Generalised Method Of Moments

Specification on moment conditions rather than the full (ML).

Classical Method of Moments

Let $\{y_1, ..., y_n\}$ be random sampled from a distribution with density $f(y; \theta)$. Assume that $E(y_t^i) = \mu_i(\theta)$, $i = 1, ..., k$. (k population moments can be calculated as functions of θ). The classical method of moments estimates θ by $\hat{\theta}$, where $\hat{\theta}$ is such that

$$\mu_i(\hat{\theta}) = \frac{1}{n}\sum_{t=1}^{n} y_t^i, \; i = 1, ..., k$$

e.g.

$$f(y_t; v) = \frac{\Gamma\left(\frac{1+v}{2}\right)}{\sqrt{\pi v}\,\Gamma\left(\frac{v}{2}\right)}(1 + \frac{y_t^2}{v})^{-\frac{1+v}{2}} \quad [t(v)]$$

$$E(y_t) = 0, \; E(y_t^2) = \frac{v}{v-2}, \; v > 2$$

$$\hat{\mu}_{2n} = \frac{1}{n}\sum_{t=1}^{n} y_t^2, \; \hat{\mu}_2 \xrightarrow{p} \mu_2$$

$$\hat{\mu}_{2n} = E(y_t^2) \Longrightarrow \hat{\mu}_{2n} = \frac{v}{v-2} \Longrightarrow \widehat{v} = \frac{2\hat{\mu}_{2n}}{\hat{\mu}_{2n} - 1}, \; \hat{\mu}_2 > 1$$

(if $\hat{\mu}_{2n} = 1$, $\hat{v}$ would be infinity: an $N(0,1)$ distribution fits the sample second moment better than any member of the t family).

Generalised Method Of Moments

In deriving an estimate of v we might use more than one moment.
If $v > 4$,

$$\mu_4 = E(y_t^4) = \frac{3v^2}{(v-2)(v-4)}$$

$$\hat{\mu}_{4n} = \frac{1}{n}\sum_{t=1}^{n} y_t^4$$

We cannot choose v so as to match both the sample second and fourth moment, but we can try choosing it so as to be as close as possible to both by minimising a criterion function such as:

$$Q = g'Wg$$

$$g = \begin{bmatrix} \hat{\mu}_{2n} - \mu_2 \\ \hat{\mu}_{4n} - \mu_4 \end{bmatrix}$$

W is a (2×2) positive definite symmetric weighting matrix reflecting the importance given to matching each of the moments.

GMM Hansen 1982

We will now present the GMM estimation procedure using Hansen (1982) notation.

Consider

w_t $= m \times 1$ a vector of variables observed at time t

θ $= k \times 1$ a vector of unknown parameters

θ_0 $= k \times 1$ a vector of true parameters values

Consider $h(\theta, w_t)$, $r \times 1$ vector-valued function, $h : (\Re^k \times \Re^m) \rightarrow \Re^r$ and suppose that θ_0 (the DGP) is characterised by the following moment conditions (orthogonality condition)

$$E[h(\theta_0, w_t)] = 0$$

The empirical moments (sample counterpart) are

$$\underset{r\times 1}{g(\theta, \eta_n)} = \frac{1}{n}\sum_{t=1}^{n} h(\theta, w_t)$$

$$\eta_n = (w_1', w_2', ..., w_n') \ (nm \times 1)$$

GMM chooses θ so as to make $g(\theta, \eta_n)$ as close as possible to zero (the population moments). That is, the GMM estimator $\hat{\theta}$ is defined as

$$\hat{\theta} = \arg\min_{\theta} Q(\theta, \eta_n)$$

where

$$Q(\theta, \eta_n) = [g(\theta, \eta_n)]' W_n [g(\theta, \eta_n)]$$

and $\{W_n\}_{n=1}^{\infty}$ is a sequence of $(r \times r)$ positive definite weighting matrices which may be a function of η_n

[if $k = r$, $g(\hat{\theta}, \eta_n) = 0$]

For the above example using this notation we have that

$$h(\theta, w_t) = \begin{bmatrix} y_t^2 - \frac{v}{v-2} \\ y_t^4 - \frac{3v^2}{(v-2)(v-4)} \end{bmatrix}$$

NB The *classical method of moments* is a special case of GMM where

- $w_t = y_t$; $\theta = \nu$; $W_n = 1$; $h(\theta, w_t) = y_t^2 - \frac{v}{v-2}$; $g(\theta, \eta_n) = \frac{1}{n}\sum_{t=1}^{n} y_t^2 - \frac{v}{v-2}$ and $Q(\theta, \eta_n) = \left(\frac{1}{n}\sum_{t=1}^{n} y_t^2 - \frac{v}{v-2}\right)^2$. Clearly the smallest value is obtained for $Q(\theta, \eta_n) = 0$ and $\widehat{v}_n = \frac{2\hat{\mu}_{2n}}{\hat{\mu}_{2n}-1}$.

Number of parameter to estimate and orthogonality conditions

In general if the number of parameters to be estimated, k, is the same that the number of orthogonality conditions r, then the objective function is minimised by setting $g(\widehat{\theta}_n, \eta_n) = 0$

If $r > k$ then how close to zero would each of the orthogonality conditions be, will depend on its weight on the weighting matrix W_n.

For any value of θ, the magnitude of the $(rx1)$ vector $g(\theta, \eta_n)$ is the sample mean of the n realizations of the $(rx1)$ random vector $h(\theta, w_t)$.

If w_t is strictly stationary and $h(\cdot)$ is continuous, the LLN implies that

$$g(\theta, \eta_n) \xrightarrow{p} E[h(\theta, w_t)]$$

Suppose that θ_0 is the only value that satisfies that $E[h(\theta_0, w_t)] = 0$, *then* $\widehat{\theta}_n$ such that it minimizes $[g(\theta, \eta_n)]' W_n [g(\theta, \eta_n)]$ gives a consistent estimate of θ_0, i.e., $\widehat{\theta}_n \xrightarrow{P} \theta_0$.

Optimal Weighting Matrix

Suppose that when evaluated at the true value θ_0, the process $\{h(\theta_0, w_t)\}_{-\infty}^{\infty}$ is strictly stationary with mean zero and autocovariance matrix given by

$$\Gamma_j = E\left\{[h(\theta_0, w_t)][h(\theta_0, w_{t-j})]'\right\}.$$

Assuming that these autocovariances are absolutely summable, define

$$S \equiv \sum_{j=-\infty}^{\infty} \Gamma_j$$

then S is the asymptotic variance of the sample mean of $h(\theta_0, w_t)$, $g(\theta_0, \eta_n)$ defined as [1]

$$S = \lim_{n\to\infty} nE\{[g(\theta, \eta_n)][g(\theta, \eta_n)]'\}$$

The optimal value for the weighting matrix W_n turns out to be S^{-1}, the inverse of the asymptotic variance. The minimum asymptotic variance for $\hat{\theta}_n$ is obtained when $\hat{\theta}_n$ is chosen to minimse

$$Q(\theta, \eta_n) = [g(\theta, \eta_n)]' S^{-1} [g(\theta, \eta_n)].$$

S can be consistently estimated by

$$\widehat{S}_n = \hat{\Gamma}_0 + \sum_{j=1}^{q} w(j, q)(\hat{\Gamma}_j + \hat{\Gamma}'_j)$$

$$\hat{\Gamma}_j = \frac{1}{n} \sum_{t=j+1}^{n} [h(\hat{\theta}, w_t)][h(\hat{\theta}, w_{t-j})]'$$

where $w(j, q)$ is a lag window

Since $\widehat{S}_n$ requires an estimate of θ_0, the following iterative procedure is used in practice which it turns out to be correct since $\widehat{S}_n = \hat{\Gamma}_0 + \sum_{j=1}^{q} w(j, q)(\hat{\Gamma}_j + \hat{\Gamma}'_j) \xrightarrow{P} S$:

1. Obtain $\hat{\theta}_0$ using an arbitrary weighting matrix such as $W_n = I_r$ and produce $\widehat{S}_n^{(0)}$

[1] If y_t is a covariance stationary process such that

- $E(y_t) = \mu$
- $\Gamma_j = E\left\{(y_t - \mu)(y_{t-j} - \mu)'\right\}.$

Then $\overline{y}_n \xrightarrow{P} \mu$ and $\lim_{n\to\infty} nE\{(\overline{y}_n - \mu)(\overline{y}_n - \mu)'\} = \sum_{j=-\infty}^{\infty} \Gamma_j$

2. Use $\widehat{S}_n^{(0)}$ to arrive at a new GMM estimate $\hat{\theta}_1$

3. Repeat the process until $\hat{\theta}_{j+1} \approx \hat{\theta}_j$

Asymptotic Properties

Let $\hat{\theta}_n$ be the value that minimizes

$$[g(\theta, \eta_n)]' \widehat{S}_n^{-1} [g(\theta, \eta_n)]$$

Thus, the GMM estimate $\hat{\theta}_n$ is the solution to the following system of non-linear equations:

$$\underbrace{\left\{ \frac{\partial g(\theta, \eta_n)}{\partial \theta'} \bigg|_{\theta=\hat{\theta}_n} \right\}}_{(kxr)} \underbrace{\widehat{S}_n^{-1}}_{(rxr)} \underbrace{\left[g\left(\hat{\theta}_n, \eta_n\right) \right]}_{(rx1)} = 0$$

Since $g(\theta, \eta_n)$ is the sample mean of a process whose population mean is zero, then $g(\cdot)$ should satisfy the central limit theorem[2]. Given that W_n is strict stationary and $h(\theta, w_n)$ is continuous

$$\Rightarrow \sqrt{n}\,[g(\theta_0, \eta_n)] \xrightarrow{L} N(0, S)$$

Let $g(\theta, \eta_n)$ be differentiable in θ for all η_n, and let $\hat{\theta}$ be the GMM with $r \geq k$. Let $\{\widehat{S}_n\}$ be a sequence of positive definite $(r \times r)$ matrices such that $\widehat{S}_n \xrightarrow{p} S > 0$. Suppose, further, that the following hold:

1. $\hat{\theta} \xrightarrow{p} \theta_0$

2. $\sqrt{n} g(\theta_0, \eta_n) \xrightarrow{d} N(0, S)$

3. for any sequence $\{\theta_n^*\}$ such that $\theta_n^* \xrightarrow{p} \theta_0$

$$p \lim_{n\to\infty} [\frac{\partial g(\theta, \eta_n)}{\partial \theta'}\bigg|_{\theta=\theta_n^*}] = p \lim_{n\to\infty} [\frac{\partial g(\theta, \eta_n)}{\partial \theta'}\bigg|_{\theta=\theta_0}] = \underset{(r\times k)}{D'}$$

with the columns of D' linearly independent
Then,

$$\begin{aligned} \sqrt{n}(\hat{\theta} - \theta_0) &\xrightarrow{d} N(0, V), \\ V &= (DS^{-1}D')^{-1} \end{aligned}$$

2

$$\sqrt{n}\,(\bar{x}_n - \mu) = \frac{1}{\sqrt{n}} \left(\sum x_i - \mu \right) \xrightarrow{d} N(0, \sigma^2)$$

$$\hat{\theta} \overset{\alpha}{\sim} N(\theta_0, \frac{1}{n}V)$$

with

$$\widehat{V} = (\hat{D}_n \widehat{S}_n^{-1} \hat{D}_n')^{-1}, \ \hat{D}_n' = \left. \frac{\partial g(\theta, \eta_n)}{\partial \theta'} \right|_{\theta=\hat{\theta}}$$

Consider the FOC

$$\underbrace{\left\{ \left. \frac{\partial g(\theta, \eta_n)}{\partial \theta'} \right|_{\theta=\hat{\theta}_n} \right\}'}_{(kxr)} \underbrace{\widehat{S}_n^{-1}}_{(rxr)} \underbrace{\left[g\left(\hat{\theta}_n, \eta_n\right) \right]}_{(rx1)} = 0$$

Note that we can do a Taylor expansion of $g\left(\hat{\theta}_n, \eta_n\right)$ around θ_0, i.e.:

$$g\left(\hat{\theta}_n, \eta_n\right) = g(\theta_0, \eta_n) + \left(\left. \frac{\partial g(\hat{\theta}_n, \eta_n)}{\partial \theta'} \right|_{\hat{\theta}_n=\theta_0} \right) \left(\hat{\theta}_n - \theta_0\right)$$

Then, substituting this expressions back in the FOC

$$\left\{ \left. \frac{\partial g(\theta, \eta_n)}{\partial \theta'} \right|_{\theta=\hat{\theta}_n} \right\}' \widehat{S}_n^{-1} \left\{ g(\theta_0, \eta_n) + \left(\left. \frac{\partial g(\hat{\theta}_n, \eta_n)}{\partial \theta'} \right|_{\hat{\theta}_n=\theta_0} \right) \left(\hat{\theta}_n - \theta_0\right) \right\} = 0$$

$$\left\{ \underset{\xrightarrow{P} D'}{\left. \frac{\partial g(\theta, \eta_n)}{\partial \theta'} \right|_{\theta=\hat{\theta}_n}} \right\}' \underset{\xrightarrow{P} S^{-1}}{\widehat{S}_n^{-1}} g(\theta_0, \eta_n) + \left\{ \underset{\xrightarrow{P} D'}{\left. \frac{\partial g(\theta, \eta_n)}{\partial \theta'} \right|_{\theta=\hat{\theta}_n}} \right\}' \underset{\xrightarrow{P} S^{-1}}{\widehat{S}_n^{-1}} \left\{ \underset{\xrightarrow{P} D'}{\left. \frac{\partial g(\hat{\theta}_n, \eta_n)}{\partial \theta'} \right|_{\hat{\theta}_n=\theta_0}} \right\} \left(\hat{\theta}_n - \theta_0\right) = 0$$

$$-\left(\left\{ \underset{\xrightarrow{P} D'}{\left. \frac{\partial g(\theta, \eta_n)}{\partial \theta'} \right|_{\theta=\hat{\theta}_n}} \right\}' \underset{\xrightarrow{P} S^{-1}}{\widehat{S}_n^{-1}} \left\{ \underset{\xrightarrow{P} D'}{\left. \frac{\partial g(\hat{\theta}_n, \eta_n)}{\partial \theta'} \right|_{\hat{\theta}_n=\theta_0}} \right\} \right)^{-1} \left\{ \underset{\xrightarrow{P} D'}{\left. \frac{\partial g(\theta, \eta_n)}{\partial \theta'} \right|_{\theta=\hat{\theta}_n}} \right\}' \underset{\xrightarrow{P} S^{-1}}{\widehat{S}_n^{-1}} g(\theta_0, \eta_n) = \left(\hat{\theta}_n - \theta_0\right)$$

We know that

$$\lim \sqrt{n} \left(\hat{\theta}_n - \theta\right) \xrightarrow{p} \underbrace{-\left(DS^{-1}D'\right)^{-1} DS^{-1}}_{C'} \sqrt{n} g(\theta_0, \eta_n)$$

But $\sqrt{n}\,(g(\theta_0, \eta_n)) \xrightarrow{L} N(0, S)$, hence

$$\sqrt{n} \left(\hat{\theta}_n - \theta\right) \xrightarrow{L} N(0, C'SC) = N\left(0, \left(DS^{-1}D'\right)^{-1}\right) = N(0, V)$$

And

$$\hat{\theta}_n \xrightarrow{L} N\left(\theta_0, \frac{\hat{V}_n}{N}\right)$$
$$\text{where } \hat{V}_n = \left[\hat{D}_n \widehat{S}_n^{-1} \hat{D}_n'\right]^{-1}$$

and the estimate of $\widehat{S}_n$ can be constructed as before.

Testing Overidentifying Restrictions:

Hansen´s J-Test

When $r > k$, we can test whether all the sample moments are as close to the population moments as we can expect if $E\left[h\left(\theta_0, w_t\right)\right] = 0$.

Knowing[3] that

$$\sqrt{n}g\left(\theta_0, \eta_n\right) \xrightarrow{L} N\left(0, S\right)$$

we can easily see that

$$\left[\sqrt{n}g\left(\theta_0, \eta_n\right)\right]' S^{-1} \left[\sqrt{n}g\left(\theta_0, \eta_n\right)\right] \xrightarrow{L} \chi^2_{(r)}$$

Notice that this holds when we evaluate $g\left(\theta, \eta_n\right)$ at θ_0, and we may well guess that it might also hold when evaluated at $\widehat{\theta}_n$. However, this is not true since the FOC implies that k different linear combination of $g\left(\widehat{\theta}_n, \eta_n\right) = 0$.

Since $g\left(\widehat{\theta}_n, \eta_n\right)$ contains $(r-k)$ non degenerate random variables, it turns out that a correct test for the identifying restrictions $r > k$ is

$$\begin{aligned} J &= \left[\sqrt{n}g\left(\widehat{\theta}_n, \eta_n\right)\right]' \widehat{S}_n^{-1} \left[\sqrt{n}g\left(\widehat{\theta}_n, \eta_n\right)\right] \\ &= n[g(\hat{\theta}_n, \eta_n)]' \widehat{S}_n^{-1} [g(\hat{\theta}_n, \eta_n)] \xrightarrow{L} \chi^2_{(r-k)} \end{aligned}$$

Unfortunately, it really fails to detect violations of the orthogonality conditions for misspecified models.

Inference with GMM

[3] Knowing that, if

$$Z_{(n\times 1)} \sim N\left(0, \Omega\right) \Longrightarrow Z'\Omega^{-1}Z \sim \chi^2_{(n)}$$

$$H_0 : r(\theta) = 0, \text{ where } r : \Theta \rightarrow \Re^q$$

$$R(\theta) = \frac{\partial r(\theta)}{\partial \theta}$$

$$W = n\hat{r}'[R(\hat{D}_n \widehat{S}_n^{-1} \hat{D}_n')^{-1} \hat{R}']\hat{r} \overset{a}{\sim} \chi^2(q)$$

Examples of GMM Estimators:

OLS

$$y_t = x_t'\beta + u_t$$

$E(x_t u_t) = 0$ justifies the use of OLS

$E[x_t(y_t - x_t'\beta_0)] = 0$ for the true value β_0

$$h(\theta, w_t) = x_t(y_t - x_t'\beta_0),\ \ w_t = (y_t, x_t')',\ \ \theta = \beta$$

$r = k$, so OLS is a just-identified GMM specification.
The GMM estimate $\hat{\beta}$ satisfies

$$g(\hat{\theta}, \eta_n) = \frac{1}{n}\sum_{t=1}^{n} x_t(y_t - x_t'\hat{\beta}) = 0$$

$$\sum_{t=1}^{n} x_t y_t = (\sum_{t=1}^{n} x_t x_t')\hat{\beta} \Longrightarrow \hat{\beta} = (\sum_{t=1}^{n} x_t x_t)^{-1}(\sum_{t=1}^{n} x_t y_t)$$

If u_t is not iid, GMM is not efficient (not as efficient as GLS). It is consistent, but the formulas for the standard errors have to be adjusted.

IV

$$y_t = x_t'\beta + u_t,\ \ E(x_t u_t) \neq 0$$

Let z_t be an $r \times 1$ vector of instruments, such that $E(z_t u_t) = 0$
The orthogonality conditions are

$$E[z_t(y_t - x_t'\beta_0)] = 0$$

This is a special case of the GMM framework, with

$$w_t = (y_t, x_t', z_t')',\ \ \theta = \beta$$

If $k = r$, the GMM estimator satisfies

$$\begin{aligned} 0 &= g(\hat{\theta}, \eta_n) = \frac{1}{n}\sum_{t=1}^{n} z_t(y_t - x_t'\hat{\beta}) \\ &\Longrightarrow \hat{\beta} = (\sum_{t=1}^{n} z_t x_t')^{-1}(\sum_{t=1}^{n} z_t y_t) \end{aligned}$$

which is the usual IV estimator.

MLE

Let y_t denote a random vector, and $Y_t = (y_1', y_2', ..., y_{t-1}')'$. Assuming that y_t are *i.i.d.*, the log-likelihood function is

$$\mathcal{L}(\theta) = \sum_{t=1}^{n} \log f(y_t; \theta)$$

Since $f(y_t; \theta)$ is a density,

$$\int_A f(y_t; \theta) dy_t = 1$$

where A is the set of possible values of y_t, and $\int dy_t$ denote multiple integration w.r.t. $dy_{1t}, dy_{2t}, ..., dy_{\nu_t}$.

By differentiating both sides w.r.t. θ, we have

$$\int_A \frac{\partial f(y_t; \theta)}{\partial \theta} dy_t = 0$$

Assuming that regularity conditions are satisfied, so that the order of differentiation and integration can be reversed.

$$\begin{aligned} \int_A \frac{\partial f(y_t; \theta)}{\partial \theta} \frac{1}{f(y_t; \theta)} f(y_t; \theta) dy_t &= 0 \Longrightarrow \\ \int_A \frac{\partial \log f(y_t; \theta)}{\partial \theta} f(y_t; \theta) dy_t &= 0 \end{aligned}$$

Let the score of the t^{th} observation be:

$$h(\theta, Y_t) = \frac{\partial \log f(y_t; \theta)}{\partial \theta}$$

$$\int_A h(\theta, Y_t) f(y_t; \theta) dy_t = 0 \Longleftrightarrow E[h(\theta, Y_t)] = 0$$

The GMM principle suggests using an estimate $\hat{\theta}$ that solves:

$$\frac{1}{n}\sum_{t=1}^{n} h(\hat{\theta}, Y_t) = 0$$

The FOC for maximisation of $\mathcal{L}(\theta)$ are:

$$\sum_{t=1}^{n} \frac{\partial \log f(y_t;\theta)}{\partial\theta} = \sum_{t=1}^{n} h(\theta, Y_t) = 0$$

$\therefore$the MLE is the same as the GMM based on the moment conditions $E[h(\theta, Y_t)] = 0$

Estimation of Dynamic RE models

The tests of rational expectations models are based on the fact that undetr this hypothesis the forecasting errors are uncorrelated with the information that the agents have available at the time of the forecast. As long as the econometrician observes a subset of the information people have actually used, RE suggest orthogonality conditions that can be used in the GMM framework.

Example Hansen-Singleton (1982)

The authors consider the problem of a representative agent (stock holder) that maximizes consumption by choosing a portfolio of assets. They assume a CIS utility function and a discount factor. They want both to estimate the unknown paramenters and test whether the theorey is concistent with the data.

Consider the following preferences: $u(c_t)$, with $\frac{\partial u(c_t)}{\partial c_t} > 0$, $\frac{\partial^2 u(c_t)}{\partial c_t^2} < 0$, where $c_t =$ consumption.

The stockholder maximizes

$$\sum_{\tau=0}^{\infty} \beta^{\tau} E\left[u\left(c_{t+\tau}\right) \middle| I_t^* \right],$$

where I_t^* is the information set at time t and $0 < \beta < 1$.

The stockholder purchases m different assets, where a dollar invested in asset i at date t will yield a gross return at date $(t+1)$ equal to $\left(1 + r_{t+1}^i\right)$. This is not known with certainty at date t.

If the stockholder takes a position in each of the m assets, the optimal portfolio is

$$u'(c_t) = \beta E\left[\left(1 + r_{t+1}^i\right) u'(c_{t+1}) \middle| I_t^* \right], \quad i = 1, 2, ..., m \tag{1}$$

Suppose:

$$u(c_t) = \frac{c_t^{1-\gamma}}{1-\gamma}, \quad \gamma > 0, \gamma \neq 1$$

Then equation(1) implies

$$\begin{aligned} c_t^{-\gamma} &= \beta E\left[\left(1 + r_{t+1}^i\right) c_{t+1}^{-\gamma} \middle| I_t^* \right] \\ 1 &= \beta E\left[\left(1 + r_{t+1}^i\right) \left(\frac{c_{t+1}}{c_t}\right)^{-\gamma} \middle| I_t^* \right] \end{aligned}$$

Then

$$E\left\{ \left[\left(1 - \beta\left(1 + r_{t+1}^i\right)\left(\frac{c_{t+1}}{c_t}\right)^{-\gamma}\right) \middle| I_t \right] \right\} = 0, \text{ where } I_t \text{ is a subset of } I_t^*$$

If we define

$$\begin{aligned} \theta &= (\beta, \gamma)' \\ w_t &= \left(r_{t+1}^1, r_{t+1}^2, ..., r_{t+1}^m, \frac{c_{t+1}}{c_t}, I_t' \right)' \end{aligned}$$

Then

$$h\Bigg(\underbrace{\theta}_{2\times 1}, w_t\Bigg)_{mx1} = \begin{bmatrix} \left[\left(1-\beta\left(1+r_{t+1}^{1}\right)\left(\frac{c_{t+1}}{c_t}\right)^{-\gamma}\right)\Big| I_t\right] \\ \vdots \\ \left[\left(1-\beta\left(1+r_{t+1}^{m}\right)\left(\frac{c_{t+1}}{c_t}\right)^{-\gamma}\right)\Big| I_t\right] \end{bmatrix}$$

(The RE hypothesis implies that $h(\cdot)$ should be uncorrelated with past values)

Hence

$$g(\theta, y_t) = \frac{1}{T}\sum_{t=1}^{T} h(\theta, w_t)$$

Therefore the function to minimize with respect to θ is

$$Q = [g(\theta, y_t)]' \widehat{S}_T^{-1} [g(\theta, y_t)]$$

where

$$\widehat{S}_T = \frac{1}{T}\sum_{t=1}^{T}\left\{[h(\theta, w_t)][h(\theta, w_t)]'\right\}$$

They used as instruments, l lags of $\frac{c_t}{c_{t+1}}$......, r_t^1....., r_t^2......r_t^m (In the paper there are only 2 assets which are returns adjusted by inflation and by exchange rates) + constants.

We have $m(m+1)l$ separete over identifying restrictions .

$$\begin{aligned} r &= m((m+1)l+1) \qquad \text{orthogonality conditions} \\ \text{and} \quad k &= 2 \end{aligned}$$

I_t are typically lagged values.

$$Test \sim \chi^2_{(m(m+1)l)}$$