Generalized Methods of Moments

Ercan Karadas

New York University
Department of Economics

Spring, 2018

Table of Contents

Table of Contents

Definitions

- **Population Moment:** a population moment γ rather simply as the expectation of some continuous function g of a random variable $\mathbf{x}$

$$\gamma = E[f(\mathbf{x})]$$

- **Sample Moment:** a sample moment γ is the sample version of the population moment in a particular random sample:

$$\hat{\gamma} = \frac{1}{N}\sum_{i=1}^{N} g(\mathbf{x})$$

- **Methods of Moments (MOM):** is merely the following proposal

  To estimate a population moment (or a function of population moments) merely use the corresponding sample moment (or functions of sample moments).

Example

- Suppose we are interested in estimating the variance of **x**

$$\begin{aligned} V(x) &= E(x^2) - [E(x)]^2 \\ \implies \hat{V}(x) &= \frac{1}{N}\sum x^2 - \left[\frac{1}{N}\sum x\right]^2 \\ &= \frac{1}{N}\sum (x-\bar{x})^2 \end{aligned}$$

  Note that the MOM estimator is biased, but it's consistent.

- Alternatively, we could have begun with the conventional definition of the pop- ulation variance and substituted sample analogs directly:

$$\begin{aligned} V(x) &= E\left[x - E(x)\right]^2 \\ \implies \hat{V}(x) &= \frac{1}{N}\sum (x-\bar{x})^2 \end{aligned}$$

Elements of GMM Estimation

- "Theory" or a priori information yields an assertion about a population orthogonality condition, which is usually of the form

$$E[f(w_t, z_t, \theta)] = 0$$

  where $f(\cdot)$ is a vector function with R elements, θ is a K-dimensional vector containing all unknown parameters, w_t is a vector of observable variables that could be endogenous or exogenous; and z_t is a vector of instruments.

- We construct the sample analog $g_T(\theta)$ to the population moment condition

$$g_T(\theta) \equiv \frac{1}{T}\sum_{t=1}^{T} f(w_t, z_t, \theta) \qquad (*)$$

  As in IV case there are different cases:
  - $R < K$: no solution
  - $R = K$: unique consistent estimator solution
  - $R > K$: cannot solve uniquely

Elements of GMM Estimation

- In the last case, the number of moment conditions is larger, so we cannot solve uniquely for the unknown parameters by setting $(*)$ to zero
- Instead, we choose our estimator for θ such that the vector of sample moments is as close as possible to zero, in the sense that a quadratic form in $g_T(\theta)$ is minimized with respect to θ:

$$\hat{\theta} \in \arg\min_{\theta} \; g_T(\theta)' \cdot W_T \cdot g_T(\theta) \qquad (**)$$

  where W_T is a positive definite matrix with $W_T \xrightarrow{p} W$ for some positive definite matrix.
- The solution to this problem provides the **generalized method of moments** or **GMM** estimator $\hat{\theta}$.

Elements of GMM Estimation

- Different weighting matrices W_T lead to different consistent estimators with different asymptotic covariance matrices
- The optimal weighting matrix, which leads to the smallest covariance matrix for the GMM estimator, is the inverse of the covariance matrix of the sample moments (Hansen, 1982). In the absence of autocorrelation it is given by
$$W^{opt} = \left\{ E\left[f(w_t, z_t, \theta)\, f(w_t, z_t, \theta)' \right] \right\}^{-1}$$
(Usually, this is best chosen to be a consistent estimate of $Var[g_T(\cdot)]^{-1}$ (For example, as in the White covariance matrix or, in the time series context, the appropriate Newey-West covariance matrix).
- In general this matrix depends upon the unknown parameter vector θ, which presents a problem that we did not encounter in the linear model.
- The solution is to adopt a multistep estimation procedure.

Elements of GMM Estimation

- In the first step we use a suboptimal choice of W_T that does not depend upon θ (e.g. the identity matrix) to obtain a first consistent estimator $\hat{\theta}_1$. Then, we can consistently estimate the optimal weighting matrix by

$$W_T^{opt} = \left\{ \frac{1}{T} \sum_{t=1}^{T} f(w_t, z_t, \hat{\theta}_1) \, f(w_t, z_t, \hat{\theta}_1)' \right\}^{-1}$$

- In the second step one obtains the asymptotically efficient (optimal) GMM estimator $\hat{\theta}$ by using W_T^{opt} in $(**)$
- The asymptotic distribution of θ is given by

$$\sqrt{T}\left(\hat{\theta} - \theta\right) \xrightarrow{d} N(0, V)$$

where the asymptotic covariance matrix V is given by

$$V = \left(D W^{opt} D'\right)^{-1}$$

and D is the $K \times R$ derivative matrix

$$D = E\left\{ \frac{\partial f(w_t, z_t, \theta)}{\partial \theta'} \right\}$$

- Intuitively, the elements in D measure how sensitive a particular moment is with respect to small changes θ
- If the sensitivity with respect to a given element in θ is large, small changes in this element lead to relatively large changes in the objective function in $(**)$ and the particular element in θ is relatively accurately estimated
- As usual, the covariance matrix V can be estimated by replacing the population moments in D and W^{opt} with their sample equivalents, evaluated at $\hat{\theta}$

Table of Contents

Model

- This example is based on Hansen and Sinleton (1982)
- In this example we will see how GMM can be used to estimate (or infer) the unknown model parameters directly from the moment conditions that are imposed by a theoretical model.
- It also illustrates how valid instruments may follow from economic theory.
- Consider an individual agent who maximizes the expected utility of current and future consumption by solving

$$\max_{\{C_{t+s}\}_{s=0}^{S}} E_t \left\{ \sum_{s=0}^{S} \delta^s U(C_{t+s}) \right\}$$

s.t.

$$C_{t+s} + q_{t+s} = w_{t+s} + (1 + r_{t+s}) q_{t+s-1}$$

E_t : the expectation operator conditional upon time t information

C_t : consumption in period t

q_{t+s} : financial wealth at the end of period $t+s$

r_{t+s} : return on financial wealth

w_{t+s} : labour income

Maximization and Euler Equation

- To solve this constrained optimization problem, you can set up the Lagrangian and proceed from there.
- But substitution method is easier in this simple example: solve for C_{t+s} in the budget constraint and substitute into the objective function to get

$$\max_{\{q_{t+s}\}_{s=0}^{S}} E_t \left\{ \sum_{s=0}^{S} \delta^s U(w_{t+s} + (1 + r_{t+s})q_{t+s-1} - q_{t+s}) \right\}$$

- FOCs with respect to q_{t+s}

$$E_t \left\{ \delta^s U'(C_{t+s})(-1) + \delta^{s+1} U'(C_{t+s+1})(1 + r_{t+1+s}) \right\} = 0$$

- Focus on $s = 0$ and simplify to get

$$E_t \left\{ \delta U'(C_{t+s+1})(1 + r_{t+1+s}) \right\} = U'(C_{t+s})$$

  This equation is known as **Euler Equation**.
- It implies that expected marginal utilities are equalized across the different periods.

Moment Conditions

- We can rewrite this equation as
$$E_t \left\{ \delta \frac{U'(C_{t+1})}{U'(C_t)}(1 + r_{t+1}) - 1 \right\} = 0 \qquad (\star)$$
- This is a (conditional) moment condition that can be exploited to estimate the unknown parameters if we make some assumption about the utility function U. We can do this by transforming $(\star)$ into a set of unconditional moment conditions
- Suppose that z_t is included in the information set at time t, so it shouldn't provide any new information about the expected value of
$$\delta \frac{U'(C_{t+1})}{U'(C_t)}(1 + r_{t+1}) - 1$$
- But this implies that we can write it as an unconditional expectation:
$$E \left\{ \left[\delta \frac{U'(C_{t+1})}{U'(C_t)}(1 + r_{t+1}) - 1 \right] z_t \right\} = 0 \qquad (\star\star)$$
- We now have a set of moment conditions that identify the unknown parameters δ and γ
- Given observations on $\{C_{t+s}, r_{t+s}, z_{t+s-1}\}_{s=0}^{S}$ we can estimate them consistently

Moment Conditions

- In the above result we used the following general result: given two random variables X_1 and X_2, $E[X_1|X_2] = 0$ implies that $E[X_1 g(X_2)] = 0$ for any function g.
- To prove $E[X_1|X_2] = 0 \implies E[X_1 g(X_2)] = 0$, just use the law of iterated expectations:

$$E[X_1 g(X_2)] = E_{X_2}\left[E[X_1 g(X_2)|X_2]\right] = E_{X_2}\left[g(X_2) E[X_1|X_2]\right] = 0$$

- To understand how we used this result, note that instead of writing the expectation conditional on time t as $E_t\{\ldots\}$, we can write it as $E\{\ldots|\mathbf{I}_t\}$, where $\mathbf{I}_t$ is the time t information set. Now, by assumption $z_t \in \mathbf{I}_t$ and therefore we can use the above result to obtain the unconditional moment condition

Moment Conditions for a Specific Utility Function

- Suppose we are given the following utility function
$$U(C_t) = \frac{C_t^{1-\gamma}}{1-\gamma}$$
- In this case, $U'(C_t) = C_t^{-\gamma}$ and therefore the unconditional moment conditions are
$$E\left\{\left[\delta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}(1+r_{t+1}) - 1\right] z_t\right\} = 0$$
- This is a set of restrictions that the theory imposes on the joint distribution of consumption growth (C_{t+1}/C_t), returns (r_t) and z_t.
- We now have a set of moment conditions that identify the unknown parameters δ and γ
- Given observations on $\{C_{t+s}/C_{t+s-1}, r_{t+s}, z_{t+s-1}\}_{s=0}^{S}$ we can estimate them consistently

Choice of Instruments

- Any stationary variable z_t realized at time t is a valid instrument
- But not all of them are equally good, of course
- For example, if z_t is uncorrelated with future returns and consumption, then it does not add any restriction since

$$E\left\{\left[\delta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}(1+r_{t+1})-1\right]z_t\right\}=0$$

$$E\left\{z_t\right\}\left\{\left[\delta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}(1+r_{t+1})-1\right]\right\}=0$$

- Therefore a good instrument z_t should forecast future returns or future consumption growth.
- Here the critical point is that the moment conditions imposes restrictions on the joint distribution of the data $\left\{C_{t+s}/C_{t+s-1}, r_{t+s}, z_{t+s-1}\right\}_{s=0}^{S}$. Therefore, even though at time t we know the realized value of z_t (so can use it as an instrument) it's still informative to predict the unrealized part of the joint distribution, in this case the future values.

Sample Moment Conditions

- Intuitively, if we had two instrument z_{1t} and z_{2t}, we simply pick δ and γ to set the sample analogs of the moment conditions equal to zero, i.e.

$$\frac{1}{T}\sum_{t=1}^{T}\left[\delta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}(1+r_{t+1})-1\right]z_{1t}=0$$
$$\frac{1}{T}\sum_{t=1}^{T}\left[\delta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}(1+r_{t+1})-1\right]z_{2t}=0$$

- In this case, there are two equations in two unknowns, so here there is a unique solution
- If more than equations, then of course (generically) cannot set all the equations equal to zero. In this case, as we saw in the general formulation, we minimize the weighted errors:

$$\min_{\delta,\gamma} g_T' \cdot W_T \cdot g_T$$

where

$$g_T=\frac{1}{T}\sum_{t=1}^{T}\left[\delta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}(1+r_{t+1})-1\right]z_t$$

and W_T is a weighting matrix.

How is this example different from OLS?

- As we discussed, in GMM framework restrictions that enable us to estimate the parameters comes from a theory and assumptions are hid in there. In the OLS framework, in contrast, the assumptions were about the specified functional form between the dependent and independent variables and the properties of the error terms
- For example, what kind of assumptions are hid in the theory part of Hansen-Singleton model?
  - We assume there is an optimizing agent with rational expectations.
  - We also assume that this optimum behavior generated the observed history of data on consumption, returns and instruments $\{C_{t+s}/C_{t+s-1}, r_{t+s}, z_{t+s-1}\}_{s=0}^{S}$.
  - Then we proceed as follow
    - Find the implications of the theoretical model on the behavior of agent (FOCs) to obtain the moment conditions
    - Given theoretical moment conditions and the observed history of data, try to *infer* the parameters of the agent's utility function
- There is no reason to believe that these assumptions are more realistic than the OLS assumptions. Therefore, we should rather think GMM as an alternative estimation method with some advantages:
  - It does not require a linear specification
  - It does not require distributional assumptions, like normality
  - It can allow for heteroskedasticity of unknown form
  - It can estimate parameters even if the model cannot be solved analytically from the first-order conditions