Fast, Robust, and Approximately Correct: Estimating Mixed Demand Systems*

Very preliminary

Bernard Salanié[†] Frank A. Wolak[‡]

May 8, 2018

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 straightforward 2SLS estimator. We study in detail the models of market shares popular in empirical IO ("macro BLP"). 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 misspecifications in the higher moments of the distribution of the random coefficients. At the least, it provides excellent starting values for more commonly used estimators.

*We are grateful to Dan Ackerberg, Steve Berry, Chris Conlon, Pierre Dubois, Jeremy Fox, Han Hong, Thierry Magnac, Ariel Pakes, Mathias Reynaert, and Pedro Souza for their useful comments. We also thank Zeyu Wang for excellent research assistance.

[†]Columbia University.

[‡]Stanford University.

Introduction

Many econometric models are estimated from moment conditions that express the orthogonality of a random unobservable term η and instruments $\boldsymbol{Z}$:

$$E\left(\eta \boldsymbol{Z}\right) = 0.$$

In structural models, the unobservable term is usually obtained by inverting the set of equations that describe the underlying economic model. That is, we start from

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

where y stands for the observed data and θ_0 for the unknown parameters, while the function G is supposed to be known. 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

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

Inversion often is a step fraught with difficulties. Even when a simple algorithm exists, inversion is still costly: it 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

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

under the structural constraints (1). This "MPEC approach" has met with some success in dynamic programming and empirical industrial organization (Su–Judd 2012, Dubé et al 2012), but it still requires constrained minimization.

We propose here 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, s_0), where s_0 is a scalar that we have reasons to think is not too far from zero. We rewrite (1) as

$$G(y, F(y, \beta_0, s_0), \beta_0, s_0) = 0.$$

We expand $s \to F(y, \beta_0, s)$ around 0:

$$F(y, \beta_0, s_0) = F(y, \beta_0, 0) + F_s(y, \beta_0, 0)s_0 + \ldots + F_{ss\ldots s}(y, \beta_0, 0)\frac{s_0^L}{L!} + O(s_0^{L+1}),$$

where the subscript s denotes a partial derivative with respect to the argument s_0.

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

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

If the true value s_0 is not too large, one may hope to obtain a satisfactory estimator with the third of these "approximate estimators." In general, this is still requires solving a nonlinear minimization problem. However, suppose the function F satisfies the following three conditions:

1. $F_s(y, \beta_0, 0) \equiv 0$
2. $F(y, \beta, 0) \equiv f_0(y) - f_1(y)\beta$ is affine in β
3. and the second derivative $F_{ss}(y, \beta, 0)$ does not depend on β.

Denote $f_2(y) = -F_{ss}(y, \beta, 0)$. Then we would minimize

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

The optimal instruments Z^* for the parameters (β_0, s_0^2) in this problem are simply (Amemiya 1974)

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

which can be estimated directly from the data. Since the optimal instruments just identify the parameters, all that is left is to run two-stage least squares, that is regress $f_0(y)$ on $f_1(y)$ and $f_2(y)/2$ with instruments Z^*.

The resulting estimators of β_0 and s_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 interesting and useful ways that we will explore later.

Conditions 1–3 extend directly to a multivariate parameter $\boldsymbol{s_0}$. They may seem very demanding. Yet as we will show, under very weak conditions the "macro-BLP" model that is the workhorse of empirical IO satisfies all three. In this application, $\boldsymbol{s_0}$ 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 6.4 a class of models with unobserved heterogeneity to which conditions 1–3 apply.

Our approach builds on "small-Σ" 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 asymptotics of k-class estimators[1] 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 sample is large enough.

The small-σ approach was used by Chesher (1991) in models with measurement error. Most directly related to us, Chesher and Santos-Silva (2002) used a second-order approximation argument to reduce a mixed multinomial logit model to a "heterogeneity adjusted" unmixed multinomial logit model in which mean utilities have additional terms. 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 their random variation; and we only use basic linear estimators.

Section 1 presents the model popularized by Berry–Levinsohn–Pakes (1995) and

[1] Which include OLS and 2SLS.

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 who are not interested in the derivation of our formulæ in fact needn't read further. The rest of the paper justifies our algorithm; studies its properties; and discusses extensions.

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, and dealing with the possible endogeneity of prices by using GMM with appropriate instruments.

To fix ideas, we define "the standard model" as follows[2]. Let J products be available on each of T markets. Each market contains an infinity of consumers who choose one of J products. Consumer i on market t derives utility

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

from choosing product j. There is also a good 0, the "outside good", whose utility for consumer i is simply u_{i0t}. The error terms $\boldsymbol{\epsilon}$ and $\boldsymbol{u}$ are independent of each other, and of the covariates $\boldsymbol{X}$ and product effects $\boldsymbol{\xi}$. The vector $\boldsymbol{u}_{it}=(u_{\mathrm{i0t}},u_{\mathrm{i1t}},\ldots,u_{\mathrm{iJt}})$ is independently and identically distributed (iid) as standard type-I Extreme Value (EV); the product effects ξ_{jt} are unknown mean zero random variables, and the random variation in preferences $\boldsymbol{\epsilon}_i$ has a mean-zero distribution which is known up to its variance-covariance matrix $\boldsymbol{\Sigma}_0$.

Some of the covariates in $\boldsymbol{X}_{jt}$ may be correlated to the product fixed effects. The usual example is a model of imperfect price competition where the prices p_{jt} $j=1,2,...,J$ for market t, which firms set depend on the value of the vector of unobservable product characteristics, $\boldsymbol{\xi}_t$.

The parameters to be estimated are the mean coefficients $\boldsymbol{\beta}_0$ and the variance-covariance matrix $\boldsymbol{\Sigma}_0$. We collect them in $\boldsymbol{\theta}_0=(\boldsymbol{\beta}_0,\boldsymbol{\Sigma}_0)$. The data available consists

[2]While some of our exposition relies on it for simplicity, our methods apply to a more general model— see section 6.4.

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 $\boldsymbol{Z}_t$. 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 $j = 1, \ldots, J$.

The augmented market shares on market t are obtained by integration over the variation in preferences: for $j = 1, \ldots, J$

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

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 $j \geq 1$ 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[3] 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

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

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 formula (1974):

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

[3] See Berry (1994).

Since the Amemiya formula relies on a consistent first-step estimate of the parameters, this is still problematic. Gandhi-Houde (2016) propose "differentiation IVs" to approximate the optimal instruments for the parameters $\mathbf{\Sigma}$ of the distribution of the random preferences $\boldsymbol{\epsilon}$. They also suggest a simple regression to detect weak instruments. Armstrong (2016) pointed out that instruments based on the characteristics of competing products achieve identification through correlation with markups. But when there are a large number of products, many models of supply predict that markups just do not have enough variation, relative to sampling error. This can give inconsistent or just uninformative estimates[4].

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 pointed algorithm must be employed to compute the implied product effects $\boldsymbol{\xi}_t$ from the observed market shares $\boldsymbol{S}_t$ and current value of θ. 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 to resort to constrained optimization instead. Their simulations suggest that this "MPEC" approach often outperforms the NFP method, sometimes by a large factor. Lee–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 MPEC.

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 θ, but still requires solving a nonlinear optimization problem to compute an estimate of θ_0. Solving a nonlinear optimization problem for a potentially large set of parameters is time-consuming and typically requires starting values in the neighborhood of the optimal solution, closed-form gradients, and careful monitoring of optimization algorithm by the analyst because 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

[4]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.

is no random variation in preferences, and becomes a coarser approximation as the amplitude of random variation 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. Most obviously, it requires a single linear 2SLS regression that can be computed in microseconds with off-the-shelf software. Moreover, it need assume very little on the distribution of the random variation in preferences $\boldsymbol{\epsilon}$ (beyond its limited amplitude), justifying the "robust" in our title. Since our estimating equation is linear, computing the optimal instruments is also straightforward.

For those who 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. It also provides useful diagnoses about how well different parameters can be identified with a particular model and dataset; and a very simple way to select between models.

2 2SLS Estimation in the Standard BLP Model

For the reader in a hurry, we give in this section a step-by-step guide to implementing our estimator in the standard model. This requires some notation. The dimensions of the vectors and matrices are as follows:

- for each $j \geq 1$ 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, $n_e \leq n_X$
- $\boldsymbol{v}$ is a row vector with n_v components
- $\boldsymbol{B}$ is an $n_e \times n_v$ matrix.

We denote $\mathcal{I}$ the set of pairs of indices (m, n) such that the variance-covariance element $\Sigma_{mn} = \text{cov}(\epsilon_{im}, \epsilon_{in})$ is *not* restricted to be zero[5]. We also assume that we use

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

all moment conditions

$$E\left(\boldsymbol{\xi}_t|\boldsymbol{Z}_{jt}\right)$$

for $j=1,\ldots,J$ and $t=1,\ldots,T$.

Our procedure runs as follows:

Algorithm 1. ***FRAC estimation of the standard 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 $(j\geq 1,t)$:*

   (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"*

$$K_{mn}^{jt}=\left(\frac{X_{jtm}}{2}-e_{tm}\right)X_{jtn}.$$

3. *(recommended) for every $j\geq 1$, regress* flexibly *$\boldsymbol{X}_{jt}$ on $\boldsymbol{Z}_{jt}$*

4. *(recommended) for every $j\geq 1$ and every (m,n) in $\mathcal{I}$, regress* flexibly *K_{mn}^{jt} on $\boldsymbol{Z}_{jt}$*

5. *(recommended) take as instruments $\hat{\boldsymbol{Z}}_{jt}$ the fitted variables in the previous flexible regressions*

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

7. *run a two-stage least squares regression of $\boldsymbol{y}$ on $\boldsymbol{X}$ with instruments $\boldsymbol{Z}$, or $\hat{\boldsymbol{Z}}$ if using steps 3–5*

8. *(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 6.*

Steps 3–5 above are meant to compute the optimal instruments for our approximate model. The optimal (Amemiya) instruments are given by a nonparametric regression; but given the curse of dimensionality, a regression that allows for a reasonably flexible specification is more realistic. It could be combined with a model selection process such as Lasso. Steps 3–5 could also be skipped, with the analyst using $\hat{\boldsymbol{Z}} = \boldsymbol{Z}$. Given earlier findings in Reynaert–Verboven 2014, we do not recommend it. Because of the well-known property of the 3SLS estimator, that misspecification of one equation of the model can lead to inconsistency in parameter estimates of all equations of the model, it is unclear whether Step 8 is worth the additional effort. We intend to explore it in further work.

3 Second-order Expansions

The rest of the paper justifies algorithm 1 and discusses extensions. We start in this section by deriving the small-σ expansions of the introduction.

We start from a specification of the 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 $j = 1, \ldots, J$; 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;
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 standard type-I EV iid.

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 $\boldsymbol{\epsilon}$ jointly normal. 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 does[6] and also imposes that all moments of $\boldsymbol{\epsilon}$ are finite-valued. The most common specification has a Gaussian $\boldsymbol{\epsilon}$ and of course obeys Assumption 1.

Assumption 1. *For some integer $L \geq 2$, all moments of order $1 \leq l \leq L+1$ of the vector $\boldsymbol{\epsilon}$ are finite; they are of order l in some non-negative scalar σ. The first moment is zero: $E\boldsymbol{\epsilon} = \mathbf{0}$.*

It will be convenient to write $\boldsymbol{\epsilon} \equiv \sigma \boldsymbol{B}\boldsymbol{v}$ with $\boldsymbol{v}$ a random vector of mean zero and variance identity, so that $\sigma\boldsymbol{B}$ is a square root of the variance-covariance matrix of $\boldsymbol{\epsilon}$: $\boldsymbol{\Sigma} = \sigma^2\boldsymbol{B}\boldsymbol{B}'$. We only use this decomposition for intermediate results; our expansions will not depend on how σ and $\boldsymbol{B}$ are normalized. Note that the dimensions of these vectors and matrices are as follows:

- $\boldsymbol{X}_j$ is a row vector with n_X components

[6] 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.

- $\boldsymbol{\beta}$ is a column vector with n_X components
- $\boldsymbol{\epsilon}$ is a row vector with n_e components
- $\boldsymbol{v}$ is a row vector with n_v components
- $\boldsymbol{B}$ is an $n_e \times n_v$ matrix.

Moreover, 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 $\boldsymbol{u}$'s have iid Type I extreme value distributions. We will show in section 6.1 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

$$\xi_j = \log \frac{S_j}{S_0} - \boldsymbol{X}_j \boldsymbol{\beta} \text{ for } j \geq 1. \tag{4}$$

This is the starting point of the contraction algorithm described in Berry et al (1995).

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[7])

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

In this equation, $\boldsymbol{X}$ regroups 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:

Definition 1. *For any J-dimensional vector $\boldsymbol{T}$ of J components, we define the scalar*

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

[7] We return to this point in section 3.

By extension, if $\boldsymbol{m}$ *is a 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 expectation, as the augmented market shares S_k for $k \geq 1$ 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 $\boldsymbol{X}_j\boldsymbol{\epsilon}$. 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 first 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

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

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

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

Proof. See Appendix **??**. □

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 since $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{B}\boldsymbol{B}'$, 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 $\boldsymbol{K}^j$ by

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

so that we can also write $\sigma^2 \frac{a_{2j}}{2} = -\operatorname{Tr}[\boldsymbol{\Sigma}\boldsymbol{K}^j]$. The matrices $\boldsymbol{K}^j$ can be constructed straightforwardly from the covariates $\boldsymbol{X}$ and the market shares $\boldsymbol{S}$. We call their elements the "artificial regressors", for reasons that will soon become clear. Given that $\boldsymbol{\Sigma}$ is symmetric,

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

Additional a priori restrictions can be accommodated very easily. It is for instance common to restrict $\boldsymbol{\Sigma}$ to be diagonal. Then only n_X artificial regressors enter in this sum; moreover,

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

If $\boldsymbol{\Sigma}$ is not diagonal, then we need to also use the artificial regressors

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

To summarize, we have:

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

$$\xi_j = \log \frac{S_j}{S_0} - \boldsymbol{X}_j \boldsymbol{\beta} - \sum_{m=1}^{n_X} \Sigma_{mm} K^j_{mm} - \sum_{m<n} \Sigma_{mn} \left(K^j_{mn} + K^j_{mn} \right) + O(\|\Sigma\|^{k/2}), \qquad (8)$$

where $k = 4$ if the model is symmetric, and $k \geq 3$ otherwise; and the artificial regressors are given by

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

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

4 2SLS Estimation

Equation (8) 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 $\xi_j = \log \frac{S_j}{S_0} - \boldsymbol{X}_j \boldsymbol{\beta} - \mathrm{Tr}[\boldsymbol{\Sigma} \boldsymbol{K}^j]$.

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 moment conditions

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

which only differs from the original model by a term of order σ^3 (or σ^4 if the model is symmetric). This can simply be done by choosing vector functions $\boldsymbol{Z}^*_{jt}$ of the instruments and running two-stage least squares: for each $j = 1, \ldots, J$, on the sample $t = 1, \ldots, T$,

- linearly regress $\boldsymbol{X}_{jt}$ and the relevant[8] variables $\boldsymbol{K}^{jt}$ on $\boldsymbol{Z}^*_{jt}$
- then linearly regress $\log(S_{jt}/S_{0t})$ on the fitted values.

These are just steps 2 and 6–7 of algorithm 1. Steps 3–5 follow the advice in Reynaert–Verboven (2014) by computing the optimal instruments. Since the model is linear, these are simply the nonparametric regressions of the covariates $\boldsymbol{X}_{jt}$ and of the artificial regressors $\boldsymbol{K}^{jt}$ on the instruments $\boldsymbol{Z}_{jt}$. Since both the covariates (obviously) and the artificial regressors can be constructing without estimating the BLP model, the computation of the optimal instruments is straightforward.

5 Properties of the 2SLS Estimation Approach

5.1 Pros and Cons

The drawback of our method is obvious: since this is only an approximate model, the resulting estimator $\hat{\boldsymbol{\theta}}$ will not converge to the true values as the number of markets T goes to infinity. We discuss this in much more detail in section 5.2. For now, let us note that this drawback is tempered by several considerations. First, asymptotics are not that relevant in empirical IO, as the number of markets available is typically small; finite-sample peformance matters more, and we will test that in section 7. More importantly, our estimator has several useful features. Let us list five of them:

1. because the estimator is linear 2SLS, computing it is extremely fast and can be done in microseconds with any of-the-shelf software.
2. we did not have to assume any distributional form for the random variation in preferences $\boldsymbol{v}$. This is a notable advantage on other methods: while they yield inconsisent estimates if the distribution of $\boldsymbol{v}$ is misspecified, our estimator remains consistent for the parameters of the approximate model.
3. computing the optimal instruments does not require any first-step estimate since the estimating equation is linear; they are given by the nonparametric regression

[8] 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.

of the elements of $\boldsymbol{X}_{jt}$ and $\boldsymbol{K}^{jt}$ on $\boldsymbol{Z}_{jt}$. That is, the optimal set of instruments $\boldsymbol{Z}^*_{jt}$ is simply

$$\left(E\left(\boldsymbol{X}_{jt}|\boldsymbol{Z}_{jt}\right), E\left(\boldsymbol{K}^{jt}|\boldsymbol{Z}_{jt}\right)\right).$$

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

5.2 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. 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 $\boldsymbol{\epsilon}$ has all moments 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 **??** in Appendix **??** plots the first four coefficients of the expansion in $(\sigma X_1)^2$ for the standard Gaussian binary model (that is, the Gaussian mixed logit) with one covariate X_1 as market shares vary between zero and one. While this simple example can only be illustrative, we find the figure encouraging as to the practical range of validity of the approximation.

5.3 Robustness

Our expansions rely on the properties of the derivatives of $L(t)$ 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 $\boldsymbol{\Sigma}$.

6 Extensions

Our technique can easily be extended to different models as long as the utility remains additive in the product effects $\boldsymbol{\xi}$. Morerover, the calculations of these and higher-order terms can easily be automated with the help of a symbolic algebra system.

6.1 Generalized Extreme Values

6.2 Higher-order terms

In Appendix **??**, we study in more detail the standard binary model, where calculations are easily done by hand or using symbolic software to any order of approximation.

More generally, return to the standard model and assume (as is often done in practice) that the ϵ_m are independent across the covariates $m = 1, \ldots, n_X$; let μ_{lm} denote the (unknown) expected value of ϵ_m^l. Tedious calculations[9] show that the

[9] Available from the authors.

second- to fourth-order terms of the expansion in σ (or, equivalently, in $\|\mathbf{\Sigma}\|^{1/2}$) are

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

with

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

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

and

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

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

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

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 $(\boldsymbol{\beta}, \boldsymbol{\mu_2}, \boldsymbol{\mu_3})$, for which it generates approximate estimates by 2SLS.

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 $\boldsymbol{\mu_2}$ (the last line). The first group suggests introducing more artificial regressors

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

whose coefficients are the μ_{4m}. The second group can only be dealt with by a nonlinear estimation method (albeit a very simple one).

6.3 Bias correction

Another way to use the third- and fourth-order terms 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 evaluate the bias due to 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

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

As the number of markets T gets large, $\hat{\boldsymbol{\theta}}_2$ converges to the solution $\boldsymbol{\theta}_2$ of $Ef_2(\boldsymbol{\theta}_2) = 0$, 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" $\boldsymbol{\xi}$. Since $E\left(\boldsymbol{Z}'\boldsymbol{\xi}(\boldsymbol{\theta}_0, \boldsymbol{X}, \boldsymbol{S}\right) = 0$, a fortiori $Ef_\infty(\boldsymbol{\theta}_0) = 0$ with

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

The dominant term in the asymptotic bias is given by expanding $Ef_\infty(\boldsymbol{\theta})$ around $\boldsymbol{\theta} = \boldsymbol{\theta}_2$; it is

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

The term in the inverse is easy to approximate:

$$E\frac{\partial f_\infty}{\partial\boldsymbol{\theta}}(\boldsymbol{\theta}_2) \simeq E\frac{\partial f_2}{\partial\boldsymbol{\theta}}(\boldsymbol{\theta}_2).$$

As for its factor:

$$
\begin{aligned}
Ef_\infty(\boldsymbol{\theta}_2) &= E(f_\infty - f_2)(\boldsymbol{\theta}_2) \\
&\simeq E\left(\frac{\partial \boldsymbol{e}_2}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta}_2)\boldsymbol{Z}\boldsymbol{W}\boldsymbol{Z}'\boldsymbol{\xi}_2(\boldsymbol{\theta}_2)\right) \\
&+ E\left(\frac{\partial \boldsymbol{\xi}_2}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta}_2)\boldsymbol{Z}\boldsymbol{W}\boldsymbol{Z}'(\boldsymbol{e}_2)(\boldsymbol{\theta}_2)\right).
\end{aligned}
$$

where $\boldsymbol{e}_2 = \boldsymbol{\xi} - \boldsymbol{\xi}_2$. It is easy to approximate since we know how to compute the higher-order terms $\boldsymbol{\xi}_3$ and $\boldsymbol{\xi}_4$, under any assumption about the skewness and kurtosis of $\boldsymbol{\epsilon}$.

6.4 Other Models with Random Coefficients

Let us return to our original structural equations (1). Now assume that the dependence of G in s comes from unobserved heterogeneity ε with scale parameter s, and that it takes the following form:

$$
G(y, \eta, \beta, s) \equiv G^*\left(y, E_\varepsilon A^*(y, \eta - f_1(y)\beta, s\varepsilon)\right) \tag{9}
$$

where ε is unobserved heterogeneity distributed independently of y and η and normalized by $E\varepsilon = 0$ and $E\varepsilon^2 = 1$. Note that the macro–BLP model takes this form, with $y = (\boldsymbol{S}, \boldsymbol{X})$, $\eta = \boldsymbol{\xi}$, $s = \sigma$, $\varepsilon = \boldsymbol{v}$, and

$$
A_j^* \equiv \Pr\left(j = \arg\max_{K=0,1,\ldots,J}\left(\boldsymbol{X}_k\boldsymbol{\beta} + \xi_k + \sigma \boldsymbol{x}_{\boldsymbol{k}}' \cdot \boldsymbol{v}\right)\right)
$$

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

Remember that $G(y, F(y, \beta, s), \beta, s) = 0$, so that $G(y, F(y, \beta, 0), \beta, 0) = 0$. Given (9), this gives $G^*(y, A^*(y, F(y, \beta, 0) - f_1(y)\beta, 0) =)0$. This can only hold if $F(y, \beta, 0) - f_1(y)\beta$ does not depend on β, which implies condition 2.

Now writing $G^*(y, E_\varepsilon A^*(y, F(y, \beta, s) - f_1(y)\beta, s\varepsilon)) = 0$ and taking derivatives with respect to s, we get

$$
\begin{aligned}
G_2^* E_\varepsilon\left(A_2^* F_s + A_3^*\varepsilon\right) &= 0 \\
G_{22}^*\left(E_\varepsilon\left(A_2^* F_s + A_3^*\varepsilon\right)\right)^2 + G_2^* E_\varepsilon\left(A_2^* F_{ss} + A_{22}^*(F_s)^2 + 2A_{23}^* F_s \varepsilon + A_{33}^*\varepsilon^2\right) &= 0
\end{aligned}
$$

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

$$G_2^* E_\varepsilon \left(A_2^* F_s(y, \beta, 0) + A_3^* \varepsilon \right) = 0,$$

where the derivatives A^* do not depend on ε since $s = 0$. It follows that $F_s(y, \beta_0, 0) = 0$ since $E\varepsilon = 0$. Therefore condition 2 also holds. Using the second equation at $s = 0$, and given that $F_s(y, \beta_0, 0) = 0$, we get

$$G_2^* E_\varepsilon \left(A_2^* F_{ss} + A_{33}^* \varepsilon^2 \right) = 0$$

so that

$$F_{ss}(y, \beta, 0) = -\frac{A_2^*}{A_{33}^*}(y, F(y, \beta, 0) - f_1(y)\beta)$$

But (9) satisfies condition 1, so that $F(y, \beta, 0) - f_1(y)\beta = f_0(y)$ and the right hand side is independent of β. Hence condition 3 holds, and any model that generates (9) can be easily estimated (approximately) by 2SLS.

Since the macro–BLP model belongs to this class, this confirms that conditions 1–3 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 Gaussian mixed model without the logit assumption in Appendix **??**.

7 Simulations

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) and the control function approach of Petrin and Train (2010). 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 "approximately correct" estimates of the means and variances of the random coefficients.

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 coefficient demand system with $T = 50$ markets and $J = 25$ products in each market, and $K = 3$ observed product characteristics. In the terms of this paper, this is a standard Gaussian model.

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

$$u_{ijt} = \beta^0 + \boldsymbol{X}_{jt}'\boldsymbol{\beta}_i^x - \beta_i^p p_{jt} + \xi_{jt} + \epsilon_{ijt}$$

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

We collect all parameters into

$$\boldsymbol{\theta} = (\bar{\beta}_1^x, \ldots, \bar{\beta}_K^x, \bar{\beta}_p, \sigma_1^2, \ldots, \sigma_K^2, \sigma_p^2)'.$$

The market share for product j is computed assuming 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

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

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

$$\beta_{ikt}^x = \bar{\beta}_k^x + \sigma_k z_{ikt} \quad \text{and} \quad \beta_{it}^p = \bar{\beta}^p + \sigma_p z_{ipt}.$$

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.

The values of $\boldsymbol{X}_t$, $\boldsymbol{p}_t$, $\boldsymbol{\xi}_t$ and the $D \times 1$ vector of instruments $\boldsymbol{Z}_{jt}$ are generated 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 $N(0, \sigma_\xi^2)$ random variables drawn independently across products and markets for different values of σ_ξ^2 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 $d = 1, 2, \ldots, D$.

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

The MPEC approach solves for θ by minimizing

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

subject to

$$s(\eta, \theta) = S$$

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 η is a $(J * T) \times 1$ vector defined by the following equation:

$$s_{jt}(\eta, \theta) = \frac{1}{N_s} \sum_{i=1}^{N_s} \frac{exp(\theta_1 + x_{1j}\beta^x_{1i} + x_{2j}\beta^x_{2i} + x_{3j}\beta^x_{3i} + p_{jt}\beta^p_i + \eta_{jt})}{1 + \sum_{k=1}^{J} exp(\beta^0 + x_{1k}\beta^x_{1i} + x_{2k}\beta^x_{2i} + x_{3k}\beta^x_{3i} + p_{kt}\beta^p_i + \eta_{kt})}$$

where each (β^x_i, β^p_i) 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 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.

7.2 A Control Function Approach

To implement the Petrin and Train (2010) control function approach, we first run a linear regression of the price p on all 42 instruments. We denote the residuals from this regression by $\hat{\varepsilon}_{jt}$.

We then solve the following maximum likelihood problem:

$$\max_{\theta, \rho} \sum_{j=0}^{J} \sum_{t=1}^{T} S_{jt} \cdot log(s'_{jt}(\theta, \rho))$$

where $j = 0$ refers to the outside product, S_{jt} is the observed market share, and s'_{jt} is defined by

$$s'_{jt}(\theta, \rho) = \frac{1}{NS} \sum_{i=1}^{NS} \frac{\exp(\beta^0 + x_{1j}\beta^x_{1i} + x_{2j}\beta^x_{2i} + x_{3j}\beta^x_{3i} + p_{jt}\beta^p_i + \rho\hat{\varepsilon}_{jt})}{1 + \sum_{k=1}^{J} \exp(\beta^0 + x_{1k}\beta^x_{1i} + x_{2k}\beta^x_{2i} + x_{3k}\beta^x_{3i} + p_{kt}\beta^p_i + \rho\hat{\varepsilon}_{kt})}$$

where β_0 and the (β^x_i, β^p_i) are generated as we did with MPEC.

7.3 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 and σ_p cannot be negative. First, we construct the instrumental variables as the MPEC approach. We then construct the artificial regressors K^1 to K^4 of Theorem 2 for each product in each market by applying

$$\bar{X}_{it} = \sum_{k=1}^{J} x_{ik} S_{kt}, i = 1, 2, 3, 4$$

$$K_i^{jt} = x_{ij}(x_{ij}/2 - \bar{X}_{it}), i = 1, 2, 3, 4$$

Note that x_4 in the above notation is the price of the product.

The next step performs an instrumental variable regression of $\log(\frac{S_{jt}}{S_{0t}})$ on 1, x_1, x_2, x_3, x_4, D_1, D_2, D_3, D_4 using all 42 instruments. If any coefficients for the last four variables is negative, we set that coefficient to 0 and rerun the regression without that variable. Keep this process until all the coefficients are positive, or all four variables are excluded from the instrumental variables regression.

7.4 Monte Carlo Simulation Results

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

Here we run the simulations for 9 scenarios obtained by varying the variance of the product effects: $\sigma^2_\xi = Var(\xi) = 0.1, 0.5, 1$ and the vector of variances of the coefficients $\beta_i = (\beta^0, \beta^x_{1i}, \beta^x_{2i}, \beta^x_{3i}, \beta^p_i)'$:

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

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

All the other parameter specifications are as described above. We group the results for the same parameters together. The results are in Tables 1–16.

7.5 Pseudo True Value for 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 value to which our 2SLS estimator converges by simulating its probability limit. The 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.

7.5.1 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, each simulation with 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. The results are in table 10. Estimates are calculated by the sample mean of the 20 simulations. Standard errors are calculated by the sample standard errors of the 20 simulations.

7.5.2 Moment-based Approach

Here we calculate the pseudo true value 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

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

With this method, we only have the estimates but cannot get the standard errors. The estimates under the 9 scenarios are in table 11. There are 1000 simulations and 10,000 markets in each simulation.

7.5.3 Computation Time Comparison

We compared the two MPEC approaches, one starting from the true value, one starting from the results of our 2SLS regression. For the MPEC starting from the true value, the average number of iterations is 1027.1. The average CPU time is 111.5 seconds to complete the estimation. For the MPEC starting from the 2SLS regression result, the average number of iterations is 1280.8. The average CPU time is 125.3 seconds. Both approaches have a 100% success rate. As for the estimates, most of them are very close: the difference is between 10^{-6} and 10^{-7}.

These results are encouraging for the use of our approach as a method for finding "approximately consistent" starting values for the MPEC and nested-fixed point estimation procedures. Given the difficulty in finding plausible starting values for these nonlinear estimation procedures, our 2SLS approach can at least serve this role, because our estimator simply involves a linear 2SLS estimation.

7.5.4 Impact of Price Elasticity

We performed a set of simulations to determine if changing the value of the price coefficient, $E(\beta^p)$ changes the distribution of the estimators. The results are in table 12.

7.6 Lognormal Distribution for β

Here instead of using a normal distribution for the consumer preference parameter, we assume a lognormal distribution for the consumer preference parameter β_i.

$$\beta_i = \bar{\beta}_i \epsilon_i$$

$$\bar{\beta}_i = [1, 1.5, 1.5, 0.5, 1]$$

$$\ln(\epsilon_i) \sim N(-0.5\sigma^2, \sigma^2)$$

We study several cases, with $\sigma = 0.3, 0.4, 0.5$ and $\xi_{jt} \sim N(0, \sqrt{0.1})$. The rest of the specification is as before.

Besides the 2SLS approach for second moment, we also introduce the third moment estimates here. Tables 13-15 plot the distributions of the estimators of the first, second, and third moments of $\beta_1 - \beta_p$. Each plot shows the distributions when we consider only the first two moments and when we consider the first three moments. Table 16 provides the corresponding summary statistics.

7.6.1 Conclusions from Monte Carlo Analysis

A number of conclusions emerge from the Monte Carlo analysis. Most are expected, but others point to directions for future research. The results presented in Tables 1 to 9 are consistent with the conclusion 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. In constrast, the control function approach appears to lead to substantial bias in the estimate of the means of the random coefficients and this bias is larger, the larger is the variance of ξ_{jt}.

The MPEC approach appears to dominate the 2SLS approach for the variance of the random coefficients. The 2SLS approach seems to be downward biased and this downward bias appears to be larger, the larger is 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. In general, the control function estimator of the variance of the random coefficients is significantly less biased than the control function estimate of the mean of the random coefficients. However, distributions of control function estimate of the variance of the random coefficients general have a larger spread than the distributions of these coefficient estimates from the MPEC estimation procedure or our 2SLS estimation procedure.

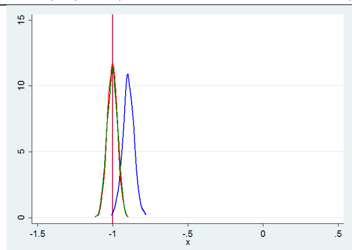
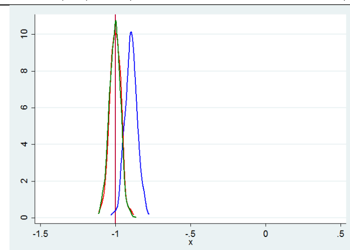
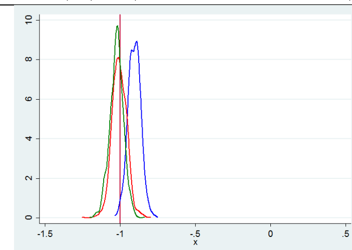
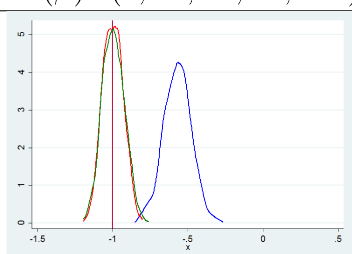
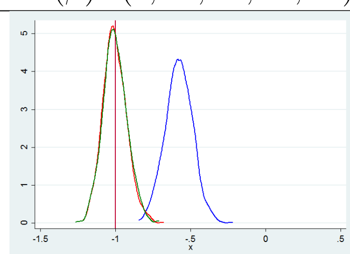
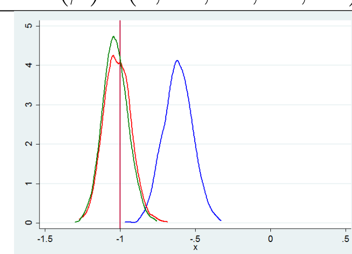
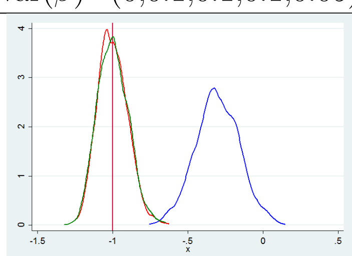
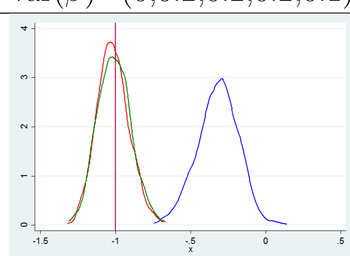
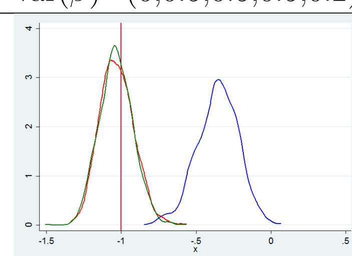
Tables 10 and 11 demonstrate that the pseudo true values from applying our 2SLS procedure are typically not substantially different from the true values for the data generation process. Based on these results it is difficult to argue that a researcher would draw economically or even statistically meaningfully different conclusions from parameters estimates obtained from our 2SLS approach relative to the MPEC approach.

Table 13 reinforces our previous conclusion about our 2SLS approach. For a range of values for the expected value of the price coefficient, our approach introduces minimal bias in the estimates of the mean of the random coefficients. In constrast, the control function approach continues to lead to significant 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 downwarded biased in general, but there are combinations of the variance of ξ_{jt} and the variance of the random coefficients that reduces the magnitude of these biases.

Table 13 compare the performance of our estimator assuming lognormally distributed random coefficients. This induces significant skewness and kurtosis into the distribution of the random coefficients. We compare the performance of our 2SLS estimator that only relies on the first two moments of the random coefficients to a 2SLS estimator that uses information for the third moment of our random coefficients. We find that for a variety of values for our one parameter lognormal distribution, the additional information in the third moment of the random coefficients does not appreciably increase the precision in our estimates of the mean and variance of the random coefficients. In fact, for some of the coefficients, we find that our procedure that uses third moment information leads to significantly less efficient estimates of the both the mean and variance of the random coefficients. This is likely due to the fact that our procedure has a difficult time precisely estimating the third moment of the random coefficients, as shown in Table 15.

List of Tables

Table 1: Estimator Distribution for E(β_1)

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)
		

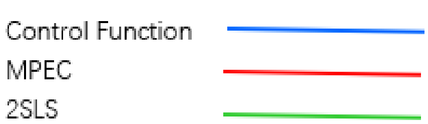
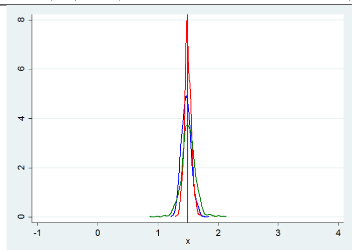
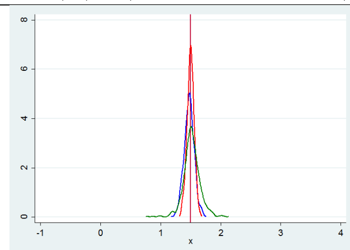
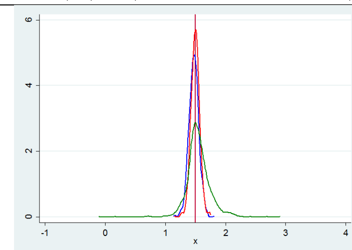
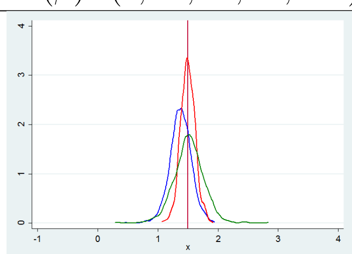
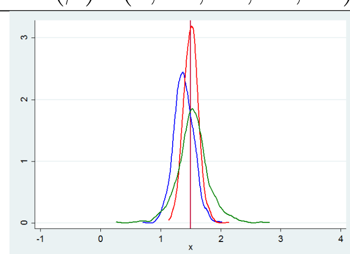
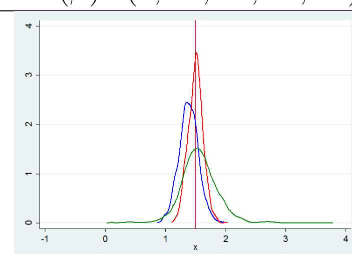
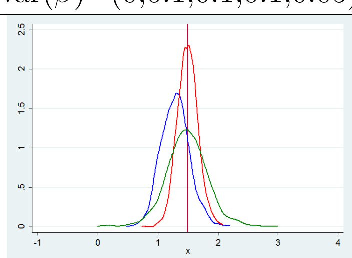
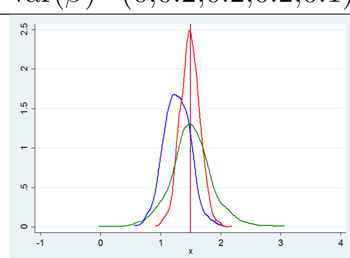
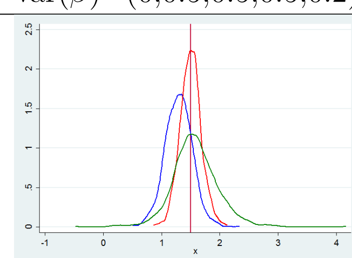


Table 2: Estimator Distribution for $E(\beta_2)$

$var(\xi)$=0.1	$var(\xi)$=0.1	$var(\xi)$=0.1
$var(\beta)$=(0,0.1,0.1,0.1,0.05)	$var(\beta)$=(0,0.2,0.2,0.2,0.1)	$var(\beta)$=(0,0.5,0.5,0.5,0.2)
		
$var(\xi)$=0.5	$var(\xi)$=0.5	$var(\xi)$=0.5
$var(\beta)$=(0,0.1,0.1,0.1,0.05)	$var(\beta)$=(0,0.2,0.2,0.2,0.1)	$var(\beta)$=(0,0.5,0.5,0.5,0.2)
		
$var(\xi)$=1.0	$var(\xi)$=1.0	$var(\xi)$=1.0
$var(\beta)$=(0,0.1,0.1,0.1,0.05)	$var(\beta)$=(0,0.2,0.2,0.2,0.1)	$var(\beta)$=(0,0.5,0.5,0.5,0.2)
		

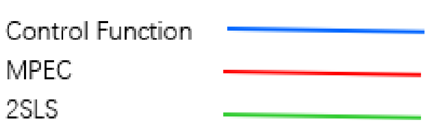
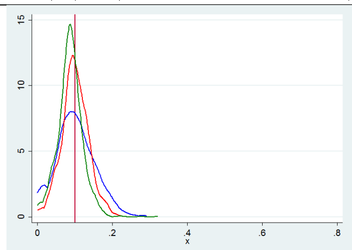
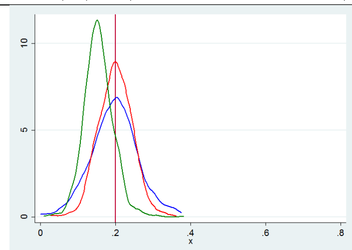
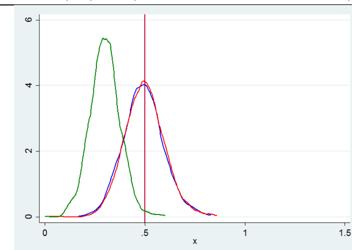
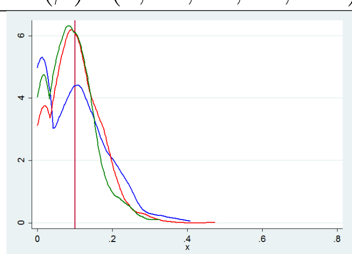
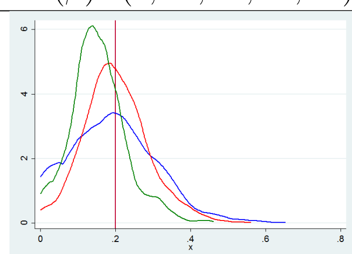
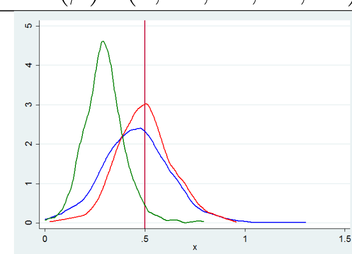
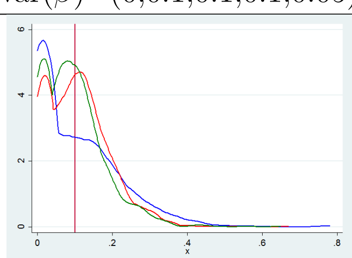
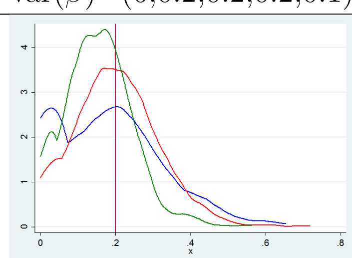
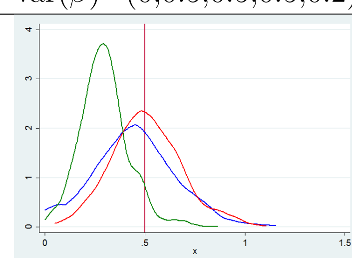


Table 3: Estimator Distribution for $var(\beta_2)$

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)
		

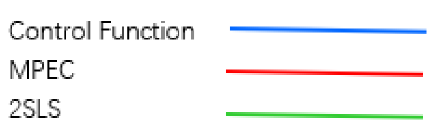
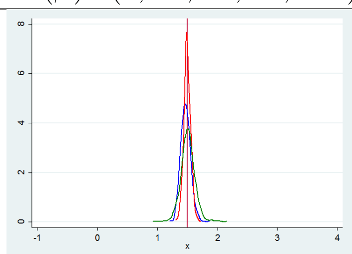
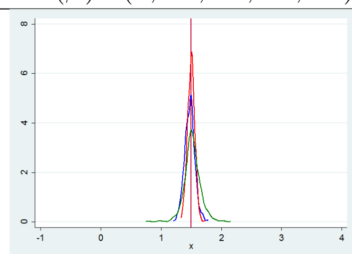
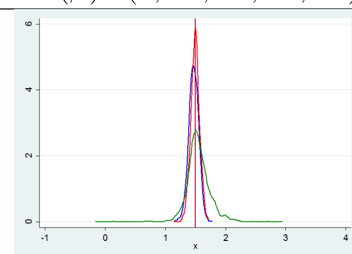
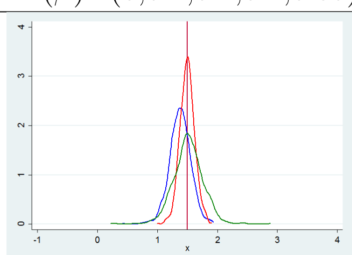
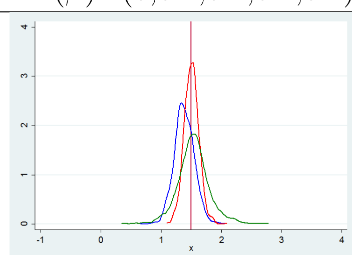
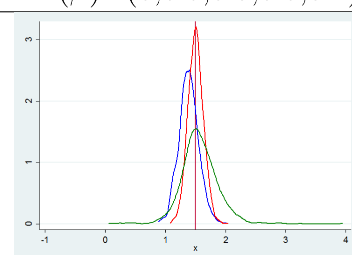
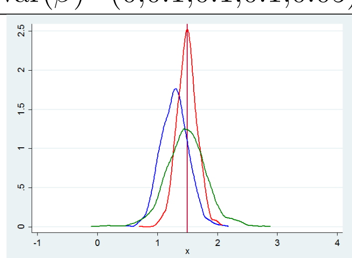
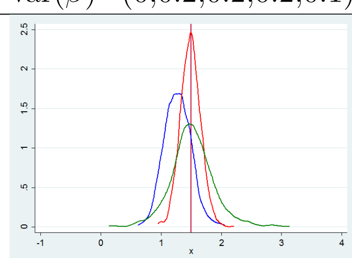
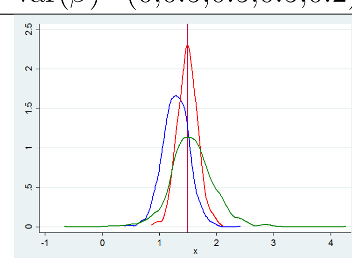


Table 4: Estimator Distribution for $E(\beta_3)$

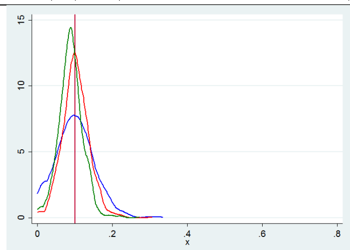
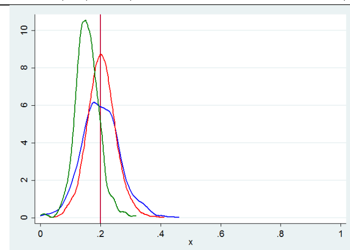
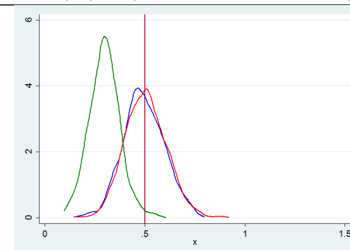
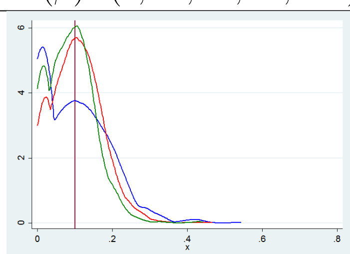
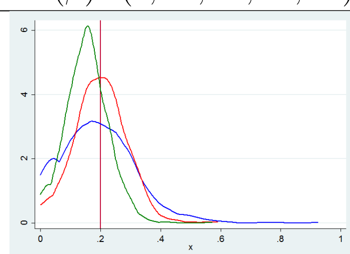
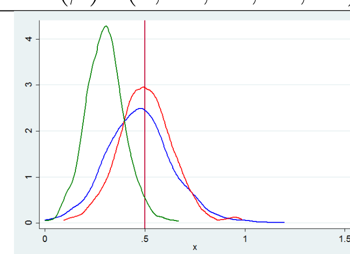
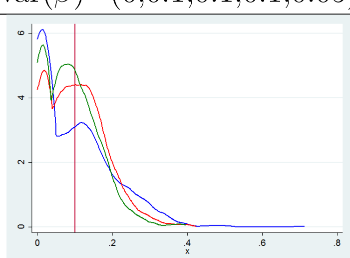
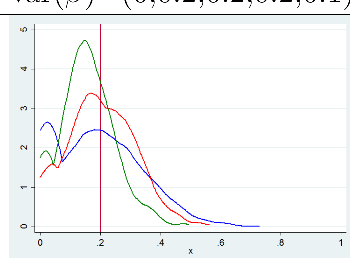
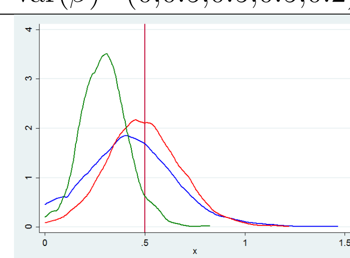
var(ξ)=0.1	var(ξ)=0.1	var(ξ)=0.1
var(β)=(0,0.1,0.1,0.1,0.05)	var(β)=(0,0.2,0.2,0.2,0.1)	var(β)=(0,0.5,0.5,0.5,0.2)
		
var(ξ)=0.5	var(ξ)=0.5	var(ξ)=0.5
var(β)=(0,0.1,0.1,0.1,0.05)	var(β)=(0,0.2,0.2,0.2,0.1)	var(β)=(0,0.5,0.5,0.5,0.2)
		
var(ξ)=1.0	var(ξ)=1.0	var(ξ)=1.0
var(β)=(0,0.1,0.1,0.1,0.05)	var(β)=(0,0.2,0.2,0.2,0.1)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

Control Function

MPEC

2SLS

Table 5: Estimator Distribution for $var(\beta_3)$

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)
		

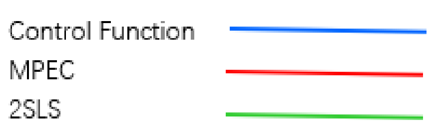
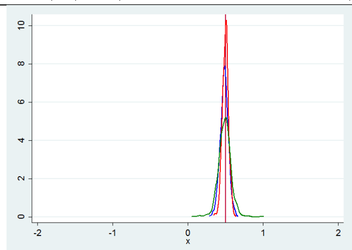
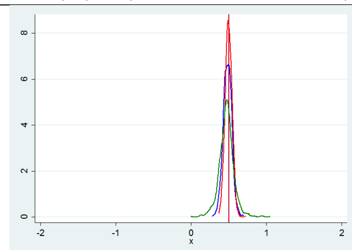
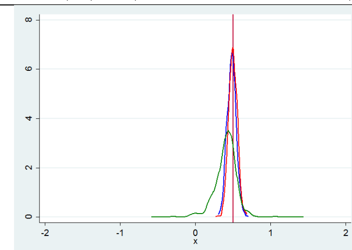
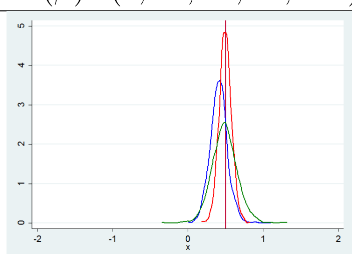
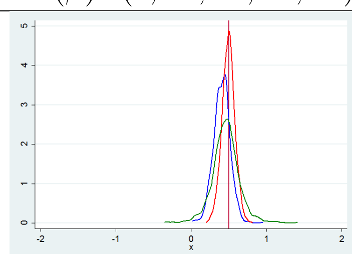
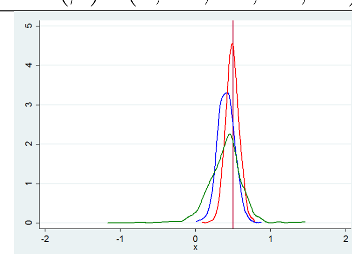
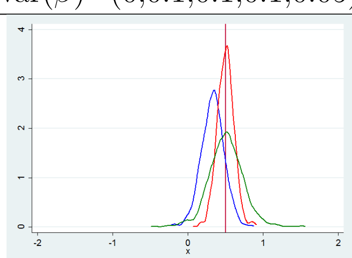
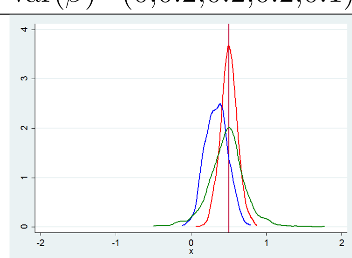
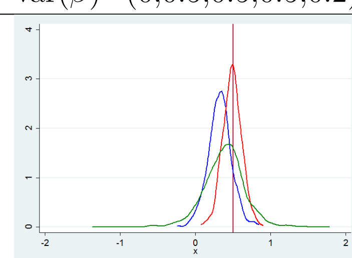


Table 6: Estimator Distribution for $E(\beta_4)$

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)
		

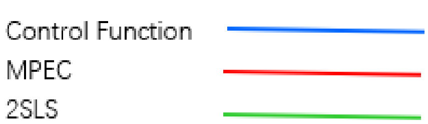
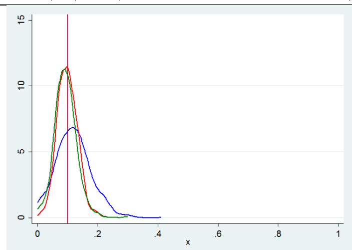
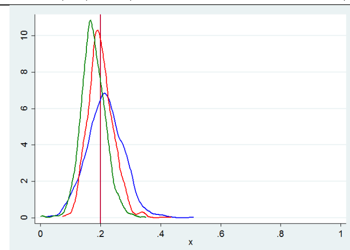
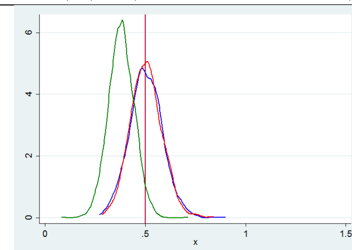
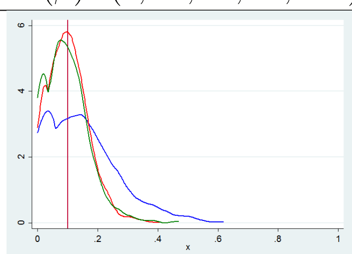
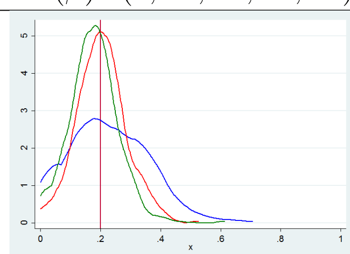
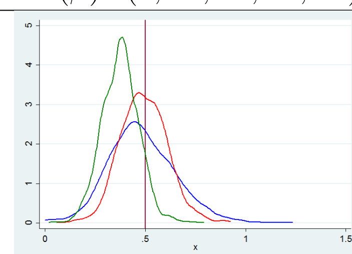
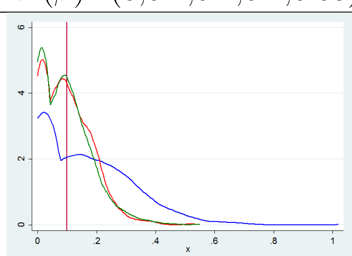
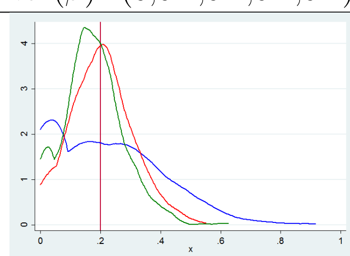
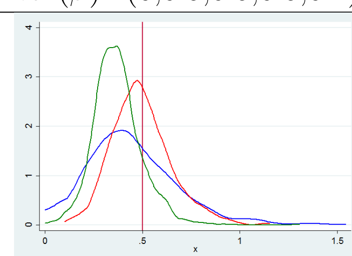


Table 7: Estimator Distribution for $var(\beta_4)$

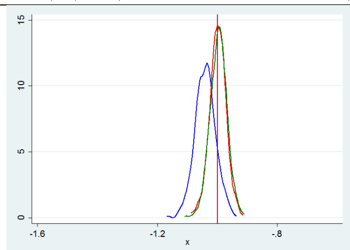
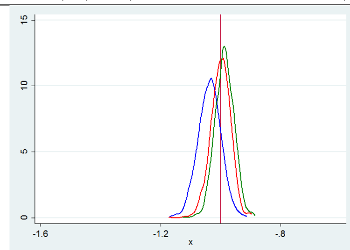
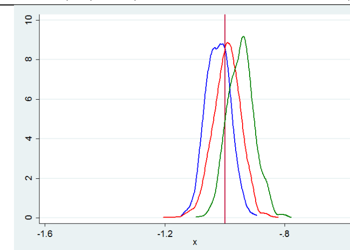
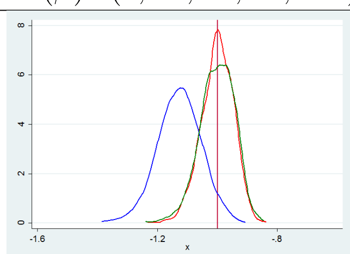
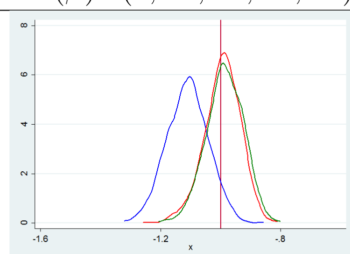
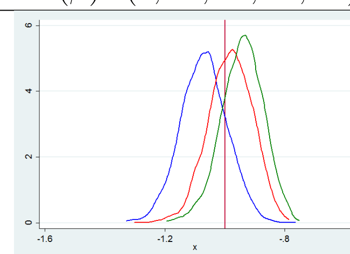
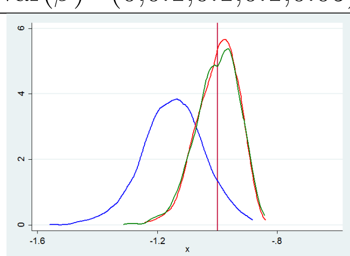
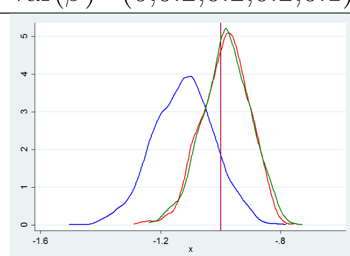
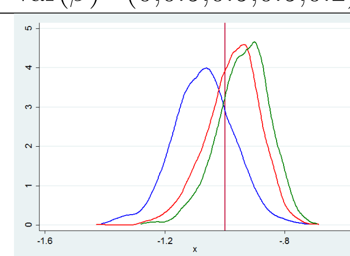
var(ξ)=0.1	var(ξ)=0.1	var(ξ)=0.1
var(β)=(0,0.1,0.1,0.1,0.05)	var(β)=(0,0.2,0.2,0.2,0.1)	var(β)=(0,0.5,0.5,0.5,0.2)
		
var(ξ)=0.5	var(ξ)=0.5	var(ξ)=0.5
var(β)=(0,0.1,0.1,0.1,0.05)	var(β)=(0,0.2,0.2,0.2,0.1)	var(β)=(0,0.5,0.5,0.5,0.2)
		
var(ξ)=1.0	var(ξ)=1.0	var(ξ)=1.0
var(β)=(0,0.1,0.1,0.1,0.05)	var(β)=(0,0.2,0.2,0.2,0.1)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

Control Function

MPEC

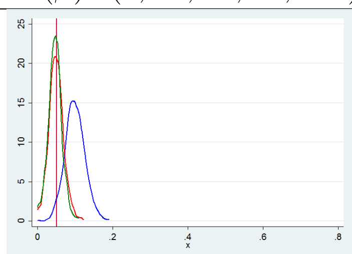
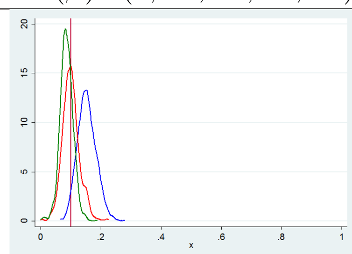
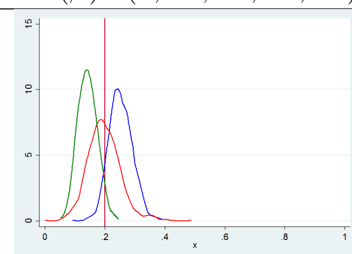
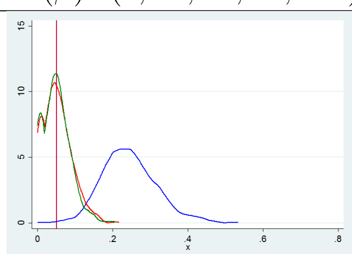
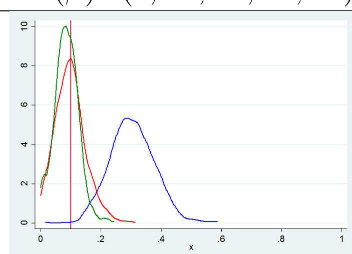
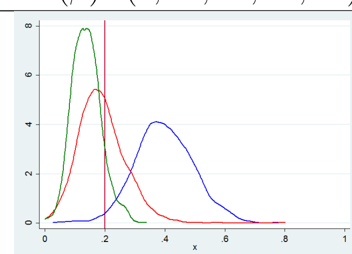
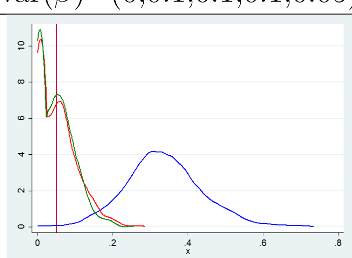
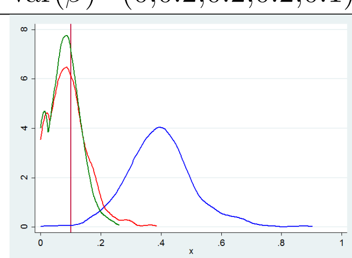
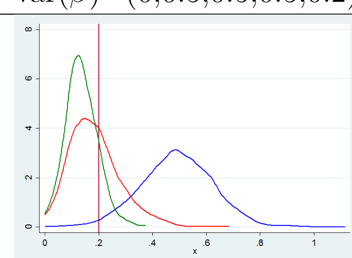
2SLS

Table 8: Estimator Distribution for $E(\beta_5)$

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: Estimator Distribution for $var(\beta_5)$

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)
		

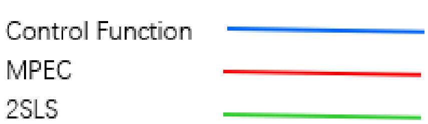


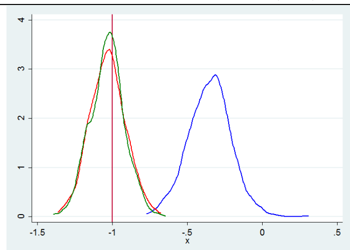
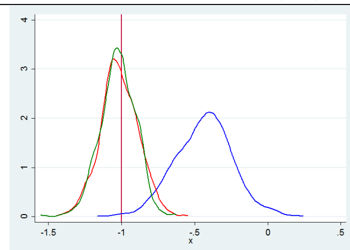
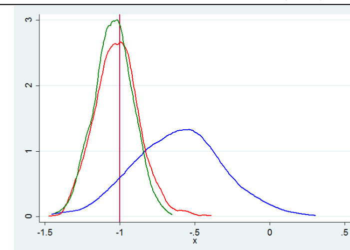
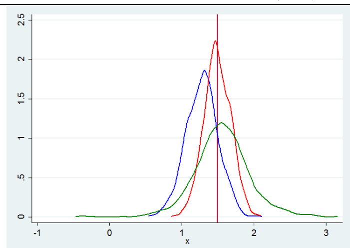
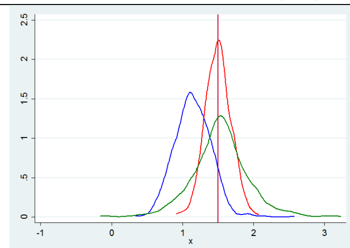
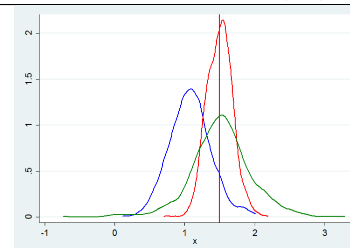
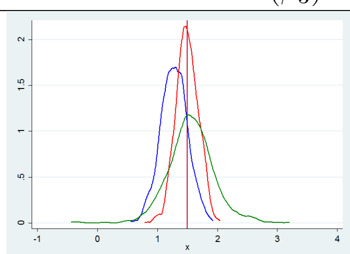
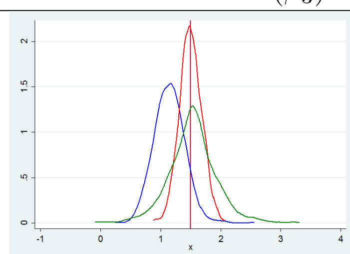
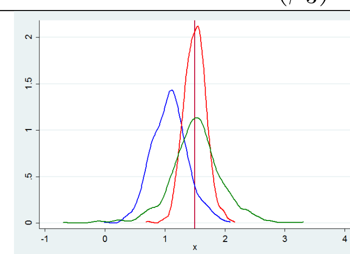
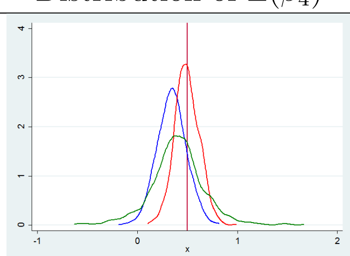
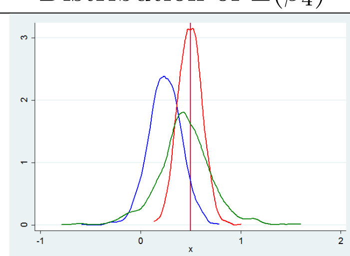
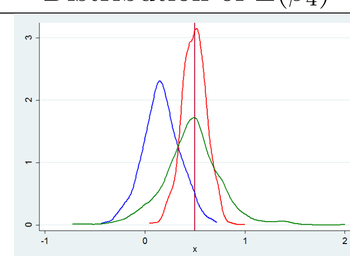
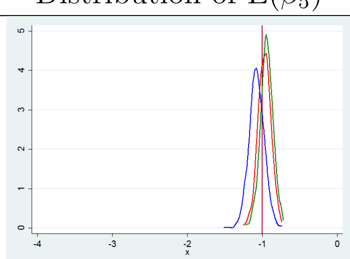
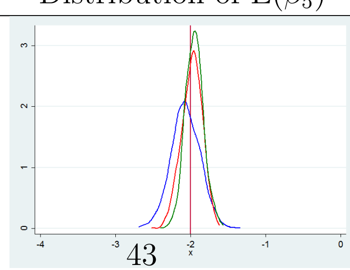
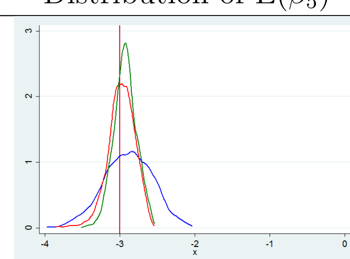
Table 10: Increasing-number-of-market Approach Pseudo True Value

Parameter	Scenarios								
True $var(\beta)$:	$(0, 0.1, 0.1, 0.1, 0.05)$			$(0, 0.2, 0.2, 0.2, 0.1)$			$(0, 0.5, 0.5, 0.5, 0.2)$		
True $var(\xi)$:	0.1	0.5	1	0.1	0.5	1	0.1	0.5	1
$E(\beta_1) = -1$	-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)
$E(\beta_2) = 1.5$	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)
$E(\beta_3) = 1.5$	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)
$E(\beta_4) = 0.5$	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)
$E(\beta_5) = -1$	-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)
$var(\beta_2)$	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)
$var(\beta_3)$	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)
$var(\beta_4)$	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)
$var(\beta_5)$	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: Moment-based Approach Pseudo True Value

Parameter	Scenarios								
True $var(\beta)$:	$(0, 0.1, 0.1, 0.1, 0.05)$			$(0, 0.2, 0.2, 0.2, 0.1)$			$(0, 0.5, 0.5, 0.5, 0.2)$		
True $var(\xi)$:	0.1	0.5	1	0.1	0.5	1	0.1	0.5	1
$E(\beta_1) = -1$	-1.01	-1.01	-1.01	-1.04	-1.04	-1.04	-1.11	-1.11	-1.12
$E(\beta_2) = 1.5$	1.49	1.49	1.49	1.48	1.48	1.48	1.43	1.43	1.43
$E(\beta_3) = 1.5$	1.49	1.49	1.49	1.48	1.48	1.48	1.43	1.43	1.43
$E(\beta_4) = 0.5$	0.496	0.496	0.496	0.486	0.486	0.486	0.455	0.455	0.455
$E(\beta_5) = -1$	-0.989	-0.988	-0.988	-0.958	-0.957	-0.955	-0.873	-0.869	-0.864
$var(\beta_2)$	0.0854	0.0855	0.0855	0.149	0.149	0.149	0.275	0.275	0.276
$var(\beta_3)$	0.0855	0.0855	0.0856	0.149	0.149	0.149	0.273	0.274	0.274
$var(\beta_4)$	0.0938	0.0938	0.0937	0.176	0.175	0.175	0.369	0.368	0.366
$var(\beta_5)$	0.0421	0.0421	0.0419	0.0685	0.0681	0.0676	0.0920	0.0906	0.0888

Table 12: First Moment Estimator Distribution with different $E(\beta_5)$

$E(\beta_5) = -1$	$E(\beta_5) = -2$	$E(\beta_5) = -3$
Distribution of E(β_1)	Distribution of E(β_1)	Distribution of E(β_1)
		
Distribution of E(β_2)	Distribution of E(β_2)	Distribution of E(β_2)
		
Distribution of E(β_3)	Distribution of E(β_3)	Distribution of E(β_3)
		
Distribution of E(β_4)	Distribution of E(β_4)	Distribution of E(β_4)
		
Distribution of E(β_5)	Distribution of E(β_5)	Distribution of E(β_5)
		

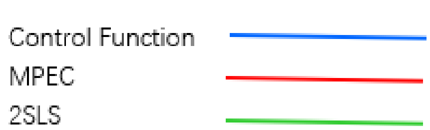
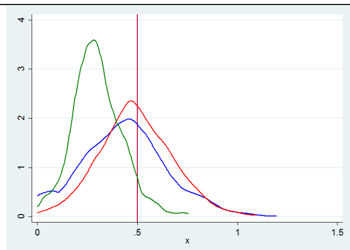
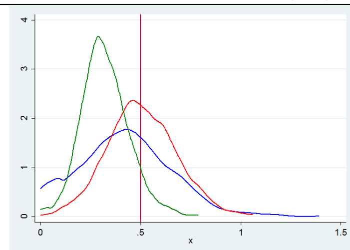
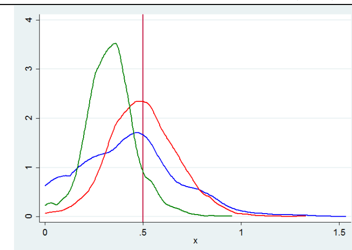
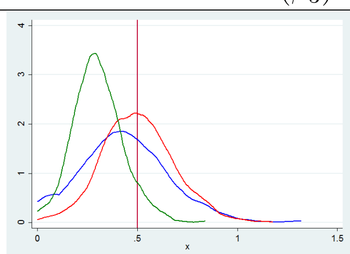
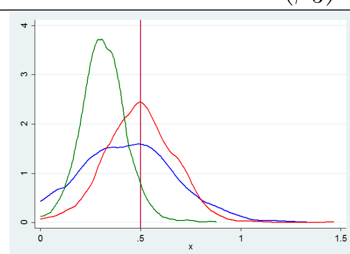
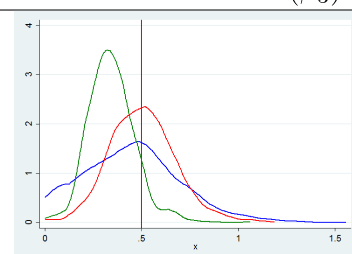
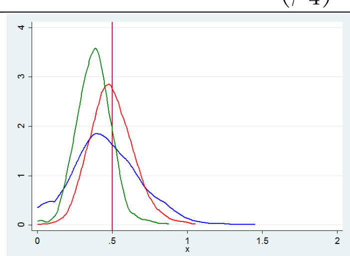
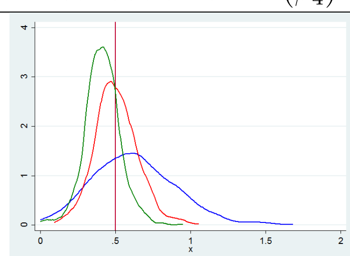
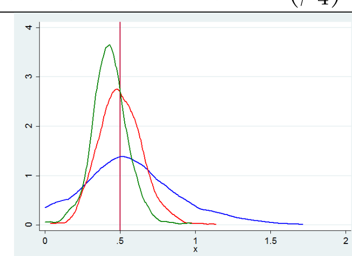
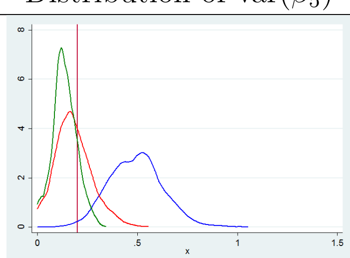
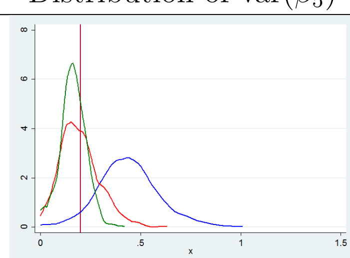
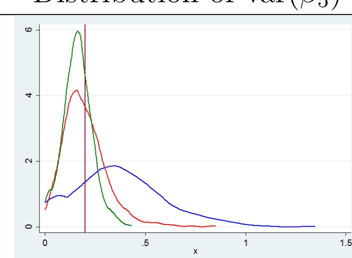


Table 13: Second Moment Estimator Distribution with different $E(\beta_5)$

$E(\beta_5) = -1$	$E(\beta_5) = -2$	$E(\beta_5) = -3$
Distribution of var(β_2)	Distribution of var(β_2)	Distribution of var(β_2)
		
Distribution of var(β_3)	Distribution of var(β_3)	Distribution of var(β_3)
		
Distribution of var(β_4)	Distribution of var(β_4)	Distribution of var(β_4)
		
Distribution of var(β_5)	Distribution of var(β_5)	Distribution of var(β_5)
		

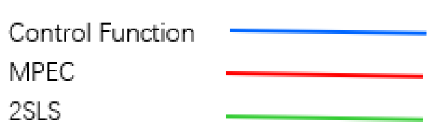
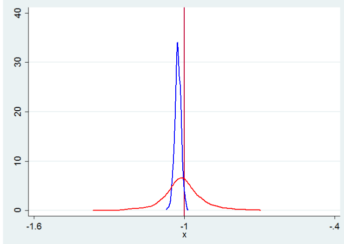
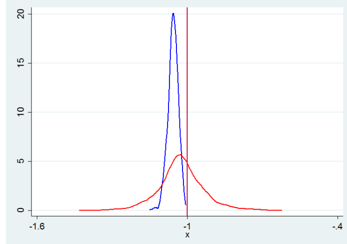
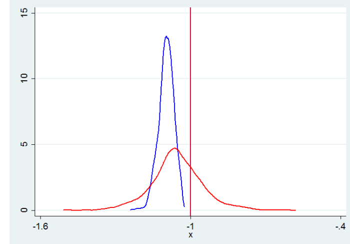
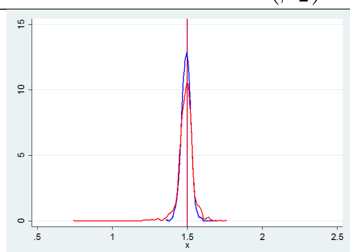
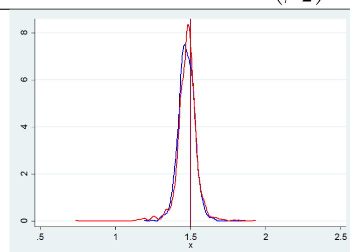
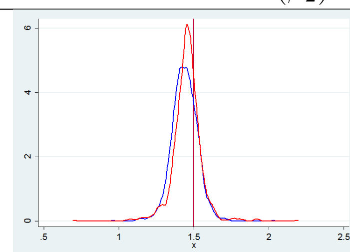
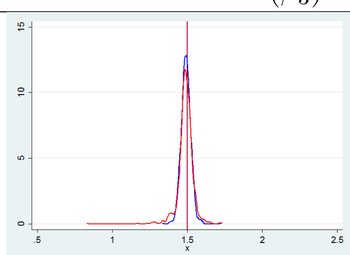
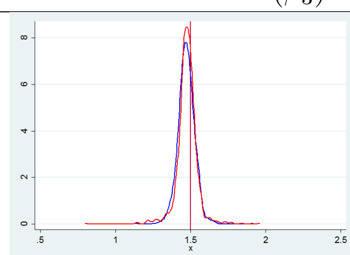
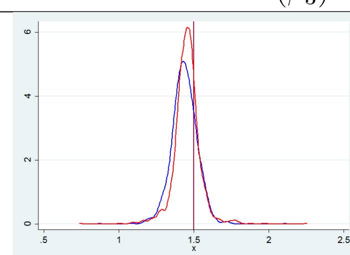
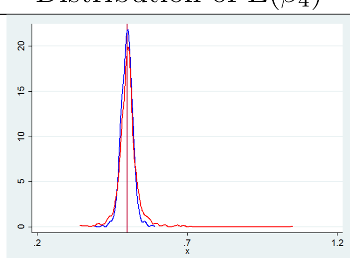
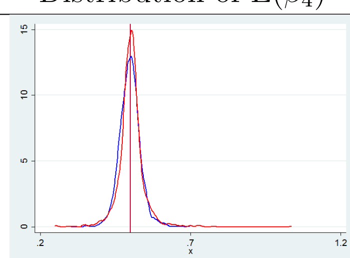
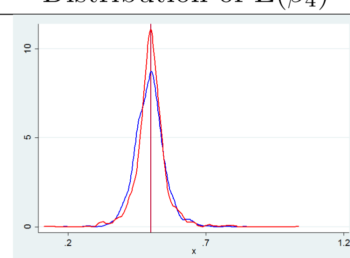
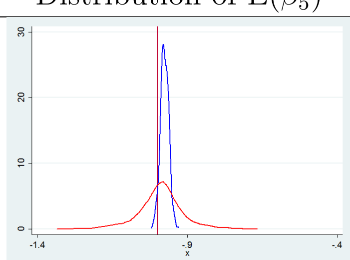
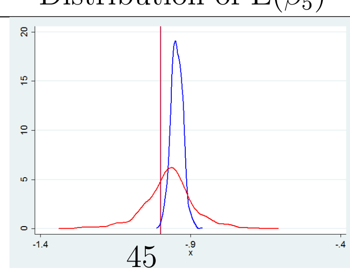
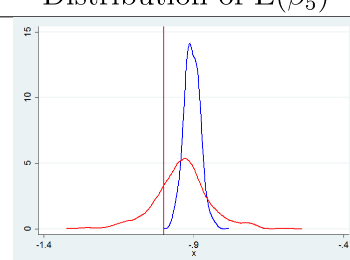


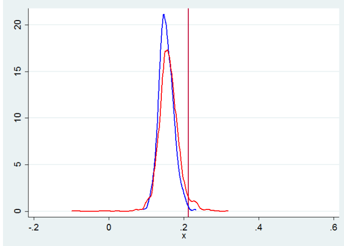
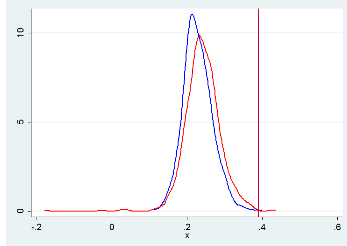
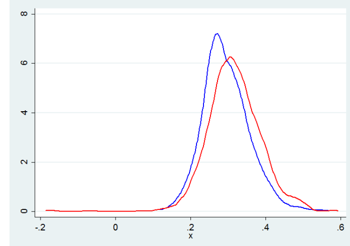
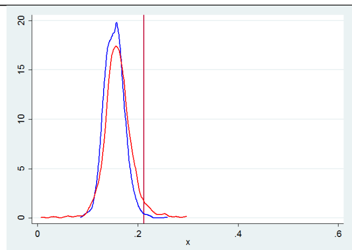
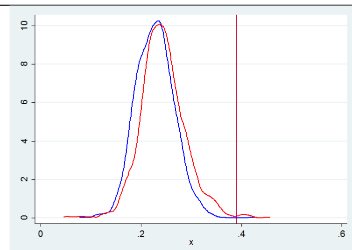
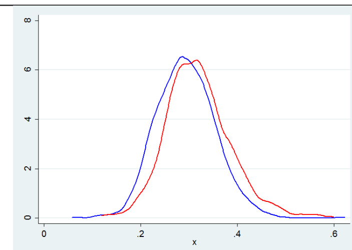
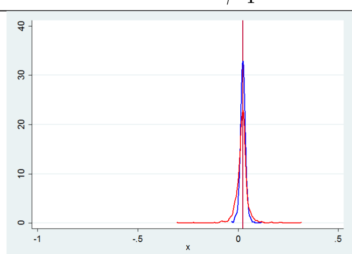
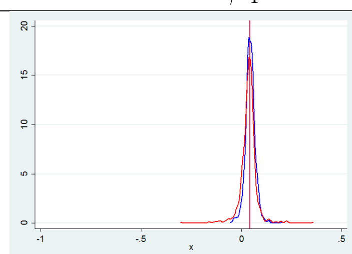
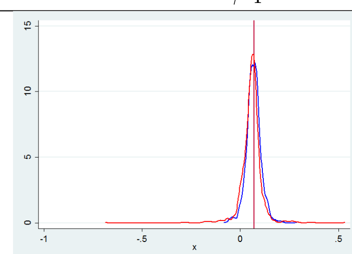
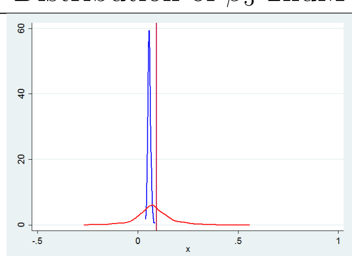
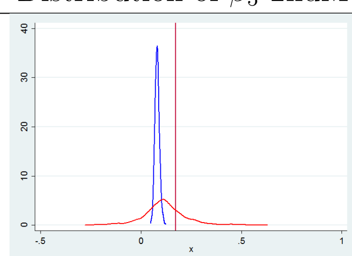
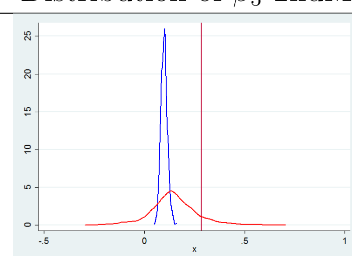
Table 14: First Moment Estimator Distribution for log normal case

$\sigma = 0.3$	$\sigma = 0.4$	$\sigma = 0.5$
Distribution of E(β_1)	Distribution of E(β_1)	Distribution of E(β_1)
		
Distribution of E(β_2)	Distribution of E(β_2)	Distribution of E(β_2)
		
Distribution of E(β_3)	Distribution of E(β_3)	Distribution of E(β_3)
		
Distribution of E(β_4)	Distribution of E(β_4)	Distribution of E(β_4)
		
Distribution of E(β_5)	Distribution of E(β_5)	Distribution of E(β_5)
		

Only include 2nd moment

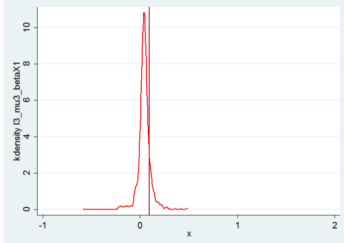
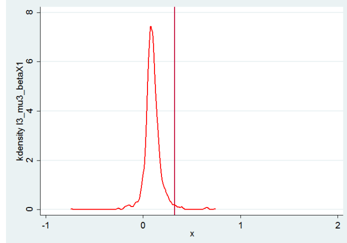
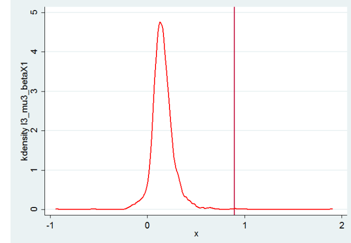
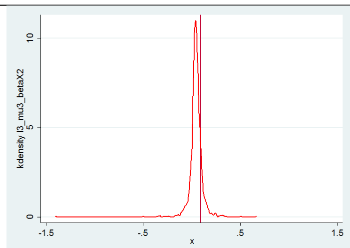
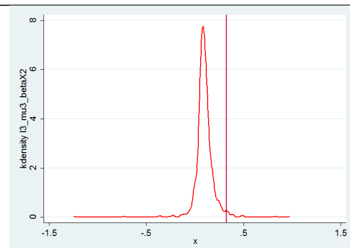
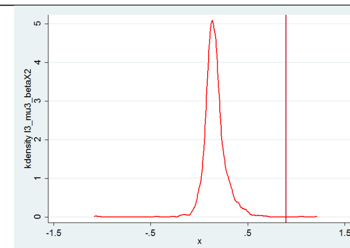
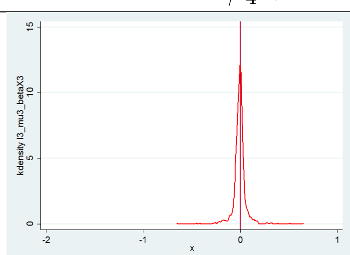
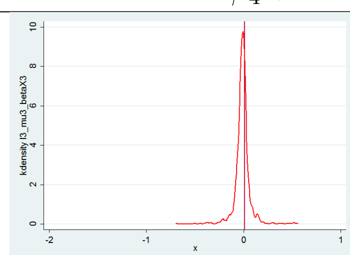
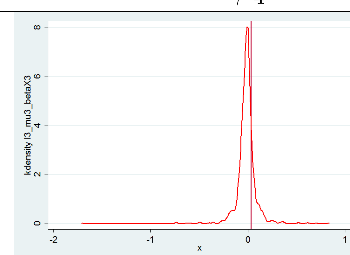
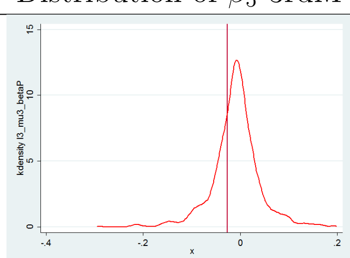
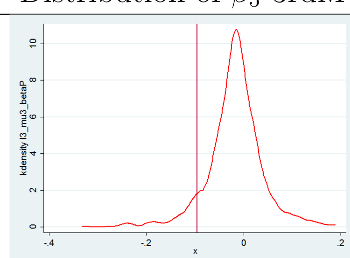
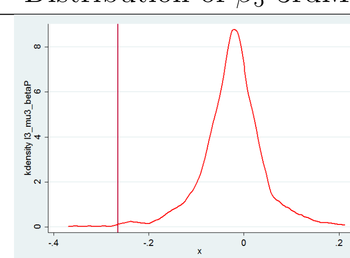
Include 3rd moment

Table 15: Second Moment Estimator Distribution for log normal case

$\sigma = 0.3$	$\sigma = 0.4$	$\sigma = 0.5$
Distribution of β_2 2ndM	Distribution of β_2 2ndM	Distribution of β_2 2ndM
		
Distribution of β_3 2ndM	Distribution of β_3 2ndM	Distribution of β_3 2ndM
		
Distribution of β_4 2ndM	Distribution of β_4 2ndM	Distribution of β_4 2ndM
		
Distribution of β_5 2ndM	Distribution of β_5 2ndM	Distribution of β_5 2ndM
		

Only include 2nd moment (blue line)
Include 3rd moment (red line)

Table 16: Third Moment Estimator Distribution for log normal case

$\sigma = 0.3$	$\sigma = 0.4$	$\sigma = 0.5$
Distribution of β_2 3rdM	Distribution of β_2 3rdM	Distribution of β_2 3rdM
		
Distribution of β_3 3rdM	Distribution of β_3 3rdM	Distribution of β_3 3rdM
		
Distribution of β_4 3rdM	Distribution of β_4 3rdM	Distribution of β_4 3rdM
		
Distribution of β_5 3rdM	Distribution of β_5 3rdM	Distribution of β_5 3rdM
		

Only include 2nd moment

Include 3rd moment

Table 17: Summary Statistics for log normal case

Parameter	Scenarios					
σ	$\sigma = 0.3$		$\sigma = 0.4$		$\sigma = 0.5$	
Moments included:	2	3	2	3	2	3
$E(\beta_1) = -1$	-1.03	-1.01	-1.06	-1.03	-1.10	-1.
	(0.0126)	(0.0801)	(0.0208)	(0.0915)	(0.0310)	(0.1
$E(\beta_2) = 1.5$	1.49	1.49	1.47	1.48	1.44	1.4
	(0.0321)	(0.0568)	(0.0557)	(0.0723)	(0.0861)	(0.09
$E(\beta_3) = 1.5$	1.49	1.49	1.47	1.47	1.44	1.4
	(0.0331)	(0.0540)	(0.0572)	(0.0698)	(0.0879)	(0.09
$E(\beta_4) = 0.5$	0.499	0.502	0.497	0.502	0.496	0.5
	(0.0201)	(0.0385)	(0.0358)	(0.0471)	(0.0560)	(0.05
$E(\beta_5) = -1$	-0.976	-0.991	-0.946	-0.972	-0.906	-0.9
	(0.0138)	(0.0761)	(0.0201)	(0.0844)	(0.0273)	(0.09
$var(\beta_2) = 0.212/0.390/0.639$	0.153	0.159	0.229	0.241	0.295	0.3
	(0.0201)	(0.0306)	(0.0390)	(0.0476)	(0.0620)	(0.07
$var(\beta_3) = 0.212/0.390/0.639$	0.153	0.160	0.228	0.244	0.294	0.3
	(0.0200)	(0.0280)	(0.0382)	(0.0448)	(0.0604)	(0.06
$var(\beta_4) = 0.0235/0.0434/0.0710$	0.0248	0.0197	0.0446	0.0357	0.0681	0.05
	(0.0141)	(0.0327)	(0.0248)	(0.0406)	(0.0376)	(0.05
$var(\beta_5) = 0.0942/0.174/0.284$	0.0579	0.0768	0.0811	0.115	0.0994	0.1
	(0.00707)	(0.0951)	(0.0112)	(0.106)	(0.0162)	(0.1
$3rdM(\beta_2) = 0.0926/0.322/0.894$		0.0437		0.0961		0.1
		(0.0672)		(0.0908)		(0.1
$3rdM(\beta_3) = 0.0926/0.322/0.894$		0.0433		0.0973		0.1
		(0.0823)		(0.101)		(0.1
$3rdM(\beta_4) = 0.00343/0.0119/0.0331$		-0.00729		-0.0120		-0.0
		(0.0699)		(0.0838)		(0.1
$3rdM(\beta_5) = -0.0274/-0.0955/-0.265$		-0.00950		-0.0177		-0.0
		(0.0499)		(0.0576)		(0.06