ECON 4160, Spring term 2014. Lecture 5

Ragnar Nymoen

University of Oslo

12 September 2014

Some references to Lecture 5

- HN Ch 12 and 14, mainly.
  Ch 13, or equivalent from other books, as self study: Standard mis-specification tests of time series models.
- DM Ch 13.
- (BN 2014, kap 6,7)

A time series of order p, AR(p) I

- In Lecture 4, we motivated the AR(1) model by appealing to the idea that conditional independence can be "created" by conditioning on Y_{t-1}.
- As a direct generalization, conditional independence my require conditioning on p lags.
- We write a time series model of order p as the stochastic difference equation

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} + \varepsilon_t \qquad (1)$$

where ϕ_0 $(j = 0, 2, \ldots, p)$ are parameters, and

$$\varepsilon_t \overset{D}{=} N(0, \sigma_\varepsilon^2) \ \forall t. \qquad (2)$$

A time series of order p, AR(p) II

A weaker model formulation is that ε_t is **white-noise**, conditional on the p lags of Y_t.

- (1) may be of interest "on its own", as a general model of single time series.
- One example is when Y_t is not a an observable variable, but a residual from OLS estimation.
  - In that interpretation (1) becomes a model of autocorrelated regression residuals, as covered in introductory econometrics courses, see also §13.3.1 in HN.
  - Estimate by NLS or feasible GLS, possibly iterated.
- When Y_t is an observable, the main motivation for using (1) is for *forecasting*.

A time series of order p, AR(p) III

- The reason for studying (1) in econometics is however, more fundamental: It gives the framework for defining the all important concepts of **dynamic stability** and **stationarity** both for individual time series and for systems of variables (for example dynamic stochastic general equilibrium models,DSGE).

AR(p) as the final equation of a system I

- We often study systems of stochastic difference equations
- The simplest case is two time series that are connected in a the first order system, without intercepts to save notation.

$$\begin{pmatrix} Y_t \\ X_t \end{pmatrix} = \begin{pmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \end{pmatrix} \begin{pmatrix} Y_{t-1} \\ X_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{yt} \\ \varepsilon_{xt} \end{pmatrix}, \qquad (3)$$

where $\begin{pmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \end{pmatrix}$ is the matrix of autoregressive coefficients and we assume that

$$\underbrace{\begin{pmatrix} \varepsilon_{yt} \\ \varepsilon_{xt} \end{pmatrix}}_{\text{„}} \overset{D}{=} N_2 \left(\mathbf{0}, \underbrace{\begin{pmatrix} \sigma_y^2 & \sigma_{yx} \\ \sigma_{yx} & \sigma_x^2 \end{pmatrix}}_{\boldsymbol{\Sigma}} \mid Y_{t-1}, X_{t-1} \right) \forall \ t \qquad (4)$$

AR(p) as the final equation of a system II

- In fact this is an example of a first order Vector Autoregressive model, **VAR** with gaussian disturbances.
- As an exercise, you can show that (3) can be reduced to the so called **final equation** for Y_{t+1}

$$Y_{t+1} = (\underbrace{\pi_{11} + \pi_{22}}_{\equiv \phi_1}) Y_t + (\underbrace{\pi_{12}\pi_{21} - \pi_{22}\pi_{11}}_{\equiv \phi_2}) Y_{t-1} \qquad (5)$$

$$+ \underbrace{\varepsilon_{yt+1} - \pi_{22}\varepsilon_{yt} + \pi_{12}\varepsilon_{xt}}_{\equiv \varepsilon_t}.$$

AR(p) as the final equation of a system III

- But the same equation must hold for Y_t so we obtain (1) for the case of $p = 2$ and $\phi = 0$ as

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t \tag{6}$$

$$\phi_1 = (\pi_{11} + \pi_{22}) \tag{7}$$

$$\phi_2 = \pi_{12}\pi_{21} - \pi_{22}\pi_{11} \tag{8}$$

$$\varepsilon_t = \varepsilon_{yt} - \pi_{22}\varepsilon_{y,t-1} + \pi_{12}\varepsilon_{xt-1} \tag{9}$$

- The omission of the intercept (which implies $\phi_0 = 0$) is only to save notation.

AR(p) as the final equation of a system IV

- Note that when ε_t is defined as in (9) we have $E(\varepsilon_t) = 0$ and

$$\begin{aligned} Var(\varepsilon_t) &= Var(\varepsilon_{y,t} - \pi_{22}\varepsilon_{y,t-1} + \pi_{12}\varepsilon_{x,t-1}) \\ &= \sigma_y^2 + \pi_{22}\sigma_{yy} + \pi_{12}\sigma_x^2 + 2\pi_{22}\pi_{12}\sigma_{yx} \end{aligned}$$

  independent of t (homoskedasticity), but

$$\begin{aligned} Cov(\varepsilon_t, \varepsilon_{t-1}) &= -\pi_{22}\sigma_y^2 + \pi_{12}\sigma_{yx} \\ Cov(\varepsilon_t, \varepsilon_{t-j}) &= 0 \; j = 2, 3, \ldots \end{aligned}$$

- In this interpretation, the disturbance ε_t in (6) is not white-noise, but a *Moving Average* (MA) process. Following custom the modelled is called ARMA(2,1).

Dynamic stability and stationarity of AR(p) I

- Consider again the AR(p) process:

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} + \varepsilon_t \qquad (10)$$

- Consider next the **homogenous version** of the difference equation:

$$Y_t^h - \phi_1 Y_{t-1}^h - \phi_2 Y_{t-2}^h - ... - \phi_p Y_{t-p}^h = 0 \qquad (11)$$

Dynamic stability and stationarity of AR(p) II

- From mathematics we know that (11) has a **global asymptotic stable solution** ($Y_t^h \to 0$ when $t \to \infty$) if and only if all the p roots (eigenvalues) of the associated characteristic polynomial

$$\lambda^p - \phi_1\lambda^{p-1} - \phi_2\lambda^{p-2} - ... - \phi_p = 0 \tag{12}$$

  are less than one in absolute value.
- From a result that is far from trivial, and which we leave for ECON 5101, we have that the same condition is necessary and sufficient for the **stationarity** of the stochastic process Y_t when it is given by (10) and ε_t is white-noise, or *any other stationary time series process* (e.g., MA(q), $q = 1, 2, ...$).
- But now we have given the condition for stationarity without a definition for stationarity...!

Stationarity defined I

For the time series $\{Y_t;\ t = 0, \pm 1, \pm 2, \pm 3, ...\}$ we define the *autocovariances* $\gamma_{j,t}$ in slightly more general way than in Lecture 4:

$$\tau_{j,t} = E[(Y_t - \mu_t)(Y_{t-j} - \mu_t)],\ \ j = 0, 1, 2, \ldots, \tag{13}$$

where$E(Y_t) = \mu_t$.
If neither μ nor $\gamma_{j,}$ depend on time t:

$$E(Y_t) = \mu, \forall\ t \tag{14}$$

and

$$E[(Y_t - \mu)(Y_{t-j} - \mu)] = \tau_j, \forall\ t,\ j. \tag{15}$$

the Y_t process $\{Y_t;\ t = 0, \pm 1, \pm 2, \pm 3, ...\}$ is **covariance stationary** (aka weakly stationary).

Stationarity defined II

For a stationary Y_t the variance is time independent

$$Var(Y_t) = \sigma_y^2 \equiv \tau_0 \text{ for } j = 0$$

and the autocovariances are symmetric backwards and forwards:
$\tau_j = \tau_{-j}$

The autocorrelation function of stationary AR(p) I

- For a stationary time series variable, the theoretical autocovariances only depend on the distance j between periods. We can regard the autocovariance as a function of j.
- The same is the case for the (theoretical) autocorrelation function (ACF). In general, it is a function of j and t:

$$\zeta_{j,t} = \{Y_t, Y_{t-j}\} = \frac{Cov(Y_t, Y_{t-j})}{Var(Y_t)} = \frac{\tau_{j,t}}{\tau_{0,t}}, \qquad (16)$$

However

$$\zeta_j = \frac{\tau_j}{\tau_0} = \zeta_{-j} \text{ for } j = 1, 2, ... \qquad (17)$$

in the stationary case.

Why is stationarity so important? I

- For an observable time series $\{Y_t;\ t = 1, 2, 3, ...T\}$, we use the empirical autocovariances,

$$\hat{\tau}_j = 1/T \sum_{t=j+1}^{T} (Y_t - \bar{Y})(Y_{t-j} - \bar{Y}),\ j = 0, 1, 2, \dots, T-1 \tag{18}$$

  where $\bar{Y} = 1/T \sum_{t=1}^{T} Y_t$.
- If the process $\{Y_t;\ t = 0, \pm 1, \pm 2, \pm 3, ...\}$ is stationary, $\hat{\tau}_j$ $(j = 0, 1, 2, \dots)$ are consistent estimators of the theoretical autocovariances.
- This in turn gives the main premise for consistent estimation of the coefficients of dynamic regression models, of which AR(p) is an example

Why is stationarity so important? II

- In short: stationary is the main premise for why we can extend the MLE and OLS based estimation and inference theory to time series data!
- Hence the importance of $-1 < \phi_1 < 1$ in the AR(1) m model
- Note that, although stationarity depends on the characteristics roots, it can be "mapped back" to the ϕ_1 and ϕ_2 parameters in the AR(2) case.

$1 - \phi_1 - \phi_2 > 0$, $1 > -\phi_1 + \phi_2$ and $1 > -\phi_2 \iff$ AR(2) is stationary

AR(2) example revisited I

$\gamma = 0,\ \phi_1 = 1,6,\ \phi_2 = -0,9$:

$$Y_t = 1.6Y_{t-1} - 0.9Y_{t-2} + \varepsilon_t, \tag{19}$$

- The characteristic equation is:

$$\lambda^2 - 1.6\lambda + 0.9 = 0$$

- The roots are a complex pair:
  - $\lambda_1 = 0.8 - 0.509\,9i$
    $\lambda_2 = 0.8 + 0.509\,9i$
- The module ("absolute value") of the roots is $|\lambda| = \sqrt{0.8^2 + 0.5^2} \approx 0.94$, inside the complex unit-circle.

Consistency and distribution I

- We now have better background for assessing the statistical properties of MLEs for AR models
- Consider the MLE for ϕ_1 that we derived in Lecture 4
- Simplify by setting $\phi_0 = 0$ in the model equation, the notations in the expression for $\widehat{\phi}_1$ can then be simplified:

$$\widehat{\phi}_1 = \frac{\sum_{t=2}^{T} Y_t Y_{t-1}}{\sum_{t=1}^{T} Y_{t-1}^2} = \sum_{t=1}^{T} \left(\frac{\phi_1 Y_{t-1}^2}{\sum_{t=1}^{T} Y_{t-1}^2} \right) + \sum_{t=1}^{T} \left(\frac{Y_{t-1}\varepsilon_t}{\sum_{t=2}^{T} Y_{t-1}^2} \right) \tag{20}$$

$\Longrightarrow$

$$E\left(\widehat{\phi}_1 - \phi_1\right) = E\left(\frac{\sum_{t=1}^{T} Y_{t-1}\varepsilon_t}{\sum_{t=1}^{T} Y_{t-1}^2} \right).$$

Consistency and distribution II

- Even if we assume $E\left(Y_{t-1}\varepsilon_t\right) = 0$, we cannot state that the denominator and numerator are independent: For example will ε_2 "be in" the numerator and (because of $Y_2 = \phi_1 + \varepsilon_2$) also in $Y_2 \times Y_2$ in the denominator.
- This means that Y_{t-1} cannot be regarded as exogenous in the econometric sense, and therefore $E\left(\widehat{\phi}_1 - \phi_1\right) \neq 0$.
- What about asymptotic properties? With reference to the Law of large numbers and Slutsky's theorem we have

$$\text{plim}\left(\widehat{\phi}_1 - \phi_1\right) = \frac{\text{plim}\,\frac{1}{T}\sum_{t=2}^{T} Y_{t-1}\varepsilon_t}{\text{plim}\,\frac{1}{T}\sum_{t=2}^{T} Y_{t-1}^2} = \frac{0}{\frac{\sigma_\varepsilon^2}{1-\phi_1^2}} = 0.$$

if $E\left(Y_{t-1}\varepsilon_t\right) = 0$ and $|\phi_1| < 1$.

Consistency and distribution III

- The zero in the numerator seems trivial since it is just a sum of terms with zero expectations, but closer inspection shows that we need that the variance of $Y_{t-1}\varepsilon_t$ is finite. The specification of the AR(1) model above is sufficient for this result.
- The denominator is due to the assumption $|\phi_1| < 1$, which entails that the variance of Y_t in (20) is finite and equal to $\sigma_\varepsilon^2/(1-\phi_1^2)$ from the solution of the AR(1) model.

Consistency and distribution IV

- The OLS/ML estimator $\widehat{\phi}_1$ is consistent, and it is approximately normal when T is large enough, see §12.7 in HN:

$$\sqrt{T}\left(\widehat{\phi}_1 - \phi_1\right) \overset{D}{\approx} N\left(0, \left(1 - \phi_1^2\right)\right) \tag{21}$$

  which entails that t-tests can be compared with critical values from the normal distribution.
- This result extend MLE estimators for the AR(1) in Lecture 4 (the model where $\phi_0 \neq 0$).

Hurwitz-bias

- In (**??**) the finite sample bias can be shown to be approximately
$$E\left(\widehat{\phi}_1 - \phi_1\right) \approx \frac{-2\phi_1}{T},$$
this is called the Hurwitz-bias after Leo Hurwitz (1958).
- In CC we can make this more concrete with a Monte-Carlo analysis.

MLE of AR(p) I

- The likelihood function of AR(p) is constructed in the same manner as for AR(1), with white-noise or gaussian disturbances (MA is a bit more complicated)
- Since the condition distribution is $E(Y_t \mid Y_{t-1}, \ldots, Y_{t-p})$

$$Y_t \overset{D}{=} N(\phi_0 + \sum_{i=1}^{p} \phi_i Y_{t-i} + \varepsilon_t, \sigma^2)$$

we have p initial values. $Y_0, \; Y_{-1}, \; \ldots, \; Y_{-(p-1)}$

- With $\mathbf{y}' = (Y_1, \; Y_2, \ldots, Y_t)$, and suitably defined $\mathbf{X}$ matrix the MLE estimators of $\phi = (\phi_0, \; \phi_1, \ldots, \phi_p)$ are given by OLS formula

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

- $\hat{\sigma}^2$ is the average of the squared residuals $\hat{\sigma}^2 = (1/T)\varepsilon'\varepsilon = \mathbf{y}'\mathbf{M}\mathbf{y}$ (cf. Lecture 3).

Lag operators I

- When we work with stochastic difference equations, it is often useful to express relationships with the use of the lag-operator L.
- The lag operator L changes the dating of a variable Y_t one or more period back in time. It works in the following way:

$$LY_t = Y_{t-1},$$
$$LLY_t = L^2 Y_t = LY_{t-1} = Y_{t-2},$$
$$L^p Y_t = Y_{t-p}.$$

- From the last property it follows that if $p = 0$, then

$$L^0 = 1,$$
$$L^0 Y_t = Y_t.$$

Lag operators II

- We also have
$$L^p L^s = L^p L^k = L^{(p+s)},$$
and
$$\left(aL^p + bL^s\right) Y_t = aL^p Y_t + bL^s Y_t = aY_{t-p} + bY_{t-s}.$$
- If we want to shift a variable forward in time, we use the forward operator L^{-1}:
$$L^{-1} Y_t = Y_{t+1}$$
and generally
$$L^{-s} = Y_{t+s}.$$

Lag operators III

- Because constants are independent of time, we have for the constant b

$$Lb = b.$$

and by induction

$$L^p b = L^{(p-1)} L b = L^{(p-1)} b = b.$$

Lag-polynomial representation of AR(p) I

- We can now write (1) more compactly as

$$\phi(L)Y_t = \phi_0 + \varepsilon_t \tag{22}$$

where is the lag polynomial of order p.

$$\phi(L)Y_t = 1 - \phi_1 L - \phi_2 L^2 - ...\phi_p L^p \tag{23}$$

and we keep the assumption of white-noise ε_t.

Lag-polynomial representation of AR(p) II

- A root of the characteristic equation associated with the lag-polynomial is:

$$1 - \phi_1 z - \phi_2 z - ...\phi_p z^p = 0 \tag{24}$$

Comparison with the characteristic equation (12) shows that

$$z = \frac{1}{\lambda}$$

meaning that the condition for stationarity can also be expressed in terms of the roots: $(z_1, z_2, ..., z_p)$:

- Y_t is stationary if all the $z-$roots are larger than one in absolute value ("outside the unit circle").

Companion form I

Consider again the VAR system (3)

$$\begin{pmatrix} Y_t \\ X_t \end{pmatrix} = \underbrace{\begin{pmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \end{pmatrix}}_{\mathbf{\Pi}} \begin{pmatrix} Y_{t-1} \\ X_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{yt} \\ \varepsilon_{xt} \end{pmatrix},$$

- Assume that ε_{yt}, ε_{xt} are two stationary series the This is secured by (4) for example.

Companion form II

- By obtaining the characteristic polynomial to $\mathbf{\Pi}$:

$$p(\lambda) = |\mathbf{\Pi} - \lambda \mathbf{I}|$$

you find that the **eigenvalues of $\mathbf{\Pi}$** are the roots of

$$|\mathbf{\Pi} - \lambda \mathbf{I}| = 0 \tag{25}$$

which is the characteristic equation associated with the final equation (5) that we derived above.

- Hence the necessary and sufficient condition for stationary of the vector $(Y_t, X_t)'$ is that the two eigenvalues of both less than one in absolute value.
- **A** is a simple example of a so called **companion form** matrix.

Companion form III

- In ECON 5101 we will show that if we have a general VAR with n time series variables and p lags, that VAR can be written as a first order system

$$\mathbf{z}_t = \mathbf{F}\mathbf{z}_{t-1} + \boldsymbol{\epsilon}_t \tag{26}$$

  where $\mathbf{z}_t$ and $\boldsymbol{\epsilon}_t$ are $1 \times np$ and the companion form matrix $\mathbf{F}$ is $np \times np$.
- For such a general VAR system, the condition for stationarity and stability is that all the np eigenvalues from

$$|\mathbf{F} - \lambda \mathbf{I}| = 0 \tag{27}$$

  are less than one in magnitude.

Companion form IV

- When we estimate a dynamic system in PcGive, the eigenvalues of the companion form are always available after estimation.

MLE of VAR(1) I

- Consider the VAR(1) made up of (3) and (4) so that ε_1, ε_2, $\ldots$,ε_T are mutually independent and normal.
- The pdf of $\mathbf{y}_t$ given $\mathbf{y}_{t-1}$ is

$$f\left(\mathbf{y}_t \mid \mathbf{y}_{t-1}\right) = \frac{1}{\sigma_y \sigma_x 2\pi \sqrt{\left(1-\rho_{xy}^2\right)}} \times \exp\left[-\frac{1}{2}\frac{\left(z_{yt}^2 - 2\rho_{xy} z_{yt} z_{xt} + z_{xt}^2\right)}{\left(1-\rho_{xy}^2\right)}\right] \tag{28}$$

MLE of VAR(1) II

where $\sigma_j = \sqrt{\sigma_j^2}$ $j = x, y$, $\rho_{xy} = \sigma_{xy}/(\sigma_x \sigma_y)$ (correlation coefficient) and

$$z_{yt} = \frac{Y_t - \mu_{Y|t-1}}{\sigma_y}$$
$$z_{xt} = \frac{X_t - \mu_{x|t-1}}{\sigma_x}$$

- where the conditional expectations are

$$\mu_{Y|t-1} = \pi_{10} + \pi_{11} Y_{t-1} + \pi_{12} X_{t-1} \quad (29)$$
$$\mu_{X|t-1} = \pi_{20} + \pi_{21} Y_{t-1} + \pi_{22} X_{t-1} \quad (30)$$

where we have included the two intercepts.

MLE of VAR(1) III

- By invoking the Markov property we can write:

$$f\left(\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_T \mid \mathbf{y}_0\right) = \prod_{t=1}^{T} f\left(\mathbf{y}_t \mid \mathbf{y}_{t-1}\right)$$

cf. page 204 in HN, which is the likelihood function for the gaussian VAR(1):

$$L_{VAR(1)} = \prod_{t=1}^{T} f\left(\mathbf{y}_t \mid \mathbf{y}_{t-1}\right) \quad (31)$$

with $f\left(\mathbf{y}_t \mid \mathbf{y}_{t-1}\right)$ given by (28)

- Consider first the case of $\pi_{ij} = 0$ for $i, j = 1, 2$ so that $\mu_{Y|t-1} = \pi_{10}$ and $\mu_{X|t-1} = \pi_{20}$. In this case the MLE are the OLS estimators $\hat{\pi}_{10} = \bar{Y}$ and $\hat{\pi}_{20} = \bar{X}$.
- The fact that $\rho_{xy} \neq 0$ in general does not change that result!

MLE of VAR(1) IV

- Which also extends to (29) and (30) in general: The MLEs of π_{10},π_{11},π_{21}, π_{20},π_{21},π_{22} are obtained by estimating each row in VAR(1) by OLS as if they were two separate regressions.
- This is a case of the SURE theorem with identical regressors.

MLE of VAR(p) I

- The result about ML estimationof the VAR by OLS on each row in the system extends to VAR(p) models:

$$\mathbf{y}_t = \sum_{i=1}^{p} \mathbf{\Pi}_i \mathbf{y}_{t-1-i} + _t$$

  where $\mathbf{\Pi}_i$ $(i = 1, 2, \ldots, p)$ are autoregressive matrices and ε_t is normal.

- We can also extend by other deterministic regressors than the intercepts. And by exogenous explanatory variables, such models are often called open-VARs or VAR-EX models

The VAR revisited I

Let us now take care to write the gaussian disturbances of the VAR (now including two intercepts)

$$\begin{pmatrix} Y_t \\ X_t \end{pmatrix} = \begin{pmatrix} \pi_{10} \\ \pi_{20} \end{pmatrix} + \begin{pmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \end{pmatrix} \begin{pmatrix} Y_{t-1} \\ X_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{yt} \\ \varepsilon_{xt} \end{pmatrix} \tag{32}$$

as conditional on period $t-1$:

$$\begin{pmatrix} \varepsilon_{yt} \\ \varepsilon_{yt} \end{pmatrix} \sim N_2 \left(\mathbf{0}, \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{pmatrix} \mid Y_{t-1}, X_{t-1} \right). \tag{33}$$

Now, (32) can be written as

$$Y_t = \mu_{y,t-1} + \varepsilon_{yt} \tag{34}$$

$$X_t = \mu_{x,t-1} + \varepsilon_{xt} \tag{35}$$

The VAR revisited II

where the **conditional expectations** $\mu_{y|t-1} \equiv E(Y_t \mid Y_{t-1}, X_{t-1})$ and $\mu_{x|t-1} \equiv E(X_t \mid Y_{t-1}, X_{t-1})$ are

$$\mu_{y,t-1} = \pi_{10} + \pi_{11} Y_{t-1} + \pi_{12} X_{t-1} \quad (36)$$

$$\mu_{x,t-1} = \pi_{20} + \pi_{21} Y_{t-1} + \pi_{22} X_{t-1}. \quad (37)$$

Interpretation: Conditional on the history of the system up to time $t-1$, Y_t and X_t are jointly normally distributed.

The conditional model for Y I

The conditional distribution for Y_t given the history **and** X_t is also normal,
In **Lecture note 3** (posted after the lecture for self-study) we show that the conditional distribution for Y_t is:

$$Y \sim N(\phi_0 + \phi_1 Y_{t-1} + \beta_0 X_t + \beta_1 X_{t-1}, \sigma^2 \mid X_t, Y_{t-1}, X_{t-1}) \quad (38)$$

which can be written in model form as

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \beta_0 X_t + \beta_1 X_{t-1} + \varepsilon_t \quad (39)$$

$$\varepsilon_t \sim N(0, \sigma^2 \mid X_t, Y_{t-1}, X_{t-1}) \quad (40)$$

The conditional model for Y II

$$\phi_0 = \pi_{10} - \frac{\sigma_{xy}}{\sigma_x^2}\pi_{20} \tag{41}$$

$$\phi_1 = \pi_{11} - \frac{\sigma_{xy}}{\sigma_x^2}\pi_{21} \tag{42}$$

$$\beta_0 = \frac{\sigma_{xy}}{\sigma_x^2} \tag{43}$$

$$\beta_1 = \pi_{12} - \frac{\sigma_{xy}}{\sigma_x^2}\pi_{22} \tag{44}$$

and

$$\sigma^2 = \sigma_y^2(1 - \phi_{xy}^2). \tag{45}$$

$$\phi_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}. \tag{46}$$

The conditional model for Y III

- Some small differences in notation apart, this is the same ADL model as in DM Ch 13.5 eq (13.58) for $p = q = 1$.
- The same ADL type model can be derived from a VAR with IID disturbances, rather than strictly normal.
- ADL(p,q) model can be derived from a VAR or order p. Consequently we must then have $p = q$ in the ADL.
- We will study such ADL models, and their estimation over the next weeks.

The conditional model for Y IV

- Finally, note that the ADL model

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \beta_0 X_t + \beta_1 X_{t-1} + \varepsilon_t \tag{47}$$

together with the second row in the VAR:

$$X_t = \pi_{20} + \pi_{21} Y_{t-1} + \pi_{22} X_{t-1} + \varepsilon_{xt} \tag{48}$$

where the two disturbances are independent, give a **regression representation** of the VAR, in terms of a **conditional model** (47) and a **marginal model** (47).

- Correspondingly, HN shows in §14.1 how the likelihood-function (31) of the VAR can be factorized into a
  - a conditional likelihood (for (47) and
  - a marginal likelihood function (for (31).

The conditional model for Y V

as long as there are no cross-equation restrictions, meaning exogeneity.

- Start with exogeneity in dynamic models in Lecture 6.