

Lecture 6: Moment Selection for GMM

Francis J. DiTraglia

March 30, 2014

1 Review of GMM

The best reference for GMM is Hall (2005). This review follows Chapter 3.

1.1 The Basic Idea

$$\hat{\theta} = \arg \min_{\theta \in \Theta} \left[\frac{1}{T} \sum_{t=1}^T f(v_t, \theta) \right]' W_T \left[\frac{1}{T} \sum_{t=1}^T f(v_t, \theta) \right]$$

1.2 Basic Assumptions for GMM

Let f be a q -vector of functions of an observable random r -vector v_t and a p -vector of parameters $\theta \in \Theta \subseteq \mathbb{R}^p$. The basic assumptions for GMM estimation are as follows.

Strict Stationarity The sequence $\{v_t: -\infty < t < \infty\}$ 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.

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

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

(ii) $E[f(v_t, \theta)] < \infty$ exists $\forall \theta \in \Theta$

(iii) $E[f(v_t, \theta)]$ is continuous on Θ

Population Moment Condition $E[f(v_t, \theta_0)] = 0$

Regularity Conditions for Derivative Matrix

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

(ii) $\theta_0 \in \text{interior}(\Theta)$ [Needed for Mean-value Expansion]

(iii) $E[\nabla_{\theta} f(v_t, \theta_0)] < \infty$ exists

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

Local Identification

Weighting Matrix The weighting matrix W_T is positive semi-definite and converges in probability to a positive definite constant matrix W .

1.3 Identifying and Overidentifying Restrictions

The “Generalized” in Generalized Method of Moments, comes from the fact that we may use more moment conditions in estimation than we have parameters to estimate. In this case we say that our estimator is “overidentified.” In the overidentified case, the GMM estimator does *not* in fact use all of the information contained in the moment conditions for estimation. This turns out to be *crucial* because it is what allows us to carry out specification tests and moment selection. The key point is that GMM estimation can be viewed as a *plain-vanilla method of moments* problem for a *transformed* set of moment conditions.

The population moment condition for GMM estimation is $E[f(v_t, \theta_0)] = 0$ where f is $q \times 1$. If f is differentiable and we can interchange expectation and derivative, then

2 Andrews (1999)

2.1 Overview

The consistency and asymptotic normality results for GMM estimation reviewed above 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 than 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 larger 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.

2.2 Notation

Talk about why identification is tricky in this setting.

2.3 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) $\kappa_T \rightarrow \infty$ as $T \rightarrow \infty$ and $\kappa_T = o(T)$

Identification Conditions

- (i) $\mathcal{M}\mathcal{Z}^0 = \{c_0\}$
- (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 2.1. *Under the preceding assumptions, $\hat{c}_T \xrightarrow{p} c_0$.*

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 $\widehat{c}_T = \arg \min_{c \in \mathcal{C}} MSC_T(c)$, so what we need to establish is

$$\lim_{T \rightarrow \infty} P[\{MSC_T(c_0) < MSC_T(c), \forall c \neq c_0\}] = 1$$

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, $\kappa_T \rightarrow \infty$. Thus,

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

and it follows that $\Delta_T(c_1, c_0) \rightarrow \infty$.

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 .

□

3 DiTraglia (2014)

4 Other Approaches