ECON 5101 ADVANCED ECONOMETRICS – TIME SERIES

Lecture note no. 6 (EB)

Erik Biørn,
Department of Economics

Version of February 14, 2011

ESTIMATION OF ARX AND VARX MODELS BY ML

This lecture note, related to **Topic 3** in the lecture plan, extends LN 3, Secs. 3 & 4. In Part A we consider a single-equation model, in Part B an interdependent model.

PART A: ESTIMATION OF AN ARX(1) MODEL

ARX(1) Model

We formulate the following ARX(1) model – using ARX(1) as an acronym for *AR(1) model with exogenous variables:*

(A.1) $Y_t = \beta X_t + \phi Y_{t-1} + c + \varepsilon_t \iff (1-\phi \mathsf{L})Y_t = \beta X_t + c + \varepsilon_t,\ |\phi| < 0, (\varepsilon_t|X_t) \sim \mathsf{IID}(0, \sigma_\varepsilon^2).$

Here β and X_t may well be extended to vectors and X_t may contain lagged values of (lag-distributions on) exogenous variables. We assume normally distributed disturbances, assuming that $\varepsilon_t = Y_t - \beta X_t - \phi Y_{t-1} - c$ has density function

$$g(\varepsilon_t; \sigma_\varepsilon) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma_\varepsilon} \exp\left\{-\frac{\varepsilon_t^2}{2\sigma_\varepsilon^2}\right\}. \tag{A.2}$$

The joint density of the observed values of $Y_T, Y_{T-1}, \ldots, Y_1, Y_0$, when we let (y_t, x_t) denote the observed value of (Y_t, X_t), and

$$\boldsymbol{x} = (x_0, x_1, \ldots, x_T),$$

can be written as the product of conditional densities as follows,

$$\begin{aligned} f(y_T, y_{T-1}, \ldots, y_1, y_0|\boldsymbol{x}) = {} & f_0(y_0|x_0) f_1(y_1|y_0, x_1) f_2(y_2|y_1, y_0, x_2) \cdots \\ & \times f_{T-1}(y_{T-1}|y_{T-2} \cdots y_2, y_1, y_0, x_{T-1}) \\ & \times f_T(y_T|y_{T-1} \cdots y_2, y_1, y_0, x_T). \end{aligned} \tag{A.3}$$

Now, for any $t = 1, \ldots, T$, conditioning y_t on $(y_{t-1}, \ldots, y_2, y_1, y_0, \boldsymbol{x})$ is equivalent to conditioning only on (y_{t-1}, x_t). Moreover, there is a one-to-one transformation from $(y_t|y_{t-1}, x_t)$ to ε_t. This, together with (A.2), implies

$$\begin{aligned} f_t(y_t|y_{t-1}, \ldots, y_2, y_1, y_0, \boldsymbol{x}) &= f_t(y_t|y_{t-1}, x_t) = g(y_t - \beta x_t - \phi y_{t-1} - c; \sigma_\varepsilon) \\ &\equiv \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma_\varepsilon} \exp\left\{-\frac{(y_t - \beta x_t - \phi y_{t-1} - c)^2}{2\sigma_\varepsilon^2}\right\}, \end{aligned} \tag{A.4}$$

$$(Y_t|Y_{t-1} = y_{t-1}, X_t = x_t) \sim \mathsf{N}\left(\beta x_t + \phi y_{t-1} + c, \sigma_\varepsilon^2\right), \qquad t = 1, \ldots, T. \tag{A.5}$$

Exact ML estimation

Now (A.1) implies, after inversion, $Y_0 = c/(1-\phi) + \beta X_0 + \sum_{i=1}^{\infty} \phi^i(\beta X_{-i} + \varepsilon_{-i})$. Hence,

$$(Y_0|X_0 = x_0) \sim \mathsf{N}\left(\beta(x_0 + \mu_{\phi x}) + \frac{c}{1-\phi}, \beta^2\sigma_{\phi x}^2 + \frac{\sigma_\varepsilon^2}{1-\phi^2}\right), \tag{A.6}$$

assuming, for simplicity (since $X_{-1}, X_{-2}, \dots$), are unobserved $[\sum_{i=1}^{\infty} \phi^i \beta X_{-i}|x_0] \sim \mathsf{N}(\mu_{\phi x}, \sigma_{\phi x}^2)$. This implies

$$\begin{aligned} f_0(y_0|x_0) &= h(y_0; \beta, \phi, c, \mu_{\phi x}, \sigma_{\phi x}, \sigma_\varepsilon) \\ &\equiv \frac{1}{\sqrt{2\pi}} \frac{1}{[\beta^2\sigma_{\phi x}^2 + \sigma_\varepsilon^2/(1-\phi^2)]^{\frac{1}{2}}} \exp\left\{-\frac{y_0 - \beta(x_0 + \mu_{\phi x}) - c/(1-\phi)}{2[\beta^2\sigma_{\phi x}^2 + \sigma_\varepsilon^2/(1-\phi^2)}\right\}. \end{aligned} \tag{A.7}$$

Using (A.3), (A.4), and (A.7) the full likelihood function can be written as

$$\mathcal{L}(\beta, \phi, c, \mu_{\phi x}, \sigma_{\phi x}, \sigma_\varepsilon) = h(y_0; \beta, \phi, c, \mu_{\phi x}, \sigma_{\phi x}, \sigma_\varepsilon) \prod_{t=1}^{T} g(y_t - \beta x_t - \phi y_{t-1} - c; \sigma_\varepsilon), \tag{A.8}$$

and the log-likelihood function becomes

$$\begin{aligned} \ln[\mathcal{L}(\beta, \phi, c, \mu_{\phi x}, \sigma_{\phi x}, \sigma_\varepsilon)] &= \ln[h(y_0; \beta, \phi, c, \mu_{\phi x}, \sigma_{\phi x}, \sigma_\varepsilon)] \\ &\quad + \sum_{t=1}^{T} \ln[g(y_t - \beta x_t - \phi y_{t-1} - c; \sigma_\varepsilon)]. \end{aligned} \tag{A.9}$$

♣ *The maximum of (A.8), or (A.9), with respect to the unknown parameters, defines the (exact) ML estimators.*

Conditional ML estimation

If we want to condition inference on y_0 and $\boldsymbol{x}$, we may instead of (A.3) *consider the density conditional on* y_0,

$$\begin{aligned} f(y_T, y_{T-1}, \dots, y_1|y_0, \boldsymbol{x}) = f_1(y_1|y_0, x_1) f_2(y_2|y_1, y_0, x_2) \cdots \\ f_{T-1}(y_{T-1}|y_{T-2}, \dots, y_2, y_1, y_0, x_{T-1}) \\ f_T(y_T|y_{T-1}, \dots, y_2, y_1, y_0, x_T) \end{aligned} \tag{A.10}$$

Then the conditional likelihood and the log-likelihood can be written as, respectively

$$\mathcal{L}^*(\beta, \phi, c, \sigma_\varepsilon) = \prod_{t=1}^{T} g(y_t - \phi y_{t-1} - c; \sigma_\varepsilon), \tag{A.11}$$

$$\ln[\mathcal{L}^*(\beta, \phi, c, \sigma_\varepsilon)] = \sum_{t=1}^{T} \ln[g(y_t - \beta x_t - \phi y_{t-1} - c; \sigma_\varepsilon)]. \tag{A.12}$$

♣ *The maximum of (A.11), or (A.12), defines the conditional (approximate) ML estimators.*

♣ It is not difficult to show that $\ln[\mathcal{L}^*(\beta, \phi, c, \sigma_\varepsilon)]$ is a monotonically decreasing function of the sum of squares based on the ARX(1) equation

$$\sum_{t=1}^{T} (y_t - \beta x_t - \phi y_{t-1} - c)^2,$$

which implies – as for the strict AR(1) model – that conditional ML estimation coincides, for the intercept and slope coefficients, with OLS estimation.

♣ The finite sample bias problem (generalizing the Hurwicz bias) remains, not only for ϕ, but also for β (because Y_{t-1} and X_t are usually correlated).

PART B: ESTIMATION OF A VARX SYSTEM BY THE ML

The model

We consider a VAR (Vector AutoRegressive) model of the first order in the endogenous N vector $\boldsymbol{y}_t$ with exogenous variables in the K vector $\boldsymbol{x}_t$ (including a one for the intercept). It may also be called a VARX model, X indicating again the inclusion of exogenous variables. We further assume that, according to economic theory, coefficient restrictions have been imposed on the system so that all N equations are identified and write the equation system for observation t as follows:

$$\text{(B.1)} \qquad \boldsymbol{B}_0\boldsymbol{y}_t = \boldsymbol{B}_1\boldsymbol{y}_{t-1} + \boldsymbol{\Gamma}\boldsymbol{x}_t + \boldsymbol{u}_t, \quad t = 1, \dots, T,$$

where $\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_T)$, $\boldsymbol{B}_0$ and $\boldsymbol{B}_1$ are $(N \times N)$, $\boldsymbol{\Gamma}$ is $(N \times K)$, $\boldsymbol{y}_t$ and $\boldsymbol{u}_t$ are $(N \times 1)$, and $\boldsymbol{x}_t$ is $(K \times 1)$. Furthermore $\boldsymbol{B}_0$ is non-singular. The disturbance vector $\boldsymbol{u}_t$ has the following properties:

$$\text{(B.2)} \qquad \mathsf{E}(\boldsymbol{u}_t|\boldsymbol{X}, \boldsymbol{y}_{t-1}, \boldsymbol{y}_{t-2}, \dots, \boldsymbol{y}_0) = 0_{N,1},$$

$$\text{(B.3)} \qquad \mathsf{E}(\boldsymbol{u}_t\boldsymbol{u}_t'|\boldsymbol{X}, \boldsymbol{y}_{t-1}, \boldsymbol{y}_{t-2}, \dots, \boldsymbol{y}_0) = \boldsymbol{\Sigma},$$

where $\boldsymbol{\Sigma}$ is $(N \times N)$ and positive definite, and

$$\text{(B.4)} \qquad \boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_T \ \text{ are independent and normally distributed.}$$

Assumptions (B.1)–(B.4) imply that $(\boldsymbol{u}_t|\boldsymbol{X}, \boldsymbol{y}_{t-1}, \boldsymbol{y}_{t-2}, \dots, \boldsymbol{y}_0)$ has density function

$$\text{(B.5)} \qquad f(\boldsymbol{u}_t|\boldsymbol{X}, \boldsymbol{y}_{t-1}, \boldsymbol{y}_{t-2}, \dots, \boldsymbol{y}_0) = (2\pi)^{-\frac{N}{2}}|\boldsymbol{\Sigma}|^{-\frac{1}{2}}e^{-\frac{1}{2}\boldsymbol{u}_t'\boldsymbol{\Sigma}^{-1}u_t}, \ \ t = 1, \dots, T,$$

♣ *In order to establish the Maximum Likelihood (ML) method, we are, in general, interested in the probability distribution of the observations on the model's endogenous variables, in the present case* $\boldsymbol{y}_t\,(t = 1, \dots, T)$, *conditional on the observations on the model's exogenous variables, in the present case* $\boldsymbol{x}_t\,(t = 1, \dots, T)$, *and the initial value of the endogenous variable,* $\boldsymbol{y}_0$. *We want to estimate the unknown parameters in the matrices* $(\boldsymbol{B}_0, \boldsymbol{B}_1, \boldsymbol{\Gamma}, \boldsymbol{\Sigma})$ *by maximizing the probability density of this joint distribution with respect to these unknown parameters, given the observations on* $\boldsymbol{y}_t$ *and* $\boldsymbol{x}_t$ $(t = 1, \dots, T)$.

The conditional distribution of the current endogenous variables

Our first problem therefore is to establish the conditional distribution of $\boldsymbol{y}_t$ by means of (B.1)–(B.5). It follows from (B.1) that $(\boldsymbol{y}_t|\boldsymbol{X}, \boldsymbol{y}_{t-1}, \boldsymbol{y}_{t-2}, \dots, \boldsymbol{y}_0)$ is a linear function of $(\boldsymbol{u}_t|\boldsymbol{X}, \boldsymbol{y}_{t-1}, \boldsymbol{y}_{t-2}, \dots, \boldsymbol{y}_0)$. Therefore, $(\boldsymbol{y}_t|\boldsymbol{X}, \boldsymbol{y}_{t-1}, \boldsymbol{y}_{t-2}, \dots, \boldsymbol{y}_0)$ also follows a multinormal distribution. From the reduced form of the model,

$$\text{(B.6)} \qquad \boldsymbol{y}_t = \boldsymbol{\Pi}_1\boldsymbol{y}_{t-1} + \boldsymbol{\Pi}_0\boldsymbol{x}_t + \boldsymbol{\epsilon}_t,$$

where

$$\text{(B.7)} \qquad \boldsymbol{\Pi}_1 = \boldsymbol{B}_0^{-1}\boldsymbol{B}_1, \ \ \boldsymbol{\Pi}_0 = \boldsymbol{B}_0^{-1}\boldsymbol{\Gamma}, \ \ \boldsymbol{\epsilon}_t = \boldsymbol{B}_0^{-1}\boldsymbol{u}_t,$$

we know that the conditional expectation vector and the conditional covariance matrix of $\boldsymbol{y}_t$ are, respectively

$$\text{(B.8)} \quad \mathsf{E}(\boldsymbol{y}_t|\boldsymbol{X}, \boldsymbol{y}_{t-1}, \boldsymbol{y}_{t-2}, \dots, \boldsymbol{y}_0) = \boldsymbol{\Pi}_1\boldsymbol{y}_{t-1} + \boldsymbol{\Pi}_0\boldsymbol{x}_t,$$

$$\text{(B.9)} \quad \mathsf{V}(\boldsymbol{y}_t|\boldsymbol{X}, \boldsymbol{y}_{t-1}, \boldsymbol{y}_{t-2}, \dots, \boldsymbol{y}_0) = \mathsf{V}(\boldsymbol{\epsilon}_t|\boldsymbol{X}, \boldsymbol{y}_{t-1}, \boldsymbol{y}_{t-2}, \dots, \boldsymbol{y}_0) = \boldsymbol{\Omega}, \qquad t = 1, \dots, T,$$

where

$$\text{(B.10)} \qquad \boldsymbol{\Omega} = \boldsymbol{B}_0^{-1}\boldsymbol{\Sigma}(\boldsymbol{B}_0')^{-1} \iff \boldsymbol{\Sigma} = \boldsymbol{B}_0\boldsymbol{\Omega}\boldsymbol{B}_0'.$$

Hence, $(\boldsymbol{y}_t|\boldsymbol{X}, \boldsymbol{y}_{t-1}, \boldsymbol{y}_{t-2}, \dots, \boldsymbol{y}_0)$ has density function

$$\text{(B.11)} \qquad \begin{aligned} g(\boldsymbol{y}_t|\boldsymbol{X}, \boldsymbol{y}_{t-1}, \boldsymbol{y}_{t-2}, \dots, \boldsymbol{y}_0) &= (2\pi)^{-\frac{N}{2}}|\boldsymbol{\Omega}|^{-\frac{1}{2}} \\ &\times \exp\{-\tfrac{1}{2}(\boldsymbol{y}_t - \boldsymbol{\Pi}_1\boldsymbol{y}_{t-1} - \boldsymbol{\Pi}_0\boldsymbol{x}_t)'\boldsymbol{\Omega}^{-1}(\boldsymbol{y}_t - \boldsymbol{\Pi}_1\boldsymbol{y}_{t-1} - \boldsymbol{\Pi}_0\boldsymbol{x}_t)\}, \end{aligned}$$

using $\exp\{z\}$ as an abbreviation for e^z when the exponent is a complex expression. It further follows from (B.1)–(B.4) that

$$\text{(B.12)} \quad (\boldsymbol{y}_1|\boldsymbol{X}, \boldsymbol{y}_0), (\boldsymbol{y}_2|\boldsymbol{X}, \boldsymbol{y}_1, \boldsymbol{y}_0), \dots, (\boldsymbol{y}_T|\boldsymbol{X}, \boldsymbol{y}_{T-1}, \dots, \boldsymbol{y}_1, \boldsymbol{y}_0) \ \text{ are stoch. independent.}$$

We have, in (B.11) expressed the density function of $(\boldsymbol{y}_t|\boldsymbol{X}, \boldsymbol{y}_{t-1}, \boldsymbol{y}_{t-2}, \dots, \boldsymbol{y}_0)$ by means of the matrices of reduced form parameters $(\boldsymbol{\Pi}_0, \boldsymbol{\Pi}_1, \boldsymbol{\Omega})$.

♣ *We want, however, to express this density function not by means of the matrices of reduced form parameters* $(\boldsymbol{\Pi}_0, \boldsymbol{\Pi}_1, \boldsymbol{\Omega})$*, but by means of the matrices of structural parameters* $(\boldsymbol{B}_0, \boldsymbol{B}_1, \boldsymbol{\Gamma}, \boldsymbol{\Sigma})$*.*

The relationships between the two sets of parameters are given by (B.7) and (B.10). From these equations we obtain

$$\begin{aligned} \boldsymbol{\Omega}^{-1} &= \boldsymbol{B}_0'\boldsymbol{\Sigma}^{-1}\boldsymbol{B}_0, \\ |\boldsymbol{\Sigma}| &= |\boldsymbol{B}_0\boldsymbol{\Omega}\boldsymbol{B}_0'| = |\boldsymbol{B}_0||\boldsymbol{\Omega}||\boldsymbol{B}_0'| = |\boldsymbol{\Omega}||\boldsymbol{B}_0|^2. \end{aligned}$$

In the exponent of the density function (B.11) the following quadratic form occurs:

$$\text{(B.13)} \qquad S_t = (\boldsymbol{y}_t - \boldsymbol{\Pi}_1\boldsymbol{y}_{t-1} - \boldsymbol{\Pi}_0\boldsymbol{x}_t)'\boldsymbol{\Omega}^{-1}(\boldsymbol{y}_t - \boldsymbol{\Pi}_1\boldsymbol{y}_{t-1} - \boldsymbol{\Pi}_0\boldsymbol{x}_t).$$

Inserting for $\boldsymbol{\Pi}_0$, $\boldsymbol{\Pi}_1$, and $\boldsymbol{\Omega}^{-1}$ in (B.13) yields

$$\text{(B.14)} \quad \begin{aligned} S_t &= (\boldsymbol{y}_t - \boldsymbol{B}_0^{-1}\boldsymbol{B}_1\boldsymbol{y}_{t-1} - \boldsymbol{B}_0^{-1}\boldsymbol{\Gamma}\boldsymbol{x}_t)'\boldsymbol{B}_0'\boldsymbol{\Sigma}^{-1}\boldsymbol{B}_0(\boldsymbol{y}_t - \boldsymbol{B}_0^{-1}\boldsymbol{B}_1\boldsymbol{y}_{t-1} - \boldsymbol{B}_0^{-1}\boldsymbol{\Gamma}\boldsymbol{x}_t) \\ &= [\boldsymbol{B}_0(\boldsymbol{y}_t - \boldsymbol{B}_0^{-1}\boldsymbol{B}_1\boldsymbol{y}_{t-1} - \boldsymbol{B}_0^{-1}\boldsymbol{\Gamma}\boldsymbol{x}_t)]'\boldsymbol{\Sigma}^{-1}[\boldsymbol{B}_0(\boldsymbol{y}_t - \boldsymbol{B}_0^{-1}\boldsymbol{B}_1\boldsymbol{y}_{t-1} - \boldsymbol{B}_0^{-1}\boldsymbol{\Gamma}\boldsymbol{x}_t)], \end{aligned}$$

i.e.,

$$\text{(B.15)} \qquad S_t = [\boldsymbol{B}_0\boldsymbol{y}_t - \boldsymbol{B}_1\boldsymbol{y}_{t-1} - \boldsymbol{\Gamma}\boldsymbol{x}_t]'\boldsymbol{\Sigma}^{-1}[\boldsymbol{B}_0\boldsymbol{y}_t - \boldsymbol{B}_1\boldsymbol{y}_{t-1} - \boldsymbol{\Gamma}\boldsymbol{x}_t].$$

We insert (B.15) and the determinant expression $|\boldsymbol{\Omega}| = |\boldsymbol{B}_0|^{-2}|\boldsymbol{\Sigma}|$ into (B.11), which implies that $(\boldsymbol{y}_t|\boldsymbol{X}, \boldsymbol{y}_{t-1}, \boldsymbol{y}_{t-2}, \dots, \boldsymbol{y}_0)$ has density function

$$\text{(B.16)} \quad \begin{aligned} g(\boldsymbol{y}_t|\boldsymbol{X}, \boldsymbol{y}_{t-1}, \boldsymbol{y}_{t-2}, \dots, \boldsymbol{y}_0) &= (2\pi)^{-\frac{N}{2}}|\boldsymbol{B}_0||\boldsymbol{\Sigma}|^{-\frac{1}{2}} \\ &\times \exp\{-\tfrac{1}{2}[\boldsymbol{B}_0\boldsymbol{y}_t - \boldsymbol{B}_1\boldsymbol{y}_{t-1} - \boldsymbol{\Gamma}\boldsymbol{x}_t)]'\boldsymbol{\Sigma}^{-1}[\boldsymbol{B}_0\boldsymbol{y}_t - \boldsymbol{B}_1\boldsymbol{y}_{t-1} - \boldsymbol{\Gamma}\boldsymbol{x}_t)]\}. \end{aligned}$$

The likelihood function and the ML problem

We now have, in (B.16), derived the conditional density function of the vector of endogenous variables in one single period, period t, given the predetermined variables in this period. In view of (B.12), the corresponding density function of all sets of observations for the T periods can be derived by multiplying the observation specific conditional density functions of all the T observations in the data set. This follows from the fact that the density function of $(\boldsymbol{y}_1,\dots,\boldsymbol{y}_T|\boldsymbol{X},\boldsymbol{y}_0)$, denoted as $G(\boldsymbol{y}_1,\dots,\boldsymbol{y}_T|\boldsymbol{X},\boldsymbol{y}_0)$, can be factorized into conditional densities as

(B.17)

$$
\begin{aligned}
G(\boldsymbol{y}_1,\dots,\boldsymbol{y}_T|\boldsymbol{X},\boldsymbol{y}_0) &= g(\boldsymbol{y}_T|\boldsymbol{X},\boldsymbol{y}_{T-1},\boldsymbol{y}_{T-2},\dots,\boldsymbol{y}_0)G(\boldsymbol{y}_1,\dots,\boldsymbol{y}_{T-1}|\boldsymbol{X},\boldsymbol{y}_0)\\
&= g(\boldsymbol{y}_T|\boldsymbol{X},\boldsymbol{y}_{T-1},\boldsymbol{y}_{T-2},\dots,\boldsymbol{y}_0)g(\boldsymbol{y}_{T-1}|\boldsymbol{X},\boldsymbol{y}_{T-2},\boldsymbol{y}_{T-3},\dots,\boldsymbol{y}_0)\\
&\qquad\qquad\qquad\qquad \times G(\boldsymbol{y}_1,\dots,\boldsymbol{y}_{T-2}|\boldsymbol{X},\boldsymbol{y}_0)\\
&= \quad\cdots\quad =\\
&= g(\boldsymbol{y}_T|\boldsymbol{X},\boldsymbol{y}_{T-1},\boldsymbol{y}_{T-2},\dots,\boldsymbol{y}_0)g(\boldsymbol{y}_{T-1}|\boldsymbol{X},\boldsymbol{y}_{T-2},\boldsymbol{y}_{T-3},\dots,\boldsymbol{y}_0)\\
&\qquad\qquad\qquad \times\cdots\times g(\boldsymbol{y}_2|\boldsymbol{X},\boldsymbol{y}_1,\boldsymbol{y}_0)g(\boldsymbol{y}_1|\boldsymbol{X},\boldsymbol{y}_0)\\
&= \prod_{t=1}^{T} g(\boldsymbol{y}_t|\boldsymbol{X},\boldsymbol{y}_{t-1},\boldsymbol{y}_{t-2},\dots,\boldsymbol{y}_0).
\end{aligned}
$$

Combining (B.16) and (B.17) gives the following joint density function of $(\boldsymbol{y}_1,\dots,\boldsymbol{y}_T|\boldsymbol{X},\boldsymbol{y}_0)$, denoted as the *likelihood function*:

(B.18)
$$
\begin{aligned}
\mathcal{L} = \mathcal{L}(\boldsymbol{y}_1,\dots,\boldsymbol{y}_T|\boldsymbol{X},\boldsymbol{y}_0;\boldsymbol{B}_0,\boldsymbol{B}_1,\boldsymbol{\Gamma},\boldsymbol{\Sigma}) &= \prod_{t=1}^{T} g(\boldsymbol{y}_t|\boldsymbol{X},\boldsymbol{y}_{t-1},\boldsymbol{y}_{t-2},\dots,\boldsymbol{y}_0)\\
= (2\pi)^{-\frac{NT}{2}}|\boldsymbol{B}_0|^T|\boldsymbol{\Sigma}|^{-\frac{T}{2}}&\\
\times \exp\Big[-\tfrac{1}{2}\sum_{t=1}^{T}(\boldsymbol{B}_0\boldsymbol{y}_t-\boldsymbol{B}_1\boldsymbol{y}_{t-1}-\boldsymbol{\Gamma}\boldsymbol{x}_t)'\boldsymbol{\Sigma}^{-1}&(\boldsymbol{B}_0\boldsymbol{y}_t-\boldsymbol{B}_1\boldsymbol{y}_{t-1}-\boldsymbol{\Gamma}\boldsymbol{x}_t)\Big].
\end{aligned}
$$

We can now describe the ML problem for our simultaneous dynamic model as follows:

♣ *Maximize $\mathcal{L}$ – defined in (B.18), with the relevant restrictions on the structural parameters in $(\boldsymbol{B}_0,\boldsymbol{B}_1,\boldsymbol{\Gamma},\boldsymbol{\Sigma})$ imposed – with respect to the unknown parameters in these four matrices, given the observations of $(\boldsymbol{y}_t,\boldsymbol{x}_t)$ for $t=1,\dots,T$. If this maximization problem has a unique solution, the corresponding parameter values are the ML estimators of the unknown structural parameters.*

Taking the (natural) logarithm of $\mathcal{L}$ we obtain the *log-likelihood-function*

(B.19)
$$
\begin{aligned}
\ln(\mathcal{L}) &= \ln\mathcal{L}(\boldsymbol{y}_1,\dots,\boldsymbol{y}_T|\boldsymbol{X},\boldsymbol{y}_0;\boldsymbol{B}_0,\boldsymbol{B}_1,\boldsymbol{\Gamma},\Omega)\\
&= \textstyle\sum_{t=1}^{T} g(\boldsymbol{y}_t|\boldsymbol{X},\boldsymbol{y}_{t-1},\boldsymbol{y}_{t-2},\dots,\boldsymbol{y}_0)\\
&= -\tfrac{NT}{2}\ln(2\pi)+T\ln|\boldsymbol{B}_0|-\tfrac{T}{2}\ln|\boldsymbol{\Sigma}|\\
&\quad -\tfrac{1}{2}\textstyle\sum_{t=1}^{T}[\boldsymbol{B}_0\boldsymbol{y}_t-\boldsymbol{B}_1\boldsymbol{y}_{t-1}-\boldsymbol{\Gamma}\boldsymbol{x}_t)]'\boldsymbol{\Sigma}^{-1}[\boldsymbol{B}_0\boldsymbol{y}_t-\boldsymbol{B}_1\boldsymbol{y}_{t-1}-\boldsymbol{\Gamma}\boldsymbol{x}_t)].
\end{aligned}
$$

Maximizing $\ln(\mathcal{L})$ is usually less complicated mathematically than maximizing $\mathcal{L}$, and the two problems are equivalent since the logarithm function is monotonically increasing. Notice that $\ln(\mathcal{L})$ is a rather complicated function of $\boldsymbol{B}_0$ in general.

APPENDIX:
DERIVED MODEL FORMS AND DYNAMIC MULTIPLIERS

The model's structural form

$$B_0\underline{y}_t = B_1\,\underline{y}_{t-1} + \Gamma\,\underline{x}_t + \underline{u}_t. \tag{C.1}$$

The model's reduced form

$$\underline{y}_t = B_0^{-1}B_1\,\underline{y}_{t-1} + B_0^{-1}\Gamma\,\underline{x}_t + B_0^{-1}\underline{u}_t, \tag{C.2}$$

or

$$\underline{y}_t = \Pi_1\,\underline{y}_{t-1} + \Pi_0\,\underline{x}_t + \underline{\varepsilon}_t. \tag{C.3}$$

The model's final form

Repeated substitution backwards in (C.3) yields

$$\begin{aligned} \underline{y}_t &= \Pi_1^{T_0}\underline{y}_{t-T_0} + \sum_{i=0}^{T_0-1}\Pi_1^i\Pi_0\underline{x}_{t-i} + \sum_{i=0}^{T_0-1}\Pi_1^i\underline{\varepsilon}_{t-i} \\ &= \Pi_1^{T_0}\underline{y}_{t-T_0} + \Pi_0\underline{x}_t + \sum_{i=1}^{T_0-1}\Pi_1^i\Pi_0\underline{x}_{t-i} + \underline{\varepsilon}_t + \sum_{i=1}^{T_0-1}\Pi_1^i\underline{\varepsilon}_{t-i}, \end{aligned} \tag{C.4}$$

where we utilize that $\Pi_1^0 = I_N$.

Assume that

$$\lim_{T_0\to\infty}\Pi_1^{T_0} = \lim_{T_0\to\infty}\left[B_0^{-1}B_1\right]^{T_0} = 0_{NN}. \tag{C.5}$$

It can be shown that

$$\left\{\begin{array}{c} \textit{The convergence condition (C.5) is satisfied} \\ \textit{if and only if all } N \textit{ eigenvalues of } \Pi_1 \\ \textit{are inside the unit circle.} \end{array}\right\} \tag{C.6}$$

The eigenvalues of Π_1 are the N (real or complex) solution values for λ in the N'th degree-equation

$$|\Pi_1 - \lambda I_N| = 0. \tag{C.7}$$

If (C.5) holds and $T_0 \to \infty$, we find from (C.3) that

$$\underline{y}_t = \Pi_0\underline{x}_t + \sum_{i=1}^{\infty}\Pi_1^i\Pi_0\underline{x}_{t-i} + \underline{\varepsilon}_t + \sum_{i=1}^{\infty}\Pi_1^i\underline{\varepsilon}_{t-i}. \tag{C.8}$$

If (C.5) holds the VAR-process is stable. The matrix equation (C.8) is the *final form of the system.* It shows how the endogenous vector $\underline{y}_t$ depends on *the entire history* of the exogenous vector, $\underline{x}_t$ and the disturbance vector.

Dynamic multipliers

$$\left\{\begin{array}{c} \Pi_0 \textit{ is the}(N\times(K+1))\textit{-matrix of} \\ \textbf{short term (or impact) multipliers} \\ \textit{showing the effect of } \underline{x}_t \textit{ on } \underline{y}_t. \end{array}\right\} \tag{C.9}$$

$$\left\{\begin{array}{c} \textit{The } (N\times(K+1))\textit{-matrices of} \\ \textbf{intermediate (interim) multipliers} \\ \textit{showing the effect of } \underline{x}_{t-i} \textit{ on } \underline{y}_t\textit{, are} \\ \Phi_i = \Pi_1^i\Pi_0, \quad i = 1,2,3,\dots\,. \end{array}\right\} \tag{C.10}$$

$$
\text{(C.11)} \qquad \left\{ \begin{array}{c} \textit{The } (N \times (K+1))\textit{-matrices of} \\ \textbf{cumulative intermediate (interim) multipliers} \\ \textit{after } \theta \textit{ periods have passed, are} \\ \Psi_\theta = \Pi_0 + \sum_{i=1}^{\theta-1} \Phi_i = \Pi_0 + \sum_{i=1}^{\theta-1} \Pi_1^i \Pi_0, \quad \theta = 1, 2, \dots . \end{array} \right\}
$$

The definitions (C.9)–(C.11) are valid regardless of whether the eigenvalue condition (C.6) is satisfied or not. Assume that

$$
\text{(C.12)} \qquad \lim_{\theta \to \infty} \Psi_\theta = \lim_{\theta \to \infty} \sum_{i=0}^{\theta-1} \Pi_1^i \Pi_0,
$$

exists. Since $\Psi_\theta \Pi_0^{-1} = \sum_{i=0}^{\theta-1} \Pi_1^i$, this requires that (C.6) is satisfied. The matrix sequence $\Psi_1 \Pi_0^{-1}, \Psi_2 \Pi_0^{-1}, \Psi_3 \Pi_0^{-1}, \dots$, and hence also the sequence $\Psi_1, \Psi_2, \Psi_3, \dots$, will converge.

Lastly we define, provided that (C.12) holds

$$
\text{(C.13)} \qquad \left\{ \begin{array}{c} \textit{The } (N \times (K+1))\textit{-matrix of} \\ \textbf{long terms multipliers is} \\ \Pi = \Psi_\infty = \Pi_0 + \sum_{i=1}^{\infty} \Phi_i = \Pi_0 + \sum_{i=1}^{\infty} \Pi_1^i \Pi_0 . \end{array} \right\}
$$

Recall that

$$
I_N + \Pi_1 + \Pi_1^2 + \Pi_1^3 + \cdots = \sum_{i=0}^{\infty} \Pi_1^i = (I_N - \Pi_1)^{-1}
$$

provided that (C.6) is satisfied.

The proof is $(I_N + \Pi_1 + \Pi_1^2 + \Pi_1^3 + \cdots)(I_N - \Pi_1) = I_N$, which can be demonstrated by multiplying the expressions in the two parentheses. Consequently,

$$
\text{(C.14)} \qquad \Pi \;=\; (I_N - \Pi_1)^{-1} \Pi_0 \;\text{ provided that }\; \lim_{T_0 \to \infty} \Pi_1^{T_0} \;=\; 0_{NN}.
$$

It also follows that

$$
\Pi = (I_N - B_0^{-1} B_1)^{-1} B_0^{-1} \Gamma = \left[B_0^{-1} (B_0 - B_1) \right]^{-1} B_0^{-1} \Gamma = (B_0 - B_1)^{-1} B_0 \, B_0^{-1} \Gamma,
$$

provided that B_0 and $B_0 - B_1$ are non-singular. Consequently,

$$
\text{(C.15)} \qquad \Pi \;=\; (B_0 - B_1)^{-1} \, \Gamma \;\text{ provided that }\; \lim_{T_0 \to \infty} \left[B_0^{-1} B_1 \right]^{T_0} \;=\; 0_{NN}.
$$

It could also have been obtained from the reduced form of the static counterpart to the model's structural form, viz.

$$
(B_0 - B_1) \, \underline{y}_t \;=\; \Gamma \, \underline{x}_t \;+\; \underline{u}_t .
$$

The model's autoregressive (ARMAX) form

We write (C.1) as

$$
\text{(C.16)} \qquad B(\mathsf{L}) \, \underline{y}_t \;=\; \Gamma \, \underline{x}_t \;+\; \underline{u}_t ,
$$

where $B(\mathsf{L})$ is a $(N \times N)$-matrix given by

$$
\text{(C.17)} \qquad B(\mathsf{L}) = B_0 - B_1 \, \mathsf{L} = \begin{bmatrix} \beta_{11}(\mathsf{L}) & \beta_{12}(\mathsf{L}) & \cdots & \beta_{1N}(\mathsf{L}) \\ \beta_{21}(\mathsf{L}) & \beta_{22}(\mathsf{L}) & \cdots & \beta_{2N}(\mathsf{L}) \\ \vdots & \vdots & & \vdots \\ \beta_{N1}(\mathsf{L}) & \beta_{N2}(\mathsf{L}) & \cdots & \beta_{NN}(\mathsf{L}) \end{bmatrix} .
$$

The typical elements (i, j) is a scalar which is linear in the lag-operator, given by

$$
\text{(C.18)} \qquad \beta_{ij}(\mathsf{L}) \;=\; \beta_{0ij} - \beta_{1ij} \mathsf{L}, \qquad i, j = 1, \dots, N.
$$

The matrix of long term multipliers (C.15) can then be written as

$$\Pi = B(1)^{-1}\Gamma, \tag{C.19}$$

since

$$B(1) = B_0 - B_1 = \begin{bmatrix} \beta_{11}(1) & \cdots & \beta_{1N}(1) \\ \vdots & & \vdots \\ \beta_{N1}(1) & \cdots & \beta_{NN}(1) \end{bmatrix}, \tag{C.20}$$

where

$$\beta_{ij}(1) \;=\; \beta_{0ij} - \beta_{1ij}, \qquad i,j = 1,\ldots,N. \tag{C.21}$$

We have, from the definition of the inverse of a matrix that

$$B^{-1} \;=\; \frac{B^*}{|B|},$$

where B^* is the adjoint matrix of B and $|B|$ is its determinant value, which is a scalar. In analogy, for a matrix whose elements are polynomials in the lag-operator, given by (C.17), we define its inverseas

$$B^{-1}(\mathsf{L}) \;=\; \frac{B^*(\mathsf{L})}{|B(\mathsf{L})|}, \tag{C.22}$$

where $B^*(\mathsf{L})$ is the adjoint of $B(\mathsf{L})$.

Let us now premultiply (C.16) by the adjoint of $B(\mathsf{L})$. This gives the following transformation of (C.1)

$$B^*(\mathsf{L})\,B(\mathsf{L})\underline{y}_t \;=\; B^*(\mathsf{L})\,\Gamma\,\underline{x}_t \;+\; B^*(\mathsf{L})\underline{u}_t. \tag{C.23}$$

Since it follows from (C.22) that $B^*(\mathsf{L}) = |B(\mathsf{L})|\,B^{-1}(\mathsf{L})$, (C.23) is equivalent to

$$|B(\mathsf{L})|\underline{y}_t \;=\; B^*(\mathsf{L})\,\Gamma\,\underline{x}_t \;+\; B^*(\mathsf{L})\underline{u}_t. \tag{C.24}$$

This is the model's *autoregressive form*, or its *ARMAX-form. It has the following notable properties*

(i) In eq. i only the i'th endogenous variable and values of it lagged N periods occur on its left hand side, i.e. eq. i contains an AR(N)-process in the i'th endogenous variable.

(ii) All N equations have the same AR part, represented by $|B(\mathsf{L})|$.

(iii) All N equations have ARMAX-form, since (a) the elements of $B^*(\mathsf{L})\underline{u}_t$, are MA(1)-processes, and (b) both current and lagged values of the exogenous variables occur on the right hand side.

Note: The derivation of the ARMAX-form, does not impose any requirements on the eigenvalues of $\Pi_1 = B_0^{-1}B_1$. It suffices for its existence that $|B(\mathsf{L})| \neq 0$.

We can thus derive the final form from the strucural form by either going (a) via the reduced form, cf. (C.1)–(C.8), or (b) via the ARMAX form, cf. (C.16)–(C.23). Of the four model forms only the final form requires the eigenvalue condition to be satisfied.

Suggested further readings (in Norwegian): Erik Biørn: *Økonometriske emner. En videreføring.* Unipub, 2008, Chapters 5 and 6.