NBER WORKING PAPER SERIES

FAST, "ROBUST", AND APPROXIMATELY CORRECT: ESTIMATING MIXED DEMAND SYSTEMS

Bernard Salanié
Frank A. Wolak

Working Paper 25726
http://www.nber.org/papers/w25726

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2019

We are grateful to Dan Ackerberg, John Asker, Steve Berry, Xiaohong Chen, Chris Conlon, 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, and Ken Wolpin 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. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Fast, "Robust", and Approximately Correct: Estimating Mixed Demand Systems
Bernard Salanié and Frank A. Wolak
NBER Working Paper No. 25726
April 2019
JEL No. L00,L13

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. We study in detail the models of product market shares and prices popular in empirical IO. Our estimator is only approximately correct, but it performs very well in practice. It is extremely fast and easy to implement, and it is “robust” to changes in the higher moments of the distribution of the random coefficients. At the very least, it provides excellent starting values for more commonly used estimators of these models.

Bernard Salanié
Department of Economics
Columbia University
1033 International Affairs Building, MC 3308
420 West 118th Street
New York, NY 10027
and CEPR
bs2237@columbia.edu

Frank A. Wolak
Department of Economics
Stanford University
Stanford, CA 94305-6072
and NBER
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:

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

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 observed endogenous 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. Then (assuming that 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

$$\|E\left(F(y, \theta)m(Z)\right)\|$$

where $m(Z)$ is a vector of measurable functions of Z.

Unless $F(y, \theta_0)$ 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

$$\|E\left(\eta m(Z)\right)\|$$

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, for instance within models of bargaining like those of Crawford and Yurokoglu (2012) or Ho and Lee (2017).

We propose an alternative that derives a linear model from a very simple series expansion. To fix ideas, suppose that θ_0 can be decomposed into a pair (β_0, σ_0), where σ_0 is a scalar that we have reasons to think is not too far from zero. We rewrite (1) as

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

We expand $\sigma \to F(y, \beta_0, \sigma)$ in a Taylor series around 0 and re-write $F(y, \beta_0, \sigma_0)$ as:

$$F(y, \beta_0, \sigma_0) = F(y, \beta_0, 0) + F_\sigma(y, \beta_0, 0)\sigma_0 + \ldots + F_{\sigma\sigma\ldots\sigma}(y, \beta_0, 0)\frac{\sigma_0^L}{L!} + O(\sigma_0^{L+1}),$$

where the subscript σ denotes a partial derivative with respect to the argument σ.

This suggests a sequence of "approximate estimators" that minimize the empirical analogs of the following norms

$$\begin{aligned}
&\| E\left(F(y, \beta, 0) m(Z)\right) \| \\
&\| E\left(\left(F(y, \beta, 0) + F_\sigma(y, \beta, 0)\sigma\right) m(Z)\right) \| \\
&\left\| E\left(\left(F(y, \beta, 0) + F_\sigma(y, \beta, 0)\sigma + F_{\sigma\sigma}(y, \beta, 0)\frac{\sigma^2}{2}\right) m(Z)\right) \right\| \\
\ldots
\end{aligned}$$

If the true value σ_0 is not too large, one may hope to obtain a satisfactory estimator with the third of these "approximate estimators." In general, this still requires solving a nonlinear minimization problem. However, suppose that the function F satisfies the following three conditions:

C1: $F_\sigma(y, \beta_0, 0) \equiv 0$

C2: $F(y, \beta, 0) \equiv f_0(y) - f_1(y)\beta$ is affine in β for known functions $f_0(\cdot)$ and $f_1(\cdot)$.

C3: the second derivative $F_{\sigma\sigma}(y, \beta, 0)$ does not depend on β.

Denote $f_2(y) \equiv -F_{\sigma\sigma}(y, \beta, 0)$. Under **C1**–**C3**, we would minimize

$$\left\| E\left(\left(f_0(y) - f_1(y)\beta - f_2(y)\frac{\sigma^2}{2}\right) m(Z)\right) \right\|.$$

Taking the parameters of interest to be (β_0, σ_0^2), this is simply a two-stage least squares regression of $f_0(y)$ on $f_1(y)$ and $f_2(y)$ with instruments $m(Z)$. As this is a linear problem, the optimal[1] instruments $m(Z)$ associated with the conditional moment restrictions $E(\eta|Z) = 0$ are simply

$$m(Z) = Z^* = \left(E\left(f_1(y)|Z\right), E\left(f_2(y)|Z\right)\right).$$

These optimal instuments could be estimated directly from the data using nonparametric regressions. Or more simply, we can include flexible functions of the columns of Z in the instruments used to compute the 2SLS estimates.

The resulting estimators of β_0 and σ_0^2 are only approximately correct, because they consistently estimate an approximation of the original model. On the other hand, they can be estimated in closed form using linear 2SLS. Moreover, because they only rely on limited features of the data generating process, they are "robust" in ways that we will explore later.

Conditions **C1**–**C3** extend directly to a multivariate parameter $\boldsymbol{\sigma_0}$. They may seem very demanding. Yet as we will show, under very weak conditions the Berry, Levinsohn, and Pakes (1995) (macro-BLP) model that is the workhorse of empirical IO satisfies all three. In this application, $\boldsymbol{\sigma_0}$ is taken to be the square root of the variance–covariance matrix $\boldsymbol{\Sigma}$ of the random coefficients in the mixed demand model. More generally, we will characterize in Section 6.1 a general class of models with unobserved heterogeneity to which conditions **C1**–**C3** apply.

Our approach builds on "small-$\boldsymbol{\Sigma}$" approximations to construct successive approximations to the inverse mapping (from market shares to product effects). 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[2] 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 er-

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

[2] Which include OLS and 2SLS.

ror. 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 "heterogeneity adjusted" unmixed multinomial logit model in which mean utilities have additional terms[3]. 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.

Section 1 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 2; readers not interested in the derivation of our formulæ in fact can jump directly to our Monte Carlo simulations in section 7. The rest of the paper justifies our algorithm (sections 3 and 4); studies its properties (section 5); and discusses a variety of extensions (section 6).

1 The macro-BLP model

Our leading example is taken from empirical IO. Much work in this area 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[4]. Let J products be available on each of T markets. Each market contains an infinity of consumers who

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

[4] While some of our exposition relies on it for simplicity, our methods apply to a more general model— see section 6.1.

choose one of J products. Consumer i in market t derives a conditional indirect utility from consuming product j equal to

$$\boldsymbol{X}_{jt}\left(\boldsymbol{\beta}_0+\boldsymbol{\epsilon}_i\right)+\xi_{jt}+u_{ijt}.$$

There is also a good 0, the “outside good”, whose utility for consumer i is typically normalized to equal u_{i0t}. The random variables $\boldsymbol{\epsilon}$ 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 vector $\boldsymbol{\epsilon}$ and $\boldsymbol{u}$ are independent of each other, and of the covariates $\boldsymbol{X}$ and product random effects $\boldsymbol{\xi}$. Berry et al. (1995) assume that the vector $\boldsymbol{u}_{it}=(u_{i0t},u_{i1t},\ldots,u_{iJt})$ is independently and identically distributed (iid) as standard type-I Extreme Value (EV); the product effects ξ_{jt} are unknown mean zero random variables conditional on a set of instruments; and the random variation in preferences $\boldsymbol{\epsilon}_i$ has a mean-zero distribution which is known up to its variance-covariance matrix $\boldsymbol{\Sigma}_0$. The $\boldsymbol{\epsilon}_i$ are often taken to be independent, identically distributed $N(0,\boldsymbol{\Sigma}_0)$ random vectors with a diagonal $\boldsymbol{\Sigma}_{\mathbf{0}}$.

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 $\boldsymbol{\beta}_0$ and the variance-covariance matrix of the random coefficients $\boldsymbol{\Sigma}_0$. We collect them in $\boldsymbol{\theta}_0=(\boldsymbol{\beta}_0,\boldsymbol{\Sigma}_0)$. 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 $\boldsymbol{\epsilon}$: for good $j\in\mathcal{J}$,

$$S_{jt}=E_{\boldsymbol{\epsilon}}\left[\frac{\exp\left(\boldsymbol{X}_{jt}\left(\boldsymbol{\beta}+\boldsymbol{\epsilon}\right)+\xi_{jt}\right)}{1+\Sigma_{k=1}^{J}\exp\left(\boldsymbol{X}_{kt}\left(\boldsymbol{\beta}+\boldsymbol{\epsilon}\right)+\xi_{kt}\right)}\right] \tag{2}$$

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 (2) defines an invertible mapping[5] in $\mathbb{R}^J$. Call $\Xi(\boldsymbol{S}_t, \boldsymbol{X}_t, \boldsymbol{\theta})$ 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:

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

over $\boldsymbol{\theta}$.

These models have proved very popular; but their implementation has faced a number of problems. Much 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:

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

As implementing the Amemiya formula relies on a consistent first-step estimate of $\boldsymbol{\theta}_0$, this is still problematic. Gandhi and Houde (2016) propose "differentiation IVs" to approximate the optimal instruments for the parameters $\boldsymbol{\Sigma}$ of the distribution of the random preferences $\boldsymbol{\epsilon}$. 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[6].

[5] See Berry (1994).

[6] Instruments that shift marginal cost directly (if available) do not need 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. The method we propose in this paper completely circumvents the need to solve a nonlinear optimization problem. 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 $\boldsymbol{\beta} + \boldsymbol{\epsilon}_i$ 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[7]. Second, our estimator needs to assume very little about the form of the distribution of the random variation in preferences $\boldsymbol{\epsilon}$ (beyond its small scale), justifying the "robust" in our title—where the quotes reflect

[7] 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—but it is constrained by linear inequalities and may be sensitive to the choice of the grid points.

our awareness that we are taking some liberties with the definition of robustness. Finally, because our estimating equation is linear, computing the "optimal" instruments for our estimator is also straightforward.

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 discuss below.

2 2SLS Estimation in the Standard BLP Model

For the reader primarily interested in applying our method, this section provides a step-by-step guide to implementing the estimator in the standard macro-BLP model. This requires some notation. The dimensions of the vectors and matrices are as follows:

- for each $j \in \mathcal{J}$ and t, $\boldsymbol{X}_{jt}$ is a row vector with n_X components
- $\boldsymbol{\beta}$ is a column vector with n_X components
- for each i, $\boldsymbol{\epsilon}_i$ is a row vector with n_e components; in the standard model, $n_e \leqslant n_X$.

We denote $\mathcal{I}$ as the set of pairs of ordered indices $(m, n \leqslant m)$ such that the variance-covariance element $\Sigma_{mn} = \text{cov}(\epsilon_{im}, \epsilon_{in})$ is *not* restricted to be zero[8]. For notational simplicity, we also 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.

Our procedure runs as follows:

[8] E.g. if $n_e = n_X$ and $\boldsymbol{\Sigma}$ is assumed to be diagonal, $\mathcal{I} = \{(1,1), \ldots, (n_X, n_X)\}$.

Algorithm 1. ***Fast Robust and Approximately Correct (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) in $\mathcal{I}$, compute the "artificial regressor" K_{mn}^{jt} as*

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

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

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 the columns of $\boldsymbol{Z}$. Define $\hat{\boldsymbol{\beta}}$ 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[9]) 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 optimal instruments $E(\boldsymbol{X}|\boldsymbol{Z})$ and $E(\boldsymbol{K}|\boldsymbol{Z})$ for our approximate model. Alternatively, these optimal instruments can be estimated by a nonparametric regressions of each the column of X on the columns of 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 unclear whether Step 4 is worth the additional effort. We intend to explore this question in future work.

[9] This step should only be considered when $T \gg J$.

It is important to note 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 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 Second-order Expansions

The rest of the paper justifies algorithm 1 and discusses extensions. We first derive the small-σ expansions of the introduction.

We start from a specification of the conditional indirect utility of variety j for consumer i on market t as

$$\boldsymbol{X}_{jt}\beta + g\left(\boldsymbol{X}_{jt}, \boldsymbol{\epsilon}_i\right) + \xi_{jt} + u_{ijt} \tag{3}$$

for $j \in \mathcal{J}$; and $U_{i0t} = u_{i0t}$. Define the vectors $\boldsymbol{u}_{it} = (u_{i0t}, u_{i1t}, \ldots, u_{iJt})$; $\boldsymbol{X}_t = (\boldsymbol{X}_{1t}, \ldots, \boldsymbol{X}_{Jt})$; and $\boldsymbol{\xi}_t = (\boldsymbol{\xi}_{1t}, \ldots, \boldsymbol{\xi}_{Jt})$. We assume that

1. the random terms $\boldsymbol{\epsilon}_i$ are i.i.d. across i with finite variance;

2. they are distributed independently of $(\boldsymbol{X}_t, \boldsymbol{\xi}_t)$;

3. $Eg(\boldsymbol{X}_{jt}, \boldsymbol{\epsilon}_i) = 0$ for all $\boldsymbol{X}_{jt}$;

4. the random vectors $\boldsymbol{u}_{it}$ are i.i.d. across i and t; and they are distributed independently of $(\boldsymbol{\epsilon}_i, \boldsymbol{X}_t, \boldsymbol{\xi}_t)$.

These assumptions are all standard, except for the third one which is only a mild extension of the usual normalization $E\boldsymbol{\epsilon}_i = 0$. They allow for any type of codependence between the product effects $\boldsymbol{\xi}_t$ and the covariates $\boldsymbol{X}_t$. Note that the additive separability between $\boldsymbol{\beta}$ and $\boldsymbol{\epsilon}$ is not as strict as it seems. If for instance we start from a multiplicative model with utilities

$$\sum_{k=1}^{n_X} X_{jtk}\beta_k\zeta_{ki} + \xi_{jt} + u_{ijt}$$

we can always redefine $\epsilon_{ki} = \beta_k(\zeta_{ki} - 1)$ to recover (3).

Our crucial assumption, which we maintain throughout, is that the utilities are affine in $\boldsymbol{\beta}$ and additive in the product effects $\boldsymbol{\xi}$ and in the idiosyncratic terms $\boldsymbol{u}$. On the other hand, we allow for any kind of distribution for $\boldsymbol{\epsilon}_i$ and $\boldsymbol{u}_{it}$. This encompasses most empirical specifications used, as well as many more. We will refer to three special cases for illustrative purposes:

1. The *standard model*, also known as the mixed multinomial logit model, has $g(\boldsymbol{X}, \boldsymbol{\epsilon}) = \boldsymbol{X}\boldsymbol{\epsilon}$; and the vector $\boldsymbol{u}_{it}$ is distributed as iid standard type-I Extreme Value random variables.

2. The *standard binary model* (or mixed logit model) further imposes $J = 1$.

3. The *standard symmetric model* is a standard model with $\boldsymbol{\epsilon}$ distributed symmetrically around $\mathbf{0}$;

4. The *standard Gaussian model* is a standard model with $\boldsymbol{\epsilon}$ normal with a diagnoral covariance matrix. It is probably the most commonly used in applications of the macro-BLP method.

5. Finally, the *standard Gaussian binary model* imposes both 2 and 4.

In order to do small-σ expansions, we need to introduce a scale parameter σ. We do this with Assumption 1, which fits the usual understanding of what a scale parameter is[10] and also imposes the restriction that all moments of $\boldsymbol{\epsilon}$ are finite-valued. The most common specification of the "macro-BLP" model has a Gaussian $\boldsymbol{\epsilon}$ and satisfies Assumption 1.

[10] In principle it should be possible to use several scale parameters, say σ_1 for one part of the variance-covariance matrix and σ_2 for another one.

Assumption 1 (Finite moments)**.** *For some integer $L \geqslant 2$, all moments of order $1 \leqslant l \leqslant L+1$ of the vector $\boldsymbol{\epsilon}$ are finite; they are of order l in some non-negative scalar σ. The first moment is zero: $E\boldsymbol{\epsilon} = \mathbf{0}$. We denote $\boldsymbol{\Sigma} = E\boldsymbol{\epsilon}\boldsymbol{\epsilon}'$ the variance-covariance matrix of $\boldsymbol{\epsilon}$, and $\boldsymbol{\mu}_l$ (for $l \geqslant 3$) its (uncentered) higher order moments.*

It will be convenient to write $\boldsymbol{\epsilon} \equiv \sigma \boldsymbol{B}\boldsymbol{v}$ with $\boldsymbol{v}$ a random vector of mean zero and covariance matrix equal to the identity matrix, so that $Var(\boldsymbol{\epsilon}) \equiv \boldsymbol{\Sigma} = \sigma^2 \boldsymbol{B}\boldsymbol{B}'$. We only use this decomposition for intermediate results. Note that $\boldsymbol{B}$ is an $n_e \times n_v$ matrix ($n_e \geqslant n_x$), where $\boldsymbol{v}$ is a row vector with n_v components. This allows for $\boldsymbol{\Sigma}$ to have a factor structure. Our final expansions do not depend on how σ and $\boldsymbol{B}$ are normalized, and we won't need to specify it.

We drop the index t from the notation in most of this section as we will only need to deal with one market at a time.

3.1 Second-order Expansions in the Standard Model

Much of the rest of the remainder of the paper focuses on the standard model, where the idiosyncratic error terms $\boldsymbol{u}$ have iid Type I extreme value distributions. We will show in section 6.4 how to extend our results to more general distributions.

Recall that in the standard model, market shares are given by (2). If the scale parameter σ was zero, inverting (2) would simply give us

$$\xi_j = \log \frac{S_j}{S_0} - \boldsymbol{X}_j \boldsymbol{\beta} \text{ for } j \in \mathcal{J}, \tag{4}$$

which is standard multinomial logit model without random coefficients. This is the starting point of the contraction algorithm described in Berry (1994).

Now let σ be positive. With $\boldsymbol{\epsilon} = \sigma \boldsymbol{B}\boldsymbol{v}$, a Taylor expansion of (4) at $\sigma = 0$ would give (assuming that the expansion is valid[11])

$$\xi_j = \log \frac{S_j}{S_0} - \boldsymbol{X}_j \boldsymbol{\beta} + \Sigma_{l=1}^{L} a_{lj}(\boldsymbol{S}, \boldsymbol{X}, \boldsymbol{\beta}) \frac{\sigma^l}{l!} + O(\sigma^{L+1}), \tag{5}$$

where the $a_{lj}(\boldsymbol{S}, \boldsymbol{X}, \boldsymbol{\beta})$ are defined below. In this equation, $\boldsymbol{X}$ collects the covariates of all products and $\boldsymbol{S}$ is the vector of market shares. Market-share weighted sums will play a crucial role in what follows:

[11] We return to this point in section 5.1.

Definition 1 (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

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

Finally, we denote $\hat{T}_j = T_j - e_{\boldsymbol{S}}\boldsymbol{T}$ and $\hat{\boldsymbol{m}}_j = \boldsymbol{m}_j - e_{\boldsymbol{S}}\boldsymbol{m}$.

Note that we are using the *observed* market shares of the J goods, so that these weighted sums are very easy to compute from the data. It is important to emphasize that the operator $e_{\boldsymbol{S}}$ is *not* an average, as the augmented market shares S_k for $k \in \mathcal{J}$ do not sum to one but to $(1 - S_0)$. Similarly, the $\hat{T}_j$ terms are not residuals, and $e_{\boldsymbol{S}}\hat{\boldsymbol{T}} \neq 0$ in general.

Our first goal is to find explicit formulæ for the coefficients a_{lj} in (5). While this can be done at a high level of generality, let us start with a result that covers a large majority of applications.

In the standard model, $g(\boldsymbol{X}_j, \boldsymbol{\epsilon})$ is simply $\boldsymbol{X}_j\boldsymbol{\epsilon}$. Using the identity $\boldsymbol{\epsilon} \equiv \sigma\boldsymbol{B}\boldsymbol{v}$, denote $\boldsymbol{x}_j = (\boldsymbol{X}_j\boldsymbol{B})'$, a vector of n_v components; and $\boldsymbol{x}$ the matrix whose J columns are $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_J)$. Then

$$g(\boldsymbol{X}_j, \boldsymbol{\epsilon}) = \sigma\boldsymbol{x}_j'\boldsymbol{v}.$$

We now use this notation to derive the second-order expansion in σ in the standard model.

Theorem 1 (Intermediate expansion in the standard model)**.** *In the standard model,*

(i) the a_{lj} coefficients only depend on $\boldsymbol{S}$ and on $\boldsymbol{x}$;

(ii) the first-order coefficients are zero: $a_{1j} \equiv 0$ for all j;

(iii) the second-order coefficients are given by

$$a_{2j} = 2\boldsymbol{x}_j \cdot e_{\boldsymbol{S}}\boldsymbol{x} - \|\boldsymbol{x}_j\|^2 = -\boldsymbol{x}_j \cdot \left(\boldsymbol{x}_j - 2\sum_{k=1}^{J} S_k\boldsymbol{x}_k\right); \qquad (6)$$

(iv) *in the standard symmetric model,* $a_{lj} = 0$ *for all* j *and odd* $l \leqslant L$*. Therefore if* $L \geqslant 3$*,*

$$\xi_j = \log \frac{S_j}{S_0} - \boldsymbol{X}_j \boldsymbol{\beta} + \frac{a_{2j}}{2} \sigma^2 + O(\sigma^4). \tag{7}$$

Proof. See Appendix A. □

3.2 The Artificial Regressors in the Standard Model

When truncated of its remainder term, equation (7) becomes linear in the parameters $(\boldsymbol{\beta}, \sigma^2)$. The coefficients a_{2j}, however, are quadratic combinations of the vectors $\boldsymbol{x}_j$, which are themselves linear in the unknown coefficients of the matrix $\boldsymbol{B}$. Fortunately, the formula that gives a_{2j} can be transformed so that it becomes linear in the coefficients of the variance-covariance matrix $\boldsymbol{\Sigma}$ of $\boldsymbol{\epsilon}$.

To see this, note that since $\boldsymbol{x}_k = \boldsymbol{B}' \boldsymbol{X}_k'$,

$$\boldsymbol{x}_k' \boldsymbol{x}_l = \boldsymbol{X}_k \boldsymbol{B} \boldsymbol{B}' \boldsymbol{X}_l'.$$

But because $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{B} \boldsymbol{B}'$, we have

$$\sigma^2 \boldsymbol{x}_k' \boldsymbol{x}_l = \sum_{m,n=1}^{n_X} \Sigma_{mn} X_{km} X_{ln} = \operatorname{Tr} \left(\boldsymbol{\Sigma} \boldsymbol{X}_l \boldsymbol{X}_k' \right)$$

where $\operatorname{Tr}(\cdot)$ is the trace operator.

Plugging this into (6) gives

$$\sigma^2 \frac{a_{2j}}{2} = \operatorname{Tr} \left(\boldsymbol{\Sigma} \left(e_{\boldsymbol{S}} \boldsymbol{X} - \frac{\boldsymbol{X}_j}{2} \right) \boldsymbol{X}_j' \right).$$

Define the $n_X \times n_X$ matrices $\boldsymbol{\mathcal{K}}^j$ by

$$\boldsymbol{\mathcal{K}}^j = \left(\frac{\boldsymbol{X}_j}{2} - e_{\boldsymbol{S}} \boldsymbol{X} \right) \boldsymbol{X}_j'$$

so that we can also write $\sigma^2 \frac{a_{2j}}{2} = -\operatorname{Tr}[\boldsymbol{\Sigma} \boldsymbol{\mathcal{K}}^j]$. The matrices $\boldsymbol{\mathcal{K}}^j$ can be constructed straightforwardly from the covariates $\boldsymbol{X}$ and the market shares $\boldsymbol{S}$. Given that $\boldsymbol{\Sigma}$ is symmetric,

$$\operatorname{Tr}[\boldsymbol{\Sigma} \boldsymbol{\mathcal{K}}^j] = \sum_{m=1}^{n_X} \Sigma_{mm} \mathcal{K}_{mm}^j + \sum_{m,n<m} \Sigma_{mn} \left(\mathcal{K}_{mn}^j + \mathcal{K}_{nm}^j \right). \tag{8}$$

Define the lower-triangular matrices $\boldsymbol{K}^j$ by

- on the diagonal: $\boldsymbol{K}^j_{mm} = \boldsymbol{\mathcal{K}}^j_{mm}$
- off the diagonal (for $n < m$): $\boldsymbol{K}^j_{mn} = \boldsymbol{\mathcal{K}}^j_{mn} + \boldsymbol{\mathcal{K}}^j_{nm}$.

We call their elements the "artificial regressors", for reasons that will soon become clear.

Additional *a priori* restrictions can be accommodated very easily. For instance, it is common to restrict $\boldsymbol{\Sigma}$ to be diagonal, as is the case in Berry et al. (1995). Then only n_X terms enter in the sum (8); moreover,

$$K^j_{mm} = \left(\frac{X_{jm}}{2} - \sum_{k=1}^J S_k X_{km}\right) X_{jm}.$$

If $\boldsymbol{\Sigma}$ is not diagonal, then we need to also use additional terms

$$K^j_{mn} = X_{jm}X_{jn} - \left(\sum_{k=1}^J S_j X_{km}\right) X_{jn} - \left(\sum_{k=1}^J S_j X_{kn}\right) X_{jm}.$$

To summarize, we have:

Theorem 2 (Final expansion in the standard model)**.** *In the standard model,*

$$\xi_j = \log \frac{S_j}{S_0} - \boldsymbol{X}_j \boldsymbol{\beta} - \sum_{m=1}^{n_X} \Sigma_{mm} K^j_{mm} - \sum_{n<m} \Sigma_{mn} K^j_{mn} + O(\|\Sigma\|^{k/2}), \tag{9}$$

where the integer $k \geqslant 3$; and $k = 4$ if the model is symmetric.

The artificial regressors are given by

$$\begin{aligned} K^j_{mm} &= \left(\frac{X_{jm}}{2} - \sum_{k=1}^J S_k X_{km}\right) X_{jm} \\ K^j_{mn} &= X_{jm}X_{jn} - \left(\sum_{k=1}^J S_j X_{km}\right) X_{jn} - \left(\sum_{k=1}^J S_j X_{kn}\right) X_{jm}. \end{aligned}$$

4 2SLS Estimation

Equation (9) is linear in the parameters of interest $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\Sigma})$, up to the remainder term. This immediately suggests neglecting the remainder term and estimating the approximate model $\xi_j = \log \frac{S_j}{S_0} - \boldsymbol{X}_j \boldsymbol{\beta} - \text{Tr}[\boldsymbol{\Sigma} \boldsymbol{K}^j]$.

More precisely, assume we are given a sample of T markets, and instruments $\boldsymbol{Z}_{jt}$ such that $E\left(\xi_{jt}|\boldsymbol{Z}_{jt}\right)$ for all j and t. Then our proposed estimator $\hat{\boldsymbol{\theta}}$ fits the approximate *linear* set of conditional moment restrictions:

$$E\left(\log\left[\frac{S_{jt}}{S_{0t}}\right] - \left(\boldsymbol{X}_{jt}\boldsymbol{\beta} + \mathrm{Tr}[\boldsymbol{\Sigma}\boldsymbol{K}^{jt}]\right)|\boldsymbol{Z}_{jt}\right) = 0 \tag{10}$$

which only differs from the original model by a term of order σ^3 (or σ^4 if the distribution of $\boldsymbol{\epsilon}$ is symmetric)[12]. This can simply be done by choosing vector functions $\boldsymbol{Z}^*_{jt}$ of the instruments and running two-stage least squares: for each $j = 1, \dots, J$ and $t = 1, \dots, T$, we run a 2SLS regression of $\log(S_{jt}/S_{0t})$ on $\boldsymbol{X}_{jt}$ and the relevant[13] variables $\boldsymbol{K}^{jt}$, using the instruments $\boldsymbol{Z}^*_{jt}$.

5 Pros and Cons of the 2SLS Estimation Approach

The drawback of our method is obvious: 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 5.1. 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 7. 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 $\boldsymbol{\epsilon}$. This is a notable advantage over other methods: while they yield inconsistent estimates if the distribution of $\boldsymbol{\epsilon}$ is misspecified, our estimator remains consistent for the parameters of the approximate model.

[12] Note that because our model is only an approximation, there may not exist a value of $\boldsymbol{\theta}$ that satisfies the conditional moment restrictions (10) in the population. Nevertheless, our 2SLS estimation procedure will still yield an estimate of the value of $\boldsymbol{\theta}$ that minimizes the probability limit of the 2SLS objective function. We explore this further in our Monte Carlo study in section 7.

[13] E.g. only the n_X variables K^{jt}_{mm} if $\boldsymbol{\Sigma}$ is restricted to be diagonal, or even a subset if some coefficients are non-random.

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 $\boldsymbol{\beta}$ and their random variation $\boldsymbol{\Sigma}$. 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 2018 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.

5.1 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 $\boldsymbol{\epsilon}$ are finite; and that the expansion is valid at order L if the moments of $\boldsymbol{\epsilon}$ are bounded to order $(L+1)$.

Characterizing this range of validity is trickier. Figure 1 uses formulæ derived in Appendix B 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.

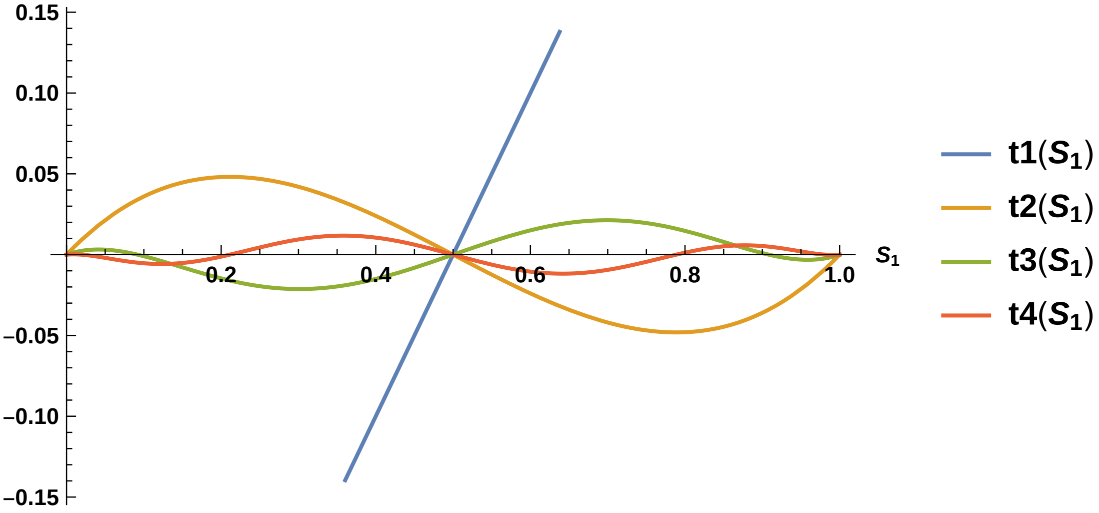


Figure 1: Coefficients $t_{1,2,3,4}(S_1)$

While this simple example can only be illustrative, we find the figure encouraging as to the practical range of validity of the approximation. We have not determined the extent to which these results generalize to the multinomial logit model.

5.2 "Robustness"

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 $\boldsymbol{\epsilon}$. This has a distinct advantage over competing methods: the lower-order moments of $\boldsymbol{\epsilon}$ 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 $\boldsymbol{\epsilon}$ has a symmetric distribution. Then the coefficients a_{1j} are still zero, and the coefficients a_{2j} are unchanged. In the absence of symmetry, the approximate model is only valid up to $O(\sigma^3)$; but running Algorithm 1 may still provide very useful estimators of the elements of $\boldsymbol{\Sigma}_0$.

6 Extensions

Our technique can be extended to other random coefficient models as long as the conditional indirect utility remains additive in the product-specific random effects $\boldsymbol{\xi}$. This is shown in section 6.1. We follow with a modification of our multinomial logit random coefficients modeling framework to account for the third and fourth moments of $\boldsymbol{\epsilon}$. We then turn to methods for improving the quality of our 2SLS estimates[14]. Finally, section 6.4 presents a nonlinear 2SLS estimation procedure for the nested logit model.

6.1 Quasi-linear Random Coefficients Models

Consider the following class of models, whose defining characteristic is that the error term $\boldsymbol{\eta}$ and the mean coefficients $\boldsymbol{\beta}$ only enter via a linear combination $\boldsymbol{\eta} - \boldsymbol{f_1}(\boldsymbol{y})\boldsymbol{\beta}$:

$$\boldsymbol{G}(\boldsymbol{y}, \boldsymbol{\eta}, \boldsymbol{\beta}, \sigma) \equiv \boldsymbol{G}^* \left(\boldsymbol{y}, E_{\boldsymbol{v}} \boldsymbol{A}^*(\boldsymbol{y}, \boldsymbol{\eta} - \boldsymbol{f_1}(\boldsymbol{y})\boldsymbol{\beta}, \sigma \boldsymbol{B}\boldsymbol{v})\right). \tag{11}$$

where $\boldsymbol{v}$ is unobserved heterogeneity distributed independently of $\boldsymbol{y}$ and $\boldsymbol{\eta}$ and normalized by $E\boldsymbol{v} = \boldsymbol{0}$ and $V\boldsymbol{v} = \boldsymbol{I}$; and both functions $\boldsymbol{G}^*$ and $\boldsymbol{A}^*$ are assumed to be known.

[14] We explore the small sample properties of several of these corrections in our Monte Carlo study.

Note that the macro–BLP model takes this form, with $\boldsymbol{y} = (\boldsymbol{S}, \boldsymbol{X})$; $\boldsymbol{f_1}(\boldsymbol{y}) = -\boldsymbol{X}$; $\boldsymbol{\eta} = \boldsymbol{\xi}$; and

$$A_j^* = \Pr\left(j = \arg\max_{J=0,1,\dots,J} \left(\boldsymbol{X}_k\boldsymbol{\beta} + \xi_k + \sigma \boldsymbol{X_k B v}\right) | \boldsymbol{X}, \boldsymbol{\xi}, \boldsymbol{v}\right)$$

so that, denoting $\boldsymbol{a}_j \equiv \boldsymbol{X_j}$ and $b_j = \boldsymbol{X}_j\boldsymbol{\beta} + \xi_j$,

$$A^*(a, b, c) \equiv \frac{\exp\left(b_j + \boldsymbol{a_j c}\right)}{1 + \sum_{k=1}^J \exp\left(b_k + \boldsymbol{a_k c}\right)};$$

and $G_j^* \equiv S_j - E_{\boldsymbol{v}} A_j^*$.

We continue to assume that $E\left(\boldsymbol{\eta}|\boldsymbol{Z}\right) = \boldsymbol{0}$. The quasi-linear structure in (11) allows this class of models to be approximately estimated by 2SLS. Denoting partial derivatives with subscripts, we have:

Theorem 3 (Expansions for quasi-linear random coefficients models)**.** *Consider a model of the class defined by* (11) *and assume that*

- $\boldsymbol{G}^*$ *is twice differentiable with respect to its second argument*
- $\boldsymbol{A}^*$ *is twice differentiable with respect to its last two arguments*
- *the matrices* $\boldsymbol{G}_2^*(\boldsymbol{y}, \boldsymbol{A}^*(\boldsymbol{y}, \boldsymbol{\eta} - \boldsymbol{f_1}(\boldsymbol{y})\boldsymbol{\beta}, \boldsymbol{0}))$ *and* $\boldsymbol{A}_2^*(\boldsymbol{y}, \boldsymbol{\eta} - \boldsymbol{f_1}(\boldsymbol{y})\boldsymbol{\beta}, \boldsymbol{0})$ *are invertible for all* $(\boldsymbol{y}, \boldsymbol{\eta}, \boldsymbol{\beta})$.

Any such model satisfies the conditions **C1**–**C3** *in the introduction. Moreover,*

- $\boldsymbol{f_1}(\boldsymbol{y})$ *appears directly in* (11)
- *the variables* $\boldsymbol{f_0}(\boldsymbol{y})$ *are defined by the system of equations*

$$\boldsymbol{G}^*(\boldsymbol{y}, \boldsymbol{A}^*(\boldsymbol{y}, \boldsymbol{f_0}(\boldsymbol{y}), \boldsymbol{0})) = \boldsymbol{0}$$

- *and the variables* $\boldsymbol{f_2}(\boldsymbol{y})$ *solve the linear system*

$$\boldsymbol{A}_{33}^*(\boldsymbol{y}, \boldsymbol{f_0}(\boldsymbol{y}), \boldsymbol{0}))\ \boldsymbol{f_2}(\boldsymbol{y}) = -\boldsymbol{A}_2^*(\boldsymbol{y}, \boldsymbol{f_0}(\boldsymbol{y}), \boldsymbol{0})).$$

Proof: See Appendix E.

As explained in the introduction, these models can be estimated by regressing $f_0(y)$ on $f_1(y)$ and $f_2(y)$ with a set of flexible functions of $\boldsymbol{Z}$ as instruments. As the macro–BLP model belongs to this class, this confirms that conditions **C1**–**C3** hold in the BLP model; we had shown it implicitly in section 3 by deriving the expansions. Note also that we did not use *any* distributional assumption on the random coefficients and the idiosyncratic shocks—although of course the terms in the expansions do depend on these distributions. We give an illustration for a one-covariate mixed binary model without any distributional assumption in Appendix B.3.

6.1.1 Examples

It is easy to generate models in the quasi-linear class (11). Starting from any Genesralized Linear Model $g(y) = \boldsymbol{X}\boldsymbol{\beta} + \eta$, we can for instance transform the right-hand side by adding additive unobserved heterogeneity and another link function:

$$g(y) = E_\varepsilon h\left(\boldsymbol{X}\boldsymbol{\beta} + \eta, \sigma\varepsilon\right).$$

When the link functions g and h are both assumed to be known, all such models obey conditions **C1**–**C3** and can therefore be studied with our method. Note that in these models $f_1(y) \equiv -\boldsymbol{X}$ and $f_2(y) = -(h_1/h_{22})(f_0(y), 0)$ where

$$f_0(y) = h(\cdot, 0)^{-1}(g(y))$$

(assuming the inverse is well-defined.)

The nested logit of section 6.4 shows that our method remains useful beyond the class of quasi-linear models, at the cost of breaking conditions **C2** and **C3** and requiring (simple) numerical optimization.

6.2 Higher-order terms

In Appendix B, 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 the ϵ_m are independent across the covariates $m = 1, \ldots, n_X$. We denote as before $\Sigma_{mm} = E(\epsilon_m^2)$, and μ_{lm} the expected value of ϵ_m^l for $l \geqslant 3$. Tedious calculations[15] show that the second- to fourth-order terms of the expansion in σ are

$$\xi_j = \log \frac{S_j}{S_0} - \boldsymbol{X}_j \beta + \sum_{l=2}^{4} A_{lj} + O(\sigma^5)$$

with

$$A_{2j} = \sum_m X_{jm} \left(e_{\boldsymbol{S}} \boldsymbol{X_m} - X_{jm}/2 \right) \Sigma_{mm};$$

$$A_{3j} = \sum_m X_{jm} \left(X_{jm} \frac{e_{\boldsymbol{S}} \boldsymbol{X}_m}{2} + \frac{e_{\boldsymbol{S}} \left(\boldsymbol{X}_m^2 \right)}{2} - \frac{X_{jm}^2}{6} - \left(e_{\boldsymbol{S}} \boldsymbol{X}_m \right)^2 \right) \mu_{3m};$$

and

$$\begin{aligned} A_{4j} = \sum_m \mu_{4m} X_{jm} \Bigg(& (e_{\boldsymbol{S}} \boldsymbol{X}_m)^3 - (e_{\boldsymbol{S}} \boldsymbol{X}_m)(e_{\boldsymbol{S}}(\boldsymbol{X}_m^2)) - X_{jm} \frac{(e_{\boldsymbol{S}} \boldsymbol{X}_m)^2}{2} - \frac{X_{jm}^3}{24} \\ & + \frac{e_{\boldsymbol{S}}(\boldsymbol{X}_m^3)}{6} + X_{jm} \frac{e_{\boldsymbol{S}}(\boldsymbol{X}_m^2)}{4} + X_{jm}^2 \frac{e_{\boldsymbol{S}} \boldsymbol{X}_m}{6} \Bigg) \\ & + \frac{A_{2j}^2}{2} + \sum_m \Sigma_{mm} X_{jm} \left(e_S(\boldsymbol{A_2 X_m}) + (e_S \boldsymbol{A_2}) \left(\frac{X_{jm}}{2} - 2(e_{\boldsymbol{S}} \boldsymbol{X}_m) \right) \right). \end{aligned}$$

First consider the third-order term A_{3j}. It is a linear function of the unknown skewnesses μ_{3m}; in fact it can be rewritten as

$$-\sum_m T_m^j \mu_{3m}$$

where we introduced new artificial regressors

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

Algorithm 1 can be adapted in the obvious way to take possible skewness of $\boldsymbol{\epsilon}$ into account. Note that the procedure remains linear in the parameters $(\boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\mu_3})$, for which it generates approximate estimates by 2SLS.

[15] Available from the authors.

The fourth-order term, on the other hand, contains terms that are linear in the μ_{4m} (the first two lines of the formula) as well as terms that are quadratic in $\boldsymbol{\Sigma}$ (the last line). The first group suggests introducing more artificial regressors

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

whose coefficients are the μ_{4m}. The second group yields

$$-\sum_{m,n=1}\Sigma_{mm}\Sigma_{nn}W^j_{mn}$$

where 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).

6.3 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.

6.3.1 Bootstrapping

Once we have approximate estimators $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\Sigma}}$, 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 $\boldsymbol{v}$ (beyond the normalization of its first two moments.)

We use Berry inversion to solve for $\hat{\boldsymbol{\xi}}_t$ in the system

$$S_{jt} = E_{\boldsymbol{v}}\,\frac{\exp\left(\boldsymbol{X}_{jt}\left(\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\Sigma}}^{1/2}\boldsymbol{v}\right) + \hat{\xi}_{jt}\right)}{1 + \Sigma^J_{k=1}\exp\left(\boldsymbol{X}_{kt}\left(\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\Sigma}}^{1/2}\boldsymbol{v}\right) + \hat{\xi}_{kt}\right)},$$

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

$$S^*_{jt} = E_{\boldsymbol{v}}\,\frac{\exp\left(\boldsymbol{X}_{jt}\left(\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\Sigma}}^{1/2}\boldsymbol{v}\right) + \xi^*_{jt}\right)}{1 + \Sigma^J_{k=1}\exp\left(\boldsymbol{X}_{kt}\left(\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\Sigma}}^{1/2}\boldsymbol{v}\right) + \xi^*_{kt}\right)}$$

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

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

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

6.3.2 Asymptotic Correction

Another way to use the third- and fourth-order terms in our expansion is as a corrective term: that is, we run 2SLS on the second-order expansion and we use the formulæ for the higher-order terms to correct for the effects of the approximation.

Denote $\boldsymbol{\theta} = (\boldsymbol{\Sigma}, \boldsymbol{\beta})$, 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 $E(\boldsymbol{\xi}_2 \boldsymbol{Z}) = 0$ with instruments $\boldsymbol{Z}$ and weighting matrix $\boldsymbol{W}$, where

$$\boldsymbol{\xi}_{2j} = \log \frac{S_j}{S_0} - \boldsymbol{X}_j \boldsymbol{\beta} - \operatorname{Tr} \boldsymbol{\Sigma} \boldsymbol{K}^{\boldsymbol{j}}. \tag{12}$$

As the number of markets T gets large, $\hat{\boldsymbol{\theta}}_2$ converges to the solution $\boldsymbol{\theta}_2$ of $Ef_2(\boldsymbol{\theta}_2) = \boldsymbol{0}$, with

$$f_2(\boldsymbol{\theta}) \equiv \frac{\partial \boldsymbol{\xi}_2}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta}, \boldsymbol{X}, \boldsymbol{S})' \boldsymbol{Z} \boldsymbol{W} \boldsymbol{Z}' \boldsymbol{\xi}_2(\boldsymbol{\theta}, \boldsymbol{X}, \boldsymbol{S}).$$

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 $\boldsymbol{\epsilon}$) that are identified using the exact $\boldsymbol{\xi}_\infty$ but not[16] 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 $Ef_\infty(\boldsymbol{\theta}_0; \boldsymbol{\lambda}_0) = \boldsymbol{0}$ with

$$f_\infty(\boldsymbol{\theta}; \boldsymbol{\lambda}_0) \equiv \frac{\partial \boldsymbol{\xi}_\infty}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta}, \boldsymbol{\lambda}_0, \boldsymbol{X}, \boldsymbol{S})' \boldsymbol{Z} \boldsymbol{W} \boldsymbol{Z}' \boldsymbol{\xi}_\infty(\boldsymbol{\theta}, \boldsymbol{\lambda}_0, \boldsymbol{X}, \boldsymbol{S}).$$

The dominant term in the asymptotic bias is given by expanding $Ef_\infty(\boldsymbol{\theta}; \boldsymbol{\lambda}_0)$ around $\boldsymbol{\theta} = \boldsymbol{\theta}_2$, keeping $\boldsymbol{\lambda}_0$ fixed. It is

$$\boldsymbol{\theta}_2 - \boldsymbol{\theta}_0 \simeq \left(E \frac{\partial f_\infty}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta}_2; \boldsymbol{\lambda}_0) \right)^{-1} Ef_\infty(\boldsymbol{\theta}_2; \boldsymbol{\lambda}_0).$$

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

Denote $\boldsymbol{X}$ the matrix with terms X_{jm} and $\boldsymbol{K}$ the matrix whose row $j = 1, \dots, J$ contains the artificial regressors K^j_{mn}. We define $\boldsymbol{e}_2(\boldsymbol{\theta}; \boldsymbol{\lambda}_0) = \boldsymbol{\xi}_\infty(\boldsymbol{\theta}; \boldsymbol{\lambda}_0) - \boldsymbol{\xi}_2(\boldsymbol{\theta})$, the approximation error on $\boldsymbol{\xi}$. Under any assumption about the parameters in $\boldsymbol{\lambda}_0$, we can compute the higher-order terms $\boldsymbol{\xi}_3, \boldsymbol{\xi}_4, \dots$ to approximate $\boldsymbol{e}_2$. If for instance we maintain the assumption that the model is symmetric, we can approximate $\boldsymbol{e}_2 \simeq \boldsymbol{\xi}_4 - \boldsymbol{\xi}_2$. Results in Robinson (1988) and in Kristensen and Salanié (2017) could be adapted to show that the resulting corrected estimator not only has a smaller bias; it is in fact asymptotically equivalent to the estimator $\hat{\boldsymbol{\theta}}_4$ based on a fourth-order expansion.

Let us suppose then that we have a reliable estimator $\hat{\boldsymbol{e}}_2(\boldsymbol{\theta}; \boldsymbol{\lambda}_0)$ of $\boldsymbol{e}_2(\boldsymbol{\theta}; \boldsymbol{\lambda}_0)$. Define $\boldsymbol{V}$ by the Cholesky decomposition $\boldsymbol{Z}\boldsymbol{W}\boldsymbol{Z}' = \boldsymbol{V}\boldsymbol{V}'$, so that $\boldsymbol{V}$ is a $(J \times J)$ matrix. We prove in Appendix D that this asymptotic correction yields the following formula:

$$\boldsymbol{\theta}_0 \simeq \boldsymbol{\theta}_2 + \begin{pmatrix} E\left(\boldsymbol{X}'\boldsymbol{V}\boldsymbol{V}'\boldsymbol{X}\right) & E\left(\boldsymbol{X}'\boldsymbol{V}\boldsymbol{V}'\boldsymbol{K}\right) \\ E\left(\boldsymbol{K}'\boldsymbol{V}\boldsymbol{V}'\boldsymbol{X}\right) & E\left(\boldsymbol{K}'\boldsymbol{V}\boldsymbol{V}'\boldsymbol{K}\right) \end{pmatrix}^{-1} \begin{pmatrix} E\left(\boldsymbol{X}'\boldsymbol{V}\boldsymbol{V}'\hat{\boldsymbol{e}}_2\right) \\ E\left(\boldsymbol{K}'\boldsymbol{V}\boldsymbol{V}'\hat{\boldsymbol{e}}_2\right) - E\left(\frac{\partial \hat{\boldsymbol{e}}_{\boldsymbol{2}}}{\partial \boldsymbol{\Sigma}}\boldsymbol{V}\boldsymbol{V}'\boldsymbol{\xi}_2\right) \end{pmatrix}.$$

To interpret this formula, note that if $\hat{\boldsymbol{e}}_2$ did not depend on $\boldsymbol{\Sigma}$ the corrective term on the right-hand-side would simply be the 2SLS estimate of the regression of $\hat{\boldsymbol{e}}_{\boldsymbol{2}}$ on $(\boldsymbol{X}, \boldsymbol{K})$ with instruments $\boldsymbol{V}$. In fact, if we are only interested in correcting the estimators of the mean coefficients $\boldsymbol{\beta}_2$, we can simply keep the corresponding part of the 2SLS estimate. The correction on $\boldsymbol{\Sigma}_2$ has an additional term as higher order terms in the expansion of $\boldsymbol{\xi}$ typically depend on $\boldsymbol{\Sigma}$. (Recall from Theorem 1.(i) that they do not depend on $\boldsymbol{\beta}$.)

6.3.3 Two-Step Estimator Based on the Asymptotic Correction

In practice, we use the estimate of $\boldsymbol{\xi}_\infty(\hat{\boldsymbol{\theta}}_2)$ computed using our 2SLS estimator and the Berry inversion as well as the residual from our 2SLS estimate

$$\boldsymbol{\xi}_{2j}(\hat{\boldsymbol{\theta}}_2) = \log \frac{S_j}{S_0} - \boldsymbol{X}_j \hat{\boldsymbol{\beta}}_2 - \operatorname{Tr} \hat{\boldsymbol{\Sigma}}_2 \boldsymbol{K}^{\boldsymbol{j}}$$

to correct our initial 2SLS estimate. The difference between $\boldsymbol{\xi}_{2j}(\boldsymbol{\theta_0})$ and $\boldsymbol{\xi}_\infty(\boldsymbol{\theta_0})$ is equal to the higher-order terms in equation (5). Computing a new dependent variable

$$y_j = \log\left[\frac{S_j}{S_0}\right] - \left(\boldsymbol{\xi}_{2j}(\boldsymbol{\theta_0}) - \boldsymbol{\xi}_\infty(\boldsymbol{\theta_0})\right)$$

and applying 2SLS would yield a consistent estimate of $\boldsymbol{\theta_0}$. Using our 2SLS estimate $\hat{\boldsymbol{\theta}}_2$ and computing

$$y_j^* = \log\left[\frac{S_j}{S_0}\right] - (\boldsymbol{\xi}_{2j}(\hat{\boldsymbol{\theta}}_2) - \boldsymbol{\xi}_\infty(\hat{\boldsymbol{\theta}}_2))$$

and applying 2SLS with this dependent variable is a feasible version of this correction[17]. In our Monte Carlo study we will explore the small sample properties of this procedure.

6.4 The Two-level Nested Logit

Campioni (2018) 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 B.3. 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 the estimating equation for the two-level nested logit model. Assume that there is a nest for good 0, and K nests $N_1, \dots, N_K$ for the varieties of the good. For $k = 1, \dots, K$, 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 $\boldsymbol{T} = (T_0, T_1, \dots, T_J)$. 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 $e_{\boldsymbol{S}}\boldsymbol{T} = \sum_{k=1}^K S_{N_k} \bar{T}_k$.

[17] Note that this procedure could be iterated.

Appendix C derives the equivalent of (6): for $j \in N_k$,

$$a_{2j} = \boldsymbol{x}_j \cdot \left(2e_{\boldsymbol{S}}\boldsymbol{x} - \frac{\boldsymbol{x}_j}{\lambda_k} + 2\frac{1-\lambda_k}{\lambda_k}\bar{\boldsymbol{x}}_k\right) - \frac{1-\lambda_k}{\lambda_k}\|\bar{\boldsymbol{x}}_k\|^2.$$

Reintroducing the market index t, the corresponding artificial regressors are

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

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

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

where as in section 3, $e_{tm} = \sum_{j=1}^{J} S_{jt}X_{jtm}$.

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

Algorithm 2. *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 product-market pair $(j \in \mathcal{J}, t)$:*
   (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*

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

   where $k(j)$ is the nest that variety j belongs to.

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

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

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

   $$K^{jt}_{mn} = X_{jt,m}X_{jt,n} - \frac{1-S_{0t}\lambda_{k(j)}}{1-S_{0t}}\frac{e_{tm}X_{jt,n} + e_{tn}X_{jt,m}}{\lambda_{k(j)}} + 2\frac{1-\lambda_{k(j)}}{\lambda_{k(j)}}\left(\bar{X}_{k(j),t,m}\bar{X}_{k(j),t,n} - \frac{\bar{X}_{k(j),t,m}X_{jt,n} + \bar{X}_{k(j),t,n}X_{jt,m}}{\lambda_{k(j)}}\right).$$

(c) define

$$y_{jt} = \log \frac{S_{N_{k(j)},t}}{S_{0t}} + \lambda_{k(j)} \log \frac{S_{jt}}{S_{N_{k(j)},t}}$$

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 43.*

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. Estimation of $(\boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$ requires numerical minimization over the $\boldsymbol{\lambda}$.

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[18]. 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.

7 Monte Carlo Analysis of the Small-Sample Performance of our Estimator

This section presents the results of a Monte Carlo study of an aggregate discrete choice demand system with random coefficients. It compares the finite sample performance of our estimator of the parameters to estimators computed using the mathematical programming with equilibrium constraints (MPEC) approach recommended by Dubé, Fox and Su (2012). We also show results demonstrating some of the "robustness" of our estimation procedure to assumptions about the distribution of the random coefficients. Specifically, we find that even if the distribution of random coefficients is misspecified, our procedure still yields very good estimates of the means and variances of the random coefficients.

[18] Technically, condition **C1** in the introduction still holds, but conditions **C2** and **C3** do not.

The basic set-up of our Monte Carlo study follows that in Dubé, Fox and Su (2012). It is a standard static aggregate discrete choice random coefficients demand system with $T = 50$ markets and $J = 25$ products in each market, and $K = 3$ observed product characteristics. Following Dubé, Fox, and Su (2012), let M_t denote the mass of consumers in market $t = 1, 2, \ldots, T$. Each product is characterized by the vector $(\boldsymbol{X}'_{jt}, \xi_{jt}, p_{jt})'$, where $\boldsymbol{X}_{jt}$ is a $K \times 1$ vector of observable attributes of product $j = 1, 2, \ldots, J$ in market t, ξ_{jt} is the vertical product characteristic of product j in market t that is observed by producers and consumers, but unobserved by the econometrician, and p_{jt} is the price of product j in market t. Collect these variables for each product into 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

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

The utility of the $j = 0$ good, the "outside" good, is equal to $u_{0jt} = \epsilon_{i0t}$. Each element of $\boldsymbol{\beta}^x_i = (\beta^x_{i1}, \ldots, \beta^x_{iK})'$ is assumed to be drawn independently from $N(\bar{\beta}^x_k, \sigma^2_k)$ distributions, and each β^p_i is assumed to be drawn independently from $N(\bar{\beta}_p, \sigma^2_p)$. We denote $\boldsymbol{\beta}_i = (\boldsymbol{\beta}^{x\prime}_i, \beta^p_i)'$.

We collect all parameters into

$$\boldsymbol{\theta} = (\beta_0, \bar{\beta}^x_1, \ldots, \bar{\beta}^x_K, \bar{\beta}_p, \sigma^2_1, \ldots, \sigma^2_K, \sigma^2_p)'.$$

Our simulations all have mean coefficients

$$(\beta_0, \bar{\beta}^x_1, \bar{\beta}^x_2, \bar{\beta}^x_3, \bar{\beta}_p) = (-1, 1.5, 1.5, 0.5, -1).$$

and varying variances $(\sigma^2_1, \sigma^2_2, \sigma^2_3, \sigma^2_p)$. We also experiment with varying $\bar{\beta}_p$.

To compute the market shares for the J products, we assume that the ϵ_{ijt} are independently and identically distributed Type I extreme value random variables, so that the probability that consumer i with random preferences $\boldsymbol{\beta}_i$ purchases good j in market t is equal to

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

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

$$\beta^x_{ikt} = \bar{\beta}^x_k + \sigma_k \zeta_{ikt} \text{ and } \beta^p_{it} = \bar{\beta}^p + \sigma_p \zeta_{ipt}.$$

We then use these draws to compute the observed market share of good j in market t as:

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

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

Consistent with the experimental design in Dubé, Fox and Su (2012), we generate the values of $\boldsymbol{X}_t$, $\boldsymbol{p}_t$, $\boldsymbol{\xi}_t$ and a vector of 6 instruments $\boldsymbol{Z}_{jt}$ as follows. First we draw $\boldsymbol{X}_t$ for all markets $t = 1, 2, \ldots, T$ from a multivariate normal distribution:

$$\begin{bmatrix} x_{1j} \\ x_{2j} \\ x_{3j} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & -0.8 & 0.3 \\ -0.8 & 1 & 0.3 \\ 0.3 & 0.3 & 1 \end{bmatrix} \right)$$

The price of good j in market t is equal to

$$p_{jt} = |0.5\xi_{jt} + e_{jt} + 1.1(x_{1j} + x_{2j} + x_{3j})|,$$

where $e_{jt} \sim N(0,1)$, distributed independently across products and markets. The ξ_{jt} are $N(0, \sigma^2_\xi)$ random variables drawn independently across products and markets for different values of σ^2_ξ described below. The data generating process for the vector of instruments is:

$$z_{jtd} \sim U(0,1) + 0.25(e_{jt} + 1.1(x_{1j} + x_{2j} + x_{3j}))$$

where $d = 1, \ldots, 6$.

For a specified value of the parameter vector θ, following this process for $T = 50$ markets yields the dataset for one Monte Carlo draw.

7.1 MPEC Approach

The MPEC approach solves a nonlinear minimization problem subject to nonlinear equilibrium constraints. The first step of the estimation process constructs the following instrumental variables for all the products in all the markets. There are 42 instruments in total; they are constructed from product characteristics x_j and excluded instruments z_{jt}:

$$1, x_{kj}, x_{kj}^2, x_{kj}^3, x_{1j}x_{2j}x_{3j}, z_{jtd}, z_{jtd}^2, z_{jtd}^3, z_{jtd}x_{1j}, z_{jtd}x_{2j}, \prod_{d=1}^{6} z_{jtd}$$

Let W denote this $(J \times T) \times 42$ matrix of instruments. In our case $J \times T = 1,250$ since $J = 25$ and $T = 50$.

The MPEC approach solves for θ by minimizing

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

subject to the "equilibrium constraints"

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

where S is the vector of observed market shares computed as described above given the values of x_t, p_t and ξ_t and $\boldsymbol{\eta}$ is a $(J \times T) \times 1$ vector defined by the following equation:

$$s_{jt}(\boldsymbol{\eta}, \boldsymbol{\theta}) = \frac{1}{N_s} \sum_{i=1}^{N_s} \frac{\exp(\theta_1 + x_{1j}\beta_{1i}^x + x_{2j}\beta_{2i}^x + x_{3j}\beta_{3i}^x + p_{jt}\beta_i^p + \eta_{jt})}{1 + \sum_{k=1}^{J} \exp(\theta_1 + x_{1k}\beta_{1i}^x + x_{2k}\beta_{2i}^x + x_{3k}\beta_{3i}^x + p_{kt}\beta_i^p + \eta_{kt})}$$

where each (β_i^x, β_i^p) is a random draw from the following normal distribution:

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

Note that θ_1 (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 the following initial point: true values for θ, and a vector of zeros for the η 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.

7.2 Our 2SLS Approach

Our 2SLS approach resorts to a slight modification of the standard linear 2SLS estimator to account for the fact that the estimates of the σ_k^2 and σ_p^2 cannot be negative. First, we construct the instrumental variables as the MPEC approach. We then construct the artificial regressors K_1, K_2, K_3, K_p of Theorem 2 for each product in each market by applying

$$\begin{aligned}\bar{X}_{it} &= \sum_{k=1}^{J} x_{ik} S_{kt} \\ K_i^{jt} &= x_{ij}(x_{ij}/2 - \bar{X}_{it})\end{aligned}$$

for $i = 1, 2, 3, p$.

The next step performs an instrumental variable regression of $y_{jt} = \log(\frac{S_{jt}}{S_{0t}})$ on 1, x_1, x_2, x_3, x_4, K_1, K_2, K_3, K_p using all 42 instruments. If any coefficient for the last four variables is negative, we set that coefficient to 0 and rerun the regression without that variable. We iterate this process until all the coefficients are positive, or all four variables are excluded from the instrumental variables regression.

In addition to this standard 2SLS estimator, we compute a corrected estimator as explained in section 6.3.3. To evaluate it, we replace y_{jt}, the dependent variables for 2SLS estimates, with $y_{jt} - (\xi_{2,jt} - \xi_{\infty,jt})$, where

- $\xi_{2,jt}$ is the residual from our initial 2SLS estimation procedure
- $\xi_{\infty,jt}$ is the value of ξ_{jt} that results from solving the equation $\boldsymbol{s}_t(\boldsymbol{\xi}_t, \hat{\boldsymbol{\theta}}) = \boldsymbol{S}_t$, where $\hat{\boldsymbol{\theta}}$ is the initial 2SLS estimate of $\boldsymbol{\theta}$.

We found that it worked as well as and often better than the bootstrap, at a considerably lower computational cost. We also experimented with using the optimal instruments, obtained by a kernel regression of $\boldsymbol{X}$ and of $\boldsymbol{K}$ on the variables $x_1, x_2, x_3, z_1, \ldots, z_6$.

7.3 Pseudo True Values for the 2SLS Approach

As explained earlier, the 2SLS estimator is not consistent for the true parameter values, as it estimates an approximate model. We constructed estimates of the pseudo true values to which our 2SLS estimators converge by simulating their probability limit. A first approach increases the number of markets and computes our 2SLS estimates for this large number of markets. The second approach computes estimates of the population values of the moments of our 2SLS estimator.

7.3.1 Increasing-number-of-markets Approach

For each simulation, we keep the size and distribution of product characteristics for each market fixed, but increase the number of markets. For each scenario, we calculate the pseudo true value (and its standard error) by 20 simulations of 100,000 markets. Note that across different simulations, we generate different product characteristics. Also, when calculating market shares, we use different random draws of β_i across different simulations, but the same random draws of β_i within a simulation. Estimates are calculated by the sample mean of the 20 simulations. Standard errors are calculated by the sample standard errors of the 20 simulations.

7.3.2 Moment-based Approach

We can also calculate the pseudo true values in a different way. We first run the first stage projection: $\hat{\Pi} = (W'W)^{-1}W'X$ for each simulation, where W is our matrix of instruments and X is our matrix of regressors. We then take the average across all the simulations to get our estimate of the population value of Π. Then in the second stage, we calculate $(W\Pi)'X$ and $(W\Pi)'Y$ for each simulation, and then take averages across all the simulations to get two matrices A and B. The final estimate is then $A^{-1}B$. In short, we have

$$\begin{gathered}
\Pi = E_{\text{all simulations}}[(W'W)^{-1}W'X] \\
A = E_{\text{all simulations}}[(W\Pi)'X] \\
B = E_{\text{all simulations}}[(W\Pi)'Y] \\
\text{Estimate} = A^{-1}B
\end{gathered}$$

With this method, we only have the estimates but cannot get the standard errors. We used 1000 simulations of 10,000 markets.

7.4 Monte Carlo Simulation Results

We used the SNOPT optimization package available from the Stanford Systems Optimization Laboratory to solve the nonlinear optimization problems for the MPEC estimator. The software employs a sparse sequential quadratic programming (SQP) algorithm with limited-memory quasi-Newton approximations to the Hessian of the Lagrangian.

We run simulations for 9 scenarios obtained by setting three values for the variance of the product random effects: $\sigma_\xi^2 = \mathrm{Var}(\xi) = 0.1, 0.5, 1$ and three values for the vector of variances of the coefficients $\beta_i = (\beta_0, \beta_{1i}^x, \beta_{2i}^x, \beta_{3i}^x, \beta_i^p)'$:

$$\mathrm{Var}(\beta_i) = (0, 0.1, 0.1, 0.1, 0.05), (0, 0.2, 0.2, 0.2, 0.1), (0, 0.5, 0.5, 0.5, 0.2).$$

Note that the square roots of the elements of $\mathrm{Var}(\beta_j)$ represent the relative values of the scale parameter σ of models 1, 2, and 5.

It is worth noting here that we explored other scenarii in which MPEC often failed to converge, even though we are starting it from the true values of the parameters. In particular, larger variances of ξ are problematic. It is also the reason why we reduced the highest value of σ_p^2 from 0.25 to 0.2.

All the other parameter specifications are as described above.

7.4.1 Distribution of the Estimates

We summarize the estimation results in Tables 1 to 9, where density plots are grouped by parameter for all scenarii. These plots suggest that if the researcher is interested in a precise estimate of the mean of the random coefficients, then using our 2SLS approach does not imply any significant bias or loss in efficiency relative to the MPEC approach.

The MPEC approach appears to dominate the 2SLS approach for the variance of the random coefficients. The 2SLS estimators of the variances have a downward bias that increases with the variance of the random coefficients. However, larger values of the variance of ξ_{jt} do seem to improve the performance of the 2SLS estimator of the variance of the random coefficients.

We also implemented the Petrin and Train (2010) control function approach. To do this, we first run a linear regression of the price p_{jt} on all 42 instruments. We denote the residuals from this regression by $\hat{\varepsilon}_{jt}$ and we include them as an additional covariate in the mean utility of product j. The middle panel of Tables 1 to 9 presents the distribution of estimated parameters for the case that $\text{Var}(\xi) = 0.5$. The middle panels of Tables 8 and 9 shows the results of this experiment for the mean and variance of the price coefficients in the central case $\text{Var}(\xi) = 0.5$. The control function approach exhibits substantial bias in the estimates of both means and variances of the random price coefficients[19].

7.4.2 Starting Values

We are giving a big advantage to MPEC in our comparisons, since we allow the algorithm to start from the true values of the parameters. This is of course infeasible in practice. With this initial boost, MPEC converges 100% of the time, after 1,030 iterations on average; the minimization takes 110 seconds on average. Our 2SLS approach provides a more realistic alternative, in which we start MPEC from the results of our 2SLS regression. This appears to work very well: MPEC converges after an average 125 seconds and 1,280 iterations, again with a 100% success rate. The resulting estimates are very close to those obtained when staring from the true values: the difference is between 10^{-6} and 10^{-7}.

These results are very encouraging for the use of our approach to find very good starting values for the MPEC and nested-fixed point estimation procedures. Given that 2SLS takes no time at all, we would strongly recommend running it before a more sophisticated algorithm.

7.4.3 Pseudo-true Values

Tables 10 and 11 demonstrate that for most scenarii and coefficients, the pseudo true values implied by our 2SLS procedure are not substantially different from the true values. Based on these results, it is difficult to argue that a researcher would draw conclusions from 2SLS estimates that differ in an economically or even statistically meaningful way from those obtained with MPEC estimates.

[19] Other experiments (not reported here) show that this bias increases with the variance of ξ_{jt}, as one would expect since the control function is inexact.

7.4.4 Price Elasticities

Based on the parameter estimates, we can estimate the own price elasticity of the demand for each product. The graphs in table 12 plot the distribution of the difference between the true price elasticity and the estimated price elasticity for the MPEC approach, our standard 2SLS approach, and our corrected 2SLS estimator. For space reasons, we only presents the results for five products: numbers 5, 10, 15, 20, 25. Our simulations have variances $(\sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_p^2) = (0, 0.4, 0.4, 0.4, 0.2)$ with $\text{Var}(\xi) = 1$. Table 12 demonstrates that our procedure recovers nearly identical mean own-price elasticities for products as the MPEC approach, although the spread for our estimates is slightly larger than in the MPEC approach.

We also performed a set of simulations (with the same variances) to determine if changing the true value of the price coefficient $\bar{\beta}_p$ changes the performance of the estimators. The results in tables 13 and 14 reinforce our previous conclusions about the 2SLS approach. For a range of values of the mean value of the price coefficient, our approach introduces minimal bias in the estimates of the means of the random coefficients. The estimates of the variances of the random coefficients for our 2SLS estimate continue to be downwards biased in general, but the bias is smaller for larger price coefficients.

7.4.5 Variable Selection Tests

Researchers in empirical IO have little guidance on the list of characteristics $\boldsymbol{X}$ they should include, or how to specify the matrix $\boldsymbol{\Sigma}$. Experimenting with different specifications is costly with the usual estimators. Our 2SLS approach, on the other hand, makes variable selection very easy. We can decide whether a characteristic simply by testing whether the corresponding covariate can be dropped from the estimating equation; and to decide whether we should allow for a random coefficient, we only need to test whether the associated artificial regressor can be dropped from the equation. We experimented with this approach to detecting random coefficients by setting $\bar{\beta}_1^x = \sigma_1^x = 0$ in the data generating process and applying standard tests that that the covariate x_1 and/or the artificial regressor K_1 has a zero coefficient in the 2SLS regression. We also performed this test using our corrected 2SLS estimates. Tables 15 to 20 give the probability that the null hypothesis is not rejected, where the null hypothesis is

- $\bar{\beta}_1^x = 0$ (Tables 15 and 16)
- $\sigma_1^x = 0$ (Tables 17 and 18)
- $\bar{\beta}_1^x = \sigma_1^x = 0$ (Tables 19 and 20).

The row labelled "2SLS with heteroskedasticity-robust standard error" is our 2SLS estimate, using a standard heteroskedasticity-robust covariance matrix to compute standard errors. The row labelled "GLS estimator and standard errors" uses Cragg's (1983) generalized least squares estimator and his recommended standard error estimates. The row labelled "2SLS with clustered standard errors" uses our 2SLS estimates with standard errors clustered at market level.

Since the null hypothesis is true, each row in Tables 15-20 would ideally contain $0.99, 0.95$, and 0.90. Clearly, our test rejects the null too often. In this particular application, this is probably better than the alternative: better to include more variables and lose some efficiency than to incur bias by leaving them out. The size distortion is smaller for tests on the means (Tables 15 and 16); it is also smaller when we use corrected estimates. The clustered standard error estimates appear to have the largest size distortions. On the whole, we take this to suggest that demonstrate that our estimator can be used to good effect in order to decide which coefficients should be modelled as random.

7.5 Lognormal Distribution for β

As explained in section 5.2, our estimating equation is the same whether the distribution of the random coefficients is normal or not. To illustrate this, we modify the data-generating process so that the consumer preference parameters β_i have a lognormal distribution:

$$\begin{aligned} \beta_i &= \bar{\beta}_i \epsilon_i \\ \bar{\beta}_i &= (1, 1.5, 1.5, 0.5, 1) \\ \ln(\epsilon_i) &\sim N(-0.5\sigma^2, \sigma^2). \end{aligned}$$

We study several cases, with $\sigma = 0.3, 0.4, 0.5$ and $\xi_{jt} \sim N(0, 0.1)$. The rest of the specification is as before. Lognormality induces significant skewness and kurtosis into the distribution of the random coefficients. The standard 2SLS approach gives us estimates of the first and second moments. We can also introduce the additional artificial regressors T of section 6.2, either to control for skewness or to estimate it. We experimented with both possibilities. Each plot in Tables 21, 22, and 23 shows the distributions when we use only X and K ("only include 2nd moment") and when we add T ("include third moment"). These three tables report the distributions of the estimates of the first, second, and third moments of $\beta_1, \beta_2, \beta_3$, and β_p. Table 24 provides the corresponding summary statistics.

For a variety of values for the parameter σ of our lognormal distribution, the 2SLS estimates are just as good as they were in the normal setup. The additional information in the third moment of the random coefficients does not appreciably increase the precision in our estimates of the means and variances of the random coefficients. In fact, for some of the coefficients, including the third-order artificial regressors T leads to significantly less efficient estimates. This is likely due to the fact that our procedure has a difficult time estimating the third moment of the random coefficients, as Table 23 shows.

7.5.1 Correction and Kernel

Our paper suggested two potential improvements to the standard 2SLS regression: our two-step asymptotic correction described in section 6.3.3, and using a kernel regression to estmate the optimal instruments $E(\boldsymbol{X}|\boldsymbol{Z})$ and $E(\boldsymbol{K}|\boldsymbol{Z})$. We compare both methods, when coefficients are normal with variances $(0, 0.5, 0.5, 0.5, 0.2)$ and when they are lognormally distributed with $\sigma = 0.4$ for $\ln(\epsilon_i) \sim N(-0.5\sigma^2, \sigma^2)$. In both cases we took $Var(\xi) = 0.1$.

Tables 25 and 26 plot the distributions of the estimators. They suggest that our correction helps reduce the bias, both for the means and the variances. This holds whether the random coefficients are normally or log-normally distributed. Using kernel regressions to approximate the optimal instruments appears to slightly reduce both the bias and the variance of some of the estimates.

Concluding Comments

Our FRAC estimation procedure applies directly to the random coefficients demand models commonly used in empirical IO. 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 is "robust" to variations on the distribution of the random coefficients. In addition, it provides straightforward tests that help in variable selection, especially as a guide to determine which coefficients in the demand system should be modeled as random.

Some of our simulation results point to directions for future research. We hope to report more general analytical results that illuminate these findings.

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.

Campioni, G. (2018), "Nonparametric Demand Estimation in Differentiated Products Markets", mimeo Yale.

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.

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

Crawford, G., and A. Yurukoglu (2012), "The Welfare Effects of Bundling in Multichannel Television Markets," *American Economic Review*, 102, 643-685.

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 (2016), "Measuring Substitution Patterns in Differentiated Products Industries", mimeo.

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.

Ho, K. and R. Lee (2017), "Insurer Competition in Health Care Markets", *Econometrica*, 85, 379-417.

Horowitz, J. and L. Nesheim (2018), "Using Penalized Likelihood to Select Parameters in a Random Coefficients Multinomial Logit Model", mimeo UCL.

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

Ketz, P. (2018), "On Asymptotic Size Distortions in the Random Coefficients Logit Model", mimeo Paris School of Economics.

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.

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 Proof of Theorem 1

Proof. We start from (2); we drop the market index t and the bold letters. Since we now denote $X\epsilon = \sigma x \cdot v$, we can rewrite (2) in the standard model as

$$S_j = E_v \frac{\exp(X_j\beta + \sigma x_j \cdot v + \xi_j)}{1 + \sum_{k=1}^J S_k \exp(X_k\beta + \sigma x_k \cdot v + \xi_k)}.$$

Given that

$$\xi_j = \log \frac{S_j}{S_0} - \boldsymbol{X}_j\boldsymbol{\beta} + a_{1j}\sigma + a_{2j}\frac{\sigma^2}{2} + O(\sigma^3),$$

we get

$$S_j = E_v \frac{S_j \exp\left(\sigma\left(x_j \cdot v + a_{1j}\right) + a_{2j}\frac{\sigma^2}{2} + O(\sigma^3)\right)}{S_0 + \sum_{k=1}^J S_k \exp\left(\sigma\left(x_k \cdot v + a_{1k}\right) + a_{2k}\frac{\sigma^2}{2} + O(\sigma^3)\right)}$$

Eliminating S_j gives

$$1 = E_{\boldsymbol{v}} \frac{\exp\left(\sigma\left(\boldsymbol{x}_j'\boldsymbol{v} + a_{1j}\right) + \frac{\sigma^2}{2}a_{2j} + O(\sigma^3)\right)}{S_0 + \sum_{k=1}^J S_k \exp\left(\sigma\left(\boldsymbol{x}_k'\boldsymbol{v} + a_{1k}\right) + \frac{\sigma^2}{2}a_{2k} + O(\sigma^3)\right)}. \tag{13}$$

In this form, Theorem 1.(i) is obvious since only the vectors $\boldsymbol{x}_k$ and market shares S_k enter the system of equations.

Now use the notation $e_S Z \equiv \sum_{k=1}^J S_k Z_k$ and $\hat{Z}_j = Z_j - e_S Z$ to rewrite (13) as

$$0 = E_v \left(\frac{\hat{V}_j}{1 + e_S V}\right) \tag{14}$$

where $V_j \equiv f_j + \alpha_j + f_j\alpha_j$, with

$$f_j \equiv \exp(\sigma x_j \cdot v) - 1 = \sigma x_j \cdot v + \frac{\sigma^2}{2}(x_j \cdot v)^2 + O_P(\sigma^3)$$

and

$$\alpha_j = \exp(a_{1j}\sigma + a_{2j}\sigma^2/2 + O(\sigma^3)) - 1 = a_{1j}\sigma + (a_{2j} + a_{1j}^2)\frac{\sigma^2}{2} + O(\sigma^3).$$

We note that f_j is $O_P(\sigma)$ and α_j is $O(\sigma)$, so that V_j is also $O_P(\sigma)$.

Now expanding (14) gives

$$\begin{aligned} O(\sigma^3) &= E_v\hat{V}_j - E_v\left(\hat{V}_j(e_S V)\right) && (15)\\ &= E_v\hat{f}_j + \hat{\alpha}_j && (16)\\ &+ \widehat{\alpha_j E_v f_j} - (e_S\alpha)E_v\hat{f}_j - \hat{\alpha}_j E_v(e_S f) - \hat{\alpha}_j(e_S\alpha) - E_v\hat{f}_j(e_S f). && (17) \end{aligned}$$

Only the terms on line (16) can be of order 1 in σ. But using $E_v v = 0$ and $E_v(x_j \cdot v)(x_k \cdot v) = x_j \cdot x_k$ gives us $E_v f_j = \frac{\sigma^2}{2}\|x_j\|^2 + O(\sigma^3)$. Therefore the only term of order 1 is in $\hat{\alpha}_j = \widehat{a}_{1j}\sigma + O(\sigma^2)$, and we must have $\widehat{a}_{1j} = 0$. We note that the "hat" operator is linear and invertible:

Lemma 1. *If $\hat{Z}_j = \hat{W}_j$ for all j and $S_0 < 1$, then $Z_j = W_j$.*

Proof. $Z_j - e_S Z = W_j - e_S W$ implies $Z_j = W_j + \lambda$, where $\lambda = e_S Z - e_S W$. But then $e_S Z = e_S W + e_S \lambda = (1 - S_0)\lambda$, so that $\lambda = (1 - S_0)\lambda = 0$. □

Applying the lemma gives $a_{1j} = 0$. As a consequence, $\alpha_j = a_{2j}\sigma^2/2 + O(\sigma^3)$; and all terms on line (17) except the last one are of order at least 3 in σ. Since

$$E_v \hat{f}_j + \hat{\alpha}_j = \frac{\sigma^2}{2}\widehat{\|x_j\|^2} + \frac{\sigma^2}{2}\widehat{a}_{2j} + O(\sigma^3)$$

and

$$E_v \hat{f}_j(e_S f) = \sigma^2 E_v(\hat{x}_j \cdot v)((e_S x) \cdot v) + O(\sigma^3) = \sigma^2(\hat{x}_j \cdot (e_S x)) + O(\sigma^3)$$

applying the lemma again gives us $(\|x_j\|^2 + a_{2j})/2 - x_j \cdot (e_S x) = 0$.

Finally, if the distribution of $\boldsymbol{v}$ is symmetric around $\mathbf{0}$ changing σ to $-\sigma$ in (2) must leave all market shares unchanged; therefore all expansions can only contain even-degree terms in σ. □

B Detailed Examination of the Mixed Logit

The standard binary model is simply a mixed logit. Applying Theorem 1 with $J = 1$ and using $S_0 + S_1 = 1$, we obtain

$$a_{21} = (2S_1 - 1)\|\boldsymbol{x}_1\|^2$$

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

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

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

The presence of the term $(1/2 - S_1)$ in this formula is a consequence of the symmetry of the distribution of $\boldsymbol{v}$ around $\mathbf{0}$ and of the logistic distribution around 0. Taken together, this implies that market shares around one half vary very little with σ. The random variation in tastes can only be identified from nonlinearities in the market shares; but since the cdf of the logistic has an inflexion point when its value is one half, market shares are essentially linear around that point. It is easy to check that this is specific to the one-product case; when $J > 1$, the mixed multinomial logit does not face any such difficulty.

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

$$E\left(\log \frac{S_1}{S_0} - \boldsymbol{X}_1 \boldsymbol{\beta} - \left(\frac{1}{2} - S_1\right) \sum_{m=1}^{n_X} \Sigma_{mm} X_{1m}^2 \;\middle|\; \boldsymbol{Z}\right) = 0. \tag{18}$$

B.1 Identification

The form of the estimating equation holds interesting insights about identification. First note that the optimal instruments are

$$\boldsymbol{f}(\boldsymbol{Z}) = E\left(\boldsymbol{X}_1 | \boldsymbol{Z}\right), E\left(\left(\frac{1}{2} - S_1\right) \boldsymbol{X}_1^2 | \boldsymbol{Z}\right)$$

where $\boldsymbol{X}_1^2$ is the vector with components X_{1m}^2. The asymptotic variance-covariance matrix of our estimator $\hat{\boldsymbol{\theta}}$ is given by the usual formula:

$$T\, V_{\text{as}} \hat{\boldsymbol{\theta}} \simeq \boldsymbol{J}^{-1} V(\xi_1 \boldsymbol{f}(\boldsymbol{Z})) \boldsymbol{J}^{-1},$$

where

$$\boldsymbol{J} = E\left(\left(\boldsymbol{X}_1, \left(\frac{1}{2} - S_1\right) \boldsymbol{X}_1^2\right) \boldsymbol{f}(\boldsymbol{Z})\right).$$

The identifying power of the (approximate) model relies on the full-rank of the matrix $\boldsymbol{J}$. Suppose for instance that after projecting (via nonparametric regression) the regressors on the instruments, the residual variation in the artificial regressor $(1/2 - S_1) X_{1m}^2$ is very well explained in a linear regression on the other covariates. 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. Of course, this can be diagnosed immediately by looking at the precision of the 2SLS estimates.

B.2 Higher-order terms

It is easy to program a symbolic algebra system to compute higher-order terms a_{lj} for $l > 2$. We show here how to compute the fourth-order term in the mixed logit model by hand. This will also illustrate the "robustness" of our method to distributional assumptions.

Assume that $\boldsymbol{\epsilon}$ 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

$$\xi_1 = \log \frac{S_1}{S_0} - \boldsymbol{X}_1\boldsymbol{\beta} + \frac{a_{21}}{2}\sigma^2 + \frac{a_{41}}{24}\sigma^4 + O(\sigma^6).$$

Define $L(t) = 1/(1+\exp(-t))$ the cdf of the logistic distribution. Note that $L' = L(1-L)$, and that higher-order derivatives follow easily:

$$\begin{aligned} L'' &= L(1-L)(1-2L) \\ L^{(3)} &= L(1-L)(1-6L+6L^2) \\ L^{(4)} &= L(1-L)(1-2L)(1-12L+12L^2). \end{aligned}$$

Since the market share of good 1 is

$$S_1 = E_{\boldsymbol{\epsilon}} L\left(\boldsymbol{X}_1(\boldsymbol{\beta}+\boldsymbol{\epsilon}) + \xi_1\right)$$

we obtain, much as in Appendix A,

$$S_1 = E_{\boldsymbol{\epsilon}} L\left(\log \frac{S_1}{S_0} + \boldsymbol{X}_1\boldsymbol{\epsilon} + \alpha_2\sigma^2 + \alpha_4\sigma^4 + O(\sigma^6)\right)$$

where we defined $\alpha_l = a_{l1}/l!$ for $l = 2, 4$.

Let a 0 subscript indicate that we take the value and derivatives of $L(t)$ at $t = \log(S_1/S_0)$. Defining $u(\boldsymbol{\epsilon}) = \boldsymbol{X}_1\boldsymbol{\epsilon} + \alpha_2\sigma^2 + \alpha_4\sigma^4 + O(\sigma^6)$ and expanding gives

$$L\left(\log \frac{S_1}{S_0} + u\right) = L_0 + L_0'u + \frac{L_0''}{2}u^2 + \frac{L_0^{(3)}}{6}u^3 + \frac{L_0^{(4)}}{24}u^4 + O(u^5).$$

Incorporating $L_0 = S_1$, $L_0' = S_1(1-S_1)$, up to $L_0^{(4)}$ gives

$$S_1 = E_{\boldsymbol{\epsilon}} \left(S_1 + S_1(1-S_1)u(\boldsymbol{\epsilon}) + S_1(1-S_1)(1-2S_1)\frac{u(\boldsymbol{\epsilon})^2}{2} \right.$$
$$\left. + S_1(1-S_1)(1-6S_1+6S_1^2)\frac{u(\boldsymbol{\epsilon})^3}{6} + S_1(1-S_1)(1-2S_1)(1-12S_1+12S_1^2)\frac{u(\boldsymbol{\epsilon})^4}{24} + O(u(\boldsymbol{\epsilon})^5) \right);$$

dividing by $S_1(1-S_1)$ yields

$$E_{\boldsymbol{\epsilon}}u + (1-2S_1)E_{\boldsymbol{\epsilon}}u^2/2 + (1-6S_1+6S_1^2)E_{\boldsymbol{\epsilon}}u^3/6 + (1-2S_1)(1-12S_1+12S_1^2)E_{\boldsymbol{\epsilon}}u^4/24 = E_{\boldsymbol{\epsilon}}O(u^5). \tag{19}$$

Finally, up to order 6 in σ:

$$\begin{aligned}
E_{\boldsymbol{\epsilon}}u &= \alpha_2\sigma^2 + \alpha_4\sigma^4 \\
E_{\boldsymbol{\epsilon}}u^2 &= \sigma^2 E_{\boldsymbol{\epsilon}} \left(\boldsymbol{X}_1\boldsymbol{\epsilon}\right)^2 + \alpha_2^2\sigma^4 = \sum_{m=1}^{n_X} \Sigma_{mm}x_{1m}^2 + \alpha_2^2\sigma^4 \\
E_{\boldsymbol{\epsilon}}u^3 &= 3\alpha_2\sigma^4 E_{\boldsymbol{\epsilon}} \left(\boldsymbol{X}_1\boldsymbol{\epsilon}\right)^2 = 3\alpha_2 \sum_{m=1}^{n_X} \Sigma_{mm}x_{1m}^2 \\
E_{\boldsymbol{\epsilon}}u^4 &= \sigma^4 E_{\boldsymbol{\epsilon}} \left(\boldsymbol{X}_1\boldsymbol{\epsilon}\right)^4 = \sum_{m=1}^{n_X} k_m x_{1m}^4.
\end{aligned}$$

Regrouping terms in σ^2 in (19) confirms that

$$\alpha_2\sigma^2 = (S_1 - 1/2)\sum_{m=1}^{n_X} \Sigma_{mm}x_{1m}^2,$$

which we knew from Theorem 1. The terms in σ^4 give us

$$\begin{aligned}
\alpha_4\sigma^4 = \alpha_2^2\sigma^4(S_1-1/2) - \alpha_2\sigma^2(1-6S_1+6S_1^2)\sum_{m=1}^{n_X}\Sigma_{mm}x_{1m}^2/2 \\
- (1-2S_1)(1-12S_1+12S_1^2)\sum_{m=1}^{n_X} k_m x_{1m}^4/24.
\end{aligned}$$

This simplifies to

$$\alpha_4\sigma^4 = \left(\frac{1}{2} - S_1\right) \times$$
$$\left(\left(\frac{1}{4} - 2S_1(1-S_1)\right)\left(\sum_{m=1}^{n_X}\Sigma_{mm}x_{1m}^2\right)^2 - \left(\frac{1}{12} - S_1(1-S_1)\right)\sum_{m=1}^{n_X} k_m x_{1m}^4 \right).$$

This formula may not seem especially enlightening, but it shows several important points. First, terms of higher orders can be computed without much difficulty. Second, each additional term adds information on lower-order moments (here σ_m^2), as well as on the moments of higher order (here k_m). 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, here represented by

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

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 is "robust" to any (well-behaved) distribution, the estimator based on this fourth-order expansion also assumes symmetry: a skewed distribution would generate terms in σ^3. 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 $\boldsymbol{\epsilon}$ has normal kurtosis. Then $k_1 = 3\Sigma_{11}^2$ and we find the simpler formula

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

Specializing further, Figure 1 in the main text plots the terms $t_k(S_1)$ for $k = 1, 2, 3, 4$ as the market share goes from zero to one when $\boldsymbol{\epsilon}$ is Gaussian.

If the components of $\boldsymbol{\epsilon}$ are independently distributed and have third moments $(s_1, \ldots, s_{n_X})$, then it is easy to see that an additional term

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

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.

B.3 Beyond Logit 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 (18) 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:

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

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

C The Two-level Nested Logit

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

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

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

$$\frac{1 + a\sigma + b\sigma^2}{1 + c\sigma + d\sigma^2} = 1 + (a - c)\sigma + (b - d - c(a - c))\sigma^2 + O(\sigma^3). \tag{20}$$

Denote $\hat{A}_{j|k} = A_j - \bar{A}_k$. Applying (20) 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

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

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 (20) 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

$$\begin{aligned}
M &= \sum_{l=1}^{K} S_{N_l} \frac{\lambda_l - 1}{\lambda_l} \|\bar{\boldsymbol{x}}_l\|^2 + \sum_{l=1}^{K} S_{N_l} \frac{\overline{\|\boldsymbol{x}\|^2}_l}{\lambda_l} - 2\|e_{\boldsymbol{S}} \boldsymbol{x}\|^2 \\
\nu_k &= \overline{\|\boldsymbol{x}\|^2}_k - 2\|\bar{\boldsymbol{x}}_k\|^2 - \lambda_k(\lambda_k - 1)\|\bar{\boldsymbol{x}}_k\|^2 - \lambda_k \overline{\|\boldsymbol{x}\|^2}_k + 2\lambda_k^2 e_{\boldsymbol{S}} \boldsymbol{x} \cdot \bar{\boldsymbol{x}}_k + 2\lambda_k \|\bar{\boldsymbol{x}}_k\|^2 - 2\lambda_k \bar{\boldsymbol{x}}_k \cdot e_{\boldsymbol{S}} \boldsymbol{x} \\
&= (1 - \lambda_k) \left(\overline{\|\boldsymbol{x}\|^2}_k - (2 - \lambda_k)\|\bar{\boldsymbol{x}}_k\|^2 - 2\lambda_k \bar{\boldsymbol{x}}_k \cdot e_{\boldsymbol{S}} \boldsymbol{x} \right) && (21) \\
\mu_j &= -\|\boldsymbol{x}_j\|^2 + 2\boldsymbol{x}_j \cdot \bar{\boldsymbol{x}}_k - 2\lambda_k \boldsymbol{x}_j \cdot (\bar{\boldsymbol{x}}_k - e_{\boldsymbol{S}} \boldsymbol{x}) \\
&= \boldsymbol{x}_j \cdot (2\lambda_k e_{\boldsymbol{S}} \boldsymbol{x} - \boldsymbol{x}_j + 2(1 - \lambda_k)\bar{\boldsymbol{x}}_k)\,. && (22)
\end{aligned}$$

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 (21) and (22) 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

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

D Our Correction Formula

Remember from section 6.3.2 that

$$
\boldsymbol{\theta}_0 \simeq \boldsymbol{\theta}_2 - \left(E\frac{\partial f_\infty}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta}_2; \boldsymbol{\lambda}_0)\right)^{-1} f_\infty(\boldsymbol{\theta}_2; \boldsymbol{\lambda}_0). \tag{23}
$$

The term in the inverse is easily proxied:

$$E\frac{\partial f_\infty}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta}_2;\boldsymbol{\lambda}_0) \simeq E\frac{\partial f_2}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta}_2) = E\left(\frac{\partial \xi_2}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta}_2)\right)' \boldsymbol{V}\boldsymbol{V}'\frac{\partial \xi_2}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta}_2),$$

since $\boldsymbol{\xi}_2$ is linear in $\boldsymbol{\theta}$. Note that this is $E\boldsymbol{\mathcal{X}}'\boldsymbol{\mathcal{X}}$, where

$$\boldsymbol{\mathcal{X}} \equiv \boldsymbol{V}'\frac{\partial \boldsymbol{\xi}_2}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta}_2) = -\boldsymbol{V}'(\boldsymbol{X},\boldsymbol{K})$$

and row $j = 1,\ldots,J$ of $(\boldsymbol{X},\boldsymbol{K})$ lists the covariates and artificial regressors for this product. It follows that

$$E\frac{\partial f_\infty}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta}_2;\boldsymbol{\lambda}_0) \simeq \begin{pmatrix} E\left(\boldsymbol{X}'\boldsymbol{V}\boldsymbol{V}'\boldsymbol{X}\right) & E\left(\boldsymbol{X}'\boldsymbol{V}\boldsymbol{V}'\boldsymbol{K}\right) \\ E\left(\boldsymbol{K}'\boldsymbol{V}\boldsymbol{V}'\boldsymbol{X}\right) & E\left(\boldsymbol{K}'\boldsymbol{V}\boldsymbol{V}'\boldsymbol{K}\right) \end{pmatrix}.$$

To the second-order in $\boldsymbol{e}_2$, $Ef_\infty(\boldsymbol{\theta}_2;\boldsymbol{\lambda}_0)$ equals

$$E\left(\frac{\partial \boldsymbol{\xi}_2}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta}_2)'\boldsymbol{V}\boldsymbol{V}'\hat{\boldsymbol{e}}_2(\boldsymbol{\theta}_2;\boldsymbol{\lambda}_0)\right) + E\left(\frac{\partial \hat{\boldsymbol{e}}_2}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta}_2;\boldsymbol{\lambda}_0)'\boldsymbol{V}\boldsymbol{V}'\boldsymbol{\xi}_2(\boldsymbol{\theta}_2)\right). \tag{24}$$

The first term in (24) is simply $E\left(\boldsymbol{\mathcal{X}}'\boldsymbol{V}'\hat{\boldsymbol{e}}_2\right)$. Going back to (23), we get

$$\boldsymbol{\theta}_0 \simeq \boldsymbol{\theta}_2 - \begin{pmatrix} E\left(\boldsymbol{X}'\boldsymbol{V}\boldsymbol{V}'\boldsymbol{X}\right) & E\left(\boldsymbol{X}'\boldsymbol{V}\boldsymbol{V}'\boldsymbol{K}\right) \\ E\left(\boldsymbol{K}'\boldsymbol{V}\boldsymbol{V}'\boldsymbol{X}\right) & E\left(\boldsymbol{K}'\boldsymbol{V}\boldsymbol{V}'\boldsymbol{K}\right) \end{pmatrix}^{-1} \left(E\left(\boldsymbol{\mathcal{X}}'\boldsymbol{V}'\hat{\boldsymbol{e}}_2\right) + E\left(\frac{\partial \hat{\boldsymbol{e}}_2}{\partial \boldsymbol{\theta}}'\boldsymbol{V}\boldsymbol{V}'\boldsymbol{\xi}_2\right)\right).$$

Finally, using Theorem 1(i), we know that $\frac{\partial \hat{\boldsymbol{e}}_2}{\partial \boldsymbol{\beta}} = \boldsymbol{0}$. Therefore

$$E\left(\boldsymbol{\mathcal{X}}'\boldsymbol{V}'\hat{\boldsymbol{e}}_2\right) + E\left(\frac{\partial \hat{\boldsymbol{e}}_2}{\partial \boldsymbol{\theta}}'\boldsymbol{V}\boldsymbol{V}'\boldsymbol{\xi}_2\right) = \begin{pmatrix} -E\left(\boldsymbol{X}'\boldsymbol{V}\boldsymbol{V}'\hat{\boldsymbol{e}}_2\right) \\ -E\left(\boldsymbol{K}'\boldsymbol{V}\boldsymbol{V}'\hat{\boldsymbol{e}}_2\right) + E\left(\frac{\partial \hat{\boldsymbol{e}}_2}{\partial \boldsymbol{\Sigma}}'\boldsymbol{V}\boldsymbol{V}'\boldsymbol{\xi}_2\right) \end{pmatrix}. \quad \square$$

E Proof of Theorem 3

We drop the bold letters in this proof to alleviate the notation, and without loss of generality we normalize $B = 1$.

Remember that $G(y, F(y,\beta,\sigma),\beta,\sigma) = 0$, so that $G(y, F(y,\beta,0),\beta,0) = 0$. Given (11), this gives $G^*(y, A^*(y, F(y,\beta,0) - f_1(y)\beta, 0) = 0$ for all β. This can only hold if $F(y,\beta,0) - f_1(y)\beta$ does not depend on β, which implies condition **C2**. Denoting $f_0(y) = F(y,\beta,0) - f_1(y)\beta$, we obtain

$$G^*(y, A^*(y, f_0(y), 0)) = 0.$$

Now writing $G^*(y, E_v A^*(y, F(y, \beta, \sigma) - f_1(y)\beta, \sigma v)) = 0$ as an identity in σ and taking derivatives with respect to σ, we get

$$\begin{aligned} G_2^* E_v \left(A_2^* F_\sigma + A_3^* v\right) &= 0 \\ G_{22}^*[E_v \left(A_2^* F_\sigma + A_3^* v\right)][E_v \left(A_2^* F_\sigma + A_3^* v\right)] & \\ + G_2^* E_v \left(A_2^* F_{\sigma\sigma} + A_{22}^*[F_\sigma, F_\sigma] + 2A_{23}^*[F_\sigma, v] + A_{33}^*[v, v]\right) &= 0. \end{aligned}$$

Fortunately, this simplifies greatly at $\sigma = 0$. The first equation gives

$$G_2^* E_v \left(A_2^* F_\sigma(y, \beta, 0) + A_3^* v\right) = 0,$$

where the derivatives A_2^* and A_3^* do not depend on v since $\sigma = 0$. It follows that $G_2^* A_2^* F_\sigma(y, \beta_0, 0) = 0$ since $Ev = 0$. Given our invertibility assumption, condition **C1** also holds. Using the second equation at $\sigma = 0$, and given that $F_\sigma(y, \beta_0, 0) = 0$, we get

$$G_2^* E_v \left(A_{22}^*[F_\sigma, F_\sigma] + 2A_{23}^*[F_\sigma, v]\right) = 0$$

so that

$$G_2^* \left(A_2^* F_{\sigma\sigma} + A_{33}^*\right) = 0.$$

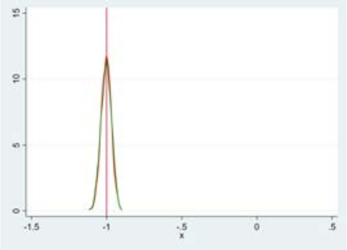
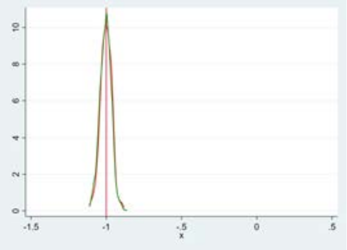
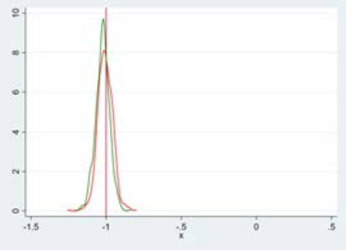
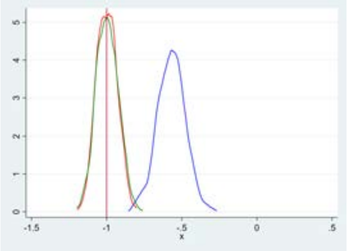
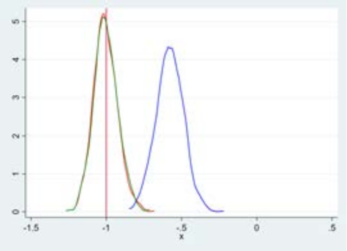
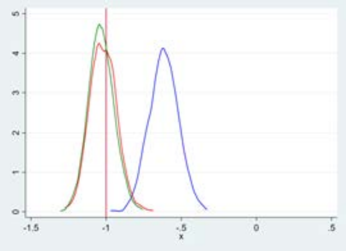
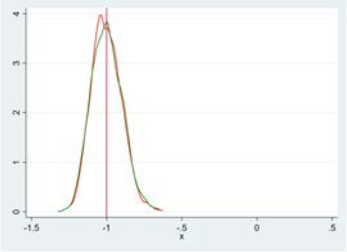
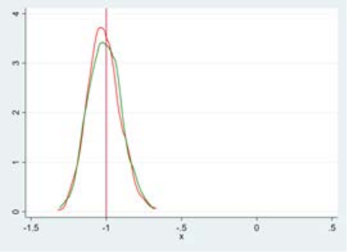
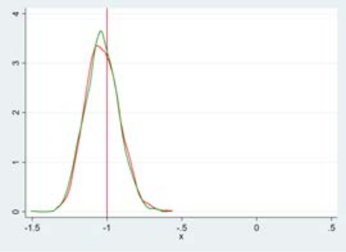
Given that G_2^* is invertible, this gives (reintroducing the arguments)

$$A_2^*(y, f_0(y), 0) F_{\sigma\sigma}(y, \beta, 0) + A_{33}^*(y, f_0(y), 0) = 0.$$

Therefore $F_{\sigma\sigma}(y, \beta, 0)$ is independent of β and condition **C3** holds. Noting that $f_2(y) = -F_{\sigma\sigma}(y, \beta, 0)$ completes the proof.

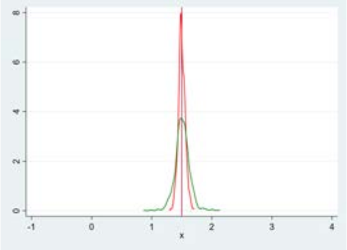
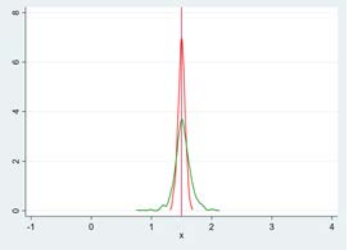
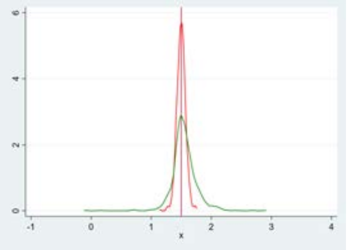
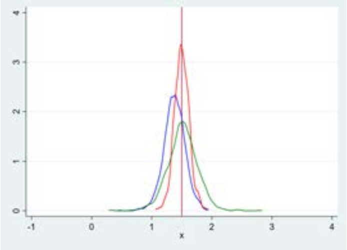
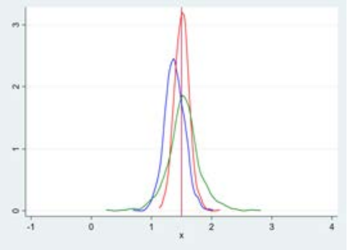
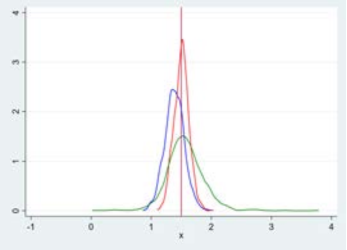
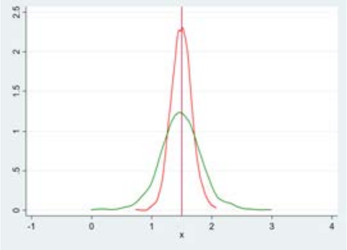
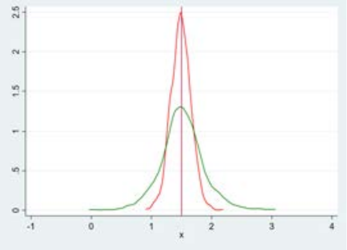
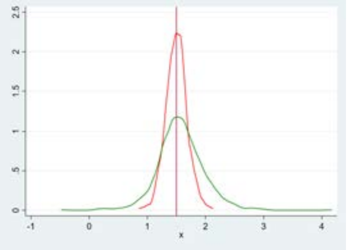
List of Tables

Table 1: Distribution of the Estimates of β_0

Var(ξ)=0.1	Var(ξ)=0.1	Var(ξ)=0.1
Var(β)=(0,0.1,0.1,0.1,0.05)	Var(β)=(0,0.2,0.2,0.2,0.1)	Var(β)=(0,0.5,0.5,0.5,0.2)
		
Var(ξ)=0.5	Var(ξ)=0.5	Var(ξ)=0.5
Var(β)=(0,0.1,0.1,0.1,0.05)	Var(β)=(0,0.2,0.2,0.2,0.1)	Var(β)=(0,0.5,0.5,0.5,0.2)
		
Var(ξ)=1.0	Var(ξ)=1.0	Var(ξ)=1.0
Var(β)=(0,0.1,0.1,0.1,0.05)	Var(β)=(0,0.2,0.2,0.2,0.1)	Var(β)=(0,0.5,0.5,0.5,0.2)
		

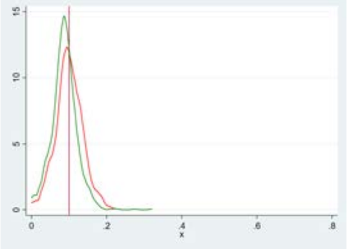
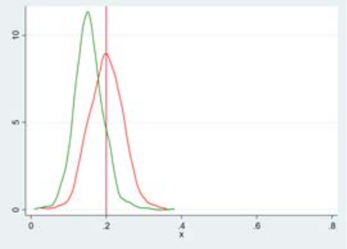
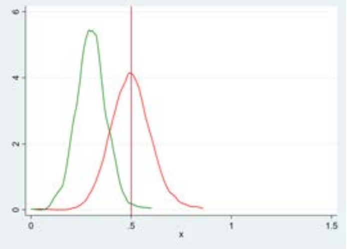
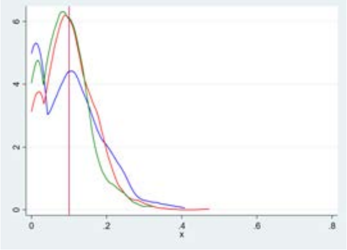
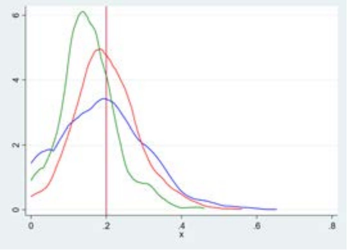
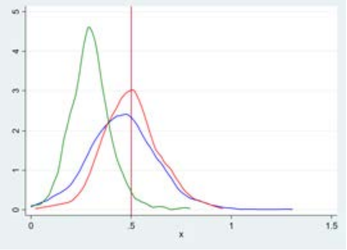
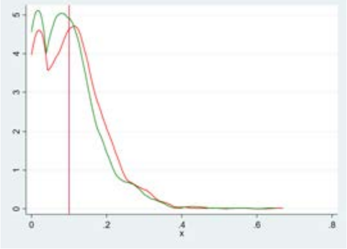
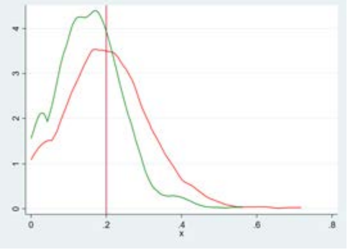
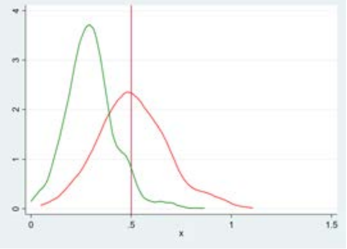
Control Function	—
MPEC	—
2SLS	—

Table 2: Distribution of the Estimates of $\bar{\beta}_1^x$

Var(ξ)=0.1	Var(ξ)=0.1	Var(ξ)=0.1
Var(β)=(0,0.1,0.1,0.1,0.05)	Var(β)=(0,0.2,0.2,0.2,0.1)	Var(β)=(0,0.5,0.5,0.5,0.2)
		
Var(ξ)=0.5	Var(ξ)=0.5	Var(ξ)=0.5
Var(β)=(0,0.1,0.1,0.1,0.05)	Var(β)=(0,0.2,0.2,0.2,0.1)	Var(β)=(0,0.5,0.5,0.5,0.2)
		
Var(ξ)=1.0	Var(ξ)=1.0	Var(ξ)=1.0
Var(β)=(0,0.1,0.1,0.1,0.05)	Var(β)=(0,0.2,0.2,0.2,0.1)	Var(β)=(0,0.5,0.5,0.5,0.2)
		

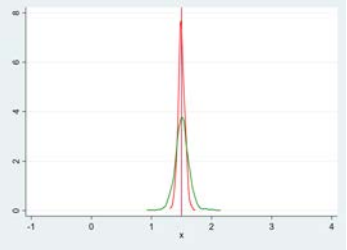
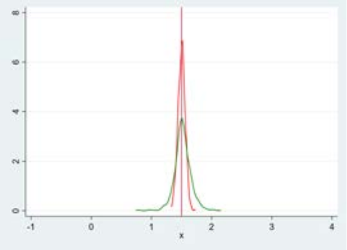
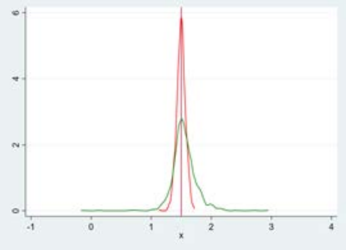
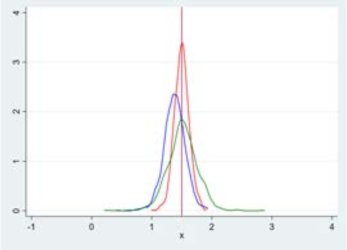
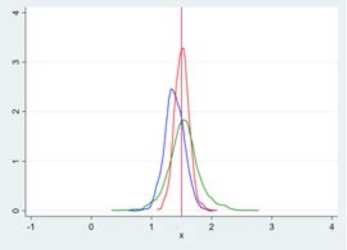
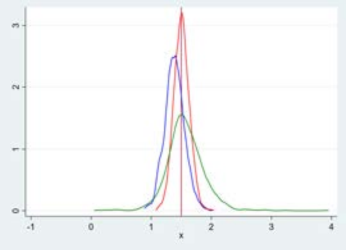
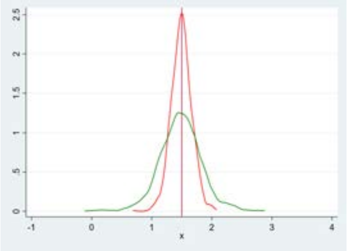
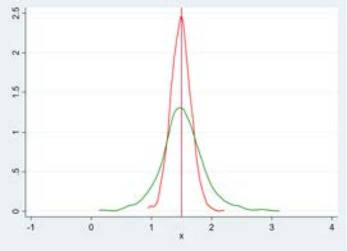
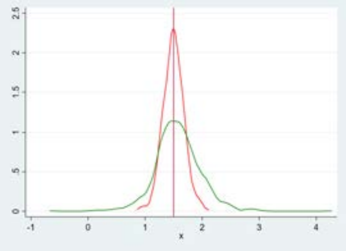
Control Function	—
MPEC	—
2SLS	—

Table 3: Distribution of the Estimates of Var(β_1^x)

Var(ξ)=0.1	Var(ξ)=0.1	Var(ξ)=0.1
Var(β)=(0,0.1,0.1,0.1,0.05)	Var(β)=(0,0.2,0.2,0.2,0.1)	Var(β)=(0,0.5,0.5,0.5,0.2)
		
Var(ξ)=0.5	Var(ξ)=0.5	Var(ξ)=0.5
Var(β)=(0,0.1,0.1,0.1,0.05)	Var(β)=(0,0.2,0.2,0.2,0.1)	Var(β)=(0,0.5,0.5,0.5,0.2)
		
Var(ξ)=1.0	Var(ξ)=1.0	Var(ξ)=1.0
Var(β)=(0,0.1,0.1,0.1,0.05)	Var(β)=(0,0.2,0.2,0.2,0.1)	Var(β)=(0,0.5,0.5,0.5,0.2)
		

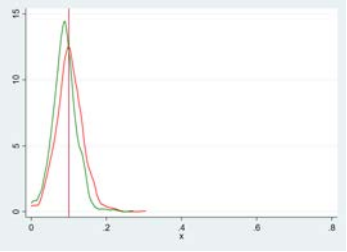
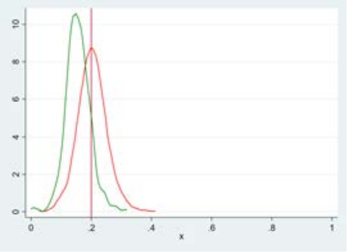
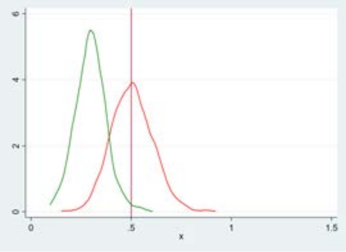
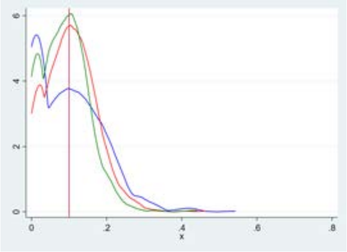
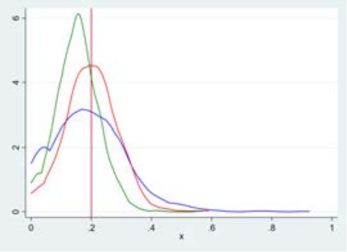
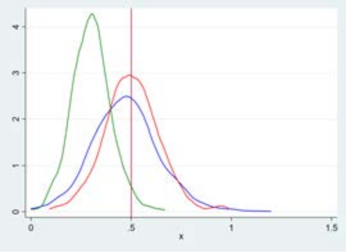
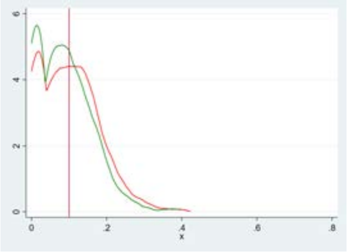
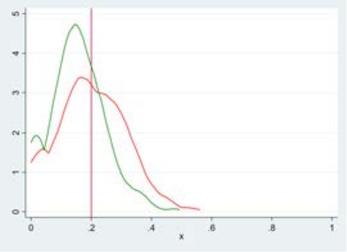
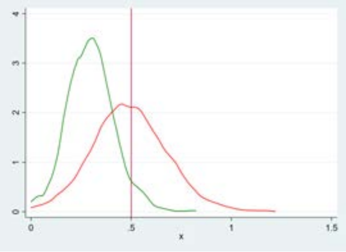
Control Function	—
MPEC	—
2SLS	—

Table 4: Distribution of the Estimates of $\bar{\beta}_2^x$

Var(ξ)=0.1	Var(ξ)=0.1	Var(ξ)=0.1
Var(β)=(0,0.1,0.1,0.1,0.05)	Var(β)=(0,0.2,0.2,0.2,0.1)	Var(β)=(0,0.5,0.5,0.5,0.2)
		
Var(ξ)=0.5	Var(ξ)=0.5	Var(ξ)=0.5
Var(β)=(0,0.1,0.1,0.1,0.05)	Var(β)=(0,0.2,0.2,0.2,0.1)	Var(β)=(0,0.5,0.5,0.5,0.2)
		
Var(ξ)=1.0	Var(ξ)=1.0	Var(ξ)=1.0
Var(β)=(0,0.1,0.1,0.1,0.05)	Var(β)=(0,0.2,0.2,0.2,0.1)	Var(β)=(0,0.5,0.5,0.5,0.2)
		

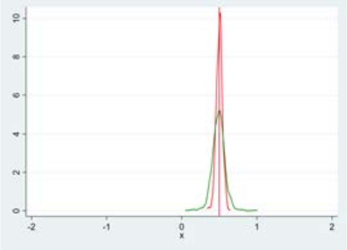
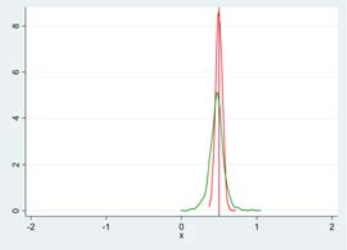
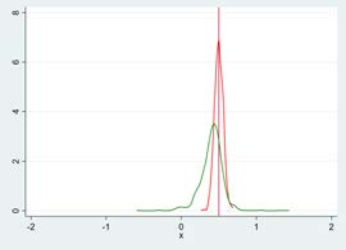
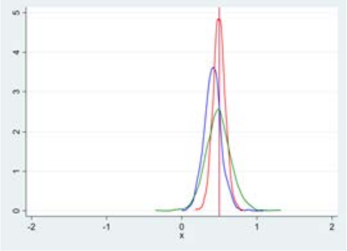
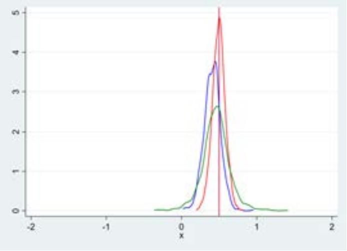
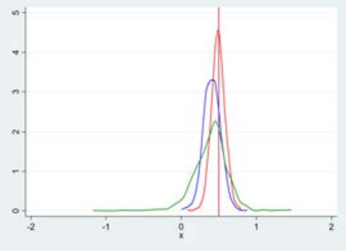
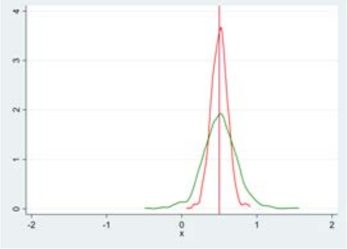
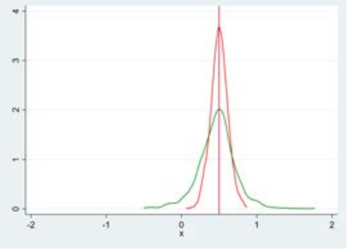
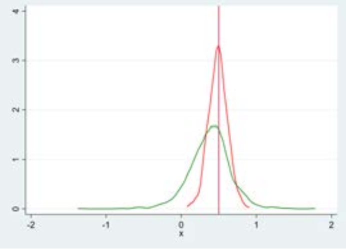
Control Function	🟦
MPEC	🟥
2SLS	🟩

Table 5: Distribution of the Estimates of Var(β_2^x)

Var(ξ)=0.1	Var(ξ)=0.1	Var(ξ)=0.1
Var(β)=(0,0.1,0.1,0.1,0.05)	Var(β)=(0,0.2,0.2,0.2,0.1)	Var(β)=(0,0.5,0.5,0.5,0.2)
		
Var(ξ)=0.5	Var(ξ)=0.5	Var(ξ)=0.5
Var(β)=(0,0.1,0.1,0.1,0.05)	Var(β)=(0,0.2,0.2,0.2,0.1)	Var(β)=(0,0.5,0.5,0.5,0.2)
		
Var(ξ)=1.0	Var(ξ)=1.0	Var(ξ)=1.0
Var(β)=(0,0.1,0.1,0.1,0.05)	Var(β)=(0,0.2,0.2,0.2,0.1)	Var(β)=(0,0.5,0.5,0.5,0.2)
		

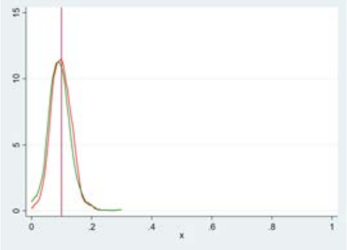
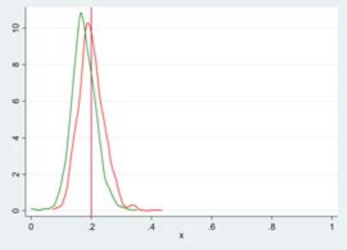
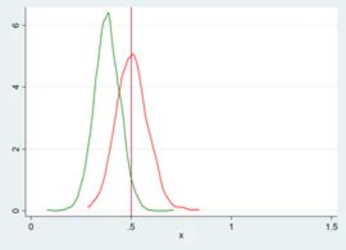
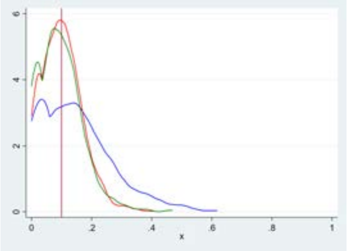
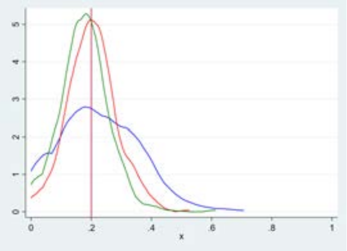
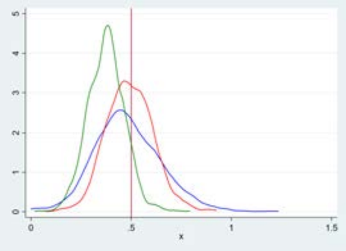
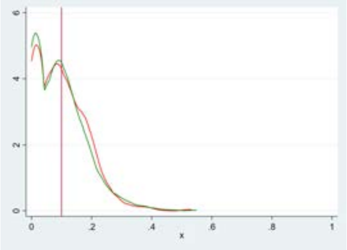
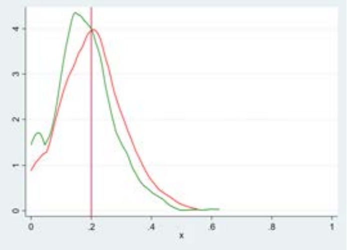
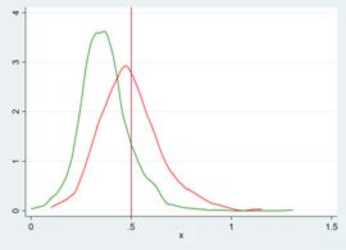
Control Function	🟦
MPEC	🟥
2SLS	🟩

Table 6: Distribution of the Estimates of $\bar{\beta}_3^x$

Var(ξ)=0.1	Var(ξ)=0.1	Var(ξ)=0.1
Var(β)=(0,0.1,0.1,0.1,0.05)	Var(β)=(0,0.2,0.2,0.2,0.1)	Var(β)=(0,0.5,0.5,0.5,0.2)
		
Var(ξ)=0.5	Var(ξ)=0.5	Var(ξ)=0.5
Var(β)=(0,0.1,0.1,0.1,0.05)	Var(β)=(0,0.2,0.2,0.2,0.1)	Var(β)=(0,0.5,0.5,0.5,0.2)
		
Var(ξ)=1.0	Var(ξ)=1.0	Var(ξ)=1.0
Var(β)=(0,0.1,0.1,0.1,0.05)	Var(β)=(0,0.2,0.2,0.2,0.1)	Var(β)=(0,0.5,0.5,0.5,0.2)
		

Control Function	—
MPEC	—
2SLS	—

Table 7: Distribution of the Estimates of Var(β_3^x)

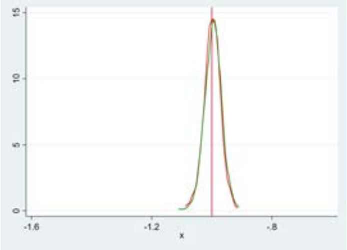
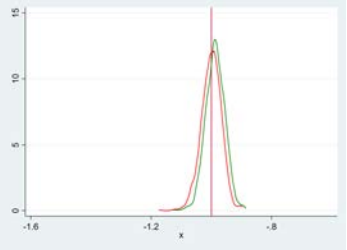
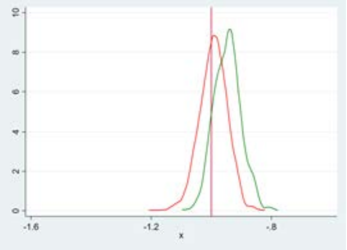
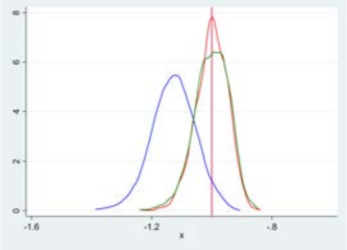
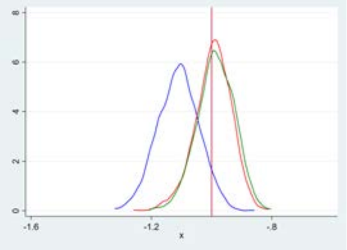
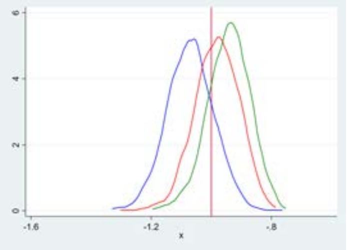
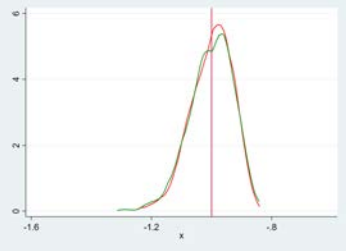
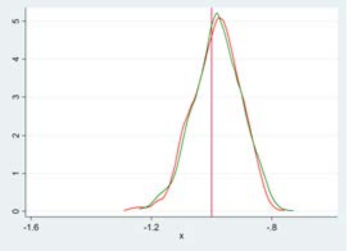
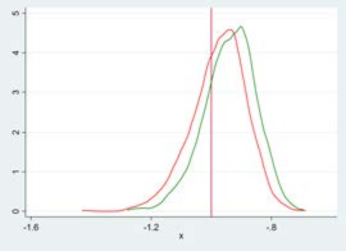
Var(ξ)=0.1	Var(ξ)=0.1	Var(ξ)=0.1
Var(β)=(0,0.1,0.1,0.1,0.05)	Var(β)=(0,0.2,0.2,0.2,0.1)	Var(β)=(0,0.5,0.5,0.5,0.2)
		
Var(ξ)=0.5	Var(ξ)=0.5	Var(ξ)=0.5
Var(β)=(0,0.1,0.1,0.1,0.05)	Var(β)=(0,0.2,0.2,0.2,0.1)	Var(β)=(0,0.5,0.5,0.5,0.2)
		
Var(ξ)=1.0	Var(ξ)=1.0	Var(ξ)=1.0
Var(β)=(0,0.1,0.1,0.1,0.05)	Var(β)=(0,0.2,0.2,0.2,0.1)	Var(β)=(0,0.5,0.5,0.5,0.2)
		

Control Function ——

MPEC ——

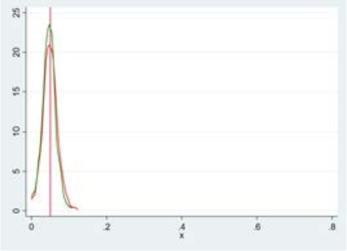
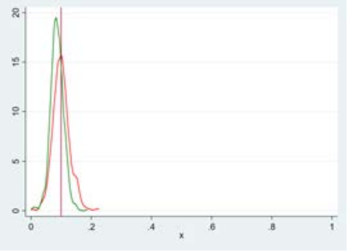
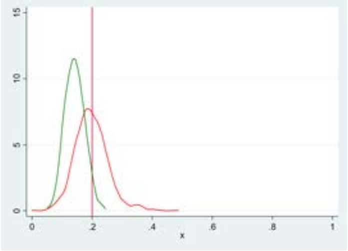
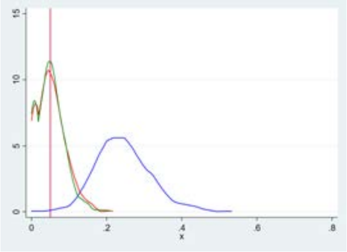
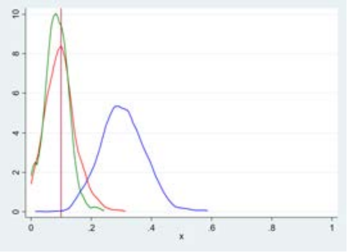
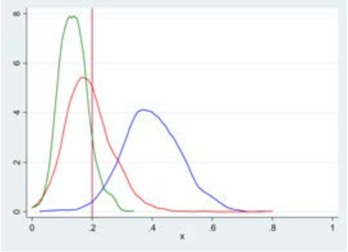
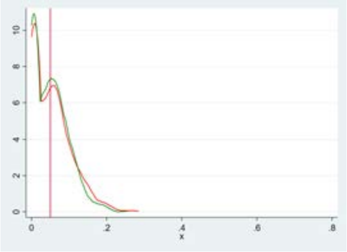
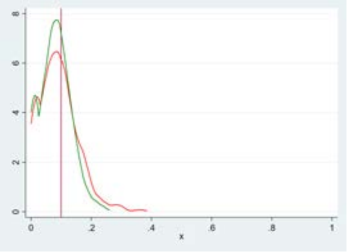
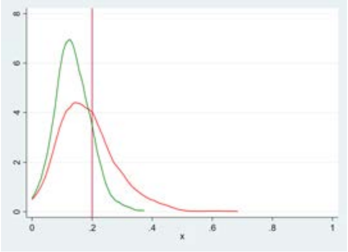
2SLS ——

Table 8: Distribution of the Estimates of $\bar{\beta}^p$

Var(ξ)=0.1	Var(ξ)=0.1	Var(ξ)=0.1
Var(β)=(0,0.1,0.1,0.1,0.05)	Var(β)=(0,0.2,0.2,0.2,0.1)	Var(β)=(0,0.5,0.5,0.5,0.2)
		
Var(ξ)=0.5	Var(ξ)=0.5	Var(ξ)=0.5
Var(β)=(0,0.1,0.1,0.1,0.05)	Var(β)=(0,0.2,0.2,0.2,0.1)	Var(β)=(0,0.5,0.5,0.5,0.2)
		
Var(ξ)=1.0	Var(ξ)=1.0	Var(ξ)=1.0
Var(β)=(0,0.1,0.1,0.1,0.05)	Var(β)=(0,0.2,0.2,0.2,0.1)	Var(β)=(0,0.5,0.5,0.5,0.2)
		

Control Function	🟦
MPEC	🟥
2SLS	🟩

Table 9: Distribution of the Estimates of Var(β^p)

Var(ξ)=0.1	Var(ξ)=0.1	Var(ξ)=0.1
Var(β)=(0,0.1,0.1,0.1,0.05)	Var(β)=(0,0.2,0.2,0.2,0.1)	Var(β)=(0,0.5,0.5,0.5,0.2)
		
Var(ξ)=0.5	Var(ξ)=0.5	Var(ξ)=0.5
Var(β)=(0,0.1,0.1,0.1,0.05)	Var(β)=(0,0.2,0.2,0.2,0.1)	Var(β)=(0,0.5,0.5,0.5,0.2)
		
Var(ξ)=1.0	Var(ξ)=1.0	Var(ξ)=1.0
Var(β)=(0,0.1,0.1,0.1,0.05)	Var(β)=(0,0.2,0.2,0.2,0.1)	Var(β)=(0,0.5,0.5,0.5,0.2)
		

Control Function	—
MPEC	—
2SLS	—

Table 10: Pseudo True Value: Increasing-number-of-markets Approach

Parameter	Scenarios								
True Var(β) :	$(0, 0.1, 0.1, 0.1, 0.05)$			$(0, 0.2, 0.2, 0.2, 0.1)$			$(0, 0.5, 0.5, 0.5, 0.2)$		
True Var(ξ) :	0.1	0.5	1	0.1	0.5	1	0.1	0.5	1
$\beta_0 = -1$	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.02	-1.03	-1.03
	(.0043)	(.0050)	(.0058)	(.011)	(.012)	(.013)	(.032)	(.035)	(.038)
$\bar{\beta}_1^x = 1.5$	1.51	1.51	1.51	1.53	1.53	1.53	1.56	1.57	1.57
	(.022)	(.023)	(.024)	(.050)	(.050)	(.050)	(.13)	(.13)	(.13)
$\bar{\beta}_2^x = 1.5$	1.51	1.51	1.51	1.52	1.52	1.52	1.55	1.56	1.56
	(.023)	(.024)	(.025)	(.048)	(.049)	(.049)	(.12)	(.12)	(.12)
$\bar{\beta}_3^x) = 0.5$	0.487	0.487	0.487	0.465	0.465	0.464	0.403	0.400	0.398
	(.022)	(.022)	(.022)	(.048)	(.047)	(.047)	(.12)	(.12)	(.11)
$\bar{\beta}^p = -1$	-0.999	-0.999	-0.999	-0.990	-0.990	-0.990	-0.954	-0.955	-0.956
	(.0086)	(.0088)	(.0090)	(.0184)	(.0186)	(.0188)	(.043)	(.044)	(.045)
Var(β_1^x)	0.0857	0.0856	0.0856	0.152	0.152	0.152	0.288	0.290	0.291
	(.011)	(.011)	(.011)	(.028)	(.027)	(.027)	(.078)	(.076)	(.075)
Var(β_2^x)	0.0863	0.0865	0.0866	0.152	0.152	0.153	0.284	0.286	0.288
	(.0086)	(.0086)	(.0087)	(.0205)	(.020)	(.020)	(.059)	(.057)	(.056)
Var(β_3^x)	0.0952	0.0949	0.0946	0.182	0.181	0.181	0.400	0.399	0.397
	(.0097)	(.010)	(.010)	(.024)	(.023)	(.023)	(.063)	(.063)	(.062)
Var(β^p)	0.0480	0.0479	0.0478	0.0888	0.088	0.088	0.148	0.147	0.147
	(.0056)	(.0057)	(.0059)	(.013)	(.013)	(.014)	(.031)	(.032)	(.033)

Table 11: Pseudo True Value: Moment-based Approach

Parameter	Scenarios								
True Var(β) :	$(0, 0.1, 0.1, 0.1, 0.05)$			$(0, 0.2, 0.2, 0.2, 0.1)$			$(0, 0.5, 0.5, 0.5, 0.2)$		
True Var(ξ) :	0.1	0.5	1	0.1	0.5	1	0.1	0.5	1
$\beta_0 = -1$	-1.01	-1.01	-1.01	-1.04	-1.04	-1.04	-1.11	-1.11	-1.12
$\bar{\beta}_1^x = 1.5$	1.49	1.49	1.49	1.48	1.48	1.48	1.43	1.43	1.43
$\bar{\beta}_2^x = 1.5$	1.49	1.49	1.49	1.48	1.48	1.48	1.43	1.43	1.43
$\bar{\beta}_3^x = 0.5$	0.496	0.496	0.496	0.486	0.486	0.486	0.455	0.455	0.455
$\bar{\beta}^p = -1$	-0.989	-0.988	-0.988	-0.958	-0.957	-0.955	-0.873	-0.869	-0.864
Var(β_1^x)	0.0854	0.0855	0.0855	0.149	0.149	0.149	0.275	0.275	0.276
Var(β_2^x)	0.0855	0.0855	0.0856	0.149	0.149	0.149	0.273	0.274	0.274
Var(β_3^x)	0.0938	0.0938	0.0937	0.176	0.175	0.175	0.369	0.368	0.366
Var(β^p)	0.0421	0.0421	0.0419	0.0685	0.0681	0.0676	0.0920	0.0906	0.0888

Table 12: Distribution of the Difference between True and Estimated Elasticity

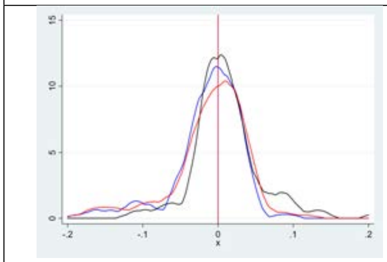
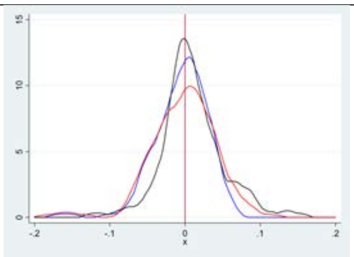
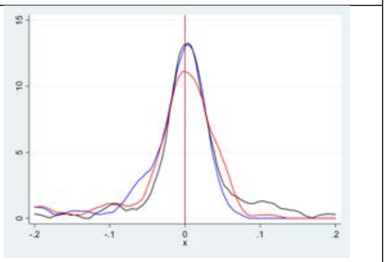
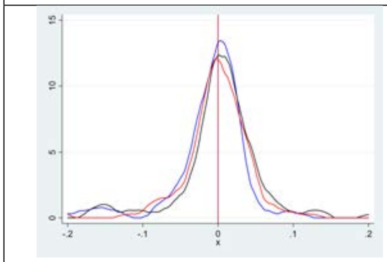
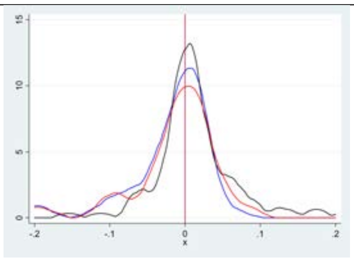
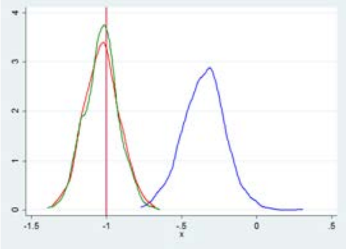
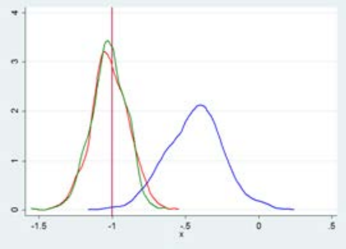
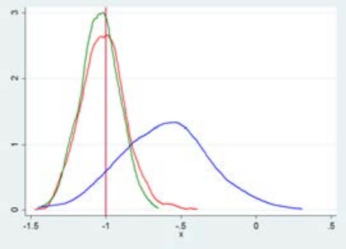
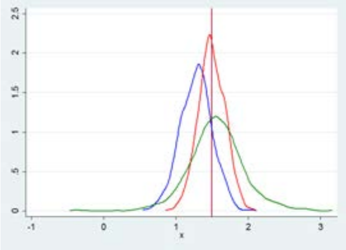
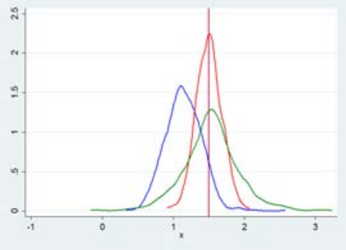
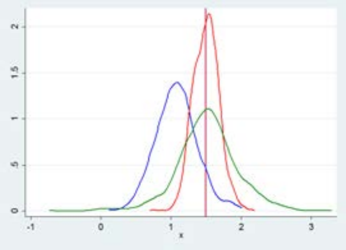
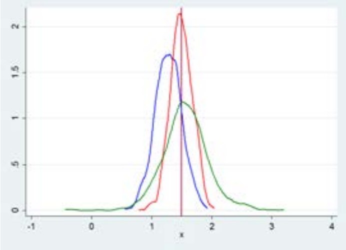
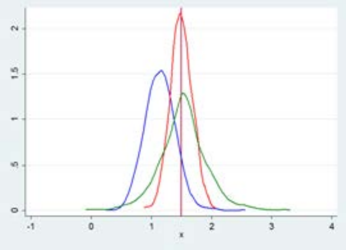
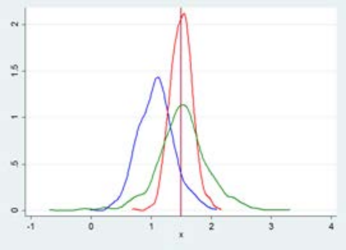
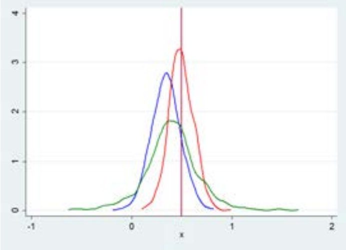
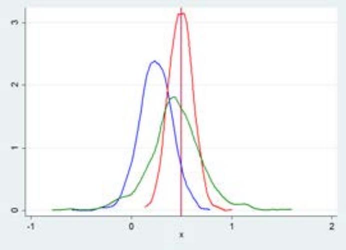
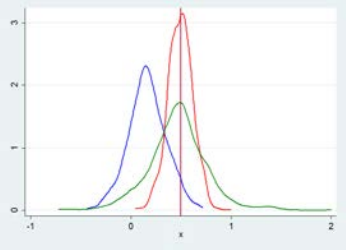
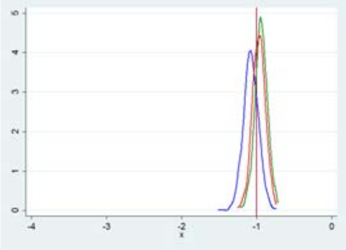
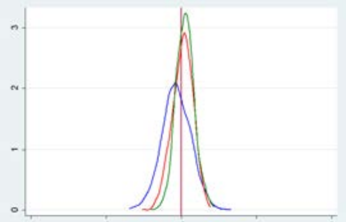
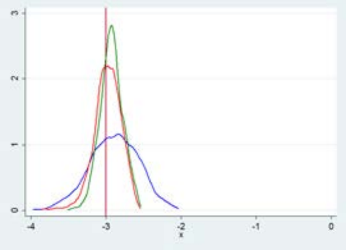
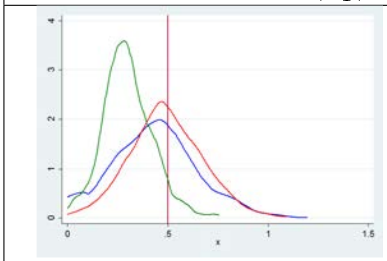
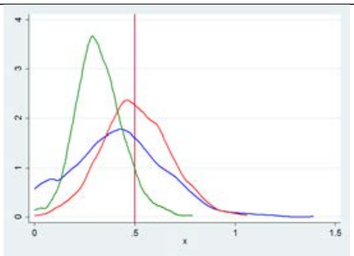
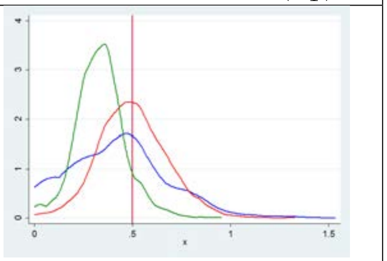
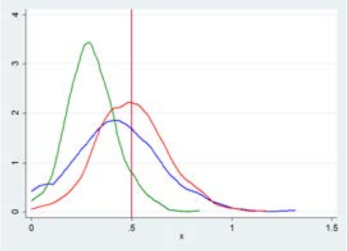
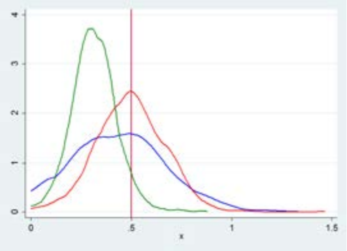
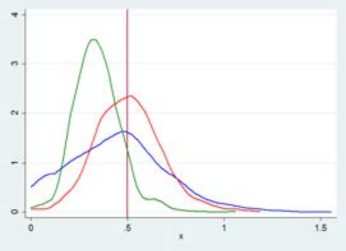
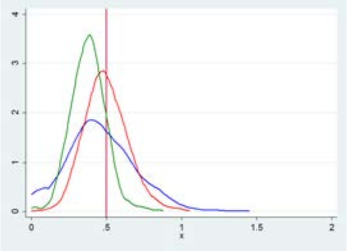
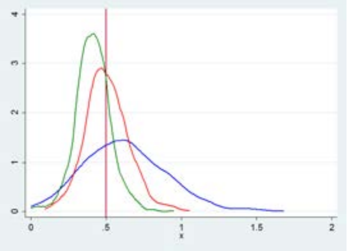
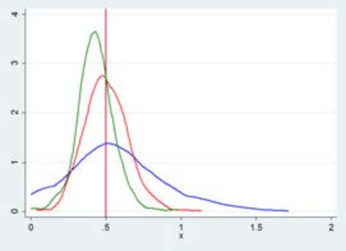
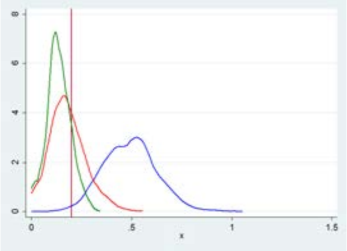
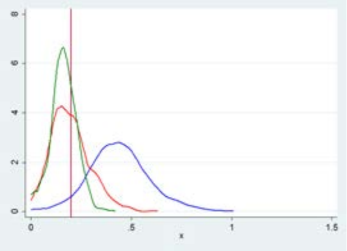
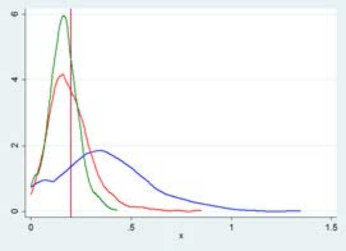
Product 5	Product 10	Product 15
		
Product 20	**Product 25**	**Legend**
		▬ MPEC ▬ Standard 2SLS ▬ Corrected 2SLS

Table 13: Distribution of the Estimates of the Means — Different $\bar{\beta}^p$

$\bar{\beta}^p = -1$	$\bar{\beta}^p = -2$	$\bar{\beta}^p = -3$
Distribution of β_0	Distribution of β_0	Distribution of β_0
		
Distribution of $\bar{\beta}^x_1$	Distribution of $\bar{\beta}^x_1$	Distribution of $\bar{\beta}^x_1$
		
Distribution of $\bar{\beta}^x_2$	Distribution of $\bar{\beta}^x_2$	Distribution of $\bar{\beta}^x_2$
		
Distribution of $\bar{\beta}^x_3$	Distribution of $\bar{\beta}^x_3$	Distribution of $\bar{\beta}^x_3$
		
Distribution of $\bar{\beta}^p$	Distribution of $\bar{\beta}^p$	Distribution of $\bar{\beta}^p$
		

Control Function	—
MPEC	—
2SLS	—

Table 14: Distribution of the Estimates of the Variances — Different $\bar{\beta}^p$

$\bar{\beta}^p = -1$	$\bar{\beta}^p = -2$	$\bar{\beta}^p = -3$
Distribution of Var(β_1^x)	Distribution of Var(β_1^x)	Distribution of Var(β_1^x)
 	 	
Distribution of Var(β_2^x)	Distribution of Var(β_2^x)	Distribution of Var(β_2^x)
 	 	
Distribution of Var(β_3^x)	Distribution of Var(β_3^x)	Distribution of Var(β_3^x)
 	 	
Distribution of Var(β^p)	Distribution of Var(β^p)	Distribution of Var(β^p)
 	 	

Control Function	🟦
MPEC	🟥
2SLS	🟩

Table 15: Testing for Zero Means — Standard 2SLS

Significance level	1%	5%	10%
2SLS with heteroskedasticity-robust standard error	0.904	0.793	0.711
GLS estimator and standard errors	0.889	0.765	0.678
2SLS with clustered standard error	0.904	0.780	0.702

Table 16: Testing for Zero Means — Corrected 2SLS

Significance level	1%	5%	10%
2SLS with heteroskedasticity-robust standard error	0.915	0.819	0.725
GLS estimator and standard errors	0.882	0.767	0.669
2SLS with clustered standard error	0.906	0.809	0.731

Table 17: Testing for Zero Variances — Standard 2SLS

Significance level	1%	5%	10%
2SLS with heteroskedasticity-robust standard error	0.746	0.625	0.556
GLS estimator and standard errors	0.738	0.618	0.542
2SLS with clustered standard error	0.740	0.617	0.547

Table 18: Testing for Zero Variances — Corrected 2SLS

Significance level	1%	5%	10%
2SLS with heteroskedasticity-robust standard error	0.792	0.688	0.626
GLS estimator and standard errors	0.773	0.680	0.613
2SLS with clustered standard error	0.775	0.679	0.620

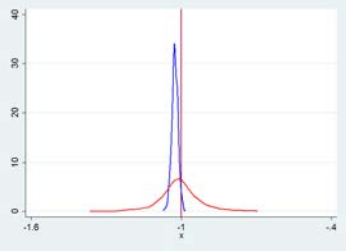
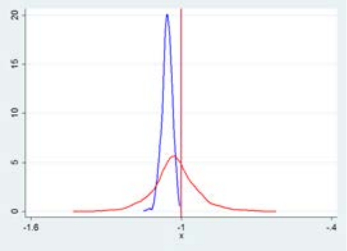
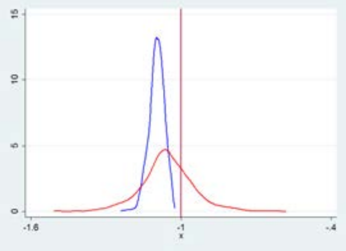
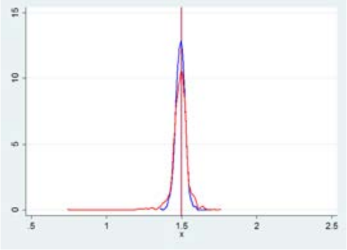
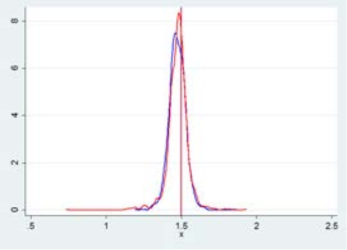
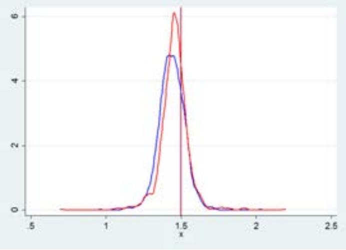
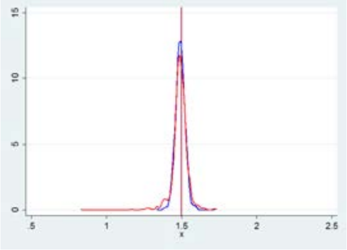
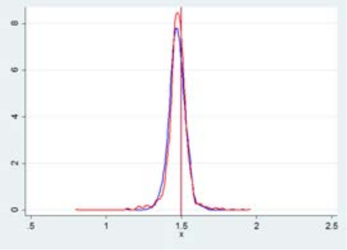
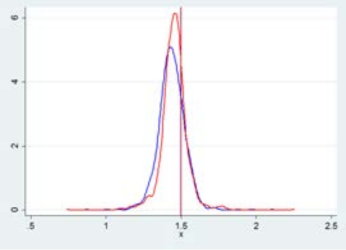
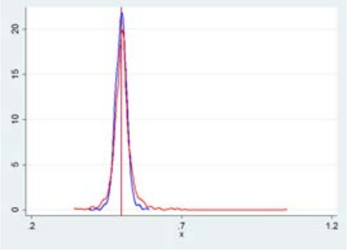
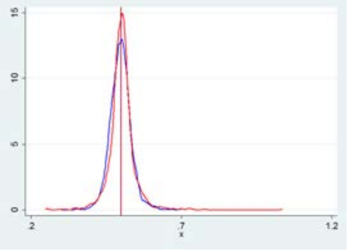
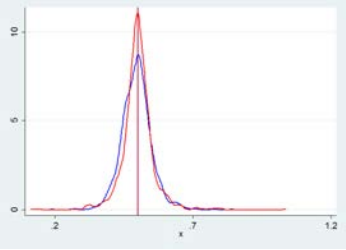
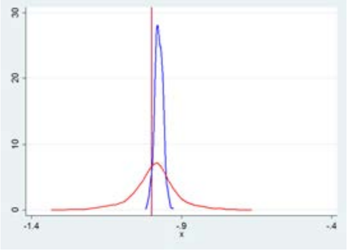
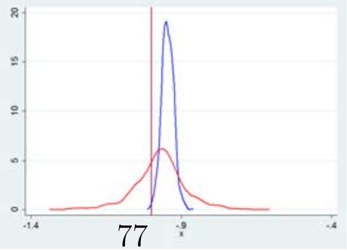
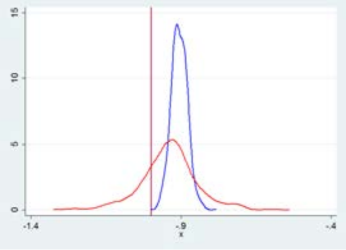
Table 19: Joint Test of Zero Means and Variances — Standard 2SLS

Significance level	1%	5%	10%
2SLS with heteroskedasticity-robust standard error	0.731	0.578	0.507
GLS estimator and standard errors	0.699	0.545	0.483
2SLS with clustered standard error	0.702	0.565	0.498

Table 20: Joint Test of Zero Means and Variances — Corrected 2SLS

Significant level	1%	5%	10%
2SLS with heteroskedasticity-robust standard error	0.756	0.642	0.568
GLS estimator and standard errors	0.738	0.631	0.548
2SLS with clustered standard error	0.749	0.617	0.546

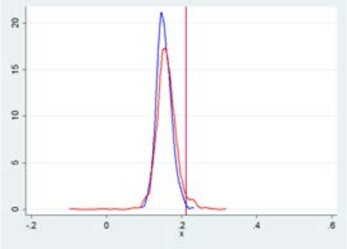
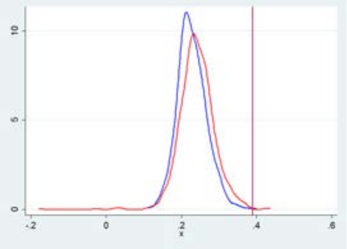
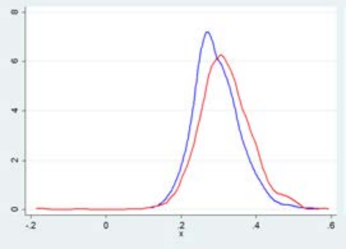
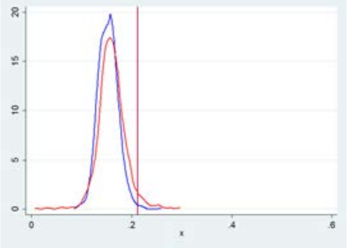
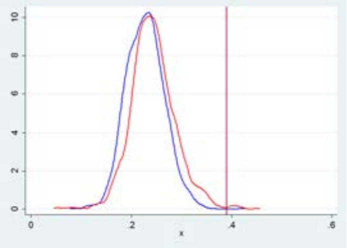
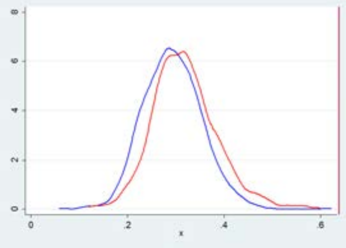
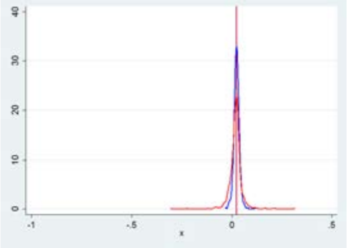
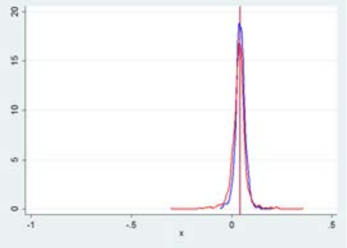
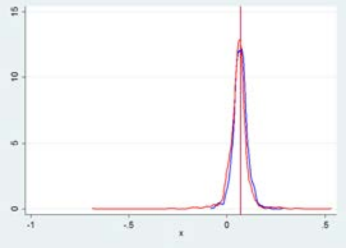
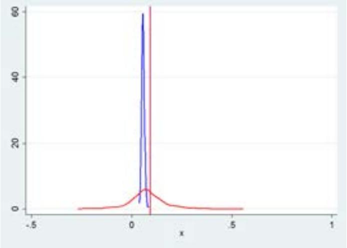
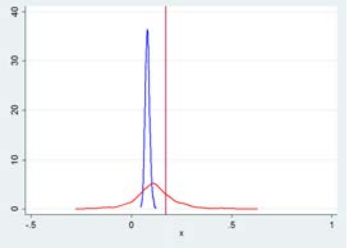
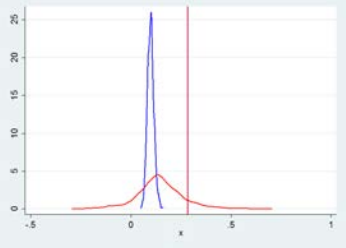
Table 21: Distribution of the Estimates of the Means (Lognormal Case)

$\sigma = 0.3$	$\sigma = 0.4$	$\sigma = 0.5$
Distribution of β_0	Distribution of β_0	Distribution of β_0
		
Distribution of $\bar{\beta}_1^x$	Distribution of $\bar{\beta}_1^x$	Distribution of $\bar{\beta}_1^x$
		
Distribution of $\bar{\beta}_2^x$	Distribution of $\bar{\beta}_2^x$	Distribution of $\bar{\beta}_2^x$
		
Distribution of $\bar{\beta}_3^x$	Distribution of $\bar{\beta}_3^x$	Distribution of $\bar{\beta}_3^x$
		
Distribution of $\bar{\beta}^p$	Distribution of $\bar{\beta}^p$	Distribution of $\bar{\beta}^p$
		

Only include 2^{nd} moment ———

Include 3^{rd} moment ———

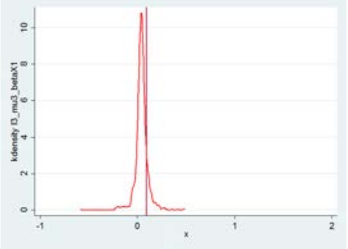
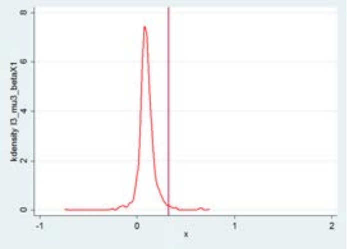
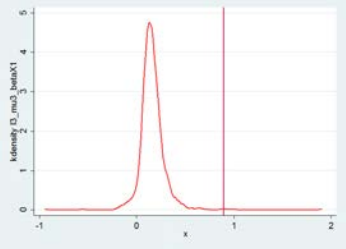
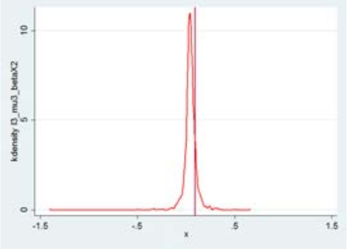
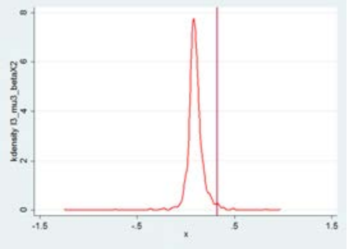
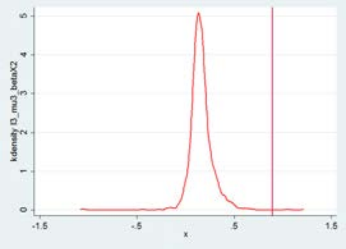
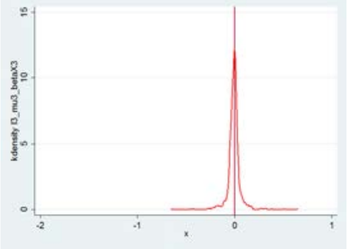
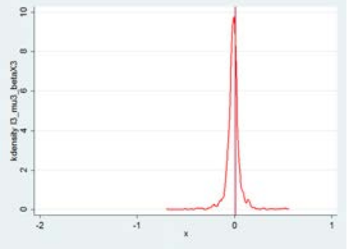
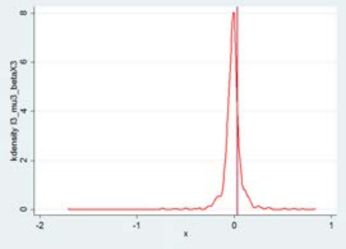
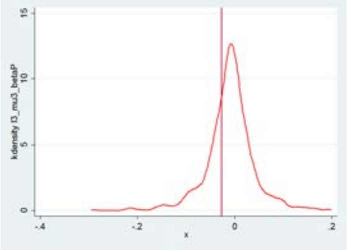
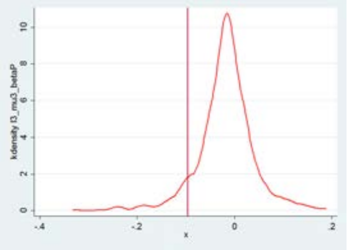
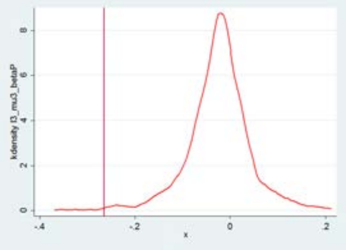
Table 22: Distribution of the Estimates of the Variances (Lognormal Case)

$\sigma = 0.3$	$\sigma = 0.4$	$\sigma = 0.5$
Distribution of Var(β_1^x)	Distribution of Var(β_1)	Distribution of Var(β_1)
		
Distribution of Var(β_2^x)	Distribution of Var(β_2)	Distribution of Var(β_2)
		
Distribution of Var(β_3^x)	Distribution of Var(β_3)	Distribution of Var(β_3)
		
Distribution of Var(β^p)	Distribution of Var(β^p)	Distribution of Var(β^p)
		

Only include 2nd moment ━━━━

Include 3rd moment ━━━━

Table 23: Distribution of the Estimates of the Third-order Moments (Lognormal Case)

$\sigma = 0.3$	$\sigma = 0.4$	$\sigma = 0.5$
Distribution of β_1^x 3rdM	Distribution of β_1^x 3rdM	Distribution of β_1^x 3rdM
		
Distribution of β_2^x 3rdM	Distribution of β_2^x 3rdM	Distribution of β_2^x 3rdM
		
Distribution of β_3^x 3rdM	Distribution of β_3^x 3rdM	Distribution of β_3^x 3rdM
		
Distribution of β^p 3rdM	Distribution of β^p 3rdM	Distribution of β^p 3rdM
		

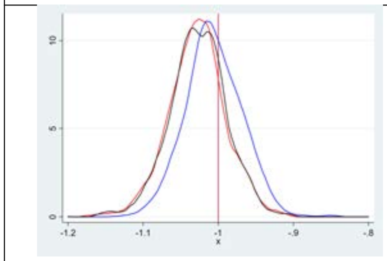
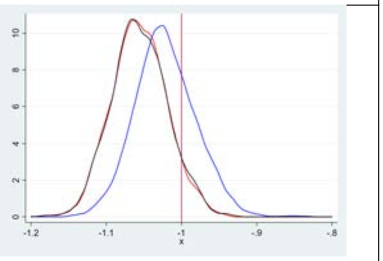
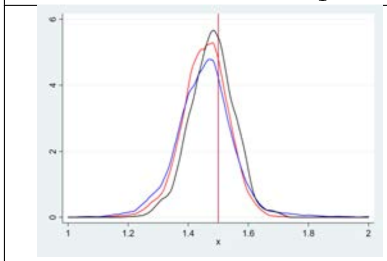
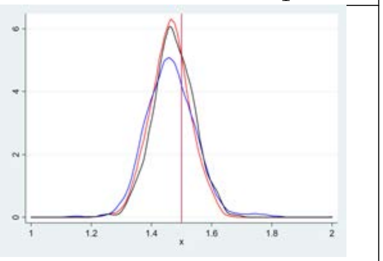
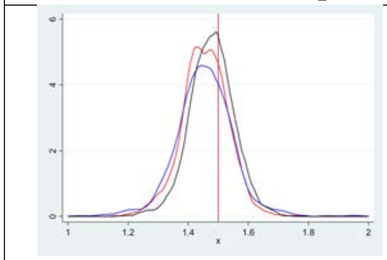
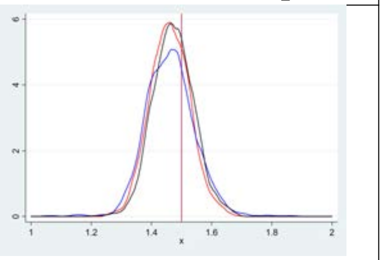
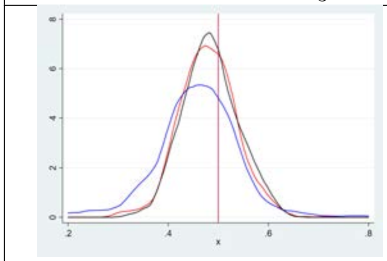
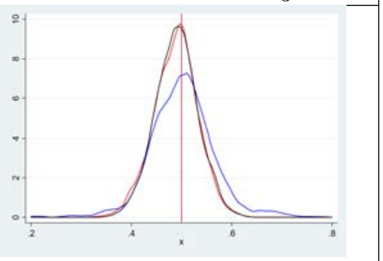
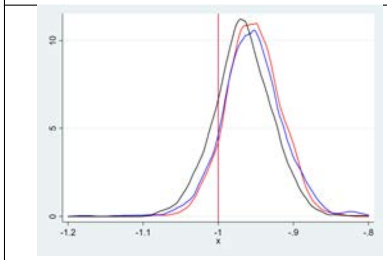
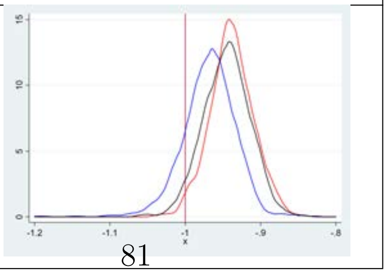
Only include 2nd moment ——

Include 3rd moment ——

Table 24: Summary Statistics for the Lognormal Case

Parameter	Scenarios					
σ	$\sigma = 0.3$		$\sigma = 0.4$		$\sigma = 0.5$	
Moments included:	2	3	2	3	2	3
$\beta_0 = -1$	-1.03	-1.01	-1.06	-1.03	-1.10	-1.06
	(0.013)	(0.080)	(0.021)	(0.092)	(0.031)	(0.107)
$\bar{\beta}_1^x = 1.5$	1.49	1.49	1.47	1.48	1.44	1.45
	(0.032)	(0.057)	(0.056)	(0.072)	(0.086)	(0.096)
$\bar{\beta}_2^x = 1.5$	1.49	1.49	1.47	1.47	1.44	1.45
	(0.033)	(0.054)	(0.057)	(0.070)	(0.088)	(0.093)
$\bar{\beta}_3^x = 0.5$	0.499	0.502	0.497	0.502	0.496	0.501
	(0.020)	(0.039)	(0.036)	(0.047)	(0.056)	(0.060)
$\bar{\beta}^p = -1$	-0.976	-0.991	-0.946	-0.972	-0.906	-0.940
	(0.014)	(0.076)	(0.020)	(0.084)	(0.027)	(0.095)
$\text{Var}(\beta_1^x) = 0.212/0.390/0.639$	0.153	0.159	0.229	0.241	0.295	0.316
	(0.020)	(0.030)	(0.039)	(0.048)	(0.062)	(0.072)
$\text{Var}(\beta_2^x) = 0.212/0.390/0.639$	0.153	0.160	0.228	0.244	0.294	0.319
	(0.020)	(0.028)	(0.038)	(0.045)	(0.060)	(0.067)
$\text{Var}(\beta_3^x) = 0.04/0.043/0.071$	0.025	0.030	0.045	0.036	0.068	0.056
	(0.014)	(0.033)	(0.025)	(0.041)	(0.038)	(0.059)
$\text{Var}(\beta^p) = 0.094/0.174/0.284$	0.0579	0.077	0.081	0.115	0.099	0.145
	(0.007)	(0.095)	(0.011)	(0.11)	(0.016)	(0.12)
$3rdM(\beta_1^x) = 0.093/0.322/0.894$		0.044		0.096		0.160
		(0.067)		(0.091)		(0.135)
$3rdM(\beta_2^x) = 0.093/0.322/0.894$		0.043		0.097		0.161
		(0.082)		(0.101)		(0.130)
$3rdM(\beta_3^x) = 0.003/0.012/0.033$		-0.0073		-0.012		-0.015
		(0.070)		(0.084)		(0.120)
$3rdM(\beta^p) = -0.027/-0.096/-0.265$		-0.0095		-0.018		-0.024
		(0.050)		(0.058)		(0.066)

Table 25: Distribution of Three Estimates of the Means — Normal and Lognormal

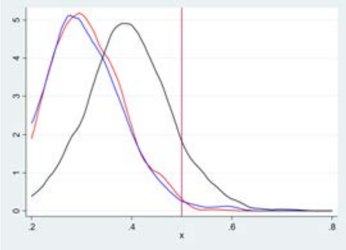
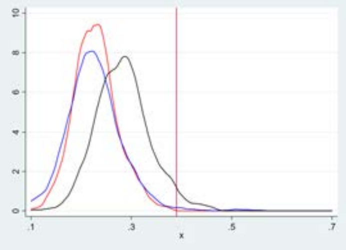
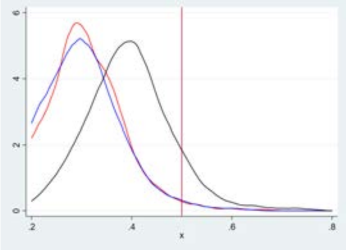
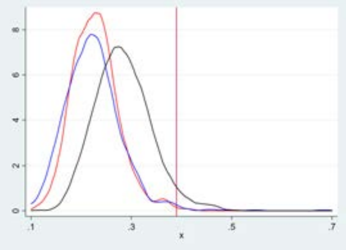
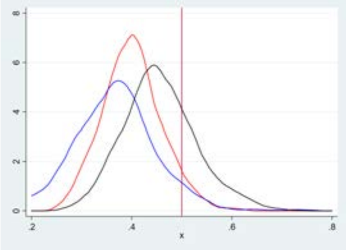
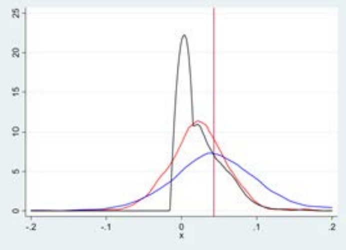
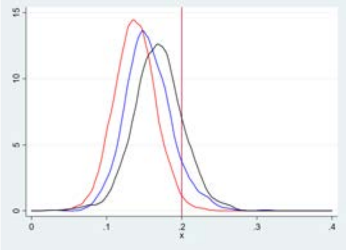
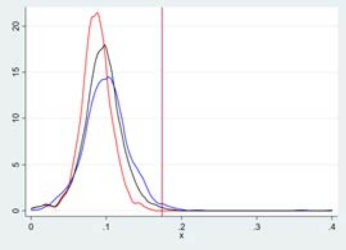
Normal	Lognormal
Distribution of β_0	Distribution of β_0
	
Distribution of β_1^x	Distribution of β_1^x
	
Distribution of β_2^x	Distribution of β_2^x
	
Distribution of β_3^x	Distribution of β_3^x
	
Distribution of β^p	Distribution of β^p
	

▬ Corrected 2SLS

▬ Standard 2SLS

▬ Kernel

Table 26: Distribution of Three Estimates of the Variances — Normal and Lognormal

Normal	Lognormal
Distribution of Var(β_1^x)	Distribution of Var(β_1^x)
	
Distribution of Var(β_2^x)	Distribution of Var(β_2^x)
	
Distribution of Var(β_3^x)	Distribution of Var(β_3^x)
	
Distribution of Var(β^p)	Distribution of Var(β^p)
	

━ Corrected 2SLS

━ Standard 2SLS

━ Kernel