol{\varepsilon}$. We then turn to methods for improving the quality of our 2SLS estimates[13]. Finally, Appendix B.1 presents a nonlinear 2SLS estimation procedure for the random coefficient nested logit model.

4.3 Higher-order terms

In Appendix A, we study in more detail the standard binary model. For this simpler case, calculations are easily done by hand for lower orders of approximation, or using symbolic software for higher orders.

More generally, return to the standard model and assume (as is often done in practice) that there are no micromoments and the ε_m are independent across the covariates $m = 1, \ldots, n_X$. We denote $\sigma_m^2 = \Sigma_{mm} = E(\varepsilon_m^2)$, and $s_m = E\varepsilon_m^3$. Simple calculations[14] show that the third-order expansion is

$$\xi_j = \log \frac{S_j}{S_0} - \bar{\boldsymbol{X}}_j \bar{\boldsymbol{\Pi}} - \sum_{m=1}^{M} \boldsymbol{K}_{mm}^j \sigma_m^2 - \sum_{m=1}^{M} T_m^j s_m$$

where the $\boldsymbol{K}_j^{mm}$ are as in Theorem 2 and we introduced new artificial regressors

$$T_m^j \equiv X_{jm} \left(\frac{X_{jm}^2}{6} + (e_{\boldsymbol{S}} \boldsymbol{X_m} - X_{jm}/2) e_{\boldsymbol{S}} \boldsymbol{X}_m - \frac{e_{\boldsymbol{S}}(\boldsymbol{X}_m^2)}{2} \right).$$

Algorithm 2 therefore can be adapted in the obvious way to take possible skewness of $\boldsymbol{\varepsilon}$ into account. Note that the procedure remains linear in the parameters $(\boldsymbol{\Pi}, \boldsymbol{\Sigma}, \boldsymbol{s})$, for which it generates approximate estimates by 2SLS.

The fourth order term has a more complicated structure—see Appendix C.3.

4.4 Correcting the 2SLS estimates

If the analyst is willing to make more distributional assumptions, she can resort to bootstrap or asymptotic corrections to improve the accuracy of our 2SLS estimators.

[13]We explore the small sample properties of several of these corrections in a Monte Carlo study in Section 6.

[14]Available from the authors.

4.4.1 Bootstrapping

Once we have approximate estimators $\hat{\mathbf{\Pi}}$ and $\hat{\mathbf{\Sigma}}$, we can use them to solve the market shares equations for estimates of the product effects $\boldsymbol{\xi}$ and bootstrap them, *provided* that we are willing to impose a distribution for $\boldsymbol{\varepsilon}$ (beyond the normalization of its first two moments.)

Denote $\boldsymbol{\zeta} = \mathbf{\Sigma}^{-1/2}\boldsymbol{\nu}$ the standardized random term. We use Berry inversion to solve for $\hat{\boldsymbol{\xi}}_t$ in the system

$$S_{jt} = E_{\boldsymbol{\zeta}} \frac{\exp\left(\bar{\boldsymbol{X}}_{jt}\hat{\bar{\mathbf{\Pi}}} + \boldsymbol{X}_{jt}\hat{\mathbf{\Sigma}}^{1/2}\boldsymbol{\zeta} + \hat{\xi}_{jt}\right)}{1 + \Sigma_{k=1}^{J}\exp\left(\bar{\boldsymbol{X}}_{kt}\hat{\bar{\mathbf{\Pi}}} + \boldsymbol{X}_{kt}\hat{\mathbf{\Sigma}}^{1/2}\boldsymbol{\zeta} + \hat{\xi}_{kt}\right)},$$

where $E_{\boldsymbol{\zeta}}$ denotes the expectation with respect to the assumed distribution of $\boldsymbol{\zeta}$. For any resample $\boldsymbol{\xi}^*$ of the $\hat{\boldsymbol{\xi}}$, we simulate the market shares from

$$S^*_{jt} = E_{\boldsymbol{\zeta}} \frac{\exp\left(\bar{\boldsymbol{X}}_{jt}\hat{\bar{\mathbf{\Pi}}} + \boldsymbol{X}_{jt}\hat{\mathbf{\Sigma}}^{1/2}\boldsymbol{\zeta} + \xi^*_{jt}\right)}{1 + \Sigma_{k=1}^{J}\exp\left(\bar{\boldsymbol{X}}_{kt}\hat{\bar{\mathbf{\Pi}}} + \boldsymbol{X}_{kt}\hat{\mathbf{\Sigma}}^{1/2}\boldsymbol{\zeta} + \xi^*_{kt}\right)}$$

and we use our 2SLS method to get new estimates $\mathbf{\Pi}^*, \mathbf{\Sigma}^*$. Finally, we compute bias-corrected estimates by e.g.

$$\mathbf{\Pi}^C = 2\hat{\mathbf{\Pi}} - \frac{1}{B}\sum_{b=1}^{B}\mathbf{\Pi}^*_{\boldsymbol{b}}.$$

More generally, the resampled estimates can be used to estimate the distribution of $\hat{\mathbf{\Pi}}$ and $\hat{\mathbf{\Sigma}}$ in the usual manner.

4.4.2 A Two-Step Estimator Based on an Asymptotic Correction

Another way to correct the estimator is to use a Newton-Raphson step to correct for the effects of the approximation. As it turns out, this can be done quite simply if one is willing to impose more structure than the second-order expansion.

Denote $\boldsymbol{\theta} = (\mathbf{\Pi}, \mathbf{\Sigma})$, and $\boldsymbol{\theta}_0$ its true value. Let $\hat{\boldsymbol{\theta}}_2$ be our 2SLS estimator based on a second-order expansion. That is, we estimate the approximate model $\hat{E}(\boldsymbol{Z}_2'\boldsymbol{\xi}_2) = \mathbf{0}$ with a set of instruments $\boldsymbol{Z}_2$ generated from $\boldsymbol{Z}$, where

$$\boldsymbol{\xi}_{2,jt} = \log\frac{S_{jt}}{S_{0t}} - \bar{\boldsymbol{X}}_{jt}\bar{\mathbf{\Pi}} - \operatorname{Tr}\mathbf{\Sigma}\boldsymbol{K}^{\boldsymbol{jt}}. \tag{16}$$

As the number of markets T gets large, $\hat{\boldsymbol{\theta}}_2$ converges to the solution $\boldsymbol{\theta}_2$ of $E(\boldsymbol{\xi}_2\boldsymbol{Z}_2) = \mathbf{0}$. Alternatively, we could have estimated the model using inversion or MPEC, with an "exact" $\boldsymbol{\xi}_\infty$. Let $\boldsymbol{\lambda}_0$ denote additional parameters of the model (such as higher-order moments of the distribution of $\boldsymbol{\varepsilon}$) that are identified using the exact $\boldsymbol{\xi}_\infty$ but not[15] with our approximate $\boldsymbol{\xi}_2$.

Since by assumption $E\left(\boldsymbol{\xi}_\infty(\boldsymbol{\theta}_0, \boldsymbol{\lambda}_0, \boldsymbol{X}, \boldsymbol{S})|\boldsymbol{Z}\right) = 0$, a fortiori $E\boldsymbol{Z}_2'\xi_\infty(\boldsymbol{\theta}_0; \boldsymbol{\lambda}_0) = \mathbf{0}$. By subtraction,

$$\begin{aligned} &E\left(\boldsymbol{Z}_2'(\boldsymbol{\xi}_2(\boldsymbol{\theta}_2) - \boldsymbol{\xi}_\infty(\boldsymbol{\theta}_2))\right) \\ +&E\left(\boldsymbol{Z}_2'(\boldsymbol{\xi}_\infty(\boldsymbol{\theta}_2) - \boldsymbol{\xi}_\infty(\boldsymbol{\theta}_0))\right) = \mathbf{0}. \end{aligned}$$

We approximate the second term with its first-order Taylor expansion

$$E\left(\boldsymbol{Z}_2'\frac{\partial\boldsymbol{\xi}_\infty}{\partial\boldsymbol{\theta}}(\boldsymbol{\theta}_2)(\boldsymbol{\theta}_2 - \boldsymbol{\theta}_0)\right).$$

Moreover, we replace the derivatives of $\boldsymbol{\xi}_\infty$ with those of $\boldsymbol{\xi}_2$, which are simply $-\boldsymbol{\mathcal{X}} \equiv -(\boldsymbol{X}, \boldsymbol{K})$ since by definition $\boldsymbol{\xi}_2(\boldsymbol{\theta}) = \boldsymbol{y} - \boldsymbol{\mathcal{X}}\boldsymbol{\theta}$. Note that these approximations are valid if $\boldsymbol{\theta}_2$ is not too far from $\boldsymbol{\theta}_0$.

This gives us

$$\left(E\boldsymbol{Z}_2'\boldsymbol{\mathcal{X}}\right)(\boldsymbol{\theta}_2 - \boldsymbol{\theta}_0) \simeq E\left(\boldsymbol{Z}_2'(\boldsymbol{\xi}_2(\boldsymbol{\theta}_2) - \boldsymbol{\xi}_\infty(\boldsymbol{\theta}_2))\right).$$

Since $E\boldsymbol{Z}_2'\boldsymbol{\xi}_2(\boldsymbol{\theta}_2) = \mathbf{0}$, we can add it on the left and subtract it on the right to obtain

$$E\boldsymbol{Z}_2'\left(\boldsymbol{\mathcal{X}}\boldsymbol{\theta}_2 + \boldsymbol{\xi}_2(\boldsymbol{\theta}_2)\right) - \left(E\boldsymbol{Z}_2'\boldsymbol{\mathcal{X}}\right)\boldsymbol{\theta}_0 \simeq -E\boldsymbol{Z}_2'\boldsymbol{\xi}_\infty(\boldsymbol{\theta}_2),$$

that is $E\boldsymbol{Z}_2'(\boldsymbol{y} - \boldsymbol{\mathcal{X}}\boldsymbol{\theta_0} + \boldsymbol{\xi}_\infty(\boldsymbol{\theta}_2)) = \mathbf{0}$.

This is simply the (limit) estimating equation for a two-stage least-squares regression. We will be recovering a corrected estimate of $\boldsymbol{\theta}_0$ by regressing the corrected left-hand side variables $\boldsymbol{y}^* = \boldsymbol{y} + \boldsymbol{\xi}_\infty(\boldsymbol{\theta}_2)$ on the same covariates $\boldsymbol{\mathcal{X}}$ we used to obtain $\hat{\boldsymbol{\theta}}_2$, with the same instruments $\boldsymbol{Z}_2$.

To evaluate $\boldsymbol{\xi}_\infty(\hat{\boldsymbol{\theta}}_2)$, we choose a distributional form for $\boldsymbol{\varepsilon}$ (typically a Gaussian) and we use Berry inversion to solve for the values $\boldsymbol{\xi}$ that rationalize the observed market shares $\boldsymbol{S}$ at the estimated parameter values $\hat{\boldsymbol{\theta}}_2$. Then we define

$$y_{jt}^* = \log\left[\frac{S_j}{S_0}\right] + \boldsymbol{\xi}_{\infty j}(\hat{\boldsymbol{\theta}}_2))$$

[15] If the only free parameters of the distribution of $\boldsymbol{\varepsilon}$ are the elements of $\boldsymbol{\Sigma}$, then $\boldsymbol{\lambda_0}$ will be empty.

and we apply our 2SLS procedure with this new dependent variable. Our Monte Carlo study in Section 6 will explore the small sample properties of this two-step procedure.

Note that just as the correction is computed at the initial two-stage least-squares estimators $\hat{\boldsymbol{\theta}}_2$, it could be computed again at the new, corrected estimates. In addition, one could replace $\boldsymbol{\xi}_\infty$ with $\boldsymbol{\xi}_p$ for some integer $p > 2$.

5 Asymptotic Performance of Our Estimators

As the sample size (the number of markets in the macro-BLP application) grows, our approximate estimator converges to a pseudo-true value. A natural way to evaluate the corresponding asymptotic bias is to run a Monte Carlo simulation with a large sample size. Since our algorithm is very fast, this can be done at little cost.

The only covariates in this simulation are 1 and the logarithm of the price $x_{jt} = \log p_{jt}$. The coefficient of $\boldsymbol{x}$ is random: it depends on a micromoment $d_i = \bar{d}_t + \tilde{d}_i$ which is normally distributed,

$$\tilde{d}_i \simeq N(0, \tau^2) \text{ and } \bar{d}_t \simeq N(0, 1),$$

and on a random shock $\varepsilon_i \simeq N(0, \sigma_0^2)$. In our previous notation,

$$\begin{aligned}\bar{\boldsymbol{X}}\bar{\boldsymbol{\Pi}} &= \beta_0 + \beta_1 x_{jt} + \pi_0 x_{jt}\bar{d}_t \\ \boldsymbol{X}\boldsymbol{\nu} &= x_{jt}(\pi_0 \tilde{d}_i + \varepsilon_i).\end{aligned}$$

Since the variance of $\bar{d}_i$ is the same on each market, the random term ν_i is distributed as $N(0, s_0^2)$, where $s_0^2 = \sigma_0^2 + \pi_0^2\tau^2$. This allows us to use the first approach described in Section 3.3.2: we will estimate π_0^2 and s_0^2 and recover an estimate of σ_0^2 by subtraction (remember that the distribution of d_i is observed, so that τ obtains directly from the data.)

The product effects $\boldsymbol{\xi}$ and the values of the instrument z_{jt} are iid draws from a standard centered normal $N(0, 1)$. The covariates $\boldsymbol{x}$ are generated as follows:

$$x_{jt} = \rho_{xz} z_{jt} + \sqrt{1 - \rho_{xz}^2}(\rho_{x\xi}\xi_{jt} + \sqrt{1 - \rho_{x\xi}^2}\zeta_{jt}).$$

where the values of ζ_{jt} are iid draws from $N(0, 1)$, independent of z_{jt} and ξ_{jt}.

This formulation implies that the R^2 of a regression of x_{jt} on z_{jt} (resp. on ξ_{jt}) is ρ^2_{xz} (resp. $(1-\rho^2_{xz})\rho^2_{x\xi}$). Therefore ρ^2_{xz} measures the strength of the instruments, and $\rho^2_{x\xi}$ is a proxy for the degree to which the price is endogeneous.

We ran a variety of simulations with $T = 5,000$ markets (close enough to infinity that the results do not change), with different parameter values and numbers of products from $J = 1$ (the mixed logit) to $J = 100$. We took the value of the standard error of the micromoment to d_i to be $\tau = 0.5$ and the strength of the instruments to be $\rho^2_{xz} = 0.5$; and we imposed $\beta_1 = 1$ for the coefficient of $\boldsymbol{x}$.

We tried all combinations of the following:

- a scenario in which we set β_0 so that the market share S_0 of the zero good fluctuates around 0.5, and one in which it fluctuates around 0.9
- a model without a micromoment ($\pi_0 = 0$) and several models with a micromoment ($\pi_0 = 0.25, 0.5, 1.0$)
- a model in which price is exogenous ($\rho_{x\xi} = 0$, in which case we use $\boldsymbol{x}$ as an instrument) and one in which it is endogeneous ($\rho^2_{x\xi} = 0.5$, with instruments $\boldsymbol{z}$).

Since ours is a small-σ approximation, we used a number of values for the variance of $\boldsymbol{\varepsilon}$: from $\sigma_0^2 = 0$ to $\sigma_0^2 = 2$. Since x_{jt} has unit variance and $\beta_1 = 1$, the R^2 of a regression of mean utilities on their covariates x_{jt} and $x_{jt}\bar{d}_t$ would be $(1+\pi_0^2)/(1+\pi_0^2(1+\tau^2)+\sigma_0^2)$. This expression decreases with σ_0^2; it decreases with π_0 iff $\sigma_0^2 \leqslant 1+\tau^2$. The R^2 varies widely across our simulation scenarii, from a minimum of 0.33 (for $\sigma_0 = 2$ and $\pi_0 = 0$) to close to 1 (for $\sigma_0 = 0$ and $\pi_0 = 0$).

5.1 Asymptotic Bias

In these very large samples, $\hat{\boldsymbol{\theta}}_2$ is a very good approximation of the pseudo-true value $\boldsymbol{\theta}_2$. Therefore the asymptotic bias of our 2SLS estimator must be close to $(\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta_0})$. We measured it in all of our siumulations, along with the asymptotic bias from the two-step "corrected" estimator we described in Section4.4.2. In addition, we computed the semiparametric efficiency bound for the exact BLP model; that is, the asymptotic variance of the most efficient estimator[16] given the moment conditions $E(\boldsymbol{\xi}_\infty|\boldsymbol{Z}) = \mathbf{0}$.

Figure 2 plots our results for a very simple model: S_0 close to 0.9, price is exogenous, there is no micromoment, and only $J = 5$ products. The three subpanels of Figure 2a plot the pseudo-true values for the three elements of $\boldsymbol{\theta}_2$ (in red) and of the corrected estimator (in green), along with the true values of $\boldsymbol{\theta}_0$ (in black). For comparison, the dashed lines plot the bounds of the 95% confidence interval for the efficient BLP estimator when the number of markets is $T = 100$.

As expected, the asymptotic bias of our estimators increases with the true value of σ. Still, even for $\sigma_0 = 2$, our 2SLS estimator of β_0 and β_1 stays well inside the 95% confidence bounds for $T = 100$; the corrected estimator does even better.

The 2SLS estimators of the variance σ^2 are biased downwards for the larger values of σ_0^2. This is not surprising as our approximation neglects the higher-order moments, which matter more as σ_0 grows. The Newton–Raphson correction cuts the bias by about half; it keeps the asymptotic bias within the 95% confidence intervals for 100 markets over the whole range of values of σ_0^2.

Since price responses are a major parameter of interest in empirical industrial organization, Figure 2b plots the mean and the dispersion across markets of the estimated semi-elasticities

$$\frac{\partial \log S_{jt}}{\partial x_{kt}}(\hat{\boldsymbol{\theta}}_2).$$

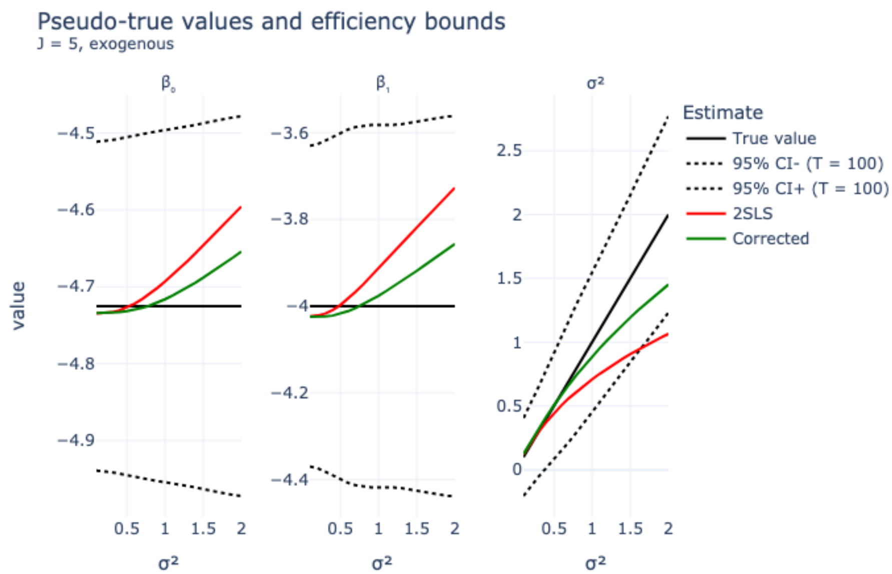
We show both the own price elasticity ($j = k = 1$) and the cross price-elasticity ($j = 1, k = 2$)[17]. Both elasticities are computed using the approximate model, and at the pseudo-true values. The black line shows the elasticities at the true parametre values, for the exact BLP model. Our estimates seem to be very reliable as long as σ_0^2 does not become too large. Once again, the correction does a very good job of reducing the (small) bias for the cross-price elasticity.

Going to the other end of the spectrum, we now add a micromoment; we make the price endogeneous; and we consider markets with $J = 100$ products. Figure 3 now has two rows, for the smallest and largest values of π_0; and four subpanels on the left

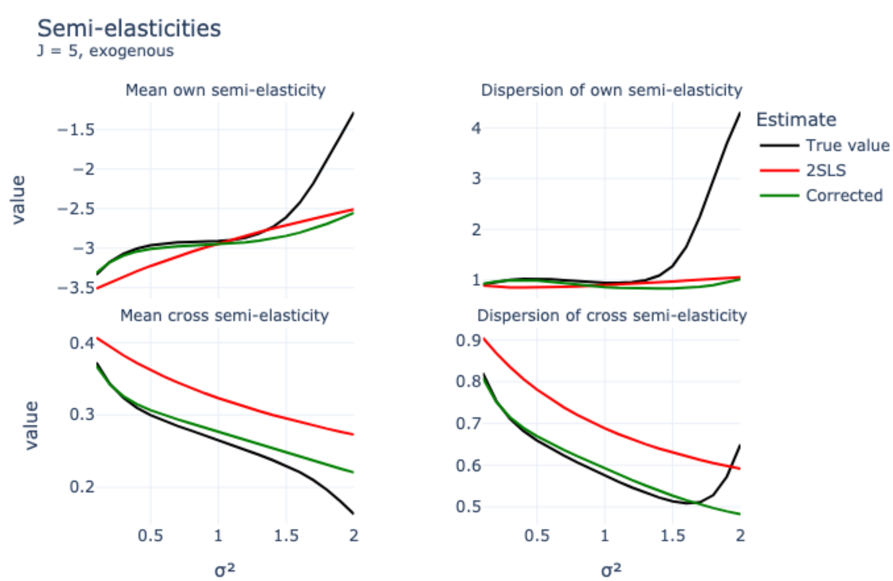
[16]This is simply the standard BLP estimator with the optimal instruments and the efficient weighting matrix.

[17]Since the model is symmetric across products, this choice of product indices is without loss of generality

Figure 2: Exogeneous price, no micromoment, 5 products



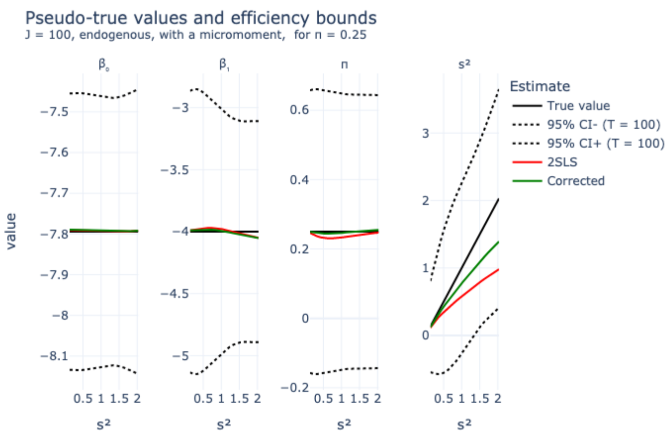
(a) Pseudo-true values

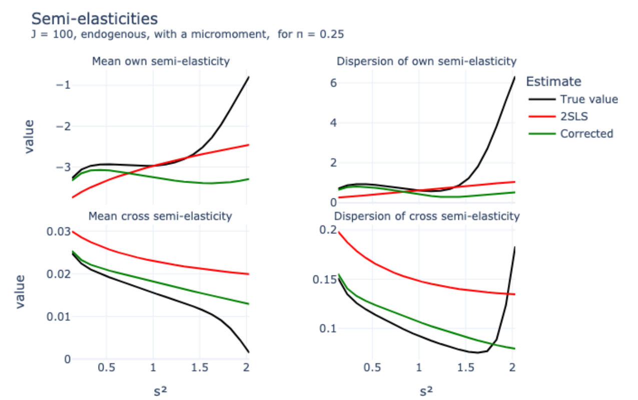


(b) Semi-elasticities

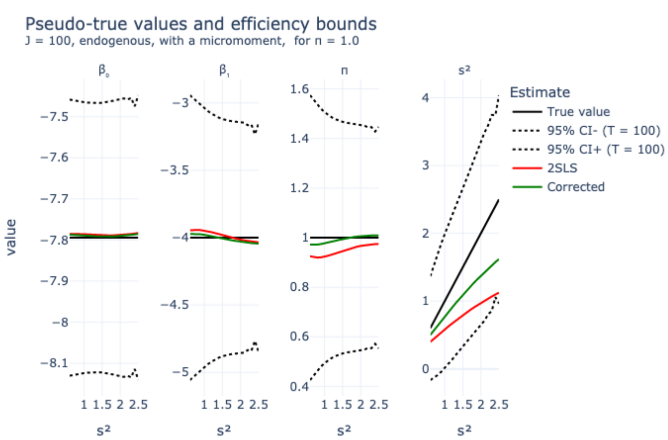
Figure 3: Endogeneous price, with a micromoment, 100 products

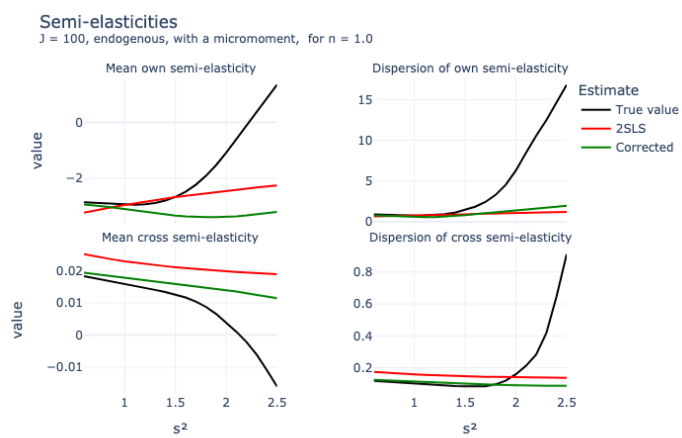
(a) $\pi_0 = 0.25$





(b) $\pi_0 = 1.0$





side as we estimate π_0 and $s_0^2 = \sigma_0^2 + \pi_0^2\tau^2$. There are obvious changes: since there are many more products, the estimates on β_0 and β_1 are very close to the true values, and the cross-price responses are smaller. Beyond that, the patterns in this figure are remarkably similar to those in Figure 2. This is constant across the many simulation runs that we did; they are all available online as an interactive Streamlit™ app at https://share.streamlit.io/bsalanie/frac.py/main/show_paper_plots.py.

6 Monte Carlo Analysis of the Small-Sample Performance of our Estimators

This section presents the results of a Monte Carlo study that explores the small-sample properties of our estimator when applied to a realistic empirical IO dataset. We compare the finite sample performance of our estimator to the one computed using the mathematical programming with equilibrium constraints (MPEC) approach recommended by Dubé, Fox and Su (2012). We adopt their basic set-up, except that we require our market shares and product prices to be result of a price-setting Nash equilibrium conditional on the realizations of the unobservables on the demand side (the unobserved product characteristics and the random preference parameters) and the supply side (the unobserved components of marginal cost).

6.1 The Data-generating Process

6.1.1 The Demand Side

We study a standard static aggregate discrete choice random coefficients demand system with $T = 50$ markets and $J = 25$ products in each market, and three observed product characteristics in addition to the price. Each product is characterized by the vector $(\boldsymbol{X}'_{jt}, \xi_{jt}, p_{jt})'$, where $\boldsymbol{X}_{jt}$ is a 3×1 vector of exogenous observable attributes of product $j = 1, 2, \ldots, J$ in market t, and p_{jt} is the price of product j in market t, which is endogenous. We assume that the ξ_{jt} are drawn independently from a $N(0, \sigma_\xi^2)$ distribution. We define the following market-specific variables: $\boldsymbol{X}_t = (\boldsymbol{X}'_{1t}, \ldots, \boldsymbol{X}'_{Jt})'$, $\boldsymbol{\xi}_t = (\xi_{1t}, \xi_{2t}, \ldots, \xi_{Jt})'$, and $\boldsymbol{p}_t = (p_{1t}, p_{2t}, \ldots, p_{Jt})'$.

The conditional indirect utility of consumer i in market t from purchasing product j is

$$\beta_0 + \boldsymbol{X}'_{jt}\boldsymbol{\beta}^x_i - \beta^p_i p_{jt} + \xi_{jt} + u_{ijt}$$

where the u_{ijt} are independently and identically distributed Type I extreme value random variables. The utility of the $j = 0$ good, the "outside" good, is equal to u_{i0t}. The vector $\boldsymbol{\beta}_i = (\beta^x_{i1}, \beta^x_{i2}, \beta^x_{i3}, \beta^p_i)'$ is assumed to be drawn from a 4-dimensional normal distribution with mean $(\bar{\beta}^x_1, \bar{\beta}^x_2, \bar{\beta}^x_3, \bar{\beta}^p)$ and a diagonal variance-covariance matrix with diagonal elements $(\sigma^2_1, \sigma^2_2, \sigma^2_3, \sigma^2_p)$.

We collect all of the demand parameters into the vector

$$\boldsymbol{\theta}_D = (\beta_0, \bar{\beta}^x_1, \bar{\beta}^x_2, \bar{\beta}^x_3, \bar{\beta}_p, \sigma^2_1, \sigma^2_2, \sigma^2_3, \sigma^2_p)'.$$

Consistent with the experimental design in Dubé, Fox and Su (2012), we generate the values of $\boldsymbol{X}_t$, $\boldsymbol{\xi}_t$ as follows. We draw $\boldsymbol{X}_t$ for all markets $t = 1, 2, \ldots, T$ and all products $j = 1, 2, \ldots, T$ independently from a 3-variate normal distribution with zero mean and variance-covariance matrix

$$\begin{pmatrix} 1 & -0.8 & 0.3 \\ -0.8 & 1 & 0.3 \\ 0.3 & 0.3 & 1 \end{pmatrix}$$

To compute the market shares for the J products, we start from the probability that consumer i with random preferences $\boldsymbol{\beta}_i$ purchases good j in market t:

$$s_{ijt}(\boldsymbol{X}_t, \boldsymbol{p}_t, \boldsymbol{\xi}_t | \boldsymbol{\beta}_i) = \frac{\exp(\beta^0 + \boldsymbol{X}'_{jt}\boldsymbol{\beta}^x_i - \beta^p_i p_{jt} + \xi_{jt})}{1 + \sum_{k=1}^{J} \exp(\beta_0 + \boldsymbol{X}'_{kt}\boldsymbol{\beta}^x_i - \beta^p_i p_{kt} + \xi_{kt})}.$$

We compute the observed market shares for all goods in market t by drawing $n_s = 1,000$ draws (ζ_{skt}) from four independent $N(0,1)$ random variables and constructing $1,000$ draws from $\boldsymbol{\beta}_i$ given $\boldsymbol{\theta}$ as follows:

$$\beta^x_{skt} = \bar{\beta}^x_k + \sigma_k \zeta_{ikt} \text{ for } k = 1, 2, 3, p.$$

We then use these draws to compute the observed market share of good j in market t for any vector of prices for market t, p_t, as:

$$S_{jt}(\boldsymbol{X}_t, \boldsymbol{p}_t, \boldsymbol{\xi}_t | \boldsymbol{\theta}) = \frac{1}{n_s} \sum_{i=1}^{n_s} s_{ijt}(\boldsymbol{X}_t, \boldsymbol{p}_t, \boldsymbol{\xi}_t | \boldsymbol{\beta}_{st})$$

given the vectors $\boldsymbol{X}_t$, $\boldsymbol{p}_t$, and $\boldsymbol{\xi}_t$ for each market t.

6.1.2 The Supply Side

Instead of a reduced form price equation that induces correlation between p_{jt} and ξ_{jt} as in Dubé, Fox, and Su (2012), we specify a cost side of the market and solve the first-order conditions for profit-maximization to compute the market clearing prices.

Let the marginal cost of good j in market t equal:

$$\mathrm{mc}_{jt} \;=\; \exp(\gamma_0 + \boldsymbol{z}'_{jt}\boldsymbol{\gamma} \;+\; \omega_{jt})$$

where as in Dubé, Fox and Su (2012), we generate the values of a vector of three instruments $\boldsymbol{Z}_{jt}$ independently across markets and products from another 3-variate centered normal distribution with variance-covariance matrix

$$\begin{pmatrix} 1 & 0.5 & -0.3 \\ 0.5 & 1 & 0.3 \\ -0.3 & 0.3 & 1 \end{pmatrix}.$$

We model price equilibrium as in Section 3.3.3, assuming for simplicity that each product j is produced by a specialized firm j. Solving the J first-order conditions for the J prices for market t yields the equilibrium vector of prices $\boldsymbol{p}_t^e$ in this market, and the corresponding market shares $S_{jt}(\boldsymbol{X}_t, \boldsymbol{p}_t^e, \boldsymbol{\xi}_t \mid \theta)$. For a specified value of the parameter vector θ. Following this process for $T = 50$ markets yields the dataset for one Monte Carlo sample.

6.1.3 Parameter Configurations

All of our simulations have mean demand coefficients

$$(\beta_0, \bar{\beta}_1^x, \bar{\beta}_2^x, \bar{\beta}_3^x, \bar{\beta}_p) = (7, 1.5, 1.5, 0.5, 4);$$

$\gamma_0 = 0.5$; and $\omega_{jt} \;\sim\; N(0, 0.2)$.

We run 12 scenarios obtained by setting

- two values for the variance of the unobserved product characteristics, $\sigma_\xi^2 = \mathrm{Var}(\xi) = 0.5, 1$
- three values for the vector of variances of the random coefficients

$$(\sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_p^2) = (0.2, 0.2, 0.2, 0.1), (0.5, 0.5, 0.5, 0.25), (1.0, 1.0, 1.0, 0.5)$$

- and two sets of values for the parameters product-specific marginal cost functions: $\boldsymbol{\gamma} = (0.1, -0.1, -0.1)'$ and $\boldsymbol{\gamma} = (0.2, -0.2, -0.2)'$.

In all of our simulation runs, we use the same 36 functions of the observed product characteristics $\boldsymbol{x}_{jt}$ and cost shifters $\boldsymbol{z}_{jt}$ to generate moment conditions. They are

$$1, x_{kjt}, x_{kjt}^2, x_{kjt}^3, (k = 1, 2, 3),$$

$$x_{1jt}x_{2jt}, x_{1jt}x_{j3t}, x_{2jt}x_{3jt}, x_{1jt}x_{2jt}x_{3jt}, z_{kjt}, z_{kjt}^2, z_{kjt}^3, (k = 1, 2, 3)$$

$$z_{1jt}z_{2jt}, z_{1jt}z_{3jt}, z_{2jt}z_{3jt}, z_{1jt}z_{2jt}z_{3jt}, z_{kjt}x_{1jt}, z_{kjt}x_{2jt}, z_{kjt}z_{3jt}, (k = 1, 2, 3)$$

Let $\boldsymbol{W}$ denote this $(J \times T) \times 36$ matrix of instruments. In our case $J \times T = 1,250$ since $J = 25$ and $T = 50$.

For both MPEC and FRAC, we estimate the 9 parameters in $\boldsymbol{\theta}_D$. For FRAC, we also estimate the 5 parameters of the supply side $\boldsymbol{\theta}_S = (\gamma_0, \gamma_1, \gamma_2, \gamma_3, \sigma_\omega^2)$.

To each of our simulation scenarii corresponds a breakdown of the variance of the endogeneous variables (market shares and prices). To apprehend it, we report simple variance decompositions. For prices, we use a linear regression to isolate the part of the variation that is explained by the instruments:

$$Vp_{jt} = VE\left(p_{jt}|W_{jt}\right) + EV(p_{jt}|W_{jt})$$

and we further break down the part that is not explained into the part that is explained by the demand-side and supply-side product effects ξ_{jt} and ω_{jt} and the unexplained part.

We use a similar method to decompose the variance in market shares into the part that is explained by the covariates $\boldsymbol{x}_{jt}$ and the instrumented price $E(p_{jt}|\boldsymbol{W}_{jt})$; the part that is explained by the randomness in the coefficients $\boldsymbol{\beta}_i$; and the part that is explained by the product effect ξ_{jt}.

We give more information on the computation of these statistics in Appendix D.

6.2 FRAC Estimation

To estimate the demand parameters $\boldsymbol{\theta}_D$ by FRAC, we construct the 4 artificial regressors K_{jt}^k for $k = 1, 2, 3, p$ in each market and for each product, and we run the two-stage least squares regression of the $J \times T$ observations $\log(S_{jt}/S_{0t})$ on the nine regressors

$$(1, x_{1jt}, x_{2jt}, x_{3jt}, p_{jt}, K^1_{jt}, K^2_{jt}, K^3_{jt}, K^p_{jt})$$

with the 36 instruments in $\boldsymbol{W}_t$.

To estimate the supply side coefficients $\boldsymbol{\theta}_S$, we then proceed as described in Section 3.3.3: using our estimator of $\boldsymbol{\theta}_D$, we obtain $\boldsymbol{\xi}$ as the solution of the market shares equation; we define the $J \times J$ diagonal matrix $\boldsymbol{\Delta}$ whose (j, j) element is equal to $-(\partial S_j)/(\partial p_j)$ for these estimates of $\boldsymbol{\theta}_D$ and $\boldsymbol{\xi}$. Then we rewrite the J first-order conditions as

$$\ln(p_{jt} - b_{jt}(\boldsymbol{p}_t, \boldsymbol{x}_t, \boldsymbol{\xi}_t \mid \boldsymbol{\theta})) \;=\; \gamma_0 + \boldsymbol{z}'_{jt}\gamma \;+\; \omega_{jt},$$

where $\boldsymbol{b} = \boldsymbol{\Delta}^{-1}\boldsymbol{S}$. Finally, we use an OLS regression of the $J \times T$ (generated) observations $\ln(p_{jt} - b_{jt}(\boldsymbol{p}_t, \boldsymbol{x}_t, \hat{\boldsymbol{\xi}}_t \mid \hat{\boldsymbol{\theta}}_D))$ on the vector $\boldsymbol{z}_{jt}$ to obtain an estimator of $\boldsymbol{\theta}_S = (\gamma_0, \boldsymbol{\gamma}', \sigma^2_\omega)'$.

The supply equation for product j and market t can be combined with demand equation for product j and market t to construct a 3SLS estimator of $\boldsymbol{\theta}_D$ and $\boldsymbol{\theta}_S$ that accounts for potential contemporaneous correlation between ξ_{jt} and ω_{jt}.

6.3 MPEC Estimation

While FRAC estimation only requires 2SLS and OLS, implementing MPEC involves solving a nonlinear optimization problem subject to nonlinear equilibrium constraints based on simulated market shares. As shown in Dubé, Fox and Su (2012), the MPEC approach consists in minimizing

$$\boldsymbol{\eta}'\boldsymbol{W}(\boldsymbol{W}'\boldsymbol{W})^{-1}\boldsymbol{W}'\boldsymbol{\eta}$$

with respect to $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$, subject to the “equilibrium constraints”

$$\boldsymbol{s}(\boldsymbol{\eta}, \boldsymbol{\theta}) = \boldsymbol{S}$$

where $\boldsymbol{S}$ is the vector of observed market shares and $\boldsymbol{s}$ represents the simulated market shares

$$s_{jt}(\boldsymbol{\eta}, \boldsymbol{\theta}) = \frac{1}{N_s} \sum_{s=1}^{N_s} \frac{\exp(\beta_{s0} + \beta^x_{s1} x_{1j} + \beta^x_{s2} x_{2j} + \beta^x_{s3} x_{3j} - \beta^p_s p_{jt} + \eta_{jt})}{1 + \sum_{k=1}^{J} \exp(\beta_{s0} + \beta^x_{s1} x_{1k} + \beta^x_{s2} x_{2k} + \beta^x_{s3} x_{3k} - \beta^p_s p_{kt} + \eta_{kt})}$$

and the ($\boldsymbol{\beta}_s$) vectors are random draws from the following normal distribution:

$$N\left(\begin{pmatrix}\theta_1\\ \theta_2\\ \theta_3\\ \theta_4\\ \theta_5\end{pmatrix}, \begin{pmatrix}0 & 0 & 0 & 0 & 0\\ 0 & \theta_6 & 0 & 0 & 0\\ 0 & 0 & \theta_7 & 0 & 0\\ 0 & 0 & 0 & \theta_8 & 0\\ 0 & 0 & 0 & 0 & \theta_9\end{pmatrix}\right).$$

Note that β_{0s} (like β_0) is not allowed to be random. For purposes of estimation, we set $N_s = 1,000$.

For each Monte Carlo simulation, we start the optimization with true values for $\boldsymbol{\theta}$, and a vector of zeros for the $\boldsymbol{\eta}$ vector. Clearly, these starting values are not feasible for empirical researchers; we use them to maximize the chances that the MPEC estimation will converge to a solution.

Demote $\boldsymbol{I}_2$ the (2×2) identity matrix. The MPEC approach with supply-side moments minimizes

$$\boldsymbol{\nu}'(\boldsymbol{I}_2 \otimes \boldsymbol{W}(\boldsymbol{W}'\boldsymbol{W})^{-1}\boldsymbol{W}'\boldsymbol{\nu}$$

with respect to $\boldsymbol{\nu} = (\boldsymbol{\eta}', \boldsymbol{\omega}')'$, $\boldsymbol{\theta}_D$, and $\boldsymbol{\theta}_S$ subject to two sets of constraints:

$$\boldsymbol{s}(\boldsymbol{\eta}, \boldsymbol{\theta}) = \boldsymbol{S}$$

and

$$\omega_{jt} \; = \; ln(p_{jt} - b_{jt}(\boldsymbol{p}_t, \boldsymbol{X}_t, \boldsymbol{\eta}_t \mid \boldsymbol{\theta})) \; - \; \boldsymbol{z}'_{jt}\boldsymbol{\gamma}$$

for all j and t.

6.4 Using FRAC for Variable and Random Coefficient Selection

The researcher often has many potential product characteristics to consider when estimating a demand system. Our estimation procedure can be used both to select variables and to decide which should have random coefficients. To do this, we change our simulation to consider tests of three hypotheses: whether $\bar{\beta}_1^x = 0$; whether $\sigma_1^2 = 0$; and the joint test of $\bar{\beta}_1^x = 0$ and $\sigma_1^2 = 0$. To compute the power functions for these tests, we follow the procedure described above for generating equilibrium prices and market shares given product characteristics and cost shifters, with the same distributions.

We also consider a test that the coefficient of price is non-random: $\sigma_p^2 = 0$. For all of these tests we compute the empirical frequency of rejection of each null hypothesis when it is true,and for economically plausible deviations from the null hypothesis. For true parameter values, we chose: $\boldsymbol{\gamma} = (0.1, -0.1, -0.1)'$,

$$(\beta_0, \bar{\beta}_1^x, \bar{\beta}_2^x, \bar{\beta}_3^x, \bar{\beta}_p) = (7, \bar{\beta}_1^x, 1.5, 0.5, 4)$$

$$(\sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_p^2) = (\sigma_1^2, 0.5, 0.5, 0.25)$$

and $\sigma_\xi^2 = 0.5$ and $\sigma_\omega^2 = 0.2$. Table 1 contains the true values for $\bar{\beta}_1^x$ and $\sigma_1^2 = 0$.

$\bar{\beta}_1^x$	σ_1^2
0	0
0.25	0
0.75	0
1.5	0
1.5	0.1
1.5	0.2
1.5	0.5

Table 1: Tests on $\boldsymbol{\beta}_{i1}^x$

The true values for σ_p^2 are $0, 0.05, 0.10, 0.25$. We keep the same values for as above with $\bar{\beta}_1^x = 1.5$ and $\sigma_1^2 = 0.5$

We perform these tests using our estimator with White (1982) model misspecification robust standard error estimates applied to both our 2SLS estimates and bias-corrected estimates.

6.5 Simulation Results

6.5.1 Estimates

On each plot, the dashed vertical purple line represents the true value of the parameter. We show four estimators: MPEC, "FRAC(D)" and "FRAC(S)" for 2SLS applied to the demand model then to the supply model, and 3SLS for the three-stage least squares estimator. In all of our simulations, we found that the three-stage least squares estimate is almost identical to the 2SLS estimate, both for demand and supply parameters.

When the randomness of the coefficients (as measured by the parameters $(\sigma_1^2, \sigma_2^2, \sigma_2^3, \sigma_p^2)$) is small, our FRAC estimators perform as well as MPEC for the mean values of the random coefficients, and actually better for the variances. Figure 4 give a representative example, for $(\sigma_1^2, \sigma_2^2, \sigma_2^3, \sigma_p^2) = (0.5, 0.5, 0.5, 0.25)$, $\sigma_\xi^2 = 1.0$, and $\gamma = (0.1, -0.1, -0.1)$. In this scenario, 15% of the variance of prices is explained at by the unobserved product effects, and 80% by the covariates and instruments. 45% of the variance of market shares is explained by the covariates and instruments and 40% by the random variation in consumer preferences.

The only scenarii in which MPEC outperforms FRAC are, not surprisingly, those where the coefficients of demand have a large variance. Figure 5 is drawn for $(\sigma_1^2, \sigma_2^2, \sigma_2^3, \sigma_p^2) = (1.0, 1.0.1.0, 0.5)$, $\sigma_\xi^2 = 0.5$, and $\gamma = (0.2, -0.2, -0.2)$. In this scenario, 50% of the variance of prices is explained at by the unobserved product effects, and 40% by the covariates and instruments. 70% of the variance of market shares is explained by the covariates and instruments and 20% by the random variation in consumer preferences.

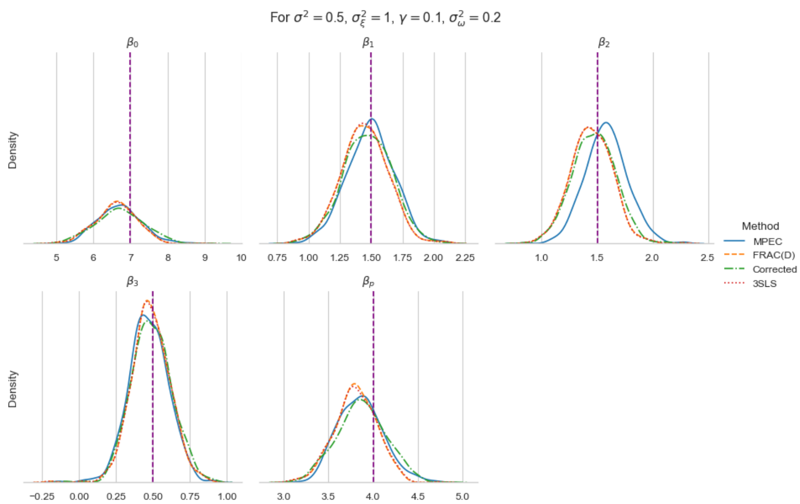
The FRAC estimates of the coefficients of supply are very reliable across all scenarii. Figure 6 plots their distribution in teh two scenarii considered above.

6.5.2 Tests

A good test should have p-values distributed uniformly over $[0, 1]$ under the null, and moving towards a mass at 1 under the alternative. Figure 7 plots the empirical cdf of the p-values of our tests for $\beta 1 = 0$ using FRAC in the small randomness scenario. The dashed vertical and horizontal lines correspond to $p = 0.05$. Whether we use second-order FRAC or the corrected version, our tests appear to have excellent

Figure 4: Distribution of the Demand Estimates for Small Randomness

(a) Means of the Random Demand Coefficients



(b) Variances of the Random Demand Coefficients

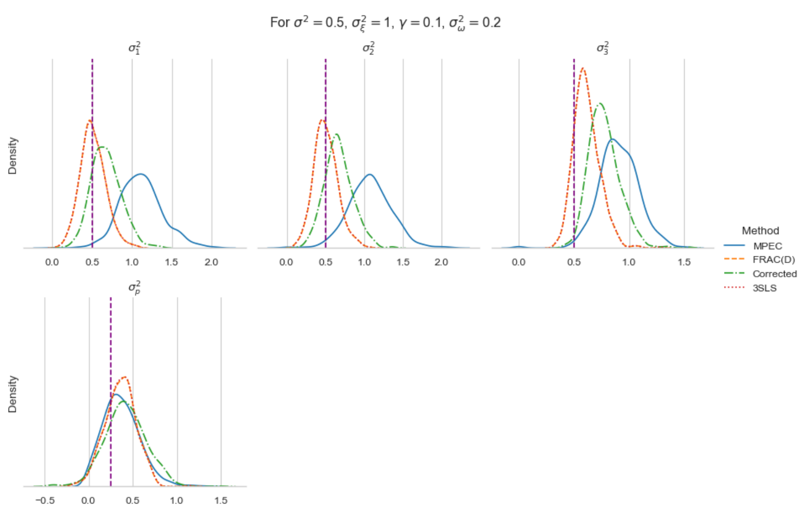
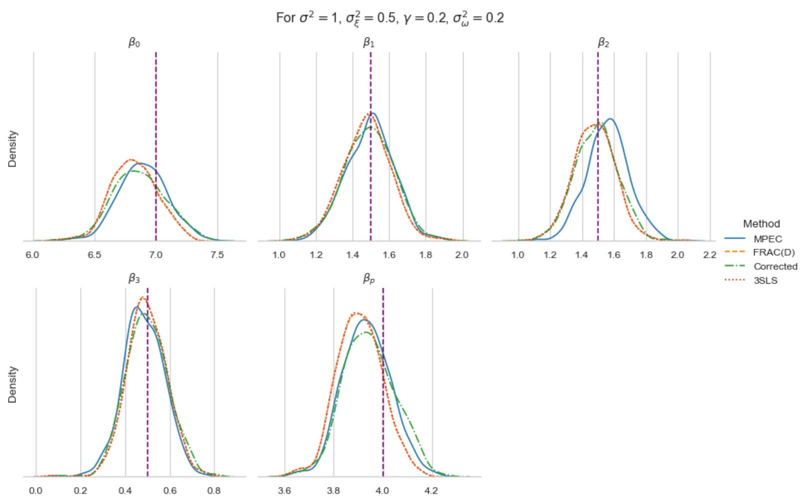


Figure 5: Distribution of the Demand Estimates for Large Randomness

(a) Means of the Random Demand Coefficients



(b) Variances of the Random Demand Coefficients

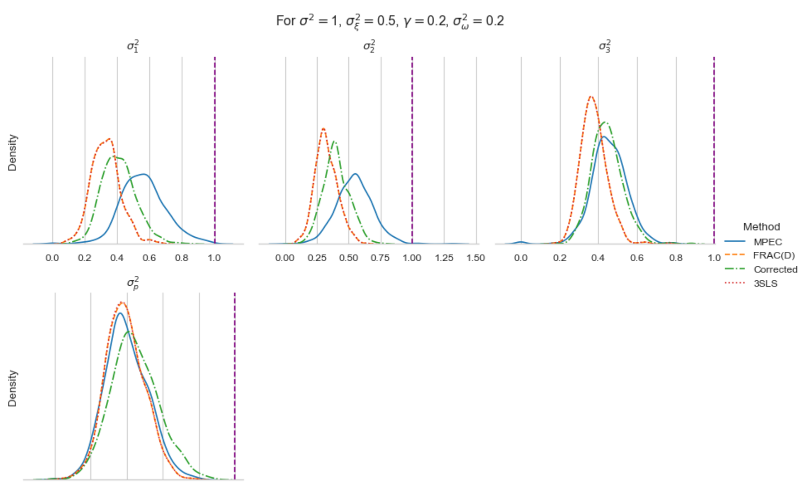
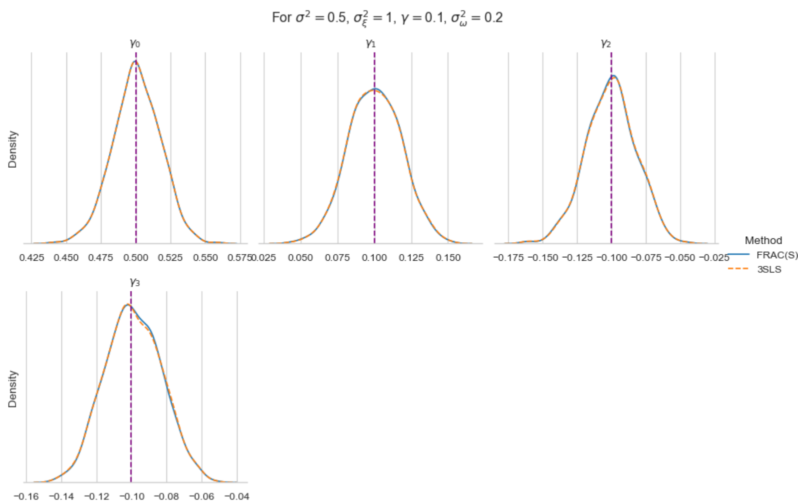


Figure 6: FRAC Estimates of the Coefficients of Supply

(a) Small Randomness



(b) Large Randomness

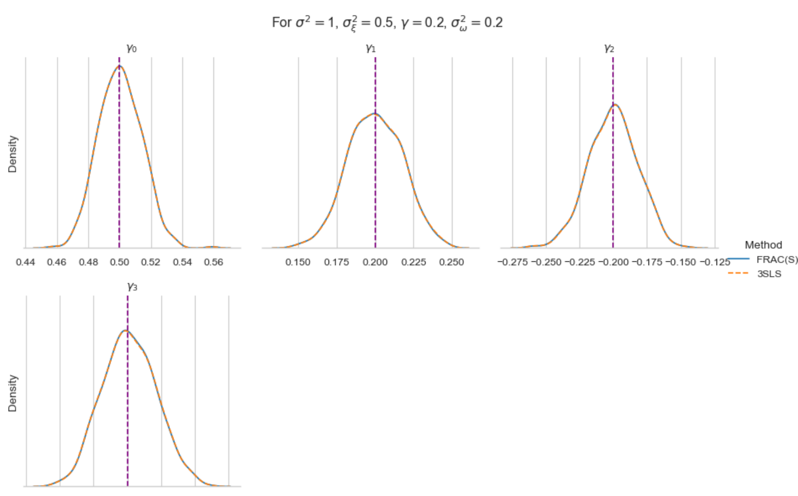
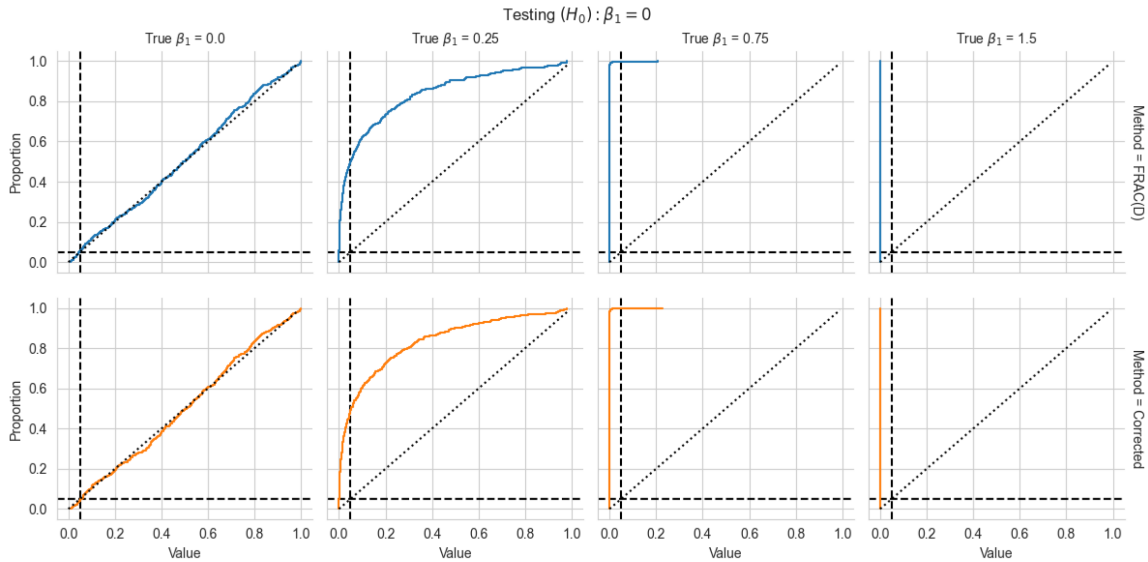


Figure 7: Test that $\beta_1 = 0$



properties.

Figures 8 and 9 plot the empirical cdf of the p-values of our tests for $\sigma_1 = 0$ and $\sigma_p = 0$, respectively. The fact that the cdf is under the diagonal for small p-values for small values of the true σ_1 or σ_p implies that our tests lack power to detect small random components. On the other hand, they seem to have sufficient power to detect economically significant randomness.

Concluding Comments

Our FRAC estimation procedure applies directly to the random coefficients demand models commonly used in empirical industrial organization. For the most part, our Monte Carlo results confirm the findings from the expansions. The 2SLS approach yields reliable estimates of the parameters of the model and of economically meaningful quantities such as price elasticities; and it does so at a very minimal cost. It does not require any assumption on the higher-order moments of the distribution of the random coefficients. In addition, it provides straightforward tests that help in

Figure 8: Test that $\sigma_1 = 0$

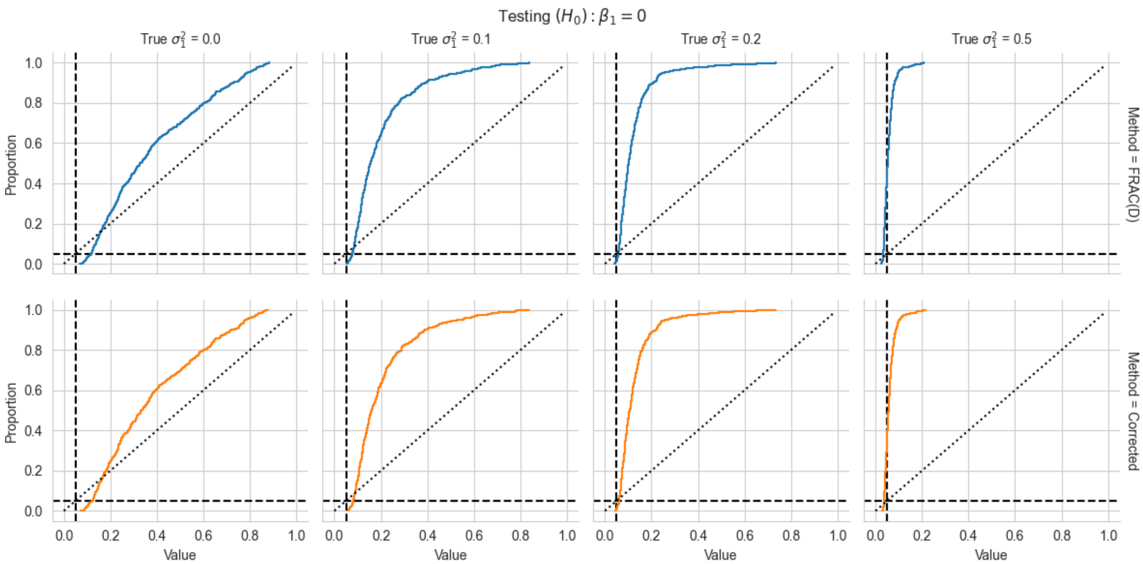
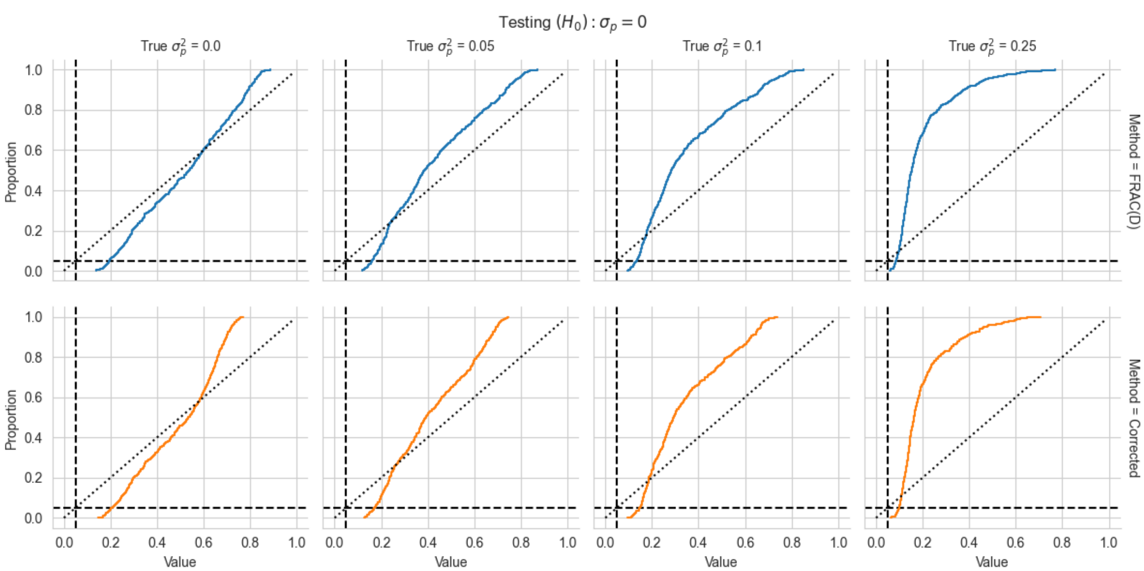


Figure 9: Test that $\sigma_p = 0$



variable selection, especially as a guide to determine which coefficients in the demand system should be modeled as random. A simple correction improves the estimates if one is willing to specify the distribution of the coefficients further.

References

Amemiya, T. (1975), "The nonlinear limited-information maximum-likelihood estimator and the modified nonlinear two-stage least-squares estimator," *Journal of Econometrics*, 3, 375–386.

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

Berry, S. (1994), "Estimating Discrete Choice Models of Product Differentiation", *Rand Journal of Economics*, 23, 242-262.

Berry, S., Levinsohn, J., and A. Pakes (1995), "Automobile Prices in Market Equilibrium", *Econometrica*, 60, 889-917.

Brand, J. (2021a), "Differences in Differentiation: Rising Variety and Markups in Retail Food Stores", mimeo UT Austin.

Brand, J. (2021b), `FRAC.jl`: A Julia package available at https://github.com/jamesbrandecon/FRAC.jl.

Compiani, G. (2021), "Market Counterfactuals and the Specification of Multi-Product Demand: A Nonparametric Approach", mimemo Chicago Booth, forthcoming in *Quantitative Economics*.

Chesher, A. (1991), "The Effect of Measurement Error", *Biometrika*, 78, 451–462.

Chesher, A. and J. Santos-Silva (2002), "Taste Variation in Discrete Choice Models", *Review of Economic Studies*, 69, 147–168.

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

Cragg, J.G. (1983), "More Efficient Estimation in the Presence of Heteroscedasticity of Unknown Form", *Econometrica*, 51, 3, 751–763.

Dubé, J.-P., Fox, J., and C.-L. Su (2012), "Improving the Numerical Performance of BLP Static and Dynamic Discrete Choice Random Coefficients Demand Estimation", *Econometrica*, 80, 2231-2267.

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

Gandhi, A. and J.-F. Houde (2020), "Measuring Substitution Patterns in Differentiated Products Industries", NBER WP 26375.

Harding, M. and J. Hausman (2007), "Using a Laplace Approximation to Estimate the Random Coefficients Logit Model by Nonlinear Least Squares", *International Economic Review*, 48, 1311–1328.

Horowitz, J. and L. Nesheim (2021), "Using Penalized Likelihood to Select Parameters in a Random Coefficients Multinomial Logit Model", *Journal of Econometrics*, 222, 44–55.

Kadane, J. (1971), "Comparison of k-class Estimators when the Variance is Small", *Econometrica*, 39, 723–737.

Ketz, P. (2019), "On Asymptotic Size Distortions in the Random Coefficients Logit Model", *Journal of Econometrics*, 212, 413–432.

Kristensen, D. and B. Salanié (2017), "Higher-order Properties of Approximate Estimators", *Journal of Econometrics*, 189–208.

Lee, J. and K. Seo (2015), "A computationally fast estimator for random coefficients logit demand models using aggregate data", *Rand Journal of Economics*, 46, 86-102.

Lu, Z., Shi, X., and J. Tao (2021), "Semi-Nonparametric Estimation of Random Coefficienta Logit Model for Aggregate Demand", mimeo.

Nevo A. (2001), "Measuring market power in the ready-to-eat cereal industry", *Econometrica* 69, 307-42.

Nevo, A., Turner, J., and J. Williams (2016), "Usage-Based Pricing and Demand for Residential Broadband", *Econometrica*, 84, 411–443.

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.

Robinson, P. (1988), "The Stochastic Difference between Econometric Statistics", *Econometrica*, 56, 531–548.

Su, C.-L. and K. Judd (2012), "Constrained Optimization Approaches to Estimation of Structural Models", *Econometrica*, 80, 2213-2230.

White, H. (1982), "Instrumental Variables Regression with Independent Observations", *Econometrica*, 50(2), 483-499.

A A Detailed Examination of the Mixed Binary Choice Model

The mixed binary choice model has $J = 1$ and

$$S_1 = E_{\boldsymbol{\beta}} F_U(\boldsymbol{X}_1 \boldsymbol{\beta})$$

where F_U is a cdf (the link funnction) and $\boldsymbol{\beta}$ is a vector of random coefficients.

A.1 The Mixed Logit

When U is distributed as a logistic, the model is simply a mixed logit. Applying Theorem 2 with $J = 1$ and using $S_0 + S_1 = 1$, we obtain

$$\boldsymbol{K}^1 = (1/2 - S_1)\boldsymbol{X}_1 \boldsymbol{X}_1'.$$

Therefore

$$\xi_1 = \log \frac{S_1}{S_0} - \boldsymbol{X}_1 \boldsymbol{\Pi} - \left(\frac{1}{2} - S_1\right) \operatorname{Tr} \boldsymbol{\Sigma} \boldsymbol{X}_1 \boldsymbol{X}_1' + O(\sigma^k)$$

where $k = 3$ in general, and $k = 4$ if the distribution of $\boldsymbol{\varepsilon}$ is symmetric around zero.

Let us focus for simplicity on the case when random variation in preferences is uncorrelated across covariates: $\boldsymbol{\Sigma}$ is the $n_X \times n_X$ diagonal matrix with elements Σ_{mm}. Then given instruments such that $E\left(\xi_1 | \boldsymbol{Z}\right) = 0$, the approximate model is

$$E\left(\log \frac{S_1}{S_0} - \boldsymbol{X}_1 \boldsymbol{\Pi} - \left(\frac{1}{2} - S_1\right) \sum_{m=1}^{n_X} \Sigma_{mm} X_{1m}^2 \;\Big|\; \boldsymbol{Z}\right) = 0. \tag{17}$$

A.1.1 Identification

The form of the estimating equation holds interesting insights about identification. Denote $\boldsymbol{\mathcal{X}} = (\boldsymbol{X}, \boldsymbol{K} = \left(\frac{1}{2} - S_1\right) \boldsymbol{X}_1^2)$ the natural and artificial regressors, where $\boldsymbol{X}_1^2$ is the vector with components X_{1m}^2. The optimal instruments are the nonparametric projections

$$\boldsymbol{Z}_2 = \left(E\left(\boldsymbol{\mathcal{X}} | \boldsymbol{Z}\right), \right)$$

Suppose for simplicity that ξ_1 is homoskedastic across markets. Then the asymptotic variance-covariance matrix of our estimator $\hat{\boldsymbol{\theta}}$ is given by the usual formula:

$$T\, V_{\text{as}}\hat{\boldsymbol{\theta}} \simeq TV(\xi_1)(\boldsymbol{Z}_2'\boldsymbol{Z}_2)^{-1}\dots$$

The identifying power of the (approximate) model crucially depends on the matrix $\boldsymbol{Z}_2'\boldsymbol{Z}_2$. Suppose for instance that the residual variation in the projected artificial regressor $E(\boldsymbol{K}_m|\boldsymbol{Z})$ is very well explained in a linear regression on the projected covariates $E(\boldsymbol{X}|\boldsymbol{Z})$ and the other $E(\boldsymbol{K}_n|\boldsymbol{Z})$. Then the estimate of Σ_{mm} will be very imprecise, and random taste variation on the characteristic X_{1m} is probably best left out of the model.

A.1.2 Higher-order terms

We show here how to compute the fourth-order term in the mixed logit model by hand. This will also illustrate how our method does not rely on distributional assumptions.

Assume that $\boldsymbol{\varepsilon}$ has a distribution that is symmetric around zero, and that its components are independent of each other with variances Σ_{mm} and fourth-order moments k_m. As before, we assume that Σ_{mm} is of order σ^2 and k_m is of order σ^4. We also assume that we can take expansions to order $L \geqslant 5$.

Since the distribution is symmetric, we already know that

$$\xi_1 = \log\frac{S_1}{S_0} - \boldsymbol{X}_1\boldsymbol{\Pi} - H_2\sigma^2 - H_4\sigma^4 + O(\sigma^6).$$

for some variables H_2 and H_4 that are deterministic functions of S_1 and $\boldsymbol{X}_1$.

We already know that

$$H_2\sigma^2 = (S_1 - 1/2)\sum_{m=1}^{n_X}\Sigma_{mm}x_{1m}^2;$$

specializing the result in Section 4.3 gives us

$$H_4\sigma^4 = \left(S_1 - \frac{1}{2}\right)\times \left(\left(\frac{1}{4} - 2S_1(1-S_1)\right)\left(\sum_{m=1}^{n_X}\Sigma_{mm}x_{1m}^2\right)^2 - \left(\frac{1}{12} - S_1(1-S_1)\right)\sum_{m=1}^{n_X}k_m x_{1m}^4\right).$$

This formula may not seem especially enlightening, but it exemplifies two important points. First, terms of higher orders can be computed without much difficulty. Second, each additional term adds information on lower-order moments (here $\Sigma_{mm} = E\varepsilon_m^2$), as well as on the moments of higher order (here $k_m = E\varepsilon_m^4$). The model remains linear in the highest-order moments; here for k_m we have new artificial regressors

$$\left(\frac{1}{2} - S_1\right)\left(\frac{1}{12} - S_1(1 - S_1)\right) x_{1m}^4.$$

On the other hand, the higher-order expansions introduce nonlinear functions of the lower-order moments, which are here quadratic functions of the variances:

$$\left(S_1 - \frac{1}{2}\right)\left(\frac{1}{4} - 2S_1(1 - S_1)\right)\left(\sum_{m=1}^{n_X} \Sigma_{mm} x_{1m}^2\right)^2,$$

and the model is not linear in these parameters any more. This could be dealt with in several ways: by nonlinear optimization (of a very simple kind), or by iterative methods. In any case, we will see in our simulations that stopping with the second-order expansion often gives results that are already very reliable.

Finally, while the estimator based on the second-order expansion has the same form for any (well-behaved) distribution, the estimator based on the fourth-order expansion above assumes symmetry: a skewed distribution would generate terms in σ^3. If the components of $\boldsymbol{\varepsilon}$ are independently distributed and have third moments $(s_1, \ldots, s_{n_X})$, then it is easy to see from Section 4.3 that an additional term

$$H_3\sigma^3 = \left(S_1(1 - S_1) - \frac{1}{6}\right) \sum_{m=1}^{n_X} s_m X_{1m}^3$$

enters the expansion. To test for skewness on covariate m, one could simply test that the regressor $\left(S_1(1 - S_1) - \frac{1}{6}\right) X_{1m}^3$ can be omitted.

Making more assumptions changes the form of the artificial regressors. To illustrate this, consider a mixed logit with one covariate only ($n_X = 1$). The expansion to order $2L$ can be written

$$\xi_1 = \log \frac{S_1}{S_0} - \beta X_1 + \sum_{k=1}^{L} t_k(S_1)\left(\Sigma_{11} X_1^2\right)^k + O(\sigma^{2L+2}).$$

Assume that $\boldsymbol{\varepsilon}$ has normal kurtosis. Then $k_1 = 3\Sigma_{11}^2$ and we find the simpler formula

$$t_2 = \alpha_4 = \left(\frac{1}{2} - S_1\right) S_1(1 - S_1).$$

It is easy to program a symbolic algebra system to compute even higher-order terms, given more distributional assumptions. This is how we generated Figure 1 in the main text, which plots the terms $t_k(S_1)$ for $k = 1, 2, 3, 4$ as the market share S_1 goes from zero to one when $\boldsymbol{\varepsilon}$ is Gaussian.

A.2 Beyond Logistic and Gaussian

The properties of the logistic function may seem to have been more central to our calculations; but in fact they are quite ancillary. Suppose that $u_{i1t} - u_{i0t}$ has some distribution with cdf Q instead of L. While the derivatives of Q may not obey the nice polynomial formulæ we used for L, it is still true that if Q is invertible and smooth then we can define functions F_k by

$$Q^{(k)}(t) = F_k(Q(t)).$$

This is all we need to carry out the expansions. One can show for instance that the factor $(S_1 - \frac{1}{2})$ that appears in (17) just needs to be replaced with

$$-\frac{F_2(S_1)}{2F_1(S_1)}.$$

Take for instance a mixed binary model with such a general distribution for $u_1 - u_0$, and a distribution of the random coefficient on the single covariate X_1 that has successive moments $0, \Sigma, \mu_3, \mu_4$. Then it is easy to derive the following fourth-order expansion, which could perhaps serve as the basis for a semiparametric estimator:

$$\begin{aligned}
\xi_2 = \log\frac{S_1}{S_0} - \Pi X_1 - \frac{F_2(S_1)}{F_1(S_1)} X_1^2 \Sigma \\
- \frac{F_3(S_1)}{F_1(S_1)} X_1^3 \mu_3 \\
+ \frac{F_2(S_1)}{F_1(S_1)} \left(3\frac{F_3(S_1)}{F_1(S_1)} - \left(\frac{F_2(S_1)}{F_1(S_1)}\right)^2\right) X_1^4 \Sigma^2 - \frac{F_4(S_1)}{F_1(S_1)} \mu_4 X_1^4 + O(\sigma^5).
\end{aligned}$$

This can be extended in the obvious way to make v heteroskedastic (just replace Σ with $E(\varepsilon^2|X_1)$ and μ_m with $E(\varepsilon^m|X_1)$ in the above formula.)

B Extensions

This appendix shows how our method applies to a nested logit with random coefficients, and to a model of count data with unobserved heterogeneity.

B.1 The Two-level Mixed Nested Logit

Compiani (2021) applies a nonparametric approach to the choice among a very large set of products. He shows that the mixed logit specification forces the price elasticity to become "too small" at high price levels. This raises the question of the appropriate choice of a distribution for the idiosyncratic terms u_{ijt}.

For the mixed logit ($J = 1$), it is very easy to compute the artificial regressors for any distribution of the idiosyncratic terms; we give the formulæ in Appendix A.2. When $J > 1$, the space of possible distributions increases dramatically. The computations also become more complicated. Finally, estimating the additional parameters of the distribution of $\boldsymbol{u}$ requires (simple) nonlinear optimization.

For illustrative purposes, we give here the estimating equations for the two-level nested logit model. Assume that there is a nest for good 0, and K nests $N_1, \ldots, N_K$ for the varieties of the good. For $k = 1, \ldots, K$, we denote λ_k the corresponding distribution parameter—with the usual interpretation that $(1-\lambda_k)$ proxies for the correlation between choices within nest k, and that the multinomial logit model obtains when all $\lambda_k = 1$.

We denote the market share of nest k by $S_{N_k} = \sum_{j \in N_k} S_j$. Take any variable $\boldsymbol{T} = (T_0, T_1, \ldots, T_J)$. We define the within-nest-k share-weighted average as

$$\bar{T}_k = \sum_{j \in N_k} \frac{S_j}{S_{N_k}} T_j.$$

Note in particular that $e_{\boldsymbol{S}}\boldsymbol{T} = \sum_{k=1}^{K} S_{N_k} \bar{T}_k$.

The following result is the equivalent of Theorem 2 for the two-level mixed nested logit. We relegate its proof to Section C.4.

Theorem 3 (The Artificial Regressors for the Mixed Nested Logit)**.** *For $j \in N_k$, the* *artificial regressors are*

$$K_{mm}^{jt} = \left(\frac{X_{jt,m}}{2} - \frac{1 - S_{0t}\lambda_k}{1 - S_{0t}} e_{tm}\right) \frac{X_{jt,m}}{\lambda_k} + \frac{1 - \lambda_k}{\lambda_k} \bar{X}_{kt,m} \left(\bar{X}_{kt,m} - \frac{2X_{jt,m}}{\lambda_k}\right)$$

and for any off-diagonal term $n < m$,

$$\begin{aligned} K_{mn}^{jt} &= X_{jt,m} X_{jt,n} - \frac{1 - S_{0t}\lambda_k}{1 - S_{0t}} \frac{e_{tm} X_{jt,n} + e_{tn} X_{jt,m}}{\lambda_k} \\ &+ 2\frac{1 - \lambda_k}{\lambda_k} \left(\bar{X}_{kt,m}\bar{X}_{kt,n} - \frac{\bar{X}_{kt,m} X_{jt,n} + \bar{X}_{kt,n} X_{jt,m}}{\lambda_k}\right) \end{aligned}$$

where $e_{tm} = \sum_{j=1}^{J} S_{jt} X_{jtm}$ as per Definition 2.

If the λ_k parameters are known, then our procedure becomes:

Algorithm 3. ***FRAC estimation of the two-level nested logit BLP model***

1. *on every market t, augment the market shares from $(s_{1t}, \ldots, s_{Jt})$ to $(S_{0t}, S_{1t}, \ldots, S_{Jt})$*
2. *for every nest k and product-market pair $(j \in N_k, t)$:*
   (a) *compute the market-share weighted covariate vector $\boldsymbol{e}_t = \sum_{l=1}^{J} S_{lt} \boldsymbol{X}_{lt}$ and the within-nest weighted average covariate vector*

   $$\bar{\boldsymbol{X}}_{k,t} = \sum_{l \in N_k} \frac{S_{lt}}{S_{N_{k,t}}} \boldsymbol{X}_{lt}$$

   (b) *for every (m, n) in $\mathcal{I}$, compute the "artificial regressor"*

   $$K_{mm}^{jt} = \left(\frac{X_{jt,m}}{2} - \frac{1 - S_{0t}\lambda_k}{1 - S_{0t}} e_{tm}\right) \frac{X_{jt,m}}{\lambda_k} + \frac{1 - \lambda_k}{\lambda_k} \bar{X}_{k,t,m} \left(\bar{X}_{k,t,m} - \frac{2X_{jt,m}}{\lambda_k}\right)$$

   and for any off-diagonal term $n < m$,

   $$\begin{aligned} K_{mn}^{jt} &= X_{jt,m} X_{jt,n} - \frac{1 - S_{0t}\lambda_k}{1 - S_{0t}} \frac{e_{tm} X_{jt,n} + e_{tn} X_{jt,m}}{\lambda_k} \\ &+ 2\frac{1 - \lambda_k}{\lambda_k} \left(\bar{X}_{k,t,m}\bar{X}_{k,t,n} - \frac{\bar{X}_{k,t,m} X_{jt,n} + \bar{X}_{k,t,n} X_{jt,m}}{\lambda_k}\right). \end{aligned}$$

   (c) *define*

   $$y_{jt} = \log \frac{S_{N_{k,t}}}{S_{0t}} + \lambda_k \log \frac{S_{jt}}{S_{N_{k,t}}}$$

3. *run a two-stage least squares regression of* $\boldsymbol{y}$ *on* $\boldsymbol{X}$ *and* $\boldsymbol{K}$, *taking as instruments a flexible set of functions of* $\boldsymbol{Z}$

4. *(optional) run a three-stage least squares (3SLS) regression across the* T *markets stacking the* J *equations for each product with a weighting matrix equal to the inverse of the sample variance of the residuals from step 3.*

If the parameters $\boldsymbol{\lambda}$ are not known, then things are slightly more complicated: the formulæ cannot be made linear in $\boldsymbol{\lambda}$, and there are no corresponding artificial regressors. Estimating $(\boldsymbol{\Pi}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$ requires numerical minimization over the $\boldsymbol{\lambda}$ parameters.

More general distributions in the GEV family could also be accommodated. As the nested logit example illustrates, there is a cost to it: the approximate model becomes nonlinear in some parameters. Note however that if there is reason to believe that the true distribution is close to the multinomial logit (say $\boldsymbol{\lambda} \simeq \mathbf{1}$ in the example above), then one can take expansions in the same way we did for the random coefficients and use a 2SLS estimate again.

B.2 Estimating Count Data Models with Heterogeneity

Consider the model defined in (4), and let the estimating equations be $E(\eta_k|Z_k) = 0$ for some set $\mathcal{K} \subset \mathbb{N}$. First we estimate the $y_k(X)$ from the observed counts.

Now define $f_{0,k}(X)$ as a solution of $y_k(X) = P_k(f_{0,k}(X_k), 0)$ (assumed to exist) for each $k \in \mathcal{K}$. Define $d_k = \eta_k + X_k\beta - f_{0,k}(X_k)$. Then we have the system

$$y_k(X) = E_\varepsilon P_k(f_{0,k}(X_k)+d_k, \sigma\varepsilon) \simeq P_k(f_{0,k}(X), 0)+\frac{\partial P_k}{\partial a}(f_{0,k}(X_k), 0)d_k+\frac{\sigma^2}{2}\frac{\partial^2 P_k}{\partial a^2}(f_{0,k}(X_k), 0)$$

where we denote $P_k = P_k(a, b)$.

Since by definition $y_k(X) = P_k(f_{0,k}(X_k), 0)$, this gives $d_k \simeq -\sigma^2 A_k$ where the artificial regressor A_k solves

$$\frac{\partial P_k}{\partial a}(f_{0,k}(X_k), 0)A_k = \frac{1}{2}\frac{\partial^2 P_k}{\partial a^2}(f_{0,k}(X_k), 0).$$

Then we have

$$0 = E(\eta_k|Z_k) = -E(X_k|Z_k)\beta + E(f_{0,k}(X_k)|Z_k) + \sigma^2 E(A_k|Z_k);$$

that is, we estimate β and σ^2 by regressing, for all subpopulations $X = (X_k)_{k \in \mathcal{K}}$ and for all $k \in \mathcal{K}$, $f_{0,k}(X_k)$ on $-X_k$ and A_k with instruments Z_k.

Once we have estimators $\hat{\beta}$ and $\widehat{\sigma^2}$, we make

$$\hat{\eta}_k = f_{0,k}(X_k) - X_k\hat{\beta} + \widehat{\sigma^2}A_k.$$

If we want to do corrected 2SLS, that works too: pick a distribution for ε, solve for the η_k that make

$$y_k(X) = E_\varepsilon P_k(\eta_k + X_k\hat{\beta}, \hat{\sigma}\varepsilon)$$

and use $f_{0,k}(X_k) + \eta_k - \hat{\eta}_k$ instead of $f_{0,k}(X_k)$.

To illustrate this, in the heterogeneous Poisson model we can take all $X_k \equiv X$, all $\eta_k \equiv \eta$, and the Poisson + heterogeneity form for P_k:

$$y_k(X) = E_\varepsilon \frac{\lambda^k \exp(-\lambda)}{k!} \text{ with } \lambda = \eta + X\beta + \sigma\varepsilon$$

but without specifying the distribution of ε beyond $E\varepsilon = 0, V\varepsilon = 1$.

This gives $P_k(a, b) = \frac{(a+b)^k \exp(-(a+b))}{k!}$, so that $y_k(X) = f_{0,k}(X)^k \exp(-f_{0,k}(X))/k!$.

There are two solutions: one with $f_{0,k}(X) < k$ and one with $f_{0,k}(X) > k$ (as long as $y_k(X) < P_k(k, 0)$, otherwise there is no solution.) We should choose the most reasonable one, perhaps by making sure that for the different k the λ_k are not too different. Call it λ_k. Since

$$\frac{\partial P_k}{\partial a} = \frac{k - \lambda_k}{\lambda_k} P_k$$

and

$$\frac{\partial^2 P_k}{\partial a^2} = \frac{(k - \lambda_k)^2 - k}{\lambda_k^2} P_k$$

we get the artificial regressor

$$A_k = \frac{(k - \lambda_k)^2 - k}{2(k - \lambda_k)\lambda_k}.$$

In this setting, 2SLS regresses λ_k on $-X$ and A_k with instruments Z.

C Proofs

C.1 Proof of Theorem 1

We drop $\boldsymbol{Y}$ from the notation since the expansion for a fixed $\boldsymbol{Y}$. Since $\boldsymbol{G}_2^*$ is invertible, there exists a vector $\boldsymbol{g}$ such that we only need to solve

$$\boldsymbol{g} = E_{\boldsymbol{v}} \boldsymbol{A}^*(\mathcal{F}(\boldsymbol{\beta}, \sigma, \boldsymbol{B}) - \boldsymbol{f}_1\boldsymbol{\beta}, \sigma \boldsymbol{B}\boldsymbol{v}). \tag{18}$$

The following lemma enumerates three properties of the function $\mathcal{F}$ at $\sigma = 0$.

Lemma 1 (Properties of the inverse $\mathcal{F}$)**.** *Any regular QLRC model has a well-defined inverse function $\mathcal{F}$ that satisfies the following:*

C1: $\mathcal{F}_\sigma(\boldsymbol{\beta}, 0, \boldsymbol{B}) \equiv \boldsymbol{0}$

C2: $\mathcal{F}(\boldsymbol{\beta}, 0, \boldsymbol{B})$ *is independent of* $\boldsymbol{B}$ *and affine in* $\boldsymbol{\beta}$.

C3: *the second derivative* $\mathcal{F}_{\sigma\sigma}(\boldsymbol{Y}, \boldsymbol{\beta}, 0, \boldsymbol{\beta})$ *does not depend on* $\boldsymbol{\beta}$.

Proof of Lemma 1. First note that at $\sigma = 0$, (18) is simply $\boldsymbol{g} = \boldsymbol{A}^*(\mathcal{F}(\boldsymbol{\beta}, 0, \boldsymbol{B}) - \boldsymbol{f}_1\boldsymbol{\beta}, \boldsymbol{0})$. Since $\boldsymbol{A}_2^*$ is invertible, the equation $\boldsymbol{g} = \boldsymbol{A}^*(\boldsymbol{f}_0, \boldsymbol{0})$ has a unique solution $\boldsymbol{f}_0$; and

$$\mathcal{F}(\boldsymbol{\beta}, 0, \boldsymbol{B}) = \boldsymbol{f}_0 + \boldsymbol{f}_1\boldsymbol{\beta}.$$

This proves **C2**. Moreover, by the Implicit Function Theorem, the function $\mathcal{F}$ is defined for small σ and it is differentiable. Writing (18) at σ and subtracting (18) at $\sigma = 0$ gives an identity in $(\sigma, \boldsymbol{B})$:

$$E_{\boldsymbol{v}} \boldsymbol{A}^*(\mathcal{F}(\boldsymbol{\beta}, \sigma, \boldsymbol{B}) - \boldsymbol{f}_1\boldsymbol{\beta}, \sigma \boldsymbol{B}\boldsymbol{v}) - \boldsymbol{A}^*(\boldsymbol{f}_0, \boldsymbol{0}) \equiv \boldsymbol{0}.$$

Taking the first derivative in σ gives

$$E_{\boldsymbol{v}} \left[\boldsymbol{A}_2^*(\mathcal{F}(\boldsymbol{\beta}, \sigma, \boldsymbol{B}) - \boldsymbol{f}_1\boldsymbol{\beta}, \sigma \boldsymbol{B}\boldsymbol{v})\mathcal{F}_\sigma(\boldsymbol{\beta}, \sigma, \boldsymbol{B}) + \boldsymbol{A}_3^*(\mathcal{F}(\boldsymbol{\beta}, \sigma, \boldsymbol{B}) - \boldsymbol{f}_1\boldsymbol{\beta}, \sigma \boldsymbol{B}\boldsymbol{v})\boldsymbol{B}\boldsymbol{v}\right] \equiv \boldsymbol{0}. \tag{19}$$

At $\sigma = 0$, this is

$$\boldsymbol{A}_2^*(\boldsymbol{f}_0, \boldsymbol{0})\mathcal{F}_\sigma(\boldsymbol{\beta}, 0, \boldsymbol{B}) + \boldsymbol{A}_3^*(\boldsymbol{f}_0, \boldsymbol{0})\boldsymbol{B} E_{\boldsymbol{v}} \boldsymbol{v} = \boldsymbol{0}.$$

Since $E_{\boldsymbol{v}}\boldsymbol{v} = \mathbf{0}$, the second term is zero. As $\boldsymbol{A}_2^*$ is invertible, $\mathcal{F}_\sigma(\boldsymbol{\beta}, 0, \boldsymbol{B})$ must be zero. This proves **C2**.

The second derivative in σ of the identity (or the first derivative of (19)) contains several terms. At $\sigma = 0$, most of them contain $\mathcal{F}_\sigma(\boldsymbol{\beta}, 0, \boldsymbol{B})$, which is zero as we just proved. The only potentially nonzero terms come from the second derivative of $\mathcal{F}$:

$$\boldsymbol{A}_2^*(\boldsymbol{f}_0, \mathbf{0})\, \mathcal{F}_{\sigma\sigma}(\boldsymbol{\beta}, 0, \boldsymbol{B}) \tag{20}$$

and from the second derivative of $\boldsymbol{A}^*$ with respect to its last argument, which is more complicated. The second term of (19) is a vector whose j-th component is

$$E_{\boldsymbol{v}} \left(\sum_{k=1}^{J} \frac{\partial \boldsymbol{A}_j^*}{\partial \varepsilon_k} \left(\mathcal{F}(\boldsymbol{\beta}, \sigma, \boldsymbol{B}) - \boldsymbol{f}_1 \boldsymbol{\beta}, \sigma \boldsymbol{B} \boldsymbol{v} \right) \sum_{m=1}^{M} B_{km} v_m \right).$$

Taking its derivative at $\sigma = 0$ gives

$$E_{\boldsymbol{v}} \left(\sum_{k,l=1}^{J} \frac{\partial^2 \boldsymbol{A}_j^*}{\partial \varepsilon_k \partial \varepsilon_l}(\boldsymbol{f}_0, \mathbf{0}) \sum_{m=1}^{M} B_{km} v_m \sum_{n=1}^{M} B_{ln} v_n \right). \tag{21}$$

We will simplify this term in the main proof. For now, it suffices to note that it does not depend on $\boldsymbol{\beta}$; combining equations (20) and (21) establishes **C3**. □

To continue with the proof of Theorem 1, let us return to equation (21). Since $E v_m v_n = \mathbb{1}(m = n)$, this is

$$\sum_{k,l=1}^{J} \frac{\partial^2 \boldsymbol{A}_j^*}{\partial \varepsilon_k \partial \varepsilon_l}(\boldsymbol{f}_0, \mathbf{0}) \sum_{m=1}^{M} B_{km} B_{lm}$$

Since the result is a scalar (for given j), we can also rewrite this term as its trace:

$$\operatorname{Tr} \left(\boldsymbol{B}' \frac{\partial^2 \boldsymbol{A}_j^*}{\partial \varepsilon \partial \varepsilon'}(\boldsymbol{f}_0, \mathbf{0}) \boldsymbol{B} \right);$$

and as $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{B}\boldsymbol{B}'$, using the cyclical property of the trace operator gives

$$\frac{1}{\sigma^2} \operatorname{Tr} \left(\frac{\partial^2 \boldsymbol{A}_j^*}{\partial \varepsilon \partial \varepsilon'}(\boldsymbol{f}_0, \mathbf{0})\, \boldsymbol{\Sigma} \right).$$

Putting things together gives, for $j = 1, \ldots, J$

$$\frac{\partial \boldsymbol{A}_j^*}{\partial \eta}(\boldsymbol{f}_0, \mathbf{0})\, \mathcal{F}_{\sigma\sigma}(\boldsymbol{\beta}, 0, \boldsymbol{B}) + \frac{1}{\sigma^2} \operatorname{Tr} \left(\frac{\partial^2 \boldsymbol{A}_j^*}{\partial \varepsilon \partial \varepsilon'}(\boldsymbol{f}_0, \mathbf{0})\, \boldsymbol{\Sigma} \right) = 0.$$

The expansion in σ therefore is

$$\begin{aligned}\mathcal{F}(\boldsymbol{\beta}, \sigma, \boldsymbol{B}) &\simeq \boldsymbol{f}_0 + \boldsymbol{f}_1\boldsymbol{\beta} + \frac{\sigma^2}{2}\mathcal{F}_{\sigma\sigma}(\boldsymbol{\beta}, 0, \boldsymbol{B}) \\ &\simeq \boldsymbol{f}_0 + \boldsymbol{f}_1\boldsymbol{\beta} - \frac{1}{2}\left(\frac{\partial \boldsymbol{A}^*}{\partial \eta}(\boldsymbol{f}_0, \boldsymbol{0})\right)^{-1} \operatorname{Tr}\left(\frac{\partial^2 \boldsymbol{A}^*}{\partial \varepsilon \partial \varepsilon'}(\boldsymbol{f}_0, \boldsymbol{0})\,\boldsymbol{\Sigma}\right)\end{aligned}$$

This completes the proof.

C.2 Proof of Theorem 2

To compute the artifical regresssors K_j^{ln}, we first evaluate the derivatives of

$$A_j^* = \frac{\exp(a_j)}{1 + \sum_{k=1}^J \exp(a_k)}.$$

Standard calculations give

$$\frac{\partial A_j^*}{\partial a_k} = A_j^*(\mathbb{1}(j = k) - A_k^*).$$

In the macro-BLP model, $a_k = \eta_k + \bar{\boldsymbol{X}}_k + \boldsymbol{X}_k\boldsymbol{\nu}$, so that

$$\frac{\partial A_j^*}{\partial \eta_k} = A_j^*(X_{jl} - \sum_{k=1}^J A_k^* X_{kl})$$

and

$$\frac{\partial A_j^*}{\partial \nu_l} = A_j^*(X_{jl} - \sum_{k=1}^J A_k^* X_{kl})$$

This yields

$$\frac{\partial^2 A_j^*}{\partial \nu_m \partial \nu_n} = A_j^* \left((X_{jm} - \sum_{k=1}^J A_k^* X_{km})(X_{jn} - \sum_{k=1}^J A_k^* X_{kn}) - \sum_{k=1}^J A_k^* X_{km} X_{kn} + \sum_{k,l=1}^J A_k^* A_l^* X_{km} X_{ln} \right)$$

Remember from Theorem 1 that these expressions have to be evaluated at $\boldsymbol{c} = \boldsymbol{0}$, where A_j^* is simply S_j. We obtain the simple formulæ:

$$\frac{\partial A_j^*}{\partial \boldsymbol{b}} = \operatorname{diag}(\boldsymbol{S}) - \boldsymbol{S}\boldsymbol{S}'$$

and

$$\frac{\partial^2 A_j^*}{\partial \nu_m \partial \nu_n} = S_j \left((X_{jm} - \sum_{k=1}^{J} S_k X_{km})(X_{2jn} - \sum_{k=1}^{J} S_k X_{2kn}) - \sum_{k=1}^{J} S_k X_{km} X_{kn} + \sum_{k,l=1}^{J} S_k S_l X_{2km} X_{ln} \right).$$

Using Definition 2, we obtain

$$\frac{\partial^2 A_j^*}{\partial \boldsymbol{c} \partial \boldsymbol{c}'} = S_j \left(\hat{\boldsymbol{X}}_j \hat{\boldsymbol{X}}_j' - e_{\boldsymbol{S}}(\boldsymbol{X}\boldsymbol{X}') + (e_{\boldsymbol{S}}\boldsymbol{X})(e_{\boldsymbol{S}}\boldsymbol{X}') \right). \tag{22}$$

It follows from Theorem (1) that the tensor $\boldsymbol{f}_2(\boldsymbol{y})$ solves the system

$$(\operatorname{diag}(\boldsymbol{S}) - \boldsymbol{S}\boldsymbol{S}')\, \boldsymbol{f}_2(\boldsymbol{y}) = -\frac{1}{2}\boldsymbol{S} \left(\hat{\boldsymbol{X}} \hat{\boldsymbol{X}}' - e_{\boldsymbol{S}}(\boldsymbol{X}\boldsymbol{X}') + (e_{\boldsymbol{S}}\boldsymbol{X})(e_{\boldsymbol{S}}\boldsymbol{X}') \right).$$

While this may seem forbidding, it decomposes into M^2 systems of J equations of the form

$$f_{2jmn} - e_{\boldsymbol{S}} f_{2mn} = \frac{1}{2} \left(-\hat{X}_{jm}\hat{X}_{jn} + e_{\boldsymbol{S}}(\boldsymbol{X}_m \boldsymbol{X}_n) - (e_{\boldsymbol{S}}\boldsymbol{X}_m)(e_{\boldsymbol{S}}\boldsymbol{X}_n) \right). \tag{23}$$

(where we have divided by S_j on both sides). Given their form, it seems natural to look for a solution of the form

$$f_{2jmn} = -\frac{1}{2}\hat{X}_{jm}\hat{X}_{jn} + d_{mn},$$

which gives

$$e_{\boldsymbol{S}} f_{2mn} = -\frac{1}{2} \left(e_{\boldsymbol{S}}(\boldsymbol{X}_m \boldsymbol{X}_n) + (1 + S_0)(e_{\boldsymbol{S}}\boldsymbol{X}_m)(e_{\boldsymbol{S}}\boldsymbol{X}_n) \right) + (1 - S_0) d_{mn}.$$

Substituting on the left hand-side of (23) gives

$$-\frac{1 + S_0}{2}(e_{\boldsymbol{S}}\boldsymbol{X}_m)(e_{\boldsymbol{S}}\boldsymbol{X}_n) + S_0 d_{mn} = -\frac{1}{2}(e_{\boldsymbol{S}}\boldsymbol{X}_m)(e_{\boldsymbol{S}}\boldsymbol{X}_n)$$

so that $d_{mn} = \frac{1}{2}(e_{\boldsymbol{S}}\boldsymbol{X}_m)(e_{\boldsymbol{S}}\boldsymbol{X}_n)$. Finally:

$$\begin{aligned} f_{2jmn} &= \frac{1}{2} \left((e_{\boldsymbol{S}}\boldsymbol{X}_m)(e_{\boldsymbol{S}}\boldsymbol{X}_n) - \hat{X}_{jm}\hat{X}_{jn} \right) \\ &= \frac{1}{2} \left(X_{jm}(e_{\boldsymbol{S}}\boldsymbol{X}_n) + X_{jn}(e_{\boldsymbol{S}}\boldsymbol{X}_m) - X_{jm}X_{jn} \right). \end{aligned}$$

Reintroducing the market index t, these are the artificial regressors K_{mn}^{jt} whose coefficients are the elements of the matrix

$$V\boldsymbol{\nu} = \boldsymbol{\Pi}\boldsymbol{V}\boldsymbol{\Pi}' + \boldsymbol{\Sigma}$$

if the variance $\boldsymbol{V}$ of the micromoments is constant across markets t (and in particular in the absence of micromoments, as $\boldsymbol{V} = \mathbf{0}$).

If we use micromoments, their observed variance-covariance matrix $\boldsymbol{V}_t$ interacts with K^{jt}_{mn} to create additional artificial regressors whose estimated coefficients are the products of the elements of $\mathbf{\Pi}$: the regression has

$$\sum_{m,n} K^{jt}_{mn}\Sigma_{mn} + \sum_{m,n,r,s} \Pi_{mr}\Pi_{sn}K^{jt}_{mn}V_{t,rs}.$$

C.3 The Fourth-order Expansion in the Standard Model

Assume that there is no micromoment and the moments of order l of $\boldsymbol{\varepsilon}$ scale as σ^l. Under these assumptions, we can write at the fourth-order

$$\xi_j = \log(S_j/S_0) - \boldsymbol{X}_j\bar{\mathbf{\Pi}} - H_{2j}\sigma^2 - H_{3j}\sigma^3 - H_{4j}\sigma^4 + O(\sigma^5), \tag{24}$$

where $\boldsymbol{H}_2, \boldsymbol{H}_3$ and $\boldsymbol{H}_4$ are deterministic functions of $\boldsymbol{X}$, $\boldsymbol{S}$, and the moments of $\boldsymbol{\varepsilon}$ up to the fourth order. Our task here is to derive formulæ for $\boldsymbol{H}_2, \boldsymbol{H}_3$ and $\boldsymbol{H}_4$. To simplify notation, we denote

$$\begin{aligned}
[\boldsymbol{A}, \boldsymbol{B}] &= E_{\boldsymbol{\varepsilon}}(\boldsymbol{A}\boldsymbol{\varepsilon})(\boldsymbol{B}\boldsymbol{\varepsilon}) \\
[\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}] &= E_{\boldsymbol{\varepsilon}}(\boldsymbol{A}\boldsymbol{\varepsilon})(\boldsymbol{B}\boldsymbol{\varepsilon})(\boldsymbol{C}\boldsymbol{\varepsilon}) \\
[\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D}] &= E_{\boldsymbol{\varepsilon}}(\boldsymbol{A}\boldsymbol{\varepsilon})(\boldsymbol{B}\boldsymbol{\varepsilon})(\boldsymbol{C}\boldsymbol{\varepsilon})(\boldsymbol{D}\boldsymbol{\varepsilon}).
\end{aligned}$$

We start by noting that given (24),

$$S_j = E_{\boldsymbol{\varepsilon}} \frac{S_j \exp(r_j)}{S_0 + \sum_{k=1}^J S_k \exp(r_k)}$$

with $r_j \equiv \boldsymbol{X}_j\boldsymbol{\varepsilon} - H_{2j}\sigma^2 - H_{3j}\sigma^3 - H_{4j}\sigma^4 + O(\sigma^5)$. Note that the first term in r_j is of first order in σ, as is $R_j \equiv \exp(r_j) - 1$.

Using the $e_{\boldsymbol{S}}$ notation and dividing through by S_j, we get

$$1 = E_{\boldsymbol{\varepsilon}} \frac{1 + R_j}{1 + e_{\boldsymbol{S}}\boldsymbol{R}}.$$

or, to the fourth order:

$$0 \simeq E_{\boldsymbol{\varepsilon}}(R_j - e_{\boldsymbol{S}}\boldsymbol{R})\left(1 - e_{\boldsymbol{S}}\boldsymbol{R} + (e_{\boldsymbol{S}}\boldsymbol{R})^2 - (e_{\boldsymbol{S}}\boldsymbol{R})^3\right). \tag{25}$$

This has the form $E_{\boldsymbol{\varepsilon}}(R_j - e_{\boldsymbol{S}}\boldsymbol{R})f(e_{\boldsymbol{S}}\boldsymbol{R}) = 0$. Applying the linear operator $e_{\boldsymbol{S}}$ to it gives $E_{\boldsymbol{\varepsilon}}(e_{\boldsymbol{S}}\boldsymbol{R} - e_{\boldsymbol{S}}e_{\boldsymbol{S}}\boldsymbol{R})f(e_{\boldsymbol{S}}\boldsymbol{R}) = 0$. Now $e_{\boldsymbol{S}}e_{\boldsymbol{S}}\boldsymbol{R} = \sum_{k=1}^{J} S_k e_{\boldsymbol{S}}\boldsymbol{R} = (1 - S_0)e_{\boldsymbol{S}}\boldsymbol{R}$, so that we obtain $E_{\boldsymbol{\varepsilon}}e_{\boldsymbol{S}}\boldsymbol{R}f(e_{\boldsymbol{S}}\boldsymbol{R}) = 0$. Therefore we must have $E_{\boldsymbol{\varepsilon}}R_j f(e_{\boldsymbol{S}}\boldsymbol{R}) = 0$. Going back to (25), we need to solve

$$E_{\boldsymbol{\varepsilon}}R_j \left(1 - e_{\boldsymbol{S}}\boldsymbol{R} + (e_{\boldsymbol{S}}\boldsymbol{R})^2 - (e_{\boldsymbol{S}}\boldsymbol{R})^3\right) = O(\sigma^5). \tag{26}$$

This requires computing the terms of degree 4 or less in σ in $E_{\boldsymbol{\varepsilon}}R_j$, $E_{\boldsymbol{\varepsilon}}R_jR_k$, $E_{\boldsymbol{\varepsilon}}R_jR_kR_l$, and $E_{\boldsymbol{\varepsilon}}R_jR_kR_lR_n$. For small σ, and to the fourth-order,

$$\begin{aligned} R_j &\simeq r_j + r_j^2/2 + r_j^3/6 + r_j^4/24 \\ &\simeq \boldsymbol{X}_j\boldsymbol{\varepsilon} \\ &- H_{2j}\sigma^2 + (\boldsymbol{X}_j\boldsymbol{\varepsilon})^2/2 \\ &- H_{3j}\sigma^3 - (\boldsymbol{X}_j\boldsymbol{\varepsilon})H_{2j}\sigma^2 + (\boldsymbol{X}_j\boldsymbol{\varepsilon})^3/6 \\ &- H_{4j}\sigma^4 + H_{2j}^2\sigma^4/2 - (\boldsymbol{X}_j\boldsymbol{\varepsilon})H_{3j}\sigma^3 - (\boldsymbol{X}_j\boldsymbol{\varepsilon})^2 H_{2j}\sigma^2/2 + (\boldsymbol{X}_j\boldsymbol{\varepsilon})^4/24 \end{aligned} \tag{27}$$

where the lines are ordered by increasing degree in σ.

Taking expectations in (27) and using $E\boldsymbol{\varepsilon} = \boldsymbol{0}$ now gives

$$\begin{aligned} E_{\boldsymbol{\varepsilon}}R_j &\simeq -H_{2j}\sigma^2 + [\boldsymbol{X}_j, \boldsymbol{X}_j]/2 \\ &\quad - H_{3j}\sigma^3 + [\boldsymbol{X}_j, \boldsymbol{X}_j, \boldsymbol{X}_j]/6 \\ &\quad - H_{4j}\sigma^4 + H_{2j}^2\sigma^4/2 - [\boldsymbol{X}_j, \boldsymbol{X}_j]H_{2j}\sigma^2/2 + [\boldsymbol{X}_j, \boldsymbol{X}_j, \boldsymbol{X}_j, \boldsymbol{X}_j]/24. \end{aligned}$$

Similarly, and keeping only terms of order 4 or less, we obtain

$$\begin{aligned} E_{\boldsymbol{\varepsilon}}R_jR_k &\simeq [\boldsymbol{X}_j, \boldsymbol{X}_k] \\ &\quad + [\boldsymbol{X}_j, \boldsymbol{X}_k, \boldsymbol{X}_k]/2 + [\boldsymbol{X}_j, \boldsymbol{X}_j, \boldsymbol{X}_k]/2 \\ &\quad + H_{2j}H_{2k}\sigma^4 \\ &\quad - \left(H_{2j}[\boldsymbol{X}_k, \boldsymbol{X}_k] + H_{2k}[\boldsymbol{X}_j, \boldsymbol{X}_j] + 2[\boldsymbol{X}_j, \boldsymbol{X}_k](H_{2j} + H_{2k})\right)\sigma^2/2 \\ &\quad + [\boldsymbol{X}_j, \boldsymbol{X}_j, \boldsymbol{X}_j, \boldsymbol{X}_k]/6 + [\boldsymbol{X}_j, \boldsymbol{X}_k, \boldsymbol{X}_k, \boldsymbol{X}_k]/6 + [\boldsymbol{X}_j, \boldsymbol{X}_j, \boldsymbol{X}_k, \boldsymbol{X}_k]/4. \end{aligned}$$

so that

$$\begin{aligned}E_{\boldsymbol{\varepsilon}} R_j e_{\boldsymbol{S}} \boldsymbol{R} \simeq & [\boldsymbol{X}_j, e_{\boldsymbol{S}} \boldsymbol{X}] \\ & + (e_{\boldsymbol{S}}[\boldsymbol{X}_j, \boldsymbol{X}, \boldsymbol{X}] + [\boldsymbol{X}_j, \boldsymbol{X}_j, e_{\boldsymbol{S}} \boldsymbol{X}])/2 \\ & + \sigma^4 H_{2j} e_{\boldsymbol{S}} \boldsymbol{H}_2 \\ & - (H_{2j} e_{\boldsymbol{S}}[\boldsymbol{X}, \boldsymbol{X}] + [\boldsymbol{X}_j, \boldsymbol{X}_j] e_{\boldsymbol{S}} \boldsymbol{H}_2 + 2[\boldsymbol{X}_j, e_{\boldsymbol{S}} \boldsymbol{X}] H_{2j} + 2 e_{\boldsymbol{S}}(\boldsymbol{H}_2[\boldsymbol{X}_j, \boldsymbol{X}]))\, \sigma^2/2 \\ & + [\boldsymbol{X}_j, \boldsymbol{X}_j, \boldsymbol{X}_j, e_{\boldsymbol{S}} \boldsymbol{X}]/6 + e_{\boldsymbol{S}}[\boldsymbol{X}_j, \boldsymbol{X}, \boldsymbol{X}, \boldsymbol{X}]/6 + e_{\boldsymbol{S}}[\boldsymbol{X}_j, \boldsymbol{X}_j, \boldsymbol{X}, \boldsymbol{X}]/4.\end{aligned}$$

For higher orders:

$$\begin{aligned}E_{\boldsymbol{\varepsilon}} R_j R_k R_l \simeq & [\boldsymbol{X}_j, \boldsymbol{X}_k, \boldsymbol{X}_l] \\ & - \sigma^2 \left([\boldsymbol{X}_j, \boldsymbol{X}_k] H_{2l} + [\boldsymbol{X}_j, \boldsymbol{X}_l] H_{2k} + [\boldsymbol{X}_k, \boldsymbol{X}_l] H_{2j}\right) \\ & + \left([\boldsymbol{X}_j, \boldsymbol{X}_k, \boldsymbol{X}_l, \boldsymbol{X}_l] + [\boldsymbol{X}_j, \boldsymbol{X}_k, \boldsymbol{X}_k, \boldsymbol{X}_l] + [\boldsymbol{X}_j, \boldsymbol{X}_j, \boldsymbol{X}_k, \boldsymbol{X}_l]\right)/2,\end{aligned}$$

so that

$$\begin{aligned}E_{\boldsymbol{\varepsilon}} R_j (e_{\boldsymbol{S}} \boldsymbol{R})^2 \simeq & [\boldsymbol{X}_j, e_{\boldsymbol{S}} \boldsymbol{X}, e_{\boldsymbol{S}} \boldsymbol{X}] \\ & - \sigma^2 \left(2[\boldsymbol{X}_j, e_{\boldsymbol{S}} \boldsymbol{X}] e_{\boldsymbol{S}} \boldsymbol{H}_2 + [e_{\boldsymbol{S}} \boldsymbol{X}, e_{\boldsymbol{S}} \boldsymbol{X}] H_{2j}\right) \\ & + \left(e_{\boldsymbol{S}}[\boldsymbol{X}_j, e_{\boldsymbol{S}} \boldsymbol{X}, \boldsymbol{X}, \boldsymbol{X}] + [\boldsymbol{X}_j, \boldsymbol{X}_j, e_{\boldsymbol{S}} \boldsymbol{X}, e_{\boldsymbol{S}} \boldsymbol{X}]/2\right);\end{aligned}$$

and finally

$$E_{\boldsymbol{\varepsilon}} R_j R_k R_l R_n \simeq [\boldsymbol{X}_j, \boldsymbol{X}_k, \boldsymbol{X}_l, \boldsymbol{X}_n]$$

and $E_{\boldsymbol{\varepsilon}} R_j (e_{\boldsymbol{S}} \boldsymbol{R})^3 \simeq [\boldsymbol{X}_j, e_{\boldsymbol{S}} \boldsymbol{X}, e_{\boldsymbol{S}} \boldsymbol{X}, e_{\boldsymbol{S}} \boldsymbol{X}]$.

Plugging all of this into (25) gives, for the terms of order 2 in σ:

$$H_{2j} \sigma^2 = [\boldsymbol{X}_j, \boldsymbol{X}_j]/2 - [\boldsymbol{X}_j, e_{\boldsymbol{S}} \boldsymbol{X}].$$

Denote $V\varepsilon = \boldsymbol{\Sigma}$. Then $[\boldsymbol{A}, \boldsymbol{B}] = \sum_{m,n} \Sigma_{mn} A_m B_n$ and

$$H_{2j} \sigma^2 = \sum_{m,n} \Sigma_{mn} X_{jm} (X_{jn}/2 - e_{\boldsymbol{S}} \boldsymbol{X}_n),$$

which is the formula given in Theorem 2 .

At order 3, we get

$$H_{3j} \sigma^3 = [\boldsymbol{X}_j, \boldsymbol{X}_j, \boldsymbol{X}_j]/6 - (e_{\boldsymbol{S}}[\boldsymbol{X}_j, \boldsymbol{X}, \boldsymbol{X}] + [\boldsymbol{X}_j, \boldsymbol{X}_j, e_{\boldsymbol{S}} \boldsymbol{X}])/2 + [\boldsymbol{X}_j, e_{\boldsymbol{S}} \boldsymbol{X}, e_{\boldsymbol{S}} \boldsymbol{X}].$$

Take the simplest case, in which the components of the vector $\boldsymbol{\varepsilon}$ are independent with respective third moments s_m. Then

$$H_{3j}\sigma^3 = \sum_m \left(X_{jm}^2/6 - e_{\boldsymbol{S}}\boldsymbol{X}_m^2/2 - X_{jm}e_{\boldsymbol{S}}\boldsymbol{X}_m/2 + (e_{\boldsymbol{S}}\boldsymbol{X}_m)^2\right) X_{jm}s_m,$$

which is the formula given in Section 4.3.

Finally, collecting the terms of order 4 gives

$$\begin{aligned}
H_{4j}\sigma^4 &= H_{2j}^2\sigma^4/2 - [\boldsymbol{X}_j,\boldsymbol{X}_j]H_{2j}\sigma^2/2 + [\boldsymbol{X}_j,\boldsymbol{X}_j,\boldsymbol{X}_j,\boldsymbol{X}_j]/24 \\
&\quad - \sigma^4 H_{2j}e_{\boldsymbol{S}}\boldsymbol{H}_2 + \left(H_{2j}e_{\boldsymbol{S}}[\boldsymbol{X},\boldsymbol{X}] + [\boldsymbol{X}_j,\boldsymbol{X}_j]e_{\boldsymbol{S}}\boldsymbol{H}_2 + 2[\boldsymbol{X}_j,e_{\boldsymbol{S}}\boldsymbol{X}]H_{2j} + 2e_{\boldsymbol{S}}(\boldsymbol{H}_2[\boldsymbol{X}_j,\boldsymbol{X}])\right)\sigma^2/2 \\
&\quad - [\boldsymbol{X}_j,\boldsymbol{X}_j,\boldsymbol{X}_j,e_{\boldsymbol{S}}\boldsymbol{X}]/6 - e_{\boldsymbol{S}}[\boldsymbol{X}_j,\boldsymbol{X},\boldsymbol{X},\boldsymbol{X}]/6 - e_{\boldsymbol{S}}[\boldsymbol{X}_j,\boldsymbol{X}_j,\boldsymbol{X},\boldsymbol{X}]/4 \\
&\quad - \sigma^2\left(2[\boldsymbol{X}_j,e_{\boldsymbol{S}}\boldsymbol{X}]e_{\boldsymbol{S}}\boldsymbol{H}_2 + [e_{\boldsymbol{S}}\boldsymbol{X},e_{\boldsymbol{S}}\boldsymbol{X}]H_{2j}\right) \\
&\quad + \left(e_{\boldsymbol{S}}[\boldsymbol{X}_j,e_{\boldsymbol{S}}\boldsymbol{X},\boldsymbol{X},\boldsymbol{X}] + [\boldsymbol{X}_j,\boldsymbol{X}_j,e_{\boldsymbol{S}}\boldsymbol{X},e_{\boldsymbol{S}}\boldsymbol{X}]/2\right) \\
&\quad - [\boldsymbol{X}_j,e_{\boldsymbol{S}}\boldsymbol{X},e_{\boldsymbol{S}}\boldsymbol{X},e_{\boldsymbol{S}}\boldsymbol{X}].
\end{aligned}$$

After replacing H_{2j} and $e_{\boldsymbol{S}}\boldsymbol{H}_2$ with their values, may terms cancel out. We finally obtain

$$\begin{aligned}
H_{4j}\sigma^4 &= [\boldsymbol{X}_j,e_{\boldsymbol{S}}\boldsymbol{X}]^2/2 + [\boldsymbol{X}_j,\boldsymbol{X}_j]^2/8 - [\boldsymbol{X}_j,\boldsymbol{X}_j][\boldsymbol{X}_j,e_{\boldsymbol{S}}\boldsymbol{X}]/2 \\
&\quad + e_{\boldsymbol{S}}\left([\boldsymbol{X}_j,\boldsymbol{X}][\boldsymbol{X},\boldsymbol{X}]\right)/2 - e_{\boldsymbol{S}}\left([\boldsymbol{X}_j,\boldsymbol{X}][e_{\boldsymbol{S}}\boldsymbol{X},\boldsymbol{X}]\right) \\
&\quad - [\boldsymbol{X}_j,\boldsymbol{X}_j,\boldsymbol{X}_j,e_{\boldsymbol{S}}\boldsymbol{X}]/6 - e_{\boldsymbol{S}}[\boldsymbol{X}_j,\boldsymbol{X},\boldsymbol{X},\boldsymbol{X}]/6 - e_{\boldsymbol{S}}[\boldsymbol{X}_j,\boldsymbol{X}_j,\boldsymbol{X},\boldsymbol{X}]/4 \\
&\quad + \left(e_{\boldsymbol{S}}[\boldsymbol{X}_j,e_{\boldsymbol{S}}\boldsymbol{X},\boldsymbol{X},\boldsymbol{X}] + [\boldsymbol{X}_j,\boldsymbol{X}_j,e_{\boldsymbol{S}}\boldsymbol{X},e_{\boldsymbol{S}}\boldsymbol{X}]/2\right) \\
&\quad - [\boldsymbol{X}_j,e_{\boldsymbol{S}}\boldsymbol{X},e_{\boldsymbol{S}}\boldsymbol{X},e_{\boldsymbol{S}}\boldsymbol{X}] + [\boldsymbol{X}_j,\boldsymbol{X}_j,\boldsymbol{X}_j,\boldsymbol{X}_j]/24
\end{aligned}$$

To illustrate, assume that the ε_m terms are independent, with variances $\Sigma_{mm} = \sigma_m^2$ and excess kurtosis $E\varepsilon_m^4 - 3\sigma_m^4 = \kappa_m$. Then $[\boldsymbol{A},\boldsymbol{B}] = \sum_m A_m B_m \sigma_m^2$ and

$$\begin{aligned}
[\boldsymbol{A},\boldsymbol{B},\boldsymbol{C},\boldsymbol{D}] = [\boldsymbol{A},\boldsymbol{B}][\boldsymbol{C},\boldsymbol{D}] + [\boldsymbol{A},\boldsymbol{C}][\boldsymbol{B},\boldsymbol{D}] + [\boldsymbol{A},\boldsymbol{D}][\boldsymbol{B},\boldsymbol{C}] \\
+ \sum_m \kappa_m A_m B_m C_m D_m.
\end{aligned}$$

The fourth-order term $H_{4j}\sigma^4$ contains both terms that are linear in the κ_m (the last three lines) and terms that are quadratic in $\boldsymbol{\Sigma}$ (the first two lines). The first group suggests introducing more artificial regressors

$$Q_m^j \equiv X_{jm}\left(\frac{X_{jm}^3}{24} - X_{jm}^2\frac{e_{\boldsymbol{S}}\boldsymbol{X}_m}{6} + X_{jm}\frac{(e_{\boldsymbol{S}}\boldsymbol{X}_m)^2}{2} - X_{jm}\frac{e_{\boldsymbol{S}}(\boldsymbol{X}_m^2)}{4}\right.$$
$$\left. + \frac{(e_{\boldsymbol{S}}\boldsymbol{X}_m)(e_{\boldsymbol{S}}(\boldsymbol{X}_m^2))}{2} - (e_{\boldsymbol{S}}\boldsymbol{X}_m)^3 - \frac{e_{\boldsymbol{S}}(\boldsymbol{X}_m^3)}{6}\right),$$

whose coefficients are the excess kurtosis parameters κ_m. The second group yields

$$\sum_{m,n}\sigma_m^2\sigma_n^2 W_{mn}^j$$

where

$$W_{mn}^j = X_{jm}\left(\frac{X_{jn}e_{\boldsymbol{S}}\boldsymbol{X}_m e_{\boldsymbol{S}}\boldsymbol{X}_n}{2} + \frac{X_{jm}X_{jn}^2}{8} - \frac{X_{jm}X_{jn}e_{\boldsymbol{S}}\boldsymbol{X}_n}{2}\right.$$
$$\left. + \frac{e_{\boldsymbol{S}}(\boldsymbol{X}_m\boldsymbol{X}_n^2)}{2} - e_{\boldsymbol{S}}(\boldsymbol{X}_m\boldsymbol{X}_n)e_{\boldsymbol{S}}\boldsymbol{X}_n\right).$$

Note that the new artificial regressors $\boldsymbol{W}$ are assigned products of the elements of $\boldsymbol{\Sigma}$. Estimating the resulting regression requires nonlinear optimization (albeit a very simple one).

C.4 Proof of Theorem 3

In the unmixed model ($\sigma = 0$) the mean utility of alternative j is $U_j = I_k + \lambda_k \log S_{j|N_k}$ if $j \in N_k$, with $I_k \equiv \log(S_{N_k}/S_0)$ and $S_{j|N_k} \equiv S_j/S_{N_k}$ .This gives

$$\xi_j^0 = -\boldsymbol{X}_j\boldsymbol{\beta} + \log(S_{N_k}/S_0) + \lambda_k \log S_{j|N_k}.$$

As in Appendix C.1, we decompose $\boldsymbol{\varepsilon} = \sigma\boldsymbol{B}\boldsymbol{v}$. We now denote $\boldsymbol{x} = \boldsymbol{B}'\boldsymbol{X}'$ so that $\boldsymbol{X}\boldsymbol{\varepsilon} = \sigma\boldsymbol{x}\cdot\boldsymbol{v}$. We write (imposing $a_{1j} = 0$ from the start as this is a general property of models with $E\boldsymbol{v} = \boldsymbol{0}$)

$$U_j(\boldsymbol{v}) = \log(S_{N_k}/S_0) + \lambda_k \log S_{j|N_k} + \sigma\boldsymbol{x}_j\cdot\boldsymbol{v} + \frac{\sigma^2}{2}a_{2j}$$

and

$$\exp(I_k(\boldsymbol{v})/\lambda_k) = \sum_{j\in N_k}\exp(U_j(\boldsymbol{v})/\lambda_k) = (S_{N_k}/S_0)^{1/\lambda_k}\bar{f}_k(\boldsymbol{v})$$

where we denote $\bar{X}_k = \sum_{j \in N_k} S_{j|N_k} X_j$ and

$$f_j(\boldsymbol{v}) = \exp\left(\frac{\sigma}{\lambda_k}\left(\boldsymbol{x}_j \cdot \boldsymbol{v} + \sigma \frac{a_{2j}}{2}\right)\right) \simeq 1 + \frac{\sigma}{\lambda_k}(\boldsymbol{x}_j \cdot \boldsymbol{v}) + \frac{\sigma^2}{2\lambda_k^2}\left(\lambda_k a_{2j} + (\boldsymbol{x}_j \cdot \boldsymbol{v})^2\right)$$

so that

$$\bar{f}_k(\boldsymbol{v}) \simeq 1 + \frac{\sigma}{\lambda_k}\bar{\boldsymbol{x}}_k \cdot \boldsymbol{v} + \frac{\sigma^2}{2\lambda_k^2}\left(\lambda_k \bar{a}_{2k} + \overline{(\boldsymbol{x} \cdot \boldsymbol{v})^2}_k\right).$$

Now using

$$S_j = E_{\boldsymbol{v}} \exp((U_j(\boldsymbol{v}) - I_k(\boldsymbol{v}))/\lambda_k) \frac{\exp(I_k(\boldsymbol{v}))}{1 + \sum_{l=1}^K \exp(I_l(\boldsymbol{v}))}$$

we get

$$1 = E_{\boldsymbol{v}}\left(\frac{f_j(\boldsymbol{v})}{\bar{f}_k(\boldsymbol{v})} \frac{\left(\bar{f}_k(\boldsymbol{v})\right)^{\lambda_k}}{S_0 + \sum_{l=1}^K S_{N_l}\left(\bar{f}_l(\boldsymbol{v})\right)^{\lambda_l}}\right).$$

We note that

$$\frac{1 + a\sigma + b\sigma^2}{1 + c\sigma + d\sigma^2} = 1 + (a-c)\sigma + (b - d - c(a-c))\sigma^2 + O(\sigma^3). \tag{28}$$

Denote $\hat{A}_{j|k} = A_j - \bar{A}_k$. Applying (28) gives

$$\frac{f_j(\boldsymbol{v})}{\bar{f}_k(\boldsymbol{v})} \simeq 1 + \frac{\sigma}{\lambda_k} C_j(\boldsymbol{v}) + \frac{\sigma^2}{2\lambda_k^2} D_j(\boldsymbol{v}).$$

with

$$C_j(\boldsymbol{v}) = \hat{\boldsymbol{x}}_{j|k} \cdot \boldsymbol{v}$$

and

$$D_j(\boldsymbol{v}) = \lambda_k \hat{a}_{2j|k} + \widehat{(\boldsymbol{x} \cdot \boldsymbol{v})^2}_{j|k} - 2(\bar{\boldsymbol{x}}_k \cdot \boldsymbol{v})(\hat{\boldsymbol{x}}_{j|k} \cdot \boldsymbol{v}).$$

Moreover,

$$\left(\bar{f}_l(\boldsymbol{v})\right)^{\lambda_l} \simeq 1 + \sigma \bar{\boldsymbol{x}}_l \cdot \boldsymbol{v} + \frac{\sigma^2}{2}\left(\frac{\lambda_l - 1}{\lambda_l}(\bar{\boldsymbol{x}}_l \cdot \boldsymbol{v})^2 + \bar{a}_{2l} + \frac{\overline{(\boldsymbol{x} \cdot \boldsymbol{v})^2}_l}{\lambda_l}\right)$$

and

$$\frac{\left(\bar{f}_k(\boldsymbol{v})\right)^{\lambda_k}}{S_0 + \sum_{l=1}^K S_{N_l}\left(\bar{f}_l(\boldsymbol{v})\right)^{\lambda_l}} \simeq \frac{1 + \sigma \bar{\boldsymbol{x}}_k \cdot \boldsymbol{v} + \frac{\sigma^2}{2}\left(\bar{a}_{2k} + \frac{\lambda_k - 1}{\lambda_k}(\bar{\boldsymbol{x}}_k \cdot \boldsymbol{v})^2 + \frac{\overline{(\boldsymbol{x} \cdot \boldsymbol{v})^2}_k}{\lambda_k}\right)}{1 + \sigma e_{\boldsymbol{S}} \boldsymbol{x} \cdot \boldsymbol{v} + \frac{\sigma^2}{2}\left(e_{\boldsymbol{S}} a_2 + \sum_{l=1}^K S_{N_l}\left(\frac{\lambda_l - 1}{\lambda_l}(\bar{\boldsymbol{x}}_l \cdot \boldsymbol{v})^2 + \frac{\overline{(\boldsymbol{x} \cdot \boldsymbol{v})^2}_l}{\lambda_l}\right)\right)}$$

where as usual $e_{\boldsymbol{S}} \boldsymbol{T} = \sum_{j=1}^J S_j T_j = \sum_{k=1}^K S_{N_k} \bar{T}_k$.

Then, using (28) again,

$$\frac{\left(\bar{f}_k(\boldsymbol{v})\right)^{\lambda_k}}{S_0 + \sum_{l=1}^{K} S_{N_l} \left(\bar{f}_l(\boldsymbol{v})\right)^{\lambda_l}} \simeq 1 + \sigma E_k(\boldsymbol{v}) + \frac{\sigma^2}{2} F_k(\boldsymbol{v})$$

with

$$E_k(\boldsymbol{v}) = (\bar{\boldsymbol{x}}_k - e_{\boldsymbol{S}}\boldsymbol{x}) \cdot \boldsymbol{v}$$

and

$$\begin{aligned} F_k(\boldsymbol{v}) = {} & \bar{a}_{2k} - e_{\boldsymbol{S}} a_2 \\ & + \frac{\lambda_k - 1}{\lambda_k} (\bar{\boldsymbol{x}}_k \cdot \boldsymbol{v})^2 - \sum_{l=1}^{K} S_{N_l} \frac{\lambda_l - 1}{\lambda_l} (\bar{\boldsymbol{x}}_l \cdot \boldsymbol{v})^2 \\ & + \frac{\overline{(\boldsymbol{x} \cdot \boldsymbol{v})^2}_k}{\lambda_k} - \sum_{l=1}^{K} S_{N_l} \frac{\overline{(\boldsymbol{x} \cdot \boldsymbol{v})^2}_l}{\lambda_l} \\ & - 2(e_{\boldsymbol{S}}\boldsymbol{x} \cdot \boldsymbol{v})((\bar{\boldsymbol{x}}_k - e_{\boldsymbol{S}}\boldsymbol{x}) \cdot \boldsymbol{v}). \end{aligned}$$

This allows us to write

$$\begin{aligned} 1 & \simeq E_{\boldsymbol{v}} \left(1 + \frac{\sigma}{\lambda_k} C_j + \frac{\sigma^2}{2\lambda_k^2} D_j\right) \left(1 + \sigma E_k + \frac{\sigma^2}{2} F_k\right) \\ & \simeq E_{\boldsymbol{v}} \left(1 + \sigma \left(\frac{C_j}{\lambda_k} + E_k\right) + \frac{\sigma^2}{2\lambda_k^2} \left(D_j + \lambda_k^2 F_k + 2\lambda_k C_j E_k\right)\right). \end{aligned}$$

We have $E_{\boldsymbol{v}} C_j = E_{\boldsymbol{v}} E_k = 0$; also,

$$\begin{aligned} ED_j & = \lambda_k \widehat{a}_{2j|k} + \|\boldsymbol{x}_j\|^2 - \overline{\|\boldsymbol{x}\|^2}_k - 2\bar{\boldsymbol{x}}_k \cdot \hat{\boldsymbol{x}}_{j|k} \\ EF_k & = \bar{a}_{2k} - e_{\boldsymbol{S}} a_2 \\ & \quad + \frac{\lambda_k - 1}{\lambda_k} \|\bar{\boldsymbol{x}}_k\|^2 - \sum_{l=1}^{K} S_{N_l} \frac{\lambda_l - 1}{\lambda_l} \|\bar{\boldsymbol{x}}_l\|^2 \\ & \quad + \frac{\overline{\|\boldsymbol{x}\|^2}_k}{\lambda_k} - \sum_{l=1}^{K} S_{N_l} \frac{\overline{\|\boldsymbol{x}\|^2}_l}{\lambda_l} \\ & \quad - 2(e_{\boldsymbol{S}}\boldsymbol{x}) \cdot (\bar{\boldsymbol{x}}_k - e_{\boldsymbol{S}}\boldsymbol{x}) \\ E(C_j E_k) & = \hat{\boldsymbol{x}}_{j|k} \cdot (\bar{\boldsymbol{x}}_k - e_{\boldsymbol{S}}\boldsymbol{x}). \end{aligned}$$

Writing $E(D_j + \lambda_k^2 F_k + 2\lambda_k C_j E_k) = 0$ gives us an equation of the form

$$\lambda_k(a_{2j} - \bar{a}_{2k}) + \lambda_k^2(\bar{a}_{2k} - e_{\boldsymbol{S}} a_2) = \lambda_k^2 M + \nu_k + \mu_j$$

where

$$
\begin{aligned}
M &= \sum_{l=1}^{K} S_{N_l} \frac{\lambda_l - 1}{\lambda_l} \|\bar{\boldsymbol{x}}_l\|^2 + \sum_{l=1}^{K} S_{N_l} \frac{\overline{\|\boldsymbol{x}\|^2}_l}{\lambda_l} - 2\|e_{\boldsymbol{S}}\boldsymbol{x}\|^2 \\
\nu_k &= \overline{\|\boldsymbol{x}\|^2}_k - 2\|\bar{\boldsymbol{x}}_k\|^2 - \lambda_k(\lambda_k - 1)\|\bar{\boldsymbol{x}}_k\|^2 - \lambda_k \overline{\|\boldsymbol{x}\|^2}_k + 2\lambda_k^2 e_{\boldsymbol{S}}\boldsymbol{x} \cdot \bar{\boldsymbol{x}}_k + 2\lambda_k \|\bar{\boldsymbol{x}}_k\|^2 - 2\lambda_k \bar{\boldsymbol{x}}_k \cdot e_{\boldsymbol{S}}\boldsymbol{x} \\
&= (1 - \lambda_k) \left(\overline{\|\boldsymbol{x}\|^2}_k - (2 - \lambda_k)\|\bar{\boldsymbol{x}}_k\|^2 - 2\lambda_k \bar{\boldsymbol{x}}_k \cdot e_{\boldsymbol{S}}\boldsymbol{x} \right) \qquad (29) \\
\mu_j &= -\|\boldsymbol{x}_j\|^2 + 2\boldsymbol{x}_j \cdot \bar{\boldsymbol{x}}_k - 2\lambda_k \boldsymbol{x}_j \cdot (\bar{\boldsymbol{x}}_k - e_{\boldsymbol{S}}\boldsymbol{x}) \\
&= \boldsymbol{x}_j \cdot (2\lambda_k e_{\boldsymbol{S}}\boldsymbol{x} - \boldsymbol{x}_j + 2(1 - \lambda_k)\bar{\boldsymbol{x}}_k) . \qquad (30)
\end{aligned}
$$

It is easy to aggregate from $a_{2j} = (1 - \lambda_k)\bar{a}_{2k} + \lambda_k e_{\boldsymbol{S}} a_2 + \lambda_k M + (\nu_k + \mu_j)/\lambda_k$ to

$$
\bar{a}_{2k} = e_{\boldsymbol{S}} a_2 + M + \frac{\nu_k + \bar{\mu}_k}{\lambda_k^2}
$$

and then to

$$
S_0 e_{\boldsymbol{S}} a_2 = (1 - S_0) M + \sum_{k=1}^{K} S_{N_k} \frac{\nu_k + \bar{\mu}_k}{\lambda_k^2},
$$

which gives

$$
\begin{aligned}
a_{2j} &= e_{\boldsymbol{S}} a_2 + M + (1 - \lambda_k) \frac{\nu_k + \bar{\mu}_k}{\lambda_k^2} + \frac{\nu_k + \mu_j}{\lambda_k} \\
&= \frac{M}{S_0} + \frac{1}{S_0} \sum_{l=1}^{K} S_{N_l} \frac{\nu_l + \bar{\mu}_l}{\lambda_l^2} + (1 - \lambda_k) \frac{\nu_k + \bar{\mu}_k}{\lambda_k^2} + \frac{\nu_k + \mu_j}{\lambda_k} \\
&= \frac{M}{S_0} + \frac{1}{S_0} \sum_{l=1}^{K} S_{N_l} \frac{\nu_l + \bar{\mu}_l}{\lambda_l^2} + \frac{\nu_k + (1 - \lambda_k)\bar{\mu}_k}{\lambda_k^2} + \frac{\mu_j}{\lambda_k}.
\end{aligned}
$$

Finally, using equations (29) and (30) we aggregate

$$
\bar{\mu}_k = 2\lambda_k \bar{\boldsymbol{x}}_k \cdot e_{\boldsymbol{S}}\boldsymbol{x} + 2(1 - \lambda_k)\|\bar{\boldsymbol{x}}_k\|^2 - \overline{\|\boldsymbol{x}\|^2}_k,
$$

which gives

$$
\nu_k + \bar{\mu}_k = 2\lambda_k^2 \bar{\boldsymbol{x}}_k \cdot e_{\boldsymbol{S}}\boldsymbol{x} + \lambda_k(1 - \lambda_k)\|\bar{\boldsymbol{x}}_k\|^2 - \lambda_k \overline{\|\boldsymbol{x}\|^2}_k
$$

and

$$
\nu_k + (1 - \lambda_k)\bar{\mu}_k = -\lambda_k(1 - \lambda_k)\|\bar{\boldsymbol{x}}_k\|^2.
$$

Putting everything together, we get

$$
\begin{aligned}
a_{2j} &= \frac{M}{S_0} + \frac{1}{S_0}\sum_{l=1}^{K} S_{N_l}\frac{\nu_l + \bar{\mu}_l}{\lambda_l^2} + \frac{\nu_k + (1-\lambda_k)\bar{\mu}_k}{\lambda_k^2} + \frac{\mu_j}{\lambda_k} \\
&= \frac{1}{S_0}\left(\sum_{l=1}^{K} S_{N_l}\frac{\lambda_l - 1}{\lambda_l}\|\bar{\boldsymbol{x}}_l\|^2 + \sum_{l=1}^{K} S_{N_l}\frac{\overline{\|\boldsymbol{x}\|^2}_l}{\lambda_l} - 2\|e_{\boldsymbol{S}}\boldsymbol{x}\|^2\right) \\
&+ \frac{2}{S_0}\|e_{\boldsymbol{S}}\boldsymbol{x}\|^2 + \frac{1}{S_0}\sum_{l=1}^{K} S_{N_l}\frac{-\overline{\|\boldsymbol{x}\|^2}_l + (1-\lambda_l)\|\bar{\boldsymbol{x}}_l\|^2}{\lambda_l} \\
&= \boldsymbol{x}_j \cdot \left(2e_{\boldsymbol{S}}\boldsymbol{x} - \frac{\boldsymbol{x}_j}{\lambda_k} + 2\frac{1-\lambda_k}{\lambda_k}\bar{\boldsymbol{x}}_k\right) - \frac{1-\lambda_k}{\lambda_k}\|\bar{\boldsymbol{x}}_k\|^2.
\end{aligned}
$$

We finally get the artificial regressors in Theorem 3 by replacing $\sigma^2\boldsymbol{x}\boldsymbol{x}'$ with $\boldsymbol{X}'\boldsymbol{\Sigma}\boldsymbol{X}$.

D Descriptive Features of the Monte Carlo Simulations of Section 6

For each simulation run, we estimate the conditional expectation of the price of product j in market t given the vector of instruments by a linear regression of p_{jt} on the 36 elements of $\boldsymbol{W}_{jt}$. Denoting $\bar{V} = \sum_{j,t} V_{jt}/(JT)$ for any variable $\boldsymbol{V}$, we use the true parameter values to compute

$$
V_1^D \;=\; \frac{1}{JT}\sum_{t=1}^{T}\sum_{j=1}^{J}(\boldsymbol{X}'_{jt}\bar{\beta}^x - \bar{\beta}^p E(p_{jt}|\boldsymbol{W}_{jt}) - (\bar{\boldsymbol{X}}'\bar{\beta}^x - \bar{\beta}^p\overline{E(p|\boldsymbol{W})}))^2;
$$

$$
V_2^D \;=\; \frac{1}{JTn_s}\sum_{t=1}^{T}\sum_{j=1}^{J}\sum_{i=1}^{n_s}(\boldsymbol{X}'_{jt}(\beta_i^x - \bar{\beta}^x) - (\beta_i^p - \bar{\beta}^p)E(p_{jt}|\boldsymbol{W}_{jt}))^2;
$$

and

$$
V_3^D \;=\; \frac{1}{JT}\sum_{t=1}^{T}\sum_{j=1}^{J}\xi_{jt}^2.
$$

We take the shares of V_1^D, V_2^D, and V_3^D in the total variance $V_1^D+V_2^D+V_3^D$ to represent the contributions of, respectively, the variation in covariates and instrumented prices; the random variation in consumer preferences; and the product effects.

For the supply side, we run an OLS regression:

$$p_{jt} = \alpha_0 + \boldsymbol{W}'_{jt}\boldsymbol{\alpha} + \xi_{jt}\alpha_\xi + \omega_{jt}\alpha_\omega + v_{jt}$$

and we define

$$V_1^S = \frac{1}{JT}\sum_{t=1}^{T}\sum_{j=1}^{J}(\boldsymbol{W}_{jt} - \bar{\boldsymbol{W}})'\hat{\boldsymbol{\alpha}})^2;$$

$$V_2^S = \frac{1}{JT}\sum_{t=1}^{T}\sum_{j=1}^{J}(\xi_{jt}\hat{\alpha}_\xi + \omega_{jt}\hat{\alpha}_\omega)^2;$$

and

$$V_3^S = \frac{1}{JT}\sum_{t=1}^{T}\sum_{j=1}^{J}\hat{v}_{jt}^2.$$

We take the shares of V_1^S, V_2^S, and V_3^S in the total variance $V_1^S + V_2^S + V_3^S$ to represent the contributions of, respectively, the variation in the cost-shifters and the demand covariates; the unobserved demand- and supply-side product effects; and the unexplained part.