Master MIF

Financial Econometrics

Nour Meddahi (Toulouse School of Economics)

Generalized Method of Moments

I. Examples

1) A Toy Example. Assume that one has an i.i.d. sample $x_1, x_2, ..., x_T$, where $x_t \sim D(\mu_0, \mu_0^2)$. We assume that $\mu_0 > 0$. We want to estimate μ_0.

A possible estimator of μ_0 is the sample average:

$$\hat{\mu}_1 = \bar{X} = \frac{1}{T}\sum_{t=1}^{T} x_t.$$

Another estimator could be derived from the sample variance (we have two roots, we take the positive one):

$$\hat{\mu}_2 = \sqrt{\frac{1}{T-1}\sum_{t=1}^{T}(x_t - \bar{X})^2}.$$

An interesting question is whether we can find a better estimator than $\hat{\mu}_1$ and $\hat{\mu}_2$? Or, saying it differently, can we **combine optimally** the moment conditions

$$E[x_t - \mu_0] = 0,\ E[x_t^2 - 2\mu_0^2] = 0$$

in order to get the most precise estimator of μ_0?

2) Dynamic Asset Pricing Models. We know that the non-arbitrage assumption implies the existence of a positive random variable m_t, called the **stochastic discount factor**, such that for any asset, one has

$$P_t = E[m_{t+1}(P_{t+1} + D_{t+1}) \mid I_t],$$

where P_t and D_t are the price and dividend of the asset at time t, while I_t is the information available at time t. Consequently, one has

$$E\left[\left(m_{t+1}\frac{(P_{t+1} + D_{t+1})}{P_t} - 1\right) \mid I_t\right] = 0.$$

There are different approaches. A common one is to write m_{t+1} in terms of factors f_{t+1}, i.e.

$$m_{t+1} = a + f'_{t+1}b.$$

The vector of factors f_{t+1} could include the market return (like the CAPM) and other factor like the Fama and French ones.

One observes the prices and dividends of many assets. Therefore, one should use them to estimate a and b.

Observe that for any $Z_t \in I_t$, one has

$$E\left[Z_t\left(m_{t+1}\frac{(P_{t+1} + D_{t+1})}{P_t} - 1\right)\right] = 0.$$

II. The Theory

We assume that we have moment conditions

$$Eg(y_t, X_t, \theta_0) = 0$$

where y_t and X_t are the endogenous and exogenous variables respectively and θ_0 is a k-dimensional vector of parameters and $g(.)$ is an L-dimensional vector, with $L \geq k$. Observe that the two previous examples are special cases.

Ideally, one would like to define the estimator $\hat{\theta}$ as the solution of the equation

$$\frac{1}{T}\sum_{t=1}^{T} g(y_t, X_t, \hat{\theta}) = 0.$$

This is the case for the OLS estimator:

$$\frac{1}{T}\sum_{t=1}^{T} x_t(y_t - x_t'\beta^{OLS}) = 0.$$

The problem is that in general one has no solution when $L > k$, i.e. when one has more equations than unknowns, a case called over-identified. When $K = L$, one says that the model is just-identified.

Instead of having the sample mean of the function $g(\cdot)$ equals to zero, the GMM minimizes a norm of the vector $\frac{1}{T}\sum_{t=1}^{T} g(y_t, X_t, \theta)$.

More precisely, for a given $L \times L$ matrix W_T that could depend on the data, assumed to be positive definite, the GMM estimator is defined as:

$$\hat{\theta}_{GMM} \equiv \text{argmin}_{\theta \in \Theta} \left(\frac{1}{T} \sum_{t=1}^{T} g(y_t, X_t, \theta) \right)' W_T \left(\frac{1}{T} \sum_{t=1}^{T} g(y_t, X_t, \theta) \right).$$

Assumptions. We need some assumptions:

1) $PlimW_T = W$ where W is positive definite.

2) $E[g(y_t, X_t, \theta)] = 0 \Longrightarrow \theta = \theta_0$ (identification assumption)

3) The following matrix has a full rank

$$G = E\left[\frac{\partial g(y_t, X_t, \theta_0)}{\partial \theta'}\right].$$

4) One has a central limit theorem

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} g(y_t, x_t, \theta^0) \ \rightarrow^d \ \mathcal{N}(0, \Omega)$$

where

$$\Omega = Plim_{T\rightarrow+\infty} Var\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} g(y_t, X_t, \theta_0)\right].$$

Observe that we allow for serial correlation in the $g(y_t, X_t, \theta_0)$.

Properties of the GMM estimator. Under these assumptions, one can show

$$\hat{\theta}_{GMM} \rightarrow^p \theta_0,$$
$$\sqrt{T}(\hat{\theta}_{GMM} - \theta_0) \rightarrow^d \mathcal{N}(0, (G'WG)^{-1}G'W\Omega WG(G'WG)^{-1}).$$

Optimal choice of W. When $L = k$, the optimal GMM estimator does not depend on W.

When $L > K$, there is a degree of freedom in the previous analysis given that the weighting matrix W could be any positive definite matrix. It is therefore interesting to characterize the matrix that leads to the smallest asymptotic variance of the estimator (the optimal GMM estimator).

One can show that the optimal choice of W is

$$W_{opt} = \Omega^{-1}.$$

The asymptotic distribution of the optimal GMM is given by

$$\sqrt{T}(\hat{\theta}_{opt} - \theta_0) \to^d \mathcal{N}(0, (G'\Omega^{-1}G)^{-1}).$$

Implementing the Optimal GMM estimator. The main problem with the optimal GMM estimator is that one needs the matrix Ω, which is often unknown. A **two-step** method is needed.

1) In the first step, one considers the weighting matrix $W_1 = Id_L$. The estimator is not optimal but is consistent. It is denoted $\hat{\theta}_1$.

One can then estimate Ω. When the data are i.i.d., an estimator of Ω is given by

$$\hat{\Omega} = \frac{1}{T}\sum_{t=1}^{T}(g(y_t, X_t, \hat{\theta}_1) - \bar{g}_1)(g(y_t, X_t, \hat{\theta}_1) - \bar{g}_1)', \quad \text{where } \bar{g}_1 = \frac{1}{T}\sum_{t=1}^{T} g(y_t, X_t, \hat{\theta}_1).$$

When there is serial correlation (potentially of unknown form), one should estimate Ω by the Newey-West method:

$$\hat{\Omega}_1 = \hat{V}ar[g(y_t, X_t, \hat{\theta}_1)]$$
$$+\sum_{l=1}^{L}\left(1 - \frac{l}{L}\right)\left(\hat{C}ov[g(y_t, X_t, \hat{\theta}_1), g(y_{t+l}, X_{t+l}, \hat{\theta}_1)] + \hat{C}ov[g(y_t, X_t, \hat{\theta}_1), g(y_{t+l}, X_{t+l}, \hat{\theta}_1)]'\right),$$

where

$$\hat{C}ov[g(y_t, X_t, \hat{\theta}_1), g(y_{t+l}, X_{t+l}, \hat{\theta}_1)] = \frac{1}{T-k}\sum_{t=1}^{n-l}(g(y_t, X_t, \hat{\theta}_1) - \bar{g}_1)(g(y_{t+l}, X_{t+l}, \hat{\theta}_1) - \bar{g}_1)'.$$

2) In the second-step, one considers the weighting matrix $W_2 = \hat{\Omega}_1^{-1}$, which leads to the optimal GMM estimator, i.e.

$$\sqrt{T}(\hat{\theta}_2 - \theta_0) \to^d \mathcal{N}(0, (G'\Omega^{-1}G)^{-1}).$$

An estimator of the variance of $\hat{\theta}_2$ is

$$\hat{V}ar[\hat{\theta}_2] = \frac{1}{T}(\hat{G}'\hat{\Omega}_1^{-1}\hat{G})^{-1}$$

where

$$\hat{G} = \frac{1}{T}\sum_{t=1}^{T} \frac{\partial g(y_t, X_t, \hat{\theta}_2)}{\partial \theta'}.$$

Observe that in practice, it is recommended to iterate the procedure few times.

Over-identification Test. When $L > k$, a simple test of the model is the over-identification test given by

$$T \left(\frac{1}{T} \sum_{t=1}^{T} g(y_t, X_t, \hat{\theta}_{opt}) \right)' \hat{\Omega}^{-1} \left(\frac{1}{T} \sum_{t=1}^{T} g(y_t, X_t, \hat{\theta}_{opt}) \right) \to^d \chi^2(L-k).$$

Concluding remarks.

1) Observe that one can test hypotheses on θ^0 by using the asymptotic distribution of $\hat{\theta}_{opt}$ and the Delta method (the same approach was followd to derive the Wald tests in the MLE case.)

2) It is not recommended to use too many moments because one has to estimate the weighting matrix.

3) The second example is defined in terms of conditional moments, while our GMM theory was based on unconditional moments. One needs to choose the instruments Z_t. There is a theory for the optimal choice of the instrument that leads to the smallest variance of the GMM estimator. It involves non-parametric estimation method. The practical device is: 1) use informative instruments, i.e. instruments that should have information about θ^0; 2) use lagged values of y_t and X_t; 3) do not use too much instruments (dimension of the weighting matrix).

4) Optimal instrument: If the model is given by

$$E[f(y_t, x_t, \theta^0) \mid x_t] = 0.$$

Then the optimal instrument (i.e. one who leads to a minimum asymptotic variance of the GMM estimator) is given by

$$z_t^* = E\left[\frac{\partial f^\top}{\partial \theta}(y_t, x_t, \theta^0) \mid x_t\right] \left(Var[f(y_t, x_t, \theta^0) \mid x_t]\right)^{-1}.$$

The problem is that often the model does not impose any restriction on $Var[f(y_t, x_t, \theta^0) \mid x_t]$, i.e. it is unknown (it could be time-varying) and sometimes $E\left[\frac{\partial f^\top}{\partial \theta}(y_t, x_t, \theta^0) \mid x_t\right]$ is unknown too.

Observe that the (unconditional) moment conditions are

$$E[z_t^* f(y_t, x_t, \theta^0)] = 0,$$

which are indeed just-identified.