EC708 Discussion 3
GMM

Yan Liu

Department of Economics
Boston University

February 12, 2021

Outline

GMM Review

Introduction

Consider the following moment restriction

$$\mathbb{E}[m(X_t, \theta)] = 0.$$

- $\dim(m) = k$ (k moments);
- $\dim(\theta) = q$ (q parameters);
- If $k = q$ (just-identified), choose $\hat{\theta}$ so that the sample analog $m_T(\hat{\theta}) = \frac{1}{T}\sum_{t=1}^{T} m(X_t, \hat{\theta}) = 0.$ $\Rightarrow$ method of moments
- If $k > q$ (over-identified), choose $\hat{\theta}$ so that $m_T(\theta)$ is as close to zero as possible. $\Rightarrow$ general method of moments (GMM)
    - Notion of closeness between two $k \times 1$ vectors A, B: $\underset{1\times k}{(A-B)'}\underset{k\times k}{W}\underset{k\times 1}{(A-B)}$ where W is symmetric positive definite.

$$\hat{\theta}_{GMM}(W_T) = \underset{\theta\in\Theta}{\arg\min}\; m_T(\theta)' W_T m_T(\theta),$$

where W_T is a positive definite weighting matrix.

Asymptotic normality:

$$\sqrt{T}(\hat{\theta}_{GMM}(W_T) - \theta_0) \xrightarrow{d} N(0, V(W)),$$

$$\text{where } V(W) = (G_0' W G_0)^{-1} G_0' W S_0 W G_0 (G_0' W G_0)^{-1}$$

with $G_0 = \mathbb{E}\left[\frac{\partial}{\partial\theta'} m(X_t, \theta)\Big|_{\theta=\theta_0}\right]$, $W = \text{plim}_{T\to\infty} W_T$, and

$$S_0 = \sum_{j=-\infty}^{\infty} \mathbb{E}[m(X_t, \theta_0) m(X_{t-j}, \theta_0)'].$$

S_0 is the long-run variance-covariance matrix.

Efficient GMM:

If we choose W_T such that $\text{plim}_{T\to\infty} W_T = S_0^{-1}$, then $\hat{\theta}_{GMM}(W_T)$ has the smallest asymptotic variance-covariance matrix $V(S_0^{-1}) = (G_0' S_0^{-1} G_0)^{-1}$.

- Intuition: assign more weight to moments with smaller variances.
- Infeasible because S_0 is unknown.

Two-step GMM:

1. Obtain $\tilde{\theta}$ by GMM using identity matrix. Compute $S_T(\tilde{\theta})$.
2. Get $\hat{\theta} = \underset{\theta\in\Theta}{\arg\min}\ m_T(\theta)' S_T^{-1}(\tilde{\theta}) m_T(\theta)$.

Continuous Updating GMM (CUGMM):

Hansen et al. (1996) consider a one-step algorithm:

$$\hat{\theta}_{CUGMM} = \underset{\theta\in\Theta}{\arg\min}\ m_T(\theta)'S_T^{-1}(\theta)m_T(\theta).$$

- CUGMM is invariant to parameter dependent normalization of the moment conditions while two-step GMM is not.
- In experiments using consumption-based CAPM, CUGMM typically has less median bias but much fatter tails than two-step GMM.
- CUGMM is not used often because its criterion function is not quadratic and thus numerically hard to solve.

CUGMM can be used for robust inference under weak identification:

- Weak identification arises when the minimal eigenvalue of $G_0'G_0$ is close to zero, relative to sampling error.
- We have

$$T(\theta_0) = Tm_T(\theta_0)' S_T^{-1}(\theta_0) m_T(\theta_0) \xrightarrow{d} \chi^2(q), \ \forall \theta_0 \text{ s.t. } \mathbb{E}[m(X_t, \theta_0)] = 0.$$

  Then one can invert $T(\theta)$ to construct the confidence region.

- For the over-identified case, see Andrews and Mikusheva (2016).

HAC (Heteroskedasticity and Autocorrelation Consistent) Estimator:

$$S_T = \hat{\Gamma}_0 + \sum_{i=1}^{T-1} w(i, B_T)(\hat{\Gamma}_i + \hat{\Gamma}_i')$$

where

- $\hat{\Gamma}_i$: the ith order autocovariance matrix. $\hat{\Gamma}_{-i} = \hat{\Gamma}_i'$.
- $w(i, B_T)$: kernel (window, weight)
    - If $w(i, B_T) = 0 \; \forall i > 0$, $S_T = \hat{\Gamma}_0$ is the Eicker-Huber-White estimator.
- B_T: bandwidth
    - For consistency, we need $B_T \to \infty$ and $B_T/T \to 0$ as $T \to \infty$.

Bartlett window (Newey and West, 1987):

$$w(i, B_T) = \begin{cases} 1 - |i|/B_T & |i| \leq B_T \\ 0 & |i| > B_T \end{cases} \quad \text{"triangle"}$$

Quadratic spectral window (Andrews, 1991):

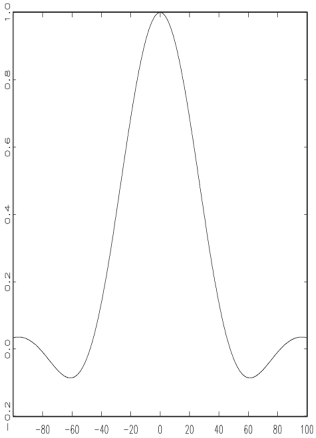
$$w(i, B_T) = 3(\sin(\delta)/\delta - \cos(\delta))/\delta^2 \quad \text{where} \quad \delta = 6\pi i/(5B_T)$$

Substantial evidence showed that the quadratic spectral window (and others) leads to estimates with better finite sample properties (lower bias and MSEs) compared to the Bartlett window. However, Stata uses the Bartlett window by default.

GMM Review

Estimation of S_0

Quadratic spectral window $(T = 100, B_T = 40)$



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 $m = X\theta + \varepsilon$.

- **Equally weighted minimum distance (EWMD):** amounts to OLS

$$\theta_{EWMD} = (X'X)^{-1}(X'm)$$

- **Optimal minimum distance (OMD):** amounts to feasible GLS

$$\theta_{OMD} = (X'\hat{\Omega}^{-1}X)^{-1}(X'\hat{\Omega}^{-1}m)$$

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

Finite Sample Properties of GMM

Altonji and Segal (1996)

Altonji and Segal (1996) conduct experiments based on independent and homoskedastic moments, moments from two different distributions, and correlated moments. They find

- OMD 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.
- OMD is almost always dominated by EWMD in RMSE, MAE (median absolute error), and the coverage rate for 90% confidence intervals.
- Comparison between OMD and "true" OMD 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)

Altonji and Segal (1996) propose an alternative split-sample estimator:

Independently weighted optimal minimum distance (IWOMD):

Randomly partition the sample into G groups of equal size. m_g uses only the data in group g and $\hat{\Omega}_{(g)}$ uses the data excluding group g. Define

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

- IWOMD is unbiased and asymptotically equivalent to OMD.
- Monte Carlo evidence indicates that IWOMD is usually dominated by EWMD in RMSE, MAE, and confidence interval coverage rates.

Finite Sample Properties of GMM

Altonji and Segal (1996)

Takeaway:

- EWMD is almost always preferable to using OMD when the optimal weighting matrix is unknown and unconstrained, especially when bias is an important concern.
- Researchers should estimate models by both OMD and EWMD, or both OMD and IWOMD, and worry about bias in OMD 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:

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

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 uses the Bartlett window with $B_T = 4$.
2. S_T^2 uses the Bartlett window with $B_T = 2$.
3. S_T^3 uses the quadratic spectral window.
4. S_T^4 imposes serial uncorrelation. $[S_T^4]_{ij} = \frac{1}{T}\sum_{t=1}^{T}(X_{it}^2 - \hat{\sigma}_i^2)(X_{jt}^2 - \hat{\sigma}_j^2)$.
5. S_T^5 imposes mutual independence on S_T^4. S_T^5 is diagonal and $[S_T^5]_{ii} = \frac{1}{T}\sum_{t=1}^{T}(X_{it}^2 - \hat{\sigma}_i^2)^2$.
6. S_T^6 imposes Gaussian distribution on S_T^5. $[S_T^6]_{ii} = 2\hat{\sigma}_i^4$.
7. S_T^7 imposes H_M on S_T^6. $[S_T^7]_{ii} = 2$ for $i \leq M$.
8. Also imposes H_M on D_T. Then D_T is diagonal, $[D_T]_{ii} = -2$ for $i \leq M$, $[D_T]_{ii} = -2\hat{\sigma}_i$ for $i > M$, and $\mathcal{W}_T^M$ reduces to $\sum_{t=1}^{M} 2(\hat{\sigma}_i - 1)^2$.

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 data t.

We can write the FOCs as

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

Z_t in principle could consist of infinite history. Let

$B(Z_t)=[B_1(Z_t)',\ldots,B_r(Z_t)']'$ denote a vector of transformations of Z_t.

By the law of iterated expectations,

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

GMM in Asset Pricing

Hansen and Singleton (1982)

We can apply the GMM approach with

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

$$X_t = (c_{t+1}, x_{1,t+1}, \ldots, x_{N,t+1}, B(Z_t)')', \text{ where } c_{t+1} = C_{t+1}/C_t.$$

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

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

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.

Hansen and Singleton (1982) used $B(Z_t)$ consisting of c_t and $x_{j,t}$ as well as several lags of c_{t-k} and $x_{j,t-k}$.

Bibliography

Altonji, J. G. and Segal, L. M. (1996). Small-sample bias in gmm estimation of covariance structures. *Journal of Business & Economic Statistics*, 14(3):353–366.

Andrews, D. W. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica: Journal of the Econometric Society*, pages 817–858.

Andrews, D. W. (1999). Consistent moment selection procedures for generalized method of moments estimation. *Econometrica*, 67(3):543–563.

Andrews, I. and Mikusheva, A. (2016). Conditional inference with a functional nuisance parameter. *Econometrica*, 84(4):1571–1612.

Burnside, C. and Eichenbaum, M. (1996). Small-sample properties of gmm-based wald tests. *Journal of Business & Economic Statistics*, 14(3):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(3):262–280.

Hansen, L. P. and Singleton, K. J. (1982). Generalized instrumental variables