1

Generalized Method of Moments (GMM)

Hüseyin Taştan[1]

[1]Department of Economics
Ph.D. Program - Advanced Econometrics
Yildiz Technical University

December 10, 2013

2

Method of Moments Estimation (MM)

MM estimators are found as a solution to the following equation system:

$$
\begin{aligned}
E(X) &= \frac{1}{n}\sum_{i=1}^{n} X_i \\
E(X^2) &= \frac{1}{n}\sum_{i=1}^{n} X_i^2 \\
\vdots &= \vdots \\
E(X^k) &= \frac{1}{n}\sum_{i=1}^{n} X_i^k
\end{aligned}
$$

Population moment conditions are equated to their sample counterparts. Note that population moment conditions are functions of unknown population parameters.

3

MM

Let $f_j(\boldsymbol{\theta}) = E(X^k)$, $j = 1, 2, \ldots, k$, MM solves the following root-finding problem:

$$
\begin{aligned}
E(X) - \frac{1}{n}\sum_{i=1}^{n} X_i &= 0 \\
E(X^2) - \frac{1}{n}\sum_{i=1}^{n} X_i^2 &= 0 \\
\vdots &= \vdots \\
E(X^k) - \frac{1}{n}\sum_{i=1}^{n} X_i^k &= 0
\end{aligned}
$$

4

Population and Sample Moments

k	Population Moments	Sample Moments
1	$\mu_1 = E(X)$	$\hat{\mu}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i$
2	$\mu_2 = E(X^2)$	$\hat{\mu}_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$
3	$\mu_3 = E(X^3)$	$\hat{\mu}_3 = \frac{1}{n}\sum_{i=1}^{n} X_i^3$
4	$\mu_4 = E(X^4)$	$\hat{\mu}_4 = \frac{1}{n}\sum_{i=1}^{n} X_i^4$
$\vdots$	$\vdots$	$\vdots$
k	$\mu_k = E(X^k)$	$\hat{\mu}_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$

MM: Example

Estimation of the degrees of freedom parameter, ν, of the Student t distribution.
Let $\{y_1, y_2, \ldots, y_T\}$ be a random sample from the Student t distribution with ν degrees of freedom. The Student t has the following density:

$$f(y_t;\nu) = \frac{\Gamma\left(\frac{(\nu+1)}{2}\right)}{(\pi\nu)^{1/2}\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{y_t^2}{\nu}\right)^{-\frac{(\nu+1)}{2}}$$

where $\Gamma(\cdot)$ is the gamma function.

MM: Example

We need at least one population moment condition as there is only one unknown parameter. If $Y \sim t(\nu)$ then for $\nu > 2$ the population mean and variance are:

$$E(Y_t) = 0, \quad E(Y_t^2) = \frac{\nu}{\nu-2}$$

Second population moment condition can be used in the estimation of ν. Equating population moment to the sample moment we obtain

$$\frac{\nu}{\nu-2} = \frac{1}{T}\sum_{t=1}^{T} Y_t^2 = \hat{\mu}_2$$

Solving for ν, the MM estimator is

$$\hat{\nu} = \frac{2\hat{\mu}_2}{\hat{\mu}_2 - 1}$$

for $\hat{\mu}_2 > 1$.

OLS and IV as MM Estimation

$\boldsymbol{x}_i = [x_{1i}, x_{2i}, \ldots, x_{ki}]^\top$: $k \times 1$ vector of explanatory variables on the ith observation
$\boldsymbol{\beta}$: $k \times 1$ vector of unknown parameters
Population regression model:

$$y_i = \boldsymbol{x}_i^\top \boldsymbol{\beta} + u_i, \quad i = 1, 2, \ldots, n$$

Let $\boldsymbol{z}_i$ be $k \times 1$ vector of observed (exogenous) instrumental variables which are uncorrelated with the ith error term. Then the population moment condition is:

$$\mathsf{E}[\boldsymbol{z}_i u_i] = \boldsymbol{0}, \quad i = 1, 2, \ldots, n$$

OLS and IV as MM Estimation

Sample Moment Conditions:

$$\begin{aligned} n^{-1}\sum_{i=1}^{n} \boldsymbol{z}_i \hat{u}_i &= n^{-1}\sum_{i=1}^{n} \boldsymbol{z}_i (y_i - \boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}}) \\ &= 0 \end{aligned}$$

Solving the system for $\hat{\boldsymbol{\beta}}$ we get

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= \left(\sum_{i=1}^{n} \boldsymbol{z}_i \boldsymbol{x}_i^\top\right)^{-1}\left(\sum_{i=1}^{n} \boldsymbol{z}_i y_i\right) \\ &= (\boldsymbol{Z}^\top \boldsymbol{X})^{-1} \boldsymbol{Z}^\top \boldsymbol{y} \end{aligned}$$

Note that this is the simple IV estimator we obtained before. This is only valid if the number of endogenous variables is exactly the same as the number of instruments (exact identification). If $\boldsymbol{Z} = \boldsymbol{X}$ we obtain the OLS estimator.

Generalized Method of Moments (GMM)

- What happens if the number of population moment conditions is greater than the number of unknown parameters? We have more information than we need. In this case we cannot use MM.
- GMM was first introduced by Lars Hansen (1982, Econometrica) (Nobel Laureate, 2013) and has been widely applied since then.
- Why did Hansen's paper have such a huge impact?
- Maximum Likelihood estimation is the best framework within the classical paradigm. Do we need GMM after all? What are the advantages of the GMM estimation?

Generalized Method of Moments (GMM)

Motivations for using GMM: (Hall, 2005,p.2)

1. *Sensitivity of the statistical properties to the distributional assumptions*: The desirable statistical properties of MLE (consistency, efficiency) are only attained if the distribution is correctly specified. Unfortunately, economic theory rarely provides the complete specification of the probability distribution of the data. Arbitrary choice of distributions may lead to biased inferences.
2. *Computational burden*: ML estimation can be computationally very burdensome. Implied likelihood function can be extremely difficult to evaluate numerically with available computer technology.

GMM

- GMM estimation uses a set of population moment conditions which are deduced from the economic (econometric) model.
- The exact conditions varies from application to application. The statistical properties of GMM estimator depends on the validity of these moment conditions.
- MM vs GMM: in the MM estimation the number of moment conditions is exactly the same as the number of unknown parameters. If the number of moment conditions is greater than the number of parameters we cannot use MM. In contrast, GMM allows the number of moment conditions to be greater than the number of parameters.
- Many classical estimation methods, such as OLS, ML, IV, GLS can be obtained as special cases of GMM.
- In that sense, GMM provides a unifying estimation framework.

GMM Population and Sample Moment Conditions

Let $p \times 1$ unknown (true) parameter vector be $\boldsymbol{\theta}_0$. Also let $\boldsymbol{y}_i$ be the vector of random variables and $f(\boldsymbol{y}_i, \boldsymbol{\theta}_0)$ be $q \times 1$ vector. For each i, the *population moment condition* can be written as

$$\mathsf{E}[f(\boldsymbol{y}_i, \boldsymbol{\theta}_0)] = \mathbf{0} \tag{1}$$

Sample analogs can be written as

$$\frac{1}{n}\sum_{i=1}^{n} f(\boldsymbol{y}_i, \boldsymbol{\theta}) = \mathbf{0} \tag{2}$$

Population and Sample Moment Conditions: Example

Let $\{y_i\}_{i=1}^n$ be a random sample from a normal distribution with mean μ and variance σ^2. Unknown population parameters are collected in 2×1 vector $\boldsymbol{\theta}_0 = [\mu_0\ \sigma_0^2]^\top$. The population moment conditions, $f(\cdot)$, and their sample analogs can be written as

$$f(y_i, \boldsymbol{\theta}_0) = \begin{bmatrix} y_i - \mu_0 \\ y_i^2 - (\sigma_0^2 + \mu_0^2) \end{bmatrix}$$

$$E[f(y_i, \boldsymbol{\theta}_0)] = \begin{bmatrix} E[y_i] - \mu_0 \\ E[y_i^2] - (\sigma_0^2 + \mu_0^2) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Sample moment conditions:

$$\frac{1}{n}\sum_{i=1}^n f(y_i, \boldsymbol{\theta}_0) = \begin{bmatrix} \frac{1}{n}\sum_{i=1}^n y_i - \mu_0 \\ \frac{1}{n}\sum_{i=1}^n y_i^2 - (\sigma_0^2 + \mu_0^2) \end{bmatrix}$$

□

GMM Estimator

The GMM estimator is defined as the solution to the following quadratic distance measure:

$$Q_n(\boldsymbol{\theta}) = \left(\frac{1}{n}\sum_{i=1}^n f(\boldsymbol{y}_i, \boldsymbol{\theta}_0)\right)^\top \boldsymbol{W}_n \left(\frac{1}{n}\sum_{i=1}^n f(\boldsymbol{y}_i, \boldsymbol{\theta}_0)\right) \tag{3}$$

where $\boldsymbol{W}_n$ is a positive definite $q \times q$ weighting matrix (which, in the limit, converges to a matrix of constants):

$$n \to \infty, \boldsymbol{W}_n \to \boldsymbol{W}^*, \text{positive definite}$$

$q \times 1$ vector of sample moment conditions:

$$m(\boldsymbol{y}_i, \boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^n f(\boldsymbol{y}_i, \boldsymbol{\theta})$$

GMM Estimator

GMM estimator can be written as:

$$\hat{\boldsymbol{\theta}}_n = \arg\min_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \boldsymbol{m}(\boldsymbol{y}_i, \boldsymbol{\theta})^\top \boldsymbol{W}_n \boldsymbol{m}(\boldsymbol{y}_i, \boldsymbol{\theta}) \tag{4}$$

For $Q_n(\boldsymbol{\theta})$ to be a meaningful metric, the weighting matrix must be positive definite and converge to a matrix of constants in the limit (Hall, 2005). This ensures that $Q_n(\boldsymbol{\theta}) \geq 0$ for any $\boldsymbol{\theta}$. At the true parameter vector $Q_0(\boldsymbol{\theta}_0) = 0$. Under global identification and positive definiteness we always have $Q_n(\boldsymbol{\theta}) \geq 0$ for $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$.

Example: Generalized IV Estimator as a Special Case of GMM

$\boldsymbol{x}_i = [x_{1i}, x_{2i}, \ldots, x_{pi}]^\top$: ith observation vector, $p \times 1$
$\boldsymbol{\beta}$ unknown parameter vector, $p \times 1$.
Linear regression model:

$$y_i = \boldsymbol{x}_i^\top \beta + u_i, \quad i = 1, 2, \ldots, n$$

We have a $q \times 1$, $q > p$, vector of instruments, $\boldsymbol{z}_i$. Population moment conditions for the ith observation:

$$\mathsf{E}[\boldsymbol{z}_i u_i] = \mathsf{E}[\boldsymbol{z}_i(y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta})] = \boldsymbol{0}$$

Sample moment conditions:

$$\begin{aligned} m(\boldsymbol{\beta}) = n^{-1}\sum_{i=1}^n \boldsymbol{z}_i \hat{u}_i &= n^{-1}\sum_{i=1}^n \boldsymbol{z}_i(y_i - \boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}}) \\ &= n^{-1}(\boldsymbol{Z}^\top \boldsymbol{y} - \boldsymbol{Z}^\top \boldsymbol{X}\boldsymbol{\beta}) \\ &= 0 \end{aligned}$$

Example: Generalized IV Estimator as a Special Case of GMM

GMM objective function:

$$Q(\boldsymbol{\beta}) = n^{-1}(\boldsymbol{Z}^\top \boldsymbol{y} - \boldsymbol{Z}^\top \boldsymbol{X}\beta)^\top \boldsymbol{W} n^{-1}(\boldsymbol{Z}^\top \boldsymbol{y} - \boldsymbol{Z}^\top \boldsymbol{X}\beta)$$

If we choose the weighting matrix as

$$\begin{aligned} \boldsymbol{W} &= \left(n^{-1}\sum_{i=1}^{n} \boldsymbol{z}_i \boldsymbol{z}_i^\top\right)^{-1} \\ &= n(\boldsymbol{Z}^\top \boldsymbol{Z})^{-1} \end{aligned}$$

the GMM objective function becomes

$$\begin{aligned} Q(\boldsymbol{\beta}) &= n^{-1}(\boldsymbol{Z}^\top \boldsymbol{y} - \boldsymbol{Z}^\top \boldsymbol{X}\boldsymbol{\beta})^\top n(\boldsymbol{Z}^\top \boldsymbol{Z})^{-1} n^{-1}(\boldsymbol{Z}^\top \boldsymbol{y} - \boldsymbol{Z}^\top \boldsymbol{X}\boldsymbol{\beta}) \\ &= n^{-1}(\boldsymbol{Z}^\top \boldsymbol{y} - \boldsymbol{Z}^\top \boldsymbol{X}\boldsymbol{\beta})^\top (\boldsymbol{Z}^\top \boldsymbol{Z})^{-1}(\boldsymbol{Z}^\top \boldsymbol{y} - \boldsymbol{Z}^\top \boldsymbol{X}\boldsymbol{\beta}) \end{aligned}$$

Example: Generalized IV Estimator as a Special Case of GMM

First order conditions:

$$\frac{\partial Q(\beta)}{\partial \beta}|_{\beta=\hat{\beta}} = n^{-1} 2\boldsymbol{X}^\top \boldsymbol{Z}(\boldsymbol{Z}^\top \boldsymbol{Z})^{-1}(\boldsymbol{Z}^\top \boldsymbol{y} - \boldsymbol{Z}^\top \boldsymbol{X}\hat{\boldsymbol{\beta}}) = 0$$

GMM estimator:

$$\hat{\boldsymbol{\beta}} = \left(\boldsymbol{X}^\top \boldsymbol{Z}(\boldsymbol{Z}^\top \boldsymbol{Z})^{-1}\boldsymbol{Z}^\top \boldsymbol{X}\right)^{-1} \boldsymbol{X}^\top \boldsymbol{Z}(\boldsymbol{Z}^\top \boldsymbol{Z})^{-1}\boldsymbol{Z}^\top \boldsymbol{y}$$

This is the usual (generalized) IV estimator which is valid when $q > p$.

Asymptotic Properties of GMM Estimator

- Under certain regularity conditions (see Hall 2005, pp 66-68) GMM estimator, $\hat{\boldsymbol{\theta}}_T$, is consistent:
$$\operatorname*{plim}_{T\to\infty} \hat{\theta}_T = \boldsymbol{\theta}_0,$$
- and asymptotically normal:
$$\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) \to N(\boldsymbol{0}, \boldsymbol{\Omega}),$$
where
$$\begin{gathered} \boldsymbol{\Omega} = \boldsymbol{M}\boldsymbol{S}\boldsymbol{M}^\top, \quad p \times p, \\ \boldsymbol{M} = (\boldsymbol{D}^\top \boldsymbol{W}\boldsymbol{D})^{-1}\boldsymbol{D}^\top \boldsymbol{W}, \quad p \times q \\ \boldsymbol{D} = \mathsf{E}\left[\frac{\partial f(\boldsymbol{y}_t, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\right], \quad q \times p \end{gathered}$$
and long run covariance matrix of moment conditions:
$$\boldsymbol{S} = \lim_{T\to\infty} \mathsf{Var}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} f(\boldsymbol{y}_t, \boldsymbol{\theta})\right], \quad q \times q$$

Asymptotic Properties of GMM Estimator

- $\boldsymbol{W}$ is $q \times q$ weighting matrix. We can obtain a simpler covariance matrix using optimal weighting matrix. Under certain regularity conditions optimal weighting matrix can be found using
$$\boldsymbol{W} = \boldsymbol{S}^{-1} = \lim_{T\to\infty} \mathsf{Var}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} f(\boldsymbol{y}_t, \boldsymbol{\theta})\right]^{-1} \tag{5}$$
(see Hall, 2005, Theorem 3.4, pp.88-89)
- Substituting optimal weighting matrix into the covariance matrix expression we obtain
$$\boldsymbol{\Omega} = (\boldsymbol{D}^\top \boldsymbol{S}^{-1}\boldsymbol{D})^{-1}\boldsymbol{D}^\top \boldsymbol{S}^{-1}\boldsymbol{S}\boldsymbol{S}^{-1\top}\boldsymbol{D}(\boldsymbol{D}^\top \boldsymbol{S}^{-1}\boldsymbol{D})^{-1}$$
- Thus, the covariance matrix of the GMM estimator is
$$\boldsymbol{\Omega} = (\boldsymbol{D}^\top \boldsymbol{S}^{-1}\boldsymbol{D})^{-1} \tag{6}$$
Note that this is only valid if we use optimal $\boldsymbol{W}$.

Asymptotic Normality of GMM Estimator

- Using the optimal weighting matrix the asymptotic normality of the GMM estimator can be written as:

$$\sqrt{T}\left(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0\right) \rightarrow N\left(\mathbf{0}, (\boldsymbol{D}^\top \boldsymbol{S}^{-1}\boldsymbol{D})^{-1}\right) \qquad (7)$$

- We can obtain an estimator for $\boldsymbol{\Omega}$ using optimal estimators for $\boldsymbol{D}$ and $\boldsymbol{S}$
- Note that this expression is only valid in large samples.
- Under exact identification (where $q = p$) covariance matrix does not depend on the optimality of the weighting matrix.
- To attain asymptotic efficiency the weighting matrix must be chosen optimally.

Asymptotic Normality of GMM Estimator

- The square root of the diagonal elements of $T^{-1}\boldsymbol{\Omega}$ gives asymptotic standard errors. Regular inference procedures are still valid.
- $100(1-\alpha)\%$ confidence intervals can be found using the formula

$$\hat{\boldsymbol{\theta}}_i \pm z_{\alpha/2}\sqrt{\hat{\boldsymbol{\Omega}}_{ii}/T}$$

  where $\hat{\boldsymbol{\Omega}}_{ii}$ is ith diagonal element of the consistent estimator of $\boldsymbol{\Omega}$.
- We can estimate $\boldsymbol{D}$ using

$$\hat{\boldsymbol{D}} = \frac{1}{T}\sum_{t=1}^{T}\frac{\partial f(\boldsymbol{y}_t, \hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}}$$

  which can be approximated by numerical differentiation methods after solving the numerical optimization problem.
- We also need to estimate the long-run covariance matrix of the moment conditions, $\boldsymbol{S}$.

Covariance Matrix of Moment Conditions

- For notational simplicity, let $\boldsymbol{f}_t = f(\boldsymbol{y}_t, \boldsymbol{\theta}_0)$ for the tth observation.
- Long-run covariance matrix is given by

$$\begin{aligned}
\boldsymbol{S} &= \lim_{T\to\infty} \mathsf{Var}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\boldsymbol{f}_t\right] \qquad (8)\\
&= \lim_{T\to\infty} \mathsf{E}\left[\left\{\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\boldsymbol{f}_t - \mathsf{E}[\boldsymbol{f}_t])\right\}\left\{\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\boldsymbol{f}_t - \mathsf{E}[\boldsymbol{f}_t])\right\}^\top\right]\\
&= \lim_{T\to\infty} \mathsf{E}\left[\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}(\boldsymbol{f}_t - \mathsf{E}[\boldsymbol{f}_t])(\boldsymbol{f}_s - \mathsf{E}[\boldsymbol{f}_s])^\top\right]
\end{aligned}$$

Covariance Matrix of Moment Conditions

- If $\boldsymbol{f}_t$ series is covariance stationary then autocovariances depend only on distance (not time).
- In this case for each t the jth autocovariance can be written as:

$$\boldsymbol{\Gamma}_j = \mathsf{E}\left[(\boldsymbol{f}_t - \mathsf{E}[\boldsymbol{f}_t])(\boldsymbol{f}_{t-j} - \mathsf{E}[\boldsymbol{f}_{t-j}])^\top\right]$$

- Using autocovariance function the long run covariance matrix can be written as (see Hamilton, Time Series Analysis, pp.279-80, and Hall, 2005, pp. 74-75):

$$\begin{aligned}
\boldsymbol{S} &= \boldsymbol{\Gamma}_0 + \lim_{T\to\infty}\left\{\sum_{j=1}^{T-1}\left(\frac{T-j}{T}\right)(\boldsymbol{\Gamma}_j + \boldsymbol{\Gamma}_j^\top)\right\} \qquad (9)\\
&= \boldsymbol{\Gamma}_0 + \sum_{i=1}^{\infty}(\boldsymbol{\Gamma}_i + \boldsymbol{\Gamma}_i^\top) \qquad (10)
\end{aligned}$$

Covariance Matrix of Moment Conditions

$$\boldsymbol{S} = \boldsymbol{\Gamma}_0 + \sum_{i=1}^{\infty}(\boldsymbol{\Gamma}_i + \boldsymbol{\Gamma}_i^\top)$$

- If $\boldsymbol{f}_t$ is **uncorrelated** with its lags, i.e., $\boldsymbol{\Gamma}_j = 0$, $j \neq 0$, the second term in the expression above drops:
$$\boldsymbol{S} = \boldsymbol{\Gamma}_0 = E[\boldsymbol{f}_t \boldsymbol{f}_t^\top]$$
- In this case a simple consistent estimator for $\boldsymbol{S}$ is:
$$\hat{\boldsymbol{S}} = \hat{\boldsymbol{\Gamma}}_0 = \frac{1}{T}\sum_{t=1}^{T}\hat{\boldsymbol{f}}_t \hat{\boldsymbol{f}}_t^\top$$
- Let $\boldsymbol{F}$ be the $T \times q$ matrix of sample moment conditions whose typical element is $\hat{\boldsymbol{f}}_{t,k}$, $k = 1, \ldots, q$, $t = 1, \ldots, T$. If moments are uncorrelated (no autocorrelation) then
$$\hat{\boldsymbol{S}} = \frac{1}{T}\boldsymbol{F}^\top \boldsymbol{F}$$

Covariance Matrix of Moment Conditions

- If the sample moments are autocorrelated (and/or heteroskedastic) the simple estimator for $\boldsymbol{S}$ will not be valid.
- There are several consistent estimators which differ in choosing the lag truncation parameter (bandwidth).
- Let the bandwidth be ν. The Hansen-White covariance matrix estimator is defined as (see Davidson and McKinnon, p.361):
$$\hat{\boldsymbol{S}}_{HW} = \hat{\boldsymbol{\Gamma}}_0 + \sum_{i=1}^{\nu}(\hat{\boldsymbol{\Gamma}}_i + \hat{\boldsymbol{\Gamma}}_i^\top) \tag{11}$$
- In this expression the population autocovariances can be consistently estimated using sample counterparts
$$\hat{\boldsymbol{\Gamma}}_j = \frac{1}{T}\sum_{t=j+1}^{T}\hat{\boldsymbol{f}}_t \hat{\boldsymbol{f}}_{t-j}$$
which is always positive definite in the limit.

- Positive (semi-) definiteness is the minimal requirement that a covariance matrix has to satisfy. However, the expression above may not satisfy positive definiteness in finite samples (it is always positive definite in the limit).
- The solution is to define a weight parameter which depends on the distance of the autocovariances - HAC estimator.
- The Heteroskedasticity and Autocorrelation Consistent (HAC) covariance matrix estimator is defined as
$$\hat{\boldsymbol{S}}_{HAC} = \hat{\boldsymbol{\Gamma}}_0 + \sum_{j=1}^{\nu}\omega_{j,\nu}(\hat{\boldsymbol{\Gamma}}_i + \hat{\boldsymbol{\Gamma}}_i^\top) \tag{12}$$
where $\omega_{j,\nu}$ is called the kernel (weight) function and ν is the lag truncation (bandwidth) parameter.
- Most frequently used kernel functions are Bartlett, suggested by Newey and West (1987); and Parzen, suggested by Gallant (1987).

HAC Covariance Matrix Estimator

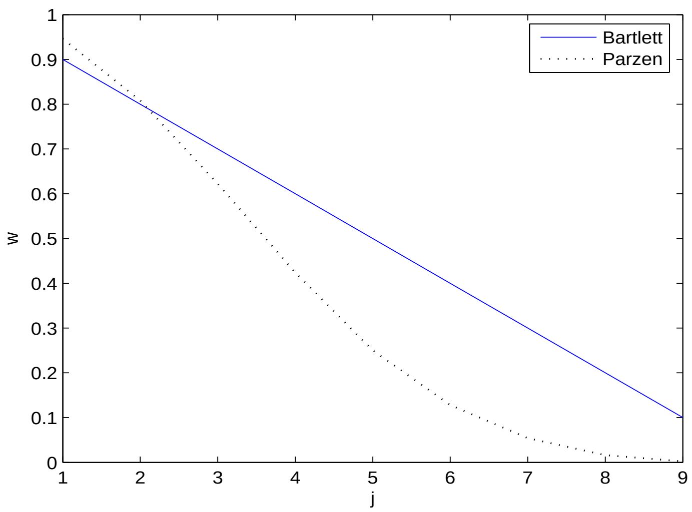
- Bartlett kernel (Newey-West, 1987)
$$\omega_{j,\nu} = \begin{cases} 1 - \frac{j}{\nu+1}, & \frac{j}{\nu+1} \leq 1\ ; \\ 0, & \frac{j}{\nu+1} > 1\ . \end{cases}$$
- For example, for $\nu = 3$ the Newey-West HAC estimator is
$$\hat{\boldsymbol{S}}_{NW} = \hat{\boldsymbol{\Gamma}}_0 + \frac{3}{4}(\hat{\boldsymbol{\Gamma}}_1 + \hat{\boldsymbol{\Gamma}}_1^\top) + \frac{1}{2}(\hat{\boldsymbol{\Gamma}}_2 + \hat{\boldsymbol{\Gamma}}_2^\top) + \frac{1}{4}(\hat{\boldsymbol{\Gamma}}_3 + \hat{\boldsymbol{\Gamma}}_3^\top)$$

HAC estimator: Parzen weights

$$\omega_{j,\nu} = \begin{cases} 1 - 6\left(\frac{j}{\nu+1}\right)^2 + 6\left(\frac{j}{\nu+1}\right)^3, & 0 \le \frac{j}{\nu+1} \le \frac{1}{2} \\ 2\left(1 - \frac{j}{\nu+1}\right)^3, & \frac{1}{2} \le \frac{j}{\nu+1} \le 1 \\ 0, & \frac{j}{\nu+1} > 1 \end{cases}$$

See the next graph for comparison. There are other HAC estimators as well (quadratic spectral kernel, VARMA covariance matrix estimator, Andrews and Monahan estimator). Another choice we have to make is the bandwidth parameter ν. The bandwidth should increase as T increases.

Bartlett and Parzen kernels for $\nu = 9$



Iterated GMM Estimator

- We saw that the optimal weighting matrix is given by the inverse of the long run covariance matrix of moment conditions: $W = \hat{S}^{-1}$
- GMM estimator is efficient if we use this expression. However, we need to find $\hat{\theta}$ before estimating W
- This implies a two-step iterated procedure for updating the GMM estimator: first, obtain $\hat{\theta}^{(1)}$ using a sub-optimal weighting matrix, such as the identity matrix $W = I$, and calculate the long run covariance matrix $\hat{S}^{(1)}$
- In the second step, re-estimate the model using $W = \hat{S}^{(1)-1}$ and obtain $\hat{\theta}^{(2)}$.
- We may iterate this procedure until a convergence criterion is satisfied.

Iterated GMM Estimator

- $i = 1$, first step:
  - Using a sub-optimal weighting matrix, e.g., $W = I_q$, estimate the parameter vector $\Longrightarrow \hat{\theta}^{(1)}$
  - Using $\hat{\theta}^{(1)}$ compute long run covariance matrix $\Longrightarrow \hat{S}^{(1)}$
- $i > 1$, following steps:
  - Determine iterated GMM convergence criterion $\Longrightarrow \epsilon$
  - Determine maximum number of iterations $\Longrightarrow maxit$
  - Using the long run covariance matrix estimate obtained in the $(i-1)$th step compute the weighting matrix $\Longrightarrow$ $W = \hat{S}^{(i-1)-1}$
  - If $\|\hat{\theta}^{(i)} - \hat{\theta}^{(i-1)}\| < \epsilon$ then stop the algorithm, iterated GMM estimate is $\Longrightarrow \hat{\theta} = \hat{\theta}^{(i)}$
  - If $\|\hat{\theta}^{(i)} - \hat{\theta}^{(i-1)}\| \ge \epsilon$ and $i < maxit$ then iterate step $i = i + 1$

33

Overidentifying Restrictions (OID) Test

- When the model is over-identified, $q > p$, population moment conditions can be decomposed into two parts: identifying restrictions, and over-identifying restrictions.
- Identifying restrictions (p dimensional) are used in the estimation of $\boldsymbol{\theta}$ and automatically satisfied in the sample.
- Over-identifying restrictions ($q - p$ dimensional) are not automatically satisfied in the sample. Also these two components are orthogonal to each other.
- Thus, over-identifying restrictions can be used to test if the moment conditions are satisfied.
- Hansen (1982) proposed the OID test statistic (also known as J-test):

$$J = TQ(\hat{\boldsymbol{\theta}}) \tag{13}$$

where $Q(\hat{\boldsymbol{\theta}})$ is the minimum value of the GMM objective function (evaluated at the solution)

34

Overidentifying Restrictions (OID) Test

- Under the null hypothesis that the model is correctly specified:

$$H_0 : \mathsf{E}[f(y_t; \theta)] = 0$$

OID test statistic has a chi-squared distribution with $q - p$ degrees of freedom:

$$J = TQ(\hat{\theta}) \to \chi^2_{q-p} \tag{14}$$

- A large value implies evidence against the null.

35

Numerical Example: t Distribution

- We examined how to estimate the d.o.f. parameter ν of the t distribution by the MM. We used only one population moment.
- Suppose that we want to use the fourth moment in the estimation as well:

$$\begin{aligned} E(Y_t^2) &= \frac{\nu}{\nu - 2} \\ E(Y_t^4) &= \frac{3\nu^2}{(\nu-2)(\nu-4)} \end{aligned}$$

- Now the $f(\cdot)$ function is

$$f(y_t, \nu) = \begin{bmatrix} y_t^2 - \frac{\nu}{\nu-2} \\ y_t^4 - \frac{3\nu^2}{(\nu-2)(\nu-4)} \end{bmatrix} \tag{15}$$

Note that $\mathsf{E}[f(y_t, \nu)] = \mathbf{0}$.

36

Numerical Example: t Distribution

- Replacing population moments with their sample counterparts we obtain:

$$m(\nu) = T^{-1} \sum_{t=1}^{T} f(y_t, \nu) \tag{16}$$

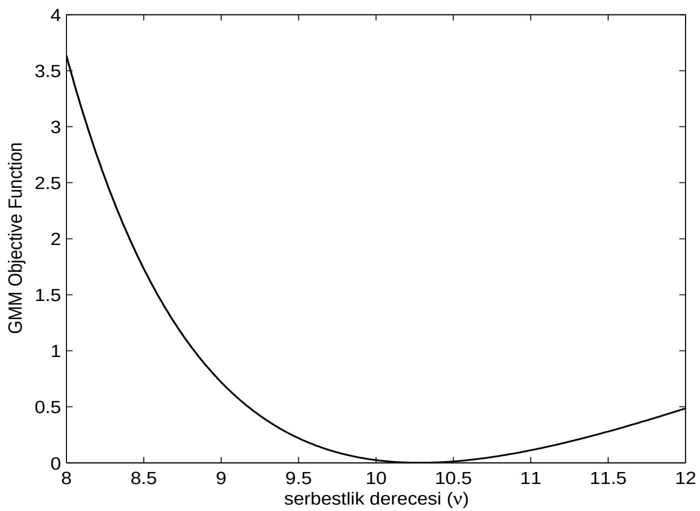
$$= \begin{bmatrix} T^{-1} \sum_{t=1}^{T} y_t^2 - \frac{\nu}{\nu-2} \\ T^{-1} \sum_{t=1}^{T} y_t^4 - \frac{3\nu^2}{(\nu-2)(\nu-4)} \end{bmatrix} \tag{17}$$

- The GMM objective function is

$$Q(\nu) = \boldsymbol{m}(\nu)^\top \boldsymbol{W} \boldsymbol{m}(\nu)$$

37

GMM Objective Function and t Distribution dof Parameter

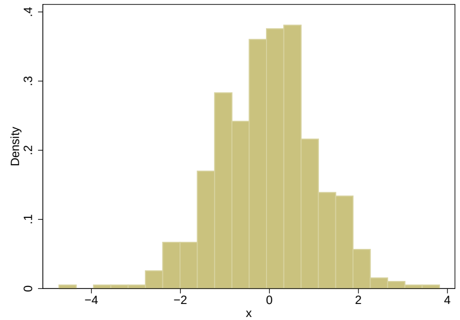


38

Numerical Example: t Distribution

Simulate 500 observations from $t(10)$ distribution and plot the histogram

```
clear
set obs 500
set seed 1111
gen x = rt(10)
hist x
```



39

Numerical Example: t Distribution

GMM Estimation in STATA: One-step GMM with identity as the weighting matrix

```
gmm (eq1: x^2-({nu=5})/({nu}-2))
    (eq2: x^4-3*{nu}^2/(({nu}-2)*({nu}-4))),
     winitial(identity) onestep

Step 1
Iteration 0:   GMM criterion Q(b) =  372.49635
Iteration 1:   GMM criterion Q(b) =  .16946279
             :
Iteration 5:   GMM criterion Q(b) =  .00141583

GMM estimation

Number of parameters =   1
Number of moments    =   2
Initial weight matrix: Identity                        Number of obs   =      500

------------------------------------------------------------------------------
             |               Robust
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         /nu |   11.13609   3.088634     3.61   0.000     5.082482    17.18971
------------------------------------------------------------------------------
Instruments for equation 1: _cons
Instruments for equation 2: _cons
```

40

Numerical Example: t Distribution

GMM Estimation in STATA: Two-step GMM with identity as the weighting matrix

```
gmm (eq1: x^2-({nu=5})/({nu}-2))
    (eq2: x^4-3*{nu}^2/(({nu}-2)*({nu}-4))),
     winitial(identity) twostep
Step 1
Iteration 0:   GMM criterion Q(b) =  372.49635
Iteration 1:   GMM criterion Q(b) =  .16946279
             :
Iteration 5:   GMM criterion Q(b) =  .00141583

Step 2
Iteration 0:   GMM criterion Q(b) =  .00113535
Iteration 1:   GMM criterion Q(b) =  .00113412
Iteration 2:   GMM criterion Q(b) =  .00113405

GMM estimation

Number of parameters =   1
Number of moments    =   2
Initial weight matrix: Identity                        Number of obs   =      500
GMM weight matrix:     Robust

------------------------------------------------------------------------------
             |               Robust
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         /nu |   11.20309   3.149931     3.56   0.000     5.029339    17.37684
------------------------------------------------------------------------------
Instruments for equation 1: _cons
Instruments for equation 2: _cons
```

Numerical Example: t Distribution

Hansen's OID test statistic:

```
. estat overid

  Test of overidentifying restriction:

  Hansen's J chi2(1) = .567022 (p = 0.4514)
```

We fail to reject the null hypothesis that the moment conditions are valid. The model is correctly specified.

Example: Consumption-Based Asset Pricing Model (Hansen and Singleton, 1982)

- Representative agent makes decisions about consumption expenditures and investment to maximize his/her expected discounted utility

$$E\left[\sum_{i=0}^{\infty} \delta^i U(c_{t+i}|\Omega_t)\right]$$

where c_t is consumption in period t, $U(\cdot)$ is a strictly concave utility function, δ is discount factor and Ω_t is the information set at time t.

- Income is either spent on consumption or invested
- N assets with maturities m_j, $j = 1, 2, \dots, n$

Example: Consumption-Based Asset Pricing Model (Hansen and Singleton, 1982)

- $q_{j,t}$ is the quantity of asset j held at the end of period t
- $p_{j,t}$ is the price of asset j at time t
- $r_{j,t}$ is the period t payoff from asset j
- Budget constraint is

$$c_t + \sum_{j=1}^{N} p_{j,t} q_{j,t} = \sum_{j=1}^{N} r_{j,t} q_{j,t-1} + w_t$$

where w_t is the labor income. We also assumed maturity of 1 period.

Example: CAPM

- The optimal path of consumption and investment satisfies

$$p_{j,t} U'(c_t) = \delta E[r_{j,t+1} U'(c_{t+1})|\Omega_t]$$

where $U'(c)$ is the marginal utility of consumption

- Utility lost by purchasing a unit of asset j must equal the value of the expected utility gained from consuming the return on the investment in period $t+1$
- Euler equation can be written as

$$E\left[\delta\left(\frac{r_{j,t+1}}{p_{j,t}}\right)\left(\frac{U'(c_{t+1})}{U'(c_t)}\right)|\Omega_t\right] - 1 = 0$$

Example: CAPM

- Assuming utility is CRRA type

$$U(c_t) = \frac{c_t^\gamma - 1}{\gamma}$$

- The Euler equation can be written as

$$E\left[\delta\left(\frac{r_{t+1}}{p_t}\right)\left(\frac{c_{t+1}}{c_t}\right)^{\gamma-1}|\Omega_t\right] - 1 = 0$$

- Parameter vector to be estimated $\theta = (\delta, \gamma)^\top$

Example: CAPM

- Letting $x_{1,t+1} = c_{t+1}/c_t$ and $x_{2,t+1} = r_{t+1}/p_t$, population moment conditions can be written as

$$E[z_t(\delta x_{1,t+1}^{\gamma-1} x_{2,t+1} - 1)] = 0,$$

  where z_t is the vector of instruments which may include 1 and lagged values of x_1 and x_2
- Thus, sample moment conditions can be written as

$$T^{-1}\sum_{t=1}^{T} f(x_1, x_2, \theta)$$

  where $f(x_1, x_2, \theta) = z_t(\delta x_{1,t+1}^{\gamma-1} x_{2,t+1} - 1)$

Example: CAPM

- GMM objective function can be written as

$$Q_T(\theta) = \left\{T^{-1}\sum_{t=1}^{T} f(x_1, x_2, \theta)\right\}^\top W_T \left\{T^{-1}\sum_{t=1}^{T} f(x_1, x_2, \theta)\right\}$$

  where W_T is the weighting matrix.
- GMM estimator of θ is

$$\hat{\theta}_T = \arg\min_{\theta\in\Theta} Q_T(\theta)$$

- Solution to this problem can be found numerically using either quasi-Newton type routines or derivative-free methods.

Example: CAPM

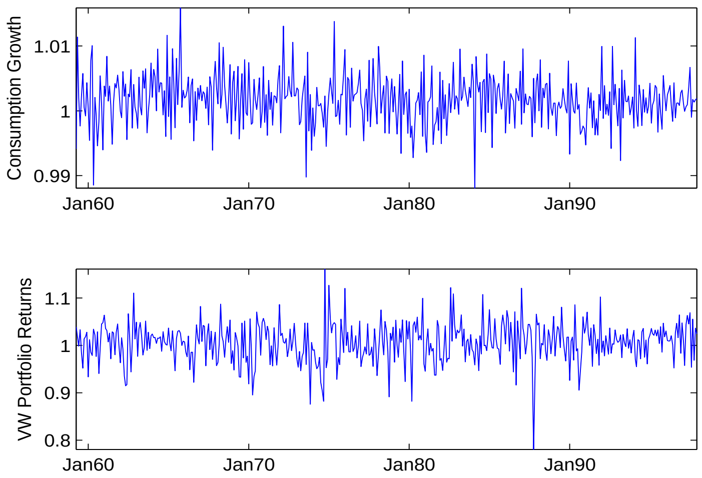
- Original data as used in Hall (2005, Generalized Method of Moments, Oxford Univ. Press, p.60)
- Monthly U.S. data for the 1959.01-1997.12 period. Value-weighted portfolio returns from NY Stock Exchange.
- Data are organized in `cbapmvwrdata.dat`
- Which can be read into MATLAB as follows

```
x = load('cbapmvwrdata.dat');
```

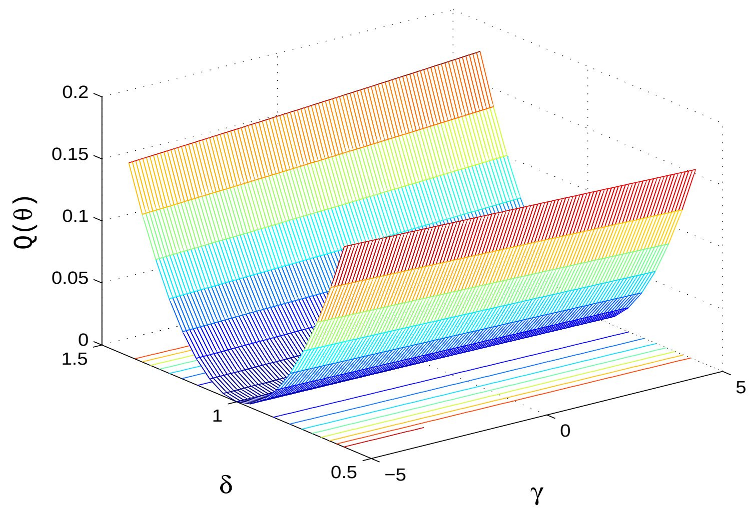
  where the first column is consumption growth (x_1) and the second column is portfolio returns (x_2).
- We also need instruments (z_t). For example, we can use the following instrument vector:

$$z_t = \begin{bmatrix} 1 & x_{1,t} & x_{1,t-1} & x_{2,t} & x_{2,t-1} \end{bmatrix}^\top$$

CBAPM US DATA



CBAPM GMM Objective Function



CBAPM GMM Estimation Results

see the mfile: `CBAPM_GMMest.m`

```
GMM ESTIMATION RESULTS
-----------------------------------------------------------------
Moment function is CBAPmom
Number of observations     =    465
Number of pop moments      =      5
Number of parameters       =      2
Number of OID restrictions =      3

    Parameters        ASE        t-stat     95% Confidence Interval
     0.666628     1.827777     0.364720   -2.915816        4.249071
     0.993592     0.003545   280.306749    0.986645        1.000540

     OID Test     p-value
     1.747997     0.626316
*****************************************************************
```


GMM Estimation of CBAPM in STATA

One-step GMM:

```
. gmm ( 1 - {delta=1}*x1lead^({gamma=1}-1)*x2lead ),
     inst(x1 x1lag x2 x2lag) onestep

Step 1
Iteration 0:   GMM criterion Q(b) =  .00012873
Iteration 1:   GMM criterion Q(b) =  .00006561
Iteration 2:   GMM criterion Q(b) =  .00006561

GMM estimation

Number of parameters =   2
Number of moments    =   5
Initial weight matrix: Unadjusted                    Number of obs   =      465

------------------------------------------------------------------------------
             |               Robust
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      /delta |   .9931797   .0043934   226.06   0.000     .9845688    1.001791
      /gamma |   .3981941   2.263423     0.18   0.860    -4.038034    4.834422
------------------------------------------------------------------------------
Instruments for equation 1: x1 x1lag x2 x2lag _cons
```

GMM Estimation of CBAPM in STATA

Two-step GMM:

```
. gmm ( 1 - {delta=1}*x1lead^({gamma=1}-1)*x2lead ),
      inst(x1 x1lag x2 x2lag) wmat(hac nw 4) twostep

Step 1
Iteration 0:   GMM criterion Q(b) =  .00012873
Iteration 1:   GMM criterion Q(b) =  .00006561
Iteration 2:   GMM criterion Q(b) =  .00006561

Step 2
Iteration 0:   GMM criterion Q(b) =  .02474876
Iteration 1:   GMM criterion Q(b) =  .02322387
Iteration 2:   GMM criterion Q(b) =  .02322387

GMM estimation

Number of parameters =   2
Number of moments    =   5
Initial weight matrix: Unadjusted                   Number of obs   =      465
GMM weight matrix:     HAC Bartlett 4

------------------------------------------------------------------------------
             |                 HAC
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      /delta |    .990941   .0043973   225.35   0.000     .9823224    .9995595
      /gamma |   .5662718   2.032626     0.28   0.781    -3.417602    4.550145
------------------------------------------------------------------------------
HAC standard errors based on Bartlett kernel with 4 lags.
Instruments for equation 1: x1 x1lag x2 x2lag _cons
```


GMM Estimation of CBAPM in STATA

Iterated GMM:

```
  :
  :
Step 9
Iteration 0:   GMM criterion Q(b) =  .02297779
Iteration 1:   GMM criterion Q(b) =  .02297779
iterative GMM weight matrix converged
iterative GMM parameter vector converged

GMM estimation

Number of parameters =   2
Number of moments    =   5
Initial weight matrix: Unadjusted                   Number of obs   =      465
GMM weight matrix:     HAC Bartlett 4

------------------------------------------------------------------------------
             |                 HAC
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      /delta |   .9904615   .0043946   225.38   0.000     .9818482    .9990749
      /gamma |   .5938478   2.031959     0.29   0.770    -3.388719    4.576415
------------------------------------------------------------------------------
HAC standard errors based on Bartlett kernel with 4 lags.
Instruments for equation 1: x1 x1lag x2 x2lag _cons
```


GMM Estimation of CBAPM in STATA

OID TEST:

```
. estat overid

  Test of overidentifying restriction:

  Hansen's J chi2(3) = 10.6847 (p = 0.0136)
```

We reject the null hypothesis that the moment conditions are valid at 5% significance level.