Primer on High-Order Moment Estimators

Toni M. Whited

July 2007

The Errors-in-Variables Model

- We will start with the classical EIV for one mismeasured regressor. The general case is in Erickson and Whited *Econometric Theory*, 2002.

$$\begin{aligned} y_i &= z_i\alpha + \chi_i\beta + u_i \\ x_i &= \gamma + \chi_i + \varepsilon_i \end{aligned}$$

- Definitions:
  - y_i is the dependent variable,
  - z_i is a **vector** of perfectly measured regressors (which includes an intercept)
  - χ_i is the mismeasured regressor
  - u_i is the regression disturbance
  - ε_i is the measurement error
  - (α, β, γ) are coefficients

Assumptions

- $(u_i, \varepsilon_i, \chi_i, z_i)$,are $i.i.d.$
- $u_i,\ \varepsilon_i,\ \chi_i$, and z_i have finite moments of every order
- u_i and ε_i are distributed **independently** of each other and of (χ_i, z_i)
- $E(u_i) = E(\varepsilon_i) = 0$
- var (χ_i, z_i) is positive definite.
- $\beta \neq 0$ and η_i is non-normally distributed.

Partialling

- Because these estimators are based on polynomials, partialling out the perfectly measured regressors is essential:

$$\begin{aligned} y_i - z_i\mu_y &= \eta_i\beta + u_i \\ x_i - z_i\mu_x &= \eta_i + \varepsilon_i \end{aligned}$$

- in which

$$\begin{aligned} \mu_x &= E\left[\left(z_i z_i'\right)^{-1}\left(z_i x_i\right)\right] \\ \mu_y &= E\left[\left(z_i z_i'\right)^{-1}\left(z_i y_i\right)\right] \\ \eta_i &= \chi_i - z_i\mu_x \end{aligned}$$

Moment Equations

- Second-Order Moment Equations

$$
\begin{aligned}
E\left[(y_i - z_i\mu_y)^2\right] &= \beta^2 E(\eta_i^2) + E(u_i^2) \\
E\left[(y_i - z_i\mu_y)(x_i - z_i\mu_x)\right] &= \beta E(\eta_i^2) \\
E\left[(x_i - z_i\mu_x)^2\right] &= E(\eta_i^2) + E(\varepsilon_i^2)
\end{aligned}
$$

- Third-Order Moment Equations

$$
\begin{aligned}
E\left[(y_i - z_i\mu_y)^2 (x_i - z_i\mu_x)\right] &= \beta^2 E(\eta_i^3) \\
E\left[(y_i - z_i\mu_y)(x_i - z_i\mu_x)^2\right] &= \beta E(\eta_i^3)
\end{aligned}
$$

- Third-Order Moment Estimator

$$
\beta = E\left[(y_i - z_i\mu_y)^2 (x_i - z_i\mu_x)\right] \Big/ E\left[(y_i - z_i\mu_y)(x_i - z_i\mu_x)^2\right]
$$

More Moment Equations

- Fourth Order Moment Equations

$$\begin{aligned} E\left[(y_i - z_i\mu_y)^3 (x_i - z_i\mu_x)\right] &= \beta^3 E\left(\eta_i^4\right) + 3\beta E\left(\eta_i^2\right) E\left(u_i^2\right) \\ E\left[(y_i - z_i\mu_y)^2 (x_i - z_i\mu_x)^2\right] &= \beta^2 \left[E\left(\eta_i^4\right) + E\left(\eta_i^2\right) E\left(\varepsilon_i^2\right)\right] + E\left(u_i^2\right) \left[E\left(\eta_i^2\right) + E\left(\varepsilon_i^2\right)\right] \\ E\left[(y_i - z_i\mu_y) (x_i - z_i\mu_x)^3\right] &= \beta \left[E\left(\eta_i^4\right) + 3E\left(\eta_i^2\right) E\left(\varepsilon_i^2\right)\right]. \end{aligned}$$

- General Formula

$$\begin{aligned} E\left[(y_i - z_i\mu_y)^r (x_i - z_i\mu_x)^{m-r}\right] &= E\left[(\eta_i\beta + u_i)^r (\eta_i + \varepsilon_i)^{m-r}\right] \\ &= \sum_{j=0}^{r} \sum_{k=0}^{m-r} \binom{r}{j} \binom{m-r}{k} \beta^{r-j} E\left(u_i^j\right) E\left(\varepsilon_i^k\right) E\left(\eta_i^{m-j-k}\right) \end{aligned}$$

Identification

- Notice from the third-order moment equations

$$
\begin{aligned}
E\left[(y_i - z_i\mu_y)^2 (x_i - z_i\mu_x)\right] &= \beta^2 E\left(\eta_i^3\right) \\
E\left[(y_i - z_i\mu_y)(x_i - z_i\mu_x)^2\right] &= \beta E\left(\eta_i^3\right)
\end{aligned}
$$

  that we can only solve for β if $\beta \neq 0$ and if $E\left(\eta_i^3\right) \neq 0$.

- These are the two identifying assumptions.
- They can be tested simply by testing whether the two left-hand-side quantities are zero.
- In general Reiersöl (1950, *Econometrica*) showed that this model is identified as long as η is not normally distributed.
- It is possible to work out identification tests for symmetric data that use fourth order moments.

Difficulties with Using High Order Moments:

- High order moments cannot be estimated with as much precision as the second order moments on which conventional regression analysis is based.
- It is important that the high order moment information be used as efficiently as possible.
- Previously, the use of high order moments has required selecting an inefficient estimator.
- A more efficient estimator can be constructed via a minimum variance combination of inefficient estimators
  - A labor-intensive technique
  - No guarantee of efficiency

Aside on GMM

- Let
  - Let $\boldsymbol{w}_i$ be an $(M \times 1)$ be an $i.i.d.$ vector of random variables for observation i.
  - $\boldsymbol{\theta}$ be an $(P \times 1)$ vector of unknown coefficients.
  - $\boldsymbol{g}(\boldsymbol{w}_i, \boldsymbol{\theta})$ be an $(L \times 1)$ vector of functions $\boldsymbol{g} : \left(\mathcal{R}^M \times \mathcal{R}^P\right) \rightarrow \mathcal{R}^L, \quad L \geq P$
- The function $\boldsymbol{g}(\boldsymbol{w}_i, \boldsymbol{\theta})$ can be nonlinear.
- Let $\boldsymbol{\theta}_0$ be the true value of $\boldsymbol{\theta}$.
- Let $\hat{\boldsymbol{\theta}}$ represent an estimate of $\boldsymbol{\theta}$. The “hat” notation applies to anything we might want to estimate.

Moment Restrictions

- GMM is based on what are generally called moment restrictions and sometimes called orthogonality conditions (The latter terminology comes from the rational expectations literature.)

$$E\left(\boldsymbol{g}\left(\boldsymbol{w}_i, \boldsymbol{\theta}_0\right)\right) = 0$$

- This condition is expressed in terms of the population. The corresponding sample moment restriction is

$$\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{g}\left(\boldsymbol{w}_i, \boldsymbol{\theta}\right) = 0$$

- What we want to do is choose $\hat{\boldsymbol{\theta}}$ to get $N^{-1}\sum_{i=1}^{N}\boldsymbol{g}\left(\boldsymbol{w}_i, \boldsymbol{\theta}\right)$ as close to zero as possible.

Criterion Function

- We minimize a quadratic form:

$$\boldsymbol{Q}_N(\boldsymbol{\theta}) = \underset{(1\times L)}{\left[N^{-1}\sum_{i=1}^{N}\boldsymbol{g}(\boldsymbol{w}_i,\boldsymbol{\theta})\right]'} \underset{(L\times L)}{\widehat{\boldsymbol{\Xi}}} \underset{(L\times 1)}{\left[N^{-1}\sum_{i=1}^{N}\boldsymbol{g}(\boldsymbol{w}_i,\boldsymbol{\theta})\right]}$$

  that converges in probability to

$$\{E[\boldsymbol{g}(\boldsymbol{w}_i,\boldsymbol{\theta})]\}'\boldsymbol{\Xi}\{E[\boldsymbol{g}(\boldsymbol{w}_i,\boldsymbol{\theta})]\}$$

- If $L = P$, then the estimator is exactly identified, and we can find $\boldsymbol{\theta}$ by solving

$$N^{-1}\sum_{i=1}^{N}\boldsymbol{g}(\boldsymbol{w}_i,\boldsymbol{\theta}) = \mathbf{0}$$

- If $L > P$, the model is overidentified and if it is nonlinear, you usually have to use numerical techniques.

Our Alternative GMM Estimator

High Order Moments

- We combine the information in the high order moments by using GMM, which is computationally convenient and efficient.
- In our application of GMM, we have:

$$
\begin{aligned}
\boldsymbol{w}_i &\equiv (y_i - z_i\mu_y, x_i - z_i\mu_x) \\
\boldsymbol{\theta} &\equiv (\beta, E(\eta_i^2), E(u_i^2), E(\varepsilon_i^2), E(\eta_i^3), \ldots) \\
\boldsymbol{g}(\boldsymbol{w}_i, \boldsymbol{\theta}) &\equiv (y_i - z_i\mu_y)^2 - \beta^2 E(\eta_i^2) + E(u_i^2) \\
&\quad \vdots \\
&\quad (y_i - z_i\mu_y)(x_i - z_i\mu_x)^2 - \beta E(\eta_i^3) \\
&\quad \vdots
\end{aligned}
$$

Our Alternative GMM Estimator

- Because all of the observables are on the left and all of the unobservables are on the right, we can just use the covariance of the observable moments as the weight matrix.
- The weight matrix does not depend on any parameters. No iterating!
- Two-step procedure: estimate μ_x and μ_y with OLS, plug these estimates into the above and apply GMM. (What else could you do?)
- Because we substitute OLS estimates of μ_x and μ_y in for their true unknown values, we have to adjust the weighting matrix.

An Aside on Two-Step GMM Estimators

- Suppose that you estimate a parameter vector $\boldsymbol{\delta}$ of dimension S via a different procedure, and then plug this estimate into a GMM estimator. How do you calculate the variance of the original parameter vector $\boldsymbol{\theta}$?

- The variance of the two-step estimator is

$$\left(\boldsymbol{G}\boldsymbol{\Omega}^{-1}\boldsymbol{G}'\right)^{-1}$$

- You can estimate Ω by

$$\widehat{\boldsymbol{\Omega}} \equiv \frac{1}{N}\sum_{i=1}^{N}\left[\boldsymbol{g}\left(\boldsymbol{w}_i,\boldsymbol{\theta}\right) - \frac{\partial \boldsymbol{g}(\boldsymbol{\theta},\boldsymbol{w}_i,\boldsymbol{\delta})}{\partial \boldsymbol{\delta}}\phi^{\boldsymbol{\delta}}(\boldsymbol{\delta},\boldsymbol{w}_i)\right]\left[\boldsymbol{g}\left(\boldsymbol{w}_i,\boldsymbol{\theta}\right) - \frac{\partial \boldsymbol{g}(\boldsymbol{\theta},\boldsymbol{w}_i,\boldsymbol{\delta})}{\partial \boldsymbol{\delta}}\phi^{\boldsymbol{\delta}}(\boldsymbol{\delta},\boldsymbol{w}_i)\right]'$$

  in which $\phi^{\boldsymbol{\delta}}$ is the influence function for $\boldsymbol{\delta}$.

- A clear derivation of this estimator is in Newey and McFadden's chapter in the 4^{th} volume of the *Handbook of Econometrics*.

Other Things to Estimate

- The R^2 of the “true” regression

$$\rho^2 = \frac{\mu_y' V_{zz} \mu_y + E\left(\eta_i^2\right) \beta^2}{\mu_y' V_{zz} \mu_y + E(\eta_i^2)\, \beta^2 + E(u_i^2)}$$

- The R^2 of the measurement equation

$$\tau^2 = \frac{\mu_x' V_{zz} \mu_x + E\left(\eta_i^2\right)}{\mu_x' V_{zz} \mu_x + E(\eta_i^2) + E(\varepsilon^2)}$$

- The vector of perfectly measured regressors

$$\alpha = \mu_y - \mu_x \beta$$

Standard Errors

- Because we substitute OLS estimates of μ_x and μ_y in for their true unknown values, we have to adjust the weighting matrix.
- You calculate the standard errors for these things by stacking the influence functions for their individual components and using the delta method.
- Recall that the influence function for a GMM estimator is

$$-\left(\boldsymbol{G}\boldsymbol{\Xi}\boldsymbol{G}\right)^{-1}\boldsymbol{G}\boldsymbol{\Xi}E\left[\boldsymbol{g}\left(\boldsymbol{w}_i,\bar{\boldsymbol{\theta}}\right)\right]$$

- Recall that many estimators fall under the umbrella of GMM.
- This formula can be used to calculate the influence functions of the components of the three things on the previous slide—μ_y, V_{zz}, …, and all of the GMM parameters.
- To calculate their joint covariance matrix, stack their influence functions and take the outer product.
- Because τ^2, α, and ρ^2 are nonlinear functions of their components, use the delta method.

Generalization and Identification

- This method can be used for multiple mismeasured regressors. You need much more data for the multiple mismeasured regressor case than for the single mismeasured regressor case.

- The moment conditions can be written in general as:

$$E\left[(y_i - z_i\mu_y)^{r_0} \prod_{j=1}^{J} (x_{ij} - z_i\mu_{xj})^{r_j}\right] = E\left[\left(\sum_{j=1}^{J} \eta_{ij}\beta_j + u_i\right)^{r_0} \prod_{j=1}^{J} (\eta_{ij}+\varepsilon_{ij})^{r_j}\right],$$

  in which $(r_0, r_1, \ldots, r_J)$ are nonnegative integers.

- This general model is identified (loosely) if all of the coefficients on the mismeasured regressors are nonzero and if at least one of the mismeasured regressors has a skewed distribution.

- The identification assumptions necessary for this model are analogous to the assumption of noncollinearity in an OLS model.

Finite Sample Performance

TABLE 1. OLS and GMM on the baseline DGP, $n = 1{,}000$

	OLS	GMM3	GMM4	GMM5	GMM6	GMM7
$E(\hat{\beta})$	0.387	1.029	1.000	0.998	0.993	0.995
$MAE(\hat{\beta})$	0.613	0.196	0.117	0.118	0.116	0.106
$P(\lvert\hat{\beta}-\beta\rvert \le 0.15)$	0.000	0.596	0.732	0.739	0.778	0.797
Size of t-test	—	0.066	0.126	0.162	0.247	0.341
$E(\hat{\alpha}_1)$	−0.845	−1.008	−1.000	−0.999	−1.000	−0.999
$MAE(\hat{\alpha}_1)$	0.155	0.069	0.055	0.055	0.057	0.054
$P(\lvert\hat{\alpha}_1-\alpha_1\rvert \le 0.15)$	0.068	0.917	0.959	0.963	0.966	0.965
Size of t-test	—	0.060	0.072	0.076	0.081	0.088
$E(\hat{\alpha}_2)$	1.155	0.994	1.001	1.001	1.003	1.003
$MAE(\hat{\alpha}_2)$	0.155	0.068	0.055	0.055	0.055	0.053
$P(\lvert\hat{\alpha}_2-\alpha_2\rvert \le 0.15)$	0.068	0.920	0.961	0.963	0.966	0.969
Size of t-test	—	0.059	0.066	0.074	0.078	0.080
$E(\hat{\alpha}_3)$	−0.846	−1.009	−1.001	−1.001	−1.000	−1.000
$MAE(\hat{\alpha}_3)$	0.154	0.069	0.055	0.055	0.055	0.053
$P(\lvert\hat{\alpha}_3-\alpha_3\rvert \le 0.15)$	0.068	0.918	0.962	0.962	0.967	0.966
Size of t-test	—	0.058	0.069	0.070	0.076	0.082
$E(\hat{\rho}^2)$	0.546	0.675	0.695	0.710	0.723	0.734
$MAE(\hat{\rho}^2)$	0.122	0.064	0.053	0.060	0.067	0.074
$P(\lvert\hat{\rho}^2-\rho^2\rvert \le 0.15)$	0.706	0.937	0.982	0.979	0.969	0.953
Size of t-test	—	0.110	0.155	0.253	0.371	0.509
Size of J-test	—	—	0.036	0.073	0.161	0.280

Finite Sample Performance

TABLE 6. OLS and GMM with two mismeasured regressors: Baseline DGP with an additional measurement error, $n = 1{,}000$

	OLS	GMM3E	GMM3O	GMM4
$E(\hat{\beta}_1)$	0.363	1.035	0.994	0.968
$MAE(\hat{\beta}_1)$	0.637	0.254	0.204	0.179
$P(\lvert\hat{\beta}_1 - \beta_1\rvert \leq 0.15)$	0.000	0.566	0.607	0.667
Size of t-test	—	0.074	0.102	0.236
$E(\hat{\beta}_2)$	−0.606	−0.996	−0.989	−0.973
$MAE(\hat{\beta}_2)$	0.394	0.155	0.155	0.084
$P(\lvert\hat{\beta}_2 - \beta_2\rvert \leq 0.15)$	0.000	0.740	0.755	0.908
Size of t-test	—	0.072	0.082	0.200
$E(\hat{\alpha}_1)$	−0.916	−1.012	−1.001	−0.997
$MAE(\hat{\alpha}_1)$	0.086	0.110	0.099	0.084
$P(\lvert\hat{\alpha}_1 - \alpha_1\rvert \leq 0.15)$	0.785	0.840	0.853	0.912
Size of t-test	—	0.076	0.095	0.111
$E(\hat{\alpha}_2)$	1.083	0.988	0.999	1.001
$MAE(\hat{\alpha}_2)$	0.085	0.109	0.098	0.084
$P(\lvert\hat{\alpha}_2 - \alpha_2\rvert \leq 0.15)$	0.785	0.842	0.859	0.914
Size of t-test	—	0.079	0.097	0.112
$E(\hat{\rho}^2)$	0.503	0.673	0.668	0.703
$MAE(\hat{\rho}^2)$	0.164	0.066	0.063	0.057
$P(\lvert\hat{\rho}^2 - \rho^2\rvert \leq 0.15)$	0.416	0.927	0.937	0.979
Size of t-test	—	0.100	0.123	0.230
Size of J-test	—	—	0.047	0.097