EC708 Discussion 5
GMM

Yan Liu

Department of Economics
Boston University

February 18, 2022

Outline

1. Hypothesis Testing with GMM
2. Finite Sample Properties of GMM
3. Finite Sample Properties of GMM-Based Wald Tests
4. GMM in Asset Pricing

Hypothesis Testing with GMM

Over-identification test

Test for moment validity in over-identified models:

$$H_0 : \mathbb{E}[m(X_t, \theta_0)] = 0, \quad H_1 : \mathbb{E}[m(X_t, \theta)] \neq 0, \ \forall \theta \in \Theta.$$

Hansen-Sargan's J-test: Let $\hat{\theta}_T$ be an efficient GMM estimator.

$$J_T = T g_T(\hat{\theta}_T)' \hat{S}_T^{-1} g_T(\hat{\theta}_T) \xrightarrow{d} \chi^2_{k-q} \quad \text{under } H_0,$$

where $k = \dim(m)$ and $q = \dim(\theta)$.

- Rejecting H_0 doesn't tell you which moments are invalid.
- Rejecting H_0 doesn't necessarily mean moments are invalid. It could also be model misspecification.
- In small samples, J-test often overrejects.

Hypothesis Testing with GMM

Test subsets of moment conditions

Partition

$$m(X_t, \theta_0) = \begin{pmatrix} \underset{k_1 \times 1}{m_1(X_t, \theta_0)} \\ \underset{k_2 \times 1}{m_2(X_t, \theta_0)} \end{pmatrix}, \quad S_0 = \begin{pmatrix} S_{0,11} & S_{0,12} \\ S_{0,21} & S_{0,22} \end{pmatrix},$$

and $g_T, \hat{S}_T, G_0, G_T$ conformably. We are interested in testing

$$H_0 : \mathbb{E}[m(X_t, \theta_0)] = 0, \quad H_1 : \mathbb{E}[m_1(X_t, \theta_0)] \neq 0,\ \mathbb{E}[m_2(X_t, \theta_0)] = 0.$$

Let $\hat{\theta}_T$ and $\tilde{\theta}_T$ denote optimal two-step GMM estimators using all moment conditions and using just $m_2(X_t, \theta)$, respectively:

$$\hat{\theta}_T = \arg\min_{\theta} g_T(\theta)' \hat{S}_T^{-1} g_T(\theta), \quad \tilde{\theta}_T = \arg\min_{\theta} g_{T,2}(\theta)' \hat{S}_{T,22}^{-1} g_{T,2}(\theta).$$

Hausman test:

- Under H_0, both estimators are consistent while $\hat{\theta}_T$ is more efficient;
- Under H_1, only $\tilde{\theta}_T$ is consistent.

If $k_1 \leq q$, a Hausman test can be formed as

$$H_T = (\tilde{\theta}_T - \hat{\theta}_T)'[\tilde{V}_T - \hat{V}_T]^{-1}(\tilde{\theta}_T - \hat{\theta}_T) \xrightarrow{d} \chi^2_{k_1} \quad \text{under } H_0,$$

where $\hat{V}_T = [G_T(\hat{\theta}_T)'\hat{S}_T^{-1}G_T(\hat{\theta}_T)]^{-1}, \tilde{V}_T = [G_{T,2}(\tilde{\theta}_T)'\hat{S}_{T,22}^{-1}G_{T,2}(\tilde{\theta}_T)]^{-1}$.

Hypothesis Testing with GMM

Test subsets of moment conditions

An alternative formulation is the **GMM specification tests** (Newey, 1985). Let $\tilde{g}_T = g_{T,1}(\hat{\theta}_T) - \hat{S}_{T,12}\hat{S}_{T,22}^{-1}g_{T,2}(\hat{\theta}_T)$ and $\tilde{S}_T$ be an estimator of its asymptotic variance. The optimal GMM test is

$$M_T = T\tilde{g}_T'\tilde{S}_T^{-1}\tilde{g}_T \xrightarrow{d} \chi^2_{k_1} \quad \text{under } H_0.$$

In fact, H_T and M_T are asymptotically equivalent.

We are interested in testing

$$H_0 : \underset{r\times 1}{h(\theta_0)} = 0, \quad \text{where } r < k.$$

Let $\hat{\theta}_T$ and $\tilde{\theta}_T$ denote the unrestricted and restricted GMM estimator:

$$\hat{\theta}_T = \arg\min_{\theta} g_T(\theta)' \hat{S}_T^{-1} g_T(\theta), \quad \tilde{\theta}_T = \arg\min_{h(\theta)=0} g_T(\theta)' \hat{S}_T^{-1} g_T(\theta).$$

Let $\hat{G}_T = G_T(\hat{\theta}_T) = \partial g_T(\hat{\theta}_T)/\partial \theta'$ and $\tilde{G}_T = G_T(\tilde{\theta}_T)$.

Hypothesis Testing with GMM

Test parametric hypothesis

Wald test: Let the Jacobian $J_h(\theta) = \frac{\partial h}{\partial \theta'}$. By the delta method, under H_0,

$$W_T = Th(\hat{\theta}_T)' \left[J_h(\hat{\theta}_T)'(\hat{G}_T' \hat{S}_T^{-1} \hat{G}_T)^{-1} J_h(\hat{\theta}_T) \right]^{-1} h(\hat{\theta}_T) \xrightarrow{d} \chi_r^2.$$

Likelihood-ratio test:

$$LR_T = Tg_T(\tilde{\theta}_T)' \hat{S}_T^{-1} g_T(\tilde{\theta}_T) - Tg_T(\hat{\theta}_T)' \hat{S}_T^{-1} g_T(\hat{\theta}_T).$$

Lagrange multiplier test/Rao's score test: Let $Q_T(\theta) = g_T(\theta) \hat{S}_T^{-1} g_T(\theta)$.

$$\begin{aligned} LM_T &= T \left(\frac{\partial Q_T(\tilde{\theta}_T)}{\partial \theta} \right)' \left(-\frac{\partial^2 Q_T(\tilde{\theta}_T)}{\partial \theta \partial \theta'} \right)^{-1} \left(\frac{\partial Q_T(\tilde{\theta}_T)}{\partial \theta} \right) \\ &= Tg_T(\tilde{\theta}_T)' \hat{S}_T^{-1} \tilde{G}_T (\tilde{G}_T' \hat{S}_T^{-1} \tilde{G}_T)^{-1} \tilde{G}_T' \hat{S}_T^{-1} g_T(\tilde{\theta}_T). \end{aligned}$$

Hypothesis Testing with GMM

Test parametric hypothesis

- W_T, LR_T, and LM_T are asymptotically equivalent $\Rightarrow$ "trinity"
- Asymptotic approximation is often more accurate for LR_T and LM_T than W_T.

Continuous Updating GMM (CUGMM):

Hansen, Heaton, and Yaron (1996) consider a one-step algorithm:

$$\hat{\theta}_{CUGMM} = \arg\min_{\theta \in \Theta} \; m_T(\theta)' \hat{S}_T^{-1}(\theta) m_T(\theta).$$

- CUGMM is invariant to how moment conditions are scaled even when scale factors are parameter-dependent while two-step GMM is not.
- CUGMM is not used often because its criterion function is not quadratic and thus numerically hard to solve.

Weak identification:

Minimal eigenvalue of $G_0'G_0$ is close to zero, relative to sampling error.

- CUGMM can be used for robust inference under weak identification: for any θ_0 such that $\mathbb{E}[m(X_t, \theta_0)] = 0$,

$$T(\theta_0) = Tm_T(\theta_0)'\hat{S}_T^{-1}(\theta_0)m_T(\theta_0) \xrightarrow{d} \chi_k^2.$$

- Invert $T(\theta)$ to form a level $1-\alpha$ confidence region:

$$CR = \{\theta : T(\theta) \leq \chi^2_{k,1-\alpha}\},$$

  where $\chi^2_{k,1-\alpha}$ is the $1-\alpha$ quantile of a χ_k^2 distribution.
- In the just-identified case, this approach performs well.
- In the over-identified case, can improve by employing other statistics (Andrews and Mikusheva, 2016).

Finite Sample Properties of GMM

- $\underset{k\times k}{S_0}$ has $\frac{k(k+1)}{2}$ distinct components. $\Rightarrow$ The asymptotic approximation ignores the variation coming from estimating them!
- In finite samples, the distribution of two-step GMM estimators is often affected by the sample variation of the estimation of S_0^{-1}.
- Inference based on the asymptotic approximation may be misleading when T is not large relative to $\frac{k(k+1)}{2}$.

Finite Sample Properties of GMM

Altonji and Segal (1996)

Altonji and Segal (1996) examine the small-sample properties of the two-step GMM.

- To isolate the weighting procedure as the sole source of bias, they focus on linear models $m = X\theta + \varepsilon$.
- They compare the OLS estimator (equally-weighted)

$$\theta_{OLS} = \arg\min_{\theta}(m - X'\theta)'(m - X'\theta) = (X'X)^{-1}(X'm)$$

and the GLS estimator (optimally-weighted)

$$\theta_{GLS} = \arg\min_{\theta}(m - X'\theta)'\hat{\Omega}^{-1}(m - X'\theta) = (X'\hat{\Omega}^{-1}X)^{-1}(X'\hat{\Omega}^{-1}m),$$

where $\hat{\Omega}$ is a consistent estimate of the covariance matrix of m.

Finite Sample Properties of GMM

Altonji and Segal (1996)

They find

- GLS is seriously downward biased in absolute value in small samples for most distributions and in relatively large samples for poorly behaved distributions such as lognormal.
- GLS is almost always dominated by OLS in RMSE, MAE (median absolute error), and the coverage rate for 90% confidence intervals.
- Comparison between GLS based on estimated $\hat{\Omega}$ and true Ω indicates that there is a large cost to having to estimate the weighting matrix that overwhelms the asymptotic efficiency gain.

Finite Sample Properties of GMM

Altonji and Segal (1996)

A debiased estimator?

- The bias in GLS arises because of a correlation between sample moments and estimated $\hat{\Omega}$.
- Altonji and Segal (1996) propose an alternative split-sample estimator. Randomly partition the sample into G groups of equal size. m_g uses only data in group g and $\hat{\Omega}_{(g)}$ uses data excluding group g. Namely,

$$\theta_{GLS(G)} = \frac{1}{G}\sum_{g=1}^{G} \arg\min_{\theta} (m_g - X\theta)'\hat{\Omega}_{(g)}^{-1}(m_g - X\theta)$$

- Sample-splitting removes the bias but worsens performance in terms of RMSE, MAE, and confidence interval coverage rates.

Finite Sample Properties of GMM

Altonji and Segal (1996)

Takeaway:

- OLS is almost always preferable to using GLS when the optimal weighting matrix is unknown and unconstrained, especially when bias is an important concern.
- Researchers should estimate models by both GLS and OLS, or both GLS and split-sample GLS, and worry about bias in GLS if the parameter estimates differ substantially.

Finite Sample Properties of GMM-Based Wald Tests

Burnside and Eichenbaum (1996)

Burnside and Eichenbaum (1996) address three questions:

1. Does the small-sample size of GMM-based tests closely approximate their asymptotic size?
2. Do joint tests of several restrictions perform as well or worse than tests of simple hypothesis?
3. How can modeling assumptions, or restrictions imposed by hypotheses themselves, be used to improve the performance of these tests?

Finite Sample Properties of GMM-Based Wald Tests

Burnside and Eichenbaum (1996)

DGP:

$$X_{it} \sim \text{i.i.d. } N(0,1), \qquad \begin{aligned} &i = 1, \ldots, J,\ J = 20 \\ &t = 1, \ldots, T,\ T = 100 \end{aligned}$$

Suppose the econometrician knows $\mathbb{E}X_{it} = 0$ and is interested in estimating the standard deviation σ_i of X_{it}.

- Moments: $\mathbb{E}(X_{it}^2 - \sigma_i^2) = 0,\ i = 1, 2, \ldots, J.$
- GMM estimator: $\hat{\sigma}_i = \left(\frac{1}{T}\sum_{t=1}^{T} X_{it}^2\right)^{1/2}.$
- Hypothesis: $H_M : \sigma_1 = \sigma_2 = \cdots = \sigma_M = 1, M \in \{1, 2, 5, 10, 20\}.$
- Wald test: $\mathcal{W}_T^M = T(\hat{\sigma} - 1)' A' (A V_T A')^{-1} A(\hat{\sigma} - 1) \xrightarrow{d} \chi^2(M)$ where $A = (I_M, 0_{M\times(J-M)}), \hat{\sigma} = (\hat{\sigma}_1, \hat{\sigma}_2, \ldots, \hat{\sigma}_J)'$.

Finite Sample Properties of GMM-Based Wald Tests

Burnside and Eichenbaum (1996)

They consider eight estimators V_T of $V_0 = (G_0' S_0^{-1} G_0)^{-1}$:

1. S_T^1, S_T^2, and S_T^3 are HAC estimators with different windows and bandwidths.
2. S_T^4 imposes serial uncorrelation. $[S_T^4]_{ij} = \frac{1}{T}\sum_{t=1}^{T}(X_{it}^2 - \hat{\sigma}_i^2)(X_{jt}^2 - \hat{\sigma}_j^2)$.
3. S_T^5 imposes mutual independence on S_T^4. S_T^5 is diagonal and $[S_T^5]_{ii} = \frac{1}{T}\sum_{t=1}^{T}(X_{it}^2 - \hat{\sigma}_i^2)^2$.
4. S_T^6 imposes Gaussian distribution on S_T^5. $[S_T^6]_{ii} = 2\hat{\sigma}_i^4$.
5. S_T^7 imposes H_M on S_T^6. $[S_T^7]_{ii} = 2$ for $i \leq M$.
6. V_T^8 also imposes H_M on the sample Jacobian G_T.

Table 1. Small-Sample Performance of Tests Using Gaussian White-Noise Data

Asymptotic size	Small sample size (%)				
	$M = 1$	$M = 2$	$M = 5$	$M = 10$	$M = 20$
	(a) Estimated S_T, $B_T = 4$				
1%	2.59	3.41	6.99	16.98	58.68
5%	7.49	9.25	15.61	30.92	73.37
10%	12.65	14.93	23.32	40.10	80.29
	(b) Estimated S_T, $B_T = 2$				
1%	2.31	2.87	4.83	9.17	28.88
5%	6.90	8.26	12.22	19.91	45.62
10%	12.03	13.62	19.32	28.55	55.88
	(c) Estimated S_T, B_T by Andrews procedure				
1%	2.27	2.91	4.71	9.06	26.64
5%	6.94	8.27	11.94	19.27	43.43
10%	11.98	13.50	19.04	27.87	53.83
	(d) Estimated S_T, no lags				
1%	2.15	2.73	4.17	6.67	17.31
5%	6.74	7.94	10.82	16.23	32.87
10%	11.79	13.22	17.43	24.10	42.51
	(e) Estimated diagonal S_T, no lags				
1%	2.15	2.67	3.33	3.88	4.71
5%	6.74	7.58	9.32	11.04	13.39
10%	11.79	13.04	15.50	17.56	21.20
	(f) Gaussianity applied to (e)				
1%	1.67	1.82	2.22	2.40	2.58
5%	5.94	6.08	7.20	7.72	8.53
10%	10.60	11.30	12.50	13.25	14.45
	(g) H_0 imposed on S_T in (f)				
1%	1.46	1.67	2.03	2.10	2.10
5%	4.61	5.33	5.97	6.58	7.26
10%	9.34	9.55	10.47	11.70	12.05
	(h) H_0 imposed on S_T in (f) and on D_T				
1%	.96	.97	.99	.96	.92
5%	5.16	4.90	5.08	5.01	4.99
10%	10.14	10.13	10.20	10.11	9.99

Finite Sample Properties of GMM-Based Wald Tests

Burnside and Eichenbaum (1996)

Findings from Monte Carlo experiments:

- The small-sample sizes of the tests using S_T^1, S_T^2, S_T^3, and S_T^4 exceed their asymptotic sizes and rise uniformly with M.
- Imposing the independence (S_T^5) and Gaussianity (S_T^6) assumptions improves size distortion. The impact becomes larger as M increases.
- Imposing additional restrictions from the null hypothesis (S_T^7) improves size distortion even further.
- Small-sample size distortion seems to be closely related to the small-sample distribution of S_T and, to a much smaller extent, G_T.

Finite Sample Properties of GMM-Based Wald Tests

Burnside and Eichenbaum (1996)

Takeaway:

- There is tendency for GMM-based Wald tests to overreject. The small-sample size increases uniformly as the dimension of joint tests increases.
- The problem is not resolved by nonparametric HAC estimators of the long-run covariance matrix.
- The analyst can improve size by imposing restrictions that emerge from the economic model or the hypothesis being tested when estimating the covariance matrix component of the Wald statistic.

GMM in Asset Pricing

Hansen and Singleton (1982)

A representative consumer chooses stochastic consumption and investment plans so as to maximize

$$\mathbb{E}_0\left[\sum_{t=0}^{\infty}\beta^t U(C_t)\right]$$

subject to budget constraints

$$C_t + \sum_{j=1}^{N} P_{jt}Q_{jt} \leq \sum_{j=1}^{N} R_{jt}Q_{j,t-M_j} + W_t.$$

- The consumer can invest in a collection of N assets with maturities M_j
- Q_{jt}, P_{jt}, R_{jt} denote the quantity, price, and payoff of asset j at date t
- W_t is real labor income at date t

GMM in Asset Pricing

Hansen and Singleton (1982)

First-order necessary conditions:

$$P_{jt}U'(C_t) = \beta^{M_j}\mathbb{E}_t[R_{j,t+M_j}U'(C_{t+M_j})]$$

- If asset j is stock, then $M_j = 1$, and $R_{j,t+1} = P_{j,t+1} + D_{j,t+1}$ where D_{jt} is the dividend per share.
- We also need to parameterize U as $U(\cdot,\gamma)$. E.g. CRRA preferences

$$U(C_t,\gamma) = \begin{cases} \frac{C_t^{1-\gamma}}{1-\gamma} & \gamma > 0, \gamma \neq 1 \\ \log(C_t) & \gamma = 1 \end{cases}.$$

Let's focus on the case with $M_j = 1\ \forall j$ and CRRA preferences.

GMM in Asset Pricing

Hansen and Singleton (1982)

- Let Z_t be the variables that represent the information available at date t. We can write the FOCs as

$$\mathbb{E}\left[\beta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}\frac{R_{j,t+1}}{P_{jt}}-1\middle|Z_t\right]=0.$$

- Z_t in principle could consist of infinite history. Let $B(Z_t)$ denote a vector of transformations of Z_t.
- By the law of iterated expectations,

$$\mathbb{E}\left[\left(\beta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}\frac{R_{j,t+1}}{P_{jt}}-1\right)\otimes B(Z_t)\right]=0.$$

GMM in Asset Pricing

Hansen and Singleton (1982)

We can apply the GMM approach with

$$m(X_t, \theta) = \left(\beta \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma} \frac{R_{j,t+1}}{P_{jt}} - 1 \right) \otimes B(Z_t),$$

$$X_t = (C_{t+1}/C_t, R_{1,t+1}/P_{1t}, \ldots, R_{N,t+1}/P_{Nt}, B(Z_t)')'.$$

$\{m(X_t, \theta)\}_{t=-\infty}^{\infty}$ is serially uncorrelated:

$$\mathbb{E}[m(X_t, \theta) m(X_{t-k}, \theta)'] = \mathbb{E}[\mathbb{E}[m(X_t, \theta) | X_{t-k}] m(X_{t-k}, \theta)'] = 0.$$

Hence, the optimal weighting matrix is $W = S_0^{-1}$, where

$$S_0 = \mathbb{E}[m(X_t, \theta_0) m(X_t, \theta_0)'].$$

GMM in Asset Pricing

Hansen and Singleton (1982)

Pratical issues:

- The number of technical instruments r should be relatively small compared to T.
  $\Rightarrow$ We can use economic reasoning or variable selection methods (e.g. Andrews (1999)) to figure out which instruments are the most important ones to keep.
- S_T can be singular: Many asset returns are highly correlated and we usually have small T and large N.

Bibliography

Altonji, J. G. and Segal, L. M. (1996), "Small-sample bias in GMM estimation of covariance structures," *Journal of Business & Economic Statistics*, 14, 353–366.

Andrews, D. W. (1999), "Consistent moment selection procedures for generalized method of moments estimation," *Econometrica*, 67, 543–563.

Andrews, I. and Mikusheva, A. (2016), "Conditional inference with a functional nuisance parameter," *Econometrica*, 84, 1571–1612.

Burnside, C. and Eichenbaum, M. (1996), "Small-sample properties of GMM-based Wald tests," *Journal of Business & Economic Statistics*, 14, 294–308.

Hansen, L. P., Heaton, J., and Yaron, A. (1996), "Finite-sample properties of some alternative GMM estimators," *Journal of Business & Economic Statistics*, 14, 262–280.

Hansen, L. P. and Singleton, K. J. (1982), "Generalized instrumental variables estimation of nonlinear rational expectations models," *Econometrica: Journal of the Econometric Society*, 1269–1286.

Newey, W. K. (1985), "Generalized method of moments specification testing," *Journal of econometrics*, 29, 229–256.