ECON 5360 Class Notes
GMM

1 Generalized Method of Moments (GMM)

I begin by outlining the classical method of moments technique (Fisher, 1925) and then proceed to generalized method of moments (Hansen, 1982).

1.1 Traditional Method of Moments

The idea is to match the population moments of a distribution to the sample moments, using as many moments as necessary to estimate the unknown parameters. Let $\{X_1, X_2, ..., X_n\}$ be a random sample from the pdf $f(x; \theta_1, ..., \theta_r)$. Also, let

$$m'_k = \frac{1}{n}\sum\nolimits_{i=1}^{n} x_i^k$$

be the k^{th} sample moment and $\mu'_k = E(X^k)$ the k^{th} population moment. The method of moments estimator for $\theta = (\theta_1, ..., \theta_r)'$ is therefore the solution to the equations

$$m'_i = \mu'_i(\theta)$$

for $i = 1, ..., r$. Method of moments can be modified to use centered, as opposed to raw, moments. While consistent, method of moments estimators are not generally efficient.

1.1.1 Example

Suppose $\{X_1, X_2, ..., X_n\}$ is a random sample from a $gamma(\alpha, \beta)$ distribution. The likelihood function

$$L(\theta) = (\Gamma(\alpha)\beta^{\alpha})^{-n} (x_1 x_2 \cdots x_n)^{\alpha-1} \exp\left(-\sum\nolimits_{i=1}^{n} x_i/\beta\right)$$

is difficult to evaluate without using numerical methods. A method of moments estimator jointly solves

$$\begin{aligned} \bar{X} &= \frac{1}{n}\sum\nolimits_i X_i = E(X) = \alpha\beta \\ \hat{\sigma}^2 &= \frac{1}{n}\sum\nolimits_i (X_i - \bar{X})^2 = E[(X - \mu_1)^2] = \alpha\beta^2 \end{aligned}$$

for $\hat{\alpha}$ and $\hat{\beta}$. This gives

$$\begin{aligned} \hat{\alpha} &= \bar{X}^2/\hat{\sigma}^2 \\ \hat{\beta} &= \hat{\sigma}^2/\bar{X}. \end{aligned}$$

1.1.2 Variance of Method of Moments Estimator

Let the sample moments be

$$\bar{g}_k = \frac{1}{n}\sum\nolimits_i g_k(X_i)$$

for $k = 1, ..., K$ and $\bar{g} = (\bar{g}_1, ..., \bar{g}_K)$ have asymptotic variance-covariance matrix V, with elements

$$V_{jk} = \frac{1}{n}\left\{\frac{1}{n}\sum\nolimits_i (g_j(X_i) - \bar{g}_j)(g_k(X_i) - \bar{g}_k)\right\}$$

where $j, k = 1, ..., K$. Now let G be the matrix

$$G = \begin{bmatrix} \frac{\partial \bar{g}_1}{\partial \theta_1} & \frac{\partial \bar{g}_1}{\partial \theta_2} & \cdots & \frac{\partial \bar{g}_1}{\partial \theta_K} \\ \frac{\partial \bar{g}_2}{\partial \theta_1} & \frac{\partial \bar{g}_2}{\partial \theta_2} & & \frac{\partial \bar{g}_2}{\partial \theta_K} \\ \vdots & & \ddots & \vdots \\ \frac{\partial \bar{g}_K}{\partial \theta_1} & \frac{\partial \bar{g}_K}{\partial \theta_2} & \cdots & \frac{\partial \bar{g}_K}{\partial \theta_K} \end{bmatrix}_{K \times K}.$$

Since the population moments $\mu(\theta)$ are typically a nonlinear function in θ, we will linearize using a first-order Taylor approximation to $\bar{g}_k = \mu_k(\theta)$ around the true value θ

$$\begin{aligned} \bar{g} &\cong \mu(\theta) + G(\theta)(\hat{\theta} - \theta) \Rightarrow \\ (\hat{\theta} - \theta) &= G^{-1}(\theta)(\bar{g} - \mu(\theta)). \end{aligned}$$

Therefore, our estimate of the asymptotic variance is

$$est.asy.var.(\hat{\theta}) = \hat{G}^{-1} V (\hat{G}^{-1})'.$$

1.1.3 Gamma Example Continued

In the gamma distribution example above, where $g_1 = X_i$ and $g_2 = (X_i - \bar{X})^2$, we have

$$\hat{G} = \begin{bmatrix} \hat{\beta} & \hat{\alpha} \\ \hat{\beta}^2 & 2\hat{\alpha}\hat{\beta} \end{bmatrix}$$

and

$$V = \frac{1}{n}\begin{bmatrix} \widehat{var}(g_1) & \widehat{cov}(g_1, g_2) \\ \widehat{cov}(g_2, g_1) & \widehat{var}(g_2) \end{bmatrix}.$$

1.2 Generalized Method of Moments

GMM extends the classical method of moments estimator to handle cases where there are more moment conditions than parameters to estimate (i.e., the model is overidentified).

1.2.1 Basic Framework

Suppose there are K parameters to estimate $\theta = (\theta_1, ..., \theta_K)'$ and $L \geq K$ moment conditions

$$E[m_l(y_i, X_i, Z_i; \theta)] = 0 \tag{1}$$

for $l = 1, ..., L$. The sample analog of (1) is

$$\bar{m}_l(y_i, X_i, Z_i; \theta) = \frac{1}{n}\sum_{i=1}^{n} m_l(y_i, X_i, Z_i; \theta) = 0$$

which will generally have a unique solution if $L = K$ and multiple solutions if $L > K$. To reconcile the multiple solutions, consider minimization of

$$q = \bar{m}(\theta)' W_n \bar{m}(\theta)$$

where $\bar{m}(\theta) = (\bar{m}_1, ..., \bar{m}_l)'$ and W_n is a positive definite weighting matrix. If $W_n = I_n$, then minimization of q is simply a least squares criterion. If $W_n \neq I_n$, then minimization of q is similar in spirit to GLS, which re-weights the observations according to the variance-covariance matrix of the errors. Again, in the spirit of GLS, Hansen (1982) shows that the optimal criterion (weighting matrix) is to minimize

$$q = \bar{m}(\theta)' \Phi^{-1} \bar{m}(\theta)$$

where

$$\Phi = Asy.Var.(\sqrt{n}\bar{m}).$$

The resulting estimator, $\hat{\theta}_{GMM}$, will have an asymptotic variance-covariance matrix equal to

$$Est.Asy.Var.(\hat{\theta}_{GMM}) = \frac{1}{n}[\Gamma' \hat{\Phi}^{-1} \Gamma]^{-1}$$

where Γ is a matrix of partial derivatives similar in spirit to G above.

1.2.2 Properties of the GMM Estimator

Assuming that the

1. parameters are identifiable,

2. empirical moments converge in probability to their population counterparts (i.e., $\bar{m}(\theta) \xrightarrow{p} 0$), and

3. the empirical moments obey the central limit theorem (i.e., $\sqrt{n}\bar{m}(\theta) \xrightarrow{d} N[0, \Phi]$),

then

$$\hat{\theta}_{GMM} \overset{asy}{\sim} N[\theta, \frac{1}{n}(\Gamma'\Phi^{-1}\Gamma)^{-1}].$$

1.2.3 Example #1. Ordinary Least Squares – Exactly Identified Case

Nearly all estimators we have covered can be posed as method of moment estimators. Consider GMM estimation of the bivariate linear regression model

$$y_i = \alpha + \beta x_i + \epsilon_i.$$

Two moment conditions arising from the Classical assumptions are

$$\begin{aligned} E[m_1(y_i, x_i; \alpha, \beta)] &= E(\epsilon_i) = 0 \\ E[m_2(y_i, x_i; \alpha, \beta)] &= E(\epsilon_i x_i) = 0. \end{aligned}$$

The sample analog of these population moment conditions are

$$\begin{aligned} \frac{1}{n}\sum\nolimits_i e_i &= \frac{1}{n}\sum\nolimits_i (y_i - \hat{\alpha} - \hat{\beta}x_i) = 0 \\ \frac{1}{n}\sum\nolimits_i e_i x_i &= \frac{1}{n}\sum\nolimits_i (y_i - \hat{\alpha} - \hat{\beta}x_i)x_i = 0, \end{aligned}$$

which are, of course, the normal equations for OLS estimation of the classical linear regression model. In this instance, the weighting matrix W is irrelevant because both moment conditions can be satisfied exactly. Therefore, we have

$$\begin{aligned} \hat{\alpha}_{GMM} &= \hat{\alpha}_{OLS} = \bar{y} - b\bar{x} \\ \hat{\beta}_{GMM} &= \hat{\beta}_{OLS} = \frac{\sum_i (y_i - \bar{y})(x_i - \bar{x})}{\sum_i (x_i - \bar{x})^2}. \end{aligned}$$

1.2.4 Example #2. Hall's Random-Walk Consumption Hypothesis

In a famous 1978 article in the *Journal of Political Economy*, Robert Hall showed that under certain conditions, consumption should be expected to follow a random walk. Consider an agent that chooses consumption

c_t to maximize discounted, expected lifetime utility

$$E_0 \sum_{t=0}^{T} (1+\rho)^{-t} 0.5\phi[\bar{c} - c_t]^2,$$

where ρ is the subjective discount rate, ϕ is a constant, and $\bar{c}$ is the bliss level of consumption, subject to

$$A_0 = \sum_{t=0}^{T} (1+r)^{-t}(c_t - w_t)$$

where A_0 is initial assets, r is the interest rate and w_t is the wage rate. Hall shows that if $\rho = r$, then consumption follows a random walk

$$c_t = c_{t-1} + \epsilon_t$$

where $E_{t-1}[\epsilon_t] = 0$. Campbell and Mankiw (1989) test Hall's hypothesis by posing a specific alternative – agents simply consume a given fraction λ of their current income (i.e., $c_t = \lambda w_t$). The two hypotheses can be nested according to

$$\begin{aligned} c_t - c_{t-1} &= \lambda(w_t - w_{t-1}) + (1-\lambda)\epsilon_t \\ \Delta c_t &= \lambda \Delta w_t + \nu_t. \end{aligned}$$

In principle, one could just run a regression of the change in consumption on the change in income and test whether the coefficient λ is different than zero. The problem is that Δw_t and ν_t are likely to be correlated so that instrumental variables need to be found. Consider using the first four lagged changes in consumption: $\Delta c_{t-1}, ..., \Delta c_{t-4}$. The moment conditions are therefore

$$\begin{aligned} E[m_1(\Delta c_t, \Delta w_t, \Delta c_{t-1}; \lambda)] &= E[\nu_t \Delta c_{t-1}] = 0 \\ E[m_2(\Delta c_t, \Delta w_t, \Delta c_{t-2}; \lambda)] &= E[\nu_t \Delta c_{t-2}] = 0 \\ E[m_3(\Delta c_t, \Delta w_t, \Delta c_{t-3}; \lambda)] &= E[\nu_t \Delta c_{t-3}] = 0 \\ E[m_4(\Delta c_t, \Delta w_t, \Delta c_{t-4}; \lambda)] &= E[\nu_t \Delta c_{t-4}] = 0. \end{aligned}$$

The sample analogs are

$$\begin{aligned} \frac{1}{T}\sum_{t=0}^{T} m_{1t}(\hat{\lambda}) &= \frac{1}{T}\sum_{t=0}^{T} (\Delta c_t - \hat{\lambda}\Delta w_t)\Delta c_{t-1} = 0 \\ \frac{1}{T}\sum_{t=0}^{T} m_{2t}(\hat{\lambda}) &= \frac{1}{T}\sum_{t=0}^{T} (\Delta c_t - \hat{\lambda}\Delta w_t)\Delta c_{t-2} = 0 \\ \frac{1}{T}\sum_{t=0}^{T} m_{3t}(\hat{\lambda}) &= \frac{1}{T}\sum_{t=0}^{T} (\Delta c_t - \hat{\lambda}\Delta w_t)\Delta c_{t-3} = 0 \\ \frac{1}{T}\sum_{t=0}^{T} m_{4t}(\hat{\lambda}) &= \frac{1}{T}\sum_{t=0}^{T} (\Delta c_t - \hat{\lambda}\Delta w_t)\Delta c_{t-4} = 0. \end{aligned}$$

The GMM estimate $\hat{\lambda}_{GMM}$ minimizes

$$q = \bar{m}(\lambda)' W_T \bar{m}(\lambda)$$

where $W_T^{-1} = \Phi$ is the asymptotic variance of $\sqrt{n}\bar{m}(\lambda)$. See Gauss example #12 for OLS, 2SLS and GMM estimates of λ.

1.2.5 Testing the Validity of the Overidentification Restrictions

In an exactly identified system, $q = 0$. In an overidentified system, the moment restrictions implied by theory will not all be satisfied exactly in the data. Therefore, $q > 0$. This observation forms the basis for a test of overidentifying restrictions. If q is substantially greater than zero, then this suggests that at least one of the overidentifying restrictions is likely to be false. Similar to the Wald test introduced in earlier chapters, we have

$$nq = [\sqrt{n}\bar{m}(\hat{\theta})]' \hat{\Phi}^{-1} [\sqrt{n}\bar{m}(\hat{\theta})] \overset{asy}{\sim} \chi^2[L - K].$$