ECON 4160, Spring term 2013. Lecture 8

GMM and FIML estimation for structural equations

Ragnar Nymoen

Department of Economics

17 Oct 2013

- References to Davidson and MacKinnon,
  - Ch 9.1-9.4 (GMM linear single equations), 9.5 (Non-linear equations, just the motivation)
  - Ch 12.4 (the section Efficient GMM for systems, and the text about 3SLS)
  - Ch 12.5 (the line of argument in the Full-Information Maximum Likelihood section) and the last part of p. 536 (test of overidentifying restrictions)
- NOBEL PRICE TO GMM!
  GMM is an extension of the GIV estimator that accounts for heteroskedasticity and autocorrelation. A famous paper by Lars Petter Hansen from 1982 (joint Nobel prize winner 2013) was the ".. the first to propose GMM estimation under that name" (Davidson and MacKinnon page 367).

- The last part of the lecture gives and intuitive explanation of full information maximum likelihood estimation, FIML, of identified linear dynamic simultaneous equations models (FIML).
- FIML is the workhorse estimation method in the multi-equation dynamic modelling part of OxMetrics/PcGive

GMM for a structural equation, motivation I

- Consider again equation # 1 in a SEM as in DM page 522:

$$\mathbf{y}_1 = \mathbf{X}_1 \beta_1 + \epsilon_\mathbf{1} \tag{1}$$

- Memo:

$$\mathbf{X}_1 = \begin{pmatrix} \mathbf{Z}_1 & : & \mathbf{Y}_1 \end{pmatrix} \tag{2}$$

$$\beta_1 = \begin{pmatrix} \beta_{11} & : & \beta_{21} \end{pmatrix}' \tag{3}$$

  - $\mathbf{y}_1$ is $n \times 1$, with observations of the variable that equation # 1 in the SEM is normalized on.
  - $\mathbf{Z}_1$ is $n \times k_{11}$ with observations of the k_{11} included predetermined or exogenous variables.
  - $\mathbf{Y}_1$, $n \times k_{12}$ holds the included endogenous explanatory variables.

GMM for a structural equation, motivation II

Change assumption about hite-noise disturbances to autocorrelated/heteroskedastic disturbances:

$$E(\epsilon_1\epsilon_1^{'}) = \mathbf{\Omega}_1. \tag{4}$$

- Consider the just-identified or over-identified case (by the rank and order conditions)
- When $E(\epsilon_1\epsilon_1^{'}) = IN(\mathbf{0},\sigma_1^2\mathbf{I}_{nxn})$ we know how to find optimal instruments: Use the reduced form to find the linear combination of exogenous and predetermined variables that give the best predictors of the variables in $\mathbf{Y}_1$.
- When we have the more general $E(\epsilon_1\epsilon_1^{'}) = \mathbf{\Omega}_1$, the best predictors for $\mathbf{Y}_1$ should take the correlation structure between the random variables in ϵ_1 into account.

GMM for a structural equation, motivation III

- Generalized (weighted) LS analogy: The regressors in the model are weighted so as to whiten the residuals to get back to the "classical assumptions"
- In the case of IV estimation, the weights must include the instruments as well—this leads to Generalized Method of Moments

GMM-criterion function and the GMM estimator I

- In the same way as for the GIV estimator, the GMM estimator of $\boldsymbol{\beta}_1$ can be found by minimization of the *GMM-criterion function*:

$$Q_{GMM}(\boldsymbol{\beta}_1, \boldsymbol{\Omega}_1, \mathbf{y}_1) = (\mathbf{y}_1 - \mathbf{X}_1\boldsymbol{\beta}_1)'\mathbf{P}_{gW_1}(\mathbf{y}_1 - \mathbf{X}_1\boldsymbol{\beta}_1). \quad (5)$$

where

$$\mathbf{P}_{gW_1} = \mathbf{W}_1(\mathbf{W}_1'\Omega_1\mathbf{W}_1)^{-1}\mathbf{W}_1'. \quad (6)$$

is the generalization of the GIV prediction maker matrix.

GMM-criterion function and the GMM estimator II

- Assume that $\boldsymbol{\Omega}_1$ is a matrix of known parameters. The first order conditions (i.e. Method of Moments!) are then compactly written as

$$\mathbf{X}_1^{'}\mathbf{P}_{{}_g W_1}(\mathbf{y}_1 - \mathbf{X}_1\hat{\beta}_{1,GMM}) = \mathbf{0} \tag{7}$$

which gives $\hat{\beta}_{1,GMM}$ as the solution

$$\hat{\beta}_{1,GMM} = (\mathbf{X}_1'\mathbf{P}_{W_{g1}}\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{P}_{W_1}\mathbf{y}_1$$

- If you "write it out", you get expression (9.10) on page 355 in DM (with some small changes in notation)

GMM-criterion function and the GMM estimator III

- The generalization of the $\mathbf{J}_1$ matrix from GIV estimation is found from:

$$\mathbf{X}_1^{'}\mathbf{P}_{gW_1} = \mathbf{J}_{g1}^{'}\mathbf{W}_1^{\prime}$$

and becomes (check)

$$\mathbf{J}_{g1} = (\mathbf{W}_1^{\prime}\Omega_1\mathbf{W}_1)^{-1}\mathbf{W}_1^{\prime}\mathbf{X}_1 \tag{8}$$

which gives the optimal weights to the instruments in the GMM case.

MM an IV summary

1. When $E(\epsilon_1\epsilon_1^{'}) = \mathbf{\Omega}_1 \neq I$, $\hat{\boldsymbol{\beta}}_{1,GMM}$ is asymptotically more efficient than $\hat{\boldsymbol{\beta}}_{1,GIV}$.
2. When $\mathbf{\Omega}_1{=}\sigma_1^2\mathbf{I}$: $\hat{\boldsymbol{\beta}}_{1,GMM} = \hat{\boldsymbol{\beta}}_{1,GIV}$ and $\mathbf{J}_{g1} = \sigma_1^{-2}\mathbf{J}_1$ (the reduced form gives the optimal iv weights).
3. $\mathbf{\Omega}_1{=}\sigma_1^2\mathbf{I}$ and exact identification: $\hat{\boldsymbol{\beta}}_{1,GMM} = \hat{\boldsymbol{\beta}}_{1,GIV} = \hat{\mathbf{ffi}}_{IV}$
4. If the structural equation is a regression model, as in a recursive system of equations, we can set $\mathbf{Z}_1 = \mathbf{X}_1$ and the optimal matrix with instruments is $\mathbf{W}_1 = \mathbf{X}_1$. So $\hat{\boldsymbol{\beta}}_{1,GMM} = \hat{\boldsymbol{\beta}}_{1,OLS}$ in this case.

Feasible GMM

- Have seen that GMM is the generalization of GIVE, to the case of *non-white-noise error terms*. Just as GLS generalized OLS estimation to that case.
- In the same way as with GLS, to make GMM a feasible of practical method, $\mathbf{\Omega}_1$ has to be estimated by a consistent estimator $\widehat{\mathbf{\Omega}}_1$. Use GIV residuals to get a first estimate.
- Can choose to iteration over $\hat{\boldsymbol{\beta}}_{1,GMM}$ and/or $\widehat{\mathbf{\Omega}}_1$.
- The covariance matrix of the feasible GMM estimator is

$$\widehat{Var}(\hat{\boldsymbol{\beta}}_{1,GMM}) = (\mathbf{X}_1\mathbf{W}_1(\mathbf{W}_1'\hat{\mathbf{\Omega}}_1\mathbf{W}_1)^{-1}\mathbf{W}_1'\mathbf{X}_1)^{-1} \tag{9}$$

and t-ratios are asymptotically $N(0,1)$ under their respective null hypotheses. So the inference procedure is as are before.

Hansen-Sargan test for GMM I

- Since the GMM criterion function only depends on $\mathbf{\Omega}_1$ trough the square matrix $\mathbf{W}_1'\mathbf{\Omega}_1\mathbf{W}_1$ it is not surprising that the IV *Specfication test* is also defined for GMM.
- It is usually interpreted as a test of the validity of the instruments (orthogonality conditions), but as Davidson and MacKinnon (DM) discuss on
  - p 338, for IV
  - p 367-368 for GMM

  the second possibility is that the *Specification test* becomes significant when an explanatory variable has incorrectly been classified as an instrument instead of (correctly) as an predetermined of exogenous variable (i.e., would go into $\mathbf{W}_1$ via $\mathbf{Z}_1$ not via $\mathbf{X}_{01}$).

Hansen-Sargan test for GMM II

- The general implication of a significant specification test is therefore *re-specification*.

GMM for non-linear equations I

- Economic theory of intertemporal decisions leads to Euler-equations that are formulated as (say) I orthogonality conditions that are similar to the moments conditions
- If the number of parameters to be estimated is less than or equal to I, we have identification.
- The moments conditions can be linear or non-linear in parameters
- The non-linear GMM can be obtained by minimization of certain quadratic forms. Chapter 9.5 is a relatively advanced chapter on non-linear GMM that you use as a reference if you apply this method.

New Keynesian Phillips curve (PCM) I

- DSGE macro models to a large extent are made up of structural equations that are
  - first order conditions of agents intertemporal optimization problems
  - price and wage equations that are based on *Calvo-pricing*: New Keynesian Phillips curves (PCM)
- We will look at a famous example of IV estimation of a PCM from the euro area.
- Let p_t be the log of a price level index. The (hybrid) PCM states that inflation, defined as $\Delta p_t \equiv p_t - p_{t-1}$, is explained by

New Keynesian Phillips curve (PCM) II

- $E_t(\Delta p_{t+1})$, expected inflation one period ahead conditional upon information available at time t, lagged inflation and a variable
- x_t, often called a forcing variable, representing excess demand or marginal costs (e.g., output gap, the unemployment rate or the wage share in logs):

$$\Delta p_t = b_{p1}^f E_t(\Delta p_{t+1}) + b_{p1}^b \Delta p_{t-1} + b_{p2} x_t + \varepsilon_{pt}, \qquad (10)$$

- ε_{pt} is assumed to be white noise

- Theory predicts (a little simplified) : $0 < b_{p1}^f < 1$ and $0 < b_{p1}^b < 1$ and $b_{p2} > 0$ if x_t is measured by the logarithm of the *wage-share*.

The NPC system

- We cannot find $E_t(\Delta p_{t+1})$ from (10) alone.
- To make progress, we need a "completing" system.
- The term 'forcing variable' suggests that x_t is exogenous, so we specify

$$\Delta p_t = b^f_{p1} E_t(\Delta p_{t+1}) + b^b_{p1} \Delta p_{t-1} + b_{p2} x_t + \varepsilon_{pt} \qquad (11)$$

$$x_t = b_{x1} x_{t-1} + \varepsilon_{xt}, \ -1 < b_{x1} < 1 \qquad (12)$$

where ε_{xt} is white-noise, and uncorrelated with ε_{pt}

The NPC in terms of observables

- A popular approach is to substitute the theoretical $E_t(\Delta p_{t+1})$ by the lead variable Δp_{t+1}.
- After substitution, the NPC is:

$$\Delta p_t = b^f_{p1}\Delta p_{t+1} + b^b_{p1}\Delta p_{t-1} + b_{p2}x_t + \nu_{pt} \qquad (13)$$

  the disturbance ν_{pt} contains the forecast error $\{\Delta p_{t+1} - E_t(\Delta p_{t+1})\}$ in addition to the white noise ε_{pt}.
- As usual with error-in-variables models, $\Delta p_{t+1}$is correlated with the ν_{pt}.
- OLS on (13) is inconsistent. Need IV estimation
- We postpone to E 5101 to show that ν_{pt} follow a first order Moving-Average process
- In principle GMM should give more efficient estimation.

Euro-area NPC I

- We replicate the results in *European inflation dynamics, Gali, Gertler and Lopez-Solado, European Economic Review (2001)*.
- Reference: *Econometric evaluation of the New Keynesian Phillips Curve*, Bårdsen, Jansen and Nymoen Oxford Bulletin of Economics and Statistics (2004)
- With the log of the wage-share ws_t as the forcing variable (x_t):

$$\Delta p_t = \underset{(0.06)}{0.60}\,\Delta p_{t+1} + \underset{(0.06)}{0.35}\,\Delta p_{t-1} + \underset{(0.03)}{0.03}\,ws_t + \underset{(0.06)}{0.08} \qquad (14)$$

$$\text{GMM},\ T = 107\ (1972\ (4)\ \text{to}\ 1997\ (4))$$
$$\chi^2_{\text{J}}(8) = 6.74\,[0.35]$$

Euro-area NPC II

- Note that $\chi^2_{\mathrm{J}}(8)$ is the $J-$form of the *Specification test*
- Note that there are 8 overidentifying restrictions, indicating that implicitly GGL had a larger NPC-system in mind.
- The results in (14) are not very robust to the details about how we estimate $\hat{\Omega}_1$. When we iterate over $\hat{\Omega}_1$:

$$\begin{aligned}\Delta p_t &= \underset{(0.052)}{0.731}\ \Delta p_{t+1} + \underset{(0.069)}{0.340}\ \Delta p_{t-1} - \underset{(0.029)}{0.042}\ ws_t \\ &\quad - \underset{(0.070)}{0.102}\end{aligned} \tag{15}$$

$$\text{GMM, } T = 107 \text{ (1971 (3) to 1998 (1))}$$

$$\chi^2_{\mathrm{J}}(8) = 7.34\,[0.50]$$

Euro-area NPC III

Lack of robustness with respect to such details need not e a problem, but it is here.

- GIV estimation results

$$\Delta p_t = \underset{(0.14)}{0.66}\, \Delta p_{t+1} + \underset{(0.12)}{0.28}\, \Delta p_{t-1} + \underset{(0.09)}{0.07}\, ws_t + \underset{(0.12)}{0.10} \tag{16}$$

$$\text{2SLS}, \ T = 104 \ (1972\ (2) \text{ to } 1998\ (1))$$
$$\chi^2_{\text{Specification}}(6) = 11.88[0.06]$$

- Misspecification tests show that there is heavy residual autocorrelation in (16).

Euro-area NPC IV

- Consistent with NPC but could also be a result misspecification: Δp_{t+1} acting as a proxy for omitted current and lagged variables.

The VAR system and dynamic SEM I

- We now switch attention back to dynamic systems and dynamic models of the system such as a the bivariate VAR

$$\underbrace{\begin{pmatrix} Y_{1t} \\ Y_{2t} \end{pmatrix}}_{\mathbf{y}_t} = \underbrace{\begin{pmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \end{pmatrix}}_{\mathbf{\Pi}} \underbrace{\begin{pmatrix} Y_{1t-1} \\ Y_{2t-1} \end{pmatrix}}_{\mathbf{y}_{t-1}} + \underbrace{\begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}}_{\boldsymbol{\varepsilon}_t}, \quad (17)$$

$$\boldsymbol{\varepsilon}_t = IN(\mathbf{0}, \mathbf{\Sigma}_{2\times 2}) \quad (18)$$

The VAR system and dynamic SEM II

- or the bivariate open-VAR (also called VARX)

$$\underbrace{\begin{pmatrix} Y_{1t} \\ Y_{2t} \end{pmatrix}}_{\mathbf{y}_t} = \underbrace{\begin{pmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \end{pmatrix}}_{\mathbf{\Pi}_1} \underbrace{\begin{pmatrix} Y_{t-1} \\ Y_{t-1} \end{pmatrix}}_{\mathbf{y}_{t-1}} + \underbrace{\begin{pmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{pmatrix}}_{\mathbf{\Gamma}_1} \underbrace{\begin{pmatrix} Z_{1t} \\ Z_{2t} \end{pmatrix}}_{\mathbf{z}_t} + \underbrace{\begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}}_{\boldsymbol{\varepsilon}_t} \tag{19}$$

The VAR system and dynamic SEM III

Generalizations to higher dimensions (more variables) and longer lags are unproblematic, as long as $\mathbf{y}_t$ and $\mathbf{z}_t$ are covariance stationary:

$$\mathbf{y}_t = \sum_{i=0}^{p} \mathbf{\Pi}_i \mathbf{y}_{t-i} + \sum_{i=0}^{q} \mathbf{\Gamma}_i \mathbf{z}_{t-i} + \varepsilon_t \quad (20)$$

$$\varepsilon_t = IN(\mathbf{0}, \mathbf{\Sigma}) \quad (21)$$

The VAR system and dynamic SEM IV

- As we have seen (!), dynamic linear structural models of (20)-(21) can be obtained by pre-multiplying (20) by a non-singular matrix **B**:

$$\mathbf{B}\mathbf{y}_t = \sum_{i=0}^{p} \mathbf{B}\mathbf{\Pi}_i \mathbf{y}_{t-i} + \sum_{i=0}^{q} \mathbf{B}\mathbf{\Gamma}_i \mathbf{z}_{t-i} + \mathbf{B}\varepsilon_t \qquad (22)$$

  so that the structural coefficients are in $\mathbf{B}$, $\mathbf{B\Pi}_i$, $\mathbf{B\Gamma}_i$ and the vector of structural disturbances are $\boldsymbol{\epsilon}_t = \mathbf{B}\boldsymbol{\varepsilon}_t$ with $E(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t^{'}) = \mathbf{\Omega}$.
- Assume just identification, or overidentification of (22).

ML estimation of the system and the SEM I

- We know already that ML estimation of the Gaussian VAR system (20)-(21) is obtained by OLS on each reduced form equation.
- The maximum likelihood estimator of the linear SEM (22) is called the **full information maximum likelihood** estimator or FIML.
- Intuitively, FIML estimators of the structural parameters **B**, $\mathbf{B\Pi}_i$, $\mathbf{B\Gamma}_i$ are obtained by "solving back" from the ML estimates of the reduced form parameters
- In the just identified case, the maximized SEM log-likelihood is exactly the same as the **unrestricted reduced form** log likelihood value L_{URF} from the OLS estimation of the (20).

ML estimation of the system and the SEM II

- In the over-identified case, the SEM restricts the maximized log likelihood value L_{RRF} through the over-identifying restrictions.
- The over-identifying can be tested by the use of the LR test-statistic

$$-2(L_{RRF} - L_{URF})$$

  which is Chi squared distributed with d.f equal to the degree of overidentification.
- PcGive reports this LR statistic as the **LR test of over-identifying restrictions** when a SEM is estimated by FIML, or any one of the other estimation methods in the *Multiple-Equation Dynamic Modelling* part of the program:
  - 2SLS

ML estimation of the system and the SEM III

- **3SLS**
- 1SLS (OLS equation by equation)

3SLS estimation I

- 3SLS is a GMM type estimator which is efficient when $E(\epsilon_t \epsilon_t^{'}) = \mathbf{\Omega}$ is not-diagonal.
  - We first estimate the SEM, equation by equation, with 2SLS
  - From the 2SLS residuals, we construct the consistent $\widehat{\mathbf{\Omega}}$.
  - Using $\widehat{\mathbf{\Omega}}$ in a GMM estimator of the structural coefficients gives the 3SLS estimator
- See p 531-532 (3SLS) and p 522-524 (GMM for contemporaneously correlated residuals)
- But since Since PcGive does an excellent FIML, little practical need for 3SLS.