Efficient Estimation with a Finite Number of Simulation Draws per Observation

Kirill Evdokimov

Princeton University

Introduction

- Method of Simulated Moments, Indirect Inference, Efficient Method of Moments...
  - S = number of draws per observation
  - S can be fixed (different from MSL)
  - $AVar = \left(G' \left(\Omega + \frac{1}{S} \Omega_{sim} \right)^{-1} G \right)^{-1}$
- "Need $S \to \infty$ to achieve efficiency"
- This paper: fixed S (e.g., $S = 1$) allows efficient estimation
  - with almost no additional computational or programming burden

Method(s) of Simulated Moments

Parameter θ is identified by

$$E\left[h\left(W_i, \theta\right)\right] = 0 \text{ iff } \theta = \theta_0.$$

and we can express

$$h\left(w, \theta\right) = \int g\left(w, \varepsilon, \theta\right) dF_\varepsilon\left(\varepsilon\right) = E_\varepsilon\left[g\left(w, \varepsilon_i, \theta\right)\right],$$

for some known f_ε (typically vectors of Uniform[0,1] or N(0,1)).

- Hard to compute the integral, but could compute g for any ε.
  - e.g., models with complicated decisions or equilibrium computation for each i, ε_i, and θ.
- Could compute various (conditional) moments of Y_i (given X_i) by simulation
- Industrial Organization, Labor Economics, ...
- McFadden (1989), Pakes and Pollard (1989) ...

Method(s) of Simulated Moments

Define

$$g_{iS}(\theta) \equiv \frac{1}{S}\sum_{s=1}^{S} g(W_i, \varepsilon_{is}, \theta)$$

$\varepsilon_{is} \in \mathbb{R}^{\dim(\varepsilon)}$ are i.i.d across i and s.

Note that

$$E[g_{iS}(\theta)] = E[g(W_i, \varepsilon_{is}, \theta)] = E[h(W_i, \theta)],$$

hence we can form $\overline{g}_S(\theta) = \frac{1}{n}\sum_{i=1}^{n} g_{iS}(\theta)$ and estimate θ using

$$\hat{Q}_{MSM}(\theta) = \overline{g}'_S(\theta)\,\Xi\,\overline{g}_S(\theta).$$

Remark: can allow non-i.i.d. ε_i (e.g., antithetic draws) and weakly dependent data.

MSM: $\hat{Q}_{MSM}(\theta) = \overline{g}_S(\theta)' \Xi \overline{g}_S(\theta)$.

Standard result:

$$\sqrt{n}\left(\widehat{\theta} - \theta_0\right) \to N\left(0, \left(G'\Xi G\right)^{-1} G'\Xi\Omega_{gg}\Xi G \left(G'\Xi G\right)^{-1}\right),$$

$$\begin{aligned}\Omega_{gg} &= E\left[g_{iS}(\theta_0) g_{iS}(\theta_0)'\right] \\ &= E\left[\left(h(W_i, \theta_0) + \Delta_{iS}\right)\left(h(W_i, \theta_0) + \Delta_{iS}\right)'\right] = \Omega_{hh} + \frac{1}{S}\Omega_{sim},\end{aligned}$$

where $\Omega_{sim} = \Omega_{\Delta\Delta}$ and

$$\Delta_{iS} = \frac{1}{S}\sum_{s=1}^{S} \Delta_{is} = \frac{1}{S}\sum_{s=1}^{S} \underbrace{g(W_i, \varepsilon_{is}, \theta_0) - h(W_i, \theta_0)}_{\equiv \Delta_{is}}$$

What are we doing wrong?

$$g_S\left(W_i, \varepsilon_{iS}, \theta\right) = \frac{1}{S}\sum_{s=1}^{S} g\left(W_i, \varepsilon_{is}, \theta\right)$$

Let $S = 1$, so $g_S\left(W_i, \varepsilon_{iS}, \theta\right) \equiv g\left(W_i, \varepsilon_i, \theta\right)$. Using

$$E\left[g\left(W_i, \varepsilon_i, \theta\right)\right] = 0$$

is inefficient, it ignores what we know about the ε_i.

What are we doing wrong?

$$g_S\left(W_i, \varepsilon_{iS}, \theta\right) = \frac{1}{S}\sum_{s=1}^{S} g\left(W_i, \varepsilon_{is}, \theta\right)$$

Let $S = 1$, so $g_S\left(W_i, \varepsilon_{iS}, \theta\right) \equiv g\left(W_i, \varepsilon_i, \theta\right)$. Using

$$E\left[g\left(W_i, \varepsilon_i, \theta\right)\right] = 0$$

is inefficient, it ignores what we know about the ε_i.

Fact

We know:

- $E\left[\eta\left(\varepsilon_i\right)\right]$ *for any* $\eta\left(\cdot\right)$
- *that* $\varepsilon_i \perp W_i$

Auxiliary and Combined Moment Conditions

Fact

We know:

- $E\left[\eta\left(\varepsilon_i\right)\right]$ *for any* $\eta\left(\cdot\right)$
- *that* $\varepsilon_i \perp W_i$

Introduce Auxiliary moment conditions g_A:

$$g_A\left(W_i, \varepsilon_i\right) = \varphi_W\left(W_i\right) \otimes \left(\varphi_\varepsilon\left(\varepsilon_i\right) - E\left[\varphi_\varepsilon\left(\varepsilon_i\right)\right]\right),$$

and Combined moment conditions g_C:

$$g_C\left(W_i, \varepsilon_i, \theta\right) \equiv \begin{pmatrix} g\left(W_i, \varepsilon_i, \theta\right) \\ g_{Ai}\left(W_i, \varepsilon_i\right) \end{pmatrix}.$$

Note that $E\left[g_{Ai}\left(W_i, \varepsilon_i\right)\right] = 0$.

Combined Moment Conditions

Combined Moment Conditions:

$$g_C\left(W_i,\varepsilon_i,\theta\right)\equiv\begin{pmatrix} g\left(W_i,\varepsilon_i,\theta\right) \\ g_{Ai}\left(W_i,\varepsilon_i\right)\end{pmatrix},\quad \Omega_{CC}=\begin{pmatrix}\Omega_{gg} & \Omega_{gA} \\ \Omega_{Ag} & \Omega_{AA}\end{pmatrix}$$

Let

$$\Sigma_g\equiv\left(G'\Omega_{gg}^{-1}G\right)^{-1}$$

and

$$\begin{aligned}\Sigma_C &\equiv \left(\begin{pmatrix} G' & O_{p\times m_A}\end{pmatrix}\Omega_{CC}^{-1}\begin{pmatrix} G \\ O_{m_A\times p}\end{pmatrix}\right)^{-1} \\ &= \left(G'\left(\Omega_{gg}-\Omega_{gA}\Omega_{AA}^{-1}\Omega_{Ag}\right)^{-1}G\right)^{-1}\end{aligned}$$

Combined Moment Conditions

For any *fixed* m_A we have $\Sigma_C \leq \Sigma_g$.

Why? We can write

$$\Omega_{gg} - \Omega_{gA}\Omega_{AA}^{-1}\Omega_{Ag} = E\left[\rho_i \rho_i'\right],$$

where

$$\rho_i \equiv g_i - E\left[g_i g_{Ai}'\right]\Omega_{AA}^{-1} g_{Ai},$$

and $g_i \equiv g_i\left(W_i, \varepsilon_i, \theta_0\right)$.

Then $E\left[\rho_i g_{Ai}'\right] = 0_{m\times m_A}$, hence

$$\Omega_{gg} = \Omega_{\rho\rho} + \Omega_{gA}\Omega_{AA}^{-1}\Omega_{Ag} \geq \Omega_{\rho\rho}.$$

Combined Moment Conditions

For any *fixed* m_A we have $\Sigma_C \leq \Sigma_g$.

Why? We can write

$$\Omega_{gg} - \Omega_{gA}\Omega_{AA}^{-1}\Omega_{Ag} = E\left[\rho_i \rho_i'\right],$$

where

$$\rho_i \equiv g_i - E\left[g_i g_{Ai}'\right]\Omega_{AA}^{-1} g_{Ai},$$

and $g_i \equiv g_i\left(W_i, \varepsilon_i, \theta_0\right)$.

Then $E\left[\rho_i g_{Ai}'\right] = 0_{m \times m_A}$, hence

$$\Omega_{gg} = \Omega_{\rho\rho} + \Omega_{gA}\Omega_{AA}^{-1}\Omega_{Ag} \geq \Omega_{\rho\rho}.$$

This method is known as the control variates approach to reducing simulation errors in Statistics.

Efficient Simulation

Remember that $E\left[g_{Ai}\left(w,\varepsilon_i\right)\right] = 0_{m_A \times 1}\ \forall w$. Let

$$\mathcal{J} \equiv \left\{\varphi : \mathcal{W} \times \mathcal{E} \rightarrow \mathbb{R}^m\text{:}\ \ E\left[\varphi\left(w,\varepsilon_i\right)\right] = 0 \text{ a.s. } \mathcal{W}\right\}.$$

Proposition

If $\left\{g_{A,j}\left(w,\varepsilon\right)\right\}_{j=1}^{\infty}$ *is a collection of functions* $g_{A,j} \in \mathcal{J}$ *that can approximate any function in* $\mathcal{J}$ *w.r.t.* $\left\|\cdot\right\|_{L^2}$*, i.e.,*

$$\lim_{m_A \to \infty} \min_{\pi \in \mathbb{R}^{m_A}} E\left[\left(\varphi\left(W_i,\varepsilon_i\right) - \pi' g_A^{m_A}\left(W_i,\varepsilon_i\right)\right)^2\right] = 0.$$

Then

$$\Omega_{\rho\rho} \to \Omega_{hh} \text{ and } \Sigma_C \to \Sigma_h \text{ as } m_A \to \infty. \qquad (*)$$

Efficient Simulation ctd.

The auxiliary moment function

$$g_A(w,\varepsilon) = \varphi_W(w) \otimes (\varphi_\varepsilon(\varepsilon) - E[\varphi_\varepsilon(\varepsilon_i)])$$

satisfies the assumptions of the previous Proposition:

Corollary

Suppose $\{g_{A,j}(w,\varepsilon)\}_{j=1}^{\infty}$ *is formed as a tensor product of* $\varphi_W(\cdot)$ *and* $(\varphi_\varepsilon(\cdot) - E[\varphi_\varepsilon(\varepsilon_i)])$*, where* φ_W *and* φ_ε *are splines, polynomials, Fourier series, wavelets, ... Then*

$$\Omega_{\rho\rho} \to \Omega_{hh} \text{ and } \Sigma_C \to \Sigma_h \text{ as } m_A \to \infty. \qquad (*)$$

Semiparametric Efficiency (almost triv. consequence)

Let $\mathcal{I}^{-1}$ denote the semiparametric efficiency bound for the parameters θ of our model.

Corollary

Suppose $\Sigma_h \to \mathcal{I}^{-1}$ as $m \to \infty$, and the conditions of the previous Proposition are satisfied. Then

$$\Sigma_C \to \mathcal{I}^{-1} \text{ as } m, m_A \to \infty.$$

We have a semiparametrically efficient estimator with one simulation draw per observation.

Large Sample Theory

Let $\overline{g}_n(\theta) : \Theta \to \mathbb{R}^K$ be a sequence of random functions and

$$S_n(\theta) \equiv \overline{g}_n(\theta)' \hat{W}(\theta) \overline{g}_n(\theta) / 2,$$

Let $\mathcal{N}$ be a small neighborhood of θ_0, $G \equiv \nabla_\theta g(\theta_0)$, and $\vartheta_n(\theta) \equiv 1 + \sqrt{n} \|\theta - \theta_0\|$.

Assumption 1: For a deterministic sequence $\alpha_n = o\left(\sqrt{n}\right)$ and $B_n \equiv B_{\alpha_n/\sqrt{n}}(\theta_0)$ the following conditions hold:

(i) $g(\theta_0) = 0$, and $\theta_0 \in int(\Theta)$, $\Theta \subset \mathbb{R}^p$;

(ii) $\exists c > 0 : \|g(\theta) - g(\theta_0)\| \geq c \|\theta - \theta_0\| \, \forall \theta \in \mathcal{N}$;

(iii) $\sup_{\theta \in \mathcal{N}} \sqrt{n} \|\overline{g}_n(\theta) - g(\theta)\| = o_p(\alpha_n)$;

Assumption 1 ctd.:

(iv) $\sup_{\theta \in B_n} \sqrt{n} \left\| g(\theta) - G(\theta - \theta_0) \right\| / \vartheta_n(\theta) = o\left(\frac{1}{\alpha_n}\right)$;

(v) (Local S.E.)
$\sup_{\theta \in B_n} \sqrt{n} \left\| \overline{g}_n(\theta) - g(\theta) - \overline{g}_n(\theta_0) \right\| / \vartheta_n(\theta) = o_p\left(\frac{1}{\alpha_n}\right)$;

(vi) *(a)* $\hat{W}(\theta)$ is symmetric and $\exists C > 0$ such that
$1/C \leq \min_{\theta \in \mathcal{N}} \lambda_{\min}\left(\hat{W}(\theta)\right)$ and $\lambda_{\max}\left(\hat{W}\right) \leq C$ w.p.a.1;
(b) $\left\| \hat{W} - W \right\|_\lambda = o_p\left(\frac{1}{\alpha_n}\right)$ for some deterministic matrices W;
(c) $\sup_{\theta \in B_n} \left(\left\| \hat{W}(\theta) - \hat{W} \right\|_\lambda \left(1 + \alpha_n^2 / \vartheta_n^2(\theta)\right)\right) = o_p(1)$ when $\hat{W}(\theta)$ depends on θ (Continuously Updating GMM);

(vii) *(a)* $\sqrt{n} G' W \overline{g}_n(\theta_0) \to_d N(0, M_V)$, for a finite nonzero matrix M_V;
(b) $G'WG \to M_{GWG}$ for a finite nonsingular M_{GWG}.

Large Sample Theory

Theorem

Suppose Assumption 1 holds, $\hat{\theta} \to_p \theta_0$*, and*
$S_n\left(\hat{\theta}\right) \leq \inf_{\theta \in \mathcal{N}} S_n\left(\theta\right) + o_p\left(n^{-1}\right)$.
Then for GMM and CUE estimators $\hat{\theta}$*:*

$$\sqrt{n}\left(\hat{\theta} - \theta_0\right) \to_d N\left(0, M_{GWG}^{-1} M_V M_{GWG}^{-1}\right) \qquad (1)$$

- Nonsmooth moment conditions (cf. Pakes and Pollard (1989))
- The number of moment conditions K can grow with n;
- $\alpha_n = o\left(n^{1/4}\right)$, so can have $K^2/n = o\left(1\right)$ (up to logs)
  - for both GMM and CUE
  - appears to be new
  - e.g., (Lipschitz) conditional moments
  - tight: if $K^2 = Cn$, equation (1) generally does not hold
- cf. Chen, Linton, Kielegom ('03), Chen and Pouzo ('09,'12,'14), Cheng and Liao ('15), Donald, Imbens, and Newey ('03)

Large Sample Theory

Theorem

Suppose $\exists \epsilon_{QB} > 0$ *such that Assumption 1 holds and* $\sup_{\theta \in \mathcal{N}} \sqrt{n} \left\| \overline{g}_n(\theta) - g(\theta) \right\| = o_p\left(\alpha_n n^{-\epsilon_{QB}}\right)$.
If in addition Assumption PRIOR holds, Theorem 1 of Chernozhukov & Hong (2003) holds.

Can be used for estimation and inference in models with nonsmooth moment conditions and increasing K.

These theorems apply to general moment condition-type models, not just MSM and the estimators in this paper.

Computational advantages:

- $S = 1$ vs $S = 10$
- works with finite m_A
- vs importance sampling
- vs quadratures
- implementation:
    - GMM and CUE, variations
    - $\hat{\Omega}_{AA}$ has a useful structure
    - Also, even for CUE we do not need to invert $\hat{\Omega}_{CC}(\theta)$ as we change θ, only $\hat{\Omega}_{gg}(\theta)$

Toy Model Illustration

Suppose $Y_i \sim N(\theta_0, 1)$, and we use MSM to estimate θ:

$$g_i(\theta) = Y_i - (\theta + \varepsilon_i), \quad \varepsilon_i \sim N(0,1)$$

Then $\sqrt{n}\left(\hat{\theta}_{MSM} - \theta_0\right) \to_d N(0,2)$, since $\overline{g}_n(\theta) = \overline{Y}_n - \theta - \bar{\varepsilon}_n$.

Combined moment conditions:

$$g_{iC}(\theta) = \begin{pmatrix} Y_i - (\theta + \varepsilon_i) \\ \varepsilon_i \end{pmatrix}, \quad \overline{g}_{C,n}(\theta) = \begin{pmatrix} \overline{Y}_n - \theta - \bar{\varepsilon}_n \\ \bar{\varepsilon}_n \end{pmatrix}$$

hence

$$\begin{aligned} \hat{Q}_{ES\text{-}MSM}(\theta) &= \overline{g}_{C,n}(\theta)' \Omega_{CC}^{-1} \overline{g}_{C,n}(\theta) \\ &= \left(\overline{Y}_n - \theta\right)^2 + \bar{\varepsilon}_n^2 + remainder_n(\theta), \end{aligned}$$

so $\sqrt{n}\left(\hat{\theta}_{ES\text{-}MSM} - \theta_0\right) \to_d N(0,1)$.

Indirect Inference Framework

Gouriéroux, Monfort, and Renault (1993), Smith (1993), Gallant and Tauchen (1996).

$$\begin{aligned} \text{Data: } & Y_i = y\left(X_i, \varepsilon_i; \theta_0\right) \\ \text{Simulation: } & Y_{is}\left(\theta\right) \equiv y\left(\varepsilon_{is}, X_i; \theta\right) \end{aligned}$$

Auxiliary model: a combination of M- and/or Z-estimators, so that

$$\text{Data:} \frac{1}{n}\sum_{i=1}^{n} \psi\left(Y_i, X_i, \hat{\beta}\right) = 0$$

$$\text{Simulation:} \frac{1}{n}\sum_{i=1}^{n} \widetilde{\psi}_i\left(\theta, \widetilde{\beta}\left(\theta\right)\right) = 0$$

where $\widetilde{\psi}_i\left(\theta, \beta\right) \equiv \frac{1}{S}\sum_{s=1}^{S} \psi\left(Y_{is}\left(\theta\right), X_i, \beta\right)$.

Indirect Inference Framework

"Score-based" estimator :

$$\hat{\theta}_{II\text{-}SC} = \arg\min_{\theta\in\Theta} \left(\frac{1}{n} \sum_{i=1}^{n} \widetilde{\psi}_i \left(\theta, \hat{\beta}\right) \right)' \Xi_{II} \left(\frac{1}{n} \sum_{i=1}^{n} \widetilde{\psi}_i \left(\theta, \hat{\beta}\right) \right)$$

Asymptotic distribution with optimal weighting matrix:

$$\sqrt{n} \left(\hat{\theta}_{II\text{-}SC} - \theta_0\right) \to_d N\left(0, \Sigma_{II}\right)$$

$$\Sigma_{II} \equiv \left(\Psi_\theta' \Omega_{\psi-\widetilde{\psi}}^{-1} \Psi_\theta\right)^{-1} = \left(\Psi_\theta' \left[\left(1 + \tfrac{1}{S}\right) J\right]^{-1} \Psi_\theta\right)^{-1}$$

where

$$J = V\left[\psi\left(Y_{is}\left(\theta_0\right), X_i, \beta_0\right) - E\left[\psi\left(Y_{is}\left(\theta_0\right), X_i, \beta_0\right) \middle| X_i\right]\right]$$

Corrected Score Function

Can write $\hat{\theta}_{II\text{-}SC}$ as GMM with "stacked" moment function

$$v_i(\theta,\beta) \equiv \begin{pmatrix} \widetilde{\psi}_i(\theta,\beta) \\ \psi_i(\beta) \end{pmatrix}.$$

Same results as for MSM hold. Fix $S = 1$ and replace ψ_{Ci} with

$$\widetilde{\psi}_{Ci}(\theta,\beta) \equiv \psi\left(y\left(X_i,\varepsilon_i,\theta\right),X_i,\beta\right) - \hat{\Omega}_{\widetilde{\psi}A}\hat{\Omega}_{AA}^{-1}g_{Ai},\ \hat{\Omega}_{\widetilde{\psi}A} \equiv E\left[\widetilde{\psi}_i g_{Ai}'\right].$$

Then, as for MSM estimator

$$\sqrt{n}\left(\hat{\theta}_{II,C} - \theta_0\right) \to_d N\left(0, \left(R_\theta' J^{-1} R_\theta\right)^{-1}\right).$$

Simple Monte Carlo

DGP:

$$Y_i = \lambda\left(\beta_1 + \beta_2 X_i - U_i\right), \quad \begin{pmatrix} X_i \\ U_i \end{pmatrix} \sim_{iid} N\left(0, \begin{pmatrix} 1 & 0 \\ 0 & \sigma^2 \end{pmatrix}\right)$$

where $\beta_1 = 0$, $\beta_2 = 1$, $\sigma = 1$.

Models:

$$L : \lambda(z) = 1 \big/ (1 + \exp(-z))$$
$$T : \lambda(z) = z1\{z > 0\}$$
$$B : \lambda(z) = 1\{z > 0\}$$

The moments:

$$g(W_i, \varepsilon_i, \theta) = \begin{pmatrix} Y_i - \lambda\left(\beta_1 + \beta_2 X_i - \sigma\varepsilon_i\right) \\ \cdots \\ Y_i^{K_Y} - \lambda^{K_Y}\left(\beta_1 + \beta_2 X_i - \sigma\varepsilon_i\right) \end{pmatrix} \otimes \varphi_X(X_i),$$

$$g_A(W_i, \varepsilon_i) = \varphi_X(X_i) \otimes \varphi_\varepsilon(\varepsilon_i).$$

	mbias	iqr	mad	RE	bias	std	rmse	RE
MSM w S=10	-0.000	0.107	0.053	1.000	-0.002	0.078	0.078	1.000
S=1								
MSM	-0.003	0.142	0.072	1.331	-0.002	0.105	0.105	1.332
ES-MSM:1	-0.000	0.102	0.051	0.953	-0.002	0.076	0.076	0.970
ES-MSM:2	-0.000	0.104	0.052	0.972	-0.002	0.078	0.078	0.994
ES-MSM:3	-0.000	0.108	0.054	1.016	-0.002	0.080	0.080	1.018
ES-MSM:4	-0.003	0.114	0.057	1.071	-0.002	0.085	0.085	1.082

Table: Model L: $\beta_1 = 0.0$, $n = 200$, $K_Z = 4$, $K_Y = 2$.

	mbias	iqr	mad	RE	bias	std	rmse	RE
MSM w S=10	0.020	0.118	0.060	1.000	0.020	0.088	0.090	1.000
S=1								
MSM	-0.001	0.159	0.080	1.348	-0.002	0.119	0.119	1.354
ES-MSM:1	0.014	0.117	0.058	0.988	0.014	0.086	0.088	0.986
ES-MSM:2	0.023	0.121	0.064	1.028	0.024	0.092	0.095	1.048
ES-MSM:3	0.024	0.129	0.066	1.095	0.024	0.096	0.099	1.098
ES-MSM:4	0.016	0.145	0.075	1.233	0.016	0.109	0.110	1.247

Table: Model L: $\beta_2 = 1.0$, $n = 200$, $K_Z = 4$, $K_Y = 2$.

	mbias	iqr	mad	RE	bias	std	rmse	RE
MSM w S=10	-0.033	0.081	0.048	1.000	-0.032	0.060	0.068	1.000
S=1								
MSM	0.002	0.108	0.054	1.336	0.004	0.081	0.081	1.356
ES-MSM:1	0.004	0.112	0.056	1.387	0.007	0.084	0.084	1.405
ES-MSM:2	-0.032	0.083	0.048	1.027	-0.030	0.062	0.069	1.040
ES-MSM:3	-0.040	0.080	0.050	0.984	-0.039	0.059	0.071	0.983
ES-MSM:4	-0.032	0.086	0.049	1.061	-0.030	0.064	0.071	1.067

Table: Model L: $\sigma = 1.0$, $n = 200$, $K_Z = 4$, $K_Y = 2$.

	mbias	iqr	mad	RE	bias	std	rmse	RE
MSM w S=10	0.001	0.073	0.036	1.000	0.001	0.055	0.055	1.000
S=1								
MSM	0.001	0.095	0.048	1.304	0.002	0.073	0.073	1.333
ES-MSM:1	0.000	0.070	0.035	0.961	0.000	0.053	0.053	0.973
ES-MSM:2	-0.000	0.072	0.036	0.984	0.000	0.053	0.053	0.980
ES-MSM:3	0.000	0.071	0.035	0.973	0.000	0.054	0.054	0.985
ES-MSM:4	-0.001	0.074	0.037	1.013	0.000	0.055	0.055	1.002

Table: Model L: $\beta_1 = 0.0$, $n = 400$, $K_Z = 4$, $K_Y = 2$.

	mbias	iqr	mad	RE	bias	std	rmse	RE
MSM w S=10	0.015	0.079	0.041	1.000	0.015	0.057	0.059	1.000
S=1								
MSM	0.002	0.106	0.053	1.345	0.002	0.078	0.078	1.373
ES-MSM:1	0.011	0.077	0.039	0.976	0.010	0.056	0.057	0.985
ES-MSM:2	0.017	0.080	0.041	1.009	0.016	0.058	0.060	1.014
ES-MSM:3	0.017	0.081	0.042	1.024	0.017	0.059	0.061	1.032
ES-MSM:4	0.013	0.083	0.042	1.052	0.012	0.061	0.062	1.076

Table: Model L: $\beta_2 = 1.0$, $n = 400$, $K_Z = 4$, $K_Y = 2$.

	mbias	iqr	mad	RE	bias	std	rmse	RE
MSM w S=10	-0.016	0.058	0.031	1.000	-0.016	0.043	0.045	1.000
S=1								
MSM	0.001	0.077	0.039	1.336	0.003	0.057	0.057	1.347
ES-MSM:1	0.003	0.079	0.039	1.365	0.005	0.059	0.059	1.375
ES-MSM:2	-0.015	0.060	0.032	1.040	-0.014	0.044	0.046	1.039
ES-MSM:3	-0.020	0.058	0.032	0.998	-0.020	0.042	0.047	0.995
ES-MSM:4	-0.016	0.058	0.031	1.008	-0.016	0.043	0.046	1.015

Table: Model L: $\sigma = 1.0$, $n = 400$, $K_Z = 4$, $K_Y = 2$.

	mbias	iqr	mad	RE	bias	std	rmse	RE
MSM w S=10	0.001	0.050	0.025	1.000	0.001	0.037	0.037	1.000
S=1								
MSM	0.001	0.069	0.034	1.384	0.001	0.051	0.051	1.374
ES-MSM:1	0.000	0.048	0.024	0.960	0.001	0.036	0.036	0.969
ES-MSM:2	0.000	0.047	0.024	0.953	0.001	0.036	0.036	0.971
ES-MSM:3	0.001	0.047	0.024	0.954	0.001	0.036	0.036	0.964
ES-MSM:4	0.001	0.048	0.024	0.968	0.001	0.036	0.036	0.971
ES-MSM:5	0.001	0.049	0.024	0.984	0.001	0.036	0.036	0.979

Table: Model L: $\beta_1 = 0.0$, $n = 800$, $K_Z = 4$, $K_Y = 2$.

	mbias	iqr	mad	RE	bias	std	rmse	RE
MSM w S=10	0.008	0.053	0.027	1.000	0.008	0.039	0.040	1.000
S=1								
MSM	-0.000	0.070	0.035	1.318	0.000	0.053	0.053	1.361
ES-MSM:1	0.005	0.053	0.026	0.995	0.005	0.038	0.039	0.986
ES-MSM:2	0.009	0.053	0.027	0.991	0.009	0.039	0.040	0.994
ES-MSM:3	0.010	0.053	0.027	1.009	0.010	0.039	0.040	1.000
ES-MSM:4	0.007	0.054	0.027	1.020	0.007	0.040	0.040	1.021
ES-MSM:5	0.006	0.056	0.028	1.061	0.006	0.040	0.041	1.043

Table: Model L: $\beta_2 = 1.0$, $n = 800$, $K_Z = 4$, $K_Y = 2$.

	mbias	iqr	mad	RE	bias	std	rmse	RE
MSM w S=10	-0.010	0.040	0.021	1.000	-0.009	0.030	0.031	1.000
S=1								
MSM	-0.000	0.055	0.027	1.354	0.001	0.040	0.040	1.358
ES-MSM:1	0.001	0.055	0.027	1.371	0.002	0.041	0.041	1.369
ES-MSM:2	-0.008	0.042	0.022	1.042	-0.007	0.031	0.032	1.041
ES-MSM:3	-0.011	0.039	0.022	0.977	-0.011	0.029	0.031	0.985
ES-MSM:4	-0.009	0.040	0.021	0.987	-0.009	0.029	0.031	0.984
ES-MSM:5	-0.009	0.039	0.021	0.974	-0.009	0.029	0.030	0.981

Table: Model L: $\sigma = 1.0$, $n = 800$, $K_Z = 4$, $K_Y = 2$.

	mbias	iqr	mad	RE	bias	std	rmse	RE
MSM w S=10	-0.048	0.160	0.085	1.000	-0.058	0.125	0.138	1.000
S=1								
MSM	-0.085	0.246	0.129	1.539	-0.106	0.194	0.221	1.555
ES-MSM:1	-0.089	0.218	0.117	1.362	-0.111	0.176	0.208	1.404
ES-MSM:2	-0.077	0.190	0.106	1.188	-0.092	0.151	0.177	1.207
ES-MSM:3	-0.071	0.179	0.100	1.119	-0.083	0.142	0.165	1.137
ES-MSM:4	-0.068	0.178	0.098	1.111	-0.081	0.141	0.163	1.129

Table: Model T: $\beta_1 = 0.0$, $n = 200$, $K_Z = 4$, $K_Y = 2$.

	mbias	iqr	mad	RE	bias	std	rmse	RE
MSM w S=10	0.043	0.173	0.091	1.000	0.050	0.133	0.142	1.000
S=1								
MSM	0.064	0.251	0.128	1.449	0.081	0.190	0.206	1.428
ES-MSM:1	0.058	0.192	0.100	1.107	0.066	0.148	0.162	1.116
ES-MSM:2	0.049	0.179	0.093	1.030	0.057	0.136	0.148	1.027
ES-MSM:3	0.054	0.185	0.097	1.067	0.061	0.141	0.154	1.065
ES-MSM:4	0.054	0.188	0.098	1.084	0.065	0.146	0.160	1.098

Table: Model T: $\beta_2 = 1.0$, $n = 200$, $K_Z = 4$, $K_Y = 2$.

	mbias	iqr	mad	RE	bias	std	rmse	RE
MSM w S=10	-0.058	0.118	0.075	1.000	-0.056	0.089	0.105	1.000
S=1								
MSM	0.083	0.200	0.110	1.689	0.107	0.180	0.209	2.011
ES-MSM:1	0.062	0.200	0.103	1.688	0.080	0.166	0.184	1.857
ES-MSM:2	-0.001	0.156	0.078	1.314	0.008	0.123	0.123	1.378
ES-MSM:3	-0.019	0.140	0.072	1.181	-0.014	0.106	0.107	1.189
ES-MSM:4	-0.029	0.133	0.070	1.125	-0.025	0.103	0.106	1.155

Table: Model T: $\sigma = 1.0$, $n = 200$, $K_Z = 4$, $K_Y = 2$.

	mbias	iqr	mad	RE	bias	std	rmse	RE
MSM w S=10	-0.020	0.111	0.056	1.000	-0.023	0.083	0.086	1.000
S=1								
MSM	-0.029	0.151	0.078	1.359	-0.035	0.113	0.119	1.362
ES-MSM:1	-0.038	0.132	0.069	1.193	-0.042	0.100	0.109	1.204
ES-MSM:2	-0.034	0.124	0.064	1.118	-0.038	0.093	0.100	1.113
ES-MSM:3	-0.029	0.120	0.060	1.077	-0.034	0.088	0.094	1.058
ES-MSM:4	-0.028	0.117	0.059	1.055	-0.033	0.087	0.093	1.049

Table: Model T: $\beta_1 = 0.0$, $n = 400$, $K_Z = 4$, $K_Y = 2$.

	mbias	iqr	mad	RE	bias	std	rmse	RE
MSM w S=10	0.020	0.116	0.058	1.000	0.024	0.086	0.089	1.000
S=1								
MSM	0.026	0.156	0.079	1.340	0.030	0.115	0.118	1.340
ES-MSM:1	0.023	0.125	0.064	1.078	0.026	0.092	0.096	1.077
ES-MSM:2	0.023	0.118	0.061	1.017	0.026	0.088	0.092	1.029
ES-MSM:3	0.024	0.117	0.061	1.002	0.026	0.088	0.092	1.026
ES-MSM:4	0.027	0.117	0.061	1.005	0.028	0.089	0.093	1.042

Table: Model T: $\beta_2 = 1.0$, $n = 400$, $K_Z = 4$, $K_Y = 2$.

	mbias	iqr	mad	RE	bias	std	rmse	RE
MSM w S=10	-0.034	0.083	0.048	1.000	-0.033	0.063	0.071	1.000
S=1								
MSM	0.030	0.120	0.061	1.434	0.035	0.090	0.096	1.425
ES-MSM:1	0.023	0.120	0.061	1.439	0.029	0.091	0.096	1.450
ES-MSM:2	-0.003	0.101	0.051	1.210	0.000	0.077	0.077	1.223
ES-MSM:3	-0.013	0.091	0.047	1.091	-0.010	0.068	0.069	1.088
ES-MSM:4	-0.017	0.087	0.046	1.044	-0.014	0.067	0.068	1.061

Table: Model T: $\sigma = 1.0$, $n = 400$, $K_Z = 4$, $K_Y = 2$.

	mbias	iqr	mad	RE	bias	std	rmse	RE
MSM w S=10	-0.008	0.078	0.040	1.000	-0.009	0.057	0.058	1.000
S=1								
MSM	-0.014	0.108	0.053	1.376	-0.016	0.079	0.081	1.384
ES-MSM:1	-0.017	0.093	0.048	1.188	-0.019	0.068	0.071	1.194
ES-MSM:2	-0.016	0.084	0.044	1.073	-0.017	0.063	0.065	1.097
ES-MSM:3	-0.014	0.080	0.041	1.021	-0.015	0.059	0.061	1.038
ES-MSM:4	-0.014	0.080	0.041	1.026	-0.015	0.059	0.061	1.029
ES-MSM:5	-0.013	0.081	0.041	1.030	-0.014	0.058	0.060	1.024

Table: Model T: $\beta_1 = 0.0$, $n = 800$, $K_Z = 4$, $K_Y = 2$.

	mbias	iqr	mad	RE	bias	std	rmse	RE
MSM w S=10	0.007	0.077	0.039	1.000	0.010	0.057	0.058	1.000
S=1								
MSM	0.010	0.105	0.052	1.362	0.013	0.078	0.079	1.373
ES-MSM:1	0.010	0.085	0.042	1.106	0.012	0.063	0.064	1.100
ES-MSM:2	0.009	0.080	0.040	1.036	0.012	0.058	0.060	1.026
ES-MSM:3	0.010	0.079	0.039	1.028	0.012	0.058	0.059	1.017
ES-MSM:4	0.011	0.080	0.040	1.033	0.013	0.058	0.059	1.020
ES-MSM:5	0.011	0.079	0.040	1.026	0.013	0.058	0.060	1.024

Table: Model T: $\beta_2 = 1.0$, $n = 800$, $K_Z = 4$, $K_Y = 2$.

	mbias	iqr	mad	RE	bias	std	rmse	RE
MSM w S=10	-0.021	0.058	0.034	1.000	-0.020	0.045	0.049	1.000
S=1								
MSM	0.013	0.081	0.042	1.406	0.015	0.062	0.064	1.389
ES-MSM:1	0.011	0.081	0.042	1.407	0.013	0.062	0.063	1.397
ES-MSM:2	-0.002	0.068	0.034	1.178	-0.002	0.052	0.052	1.173
ES-MSM:3	-0.007	0.062	0.031	1.067	-0.006	0.047	0.047	1.051
ES-MSM:4	-0.009	0.060	0.030	1.046	-0.009	0.046	0.046	1.026
ES-MSM:5	-0.011	0.059	0.030	1.024	-0.010	0.045	0.047	1.021

Table: Model T: $\sigma = 1.0$, $n = 800$, $K_Z = 4$, $K_Y = 2$.

	mbias	iqr	mad	RE	bias	std	rmse	RE
MSM w S=10	-0.003	0.163	0.082	1.000	-0.004	0.126	0.126	1.000
S=1								
MSM	0.000	0.204	0.102	1.249	-0.003	0.153	0.153	1.213
ES-MSM:1	-0.001	0.179	0.090	1.099	-0.002	0.136	0.136	1.078
ES-MSM:2	-0.003	0.176	0.087	1.079	-0.004	0.133	0.133	1.061
ES-MSM:3	-0.003	0.176	0.088	1.077	-0.004	0.134	0.134	1.062
ES-MSM:4	-0.002	0.179	0.090	1.096	-0.004	0.136	0.136	1.082

Table: Model B: $\beta_1 = 0.0$, $n = 200$, $K_Z = 4$, $K_Y = 1$.

	mbias	iqr	mad	RE	bias	std	rmse	RE
MSM w S=10	0.139	0.259	0.157	1.000	0.159	0.205	0.259	1.000
S=1								
MSM	0.093	0.290	0.150	1.118	0.111	0.225	0.251	1.099
ES-MSM:1	0.100	0.275	0.144	1.061	0.120	0.214	0.245	1.045
ES-MSM:2	0.131	0.271	0.156	1.045	0.154	0.211	0.262	1.032
ES-MSM:3	0.147	0.273	0.165	1.054	0.170	0.215	0.274	1.051
ES-MSM:4	0.159	0.284	0.176	1.098	0.178	0.221	0.284	1.081

Table: Model B: $\beta_2 = 1.0$, $n = 200$, $K_Z = 4$, $K_Y = 1$.

	mbias	iqr	mad	RE	bias	std	rmse	RE
MSM w S=10	0.002	0.110	0.055	1.000	0.001	0.081	0.081	1.000
S=1								
MSM	0.000	0.137	0.069	1.241	0.002	0.101	0.101	1.248
ES-MSM:1	0.001	0.119	0.060	1.079	0.001	0.089	0.089	1.094
ES-MSM:2	-0.001	0.116	0.058	1.048	0.001	0.085	0.085	1.055
ES-MSM:3	0.000	0.114	0.057	1.032	0.001	0.085	0.085	1.044
ES-MSM:4	0.001	0.112	0.056	1.017	0.001	0.084	0.084	1.041

Table: Model B: $\beta_1 = 0.0$, $n = 400$, $K_Z = 4$, $K_Y = 1$.

	mbias	iqr	mad	RE	bias	std	rmse	RE
MSM w S=10	0.066	0.152	0.087	1.000	0.072	0.116	0.137	1.000
S=1								
MSM	0.038	0.187	0.095	1.237	0.045	0.140	0.147	1.206
ES-MSM:1	0.042	0.176	0.088	1.158	0.048	0.129	0.138	1.108
ES-MSM:2	0.058	0.163	0.088	1.073	0.065	0.123	0.139	1.059
ES-MSM:3	0.065	0.161	0.091	1.059	0.072	0.124	0.143	1.062
ES-MSM:4	0.071	0.164	0.093	1.083	0.077	0.125	0.146	1.070

Table: Model B: $\beta_2 = 1.0$, $n = 400$, $K_Z = 4$, $K_Y = 1$.

	mbias	iqr	mad	RE	bias	std	rmse	RE
MSM w S=10	0.000	0.072	0.036	1.000	0.001	0.054	0.054	1.000
S=1								
MSM	0.001	0.097	0.049	1.341	0.000	0.071	0.071	1.328
ES-MSM:1	0.000	0.081	0.040	1.117	0.001	0.060	0.060	1.119
ES-MSM:2	0.001	0.076	0.038	1.046	0.000	0.057	0.057	1.064
ES-MSM:3	0.002	0.073	0.037	1.008	0.001	0.056	0.056	1.041
ES-MSM:4	0.001	0.074	0.037	1.021	0.001	0.056	0.056	1.040
ES-MSM:5	0.002	0.074	0.037	1.020	0.001	0.056	0.056	1.034

Table: Model B: $\beta_1 = 0.0$, $n = 800$, $K_Z = 4$, $K_Y = 1$.

	mbias	iqr	mad	RE	bias	std	rmse	RE
MSM w S=10	0.032	0.102	0.055	1.000	0.035	0.076	0.084	1.000
S=1								
MSM	0.017	0.130	0.065	1.278	0.020	0.097	0.099	1.273
ES-MSM:1	0.017	0.121	0.060	1.186	0.022	0.088	0.091	1.161
ES-MSM:2	0.027	0.113	0.057	1.112	0.031	0.084	0.089	1.097
ES-MSM:3	0.030	0.112	0.057	1.106	0.034	0.083	0.090	1.091
ES-MSM:4	0.032	0.110	0.058	1.084	0.037	0.083	0.091	1.091
ES-MSM:5	0.031	0.111	0.058	1.095	0.037	0.083	0.091	1.094

Table: Model B: $\beta_2 = 1.0$, $n = 800$, $K_Z = 4$, $K_Y = 1$.

Literature

- McFadden (1989), Pakes and Pollard (1989), Andrews (1994), Newey and McFadden (1994), Chen, Linton, and Van Keilegom (2003), Armstrong, Gallant, Hong, and Huiyu (2012), Chen and Pouzo (2012) ...
- Smith (1990, 1993), Gouriéroux, Monfort, and Renault (1993), Gallant and Tauchen (1996), Keane and Smith (2003); Gallant and Long (1997); Carrasco and Florens(2002); Lerman and Manski (1981), Lee and Song (2013)...
- Lee (1992,1995), Laffont, Ossard, and Vuong (1995), Kristensen and Salanié (2013) ...
- Geweke (1996), Chernozhukov and Hong (2004); Hajivassiliou, McFadden, and Ruud (1996), Ackerberg (2009)
- Control variates approach
- Brown and Newey (1998), Prokhorov and Schmidt (2009), Graham (2011); Newey (1985)