Economic modelling and forecasting

2-6 February 2015

The generalised method of moments

Ole Rummel

Adviser, CCBS at the Bank of England

ole.rummel@bankofengland.co.uk

BANK OF ENGLAND

Outline

- Classical estimation methodologies
- Why GMM?
- Moments and the method of moments
- Orthogonality or moment conditions
- The generalised method of moments (GMM)
- HAC estimators and the GMM weighting matrix
- Testing restrictions with GMM
- The choice of instruments and the efficiency bound
- Summary

Estimation methodologies

- There are a number of **classical** methodologies for estimation and inference:
  - (ordinary) least squares (OLS) – and its extensions such as two-stage least squares, generalised least squares and instrumental variables; and
  - maximum-likelihood (ML) – which is the best available estimator within the classical paradigm
- Why would I want to introduce yet another approach to estimation – the **generalised method of moments** (GMM)?

Reasons for GMM

- OLS has some stringent underlying assumptions
- Two particular problems with ML have motivated the use of GMM:
  - the sensitivity of statistical properties to the distributional assumption underlying ML estimation; and
  - the computational burden
- The GMM framework provides a computationally convenient method of performing estimation and inference while avoiding often unwanted or unnecessary assumptions, such as specifying a particular distribution (i.e., the likelihood function) for the errors
- Finally, GMM is a broadly applicable parameter estimation methodology that nests many of the classical (frequentist) estimation techniques

Definition of population moments

- We define a **population** moment, v, as the expectation of some continuous function $m(\bullet)$ of a (discrete) random variable, x, describing the population of interest:

$$v = \mathrm{E}[m(x)] \qquad (1)$$

- The notion of a moment is fundamental for describing features of a population (Wooldridge (2001))
- The population mean, v_1, or first moment about the origin, measures central tendency and is given by:

$$v_1 = \mathrm{E}[x] \qquad (2)$$

in which case $m(\bullet)$ is the identity function

Higher moments and functions of moments

- The (uncentred) second population moment is given by:

$$v_2 = \mathrm{E}[x^2] \tag{3}$$

- The population variance of x is a measure of spread in a distribution
- It is defined as the second moment of x centred about its mean and can be expressed as a function of the first two population moments, v_1 and v_2:

$$\begin{aligned}\mathrm{Var}(x) &= \mathrm{E}[x^2] - (\mathrm{E}[x])^2 \\ &= v_2 - v_1^2 \end{aligned} \tag{4}$$

- Functions of moments, such as $\mathrm{Var}(x)$, are also called moments

Definition of sample moments

- So far, we have been concerned with **population** moments (as in (1) through (4))
- Since we can rarely obtain information on an entire population, we use a (random) sample $\{x_i: i = 1, \dots, T\}$ from the population to estimate population moments
- In estimation, we therefore need to define the **sample** moment
- The sample moment is merely the sample version of the population moment in a particular random sample of size T:

$$\hat{v} = \frac{1}{T}\sum_{t=1}^{T} m(x_t) \qquad (5)$$

where we replace the expectation by the sample average

Orthogonality or moment conditions (1)

- The population moment conditions can also be defined to set expectations of functions of the data, x, and unknown parameters, θ_0, to particular (known) values, frequently zero:

$$E[m(x,\theta_0)] = 0 \qquad (6)$$

- One simple restriction involves the mean μ of data x_t:

$$E[x_t] = \mu \qquad (7)$$

giving the population orthogonality or moment condition:

$$E[x_t - \mu] = 0 \qquad (8)$$

where $m(x,\theta_0) = x_t - \mu$ and θ_0 contains one unknown parameter, such that $\theta_0 = \mu$

Orthogonality or moment conditions (2)

- The sample counterpart to (8) is:

$$E[m(x,\theta_0)] = \frac{1}{T}\sum_{t=1}^{T}(x_t - \mu)$$
$$= \frac{1}{T}\sum_{t=1}^{T}x_t - \frac{1}{T}\sum_{t=1}^{T}\mu = \frac{1}{T}\sum_{t=1}^{T}x_t - \mu = 0$$
$$\mu = \frac{1}{T}\sum_{t=1}^{T}x_t \qquad (9)$$

- This is the classical **methods of moments** estimator of μ
- Under random sampling, this estimator will be unbiased and consistent for μ regardless of other features of the underlying population

Orthogonality or moment conditions (3)

- A moment restriction for the variance (σ^2) is:

$$E[(x_t - \mu)^2] = \sigma^2 \qquad (10)$$

- Note that (10) now contains two unknowns, such that $\theta_0 = (\mu, \sigma)$
- This means that **two** moment conditions are now needed
- We therefore have to also include the moment condition for the mean to estimate the variance:

$$E\begin{bmatrix} x_t - \mu \\ (x_t - \mu)^2 - \sigma^2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad (11)$$

resulting in the system (11) containing two equations in the **two** unknowns, μ and σ

Orthogonality or moment conditions (4)

- Similarly, a covariance moment restriction would be:

$$E[(x_t - \mu_x)(y_t - \mu_y)] = \sigma_{x,y} \quad (12)$$

which involves three unknowns, $\theta_0 = (\mu_x, \mu_y, \sigma_{x,y})$, resulting in the system:

$$E\begin{bmatrix} x_t - \mu_x \\ y_t - \mu_y \\ (x_t - \mu_x)(y_t - \mu_y) - \sigma_{x,y} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad (13)$$

Example: estimating an MA(1) process

- The parameters of a moving-average (MA) model cannot be estimated by OLS, but maximum-likelihood and non-linear least-squares estimators are available
- For an MA(1) process, $y_t = \varepsilon_t + \theta\varepsilon_{t-1}$, we have:

$$E[y_t^2] = E[(\varepsilon_t + \theta\varepsilon_{t-1})^2] = \sigma_\varepsilon^2(1 + \theta^2) \qquad (14)$$

$$E[y_t y_{t-1}] = E[(\varepsilon_t + \theta\varepsilon_{t-1})(\varepsilon_{t-1} + \theta\varepsilon_{t-2})] = \theta\sigma_\varepsilon^2 \qquad (15)$$

which are the variance and the first autocovariance respectively

- The two sample moment conditions could therefore be:

$$\begin{pmatrix} \frac{1}{T}\sum_{t=1}^{T} y_t^2 - \sigma_\varepsilon^2(1+\theta^2) \\ \frac{1}{T}\sum_{t=1}^{T} y_t y_{t-1} - \sigma_\varepsilon^2\theta \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \qquad (16)$$

Orthogonality or moment conditions (5)

- GMM extends the classical set-up in two ways:
  - it formally treats the problem of having two or more moment conditions which have information about unknown parameters; and
  - it permits the use of quantities other than sample moments in estimation
- The generalised method of moments (GMM) also uses moment or orthogonality conditions and a key building block of GMM estimation, inference and testing is the specification of the appropriate moment, or orthogonality, condition, $m(x,\theta_0)$
- An important approach to model specification testing is to base tests directly on certain conditions that the error terms of a model should satisfy

Orthogonality or moment conditions (6)

- The moment conditions in GMM are therefore commonly based on the error terms from an economic model
- The basic idea is that if a model is correctly specified, many random quantities which are **functions** of the error terms should have an expectation of zero
- The specification of a model sometimes allows a stronger conclusion, according to which such functions of the error terms have zero expectation conditional on some **information set**

Orthogonality or moment conditions (7)

- Consider a general model for a **stationary** series, y_t, of the form:

$$\underset{(T\times 1)}{y} = f(\underset{(T\times k)}{X}; \underset{(k\times 1)}{\theta_0}) + \underset{(T\times 1)}{\varepsilon} \qquad (17)$$

  where $f(X;\theta_0)$ can be a **non-linear** function

- This $f(X;\theta_0)$ is specified to be the mean of y_t conditional on some **information set**, Ω_t:

$$E[y_t|\Omega_t] = f(X;\theta_0) \qquad (18)$$

- This information set, Ω_t, contains the information that is available and potentially used at the time of estimation – ideally, it contains all the information that is observed and known at time T
- We usually assume that the information set at time T contains the values of y_t (t = 1, … , T) and the X_t's (t = 1, … , T) and therefore also all of their lags

Orthogonality or moment conditions (8)

- If this model is specified correctly, then the conditional expectation of ε_t is given by:

$$E[\varepsilon_t|\Omega_t] = E[y_t - f(X;\theta_0)|\Omega_t] = 0 \qquad (19)$$

- If k denotes the number of parameters in (17), it is clear that we need at least k moment conditions in order to define a full set of parameter estimates, θ_0
- But (19) seems to provide no more than one moment condition
- The way out of this constitutes one of the most important features of GMM

Orthogonality or moment conditions (9)

- It turns out that the conditional nature of the expectation (19) produces further implication for the model
- Not only is the moment condition (19) required to have a mean of zero…
- …but the **conditional** nature of moment condition (19) also implies that ε_t should be orthogonal (uncorrelated) to any variable that belongs to Ω_t – hence, for any **vector** $Z_t = (z_{1t}, z_{2t}, \ldots , z_{qt})$ such that $Z_t \in \Omega_t$, the unconditional moments $E[Z_t \varepsilon_t] = E[Z_t(y_t - f(X;\theta_0))]$ should also be zero
- The corresponding empirical (sample) moments are:

$$\frac{1}{T}\sum_{t=1}^{T} Z_t \hat{\varepsilon}_t = \frac{1}{T} Z' \hat{\varepsilon} = 0 \qquad (20)$$

Orthogonality or moment conditions (10)

- In making use of the fact that E[$Z'\varepsilon$] = 0, the q elements of $Z_t = (z_{1t}, z_{2t}, \dots, z_{qt})$ are referred to as **instruments**
- In the sample, the model errors are:

$$\varepsilon(y, X; \hat{\theta}_0) = y - f(X; \hat{\theta}_0) \tag{21}$$

giving the moment conditions:

$$E[m(y, X; \hat{\theta}_0)] = \frac{1}{T}\sum_{t=1}^{T} Z_t \varepsilon(y_t, x_t; \hat{\theta}_0) = \frac{1}{T} Z'\varepsilon(y, X; \hat{\theta}_0) = 0 \tag{22}$$

- In time series models, we could therefore add moment conditions by assuming that past values of explanatory variables, or even past values of the dependent variable, are uncorrelated with the error term, even if they do not appear in the model

Orthogonality or moment conditions (11)

- Based on these instruments, we can add as many moment conditions as are required ($q = k$)
- At the very least, the regressors X_t belong to the information set Ω_t and there are precisely k of them
- We may therefore use the k regressors in X_t to define the k unconditional moment conditions:

$$\mathrm{E}[Z'\varepsilon] = \mathrm{E}[X'\varepsilon] = 0 \qquad (23)$$

The OLS estimator (1)

- The usefulness of GMM comes from the fact that the object of interest in many estimation exercises is simply a function of moments
- If the linear regression model $y = X\beta + \varepsilon$ is correctly specified, then

$$E[\varepsilon] = 0 \quad \text{and} \quad E[X'\varepsilon] = 0 \qquad (24)$$

- This allows us to set up a moment condition that can be used to obtain an estimate of β

The OLS estimator (2)

- In the **population**:

$$E[X'(y - X\beta)] = 0 \qquad (25)$$

where we have used the fact that $\varepsilon = y - X\beta$

- The method of moments suggests replacing the left-hand side of the above equation with its **sample** equivalent:

$$(1/T)X'(y - X\beta) \qquad (26)$$

- How can we proceed to obtain an estimate of β?

The OLS estimator (3)

- We know that the true β sets the **population** moment (25) equal to zero in expectation, so it seems reasonable to assume that a good choice of $\hat{\beta}$ would be one that sets the **sample** moment to zero:

$$(1/T)X'(y - X\hat{\beta}) = 0 \qquad (27)$$

- This is a set of k simultaneous equations with k unknown parameters and, provided that X has full column rank, the GMM estimator of β is:

$$\hat{\beta} = (X'X)^{-1}X'y \qquad (28)$$

Interim summary (1)

- In order to identify the *k*-vector of parameters, θ_0, *w*e need $q \geq k$ (independent) moment or orthogonality conditions
- The population version of each of these restrictions (i = 1, … , q) is of the form:

$$E[m_i(y, X; \theta_0)] = 0 \qquad (29)$$

- The sample analogue to (29) for each i = 1, … , q is:

$$E[m_i(y, X; \hat{\theta}_0)] = \frac{1}{T}\sum_{t=1}^{T} m_{it}(y_t, x_t; \hat{\theta}_0) = 0 \qquad (30)$$

Interim summary (2)

- Many statistical and economic models feature q moment or orthogonality conditions of the form:

$$E[m(y,X;\theta_0)] = \begin{pmatrix} E[m_1(y,X;\theta_0)] \\ \vdots \\ E[m_q(y,X;\theta_0)] \end{pmatrix} = 0 \qquad (31)$$

  from which we want to estimate the $(k \times 1)$ $(k \leq q)$ vector of true parameters, θ_0

- GMM requires not only that $E[m(y,X;\theta_0)] = 0$ at the 'true' parameter value, θ_0, but also that $E[m(y,X;\theta)] \neq 0$ for $\theta \neq \theta_0$ for identification

Shortcomings of the method of moments

- Clearly, some modification is needed in order to produce estimates of k parameters based on more than k population moment conditions (incidentally, this problem is not specific to GMM estimation alone):
  - condition (31) may not have an exact solution for θ_0 when there are more moment conditions, q, than there are parameters in θ_0 ($q > k$)
- An important feature of GMM is that it allows more moment conditions than there are parameters to estimate – that is, it allows the parameters to be **overidentified**
- The modification suggested by Hansen (1982) re-formulates the generalised method of moment approach as an (equivalent) minimisation problem

Advantages of GMM

- The key difference is that the method of moments is defined as the solution to a set of moment conditions (and this solution only exists if there are as many moment conditions as parameters, i.e., $q = k$), whereas GMM is defined in terms of a minimisation, which can be performed for any number of moment conditions ($q \geq k$)
- We therefore reformulate the problem as one of choosing a θ so that the sample moments $m(y,X;\theta)$ are as close to zero as possible
- But in order to implement such a strategy and to define what ‘close’ means, it is necessary to specify an appropriate minimand as a measure of distance

The theory behind GMM: the quadratic form

- Hansen (1982) proposed a **quadratic form** in the sample moment conditions as a measure of distance
- GMM therefore chooses the parameters, θ, which minimise the quadratic form:

$$J(\theta) = m(y,X;\theta)'Wm(y,X;\theta) \quad (32)$$

  where θ is a k-vector of parameters, $m(y,X;\theta)$ is a q-vector of orthogonality or moment conditions as in (31) and W is a $(q \times q)$ symmetric, positive definite **weighting matrix** that may (and generally will) depend on the data
- In other words, GMM weights the q sample moment conditions to obtain an asymptotically optimal estimator

The GMM estimator (1)

- The GMM estimator is defined as the value of θ that minimises the weighted quadratic form (32), which is equivalent to a loss function
- In fact, $J(\theta)$ is a non-negative measure of the 'length' of the vector $m(y,X;\theta)$
- For example, when $W = I_n$, then $J(\theta)$ is the square of the Euclidian length or distance of $m(y,X;\theta)$:

$$J(\theta) = m(y,X;\theta)'m(y,X;\theta)$$

$$= m_1(y,X;\theta)^2 + m_2(y,X;\theta)^2 + ... + m_q(y,X;\theta)^2$$

The GMM estimator (2)

- Equation (32) written out looks like:

$$J(\theta) = \begin{pmatrix} \frac{1}{T}\sum_{t=1}^{T} m_{1t}(y, X; \theta) \\ \vdots \\ \frac{1}{T}\sum_{t=1}^{T} m_{qt}(y, X; \theta) \end{pmatrix}' \begin{pmatrix} w_{11} & \cdots & w_{1q} \\ \vdots & \ddots & \vdots \\ w_{q1} & \cdots & w_{qq} \end{pmatrix} \begin{pmatrix} \frac{1}{T}\sum_{t=1}^{T} m_{1t}(y, X; \theta) \\ \vdots \\ \frac{1}{T}\sum_{t=1}^{T} m_{qt}(y, X; \theta) \end{pmatrix}$$

(1×1) $(1 \times q)$ $(q \times q)$ $(q \times 1)$

where θ is a k-vector of parameters and W is some $(q \times q)$ symmetric, positive definite weighting matrix

- The role of the weighting matrix W is to determine the relative importance of the various moment conditions

The weighting matrix *W* (1)

- In order to operationalise the GMM estimator, q, the number of moment conditions, will be required to greater than or equal to p, the number of unknown parameters
- If there are as many moment conditions as parameters, the moments will all be perfectly matched and the objective function $J(\theta)$ in (32) will have a value of zero – this is referred to as the **just-identified** case
- In the situation where there are more moment conditions than parameters (the **overidentified** case), not all the moment restrictions will be satisfied (meaning that $J(\theta)$ in (32) will be greater than zero)

The weighting matrix *W* (2)

- Hansen (1982) and White (1982) showed that the optimal choice for the weighting matrix *W* is a heteroskedasticity and autocorrelation consistent (HAC) estimate of the inverse of the asymptotic variance-covariance matrix $E[m(\theta_0)m(\theta_0)']^{-1} = S^{-1}$ of the moment conditions (where θ_0 is the true value of the parameter θ)
- Any weighting matrix other than S^{-1} will give a less efficient GMM estimator
- The description of the optimal weighting matrix is somewhat circular – before we can estimate the unknown parameter vector, θ, we need an estimate of the matrix S, and before we can estimate the matrix S, we need an estimate of the unknown parameter vector, θ

The weighting matrix *W* (3)

- We therefore follow an iterative or multi-step estimation process:
  - estimate the model **suboptimally** with some simple (symmetric, positive definite) weighting matrix that does not depend on θ (the identity matrix is typically a good choice), i.e. $W_{(0)} = I_q$;
  - this gives a consistent estimate of the parameters $\theta_{(1)}$ which can then be used to produce an initial (consistent) estimate $S_{(1)}$;
  - as advocated by Cochrane (2001), we can – and frequently should – stop here; but
  - we can also use the consistent $S_{(1)}$ in a second step to define a new weighting matrix as $W_{(1)} = [S_{(1)}]^{-1}$ (this is the classic **two-step GMM estimator**)

The weighting matrix *W* (4)

- The process can be iterated further by calculating $W_{(2)}$, then minimising again to find $\hat{\theta}_{(2)}$ and continuing in this vein, i.e., continuously iterating between θ and S using the most recent values of θ available – in general, iterating to end with $\hat{\theta}_{(n)}$ is called *n*-stage GMM
- You can also iterate until the change in the objective function (or in the estimator $\hat{\theta}_{(n)}$) is sufficiently small, i.e., up until the point some convergence criterion has been met
- Iteration until convergence eliminates any dependence on the initial weighting matrix and appears to improve the finite-sample performance of the GMM estimator when the number of parameters is large

The weighting matrix *W* (5)

- The final, and most complicated, type of GMM estimation is the continuously updating estimator (CUE)
- Instead of iterating between estimation of θ and S, this estimator parameterises S as a function of θ
- In this approach, θ is found as the minimum of:

$$J(\theta) = m(y,X;\theta)'[S(\theta)]^{-1}m(y,X;\theta) \qquad \text{(X)}$$

- Note that θ now appears in all three terms of the minimisation problem
- **If** the continuously updating estimator converges, it is generally regarded to have the best small-sample properties among the different GMM estimators

Estimating the variance-covariance matrix *S*

- Estimation of the asymptotic or long-run variance-covariance matrix S of the moment conditions is important and often has a significant impact on tests of either model or individual coefficients
- Residual-based estimates of *S* will need to confront two issues:
  - heteroskedasticity; and
  - serial- or auto-correlation
- Nowadays it has become common to construct heteroskedasticity and autocorrelation consistent (HAC) covariance matrices
- HAC estimation requires two choices:
  - the bandwidth; and
  - the kernel

Regression with correlated errors (1)

- In regression models with correlated errors, classical (and robust) standard errors, given by $\Sigma = \sigma^2(X'X)^{-1}$, are no longer appropriate
- The asymptotic variance of least-squares estimators needs to be multiplied by an adjustment factor for serial correlation, *f*, which is equal to:

$$f = 1 + 2\sum_{j=1}^{\infty} \rho(j)$$

  where the $\rho(j)$'s are autocorrelations
- The estimation of *f* for variances and standard errors under autocorrelation is called **heteroskedasticity and autocorrelation consistent** (HAC) variance estimation
- Having estimated the adjustment factor, *f*, we multiply conventional variance estimates by this estimate of *f*

Regression with correlated errors (2)

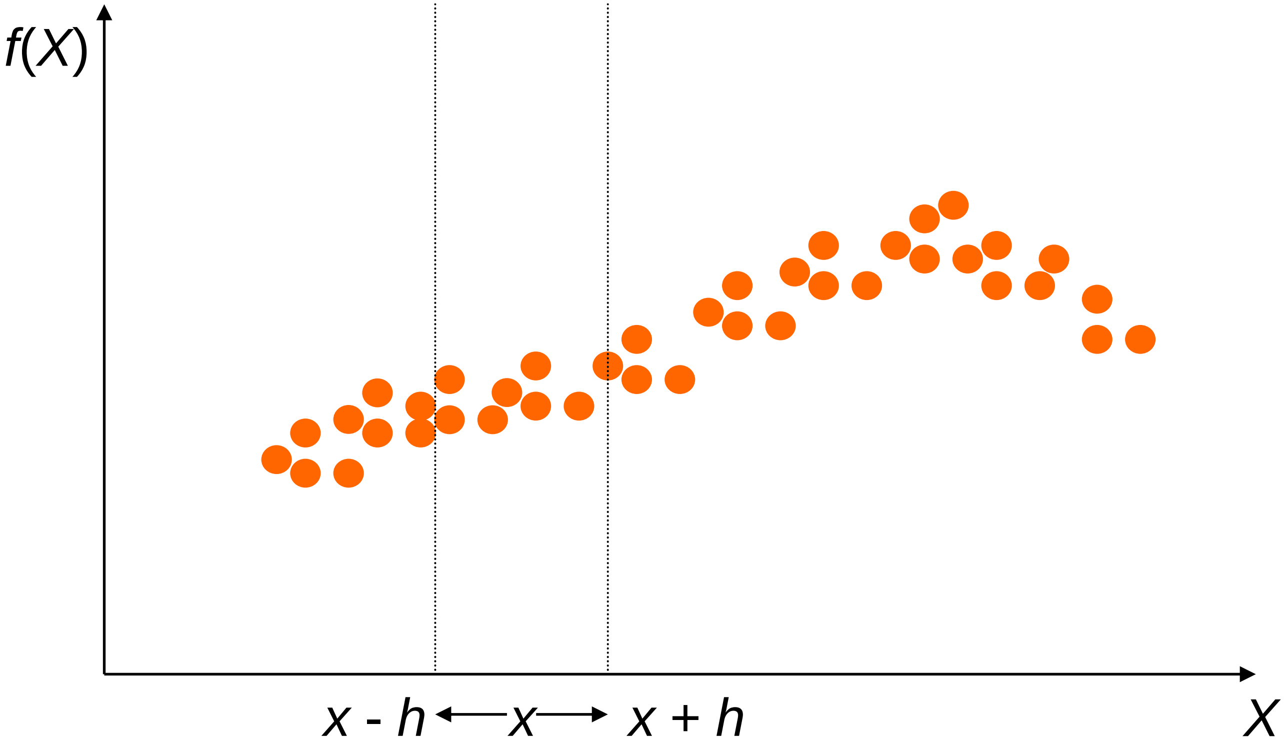
- Using the least-squares residuals, we estimate the $\rho(j)$ by the sample autocorrelations
- But in a limited sample of length T we cannot estimate all autocorrelations well
- We therefore use the (weighted) HAC estimator due to Newey and West (1987):

$$\hat{f} = 1 + 2\sum_{j=1}^{m}\left(\frac{m-j}{m}\right)\hat{\rho}(j)$$

where m is some **truncation parameter**

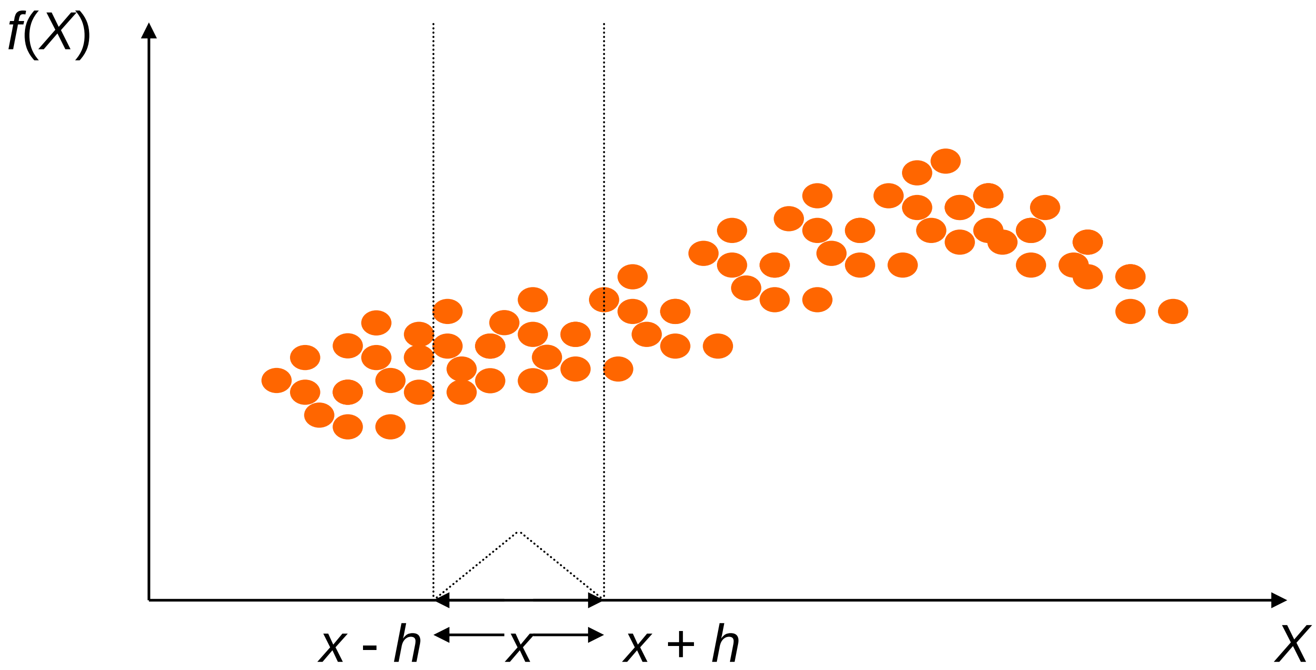
- This weighted estimator is always positive and changes smoothly in the truncation parameter, m
- More sophisticated data-dependent methods to pick m have also been developed

HAC estimators: the bandwidth



- We only use data points in the region x - h to x + h, where h is the **bandwidth** of the estimator

HAC estimators: the kernel



- The weight attached to each point in the interval need not necessarily be the same – this is controlled by the **kernel** function: for example, the triangular kernel puts less weight on points further from *x*

The overidentifying test (1)

- If the model is **just-identified**, there is one parameter for each restriction so the restrictions are jointly satisfied by construction, such that $J(\theta)$ in (32) is zero, the choice of weighting matrix W does not matter and only one step is needed
- If the model is **overidentified** (the number of moment restrictions, q, is greater than the number of parameters, k), it will not be possible to set every moment condition to zero, meaning that $J(\theta)$ in (32) will be greater than zero
- The question is, how far from zero is it?

The overidentifying test (2)

- Under the null hypothesis that the required number of k restrictions are valid, the test of the $(q - k)$ overidentifying restrictions (also known as the J test) is distributed as:

$$TJ(\theta) \sim \chi^2(q - k) \qquad (33)$$

  where T is the number of observations and $J(\theta)$ is the minimised value of the objective function (32)
- When overidentified models are estimated by GMM, it is customary to report the J statistic (33) as a general test of model adequacy and/or misspecification
- If $TJ(\theta)$ exceeds the $\chi^2(q - k)$ critical value, we reject the model, but it is not clear, from this information alone, what is wrong with the model

The overidentifying test (3)

- The test of overidentifying restrictions is often misunderstood
- It is **not** a test for whether all the instruments are 'valid'
- Instead the test answers the following question: given that a subset k of the q instruments is valid and exactly identifies the coefficients, are the $(q - k)$ 'extra' instruments valid?
- In other words, are the overidentifying restrictions (also) satisfied?

The overidentifying test (4)

- The small sample evidence for the reliability of overidentifying tests is mixed:
  - results depend on *W*, on whether a fully iterative or a two-step GMM method is used, on whether instruments carry ‘good’ information, on how many instruments there are, etc.
- It is useful to experiment with various alternatives – in particular, with various estimates of *W* and of instruments – before deriving conclusions about parameter estimates and the quality of the model

Choice of instruments

- In many applications, there is a large (possibly infinite) set of candidate instruments and hence a large class of consistent, asymptotically normal GMM estimators
- This leads to both theoretical and practical questions about instrument choice
- In practice, results can be sensitive to the instrument set…
- …and the choice of instruments involves judgement and generally cannot be reduced to a mechanical rule

The efficiency bound

- In general, the more instruments are used, the more precise the GMM estimator will be
- Despite this, the asymptotic covariance matrix of the GMM estimator which can be constructed from instruments contained in the information set is bounded from below by Hansen's (1985) **efficiency bound**

Hypothesis testing

- Estimates of the parameters and of the standard errors are biased in small samples
- In particular, estimates of the standard errors are, in general, downward biased (in other words, *t*-statistics have long and fat tails) – the test rejects too often
- Tests of hypotheses when *T* is small should therefore be undertaken with caution

The small-sample performance of GMM

- GMM's lack of structure means that it is widely applicable, although this generality comes at the cost of a number of issues, the most important of which is questionable small-sample performance
- Experimental evidence on simple specifications suggests that, with $T = 300$, GMM estimators obtained with good W estimates and good instruments approximate the true values in distribution
- …but $T = 300$ is a large number…
- …and convergence is slow!

Summary (1)

- Many commonly used estimators in econometrics, including OLS and instrumental variables, are derived most naturally using the method of moments
- GMM is essential to sophisticated macroeconometrics applications, including (nonlinear) rational expectations models and dynamic unobserved effects panel data models
- Many dynamic optimising models imply conditional moment restrictions (Euler equations) that can be used to construct unconditional moment equations (new Keynesian Phillips curves, forward-looking Taylor rules, etc.)
- Many forecasting equations imply orthogonality relations that can be used for moment construction

Summary (2)

- Complicated dynamics can give rise to intractable likelihoods but feasible moment relationships
- The GMM framework provides a computationally convenient method of performing inference without the need to specify the likelihood function
- Models that are linear in the variables but subject to nonlinear restrictions on the parameters are naturally estimated by GMM

References and further reading (1)

Andrews, D W K (1991), ‘Heteroskedasticity and autocorrelation consistent covariance matrix estimation’, *Econometrica*, Vol. 59, No. 3, pages 817-58.

Andrews, D W K (1993), ‘Tests for parameter instability and structural change with unknown change point’, *Econometrica*, Vol. 61, No. 4, pages 821-56.

Andrews, D W K and Fair, R (1988), ‘Inference in econometric models with a structural change’, *Review of Economic Studies*, Vol. 55, No. 4, pages 615-40.

Cochrane, J H (2001), *Asset pricing*, Princeton, Princeton University Press.

References and further reading (2)

Hall, A R (1993), 'Some aspects of generalised methods of moments estimation', Chapter 15, *Handbook of Statistics*, Vol. 11, Amsterdam, Elsevier.

Hall, A R (2005), *Generalised method of moments*, Oxford, Oxford University Press.

Hall, A R and Sen, A (1999), 'Structural stability testing in models estimated by generalised methods of moments', *Journal of Business & Economic Statistics*, Vol. 17, No. 3, pages 335-48.

Hansen, L P (1982), 'Large sample properties of generalised method of moments estimators', *Econometrica*, Vol. 50, No. 4, pages 1029-54.

References and further reading (3)

Hansen, L P (1985), ‘A method for calculating bounds on the asymptotic covariance matrices of generalized-method of moments estimators’, *Journal of Econometrics*, Vol. 30, No. 1-2, pages 203-38.

Hansen, B E and West, K D (2002), ‘Generalised method of moments and macroeconomics’, *Journal of Business & Economic Statistics*, Vol. 20, No. 4, pages 460-9.

Ogaki, M (1993), ‘Generalised method of moments: econometric applications’, Chapter 15, *Handbook of Statistics*, Vol. 11, Amsterdam, Elsevier.

Sen, A and Hall, A R (1999), ‘Two further aspects of some new tests for structural stability’, *Structural Change and Economic Dynamics*, Vol. 10, No. 3-4, pages 431-43.

References and further reading (4)

White, H (1982), 'Instrumental variables regression with independent observations', *Econometrica*, Vol. 50, No. 2, pages 483-99.

Wooldridge, J M (2001), 'Applications of generalised method of moments estimation', *Journal of Economic Perspectives*, Vol. 15, No. 4, pages 87-100.