ØAMET4100 · Spring 2019
Lecture Note 11

Instructor: Fenella Carpena

April 4, 2019

This lecture note provides a review of time series data (Stock & Watson, Chapter 16). This lecture note is not intended to be a comprehensive review of lecture or the textbook, since there is a lot more material than we have time to cover. However, I have tried to focus on the concepts which I believe are necessary to be successful in our class.

1 Vector Autoregressions

A **vector autoregressive (VAR)** model is useful when one is interested in predicting multiple time series variables using a single model. At its core, the VAR model is an extension of the autoregressive models we have dealt with in Chapter 14.

The VAR model extends the idea of univariate autoregression to k time series regressions, where the lagged values of **all** k series appear as regressors. Put differently, in a VAR model we regress a **vector** of time series variables on lagged vectors of these variables. As with AR(p) models, the lag order is denoted by p so the VAR(p) model of two variables X_t and Y_t (i.e., $k = 2$) is given by the equations

$$\begin{aligned} Y_t =& \beta_{10} + \beta_{11} Y_{t-1} + \cdots + \beta_{1p} Y_{t-p} + \gamma_{11} X_{t-1} + \cdots + \gamma_{1p} X_{t-p} + u_{1t}, \\ X_t =& \beta_{20} + \beta_{21} Y_{t-1} + \cdots + \beta_{2p} Y_{t-p} + \gamma_{21} X_{t-1} + \cdots + \gamma_{2p} X_{t-p} + u_{2t}. \end{aligned}$$

The β's and γ's can be estimated using OLS on each equation. The assumptions for VARs are the time series assumptions presented last lecture (see Key Concept 14.6 of the textbook) applied to each of the equations. When these assumptions hold, the OLS estimators of the VAR coefficients are consistent and jointly normal in large samples so that the usual inferential methods such as confidence intervals and t-statistics can be used.

How do we conduct hypothesis tests in VAR models? As mentioned above, when the time series assumptions hold, we can apply the usual methods (e.g., the 95% confidence interval can be constructed as the estimated coefficient ± 1.96 times the standard error).

But we also have a new aspect of hypothesis testing in VARs. Specifically, the structure of VARs allows us to jointly test restrictions across multiple equations. For instance, it may be of interest to test whether the coefficients on all regressors of the lag p are zero (i.e., which is testing that the lag order $p - 1$ is correct). This corresponds to the null hypothesis

$$H_0 : \beta_{1p} = 0, \beta_{2p} = 0, \gamma_{1p} = 0, \gamma_{2p} = 0,$$

which involves coefficients from *both* equations in the VAR. The alternative hypothesis is that at least one of these coefficients is not zero. The above test can then be carried out using an F-stat. The explicit formula for such a test statistic is rather complicated but fortunately such computations are easily done using software (e.g., Stata).

How do we choose the lag length in VAR? Apart from using the F-stat to determine the lag length, another way is to use information criteria like the BIC which we studied last week. Just as in the case of a single autoregressive equation, for a multiple equation model we choose the specification which has the smallest $BIC(p)$, where

$$BIC(p) = \log\left[\det(\widehat{\Sigma}_u)\right] + k(kp+1)\frac{\log(T)}{T}. \tag{1}$$

The term $\widehat{\Sigma}_u$ denotes the estimate of the $k \times k$ covariance matrix of the VAR errors and $\det(\cdot)$ denotes the determinant. The AIC is computed using the same equation as above but replacing the term "ln(T)" with "2." Note that $k(kp+1)$ is the total number of regression coefficients in the VAR (i.e., we have k equations, each with an intercept and p lags of each of the k variables).

How many variables should be included in the VAR? As with autoregressions with just one equation, we should think carefully about which variables to include in a VAR. Adding unrelated variables reduces the forecast accuracy by increasing the estimation error. This is particularly important because the number of parameters to be estimated grows quadratically with the number of variables modeled by the VAR. For example, a VAR with 5 variables and 4 lags will have a total of 105 coefficients (i.e., 21 coefficients in each of the 5 equations).

2 Multiperiod Forecasts

So far, we have focused on forecasts one period in advance, but it is also possible to use the models we have studied to make forecasts more than one period into the future. We will discuss two methods for making multiperiod forecasts: (1) **iterated** forecasts, where a one-period ahead model is iterated forward one period at a time, and (2) **direct** forecasts, where we use a regression in which the dependent variable is the multiperiod variable that we want to predict.

2.1 Iterated Forecasts

The key idea with iterated forecasting is that we first use the model to make a forecast that is one period ahead (i.e., for $T+1$) using data until period T. Then, we treat that forecasted value as data to make a forecast for the next period $T+2$. The following example illustrates how this works.

Example. Suppose that we have the following AR(1) model of the GDP growth rate $GDPGR$, which we obtain using quarterly data from 1962:Q1 to 2012:Q4.

$$\widehat{GDPGR}_t = \underset{(0.35)}{1.99} + \underset{(0.08)}{0.34}\, GDPGR_{t-1}.$$

The following are the values of $GDPGR$ from 2012:Q1 to 2012:Q4.

Date	2012:Q1	2012:Q2	2012:Q3	2012:Q4
$GDPGR$	3.64	1.20	2.75	0.15

How can we use the above AR(1) model to forecast the GDP growth rate in 2013:Q2? We need to proceed in the two steps. First, we need to obtain the forecasted GDP growth rate for 2013:Q1.

$$\widehat{GDPGR}_{2013:Q1|2012:Q4} = 1.99 + 0.34 * 0.15 = 2.04.$$

This gives us the forecast for one period into the future. Then, we need to substitute this forecast back into the AR(1) equation again to obtain the two-period ahead forecast.

$$\widehat{GDPGR}_{2013:Q2|2012:Q4} = 1.99 + 0.34 * \widehat{GDPGR}_{2013:Q1|2012:Q4} = 1.99 + 0.34 * 2.04 = 2.68.$$

In sum, our forecast for the GDP growth rate in 2013:Q1 is 2.68.

Iterated Forecasts in AR(p) Model. The iterated forecasting strategy shown in the above example can be extended to an AR(p) by replacing Y_{T+1} with its forecast $\widehat{Y_{T+1|T}}$, and then treating that forecast as data for the AR(p) forecast of Y_{T+2}.

Iterated Forecasts in VAR Model. Iterated forecasts in a VAR model also follows the same approach as in the above example. The main point to be careful with is that the forecast of one variable two periods into the future (i.e., $T + 2$) depends on all variables in the VAR in period $T + 1$.

2.2 Direct Forecasts

Direct forecasts are computed without iteration, using one regression where the dependent variable is the multiperiod-ahead variable to be forecasted and the regressors are the predicted variables. For instance, a direct multiperiod forecast two periods into the future based on two lags each of Y_t and X_t is computed by first estimating the regression

$$Y_t = \beta_0 + \beta_1 Y_{t-2} + \beta_2 Y_{t-3} + \delta_1 X_{t-2} + \delta_2 X_{t-3} + u_t$$

and then using the estimated coefficients to make the forecast of Y_{T+2} using data through period T. If, instead, you wanted to make a forecast five periods into the future using the direct forecasting method, then the regression you would estimate is

$$Y_t = \beta_0 + \beta_1 Y_{t-5} + \beta_2 Y_{t-4} + \delta_1 X_{t-5} + \delta_2 X_{t-4} + u_t.$$

Which method should you use for multiperiod forecasting? In most applications, the iterated forecasting method is recommended for two reasons. First, if the underlying one-period-ahead model (e.g., the AR or the VAR that is used to compute the iterated forecast) is specified correctly, then the coefficients are estimated more efficiently in the one-period-ahead model than the multiperiod-ahead model. Second, iterated forecasts are produced using the same regression model, so they tend to be less erratic than direct forecasts.

Nevertheless, there are some circumstances when direct forecasts may be preferable to iterated forecasting. This can happen when especially in VARs when: (1) you have reason to believe that one of the equations in the VAR is not specified correctly, and (2) there are many predictors, in which case the VAR has many estimated coefficients which may cause unreliability.

3 Orders of Integration

Last week, we discussed a random walk model with drift which is written as

$$Y_t = \beta_0 + Y_{t-1} + u_t. \tag{2}$$

We also learned last week that if Y_t follows a random walk, then ΔY_t is stationary. Although the random walk describes the long-run movements of many economic time series, some economic time series have smoother trends (i.e., varies less from one period to the next) than a random walk. We need a different model to describe the trends of such series.

One model of a smooth trend is

$$\Delta Y_t = \beta_0 + \Delta Y_{t-1} + u_t \tag{3}$$

where u_t is a serially uncorrelated error term. This model states that the first difference of a series is a random walk. Consequently, the series of second differences of Y_t (i.e., $\Delta Y_t - \Delta Y_{t-1}$) is stationary.

The above gives rise to special terminology to distinguish between the models in equations (**??**) and (**??**), which is summarized as follows.

- A series that does not have a stochastic trend and is stationary is said to be **integrated of order zero**, written I(0).
- A series that has a random walk trend is said to be **integrated of order one**, written as I(1). The first difference of the series (e.g., ΔY_t) is stationary.
- A series that has a trend of the form in equation (**??**) is said to be **integrated of order two**, written as I(2). The second difference of this series (e.g., $\Delta Y_t - \Delta Y_{t-1}$) is stationary.
- The **order of integration** is the number of times that the series needs to be differenced for it to be stationary.

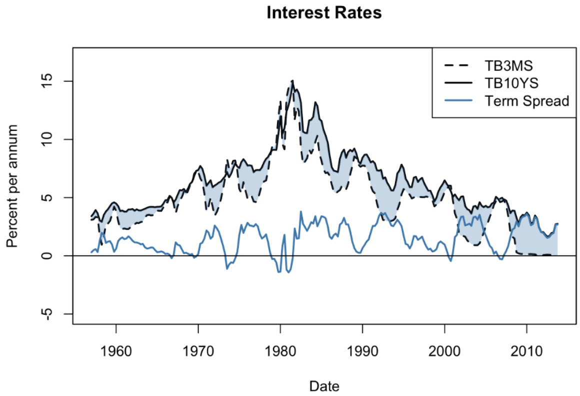
4 Cointegration

4.1 Definition

Two or more series that have a common stochastic trend are said to be **cointegrated**. In this case, we can still use regressions to understand the relationships among the time series variables, but some new methods are needed.

The formal (mathematical) definition of **cointegration** is the following: when X_t and Y_t are $I(1)$ and if there is a θ such that $Y_t - \theta X_t$ is $I(0)$, X_t and Y_t are said to be **cointegrated.** Put differently, cointegration of X_t and Y_t means that they both have the same or a common stochastic trend, and this trend can be eliminated by taking a difference of the series such that the resulting series is stationary.

As an example, consider the the relation between the interest rates of the U.S. 3-month treasury bills (a short-term interest rate, seen in the dashed line) and the U.S. 10-year treasury bonds (a long-term interest rate, seen in the solid black line). The figure below plots these interest rates over time together with the term spread (i.e., the difference between the 10-year and 3-month rates, seen in the blue line).



The plot suggests that long-term and short-term interest rates are cointegrated: both series of the interest rates seem to have the same long-run behavior. They share a common stochastic trend. The term spread seems to be stationary. In fact, theory suggests that an investor should earn the same amount of interest by investing in 40 consecutive 3-month treasury bills versus investing in one 10-year bond today (i.e., the "expectations theory" of the term structure). Theoretically, the cointegrating coefficient θ should be 1. This theory is consistent with the visual result we can see in the above graph.

What is the implication of cointegration for our forecasting models? If X_t and Y_t are cointegrated, we can eliminate the trend by including the term $Y_t - \theta X_t$ in our regression model. Specifically, the first differences of X_t and Y_t can be modeled using a VAR, augmented by including $Y_{t-1} - \theta X_{t-1}$ as an additional regressor:

$$\begin{aligned} Y_t =& \beta_{10} + \beta_{11} Y_{t-1} + \cdots + \beta_{1p} Y_{t-p} + \gamma_{11} X_{t-1} + \cdots + \gamma_{1p} X_{t-p} + \alpha_1 (Y_{t-1} - \theta X_{t-1}) + u_{1t}, \\ X_t =& \beta_{20} + \beta_{21} Y_{t-1} + \cdots + \beta_{2p} Y_{t-p} + \gamma_{21} X_{t-1} + \cdots + \gamma_{2p} X_{t-p} + \alpha_1 (Y_{t-1} - \theta X_{t-1}) + u_{2t}. \end{aligned}$$

Here, the term $Y_{t-1} - \theta X_{t-1}$ is called the **error correction term** and the combined model in the above two equations is called a **vector error correction (VECM)**.

4.2 Testing for Cointegration

There are three ways to determine whether two variables are plausibly cointegrated.

1. **Use economic theory and expert knowledge of the variables to decide whether cointegration is plausible.** As an example, in the above discussion of 3-month and 10-year interest rates, economic theory suggests that in equilibrium, 40 consecutive 3-month treasury bills should give us the same earnings as one 10-year bond. Therefore, theory suggests that the two interest rates are cointegrated.
2. **Graph the series.** We can check visually whether two time series have the same stochastic trend. Again, this is what we did in the above example with 3-month and 10-year interest rates.

3. **Statistical tests of cointegration.** Following the definition of cointegration we talked about in the previous section, it seems natural to construct a test for cointegration of two series in the following manner: if two series X_t and Y_t are cointegrated, the series obtained by taking the difference $Y_t - \theta X_t$ must be stationary. If the series are not cointegrated, $Y_t - \theta X_t$ is nonstationary. This is an assumption that can be tested using a unit root test. We have to distinguish between two cases: whether θ is known or unknown.

Case 1: θ is known. Knowledge of θ enables us to compute differences $z_t = Y_t - \theta X_t$ so that Dickey-Fuller unit root tests can be applied to z_t. For these tests, the critical values are the critical values of the ADF test (i.e., Table 14.4 of the textbook, reproduced below).

TABLE 14.4 Large-Sample Critical Values of the Augmented Dickey–Fuller Statistic

Deterministic Regressors	10%	5%	1%
Intercept only	−2.57	−2.86	−3.43
Intercept and time trend	−3.12	−3.41	−3.96

Case 2: θ is unknown. If θ is unknown, it must be estimated before the unit root test can be applied. This is done by estimating the regression

$$Y_t = \alpha + \theta X_t + z_t$$

using OLS. Then, a Dickey-Fuller test is used for testing the hypothesis that z_t is a nonstationary series. This is known as the Engle-Granger Augmented Dickey-Fuller test for cointegration (or EG-ADF test) after Engle (1987). The critical values for this test are special as the associated null distribution is non-normal and depends on the number of $I(1)$ variables used as regressors in the above regression. The critical values are show in Table 16.2 of the textbook, which is reproduced below.

TABLE 16.2 Critical Values for the Engle–Granger ADF Statistic

Number of *X*'s in Equation (16.24)	10%	5%	1%
1	−3.12	−3.41	−3.96
2	−3.52	−3.80	−4.36
3	−3.84	−4.16	−4.73
4	−4.20	−4.49	−5.07

4.3 Estimating Cointegration Coefficients

If θ is unknown and we believe that two time series are cointegrated, how do we find θ? One estimator of θ that is simple to use in practice is the **dynamic OLS (DOLS) estimator** proposed by Stock and Watson (1993). The DOLS estimator is based on the following regression

$$Y_t = \beta_0 + \theta X_t + \sum_{j=-p}^{p} \delta_j \Delta X_{t-j} + u_t.$$

This equation just means that we regress Y on X, plus the past, present, and future values of δX. The DOLS estimator of θ is then the OLS estimate of θ in the above regression. Intuitively, one way to interpret this equation is that θ can be though of as the long-run effect on Y of a change in X.