3. The Generalized Method of Moments

The Generalized Method of Moments, as the name suggest, can be thought of just as a generalization of the classical MM.

A key in the GMM is a set of population moment conditions that are derived from the assumptions of the econometric model.

Example 3.1 In classical linear regression

$$y_t = \mathbf{x}_t'\beta + e_t,$$

where $\mathbf{x}_t = (x_{1t}, \ldots, x_{mt})'$ is a m-vector of explanatory variables and $\boldsymbol{\beta}$ is an m-vector of regression coefficients, and e_t is an error term.

The moment conditions are:

(i) $\mathbb{V}\text{ar}[e_t] = \sigma^2$ a constant for all t
(ii) $\mathbb{E}[(y_t - \mathbf{x}_t'\boldsymbol{\beta})x_t] = \mathbb{E}[e_t\mathbf{x}_t] = 0$ for all t
(iii) $\mathbb{E}[e_t e_u] = 0$ for all $t \neq u$,

of which (*ii*) is the key condition in estimating $\boldsymbol{\beta}$.

Given data on the observable variables the GMM finds values for the model parameters such that corresponding sample moment conditions are satisfied as closely as possible.

Example 3.2. (Ex 3.1 continued). Using only the moment conditions (*ii*), given T observations, the implied sample moment is

$$\frac{1}{T}\sum_{t=1}^{T}(y_t - \mathbf{x}_t'\beta)\mathbf{x}_t,$$

and the task is to select a value $\widehat{\beta}$ of β such that

$$\frac{1}{T}\sum_{t=1}^{T}(y_t - \mathbf{x}_t'\widehat{\beta})\mathbf{x}_t \approx 0.$$

In fact, provided $T \geq m$, selecting $\widehat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ (OLS estimator), the empirical moment condition is exactly satisfied. Note that

$$\sum_{t=1}^{T}(y_t - \mathbf{x}_t'\beta)\mathbf{x}_t = \mathbf{X}'\mathbf{y} - (\mathbf{X}'\mathbf{X})\beta,$$

where $\mathbf{X}$ is the $(T \times m)$ data matrix with $\mathbf{x}_t'$ in line t, and $y_t = (y_1, \ldots, y_T)'$. Thus the OLS and GMM estimators are here the same.

Example 3.3. (Example 2.2 continued): Suppose $\nu > 4$, then we can calculate

$$\mu_4 = E\left[y^4\right] = \frac{3\nu^2}{(\nu-2)(\nu-4)}$$

and the corresponding sample moment is

$$\hat{\mu}_4 = \frac{1}{T}\sum_{t=1}^{T} y_t^4.$$

Thus the moment conditions implied by the model are

$$\mathbb{E}\left[\begin{pmatrix} y^2 - \frac{\nu}{\nu-2} \\ y^4 - \frac{3\nu^2}{(\nu-2)(\nu-4)} \end{pmatrix}\right] = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Generally we cannot find a single value for ν that satisfies exactly the corresponding sample moment conditions

$$\frac{1}{T}\sum_{t=1}^{T}\begin{pmatrix} y_t^2 - \frac{\nu}{\nu-2} \\ y_t^4 - \frac{3\nu^2}{(\nu-2)(\nu-4)} \end{pmatrix} = \begin{pmatrix} \hat{\mu}_2 - \frac{\nu}{\nu-2} \\ \hat{\mu}_4 - \frac{3\nu^2}{(\nu-2)(\nu-4)} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

(two equations and one unknown).

However, we can choose ν so that both equations are satisfied as closely as possible. The closeness is measured in terms of (weighted) squared errors (c.f. Least Squares), resulting to minimizing criterion function

$$Q(\nu; \mathbf{y}) = \mathbf{g}_T' \mathbf{W} \mathbf{g}_T,$$

where $\mathbf{y} = (y_1, \ldots, y_T)'$,

$$\mathbf{g}_T = \mathbf{g}_T(\nu; \mathbf{y}) = \begin{pmatrix} \hat{\mu}_2 - \frac{\nu}{\nu-2} \\ \hat{\mu}_4 - \frac{3\nu^2}{(\nu-2)(\nu-4)} \end{pmatrix} = \begin{pmatrix} g_{1,T} \\ g_{2,T} \end{pmatrix},$$

and $\mathbf{W}$ is a suitable weighting matrix (2×2).

If we choose $\mathbf{W} = \mathbf{I}$, the identity matrix, the solution is a kind of nonlinear least squares. Improved results are, however, achieved if the "less noisy" moment conditions are weighted more than the "noisier" ones. Matrix $\mathbf{W}$ serves for this purpose.

It turns out that an optimal weighting matrix is the inverse of

$$\mathbf{S} = \lim_{T\to\infty} T \cdot E\left[(\mathbf{g}_T(\boldsymbol{\theta}_0;\mathbf{y}))\,(\mathbf{g}_T(\boldsymbol{\theta}_0;\mathbf{y}))'\right],$$

where $\mathbb{E}\left[(\mathbf{g}_T(\boldsymbol{\theta}_0;\mathbf{y}))\,(\mathbf{g}_T(\boldsymbol{\theta}_0;\mathbf{y}))'\right]$ is the covariance matrix of the average error terms (residuals)

$$g_{1,T} = \frac{1}{T}\sum_{t=1}^{T}\left(y_t^2 - \frac{\nu_0}{\nu_0 - 2}\right) = \frac{1}{T}\sum_{t=1}^{T} u_{1,t}$$

and

$$g_{2,T} = \frac{1}{T}\sum_{t=1}^{T}\left(y_t^4 - \frac{3\nu_0^2}{(\nu_0 - 2)(\nu_0 - 4)}\right) = \frac{1}{T}\sum_{t=1}^{T} u_{2,t},$$

where

$$u_{1,t} = y_t^2 - \frac{\nu_0}{\nu_0 - 2}$$

and

$$u_{2,t} = y_t^4 - \frac{3\nu_0^2}{(\nu_0 - 2)(\nu_0 - 4)}.$$

Hansen's two step GMM procedure*

Let $\mathbf{x}_t$ be an $s \times 1$ vector of variables that are observed at date t, let $\boldsymbol{\theta}$ denote the $m \times 1$ unknown parameter vector, and let $\mathbf{u}_t = \mathbf{u}(\mathbf{x}_t; \boldsymbol{\theta})$ be an $r \times 1$ covariance stationary† vector valued function, such that for true parameter value $\boldsymbol{\theta}_0$

$$\mathbb{E}\left[\mathbf{u}_t\right] = \mathbb{E}\left[\mathbf{u}(\mathbf{x}_t; \boldsymbol{\theta}_0)\right] = 0. \tag{1}$$

In GMM function $\mathbf{u}(\mathbf{x}; \boldsymbol{\theta})$ define the *moment* or more generally the *orthogonality* conditions of the model (sometimes called also the residuals of the model).

*Hansen, L. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, 50, 1029–1054.

†(1) $\mathbb{E}[\mathbf{u}_t] = \mu$ for all t
(2) $\mathbb{C}\text{ov}[\mathbf{u}_t, \mathbf{u}_{t+j}] = \mathbb{E}\left[(\mathbf{u}_t - \mu)(\mathbf{u}_{t+j} - \mu)'\right] = \mathbf{S}_j$ for all t. Note: $\mathbb{C}\text{ov}[\mathbf{u}_t, \mathbf{u}_{t-j}] = \mathbf{S}_{-j} = \mathbf{S}_j'$.

A sample counterpart of the expected value in (1) is the sample average

$$(2) \qquad \mathbf{g}_T(\boldsymbol{\theta}) = \frac{1}{T}\sum_{t=1}^{T} \mathbf{u}(\mathbf{x}_t;\boldsymbol{\theta}) = \frac{1}{T}\sum_{t=1}^{T} \mathbf{u}_t.$$

For the true parameter value $\boldsymbol{\theta}_0$, $\mathbf{g}_T(\boldsymbol{\theta}_0)$ measures the average sampling error with $\mathbb{E}\left[\mathbf{g}_T(\boldsymbol{\theta}_0)\right] = 0$.

Consequently an estimator of $\boldsymbol{\theta}$ is selected such that $\mathbf{g}_T(\boldsymbol{\theta})$ becomes as close as possible to zero.

In the case where there are equally many moment conditions as estimated parameters a unique estimator for $\boldsymbol{\theta}_0$ can be selected such that the average sampling error (2) becomes exactly equally to zero.

In the general case, where there are more moment conditions than parameters the estimator for $\boldsymbol{\theta}_0$ is a "compromise" that makes (2) close to zero.

This is achieved in the GMM by selecting the estimator for $\boldsymbol{\theta}_0$ such that the sampling error with respect to the estimated value is as small as possible in the (generalized) least squares sense.

That is, the GMM estimator $\widehat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}_0$ is the value of $\boldsymbol{\theta}$ that minimizes

$$Q(\boldsymbol{\theta}) = \mathbf{g}_T(\boldsymbol{\theta})'\mathbf{W}\mathbf{g}_T(\boldsymbol{\theta}). \tag{3}$$

where the prime denotes matrix transposition and $\mathbf{W}$ is a suitably chosen weighting matrix.

Choosing the weighting matrix $\mathbf{W}$:

The weighting matrix $\mathbf{W}$ determines how each moment condition is weighted in the estimation.

The principle is that more accurate (less noisy) moment conditions should be weighted more than the less accurate (more noisy or uncertain) ones.

The accuracy of the moment conditions can be measured by the variance covariance matrix (recall, $\mathbb{E}[\mathbf{u}_t] = \mathbf{0}$ and $\mathbb{E}[\mathbf{g}_T(\boldsymbol{\theta})] = 0$ for the true parameter value)

$$\begin{aligned}
\mathbb{C}\text{ov}\left[\mathbf{g}_T(\boldsymbol{\theta})\right] &= \mathbb{E}\left[\mathbf{g}_T(\boldsymbol{\theta})\mathbf{g}_T(\boldsymbol{\theta})'\right] \\
&= \tfrac{1}{T^2}\mathbb{E}\left[\left(\textstyle\sum_{t=1}^T \mathbf{u}_t\right)\left(\textstyle\sum_{t=1}^T \mathbf{u}_t\right)'\right] \qquad (4)\\
&= \tfrac{1}{T^2}\textstyle\sum_{s=1}^T\sum_{t=1}^T \mathbb{E}\left[\mathbf{u}_s\mathbf{u}_t'\right]
\end{aligned}$$

Let $j = |t - s|$, then under the assumption of stationarity of $\mathbf{u}_t$

(5) $$\mathbb{E}\left[\mathbf{u}_t \mathbf{u}'_{t+j}\right] = \mathbf{S}_j = \mathbf{S}'_{-j}$$

for all t. Thus in (4) we can write

(6) $$\sum_{s=1}^{T} \sum_{t=1}^{T} \mathbb{E}\left[\mathbf{u}_s \mathbf{u}'_t\right] = \sum_{j=-T}^{T} (T - |j|)\mathbf{S}_j,$$

or by using (5)

(7) $$\sum_{s=1}^{T} \sum_{t=1}^{T} \mathbb{E}\left[\mathbf{u}_s \mathbf{u}'_t\right] = T\,\mathbf{S}_0 + \sum_{j=1}^{T}(T - j)(\mathbf{S}_j + \mathbf{S}'_j).$$

Under the stationarity and some technical assumptions (see, Hansen 1982), it can be shown that

$$(8)\qquad \lim_{k\to\infty} \sum_{j=-k}^{k} \mathbf{S}_j = \mathbf{S},$$

where $\mathbf{S}$ is a positive definite matrix, called the <u>long run covariance matrix</u> of $\mathbf{u}_t$.

Thus, because $(T - |j|)/T \to 1$ as $T \to \infty$

$$(9)\qquad \mathbb{C}\text{ov}\left[\sqrt{T}\mathbf{g}_T\right] \to \sum_{j=-\infty}^{\infty} \mathbf{S}_j = \mathbf{S},$$

as $T \to \infty$.

Let $\hat{\mathbf{u}}_t$ denote observations on $\mathbf{u}_t$s, $t = 1, \ldots, T$. Then the autocovariance matrices $\mathbf{S}_j$ are estimated by

$$(10) \qquad \hat{\mathbf{S}}_j = \frac{1}{T} \sum_{t=j+1}^{T} \hat{\mathbf{u}}_t \hat{\mathbf{u}}'_{t-j},$$

$j = 0, 1, \ldots, \ell$, where ℓ is the selected maximum lag length. The long-run covariance matrix is estimated then by

$$(11) \qquad \hat{\mathbf{S}} = \hat{\mathbf{S}}_0 + \sum_{j=1}^{\ell} w_j (\hat{\mathbf{S}}_j + \hat{\mathbf{S}}'_j),$$

where w_js are weights. If $w_j \equiv 1$ then all lag lengths are equally weighted.

Usually, however, the more distant lags are weighted less. One popular weighting scheme is the Bartlett weights

$$(12) \qquad w_j = 1 - j/(\ell + 1).$$

The two step procedure:

Step 1: Set $\mathbf{W} = \mathbf{I}$, the identity matrix and solve the (nonlinear) least squares problem

$$\hat{\boldsymbol{\theta}}^{(1)} = \underset{\theta}{\text{argmin}}\, \mathbf{g}_T(\boldsymbol{\theta})'\mathbf{g}_T(\boldsymbol{\theta}). \tag{13}$$

Step 2: Compute

$$\hat{\mathbf{u}}_t = \mathbf{u}(\mathbf{x}_t; \hat{\boldsymbol{\theta}}^{(1)}) \tag{14}$$

and estimate $\mathbf{S}_j$ as

$$\hat{\mathbf{S}}_j = \frac{1}{T} \sum_{t=j+1}^{T} \hat{\mathbf{u}}_t \hat{\mathbf{u}}_{t+j}, \tag{15}$$

$j = 0, 1, \ldots, \ell$.

Estimate $\mathbf{S}$ by

$$\hat{\mathbf{S}} = \hat{\mathbf{S}}_0 + \sum_{j=1}^{\ell} w_j (\hat{\mathbf{S}}_j + \hat{\mathbf{S}}_j'). \tag{16}$$

Select $\mathbf{W} = \widehat{\mathbf{S}}^{-1}$ and obtain the second step estimate

$$\widehat{\boldsymbol{\theta}}^{(2)} = \underset{\theta}{\text{argmin}}\, \mathbf{g}_T(\boldsymbol{\theta})'\mathbf{W}\mathbf{g}_T(\boldsymbol{\theta}). \tag{17}$$

Remark 3.1: An iterative GMM iterates Step 2 until convergence of the θ-estimate, i.e., when after k^{th} step $\widehat{\boldsymbol{\theta}}^{(k+1)} \approx \widehat{\boldsymbol{\theta}}^{(k)}$.

Instruments

Suppose our suggested model is

$$y_t = f(\mathbf{x}_t;\theta) + e_t, \tag{18}$$

where $f(\cdot)$ is some function, $\mathbf{x}$ is some background information, θ contains the model parameters, and $e_t = y_t - f(\mathbf{x}_t;\theta)$ is the residual with

$$\mathbb{E}[e_t] = \mathbb{E}[y_t - f(\mathbf{x}_t;\theta_0)] = 0. \tag{19}$$

Thus this implies one moment or orthogonality condition.

In order $f(\mathbf{x}_t; \theta)$ to capture all the systematic variation in y, the residual should be uncorrelated with all potential additional variables (predictors).

Let $\mathbf{z}_t$ be a vector of these potential predictors observable at time t, the residuals $y_t - f(\mathbf{x}_t; \theta_0)$ should be uncorrelated with the components of $\mathbf{z}_t$.

Technically this can be denoted as

$$\mathbb{E}\left[(y - f(x;\theta)) \otimes \mathbf{z}\right] = 0, \tag{20}$$

where $\otimes$ is the Kronecker product,*

*E.g., $\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ and $\mathbf{B} = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}$ then

$$\begin{aligned} \mathbf{A} \otimes \mathbf{B} &= \begin{pmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} \\ a_{12}\mathbf{B} & a_{22}\mathbf{B} \end{pmatrix} \\ &= \begin{pmatrix} a_{11}b_{11} & a_{11}b_{12} & a_{12}b_{11} & a_{12}b_{12} \\ a_{11}b_{21} & a_{11}b_{22} & a_{12}b_{21} & a_{12}b_{22} \\ a_{21}b_{11} & a_{21}b_{12} & a_{22}b_{11} & a_{22}b_{12} \\ a_{21}b_{21} & a_{21}b_{22} & a_{22}b_{21} & a_{22}b_{22} \end{pmatrix}. \end{aligned} \tag{21}$$

Note that (20) includes condition (20) (defining e.g. the first component of $\mathbf{z}_t$, $z_{1t} = 1$).

Variables in $\mathbf{z}_t$ are called <u>instrument</u> variables.

The $\mathbf{u}$-function in equation (1) generalizes now

$$\mathbf{u}_t = \mathbf{u}_t(\boldsymbol{\theta}) = (y_t - f(\mathbf{x}_t; \boldsymbol{\theta})) \otimes \mathbf{z}_t, \tag{22}$$

with

$$\mathbb{E}[\mathbf{u}_t] = 0. \tag{23}$$

Example 3.4 Let $\mathbf{X}_t = (X_t^a, X_t^b)' = (1 + R_t^a, 1 + R_t^b)'$ returns of two assets, $M_t(\theta)$ a stochastic discount factor, and $\mathbf{z}_t = (1, z_t)'$ an instrument. Then

$$\mathbf{u}_t = (M_{t+1}(\boldsymbol{\theta})\mathbf{X}_{t+1} - 1) \otimes \mathbf{z}_t = \begin{pmatrix} M_{t+1}(\theta)\, X_{t+1}^a - 1 \\ M_{t+1}(\theta)\, X_{t+1}^a z_t - z_t \\ M_{t+1}(\theta)\, X_{t+1}^b - 1 \\ M_{t+1}(\theta)\, X_{t+1}^b z_t - z_t \end{pmatrix} \tag{24}$$

with

$$\mathbb{E}\left[\mathbf{u}_t\right] = \mathbb{E}\left[\begin{pmatrix} M_{t+1}(\theta)\, X_{t+1}^a - 1 \\ M_{t+1}(\theta)\, X_{t+1}^a z_t - z_t \\ M_{t+1}(\theta)\, X_{t+1}^b - 1 \\ M_{t+1}(\theta)\, X_{t+1}^b z_t - z_t \end{pmatrix}\right] = \mathbf{0}. \tag{25}$$

If $m = r$, where m is the dimension of θ and r is the number of moment conditions, then there are equally many parameters to be estimated as orthogonality conditions, and

$$\mathbf{g}(\widehat{\boldsymbol{\theta}}) = \frac{1}{T}\sum_{t=1}^{T} \mathbf{u}_t = 0 \tag{26}$$

can be solved exactly. We say that the problem is exactly identified.

Example 3.5: OLS as an exactly identified GMM (Example 3.1 revisited).

Assume again the standard regression model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}, \tag{27}$$

where

$$\mathbf{y} = (y_1, \ldots, y_T)'$$

is an $(T \times 1)$ vector (prime denotes the transposition), $\mathbf{X}$ is an $(T \times m)$, $m < T$, full rank matrix,

$$\boldsymbol{\beta} = (\beta_1, \ldots, \beta_m)'$$

is a $(m \times 1)$ parameter vector, and

$$\mathbf{e} = (e_1, \ldots, e_T)'$$

is an $(T \times 1)$ error vector with $\mathbb{E}[\mathbf{e}] = 0$ and $\mathbb{C}\text{ov}[\mathbf{e}] = \sigma_e^2\mathbf{I}$, with $\mathbf{I}$ an $(T \times T)$ identity matrix, and $\mathbf{e}$ is independent of $\mathbf{X}$.

The orthogonality conditions are

$$\mathbb{E}[\mathbf{u}_t] = E[(y_t - \mathbf{x}_t'\boldsymbol{\beta})\mathbf{x}_t] = \mathbb{E}[e_t\mathbf{x}_t] = 0, \tag{28}$$

where $\mathbf{x}_t'$ is the t^{th} row of the $\mathbf{X}$ matrix, $t = 1, \ldots, T$.

Now $m = \dim(\boldsymbol{\beta}) = \dim(\mathbf{u}_t)$. Thus the problem is exactly identified.

The sample counterpart of (28) is

$$\mathbf{g}_T(\boldsymbol{\beta}) = \frac{1}{T}\sum_{t=1}^{T}(y_t - \mathbf{x}_t'\boldsymbol{\beta})\mathbf{x}_t = \frac{1}{T}\mathbf{X}'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}). \tag{29}$$

Setting this equal to zero, multiplying by the T, and arranging terms (denote the beta satisfying the equation by $\widehat{\boldsymbol{\beta}}$) yields

$$\mathbf{X}'\mathbf{y} - \mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{0}, \tag{30}$$

and we get finally

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}, \tag{31}$$

the same as the OLS estimator.

Remark 3.2: Under the assumptions of the above example $\mathbf{x}_t$ and e_t are independent, which implies that $\mathbb{E}[\mathbf{x}_t e_t] = \mathbb{E}[\mathbf{x}_t]\mathbb{E}[e_t] = 0$, and hence

(32) $$\mathbb{E}[\mathbf{u}_t\mathbf{u}_t'] = \mathbb{E}[e_t^2]\mathbb{E}[\mathbf{x}_t\mathbf{x}_t'] = \sigma_e^2\mathbb{E}[\mathbf{x}_t\mathbf{x}_t'].$$

and

(33) $$\mathbb{E}[\mathbf{u}_t\mathbf{u}_{t+j}'] = \mathbb{E}[\mathbf{x}_t\mathbf{x}_{t+j}]\mathbb{E}[e_t e_{t+j}] = 0$$

because of no autocorrelation in the residuals.

These imply

(34) $$\mathbf{S} = \mathbf{S}_0 = \sigma_e^2\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[\mathbf{x}_t\mathbf{x}_t'].$$

Estimating $\sum_{t=1}^{T}\mathbb{E}[\mathbf{x}_t\mathbf{x}_t']$ by

(35) $$\mathbf{X}'\mathbf{X} = \sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'$$

minimization of

(36) $$\mathbf{g}_T(\boldsymbol{\beta})'\mathbf{W}\mathbf{g}_T(\boldsymbol{\beta})$$

with

(37) $$\mathbf{W} = \frac{T}{\sigma_e^2}(\mathbf{X}'\mathbf{X})^{-1}$$

results exactly again the OLS estimator. However, the solution is obtained directly by solving $\mathbf{g}_T(\boldsymbol{\beta}) = 0$, because the problem is exactly identified.

It may be further noted that in this exactly identified case the solution is independend of of the $(m \times m)$ p.d. matrix $\mathbf{W}$, because the first order condition of the minimum of (36) is

$$\partial \left(\mathbf{g}_T(\boldsymbol{\beta})'\mathbf{W}\mathbf{g}_T(\boldsymbol{\beta}) \right) / \partial \boldsymbol{\beta} = 2\mathbf{D}'\mathbf{W}\mathbf{g}_T(\boldsymbol{\beta}) = 0, \tag{38}$$

where $\mathbf{D} = \partial \mathbf{g}_T(\boldsymbol{\beta})/\partial\boldsymbol{\beta} = -\mathbf{X}'\mathbf{X}/T$. Multiplying both sides of (38) from the left by the inverse of $\mathbf{D}'\mathbf{W}$ (which is p.d.) leads again to the OLS solution.

Remark 3.3: Even in the case of exact identification one can utilize the flexibility and (plausible) robustness of GMM by giving up the heteroscedasticity and no-autocorrelation assumptions in standard OLS.

Example 3.6: Regression with autocorrelated residuals of unknown form.

If in Example 3.5 we allow autocorrelation of unspecified form in the regression errors

$$e_t = y_t - \mathbf{x}_t'\boldsymbol{\beta},$$

but assuming, however, stationarity of the moment condition errors

$$\mathbf{u}_t = (y_t - \mathbf{x}_t'\boldsymbol{\beta})\mathbf{x}_t = e_t\mathbf{x}_t,$$

we get

$$\mathbf{S}_j = \mathbb{E}[\mathbf{u}_t\mathbf{u}_{t+j}'] = \mathbb{E}[e_te_{t+j}\mathbf{x}_t\mathbf{x}_{t+j}']. \tag{39}$$

Using the OLS residuals, $\hat{e}_t = y_t - \mathbf{x}_t'\hat{\boldsymbol{\beta}}_{\text{OLS}}$, from the first step, we can estimate $\mathbf{S}$ in the second step by

$$\hat{\mathbf{S}} = \hat{\mathbf{S}}_0 + \sum_{j=1}^{\ell} w_j(\hat{\mathbf{S}}_j + \hat{\mathbf{S}}_j'), \tag{40}$$

where

$$\hat{\mathbf{S}}_j = \frac{1}{T}\sum_{t=j+1}^{T} \hat{e}_t\hat{e}_{t-j}\mathbf{x}_t\mathbf{x}_{t-j}', \tag{41}$$

$j = 0, 1, \ldots, \ell$.

Note that this accounts for both the unspecified autocorrelation and heteroscedasticity.

If $m < r$ then there are more orthogonality conditions than parameters to estimate. The above equality does not hold exactly in sample data. It is said that the model is over identified, and one can empirically test the over identification constraints. If the hypothesis is rejected it indicates that the data does not support the estimated model.

The null hypothesis

$$H_0 : \mathbb{E}[\mathbf{u}_t(\boldsymbol{\theta})] = 0 \tag{42}$$

the over identification conditions can be tested with the statistic

$$J = \mathbf{g}(\hat{\boldsymbol{\theta}})'\mathbf{W}\mathbf{g}(\hat{\boldsymbol{\theta}}_T), \tag{43}$$

where TJ is asymptotically χ^2_{r-m} distributed under the null hypothesis.

Some properties of the GMM estimators

Because the GMM estimators are sums of (approximately) independent random variables the Central Limit Theorem implies that

$$\hat{\boldsymbol{\theta}} \sim AN(\boldsymbol{\theta}, \hat{\mathbf{V}}/T), \tag{44}$$

where

$$\hat{\mathbf{V}} = \left(\hat{\mathbf{D}}\hat{\mathbf{S}}^{-1}\hat{\mathbf{D}}'\right)^{-1}, \tag{45}$$

where

$$\hat{\mathbf{D}} = \left.\frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\theta=\hat{\theta}}. \tag{46}$$

Furthermore under general regularity conditions, it can be shown that

$$\hat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}_0, \tag{47}$$

i.e. they are consistent.