What is Generalized Method of Moments?

Alastair R. Hall

The University of Manchester

April 29, 2010

Prepared for the "methods@manchester: research methods in social sciences" seminar series

Introduction

- Econometric analysis involves the use of statistical methods to analyze economic data, and starts with some economic phenomenon of interest.
- Economic theory postulates an explanation that is most often expressed via a set of mathematical equations involving:
  - economic variables;
  - certain constants, known as parameters, that reflect aspects of economic environment.
- Interpretation of parameters is known, but their specific value is not $\rightarrow$ use of statistical methods to estimate parameters based on observed economic data.

Introduction

Various statistical estimation methods available: want method whose implementation does not require the imposition of restrictions on the statistical behaviour of the economic variables beyond those implied by the economic model.

Restrictions implied by economic theory take the form of population moment conditions.

Generalized Method of Moments is a statistical method that combines economic data with the information in population moment conditions to produce estimates of the unknown parameters.

Population moment conditions

The term population moment appears to date from the work of Adolphe Quetelet in early-mid 19th century.

r^{th} population moment is $\nu_r = E[v_t^r]$, (population average of v_t^r).

This concept can be generalized to yield what is refered to as a population moment condition, that is:

$$E[f(v_t, \theta_0)] = 0 \qquad \text{for all } t$$

where v_t denotes (in our context) economic data and θ_0 denotes an unknown parameter vector.

The analogous sample moment is $T^{-1}\sum_{t=1}^{T} f(v_t, \theta_0)$

Method of Moments and GMM

Method of Moments, Pearson (1893, 1894, 1895)
Example: Consider the case where v_t has a normal distribution with unknown mean θ_0 and a (known) variance equal to one. Want to estimate θ_0.

- population moment condition: $E[v_t] - \theta_0 \; = \; 0$
- estimate θ_0 by solving sample moment condition:

$$T^{-1}\sum_{t=1}^{T} v_t - \hat{\theta}_T \; = \; 0$$

$\rightarrow$ Method of Moments estimator: $\hat{\theta}_T = T^{-1}\sum_{t=1}^{T} v_t$.

Method of Moments and GMM

Method of Moments is intuitive but not without its weaknesses.

Focus here here on what happens if want to use information from second moment

$$E[v_t^2] - \theta_0^2 - 1 = 0$$

Now have two sample moment conditions,

$$T^{-1}\sum_{t=1}^{T} v_t - \hat{\theta}_T = 0, \qquad T^{-1}\sum_{t=1}^{T} v_t^2 - \hat{\theta}_T^2 - 1 = 0$$

in one unknown, $\hat{\theta}_T$.

In general no solution $\rightarrow$ GMM.

Generalized Method of Moments

Generalized Method of Moments (GMM), Hansen (1982).

GMM estimator of θ_0 based on $E[f(v_t, \theta_0)] = 0$ is defined to be

$$\hat{\theta}_T \;=\; argmin_{\theta \in \Theta} Q_T(\theta)$$

where

$$Q_T(\theta) \;=\; T^{-1} \sum_{t=1}^{T} f(v_t, \theta)' W_T T^{-1} \sum_{t=1}^{T} f(v_t, \theta)$$

where W_T is known as the weighting matrix.

Generalized Method of Moments

Let $f(.)$ be $q \times 1$ and θ be $p \times 1$.

- If $q = p$ then GMM = Method of Moments based on $E[f(v_t, \theta_0)] = 0$.
- If $q > p$ then GMM estimator is value of θ closest to solving sample moment condition.
- $Q_T(\theta)$ is measure of "closeness to zero": for this measure to make sense need W_T to satisfy certain restrictions so that *inter alia* $Q_T(\theta) \geq 0$.

For example earlier:

$$f(v_t, \theta) = \begin{bmatrix} v_t - \theta \\ v_t^2 - \theta^2 - 1 \end{bmatrix}$$

Other statistical antecedents to GMM

- Method of Minimum Chi-square, Neyman & Pearson (1928)
  - solved problem of how to estimate when more moment conditions than unknown parameters in context of specific modelling context; potential generality of approach only begun to be recognized in late 1940's and 1950's.
- Instrumental Variables (IV) estimation
  - linear models: Wright (1928), Reiersøl (1945), Sargan (1958).
  - nonlinear models: Amemiya (1974), Jorgenson and Laffont (1974).

Example (i): returns to education, Angrist & Krueger (1991)

$$ln[w_t] = \theta_1 + \theta_e ed_t + controls + u_t$$

where

- t denotes the index for the individuals in the sample;
- w_t equals the weekly wage;
- ed_t is the number of years of education;
- *controls* include: birth year, state of residence.
- u_t is the error term.

Interested in θ_e - complication ed_t is likely correlated with u_t

Example (i) cont.d

Angrist and Krueger (1991) argue following population moment condition is valid:

$$E[z_t u_t(\theta_0)] = 0$$

where $u_t(\theta) = (ln[w_t] - \theta_1 - \theta_e ed_t - controls)$, z_t is a vector of variables including the *controls* and:

- a dummy variables indicating quarter of birth

Example (ii): Asset pricing, Hansen & Singleton (1982)

Representative agent makes consumption/investment decisions to maximize his/her expected discounted utility

$$E[\sum_{i=0}^{\infty} \delta^i U(c_{t+i})|\Omega_t]$$

subject to

$$c_t + p_t q_t = r_t q_{t-1} + w_t$$

for all t, where

- the utility function is $U(c) = (c^\gamma - 1)/\gamma$
- c_t is consumption
- Ω_t is the information set available to the agent in period t
- p_t, q_t, r_t are the price, quantity and return on the asset
- w_t is labour income

Example (ii) cont.d

Two parameters to be estimated: $\theta_0 = (\gamma_0, \delta_0)'$.

Model implies following population moment condition:

$$E[e_t(\gamma_0, \delta_0)\, z_t] = 0$$

where

- $e_t(\gamma, \delta) \;=\; \delta(r_{t+1}/p_t)(c_{t+1}/c_t)^{\gamma-1} - 1$
- (any vector) $z_t \in \Omega_t$.

Outline of rest of talk

- Practical issues
- Information in moment conditions
- Inference procedures
- Issues on the research frontier
- Additional resources

Practical issues

Calculation of GMM is performed on the computer.

- In some cases, we can solve analytically for $\hat{\theta}_T$; *e.g.* linear models.
- In other cases, such as nonlinear models, must use numerical optimization routines. (See webpage for link to GMM toolbox in MATLAB.)

Information in moment conditions

Key idea: population moment conditions provide information about unknown parameters.

This raises three important questions:

- Does the population moment condition provide enough information for estimation to be "successful"? $\longrightarrow$ issue of identification.
- How can we extract the most information from a given population moment condition? $\longrightarrow$ issue of optimal choice of weighting matrix.
- What is the most informative choice of moment condition to use? $\longrightarrow$ issue of optimal choice of moment condition.

Population moment condition and identification

Consider behaviour of $E[f(v_t, \theta)]$ as function of θ.

Population Moment Condition: $E[f(v_t, \theta_0)] = 0$, this represents the information upon which estimation is based.

But for this estimation to be "successful" this information must be a unique property of θ_0

$\rightarrow$ Identification condition: $E[f(v_t(\theta)] \neq 0$ for all $\theta \neq \theta_0$

This can be hard to verify *a priori*. But sometimes more revealing conditions for identification can be derived, *e.g.* in returns to education example, quarter of birth must be related to level of education.

Optimal choice of weighting matrix

To characterize the optimal choice of weighting matrix, need to summarize the large sample properties of GMM. Let $W = plimW_T$.

Large sample distribution of $\hat{\theta}_T$ is normal, centred on θ_0 with variance V.

- If $q = p$ then V is independent of W.
- If $q > p$ then V depends on W, and so optimal choice is value that minimizes the large sample variance of $\hat{\theta}_T$ which is:

$$W = S^{-1}$$

  where S is the population variance of the sample moment.

Optimal choice of weighting matrix

- Practical issue: need W_T that converges to S^{-1}.
- Obvious choice: put $W_T = \hat{S}_T^{-1}$ where $plim \hat{S}_T = S$.
- Problem: need estimate of θ_0 to construct $\hat{S}_T$.
- Solution: multi-step estimation:
  - *Step 1*: use sub-optimal $W_T \rightarrow \hat{\theta}_T(1) \rightarrow \hat{S}_T^{-1}(1)$.
  - *Step 2*: use $W_T = \hat{S}_T^{-1}(1) \rightarrow \hat{\theta}_T(2)=$ two-step GMM estimator.
  - Can continue this process $\rightarrow$ iterated GMM estimator.

Optimal choice of moment condition

Within classical paradigm, the asymptotically efficient estimator is Maximum Likelihood (ML).

It can be shown that ML is GMM based on the score function.

Thus the optimal choice of moment condition is the score function associated with the true underlying probability distribution of the data.

While true, this is not practical solution in economic models because true probability distribution is unknown.

Some work on optimal choice within certain classes of moment condition (*e.g.* optimal instrument) but these are often complicated to implement and rarely used.

Inference procedures

Three broad inference questions naturally arise:

- Is the model correctly specified?
- Does the model satisfy restrictions implied by economic/statistical theory?
- Which of two competing models is correct?

Answers to all three involve hypothesis tests involving either the estimated sample moment or the parameter estimators.
Various procedure have been developed. Here we concentrate on the most famous, the overidentifying restrictions test that can be applied when $q > p$.

The overidentifying restrictions test

GMM rests crucially on population moment condition. Therefore, wish to test the null hypothesis

$$H_0 : E[f(v_t, \theta_0)] = 0$$

Overidentifying restrictions test statistic is:

$$J_T = T \times (\text{ second-step GMM minimand evaluated at } \hat{\theta}_T)$$

Under H_0: J_T has large sample χ^2_{q-p} distribution.

Issues

So far, all inference based on large sample (first order asymptotic) theory.

Two natural questions:

- How well does this theory approximate finite sample behaviour?
- What factors affect the quality of this approximation?

Numerous studies, both analytical and simulation based, in the literature.

Findings

What has been learnt? Sometimes GMM works well and sometimes not!

Evidence from all types of study indicates following factors appear to play an important role in determining the quality of the asymptotic approximation:

- the functional form of $f(v_t, \theta_0)$;
- the degree of overidentification, $q - p$;
- the interrelationship between the elements of $f(v_t, \theta_0)$;
- the quality of the identification.

Can we improve the performance of inference procedures?

Two broad responses that stay within GMM framework

1. methods of moment selection - aim is to pick out which moments work "best" for theory based on first order asymptotics

2. alternative approximations to large sample behaviour involving

   2.1 weak identification

   2.2 bootstrap

   2.3 many moment conditions (instrument) asymptotics

Third response: use something else such as Generalized Empirical Likelihood estimation.

Resources

- Alastair R. Hall, 2005, Generalized Method of Moments, Oxford University Press, Oxford, UK.
- Kostas Kyriakoulis, GMM Toolbox for MATLAB. (The help files and examples for this toolbox are linked to Hall, 2005.)

Additional resource:

- Generalized Method of Moments Estimation, Laszlo Matyas (ed.), 1999, Cambridge University Press, Cambridge, UK.