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

Bernard Salanié   Frank Wolak

UCL
8 May 2018

The Goal

Help simplify estimation of a class of models that integrate over unobserved heterogeneity

including the standard models of empirical IO that only use information on market shares:

macro BLP.

Several questions:

1. How much information is there really in the data? (practical identification)
2. Can we diagnose/anticipate problems and alleviate them? (specification)
3. Are there simpler ways than GMM or MLE to estimate the parameters? (estimation)

With FRAC

The answers are **Yes, yes, and yes**

We use approximate models, leading to

- Fast 2SLS estimates of the parameters
- that are Approximately Correct
- and (approximately) Robust to misspecification of higher moments
- and provide simple diagnoses of underdentification.

The Starting Point

Start with a structural parametric model $G(y, \eta, \theta_0)$ (omitting covariates)
with a (unique) inverse $\eta = F(y, \theta_0)$

and we assume moment conditions $E(\eta Z) = 0$.

Usually estimated by GMM, minimizing

$$\left\| \sum_i F(y_i, \theta) Z_i \right\|_{\hat{W}}.$$

Often tricky: model overspecified, badly identified, numerical difficulties...

The Idea

If the underlying model integrates over unobserved heterogeneity with unknown parameters s_0, split

$$\theta_0 = (\beta_0, s_0)$$

and take Taylor expansions around $s = 0$ for fixed β: small-σ analysis

stop at a reasonable order and estimate the resulting (hopefully) simple approximate model.

Cf Kadane 1971, Chesher 1991, and especially Chesher and Santos–Silva 2002 (MLE in mixed multinomial logit with exogenous covariates).

Empirical IO: the standard model

Since Berry–Levinsohn–Pakes 1995: demand and loosely specified supply

- demand = mixed multinomial logit:
  the classic demand side in many empirical investigations (IO, transport, demand systems . . .)
  circumvents well-known limitations of unmixed logit
- (typically) aggregate version: we observe choice probabilities for groups of consumers (markets)
- supply: product effects are orthogonal to well-chosen instruments.

Gives a GMM estimator.

Plus Notation

Utility of variety $j = 1, \ldots, J$ for consumer i in market $t = 1, \ldots, T$ is

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

with

- $\boldsymbol{u}_i$ a vector of iid standard type I EV (parameter-free)
- $\boldsymbol{\epsilon}_i$ iid across consumers, distribution known up to parameters Σ_0.

ξ_t is a vector of product effects that shift the demand of all consumers in market t,
and we assume

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

We observe the market shares

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

Macro-BLP and our General Framework

Define $y_{jt} = \log(S_{jt}/S_{0t})$

and artificial regressors (m, n index components of the covariate vectors)

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

with $e_{tm} = \sum_{j=1}^{J} X_{jtm}/J$.

Estimate the optimal instruments

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

Approximate Estimation

Run a Fast two-stage least squares regression of $\boldsymbol{y}_{jt}$ on $\boldsymbol{X}_{jt}, \boldsymbol{K}_{jt}$ with instruments $\hat{\boldsymbol{Z}}_{jt}$

The estimators $\hat{\beta}, \hat{\Sigma}$ are Approximately Correct.
More precisely: the error is $O_P(\|\Sigma\|^{3/2})$, and in fact $O_P(\|\Sigma\|^2)$ if the randomness in the coefficients is symmetric.

The 2SLS estimators are also Robust in that they are equally Approximately Correct independently of other features of the distribution of ϵ.

They can also be adapted to different specifications of the idiosyncratic $\boldsymbol{u}$ (e.g. nested logit—then we need NL2SLS.)

When it Works

Suppose the structural form of the model $G(y,\eta,\theta)=0$ is

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

Here ε is the unobserved heterogeneity, with $E\varepsilon = 0$; and y has all observables (or functions of).

E.g. for macro-BLP: $y = \left(S_j, \boldsymbol{X}_j\right))_j$ and $\eta = \xi$ and $s = \Sigma^{1/2}$ gives

$$G_j = S_j - E_{\varepsilon}A_j^{*}(\boldsymbol{X}, \xi + \boldsymbol{X}\beta, s\varepsilon)$$

with

$$A_j^{*}(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}) = \frac{\exp(b_j + c_j)}{1 + \sum_{k=1}^{J}\exp(b_k + c_k)}.$$

Why it Works

With this form, the inverse $\eta = F(y, \beta, s)$ given by
$G(y, F(y, \beta, s), \beta, s) = 0$
has three properties:

1. $F_s(y, \beta, 0) \equiv 0$
2. $F(y, \beta, 0) - f_1(y)\beta$ does not depend on β; call it $f_0(y)$
3. $F_{ss}(y, \beta, 0)$ does not depend on β; call it $-f_2(y)$.

Then $F(y, \beta, s) \simeq f_0(y) - f_1(y)\beta - f_2(y)s^2/2$ and writing
$E(\eta Z) = 0$ gives

$$E(f_0(y)Z) \simeq E(f_1(y)Z)\beta + \frac{E(f_2(y)Z)}{2}s^2$$

nicely linear in (β, s^2).

How it Works in a Given Model

$f_1(y)$ is from the structural form (e.g. it is $\boldsymbol{X}$ in macro BLP)

for $f_0(y)$, need to solve

$$G^*(y, E_\varepsilon A^*(y, f_0(y), 0)) = 0$$

e.g. in macro BLP:

$$S_j = \frac{\exp(f_{0j})}{1 + \sum_{k=1}^{J} \exp(f_{0k})}$$

gives $f_{0j} = \log(S_j/S_0)$

How it Works, 2

The hardest part:

$$f_2(y) \equiv \left(\left(A^*_{33}\right)^{-1} A^*_2\right)(y, f_0(y), 0)$$

It generates the artificial regressors K^j in macro BLP;
in general it depends on the properties of A^* and on f_0

not those of ε; (again, Robustness)
and only via f_0 for G^*.

Optimal approximate instruments

In BLP, we need to compute

$$W = E\left(\frac{\partial \xi}{\partial \theta} | Z\right)$$

which requires a prior estimate of θ, including the distribution of the random coefficients.

Here, at order 2

$$\frac{\partial \xi}{\partial \boldsymbol{\beta}, \boldsymbol{\Sigma}} = (\boldsymbol{X}, \boldsymbol{K})$$

makes it very easy:

$$\hat{\boldsymbol{Z}} = (E(\boldsymbol{X}|\boldsymbol{Z}), E(\boldsymbol{K}|\boldsymbol{Z}))\,.$$

How are the parameters identified?

Much easier to answer in the approximate 2SLS framework, say at order 2:

The identification of (β, Σ) relies on the variance covariance of

$$\begin{pmatrix} E(\boldsymbol{X}|\boldsymbol{Z}) \\ E(\boldsymbol{K}|\boldsymbol{Z}) \end{pmatrix}$$

being well-conditioned.

Easy to compute with standard software.
Can suggest how hard it will be to identify a given parameter of interest,
even without running any estimation.

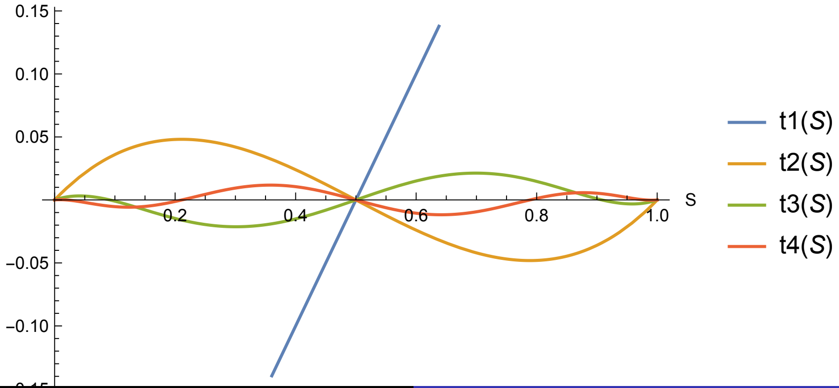
How it Extends

1. higher order expansions: give better approximations (within a radius) and
   - third order s^3 allow to recover the skewness of ϵ; still 2SLS
   - fourth order gives kurtosis, with NL2SLS
2. models with more complex A^* (e.g. some nested logits give rise to NL2SLS)

But Does it Work?

Teaser: for the mixed normal logit ($J = 1$) with one covariate, define $d = \sigma X$; then

$$\log \frac{S}{1-S} = \beta_0 + \beta_1 X + \sum_{i=1}^{\infty} t_i(S) d^{2i}$$



Robustness in the Mixed Normal Logit

We did not use much of the properties of the logistic cdf L and normal cdf Φ: only

- the fact that $L^{-1}(S) = \log(S/(1-S))$
- the form of the P_k in $L^{(k)}(t) = P_k(L(t))$
- $E\varepsilon = 0$ and $V\varepsilon = 1$
- and $E\varepsilon^3 = 0$ and $E\varepsilon^4 = 3$ (for t_2 and above)
- and $E\varepsilon^5 = 0$ and $E\varepsilon^6 = 15$ (for t_3 and above), etc

Robustness in the Mixed Non-Normal Non-Logit

For any L and Φ,
if we normalize $E\varepsilon = 0$ and $V\varepsilon = 1$:

$$\xi = L^{-1}(S) - (\beta_0 + \beta_1 X) + \frac{P_2(S)}{2P_1(S)} E(X\varepsilon)^2 + \frac{P_3(S)}{6P_1(S)} E(X\varepsilon)^3 + \dots$$

A third order 2SLS method would regress $\log(S/(1-S))$ on

$$\mathcal{X} \equiv \left(1, X, X^2 \frac{P_2(S)}{2P_1(S)}, X^3 \frac{P_3(S)}{6P_1(S)}\right)$$

with instruments = the projections $E(\mathcal{X}|Z)$.

ξ_4 and beyond

Using higher order approximations makes things a tiny bit harder:

1. successive powers of σ_ε^2 make it nonlinear IV
2. optimal instruments depend on value of σ_ε^2

But we can build on lower order approximations.

How good are the approximations?

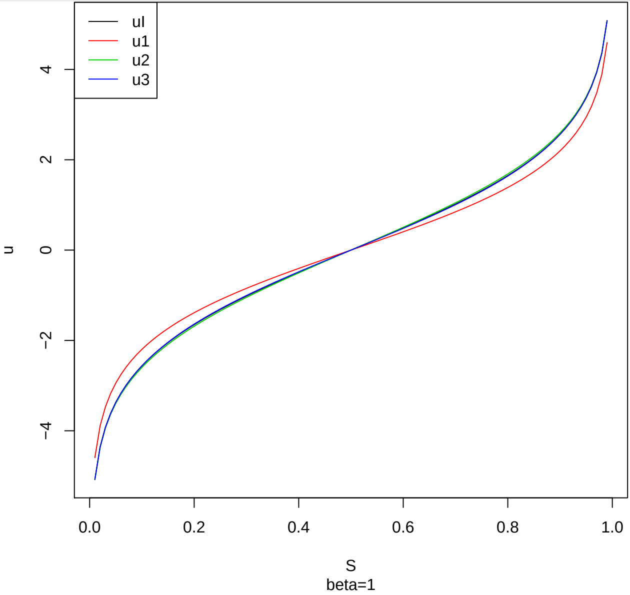
Define a function $u(S,\beta)$ by

$$\int L(u(S,\beta) - \beta\varepsilon)\phi(\varepsilon)d\varepsilon \equiv S.$$

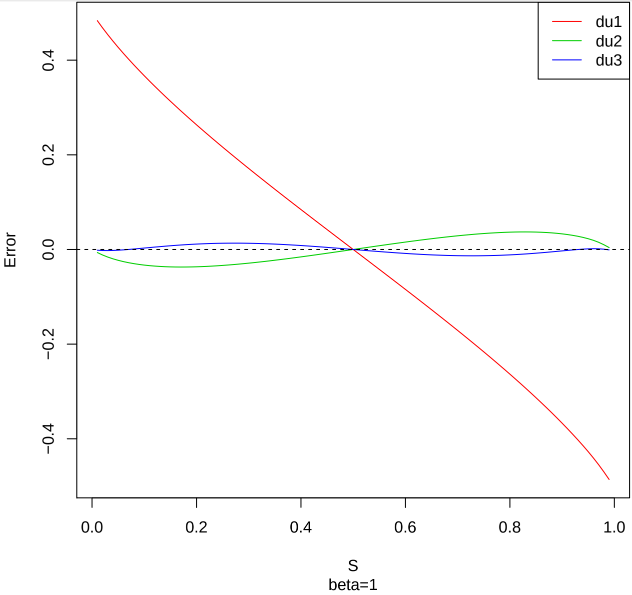
We have $\xi = u(S, \sigma_\varepsilon p) - (a + bp)$ with

1. $u_1(S,\beta) = \log S/(1 - S)$
2. $u_2 = u_1 + (S - 1/2)\beta^2$
3. $u_3 = u_2 - S(1 - S)(S - 1/2)\beta^4$
4. $u_I =$ from Berry inversion.

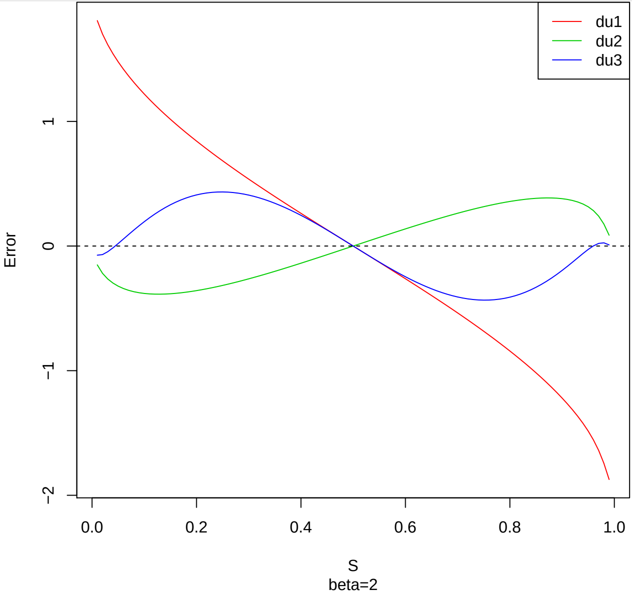
Comparing the u_k's: $\beta = 1$



Comparing the errors $u_k - u_l$: $\beta = 1$



Comparing the errors $u_k - u_l$: $\beta = 2$



Monte Carlo on Standard Macro BLP

Dubé, Fox and Su (2012) design.

$T = 50$ markets and $J = 25$ products in each market

3 observed product characteristics; one (price) is endogenous.

42 instruments (including also covariates and prices in other markets.)

We compare:

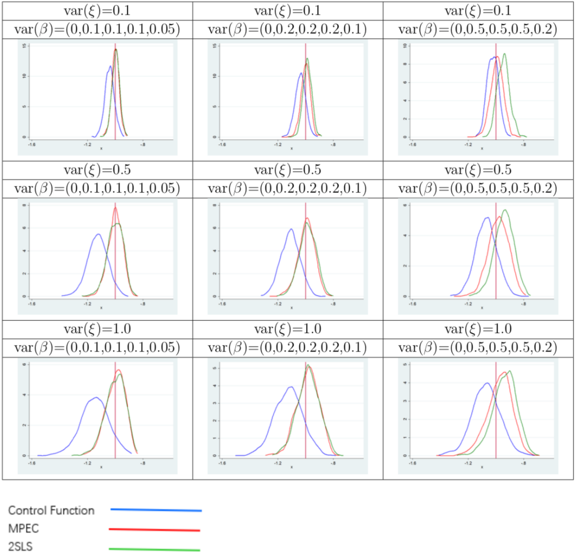
- MPEC (Su and Judd, Dubé–Fox–Su) starting from the true values of the parameters
- the "control function" aproach of Petrin–Train 2010 same
- our 2SLS estimators no need for starting values.

for various values of $V\xi$, $V\beta$

Basic Findings

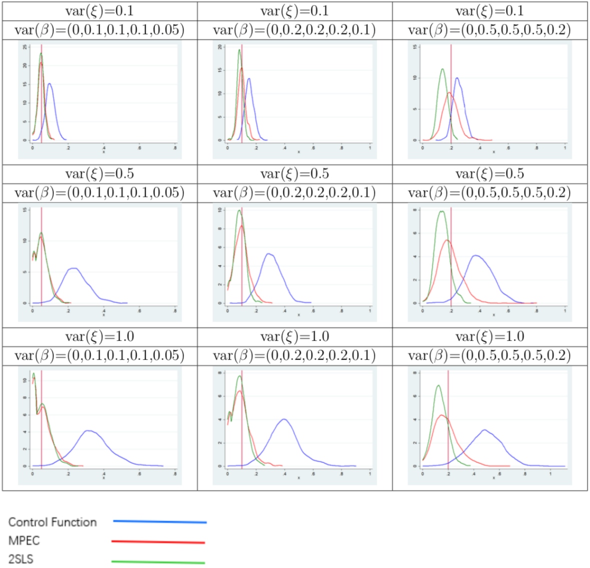
- estimators of the means $E\beta$ of the random coefficients:
  2SLS $\simeq$ MPEC $\gg$ PT
  PT has a large bias that grows with $V\xi$
  ▸ Estimates of means
- estimators of the variances $V\beta$:
  MPEC > 2SLS >< PT
  2SLS has a downward bias that increases with $V\beta$ and decreases with $V\xi$
  PT has less bias but more variance
  ▸ Estimates of variances

Mean of price coefficient



◂ Findings

Variance of price coefficient



Findings

More findings, and conclusion

- experiments with lognormal ϵ show that
  - the second order approach is quite robust to skewness in ϵ
  - using the third order expansion does not help (not enough information to estimate skewness)
- 2SLS provides great starting values for MPEC:
  - convergence to the same estimates
  - at a very minimal cost, +10% over (infeasible) true values.