ECON 4160: Econometrics–Modelling and Systems Estimation:

Computer Class # 5

André Anundsen

Department of Economics, University of Oslo

October 21, 2013

Outline

Data sets for today:

- *exact_identification.in7*
- *weak_identification.in7*
- *over_identification.in7*

Batch files for today:

- *Exact identification IV.fl*
- *Weak identification IV.fl*
- *Over identification IV.fl*
- *VAR.fl*
- *Exactly identified FIML.fl*
- *Over identified FIML.fl*

Recursive graphs of parameters I

Consider the model

$$y_t = \beta_1 + \beta_2 x_t + \varepsilon_t, \;\; t = 1, 2, \ldots, T, \tag{1}$$

As we have seen, a simple and intuitive method for assessing the hypothesis that the parameters β_1, β_2, $\sigma = \sqrt{Var(\varepsilon_t)}$ are stable, is to graph the output from a recursive estimation.

Recursive graphs of parameters II

Recursive estimation works in the following way:

1. Estimate on a short sample: $t = 1, 2, ..., T_1,\ T_1 < T$
2. Extend the sample with single observation, and estimate the model on the sample $t = 1, 2, ..., T_1 + 1,\ T_1 + 1 < T$
3. Continue until all observations are part of the sample, i.e. $t = 1, 2, ..., T$
4. The sequence of estimates, $\hat{\beta}_1(j)$, $\hat{\beta}_2(j)$, and $\hat{\sigma}(j)$ $(j = T_1, T_1 + 1, \ldots, T)$, can be plotted against time

Pc-Give plots e.g., $\hat{\beta}_2(j)$ together with $\pm 2\sqrt{Var(\hat{\beta}_1(j))}$, and $\hat{\sigma}(j)$ together with the *1-step residuals*, which are

$$y_j - \hat{\beta}_1(j) - \hat{\beta}_2(j)x_j, \quad j = T_1, T_1 + 1, ..., T$$

Chow-tests of parameter stability

- If we have a hypothesis about when a structural break occurs, we can test that hypothesis
- Let T_1 denote the last period with the "old" regime and let $T_1 + 1$ denote the first period of the "new" regime:

$$y_t = \beta_1 + \beta_2 x_t + \varepsilon_t, \ \ t = 1, 2, \ldots, T_1 \text{ and}$$
$$y_s = \gamma_1 + \gamma_2 x_s + \epsilon_s, \ \ s = T_1 + 1, 2, \ldots, T.$$

then

$$H_0\text{: } \beta_1 = \gamma_1, \beta_2 = \gamma_2 \text{ vs } H_1\text{: } \beta_1 \neq \gamma_1, \beta_2 \neq \gamma_2.$$

- There are two well known statistics for these cases, both due to Chow (1960) and referred to as *Chow tests*.

2-sample Chow-test

Let SSE be the Sum of Squared Errors (which of course corresponds to RSS or SSR):

- SSE_1 is for the first sample ($t = 1, 2, .., T_1$)
- SSE_2 is for the second sample ($t = T_1 + 1, 2, .., T_1$)
- $SSE_U = SSE_1 + SSE_2$
- SSE_R is the SSE when the whole sample is used, i.e. under H_0

The test statistic in our example is:

$$F_{Chow2} = \frac{SSE_R - SSE_U}{SSE_U} \cdot \frac{(T-4)}{2} \sim F(2, T-4).$$

In general, it is:

$$F_{Chow2} = \frac{SSE_R - SSE_U}{SSE_U} \cdot \frac{T-2K}{K} \sim F(K, T-2K).$$

where K is the number of parameters

Insufficient observations Chow-test

- Consider $T - T_1 < K$ (the sample for the second period), where we remember that K is the number of parameters in the model. In this case we have:
  - The same SSE_R as for the 2-sample Chow-test (i.e. based on the full sample)
  - But SSE_U is only based on the first T_1 observations

This "predictive" Chow-test is given as

$$F_{ChowP} = \frac{SSE_R - SSE_U}{SSE_U} \cdot \frac{T_1 - K}{T - T_1} \sim F(T - T_1, T_1 - K)$$

Insufficient observations Chow-test

- If we have no clear idea about the dating of a regime shift, a graph with the whole sequence of predictive Chow tests is useful
- Chow tests rely on constant and equal variances of the disturbances. Hence, it is good practice to plot the recursively estimated $\hat{\sigma}$

Recursive Chow-test: 1-step test

The Test-Menu Graphics in PcGive contains three version of the F_{ChowP} test!

- *1-step test*: Both T_1 and T is increased by one observation sequentially, so that $N = T - T_1$ is always 1. Naturally, the test statistic is:
  -

$$F_{ChowP} = \frac{SSE_R - SSE_U}{SSE_U} \cdot \frac{T_1 - K}{T - T_1}$$
$$= \frac{SSE_R - SSE_U}{SSE_U} \cdot \frac{T_1 - K}{1} \sim F(1, T_1 - K)$$

Example:

1. Estimate (1) on the sample $t = 1, ..., 10$ and $t = 1, ..., 11$ and calculate F_{ChowP}
2. Estimate (1) on the sample $t = 1, ..., 11$ and $t = 1, ..., 12$ and calculate F_{ChowP}
3. Continue like this! The graph plots the sequence of F_{ChowP}

Recursive Chow-test: Break-point Chow test

- *Break-point Chow test*, called *N-dn CHOWS* (N down Chow Statistic) in PcGive Here T_1 is increased by one observation, then by 2 and so on until $T_1 = T$ (T is always fixed!). Hence, $N = T - T_1$ is decreasing (N down) through the sequence of tests. The test statistic is:

$$F_{ChowP} = \frac{SSE_R - SSE_U}{SSE_U} \cdot \frac{T_1 - K}{T - T_1} \sim F(T - T_1, T_1 - K)$$

Example

1. Estimate (1) on the sample $t = 1, ..., 10$ and $t = 1, ..., 200$ and calculate F_{ChowP}
2. Estimate (1) on the sample $t = 1, ..., 11$ and $t = 1, ..., 200$ and calculate F_{ChowP}
3. Continue like this! The graph plots the sequence of F_{ChowP}

Recursive Chow-test: Forecast Chow test

The final version is the:

- *Forecast Chow test*, called *N-up CHOWS* (N up Statistic): Here T_1 is kept fixed and $T - T_1$ is first 1, then 2, and so on. Hence, $N = T - T_1$ is increasing through the sequence of tests (N up). The test statistic is:

$$F_{ChowP} = \frac{SSE_R - SSE_U}{SSE_U} \cdot \frac{T_1 - K}{T - T_1} \sim F(T - T_1, T_1 - K)$$

Example

1. Estimate (1) on the sample $t = 1, ..., 10$ and $t = 1, ..., 199$ and calculate F_{ChowP}
2. Estimate (1) on the sample $t = 1, ..., 10$ and $t = 1, ..., 200$ and calculate F_{ChowP}
3. Continue like this! The graph plots the sequence of F_{ChowP}

- All the Recursive Chows are scaled by (1-off) critical values. (1% level is the default) so that the critical values becomes a straight line at unity

You can easily calculate the different recursive Chow test statistics yourself by using OLS and then comparing what you get to what the program tells you (choose *Write results instead of graphing*). This is a nice way to become familiar with the tests

The Order Condition for identification

Suppose you have a system of N endogenous variables and K exogenous variables. Equation i in that system is exactly identified if the number of exogenous variables omitted from that the equation $K - K_i$ (K_i is the number of exogenous variables in equation i) is equal to the number of endogenous variables in that equation, N_i, less one, i.e.

$$K - K_i = N_i - 1 \text{ ,or equivalently } (K - K_i) + (N - N_i) = N - 1$$

In a similar way we get that equation is overidentified if

$$K - K_i > N_i - 1 \text{ ,or equivalently } (K - K_i) + (N - N_i) > N - 1$$

And it is not identified if:

$$K - K_i < N_i - 1 \text{ ,or equivalently } (K - K_i) + (N - N_i) < N - 1$$

Recap: The rank condition for identification I

- The order condition is necessary, but not sufficient for identification
- If the coefficient of the omitted exogenous variable in an equation *i* of the system *is zero also in the equation(s)* where it is not omitted, its omission does not help for identification of equation *i*
- The omitted variable is in that case not a part of the statistical system (the reduced form), and it is not a valid instrument

Recap: The rank condition for identification II

- The rank condition says that an equation in a system of N equations is identified *if and only if we can construct at least one non-zero* $(N-1) \times (N-1)$ *determinant from the coefficients excluded from that equation, but that are still contained in other equations in the system*
  Another way to think about this is that the matrix of the coefficients on the variables excluded from that equation should have rank $= N-1$

Recap: The rank condition for identification III

An intuitive way of understanding the rank condition: *All variables excluded from an equation in a system must appear in at least one other equation in that system (non-zero columns), i.e. it must be part of the reduced form. In addition, at least one of the variables excluded from that equation must appear in all other equations in the system*

What do we take away form this?

- The *Order condition for identification* is a necessary, but not sufficient, condition for identification.
- Any conclusion about identification on the basis of the Order condition is therefore provisional, until the Rank condition has been checked (in practice that is often a small step though)
- The Rank condition is sufficient for identification
- Note that the main distinction is between non-identified and identified.

- The amount of (or degree of overidentification) is another issue: An over-identified equation is identified so there are no principal difference between identification and over-identification.
- What we can gain from overidentification is twofold:
  - We have more than one consistent estimator of the structural parameters—these can be weighted together in an efficient way: Achieved by 2SLS in the case of independent structural disturbances.
  - The validity of the overidentifying restrictions can be tested.

Consider the following simultaneous (demand-supply) system:

$$p_t = \beta_{10} + \beta_{11} x_t + \gamma_{11} z_{1t} + \gamma_{13} z_{3t} + \varepsilon_{1t} \quad (2)$$

$$x_t = \beta_{20} + \beta_{22} p_t + \gamma_{22} z_{2t} + \gamma_{23} z_{3t} + \varepsilon_{2t} \quad (3)$$

where the z's are exogenous and the disturbances have the following joint distribution

$$\boldsymbol{\varepsilon}_t = \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \sim N\left(\mathbf{0}_{2\times 1}, \boldsymbol{\Omega} = \begin{pmatrix} \omega_1^2 & \omega_{12} \\ \omega_{12} & \omega_2^2 \end{pmatrix}\right)$$

Using our knowlegde about identification, we see that:

1. Equation (2) is exactly identified
2. Equation (3) is exactly identified

Suppose we are interested only in estimating the inverted demand function (equation 2). How would you proceed?

1. We first acknowledge that OLS estimator is inconsistent, since $cov(x_t, \varepsilon_{1t}) \neq 0$
2. But you know from Lecture # 7 that we can overcome this challenge by considering the 2SLS (which – in the case of exact identification – is equal to the IV and the ILS estimator)

What do we require for a (vector of) variable(s), w_t, to work as an instrument for x_t in (2)?

1. Instrument relevance: $cov(x_t, w_t) \neq 0$
2. Must solve the problem: $cov(\varepsilon_{1t}, w_t) = 0$
3. At least one of the instruments must not be part of (2), i.e. we need identification!

Given this, what are the relevant candidate(s)?

We remember that the system reads:

$$p_t = \beta_{10} + \beta_{11} x_t + \gamma_{11} z_{1t} + \gamma_{13} z_{3t} + \varepsilon_{1t}$$
$$x_t = \beta_{20} + \beta_{22} p_t + \gamma_{22} z_{2t} + \gamma_{23} z_{3t} + \varepsilon_{2t}$$

We see that the only natural candidate is z_{2t}, since:

- It is part of the supply equation (normalized on x_t), but not the demand equation (normalized on p_t)
- By exogeneity, it is uncorrelated with ε_{1t}

Message!

Identification is a system property, so:

Important to represent the system when looking for instruments!

I have generated some artificial data with the following structural parameters

$$\beta_{10} = 1.2, \beta_{20} = 0.8$$
$$\beta_{11} = -1.6$$
$$\beta_{22} = 1.4$$
$$\gamma_{11} = 2$$
$$\gamma_{22} = 1.5$$
$$\gamma_{13} = 1.3, \gamma_{23} = -0.5$$

and

$$\sigma_1^2 = 1, \sigma_2^2 = 1, \sigma_{12} = 0.7$$

Thus, our system looks like this:

$$p_t = 1.2 - 1.6x_t + 2z_{1t} + 1.3z_{3t} + \varepsilon_{1t}$$
$$x_t = 0.8 + 1.4p_t + 1.5z_{2t} - 2.5z_{3t} + \varepsilon_{2t}$$

$$\boldsymbol{\varepsilon}_t \sim N\left(\mathbf{0}_{2\times 1}, \begin{pmatrix} 1 & 0.7 \\ 0.7 & 1 \end{pmatrix}\right)$$

Use the data *exact_identification.in7* and estimate the inverted demand equation using IV estimation in PcGive. Explore also the batch-file *Exact identification IV.fl*

Weak identification

I have made a small change to the DGP, so that $\gamma_{22} = 0.2$ instead of $\gamma_{22} = 1.5$, i.e. it has a relatively small impact on x_t (the variable it is acting as an instrument for!). In this case, our system looks like this:

$$p_t = 1.2 - 1.6x_t + 2z_{1t} + 1.3z_{3t} + \varepsilon_{1t}$$
$$x_t = 0.8 + 1.4p_t + 0.2z_{2t} - 2.5z_{3t} + \varepsilon_{2t}$$

Estimate the inverted demand equation by IV using the dataset *weak_identification.in7*! What do you find? You can also look at the batch file *Weak identification IV.fl*

The Sargan-Hansen Specification test

- In the case of over-identification, we can test the assumption of independence between the instruments and the disturbances of the structural equation we estimate
- In the IVE output from PcGive, this test is found as the

Specification test

It is calculated as

$$\text{number of observations} * R^2_{IVres}$$

where R^2_{IVres} is the R^2 from the regression between the residuals from the IV estimated structural equation and the variables that used as instruments (an auxiliary regression)

The Sargan-Hansen Specification test

- Under the H_0 of correct specification, number of observations * R^2_{IVres} is approximately χ^2 with degrees of freedom equal to the number of overidentifying instruments
- This test is one of the most cited in single equation IV estimation, see for example Stata output where it is referred to as *Hansen J statstic [Overidentifying test of all instruments]* (Nobel Lareute Hansen)
- It is basically the same test as in Give (Bårdsen and Nymoen (2011), p 231 contains a clarification), which is attributed to an earlier contribution by Denis Sargan
- Since the test is based on residuals, we include the intercept both in the structural equation and in the auxiliary regression

The Sargan-Hansen Specification test

- Take care to note that also the exogenous (non-IV) variable(s) of the structural equation under consideration is included in the auxiliary regresson, since it is part of the reduced form, and hence part of the instrument list!

The overidentified case

I have now generated some artificial data with the following structural parameters

$$\beta_{10} = 1.2, \beta_{20} = 0.8$$
$$\beta_{11} = -1.6$$
$$\beta_{22} = 1.4$$
$$\gamma_{11} = 2$$
$$\gamma_{22} = 1.5$$
$$\gamma_{13} = 0, \gamma_{23} = -0.5$$

and

$$\sigma_1^2 = 1, \sigma_2^2 = 1, \sigma_{12} = 0.7$$

Thus, our system looks like this:

$$p_t = 1.2 - 1.6x_t + 2z_{1t} + \varepsilon_{1t}$$
$$x_t = 0.8 + 1.4p_t + 1.5z_{2t} - 2.5z_{3t} + \varepsilon_{2t}$$

$$\boldsymbol{\varepsilon}_t =\sim N\left(\mathbf{0}_{2\times 1}, \begin{pmatrix} 1 & 0.7 \\ 0.7 & 1 \end{pmatrix}\right)$$

Estimate this system using 2SLS (IVE) estimation in PcGive and the data *over_identifcation.in7*. Do we reject the overidentifying restriction? Play around with *Over identifcation IV.fl* as well

We find that R^2 from the auxiliary regression of the residuals on the IV vector is 3.80E-05. Multiplying this by the number of observations (200), we get

$$200 * 3.80E - 05 = 7.59E - 03 = 0.00759$$

and from the program, we have

$$\text{Specification test:}\chi^2(1) = 0.0075919[0.9306]$$

from the IV estimation. We see that the two are the same!

Note that the system (where we are now back at the case with $\gamma_{13} \neq 0$):

$$p_t = \beta_{10} + \beta_{11} x_t + \gamma_{11} z_{1t} + \gamma_{13} z_{3t} + \varepsilon_{1t}$$
$$x_t = \beta_{20} + \beta_{22} p_t + \gamma_{22} z_{2t} + \gamma_{23} z_{3t} + \varepsilon_{2t}$$

may also be expressed as:

$$-\beta_{11} x_t + p_t = \beta_{10} + \gamma_{11} z_{1t} + \gamma_{13} z_{3t} + \varepsilon_{1t}$$
$$x_t - \beta_{22} p_t = \beta_{20} + \gamma_{22} z_{2t} + \gamma_{23} z_{3t} + \varepsilon_{2t}$$

This can also be written on matrix form:

$$\begin{pmatrix} -\beta_{11} & 1 \\ 1 & -\beta_{22} \end{pmatrix} \begin{pmatrix} x_t \\ p_t \end{pmatrix} = \begin{pmatrix} \beta_{10} \\ \beta_{20} \end{pmatrix} + \begin{pmatrix} \gamma_{11} & 0 & \gamma_{13} \\ 0 & \gamma_{22} & \gamma_{23} \end{pmatrix} \begin{pmatrix} z_{1t} \\ z_{2t} \\ z_{3t} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix}$$

or more compact:

$$\mathbf{B}\mathbf{y}_t = \mathbf{A} + \mathbf{\Gamma}\mathbf{Z}_t + \boldsymbol{\varepsilon}_t \tag{4}$$

where the various matrices are defined in the obvious way

From the SEM to the VARX

Note that by pre-multiplying by $\mathbf{B}^{-1}$, we have:

$$\mathbf{y}_t = \mathbf{B}^{-1}\mathbf{A} + \mathbf{B}^{-1}\mathbf{\Gamma}\mathbf{Z}_t + \mathbf{B}^{-1}\boldsymbol{\varepsilon}_t$$

or

$$\mathbf{y}_t = \boldsymbol{\mu} + \mathbf{\Phi}\mathbf{Z}_t + \boldsymbol{\epsilon}_t \tag{5}$$

with $\boldsymbol{\mu} = \mathbf{B}^{-1}\mathbf{A}, \mathbf{\Phi} = \mathbf{B}^{-1}\mathbf{\Gamma}, \boldsymbol{\epsilon}_t = \mathbf{B}^{-1}\boldsymbol{\varepsilon}_t$. But this is just a $VARX(0,0)$ model, i.e. a VAR with exogenous variables and with no lags! Furthermore, we see that:

$$E(\boldsymbol{\epsilon}_t) = E(\mathbf{B}^{-1}\boldsymbol{\varepsilon}_t) = \mathbf{B}^{-1}E(\boldsymbol{\varepsilon}_t) = \mathbf{0}_{2\times 1}$$

$$E\left(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t'\right) = E\left(\left(\mathbf{B}^{-1}\boldsymbol{\varepsilon}_t\right)\left(\mathbf{B}^{-1}\boldsymbol{\varepsilon}_t\right)'\right) = \mathbf{B}^{-1}E\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\right)\left(\mathbf{B}^{-1}\right)' = \mathbf{B}^{-1}\mathbf{\Omega}\left(\mathbf{B}^{-1}\right)' = \mathbf{\Sigma}$$

and hence

$$\boldsymbol{\epsilon}_t \sim MVN(\mathbf{0}, \mathbf{\Sigma}) \tag{6}$$

- We know that ML estimation of the Gaussian VAR system (5)-(6) is obtained by OLS on each of the reduced form equations
- The maximum likelihood estimator of the linear SEM (4) is called the **full information maximum likelihood** estimator or FIML.
- Intuitively, FIML estimators of the structural parameters **B**, **A** and $\mathbf{\Gamma}$ are obtained by "solving back" from the ML estimates of the reduced form parameters
- In the just identified case, the maximized SEM log-likelihood from (4) is exactly the same as the **unrestricted reduced form** log-likelihood obtained from the OLS estimation of (5),L_{URF}.

Explore this by estimating both the SEM and the VAR using the data *exactly_identified.in7*. Also play around with the batch files *VAR.fl* and *Exactly identified FIML.fl*

Remember, this is the DGP:

$$p_t = 1.2 - 1.6x_t + 2z_{1t} + 1.3z_{3t} + \varepsilon_{1t}$$
$$x_t = 0.8 + 1.4p_t + 1.5z_{2t} - 2.5z_{3t} + \varepsilon_{2t}$$

So, FIML is doing a quite decent job!

In both cases, we find $LogLik = -259.336815$

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

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

  which is χ^2 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:
  1. 2SLS
  2. **3SLS**
  3. 1SLS (OLS equation by equation)

- 3SLS is a GMM type estimator which is efficient when $E(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t^{'}) = \boldsymbol{\Omega}$ is not-diagonal.
  - We first estimate the SEM, equation by equation, with 2SLS
  - From the 2SLS residuals, we construct the consistent $\widehat{\boldsymbol{\Omega}}$.
  - Using $\widehat{\boldsymbol{\Omega}}$ 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.

Use the data *over_identified.in7* to test your new modelling skills!! In particular, consider first the unrestricted case ($\gamma_{13} \neq 0$):

$$p_t = \beta_{10} + \beta_{11} x_t + \gamma_{11} z_{1t} + \gamma_{13} z_{3t} + \varepsilon_{1t}$$
$$x_t = \beta_{20} + \beta_{22} p_t + \gamma_{22} z_{2t} + \gamma_{23} z_{3t} + \varepsilon_{2t}$$

and then the restricted case:

$$p_t = \beta_{10} + \beta_{11} x_t + \gamma_{11} z_{1t} + \varepsilon_{1t}$$
$$x_t = \beta_{20} + \beta_{22} p_t + \gamma_{22} z_{2t} + \gamma_{23} z_{3t} + \varepsilon_{2t}$$

Estimate by FIML using the data set *over_identification*. What does the test for overidentifying restrictions suggest? Also play around with the batch file *Over identified FIML.fl*

We find that the unrestricted likelihood is:

$$L_{URF} = -259.336815$$

We find that the restricted likelihood is:

$$L_{RRF} = -259.340611$$

and thus:

$$-2(L_{RRF} - L_{URF}) = -2\left(-259.340611 - (-259.336815)\right) = 0.007592$$

and from the program, we have:
LR test of over-identifying restrictions: $\chi^2(1) = 0.0075920[0.9306]$

They are equal!