Advanced Quantitative Methods: Generalized Methods of Moments

Ingmar Nolte

Room: C48

I.Nolte@lancaster.ac.uk

Office Hours: whenever you like

Webpage:

http://www.lancaster.ac.uk/staff/nolte/

Password: 42GARCH

Overview:

- General Setup
- Estimation
- Large Sample Properties
- Which $\hat{W}$? – Efficient Estimation
- Multiple Equation GMM

Motivation:

- Endogeneity Bias (e.g. Demand & Supply example, Measurement Errors, Hayashi (2000) Ch. 3.1).
- A regressor is **endogenous** if it is not **predetermined**, i.e. $\mathrm{E}[\varepsilon_t X_t] \neq 0$, i.e the orthogonality assumption does not hold.
- Implication: biased & inconsistent estimators (obvious from Asymptotic Theory Lecture).
- **Instrumental Variables**: A **predetermined** variable that is **correlated** with the endogenous regressor.

General Setup:

Assumptions:

i) The equation to be estimated is $Y_t = X_t'\beta + \varepsilon_t, \quad t = 1, \ldots, T, \quad \beta \in \mathbb{R}^k$ **(Linearity)**

ii) Let Z_t denote the ℓ dimensional vector of instruments and let W_t be the unique and nonconstant elements of $\{Y_t, X_t, Z_t\}$. W_t is a stationary ergodic process. **(Ergodic Stationarity)**

iii) All ℓ variables in Z_t are predetermined in the sense that they are orthogonal to the current error term, i.e. $\mathrm{E}[Z_t \varepsilon_t] = 0$ for all t. It is often written as $\mathrm{E}[Z_t(Y_t - X_t'\beta)] = 0$ or $\mathrm{E}[g_t] = 0$ with $g_t \equiv g(W_t, \beta) \equiv Z_t \varepsilon_t$. **(Orthogonality Conditions)**

iv) The $\ell \times k$ matrix $\mathrm{E}[Z_t X_t']$ is of full column rank (and the cross moments exist and are finite). We denote that matrix by Σ_{zx}. [**Remark:** We need that the instruments and the endogenous regressors are correlated].**(Rank Condition for Identification)**

v) Since $\mathrm{rk}(\Sigma_{zx}) < k$ if $\ell < k$ a necessary condition is that $\ell \geq k$, i.e. number of orthogonality conditions $\geq$ number of parameters. **(Order Condition for Identification)**

- **over-identification**: rank condition satisfied and $\ell > k$
- **exact identification**: rank condition satisfied and $\ell = k$
- **under-identification**: $\ell < k$

vi) g_t is a martingale difference sequence. The $\ell \times \ell$ matrix $\mathrm{E}[g_t g_t']$ is non-singular. **(Assumptions for Asymptotic Normality)**

Remark:

i) **(Rank Condition for Identification)**: In general:

The k dim. vector β is a solution to $\mathrm{E}[g_t] = \mathrm{E}[g(W_t, \tilde{\beta})] = 0_{\ell \times 1}$. **Identification**, however, refers to β being the **only** solution.

$$
\begin{aligned}
\mathrm{E}[g_t] &= 0 \\
\Sigma_{zx}\tilde{\beta} &= \sigma_{zy}, \quad \text{with } \sigma_{zy} \equiv \mathrm{E}[Z_t Y_t]
\end{aligned}
$$

A sufficient and necessary condition for a unique solution is that Σ_{zx} is a full column rank.

ii) **(Asymptotic Normality)**: We follow the outline in Hayashi (2000), Ch. 3.2., which corresponds to the Central Limit Theorem for Martingale Difference Sequences covered in the Asymptotic Theory Lecture. The other cases covered there follow trivially and similarly.

Estimation:

Principle: Choose the parameter estimate of β so that the sample moments corresponding to the population moments $\mathrm{E}[g_t]$ are zero.

Let

$$g_T(\tilde{\beta}) \equiv \frac{1}{T}\sum_{t=1}^{T} g_t(W_t, \tilde{\beta})$$

denote the sample mean of the orthogonality conditions (sample moments function) for an arbitrary estimator $\tilde{\beta}$.

We obtain:

$$\begin{aligned} g_T(\tilde{\beta}) &= \frac{1}{T}\sum_{t=1}^{T} Z_t(Y_t - X_t'\tilde{\beta}) \\ &= \frac{1}{T}\sum_{t=1}^{T} Z_t Y_t - \left(\frac{1}{T}\sum_{t=1}^{T} Z_t X_t'\right)\tilde{\beta} \end{aligned}$$

Denoting $s_{zy} \equiv \frac{1}{T}\sum_{t=1}^{T} Z_t Y_t$ and $S_{zx} \equiv \frac{1}{T}\sum_{t=1}^{T} Z_t X_t'$ and using $g_T(\tilde{\beta}) = 0$, we obtain

$$S_{zx}\tilde{\beta} = s_{zy}$$

which is a system of ℓ linear equations in k unknowns, derived from the sample moment conditions.

Case I: Method of Moments: $\ell = k$

The system is exactly identified and Σ_{zx} is square and invertible. Since S_{zx} converges to Σ_{zx} with probability one it follows that S_{zx} is invertible for T large enough with probability one and we obtain

$$\hat{\beta}_{IV} = S_{zx}^{-1} s_{zy} = \left(\frac{1}{T} \sum_{t=1}^{T} Z_t X_t' \right)^{-1} \frac{1}{T} \sum_{t=1}^{T} Z_t Y_t,$$

the **instrumental variable estimator** of β.

Case II: Generalized Method of Moments (GMM): $\ell > k$

The system is over-identified and we cannot choose $\tilde{\beta}$ so that $g_T(\tilde{\beta})$ is zero, but we can choose $\tilde{\beta}$ so that $g_T(\tilde{\beta})$ is as close to zero as possible.

Close: We understand close with respect to a quadratic form distance measure between two vectors a and b given by $(a-b)'\hat{W}(a-b)$, where $\hat{W}$ is a symmetric positive definite weighting matrix, defining the distance.

Definition: GMM estimator:

Let $\hat{W}$ be an $\ell \times \ell$ symmetric, positive definite matrix, that may depend on the sample size T and is such that $\hat{W} \xrightarrow{p} W$, where W is symmetric and positive definite. The GMM estimator of β is given by

$$\hat{\beta} \equiv \hat{\beta}(\hat{W}) \equiv \underset{\tilde{\beta}}{\operatorname{argmin}}\, J(\tilde{\beta}, \hat{W}),$$

with

$$J(\tilde{\beta}, \hat{W}) \equiv T g_T(\tilde{\beta})' \hat{W} g_T(\tilde{\beta}).$$

In the linear regression model case $g_T(\tilde{\beta})'$ is linear in $\tilde{\beta}$ and the objective function is quadratic in $\tilde{\beta}$, so that

$$J(\tilde{\beta}, \hat{W}) \equiv T(s_{zy} - S_{zx}\tilde{\beta})'\hat{W}(s_{zy} - S_{zx}\tilde{\beta}),$$

the first order conditions w.r.t. $\tilde{\beta}$ then become

$$S'_{zx}\hat{W}s_{zy} = S'_{zx}\hat{W}S_{zx}\tilde{\beta}.$$

Since S_{zx} is of full column rank (for T large enough a.s.) and $\hat{W}$ is positive definite, the $k \times k$ matrix $S'_{zx}\hat{W}S_{zx}$ is nonsingular and thus

$$\hat{\beta} = \left(S'_{zx}\hat{W}S_{zx}\right)^{-1} S'_{zx}\hat{W}s_{zy}.$$

If $\ell = k$ S_{zx} is a square matrix and we obtain the IV estimator.

Sampling Error

Multiplying $Y_t = X_t'\beta + \varepsilon_t$ from the left by Z_t and taking averages yields

$$s_{zy} = S_{zx}\tilde{\beta} + \bar{g}$$

where $\bar{g} = \frac{1}{T}\sum_{t=1}^{T} Z_t\varepsilon_t = \frac{1}{T}\sum_{t=1}^{T} g_t(W_t, \beta) = g_T(\beta)$. Thus, we get

$$\hat{\beta} - \beta = \left(S_{zx}'\hat{W}S_{zx}\right)^{-1} S_{zx}'\hat{W}\bar{g}.$$

Large-Sample Properties:

The GMM estimator is denoted by $\hat{\beta}_T = \hat{\beta} = \hat{\beta}(\hat{W})$ for any choice of $\hat{W}$ and the asymptotic theory stated below is valid for an arbitrary choice of $\hat{W}$.

Proposition: Asymptotic Behaviour

- Under assumptions i) to iv): $\hat{\beta}_T \xrightarrow{p} \beta$ **(Consistency)**.
- If in addition vi) holds: **(Asymptotic Normality)**

$$D_T^{-1/2}\sqrt{T}(\hat{\beta}_T - \beta) \overset{asy}{\sim} N(0, I),$$

with

$$D_T \equiv (\Sigma'_{zx} W \Sigma_{zx})^{-1} \Sigma'_{zx} W V W \Sigma_{zx} (\Sigma'_{zx} W \Sigma_{zx})^{-1},$$

where $\Sigma_{zx} = \mathrm{E}[Z_t X_t'], V = \mathrm{E}[g_t g_t'] = \mathrm{E}[\varepsilon_t^2 Z_t Z_t'], W = \operatorname{plim} \hat{W}$

- If in addition there exists a matrix $\hat{V}_T$ positive semi-definite and symmetric such that $\hat{V}_T - V \xrightarrow{p} 0$[a].

[a]For this to hold we need the assumption vii): $\mathrm{E}[(Z_{ti}X_{tj})^2] < \infty$ for all $i = 1, \dots, \ell$ and $j = 1, \dots, k$. **(Fourth Moments Condition)**

Then $\hat{D}_T - D \xrightarrow{p} 0$, with

$$\hat{D}_T \equiv (S'_{zx}\hat{W}S_{zx})^{-1}S'_{zx}\hat{W}\hat{V}_T\hat{W}S_{zx}(S'_{zx}\hat{W}S_{zx})^{-1}.$$

- With i), ii) and $\mathrm{E}[X_t X_t']$ existing and finite we obtain $T^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_t^2 \xrightarrow{p} \mathrm{E}[\varepsilon_t^2]$, where $\hat{\varepsilon}_t = Y_t - X_t'\hat{\beta}$ **(Consistent Estimation of Error Covariance)**

Robust: t-ratio and Wald F-statistics

Under assumptions i) to vi) and a consistent estimator $\hat{V}_T$ of V and $\hat{D} \equiv \hat{D}_T$ given as above we obtain

- Under the null hypothesis $\beta_j = r$,

$$t_j = \frac{\sqrt{T}(\hat{\beta}_j - r)}{\sqrt{\hat{D}_{jj}}} = \frac{\hat{\beta}_j - r}{\hat{SE}_j} \overset{asy}{\sim} N(0,1),$$

  with $\hat{SE}_j \equiv \sqrt{T^{-1}\hat{D}_{jj}}$ (robust standard errors).

- Under the null hypothesis $R\beta = r$, (m linear restrictions)

$$F = T \cdot (R\hat{\beta} - r)'(R\hat{D}R)^{-1}(R\hat{\beta} - r) \overset{asy}{\sim} \chi^2_{(m)}$$

- Under the null hypothesis $a(\beta) = 0$, (m non-linear restrictions), with $A(\beta)$ the matrix of first derivatives of $a(\beta)$

$$F = T \cdot a(\hat{\beta})'(A(\hat{\beta})\hat{D}A(\hat{\beta}))^{-1}a(\hat{\beta}) \overset{asy}{\sim} \chi^2_{(m)}$$

Which $\hat{W}$? – Efficient Estimation

Proposition: Optimal Choice of $\hat{W}$

A lower bound for the asymptotic variance $D_T = D_T(\hat{W})$ of the GMM estimator is given by

$$(\Sigma_{zx}' V^{-1} \Sigma_{zx}')^{-1}$$

and it is achieved if $\hat{W}$ is chosen such that $W (= \operatorname{plim} \hat{W}) = V^{-1}$. Hence, we get:

$$(\Sigma_{zx}' W \Sigma_{zx})^{-1} \Sigma_{zx}' W V W \Sigma_{zx} (\Sigma_{zx}' W \Sigma_{zx})^{-1} \geq (\Sigma_{zx}' V^{-1} \Sigma_{zx}')^{-1}.$$

Since V is usually unknown, we estimate V by $\hat{V} = \hat{V}_T$ and call that GMM estimator the **efficient GMM estimator** for which we replaced the weighting matrix $\hat{W}$ by $\hat{V}^{-1}$ and we obtain:

$$\begin{aligned} \hat{\beta} &= \left(S'_{zx} \hat{V}^{-1} S_{zx} \right)^{-1} S'_{zx} \hat{V}^{-1} s_{zy}, \\ D_T(\hat{V}^{-1}) &= (\Sigma'_{zx} V^{-1} \Sigma_{zx})^{-1}, \\ \hat{D}_T(\hat{V}^{-1}) &= (S'_{zx} \hat{V}^{-1} S_{zx})^{-1}. \end{aligned}$$

Two Step Efficient GMM

- Step 1: Choose a matrix $\hat{W}$ and compute the GMM estimator $\hat{\beta}(\hat{W})$. Standard choices for $\hat{W}$ are the $\hat{W} = I$ or $\hat{W} = S_{zz}^{-1}$ (yields the two step LS estimator). Calculate the residuals $\hat{\varepsilon}_t = Y_t - X_t'\hat{\beta}$ and compute a consistent estimator $\hat{V}$ of V.
- Step 2: Compute the GMM estimator $\hat{\beta}(\hat{V}^{-1})$.

Testing Overidentifying Restrictions

Hansen's test of overidentifying restrictions

Suppose there is a consistent estimator $\hat{V}$ of V. Under assumptions i) to vi) & [vii)]

$$J(\tilde{\beta}, \hat{V}^{-1}) \equiv T g_T(\tilde{\beta})' \hat{V}^{-1} g_T(\tilde{\beta}) \overset{asy}{\sim} \chi^2_{(\ell - k)}.$$

Remark:

- Specification Test testing whether **all** assumptions i) to vii) are satisfied. If the test statistic is very high, it means that indication that the orthogonality conditions **or** the other assumptions **or** both do not hold! Only if we can rule out failure of the other assumptions we can interpret the test as a "test for endogeneity of the instruments".

Testing a subset of instruments for orthogonality

Suppose assumptions i) to vi) & [vii)] hold. Let Z_{1t} be a subset (ℓ_1 dimensional) of the instrument vector Z_t. We also assume that the rank condition for identification also holds for this subset of instruments. Let J denote the usual J-statistic based on Z_t and J_1 the J-statistic derived only using Z_{1t}. Then, for any consistent estimators $\hat{V}$ and $\hat{V}_{11}$ of V and V_{11}, with V_{11} the associated submatrix in V, we obtain

$$J - J_1 \overset{asy}{\sim} \chi^2_{(\ell-\ell_1)}.$$

Likelihood Ratio Type Tests

$H_0 : a(\beta) = 0$, (m non-linear restrictions)

Let $\hat{\beta}_R$ denote the restricted efficient GMM estimator obtain by:

$$\hat{\beta}_R \equiv \hat{\beta}_R(\hat{V}^{-1}) \equiv \underset{\tilde{\beta}}{\operatorname{argmin}}\, J(\tilde{\beta}, \hat{V}^{-1}), \quad \text{s.t. } H_0,$$

then

$$\mathcal{LR} \equiv J(\hat{\beta}_R, \hat{V}^{-1}) - J(\hat{\beta}, \hat{V}^{-1}) \overset{asy}{\sim} \chi^2_{(m)}$$

Remarks:

- The Wald and the LR statistics are asymptotically equivalent, i) the distributions are the same and ii) their numerical difference converges in probability to zero.
- if $H_0 : a(\beta) = 0$ is linear then the statistics are numerically equal.
- The Wald statistic is not invariant to the way $a(\beta) = 0$ is written while the LR statistic is (small sample problem).
- Computation of LR requires a nonlinear optimization program.

- We don't need the efficiency presumption $\hat{W} = V^{-1}$ for the Wald statistic.
- The same estimate of V should be used in both J's in the LR statistic.

Conditional Homoscedasticity

Under the assumption of conditional homoscedasticity $V = \sigma^2 \Sigma_{zz}$ and the efficient GMM estimator becomes the 2SLS one by choosing $\hat{W} = S_{zz}$ or $\hat{W} = \hat{\sigma}^2 S_{zz}$:

$$\hat{\beta}(S_{zz}) = \left(S'_{zx} S_{zz} S_{zx}\right)^{-1} S'_{zx} S_{zz} s_{zy}.$$

It can also be obtained as the limited information ML estimator.

Multiple Equation GMM:

Hayashi, F (2000), Ch. 4.2

We consider now:

$$Y_{th} = X'_{th}\beta_h + \varepsilon_{th}, \quad \text{for } t = 1, \ldots, T;\ h = 1, \ldots, p$$

with X_{th} the k_h dimensional vector of regressors, and Z_{th} the ℓ_h dimensional vector of instruments.

Orthogonality conditions: $\mathrm{E}[g_t] = 0$ with

$$g_t \equiv \begin{pmatrix} Z_{t1}\varepsilon_{t1} \\ \vdots \\ Z_{tp}\varepsilon_{tp} \end{pmatrix}_{(\sum \ell_h \times 1)}$$

Weighting Matrix: $\hat{W}$ of dimension $\sum \ell_h \times \sum \ell_h$.

Asymptotic Covariance V:

$$V = \mathrm{E}[g_t g_t'] = \begin{pmatrix} \mathrm{E}[\varepsilon_{t1}\varepsilon_{t1}Z_{t1}Z_{t1}'] & \cdots & \mathrm{E}[\varepsilon_{t1}\varepsilon_{tp}Z_{t1}Z_{tp}'] \\ \vdots & & \vdots \\ \mathrm{E}[\varepsilon_{tp}\varepsilon_{t1}Z_{tp}Z_{t1}'] & \cdots & \mathrm{E}[\varepsilon_{tp}\varepsilon_{tp}Z_{tp}Z_{tp}'] \end{pmatrix}$$

Remark: Assumptions: The same assumptions apply as in the single equation GMM section, now "equation-by-equation"wise. The assumptions about ergodicity and stationarity are strengthened to ergodicity and stationarity of the joint processes, where applicable.

Single vs. Multiple Equation GMM

Idea: Compute the GMM estimator of β_h equation-by-equation. What changes?

Answer: The weighting matrix: $\hat{W}$. The multiple eq. GMM $\hat{W}$ can be constructed as the stacked single equation $\hat{W}$s, but this is a special case.

And ... in general there will be an impact on the efficiency of the estimator.

But:

- If all equations are exactly identified, then equation-by-equation GMM and multiple GMM are numerical the same. (IV case).
- If at least one equation is over-identified, but the equations are unrelated then the **efficient** equation-by-equation GMM and the **efficient** multiple GMM estimator are asymptotically equivalent.

Unrelated:

$$\mathrm{E}[\varepsilon_{th}\varepsilon_{th'} Z_{th} Z'_{th'}] = 0 \quad \text{for all } h \neq h' (= 1, \ldots, p)$$

Special Cases:

Conditional Homoscedasticity: $\mathrm{E}[\varepsilon_{th}\varepsilon_{th'}|Z_{th}Z'_{th'}] = \sigma_{hh'}$

Full-Information Instrumental Variable Efficient (FIVE):

$$\hat{V} = \begin{pmatrix} \hat{\sigma}_{11}\hat{\mathrm{E}}[Z_{t1}Z'_{t1}] & \cdots & \hat{\sigma}_{1p}\hat{\mathrm{E}}[Z_{t1}Z'_{tp}] \\ \vdots & & \vdots \\ \hat{\sigma}_{p1}\hat{\mathrm{E}}[Z_{tp}Z'_{t1}] & \cdots & \hat{\sigma}_{pp}\hat{\mathrm{E}}[Z_{tp}Z'_{tp}] \end{pmatrix},$$

with $\hat{\sigma}_{hh'} = T^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_{th}\hat{\varepsilon}_{th'}$ and $\hat{\mathrm{E}}[Z_{th}Z'_{th'}] = T^{-1}\sum_{t=1}^{T} Z_{th}Z'_{th'}$ (dim: $\ell_h \times \ell_{h'}$). As the initial estimator usually the 2SLS one is used.

If in addition the **same instruments** $Z_t (= Z_{1t} = \cdots = Z_{pt})$ are used

Three Stage Least Squares (3SLS):

$$\hat{V} = \hat{\Sigma} \otimes \hat{\mathrm{E}}[Z_t Z_t']$$

with $\hat{\Sigma} = T^{-1} \sum_{t=1}^{T} \hat{\varepsilon}_t \hat{\varepsilon}_t'$ (dim: $p \times p$)
and $\hat{\mathrm{E}}[Z_t Z_t'] = T^{-1} \sum_{t=1}^{T} Z_t Z_t'$ (dim: $\ell \times \ell$)
It holds that:

$$\hat{V}^{-1} = \hat{\Sigma}^{-1} \otimes \hat{\mathrm{E}}[Z_t Z_t']^{-1}$$

As the initial estimator usually the 2SLS one is used. Moreover, we obtain now:

$$\hat{\beta}_{3SLS} = \begin{pmatrix} \hat{\sigma}_{11}\hat{A}_{11} & \cdots & \hat{\sigma}_{1p}\hat{A}_{1p} \\ \vdots & & \vdots \\ \hat{\sigma}_{p1}\hat{A}_{p1} & \cdots & \hat{\sigma}_{pp}\hat{A}_{pp} \end{pmatrix}^{-1} \begin{pmatrix} \hat{\sigma}_{11}\hat{c}_{11} & \cdots & \hat{\sigma}_{1p}\hat{c}_{1p} \\ \vdots & & \vdots \\ \hat{\sigma}_{p1}\hat{c}_{p1} & \cdots & \hat{\sigma}_{pp}\hat{c}_{pp} \end{pmatrix}$$

with

$$\begin{aligned} \hat{A}_{hh'} &\equiv \hat{\mathrm{E}}[X_{th}Z_t']\hat{\mathrm{E}}[Z_tZ_t']^{-1}\hat{\mathrm{E}}[Z_tX_{th'}'] \\ \hat{c}_{hh'} &\equiv \hat{\mathrm{E}}[X_{th}Z_t']\hat{\mathrm{E}}[Z_tZ_t']^{-1}\hat{\mathrm{E}}[Z_tY_{th'}] \end{aligned}$$

If in addition the **regressors are a subset of the instruments**, i.e $Z_t = \bigcup_h \{X_{th}\}$ or $\mathrm{E}[X_{th}\varepsilon_{th'}] = 0$, which means there are a priori predetermined and satisfy cross-orthogonality conditions, then:

Seemingly Unrelated Regression (SUR):

$$\begin{aligned} \hat{A}_{hh'} &\equiv \hat{\mathrm{E}}[X_{th}X'_{th'}] \\ \hat{c}_{hh'} &\equiv \hat{\mathrm{E}}[X_{th}Y_{th'}] \end{aligned}$$

OLS vs. SUR

- all equations are exactly identified, then the regressors are the same for all equations and the equation-by-equation OLS and SUR are numerical the same.
- If at least one equation is over-identified, then SUR is more efficient unless the equations are unrelated.

Outlook/Issues:

- Panel Data Models
- Simulation Based Approaches
- Weak Instruments or Weak Identification
- Misspecification

Readings:

- Hayashi, F (2000): *Econometrics*, Princeton University Press, Princeton. Ch.s 3 & 4.
- Hansen L, (1982): Large Sample Properties of Generalized Method of Moments Estimators, *Econometrica*, Vol. 50, No. 4, pp. 1029-1054.
- Hall A, (2005): *Generalized Method of Moments*, Oxford University Press, New York.
- Matyas L, (1999): *Generalized Method of Moments Estimation*, Cambridge University Press.