Six notes on Basic Econometric topics

Supplement to:

Greene: Econometric Analysis (6th or 7th Edition)

Translated excerpts from Biørn: Økonometriske emner

For Master course:

ECON4160 ECONOMETRICS – MODELLING AND SYSTEMS ESTIMATION

CONTENTS

Erik Biørn

Department of Economics,
University of Oslo

Version of July 5, 2011

ECON4160 ECONOMETRICS – MODELLING AND SYSTEMS ESTIMATION

Lecture note A:

EXOGENEITY AND AUTONOMY

Erik Biørn
Department of Economics

Version of July 5, 2011

1. A model with stochastic regressors

Consider the regression equation

$$\text{(A-1)} \qquad y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i, \qquad i = 1, \dots, n,$$

where $(y_i, x_{1i}, x_{2i}, u_i)$ are all assumed to be stochastic, They then have a joint probability distribution. We assume in the simplest, 'classical' regression model that

$$\text{(A-2)} \qquad \mathsf{E}(u_i | x_{1i}, x_{2i}) = 0, \qquad i = 1, \dots, n,$$

$$\text{(A-3)} \qquad \mathsf{E}(u_i u_j | x_{1i}, x_{2i}, x_{1j}, x_{2j}) = \begin{cases} \sigma^2 & \text{for } j = i, \\ 0 & \text{for } j \neq i, \end{cases} \qquad i, j = 1, \dots, n.$$

It follows from these assumptions that

$$\text{(A-4)} \quad \mathsf{E}(y_i | x_{1i}, x_{2i}) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \mathsf{E}(u_i | x_{1i}, x_{2i}) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i},$$

$$\text{(A-5)} \quad \text{var}(y_i | x_{1i}, x_{2i}) = \text{var}(u_i | x_{1i}, x_{2i}) = \sigma^2,$$

$$\text{(A-6)} \quad \text{cov}(y_i, y_j | x_{1i}, x_{2i}, x_{1j}, x_{2j}) = \text{cov}(u_i, u_j | x_{1i}, x_{2i}, x_{1j}, x_{2j}) = 0, \qquad i, j = 1, \dots, n, \; j \neq i.$$

In deriving (A-4)–(A-6) from (A-1)–(A-3), we exploit the following:

(a) *When considering a conditional distribution, we can proceed and reason as if the variables on which we condition are constant, non-stochastic parameters.*

(b) The expectation of a non-stochastic entity is the entity itself.

(c) The variance of a non-stochastic entity and the covariance between two non-stochastic entities are zero.

Equations (A-4) and (A-5) summarize our assumptions about the joint distribution of (y_i, x_{1i}, x_{2i}).

Let us represent this distribution by the density function

$$\text{(A-7)} \qquad f(y_i, x_{1i}, x_{2i}) = f_y(y_i | x_{1i}, x_{2i})\, g(x_{1i}, x_{2i}), \qquad i = 1, \dots, n,$$

where $f_y(y_i | x_{1i}, x_{2i})$ is the conditional density function of y_i, given (x_{1i}, x_{2i}), and $g(x_{1i}, x_{2i})$ is the simultaneous (marginal) density function of (x_{1i}, x_{2i}). THE ESSENCE of the model description above is that we postulate that the conditional distribution $(y_i|x_{1i}, x_{2i})$ has certain properties, that is, that the function $f_y(\cdot)$ has certain properties, but do not postulate anything about the function $g(\cdot)$.

2. Formal definitions of exogeneity in relation to regression models

In economic theory we say that a variable is exogenous if it is 'determined outside the model'. This is interesting also in econometrics, but it is too vague and imprecise.

We will now refer **four alternative definitions** which can be used for regression models: We consider the RHS (Right Hand Side) variables in (A-1) as stochastic and say that the disturbance, u_i, has zero expectation and that the x_{ki}s are exogenous relative to the regression equation (A-1) if

Definition (i): $\mathsf{E}(u_i|x_{1i}, x_{2i}) = 0$, $i = 1, \dots, n$.

Definition (ii): $\mathsf{E}(u_i) = 0$ *and* $\text{cov}(u_i, x_{ki}) = 0$, $i = 1, \dots, n;\ k = 1, 2$.

Definition (iii): $\mathsf{E}(u_i) = 0$, $i = 1, \dots, n$, *and* $u = (u_1, u_2, \dots, u_n)'$ *are stochastically independent of* $X = (x_{11}, x_{21}, \dots, x_{1n}, x_{2n})$.

Definition (iv): *In* $f(y_i, x_{1i}, x_{2i}) = f_y(y_i|x_{1i}, x_{2i})\, g(x_{1i}, x_{2i})$ *[cf. (A-7)], the distribution represented by the density function* $g(x_{1i}, x_{2i})$ *is determined outside the model. In particular, the parameters* $(\beta_0, \beta_1, \beta_2, \sigma^2)$*, which describe* $f_y(\cdot)$*, are assumed not to be included among the parameters of* $g(\cdot)$*.* Also: The variation of the parameters in $g(\cdot)$ does not impose any restriction on the possible variation of $(\beta_0, \beta_1, \beta_2, \sigma^2)$ in $f_y(\cdot)$.

These are four different, competing definitions. Definition (iv) is of a different nature than Definitions (i)–(iii), while Definitions (i)–(iii) are relatively closely related.

Definition (iii) is the most restrictive. If exogeneity according to definition (iii) is satisfied, then exogeneity according to Definition (i) will also hold, because stochastic independence between two variables implies that their conditional and marginal distributions coincide Therefore $\mathsf{E}(u_i|X) = 0$ is a necessary (but not sufficient) condition for both $\mathsf{E}(u_i) = 0$ and stochastic independence of u_i and X. Hence (iii) implies (i). Moreover, the conditions in Definition (i) are stronger than in Definition (ii). This follows from the theorem of double expectation, as can be seen as follows: Assume that (i) is satisfied. It then follows, first, that

$$\mathsf{E}(u_i) = \mathsf{E}[\mathsf{E}(u_i|x_{1i}, x_{2i})] = \mathsf{E}(0) = 0,$$

and second, that

$$\text{cov}(u_i, x_{ki}) = \mathsf{E}(u_i x_{ki}) = \mathsf{E}[\mathsf{E}(u_i\, x_{ki}|x_{1i}, x_{2i})] = \mathsf{E}[x_{ki}\mathsf{E}(u_i|x_{1i}, x_{2i})] = 0.$$

Consequently Definition (i) implies that Definition (ii) also holds.

Our conclusion therefore is

$$\textbf{Definition (iii)} \implies \textbf{Definition (i)} \implies \textbf{Definition (ii)}.$$

In this course we will mostly stick to Definition (i) – also when considering more complicated models (generalized regression models, systems of regression models, simultaneous equation systems, etc.).

3. Remark on the concept of autonomy

Consider the decomposition (A-7). It shows that the form of $f(y_i, x_{1i}, x_{2i})$ may undergo changes either by changes in $f_y(y_i|x_{1i}, x_{2i})$ or by changes in $g(x_{1i}, x_{2i})$ (or both). If a change in $g(\cdot)$ induces a change in $f(\cdot)$ while $f_y(\cdot)$ is unchanged, then we say that $f_y(\cdot)$ is autonomous with respect to the change in $g(\cdot)$. This term was used by some of the 'founding fathers' of Econometrics (Frisch, Tinbergen, Koopmans, Haavelmo) 'Structural invariance' may be a more modern term.

Example 1: Assume that (A-1)–(A-6) is an econometric model of the consumption function: y = consumption, x_1 = income, x_2 = wealth. Then $f_y(y_i|x_{1i}, x_{2i})$ represents the conditional distribution of consumption, given income and wealth, its expectation being the (expected) consumption function, and $g(x_{1i}, x_{2i})$ represents the joint distribution of income and wealth. Then the parameters of the consumption function may be invariant to changes in the distribution of income and wealth, for instance induced by changes in the tax system. We then say that equation (A-1) is autonomous to this change in the income-wealth distribution. Equation (A-1) may be autonomous to some such changes, but not to others. (Could you find examples?)

Example 2: Assume that (A-1)–(A-6) is an econometric model of the log of a Cobb-Douglas production function: y = log-output, x_1 = log-labour input, x_2 = log-capital input. Then $f_y(y_i|x_{1i}, x_{2i})$ represents the conditional distribution of log-output, given log-labour and log-capital input, its expectation being the (expected) production function, and $g(x_{1i}, x_{2i})$ represents the joint distribution of the log-inputs. Then the parameters of the production may be, or may not be, invariant to changes in the distribution, for instance induced by changes in output and input prices, excise taxes etc. We then say that equation (A-1) is autonomous to, or not autonomous to, this change in the log-input distribution. Equation (A-1) may be autonomous to some such changes, but not to others. (Could you find examples?)

Further readings:

Greene: Econometric Analysis. B.8 and B.9.

Engle, Hendry & Richard: Exogeneity. *Econometrica* 51 (1983), 277–304.

Aldrich: Autonomy. *Oxford Economic Papers*, 41 (1989), 15–34.

ECON4160 ECONOMETRICS – MODELLING AND SYSTEMS ESTIMATION

Lecture note B:

SYSTEMS OF REGRESSION EQUATIONS

Erik Biørn
Department of Economics

Version of July 5, 2011

1. What is a system of regression equations?

A system of linear regression equations is a model with the following characteristics:

(i) *The model has at least two (linear) equations.*

(ii) *Each equation has one and only one endogenous variable, which is the equation's LHS (Left Hand Side) variable.*

(iii) *Each equation has one or more exogenous variables. They are the equation's RHS (Right Hand Side) variables. They may be specific for a single equation or occur in several equations.*

(iv) *Each equation has a disturbance which is uncorrelated with all RHS variables in all the equations.*

(v) *The disturbances of different equations may be correlated.*

(vi) *Restrictions on coefficients in different equations may be imposed.*

2. Model with two regression equations

Consider the model

$$\text{(B-1)}\qquad \begin{aligned} y_{1i} &= \beta_1\, x_{1i} + u_{1i}, \\ y_{2i} &= \beta_2\, x_{2i} + u_{2i}, \end{aligned} \qquad i = 1, \dots, n,$$

where u_{1i} and u_{2i} are disturbances, y_{1i} and y_{2i} are endogenous variables and x_{1i} and x_{2i} are exogenous variables. We assume that the four latter variables are measured from their means, so that we can specify the equations without intercepts. (Explain why.) No restriction are imposed on the coefficients β_1 and β_2.

Considering the x's as stochastic, we make the following assumptions:

$$\text{(B-2)}\qquad \mathsf{E}(u_{ki}|\boldsymbol{X}) = 0,$$

$$\text{(B-3)}\qquad \mathsf{E}(u_{ki}u_{rj}|\boldsymbol{X}) = \begin{cases} \sigma_{kr} & \text{for } j = i, \\ 0 & \text{for } j \neq i, \end{cases} \qquad k, r = 1, 2;\ i, j = 1, \dots, n,$$

where $\boldsymbol{X} = (x_{11}, \dots, x_{1n}, x_{21}, \dots, x_{2n})$. Here, (B-2) formally expresses that x_1 *and* x_2*, are exogeneous relatively to both regression equations.* Using the rule of iterated expectations, it follows that

$$\text{(B-4)}\qquad \mathsf{E}(u_{ki}) = 0,$$

$$\text{(B-5)}\qquad \text{cov}(u_{ki}, x_{rj}) = 0,$$

$$\text{(B-6)}\qquad \text{var}(u_{ki}) = \sigma_{kk},$$

$$\text{(B-7)}\qquad \text{cov}(u_{ki}, u_{rj}) = \begin{cases} \sigma_{kr}, & j = i, \\ 0, & j \neq i. \end{cases} \qquad \begin{aligned} &k, r = 1, 2, \\ &i, j = 1, \dots, n. \end{aligned}$$

Note that, by (B-7), we allow for the disturbances in the two equations in (B-1) being correlated.

3. Compressing the two equations into one equation

We now perform the following trick: *We combine the two equations in (B-1), each with one RHS variable and* n *observations, into one equation with two RHS variables and* $2n$ *observations.* Technically, we do this by defining three new variables, y_i*, x^*_{1i} and x^*_{2i}, and new disturbances u^*_i in the following way:

i	y^*_i	x^*_{1i}	x^*_{2i}	u^*_i
1	y_{11}	x_{11}	0	u_{11}
2	y_{12}	x_{12}	0	u_{12}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	y_{1n}	x_{1n}	0	u_{1n}
$n+1$	y_{21}	0	x_{21}	u_{21}
$n+2$	y_{22}	0	x_{22}	u_{22}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$2n$	y_{2n}	0	x_{2n}	u_{2n}

Then (B-1) can be written as

(B-8) $$y^*_i = \beta_1 x^*_{1i} \ + \ \beta_2 x^*_{2i} \ + \ u^*_i, \qquad i = 1, \dots, 2n.$$

Defining the following vectors and matrices:

$$\boldsymbol{y} = \begin{bmatrix} y^*_1 \\ y^*_2 \\ \vdots \\ y^*_{2n} \end{bmatrix} = \begin{bmatrix} y_{11} \\ \vdots \\ y_{1n} \\ y_{21} \\ \vdots \\ y_{2n} \end{bmatrix}, \quad \boldsymbol{u} = \begin{bmatrix} u^*_1 \\ u^*_2 \\ \vdots \\ u^*_{2n} \end{bmatrix} = \begin{bmatrix} u_{11} \\ \vdots \\ u_{1n} \\ u_{21} \\ \vdots \\ u_{2n} \end{bmatrix},$$

$$\boldsymbol{X} = \begin{bmatrix} x^*_{11} & x^*_{21} \\ x^*_{12} & x^*_{22} \\ \vdots & \vdots \\ x^*_{1n} & x^*_{2n} \\ x^*_{1,n+1} & x^*_{2,n+1} \\ x^*_{1,n+2} & x^*_{2,n+2} \\ \vdots & \vdots \\ x^*_{1,2n} & x^*_{2,2n} \end{bmatrix} = \begin{bmatrix} x_{11} & 0 \\ x_{12} & 0 \\ \vdots & \vdots \\ x_{1n} & 0 \\ 0 & x_{21} \\ 0 & x_{22} \\ \vdots & \vdots \\ 0 & x_{2n} \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix},$$

of dimensions $(2n \times 1)$, $(2n \times 1)$, $(2n \times 2)$ and (2×1), respectively, (B-8) can be written compactly as

(B-9) $$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}.$$

4. The composite disturbance covariance matrix

Equation (B-8) looks like *a generalized regression equation with two RHS variables*, based on $2n$ observations, with the following variance-covariance structure:

$$\text{(B-10)} \qquad \mathsf{E}(u_i^*|\boldsymbol{X}) = 0, \qquad i = 1,\dots,2n,$$

$$\text{(B-11)} \qquad \mathsf{E}(u_i^* u_j^*|\boldsymbol{X}) = \begin{cases} \sigma_{11} & \text{for} \quad i=j=1,\dots,n, \\ \sigma_{22} & \text{for} \quad i=j=n+1,\dots,2n, \\ \sigma_{12} & \text{for} \quad i=1,\dots,n;\ j=i+n \text{ and} \\ & \phantom{\text{for}} \quad i=n+1,\dots,2n;\ j=i-n, \\ 0 & \text{otherwise.} \end{cases}$$

Here (B-11) expresses *that we formally have (i) a particular kind of heteroskedasticity, in that the variances can take two different values, and (ii) a particular kind of autocorrelation* in the composite disturbance vector, in that disturbances with distance n observations are correlated.

We can write (B-2) and (B-3) in matrix notation as

$$\text{(B-12)} \qquad \mathsf{E}(\boldsymbol{u}|\boldsymbol{X}) = \boldsymbol{0}_{2m},$$

and

$$\text{(B-13)} \qquad \mathsf{E}(\boldsymbol{u}\boldsymbol{u}'|\boldsymbol{X}) = \boldsymbol{V},$$

where

$$\boldsymbol{V} = \begin{bmatrix} \sigma_{11} & 0 & \cdots & 0 & \sigma_{12} & 0 & \cdots & 0 \\ 0 & \sigma_{11} & \cdots & 0 & 0 & \sigma_{12} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{11} & 0 & 0 & \cdots & \sigma_{12} \\ \sigma_{12} & 0 & \cdots & 0 & \sigma_{22} & 0 & \cdots & 0 \\ 0 & \sigma_{12} & \cdots & 0 & 0 & \sigma_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{12} & 0 & 0 & \cdots & \sigma_{22} \end{bmatrix} = \begin{bmatrix} \sigma_{11}\boldsymbol{I}_n & \sigma_{12}\boldsymbol{I}_n \\ \sigma_{12}\boldsymbol{I}_n & \sigma_{22}\boldsymbol{I}_n \end{bmatrix}.$$

It exemplifies a *partitioned matrix*, which is a matrix with a block structure.

5. Estimation of the composite equation by GLS

We apply GLS on (B-9), assuming that the σ's are known. The normal equations for the estimators, which we denote as $\widetilde{\beta} = (\widetilde{\beta}_1, \widetilde{\beta}_2)'$, have, in matrix notation, the form

$$\text{(B-14)} \qquad (\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})\widetilde{\boldsymbol{\beta}} = \boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{y}.$$

We need an expression for the inverse of V and the matrix products in the normal equations. It is not difficult to show that

$$\boldsymbol{V}^{-1} = (\sigma_{11}\sigma_{22} - \sigma_{12}^2)^{-1} \begin{bmatrix} \sigma_{22}\boldsymbol{I}_n & -\sigma_{12}\boldsymbol{I}_n \\ -\sigma_{12}\boldsymbol{I}_n & \sigma_{11}\boldsymbol{I}_n \end{bmatrix},$$

which exists if $\sigma_{11}\sigma_{22} > \sigma_{12}^2$, i.e., if the disturbances in the two equations are not perfectly correlated. One can verify this by showing that $\boldsymbol{V}\boldsymbol{V}^{-1} = \boldsymbol{I}_{2n}$. (If $\sigma_{11}\sigma_{22} = \sigma_{12}^2$, then V is singular and GLS collapses.) Notice that $\boldsymbol{V}$ can be considered as obtained from the (2×2)-covariance matrix

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}$$

when each element is "inflated" by the identity matrix $\boldsymbol{I}_n$. We may write this as $\boldsymbol{V} = \boldsymbol{\Sigma} \otimes \boldsymbol{I}_n$, where $\otimes$ is an operator which performs such operations, often denoted as *Kronecker-products* [see Greene: Appendix A.5.5]. Correspondingly, the inverse $\boldsymbol{V}^{-1}$ has the (2×2) covariance matrix

$$\boldsymbol{\Sigma}^{-1} = (\sigma_{11}\sigma_{22} - \sigma_{12}^2)^{-1} \begin{bmatrix} \sigma_{22} & -\sigma_{12} \\ -\sigma_{12} & \sigma_{11} \end{bmatrix}$$

after having "inflated" each element by $\boldsymbol{I}_n$; we have $\boldsymbol{V}^{-1} = \boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n^{-1} = \boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n$.

Using the definitions of $\boldsymbol{X}$ and $\boldsymbol{y}$ and the expression for $\boldsymbol{V}^{-1}$, we find that the normal equations for GLS, (B-14), can be expressed in terms of the variables in the original regression equations, $(x_{1i}, x_{2i}, y_{1i}, y_{2i})$, as follows

$$\begin{aligned}(\sigma_{11}\sigma_{22} - \sigma_{12}^2)^{-1} &\begin{bmatrix} \sigma_{22}\sum x_{1i}^2 & -\sigma_{12}\sum x_{1i}x_{2i} \\ -\sigma_{12}\sum x_{1i}x_{2i} & \sigma_{11}\sum x_{2i}^2 \end{bmatrix} \begin{bmatrix} \widetilde{\beta}_1 \\ \widetilde{\beta}_2 \end{bmatrix} \\ &= (\sigma_{11}\sigma_{22} - \sigma_{12}^2)^{-1} \begin{bmatrix} \sigma_{22}\sum x_{1i}y_{1i} - \sigma_{12}\sum x_{1i}y_{2i} \\ -\sigma_{12}\sum x_{2i}y_{1i} + \sigma_{11}\sum x_{2i}y_{2i} \end{bmatrix},\end{aligned}$$

where $\sum$ denotes summation across i from 1 to n. Multiply on both sides of the equality sign by $(\sigma_{11}\sigma_{22} - \sigma_{12}^2)$ and divide by n. This gives

$$\text{(B-15)} \qquad \begin{aligned} \sigma_{22}m_{11}\widetilde{\beta}_1 - \sigma_{12}m_{12}\widetilde{\beta}_2 &= \sigma_{22}m_{y11} - \sigma_{12}m_{y21}, \\ -\sigma_{12}m_{12}\widetilde{\beta}_1 + \sigma_{11}m_{22}\widetilde{\beta}_2 &= -\sigma_{12}m_{y12} + \sigma_{11}m_{y22}, \end{aligned}$$

where (recall that the variables are measured from their means):

$$m_{kr} = M[x_k, x_r] = \tfrac{1}{n}\textstyle\sum_{i=1}^n x_{ki}x_{ri}, \qquad m_{ykr} = M[y_k, x_r] = \tfrac{1}{n}\textstyle\sum_{i=1}^n y_{ki}x_{ri}.$$

6. The GLS estimators

The solution to (B-16), giving the GLS estimators, exists if the determinant value of (B-15), which is

$$\sigma_{11}\sigma_{22}m_{11}m_{22} - \sigma_{12}^2 m_{12}^2,$$

is different from zero. Since $m_{11}m_{22} \geq m_{12}^2$ always holds, $\sigma_{11}\sigma_{22} > \sigma_{12}^2$ (i.e., the disturbances are not perfectly correlated) will be sufficient to ensure a positive determinant value. The explicit expressions for the GLS estimators are

$$\text{(B-16)}\qquad \begin{aligned} \widetilde{\beta}_1 &= \frac{(\sigma_{22}m_{y11}-\sigma_{12}m_{y21})(\sigma_{11}m_{22})+(-\sigma_{12}m_{y12}+\sigma_{11}m_{y22})(\sigma_{12}m_{12})}{\sigma_{11}\sigma_{22}m_{11}m_{22} - \sigma_{12}^2 m_{12}^2}, \\ \widetilde{\beta}_2 &= \frac{(\sigma_{22}m_{y11}-\sigma_{12}m_{y21})(\sigma_{12}m_{12})+(-\sigma_{12}m_{y12}+\sigma_{11}m_{y22})(\sigma_{22}m_{11})}{\sigma_{11}\sigma_{22}m_{11}m_{22} - \sigma_{12}^2 m_{12}^2}. \end{aligned}$$

In general, both estimators are functions of (i) all the three *empirical* second order moments of the x's, (ii) all the four *empirical* covariances between y's and x's and (iii) alle the three *theoretical* second order moments of the two disturbances. In the general case we are only able to perform the estimation if σ_{11}, σ_{22} and σ_{12} (or at least their ratios) are known.

In general, GLS applied on a generalized regression model, which the present two-equation model rearranged into one equation is, is MVLUE. This optimality therefore also carries over to our case, provided that the disturbance variances and covariances are known.

7. Estimation of the σ's and Feasible GLS

When the σ's are unknown, we can proceed as follows:

1. Estimate the regression equations separately by OLS, giving

$$\widehat{\beta}_1 = \frac{m_{y11}}{m_{11}} = \frac{M[y_1, x_1]}{M[x_1, x_1]}, \qquad \widehat{\beta}_2 = \frac{m_{y22}}{m_{22}} = \frac{M[y_2, x_2]}{M[x_2, x_2]}.$$

The corresponding residuals are $\widehat{u}_{ki} = y_{ki} - \widehat{\beta}_k x_{ki}$, $k = 1, 2$. Use these in forming the estimates

$$\widehat{\sigma}_{11} = \tfrac{1}{n}\textstyle\sum_{i=1}^n \widehat{u}_{1i}^2, \quad \widehat{\sigma}_{22} = \tfrac{1}{n}\sum_{i=1}^n \widehat{u}_{2i}^2, \quad \widehat{\sigma}_{12} = \tfrac{1}{n}\sum_{i=1}^n \widehat{u}_{1i}\widehat{u}_{2i}.$$

They are consistent because

$$\text{plim}(\widehat{\sigma}_{kr}) = \text{plim}(M[\widehat{u}_k, \widehat{u}_r]) = \text{plim}(M[u_k, u_r]) = \sigma_{kr}.$$

Exercise: Show this, by exploiting the consistency of the OLS estimators in stage 1.

2. Replace σ_{11}, σ_{22} and σ_{12} i (B-16) by $\widehat{\sigma}_{11}$, $\widehat{\sigma}_{22}$ and $\widehat{\sigma}_{12}$ and use the expressions obtained as estimators of β_1 and β_2. This gives the FGLS (Feasible GLS) estimators. They differ from GLS in finite samples, but the two sets of estimators converge when $n \to \infty$.

8. Two special cases

In general, GLS applied on (B-8) will give a gain over OLS applied separately on the two equations (B-1), in the form of *lower variances of the estimators.* We know that GLS applied on a generalized regression model is MVLUE. In two cases, however, GLS applied in this way will degenerate to simple application of OLS on each of the two equations.

(i) Uncorrelated disturbances: Assume that $\sigma_{12} = 0$. Then (B-16) can be simplified to

$$\widetilde{\beta}_1 = \frac{m_{y11}}{m_{11}} = \frac{M[y_1, x_1]}{M[x_1, x_1]} = \frac{\sum y_{1i} x_{1i}}{\sum x_{1i}^2},$$

$$\widetilde{\beta}_2 = \frac{m_{y22}}{m_{22}} = \frac{M[y_2, x_2]}{M[x_2, x_2]} = \frac{\sum y_{2i} x_{2i}}{\sum x_{2i}^2}.$$

IN THIS CASE, APPLICATION OF GLS ON THE COMPOSITE EQUATION DEGENERATES TO APPLYING OLS ON EACH OF THE ORIGINAL EQUATIONS SEPARATELY.

(ii) Identical regressors: Assume that $x_{1i} = x_{2i} = x_i$ for $i = 1, \dots, n$. Then we have

$$m_{11} = m_{22} = m_{12} = \tfrac{1}{n} \textstyle\sum_{i=1}^n x_i^2 = m_{xx} = M[x, x],$$

where $m_{xx} = M[x, x]$ is defined by the fourth equality. We also have

$$m_{y11} = m_{y12} = \tfrac{1}{n} \textstyle\sum_{i=1}^n y_{1i} x_i = m_{y1x} = M[y_1, x],$$
$$m_{y21} = m_{y22} = \tfrac{1}{n} \textstyle\sum_{i=1}^n y_{2i} x_i = m_{y2x} = M[y_2, x].$$

Then (B-16) can be simplified to, provided that $\sigma_{11}\sigma_{22} > \sigma_{12}^2$,

$$\widetilde{\beta}_1 = \frac{m_{y1x}}{m_{xx}} = \frac{M[y_1, x]}{M[x, x]} = \frac{\sum y_{1i} x_i}{\sum x_i^2},$$
$$\widetilde{\beta}_2 = \frac{m_{y2x}}{m_{xx}} = \frac{M[y_2, x]}{M[x, x]} = \frac{\sum y_{2i} x_i}{\sum x_i^2}.$$

ALSO IN THIS CASE, APPLICATION OF GLS ON THE COMPOSITE EQUATION DEGENERATES TO APPLYING OLS ON EACH OF THE ORIGINAL EQUATIONS SEPARATELY.

ECON4160 ECONOMETRICS – MODELLING AND SYSTEMS ESTIMATION

Lecture note C:

ASYMPTOTIC CONCEPTS AND RESULTS. STOCHASTIC CONVERGENCE

Erik Biørn
Department of Economics

Version of July 5, 2011

There are several kinds of stochastic convergence that we are concerned with in econometrics. Below we will present some of them.

1. Probability limit. Convergence in probability

Let $\widehat{\theta}_{(n)}$ be a stochastic variable constructed from n observations.

If a θ exists such that

$$\lim_{n\to\infty} P(|\widehat{\theta}_{(n)} - \theta| < \delta) = 1 \quad \text{for all } \delta > 0$$
$$\Longleftrightarrow$$
$$\lim_{n\to\infty} P(|\widehat{\theta}_{(n)} - \theta| > \delta) = 0 \quad \text{for all } \delta > 0,$$

then $\widehat{\theta}_{(n)}$ is said to have probability limit θ. An equivalent statement is that $\widehat{\theta}_{(n)}$ converges towards θ in probablity.

In compact notation this is written as

$$\operatorname*{plim}_{n\to\infty}(\widehat{\theta}_{(n)}) = \theta \quad \text{or} \quad \widehat{\theta}_{(n)} \xrightarrow{p} \theta.$$

using p as abbreviation for probability. Often, we write $\text{plim}_{n\to\infty}$ simply as plim. That $\widehat{\theta}_{(n)}$ converges towards θ in probability means that the probability that $\widehat{\theta}_{(n)}$ deviates from θ by less than a small positive entity δ goes to one when the number of observations goes to infinity. The probability mass (or the density function if it exists) of $\widehat{\theta}_{(n)}$ is gradually more strongly concentrated around the point θ as the number of observations increases.

2. Consistency

Assume that $\widehat{\theta}_{(n)}$ is a statistic (a KNOWN function of OBSERVABLE stochastic variables), based on n observations. Let us use it as an estimator for the parameter θ in a probability distribution or in an econometric model. (For notational simplicity, we use θ to symbolize both the parameter and its value.)

If $\widehat{\theta}_{(n)}$ converges towards θ in probability when the number of observations goes to infinity, i.e., if

$$\operatorname*{plim}_{n\to\infty}(\widehat{\theta}_{(n)}) = \theta,$$

then $\widehat{\theta}_{(n)}$ is said to be a consistent estimator of θ, or briefly, to be consistent for θ.

We usually require of an estimator for a parameter that is consistent. This is often regarded as a *minimum requirement.* If the estimator does not converge towards the true value even in the most favourable data situation imaginable (i.e., with an infinite sample size), then it is hardly useful.

3. Slutsky's Theorem

Slutsky's theorem is concerned with a very important and very useful property of *probability limits of functions of stochastic variables*:

> *Assume that the probability limits of the stochastic variables* $\widehat{\lambda}_1, \dots, \widehat{\lambda}_K$*, based on* n *observations, exist and are equal to* $\lambda_1, \dots, \lambda_K$*, respectively. Let* $g(\widehat{\lambda}_1, \dots, \widehat{\lambda}_K)$ *be a function of these stochastic variables which is continuous in the point* $g(\lambda_1, \dots, \lambda_K)$*. Then we have*
>
> $$\operatorname*{plim}_{n\to\infty} g(\widehat{\lambda}_1, \dots, \widehat{\lambda}_K) = g[\text{plim}(\widehat{\lambda}_1), \dots, \text{plim}(\widehat{\lambda}_K)] = g(\lambda_1, \dots, \lambda_K).$$

Briefly, we can "move the plim operator inside the function symbol". The following examples illustrate applications of Slutsky's theorem for $K = 2$:

$$\begin{aligned}
&\text{plim}(\widehat{\lambda}_1 + \widehat{\lambda}_2) = \text{plim}(\widehat{\lambda}_1) + \text{plim}(\widehat{\lambda}_2),\\
&\text{plim}(\widehat{\lambda}_1\widehat{\lambda}_2) = \text{plim}(\widehat{\lambda}_1)\text{plim}(\widehat{\lambda}_2),\\
&\text{plim}\left(\frac{\widehat{\lambda}_1}{\widehat{\lambda}_2}\right) = \frac{\text{plim}(\widehat{\lambda}_1)}{\text{plim}(\widehat{\lambda}_2)} \qquad (\text{plim}(\widehat{\lambda}_2) \neq 0),\\
&\text{plim}[\ln(\widehat{\lambda}_1) + \ln(\widehat{\lambda}_2)] = \ln[\text{plim}(\widehat{\lambda}_1)] + \ln[\text{plim}(\widehat{\lambda}_2)].
\end{aligned}$$

4. Convergence in mean and asymptotic unbiasedness

Let $\widehat{\theta}_{(n)}$ be a stochastic variable based on n observations. We use is as estimator of the parameter θ

> *If*
>
> $$\lim_{n\to\infty} \mathsf{E}(\widehat{\theta}_{(n)}) = \theta$$
>
> *then* $\widehat{\theta}_{(n)}$ *is said to converge towards* θ *in expectation, or we say that the estimator is asymptotically unbiased for* θ*.*

Consistency is conceptually different from asymptotic unbiasedness. First, it may be cases in which an estimator is consistent even if its asymptotic expectation does not exist. Second, an estimator may have an *expectation* which converges towards the true parameter value, while the *probability mass* (or the density it is exists) of the estimator will not be more and more concentrated around this value as the sample size increases.

5. Convergence in quadratic mean

Let $\widehat{\theta}_{(n)}$ be a stochastic variable based on n observations.

If there exists a θ such that

$$\lim_{n\to\infty} \mathsf{E}(\widehat{\theta}_{(n)}) = \theta \text{ and } \lim_{n\to\infty} \mathsf{E}(\widehat{\theta}_{(n)} - \theta)^2 = 0,$$

then $\widehat{\theta}_{(n)}$ is said to converge towards θ in quadratic mean.

In compact notation, this is written as

$$\widehat{\theta}_{(n)} \stackrel{qm}{\to} \theta,$$

where *qm* is an abbreviation of quadratic mean.

An example of an estimator which converges to the true parameter value in quadratic mean is an unbiased or an asymptotically unbiased estimator whose variance goes to zero when the number of observations goes to infinity.

Convergence in quadratic mean is a stronger claim than convergence in probability, formally

$$\widehat{\theta}_{(n)} \stackrel{qm}{\to} \theta \implies \widehat{\theta}_{(n)} \stackrel{p}{\to} \theta,$$

while the opposite does not necessarily hold. The proof of this is based on *Chebychev's inequality*, which in the present case says that, for any positive constant ϵ, we have

$$P(|\widehat{\theta}_{(n)} - \theta| > \epsilon) \leq \frac{\mathsf{E}[(\widehat{\theta}_{(n)} - \theta)^2]}{\epsilon^2}.$$

Thus there may exist consistent estimators which do not converge to the true parameter value in quadratic mean. The variance of a consistent estimator does not even need to exist. On the other hand, it is *sufficient* for an estimator being consistent that it is unbiased and its variance goes to zero.

6. Convergence in distribution

Let z_n be a stochastic variable based on n observations, with cumulative distribution function (CDF) $F_n(z_n)$, and let z be another stochastic variable with CDF $F(z)$.

If

$$\lim_{n\to\infty} |F_n(z_n) - F(z)| = 0$$

in all points where F is continuous, then we say that z_n converges towards z in distribution. It is written

$$z_n \stackrel{d}{\to} z,$$

letting d be an abbreviation of distribution.

It can be shown that *convergence in distribution is a stronger claim than convergence in probability*, in the following sense:

$$z_n \xrightarrow{d} z \implies z_n - z \xrightarrow{p} 0.$$

See also Greene: Theorem D.17.

7. Convergence in moments

Convergence in moments refers to the very useful property of probability limits of empirical moments of stochastic variables that they, under certain relatively wide conditions, converge in probability towards the corresponding theoretical moments, if the latter exist. Most frequently, this property is used for first-order moments (means and expectations) and second-order moments (empirical and theoretical variances and covariances). We can formulate this property as follows:

1. *Assume that at $x_1, \dots, x_n$ are stochastic variables with common expectation μ_x and finite variance. Let $\bar{x} = (1/n)\sum_{i=1}^{n} x_i$. Then*

$$\operatorname*{plim}_{n\to\infty}(\bar{x}) = \mathsf{E}(x_i) = \mu.$$

2. *Assume that at $x_1, \dots, x_n$ and $y_1, \dots, y_n$ are stochastic variables with theoretical covariance $\text{cov}(x_i, y_i) = \sigma_{xy}$. Let $M[x, y] = (1/n)\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$ be their empirical covariance. Then the following holds under certain additional conditions (inter alia, about the higher-order moments of the distribution*:

$$\operatorname*{plim}_{n\to\infty}(M[x, y]) = \text{cov}(x_i, y_i) = \sigma_{xy}.$$

If, in particular, $y_i = x_i$, a similar property holds for the variance.

An example of the result in 1 is that the mean of disturbances with a common expectation of zero and finite variances converges to zero in probability. An example of application of the result in 2 is that the empirical covariance between the disturbance and an exogenous variable in a regression equation has probability limit zero. Another application is using empirical second-order moments as consistent estimators of corresponding theoretical moments.

8. An example illustrating the distinction between asymptotic unbiasedness, convergence in quadratic mean, and consistency

In the previous sections, the following asymptotic properties of estimators of parameters in econometric models were defined: asymptotic unbiasedness, convergence in quadratic mean, and consistency. I have experienced that students (and other people as well) often have some difficulty in distinguishing intuitively between these three concepts – and in keeping the distinction in their mind. This note provides an example – admittedly particular – illustrating the distinction.

8.1. The distribution of the estimator

Let $\widehat{\theta}_n$ be an estimator of the parameter θ, based on n observations. Assume that it has the following (particular) discrete distribution, characterized by the parameters (θ, r, q):

$$\text{(C-17)} \qquad \begin{array}{lll} \widehat{\theta}_n = \theta & \text{with} & P(\widehat{\theta}_n = \theta) = p = 1-n^{-r}, \\ \widehat{\theta}_n = n^q & \text{with} & P(\widehat{\theta}_n = n^q) = 1-p = n^{-r}, \end{array} \qquad 0 < r < \infty;\ 0 < q < \infty.$$

Hence $\widehat{\theta}_n$ takes the 'true' value with probability, p, which goes to one when the number of observations goes to infinity. On the other hand, there is a small probability, $1-p$ (going to zero when the number of observations goes to infinity) that the estimator takes a value which grows with the number of observations. It turns out that the various asymptotic properties depend on the relative magnitude of r and q, i.e., how fast the outlier in the distribution of the estimator, n^q, goes to infinity and how fast the probability of this outcome, n^{-r} goes to zero.

8.2. Formal properties of the estimator

We find from (C-17) and the definition of an expectation and a variance that

$$\text{(C-18)} \qquad \begin{aligned} \mathsf{E}(\widehat{\theta}_n) &= \theta\, P(\widehat{\theta}_n = \theta) + n^q\, P(\widehat{\theta}_n = n^q) \\ &= \theta(1-n^{-r}) + n^q n^{-r} = \theta + (n^q-\theta)n^{-r}, \end{aligned}$$

$$\text{(C-19)} \qquad \begin{aligned} \mathsf{E}[(\widehat{\theta}_n-\theta)^2] &= (\theta-\theta)^2\, P(\widehat{\theta}_n = \theta) + (n^q-\theta)^2\, P(\widehat{\theta}_n = n^q) \\ &= (\theta-\theta)^2(1-n^{-r}) + (n^q-\theta)^2 n^{-r} = (n^q-\theta)^2 n^{-r}, \end{aligned}$$

$$\text{(C-20)} \qquad \begin{aligned} \mathsf{V}(\widehat{\theta}_n) &= \mathsf{E}[\widehat{\theta}_n-\mathsf{E}(\widehat{\theta}_n)]^2 = \mathsf{E}[(\widehat{\theta}_n-\theta)^2] - [\mathsf{E}(\widehat{\theta}_n)-\theta]^2 \\ &= (n^q-\theta)^2 n^{-r}(1-n^{-r}). \end{aligned}$$

We also find

$$\text{(C-21)} \qquad \lim_{n\to\infty} P(|\widehat{\theta}_{(n)}-\theta| < \delta) = \lim_{n\to\infty} P(\widehat{\theta}_n = \theta) = \lim_{n\to\infty} (1-n^{-r}), \quad \forall\, \delta > 0.$$

8.3. Implications of the properties

Equations (C-18)–(C-21) imply, respectively,

$$\text{(C-22)} \qquad \lim_{n\to\infty} \mathsf{E}(\widehat{\theta}_n) = \begin{cases} = \theta & \text{if} \ \ 0 < q < r, \\ = \theta+1 & \text{if} \ \ 0 < q = r, \\ \text{does not exist} & \text{if} \ \ 0 < r < q. \end{cases}$$

$$\text{(C-23)} \qquad \lim_{n\to\infty} \mathsf{E}[(\widehat{\theta}_n - \theta)^2] = \begin{cases} = 0 & \text{if} \ \ 0 < q < \frac{1}{2}r, \\ = 1 & \text{if} \ \ 0 < q = \frac{1}{2}r, \\ \text{does not exist} & \text{if} \ \ 0 < \frac{1}{2}r < q. \end{cases}$$

$$\text{(C-24)} \qquad \lim_{n\to\infty} \mathsf{V}(\widehat{\theta}_n) = \begin{cases} = 0 & \text{if} \ \ 0 < q < \frac{1}{2}r, \\ = 1 & \text{if} \ \ 0 < q = \frac{1}{2}r, \\ \text{does not exist} & \text{if} \ \ 0 < \frac{1}{2}r < q. \end{cases}$$

(C-25) $$\operatorname*{plim}_{n\to\infty}(\widehat{\theta}_n) = \theta \quad \text{if} \quad r > 0, \quad \text{regardless of} \quad q$$

We see that

♣ An estimator may be asymptotically unbiased even if its asymptotic variance does not exist.
Example: $q = 2,\ r = 3$.

♣ An estimator may be consistent even if its asymptotic expectation and its asymptotic variance does not exist.
Example: $q = 3,\ r = 2$.

♣ An estimator may be consistent if its asymptotic expectation exists, even if its asymptotic variance does not exist.
Example: $q = 2,\ r = 3$.

♣ Leave for a moment the above example and try to recall, when the number of observations grows towards infinity, the visual picture of the density function of a **continuously** distributed estimator (it may for instance have a bell-shape) which is (a) asymptotically unbiased, and (b) consistent.

8.4. Conclusion

(i) The estimator $\widehat{\theta}_n$ is asymptotically unbiased for θ if $q < r$.
Example: $q = 1,\ r = 2$.

(ii) The estimator $\widehat{\theta}_n$ converges to θ in quadratic mean if $q < \frac{1}{2}r$.
Example: $q = 1,\ r = 3$.

(iii) The estimator $\widehat{\theta}_n$ has a variance which converges to zero if $q < \frac{1}{2}r$.
Example: $q = 1,\ r = 3$.

(iv) The estimator $\widehat{\theta}_n$ is consistent for θ if $r > 0$.
Example: $q = 1, 2, 3, \dots\ ,\ r = 1$.

Further readings:

W.H. Greene: Econometric Analysis. Appendix D.

B. McCabe and A. Tremayne: Elements of modern asymptotic theory with statistical applications. Manchester University Press, 1993.

ECON4160 ECONOMETRICS – MODELLING AND SYSTEMS ESTIMATION

Lecture note D:

CONCEPTS RELATED TO SIMULTANEOUS EQUATION SYSTEMS

Erik Biørn
Department of Economics

Version of July 5, 2011

In this lecture note some basic terms related to econometric simultaneous equation systems are explained.

1. The structural form and related concepts

(i) *A* STRUCTURAL RELATIONSHIP *or* STRUCTURAL EQUATION *is a relationship representing:*

- *(a) economic agents' or sectors' behaviour,*
- *(b) technological conditions,*
- *(c) institutional conditions, or*
- *(d) definitional equations or equilibrium equations.*

Each structural equation represents a specific theory which is assumed to be valid independently of the fact that it is an element in a simultaneous model together with other structural equations.

Basic to postulating any structural equation is, in principle, an imagined *experiment*, more or less precisely described by the underlying economic theory. Examples are supply functions obtained from producer theory under price taking behaviour, or demand functions obtained from consumer theory under price taking behaviour.

A structural equation has a certain *degree of autonomy* (independent existence, or *structural invariance*), vis-à-vis changes in other structural equations. If a equation possesses a high degree of autonomy, this signifies that it represents a good theory. Relationships in macro theory which have been argued to have a low (or not particularly high) degree of autonomy against shocks in the rest of the economy are the Keynesian investment function and Phillips curve relations. (Could you give some reasons for this?) We can imagine changes occurring in the economic structure which 'destroy' one (or a few) relationship(s) without affecting the others. A demand function may, for instance, be destroyed by sudden or gradual changes in preferences or in demographic characteristics without changing the supply function. A supply function may, for instance, be destroyed by sudden or gradual changes in the production technology or in the supply side conditions without changing the demand function. Etc.

(ii) *By a* STRUCTURAL COEFFICIENT *we understand a coefficient occurring in a structural equation.* Examples are the price coefficients β_1 and β_2 in the market models in Lecture note E, Section 1, below.

(iii) *By a* STRUCTURAL PARAMETER *we understand a structural coefficient* OR *a parameter describing the distribution of the disturbance of a structural equation.*

Examples are β_1, β_2, σ_u^2, and σ_v^2 in the market models in Lecture note E below.

(iv) *By a* STRUCTURAL MODEL OR A MODEL'S STRUCTURAL FORM (ABBREVIATED SF) *we mean a set of structural equation, with as high degree of autonomy as possible, put together into a determined system with specified endogenous and exogenous variables. A structural model has the same number of equations as its number of endogenous variables.*

2. The reduced form and related concepts

(i) *By a structural model's* REDUCED FORM (ABBREVIATED RF) *we understand the set of equations which arise when we solve the model's equations such that each of its endogenous variables is expressed as a function of the model's exogenous variables (if such variables occur in the model) and the disturbances.*

(ii) Coefficient combinations occurring as intercepts or as coefficients of exogenous variables in a model's reduced form, are denoted as COEFFICIENTS IN THE REDUCED FORM, OR SIMPLY, REDUCED-FORM-COEFFICIENTS. They are impact coefficients (impact multipliers). Confer shift analyses in simple market theory and multipliers in (Keynesian) macro models. Examples of reduced-form-coefficients in the market models in Lecture note E below are, in the case with no exogenous variables:

$$\mu_y = \frac{\beta_1\alpha_2 - \beta_2\alpha_1}{\beta_1 - \beta_2} \quad \text{and} \quad \mu_p = \frac{\alpha_2 - \alpha_1}{\beta_1 - \beta_2},$$

and, in the case with two exogenous variables: $\Pi_{11}, \Pi_{12}, \Pi_{21}.\Pi_{22}$

(iii) *By a* PARAMETER IN THE REDUCED FORM, OR SIMPLY A REDUCED-FORM-PARAMETER*, we understand a reduced-form-coefficient* OR *a parameter describing the distribution of the disturbances in a model's reduced form.* Examples of reduced forms (in cases without and with exogenous variables) for simple market models are given in Lecture note E.

(iv) *Each endogenous variable has one RF equation.* The endogenous variable is its LHS variable. All exogenous variables in the model occur as RHS variables in the RF, in general.

(v) The RF disturbances are all linear combinations of the SF disturbances. The coefficients in these linear combinations are functions of the coefficients of the *endogenous* variables in the model's SF.

(vi) The RF is, formally, *a system of regression equations with the same set of exogenous variables*, viz. all the model's exogenous variables.

(vii) The RF coefficients are non-linear functions of the SF coefficients; cf. (ii) above.

(viii) The RF equations have *lower degree of autonomy* than the SF equations; cf. (ii) above. They are *confluent relationships*, they "flow together", because structural coefficients from several equations are 'mixed up' in the same RF equation.

ECON4160 ECONOMETRICS – MODELLING AND SYSTEMS ESTIMATION

Lecture note E

IDENTIFICATION PROBLEMS 1: SIMPLE MARKET MODELS

Erik Biørn
Department of Economics

Version of July 5, 2011

1. A market model without exogenous variables

Let us consider the market for a single consumption commodity, and let y_t denote the market quantity traded and p_t the market price. The market model has a supply function (S) and a demand function (D) and is:

$$\text{(E-1)} \qquad y_t = \alpha_1 + \beta_1 p_t + u_t, \qquad (S)$$

$$\text{(E-2)} \qquad y_t = \alpha_2 + \beta_2 p_t + v_t, \qquad (D)$$

$$\text{(E-3)} \qquad \mathsf{E}(u_t) = \mathsf{E}(v_t) = 0,$$

$$\text{(E-4)} \qquad \text{var}(u_t) = \sigma_u^2, \quad \text{var}(v_t) = \sigma_v^2, \quad \text{cov}(u_t, v_t) = \sigma_{uv},$$

$$\text{(E-5)} \qquad \text{cov}(u_t, u_s) = \text{cov}(v_t, v_s) = \text{cov}(u_t, v_s) = 0, \quad s \neq t, \quad t, s = 1, \dots, T.$$

The observations generated from it are T points scattered around the point of intersection between the lines $y_t = \alpha_1 + \beta_1 p_t$ and $y_t = \alpha_2 + \beta_2 p_t$. Some more terminology related to simultaneous equation systems in econometrics is given in lecture note D.

The essential question when deciding whether we have identification of the model's parameters (or some of them) and hence are able (in one way or another, not discussed here) to estimate its parameters 'in a proper way', is: *What do we know, from (E-1)–(E-5), about the joint probability distribution of the two observed endogenous variables y_t and p_t?.* This is the fundamental question we seek to answer to decide whether or not we have identification of the equations of this stochastic market model. In order to answer it we first solve (E-1) and (E-2) with respect to the endogenous variables, giving

$$\text{(E-6)} \qquad y_t = \frac{\beta_1 \alpha_2 - \beta_2 \alpha_1}{\beta_1 - \beta_2} + \frac{\beta_1 v_t - \beta_2 u_t}{\beta_1 - \beta_2},$$

$$\text{(E-7)} \qquad p_t = \frac{\alpha_2 - \alpha_1}{\beta_1 - \beta_2} + \frac{v_t - u_t}{\beta_1 - \beta_2}.$$

Utilizing (E-3)–(E-5), we obtain

$$\text{(E-8)} \qquad \mathsf{E}(y_t) = \frac{\beta_1\alpha_2 - \beta_2\alpha_1}{\beta_1 - \beta_2} = \mu_y,$$

$$\text{(E-9)} \qquad \mathsf{E}(p_t) = \frac{\alpha_2 - \alpha_1}{\beta_1 - \beta_2} = \mu_p,$$

$$\text{(E-10)} \qquad \text{var}(y_t) = \frac{\beta_1^2\sigma_v^2 - 2\beta_1\beta_2\sigma_{uv} + \beta_2^2\sigma_u^2}{(\beta_1 - \beta_2)^2} = \sigma_y^2,$$

$$\text{(E-11)} \qquad \text{var}(p_t) = \frac{\sigma_v^2 - 2\sigma_{uv} + \sigma_u^2}{(\beta_1 - \beta_2)^2} = \sigma_p^2,$$

$$\text{(E-12)} \qquad \text{cov}(y_t, p_t) = \frac{\beta_1\sigma_v^2 - (\beta_1 + \beta_2)\sigma_{uv} + \beta_2\sigma_u^2}{(\beta_1 - \beta_2)^2} = \sigma_{yp},$$

where $\mu_y, \mu_p, \sigma_y^2, \sigma_p^2, \sigma_{yp}$ are defined by the last equalities of the five equations.

The simultaneous distribution of the endogenous variables (y_t, p_t) is described by their first- and second-order moments, i.e., the expectations μ_y, μ_p, the variances σ_y^2, σ_p^2, and the covariance σ_{yp}. This is the distribution of our observed (endogenous) variables.

Remark: Strictly speaking, we here make the additional assumption that we *confine attention to the first- and second-order moments of the distribution.* We thus disregard information we might obtain by considering also the higher-order moments of (y_t, p_t). However, if we in addition to (E-3)–(E-5) assume that the simultaneous distribution of (u_t, v_t) is *binormal*, so that also (y_t, p_t) become binormally distributed (as two linear combinations of binormal variables are also binormal), we are unable to obtain any additional information by utilizing the higher-order moments. The reason is that the binormal distribution (and, more generally, the multinormal distribution) is completely described by its first- and second-order moments. On the binomal distribution, see Greene: Appendix B.9.

Hence, what we, under the most favourable circumstances we can imagine, are able to 'extract' from our T observations (y_t, p_t), is to estimate this distribution, i.e., to determine the five parameters

$$\mu_y, \mu_p, \sigma_y^2, \sigma_p^2, \sigma_{yp}.$$

We could, for instance, estimate these (theoretical) moments by the corresponding empirical moments,

$$\bar{y}, \bar{p}, M[y, y], M[p, p], M[y, p],$$

utilizing the convergence in moments property. This exemplifies the so-called *Method of Moments*. If $T \to \infty$, we are able to completely determine this distribution. *This is the most favourable data situation we can imagine.*

Now, the model (E-1)–(E-5) has seven unknown parameters,

$$\alpha_1, \alpha_2, \beta_1, \beta_2, \sigma_u^2, \sigma_v^2, \sigma_{uv}.$$

The crucial question then becomes: *Can we, from knowledge of $\mu_y, \mu_p, \sigma_y^2, \sigma_p^2, \sigma_{yp}$ deduce the values of these seven parameters, while utilizing (E-8)–(E-12)?* The answer is *no* in general; we have too few equations, only five. *In the absence of additional information, we have an identification problem*: Neither the supply nor the demand function is identifiable.

We next present three examples in which we possess such additional information

2. Three examples with identification from additional information

Example 1: Assume that there is *no disturbance in the supply function*: $u_t = 0 \Longrightarrow \sigma_u^2 = \sigma_{uv} = 0$. Then all observations will be different and will lie on the supply function, provided that $\sigma_v^2 > 0$. Intuitively, we should then expect the supply function to be identifiable. Algebra based on (E-8)–(E-12) will confirm this. We find

$$\text{(E-13)} \qquad \beta_1 = \frac{\sigma_y^2}{\sigma_{yp}} = \frac{\sigma_{yp}}{\sigma_p^2},$$

$$\text{(E-14)} \qquad \alpha_1 = \mu_y - \frac{\sigma_y^2}{\sigma_{yp}}\mu_p = \mu_y - \frac{\sigma_{yp}}{\sigma_p^2}\mu_p.$$

The coefficient and the intercept of the supply function can be calculated when knowing the joint distribution of market price and traded quantity. The demand function is not identifiable. Note that in this case, there are two different ways of calculating β_1 and α_1. Even if we now have five equations and five unknowns in (E-8)–(E-12), we are unable to solve for β_2 and α_2. At the same time, there are two ways of expressing α_1 and β_1.

A symmetric example, which you are invited to elaborate, is that in which the demand function does not contain a disturbance.

Example 2: Assume next that our theory tells us that the supply is *price inelastic*, i.e., $\beta_1 = 0$, and that the two disturbances are uncorrelated, i.e., $\sigma_{uv} = 0$. It then follows from (E-8)–(E-12) that

$$\beta_2 = \frac{\sigma_y^2}{\sigma_{yp}},$$

$$\sigma_u^2 = \sigma_y^2, \qquad \sigma_v^2 = \sigma_p^2\left(\frac{\sigma_y^2}{\sigma_{yp}}\right)^2 - \sigma_y^2, \alpha_1 = \mu_y, \qquad \alpha_2 = \mu_y - \frac{\sigma_y^2}{\sigma_{yp}}\mu_p.$$

We are therefore able, when knowing the parameters of the distribution of (y_t, p_t), to compute all the five remaining unknown parameters in the market model. The additional information we have, is crucial.

Since the price coefficient of the demand function emerges as the ratio between the variance of the quantity and the covariance between quantity and price, a useful procedure seems to be to *estimate* β_2 by

$$\widehat{\beta}_2 = \frac{M[y,y]}{M[y,p]}.$$

This corresponds to estimating $1/\beta_2$ by regressing the market price on the quantity traded and afterwards invert the expression to estimate β_2. Using Slutsky's Theorem in combination with convergence in moments, we can show that $\widehat{\beta}_2$ is consistent for β_2.

A symmetric example, which you are invited to elaborate, is that in which the demand side is price inelastic

EXAMPLE 3: We now make the additional assumption that the country we are considering is a *small country* and that the demand side is characterized by the price being given in the world market regardless of the supply conditions in our small country. We then omit (E-2) and replace (E-3)–(E-5) by

$$\mathsf{E}(u_t) = 0, \qquad \text{var}(u_t) = \sigma_u^2, \qquad \text{cov}(p_t, u_t) = 0,$$

using the last assumption to make precise our economic insight that p_t is exogenously determined in the world market. From (E-1) we then obtain

$$\begin{aligned} \mu_y &= \alpha_1 + \beta_1 \mu_p, \\ \sigma_y^2 &= \beta_1^2 \sigma_p^2 + \sigma_u^2, \\ \sigma_{yp} &= \beta_1 \sigma_p^2, \end{aligned}$$

giving

$$\beta_1 = \frac{\sigma_{yp}}{\sigma_p^2}, \qquad \alpha_1 = \mu_y - \frac{\sigma_{yp}}{\sigma_p^2}\mu_p, \qquad \sigma_u^2 = \sigma_y^2 - \left(\frac{\sigma_{yp}}{\sigma_p^2}\right)^2 \sigma_p^2.$$

In this case, the supply function can be identified. We see, furthermore, that the supply price coefficient emerges as the ratio between the covariance between quantity and price and the variance of the price. In this situation, it then can be convenient to estimate β_1 by

$$\widehat{\beta}_1 = \frac{M[y,p]}{M[p,p]}.$$

This corresponds to estimating the supply price coefficient by regressing the traded quantity on the market price. Using Slutsky's Theorem in combination with convergence in moments, we can show that $\widehat{\beta}_1$ is consistent for β_1.

A symmetric example, which you are invited to elaborate, is that in which the supply side is characterized by the small country property.

3. A market model with two exogenous variables

We now consider a more general market model, with the following structural form (SF) equations:

$$\text{(E-15)} \qquad y_t = \alpha_1 + \beta_1 p_t + \gamma_{11} z_t + u_t, \qquad (S)$$

$$\text{(E-16)} \qquad y_t = \alpha_2 + \beta_2 p_t + \gamma_{22} R_t + v_t, \qquad (D)$$

where (E-15) is the supply function (S) and (E-16) is the demand function (D) $(t = 1, \ldots, T)$, and where z_t and R_t are exogenous variables. The observable variables are (y_t, p_t, z_t, R_t). We can think of (E-15) and (E-16) as having been obtained by putting, according to our theory, $\gamma_{12} = \gamma_{21} = 0$ since R_t plays no role for the supply and z_t plays no role for the demand. Further, we assume, reflecting our exogeneity assumptions that

$$\text{(E-17)} \qquad \mathsf{E}(u_t|z_t, R_t) = 0,$$
$$\text{(E-18)} \qquad \mathsf{E}(v_t|z_t, R_t) = 0,$$
$$\text{(E-19)} \qquad \text{var}(u_t|z_t, R_t) = \sigma_u^2,$$
$$\text{(E-20)} \qquad \text{var}(v_t|z_t, R_t) = \sigma_v^2,$$
$$\text{(E-21)} \qquad \text{cov}(u_t, v_t|z_t, R_t) = \sigma_{uv},$$
$$\text{(E-22)} \qquad \text{cov}(u_t, u_s|z_t, R_t, z_s, R_s) = 0, \qquad s \neq t,$$
$$\text{(E-23)} \qquad \text{cov}(v_t, v_s|z_t, R_t, z_s, R_s) = 0, \qquad s \neq t,$$
$$\text{(E-24)} \qquad \text{cov}(u_t, v_s|z_t, R_t, z_s, R_s) = 0, \qquad s \neq t, \quad t, s = 1, \dots, T.$$

More generally (and often more conveniently) we could assume that (E-17)–(E-24) hold conditionally on all z and R values in the data set, $z = (z_1, \dots, z_T)$ and $R = (R_1, \dots, R_T)$.

Using the law of double expectations, we derive

$$\text{(E-25)} \qquad \mathsf{E}(u_t) = \mathsf{E}(v_t) = 0,$$
$$\text{(E-26)} \qquad \text{var}(u_t) = \sigma_u^2, \quad \text{var}(v_t) = \sigma_v^2, \quad \text{cov}(u_t, v_t) = \sigma_{uv},$$
$$\text{(E-27)} \quad \text{cov}(u_t, u_s) = \text{cov}(v_t, v_s) = \text{cov}(u_t, v_s) = 0, \quad s \neq t,$$
$$\text{(E-28)} \quad \text{cov}(z_t, u_t) = \text{cov}(z_t, v_t) = \text{cov}(R_t, u_t) = \text{cov}(R_t, v_t) = 0, \quad t, s = 1, \dots, T.$$

4. Elements of the econometric model description

The description of our structural model (E-15)–(E-24) contains:

(i) A determined system of structural equations with specification of endogenous and exogenous variables, disturbances and unknown structural coefficients,

(ii) Normalization restrictions, exemplified by one of the variables in each equation, y_t, having a known coefficient, set to unity. If no normalization had been imposed, we could have "blown up" all of the equation's coefficients by a common unspecified factor and would have had no chance of identifying any of its coefficients.

(iii) Exclusion restrictions, exemplified by R_t not occurring the supply function ($\gamma_{12} = 0$) and z_t not occurring in the demand function ($\gamma_{21} = 0$).

(iv) A specification of the properties of the disturbance distribution.

5. The structural model expressed in conditional expectations

From (E-15)–(E-18) we obtain

$$\text{(E-29)} \qquad \mathsf{E}(y_t|z_t, R_t) = \alpha_1 + \beta_1 \mathsf{E}(p_t|z_t, R_t) + \gamma_{11} z_t, \qquad (T)$$

$$\text{(E-30)} \qquad \mathsf{E}(y_t|z_t, R_t) = \alpha_2 + \beta_2 \mathsf{E}(p_t|z_t, R_t) + \gamma_{22} R_t, \qquad (E)$$

The model's reduced form (RF) is

$$\text{(E-31)} \qquad y_t = \frac{\beta_1\alpha_2 - \beta_2\alpha_1}{\beta_1 - \beta_2} - \frac{\beta_2\gamma_{11}}{\beta_1 - \beta_2} z_t + \frac{\beta_1\gamma_{22}}{\beta_1 - \beta_2} R_t + \frac{\beta_1 v_t - \beta_2 u_t}{\beta_1 - \beta_2},$$

$$\text{(E-32)} \qquad p_t = \frac{\alpha_2 - \alpha_1}{\beta_1 - \beta_2} - \frac{\gamma_{11}}{\beta_1 - \beta_2} z_t + \frac{\gamma_{22}}{\beta_1 - \beta_2} R_t + \frac{v_t - u_t}{\beta_1 - \beta_2},$$

or, in abbreviated notation,

$$\text{(E-33)} \qquad y_t = \Pi_{10} + \Pi_{11} z_t + \Pi_{12} R_t + \varepsilon_{1t},$$

$$\text{(E-34)} \qquad p_t = \Pi_{20} + \Pi_{21} z_t + \Pi_{22} R_t + \varepsilon_{2t}.$$

Here we have

$$\text{(E-35)} \qquad \varepsilon_{1t} = \frac{\beta_1 v_t - \beta_2 u_t}{\beta_1 - \beta_2},$$

$$\text{(E-36)} \qquad \varepsilon_{2t} = \frac{v_t - u_t}{\beta_1 - \beta_2}$$

and

$$\Pi_{10} = \frac{\beta_1\alpha_2 - \beta_2\alpha_1}{\beta_1 - \beta_2}, \qquad \Pi_{11} = -\frac{\beta_2\gamma_{11}}{\beta_1 - \beta_2}$$

etc. It follows that

$$\text{(E-37)} \qquad \mathsf{E}(\varepsilon_{1t}|z_t, R_t) = 0,$$

$$\text{(E-38)} \qquad \mathsf{E}(\varepsilon_{2t}|z_t, R_t) = 0,$$

which implies that the RF disturbances are uncorrelated with the RHS variables in the RF.

6. The reduced form and the distribution of the endogenous variables conditional on the exogenous variables

Why is the model's RF useful when examining whether or not each of the SF equations are identifiable. The answer is:

WE CAN USE THE MODEL'S RF TO DESCRIBE THE PROBABILITY DISTRIBUTION OF THE MODEL'S (OBSERVABLE) ENDOGENOUS VARIABLES CONDITIONAL ON ITS (OBSERVABLE) EXOGENOUS VARIABLES.

From (E-33)–(E-38) we find

$$\text{(E-39)}\qquad \mathsf{E}(y_t|z_t,R_t) = \Pi_{10} + \Pi_{11}\,z_t + \Pi_{12}\,R_t,$$

$$\text{(E-40)}\qquad \mathsf{E}(p_t|z_t,R_t) = \Pi_{20} + \Pi_{21}\,z_t + \Pi_{22}\,R_t,$$

$$\text{(E-41)}\qquad \begin{aligned}\text{var}(y_t|z_t,R_t) &= \text{var}(\varepsilon_{1t}|z_t,R_t)\\ &= \frac{\beta_1^2\sigma_v^2 - 2\beta_1\beta_2\sigma_{uv} + \beta_2^2\sigma_u^2}{(\beta_1-\beta_2)^2} = \omega_{11},\end{aligned}$$

$$\text{(E-42)}\qquad \begin{aligned}\text{var}(p_t|z_t,R_t) &= \text{var}(\varepsilon_{2t}|z_t,R_t)\\ &= \frac{\sigma_v^2 - 2\sigma_{uv} + \sigma_u^2}{(\beta_1-\beta_2)^2} = \omega_{22},\end{aligned}$$

$$\text{(E-43)}\qquad \begin{aligned}\text{cov}(y_t,p_t|z_t,R_t) &= \text{cov}(\varepsilon_{1t},\varepsilon_{2t}|z_t,R_t)\\ &= \frac{\beta_1\sigma_v^2 - (\beta_1+\beta_2)\sigma_{uv} + \beta_2\sigma_u^2}{(\beta_1-\beta_2)^2} = \omega_{12},\end{aligned}$$

$$\text{(E-44)}\qquad \text{cov}(y_t,y_s|z_t,R_t,z_s,R_s) = \text{cov}(\varepsilon_{1t},\varepsilon_{1s}|z_t,R_t,z_s,R_s) = 0,\quad s\neq t,$$

$$\text{(E-45)}\qquad \text{cov}(p_t,p_s|z_t,R_t,z_s,R_s) = \text{cov}(\varepsilon_{2t},\varepsilon_{2s}|z_t,R_t,z_s,R_s) = 0,\quad s\neq t,$$

$$\text{(E-46)}\qquad \text{cov}(y_t,p_s|z_t,R_t,z_s,R_s) = \text{cov}(\varepsilon_{1t},\varepsilon_{2s}|z_t,R_t,z_s,R_s) = 0,\quad s\neq t,$$

where ω_{11}, ω_{22} and ω_{12} are defined by the last equalities in (E-41)–(E-43).

From our data (y_t,p_t,z_t,R_t), $t=1,\dots,T,$, we can, more or less precisely, determine (estimate) the probability distribution of the model's endogenous variables (y_t,p_t) conditional on its exogenous variables (z_t,R_t). This distribution is characterized by the Π_{ij}'s and the ω_{ij}'s. *We can determine this distribution perfectly if we have an infinite sample. This is the most favourable data situation we can imagine.*

The crucial question we must answer to decide whether we have identification of (S) and (D) therefore reduces to the following: Knowing the Π's and the ω's, are we able to determine the values of the unknown α's, β's, γ's and/or σ's?.

In our example we have six structural coefficients

$$\alpha_1,\alpha_2,\beta_1,\beta_2,\gamma_{11},\gamma_{22},$$

and six reduced-form coefficients,

$$\Pi_{10},\Pi_{11},\Pi_{12},\Pi_{20},\Pi_{21},\Pi_{22}.$$

Furthermore, we have three second-order moments of the SF disturbances

$$\sigma_u^2,\sigma_v^2,\sigma_{uv},$$

and the same number of second-order moments of the RF disturbances,

$$\omega_{11},\omega_{22},\omega_{12}.$$

Given the Π's and the ω's, we thus have nine unknown SF parameters. We also have nine equations, namely those which emerge when utilizing the coefficient equalities between (E-31)–(E-32) and (E-33)–(E-34), giving six equations, and (E-41)–(E-43), giving three equations.

We can state the following:

(i) If no restrictions are imposed on $\sigma_u^2, \sigma_v^2, \sigma_{uv}$, the number of σ's will be equal to the number of ω's.

(ii) Given β_1 and β_2, the relationships between the σ's and the ω's are linear.

Therefore:

(iii) All σ's will certainly be identifiable if all β's are identifiable.

(iv) We may, if we only are interested in the identifiability of the SF COEFFICIENTS, i.e., the α's, the β's and the γ's, disregard all equations in (E-39)–(E-43) which contain the second-order moments, i.e., (E-41)–(E-43), and confine attention to the equations derived from the first-order moments.

7. Further remarks on the identification of the supply function

Consider the model's supply function. Inserting (E-39)–(E-40) into (E-29), we get

$$\Pi_{10} + \Pi_{11} z_t + \Pi_{12} R_t = \alpha_1 + \beta_1(\Pi_{20} + \Pi_{21} z_t + \Pi_{22} R_t) + \gamma_{11} z_t. \tag{E-47}$$

This shall hold as an *identity in the exogenous variables.* Recall that the exogenous variables are determined outside the model, and hence there would have been a violation of the model's logic if (E-47) were to restrict their variation. For this reason, there must be pair-wise equality between the coefficients of the LHS (Left Hand Side) and the RHS (Right hand side) of (E-47). We therefore have:

$$\Pi_{10} = \alpha_1 + \beta_1 \Pi_{20}, \tag{E-48}$$

$$\Pi_{11} = \beta_1 \Pi_{21} + \gamma_{11}, \tag{E-49}$$

$$\Pi_{12} = \beta_1 \Pi_{22}, \tag{E-50}$$

corresponding to respectively, the intercept, z_t, and R_t. If we know the six Π's we can then solve (E-48)–(E-50) and determine the SF coefficients in the supply function as follows:

$$\beta_1 = \frac{\Pi_{12}}{\Pi_{22}}, \quad \gamma_{11} = \Pi_{11} - \frac{\Pi_{12}}{\Pi_{22}} \Pi_{21}, \quad \alpha_1 = \Pi_{10} - \frac{\Pi_{12}}{\Pi_{22}} \Pi_{20}. \tag{E-51}$$

Hence, all the coefficients of the supply function are identifiable.

8. A formal description of the identification problem in general

Our problem is, in general, of the following form:

(i) Let $(q_{1t}, \dots, q_{Nt})$ be N observable endogenous variable and let $(x_{1t}, \dots, x_{Kt})$ be K observable exogenous variables.

(ii) Let furthermore $(\lambda_1, \dots, \lambda_Q)$ be the model's Q unknown SF parameters. Assume, however, that the model, when regard is paid to all of its restrictions, implies that $(q_{1t}, \dots, q_{Nt})$ and $(x_{1t}, \dots, x_{Kt})$ have a simultaneous distribution which could be fully described by the $P + R$ parameters $(\theta_1, \dots, \theta_P)$ and $(\tau_1, \dots, \tau_R)$. We assume that the density of this distribution exists and denote it as $f_t(q_{1t}, \dots, q_{Nt}, x_{1t}, \dots, x_{Kt}; \theta_1, \dots, \theta_P, \tau_1, \dots, \tau_R)$.

(iii) We assume that $(\theta_1, \dots, \theta_P)$ and $(\tau_1, \dots, \tau_R)$ have no elements in common.

(iv) We assume that the joint density of all the exogenous and endogenous variables can be written as the product of a marginal density of the exogenous variables and a conditional density of the endogenous variables on the exogenous variables as follows:

$$\begin{aligned}&f_t(q_{1t}, \dots, q_{Nt}, x_{1t}, \dots, x_{Kt}; \theta_1, \dots, \theta_P, \tau_1, \dots, \tau_R)\\ =\ & h(q_{1t}, \dots, q_{Nt}|x_{1t}, \dots, x_{Kt}; \theta_1, \dots, \theta_P)\phi_t(x_{1t}, \dots, x_{Kt}; \tau_1, \dots, \tau_R),\end{aligned}$$

where $h(\cdot)$ is independent of t. Its form is fully specified when $(\theta_1, \dots, \theta_P)$ is given. The density of the exogenous variables, $\phi_t(\cdot)$, only depends on the parameter vector $(\tau_1, \dots, \tau_R)$, but can change with t.

We now, have an economic theory, stated in our structural model, which implies that the θ's can be expressed by the λ's as follows:

$$\begin{aligned}\theta_1 &= g_1(\lambda_1, \lambda_2, \dots, \lambda_Q),\\ \theta_2 &= g_2(\lambda_1, \lambda_2, \dots, \lambda_Q),\\ &\vdots\\ \theta_P &= g_P(\lambda_1, \lambda_2, \dots, \lambda_Q),\end{aligned} \tag{E-52}$$

where the functions $g_1(\cdot), \dots, g_P(\cdot)$ are *completely known from the theory.* We consequently can compute all θ's when all λ's are given. The fundamental question is: *Can we go the opposite way?* Three answers to this question can be given:

(i) If the parameter λ_i can be computed from (E-52), λ_i is identifiable.

(ii) If λ_i cannot be computed in this way, it is non-identifiable.

(iii) If the parameters $\lambda_1, \dots, \lambda_Q$ are known and can be computed from (E-52) when $\theta_1, \dots, \theta_P$ are known, all the model's SF parameters are identifiable.

(iv) A NECESSARY *condition for all the model's SF parameters being identifiable is that $P \geq Q$, i.e., that the number of equations in (E-52) is at least as large as the number of unknowns.*

So far we have considered the formalities of the general case. Let us concretize by referring to the market model example in Sections 3–7.

We then in particular have $N = 2$, $K = 2$, $P = Q = 9$. Moreover,

$(q_{1t}, q_{2t}, \dots, q_{Nt})$	correspond to	(y_t, p_t),
$(x_{1t}, x_{2t}, \dots, x_{Kt})$	correspond to	(z_t, R_t),
$(\theta_1, \theta_2, \dots, \theta_P)$	correspond to	$(\Pi_{10}, \Pi_{11}, \Pi_{12}, \Pi_{20}, \Pi_{21}, \Pi_{22}, \omega_{11}, \omega_{22}, \omega_{12})$,
$(\lambda_1, \lambda_2, \dots, \lambda_Q)$	correspond to	$(\alpha_1, \alpha_2, \beta_1, \beta_2, \gamma_{11}, \gamma_{22}, \sigma_u^2, \sigma_v^2, \sigma_{uv})$,
$(\tau_1, \tau_2, \dots, \tau_R)$		are unspecified and irrelevant.

The equation system (E-52) corresponds to the systems of coefficient restrictions which follows by comparing (E-31)–(E-32) with (E-33)–(E-34) and from (E-41)–(E-43), that is

$$\Pi_{10} = \frac{\beta_1\alpha_2 - \beta_2\alpha_1}{\beta_1 - \beta_2}, \quad \Pi_{11} = -\frac{\beta_2\gamma_{11}}{\beta_1 - \beta_2}, \quad \omega_{11} = \frac{\beta_1^2\sigma_v^2 - 2\beta_1\beta_2\sigma_{uv} + \beta_2^2\sigma_u^2}{(\beta_1 - \beta_2)^2}$$

etc.

We now want to obtain a rather general criterion for examining the identification status of an equation, without having to be involved in solving a system like (E-52) each time.

9. The order condition for identification

Consider an equation in a simultaneous model and let:

K = No. of *exogenous* variables in the complete *model*.
N = No. of *endogenous* variables in the complete *model*.
K_1 = No. of *exogenous* variables included in the actual *equation*.
N_1 = No. of *endogenous* variables included in the actual *equation*.
K_2 = No. of *exogenous* variables excluded from the actual *equation*.
N_2 = No. of *endogenous* variables excluded from the actual *equation*.

We have

$$K_1 + K_2 = K, \qquad N_1 + N_2 = N. \tag{E-53}$$

Let H denote the number of equations between the RF coefficients and the SF coefficients in the actual equation, of the form (E-48)–(E-50). This number is one more than the number of exogenous variables, as the intercept term also occupies one equation, i.e.,

$$H = K + 1.$$

In the structural equation considered, we have

$$M_N = N_1 - 1$$

unknown coefficients of the included endogenous variables, as one coefficient must be deducted because of the normalization restriction. In the structural equation considered, we have

$$M_K = K_1 + 1$$

unknown coefficients belonging to the exogenous variables including the intercept term.

The order condition for identification of the equation considered is the condition that we have a sufficient number of equations between the coefficients. It can be formulated as

$$H \geq M_N + M_K \quad \Longleftrightarrow \quad N_1 + K_1 \leq K + 1. \tag{E-54}$$

Using (E-53), we find that the order condition is equivalent to the following condition, which is easier to remember:

$$K_2 + N_2 \geq N - 1, \tag{E-55}$$

i.e.,

$$\left\{ \begin{array}{c} \textit{The total number of variables excluded from the equation} \\ \textit{should be at least as large as} \\ \textit{the number of equations in the model minus 1.} \end{array} \right\}$$

Note that:

1. The order condition is necessary for identification when no restrictions have been imposed on the SF disturbance variances and covariances. There exists a corresponding rank condition (knowledge of this is not required in this course) which is both necessary and sufficient.

2. If at least one restriction has been imposed on these variances and covariances, then the order and rank conditions are no longer necessary, and the problem should be examined specifically. Simple examples of the latter kind of examination are Examples 1–3 in Section 2.

10. Indirect Least Squares (ILS) estimation in exactly identified and overidentified equations: Examples and exercises

Consider a simple model of a commodity market with SF (Structural Form) equations

$$y_t = \alpha_1 + \beta_1 p_t + \gamma_{11} z_t + u_{1t}, \tag{E-56}$$

$$y_t = \alpha_2 + \beta_2 p_t + \gamma_{22} R_t + \gamma_{23} q_t + u_{2t}, \tag{E-57}$$

where (E-56) is the supply function and (E-57) is the demand function. The endogenous variables are (y_t, p_t), the exogenous variable are (z_t, R_t, q_t). The SF disturbances are (u_{1t}, u_{2t}). Its RF (Reduced Form) equations are

$$y_t = \Pi_{10} + \Pi_{11} z_t + \Pi_{12} R_t + \Pi_{13} q_t + \varepsilon_{1t}, \tag{E-58}$$

$$p_t = \Pi_{20} + \Pi_{21} z_t + \Pi_{22} R_t + \Pi_{23} q_t + \varepsilon_{2t}, \tag{E-59}$$

where

$$y_t = \frac{\beta_1\alpha_2 - \beta_2\alpha_1}{\beta_1 - \beta_2} - \frac{\beta_2\gamma_{11}}{\beta_1 - \beta_2} z_t + \frac{\beta_1\gamma_{22}}{\beta_1 - \beta_2} R_t + \frac{\beta_1\gamma_{23}}{\beta_1 - \beta_2} q_t + \frac{\beta_1 u_{2t} - \beta_2 u_{1t}}{\beta_1 - \beta_2}, \tag{E-60}$$

$$p_t = \frac{\alpha_2 - \alpha_1}{\beta_1 - \beta_2} - \frac{\gamma_{11}}{\beta_1 - \beta_2} z_t + \frac{\gamma_{22}}{\beta_1 - \beta_2} R_t + \frac{\gamma_{23}}{\beta_1 - \beta_2} q_t + \frac{u_{2t} - u_{1t}}{\beta_1 - \beta_2}. \tag{E-61}$$

From (E-56)–(E-57) and (E-58)–(E-59), utilizing the zero conditional expectations for the disturbances, we find

$$\mathsf{E}(y_t|z_t, R_t, q_t) = \alpha_1 + \beta_1 \mathsf{E}(p_t|z_t, R_t, q_t) + \gamma_{11} z_t, \tag{E-62}$$

$$\mathsf{E}(y_t|z_t, R_t, q_t) = \alpha_2 + \beta_2 \mathsf{E}(p_t|z_t, R_t, q_t) + \gamma_{22} R_t + \gamma_{23} q_t, \tag{E-63}$$

$$\mathsf{E}(y_t|z_t, R_t, q_t) = \Pi_{10} + \Pi_{11}\, z_t + \Pi_{12}\, R_t + \Pi_{13}\, q_t, \tag{E-64}$$

$$\mathsf{E}(p_t|z_t, R_t, q_t) = \Pi_{20} + \Pi_{21}\, z_t + \Pi_{22} R_t + \Pi_{23}\, q_t. \tag{E-65}$$

Inserting (E-64)–(E-65) into (E-62) and into (E-63), we get, respectively

$$\text{(E-66)}\quad \Pi_{10} + \Pi_{11} z_t + \Pi_{12} R_t + \Pi_{13} q_t = \alpha_1 + \beta_1(\Pi_{20} + \Pi_{21} z_t + \Pi_{22} R_t + \Pi_{23} q_t) + \gamma_{11} z_t,$$

$$\text{(E-67)}\quad \Pi_{10} + \Pi_{11} z_t + \Pi_{12} R_t + \Pi_{13} q_t = \alpha_2 + \beta_2(\Pi_{20} + \Pi_{21} z_t + \Pi_{22} R_t + \Pi_{23} q_t) + \gamma_{22} R_t + \gamma_{23} q_t.$$

THESE EQUATIONS SHALL HOLD AS IDENTITIES IN THE EXOGENOUS VARIABLES. Recall that the exogenous variables are determined outside the model, and hence there would have been a violation of the logic of the structural model we have postulated if (E-66) and (E-67) had restricted their variation. For this reason, there must be pair-wise equality between the coefficients of the LHS and the RHS of (E-66) and those of (E-67).

We therefore have from equation (E-66) [We confine attention to this; equation (E-67) can be treated similarly – you may do this as an exercise]

$$\text{(E-68)}\quad \begin{aligned} \Pi_{10} &= \alpha_1 + \beta_1 \Pi_{20}, \\ \Pi_{11} &= \beta_1 \Pi_{21} + \gamma_{11}, \\ \Pi_{12} &= \beta_1 \Pi_{22}, \\ \Pi_{13} &= \beta_1 \Pi_{23}, \end{aligned}$$

corresponding to, respectively, the intercept, z_t, R_t, and q_t. This implies

$$\text{(E-69)}\quad \begin{aligned} \beta_1 &= \frac{\Pi_{12}}{\Pi_{22}} = \frac{\Pi_{13}}{\Pi_{23}}, \\ \gamma_{11} &= \Pi_{11} - \frac{\Pi_{12}}{\Pi_{22}} \Pi_{21} = \Pi_{11} - \frac{\Pi_{13}}{\Pi_{23}} \Pi_{21}, \\ \alpha_1 &= \Pi_{10} - \frac{\Pi_{12}}{\Pi_{22}} \Pi_{20} = \Pi_{10} - \frac{\Pi_{13}}{\Pi_{23}} \Pi_{20}. \end{aligned}$$

Remark: Equations (E-68) and (E-69) can also be obtained by comparing (E-58)–(E-59) with (E-60)–(E-61).

EXERCISE 1: Assume that we know that $\gamma_{23} = 0$. Then the supply function (E-56) is EXACTLY IDENTIFIED. Why will this imply $\Pi_{13} = \Pi_{23} = 0$? An OLS estimation of (E-58) and (E-59) has given [The example has been constructed!]

$$\begin{aligned} &\widehat{\Pi}_{10} = 265, \quad \widehat{\Pi}_{11} = 2.1, \quad \widehat{\Pi}_{12} = 58.5, \\ &\widehat{\Pi}_{20} = 5, \quad \widehat{\Pi}_{21} = -0.3, \quad \widehat{\Pi}_{22} = 4.5. \end{aligned}$$

Interpret these estimates. Show, by using the relationships between the SF coefficients $(\alpha_1, \alpha_2, \beta_1, \beta_2, \gamma_{11}, \gamma_{22})$ and the RF coefficients $(\Pi_{10}, \Pi_{11}, \Pi_{12}, \Pi_{20}, \Pi_{21}, \Pi_{22})$ obtained from (E-68)–(E-69), and the corresponding equations relating to equation (E-57), that the derived estimates of the SF coefficients in the supply function are

$$\begin{aligned} &\widehat{\alpha}_1 = 200, \quad \widehat{\beta}_1 = 13, \quad \widehat{\gamma}_{11} = 6, \\ &\widehat{\alpha}_2 = 300, \quad \widehat{\beta}_2 = -7, \quad \widehat{\gamma}_{22} = 90. \end{aligned}$$

Interpret these estimates. Are the estimators of the RF coefficients (i) unbiased, (ii) consistent? Are the estimators of the SF coefficients(i) unbiased, (ii) consistent?

The estimation method just described exemplifies the Indirect Least Squares (ILS) method.

ILS means: First estimate the RF coefficients by applying OLS on each RF equation separately. Next derive estimators of the SF coefficients by exploiting the relationships between the the RF coefficients and the the SF coefficients, as exemplified by (E-68)–(E-69).

Exercise 2: Assume that $\gamma_{23} \neq 0$ and unknown. Then the supply function (E-56) is OVERIDENTIFIED. Why will this imply $\Pi_{13} \neq 0, \Pi_{23} \neq 0$? An OLS estimation of (E-58) and (E-59) has given [Again, the example has been constructed]

$$\widehat{\Pi}_{10} = 265, \quad \widehat{\Pi}_{11} = 2.1, \quad \widehat{\Pi}_{12} = 58.5, \quad \widehat{\Pi}_{13} = 18,$$
$$\widehat{\Pi}_{20} = 5, \quad \widehat{\Pi}_{21} = -0.3, \quad \widehat{\Pi}_{22} = 4.5, \quad \widehat{\Pi}_{23} = 2.$$

Interpret these estimates. Show, by using the relationships between the SF coefficients $(\alpha_1, \alpha_2, \beta_1, \beta_2, \gamma_{11}, \gamma_{22}, \gamma_{23})$ and the RF coefficients $(\Pi_{10}, \Pi_{11}, \Pi_{12}, \Pi_{13}, \Pi_{20}, \Pi_{21}, \Pi_{22}, \Pi_{23})$ obtained from (E-68)–(E-69) that we get two sets of ILS estimates of the SF coefficients in (E-56), denoted as (a) and (b):

$$(a) \qquad \widehat{\alpha}_1 = 200, \quad \widehat{\beta}_1 = 13, \quad \widehat{\gamma}_{11} = 6,$$
$$(b) \qquad \widetilde{\alpha}_1 = 210, \quad \widetilde{\beta}_1 = 9, \quad \widetilde{\gamma}_{11} = 4.8,$$

and, by a similar argument, one set of ILS estimates of the SF coefficients in (E-57):

$$(a*) \qquad \widehat{\alpha}_2 = 300, \quad \widehat{\beta}_2 = -7, \quad \widehat{\gamma}_{22} = 90, \quad \widehat{\gamma}_{23} = 32.$$

Interpret these estimates.

Exercise 3:

[1] Explain why we get two sets of estimates of the coefficients in (E-56) in Exercise 2, but only one set of estimates in Exercise 1.

[2] Explain why we get one set of estimates of the coefficients in (E-57) in both Exercise 2 and Exercise 1.

[3] Are the estimators of the RF coefficients (i) unbiased, (ii) consistent? Are the estimators of the SF coefficients (i) unbiased, (ii) consistent?

[4] How would (E-66)–(E-69) have been changed and how would the estimation of (E-56) have proceeded if we had NOT known (from economic theory) that R_t and q_t were excluded from this equation?

ECON4160 ECONOMETRICS – MODELLING AND SYSTEMS ESTIMATION

Lecture note F:

IDENTIFICATION PROBLEMS 2: MEASUREMENT ERROR MODELS

Erik Biørn
Department of Economics

Version of July 5, 2011

In reading the lecture note, you may take advantage in reading it in parallel with **Lecture note E**, section 8.

1. Model with random measurement error in the exogenous variable

We specify the following deterministic relationship between the endogenous variable Y_i^* and the exogenous variable X_i^*:

$$Y_i^* = \alpha + \beta X_i^*, \qquad i = 1, \dots, n, \tag{F-1}$$

where α and β are unknown coefficients. Neither Y_i^* nor X_i^* is observable. The observations are given by

$$Y_i = Y_i^* + u_i, \tag{F-2}$$

$$X_i = X_i^* + v_i, \qquad i = 1, \dots, n. \tag{F-3}$$

We denote u_i as the measurement error in Y_i and v_i as the measurement error in X_i. We further denote Y_i^* and X_i^* as *latent structural variables.* A concrete interpretation may be: Y_i, Y_i^*, and u_i are, respectively, observed, permanent (normal), and transitory (irregular) consumption, and X_i, X_i^*, and v_i are, respectively, observed, permanent (normal), and transitory (irregular) income (windfall gains). The basic hypotheses then are that neither permanent, nor transitory consumption respond to transitory income.

Eliminating Y_i^* from (F-1) by using (F-2) we get

$$Y_i = \alpha + \beta X_i^* + u_i, \qquad i = 1, \dots, n. \tag{F-4}$$

We may interpret u_i as also including a disturbance in (F-1).

We assume that

$$\mathsf{E}(u_i|X_i^*) = \mathsf{E}(v_i|X_i^*) = 0, \qquad i = 1, \dots, n. \tag{F-5}$$

Regardless of which value X_i^* takes, the measurement error has zero expectation. On average, we measure X correctly, regardless of how large the value we want to measure is. It follows from (F-5) that

$$\mathsf{E}(u_i) = \mathsf{E}(v_i) = 0, \tag{F-6}$$

$$\mathrm{cov}(u_i, X_i^*) = \mathrm{cov}(v_i, X_i^*) = 0, \qquad i = 1, \dots, n. \tag{F-7}$$

We further assume

$$\begin{array}{llll}
\text{(F-8)} & \operatorname{var}(u_i) = \sigma_u^2, & \operatorname{cov}(u_i, u_j) = 0, \quad j \neq i, & \\
\text{(F-9)} & \operatorname{var}(v_i) = \sigma_v^2, & \operatorname{cov}(v_i, v_j) = 0, \quad j \neq i, & \\
\text{(F-10)} & \operatorname{cov}(u_i, v_i) = 0, & \operatorname{cov}(u_i, v_j) = 0, \quad j \neq i, & i, j = 1, \dots, n.
\end{array}$$

We eliminate Y_i^* and X_i^* from (F-1) by using (F-2)–(F-3) and get

$$Y_i - u_i \;=\; \alpha + \beta(X_i - v_i),$$

i.e.,

$$\text{(F-11)} \qquad Y_i \;=\; \alpha + \beta X_i + u_i - \beta v_i \;=\; \alpha + \beta X_i + w_i, \qquad i = 1, \dots, n,$$

where

$$\text{(F-12)} \qquad w_i \;=\; u_i - \beta v_i, \qquad i = 1, \dots, n.$$

Which are the properties of the composite error w_i? Using (F-5)–(F-10), we find

$$\begin{aligned}
\mathsf{E}(w_i) &= \mathsf{E}(w_i|X_i^*) = \mathsf{E}(w_i|Y_i^*) = 0, \\
\operatorname{var}(w_i) &= \mathsf{E}(w_i^2) = \sigma_u^2 + \beta^2\sigma_v^2, \\
\operatorname{cov}(w_i, w_j) &= \mathsf{E}(w_i w_j) = 0, \quad j \neq i, \quad i, j = 1, \dots, n,
\end{aligned}$$

i.e., w_i has zero expectation, varies randomly in relation to X_i^* and Y_i^*, and is homoskedastic and non-autocorrelated. Further,

$$\begin{array}{lrcl}
\text{(F-13)} & \operatorname{cov}(w_i, X_i^*) &=& \mathsf{E}(w_i X_i^*) = \mathsf{E}[(u_i - \beta v_i)X_i^*] = 0, \\
\text{(F-14)} & \operatorname{cov}(w_i, Y_i^*) &=& \mathsf{E}(w_i Y_i^*) = \mathsf{E}[(u_i - \beta v_i)Y_i^*] = 0.
\end{array}$$

But, we have

$$\begin{array}{lrcl}
\text{(F-15)} & \operatorname{cov}(w_i, X_i) &=& \mathsf{E}(w_i X_i) \;=\; \mathsf{E}[(u_i - \beta v_i)(X_i^* + v_i)] \;=\; -\beta\sigma_v^2, \\
\text{(F-16)} & \operatorname{cov}(w_i, Y_i) &=& \mathsf{E}(w_i Y_i) \;=\; \mathsf{E}[(u_i - \beta v_i)(Y_i^* + u_i)] \;=\; \sigma_u^2.
\end{array}$$

Thus, the composite error in the equation between the observed variables, (F-11), is correlated with both the LHS and the RHS variables: $\operatorname{cov}(w_i, Y_i) > 0$, while $\operatorname{cov}(w_i, X_i)$ has a sign which is the opposite of the sign of β, provided that $\sigma_u^2 > 0$ and $\sigma_v^2 > 0$.

2. The probability distribution of the observable variables

Let us express the (theoretical) first- and second-order moments of X_i and Y_i by means of the model's structural parameters and the parameters in the distribution of the latent exogenous variable. From (F-1)–(F-3) and (F-6) we find

$$\begin{array}{lrcl}
\text{(F-17)} & \mathsf{E}(X_i) &=& \mathsf{E}(X_i^*), \\
\text{(F-18)} & \mathsf{E}(Y_i) &=& \mathsf{E}(Y_i^*) = \alpha + \beta\mathsf{E}(X_i^*).
\end{array}$$

Further, (F-1)–(F-3) and (F-7)–(F-10) imply

$$
\begin{aligned}
&\text{(F-19)} \qquad & \text{var}(X_i) &= \text{var}(X_i^* + v_i) = \text{var}(X_i^*) + \sigma_v^2, \\
&\text{(F-20)} \qquad & \text{var}(Y_i) &= \text{var}(Y_i^* + u_i) = \text{var}(Y_i^*) + \sigma_u^2 = \beta^2 \text{var}(X_i^*) + \sigma_u^2, \\
&\text{(F-21)} \qquad & \text{cov}(X_i, Y_i) &= \text{cov}(Y_i^* + u_i, X_i^* + v_i) \\
& & &= \text{cov}(Y_i^*, X_i^*) + \text{cov}(u_i, v_i) = \beta \text{var}(X_i^*) = \frac{1}{\beta} \text{var}(Y_i^*).
\end{aligned}
$$

3. Identification

We first make the following assumption about the distribution of the model's latent, exogenous variable, X_i^*: *The expectation and the variance is the same for all observation units.* Precisely,

$$
\begin{aligned}
&\text{(F-22)} \qquad & \mathsf{E}(X_i^*) &= \mu_{X*}, \\
&\text{(F-23)} \qquad & \text{var}(X_i^*) &= \sigma_{X*}^2, \qquad i = 1, \ldots, n.
\end{aligned}
$$

This is stronger type of assumptions than we normally use in regression models and simultaneous equation systems, and they may not always be realistic. The model obtained sometimes is referred to as a *structural model*, because we impose a structure on the probability distribution of the exogenous variable.

It follows from (F-22)–(F-23) and (F-17)–(F-21) that the distribution of the observable variables is described by

$$
\begin{aligned}
&\text{(F-24)} \qquad & \mathsf{E}(X_i) &= \mu_{X*} = \mu_X, \\
&\text{(F-25)} \qquad & \mathsf{E}(Y_i) &= \alpha + \beta\mu_{X*} = \mu_Y, \qquad i = 1, \ldots, n,
\end{aligned}
$$

where μ_X and μ_Y are defined by the two last equalities, and

$$
\begin{aligned}
&\text{(F-26)} \qquad & \text{var}(X_i) &= \sigma_{X*}^2 + \sigma_v^2 = \sigma_X^2, \\
&\text{(F-27)} \qquad & \text{var}(Y_i) &= \beta^2\sigma_{X*}^2 + \sigma_u^2 = \sigma_Y^2, \\
&\text{(F-28)} \qquad & \text{cov}(X_i, Y_i) &= \beta\sigma_{X*}^2 = \sigma_{XY}, \qquad i = 1, \ldots, n,
\end{aligned}
$$

where σ_X^2, σ_Y^2 and σ_{XY} are defined by the three last equalities. The information we, at most, can 'extract' from a set of observations (Y_i, X_i), $i = 1, \ldots, n$, is to determine this distribution completely. Confining attention to the first-order and second-order moments, this means that we, at most, can determine the following five first- and second-order moments

$$\mu_X,\ \mu_Y,\ \sigma_X^2,\ \sigma_Y^2,\ \sigma_{XY}.$$

They can, for instance, be estimated by $(\bar{X}, \bar{Y}, M[X,X], M[Y,Y], M[X,Y])$, utilizing convergence in moments. However, the model contains the following six structural parameters:

$$\alpha,\ \beta,\ \mu_{X*},\ \sigma_u^2,\ \sigma_v^2,\ \sigma_{X*}^2.$$

They are connected with the five first- and second-order moments of the observable variables via (F-24)–(F-28).

It follows from (F-24) that μ_{X*} is always identifiable. Using (F-24) to eliminate μ_{X*}, we get the following system with four equations:

$$\text{(F-29)}\qquad \begin{aligned} \mu_Y &= \alpha + \beta\mu_X, \\ \sigma_X^2 &= \sigma_{X*}^2 + \sigma_v^2, \\ \sigma_Y^2 &= \beta^2\sigma_{X*}^2 + \sigma_u^2, \\ \sigma_{XY} &= \beta\sigma_{X*}^2. \end{aligned}$$

It determines the possibilities for identification of the remaining five structural parameters, α, β, σ_{X*}^2, σ_u^2 and σ_v^2. The mathematical interpretation of the identification problem is that we only have four equations in five unknowns, when considering μ_Y, μ_X, σ_Y^2, σ_X^2 and σ_{XY}^2 as known entities.

4. Identification under additional information. Five examples

In all the following examples, we impose one and only one additional restriction and show that can then solve the identification problem.

Example 1: The intercept α is known. Then (F-29) gives

$$\begin{aligned} \beta &= \frac{\mu_Y - \alpha}{\mu_X}, \\ \sigma_{X*}^2 &= \sigma_{XY}\frac{\mu_X}{\mu_Y - \alpha}, \\ \sigma_v^2 &= \sigma_X^2 - \sigma_{XY}\frac{\mu_X}{\mu_Y - \alpha}, \\ \sigma_u^2 &= \sigma_Y^2 - \sigma_{XY}\frac{\mu_Y - \alpha}{\mu_X}. \end{aligned}$$

If, in particular, $\alpha = 0$, i.e., the structural equation goes through the origin, then

$$\beta = \frac{\mu_Y}{\mu_X},$$

etc. We have full identification.

Question 1: How would you estimate β consistently in the last case?

Example 2: The measurement error variance of X, σ_v^2, is known. Then (F-29) gives

$$\begin{aligned} \beta &= \frac{\sigma_{XY}}{\sigma_X^2 - \sigma_v^2}, \\ \alpha &= \mu_Y - \frac{\sigma_{XY}}{\sigma_X^2 - \sigma_v^2}\mu_X, \\ \sigma_{X*}^2 &= \sigma_X^2 - \sigma_v^2, \\ \sigma_u^2 &= \sigma_Y^2 - \frac{\sigma_{XY}^2}{\sigma_X^2 - \sigma_v^2}. \end{aligned}$$

If, in particular, $\sigma_v^2 = 0$, i.e., X_i has no measurement error, then

$$\beta = \frac{\sigma_{XY}}{\sigma_X^2},$$

etc. We again have full identification.

Question 2: How would you estimate β consistently in the last case?

Example 3: The measurement error variance of Y, σ_u^2, is known. Then (F-29) gives

$$\begin{aligned} \beta &= \frac{\sigma_Y^2 - \sigma_u^2}{\sigma_{XY}}, \\ \alpha &= \mu_Y - \frac{\sigma_Y^2 - \sigma_u^2}{\sigma_{XY}}\mu_X, \\ \sigma_{X*}^2 &= \frac{\sigma_{XY}^2}{\sigma_Y^2 - \sigma_u^2}, \\ \sigma_v^2 &= \sigma_X^2 - \frac{\sigma_{XY}^2}{\sigma_Y^2 - \sigma_u^2}. \end{aligned}$$

If, in particular, $\sigma_u^2 = 0$, i.e., Y_i has no measurement error (and the equation has no disturbance), then

$$\beta \;=\; \frac{\sigma_Y^2}{\sigma_{XY}},$$

etc. We again have full identification.

QUESTION 3: How would you estimate β consistently in the last case?

EXAMPLE 4: THE RATIO BETWEEN THE MEASUREMENT ERROR VARIANCE AND THE VARIANCE OF THE TRUE X IS KNOWN, AND EQUAL TO c: $\sigma_v^2 = c\,\sigma_{X*}^2$. This is equivalent to assuming that the measurement error variance is $c/(1+c)$ times the variance of the observed X. From (F-29) it follows that

$$\begin{aligned}
\beta &= \frac{\sigma_{XY}}{\sigma_X^2}(1+c),\\
\alpha &= \mu_Y - \frac{\sigma_{XY}}{\sigma_X^2}(1+c)\mu_X,\\
\sigma_{X*}^2 &= \frac{1}{1+c}\sigma_X^2,\\
\sigma_v^2 &= \frac{c}{1+c}\sigma_X^2,\\
\sigma_u^2 &= \sigma_Y^2 - \frac{\sigma_{XY}^2}{\sigma_X^2}(1+c).
\end{aligned}$$

We again have full identification.

QUESTION 4: How would you estimate β consistently in this case if you knew *a priori* that $c = 0.1$?

EXAMPLE 5: THE RATIO BETWEEN THE MEASUREMENT ERROR AND THE VARIANCE OF THE TRUE Y IS KNOWN AND EQUAL TO d: $\sigma_u^2 = d\,\beta^2\sigma_{X*}^2$. This is equivalent to assuming that the measurement error variance is $d/(1+d)$ times the variance of the observed Y. From (F-29) it now follows that

$$\begin{aligned}
\beta &= \frac{\sigma_Y^2}{\sigma_{XY}}\frac{1}{1+d},\\
\alpha &= \mu_Y - \frac{\sigma_Y^2}{\sigma_{XY}}\frac{1}{1+d}\mu_X,\\
\sigma_{X*}^2 &= \frac{\sigma_{XY}^2}{\sigma_Y^2}(1+d),\\
\sigma_v^2 &= \sigma_X^2 - \frac{\sigma_{XY}^2}{\sigma_Y^2}(1+d),\\
\sigma_u^2 &= \frac{d}{1+d}\sigma_Y^2.
\end{aligned}$$

We again have full identification.

QUESTION 5: How would you estimate β consistently in this case if you knew *a priori* that $d = 0.2$?

IMPORTANT REMINDER: THE **order condition** FOR IDENTIFICATION DEVELOPED IN **Lecture note E**, SECTION 9, PRESUMES THAT ALL (STRUCTURAL) VARIABLES ARE OBSERVABLE. THEREFORE, IT CANNOT BE USED IN EXAMINING THE IDENTIFICATION STATUS OF PARAMETERS IN MEASUREMENT ERROR MODELS!