Forecasting with Regression Models

Ka-fu WONG

University of Hong Kong

March 29, 2019

Contents

4. Example: Predicting the Price of A share 87

1. General discussion of regression-based forecast

Consider a schematic simple linear regression model:

$$y_t = \beta_0 + \beta_1 x_t + \epsilon_t \qquad \epsilon_t \sim WN(0, \sigma^2) \tag{1.1}$$

where y_t is usually called the *endogenous* variable and x_t the *exogenous* variable or explanatory variable.

For instance, y_t could be Hong Kong's GDP growth rate and x_t China's GDP growth rate.

Assuming that the parameters are known, how can we use the model (1.1) to forecast y_{T+h}?

Note that the h-step-ahead forecast of y_{T+h} can be computed only if we know *h-step-ahead value of x_{T+h} or an estimate* of the h-step-ahead value of x_{T+h} given time T information, usually denoted as $\hat{x}_{T+h,T}$.

Let x^*_{T+h} denote some assumed h-step-ahead value of x_{T+h} (either known at time T with certainty or forecasted based on some other models and information available at time T), the h-step ahead forecast of y_{T+h} is:

$$y_{T+h}|x^*_{T+h} = \beta_0 + \beta_1 x^*_{T+h}$$

because $\epsilon_t \sim WN(0, \sigma^2)$ and hence $E[\epsilon_t \mid x^*_{T+h}] = 0$.

The model with *only time trend and seasonal components* is a perfect example in which the h-step-ahead value of x_{T+h} is *known at time T*. For instance, in a model with only time trend, $x_t = t$ and $x_{T+h} = T + h$.

x_{T+h} is a *policy variable* (say, interest rate target for a central bank), then it is also known at time T. Sometimes, the policy-maker may want to see the implied forecast when different targets of the policy variable are used, we could call this exercise as a *scenario analysis*.

Generally, however, at the time of making a forecast (say, period T), x_{T+h} is usually not known. Standing at time T, we have only the observations, $(x_1, y_1), (x_2, y_2), ..., (x_T, y_T)$. *If we do not want to model and forecast x_{T+h}*, the h-step-ahead forecast of y_{T+h} can still be produced with a modification of the model:

(1) 1-step-ahead ($h = 1$):

(a) Consider the regression model:

$$y_t = \beta_0 + \beta_1 x_{t-1} + \epsilon_t \quad t = 2, 3, 4, 5, ..., T$$

(b) Produce

- the forecast of

$$\hat{y}_{T+1} = \beta_0 + \beta_1 x_T$$

if β_0 and β_1 are known, or

- the forecast of

$$\hat{y}_{T+1} = \hat{\beta}_0 + \hat{\beta}_1 x_T$$

if β_0 and β_1 are unknown.

(2) 2-step-ahead ($h = 2$):

(a) Consider the regression model:

$$y_t = \beta_0 + \beta_1 x_{t-2} + \epsilon_t, \quad t = 3, 4, 5, ..., T$$

(b) Produce

- if β_0 and β_1 are known

$$\hat{y}_{T+1} = \beta_0 + \beta_1 x_{T-1}$$

$$\hat{y}_{T+2} = \beta_0 + \beta_1 x_T$$

- if β_0 and β_1 are unknown

$$\hat{y}_{T+1} = \hat{\beta}_0 + \hat{\beta}_1 x_{T-1}$$

$$\hat{y}_{T+2} = \hat{\beta}_0 + \hat{\beta}_1 x_T$$

(3) 3-step-ahead ($h = 3$):

(a) Consider the regression model:

$$y_t = \beta_0 + \beta_1 x_{t-3} + \epsilon_t, \quad t = 4, 5, ..., T$$

(b) Produce

- if β_0 and β_1 are known

$$\hat{y}_{T+1} = \beta_0 + \beta_1 x_{T-2}$$

$$\hat{y}_{T+2} = \beta_0 + \beta_1 x_{T-1}$$

$$\hat{y}_{T+3} = \beta_0 + \beta_1 x_T$$

- if β_0 and β_1 are unknown

$$\hat{y}_{T+1} = \hat{\beta}_0 + \hat{\beta}_1 x_{T-2}$$

$$\hat{y}_{T+2} = \hat{\beta}_0 + \hat{\beta}_1 x_{T-1}$$

$$\hat{y}_{T+3} = \hat{\beta}_0 + \hat{\beta}_1 x_T$$

(4) h-step-ahead:

(a) Consider the regression model:

$$y_t = \beta_0 + \beta_1 x_{t-h} + \epsilon_t, \quad t = h+1, ..., T$$

(b) Produce

- if β_0 and β_1 are known

$$\hat{y}_{T+1} = \beta_0 + \beta_1 x_{T-h+1}$$

$$\hat{y}_{T+2} = \beta_0 + \beta_1 x_{T-h+2}$$

$$\cdots\cdots$$

$$\hat{y}_{T+h} = \beta_0 + \beta_1 x_T$$

- if β_0 and β_1 are unknown

$$\hat{y}_{T+1} = \hat{\beta}_0 + \hat{\beta}_1 x_{T-h+1}$$

$$\hat{y}_{T+2} = \hat{\beta}_0 + \hat{\beta}_1 x_{T-h+2}$$

$$\ldots\ldots$$

$$\hat{y}_{T+h} = \hat{\beta}_0 + \hat{\beta}_1 x_T$$

As the discussion shows, when x_{T+h} is not known, long-range forecast (i.e., large h) is *feasible only with a loss in the number of observations* used in the estimation of the parameters.

A *smaller* number of observations used in the parameter estimation implies an *increase* in the parameter uncertainty, and hence an *increase* in the uncertainty in the forecast.

Thus, when we need to provide *long-range forecast*, we may want to model x_t explicitly in order to avoid the increase of forecast uncertainty due to a loss in the number of observations.

In the context of forecasting Hong Kong's GDP growth rate (y_t) four quarters ahead using China's GDP growth rate (x_t) and quarterly data, we may *postulate* Hong Kong's GDP growth rate (y_t) depends on China's GDP growth rate (x_t), i.e.,

$$y_t = \beta_0 + \beta_1 x_t + \epsilon_t$$

If we are *not willing* to build a forecasting model of x_t, we will have to restrict our model to

$$y_t = \beta_0 + \beta_1 x_{t-4} + \epsilon_t$$

Alternatively, if we are willing to build a self-contained forecasting model of x_t, say x_t to consist of a trend, seasonal components and some ARMA components, we can entertain the postulation that Hong Kong's GDP growth rate (y_t) depends on China's contemporaneous GDP growth rate (x_t), i.e.,

$$y_t = \beta_0 + \beta_1 x_t + \epsilon_t$$

Note that x_t is not the object of our forecast. We build a model of x_t only to help us forecast x_{T+h} (i.e., $x_{T+h,T}$) and hence y_{T+h} (i.e., $y_{T+h,T}$).

$$x_1, x_2, ..., x_T \longrightarrow \hat{x}_{T+h,T} \longrightarrow \hat{y}_{T+h,T}$$

2. Regression models with distributed lags

Of course, there is no need to restrict the forecasting model to those with one lag of x_t on the right-hand side:

$$y_t = \beta_0 + \beta_1 x_{t-1} + \epsilon_t$$

Sometimes, we believe that the *additional lags* of x_t might help forecast y_t. That is, we may want to consider regression models like

$$y_t = \beta_0 + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \epsilon_t$$

$$\epsilon_t \sim WN(0, \sigma^2) \quad t = 3, 4, 5, ..., T$$

or

$$y_t = \beta_0 + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \beta_3 x_{t-3} + \epsilon_t$$

$$\epsilon_t \sim WN(0, \sigma^2) \quad t = 4, 5, 6, ..., T$$

or more generally

$$y_t = \beta_0 + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \beta_3 x_{t-3} + ... + \beta_m x_{t-m} + \epsilon_t$$

$$\epsilon_t \sim WN(0, \sigma^2) \quad t = m+1, m+2, ... T$$

or written compactly as

$$y_t = \beta_0 + \sum_{i=1}^{m} \beta_i x_{t-i} + \epsilon_t$$

$$\epsilon_t \sim WN(0, \sigma^2) \quad t = m+1, m+2, ..., T$$

The problem of this kind of models is that as we include *more* lags of x_t (i.e., larger m), we have to estimate a *larger* number of parameters (a total of $m+1$ parameters, $\beta_0, \beta_1, \beta_2, ..., \beta_m$), with a *smaller* number of observations ($T-m$). Consequently, the parameter uncertainty, and hence forecast uncertainty, *increases* with m.

$$m \uparrow \quad \longrightarrow \quad T - m \downarrow \quad \longrightarrow \quad \sigma_h^2 \uparrow$$

2.1. **Polynomial distributed lags.** Fortunately, the number of parameters to be estimated can be greatly reduced *if we are willing to make certain assumption* about the relationship among the parameters on the lags of x_t, i.e., $(\beta_1, \beta_2, ..., \beta_m)$. As an extreme example, one might be willing to assume that

$$\beta_1 = \beta_2 = ... = \beta_m.$$

In this case, the number of parameters to be estimated is reduced *from* $m+1$ *to* 2 (i.e., β_0 and β_1), i.e.,

$$y_t = \beta_0 + \beta_1 \sum_{i=1}^{m} x_{t-i} + \epsilon_t$$

$$\epsilon_t \sim WN(0, \sigma^2)$$

Or, one might be willing to assume

$$\beta_i = b_0 + b_1 i, \qquad i = 1, ..., m$$

$$y_t = \beta_0 + b_0 \sum_{i=1}^{m} x_{t-i} + b_1 \sum_{i=1}^{m} i x_{t-i} + \epsilon_t$$

$$\epsilon_t \sim WN(0, \sigma^2)$$

In this case, the number of parameters to be estimated is reduced *from* $m+1$ *to* 3 (i.e., β_0, b_0 and b_1).

Or, one might be willing to assume

$$\beta_i = b_0 + b_1 i + b_2 i^2, \qquad i = 1, ..., m$$

$$y_t = \beta_0 + b_0 \sum_{i=1}^{m} x_{t-i} + b_1 \sum_{i=1}^{m} i x_{t-i} + b_2 \sum_{i=1}^{m} i^2 x_{t-i} + \epsilon_t$$

$$\epsilon_t \sim WN(0, \sigma^2)$$

In this case, the number of parameters to be estimated is reduced *from $m + 1$ to 4* (i.e., β_0, b_0, b_1 and b_2).

This approach may be generalized to assume β_i are related to i in the form of a *polynomial of n-th order in i*.

$$\beta_i = b_0 + b_1 i + b_2 i^2 + b_3 i^3 + \ldots + b_n i^n, \qquad i = 1, \ldots, m$$

$$y_t = \beta_0 + b_0 \sum_{i=1}^{m} x_{t-i} + b_1 \sum_{i=1}^{m} i x_{t-i} + b_2 \sum_{i=1}^{m} i^2 x_{t-i} + \ldots + b_n \sum_{i=1}^{m} i^n x_{t-i} + \epsilon_t$$

$$\epsilon_t \sim WN(0, \sigma^2)$$

In this case, the number of parameters to be estimated is reduced *from $m+1$ to $n+2$* (i.e., β_0, b_0, b_1, ..., and b_n). Of course, such model is meant to be an approximation and *n is usually made small relative to m*.

Because β_i are assumed related to i in the form of a polynomial of n-th order in i, this kind of model is called models with *polynomial distributed lags*.

2.2. **Rational distributed lags.** It is natural to extend the distributed lags model to other forms. To reduce typing, let's assume y_t to have mean zero and write the generic distributed lags model as

$$y_t = B(L)x_t + \epsilon_t \qquad \epsilon_t \sim WN(0, \sigma^2)$$

where $B(L)$ is a *m-degree polynomial of lag operators*, i.e.,

$$B(L) = \beta_0 + \beta_1 L + \beta_2 L^2 + ... + \beta_m L^m$$

A further extension is to allow a complicated polynomial of lag operators, such as,

$$y_t = \frac{B(L)}{A(L)} x_t + \epsilon_t \qquad \epsilon_t \sim WN(0, \sigma^2).$$

where $A(L)$ is another polynomial of lag operators.

It might look scary at first to have a *ratio of two polynomials of lag operators*. It really means

$$A(L) y_t = B(L) x_t + A(L) \epsilon_t \qquad \epsilon_t \sim WN(0, \sigma^2).$$

Less scary. But still scary!!

Let's consider several illustrative examples:

(1) If $A(L) = 1$, we have

$$y_t = B(L)x_t + \epsilon_t \qquad \epsilon_t \sim WN(0, \sigma^2)$$

(2) If $A(L) = 1$, $B(L) = \beta_0 + \beta_1 L$, we have

$$y_t = \beta_0 x_t + \beta_1 x_{t-1} + \epsilon_t \qquad \epsilon_t \sim WN(0, \sigma^2)$$

(3) If $A(L) = \alpha_0 + \alpha_1 L$, $B(L) = \beta_0 + \beta_1 L$, we have

$$
\begin{aligned}
A(L)y_t &= B(L)x_t + A(L)\epsilon_t \\
(\alpha_0 + \alpha_1 L)y_t &= (\beta_0 + \beta_1 L)x_t + (\alpha_0 + \alpha_1 L)\epsilon_t \\
\alpha_0 y_t + \alpha_1 y_{t-1} &= \beta_0 x_t + \beta_1 x_{t-1} + \alpha_0 \epsilon_t + \alpha_1 \epsilon_{t-1} \\
\alpha_0 y_t &= -\alpha_1 y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + \alpha_0 \epsilon_t + \alpha_1 \epsilon_{t-1} \\
y_t &= -\frac{\alpha_1}{\alpha_0} y_{t-1} + \frac{\beta_0}{\alpha_0} x_t + \frac{\beta_1}{\alpha_0} x_{t-1} + \epsilon_t + \frac{\alpha_1}{\alpha_0} \epsilon_{t-1}
\end{aligned}
$$

which is really an *$ARMA(1,1)$ model with lags of additional exogenous variables x_t* and *some restrictions on the parameters* (coefficients of y_{t-1} and ϵ_{t-1} are restricted to be same in magnitude and opposite in sign).

Because in this kind of models,

$$
y_t = \frac{B(L)}{A(L)} x_t + \epsilon_t \qquad \epsilon_t \sim WN(0, \sigma^2).
$$

the coefficient on x_t is a *ratio of two polynomials of lag operators*, they are called models with *rational distributed lags*.

As a further extension, we may have

$$(2.1) \qquad A(L)y_t = B(L)x_t + C(L)\epsilon_t \qquad \epsilon_t \sim WN(0, \sigma^2)$$

where $A(L)$, $B(L)$ and $C(L)$ are three polynomials of lag operators.

This specification is so general that it includes many other models as special cases.

(1) $A(L) = 1$, $C(L) = 1$: model of standard distributed lags

(2) $A(L) = C(L)$: model of rational distributed lags

(3) $B(L) = 0$, $C(L) = 1$: univariate AR model

(4) $A(L) = 1$, $B(L) = 0$: univariate MA model

(5) $B(L) = 0$: univariate ARMA model

Why would anyone be interested in writing the general but scary form of equation 2.1? This very general form is often *used by econometricians to prove the statistical property of coefficient estimators*. Once the property of this general case is shown, we would not need to show separately the other special cases, such as the univariate AR models.

For practitioners, we have to *check whether their proof include the models we have in mind*, so that we will feel safe to go ahead with using our mdoel to produce forecast.

3. Vector Autoregressions

Suppose our objective is to forecast y_{T+h}. Sometimes, we know that, *in addition to* the past values of y_t, *past values of x_t is also useful* in forecasting y_t, say,

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 x_{t-1} + \epsilon_t$$

Let say we are interested in forecasting the variable at a rather long horizon (i.e., h large). In this case, we would have to model x_t. Suppose, in building a model of x_t, we find it reasonable to model x_t to depend on *past values of y_t, in addition to* the past values of x_t, say,

$$x_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 x_{t-1} + v_t.$$

There are many examples of which y_t and x_t are reasonably modelled in this manner.

For instance, y_t is the exchange rate of Japanese yen per US dollar, and x_t is the trade balance between US and Japan. Import and export, and hence trade balance, drive the supply and demand for the two currencies, and hence the equilibrium exchange rate. Since nominal exchange rate determines the terms of trade, nominal exchange rate determines import and export, and hence the trade balance.

The kind of model might look complicated:

- To forecast y_{T+h}, we need to forecast x_{T+h-1};
- To forecast x_{T+h-1}, we need to forecast y_{T+h-2} and so on.

Complicated as it may seem, when we write the two models jointly in matrix form, the complexity is greatly reduced.

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} \beta_0 \\ \alpha_0 \end{pmatrix} + \begin{pmatrix} \beta_1 & \beta_2 \\ \alpha_1 & \alpha_2 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_t \\ v_t \end{pmatrix}$$

$$z_t = \gamma_0 + \gamma_1 z_{t-1} + \eta_t \tag{3.1}$$

where

$$z_t \equiv \begin{pmatrix} y_t \\ x_t \end{pmatrix}, \quad \gamma_0 \equiv \begin{pmatrix} \beta_0 \\ \alpha_0 \end{pmatrix}, \quad \gamma_1 \equiv \begin{pmatrix} \beta_1 & \beta_2 \\ \alpha_1 & \alpha_2 \end{pmatrix}, \quad \eta_t \equiv \begin{pmatrix} \epsilon_t \\ v_t \end{pmatrix}.$$

Thus, we have a autoregressive model of z_t. It differs from the univariate autoregression we have studied earlier only in the number of elements in z_t. Since z_t is a vector, we call this kind of models vector autoregressions, or VAR in short.

Some basic matrix operations.

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} \epsilon_t \\ v_t \end{pmatrix}$$

means

$$y_t = \epsilon_t$$

$$x_t = v_t$$

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} \beta_0 \\ \alpha_0 \end{pmatrix} + \begin{pmatrix} \epsilon_t \\ v_t \end{pmatrix}$$

means

$$y_t = \beta_0 + \epsilon_t$$
$$x_t = \alpha_0 + v_t$$

$$\begin{pmatrix} \beta_1 & \beta_2 \\ \alpha_1 & \alpha_2 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} = \begin{pmatrix} \beta_1 y_{t-1} + \beta_2 x_{t-1} \\ \alpha_1 y_{t-1} + \alpha_2 x_{t-1} \end{pmatrix}$$

Identity matrix:

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} = \begin{pmatrix} 1 y_{t-1} + 0 x_{t-1} \\ 0 y_{t-1} + 1 x_{t-1} \end{pmatrix} = \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix}$$

$$\begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} = \begin{pmatrix} 3 y_{t-1} + 0 x_{t-1} \\ 0 y_{t-1} + 2 x_{t-1} \end{pmatrix} = \begin{pmatrix} 3 y_{t-1} \\ 2 x_{t-1} \end{pmatrix}$$

$$\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}^2 = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}$$

$$= \begin{pmatrix} a \times a + 0 \times 0 & a \times 0 + 0 \times b \\ 0 \times a + b \times 0 & 0 \times 0 + b \times b \end{pmatrix} = \begin{pmatrix} a^2 & 0 \\ 0 & b^2 \end{pmatrix}$$

$$\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}^3 = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} = \begin{pmatrix} a^3 & 0 \\ 0 & b^3 \end{pmatrix}$$

$$\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}^n = \begin{pmatrix} a^n & 0 \\ 0 & b^n \end{pmatrix}$$

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} \beta_0 \\ \alpha_0 \end{pmatrix} + \begin{pmatrix} \beta_1 & \beta_2 \\ \alpha_1 & \alpha_2 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_t \\ v_t \end{pmatrix}$$

means

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 x_{t-1} + \epsilon_t$$
$$x_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 x_{t-1} + v_t$$

Inverse:

$$\begin{pmatrix} \beta_1 & \beta_2 \\ \alpha_1 & \alpha_2 \end{pmatrix} \begin{pmatrix} \beta_1 & \beta_2 \\ \alpha_1 & \alpha_2 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$\begin{pmatrix} \beta_1 & \beta_2 \\ \alpha_1 & \alpha_2 \end{pmatrix}^{-1} \begin{pmatrix} \beta_1 & \beta_2 \\ \alpha_1 & \alpha_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

A schematic matrix is often labelled as

$$\Phi = \begin{pmatrix} \Phi(1,1) & \Phi(1,2) & \Phi(1,3) & \Phi(1,4) \\ \Phi(2,1) & \Phi(2,2) & \Phi(2,3) & \Phi(2,4) \\ \Phi(3,1) & \Phi(3,2) & \Phi(3,3) & \Phi(3,4) \end{pmatrix}$$

so that $\Phi(2,3)$ is the second-row and third-column element of Φ.

Additional reading (introduction of matrices):

- http://web.mit.edu/2.14/www/Handouts/Matrices.pdf
- https://www.khanacademy.org/math/precalculus/precalc-matrices

Recall that a univariate autoregression is a single-equation, single-variable linear model in which the current value of a variable is explained by its own lagged values.

A VAR is an n-equation, n-variable linear model in which each variable is in turn explained by its own lagged values, plus the past values of the remaining $n-1$ variables. VAR can be used more broadly to:

(1) describe and summarize macroeconomic data,

(2) make macroeconomic forecasts,

(3) quantify what we do or do not know about the true structure of the macroeconomy, and

(4) advise (and sometimes become) macroeconomic policymakers.

VARs come in three varieties:
(a) reduced form,
(b) recursive form, and
(c) structural form.

In our case, the *focus is* on making macroeconomic *forecasts*. For this purpose, we will consider *only reduced form VAR*.

Suppose we are going to model in a VAR the *exchange rate* of Japanese yen per US dollar, and the *trade balance* between US and Japan.

z_{1t} could be the exchange rate of Japanese yen per US dollar, and z_{2t} the trade balance between US and Japan.

Suppose we believe that trade balance will have contemporaneous impact on exchange rate but exchange rate has impact on trade balance only with a lag. We would want to write

$$z_{1t} = \gamma_0 + \gamma_1 z_{1t-1} + \gamma_2 z_{2t} + \gamma_3 z_{2t-1} + \epsilon_{1t}$$
$$z_{2t} = \beta_0 + \beta_1 z_{1t-1} + \beta_2 z_{2t-1} + \epsilon_{2t}$$

and in matrix form

$$\begin{pmatrix} z_{1t} \\ z_{2t} \end{pmatrix} = \begin{pmatrix} \gamma_0 \\ \beta_0 \end{pmatrix} + \begin{pmatrix} 0 & \gamma_2 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} z_{1t} \\ z_{2t} \end{pmatrix} + \begin{pmatrix} \gamma_1 & \gamma_3 \\ \beta_1 & \beta_2 \end{pmatrix} \begin{pmatrix} z_{1t-1} \\ z_{2t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix}$$

We can then further rewrite into

$$\begin{pmatrix} 1 & -\gamma_2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} z_{1t} \\ z_{2t} \end{pmatrix} = \begin{pmatrix} \gamma_0 \\ \beta_0 \end{pmatrix} + \begin{pmatrix} \gamma_1 & \gamma_3 \\ \beta_1 & \beta_2 \end{pmatrix} \begin{pmatrix} z_{1t-1} \\ z_{2t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix}$$

That is,

$$z_{1t} - \gamma_2 z_{2t} = \gamma_0 + \gamma_1 z_{1t-1} + \gamma_3 z_{2t-1} + \epsilon_{1t}$$
$$z_{2t} = \beta_0 + \beta_1 z_{1t-1} + \beta_2 z_{2t-1} + \epsilon_{2t}$$

In condense matrix notations,

$$B_0 z_t = C + B_1 z_{t-1} + \epsilon_t$$

Premultiply through by B_0^{-1},

$$z_t = B_0^{-1} C + B_0^{-1} B_1 z_{t-1} + B_0^{-1} \epsilon_t$$
$$z_t = \Phi_0 + \Phi_1 z_{t-1} + \eta_t$$

$$\begin{pmatrix} z_{1t} \\ z_{2t} \end{pmatrix} = \begin{pmatrix} \Phi_0(1) \\ \Phi_0(2) \end{pmatrix} + \begin{pmatrix} \Phi_1(1,1) & \Phi_1(1,2) \\ \Phi_1(2,1) & \Phi_1(2,2) \end{pmatrix} \begin{pmatrix} z_{1t-1} \\ z_{2t-1} \end{pmatrix} + \begin{pmatrix} \eta_{1t} \\ \eta_{2t} \end{pmatrix}$$

Again, in our case, the *focus is* on making macroeconomic *forecasts*. For this purpose, we will consider *only reduced form VAR*.

Although we will briefly introduce several elements of VAR (*impulse response function* and *variance decomposition*) that are usually important for other empirical studies, we have to emphasize that these elements are usually *not essential* for the the purpose of forecasting.

For a concise discussion, see the brief tutorial by Stock and Watson (2001)[1] and the references therein.

[1]Stock, James H. and Mark W. Watson (2001): "Vector Autoregressions," *Journal of Economic Perspectives*, Vol. 15, No. 4, pp.101-115.

A n-variable reduced form VAR of order p writes each variable as a linear function of its own past values, the past values of all other variables up to p lags, and a serially uncorrelated error term.

Suppose we are interested in forecasting Japanese yen's exchange rate against US dollar and Japan's balance with the US above. Each of the n equations can be estimated by ordinary least squares regression.

The number of lagged values to include in each equation can be determined by a number of different methods such as *AIC and SIC*. Just like the univariate AR, the error terms in these regressions are the "surprise" movements in the variables after taking its past values into account.

Generally, we have the *n-variable VAR(p)* model

$$z_t = \Phi_0 + \Phi_1 z_{t-1} + \Phi_2 z_{t-2} + \ldots + \Phi_p z_{t-p} + \eta_t$$

where z_t is a $n \times 1$ vector

$$z_t \equiv \begin{pmatrix} z_{1t} \\ z_{2t} \\ \ldots \\ z_{nt} \end{pmatrix}$$

η_t is a $n \times 1$ vector

$$\eta_t \equiv \begin{pmatrix} \eta_{1t} \\ \eta_{2t} \\ \ldots \\ \eta_{nt} \end{pmatrix}$$

and Φ_k is a $n \times n$ matrix, where the (row i, column j) element of the matrix is usually denoted as $\Phi_k(i, j)$.

It would be useful to map this general model to a specific example. Suppose we are going to model in a VAR the *exchange rate* of Japanese yen per US dollar, and the *trade balance* between US and Japan.

In this case, $m = 2$. z_{1t} could be the exchange rate of Japanese yen per US dollar, and z_{2t} the trade balance between US and Japan. If only variables *lagged one period* are useful in forecasting z_t, we would have

$$\begin{pmatrix} z_{1t} \\ z_{2t} \end{pmatrix} = \begin{pmatrix} \Phi_0(1) \\ \Phi_0(2) \end{pmatrix} + \begin{pmatrix} \Phi_1(1,1) & \Phi_1(1,2) \\ \Phi_1(2,1) & \Phi_1(2,2) \end{pmatrix} \begin{pmatrix} z_{1t-1} \\ z_{2t-1} \end{pmatrix} + \begin{pmatrix} \eta_{1t} \\ \eta_{2t} \end{pmatrix}$$

That is, a 2-variable VAR(1) model.

If variables up to *lagged two periods* are useful in forecasting z_t, we would have

$$\begin{aligned}
\begin{pmatrix} z_{1t} \\ z_{2t} \end{pmatrix} &= \begin{pmatrix} \Phi_0(1) \\ \Phi_0(2) \end{pmatrix} \\
&\quad + \begin{pmatrix} \Phi_1(1,1) & \Phi_1(1,2) \\ \Phi_1(2,1) & \Phi_1(2,2) \end{pmatrix} \begin{pmatrix} z_{1t-1} \\ z_{2t-1} \end{pmatrix} \\
&\quad + \begin{pmatrix} \Phi_2(1,1) & \Phi_2(1,2) \\ \Phi_2(2,1) & \Phi_2(2,2) \end{pmatrix} \begin{pmatrix} z_{1t-2} \\ z_{2t-2} \end{pmatrix} + \begin{pmatrix} \eta_{1t} \\ \eta_{2t} \end{pmatrix}
\end{aligned}$$

That is, a 2-variable VAR(2) model.

If variables up to *lagged three periods* are useful in forecasting z_t, we would have

$$
\begin{aligned}
\begin{pmatrix} z_{1t} \\ z_{2t} \end{pmatrix} &= \begin{pmatrix} \Phi_0(1) \\ \Phi_0(2) \end{pmatrix} \\
&+ \begin{pmatrix} \Phi_1(1,1) & \Phi_1(1,2) \\ \Phi_1(2,1) & \Phi_1(2,2) \end{pmatrix} \begin{pmatrix} z_{1t-1} \\ z_{2t-1} \end{pmatrix} \\
&+ \begin{pmatrix} \Phi_2(1,1) & \Phi_2(1,2) \\ \Phi_2(2,1) & \Phi_2(2,2) \end{pmatrix} \begin{pmatrix} z_{1t-2} \\ z_{2t-2} \end{pmatrix} \\
&+ \begin{pmatrix} \Phi_3(1,1) & \Phi_3(1,2) \\ \Phi_3(2,1) & \Phi_3(2,2) \end{pmatrix} \begin{pmatrix} z_{1t-3} \\ z_{2t-3} \end{pmatrix} + \begin{pmatrix} \eta_{1t} \\ \eta_{2t} \end{pmatrix}
\end{aligned}
$$

That is, a 2-variable VAR(3) model.

If variables up to *lagged four periods* are useful in forecasting z_t, we would have

$$
\begin{aligned}
\begin{pmatrix} z_{1t} \\ z_{2t} \end{pmatrix} &= \begin{pmatrix} \Phi_0(1) \\ \Phi_0(2) \end{pmatrix} \\
&+ \begin{pmatrix} \Phi_1(1,1) & \Phi_1(1,2) \\ \Phi_1(2,1) & \Phi_1(2,2) \end{pmatrix} \begin{pmatrix} z_{1t-1} \\ z_{2t-1} \end{pmatrix} \\
&+ \begin{pmatrix} \Phi_2(1,1) & \Phi_2(1,2) \\ \Phi_2(2,1) & \Phi_2(2,2) \end{pmatrix} \begin{pmatrix} z_{1t-2} \\ z_{2t-2} \end{pmatrix} \\
&+ \begin{pmatrix} \Phi_3(1,1) & \Phi_3(1,2) \\ \Phi_3(2,1) & \Phi_3(2,2) \end{pmatrix} \begin{pmatrix} z_{1t-3} \\ z_{2t-3} \end{pmatrix} \\
&+ \begin{pmatrix} \Phi_4(1,1) & \Phi_4(1,2) \\ \Phi_4(2,1) & \Phi_4(2,2) \end{pmatrix} \begin{pmatrix} z_{1t-4} \\ z_{2t-4} \end{pmatrix} + \begin{pmatrix} \eta_{1t} \\ \eta_{2t} \end{pmatrix}
\end{aligned}
$$

That is, a 2-variable VAR(4) model.

3.1. **Vector Wold Representation.** Like univariate AR models, under some conditions on the coefficient matrix, we can use repeated substitution to obtain a Wold Representation. Take a simple VAR(1) for example.

$$z_t = \Phi z_{t-1} + \eta_t$$

Using repeated substitution, we get

$$z_t = \eta_t + \Phi\eta_{t-1} + \Phi^2\eta_{t-2} + \Phi^3\eta_{t-3} + \ldots$$

A note on eigendecomposition

Suppose the eigenvectors of A form a basis, or equivalently A has n linearly independent eigenvectors $v_1, v_2, \ldots, v_n$ with associated eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. The eigenvalues need not be distinct. Define a square matrix Q whose columns are the n linearly independent eigenvectors of A,

$$Q = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix}.$$

Define a diagonal matrix Λ where each diagonal element Λ_{ii} is the eigenvalue associated with the i-th column of Q.

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & .. & \lambda_n \end{pmatrix}$$

Then

$$AQ = Q\Lambda$$

Because the columns of Q are linearly independent, Q is invertible. Right multiplying both sides of the equation by Q^{-1},

$$A = Q\Lambda Q^{-1}$$

$$A^2 = A \times A = Q\Lambda Q^{-1}Q\Lambda Q^{-1} = Q\Lambda^2 Q^{-1}$$

$$A^3 = A^2 \times A = Q\Lambda^2 Q^{-1}Q\Lambda Q^{-1} = Q\Lambda^3 Q^{-1}$$

$$A^k = Q\Lambda^k Q^{-1}$$

Recall

$$\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}^k = \begin{pmatrix} a^k & 0 \\ 0 & b^k \end{pmatrix}$$

$$\Lambda^k = \begin{pmatrix} \lambda_1^k & 0 & 0 & \dots & 0 \\ 0 & \lambda_2^k & 0 & \dots & 0 \\ 0 & 0 & \lambda_3^k & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & .. & \lambda_n^k \end{pmatrix}$$

If the largest absolute value of eigenvalues is smaller than 1, the k-th power of an $n \times n$ matrix A will approach a corresponding zero matrix

$$\lim_{k \to \infty} A^k = 0$$

That is, the spectral radius

$$\rho(A) = \max\{|\lambda_1|, ..., |\lambda_n|\} < 1$$

where λ_i's are the eigenvalues of the matrix A.

Such Wold representation can also be obtained using lag operators as in the univariate AR case.

$$
\begin{aligned}
(I - \Phi L)z_t &= \eta_t \\
z_t &= (I - \Phi L)^{-1}\eta_t \\
z_t &= (I + \Phi L + \Phi^2 L^2 + \dots)\eta_t
\end{aligned}
$$

where I is an identity matrix

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

3.2. **Impulse response.** As in the case of univariate AR, we can also talk about the impulse response function.

In the univariate case, with the following Wold representation:

$$y_t = \epsilon_t + b_1\epsilon_{t-1} + b_2\epsilon_{t-2} + b_3\epsilon_{t-3} + ... + b_h\epsilon_{t-h} + ... \tag{3.2}$$

or

$$y_{t+h} = \epsilon_{t+h} + b_1\epsilon_{t+h-1} + b_2\epsilon_{t+h-2} + ... + b_h\epsilon_t + ... \tag{3.3}$$

Referring to equation 3.3, the impact of one unit shock at time t on y at time $t+h$ is thus

$$\frac{\partial y_{t+h}}{\partial \epsilon_t} = b_h$$

Of course, we can easily see that b_h is simply the coefficient of ϵ_{t-h} in equation 3.2.

$$\frac{\partial y_t}{\partial \epsilon_{t-h}} = b_h$$

This impact of one unit shock at time t on y at time $t+h$ is also called the impulse response. The shock is the impulse. The impact is the response.

Impulse (ϵ_t)	$\longrightarrow$	Response (y_{t+h})	
1		1	$(h=0)$
		b_1	$(h=1)$
		b_2	$(h=2)$
		$\cdots$	$\cdots$
		$\cdots$	$\cdots$
		b_h	$(h=h)$
		$\cdots$	$\cdots$
		$\cdots$	$\cdots$

$$
\begin{array}{ccll}
\text{Impulse } (\epsilon_{t-h}) & \longrightarrow & \text{Response } (y_t) & \\
1 & & 1 & (h=0) \\
 & & b_1 & (h=1) \\
 & & b_2 & (h=2) \\
 & & \cdots & \cdots \\
 & & \cdots & \cdots \\
 & & b_h & (h=h) \\
 & & \cdots & \cdots \\
 & & \cdots & \cdots
\end{array}
$$

Often we would like to talk about the response of y_{t+h} to a one standard deviation shock of ϵ_t. It is simply the coefficient of ϵ_{t-h} in the Wold representation multiplied by the standard deviation of ϵ_t, i.e., $b_h\sigma$.

$$
\begin{array}{ccccc}
\text{Impulse } (\epsilon_t) & \longrightarrow & \text{Response } (y_{t+h}) & & \\
\sigma & & \sigma & & (h=0) \\
 & & \sigma b_1 & & (h=1) \\
 & & \sigma b_2 & & (h=2) \\
 & & \dots & & \dots \\
 & & \dots & & \dots \\
 & & \sigma b_h & & (h=h) \\
 & & \dots & & \dots \\
 & & \dots & & \dots
\end{array}
$$

$$
\begin{array}{cccc}
\text{Impulse } (\epsilon_{t-h}) & \longrightarrow & \text{Response } (y_t) & \\
\sigma & & \sigma & (h=0) \\
 & & \sigma b_1 & (h=1) \\
 & & \sigma b_2 & (h=2) \\
 & & \cdots & \cdots \\
 & & \cdots & \cdots \\
 & & \sigma b_h & (h=h) \\
 & & \cdots & \cdots \\
 & & \cdots & \cdots
\end{array}
$$

Why do we want to consider the response of y_{t+h} to a one standard deviation shock of ϵ_t? The reason is that one unit shock of ϵ_t can be

- *a small shock when the standard deviation is large*, and
- *a big shock when the standard deviation is small*.

In particular, the response of one unit shock is "unit dependent".

- the response of exchange rate to a one-dollar increase in trade balance (say, when the unit of trade balance is dollars).
- the response of exchange rate to a one-thousand-dollar increase in trade balance (say, when the unit of trade balance is thousand dollars).
- the response of exchange rate to a one-million-dollar increase in trade balance (say, when the unit of trade balance is million dollars).

By considering a one standard deviation shock, we make the "size" of the shocks comparable across models and specifications.

The *slight complication* in a n-variable VAR is that we have a vector of n shocks, each corresponding to different variables.

Response

		y_{1t+h}	y_{2t+h}	....	y_{nt+h}
	ϵ_{1t}				
shock	ϵ_{2t}				
	....				
	ϵ_{nt}				

Take the example of the exchange rate of Japanese yen per US dollar and the trade balance between US and Japan discussed earlier, the exchange rate shock is η_{1t} and the trade balance shock is η_{2t}.

	z_{1t}	z_{2t}	z_{1t+1}	z_{2t+1}	z_{1t+2}	z_{2t+2}	$\dots$	z_{1t+h}	z_{2t+h}	$\dots$
η_{1t}										
η_{2t}										

One can trace the impact of a shock or innovation on the two variables as we do in the usual univariate ARMA models.

- First, we can trace the path of z_{1t} and z_{2t} when η_{1t} and η_{2t} are all zero at all t. Call this path (1).

$$\begin{pmatrix} z^{\#}_{1t} \\ z^{\#}_{2t} \end{pmatrix} = \begin{pmatrix} \Phi_0(1) \\ \Phi_0(2) \end{pmatrix} + \begin{pmatrix} \Phi_1(1,1) & \Phi_1(1,2) \\ \Phi_1(2,1) & \Phi_1(2,2) \end{pmatrix} \begin{pmatrix} z_{1t-1} \\ z_{2t-1} \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} z^{\#}_{1t+1} \\ z^{\#}_{2t+1} \end{pmatrix} = \begin{pmatrix} \Phi_0(1) \\ \Phi_0(2) \end{pmatrix} + \begin{pmatrix} \Phi_1(1,1) & \Phi_1(1,2) \\ \Phi_1(2,1) & \Phi_1(2,2) \end{pmatrix} \begin{pmatrix} z^{\#}_{1t} \\ z^{\#}_{2t} \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} z^{\#}_{1t+2} \\ z^{\#}_{2t+2} \end{pmatrix} = \begin{pmatrix} \Phi_0(1) \\ \Phi_0(2) \end{pmatrix} + \begin{pmatrix} \Phi_1(1,1) & \Phi_1(1,2) \\ \Phi_1(2,1) & \Phi_1(2,2) \end{pmatrix} \begin{pmatrix} z^{\#}_{1t+1} \\ z^{\#}_{2t+1} \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

- Then we trace the path of z_{1t} and z_{2t} when η_{1t} and η_{2t} are all zero at all t, except at $t = t^*$ when η_{1t} *is set to 1* (i.e., one unit shock). Call this path (2).

$$\begin{pmatrix} z_{1t}^* \\ z_{2t}^* \end{pmatrix} = \begin{pmatrix} \Phi_0(1) \\ \Phi_0(2) \end{pmatrix} + \begin{pmatrix} \Phi_1(1,1) & \Phi_1(1,2) \\ \Phi_1(2,1) & \Phi_1(2,2) \end{pmatrix} \begin{pmatrix} z_{1t-1} \\ z_{2t-1} \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} z_{1t+1}^* \\ z_{2t+1}^* \end{pmatrix} = \begin{pmatrix} \Phi_0(1) \\ \Phi_0(2) \end{pmatrix} + \begin{pmatrix} \Phi_1(1,1) & \Phi_1(1,2) \\ \Phi_1(2,1) & \Phi_1(2,2) \end{pmatrix} \begin{pmatrix} z_{1t}^* \\ z_{2t}^* \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} z_{1t+2}^* \\ z_{2t+2}^* \end{pmatrix} = \begin{pmatrix} \Phi_0(1) \\ \Phi_0(2) \end{pmatrix} + \begin{pmatrix} \Phi_1(1,1) & \Phi_1(1,2) \\ \Phi_1(2,1) & \Phi_1(2,2) \end{pmatrix} \begin{pmatrix} z_{1t+1}^* \\ z_{2t+1}^* \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

- The difference in z_{1t} in path (1) (i.e., $z^{\#}_{1t+h}$) and path (2) (i.e., z^{*}_{1t+h}) is the responses of z_{1t} to a one unit shock of η_{1t}.
- Similarly, the difference in z_{2t} in path (1) (i.e., $z^{\#}_{2t+h}$) and path (2) (i.e., z^{*}_{2t+h}) is the responses of z_{2t} to a one unit shock of η_{1t}.

$$\frac{\partial z_{1t+h}}{\partial \eta_{1t}} \quad \text{and} \quad \frac{\partial z_{2t+h}}{\partial \eta_{1t}}$$

We can also compute the impulse response of the two variables to the innovation η_{2t} in a similar manner.

- First, we can trace the path of z_{1t} and z_{2t} when η_{1t} and η_{2t} are all zero at all t. Call this path (1).

$$\begin{pmatrix} z_{1t}^{\#} \\ z_{2t}^{\#} \end{pmatrix} = \begin{pmatrix} \Phi_0(1) \\ \Phi_0(2) \end{pmatrix} + \begin{pmatrix} \Phi_1(1,1) & \Phi_1(1,2) \\ \Phi_1(2,1) & \Phi_1(2,2) \end{pmatrix} \begin{pmatrix} z_{1t-1} \\ z_{2t-1} \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} z_{1t+1}^{\#} \\ z_{2t+1}^{\#} \end{pmatrix} = \begin{pmatrix} \Phi_0(1) \\ \Phi_0(2) \end{pmatrix} + \begin{pmatrix} \Phi_1(1,1) & \Phi_1(1,2) \\ \Phi_1(2,1) & \Phi_1(2,2) \end{pmatrix} \begin{pmatrix} z_{1t}^{\#} \\ z_{2t}^{\#} \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} z_{1t+2}^{\#} \\ z_{2t+2}^{\#} \end{pmatrix} = \begin{pmatrix} \Phi_0(1) \\ \Phi_0(2) \end{pmatrix} + \begin{pmatrix} \Phi_1(1,1) & \Phi_1(1,2) \\ \Phi_1(2,1) & \Phi_1(2,2) \end{pmatrix} \begin{pmatrix} z_{1t+1}^{\#} \\ z_{2t+1}^{\#} \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

- Then we trace the path of z_{1t} and z_{2t} when η_{1t} and η_{2t} are all zero at all t, except at $t = t^*$ when η_{2t} *is set to 1* (i.e., one unit shock). Call this path (2).

$$\begin{pmatrix} z^{\%}_{1t} \\ z^{\%}_{2t} \end{pmatrix} = \begin{pmatrix} \Phi_0(1) \\ \Phi_0(2) \end{pmatrix} + \begin{pmatrix} \Phi_1(1,1) & \Phi_1(1,2) \\ \Phi_1(2,1) & \Phi_1(2,2) \end{pmatrix} \begin{pmatrix} z_{1t-1} \\ z_{2t-1} \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

$$\begin{pmatrix} z^{\%}_{1t+1} \\ z^{\%}_{2t+1} \end{pmatrix} = \begin{pmatrix} \Phi_0(1) \\ \Phi_0(2) \end{pmatrix} + \begin{pmatrix} \Phi_1(1,1) & \Phi_1(1,2) \\ \Phi_1(2,1) & \Phi_1(2,2) \end{pmatrix} \begin{pmatrix} z^{\%}_{1t} \\ z^{\%}_{2t} \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} z^{\%}_{1t+2} \\ z^{\%}_{2t+2} \end{pmatrix} = \begin{pmatrix} \Phi_0(1) \\ \Phi_0(2) \end{pmatrix} + \begin{pmatrix} \Phi_1(1,1) & \Phi_1(1,2) \\ \Phi_1(2,1) & \Phi_1(2,2) \end{pmatrix} \begin{pmatrix} z^{\%}_{1t+1} \\ z^{\%}_{2t+1} \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

- The difference in z_{1t} in path (1) (i.e., $z^{\#}_{1t+h}$) and path (2) (i.e., $z^{\%}_{1t+h}$) is the responses of z_{1t} to a one unit shock of η_{2t}.
- Similarly, the difference in z_{2t} in path(1) (i.e., $z^{\#}_{2t+h}$) and path (2) (i.e., $z^{\%}_{2t+h}$) is the responses of z_{2t} to a one unit shock of η_{2t}.

$$\frac{\partial z_{1t+h}}{\partial \eta_{2t}} \quad \text{and} \quad \frac{\partial z_{2t+h}}{\partial \eta_{2t}}$$

Similar to the univariate case, if we have a wold respresentation of z_t, the impulse response is simply the corresponding coefficients in the representation. Suppose z_t consists of two elements (i.e., $n = 2$) and has the following general Wold representation.

$$z_t = \eta_t + B_1\eta_{t-1} + B_2\eta_{t-2} + B_3\eta_{t-3} + \ldots$$

or

$$\begin{aligned}\begin{pmatrix} z_{1t} \\ z_{2t} \end{pmatrix} = \begin{pmatrix} \eta_{1t} \\ \eta_{2t} \end{pmatrix} &+ \begin{pmatrix} B_1(1,1) & B_1(1,2) \\ B_1(2,1) & B_1(2,2) \end{pmatrix} \begin{pmatrix} \eta_{1t-1} \\ \eta_{2t-1} \end{pmatrix} \\ &+ \begin{pmatrix} B_2(1,1) & B_2(1,2) \\ B_2(2,1) & B_2(2,2) \end{pmatrix} \begin{pmatrix} \eta_{1t-2} \\ \eta_{2t-2} \end{pmatrix} \\ &+ \begin{pmatrix} B_3(1,1) & B_3(1,2) \\ B_3(2,1) & B_3(2,2) \end{pmatrix} \begin{pmatrix} \eta_{1t-3} \\ \eta_{2t-3} \end{pmatrix} \\ &+ \ldots\end{aligned}$$

The following table computes the responses of z_t to the shocks of η_t.

Response of z_{t+h} to one-unit shock of η_{1t}

Horizon (h)	z_{1t+h}	z_{2t+h}
0	1	0
1	$B_1(1,1)$	$B_1(2,1)$
2	$B_2(1,1)$	$B_2(2,1)$
3	$B_3(1,1)$	$B_3(2,1)$
$\vdots$	$\vdots$	$\vdots$

Response of z_{t+h} to one-unit shock of η_{2t}

Horizon (h)	z_{1t+h}	z_{2t+h}
0	0	1
1	$B_1(1,2)$	$B_1(2,2)$
2	$B_2(1,2)$	$B_2(2,2)$
3	$B_3(1,2)$	$B_3(2,2)$
$\vdots$	$\vdots$	$\vdots$

The response of z_t due to a one-standard-deviation shock of η_t can be obtained by simply multiplying the corresponding entries with the standard deviation of η_{1t} and η_{2t}.

3.3. **Variance decomposition.** Note that the concept of variance decomposition does not occur in the univariate case because theoretically all the squared prediction errors are *due to the innovation of the variable*.

In an VAR, we have two or more variables. Naturally, we would be interested in knowing the percentage of the expected h-period ahead squared prediction errors of a variable *attributed to* an innovation of another variable.

	z_{1t}	z_{2t}	z_{1t+1}	z_{2t+1}	z_{1t+2}	z_{2t+2}	$\ldots$	z_{1t+h}	z_{2t+h}	$\ldots$
η_{1t}										
η_{2t}										

3.4. **Granger causality.** Recall the reason for modelling the vector of variables together is that in order to forecast the variable z_1, we need to forecast z_2, and when we try to build a model to forecast z_2, we find that z_2 depends on z_1.

If the history (i.e. lagged observations) of variable z_1 does not help predict the future values of variable z_2 (given lagged values of z_2 and lagged values of other variables), there is no need to consider z_1 and z_2 together.

Let's consider two scenarios.

In the first scenario, suppose our objective is to forecast z_1 (say, the exchange rate of Japanese yen per US dollar), and we may know that z_2 (say, the trade balance between Japan and the US) does not help improve the forecast once the history of z_1 is included in the information set.

$$z_{1t} = \alpha_0^1 + \alpha_1^1 z_{1t-1} + 0 \times z_{2t-1} + \epsilon_{1t}$$

$$z_2 \quad \longrightarrow \quad z_1 \quad ?$$

Do we need to model z_2 to depend on the history of z_2 and z_1? There is obviously no need. In fact, the model of z_2 will not help.

Of course, we can *insist* in modelling z_1 and z_2 in a VAR, but it *only adds to the forecast uncertainty* through the uncertainty in the estimation of some unnecessary parameters.

In the second scenario, suppose our objective is to forecast z_1 (say, the exchange rate of Japanese yen per US dollar), and we may know that z_2 (say, the trade balance between Japan and the US) can help improve the forecast even after the history of z_1 is included in the information set.

$$z_{1t} = \alpha_0^1 + \alpha_1^1 z_{1t-1} + \beta_1^1 \times z_{2t-1} + \epsilon_{1t}$$

$$z_1 \overset{?}{\longrightarrow} z_2$$

If z_2 can be modelled as a univariate model and z_1 is known to be useless in forecasting z_2,

$$z_{2t} = \alpha_0^2 + 0 \times z_{1t-1} + \beta_1^2 \times z_{2t-1} + \epsilon_{1t}$$

we can forecast z_1 using the simple regression-based forecast discussed earlier. That is, use the univariate model of z_2 to produce forecast of z_{2t+h}; plug $\hat{z}_{2t+h}$ into the regression model of z_{1t} on the lags of z_{1t} and z_{2t}

$$\hat{z}_{1t+h+1} = \alpha_0^1 + \alpha_1^1 z_{1t+h} + \beta_1^1 \times \hat{z}_{2t+h}$$

Can we still model the two variables as a VAR when z_1 is known useless in forecasting z_2? Yes, we could but it is going to introduce the *additional forecast uncertainty* due to the estimation of the *additional coefficients* that are known to be zero.

$$\text{MSPE}\ (\hat{\beta}, \hat{\gamma} \mid \beta \neq 0, \gamma = 0) > \text{MSPE}\ (\hat{\beta} \mid \beta \neq 0, \gamma = 0)$$

Simulation exercise:

- Simulate T (say, $T = 100$) observations of a linear trend model $y_t = \beta_0 + \beta_1 t + \epsilon_t$, $\epsilon \sim N(0,1)$, $\beta_0 = 3$, $\beta_1 = 0.5$.
- Estimate the linear trend model $y_t = \beta_0 + \beta_1 t + \epsilon_t$ using the first R (say, $R \approx 0.6T$) observations and produce one-step-ahead forecast
  $\hat{y}^A_{R+1,R} = \hat{\beta}_0 + \hat{\beta}_1(R+1)$.
  Repeat the forecast exercise recursively. Compute the mean squared prediction error, denoted as $MSPE^A$.
- Estimate the quadratic trend model $y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \epsilon_t$ using the first R observations and produce one-step-ahead forecast
  $\hat{y}^B_{R+1,R} = \hat{\beta}_0 + \hat{\beta}_1(R+1) + \hat{\beta}_2(R+1)^2$.
  Repeat the forecast exercise recursively. Compute the mean squared prediction error, denoted as $MSPE^B$.
- Observe the difference between $MSPE^A$ and $MSPE^B$
- Repeat the exercise with a different T ($T = 100, 200, ..., 1000$) and still use $R \approx 0.6T$. Observe how the difference between $MSPE$s changes with T.
- Repeat the exercise with a fixed T (say, $T = 1000$) and still use $R \approx \gamma T$ for $\gamma = 0.1, 0.2, ..., 0.9$. Observe how the difference between $MSPE$s changes with γ.

Consequently, it is important to perform a test to check whether the history of z_2 helps forecast z_1 and whether the history of z_1 help forecast z_2. Specifically, if the history (i.e. lagged observations) of variable x does not help predict the future values of variable y (given lagged values of y and lagged values of other variables), we say that x does not Granger-cause y.[2]

[2] Granger, C.W.J. (1969): "Investigating causal relations by econometric models and cross-spectral methods." *Econometrica* 37 (3), 424–438.

For instance, consider the VAR(1) model.

$$\begin{pmatrix} z_{1t} \\ z_{2t} \end{pmatrix} = \begin{pmatrix} \Phi_0(1) \\ \Phi_0(2) \end{pmatrix} + \begin{pmatrix} \Phi_1(1,1) & \Phi_1(1,2) \\ \Phi_1(2,1) & \Phi_1(2,2) \end{pmatrix} \begin{pmatrix} z_{1t-1} \\ z_{2t-1} \end{pmatrix} + \begin{pmatrix} \eta_{1t} \\ \eta_{2t} \end{pmatrix}$$

$$z_{1t} = \Phi_0(1) + \Phi_1(1,1) z_{1t-1} + \Phi_1(1,2) z_{2t-1} + \eta_{1t}$$

$$z_{2t} = \Phi_0(2) + \Phi_1(2,1) z_{1t-1} + \Phi_1(2,2) z_{2t-1} + \eta_{2t}$$

(1) The Granger-causality test of the null of z_2 does not Granger-cause z_1 is done by

$$H_0\colon\ \Phi_1(1,2) = 0 \text{ versus } H_1\colon\ \Phi_1(1,2) \neq 0$$

$$z_{1t} = \Phi_0(1) + \Phi_1(1,1)z_{1t-1} + \Phi_1(1,2)z_{2t-1} + \eta_{1t}$$

(2) The Granger-causality test of the null of z_1 does not Granger-cause z_2 is done by

$$H_0\colon\ \Phi_1(2,1) = 0 \text{ versus } H_1\colon\ \Phi_1(2,1) \neq 0$$

$$z_{2t} = \Phi_0(2) + \Phi_1(2,1)z_{1t-1} + \Phi_1(2,2)z_{2t-1} + \eta_{2t}$$

In this case, since only one parameter is involved, the test can be conducted using t-statistics.

Consider the VAR(2) model.

$$\begin{pmatrix} z_{1t} \\ z_{2t} \end{pmatrix} = \begin{pmatrix} \Phi_0(1) \\ \Phi_0(2) \end{pmatrix} + \begin{pmatrix} \Phi_1(1,1) & \Phi_1(1,2) \\ \Phi_1(2,1) & \Phi_1(2,2) \end{pmatrix} \begin{pmatrix} z_{1t-1} \\ z_{2t-1} \end{pmatrix} + \begin{pmatrix} \Phi_2(1,1) & \Phi_2(1,2) \\ \Phi_2(2,1) & \Phi_2(2,2) \end{pmatrix} \begin{pmatrix} z_{1t-2} \\ z_{2t-2} \end{pmatrix} + \begin{pmatrix} \eta_{1t} \\ \eta_{2t} \end{pmatrix}$$

$$z_{1t} = \Phi_0(1) + \Phi_1(1,1)z_{1t-1} + \Phi_1(1,2)z_{2t-1} + \Phi_2(1,1)z_{1t-2} + \Phi_2(1,2)z_{2t-2} + \eta_{1t}$$

$$z_{2t} = \Phi_0(2) + \Phi_1(2,1)z_{1t-1} + \Phi_1(2,2)z_{2t-1} + \Phi_2(2,1)z_{1t-2} + \Phi_2(2,2)z_{2t-2} + \eta_{2t}$$

The Granger-causality test of the null of z_2 does not Granger-cause z_1 is done by
H_0: $\Phi_1(1,2) = 0, \Phi_2(1,2) = 0$ and versus H_1: $\Phi_1(1,2) \neq 0$ or $\Phi_2(1,2) \neq 0$

$$z_{1t} = \Phi_0(1) + \Phi_1(1,1)z_{1t-1} + \Phi_1(1,2)z_{2t-1} + \Phi_2(1,1)z_{1t-2} + \Phi_2(1,2)z_{2t-2} + \eta_{1t}$$

The Granger-causality test of the null of z_1 does not Granger-cause z_2 is done by
H_0: $\Phi_1(2,1) = 0, \Phi_2(2,1) = 0$ and versus H_1: $\Phi_1(2,1) \neq 0$ or $\Phi_2(2,1) \neq 0$

$$z_{2t} = \Phi_0(2) + \Phi_1(2,1)z_{1t-1} + \Phi_1(2,2)z_{2t-1} + \Phi_2(2,1)z_{1t-2} + \Phi_2(2,2)z_{2t-2} + \eta_{2t}$$

In this case, since more than one parameters are involved, the test cannot be conducted using t-statistics, but can be conducted using *log-likelihood ratio test* or *Wald test*.

3.5. **Estimation.** How to estimate the coefficients in a VAR model? Do we have to estimate the n equations jointly?

It is relative easy to estimate the coefficients of each equation by OLS. If we have the *same set of variables in the right hand side of each equations, OLS estimation of equation by equation is as efficient as the maximum likelihood*.

For instance, consider the 2-variable VAR(1) model.

$$\begin{pmatrix} z_{1t} \\ z_{2t} \end{pmatrix} = \begin{pmatrix} \Phi_0(1) \\ \Phi_0(2) \end{pmatrix} + \begin{pmatrix} \Phi_1(1,1) & \Phi_1(1,2) \\ \Phi_1(2,1) & \Phi_1(2,2) \end{pmatrix} \begin{pmatrix} z_{1t-1} \\ z_{2t-1} \end{pmatrix} + \begin{pmatrix} \eta_{1t} \\ \eta_{2t} \end{pmatrix}$$

We would estimate the coefficients of the following equations separately by OLS

- $z_{1t} = \Phi_0(1) + \Phi_1(1,1)z_{1t-1} + \Phi_1(1,2)z_{2t-1} + \eta_{1t}$
- $z_{2t} = \Phi_0(2) + \Phi_1(2,1)z_{1t-1} + \Phi_1(2,2)z_{2t-1} + \eta_{2t}$

When the set of variables *differ* across equations, OLS estimation of equation by equation is *no longer efficient*. To obtain efficient estimators, we will need to consider maximum likelihood or a seemingly uncorrelated regression estimation (i.e., SUR model).

For such efficiency, we sometimes include variables that are known to have zero coefficients (say, concluded via a Granger causality test).

3.6. **Model Selection.** To choose the appropriate lag length, we can use the criteria of AIC and SIC modified to suit the vector context.

R also reports other selection critieria. For instance, R also reports a sequential modified likelihood ratio (LR) test.[3] Starting from the pre-specified maximum lag, R tests the hypothesis that the coefficients on last lag are jointly zero based on some modified LR statistics. The modified LR statistics is compared to the 5% critical values starting from the maximum lag, and decreasing the lag one at a time until we first get a rejection of the null that the coefficients on last lag are jointly zero.

$$\text{VAR}(10) \text{ vs VAR}(9) \longrightarrow \text{VAR}(9) \text{ vs VAR}(8) \longrightarrow \text{VAR}(8) \text{ vs VAR}(7) \longrightarrow \ldots.$$

Refer to Hamilton (1994) and Lutkepohl (1993) for additional discussion.

[3]By default, R reports five criteria: (1) sequential modified LR test, (2) final prediction error, (3) Akaike information criterion, (4) Schwarz information criterion and (5) Hannan-Quinn information criterion.

3.7. **VARMA.** Like univariate time series models, building a ARMA model in the vector of variables is possible. However, the estimation requires maximum likelihood. Estimation may occasionally crash, partly due to the use of numerical optimization. Since MA can be approximated by AR, most researchers will find it convenient to use VAR instead of VARMA.

Refer to Lutkepohl (1993) for additional discussion of VARMA.

3.8. **Mapping into VAR(1).**

(1) Consider a univariate AR(2) process:

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t$$

It can be rewritten as 2-variable VAR of order 1:

$$\begin{pmatrix} y_t \\ y_{t-1} \end{pmatrix} = \begin{pmatrix} \phi_0 \\ 0 \end{pmatrix} + \begin{pmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ y_{t-2} \end{pmatrix} + \begin{pmatrix} \epsilon_t \\ 0 \end{pmatrix}$$

(2) Consider a univariate AR(3) process:

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + \epsilon_t$$

It can be rewritten as 3-variable VAR of order 1:

$$\begin{pmatrix} y_t \\ y_{t-1} \\ y_{t-2} \end{pmatrix} = \begin{pmatrix} \phi_0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} \phi_1 & \phi_2 & \phi_3 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ y_{t-2} \\ y_{t-3} \end{pmatrix} + \begin{pmatrix} \epsilon_t \\ 0 \\ 0 \end{pmatrix}$$

(3) Consider a k-variable VAR(2) process:

$$z_t = \Phi_0 + \Phi_1 z_{t-1} + \Phi_2 z_{t-2} + \eta_t$$

It can be rewritten as $2k$-variable VAR of order 1:

$$\begin{pmatrix} z_t \\ z_{t-1} \end{pmatrix} = \begin{pmatrix} \Phi_0 \\ 0 \end{pmatrix} + \begin{pmatrix} \Phi_1 & \Phi_2 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} z_{t-1} \\ z_{t-2} \end{pmatrix} + \begin{pmatrix} \eta_t \\ 0 \end{pmatrix}$$

VAR(1) is most studied. Mapping models into VAR(1) allows us to adopt the previous VAR(1) related results to make conclusion or estimation of these models.

4. Example: Predicting the Price of A share

ZTE Corporation is listed in mainland (A-share code: 000063) and Hong Kong (H-share code: 763).

The prices of A and H shares are supposedly driven by the similar fundamental factors and thus likely move together. Thus, VAR is an appropriate tool in capturing the "co-movement".

Data daily from 2010/1/1 to 2013/5/20, obtained from Bloomberg.

After removing all the non-trading days in at least one of the markets, we are left with 781 observations.

Estimation and selection use the first 600 observations. The remaining are for out of sample comparison.

We will work on the log share prices.

Benchmark no change model:

$$E(y_{T+1} \mid \Omega_T) = y_T$$

$$\begin{aligned} E(y_{T+2} \mid \Omega_T) &= E\left[E(y_{T+2} \mid \Omega_{T+1}) \mid \Omega_T\right] \\ &= E\left[y_{T+1} \mid \Omega_T\right] \\ &= y_T \end{aligned}$$

$$\begin{aligned} E(y_{T+3} \mid \Omega_T) &= E\left\{E\left[E(y_{T+3} \mid \Omega_{T+2}) \mid \Omega_{T+1}\right] \mid \Omega_T\right\} \\ &= E\left\{E\left[y_{T+2} \mid \Omega_{T+1}\right] \mid \Omega_T\right\} \\ &= E\left\{y_{T+1} \mid \Omega_T\right\} \\ &= y_T \end{aligned}$$

and so on.

FIGURE 4.1. Time series plot of log share prices of A and H

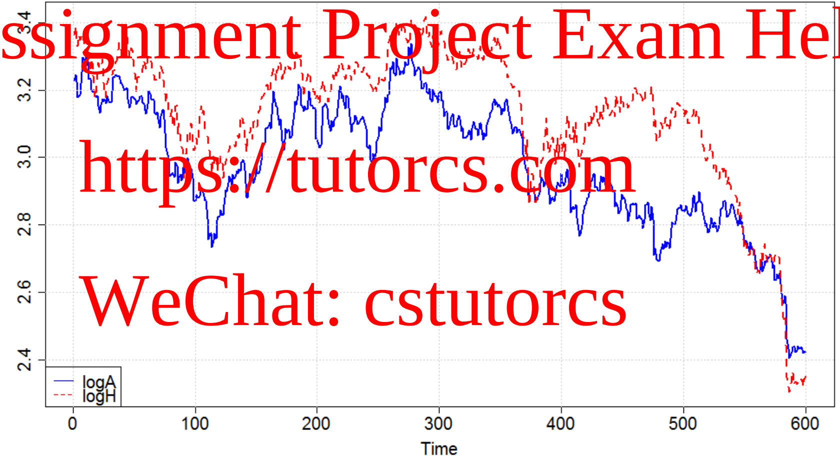


FIGURE 4.2. ACF of log share prices of A

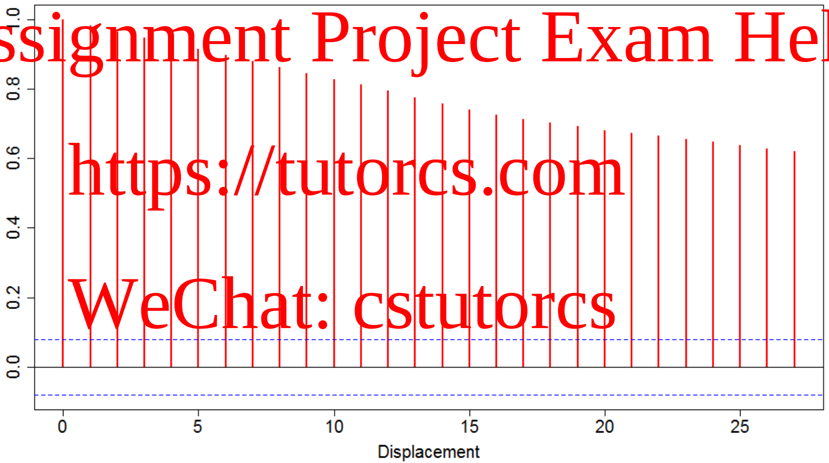


FIGURE 4.3. PACF of log share prices of A

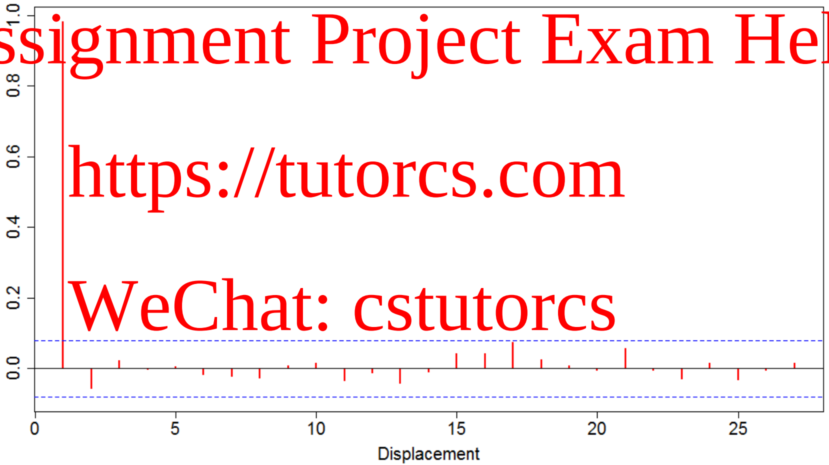


FIGURE 4.4. ACF of log share prices of H

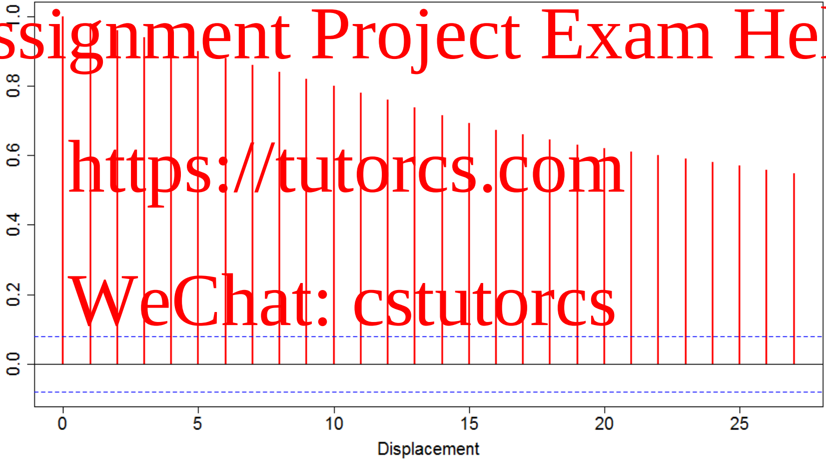


FIGURE 4.5. PACF of log share prices of H

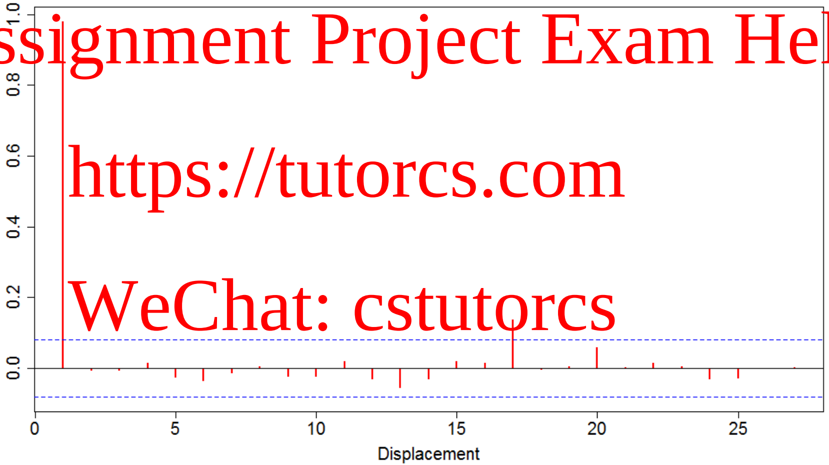


FIGURE 4.6. Cross-correlation Fucntion of log share prices of A and H

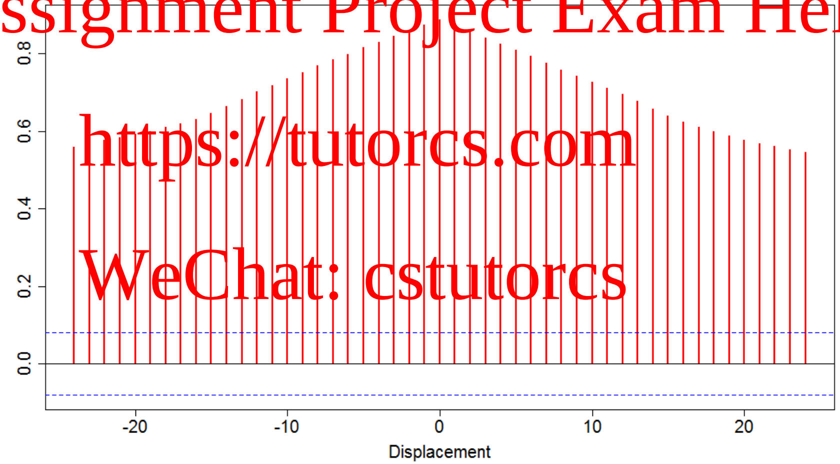


TABLE 4.1. Selection Criteria of VAR Order

n	1	2	3	4	5	6	7	8
AIC(n)	-14.900	-14.905	-14.896	-14.889	-14.880	-14.872	-14.860	-14.850
HQ(n)	-14.877	-14.871	-14.850	-14.832	-14.811	-14.792	-14.767	-14.746
SC(n)	-14.841	-14.817	-14.778	-14.741	-14.703	-14.665	-14.623	-14.583
FPE(n)	3.379	3.363	3.394	3.417	3.448	3.476	3.520	3.555

Thus, restricting ourselves to a maximum allowable order of 8, the implied VAR order is either 1 or 2. We choose the order of 1 because it is consistent with the pattern of ACF and PACF of the individual univariate series. Given the larger amount of data and the small VAR system (only two variables), increasing the order from 1 to 2 does not cause much loss in estimation efficiency. Indeed, we also consider the order of 2. The major conclusion does not change much. Here, we report only result of VAR(1).

Estimation results for equation logA (standard errors in parentheses):

$$\log A_t = \underset{(0.0113)}{0.9749} \log A_{t-1} + \underset{(0.0101)}{0.0228} \log H_{t-1} + \underset{(0.0158)}{0.0021}$$

Estimation results for equation logH:

$$\log H_t = \underset{(0.0131)}{0.0115} \log A_{t-1} + \underset{(0.0117)}{0.9923} \log H_{t-1} - \underset{(0.0181)}{0.0118}$$

FIGURE 4.7. ACF of residuals of log share prices of A

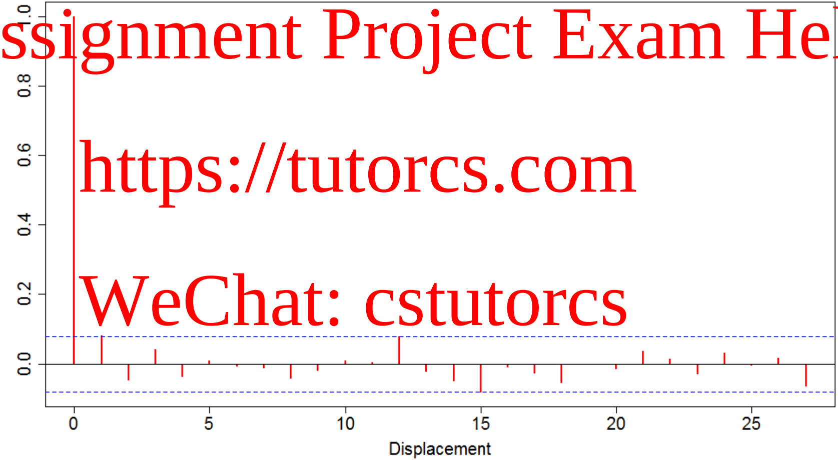


FIGURE 4.8. PACF of residuals log share prices of A

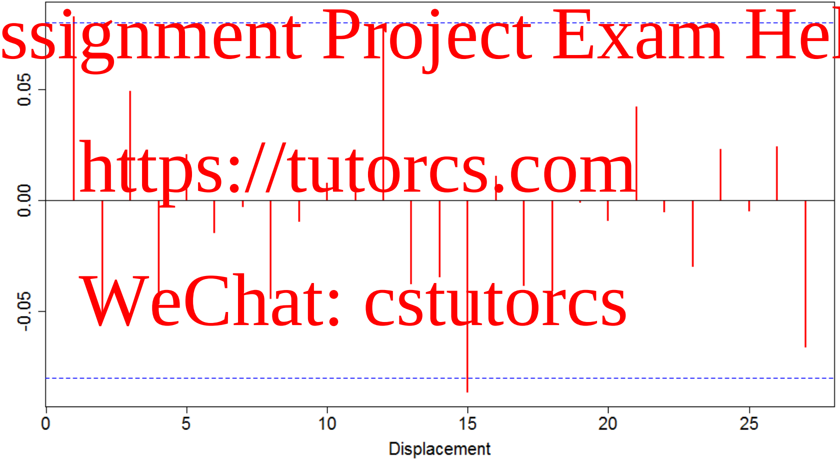


FIGURE 4.9. ACF of residuals log share prices of H

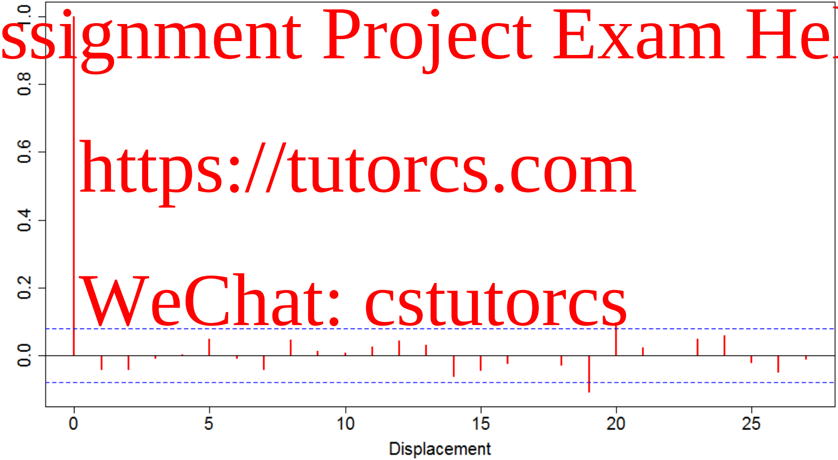


FIGURE 4.10. PACF of residuals of log share prices of H

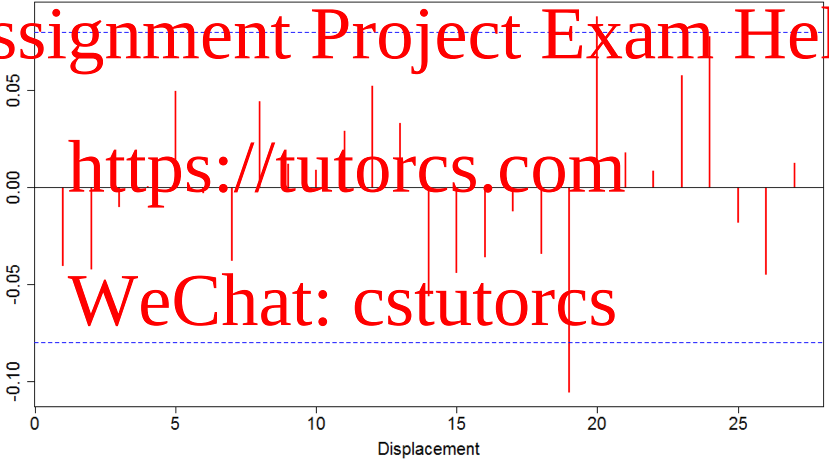
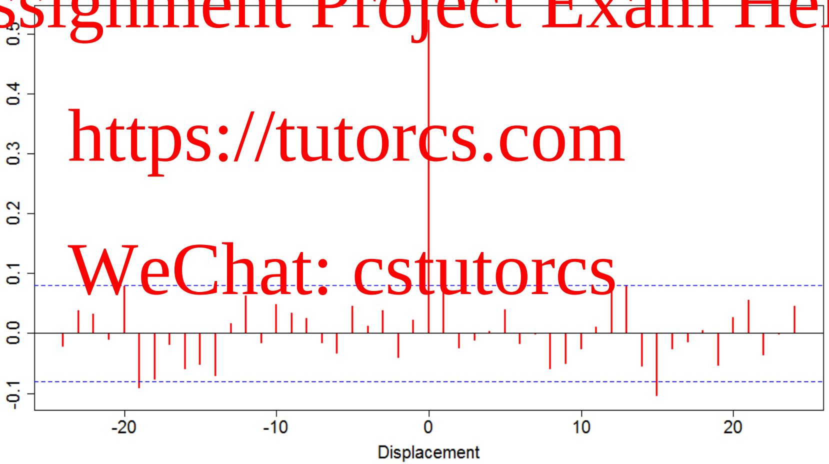


FIGURE 4.11. Cross-correlation Fucntion of residuals log share prices of A and H



Granger Casuality Tests

(1) Granger causality H_0: logA do not Granger-cause logH

$$\log H_t = \underset{(0.0131)}{0.0115} \log A_{t-1} + \underset{(0.0117)}{0.9923} \log H_{t-1} + \underset{(0.0181)}{0.0118}$$

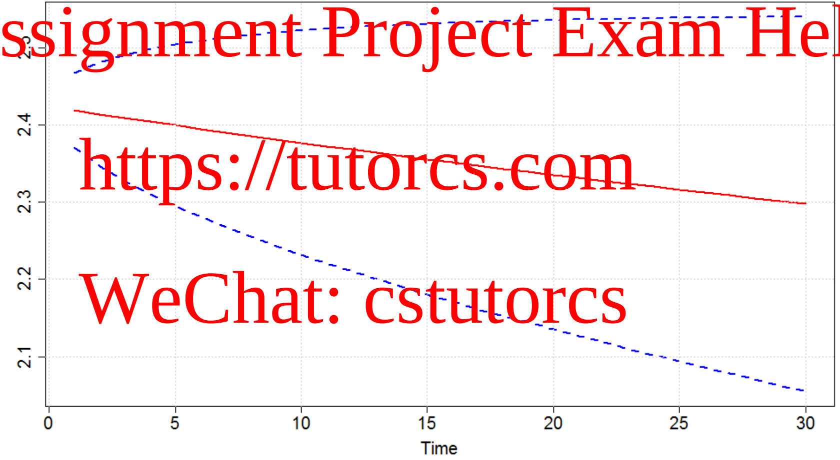
F-Test = 0.7697, df1 = 1, df2 = 1192, p-value = 0.3805

(2) Granger causality H0: logH do not Granger-cause logA

$$\log A_t = \underset{(0.0113)}{0.9749} \log A_{t-1} + \underset{(0.0101)}{0.0228} \log H_{t-1} + \underset{(0.0158)}{0.0021}$$

F-Test = 5.1044, df1 = 1, df2 = 1192, p-value = 0.02405

FIGURE 4.12. Forecast of log share prices of A based on VAR(1)



Out-of-sample comparison is based on "recursive scheme".

(1) Estimate with the observations 1 to R. Produce a h-step-ahead forecast, compute the forecast error.

(2) Estimate with the observations 1 to $R+1$. Produce a h-step-ahead forecast, compute the forecast error.

(3) Estimate with the observations 1 to $R+2$. Produce a h-step-ahead forecast, compute the forecast error.

...

Repeat with an increasing sample size until it is no longer possible to compute the forecast error.

The last estimation will use observations 1 to $T-h$.

FIGURE 4.13. Forecast Errors of VAR and No Change model ($h = 1$)

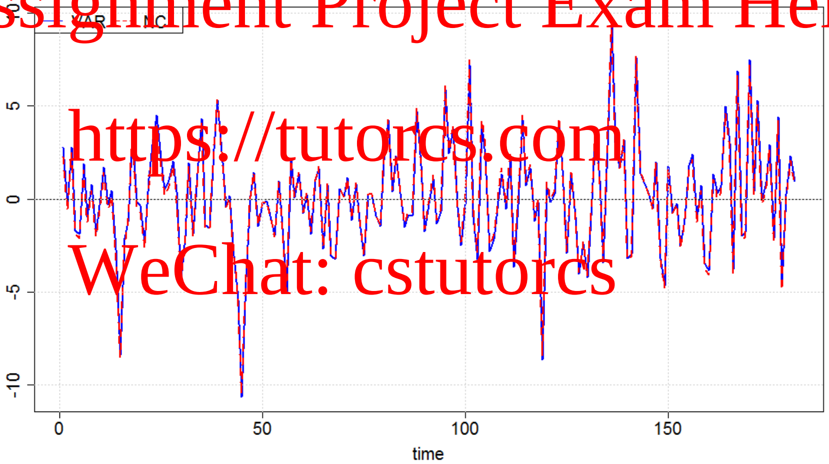


FIGURE 4.14. Forecast Errors of VAR and No Change model ($h = 3$)

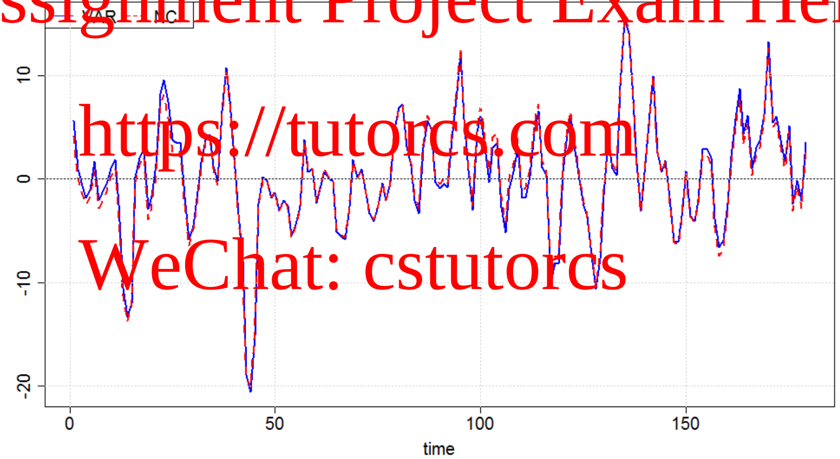


FIGURE 4.15. Forecast Errors of VAR and No Change model ($h = 6$)

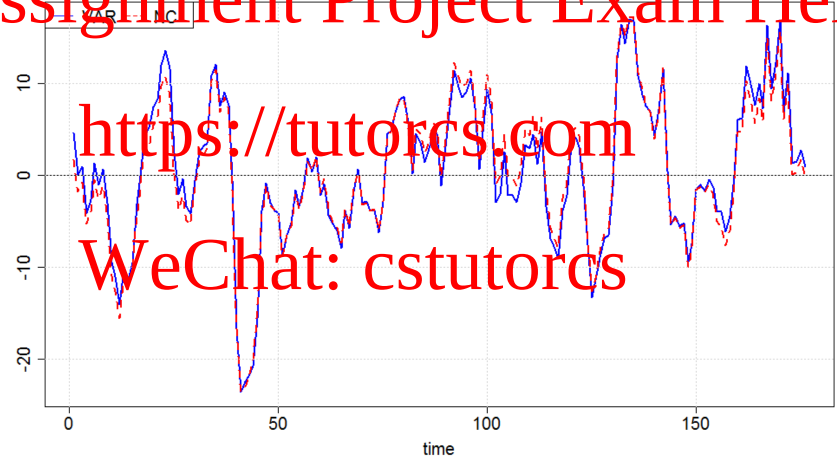


FIGURE 4.16. Forecast Errors of VAR and No Change model ($h = 12$)

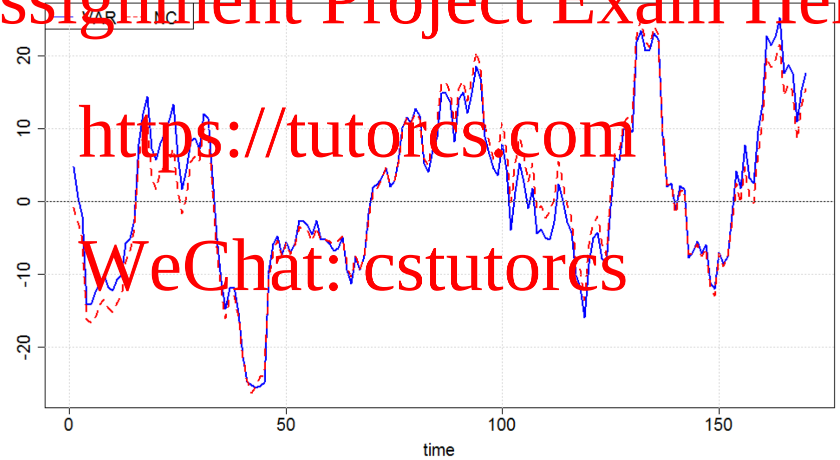


FIGURE 4.17. Forecast Errors of VAR and No Change model ($h = 24$)

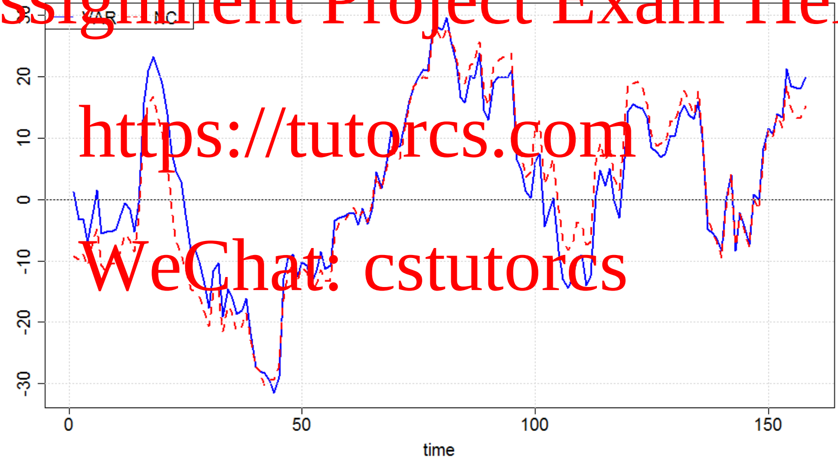


TABLE 4.2. A comparison based on MPE, RMSPE and MAPE

	$h=1$	$h=3$	$h=6$	$h=12$	$h=24$
MPE (VAR)	0.120	0.318	0.623	1.215	2.203
MPE (NC)	0.088	0.226	0.449	0.883	1.669
RMSPE (VAR)	2.898	5.389	7.746	11.019	13.803
RMSPE (NC)	2.888	5.355	7.707	10.97	14.408
MAPE (VAR)	2.170	3.989	6.068	9.089	11.378
MAPE (NC)	2.147	3.956	6.038	8.920	12.367

$$MPE = \frac{1}{n}\sum_{i=1}^{n} e_i \qquad RMSPE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} e_i^2} \qquad MAPE = \frac{1}{n}\sum_{i=1}^{n} \mid e_i \mid$$

Example: Predicting the Price of A share

TABLE 4.3. Correlation between the Predicted Change of VAR and Actual Change

	$h = 1$	$h = 3$	$h = 6$	$h = 12$	$h = 24$
Correlation	-0.0137	0.0038	0.0492	0.0901	0.3077
Directional accuracy	0.4530	0.4804	0.5341	0.5588	0.6772

The *correlation* of the *actual change* of log A Share price and the *predicted change*.

- If the correlation is equal to 1, the prediction is perfect.
- If the correlation is equal to 0, the prediction is useless.
- If the correlation is positive, the prediction is of some use.
- If the correlation is negative, the prediction is not to be trusted.

The *directional accuracy* (i.e., percentage of correct directional forecast)

$$\frac{\#[\text{sign}(\text{actual } \Delta \text{ at } h \text{ step}) = \text{sign}(\text{predicted } \Delta \text{of VAR at } h \text{ step})]}{\text{Total number of } h \text{ step prediction}}$$

If directional accuracy is larger than 0.5, the model is doing better than flipping a fair coin.