Simple Estimation of Semiparametric Models with Measurement Errors

Kirill Evdokimov
Andrei Zeleneev

January 2020

Universitat Pompeu Fabra & Barcelona GSE
Princeton University

Framework: EIV for general GMM models

- General moment condition model

$$\mathbb{E}\left[g(X_i^*, S_i, \theta)\right] = 0 \quad \text{iff} \quad \theta = \theta_0$$

- X_i^* is mismeasured:

$$X_i = X_i^* + \varepsilon_i,$$

  - We first consider classical measurement error ε_i, i.e.: $\varepsilon_i \perp (X_i^*, S_i), \quad \mathbb{E}[\varepsilon_i] = 0$
  - Then discuss non-classical, multivariate, and serially correlated measurement errors.
- Leading example: nonlinear regression

$$Y_i = \rho(X_i^*, W_i, \theta_0) + U_i, \quad \mathbb{E}\left[U_i | X_i^*, W_i, Z_i\right] = 0$$
$$g(x, w, y, z, \theta) = (\rho(x, w, \theta) - y)h(x, w, z)$$

- Logit/Probit/Tobit,CES production function ...
- but also multi-equation models, structural models, ...

Framework: Moderate Measurement Errors

- Most of the literature treats $\sigma^2 = \mathbb{E}\left[\varepsilon_i^2\right]$ as fixed
  - Too pessimistic and not representative of most empirical settings
  - Estimation of an infinite dimensional nuisance parameter and/or numerical simulation are required
  - The asymptotic theory assumes σ^2 is bounded away from zero
  - Hausman, Newey, Ichimura, and Powell (1991), Newey (2001), Li (2002), Schennach (2004, 2007), Hu and Schennach (2008), and others
- We consider an alternative asymptotic framework modeling $\sigma_n^2 \equiv \mathbb{E}\left[\varepsilon_{in}^2\right]$ as shrinking to zero
  - Perhaps a better match for the problem
  - Main focus is on using instruments to ID the problem
  - Simple estimation procedure: GMM
  - Advantageous in terms of the quality of point estimates
  - Valid inference

Estimation: Motivation

Taylor expansion of $g(X_i, S_i, \theta)$ (in the spirit of Chesher, 1991): $\sigma_n^2 \equiv \mathbb{E}\left[\varepsilon_{in}^2\right] \to 0$

$$g(X_i, S_i, \theta) = g(X_i^*, S_i, \theta) + g_x^{(1)}(X_i^*, S_i, \theta)\varepsilon_{in} + \frac{1}{2} g_x^{(2)}(X_i^*, S_i, \theta)\varepsilon_{in}^2 + O_p(\sigma_n^3),$$

$$\mathbb{E}[g(X_i, S_i, \theta)] = \mathbb{E}[g(X_i^*, S_i, \theta)] + \frac{\sigma_n^2}{2}\mathbb{E}[g_x^{(2)}(X_i^*, S_i, \theta)] + O(\sigma_n^3),$$

where $g_x^{(k)}(x, r, \theta) \equiv \dfrac{\partial^k}{\partial x^k} g(x, r, \theta)$

Therefore, $\mathbb{E}\left[g(X_i, S_i, \theta_0)\right] = O(\sigma_n^2)$, and the standard estimator is:

- Consistent if $\sigma_n^2 \to 0$
- Asymptotically biased if $\sqrt{n}\sigma_n^2 \to C \in (0, +\infty)$
- Not $\sqrt{n}$-consistent if $\sqrt{n}\sigma_n^2 \to \infty$

SME Estimator

- Define the corrected moment function:

$$\psi\left(X_i, R_i, \theta, \gamma\right) \equiv g\left(X_i, R_i, \theta\right) - \gamma g_x^{(2)}\left(X_i, R_i, \theta\right)$$

Assumption (Moderate Measurement Error, $K = 2$)

$\sigma_n^2 = o(n^{-1/3})$

- $\mathbb{E}\left[\psi(X_i, S_i, \theta_0, \gamma_{0n})\right] = O(\sigma_n^3) = o(n^{-1/2})$, where $\gamma_{0n} \equiv \sigma_n^2/2$

- SME estimator:

$$(\hat{\theta}, \hat{\gamma}) = \underset{\theta\in\Theta,\gamma\in\Gamma}{\text{argmin}}\, \hat{Q}(\theta, \gamma), \quad \hat{Q}(\theta, \gamma) = \overline{\psi}(\theta, \gamma)'\hat{\Xi}(\theta, \gamma)\overline{\psi}(\theta, \gamma)$$

- Higher order expansion is more subtle: one needs to correct the correction terms.

Monte Carlo: Point Estimation

From Schennach (2007, Ecta):

$$\begin{aligned} Y_i &= \rho(X_i^*, \theta_0) + U_i \\ X_i^* &= Z_i + \eta_i \\ X_i &= X_i^* + \varepsilon_i \end{aligned} \qquad \begin{pmatrix} Z_i \\ \eta_i \\ U_i \\ \varepsilon_i \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1/4 & 0 & 0 \\ 0 & 0 & 1/4^{\dagger} & 0 \\ 0 & 0 & 0 & 1/4 \end{pmatrix} \right)$$

$$g(y, x, z, \theta) = (\rho(x, \theta) - y) h(x, z)$$

The ratio of the standard deviations is $\sigma_\varepsilon / \sigma_{X^*} \approx 0.45$ “fairly large”, $n = 1000$

Specifications:

- Polynomial
- Rational Fraction
- Probit

Monte Carlo: Polynomial Specification

$$\rho(x,\theta) = \theta_1 + \theta_2 x + \theta_3 x^2 + \theta_4 x^3, \quad \theta_0 = (1, 1, 0, -0.5)'$$

	Bias				Std. Dev.				RMSE				
	θ_1	θ_2	θ_3	θ_4	θ_1	θ_2	θ_3	θ_4	θ_1	θ_2	θ_3	θ_4	All
OLS	-0.00	-0.43	0.00	0.21	0.07	0.13	0.06	0.04	0.07	0.45	0.06	0.22	0.51
SS07	-0.05	-0.07	-0.02	0.05	0.17	0.19	0.24	0.05	0.17	0.20	0.24	0.07	0.36
$\text{SME}_{a,K=2}$	-0.00	0.10	0.00	0.00	0.10	0.23	0.10	0.08	0.10	0.25	0.10	0.08	0.29
$\text{SME}_{b,K=2}$	-0.00	-0.17	0.00	0.12	0.08	0.20	0.08	0.07	0.08	0.26	0.08	0.14	0.32
$\text{SME}_{a,K=4}$	-0.00	0.11	0.00	-0.01	0.10	0.26	0.10	0.08	0.10	0.29	0.10	0.08	0.33
$\text{SME}_{b,K=4}$	0.00	0.00	-0.00	0.02	0.09	0.21	0.10	0.08	0.09	0.21	0.10	0.08	0.27

$\sigma_\varepsilon/\sigma_{X^*} \approx 0.45$, $Y_i = \rho(X_i^*, \theta_0) + U_i$, $U_i \sim N(0, 1/4)$

Monte Carlo: Rational Fraction Specification

$$\rho(x,\theta) = \theta_1 + \theta_2 x + \frac{\theta_3}{(1+x^2)^2}, \qquad \theta_0 = (1,1,2)'$$

	Bias			Std. Dev.			RMSE			
	θ_1	θ_2	θ_3	θ_1	θ_2	θ_3	θ_1	θ_2	θ_3	All
OLS	0.339	-0.167	-0.644	0.040	0.020	0.076	0.341	0.168	0.648	0.752
SS07	0.107	0.117	-0.150	0.146	0.139	0.328	0.181	0.182	0.361	0.443
$\text{SME}_{a,K=2}$	-0.004	-0.018	0.014	0.062	0.026	0.139	0.062	0.032	0.139	0.156
$\text{SME}_{b,K=2}$	0.035	-0.061	-0.024	0.065	0.030	0.138	0.074	0.068	0.140	0.172
$\text{SME}_{a,K=4}$	0.002	-0.002	0.005	0.073	0.031	0.175	0.073	0.031	0.175	0.192
$\text{SME}_{b,K=4}$	0.014	-0.002	-0.025	0.062	0.031	0.154	0.063	0.031	0.156	0.172

$\sigma_\varepsilon/\sigma_{X^*} \approx 0.45$, $Y_i = \rho(X_i^*, \theta_0) + U_i$, $U_i \sim N(0, 1/4)$

Monte Carlo: Probit Model

$$\rho(x,\theta) = \frac{1}{2}(1 + \text{erf}(\theta_1 + \theta_2 x)), \quad \theta_0 = (-1,2)'$$

	Bias		Std. Dev.		RMSE		
	θ_1	θ_2	θ_1	θ_2	θ_1	θ_2	All
NLLS	0.38	-0.97	0.06	0.08	0.39	0.98	1.05
SS07	0.05	-0.06	0.39	0.53	0.39	0.53	0.69
SME$_{a,K=2}$	0.11	-0.31	0.16	0.31	0.20	0.44	0.48
SME$_{b,K=2}$	0.05	-0.25	0.23	0.42	0.23	0.49	0.54
SME$_{a,K=4}$	-0.05	0.09	0.28	0.54	0.29	0.55	0.62
SME$_{b,K=4}$	-0.02	0.01	0.25	0.48	0.25	0.48	0.54

$\sigma_\varepsilon/\sigma_{X^*} \approx 0.45$, $Y_i = \rho(X_i^*,\theta_0) + U_i$, $U_i = 1 - \rho(X_i^*,\theta_0)$ w.p. $\rho(X_i^*,\theta_0)$, and $-\rho(X_i^*,\theta_0)$ o/w

Monte Carlo: Logit Design

Logit model with 4 covariates and intercept:

$$
\begin{aligned}
Y_i &= \mathbb{1}\{\theta_1 X_i^* + \theta_2 W_{1i} + \theta_3 W_{2i} + \theta_4 W_{3i} + \theta_5 + U_i > 0\}, \quad U_i \sim \mathit{Logistic}, \\
X_i &= X_i^* + \sigma_\varepsilon \varepsilon_i, \quad X_i^* = (Z_i + e_i^x)/\sqrt{2}, \\
W_{1i} &= \rho X_i^* + \sqrt{1-\rho^2} e_{1i}^w, \quad W_{2i} = e_{2i}^w, \quad W_{3i} = e_{3i}^w.
\end{aligned}
$$

where $\varepsilon_i, e_i^x, e_{1i}^w, e_{2i}^w, e_{3i}^w$ are $N(0,1)$ and independent of each other, and

$$
Z_i = \begin{cases} 1, & \text{with prob. } 0.5; \\ -1, & \text{with prob. } 0.5. \end{cases}
$$

Parameters are $(\theta_1, \theta_2, \theta_3, \theta_4, \theta_5) = (1, 0, 0.5, 0, 1)$, $n = 1000$

The moment function is $(\Lambda(\cdot) - Y_i)h_i$, where $h_i = (1, Z_i, X_i, X_i^2, X_i^3, W_{1i}, W_{2i}, W_{3i})'$.

Monte Carlo: Logit Design

	MLE					SME				
	Bias	Med bias	Std	IQR/1.35	Size	Bias	Med bias	Std	IQR/1.35	Size
					$\sigma_\varepsilon = 0.0$					
X	0.0001	0.0002	0.0259	0.0251	4.6	0.0177	0.0151	0.0356	0.0331	6.5
W_1	0.0004	-0.0003	0.0251	0.0238	5.4	-0.0112	-0.0103	0.0320	0.0302	7.2
θ_1	0.0082	0.0084	0.1353	0.1310	4.8	0.1078	0.0899	0.1911	0.1736	6.1
θ_2	0.0018	-0.0017	0.1291	0.1219	5.4	-0.0585	-0.0523	0.1668	0.1557	7.0
					$\sigma_\varepsilon = 0.4$					
X	-0.0610	-0.0607	0.0212	0.0211	81.3	-0.0097	-0.0133	0.0430	0.0426	4.9
W_1	0.0484	0.0484	0.0223	0.0212	57.9	0.0090	0.0111	0.0377	0.0380	6.4
θ_1	-0.3125	-0.3119	0.1084	0.1065	81.7	-0.0387	-0.0546	0.2258	0.2241	4.8
θ_2	0.2454	0.2429	0.1134	0.1062	57.4	0.0455	0.0569	0.1944	0.1954	6.6

Estimation and size, $\rho = 0.8$

Estimator: General Case

Assumption (Moderate Measurement Error)

$\sigma_n^2 = o\left(n^{-1/(K+1)}\right)$

$$\psi(X_i, S_i, \theta, \gamma) \equiv g(X_i, S_i, \theta) - \sum_{k=2}^{K} \gamma_k g_x^{(k)}(X_i, S_i, \theta), \quad \gamma = (\gamma_2, \ldots, \gamma_K)'$$

- Under smoothness conditions, $\mathbb{E}[\psi(X_i, S_i, \theta_0, \gamma_{0n})] = o(n^{-1/2})$
- γ_{0n} is determined by the moments of ε_{in}
- Need to ensure that the Taylor's expansion remainder is negligible
- For the polynomial specification, the expansion is exact
- Hong and Tamer (2003): if $\varepsilon_i \sim Laplace$, the expansion is exact with $K = 2$

Assumption: Moment Function

Assumption (Lipschitz-Polynomial). For some functions $b_j(x,r,\theta)$ for $j \in \{1,\ldots,J\}$ s.t., $\forall x, x' \in \mathcal{X}$ and $\forall (r,\theta) \in \mathcal{R} \times \Theta$,

$$\left\| g_x^{(K)}(x',r,\theta) - g_x^{(K)}(x,r,\theta) \right\| \leqslant \sum_{j=1}^{J} b_j(x,r,\theta)|x'-x|^j,$$

and $\mathbb{E}\left[\sup_{\theta\in\Theta} b_j(X_i^*, S_i, \theta)\right] < C$ for $j \in \{1,\ldots,J\}$

- Key to show $\mathbb{E}\left[\psi(X_i, S_i, \theta_0, \gamma_{0n})\right] = O(\sigma_n^{K+1}) = o(n^{-1/2})$
- Satisfied in most models
- When $|\varepsilon_{in}/\sigma_n|$ has M bounded moments, can allow $J = M - K$
- A similar condition is imposed on $\nabla_\theta g_x^{(K)}(x,r,\theta)$

Asymptotic Normality

- Denote $\hat{\beta} \equiv (\hat{\theta}', \hat{\gamma}')'$, $\beta_{0n} \equiv (\theta_0', \gamma_{0n}')'$

Theorem (Asymptotic Normality)

Under standard assumptions,

$$n^{1/2}\Sigma^{-1/2}(\hat{\beta} - \beta_{0n}) \xrightarrow{d} N(0, I_{p+K-1}),$$

$$\Sigma = (\Psi^{*\prime}\Xi\Psi^*)^{-1}\Psi^{*\prime}\Xi\Omega^*\Xi\Psi^*(\Psi^{*\prime}\Xi\Psi^*)^{-1}$$

$$\Omega^* \equiv \mathbb{E}\left[g(X_i^*, S_i, \theta_0)g(X_i^*, S_i, \theta_0)'\right], \quad \Psi^* \equiv \mathbb{E}\left[\nabla_\beta \psi(X_i^*, S_i, \theta_0, 0)\right]$$

- For asymptotic normality, $\gamma_{0n} \to \gamma_0 = 0 \in \text{int}\,(\Gamma)$ or $n^{1/2}\gamma_{0n} \to \infty$
- Σ can be consistently estimated, the standard inference tools apply

$$n^{-1}\sum_{i=1}^{n}\psi_i(\hat{\beta})\psi_i(\hat{\beta})' = \Omega^* + o_{p,n}(1), \quad n^{-1}\sum_{i=1}^{n}\nabla_\beta\psi_i(\hat{\beta}) = \Psi^* + o_{p,n}(1)$$

Boundary "Problem"

- $\gamma_{0n} = \sigma_n^2/2 \geqslant 0$, reasonable to restrict $\gamma \in [0, \overline{\gamma}]$
- We model $\gamma_{0n} \to 0$
- Asymptotic normality of $\hat{\theta}$ (and $\hat{\gamma}$) only if
  - Modify the parameter space: int $(\Gamma) \ni 0$
  - Or hope σ_n^2 is large enough: $n^{1/2}\sigma_n^2 \to \infty$
- Under strong ID, an asymptotically normal estimator of θ_0 is available, and the standard tests are valid but no longer (locally) efficient
- Make use of the tests developed by Elliot, Müller, and Watson (2015) – this approach exploits that $\gamma_{0n} \geq 0$ and is nearly optimal under strong ID, and of Ketz (2014).
- In nonparametric models: $f_\varepsilon(x) \to \delta(x)$

Summary

- $\sigma_\varepsilon \to 0$ is a better approximation
- Can make use of instruments (even discrete!) to ID the problem
- Simple estimation via GMM
- Advantageous in terms of the quality of point estimates
- Valid inference
- Handle/Leverage boundary concerns
- Feasible to handle multivariate X
- Dependence, Panels
- Non-classical measurement errors