# 1 Cointegration and Error Correction Model

This part discusses a new theory for a regression with nonstationary unit root variables. In general, this should require a different treatment from a conventional regression with stationary variables, which has been covered so far. In particular, we focus on a class of the linear combination of the unit root processes known as cointegrated process.

## 1.1 Stylized Facts about Economic Time Series

Casual inspection of most economic time series data such as GNP and prices reveals that these series are non stationary. We can characterize some of the key feature of the various series as follows:

1. Most of the series contain a clear trend. In general, it is hard to distinguish between trend stationary and difference stationary processes.

2. Some series seem to meander. For example, the pound/dollar exchange rate shows no particular tendency to increase or decrease. The pound seems to go through sustained periods of appreciation and then depreciation with no tendency to revert to a long-run mean. This type of random walk behavior is typical of unit root series.

3. Any shock to a series displays a high degree of persistence. For example, the UK industrial production plummeted in the late 1970s and not returning to its previous level until mid 80s. Overall the general consensus is at least empirically that most macro economic time series follow a unit root process.

4 Some series share co-movements with other series. For example, short- and long-term interest rate, though meandering individually, track each other quite closely maybe due to the underlying common economic forces. This phenomenon is called cointegration.

A note on notations: It is widely used that the unit root process is called an integrated of order 1 or for short $I(1)$ process. On the other hand, a stationary process is called an $I(0)$ process.[1]

[1] In this regard we can define $I(d)$ process, and $d$ is a number of differencing to render the series stationary.

## 1.2 Spurious Regression

Suppose that two $I(1)$ processes, $y_t$ and $x_t$, are independently distributed. We now consider the following simple regression:

$$y_t = \beta x_t + error.$$

Clearly, there should be no systematic relationship between $y$ and $x$, and therefore, we should expect that an OLS estimate of $\beta$ should be close to zero, or insignificantly different from zero, at least as the sample size increases. But, as will be shown below, this is not the case. This phenomenon originated from Yule (1926) was called "a nonsense correlation."

**Example 1** *There are some famous examples for spurious correlation. One is that of Yule (1926, Journal of the Royal Statistical Society), reporting a correlation of 0.95 between the proportion of Church of England marriages to all marriages and the mortality rate over the period 1866-1911. Yet another is the econometric example reported by Hendry (1980, Economica) between the price level and the cumulative rainfall in the UK.*[2]

[2] *This relation proved resilient to many econometric diagnostic tests and was humorously advanced as a new theory of inflation.*

As we have come to understand in recent years, it is commonality of (stochastic) trending mechanisms in data that often leads to these spurious relations. What makes the phenomenon dramatic is that it occurs even when the data are otherwise independent.

In a prototypical spurious regression the fitted coefficients are statistically significant when there is no true relationship between the dependent variable and the regressors. Using Monte Carlo simulations Granger and Newbold (1974, *Journal of Econometrics*) showed this phenomenon. Phillips (1986, *Journal of Econometrics*) derived an analytic proof. These results are summarized in the following theorem:

**Theorem 1** ***(Spurious Regression)*** *Suppose that $y$ and $x$ are independent $I(1)$ variables generated respectively by*

$$y_t = y_{t-1} + \varepsilon_t,$$

$$x_t = x_{t-1} + e_t,$$

*where $\varepsilon_t \sim iid\,(0, \sigma_\varepsilon^2)$ and $e_t \sim iid\,(0, \sigma_e^2)$, and $\varepsilon_t$ and $e_t$ are independent of* *each other. Consider the regression,*

$$y_t = \beta x_t + u_t. \tag{1}$$

*Then, as $T \to \infty$,*

*(a) The OLS estimator of $\beta$ obtained from (1), denoted $\hat{\beta}$, does not converge to (true value of) zero.*

*(b) The t-statistic testing $\beta = 0$ in (1) diverges to $\pm$infinity.*

In sum, in the case of spurious regression, $\hat{\beta}$ takes any value randomly, and its t-statistic always indicates significance of the estimate. Though a formal testing procedure will be needed to detect evidence of the spurious regression or cointegration (see below), one useful guideline is that we are likely to face with the spurious relation when we find a highly significant t-ratio combined with a rather low value of $R^2$ and a low value of the Durbin-Watson statistic.

## 1.3 Cointegration

Economic theory often suggests that certain subset of variables should be linked by a long-run equilibrium relationship. Although the variables under consideration may drift away from equilibrium for a while, economic forces or government actions may be expected to restore equilibrium.

**Example 2** *Consider the market for tomatoes in two parts of a country, the north and the south with prices $p_{nt}$ and $p_{st}$ respectively. If these prices are equal the market will be in equilibrium. So $p_{nt} = p_{st}$ is called an attractor. If the prices are unequal it will be possible to make a profit by buying tomatoes in one region and selling them in the other. This trading mechanism will be inclined to equate prices again, raising prices in the buying region and lowering them in selling region.*

When the concept of equilibrium is applied to $I(1)$ variables, cointegration occurs; that is, cointegration is defined as a certain stationary linear combination of multiple $I(1)$ variables.

**Example 3** *Consider the consumption spending model. Although both consumption and income exhibit a unit root, over the long run consumption tends to be a roughly constant proportion of income, so that the difference between the log of consumption and log of income appears to be a stationary process.*

**Example 4** *Another well-known example is the theory of Purchasing Power Parity (PPP). This theory holds that apart from transportation costs, goods* *should sell for the same effective price in two countries. Let* $P_t$ *denote the index of the price level in US (in dollars per good),* $P_t^*$ *denote the price index for UK (in pounds per good), and* $S_t$ *the rate of exchange between the currencies (in dollars per pound). Then the PPP holds that* $P_t = S_t P_t^*$*, taking logarithm,* $p_t = s_t + p_t^*$*, where* $p_t = \ln(P_t)$*,* $s_t = \ln(S_t)$*,* $p_t^* = \ln(P_t^*)$*. In practice, errors in measuring prices, transportation costs and differences in quality prevent PPP from holding exactly at every date* $t$*. A weaker version of the hypothesis is that the variable* $z_t$ *defined by* $z_t = p_t - s_t - p_t^*$ *is stationary, even though the individual elements* $(p_t, s_t, p_t^*)$ *are all* $I(1)$*.*

Cointegration brings with it two obvious econometric questions. The first is how to estimate the cointegrating parameters and the second is how to test whether two or more variables are cointegrated or spurious.

We first examine estimation of the cointegrating regression. Consider the simple time series regression,[3]

$$y_t = \beta x_t + u_t, \tag{2}$$

[3] The deterministic regressors such as intercept and a linear time trend can be easily accommodated in the regression without changing the results in what follows.

where $x_t$ is an $I(1)$ variable given by

$$x_t = x_{t-1} + e_t. \tag{3}$$

Since $x_t$ is $I(1)$, it follows that $y_t$ is $I(1)$. But, for $y_t$ and $x_t$ to be cointegrated, their linear combination, $u_t = y_t - \beta x_t$ must be stationary. Thus, we assume:

**Assumption 2.1.** $u_t$ is *iid* process with mean zero and variance $\sigma_u^2$.

**Assumption 2.2.** $e_t$'s are stationary and independently distributed of $u_t$.

Assumption 2.1 ensures that there exists a stationary cointegrating relationship between $y$ and $x$. Assumption 2.2 implies that $x_t$ is exogenous, *i.e.*, $E(x_t u_t) = 0$, which has been one of the standard assumption.1

**Theorem 2** ***(Cointegrating Regression)*** *Consider the OLS estimator of* $\beta$ *obtained from (2). Under Assumptions 2.1 and 2.2, as* $T \to \infty$*,*

*(a) The OLS estimator* $\hat{\beta}$ *is consistent; that is,* $\hat{\beta} \to_p \beta$*, and* $T\left(\hat{\beta} - \beta\right)$ *has an asymptotic normal distribution.*

*(b) The t-statistic testing* $\beta = \beta_0$ *in (2) converges to a standard normal random variable.*

This theorem clearly indicates that most estimation and inference results as obtained for the regression with stationary variables can be extended to the cointegrating regression.

The result in theorem can be readily extended to a multiple cointegrating regression with $k$ regressors:

$$y_t = \boldsymbol{\beta}'\mathbf{x}_t + u_t,$$

where $\mathbf{x}_t = (x_{1t}, x_{2t}, \ldots, x_{kt})'$ is a $k$-dimensional $I(1)$ regressors given by

$$\mathbf{x}_t = \mathbf{x}_{t-1} + \mathbf{e}_t,$$

where $\mathbf{e}_t = (e_{1t}, e_{2t}, \ldots, e_{kt})'$ are $k$-dimensional stationary disturbances, and independently distributed of $u_t$. In this case we need one additional condition:

**Assumption 2.3.** $\mathbf{x}_t$'s not cointegrated among themselves.

Violation of Assumption 2.3 means that there may be more than one cointegrating relations. Such a case cannot be covered in a single equation approach, and will be covered in the system VAR approach to cointegration. Assumption 2.3 is similar to the multicollinearity assumption made in a multiple regression with stationary regressors.

In this more general case the OLS estimator of $\boldsymbol{\beta}$ is also consistent, and has an asymptotic normal distribution such that multiple restrictions on $\boldsymbol{\beta}$ can be tested in a standard way using the Wald statistic, which is asymptotically $\chi^2$ distributed.

### 1.3.1 Residual-based Test for Cointegration

We find that the fundamentally different conclusion is made between spurious regression and cointegration. Therefore, the detection of cointegration is very important in practice prior to estimation.

One of most popular tests for (a single) cointegration has been suggested by Engle and Granger (1987, *Econometrica*). Consider the multiple regression:

$$y_t = \boldsymbol{\beta}'\mathbf{x}_t + u_t, \ t = 1, ..., T, \tag{4}$$

where $\mathbf{x}_t = (x_{1t}, x_{2t}, \ldots, x_{kt})'$ is the $k$-dimensional $I(1)$ regressors. Notice that for $y_t$ and $\mathbf{x}_t$ to be cointegrated, $u_t$ must be $I(0)$. Otherwise it is spurious. Thus, a basic idea behind is to test whether $u_t$ is $I(0)$ or $I(1)$.

The Engle and Granger cointegration test is carried out in two steps:[4]

[4] Of course you need to pretest the individual variables for unit roots. By definition of

1. Run the OLS regression of (4) and obtain the residuals by

$$\hat{u}_t = y_t - \hat{\boldsymbol{\beta}}' \mathbf{x}_t,\ t = 1, ..., T,$$

where $\hat{\boldsymbol{\beta}}$ are the OLS estimate of $\boldsymbol{\beta}$.

2. Apply a unit root test to $\hat{u}_t$ by constructing an AR(1) regression for $\hat{u}_t$:

$$\hat{u}_t = \phi \hat{u}_{t-1} + \varepsilon_t. \tag{5}$$

That is, do the DF t-test of $H_0 : \phi = 1$ against $H_1 : \phi < 1$ in (5).[5]

This is called the residual-based EGDF cointegration test. Strictly, it is the test of no-cointegration, because the null of unit root in $\hat{u}_t$ implies that there is no-cointegration between $y$ and $\mathbf{x}$. So if you reject $H_0 : \phi = 1$ in (5), you may conclude that there is a cointegration and *vice versa*.

Notice that the asymptotic distribution of t-statistic for $\phi = 1$ in (5) is also non-standard, but more importantly different from that of the univariate DF unit test. Main difference stems from the fact that one needs to allow

cointegration you need to ascertain that all the variable involved are $I(1)$.

[5]Since $\hat{u}_t$ is a zero-mean residual process, there is no need to include an intercept term here.

for estimation uncertainty through $\hat{\boldsymbol{\beta}}$ in the first step. The resulting test distribution thus depends on the dimension of the regressors, $k$.

As in the case of the univariate unit root test, there are the three specifications with different deterministic components:

$$y_t = \boldsymbol{\beta}'\mathbf{x}_t + v_t, \tag{6}$$

$$y_t = a_0 + \boldsymbol{\beta}'\mathbf{x}_t + v_t, \tag{7}$$

$$y_t = a_0 + a_1 t + \boldsymbol{\beta}'\mathbf{x}_t + v_t, \tag{8}$$

where $a_0$ is an intercept and $a_1$ is the linear trend coefficient. The three sets of critical values of the EGDF tests have been provided by Engle and Yoo (1987, *Journal of Econometrics*) and Phillips and Ouliaris (1990, *Econometrica*) for different values of $k$.[6]

Since the serial correlation is also often a problem in practice, it is common to use an augmented version of EGDF test; that is, extend (5) to

$$\Delta\hat{u}_t = \varphi\hat{u}_{t-1} + \sum_{i=1}^{p-1}\gamma_i\Delta\hat{u}_{t-1} + \varepsilon_t,$$

[6] For example, Microfit provides the corresponding critical values when estimating and testing.

and do the t-test for $\varphi = 0$.

Notice that all the problems that afflict the unit root tests also afflict the residual-based cointegration tests. In particular, the asymptotic critical values may be seriously misleading in small samples. Unfortunately, the cointegration tests are often severely lacking in power especially because of the imprecision or uncertainty of estimating $\beta$ in the first step. Thus, failure to reject the null of no-cointegration is common in application, which may provide only weak evidence that two or more variables are not cointegrated.

## 1.4 Error Correction Model

The cointegrating regression so far considers only the long-run property of the model, and does not deal with the short-run dynamics explicitly.[7] Clearly, a good time series modelling should describe both short-run dynamics and the long-run equilibrium simultaneously. For this purpose we now develop an error correction model (ECM). Although ECM has been popularized after Engle and Granger, it has a long tradition in time series econometrics dating back to Sargan (1964) or being embedded in the London School of Economics

[7] Here the long-run relationship measures any relation between the level of the variables under consideration while the short-run dynamics measure any dynamic adjustments between the first-difference's of the variables.

tradition.

To start, we define the error correction term by

$$\xi_t = y_t - \beta x_t,$$

where $\beta$ is a cointegrating coefficient. In fact, $\xi_t$ is the error from a regression of $y_t$ on $x_t$. Then an ECM is simply defined as

$$\Delta y_t = \alpha \xi_{t-1} + \gamma \Delta x_t + u_t, \tag{9}$$

where $u_t$ is *iid*. The ECM equation (9) simply says that $\Delta y_t$ can be explained by the lagged $\xi_{t-1}$ and $\Delta x_t$. Notice that $\xi_{t-1}$ can be thought of as an equilibrium error (or disequilibrium term) occurred in the previous period. If it is non-zero, the model is out of equilibrium and *vice versa*.

**Example 5** *Consider the simple case where $\Delta x_t = 0$. Suppose that $\xi_{t-1} > 0$, which means that $y_{t-1}$ is too high above its equilibrium value, so in order to restore equilibrium, $\Delta y_t$ must be negative. This intuitively means that the error correction coefficient $\alpha$ must be negative such that (9) is dynamically stable. In other words, if $y_{t-1}$ is above its equilibrium, then it will start falling in the next period and the equilibrium error will be corrected in the model,* *hence the term error correction model.*

Notice that $\beta$ is called the long-run parameter, and $\alpha$ and $\gamma$ are called short-run parameters. Thus the ECM has both long-run and short-run properties built in it. The former property is embedded in the error correction term $\xi_{t-1}$ and the short-run behavior is partially but crucially captured by the error correction coefficient, $\alpha$. All the variables in the ECM are stationary, and therefore, the ECM has no spurious regression problem.

**Example 6** *Express the (static) cointegrating model (2) in an ECM form. First, notice that*

$$y_t - (y_{t-1} - y_{t-1}) = \beta x_t + \beta (x_{t-1} - x_{t-1}) + u_t,$$

$$\Delta y_t = -(y_{t-1} - \beta x_{t-1}) + \beta \Delta x_t + u_t.$$

*Therefore, the associated error correction model becomes*

$$\Delta y_t = -\xi_{t-1} + \beta \Delta x_t + u_t,$$

*where the error correction coefficient is -1 by construction, meaning the perfect adjustment or error correction is made every period, which is unduly* *restrictive and unlikely to happen in practice.*

In general, the error correction term $\xi_{t-1}$ is unknown *a priori*, and needs to be estimated. In the case of cointegration the following Engle and Granger two-step procedure can be used:

1. Run a (cointegrating) regression of $y$ on $x$ and save the residuals, $\hat{\xi}_t = y_t - \hat{\beta} x_t$.

2. Run an ECM regression of $\Delta y$ on $\hat{\xi}_t$ and $\Delta x$,

$$\Delta y_t = \alpha \hat{\xi}_{t-1} + \gamma \Delta x_t + u_t.$$

**Example 7** *ECM model is also used in the Present Value (PV) model for stocks. A PV model relates the price of a stock to the discounted sum of its expected future dividends. It's noted that 1. stock prices and dividends must be cointegrated if dividends are I(1); 2. Persistent movements in dividends have much larger effects on price than transitory movements. Thus an approximate PV model with time-varying expected returns are introduced, an ECM model. This is also called the dynamic Gordon model in finance.*

**Remark 1** *We only cover single equation modelling and testing for 1 coin-*

*tegration relationship. When there are more than 2 variables, there could be more than 1 cointegration relations among them. Then a single equation analysis cannot provide information about the number of cointegrating relations and cannot test several cointegration relations jointly. So we rely on a system analysis which could solve the two problems. Namely, we use a VAR (Vector Autoregressive) model and use Johansen's test to test the reduced rank restriction on the coefficients of a VAR.*

# 2

## 2.1 ARDL Modelling Approach to Cointegration Analysis

In time series analysis the explanatory variable may influence the dependent variable with a time lag. This often necessiates the inclusion of lags of the explanatory variable in the regression. Furthermore, the dependent variable may be correlated with lags of itself, suggesting that lags of the dependent variable should be included in the regression as well. These considerations motivate the commonly used ARDL$(p, q)$ model defined as follow:[8]

$$y_t = \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \theta_0 x_t + \theta_1 x_{t-1} + \cdots + \theta_q x_{t-q} + u_t, \tag{10}$$

where $u_t \sim iid\,(0, \sigma^2)$.

In the case where the variables of interest are trend stationary, the general practice has been to de-trend the series and to model the de-trended series as stationary distributed lag or autoregressive distributed lag (ARDL) models. Estimation and inference concerning the properties of the model are then carried out using standard asymptotic normal theory.

However, the analysis becomes more complicated when the variables are difference-stationary, or $I(1)$.The recent literature on cointegration is concerned with the analysis of the long run relations between $I(1)$ variables, and its basic premise is, at least implicitly, that in the presence of $I(1)$ variables the traditional ARDL approach seems no longer applicable.[9]

Pesaran and Shin (1999) recently re-examined the traditional ARDL ap-

[8] For convenience we do not include the deterministic regressors such as constant and linear time trend.

[9] Consequently, a large number of alternative estimation and hypothesis testing procedures have been specifically developed for the analysis of $I(1)$ variables. See Phillips and Hansen (1990, *Review of Economic Studies*) and Phillips and Loretan (1991, *Review of Economic Studies*).

proach for an analysis of a long run relationship when the underlying variables are $I(1)$, and find:

1. The ARDL-based estimators of the long-run coefficients are also consistent, and have an asymptotic normal distribution.

2. Valid inferences via the Wald on the cointegrating parameters can be made using the standard $\chi^2$ asymptotic distribution.

We illustrate this approach using an ARDL(1,1) regression with an $I(1)$ regressor,

$$y_t = \phi y_{t-1} + \theta_0 x_t + \theta_1 x_{t-1} + u_t,\ t = 1, ..., T, \tag{11}$$

where $\phi$, $\theta_0$ and $\theta_1$ are unknown parameters, and $x_t$ is an $I(1)$ process generated by

$$x_t = x_{t-1} + e_t, \tag{12}$$

We make the following assumptions :

**A1.** $u_t \sim iid(0, \sigma_u^2)$.

**A2.** $e_t$ is a general linear stationary process.

**A3.** $u_t$ and $e_t$ are uncorrelated for all lags such that $x_t$ is strictly exogenous with respect to $u_t$.

**A4.** (**Stability Condition**) $|\phi| < 1$, so that the model is dynamically stable.

Assumption A4 is similar to the stationarity condition for an AR(1) process and implies that there exists a stable long-run relationship between $y_t$ and $x_t$.[10]

Suppose that a long-run equilibrium occurs at $y_t = y^*$, $x_t = x^*$ and $u_t = 0$ for all $t$. Then, (11) becomes

$$y^* = \phi y^* + \theta_0 x^* + \theta_1 x^*,$$

and therefore,

$$y^* = \beta x^*,$$

where $\beta = \frac{\theta_0+\theta_1}{1-\phi}$ is called the long-run multiplier.

We derive an ECM directly from (11). First, transform (11) to

$$\begin{aligned} \Delta y_t &= -(1-\phi)\, y_{t-1} + (\theta_0 + \theta_1)\, x_{t-1} + \theta_0 \Delta x_t + u_t \qquad (13) \\ &= \alpha y_{t-1} + \theta x_{t-1} + \psi \Delta x_t + u_t, \end{aligned}$$

where $\alpha = -(1-\phi)$, $\theta = \theta_0 + \theta_1$ and $\psi = \theta_0$. Noting that $\alpha < 0$ under

[10] If $\phi = 1$, then there would be no long-run relationship.

Assumption A4, we further obtain

$$\Delta y_t = \alpha \left( y_{t-1} - \beta x_{t-1} \right) + \psi \Delta x_t + u_t, \tag{14}$$

which is an ECM. $\beta$ can also be regarded as a long-run cointegrating parameter. This shows that the ARDL specification is indeed useful to characterize both the long-run and short-run behavior. Notice that (11), (13) and (14) are the same but represented differently. We may call (13) and (14) as unrestricted ECM and (restricted) ECM, respectively. From (13) the long-run parameter is also obtained by

$$\beta = -\frac{\theta}{\alpha}. \tag{15}$$

**Theorem 3** *(**ARDL**) Consider the unrestricted error correction model, (13). Under assumptions A1 - A4,*

*(a) The ARDL-based estimators of the long-run parameter, given by $\hat{\beta} = -\hat{\theta}/\hat{\alpha}$, converge to $\beta = -\theta/\alpha$ as $T \to \infty$ and $T\left(\hat{\beta} - \beta\right)$ has the limiting normal distribution.*

*(b) The t-statistic testing $\beta = \beta_0$ converges to the standard normal distribution.*

We also find that the OLS estimators of all of the short-run parameters, $\alpha$, $\theta$, $\psi$ are $\sqrt{T}$-consistent and have the normal distribution.

In a more general case where $u_t$ and $e_t$ in (13) are serially correlated, the ARDL specification needs to be augmented with adequate number of lagged changes in the dependent and the independent variables before estimation and inference is carried out: that is, we now consider the ARDL$(p, q)$ or unrestricted ECM model,

$$\Delta y_t = \alpha y_{t-1} + \theta x_{t-1} + \sum_{j=1}^{p-1} \gamma_j \Delta y_{t-j} + \sum_{j=0}^{q-1} \psi_j \Delta x_{t-j} + u_t. \tag{16}$$

To estimate an ECM based on an ARDL approach, Pesaran and Shin (1999) suggested the following three-step procedure:

1. Use model selection criteria such as AIC and SBC, and choose the orders of the ARDL model.

2. Run an ARDL$(p, q)$ or unrestricted ECM regression (16),then save the error correction term by $\hat{\xi}_t = y_t - \hat{\beta} x_t$, where $\hat{\beta} = -\frac{\hat{\theta}}{\hat{\alpha}}$ and $\hat{\alpha}$ and $\hat{\theta}$ are the OLS estimators obtained from (16).

3. Run an ECM regression of $\Delta y$ on $\hat{\xi}_t$ , $\Delta x$, $p-1$ lagged $\Delta y$'s and $q-1$ lagged $\Delta y$'s,

$$\Delta y_t = \alpha \hat{\xi}_{t-1} + \sum_{j=1}^{p-1} \gamma_j \Delta y_{t-j} + \sum_{j=0}^{q-1} \psi_j \Delta x_{t-j} + u_t.$$

## 2.2 Vector Autoregressive Model

1. Given $y_t = \begin{bmatrix} x_t \\ z_t \end{bmatrix}$ and $e_t = \begin{bmatrix} u_t \\ v_t \end{bmatrix} \sim iidN(0, \Sigma)$

   $y_t = \varphi y_{t-1} + e_t \Longrightarrow \begin{bmatrix} x_t \\ z_t \end{bmatrix} = \begin{bmatrix} \varphi_{xx} & \varphi_{xz} \\ \varphi_{zx} & \varphi_{zz} \end{bmatrix} \begin{bmatrix} x_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} u_t \\ v_t \end{bmatrix}$

   In general, $A(L)y_t = B(L)e_t$ where

   $A(L) = 1 - \Phi_1 L - \Phi_2 L^2 - ..... - \Phi_p L^p$

   $B(L) = 1 + \Theta_1 L + \Theta_2 L^2 + ..... + \Theta_p L^p$

   $\Phi_j = \begin{bmatrix} \varphi_{j,xx} & \varphi_{j,xz} \\ \varphi_{j,zx} & \varphi_{j,zz} \end{bmatrix}$

   Given an $AR(1)$ representation,

   $A(L)y_t = e_t \Longrightarrow y_t = A(L)^{-1}e_t \Longrightarrow \begin{bmatrix} a_{xx}(L) & a_{xz}(L) \\ a_{zx}(L) & a_{zz}(L) \end{bmatrix} \begin{bmatrix} x_t \\ z_t \end{bmatrix} =$

$$\begin{bmatrix} u_t \\ v_t \end{bmatrix}$$

$$\Longrightarrow \begin{bmatrix} x_t \\ z_t \end{bmatrix} = [a_{xx}(L)a_{zz}(L) - a_{xz}(L)a_{zx}(L)]^{-1} \begin{bmatrix} a_{xx}(L) & -a_{xz}(L) \\ -a_{zx}(L) & a_{zz}(L) \end{bmatrix} \begin{bmatrix} u_t \\ v_t \end{bmatrix}$$

Similarly a process like an $ARMA(2,1)$ such as:

$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + u_t + \theta_1 u_{t-1}$ can be represented as:

$$\begin{bmatrix} y_t \\ y_{t-1} \\ u_t \end{bmatrix} = \begin{bmatrix} \varphi_1 & \varphi_2 & \theta_1 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ y_{t-2} \\ u_{t-1} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} w_t \Longrightarrow x_t = Ax_{t-1} + Cw_t$$

It is sometimes convenient to define the $C$ matrix in such a way so the variance-covariance matrix of the shocks is the identity matrix. Then

$$C = \begin{bmatrix} \sigma_e \\ 0 \\ \sigma_e \end{bmatrix} \text{ and } E(w_t w_t^{/}) = I$$

### 2.2.1 *Forecast*

We start with a vector $AR(1)$ representation which can have a $MA(\infty)$ representation as:

$$y_t = \sum_{j=0}^{\infty} A^j C w_{t-j}$$

$$E_t(y_{t+k}) = A^k y_t$$

$y_{t+1} - E_t(y_{t+1}) = Cw_{t+1}$

$V_t(y_{t+1}) = CC^{/}$

$y_{t+2} - E_t(y_{t+2}) = Cw_{t+2} + ACw_{t+1}$

$V_t(y_{t+2}) = CC^{/} + ACC^{/}A^{/}$

$V_t(y_{t+j}) = \sum_{j=0}^{k-1} A^j CC^{/} A^{j/}$

$E_t(y_{t+k}) = AE_t(y_{t+k-1})$

$V_t(y_{t+k}) = CC^{/} + AV_t(y_{t+k-1})A^{/}$

Suppose, $y_t = \begin{bmatrix} x_t \\ z_t \end{bmatrix}$, then

$$\begin{aligned} x_t &= \varphi_{xx1}x_{t-1} + \varphi_{xx2}x_{t-1} + ... + \varphi_{xz1}z_{t-1} + \varphi_{xz2}x_{t-2} + ... + e_{xt} \qquad (17) \\ z_t &= \varphi_{zx1}x_{t-1} + \varphi_{zx2}x_{t-1} + ... + \varphi_{zz1}z_{t-1} + \varphi_{zz2}x_{t-2} + ... + e_{zt} \end{aligned}$$

$AR(1)$ mapping of the above equation gives:

$$\begin{bmatrix} x_t \\ z_t \\ x_{t-1} \\ z_{t-1} \\ .. \end{bmatrix} = \begin{bmatrix} \varphi_{xx1} & \varphi_{xz1} & \varphi_{xx2} & \varphi_{xz2} & .. \\ \varphi_{zx1} & \varphi_{zz1} & \varphi_{zx2} & \varphi_{zz2} & .. \\ 1 & 0 & 0 & . & .. \\ 0 & 1 & 0 & . & .. \\ .. & .. & .. & .. & .. \end{bmatrix} \begin{bmatrix} x_{t-1} \\ z_{t-1} \\ x_{t-2} \\ z_{t-2} \\ .. \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \\ .. & .. \end{bmatrix} \begin{bmatrix} e_{xt} \\ e_{zt} \end{bmatrix}$$

$\Longrightarrow y_t = Ay_{t-1} + e_t$

We can also start with:

$y_t = \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + .. + e_t$

$$\begin{bmatrix} y_t \\ y_{t-1} \\ y_{t-2} \\ .. \end{bmatrix} = \begin{bmatrix} \Phi_1 & \Phi_2 & .. & .. \\ I & 0 & .. & .. \\ 0 & I & .. & .. \\ .. & .. & .. & .. \end{bmatrix} \begin{bmatrix} y_{t-1} \\ y_{t-2} \\ y_{t-3} \\ .. \end{bmatrix} + \begin{bmatrix} I \\ 0 \\ 0 \\ .. \end{bmatrix} [e_t]$$

Given the above $AR(1)$ representation, we can forecats both $x$ and $z$. We can also add a small change in the above formulation by choosing the $C$ matrix such that the shocks are orthogonal to each other. This implies that $E(ee^{/}) = I$.

### 2.2.2 *Impulse Response*

The impulse response is that path $y$ follows if it is affected by a single unit shock in $e_t$, i.e., $e_{t-j} = 0, e_t = 1, e_{t+j} = 0$. It is interesting as it allows us to think about 'cause' and 'effects'. For example, we may want to compute the response of stock price to a shock in interest rate in Stock-Price - Interest Rate VAR model and interpret the result as the effect of interest rate on stock price.

$AR(1)$ Model: $y_t = \varphi y_{t-1} + e_t \Longrightarrow y_t = \sum\limits_{j=0}^{\infty} \varphi^j e_{t-j}$

$e_t : 0 \quad 0 \quad 1 \quad 0 \quad 0 \quad 0 \quad 0$

$y_t : 0 \quad 0 \quad 1 \quad \varphi \quad \varphi^2 \quad \varphi^3 \quad \varphi^4$

$MA(\infty)$ Model: $y_t = \sum_{j=0}^{\infty} \theta_j u_{t-j}$

$u_t : 0 \quad 0 \quad 1 \quad 0 \quad 0 \quad 0 \quad 0$

$y_t : 0 \quad 0 \quad 1 \quad \theta_1 \quad \theta_2 \quad \theta_3 \quad \theta_4$

Given the above formulation, the vector process works in the same way. If an vector $MA(\infty)$ Model has the following representation:

$y_t = \Theta(L)u_t$ where $u_t \equiv [u_{xt}, u_{zt}]^/$ and $\Theta(L) = \theta_0 + \theta_1 L + ....$, then $\{\theta_0, \theta_1, ...\}$ define the impulse-response functions.

$$\Theta(L) = \begin{bmatrix} \theta_{xx}(L) & \theta_{xz}(L) \\ \theta_{zx}(L) & \theta_{zz}(L) \end{bmatrix}$$

Therefore, $\theta_{xx}(L)$ gives the response of $x_{t+k}$ to a unit shock in $u_{xt}$, and $\theta_{xz}(L)$ gives the response of $x_{t+k}$ to a unit shock in $u_{zt}$. In practice, however. we map this to a vector $AR(1)$ representation as it is more convenient to calculate the imp[ulse response.

$y_t = Ay_{t-1} + Ce_t$ and therefore the impulse-respponse function is: $C$, $AC$, $A^2C$,....$A^kC$,.....

### 2.2.3 *Orthogonalization*

$$
\begin{aligned}
A(L)y_t &= e_t, A(0) = I, E(e_t e_t^{/}) = \Sigma && (18)\\
y_t &= B(L)e_t, B(0) = I, E(e_t e_t^{/}) = \Sigma \text{ where } B(L) = A(L)^{-1}
\end{aligned}
$$

Suppose we want to examine the response of $y_t$ to unit movements in $e_{xt}$ and $e_{zt} + 0.3e_{xt}$. We call the new shock as $\eta_{1t} = e_{xt}$ and $\eta_{2t} = e_{zt} + 0.3e_{xt}$. We define $\eta_t = Qe_t$ where $Q = \begin{bmatrix} 1 & 0 \\ 0.3 & 1 \end{bmatrix}$. Then we get the following:

$y_t = B(L)Q^{-1}Qe_t = C(L)\eta_t$ where $C(L) = B(L)Q^{-1}$; $C(L)$ gives the response of $y_t$ to the new shocks $\eta_t$. Note that $y_t = B(L)e_t$ and $y_t = C(L)\eta_t$ are observationally equivalent, since they produce the same series $y_t$. We, then, need to decide which linear combinations is the most interesting. In order to do so, we state a set of assumptions, called orthogonalization assumptions, which uniquely determine the linear combinations of shocks that we find most interesting.

**Orthogonal Shocks:** The first assumption is that the shocks should be orthogonal. Let us start with the example: 'effect of interest rate on stock prices'. If the interest rate shock is correlated with the stock price shock, then we we do not know if the response that we observe is the response of stock price to interest rate, or to an announcement shock (in terms of dividends) that happen to come at the same time as the interest rate shock (it may be because the Reserve Bank of India sees the announcement shock and accomodates it). Additionally, it is convenient to rescale the shocks so that they have a unit variance. Thus we want to choose $Q$ such that $E(\eta_t\eta_t^{/}) = I$. To do this we need a $Q$ such that $Q^{-1}Q^{-1/} = \Sigma$, $E(\eta_t\eta_t^{/}) = E(Qe_te_t^{/}Q^{/}) = Q\Sigma Q^{/} = I$. However, there may be many different $Q$'s that can act as the square root matrices for $\Sigma$. Given one such $Q$, we can have another defined as $Q^* = MQ$ where $M$ is the orthogonal matrix $MM^{/} = I$, $Q^*Q^{*/} = MQQ^{/}M^{/} = MM^{/} = I$. Since $C(L) = B(L)Q^{-1}$, we specify a desired property of $C(L)$, which help us to pin down $Q$.

**Sims' Orthogonalization:** Sims (1980) suggested that we specify properties of $C(0)$ that gives the instantaneous response of each variable to each orthogonalized shock $\eta$. Sims suggested a lower triangular $C(0)$:

$$\begin{bmatrix} x_t \\ z_t \end{bmatrix} = \begin{bmatrix} C_{0xx} & 0 \\ C_{0zx} & C_{0zz} \end{bmatrix} \begin{bmatrix} \eta_{xt} \\ \eta_{zt} \end{bmatrix} + C_1\eta_{t-1} + ...$$

The above spefication implies that $\eta_{2t}$ does not affect the first variable, $x_t$, contemporaneously. Both shocks can affect $z_t$ contemporaneously. In the orginal VAR-model, $A(0) = I$ implying that contemporaneous values of each variable do not appear in the other varoable's equation. A lower triangular matrix $C(0)$ implies that contemporaneous $x_t$ appears in the $z_t$ equation, but $z_t$ does not appear in the $x_t$ equation.

Let us denote the orthogonalized autoregressive representation: $D(L)y_t = \eta_t$ where $D(L) = C(L)^{-1}$. As $C(0)$ is lower traingular $D(0)$ is also lower traingular.

$$\begin{bmatrix} D_{0xx} & 0 \\ D_{0zx} & D_{0zz} \end{bmatrix} \begin{bmatrix} x_t \\ z_t \end{bmatrix} + D_1 y_{t-1} + ... = \eta_t$$

Assuming an $AR(1)$ process we get:

$$\begin{aligned} D_{0xx} x_t &= -D_{1xx} x_{t-1} - D_{1xz} z_{t-1} + \eta_{xt} \qquad (19) \\ D_{0zz} z_t &= -D_{0zx} x_t - D_{1zx} x_{t-1} - D_{1zz} z_{t-1} + \eta_{zt} \end{aligned}$$

Note the above is numerically equivalent to estimate the system by OLS with contemporaneous $x_t$ in the $z_t$ equation, but not the vice versa, and then scaling each equation so that the eroor variance is one. Recall that OLS estimates produce residuals that are uncorrelated with the RHS variables by construction. Hence if we ran OLS as:

$$
\begin{aligned}
x_t &= a_{1xx}x_{t-1} + .. + a_{1xz}z_{t-1} + \eta_{xt} \\
z_t &= a_{0zx}x_t + a_{1zx}x_{t-1} + .. + a_{1zz}z_{t-1} + \eta_{zt}
\end{aligned} \tag{20}
$$

The first OLS residual is $\eta_{xt} = x_t - E(x_t \mid x_{t-1}, .., z_{t-1})$ and hence a linear combination of $\{x_t, x_{t-1}, .., z_{t-1}, ...\}$. OLS residuals are orthogonal to the RHS variables, so $\eta_{zt}$ is orthogonal to any linear combination of $\{x_t, x_{t-1}, .., z_{t-1}, ...\}$ by construction. Hence, $\eta_{xt}$ and $\eta_{zt}$ are uncorrelated with each other. The term $a_{0zx}$ captures all of the contemporaneous correlation of news in $x_t$ and that in $x_t$.[11]

### 2.2.4 *Variance Decomposition*

The purpose is to compute an accounting of the forecast error variance in the orthogonalized system: what percent of the k-step ahead forecast error variance is due to which variable.

$y_t = C(L)\eta_t$, $E(\eta_t\eta_t^{/}) = I$

The 1-step ahead forecast error variance is:

[11] $C(L) = B(L)Q^{-1}$
$C(0) = B(0)Q^{-1}$
This implies that the Choleski decomposition produces a lower triangular $Q$.

$$e_t = y_{t+1} - E_t(y_{t+1}) = C(0)\eta_{t+1} = \begin{bmatrix} c_{xx,0} & c_{xz,0} \\ c_{zx,0} & c_{zz,0} \end{bmatrix} \begin{bmatrix} \eta_{x,t+1} \\ \eta_{z,t+1} \end{bmatrix}$$

where

$C(L) = C(0) + C(1)L + C(2)L^2 + ...$ and the elements of $C(L)$ as $c_{xx,0} + c_{xx,1}L + c_{xx,2}L^2 + ....$ Since, the $\eta$s are uncorrelated and have unit variance:

$V_t(x_{t+1}) = c^2_{xx,0}\sigma^2(\eta_x) + c^2_{xy,0}\sigma^2(\eta_z) = c^2_{xx,0} + c^2_{xy,0}$ and similarly for $z$. Thus $c^2_{xx,0}$ gives the amount of 1-step ahead forecast error variance of $x$ due to the $\eta_x$ shock and $c^2_{xx,0}$ due to the $\eta_z$ shock. One in general reports them as fractions: $\frac{c^2_{xx,0}}{c^2_{xx,0}+c^2_{xx,0}}, \frac{c^2_{zz,0}}{c^2_{zx,0}+c^2_{xx,0}}$. Hence, we can write:

$V_t(x_{t+1}) = C_0C_0^{/}$, $I_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ and $I_2 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$

Thus, the part of the 1-step ahead forecast error variance due to the first ($x$) shock is $C_0I_1C_0^{/}$ and the part due to the second ($z$) shock is $C_0I_2C_0^{/}$. We also get: $C_0C_0^{/} = C_0I_1C_0^{/} + C_0I_2C_0^{/}$ . If we think $I_m$ as a new covariance matrix where all shocks but the m-th are turned off then the total variance of forecast errors must be equal to the part due to the m-th shock, given by $C_0I_mC_0^{/}$ . Therefore,

$y_{t+k} - E_t(y_{t+k}) = C(0)\eta_{t+k} + C(1)\eta_{t+k-1} + ... + C(k-1)\eta_{t+1}$

$V_t(y_{t+k}) = C_0C_0^{/} + C_1C_1^{/} + .... + C_{k-1}C_{k-1}^{/} \Longrightarrow V_{k,m} = \sum_{j-0}^{k-1} C_jI_mC_j^{/}$ is the variance of $k$-step ahead forecast due to the m-th shock, and the variance is the sum of these components, $V_t(y_{t+k}) = \sum_m V_{k,m}$. On the other hand, note $\eta_t = Qe_t$, therefore we can get:

$$y_t = \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + .. + e_t$$

$$\begin{bmatrix} y_t \\ y_{t-1} \\ y_{t-2} \\ .. \end{bmatrix} = \begin{bmatrix} \Phi_1 & \Phi_2 & .. & .. \\ I & 0 & .. & .. \\ 0 & I & .. & .. \\ .. & .. & .. & .. \end{bmatrix} \begin{bmatrix} y_{t-1} \\ y_{t-2} \\ y_{t-3} \\ .. \end{bmatrix} + \begin{bmatrix} Q^{-1} \\ 0 \\ 0 \\ .. \end{bmatrix} [\eta_t]$$

$$\Longrightarrow y_t = Ay_{t-1} + C\eta_t,\ E(\eta_t \eta_t^{/}) = I$$

The impulse-response function is $C$, $AC$, $A^2C$,..... and can be found recursively from:

$$IR_0 = C,\ IR_j = AIR_{j-1}$$

If $Q^{-1}$ is lower diagonal, then only the 1st shock affects the 1st variable:

$$V_t(y_{t+k}) = \sum_{j-0}^{k-1} A^j CC^{/} A^{j/} \Longrightarrow V_{k,m} = \sum_{j-0}^{k-1} A^j CI_m C^{/} A^{j/}$$

**Causality** $Y = X\beta + e$ does not imply "$x$ causes" $y$. The most natural definition of cause: "cause should preced effects. However, this may not be true in Time-Series.

**Definition 1** *: Granger Causality*

$x_t$ Granger causes $y_t$ if $x_t$ helps to forecast $y_t$ , given past $y_t$. Consider the following:

$$\begin{aligned} y_t &= a(L)y_{t-1} + b(L)x_{t-1} + u_t \\ x_t &= c(L)y_{t-1} + d(L)x_{t-1} + v_t \end{aligned}$$

Our definition inplies that $x_t$ does not Granger cause $y_t$ if $b(L) = 0$, i.e. the above system of equation s can be represented as:

$$\begin{aligned} y_t &= a(L)y_{t-1} + u_t \\ x_t &= c(L)y_{t-1} + d(L)x_{t-1} + v_t \end{aligned}$$

Alternatively,

$$\begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} a(L) & b(L) \\ c(L) & d(L) \end{bmatrix} \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + \begin{bmatrix} u_t \\ v_t \end{bmatrix}$$

$$\Longrightarrow \begin{bmatrix} I - La(L) & -Lb(L) \\ -Lc(L) & I - Ld(L) \end{bmatrix} \begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} u_t \\ v_t \end{bmatrix}$$

$$\Longrightarrow \begin{bmatrix} a^*(L) & b^*(L) \\ c^*(L) & d^*(L) \end{bmatrix} \begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} u_t \\ v_t \end{bmatrix}$$

Thus $x$ does not Granger cause $y$ iff $b^*(L) = 0$.

We can also have a $MA$ representation to explain the concept of Granger causality.

$$\begin{bmatrix} y_t \\ x_t \end{bmatrix} = \frac{1}{a^*(L)d^*(L)-b^*(L)c^*(L)} \begin{bmatrix} d^*(L) & -b^*(L) \\ -c^*(L) & a^*(L) \end{bmatrix} \begin{bmatrix} u_t \\ v_t \end{bmatrix}$$

This implies that $x$ does not Granger cause $y$ if the Wold moving average matrix lag polynomial is lower triangular. If $x$ does not Granger cause $y$, then y is a function of its own shocks only, and does not respond to $x$ shocks. $x$ on the other hand, is a function of both $y$ shocks and $x$ shocks. This also implies:

$x$ does not Granger cause $y$ if the $y$'s bivariate Wold representation is the same as the univariate Wold representation. Therefore, $x$ does not Granger cause $y$ if the projection of $y$ on past $y$ and $x$ is the same as the projection of $y$ on past $y$ alone.

Suppose,

$$\begin{aligned} y_t &= e(L)\eta_t \text{ where } \eta_t = y_t - P(y_t|y_{t-1}, y_{t-2}, ...) \\ x_t &= f(L)\lambda_t \text{ where } \lambda_t = x_t - P(x_t|x_{t-1}, x_{t-2}, ...) \end{aligned}$$

In the above formulation, $x_t$ does not Granger cause $y_t$ if $E(\lambda_t \eta_{t+j}) = 0$ for all $j > 0$. This implies that the univariate innovations of $x_t$ are uncorrelated with the univariate innovations in $y_t$. Our original idea that $x_t$ Granger causes $y_t$ if its movements preced those of $y_t$ was true iff it applies to innovations, not the level of the series. If $x_t$ does not Granger cause $y_t$ then

$$x_t = c(L)u_t + d(L)v_t$$

However, the univariate representation of $x_t$ is $x_t = f(L)\lambda_t$. This implies:

$$x_t = c(L)u_t + d(L)v_t = f(L)\lambda_t$$

Thus, $\lambda_t$ is a linear combination of current and past $u_t$ and $v_t$. Since $u_t$ is the bivariate error:

$$E(u_t|y_{t-1}, y_{t-2}, ..., x_{t-1}, x_{t-2}, ...) = E(u_t|u_{t-1}, u_{t-2}, ..., v_{t-1}, v_{t-2}, ..) = 0$$

This implies that $u_t$ is uncorrelated with lagged $u_t$ and $v_t$ and therefore lagged $\lambda_t$. If $E(\lambda_t \eta_{t+j}) = 0$, then past $\eta$ does not help to forecast $\lambda$, and thus past $\eta$ does not help forecasting $y$ given past $y$. Since $x_t = f(L)\lambda_t$, this implies that past observations on $x$ do not help in forecasting y given past $y$.

**Facts**

$x$ does not Granger cause $y$ if:

- Past $x$ do not help in forecasting $y$ given past $y$. In a regression of $y$ on past $y$ and past $x$, the coefficeints of the past x are all zero.
- The autoregressive representation is lower triangular.
- Univariate innovations in $x$ are not correlated with subsequent innovations in $y$.