

4. Generalized Method of Moments (GMM)

4.1 Background

Observations $y_i, \mathbf{x}_i, i = 1, \dots, n$.

$$(1) \quad y_i = g(\mathbf{x}_i; \boldsymbol{\theta}) + u_i$$

where g is a function ("model between y and x variables"), $\boldsymbol{\theta}$ contains the model parameters, and u_i is a random error term.

Least squares:

$$(2) \quad \hat{\boldsymbol{\theta}}_{LS} = \arg \min_{\boldsymbol{\theta}} \sum_{i=1}^n (y_i - g(\mathbf{x}_i; \boldsymbol{\theta}))^2.$$

Maximum Likelihood:

$$(3) \quad \hat{\theta}_{ML} = \arg \max_{\theta} L(\theta; \mathbf{y}, \mathbf{x}),$$

where $L(\theta; \mathbf{y}, \mathbf{x})$ is the *likelihood function* of θ .

4.2. Classical Method of Moments

Let θ be a m -vector of parameters that characterize the distribution random variable y .

The k^{th} moment (provided it exists) is defined as

$$(4) \quad \mu_k(\theta) = E[y^k].$$

Suppose we have sample of size T from y with observations y_1, y_2, \dots, y_T (considered as T independent random variables).

The corresponding sample moments are

$$(5) \quad \hat{\mu}_k = \frac{1}{T} \sum_{t=1}^T y_t^k.$$

With the method of moments one estimates the components of θ by simply equating the first m population moments $\mu_k(\theta)$ with the corresponding sample moments $\hat{\mu}_k$, and solving for the components in the parameter vector θ .

Example 4.1. Normal distribution:

y_1, \dots, y_T sample from $N(\mu, \sigma^2)$.

$$\theta = (\mu, \sigma^2),$$

$$(6) \quad \mu_1(\theta) = E[y] = \mu.$$

From $\sigma^2 = \text{Var}[y] = E[(y - \mu)^2] = E[y^2] - \mu^2$,

$$(7) \quad \mu_2(\theta) = E[y^2] = \sigma^2 + \mu^2.$$

Equate these with the sample moments,

$$(8) \quad \mu_1(\hat{\theta}) = \frac{1}{T} \sum_{t=1}^T y_t$$

and

$$(9) \quad \mu_2(\hat{\theta}) = \frac{1}{T} \sum_{t=1}^T y_t^2.$$

That is

$$(10) \quad \hat{\mu} = \frac{1}{T} \sum_{t=1}^T y_t = \bar{y}$$

and

$$(11) \quad \hat{\sigma}^2 + \hat{\mu}^2 = \frac{1}{T} \sum_{t=1}^T y_t^2.$$

Arranging terms, we get for $\hat{\sigma}^2$

$$(12) \quad \hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^T y_t^2 - \hat{\mu}^2 = \frac{1}{T} \sum_{t=1}^T y_t^2 - \bar{y}^2.$$

Recalling, $\sum_{t=1}^T (y_t - \bar{y})^2 = \sum_{t=1}^T y_t^2 - T\bar{y}$, we get finally

$$(13) \quad \hat{\mu} = \bar{y}$$

and

$$(14) \quad \hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^T (y_t - \bar{y})^2$$

as the MM estimators of μ and σ^2 .

Example 4.2. The t -distribution.

Suppose y is a random variable following a t -distribution with ν degrees of freedom.

The density is then

$$(15) \quad f(y_t; \nu) = \frac{\Gamma((\nu + 1)/2)}{\sqrt{\pi\nu}\Gamma(\nu/2)} [1 + (y_t/\nu)^2]^{-\frac{1}{2}(\nu+1)},$$

where

$$(16) \quad \Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt$$

is the gamma function, with the property that, if n is an integer (≥ 0) then $\Gamma(n+1) = n! = n \cdot (n-1) \cdots 2 \cdot 1$, note that by definition $0! = 1$.

From sample y_1, \dots, y_T , the MM estimator is obtained as follows.

Provided that $\nu > 2$ the expected value of a t -distributed random variable is zero ($E[Y_t] = 0$) and variance

$$(17) \quad \mu_2 = E \left[(Y_t - E[Y_t])^2 \right] = E [Y_t^2] = \frac{\nu}{\nu - 2},$$

which at the same time is the second (population) moment of the distribution.

Equate this again with the second sample moment and solve for ν to obtain

$$(18) \quad \hat{\nu} = \frac{2\hat{\mu}_2}{\hat{\mu}_2 - 1}$$

provided that $\hat{\mu}_2 > 1$, where $\hat{\mu}_2 = (1/T) \sum_{t=1}^T y_t^2$.

Otherwise the estimate does not exist.

The moment estimators are functions of averages of random variables.

The Law of Large Numbers implies that these converge toward the corresponding theoretical moments.

Thus, the moment estimators are consistent.

Furthermore, (under general conditions) the CLT implies that the asymptotic distributions of MM estimators are normal.

4.3. 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 4.3 In classical linear regression

$$(19) \quad y_t = \mathbf{x}'_t \boldsymbol{\beta} + e_t,$$

where $\mathbf{x}_t = (x_{1t}, \dots, x_{mt})'$ 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:

$$(20) \quad \begin{array}{l} (i) \quad E[e_t] = 0 \text{ a constant for all } t \\ (ii) \quad E[(y_t - \mathbf{x}'_t \boldsymbol{\beta}) \mathbf{x}_t] = E[e_t \mathbf{x}_t] = 0 \text{ for all } t \\ (iii) \quad E[e_t e_u] = 0 \text{ for all } t \neq u, \end{array}$$

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 4.4. (Ex 4.3 continued). Given T observations the sample moment corresponding the theoretical moment $E[y_t - \mathbf{x}'_t\boldsymbol{\beta}]$ in the moment conditions (ii) is

$$(21) \quad \frac{1}{T} \sum_{t=1}^T (y_t - \mathbf{x}'_t\boldsymbol{\beta})\mathbf{x}_t,$$

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

$$(22) \quad \frac{1}{T} \sum_{t=1}^T (y_t - \mathbf{x}'_t\hat{\boldsymbol{\beta}})\mathbf{x}_t = 0.$$

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

Note that

$$(23) \quad \sum_{t=1}^T (y_t - \mathbf{x}'_t\boldsymbol{\beta})\mathbf{x}_t = \mathbf{X}'\mathbf{y} - (\mathbf{X}'\mathbf{X})\boldsymbol{\beta},$$

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

Example 4.5. (Example 4.2 continued): Suppose $\nu > 4$, then we can calculate

$$(24) \quad \mu_4 = E[y^4] = \frac{3\nu^2}{(\nu-2)(\nu-4)}$$

and the corresponding sample moment is

$$(25) \quad \hat{\mu}_4 = \frac{1}{T} \sum_{t=1}^T y_t^4.$$

Thus the moment conditions implied by the model are

$$(26) \quad E \left[\begin{pmatrix} y^2 - \frac{\nu}{\nu-2} \\ y^4 - \frac{\nu}{(\nu-2)(\nu-4)} \end{pmatrix} \right] = 0.$$

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

$$(27) \quad \frac{1}{T} \sum_{t=1}^T \begin{pmatrix} y_t^2 - \frac{\nu}{\nu-2} \\ y_t^4 - \frac{3\nu^2}{(\nu-2)(\nu-4)} \end{pmatrix} = \begin{pmatrix} \hat{\mu}_2 - \frac{\nu}{\nu-2} \\ \hat{\mu}_4 - \frac{3\nu^2}{(\nu-2)(\nu-4)} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

(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

$$(28) \quad Q(\nu; \mathbf{y}) = \mathbf{g}'_T \mathbf{W} \mathbf{g}_T,$$

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

$$(29) \quad \mathbf{g}_T = \mathbf{g}_T(\nu; \mathbf{y}) = \begin{pmatrix} \hat{\mu}_2 - \frac{\nu}{\nu-2} \\ \hat{\mu}_4 - \frac{3\nu^2}{(\nu-2)(\nu-4)} \end{pmatrix} = \begin{pmatrix} g_{1,T} \\ g_{2,T} \end{pmatrix},$$

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

$$(30) \quad \mathbf{S} = \lim_{T \rightarrow \infty} T \cdot E [(\mathbf{g}_T(\boldsymbol{\theta}_0; \mathbf{y})) (\mathbf{g}_T(\boldsymbol{\theta}_0; \mathbf{y}))'],$$

where $E [(\mathbf{g}_T(\boldsymbol{\theta}_0; \mathbf{y})) (\mathbf{g}_T(\boldsymbol{\theta}_0; \mathbf{y}))']$ 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 $\boldsymbol{\theta}_0$

$$(31) \quad \mathbf{E}[\mathbf{u}_t] = \mathbf{E}[\mathbf{u}(\mathbf{x}_t; \boldsymbol{\theta}_0)] = \mathbf{0}.$$

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) $\mathbf{E}[\mathbf{u}_t] = \boldsymbol{\mu}$ for all t

(2) $\text{Cov}[\mathbf{u}_t, \mathbf{u}_{t+j}] = \mathbf{E}[(\mathbf{u}_t - \boldsymbol{\mu})(\mathbf{u}_{t+j} - \boldsymbol{\mu})'] = \mathbf{S}_j$ for all t . Note: $\text{Cov}[\mathbf{u}_t, \mathbf{u}_{t-j}] = \mathbf{S}_{-j} = \mathbf{S}'_j$.

A sample counterpart of the expected value in (31) is the sample average

$$(32) \quad \mathbf{g}_T(\boldsymbol{\theta}) = \frac{1}{T} \sum_{t=1}^T \mathbf{u}(\mathbf{x}_t; \boldsymbol{\theta}) = \frac{1}{T} \sum_{t=1}^T \mathbf{u}_t.$$

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

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 (32) becomes exactly equal to zero.

In the general case, where there are more moment conditions than parameters the estimator for θ_0 is a "compromise" that makes (32) close to zero.

This is achieved in the GMM by selecting the estimator for θ_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 $\hat{\theta}$ of θ_0 is the value of θ that minimizes

$$(33) \quad Q(\theta) = g_T(\theta)'Wg_T(\theta).$$

where the prime denotes matrix transposition and 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, $\mathbf{E}[\mathbf{u}_t] = \mathbf{0}$ and $\mathbf{E}[\mathbf{g}_T(\boldsymbol{\theta})] = \mathbf{0}$ for the true parameter value)

$$\begin{aligned} \text{Cov}[\mathbf{g}_T(\boldsymbol{\theta})] &= \mathbf{E}[\mathbf{g}_T(\boldsymbol{\theta})\mathbf{g}_T(\boldsymbol{\theta})'] \\ (34) \quad &= \frac{1}{T^2} \mathbf{E} \left[\left(\sum_{t=1}^T \mathbf{u}_t \right) \left(\sum_{t=1}^T \mathbf{u}_t \right)' \right] \\ &= \frac{1}{T^2} \sum_{s=1}^T \sum_{t=1}^T \mathbf{E}[\mathbf{u}_s \mathbf{u}_t'] \end{aligned}$$

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

$$(35) \quad \mathbb{E} [\mathbf{u}_t \mathbf{u}'_{t+j}] = \mathbf{S}_j = \mathbf{S}'_{-j}$$

for all t . Thus in (34) we can write

$$(36) \quad \sum_{s=1}^T \sum_{t=1}^T \mathbb{E} [\mathbf{u}_s \mathbf{u}'_t] = \sum_{j=-T}^T (T - |j|) \mathbf{S}_j,$$

or by using (35)

$$(37) \quad \sum_{s=1}^T \sum_{t=1}^T \mathbb{E} [\mathbf{u}_s \mathbf{u}'_t] = T \mathbf{S}_0 + \sum_{j=1}^T (T - j) (\mathbf{S}_j + \mathbf{S}'_j).$$

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

$$(38) \quad \lim_{k \rightarrow \infty} \sum_{j=-k}^k \mathbf{S}_j = \mathbf{S},$$

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

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

$$(39) \quad \text{Cov} [\sqrt{T} \mathbf{g}_T] \rightarrow \sum_{j=-\infty}^{\infty} \mathbf{S}_j = \mathbf{S},$$

as $T \rightarrow \infty$.

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

$$(40) \quad \hat{\mathbf{S}}_j = \frac{1}{T} \sum_{t=j+1}^T \hat{\mathbf{u}}_t \hat{\mathbf{u}}'_{t-j},$$

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

$$(41) \quad \hat{\mathbf{S}} = \hat{\mathbf{S}}_0 + \sum_{j=1}^{\ell} w_j (\hat{\mathbf{S}}_j + \hat{\mathbf{S}}'_j),$$

where w_j s 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

$$(42) \quad w_j = 1 - j/(\ell + 1).$$

The two step procedure:

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

$$(43) \quad \hat{\boldsymbol{\theta}}^{(1)} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \mathbf{g}_T(\boldsymbol{\theta})' \mathbf{g}_T(\boldsymbol{\theta}).$$

Step 2: Compute

$$(44) \quad \hat{\mathbf{u}}_t = \mathbf{u}(\mathbf{x}_t; \hat{\boldsymbol{\theta}}^{(1)})$$

and estimate \mathbf{S}_j as

$$(45) \quad \hat{\mathbf{S}}_j = \frac{1}{T} \sum_{t=j+1}^T \hat{\mathbf{u}}_t \hat{\mathbf{u}}_{t+j}'$$

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

Estimate \mathbf{S} by

$$(46) \quad \hat{\mathbf{S}} = \hat{\mathbf{S}}_0 + \sum_{j=1}^{\ell} w_j (\hat{\mathbf{S}}_j + \hat{\mathbf{S}}_j').$$

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

$$(47) \quad \hat{\boldsymbol{\theta}}^{(2)} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \mathbf{g}_T(\boldsymbol{\theta})' \mathbf{W} \mathbf{g}_T(\boldsymbol{\theta}).$$

Remark 4.1: An iterative GMM iterates Step 2 until convergence of the $\boldsymbol{\theta}$ -estimate, i.e., when after k^{th} step $\hat{\boldsymbol{\theta}}^{(k+1)} \approx \hat{\boldsymbol{\theta}}^{(k)}$.

Instruments

Suppose our suggested model is

$$(48) \quad y_t = f(\mathbf{x}_t; \theta) + e_t,$$

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

$$(49) \quad E[e_t] = E[y_t - f(\mathbf{x}_t; \theta_0)] = 0.$$

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

$$(50) \quad \mathbb{E} [(y - f(x; \theta)) \otimes \mathbf{z}] = 0,$$

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

$$(51) \quad \begin{aligned} \mathbf{A} \otimes \mathbf{B} &= \begin{pmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} \end{pmatrix} \\ &= \begin{pmatrix} a_{11}b_{11} & a_{11}b_{12} & a_{12}b_{11} & a_{12}b_{12} \\ a_{11}b_{21} & a_{11}b_{22} & a_{12}b_{21} & a_{12}b_{22} \\ a_{21}b_{11} & a_{21}b_{12} & a_{22}b_{11} & a_{22}b_{12} \\ a_{21}b_{21} & a_{21}b_{22} & a_{22}b_{21} & a_{22}b_{22} \end{pmatrix}. \end{aligned}$$

Note that (50) includes condition (??) (defining e.g. the first component of \mathbf{z}_t , $z_{1t} = 1$).

Variables in \mathbf{z}_t are called instrument variables.

The \mathbf{u} -function in equation (28) generalizes now

$$(52) \quad \mathbf{u}_t = \mathbf{u}_t(\boldsymbol{\theta}) = (y_t - f(\mathbf{x}_t; \boldsymbol{\theta})) \otimes \mathbf{z}_t,$$

with

$$(53) \quad \mathbf{E}[\mathbf{u}_t] = \mathbf{0}.$$

Example 4.6 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

$$(54) \mathbf{u}_t = (M_{t+1}(\theta)\mathbf{X}_{t+1} - 1) \otimes \mathbf{z}_t = \begin{pmatrix} M_{t+1}(\theta) X_{t+1}^a - 1 \\ M_{t+1}(\theta) X_{t+1}^a z_t - z_t \\ M_{t+1}(\theta) X_{t+1}^b - 1 \\ M_{t+1}(\theta) X_{t+1}^b z_t - z_t \end{pmatrix}$$

with

$$(55) \quad \mathbb{E}[\mathbf{u}_t] = \mathbb{E} \left[\begin{pmatrix} M_{t+1}(\theta) X_{t+1}^a - 1 \\ M_{t+1}(\theta) X_{t+1}^a z_t - z_t \\ M_{t+1}(\theta) X_{t+1}^b - 1 \\ M_{t+1}(\theta) X_{t+1}^b z_t - z_t \end{pmatrix} \right] = \mathbf{0}.$$

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

$$(56) \quad \mathbf{g}(\hat{\theta}) = \frac{1}{T} \sum_{t=1}^T \mathbf{u}_t = 0$$

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

Example 4.7: OLS as an exactly identified GMM (Example 4.4 revisited).

Assume again the standard regression model

$$(57) \quad \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e},$$

where

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

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

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

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

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

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

The orthogonality conditions are

$$(58) \quad E[\mathbf{u}_t] = E[(y_t - \mathbf{x}'_t \boldsymbol{\beta}) \mathbf{x}_t] = E[e_t \mathbf{x}_t] = 0,$$

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

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

The sample counterpart of (58) is

$$(59) \quad \mathbf{g}_T(\boldsymbol{\beta}) = \frac{1}{T} \sum_{t=1}^T (y_t - \mathbf{x}'_t \boldsymbol{\beta}) \mathbf{x}_t = \frac{1}{T} \mathbf{X}'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}).$$

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

$$(60) \quad \mathbf{X}'\mathbf{y} - \mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{0},$$

and we get finally

$$(61) \quad \hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y},$$

the same as the OLS estimator.

Remark 4.2: Under the assumptions of the above example \mathbf{x}_t and e_t are independent, which implies that $E[\mathbf{x}_t e_t] = E[\mathbf{x}_t]E[e_t] = 0$, and hence

$$(62) \quad E[\mathbf{u}_t \mathbf{u}_t'] = E[e_t^2] E[\mathbf{x}_t \mathbf{x}_t'] = \sigma_e^2 E[\mathbf{x}_t \mathbf{x}_t'].$$

and

$$(63) \quad E[\mathbf{u}_t \mathbf{u}_{t+j}'] = E[\mathbf{x}_t \mathbf{x}_{t+j}'] E[e_t e_{t+j}] = 0$$

because of no autocorrelation in the residuals.

These imply

$$(64) \quad \mathbf{S} = \mathbf{S}_0 = \sigma_e^2 \frac{1}{T} \sum_{t=1}^T E[\mathbf{x}_t \mathbf{x}_t'].$$

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

$$(65) \quad \mathbf{X}'\mathbf{X} = \sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t'$$

minimization of

$$(66) \quad \mathbf{g}_T(\boldsymbol{\beta})' \mathbf{W} \mathbf{g}_T(\boldsymbol{\beta})$$

with

$$(67) \quad \mathbf{W} = \frac{T}{\sigma_e^2} (\mathbf{X}'\mathbf{X})^{-1}$$

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 independent of the $(m \times m)$ p.d. matrix \mathbf{W} , because the first order condition of the minimum of (66) is

$$(68) \quad \partial (\mathbf{g}_T(\boldsymbol{\beta})' \mathbf{W} \mathbf{g}_T(\boldsymbol{\beta})) / \partial \boldsymbol{\beta} = 2\mathbf{D}' \mathbf{W} \mathbf{g}_T(\boldsymbol{\beta}) = 0,$$

where $\mathbf{D} = \partial \mathbf{g}_T(\boldsymbol{\beta}) / \partial \boldsymbol{\beta} = -\mathbf{X}' \mathbf{X} / T$. Multiplying both sides of (68) from the left by the inverse of $\mathbf{D}' \mathbf{W}$ (which is p.d.) leads again to the OLS solution.

Remark 4.3: 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 4.8: 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

$$(69) \quad \mathbf{S}_j = \mathbf{E}[\mathbf{u}_t \mathbf{u}'_{t+j}] = \mathbf{E}[e_t e_{t+j} \mathbf{x}_t \mathbf{x}'_{t+j}].$$

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

$$(70) \quad \hat{\mathbf{S}} = \hat{\mathbf{S}}_0 + \sum_{j=1}^{\ell} w_j (\hat{\mathbf{S}}_j + \hat{\mathbf{S}}'_j),$$

where

$$(71) \quad \hat{\mathbf{S}}_j = \frac{1}{T} \sum_{t=j+1}^T \hat{e}_t \hat{e}_{t-j} \mathbf{x}_t \mathbf{x}'_{t-j},$$

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

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

$$(72) \quad H_0 : E[\mathbf{u}_t(\boldsymbol{\theta})] = 0$$

the over identification conditions can be tested with the statistic

$$(73) \quad J = \mathbf{g}(\hat{\boldsymbol{\theta}})' \mathbf{W} \mathbf{g}(\hat{\boldsymbol{\theta}}_T),$$

where TJ is asymptotically χ_{r-m}^2 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

$$(74) \quad \hat{\theta} \sim AN(\theta, \hat{V}/T),$$

where

$$(75) \quad \hat{V} = (\hat{D}\hat{S}^{-1}\hat{D}')^{-1},$$

where

$$(76) \quad \hat{D} = \left. \frac{\partial \mathbf{g}(\theta)}{\partial \theta} \right|_{\theta=\hat{\theta}}.$$

Furthermore under general regularity conditions, it can be shown that

$$(77) \quad \hat{\theta} \xrightarrow{p} \theta_0,$$

i.e. they are consistent.

4.4. Examples

Example 4.9: Normality of SP500 (daily) returns, y_t . If normality holds, the following (first four) moment conditions hold:

$$(78) \quad \begin{aligned} E[y_t - \mu] &= 0 \\ E[(y_t - \mu)^2 - \sigma^2] &= 0 \\ E[(y_t - \mu)^3 / \sigma^3] &= 0 \\ E[(y_t - \mu)^4 / \sigma^4 - 3] &= 0 \end{aligned}$$

where $\mu = E[y_t]$, $\sigma^2 = \text{Var}[y_t]$. Data is obtained from finance.yahoo.com web site with sample period Jan 2, 1995 to May 19, 2005.

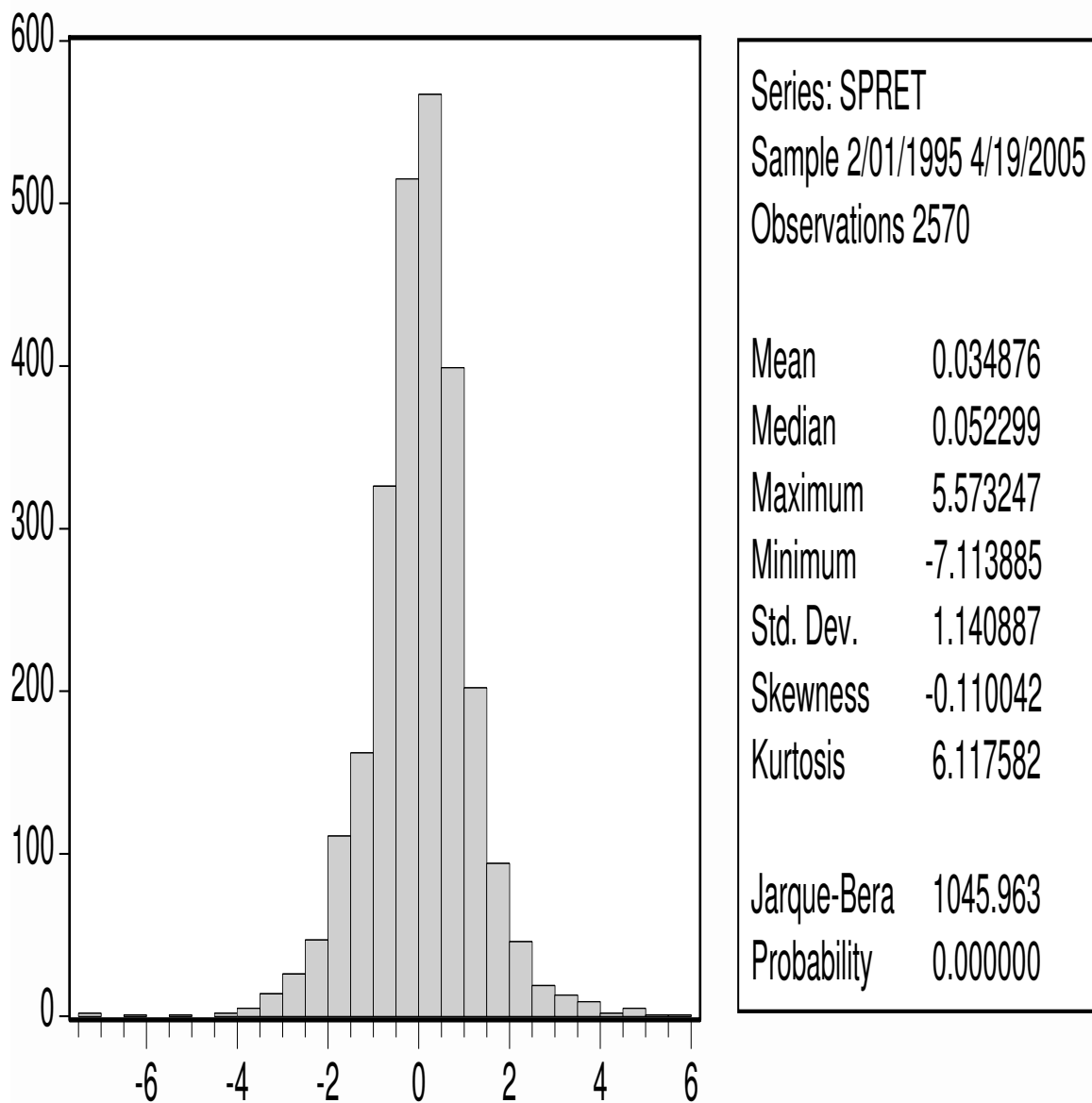
In EViews open Object → New object ... → System and write commands (c(1) = mean, c(2) = standard deviation):

```
@inst c
param c(1) 0 c(2) 1.0
spret - c(1)
(spret - c(1))^2 - c(2)^2
((spret - c(1))/c(2))^3
((spret - c(1))/c(2))^4 - 3
```

where **spret** is the SP500 return.

On the `parameter` line 0 and 1.0 are initial values for the numerical process in the the estimation.

Sample statistics



Select Estimate → GMM to get results

System: GMM_NORMALITY Estimation Method: Generalized Method of Moments Date: 04/20/05 Time: 02:51 Sample: 2/01/1995 4/19/2005 Included observations: 2570 Total system (balanced) observations 10280 Kernel: Bartlett, Bandwidth: Fixed (8), No prewhitening Iterate coefficients after one-step weighting matrix Convergence not achieved after: 1 weight matrix, 506 total coef iterations				
	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	0.020709	0.020425	1.013894	0.3107
C(2)	1.286071	0.018708	68.74604	0.0000
Determinant residual covariance		74161.73		
J-statistic		0.050103		
Equation: SPRET - C(1)				
Instruments: C				
Observations: 2570				
S.E. of regression	1.140975	Sum squared resid	3344.388	
Durbin-Watson stat	2.022964			
Equation: (SPRET - C(1))^2 - C(2)^2				
Instruments: C				
Observations: 2570				
S.E. of regression	2.964218	Sum squared resid	22563.95	
Durbin-Watson stat	1.594675			
Equation: ((SPRET - C(1))/C(2))^3				
Instruments: C				
Observations: 2570				
S.E. of regression	7.010289	Sum squared resid	126202.2	
Durbin-Watson stat	2.244303			
Equation: ((SPRET - C(1))/C(2))^4 - 3				
Instruments: C				
Observations: 2570				
S.E. of regression	31.77565	Sum squared resid	2592889.	
Durbin-Watson stat	1.756577			

Note that the J -statistic in EViews is not multiplied by number of observations. Thus

$$J = 2570 \times 0.050103 \approx 128.8,$$

which is highly statistically significant ($df = 4 - 2 = 2$), and thus rejects the normality hypothesis.

Let us next test whether a t -distribution with location parameter μ , scale parameter σ^2 , and degrees of freedom parameter ν fits better. The density function is

$$(79) \quad f(y) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu\sigma^2} \Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{(y-\mu)^2}{\sigma^2(\nu-2)}\right)^{-\frac{1}{2}(\nu+1)}$$

with

$$E[y] = \mu,$$

$$\text{Var}[y] = \frac{\nu\sigma^2}{\nu-2},$$

and

$$E[(y-\mu)^4] = \frac{3\nu^2\sigma^4}{(\nu-2)(\nu-4)}.$$

The implied moment conditions are

$$E[y - \mu] = 0$$

$$E\left[(y - \mu)^2 - \sigma^2 \frac{\nu}{\nu-2}\right] = 0$$

$$E[(y - \mu)^3] = 0$$

$$E\left[(y - \mu)^4 - \frac{3\nu^2\sigma^4}{(\nu-2)(\nu-4)}\right] = 0$$

EViews estimation produces with commands (c(1) = mean, c(2) = scale, c(3) = df)

```
@inst c
param c(1) 0 c(2) 1.0 c(3) 7.0
sprete - c(1)
(sprete - c(1))^2 - c(2)^2*c(3)/(c(3)-2)
(sprete - c(1))^3
(sprete - c(1))^4 - (3*c(2)^4)*(c(3)^2)/((c(3)-2)*(c(3)-4))
```

System: GMM_T				
Estimation Method: Generalized Method of Moments				
Date: 04/25/05 Time: 00:29				
Sample: 2/01/1995 4/19/2005				
Included observations: 2570				
Total system (balanced) observations 10280				
Kernel: Bartlett, Bandwidth: Fixed (8), No prewhitening				
Iterate coefficients after one-step weighting matrix				
Convergence achieved after: 1 weight matrix, 8 total coef iterations				
	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	0.040571	0.019412	2.089975	0.0366
C(2)	0.933509	0.029089	32.09109	0.0000
C(3)	6.132127	0.436527	14.04754	0.0000
Determinant residual covariance		2346223.		
J-statistic		0.000233		
Equation: SPRET - C(1)				
Instruments: C				
Observations: 2570				
S.E. of regression	1.140902	Sum squared resid	3343.955	
Durbin-Watson stat	2.023225			
Equation: (SPRET - C(1))^2 - C(2)^2*C(3)/(C(3)-2)				
Instruments: C				
Observations: 2570				
S.E. of regression	2.945789	Sum squared resid	22275.58	
Durbin-Watson stat	1.616382			
Equation: (SPRET - C(1))^3				
Instruments: C				
Observations: 2570				
S.E. of regression	14.94928	Sum squared resid	574122.4	
Durbin-Watson stat	2.241108			
Equation: (SPRET - C(1))^4 - (3*C(2)^4)*(C(3)^2)/((C(3)-2)*(C(3)-4))				
Instruments: C				
Observations: 2570				
S.E. of regression	87.50198	Sum squared resid	19654482	
Durbin-Watson stat	1.761688			

The J -statistic is $J = T \times J_{EViews} = 2570 \times 0.000233 = 0.598$ with p -value 0.439 ($df = 1$), which indicates

the return distribution seems to behave like a the t -distribution at least up to the first four moments with estimates $\hat{\mu} = 0.041$, $\hat{\sigma} = 0.934$, and $\hat{\nu} = 6.132$.

Example 4.4 Consumption based asset pricing model:
(Ferson, W. and Harvey, C. (1992), Seasonality and
Consumption-Based Asset Pricing, *Journal of Finance*,
47, 511-552.)*

Source: <http://support.sas.com/rnd/app/examples/ets/harvey/index.htm>

Asset pricing is one of the fundamental questions in financial economics.

One approach is that individuals hold assets to optimize intertemporal consumption.

Assume that the current price of a security is p_t and the future payoff (price plus dividends) is p_{t+1} . Let furthermore q denote the amount of the investor chooses to buy of the asset and w_t denote his/her potential to consume (wage plus other available wealth).

*See also, Hansen and Singleton (1982). Generalized instrumental variables method of nonlinear rational expectations models. *Econometrica* 50, 1269–1286. Errata: *Econometrica* 52, 267–268.

Then the problem of the investor is to balance today's consumption and saving in order to maximize his/her utility.

Thus, the problem then is to maximize (time-additive) expected discounted utility

$$(80) \quad U(c_t, c_{t+1}) = u(c_t) + E_t [\beta u(c_{t+1})],$$

subject to budget constraints

$$(81) \quad \begin{aligned} c_t &= w_t - qp_t \\ c_{t+1} &= w_{t+1} + qp_{t+1}, \end{aligned}$$

where β is the *subjective discount factor*.

The control variable is q , the amount of shares.

Substituting c_t and c_{t+1} in (80) by the right hand sides of (81), the first order condition for the maximum is obtained by setting the derivative of (80) w.r.t. q to zero, i.e.,

$$(82) \quad \frac{du}{dq} = -p_t u'(c_t) + E_t [\beta u'(c_{t+1}) p_{t+1}] = 0,$$

from which

$$(83) \quad p_t = E_t \left[\beta \frac{u'(c_{t+1})}{u'(c_t)} p_{t+1} \right].$$

Using a convenient power utility form,

$$(84) \quad u(c_t) = \frac{c_t^{1-\alpha} - 1}{1-\alpha}.$$

The limit $\alpha \rightarrow 1$ is

$$(85) \quad u(c_t) = \log(c_t).$$

$$(86) \quad u'(c_t) = c_t^{-\alpha}.$$

Thus,

$$(87) \quad p_t = \mathbb{E}_t \left[\beta \left(\frac{c_{t+1}}{c_t} \right)^{-\alpha} p_{t+1} \right],$$

or

$$(88) \quad 1 = \mathbb{E}_t \left[\beta \left(\frac{c_{t+1}}{c_t} \right)^{-\alpha} (1 + R_{t+1}) \right],$$

where

$$(89) \quad R_{t+1} = \frac{P_{t+1} - P_t}{P_t}$$

is the one period return of the asset.

SAS estimation results:

<http://support.sas.com/rnd/app/examples/ets/harvey/index.htm>