Lecture 10: Generalised Method of Moments

Econometric Methods

Andrzej Torój

SGH Warsaw School of Economics – Institute of Econometrics

Outline

Outline

Instrumental variables: general idea

- OLS estimaton based on the general underlying assumption that $E\left(\mathbf{X}_{T\times k}^{T}\varepsilon_{T\times 1}\right)=\mathbf{0}_{k\times 1}$ (by Gauss-Markov).
- It may be broken i.a. for the following reasons:
    - just by construction of the economic model;
    - two-way causality between $\boldsymbol{y}_t$ and a subset of $\boldsymbol{x}_t$;
    - non-random sample selection.
- Solution: find variables that are truly orthogonal to $\varepsilon_{T\times 1}$ ("instrumental variables").

Instrumental variables: general idea

- OLS estimaton based on the general underlying assumption that $E\left(\mathbf{X}_{T\times k}^{T}\varepsilon_{T\times 1}\right)=\mathbf{0}_{k\times 1}$ (by Gauss-Markov).
- It may be broken i.a. for the following reasons:
  - just by construction of the economic model;
  - two-way causality between $\boldsymbol{y}_t$ and a subset of $\boldsymbol{x}_t$;
  - non-random sample selection.
- Solution: find variables that are truly orthogonal to $\varepsilon_{T\times 1}$ ("instrumental variables").

Instrumental variables: general idea

- OLS estimaton based on the general underlying assumption that $E\left(\mathbf{X}_{T\times k}^{T}\varepsilon_{T\times 1}\right)=\mathbf{0}_{k\times 1}$ (by Gauss-Markov).
- It may be broken i.a. for the following reasons:
    - just by construction of the economic model;
    - two-way causality between $\boldsymbol{y}_t$ and a subset of $\boldsymbol{x}_t$;
    - non-random sample selection.
- Solution: find variables that are truly orthogonal to $\varepsilon_{T\times 1}$ ("instrumental variables").

OLS and IV: alternative approach

- OLS: residuals uncorrelated with regressors, $E\left(\mathbf{X}^T_{T\times k}\varepsilon_{T\times 1}\right)=\mathbf{0}_{k\times 1}$
  - these are k "moment conditions" from which we infer the estimates
  - $\mathbf{X}^T\mathbf{y}=\mathbf{X}^T\mathbf{X}\beta+\underbrace{\mathbf{X}^T\varepsilon}_{\mathbf{0}}\Rightarrow\hat{\beta}^{OLS}=\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y}$
- IV: residuals uncorrelated with instruments, $E\left(\mathbf{Z}^T_{T\times k}\varepsilon_{T\times 1}\right)=\mathbf{0}_{k\times 1}$
  - these are k "moment conditions" from which we infer the estimates
  - $\mathbf{Z}^T\mathbf{y}=\mathbf{Z}^T\mathbf{X}\beta+\underbrace{\mathbf{Z}^T\varepsilon}_{\mathbf{0}}\Rightarrow\hat{\beta}^{OLS}=\left(\mathbf{Z}^T\mathbf{X}\right)^{-1}\mathbf{Z}^T\mathbf{y}$
- what if we get more than $l\,(>k)$ "moment conditions", i.e. more than we actually need? $\mathbf{Z}^T_{T\times l}\varepsilon_{T\times 1}=\mathbf{0}_{l\times 1}$
  - there are $l\,(>k)$ "moment conditions", so not all of them can be fulfilled by modifying β (only k parameters)
  - $\mathbf{Z}^T\mathbf{y}=\mathbf{Z}^T\mathbf{X}\beta+\underbrace{\mathbf{Z}^T\varepsilon}_{\mathbf{0}}\Rightarrow$ but we won't invert $\mathbf{Z}^T\mathbf{X}$ (not a square matrix this time!)

OLS and IV: alternative approach

- OLS: residuals uncorrelated with regressors, $E\left(\mathbf{X}^T_{T\times k}\varepsilon_{T\times 1}\right)=\mathbf{0}_{k\times 1}$
  - these are k "moment conditions" from which we infer the estimates
  - $\mathbf{X}^T\mathbf{y}=\mathbf{X}^T\mathbf{X}\beta+\underbrace{\mathbf{X}^T\varepsilon}_{\mathbf{0}}\Rightarrow\hat{\beta}^{OLS}=\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y}$
- IV: residuals uncorrelated with instruments, $E\left(\mathbf{Z}^T_{T\times k}\varepsilon_{T\times 1}\right)=\mathbf{0}_{k\times 1}$
  - these are k "moment conditions" from which we infer the estimates
  - $\mathbf{Z}^T\mathbf{y}=\mathbf{Z}^T\mathbf{X}\beta+\underbrace{\mathbf{Z}^T\varepsilon}_{\mathbf{0}}\Rightarrow\hat{\beta}^{OLS}=\left(\mathbf{Z}^T\mathbf{X}\right)^{-1}\mathbf{Z}^T\mathbf{y}$
- what if we get more than $l\,(>k)$ "moment conditions", i.e. more than we actually need? $\mathbf{Z}^T_{T\times l}\varepsilon_{T\times 1}=\mathbf{0}_{l\times 1}$
  - there are $l\,(>k)$ "moment conditions", so not all of them can be fulfilled by modifying β (only k parameters)
  - $\mathbf{Z}^T\mathbf{y}=\mathbf{Z}^T\mathbf{X}\beta+\underbrace{\mathbf{Z}^T\varepsilon}_{\mathbf{0}}\Rightarrow$ but we won't invert $\mathbf{Z}^T\mathbf{X}$ (not a square matrix this time!)

OLS and IV: alternative approach

- OLS: residuals uncorrelated with regressors, $E\left(\mathbf{X}^T_{T\times k}\varepsilon_{T\times 1}\right) = \mathbf{0}_{k\times 1}$
  - these are k "moment conditions" from which we infer the estimates
  - $\mathbf{X}^T\mathbf{y} = \mathbf{X}^T\mathbf{X}\beta + \underbrace{\mathbf{X}^T\varepsilon}_{\mathbf{0}} \Rightarrow \hat{\beta}^{OLS} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y}$
- IV: residuals uncorrelated with instruments, $E\left(\mathbf{Z}^T_{T\times k}\varepsilon_{T\times 1}\right) = \mathbf{0}_{k\times 1}$
  - these are k "moment conditions" from which we infer the estimates
  - $\mathbf{Z}^T\mathbf{y} = \mathbf{Z}^T\mathbf{X}\beta + \underbrace{\mathbf{Z}^T\varepsilon}_{\mathbf{0}} \Rightarrow \hat{\beta}^{OLS} = \left(\mathbf{Z}^T\mathbf{X}\right)^{-1}\mathbf{Z}^T\mathbf{y}$
- what if we get more than $l\,(> k)$ "moment conditions", i.e. more than we actually need? $\mathbf{Z}^T_{T\times l}\varepsilon_{T\times 1} = \mathbf{0}_{l\times 1}$
  - there are $l\,(> k)$ "moment conditions", so not all of them can be fulfilled by modifying β (only k parameters)
  - $\mathbf{Z}^T\mathbf{y} = \mathbf{Z}^T\mathbf{X}\beta + \underbrace{\mathbf{Z}^T\varepsilon}_{\mathbf{0}} \Rightarrow$ but we won't invert $\mathbf{Z}^T\mathbf{X}$ (not a square matrix this time!)

Limitations of IV estimation

- $\beta^{IV} = \left(\mathbf{Z}^T\mathbf{X}\right)^{-1}\mathbf{Z}^T\mathbf{y}$
  - $\mathbf{Z}$ must have the same number of columns as $\mathbf{X}$ for this operation to be feasible
- in IV, there must be as many instrumental variables as regressors (some of which can instrumentalise themselves)
- what are $\mathbf{Z}$?
  - uncorrelated with ε, correlated with $\mathbf{X}$
  - no reason to assume that there is a limited number of such variables

Limitations of IV estimation

- $\beta^{IV} = \left(\mathbf{Z}^T\mathbf{X}\right)^{-1}\mathbf{Z}^T\mathbf{y}$
  - $\mathbf{Z}$ must have the same number of columns as $\mathbf{X}$ for this operation to be feasible
- in IV, there must be as many instrumental variables as regressors (some of which can instrumentalise themselves)
- what are $\mathbf{Z}$?
  - uncorrelated with ε, correlated with $\mathbf{X}$
  - no reason to assume that there is a limited number of such variables

Limitations of IV estimation

- $\beta^{IV} = \left(\mathbf{Z}^T\mathbf{X}\right)^{-1}\mathbf{Z}^T\mathbf{y}$
  - $\mathbf{Z}$ must have the same number of columns as $\mathbf{X}$ for this operation to be feasible
- in IV, there must be as many instrumental variables as regressors (some of which can instrumentalise themselves)
- what are $\mathbf{Z}$?
  - uncorrelated with ε, correlated with $\mathbf{X}$
  - no reason to assume that there is a limited number of such variables

Outline

Generalized Method of Moments

So let's take yet another perspective...

- OLS minimises the quadratic form $\left[\mathbf{X}^T\varepsilon(\beta)\right]^T\left[\mathbf{X}^T\varepsilon(\beta)\right]$
  - wrt. $\beta_{k\times 1}$ (down to zero!)
- IV minimises the quadratic form $\left[\mathbf{Z}^T\varepsilon(\beta)\right]^T\left[\mathbf{Z}^T\varepsilon(\beta)\right]$
  - wrt. $\beta_{k\times 1}$ with $\mathbf{Z}_{k\times T}$ (down to zero!)
- IV cannot minimise the quadratic form $\left(\mathbf{Z}^T\varepsilon\right)^T\left(\mathbf{Z}^T\varepsilon\right)$
  - wrt. $\beta_{k\times 1}$ with $\mathbf{Z}_{l\times T}$ down to zero, i.e. it is impossible to perfectly fulfil all the moment conditions
  - cannot solve l equations (moment conditions) for $k < l$ unknowns
- ...so take quadratic form $\left(\mathbf{Z}^T\varepsilon\right)^T \mathbf{W} \left(\mathbf{Z}^T\varepsilon\right)$ instead!: $\mathbf{W}_{l\times l}$
  - weigh the squared differences between left-hand and right-hand side and minimise the sum!

Generalized Method of Moments

So let's take yet another perspective...

- OLS minimises the quadratic form $\left[\mathbf{X}^T\varepsilon(\beta)\right]^T\left[\mathbf{X}^T\varepsilon(\beta)\right]$
  - wrt. $\beta_{k\times 1}$ (down to zero!)
- IV minimises the quadratic form $\left[\mathbf{Z}^T\varepsilon(\beta)\right]^T\left[\mathbf{Z}^T\varepsilon(\beta)\right]$
  - wrt. $\beta_{k\times 1}$ with $\mathbf{Z}_{k\times T}$ (down to zero!)
- IV cannot minimise the quadratic form $\left(\mathbf{Z}^T\varepsilon\right)^T\left(\mathbf{Z}^T\varepsilon\right)$
  - wrt. $\beta_{k\times 1}$ with $\mathbf{Z}_{l\times T}$ down to zero, i.e. it is impossible to perfectly fulfil all the moment conditions
  - cannot solve l equations (moment conditions) for $k<l$ unknowns
- ...so take quadratic form $\left(\mathbf{Z}^T\varepsilon\right)^T\mathbf{W}\left(\mathbf{Z}^T\varepsilon\right)$ instead!: $\mathbf{W}_{l\times l}$
  - weigh the squared differences between left-hand and right-hand side and minimise the sum!

Generalized Method of Moments

So let's take yet another perspective...

- OLS minimises the quadratic form $\left[\mathbf{X}^T\varepsilon(\beta)\right]^T\left[\mathbf{X}^T\varepsilon(\beta)\right]$
  - wrt. $\beta_{k\times 1}$ (down to zero!)
- IV minimises the quadratic form $\left[\mathbf{Z}^T\varepsilon(\beta)\right]^T\left[\mathbf{Z}^T\varepsilon(\beta)\right]$
  - wrt. $\beta_{k\times 1}$ with $\mathbf{Z}_{k\times T}$ (down to zero!)
- IV cannot minimise the quadratic form $\left(\mathbf{Z}^T\varepsilon\right)^T\left(\mathbf{Z}^T\varepsilon\right)$
  - wrt. $\beta_{k\times 1}$ with $\mathbf{Z}_{l\times T}$ down to zero, i.e. it is impossible to perfectly fulfil all the moment conditions
  - cannot solve l equations (moment conditions) for $k < l$ unknowns
- ...so take quadratic form $\left(\mathbf{Z}^T\varepsilon\right)^T\mathbf{W}\left(\mathbf{Z}^T\varepsilon\right)$ instead!: $\mathbf{W}_{l\times l}$
  - weigh the squared differences between left-hand and right-hand side and minimise the sum!

Generalized Method of Moments

So let's take yet another perspective...

- OLS minimises the quadratic form $\left[\mathbf{X}^T\varepsilon(\beta)\right]^T\left[\mathbf{X}^T\varepsilon(\beta)\right]$
  - wrt. $\beta_{k\times1}$ (down to zero!)
- IV minimises the quadratic form $\left[\mathbf{Z}^T\varepsilon(\beta)\right]^T\left[\mathbf{Z}^T\varepsilon(\beta)\right]$
  - wrt. $\beta_{k\times1}$ with $\mathbf{Z}_{k\times T}$ (down to zero!)
- IV cannot minimise the quadratic form $\left(\mathbf{Z}^T\varepsilon\right)^T\left(\mathbf{Z}^T\varepsilon\right)$
  - wrt. $\beta_{k\times1}$ with $\mathbf{Z}_{l\times T}$ down to zero, i.e. it is impossible to perfectly fulfil all the moment conditions
  - cannot solve l equations (moment conditions) for $k<l$ unknowns
- ...so take quadratic form $\left(\mathbf{Z}^T\varepsilon\right)^T\mathbf{W}\left(\mathbf{Z}^T\varepsilon\right)$ instead!: $\mathbf{W}_{l\times l}$
  - weigh the squared differences between left-hand and right-hand side and minimise the sum!

How to obtain $\mathbf{W}$?

- in general: any symmetric, positive-definite matrix sized $l \times l$
- according to ***Hansen (1982)***, you should assign the lowest weights to the moment conditions with lowest precision of estimation, which implies
$$\mathbf{V} = \frac{1}{T}\mathbf{Z}^T\mathbf{\Omega}\mathbf{Z}$$
- with $\mathbf{\Omega}$ – variance-covariance matrix of the estimator
  - under white noise: $\hat{\mathbf{\Omega}} = \hat{\sigma}^2\boldsymbol{I}$
  - under non-spherical disturbances – White or Newey-West versions (see: serial correlation / heteroskedasticity lectures)

GMM as iterative procedure

- $\hat{\varepsilon}$ unknown in advance!
  1. estimate under the assumption of white noise
  2. calculate the residuals
  3. update $\hat{\mathbf{W}}$
- various implementations of this algorithm in the software (i.a. as a two-step or iterative estimator – cf. GLS)

Derivation of GMM estimator (1)

Let us minimise the quadratic form with respect to $\boldsymbol{\beta}$:

$$\begin{aligned}\varepsilon^T \mathbf{ZWZ}^T \varepsilon &= (\mathbf{y}-\boldsymbol{\beta}\mathbf{X})^T \mathbf{ZWZ}^T (\mathbf{y}-\boldsymbol{\beta}\mathbf{X}) = \\ &= \mathbf{y}^T\mathbf{ZWZ}^T\mathbf{y} - \mathbf{y}^T\mathbf{ZWZ}^T\mathbf{X}\boldsymbol{\beta} \\ &\quad -\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{ZWZ}^T\mathbf{y} + \boldsymbol{\beta}^T\mathbf{X}^T\mathbf{ZWZ}^T\mathbf{X}\boldsymbol{\beta} \\ &\rightarrow \mathit{min}\end{aligned}$$

From matrix calculus:

$$\frac{\partial \mathbf{x}^T\mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = \mathbf{x}^T\left(\mathbf{A}^T + \mathbf{A}\right)$$

$$\frac{\partial \mathbf{x}^T\mathbf{a}}{\partial \mathbf{x}} = \mathbf{a}^T$$

Compute the derivative and set it equal to 0.

Derivation of GMM estimator (2)

- $-y^T ZWZ^T X - y^T ZWZ^T X + \beta^T \left(X^T ZWZ^T X + X^T ZWZ^T X \right) = 0$
- $\beta^T \left(X^T ZWZ^T X \right) = y^T ZWZ^T X$
- $\left(X^T ZWZ^T X \right) \beta = X^T ZWZ^T y$
- $\hat{\beta}^{GMM} = \left(X^T ZWZ^T X \right)^{-1} X^T ZWZ^T y$

Nonlinear GMM:

– you can also (numerically) minimise the quadratic form for a nonlinear model;
– it is advisable to write the moment conditions "as close to linear as possible".

Derivation of GMM estimator (2)

- $-\mathbf{y}^T\mathbf{ZWZ}^T\mathbf{X} - \mathbf{y}^T\mathbf{ZWZ}^T\mathbf{X} + \beta^T\left(\mathbf{X}^T\mathbf{ZWZ}^T\mathbf{X} + \mathbf{X}^T\mathbf{ZWZ}^T\mathbf{X}\right) = 0$
- $\beta^T\left(\mathbf{X}^T\mathbf{ZWZ}^T\mathbf{X}\right) = \mathbf{y}^T\mathbf{ZWZ}^T\mathbf{X}$
- $\left(\mathbf{X}^T\mathbf{ZWZ}^T\mathbf{X}\right)\beta = \mathbf{X}^T\mathbf{ZWZ}^T\mathbf{y}$
- $\hat{\beta}^{GMM} = \left(\mathbf{X}^T\mathbf{ZWZ}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{ZWZ}^T\mathbf{y}$

Nonlinear GMM:

– you can also (numerically) minimise the quadratic form for a nonlinear model;
– it is advisable to write the moment conditions "as close to linear as possible".

Derivation of GMM estimator (2)

- $-\mathbf{y}^T\mathbf{Z}\mathbf{W}\mathbf{Z}^T\mathbf{X} - \mathbf{y}^T\mathbf{Z}\mathbf{W}\mathbf{Z}^T\mathbf{X} + \boldsymbol{\beta}^T\left(\mathbf{X}^T\mathbf{Z}\mathbf{W}\mathbf{Z}^T\mathbf{X} + \mathbf{X}^T\mathbf{Z}\mathbf{W}\mathbf{Z}^T\mathbf{X}\right) = 0$
- $\boldsymbol{\beta}^T\left(\mathbf{X}^T\mathbf{Z}\mathbf{W}\mathbf{Z}^T\mathbf{X}\right) = \mathbf{y}^T\mathbf{Z}\mathbf{W}\mathbf{Z}^T\mathbf{X}$
- $\left(\mathbf{X}^T\mathbf{Z}\mathbf{W}\mathbf{Z}^T\mathbf{X}\right)\boldsymbol{\beta} = \mathbf{X}^T\mathbf{Z}\mathbf{W}\mathbf{Z}^T\mathbf{y}$
- $\hat{\boldsymbol{\beta}}^{GMM} = \left(\mathbf{X}^T\mathbf{Z}\mathbf{W}\mathbf{Z}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{Z}\mathbf{W}\mathbf{Z}^T\mathbf{y}$

Nonlinear GMM:

– you can also (numerically) minimise the quadratic form for a nonlinear model;
– it is advisable to write the moment conditions “as close to linear as possible”.

Derivation of GMM estimator (2)

- $-\mathbf{y}^T\mathbf{Z}\mathbf{W}\mathbf{Z}^T\mathbf{X} - \mathbf{y}^T\mathbf{Z}\mathbf{W}\mathbf{Z}^T\mathbf{X} + \beta^T\left(\mathbf{X}^T\mathbf{Z}\mathbf{W}\mathbf{Z}^T\mathbf{X} + \mathbf{X}^T\mathbf{Z}\mathbf{W}\mathbf{Z}^T\mathbf{X}\right) = 0$
- $\beta^T\left(\mathbf{X}^T\mathbf{Z}\mathbf{W}\mathbf{Z}^T\mathbf{X}\right) = \mathbf{y}^T\mathbf{Z}\mathbf{W}\mathbf{Z}^T\mathbf{X}$
- $\left(\mathbf{X}^T\mathbf{Z}\mathbf{W}\mathbf{Z}^T\mathbf{X}\right)\beta = \mathbf{X}^T\mathbf{Z}\mathbf{W}\mathbf{Z}^T\mathbf{y}$
- $\hat{\beta}^{GMM} = \left(\mathbf{X}^T\mathbf{Z}\mathbf{W}\mathbf{Z}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{Z}\mathbf{W}\mathbf{Z}^T\mathbf{y}$

Nonlinear GMM:

– you can also (numerically) minimise the quadratic form for a nonlinear model;
– it is advisable to write the moment conditions "as close to linear as possible".

Derivation of GMM estimator (2)

- $-\mathbf{y}^T\mathbf{Z}\mathbf{W}\mathbf{Z}^T\mathbf{X} - \mathbf{y}^T\mathbf{Z}\mathbf{W}\mathbf{Z}^T\mathbf{X} +$
  $\beta^T\left(\mathbf{X}^T\mathbf{Z}\mathbf{W}\mathbf{Z}^T\mathbf{X} + \mathbf{X}^T\mathbf{Z}\mathbf{W}\mathbf{Z}^T\mathbf{X}\right) = 0$
- $\beta^T\left(\mathbf{X}^T\mathbf{Z}\mathbf{W}\mathbf{Z}^T\mathbf{X}\right) = \mathbf{y}^T\mathbf{Z}\mathbf{W}\mathbf{Z}^T\mathbf{X}$
- $\left(\mathbf{X}^T\mathbf{Z}\mathbf{W}\mathbf{Z}^T\mathbf{X}\right)\beta = \mathbf{X}^T\mathbf{Z}\mathbf{W}\mathbf{Z}^T\mathbf{y}$
- $\hat{\beta}^{GMM} = \left(\mathbf{X}^T\mathbf{Z}\mathbf{W}\mathbf{Z}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{Z}\mathbf{W}\mathbf{Z}^T\mathbf{y}$

Nonlinear GMM:

– you can also (numerically) minimise the quadratic form for a nonlinear model;
– it is advisable to write the moment conditions "as close to linear as possible".

GMM: special cases

1. $Z_{T\times k}$, $X_{T\times k}$
   $\hat{\beta}^{GMM} = (X^T ZWZ^T X)^{-1} X^T ZWZ^T y =$
   $(Z^T X)^{-1} W^{-1} (X^T Z)^{-1} X^T ZWZ^T y = (Z^T X)^{-1} Z^T y = \hat{\beta}^{IV}$
2. $Z = X$
   $\hat{\beta}^{GMM} = (X^T XWX^T X)^{-1} X^T XWX^T y =$
   $(X^T X)^{-1} W^{-1} (X^T X)^{-1} X^T XWX^T y = (X^T X)^{-1} X^T y =$
   $\hat{\beta}^{OLS}$

GMM: special cases

1. $\mathbf{Z}_{T\times k}$, $\mathbf{X}_{T\times k}$
$$\hat{\beta}^{GMM} = (\mathbf{X}^T\mathbf{Z}\mathbf{W}\mathbf{Z}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Z}\mathbf{W}\mathbf{Z}^T\mathbf{y} = (\mathbf{Z}^T\mathbf{X})^{-1}\mathbf{W}^{-1}(\mathbf{X}^T\mathbf{Z})^{-1}\mathbf{X}^T\mathbf{Z}\mathbf{W}\mathbf{Z}^T\mathbf{y} = (\mathbf{Z}^T\mathbf{X})^{-1}\mathbf{Z}^T\mathbf{y} = \hat{\beta}^{IV}$$
2. $\mathbf{Z} = \mathbf{X}$
$$\hat{\beta}^{GMM} = (\mathbf{X}^T\mathbf{X}\mathbf{W}\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}\mathbf{W}\mathbf{X}^T\mathbf{y} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{W}^{-1}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}\mathbf{W}\mathbf{X}^T\mathbf{y} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} = \hat{\beta}^{OLS}$$

Variance-covariance of GMM estimator

- GMM estimator: consistent, asymptotically normally distributed
- asymptotic variance-covariance estimator in a linear model:

$$Var\left(\beta^{GMM}\right) = \frac{1}{T}\left[\frac{1}{T}\left(\mathbf{X}^T\mathbf{Z}\right)\mathbf{W}^{-1}\frac{1}{T}\left(\mathbf{Z}^T\mathbf{X}\right)\right]^{-1}$$

J statistic

- the non-zero value of the minimised quadratic form is interpretable
  - how far are we from fulfilling the (excessive) orthogonality conditions?
  - the lower the value, the lower the "distance" to perfect fulfilment of excessive conditions
- divided by T, it is χ^2-distributed with degrees of freedom equal to $l - k$ (instruments in excess of parameters)

J-test of orthogonality

$$J\left(\hat{\beta}, \hat{\mathbf{\Omega}}^{-\mathbf{1}}\right) = \frac{1}{T}\varepsilon\left(\hat{\beta}\right)^T \mathbf{Z}\hat{\mathbf{\Omega}}^{-\mathbf{1}}\mathbf{Z}^T\varepsilon\left(\hat{\beta}\right)$$

H_0 : all orthogonality conditions fulfilled

H_1 : some orthogonality conditions not fulfilled

Outline

Hybrid Phillips curve

- *Gali, Gertler (1999)*:
  $$\pi_t = \gamma_b \pi_{t-1} + \gamma_f E_t \pi_{t+1} + \lambda mc_t + \varepsilon_t$$
  where π_t – inflation rate, mc_t – real marginal cost, ε_t – error term.
- Why GMM?
  - there is an unobservable variable on the right-hand side, $E_t \pi_{t+1}$
  - we can just replace it with observable $\pi_{t+1} = E_t \pi_{t+1} + v_{t+1}$, where v_t – expectations error
    $$\pi_t = \gamma_b \pi_{t-1} + \gamma_f \pi_{t+1} (v_{t+1}) + \lambda mc_t + \varepsilon_t$$
  - but v_{t+1} clearly not independent from ε_t – inconsistency!

Hybrid Phillips curve

- *Gali, Gertler (1999)*:
  $$\pi_t = \gamma_b \pi_{t-1} + \gamma_f E_t \pi_{t+1} + \lambda mc_t + \varepsilon_t$$
  where π_t – inflation rate, mc_t – real marginal cost, ε_t – error term.
- Why GMM?
  - there is an unobservable variable on the right-hand side, $E_t \pi_{t+1}$
  - we can just replace it with observable $\pi_{t+1} = E_t \pi_{t+1} + v_{t+1}$, where v_t – expectations error
    $$\pi_t = \gamma_b \pi_{t-1} + \gamma_f \pi_{t+1} \left(v_{t+1}\right) + \lambda mc_t + \varepsilon_t$$
  - but v_{t+1} clearly not independent from ε_t – inconsistency!

Orthogonality conditions

- $E_t\left[\varepsilon_t \mathbf{z}_t\right] = E_t\left[\left(\pi_t - \gamma_b \pi_{t-1} - \gamma_f E_t \pi_{t+1} - \lambda mc_t\right)\mathbf{z}_t\right] = E_t\left[\left(\pi_t - \gamma_b \pi_{t-1} - \gamma_f \pi_{t+1} - \lambda mc_t\right)\mathbf{z}_t\right] = \mathbf{0}$
    - the expected value on the orthogonality condition allows to drop the expected value on π_{t+1}
- in this context, we interpret the instrument set $\mathbf{Z}$ as **variables that allow to forecast inflation one period ahead without systematic errors**
- there should be more than 3 instruments to use GMM here

Example

(Imperfect) replication of Gali-Gertler results:

- Read the paper by Gali and Gertler.
- Consider the following set of variables as linear one-period-ahead predictors:
  - *inflation (4 lags), log real ULC (1 lag), output gap (1 lag), short- vs long-term interest rate spread (1 lag), log-differences of wage index (1 lag), log-differences of commodity price index (1 lag)*
- Define the instruments and the initial weight matrix $\mathbf{W}$.
- Estimate the model. Discuss the results. Do they confirm Your intuition? If not, look for the (economic) solution in the article.