Lecture 6: Moment Selection for GMM

Francis J. DiTraglia

March 31, 2014

1 Review of Generalized Method of Moments

The best all-around reference for for GMM is Hall (2005). These notes draw on chapters 3–7 of his book and use essentially the same notation.

1.1 Key Assumptions

Let f be a q-vector of functions of an observable random r-vector v_t and a p-vector of parameters $\theta \in \Theta \subseteq \mathbb{R}^p$ where Θ is compact. The GMM estimator is defined as follows:

$$
\begin{aligned}
\bar{g}_T(\theta) &= \frac{1}{T}\sum_{t=1}^{T} f(v_t, \theta) \\
\widehat{\theta}_T &= \underset{\theta \in \Theta}{\arg\min}\; \bar{g}_T(\theta)' W_T \bar{g}_T(\theta)
\end{aligned}
$$

The basic assumptions required for GMM estimation are as follows.

Strict Stationarity The sequence $\{v_t \colon -\infty < t < \infty\}$ of random r-vectors is a strictly stationary process with sample space $\mathcal{V} \subseteq \mathbb{R}^r$. Importantly, this implies that the expectations of *any* functions of v_t are constant over time.

Population Moment Condition $E[f(v_t, \theta_0)] = 0$ for some $\theta_0 \in \text{interior}(\Theta)$.

Global Identification For any $\widetilde{\theta} \in \Theta$ such that $\widetilde{\theta} \neq \theta_0$, $E[f(v_t, \widetilde{\theta})] \neq 0$.

Weighting Matrix The $(q \times q)$ weighting matrix W_T is positive semi-definite and converges in probability to a postitive definite constant matrix W.

1.2 Regularity Conditions

Regularity Conditions for Moment Functions The q moment functions $f : \mathcal{V} \times \Theta \to \mathbb{R}^q$ satisfy the following conditions:

(i) f is v_t-almost surely continuous on Θ

(ii) $E[f(v_t, \theta)] < \infty$ exists and is continuous on Θ

Regularity Conditions for Derivative Matrix

(i) The $q \times p$ matrix $\nabla_{\theta'} f(v_t, \theta)$ exists and is v_t-almost continuous on Θ

(ii) $E[\nabla_\theta f(v_t, \theta_0)] < \infty$ exists and is continuous in a neighborhood N_ϵ of θ_0

(iii) $\sup\limits_{\theta \in N_\epsilon} \left|\left| T^{-1} \sum_{t=1}^{T} \nabla_\theta f(v_t, \theta) - E[\nabla_\theta f(v_t, \theta)] \right|\right| \xrightarrow{p} 0$

Regularity Conditions for Variance of Sample Moment Conditions

(i) $E[f(v_t, \theta_0) f(v_t, \theta_0)']$ exists and is finite.

(ii) $\lim\limits_{T \to \infty} Var\left[\sqrt{T} \bar{g}_T(\theta_0)\right] = S$ exists and is a finite, positive definite matrix.

1.3 Basic Asymptotic Results

Under the set of assumptions given above, we obtain the following:

Consistency of GMM Estimator $\widehat{\theta}_T \xrightarrow{p} \theta_0$

Asymptotic Normality of GMM Estimator $\sqrt{T}(\widehat{\theta}_T - \theta_0) \xrightarrow{d} \mathcal{N}(0, MSM')$

$$\begin{aligned} M &= (G_0 W G_0)^{-1} G_0' W \\ G_0 &= E[\nabla_{\theta'} f(v_t, \theta_0)] \end{aligned}$$

2 The J-test Statistic

The J-test statistic is given by

$$J_T = T\ \bar{g}_T(\widehat{\theta}_T')\ \widehat{S}^{-1}\ \bar{g}_T(\widehat{\theta}_T)$$

where $\widehat{S}$ is a consistent estimator of S, the long-run variance matrix of the GMM sample moment conditions. We will need to consider the asymptotic behavior of this quantity in two settings: when the population moment condition is satisfied, and when it is violated.

2.1 Correct Specification

Earlier in this document we reviewed the basic asymptotic results for GMM estimation under standard regularity conditions *assuming the population moment condition is correct.* Our main findings were that, regardless of weighting matrix, GMM is consistent and asymptotically normal. The particular choice of W_T *only* affects the asymptotic variance of the estimator. To study the behavior of the J-test in this setting, we need to examine an asymptotic expansion for the estimated sample moment. Using Taylor Expansion arguments, we can show that

$$W_T^{1/2}\sqrt{T}\bar{g}_T(\widehat{\theta}_T) = [I_q - P(\theta_0)]\, W^{1/2}\sqrt{T}\bar{g}_T(\theta_0) + o_p(1)$$

where

$$
\begin{aligned}
P(\theta_0) &= F(\theta_0)\left[F(\theta_0)'F(\theta_0)\right]^{-1}F(\theta_0)' \\
F(\theta_0) &= W^{1/2}E[\nabla_\theta f(v_t,\theta_0)]
\end{aligned}
$$

The matrix $P(\theta_0)$ is called the *identifying restrictions* and corresponds to the particular projection of $W^{1/2}E[f(v_t,\theta)]$ actually used in GMM estimation. Its orthogonal complement, $N = I_q - P(\theta_0)$, is called the *overidentifying restrictions.* The expansion just stated shows that the asymptotic behavior of the estimated sample moment is *entirely governed by the overidentifying restrictions.* Via a CLT for $\sqrt{T}\bar{g}_T(\theta_0)$, it follows that

$$
W_T^{1/2}\sqrt{T}\bar{g}_T(\widehat{\theta}_T) \xrightarrow{d} \mathcal{N}(0, NW^{1/2}SW^{1/2}N')
$$

Note that the $N = I_q - P(\theta_0)$ has rank $q-p$ since it is the orthogonal complement of the rank p projection matrix $P(\theta_0)$. Hence, in the limit we obtain a *singular normal distribution*, that is a q-dimensional random vector that concentrates on a $(q-p)$-dimensional subspace of $\mathbb{R}^q$. Substituting the efficient weighting matrix $\widehat{S}^{-1}$ we find that $J_T \xrightarrow{d} \chi^2_{q-p}$ by the Continuous Mapping Theorem, *assuming that the population moment condition is correct.*

2.2 Incorrect Specification

When the population moment condition $E[f(v_t,\theta)] = 0$ is *false* for all $\theta \in \Theta$, the situation is completely different. In this case the probability limit of $\widehat{\theta}_T$ in general *will* depend on the choice of weighting matrix and the rate of convergence depends on the rate at which W_T converges to W. Unsurprisingly, this leads to very different behavior for the J-test statistic. So exactly in what sense is $E[f(v_t,\theta)] = 0$ false? For now we'll consider **fixed mis-specification**. Specifically we'll suppose that

$$
E[f(v_t,\theta)] = \mu(\theta), \quad ||\mu(\theta)|| > 0 \quad \forall\, \theta \in \Theta
$$

Note that this situation can only occur if $q > p$ since we can always solve the population moment conditions *exactly* for θ in the just-identified case.

Let $\widehat{S}$ be an estimator of the variance matrix of the moment conditions and let W be the probability limit of $\widehat{S}^{-1}$. Then, if μ_* is the probability limit of $\bar{g}_T(\widehat{\theta})$, we have

$$\frac{1}{T} J_T = \bar{g}_T(\widehat{\theta}_T)' \widehat{S}^{-1} \bar{g}_T(\widehat{\theta}_T) = \mu_*' W \mu_* + o_p(1)$$

In other words $J_T = T\mu_*' W \mu_* + o_p(T)$. Thus, under fixed mis-specification the J-test statistic *diverges at rate* T. Thus, J_T provides a consistent test against the alternative hypothesis of fixed mis-specification.

3 Andrews' GMM Moment Selection Criteria

The consistency and asymptotic normality results for GMM estimation rely on the assumption that the moment conditions used in estimation are correct. That is, they assume that $E[f(v_t, \theta_0)] = 0$. But what if we are unsure of this assumption? In many real-world applications we have a fairly large collection of moment functions, the q elements of f, some of which may have been derived under different economic or statistical assumptions that others. It could easily be the case that only *some* of the moment functions in f satisfy the moment conditions, while others do not. To take a simple example, we may have a collection of instrumental variables that arise from different sources or different assumptions on the DGP. Perhaps only some of these instruments are truly exogenous but we are unsure which. Andrews (1999) proposes a family of *moment selection criteria* (MSC) for this situation, in which the aim is to consistently select *any and all* elements of f that satisfy the moment condition, and eliminate those that do not.

Roughly speaking, the intuition is as follows. When we studied AIC, BIC and friends, we discussed how the maximized log-likelihood measures model fit but unfairly advantages models with more parameters. The various model selection criteria we examined amounted to adding some kind of "penalty" term to correct for this by *penalizing* more complicated models. In a similar vein, so long as we have more moment conditions than parameters, the J-test statistic provides a measure of how well the data "fit" the moment conditions: the bigger the statistic, the greater the evidence that the moment conditions are violated. The problem is that J-test statistic tends to increase as we add additional moment conditions *even if they are correct*. Thus, if we simply compared J-statistics, we would be led to select *too few* moment conditions. To correct for this, Andrews (1999) considers a variety of "bonus terms" that *reward* estimators based on a lager number of moment conditions. Using this idea, he derives GMM analogues of AIC, BIC and the Hannan-Quinn information criterion, and studies the conditions under which a bonus term will yield consistent moment selection.

3.1 Notation

Let f_{max} be a $(q \times 1)$ vector containing all of the moment functions under consideration. Let c be a *selection vector*, a $(q \times 1)$ vector of ones and zeros indicating which elements of f_{max} we use in estimation for a *particular candidate specification*. Let $\mathcal{C}$ denote the set of all candidates and $|c|$ denote the number of moment conditions used to estimate candidate c. Naturally, we require that there are at least as many moment conditions as parameters to estimate.

Let $\widehat{\theta}_T(c)$ be the (efficient two-step) GMM estimator based on $E[f(v_t, \theta, c)] = 0$ and let $V_\theta(c) = \left[G_0(c)S(c)^{-1}G_0(c)\right]^{-1}$ where $G_0(c) = E[\nabla'_\theta f(v_t, \theta_0; c)]$ and $S(c) = \lim_{T\to\infty} Var\left[\frac{1}{\sqrt{T}}\sum_{t=1}^T f(v_t, \theta_0; c)\right]$

So how should we choose c? Talk about why identification is tricky in this setting. Explain about asymptotic efficiency. Adding moment valid moment conditions can never increase asymptotic variance (give reference or maybe even the proof) so it makes sense to use any and all correctly specified moment conditions.

Define $J_T(c) = \left[\frac{1}{\sqrt{T}}\sum_{t=1}^T f(v_t, \widehat{\theta}_T(c); c)\right]' \widehat{S}_T(c)^{-1} \left[\frac{1}{\sqrt{T}}\sum_{t=1}^T f(v_t, \widehat{\theta}_T(c); c)\right]$

Consider moment selection criteria of the form $\text{MSC}(c) = J_T(c) - B(T, |c|)$ where B is a "bonus term."

Select according to $\widehat{c}_T = \underset{c \in \mathcal{C}}{\arg\min}\ \text{MSC}(c)$

Distribution of J-test under correct specification.

Rate of divergence of J-test statistic under mis-specification. Centered covariance matrix estimator.

3.2 Consistent Selection

Regularity Conditions for the J-test Statistic

(i) If $E[f(v_t, \theta; c)] = 0$ for a unique $\theta \in \Theta$, then $J_T(c) \xrightarrow{d} \chi^2_{|c|-p}$

(ii) If $E[f(v_t, \theta; c)] \neq 0$ for a *all* $\theta \in \Theta$ then $T^{-1} J_T(c) \xrightarrow{p} a(c)$, a finite, positive constant that may depend on c.

Regularity Conditions for Bonus Term

(i) $h(\cdot)$ is strictly increasing

(ii) $\kappa_T \to \infty$ as $T \to \infty$ and $\kappa_T = o(T)$

Identification Conditions

(i) $\mathcal{MZ}^0 = \{c_0\}$

(ii) $E[f(v_t, \theta_0; c_0)] = 0$ and $E[f(v_t, \theta; c_0)] \neq 0$ for any $\theta \neq \theta_0$

Theorem 3.1. *Under the preceding assumptions, $\widehat{c}_T \xrightarrow{p} c_0$.*

Proof. We're trying to show that the moment conditions $\widehat{c}_T$ selected by our criterion are consistent for the maximal set c_0 of correct moment conditions. By definition $\widehat{c}_T = \underset{c \in \mathcal{C}}{\arg\min}\ MSC_T(c)$, so we need to show that

$$\lim_{T \to \infty} P\left[\{MSC_T(c) - MSC_T(c_0) > 0,\ \forall c \neq c_0\}\right] = 1$$

To simplify the notation, define

$$\begin{aligned}\Delta_T(c, c_0) &= MSC_T(c) - MSC_T(c_0) \\ &= [J_T(c) - h(|c|)\kappa_T] - [J_T(c_0) - h(|c_0|)\kappa_T] \\ &= [J_T(c) - J_T(c_0)] + \kappa_T\,[h(|c_0|) - h(|c|)]\end{aligned}$$

Now, we are interested in $\Delta_T(c, c_0)$ *only* for situations in which $c \neq c_0$. Subject to this restriction, there are two cases, which we consider in turn.

Case 1 Consider $c_1 \neq c_0$ such that $E[f(v_t, \theta_1; c_1)] = 0$ *for a unique* θ_1. In this case the first Regularity Condition for the J-test Statistic applies to *both* c_1 *and* c_0 and we have

$$J_T(c_1) - J_T(c_0) \xrightarrow{d} \chi^2_{|c_1|-p} - \chi^2_{|c_0|-p} = O_p(1)$$

By the first Identification Condition, c_0 is the *unique* maximal set of correct moment conditions. Hence $|c_0| > |c_1|$. Now, by the first Regularity Condition for the Bonus Term, h is strictly increasing. It follows that $h(|c_0|) - h(|c_1|) > 0$. By the second Regularity Condition for the Bonus Term, $\kappa_T \to \infty$. Thus,

$$\kappa_T\,[h(|c_0|) - h(|c|)] \to \infty$$

It follows that $\Delta_T(c_1, c_0) \to \infty$ and we obtain our desired result.

Case 2 Consider $c_2 \neq c_0$ such that $E[f(v_t, \theta; c_2)] \neq 0$ for any $\theta \in \Theta$. In this case, the *first* Regularity Condition for the J-test Statistic applies to c_0, while the *second* applies to c_2 so we have

$$T^{-1}\,[J_T(c_2) - J_T(c_0)] = a(c_2) + o_p(1) - T^{-1}O_p(1)$$

Now, whatever the value $[h(|c_0|) - h(|c|)]$ happens to be, it is definitely finite since h is strictly increasing by the first Regularity Condition for the Bonus

Term, and both $|c|$ and $|c_0|$ are finite. By the second Regularity Condition for the Bonus Term, $\kappa_T = o(T)$. Hence,

$$T^{-1}\kappa_T\left[h(|c_0|) - h(|c|)\right] = o(1)$$

Putting the pieces together, we have

$$\begin{aligned} T^{-1}\Delta_T(c_2, c_0) &= a(c_2) + o_p(1) - T^{-1}O_p(1) + o(1) \\ &= a(c_2) + o_p(1) \end{aligned}$$

By the second Regularity Condition for the J-test Statistic, $a(c_2) > 0$. Thus, $T^{-1}\Delta_T(c_2, c_0) > 0$ with probability approaching one as $T \to \infty$. It follows that $\Delta_T(c_2, c_0) \to \infty$ with probability approaching one as $T \to \infty$, as required. $\square$

3.3 Which Criteria Are Consistent?

Recall from above that:

$$\begin{aligned} \text{GMM-BIC}(c) &= J_T(c) - (|c| - p)\log(T) \\ \text{GMM-HQ}(c) &= J_T(c) - 2.01\,(|c| - p)\log(\log(T)) \\ \text{GMM-AIC}(c) &= J_T(c) - 2\,(|c| - p)\log(T) \end{aligned}$$

We see immediately that GMM-AIC does *not* satisfy the necessary conditions for consistency, since $\kappa_T = 2$ does not diverge as $T \to \infty$. In constrast, both the GMM-BIC and GMM-HQ diverge as $T \to \infty$, so we simply need to check the requirement that $\kappa_T = o(T)$. For GMM-BIC we have

$$\lim_{T\to\infty} \frac{\log T}{T} = \lim_{T\to\infty} \frac{1}{T} = 0$$

by l'Hôpital's rule, and similarly for GMM-HQ

$$\lim_{T\to\infty} \frac{\log\log T}{T} = \lim_{T\to\infty} \frac{1}{\log T} = 0$$

Thus both GMM-BIC and GMM-HQ provide consistent moment selection.

3.4 Asymptotics for GMM-AIC

We saw in the previous subsection that GMM-AIC does not satisfy the sufficient conditions for consistent moment selection. The question remains: how does this criterion behave in the limit? To answer this question, we revisit the proof of consistent selection from above. It turns out that GMM-AIC behaves *differently* in the two cases considered in the proof. Combining them, we will see that GMM-AIC is *not* a consistent moment selection criterion.

Case 2 In this case, we examined $c_2 \neq c_0$ such that $E[f(v_t, \theta; c_2)] \neq 0$ for any $\theta \in \Theta$. In other words, the moment conditions indexed by c_2 are *not* satisfied for *any* parameter value θ. Asymptotically, GMM-AIC will *never* select such a set of moment conditions. To see why, recall that $\kappa_T = 2$ for GMM-AIC. Although it does not diverge, this choice of κ_T is *still* $o(T)$. Thus, the argument from Case 2 *still applies* to the GMM-AIC. We did not in fact use the assumption that κ_T diverges in the proof of this case!

Case 1 In this case, we examined $c_1 \neq c_0$ such that $E[f(v_t, \theta_1; c_1)] = 0$ *for a unique* θ_1. In other words, we considered a situation in which there *is* a parameter vector θ_1 at which the moment conditions indexed by c_1 are satisfied. Now, the difference of J-test statistics continues to be $O_p(1)$ regardless of the choice of κ_T, provided the regularity conditions are satisfied. Thus, substituting $\kappa_T = 2$, we have

$$\Delta_T(c_1, c_0) = O_p(1) + 2\left[h(|c_0|) - h(|c|)\right]$$

But since the second term is a *constant*, this is simply $\Delta_T(c_1, c_0) = O_p(1)$. In other words, the GMM-AIC is a *random variable*, even in the limit as $T \to \infty$.

So where does this leave us? In Case 2 GMM-AIC cosistently selects c_0, but in Case 1 GMM-AIC is *random even in the limit*. Putting these two results together, we see that, although it will never select a set of false moment conditions, GMM-AIC chooses *randomly* among the set of correct moment conditions. In other words, it will not necessarily select c_0 as $T \to \infty$.

3.5 Problems with Andrews' Approach

Irrelevant moment conditions: Hall & Peixe. Identification condition is stronger than it sounds, can easily fail in simple examples. Another issue worth pointing out is that the GMM-AIC isn't *really* an AIC-type criterion: it simply has a penalty term that looks like AIC. But remember that there was a particular *reason* why the AIC took the form that it did. This isn't how Andrews proceeds, since the paper is mainly interested in sufficient conditions for consistent selection. Comparatively rare that we are actually interested in determining whether our moment conditions are correct. More typically, the goal is to estimate θ. This motivates FMSC, descibed in the next section.

4 The Focused Moment Selection Criterion

DiTraglia (2014). The motivation is simply that it's comparatively rare that we are actually interested in determining whether our moment conditions are correct. More typically, the goal is to estimate θ.

Add asymptotics for Andrews criteria under my asymptotics.