Lecture note 4

From the Haavelmo distribution to the VAR and to ADL models.

The Haavelmo distribution and sequential factorization

In lecture 3, and in Lecture note 3, and in CC2, we learned that if we accept that the VAR(1) with normally distributed (Gaussian) disturbances is conditioned on the history of the system up to period "t-1", we can show that the ADL model with one lag in both Y and X can be derived as a valid conditional model of the VAR.

For those interested, we now give the more general argument for why we can regard not only VAR(1), but also the VAR(p) as a model that is conditioned on history of the system.

We start by the letting the vector $\mathbf{y}_t$ consist of $(k+1)$ stationary variables. There are T such vector time series and we can write the whole sample as a $(k+1) \times T$ matrix: $\mathbf{Y}_T^1 = (\mathbf{y}_1, \ldots, \mathbf{y}_T)$. Consistent with this, we write $\mathbf{Y}_{t-1}^1 = (\mathbf{y}_1, \ldots, \mathbf{y}_{t-1})$, and so on. We do not know when the process begins, but we assume that there are p starting vaues which we collect in $\mathbf{Y}_0 \equiv \left(\mathbf{y}_{-(p+1)}, \ldots, \mathbf{y}_{-1}, \mathbf{y}_0\right)$, so that we can write the whole history of the time series variables as $\mathbf{Y}_{T-1} \equiv \left(\mathbf{Y}_0, \mathbf{Y}_{T-1}^1\right)$. The probability density function of the variables in $\mathbf{Y}_T$, for the given initial values are therefore:

$$f\left(\mathbf{Y}_T^1 \mid \mathbf{Y}_0, \boldsymbol{\theta}\right).$$

This probability density function has become known as the **Haavelmo distribution** in moden econometrics. This shows the recognition of the seminal methodological contribution Haavelmo gave in the *Probability Approach to Econometrics* in 1944. The relevance for us is that we can factorize the Haavelmo distribution just as any other pdf:

$$f\left(\mathbf{Y}_T^1 \mid \mathbf{Y}_0; \boldsymbol{\theta}\right) = g_T\left(\mathbf{y}_T \mid \mathbf{Y}_{T-1}^1, \mathbf{Y}_0; \boldsymbol{\theta}\right) g_{T-1}\left(\mathbf{Y}_{T-1}^1 \mid \mathbf{Y}_0; \boldsymbol{\theta}\right). \tag{1}$$

where $g_T\left(\mathbf{y}_T \mid \mathbf{Y}_{T-1}, \mathbf{Y}_0; \boldsymbol{\theta}\right)$ is a conditional pdf relative to the Haavelmo distribution, and $g_{T-1}\left(\mathbf{Y}_{T-1}^1 \mid \mathbf{Y}_0; \boldsymbol{\theta}\right)$ is a marginal pdf (also in that relative sense).

In the following, we can think of the initial conditions in $\mathbf{Y}_0$ as fixed numbers. Since it is a conditional pdf, $g_T\left(\mathbf{y}_T \mid \mathbf{Y}_{T-1}^1, \mathbf{Y}_0; \boldsymbol{\theta}\right)$ motivates a regression between the $(k+1)$-dimensional vector $\mathbf{y}_T$ and $\mathbf{Y}_{T-1}^1$. If we do a second factorization in (1), this time with resepct to $T-2$ (1), we get a second conditonal distribution $g_{T-1}\left(\mathbf{y}_{T-1} \mid \mathbf{Y}_{T-2}^1 \mathbf{Y}_0; \boldsymbol{\theta}\right)$, which defines the regression between $\mathbf{y}_{T-1}$ and $\mathbf{Y}_{T-2}^1$. We can repeat this factorization sequentially, all the way back to $t=1$, which is a regression between $\mathbf{y}_1$ and the p initial conditions. Since our goal is to represent all these regressions by one common model, we assume that there are p lags in each of other regressions as well. This is the same as assuming that only p lags are relevant for predicting each $\mathbf{y}_t$. We then have

$$E\left(\mathbf{y}_t \mid \mathbf{Y}_{t-1}, \boldsymbol{\theta}\right) = E\left(\mathbf{y}_t \mid \mathbf{Y}_{t-1}^{t-p}, \boldsymbol{\theta}\right), t = 1, 2, \ldots, T$$

where $\mathbf{Y}_{t-1}^{t-p} = (\mathbf{y}_{t-p}, \ldots, \mathbf{y}_{t-1})$[1].

We will assume that the multivariate conditional expectation is a linear function, and from before we know that this is particularly the case when the Haavelmo distribution is a Gaussian, and we can write the expectation as:

$$E\left(\mathbf{y}_t \mid \mathbf{Y}_{t-1}^{t-p}, \boldsymbol{\theta}\right) = \sum_{i=1}^{p} \boldsymbol{\Pi}_i \mathbf{y}_{t-i} \tag{2}$$

[1] This is knows as a Markov-process of degree p.

where we let $\mathbf{\Pi}$ replace $\boldsymbol{\theta}$ as symbol for the parameters, to underline that linearity in the parameters of the expectation function is a separate assumption.

In exactely the same way as the VAR(1) case (cf. Lecture 3 and Lecture note 1) we express the variables as deviations from the expectations in (2):

$$\boldsymbol{\epsilon}_t = \mathbf{y}_t - E\left(\mathbf{y}_t \mid \mathbf{Y}_{t-1}^{t-p}, \mathbf{\Pi}\right) = \mathbf{y}_t - \sum_{i=1}^{p} \mathbf{\Pi}_i \mathbf{y}_{t-i}.$$

and we can write the distributional properties for the variables i $\mathbf{y}_t$ in "model form" as:

$$\mathbf{y}_t = \mathbf{\Pi}_1 \mathbf{y}_{t-1} + \mathbf{\Pi}_2 \mathbf{y}_{t-2} + ... + \mathbf{\Pi}_p \mathbf{y}_{t-p} + \boldsymbol{\epsilon}_t \tag{3}$$

which is an n-dimensional VAR of order p,VAR(p). $\boldsymbol{\epsilon}_t$is multivariate Gaussian:

$$\boldsymbol{\epsilon}_t \sim IN(\mathbf{0}, \mathbf{\Sigma}). \tag{4}$$

From VAR(p) to the ADL(p,q) model.

In lecture 3, and in *Lecture note 3*, we studied the case of $n = 2$ and $p = 1$, and showed that the ADL(1) model can be interpreted as a conditional model for one variable, given the other, obtained from that VAR.

The same argument holds for the general VAR(p) in (3), and gives rise to general ADL models as well. For example if we set $\mathbf{y}_t = (Y_t, X_{1t}, \ldots, X_{kt})$ and choose Y_t as the regressor, the derived conditional model will be an ADL with k distributed lag polynomials, each of length p, and an autoregressive lag polynomial of degree p:

$$\phi(L)Y_t = \phi_0 + \sum_{j=0}^{k} \beta_j(L) X_{jt} + \varepsilon_t \tag{5}$$

where

$$\phi(L) = 1 - \sum_{i=1}^{p} \phi_i L^i \tag{6}$$

$$\beta_j(L) = \sum_{i=0}^{p} \beta_{ji} L^i, \; j = 1, 2, \ldots k \tag{7}$$

By the assumptions of the model, all the X variables are exogenous (as in any Gaussian or IID regression model), but Y_{t-1}, Y_{t-2}, …,Y_{t-p} are predetermined variables.

In Davidson and MacKinnon, the general ADL model is written as $ADL(p, q)$ where p is the lag order for the regressand and q are the number of lags in X. $q \neq p$ does not create any logical problems for the derivation above, it is just that $p = q$ saves notation.

Constants and dummies of various sorts

ADL models also often include deterministic dummies and trends. They reason why such dummies are absent in the derivations of the ADL in Lect 3 and aboves simply that we save notation by not including them in the VAR. If they *had* been in the VAR they would logically also end up in the conditional ADL model!

In practice we often use VAR models that include non-modelled (exogenous) variables. VARs of this type are referred to as **open VARs** or **VAR-EX**. Again, if you start with a VAR-EX model with e.g.. one exogoenous Z variable, and go through the same line of tought as in *Lecture 3* and above, that Z variable will logically "end up" in the dervied ADL