Simple Estimation of Semiparametric Models with Measurement Errors

Kirill Evdokimov
Andrei Zeleneev

Princeton University

May 17, 2018

Framework

- General moment condition model

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

- X_i^* is mismeasured with a (classical) measurement error ε_i

$$X_i = X_i^* + \varepsilon_i, \quad \varepsilon_i \perp (X_i^*, R_i), \quad \mathbb{E}[\varepsilon_i] = 0$$

- Leading example: nonlinear regression

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

- For example, Logit/Probit/Tobit

Framework: Small and Moderate Measurement Error

- 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 is 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 suggest an alternative asymptotic framework modeling $\sigma_n^2 \equiv \mathbb{E}\left[\varepsilon_{in}^2\right]$ as shrinking to zero
  - Simple estimation procedure: GMM
  - Advantageous in terms of the quality of point estimates
  - ID robust and powerful inference

Estimation: Motivation

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

$$
\begin{aligned}
g(X_i, R_i, \theta) &= g(X_i^*, R_i, \theta) + g_x^{(1)}(X_i^*, R_i, \theta)\varepsilon_{in} + \frac{1}{2} g_x^{(2)}(X_i^*, R_i, \theta)\varepsilon_{in}^2 + O_p(\sigma_n^3), \\
\mathbb{E}[g(X_i, R_i, \theta)] &= \mathbb{E}[g(X_i^*, R_i, \theta)] \qquad + \frac{\sigma_n^2}{2}\mathbb{E}[g_x^{(2)}(X_i^*, R_i, \theta)] + O(\sigma_n^3),
\end{aligned}
$$

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

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

- 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 (Small Measurement Error)

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

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

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

- Higher order expansion is a bit more tricky: one needs to correct the correction terms. Details are in the paper.

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
SME	-0.00	-0.05	0.00	0.07	0.09	0.22	0.10	0.08	0.09	0.22	0.10	0.10	0.28

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

Monte Carlo: Rational Fraction Specification

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

	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
SME	0.023	-0.004	-0.031	0.056	0.029	0.123	0.060	0.029	0.127	0.143

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

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	-0.09	0.15	0.28	0.56	0.29	0.58	0.65

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

Estimator: General Case

Assumption (Small/Moderate Measurement Error)

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

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

- Under smoothness conditions, $\mathbb{E}[\psi(X_i, R_i, \theta_0, \gamma_{0n})] = o(n^{-1/2})$
- γ_{0n} is determined by the moments of ε_{in}
- Need to ensure that the Taylor's expansion reminder 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 $\mathbb{E}\left[\sup_{\theta \in \Theta} b_j(X_i^*, R_i, \theta)\right] < C$ for $j \in \{1, \ldots, J\}$

- Key to show $\mathbb{E}\left[\psi(X_i, R_i, \theta_0, \gamma_{0n})\right] = O(\sigma_n^{K+1}) = o(n^{-1/2})$
- Satisfied in the most of empirically relevant models including Logit/Tobit/Probit
- If M moments of $|\varepsilon_{in}/\sigma_n|$ exist, $J = M - K$ is allowed
- A similar condition is also 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 (for the strongly identified models),

$$n^{1/2}\Omega^{-1/2}(\hat{\beta} - \beta_{0n}) \xrightarrow{d} N(0, I_{p+K-1}),$$
$$\Omega = (\Psi^{*\prime}\Xi\Psi^*)^{-1}\Psi^{*\prime}\Xi\Sigma^*\Xi\Psi^*(\Psi^{*\prime}\Xi\Psi^*)^{-1}$$
$$\Sigma^* \equiv \mathbb{E}\left[g(X_i^*, R_i, \theta_0)g(X_i^*, R_i, \theta_0)'\right], \quad \Psi^* \equiv \mathbb{E}\left[\nabla_\beta \psi(X_i^*, R_i, \theta_0, 0)\right]$$

- Strong ID: $\lambda_{min}\left(\Psi^{*\prime}\Xi\Psi^*\right) > C > 0$
- For asymptotic normality, $\gamma_{0n} \to \gamma_0 = 0 \in \text{int}\,(\Gamma)$ or $n^{1/2}\gamma_{0n} \to \infty$
- Ω can be consistently estimated, the standard inference tools apply

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

Summary

1. Simple estimation for nonlinear models
   - GMM
2. In particular can handle
   - panel data models
   - weakly-dependent data
   - serially correlated dependent
   - non-classical measurement errors
3. Inference is simply GMM inference (standard/textbook procedures apply)

Summary

1. Simple estimation for nonlinear models
   - GMM
2. In particular can handle
   - panel data models
   - weakly-dependent data
   - serially correlated dependent
   - non-classical measurement errors
3. Inference is simply GMM inference (standard/textbook procedures apply)
4. In practice, the measurement error can be very small, relative to the sampling variability. And we should not rule out the possibility that it is absent.
   - Econometricians/Inference methods should take this into account
   - We address this concern as the parameter on the boundary problem: $\sigma_\varepsilon^2 \geqslant 0$

Advertising another paper of ours:

Issues of Nonstandard Inference in Measurement Error Models

- This paper points out that the nuisance parameter (measurement error distribution) can be weakly ID-ed or not ID-ed even if the instruments are strong

Example: GLM

$$
\begin{aligned}
Y_i &= \rho(\theta_{01} X_i^* + \theta_{02}' W_i) + U_i, \quad \mathbb{E}\left[U_i | X_i^*, W_i, Z_i\right] = 0 \\
X_i^* &= m(Z_i) + \eta_i \\
X_i &= X_i^* + \varepsilon_i = m(Z_i) + \eta_i + \varepsilon_i
\end{aligned}
$$

- If $\theta_{01} = 0$, $f_\varepsilon(\cdot)$ (and σ_ε^2) are not identified!
  - this is a *feature of the problem*, not of an estimation method.
- Not even clear if the existing estimators are consistent or what are the rates of convergence
- This paper:
  - Establishes uniform $\sqrt{n}$-consistency of $\hat{\theta}$ without assuming identifiability of nuisance parameter
  - Develops a simple (but powerful and ID robust) inference procedure