Evaluating and Combining Forecasts

Ka-fu WONG

University of Hong Kong

April 26, 2019

Contents

1. The Evaluation Situation

Imagine a company is about to hire one forecaster among two candidates. Suppose his major duty is to produce one-step ahead forecast of y_t.

- What kind of quality do we want him to posses?
- How do we evaluate and compare their quality of forecast?

To answer these questions, we have to first understand that there are three types of forecast schemes.

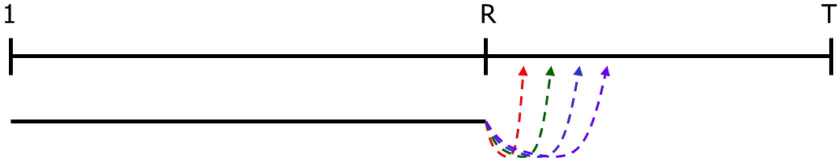
2. Forecast Schemes

Suppose we have T observations in our sample. We mimic an analyst starting to produce h-step-ahead forecasts at time R. $P \equiv T - R$ observations are reserved for out-of-sample comparison purpose.

2.1. **Fixed scheme.** Suppose we use all the information up to time R to produce forecast of y_t for the coming P periods, i.e., y_{R+1}, y_{R+2}, ..., y_T. (Note $R+P = T$.) We will use

- $y_{R+h,R}$ to denote a forecast of y_{R+h} using all information up to time R, and
- $e_{R+h,R}$ to denote the corresponding forecast errors, i.e., $e_{R+h,R} = y_{R+h} - y_{R+h,R}$.

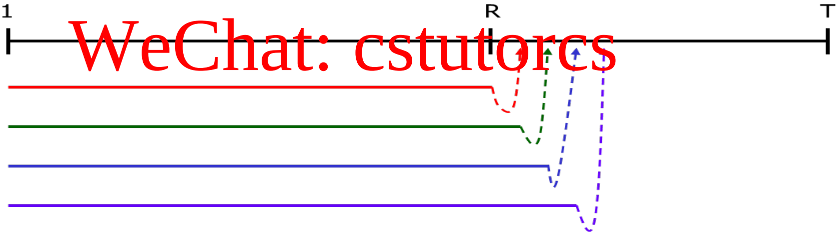
Thus, the forecasts and the forecast errors are $y_{R+1,R}$, $y_{R+2,R}$, $y_{R+3,R}\ldots$, $y_{T,R}$, and $e_{R+1,R}$, $e_{R+2,R}$, $e_{R+3,R}\ldots$, $e_{T,R}$. This scheme of producing the forecast is often called the fixed scheme because a *fixed* sample is used to produce forecasts of different horizons.



Est. sample		1-step-ahead forecast		Forecast error
$y_1, ..., y_R$	$\longrightarrow$	$y_{R+1,R}$	$\longrightarrow$	$e_{R+1,R} = y_{R+1} - y_{R+1,R}$
Est. sample		2-step-ahead forecast		Forecast error
$y_1, ..., y_R$	$\longrightarrow$	$y_{R+2,R}$	$\longrightarrow$	$e_{R+2,R} = y_{R+2} - y_{R+2,R}$
	...	...		...
Est. sample		P-step-ahead forecast		Forecast error
$y_1, ..., y_R$	$\longrightarrow$	$y_{R+P,R}$	$\longrightarrow$	$e_{R+P,R} = y_{R+P} - y_{R+P,R}$

2.2. **Recursive scheme.** Suppose we use all the information up to time R to produce forecast of y_t one period ahead ($h = 1$), i.e., y_{R+1}, and *use all the information up to time $R+1$ to produce forecast of y_t one period ahead, i.e., y_{R+2}*, and so on. We will use

- $y_{R+1,R}$ to denote a forecast of y_{R+1} using all information up to time R, and
- $e_{R+1,R}$ to denote the corresponding forecast errors, i.e., $e_{R+1,R} = y_{R+1} - y_{R+1,R}$.



Similarly, suppose we use all the information up to time R to produce forecast of y_t two period ahead ($h = 2$), i.e., y_{R+2}, and *use all the information up to time $R+1$ to produce forecast of y_t one period ahead, i.e., y_{R+3}*, and so on. We will use

- $y_{R+2,R}$ to denote a forecast of y_{R+2} using all information up to time R, and
- $e_{R+2,R}$ to denote the corresponding forecast errors, i.e., $e_{R+2,R} = y_{R+2} - y_{R+2,R}$.

Thus, the one-step-ahead forecasts and the forecast errors are $y_{R+1,R}$, $y_{R+2,R+1}$, $y_{R+3,R+2}$..., $y_{T,T-1}$, and $e_{R+1,R}$, $e_{R+2,R+1}$, $e_{R+3,R+2}$..., $e_{T,T-1}$. Again, note that $T = R + P$.

The forecast horizon of course can be set larger than 1.

The two-step-ahead forecasts and the forecast errors are $y_{R+2,R}$, $y_{R+3,R+1}$, $y_{R+4,R+2}$..., $y_{T,T-2}$, and $e_{R+2,R}$, $e_{R+3,R+1}$, $e_{R+4,R+2}$..., $e_{T,T-2}$.

This scheme of producing the forecast is often called the *recursive* scheme because a *recursively larger sample is used* to produce forecasts of a fixed horizon.

$$
\begin{array}{ccccc}
 & & \text{1-step-ahead} & & \\
\text{Est. sample} & & \text{forecast} & & \text{Forecast error} \\
y_1, \ldots, y_R & \longrightarrow & y_{R+1,R} & \longrightarrow & e_{R+1,R} = y_{R+1} - y_{R+1,R} \\
y_1, \ldots, y_{R+1} & \longrightarrow & y_{R+2,R+1} & \longrightarrow & e_{R+2,R+1} = y_{R+2} - y_{R+2,R+1} \\
y_1, \ldots, y_{R+2} & \longrightarrow & y_{R+3,R+2} & \longrightarrow & e_{R+3,R+2} = y_{R+3} - y_{R+3,R+2} \\
 & \ldots & \ldots & & \ldots \\
y_1, \ldots, y_{T-1} & \longrightarrow & y_{T,T-1} & \longrightarrow & e_{T,T-1} = y_T - y_{T,T-1} \\
 & & & & \uparrow \\
 & & & & P \text{ of them}
\end{array}
$$

Est. sample		2-step-ahead forecast		Forecast error
$y_1, \ldots, y_R$	$\longrightarrow$	$y_{R+2,R}$	$\longrightarrow$	$e_{R+2,R} = y_{R+2} - y_{R+2,R}$
$y_1, \ldots, y_{R+1}$	$\longrightarrow$	$y_{R+3,R+1}$	$\longrightarrow$	$e_{R+3,R+1} = y_{R+3} - y_{R+3,R+1}$
$y_1, \ldots, y_{R+2}$	$\longrightarrow$	$y_{R+4,R+2}$	$\longrightarrow$	$e_{R+4,R+2} = y_{R+4} - y_{R+4,R+2}$
	$\ldots$	$\ldots$		$\ldots$
$y_1, \ldots, y_{T-2}$	$\longrightarrow$	$y_{T,T-2}$	$\longrightarrow$	$e_{T,T-2} = y_T - y_{T,T-2}$
				$\uparrow$
				$(P-1)$ of them

Est. sample		h-step-ahead forecast		Forecast error
$y_1, \ldots, y_R$	$\longrightarrow$	$y_{R+h,R}$	$\longrightarrow$	$e_{R+h,R} = y_{R+h} - y_{R+h,R}$
$y_1, \ldots, y_{R+1}$	$\longrightarrow$	$y_{R+1+h,R+1}$	$\longrightarrow$	$e_{R+1+h,R+1} = y_{R+1+h} - y_{R+1+h,R+1}$
$y_1, \ldots, y_{R+2}$	$\longrightarrow$	$y_{R+2+h,R+2}$	$\longrightarrow$	$e_{R+2+h,R+2} = y_{R+2+h} - y_{R+2+h,R+2}$
	...	...		...
$y_1, \ldots, y_{T-h}$	$\longrightarrow$	$y_{T,T-h}$	$\longrightarrow$	$e_{T,T-h} = y_T - y_{T,T-h}$
				$\uparrow$
				$(P-h+1)$ of them

2.3. **Rolling scheme.** Suppose we fix the size of the sample used for estimation. That is, we use the information from period 1 up to period R to produce a forecast of y_t one period ahead, i.e., y_{R+1}. We will use

- $y_{R+1,R}$ to denote a forecast of y_{R+1} using the information from period 1 up to time R, and
- $e_{R+1,R}$ to denote the corresponding forecast errors, i.e., $e_{R+1,R} = y_{R+1} - y_{R+1,R}$.

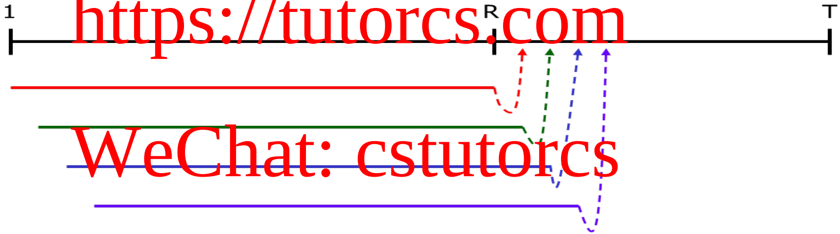
We *use the information from period 2 up to time $R+1$ to produce a forecast of y_t one period ahead, i.e., y_{R+2}*,

- $y_{R+2,R+1}$ to denote a forecast of y_{R+2} using the information from period 2 up to time $R+1$, and
- $e_{R+2,R+1}$ to denote the corresponding forecast errors, i.e., $e_{R+2,R+1} = y_{R+2} - y_{R+2,R+1}$.

and so on.

Despite the risk of confusing readers, we have chosen to use the same notations as in the recursive scheme in order to avoid additional subscripts or superscripts. Readers please take note.

This scheme of producing the forecast is often called the *rolling* scheme because *a sample of fixed size rolls along* to produce forecasts of a fixed horizon.



Est. sample		1-step-ahead forecast		Forecast error
$y_1, \ldots, y_R$	$\longrightarrow$	$y_{R+1,R}$	$\longrightarrow$	$e_{R+1,R} = y_{R+1} - y_{R+1,R}$
$y_2, \ldots, y_{R+1}$	$\longrightarrow$	$y_{R+2,R+1}$	$\longrightarrow$	$e_{R+2,R+1} = y_{R+2} - y_{R+2,R+1}$
$y_3, \ldots, y_{R+2}$	$\longrightarrow$	$y_{R+3,R+2}$	$\longrightarrow$	$e_{R+3,R+2} = y_{R+3} - y_{R+3,R+2}$
	...	...		...
$y_{T-R}, \ldots, y_{T-1}$	$\longrightarrow$	$y_{T,T-1}$	$\longrightarrow$	$e_{T,T-1} = y_T - y_{T,T-1}$
				$\uparrow$
				P of them

Est. sample		2-step-ahead forecast		Forecast error
$y_1, ..., y_R$	$\longrightarrow$	$y_{R+2,R}$	$\longrightarrow$	$e_{R+2,R} = y_{R+2} - y_{R+2,R}$
$y_2, ..., y_{R+1}$	$\longrightarrow$	$y_{R+3,R+1}$	$\longrightarrow$	$e_{R+3,R+1} = y_{R+3} - y_{R+3,R+1}$
$y_3, ..., y_{R+2}$	$\longrightarrow$	$y_{R+4,R+2}$	$\longrightarrow$	$e_{R+4,R+2} = y_{R+4} - y_{R+4,R+2}$
	...	...		...
$y_{T-R-1}, ..., y_{T-2}$	$\longrightarrow$	$y_{T,T-2}$	$\longrightarrow$	$e_{T,T-2} = y_T - y_{T,T-2}$
				$\uparrow$
				$(P-1)$ of them

$$
\begin{array}{ccccc}
\text{Est. sample} & & h\text{-step-ahead forecast} & & \text{Forecast error} \\
y_1, \ldots, y_R & \longrightarrow & y_{R+h,R} & \longrightarrow & e_{R+h,R} = y_{R+h} - y_{R+h,R} \\
y_2, \ldots, y_{R+1} & \longrightarrow & y_{R+1+h,R+1} & \longrightarrow & e_{R+1+h,R+1} = y_{R+1+h} - y_{R+1+h,R+1} \\
y_3, \ldots, y_{R+2} & \longrightarrow & y_{R+2+h,R+2} & \longrightarrow & e_{R+2+h,R+2} = y_{R+2+h} - y_{R+2+h,R+2} \\
 & \cdots & \cdots & & \cdots \\
y_{T-R-h+1}, \ldots, y_{T-h} & \longrightarrow & y_{T,T-h} & \longrightarrow & e_{T,T-h} = y_T - y_{T,T-h} \\
 & & & & \uparrow \\
 & & & & (P-h+1) \text{ of them}
\end{array}
$$

2.4. **Fixed, Recursive or Rolling.**

It is easy to see that the recursive scheme is similar to the way forecasters produce their forecasts and how they are likely evaluated.

Note that in the rolling scheme, the information at the beginning of the last sample is ignored as we roll along. Why do we want to throw away some information when such information is in fact available to us? One reason is that rolling scheme can *help avoid the influence of structural breaks* in the data.

With a long time series, structural breaks are common. So, we may choose to ignore data before the structural break. If, however, we do not know when the structural breaks happen, we may want to constantly throw away some old information.

In the following discussions of forecast evaluation, we will *focus* on the forecast due to the *recursive* scheme or the *rolling* scheme.

3. Optimality of forecast

There are four key properties of optimal forecast.

(1) Optimal forecasts are unbiased

(2) Optimal forecasts have 1-step-ahead forecast errors that are white noise

(3) Optimal forecasts have h-step-ahead forecast errors that are at most $MA(h-1)$

(4) Optimal forecasts have h-step-ahead forecast errors with variances that are non-decreasing in h and that converge to the unconditional variance of the process.

To fix ideas for the discussion below, suppose we are interested in forecasting a *covariance stationary* series with the following *underlying Wold representation*

$$
\begin{aligned}
y_t &= \mu + \epsilon_t + b_1\epsilon_{t-1} + b_2\epsilon_{t-2} + \dots \\
\epsilon_t &\sim WN(0, \sigma^2)
\end{aligned}
$$

where the mean of y_t is μ, nonzero.

Of course, we can easily extend the discussion to more complicated data generating processes (DGP) with trend and seasonality components. In that case, all we have to do is to modify μ to allow it to have trend and seasonality components.

3.1. **Unbiasedness.** A h-step-ahead forecast $y_{t+h,t}$ is unbiased if

$$y_{t+h,t} = E(y_{t+h} \mid y_t, y_{t-1}, y_{t-2}, \ldots)$$

or equivalently

$$\begin{aligned} E(e_{t+h,t} \mid y_t, y_{t-1}, \ldots) &= E[(y_{t+h} - y_{t+h,t}) \mid y_t, y_{t-1}, \ldots] \\ &= E(y_{t+h} \mid y_t, y_{t-1}, \ldots) - E(y_{t+h,t} \mid y_t, y_{t-1}, \ldots) \\ &= E(y_{t+h} \mid y_t, y_{t-1}, \ldots) - E(y_{t+h} \mid y_t, y_{t-1}, \ldots) \\ &= 0 \end{aligned}$$

where the third line in the above equation is due to the *law of iterated expectation* that $E(E(A \mid B) \mid B) = E(A \mid B)$.

Again, by law of iterated expectation,

$$E(e_{t+h,t} \mid y_t, y_{t-1}, y_{t-2}, \ldots) = 0 \text{ implies } E(e_{t+h,t}) = 0.$$

In producing the forecast, it is unlikely we will know the DGP. That is, we may *not know* what the conditional expectation looks like, or whether our forecast is the same as the conditional expectation, $E(y_{t+h} \mid y_t, y_{t-1}, y_{t-2}, ...)$. In this case, we often want to check the unbiasedness of our forecast.

Suppose we have a set of *1-step-ahead* forecast errors, $e_{R+1,R}$, $e_{R+2,R+1}$, ..., $e_{T,T-1}$ from some forecaster or forecasting model. How do we test whether the forecaster is unbiased?

One possibility is to compute the average of the P forecast errors and test:

H_0: $E(e_{t+1,t}) = 0$ *versus* H_1: $E(e_{t+1,t}) \neq 0$

using the usual hypothesis testing procedures.

An alternative is to run a regression of

$$e_{t+1,t} = \alpha + v_t \qquad t = R, R+1, ..., T-1$$

where α is a constant and v_t the residual. Then, it is easy to perform the usual test

$$H_0\text{: } \alpha = 0 \text{ versus } H_1\text{: } \alpha \neq 0$$

using a *heteroskedasticity consistent standard errors* of $\hat{\alpha}$ ($\hat{\alpha}$=the estimator of α).

Suppose we have a set of *2-step-ahead* forecast errors, $e_{R+2,R}$, $e_{R+3,R+1}$, ..., $e_{T,T-2}$ from some forecaster or forecasting model. How do we test whether the forecaster is unbiased?

One possibility is to compute the average of the $P-1 = T-R-1$ forecast errors and test

$$H_0\text{: } E(e_{t+2,t}) = 0 \textit{ versus } H_1\text{: } E(e_{t+2,t}) \neq 0$$

using the usual hypothesis testing procedures.

An alternative is to run a regression of

$$e_{t+2,t} = \alpha + v_t \qquad t = R, R+1, ..., T-2$$

where α is a constant and v_t the residual. Then, it is easy to perform the usual test

$$H_0\text{: } \alpha = 0 \textit{ versus } H_1\text{: } \alpha \neq 0$$

using a *heteroskedasticity consistent standard errors* of $\hat{\alpha}$ ($\hat{\alpha}$=the estimator of α).

More generally, if we have a set of *h-step-ahead* forecast errors, $e_{R+h,R}$, $e_{R+h+1,R+1}$, ..., $e_{T,T-h}$ from some forecaster or forecasting model. To test whether the forecaster is unbiased, one possibility is to compute the average of the $P-h+1 = T-R-h+1$ forecast errors and test

$$H_0 : E(e_{t+h,t}) = 0 \text{ versus } H_1 : E(e_{t+h,t}) \neq 0$$

using the usual hypothesis testing procedures.

An alternative is to run a regression of

$$e_{t+h,t} = \alpha + v_t \qquad t = R, R+1, ..., T-h$$

where α is a constant and v_t the residual. Then, it is easy to perform the usual test

$$H_0 : \alpha = 0 \text{ versus } H_1 : \alpha \neq 0$$

using a heteroskedasticity consistent standard errors of $\hat{\alpha}$ ($\hat{\alpha}$=the estimator of α).

3.2. **1-step-ahead forecast errors that are white noise.** As discussed earlier, when the underlying process is

$$\begin{aligned} y_t &= \mu + \epsilon_t + b_1\epsilon_{t-1} + b_2\epsilon_{t-2} + \dots \\ \epsilon_t &\sim WN(0, \sigma^2) \end{aligned}$$

and hence

$$y_{t+1} = \mu + \epsilon_{t+1} + b_1\epsilon_t + b_2\epsilon_{t-1} + \dots$$

The optimal 1-step-ahead forecast is

$$y_{t+1,t} = E(y_{t+1} \mid y_t, y_{t-1}, y_{t-2}, \dots) = \mu + b_1\epsilon_t + b_2\epsilon_{t-1} + b_3\epsilon_{t-2} + \dots$$

and the forecast error

$$e_{t+h,t} = y_{t+1} - y_{t+1,t} = \epsilon_{t+1}$$

That is, if the forecast is optimal, the 1-step-ahead forecast is supposedly white noise.

When the forecast errors are not white noise, there is still room for improvement by including additional ARMA terms in the model. Thus, we often want to check whether 1-step-ahead forecast errors are white noise.

To *check* whether the *1-step-ahead* forecast errors are *white noise*, we can regress the errors on a constant term and then check

- the correlogram (ACF and PACF),
- the Durbin-Watson (for first autocorrelation), and/or
- the Box-Pierce and Ljung-Box statistics

of the residuals.

3.3. **h-step-ahead forecast errors are at most $MA(h-1)$.** Recall if we have

$$\begin{aligned} y_t &= \mu + \epsilon_t + b_1\epsilon_{t-1} + b_2\epsilon_{t-2} + \ldots \\ y_{t+h} &= \mu + \epsilon_{t+h} + b_1\epsilon_{t+h-1} + b_2\epsilon_{t+h-2} + \ldots + b_{h-1}\epsilon_{t+1} + b_h\epsilon_t + b_{h+1}\epsilon_{t-1} + \ldots \end{aligned}$$

the h-step-ahead forecast error is

$$e_{t+h,t} = y_{t+h} - y_{t+h,t} = \epsilon_{t+h} + b_1\epsilon_{t+h-1} + b_2\epsilon_{t+h-2} + \ldots + b_{h-1}\epsilon_{t+1}$$

h	$e_{t+h,t}$
1	ϵ_{t+1}
2	$\epsilon_{t+2} + b_1\epsilon_{t+1}$
3	$\epsilon_{t+3} + b_1\epsilon_{t+2} + b_2\epsilon_{t+1}$
4	$\epsilon_{t+4} + b_1\epsilon_{t+3} + b_2\epsilon_{t+2} + b_3\epsilon_{t+1}$
...	...

That is, a $MA(h-1)$ structure is expected in the forecast error for optimal forecast. The $MA(h-1)$ structure implies a cutoff in the forecast error's auto-correlation function beyond displacement $(h-1)$. To check this property visually, we can look at the plots of autocorrelation function.

Statistically, we can run a regression of $e_{t+h,t}$ on a constant term *allowing for $MA(q)$ disturbances with $q > (h-1)$*, and test whether the moving-average parameters beyond lag $h-1$ are zero.

Suppose we allow $q = h$. That means we run the regression

$$e_{t+h,t} = \alpha + w_t + \theta_1 w_{t-1} + \ldots + \theta_{h-1} w_{t-h+1} + \theta_h w_{t-h}$$

and *test* $\theta_h = 0$.

3.4. h-step-ahead forecast errors with variances that are non-decreasing in h. As discussed earlier, when the underlying process is

$$\begin{aligned} y_t &= \mu + \epsilon_t + b_1\epsilon_{t-1} + b_2\epsilon_{t-2} + \ldots \\ \epsilon_t &\sim WN(0, \sigma^2) \end{aligned}$$

and we have the optimal forecast of

$$y_{t+h,t} = E(y_{t+h} \mid y_t, y_{t-1}, y_{t-2}, \ldots) = \mu + b_h\epsilon_t + b_{h+1}\epsilon_{t-1} + b_{h+2}\epsilon_{t-2} + \ldots$$

the forecast error

$$e_{t+h,t} = y_{t+h} - y_{t+h,t} = \epsilon_{t+h} + b_1\epsilon_{t+h-1} + b_2\epsilon_{t+h-2} + \ldots + b_{h-1}\epsilon_{t+1}$$

the *variance of the h-step-ahead forecast error* is

$$\begin{aligned} Var(e_{t+h,t}) &= Var(\epsilon_{t+h} + b_1\epsilon_{t+h-1} + \ldots + b_{h-1}\epsilon_{t+1}) \\ &= Var(\epsilon_{t+h}) + Var(b_1\epsilon_{t+h-1}) + \ldots + Var(b_{h-1}\epsilon_{t+1}) \\ &\quad +\text{a bunch of covariances} \\ &= \sigma^2 + b_1^2\sigma^2 + \ldots + b_{h-1}^2\sigma^2 \end{aligned}$$

That is,

$$\sigma_h^2 = \sigma^2 \left(1 + \sum_{i=1}^{h-1} b_i^2 \right)$$

which is increasing in h, or at least *non-decreasing in h*.

h	σ_h^2
1	σ^2
2	$\sigma^2(1 + b_1^2)$
3	$\sigma^2(1 + b_1^2 + b_2^2)$
4	$\sigma^2(1 + b_1^2 + b_2^2 + b_3^2)$
...	...

3.5. **Unforecastability principle.** Optimal forecast errors should be unforecastable on the basis of information available at the time the forecast was made.

To assess the forecastability, we can run the following regression

$$e_{t+h,t} = \alpha_0 + \sum_{i=1}^{k} \alpha_i x_{it} + v_t \qquad t = R, R+1, ..., T-h$$

where x_{it} *are information available at time t when the forecast was made*.

Unforecastability implies

$$\alpha_0 = \alpha_1 = \alpha_2 = ... = \alpha_k = 0$$

The simplest piece of information that is available at time t is the h-step-ahead forecast made at time t, $y_{t+h,t}$. Thus, a simple test is to run the *Mincer-Zarnowitz regression* of

$$e_{t+h,t} = \alpha_0 + \alpha_1 y_{t+h,t} + v_t \qquad t = R, R+1, \ldots, T-h \tag{3.1}$$

Since unbiasedness and unforecastability implies $\alpha_0 = \alpha_1 = 0$, we will test for H_0: $(\alpha_0, \alpha_1) = (0, 0)$ versus H_1: $(\alpha_0, \alpha_1) \neq (0, 0)$.

Using $e_{t+h,t} = y_{t+h} - y_{t+h,t}$, we can rewrite equation (3.1) into

$$\begin{aligned} y_{t+h} - y_{t+h,t} &= \alpha_0 + \alpha_1 y_{t+h,t} + v_t \\ y_{t+h} &= \alpha_0 + (\alpha_1 + 1) y_{t+h,t} + v_t \\ y_{t+h} &= \beta_0 + \beta_1 y_{t+h,t} + v_t \end{aligned} \tag{3.2}$$

and we will test for H_0: $(\beta_0, \beta_1) = (0, 1)$ versus H_1: $(\beta_0, \beta_1) \neq (0, 1)$.

4. Measures of Forecast Accuracy

There are several common measures of forecast accuracy. Note that forecast errors are sometimes called prediction errors and the two terms are often used interchangeably.

(1) Mean forecast errors

$$MFE = \frac{1}{P-h+1}\sum_{t=R}^{T-h} e_{t+h,t}$$

(2) Mean squared forecast errors

$$MSFE = \frac{1}{P-h+1}\sum_{t=R}^{T-h} e^2_{t+h,t}$$

(3) Root mean squared forecast errors

$$RMSFE = \sqrt{MSFE} = \sqrt{\frac{1}{P-h+1}\sum_{t=R}^{T-h} e_{t+h,t}^2}$$

Let the variance of forecast errors be

$$VFE = \frac{1}{P-h+1}\sum_{t=R}^{T-h} (e_{t+h,t} - ME)^2$$

We can verify that MSFE are related to MFE and VFE

$$MSFE = MFE^2 + VFE$$

Sometimes, we might be concern the prediction errors *as a percentage of the realized data*, i.e., the percentage prediction errors,

$$p_{t+h,t} = \frac{(y_{t+h} - y_{t+h,t})}{y_{t+h}}$$

We can then define the various measures of the forecast accuracy similarly:

(1) Mean percentage forecast errors

$$MPFE = \frac{1}{P-h+1}\sum_{t=R}^{T-h} p_{t+h,t}$$

(2) Mean squared percentage forecast errors

$$MSPFE = \frac{1}{P-h+1}\sum_{t=R}^{T-h} p_{t+h,t}^2$$

(3) Root mean squared percentage forecast errors

$$RMSPFE = \sqrt{MSPFE} = \sqrt{\frac{1}{P-h+1}\sum_{t=R}^{T-h} p_{t+h,t}^2}$$

5. Meese and Rogoff (1983)

In 1983, Richard Meese and Kenneth Rogoff[1] reported a striking result that sparked a series of research on the comparison of forecast accuracy (sometimes known as predictive accuracy).

Before the research by Meese and Rogoff, most studies of exchange rate models claimed success based on the significance of coefficients and R-squares of the regressions. R-square is essentially a measure of in-sample accuracy. Meese and Rogoff took a different route. They used *out-of-sample* forecast performance as a *metric* to compare several popular structural models at the time with some simple time series models.

[1] Meese, Richard A. and Kenneth Rogoff (1983): "Empirical Exchange Rate Models of the Seventies: Do they fit out of sample?" *Journal of International Economics*, 14: 3-24.

5.1. **Structural models.** The structural models considered can be summarized as in a regression model

$$s = a_0 + a_1(m - \dot{m}) + a_2(y - \dot{y}) + a_3(r_s - \dot{r}_s) + a_4(\pi^e - \dot{\pi}^e) + a_5\overline{TB} + a_6\dot{\overline{TB}} + u$$

where

- s is the logarithm of the dollar price of foreign currency,
- $(m - \dot{m})$ the logarithm of the ratio of the US money supply to the foreign money supply,
- $(y - \dot{y})$ the logarithm of the ratio of the US to the foreign real income,
- $(r_s - \dot{r}_s)$ the short-term interest rate differential,
- $(\pi^e - \dot{\pi}^e)$ the expected long-run inflation rate differential,
- $\overline{TB}$ the cumulative US trade balance,
- $\dot{\overline{TB}}$ the cumulative foreign trade balance.

As noted by Meese and Rogoff,

- all models posit that the exchange rate exhibits *first-degree homogeneity in the relative money supplies*, i.e., $a_1 = 1$.
- The Frenkel-Bilson model assumes *purchasing power parity*, i.e., $a_4 = a_5 = a_6 = 0$.
- The Dornbusch-Frankel model allows for domestic price adjustment and consequent *deviations from purchasing power parity*, i.e., $a_5 = a_6 = 0$.
- The Hooper-Morton model imposes *no constraints on the coefficients*.

5.2. **Time series models.** Univariate and multivariate time series models were considered. A variety of prefiltering techniques (differencing, deseasonalizing, and de-trending) were applied to the data. Lag lengths were selected using standard model selection criteria such as AIC and SIC. Generally, the univariate and multivariate time series models can be written as

$$s_t = a_{i1}s_{t-1} + a_{i2}s_{t-2} + \ldots + a_{in}s_{t-n} + B_{i1}^{'}X_{t-1} + B_{i2}^{'}X_{t-2} + \ldots + B_{in}^{'}X_{t-n} + u_{it}$$

where

- s_t is the i-th variable in the multivariate time series model and
- the subscript i is used to index the equation i coefficients,
- X_t is the additional vector of explanatory variables included in the univariate model or a vector of explanatory variables in a multivariate models (i.e., VAR).

5.3. Models that do not require estimation.

Random walk model:

$$s_t = s_{t-1} + u_t$$

The h-step-ahead forecast is

$$s_{t+h,t} = s_t$$

Forward rate model:

$$s_{t+h} = f_t(h) + u_t$$

where $f_t(h)$ is the h-step-ahead forward exchange rate, known at time t. Thus, the h-step-ahead forecast is

$$s_{t+h,t} = f_t(h)$$

5.4. **Forecast comparison.** Meese and Rogoff reported root mean squared forecast errors, mean absolute forecast errors and mean forecast errors. Their root mean squared forecast errors are reproduced below.

Exchange Rate	Horizon	Model: RW	FR	AR(p)	VAR(p)	FB	DF	HM
$/Mark	1 month	3.72	3.20	3.51	5.40	3.17	3.65	3.50
	6 months	8.71	9.03	12.40	11.83	9.64	12.03	9.95
	12 months	12.98	12.60	22.53	15.06	16.12	18.87	15.69
$/Yen	1 month	3.68	3.72	4.46	7.76	4.11	4.40	4.20
	6 months	11.58	11.93	22.04	18.90	13.38	13.94	11.94
	12 months	18.31	18.95	52.18	22.98	18.55	20.41	19.20
$/pound	1 month	2.56	2.67	2.79	5.56	2.82	2.90	3.03
	6 months	6.45	7.23	7.27	12.97	8.90	8.88	9.08
	12 months	9.96	11.62	13.35	21.28	14.62	13.66	14.57
Trade-	1 month	1.99	NA	2.72	4.10	2.40	2.50	2.74
weighted	6 months	6.09	NA	6.82	8.91	7.07	6.49	7.11
dollar	12 months	8.65	14.24	11.14	10.96	11.40	9.80	10.35

Meese and Rogoff (1983)

		Model:						
Exchange Rate	Horizon	RW	FR	AR(p)	VAR(p)	FB	DF	HM
$/Mark	1 month	1.00	0.86	0.94	1.45	0.85	0.98	0.94
	6 months	1.00	1.04	1.42	1.36	1.11	1.38	1.14
	12 months	1.00	0.97	1.74	1.16	1.24	1.45	1.21
$/Yen	1 month	1.00	1.01	1.21	2.11	1.12	1.20	1.14
	6 months	1.00	1.03	1.90	1.63	1.16	1.20	1.03
	12 months	1.00	1.03	2.85	1.26	1.01	1.11	1.05
$/pound	1 month	1.00	1.04	1.09	2.17	1.10	1.13	1.18
	6 months	1.00	1.12	1.13	2.01	1.38	1.38	1.41
	12 months	1.00	1.17	1.34	2.14	1.47	1.37	1.46
Trade-	1 month	1.00	NA	1.37	2.06	1.21	1.26	1.38
weighted	6 months	1.00	NA	1.12	1.46	1.16	1.07	1.17
dollar	12 months	1.00	1.65	1.29	1.27	1.32	1.13	1.20

From the table, it is easy to see that except for $/mark at 1-month-ahead forecast, random walk forecast achieved the smaller root mean squared forecast errors.

Exchange Rate	Horizon	Model: RW	FR	AR(p)	VAR(p)	FB	DF	HM
$/Mark	1 month	1.16	1.00	1.10	1.69	0.99	1.14	1.09
	6 months	0.96	1.00	1.37	1.31	1.07	1.33	1.10
	12 months	1.03	1.00	1.79	1.20	1.23	1.50	1.25
$/Yen	1 month	0.99	1.00	1.20	2.09	1.10	1.18	1.13
	6 months	0.97	1.00	1.85	1.58	1.12	1.17	1.00
	12 months	0.97	1.00	2.75	1.21	0.98	1.08	1.01
$/pound	1 month	0.96	1.00	1.04	2.08	1.06	1.09	1.13
	6 months	0.89	1.00	1.01	1.79	1.23	1.23	1.26
	12 months	0.86	1.00	1.15	1.83	1.26	1.18	1.25
Trade-	1 month	NA	NA	NA	NA	NA	NA	NA
weighted	6 months	NA	NA	NA	NA	NA	NA	NA
dollar	12 months	0.61	1.00	0.78	0.77	0.80	0.69	0.73

Meese and Rogoff (1983)

This finding is shocking. Random walk forecast costs almost nothing to produce, yet it out-performs the other models that are costly to produce.

> *This study compares the out-of-sample forecasting accuracy of various structural and time series exchange rate models. We find that a random walk model performs as well as any estimated model at one to twelve month horizons for the dollar/pound, dollar/mark, dollar/yen and trade weighted dollar exchange rates. The candidate structural models include the flexible-price (Frenkel-Bilson) and sticky-price (Dornbusch-Frankel) monetary models, and a sticky-price model which incorporates the current account (Hooper-Morton). The structural models perform poorly despite the fact that we based their forecasts on actual realized values of future explanatory variables.* (Meese and Rogoff, 1983, abstract)

6. Why is random walk a good approximation?

Our experience suggests that macroeconomic forecasts usually outperform naive random walk benchmark. However, it appears not the case for financial forecasts.

Meese and Rogoff (1983) showed that random walk forecast is at least as good as the forecasts of exchange rates from other time series or structural models.

Why is random walk a good approximation? It turns out that theory can be derived to yield time series that are approximately random walk.

For instance, using a infinitely lived representative agent model with rational expectation, Hall (1987)[2] was able to derive the Euler's equation relating the marginal utilities of today's consumption and tomorrow's consumption:

$$
\begin{aligned}
E\left[u'(c_{t+1}) \mid I_t\right] &= \frac{1+\delta}{1+r} u'(c_t) \\
u'(c_{t+1}) &= \frac{1+\delta}{1+r} u'(c_t) + e_{t+1}
\end{aligned}
$$

[2] Hall, Robert E. (1987): "Consumption," NBER Working Paper No. 2265.

If utility function is quadratic in consumption, marginal utility will be linear in consumption, i.e., $u'(c_t) = ac_t + b$, and hence

$$
\begin{aligned}
ac_{t+1} + b &= \frac{1+\delta}{1+r}(ac_t + b) + e_{t+1} \\
ac_{t+1} &= \frac{1+\delta}{1+r}ac_t + \left(\frac{1+\delta}{1+r} - 1\right) b + e_{t+1} \\
ac_{t+1} &= \frac{1+\delta}{1+r}ac_t + \left(\frac{\delta - r}{1+r}\right) b + e_{t+1} \\
c_{t+1} &= \left(\frac{1+\delta}{1+r}\right) c_t + \left(\frac{\delta - r}{1+r}\right) \frac{b}{a} + \frac{e_{t+1}}{a}
\end{aligned}
$$

which is AR(1) in c_t. *When the discount factor δ is close to interest rate r*, the c_t will be close to a random walk.

Why is random walk a good approximation?

Efficient market hypothesis, e.g., Fama (1965)[3], may generate similar result. With rational expectation, we have

$$E\left[P_t \mid I_{t-1}\right] = (1+r)P_{t-1}$$

$$P_t = (1+r)P_{t-1} + e_t$$

Hence, P_t follows non-stationary AR(1) with an explosive root (since $r > 0$). *When r is close to zero*, P_t looks like random walk.

[3] Fama, Eugene (1965). "The Behavior of Stock Market Prices," *Journal of Business* 38: 34–105.

Why is random walk a good approximation?

7. Statistical Comparison of Forecast Accuracy

Some researchers argue that the findings based on simple forecast comparison as in Meese and Rogoff (1983) can be due to luck. Indeed, comparison of forecast accuracy has to be based on sample information. Consequently, we must take into account of the *statistical distribution of the statistic* (i.e., sampling error) in drawing conclusion.

To fix ideas, suppose we have two sets of h-step ahead forecast errors, $e^a_{t+h,t}$ from model A and $e^b_{t+h,t}$ from model B, $t = R, ..., T - h$. Let the loss function or the criteria of comparison be denoted $L(e_{t+h,t})$. An example of such loss function is squared errors

$$L(e_{t+h,t}) = e^2_{t+h,t} \qquad t = R, ..., T - h$$

We would like to test the null hypothesis of no difference in the performance of the two models on average versus there is a difference between the two models, i.e.,

$$H_0\colon\ E(L(e^a_{t+h,t})) = E(L(e^b_{t+h,t})) \text{ versus } H_1\colon\ E(L(e^a_{t+h,t})) \neq E(L(e^b_{t+h,t}))$$

If the loss function is squared errors, we would use the sample mean squared errors for the comparison

$$MSE = \frac{1}{T-R-h+1}\sum_{t=R}^{T-h} e^2_{t+h,t}$$

Often, we rewrite the hypotheses as

$$H_0\colon\ E(L(e^a_{t+h,t})) - E(L(e^b_{t+h,t})) = 0 \text{ versus } H_1\colon\ E(L(e^a_{t+h,t})) - E(L(e^b_{t+h,t})) \neq 0$$

or

$$H_0\colon\ E(L(e^a_{t+h,t}) - L(e^b_{t+h,t})) = 0 \text{ versus } H_1\colon\ E(L(e^a_{t+h,t}) - L(e^b_{t+h,t})) \neq 0$$

Define d_t as the difference in the squared losses of the two models.

$$d_t = L(e^a_{t+h,t}) - L(e^b_{t+h,t})$$

The hypotheses can be rewritten as

$$H_0 : E(d_t) = 0 \text{ versus } H_1 : E(d_t) \neq 0$$

The time series version of Central Limit Theorem suggests that

$$\sqrt{T}(\bar{d} - \mu) \sim N(0, f)$$

where f is the covariance of $\sqrt{T}(\bar{d} - \mu)$.

The covariance f has to be estimated and the test can be based on

$$B = \frac{\bar{d}}{\sqrt{\frac{\hat{f}}{T}}} \sim N(0, 1) \qquad \hat{f} = \sum_{\tau=-M}^{M} \hat{\gamma}_d(\tau) \qquad M = T^{1/3}$$

where $\hat{\gamma}_d(\tau)$ is sample auto-covariance of d at τ displacement. *Knowing the distribution of $\bar{d}$, we can perform the usual hypothesis test.*

Performing the test using this procedure is cumbersome and prohibitive. Fortunately, like the usual test of the zero population mean, we can implement the test using a simple regression

$$d_t = \alpha + v_t$$

and test

$$H_0\text{: } \alpha = 0 \text{ versus } H_1\text{: } \alpha \neq 0$$

On the use of regression to implement various kinds of tests of forecast accuracy, see West and McCracken (1998).[4]

[4] West, Kenneth and Michael W. McCracken (1998): "Regression Based Tests of Predictive Ability," *International Economic Review*, 39: 817-840.

8. Forecast Encompassing

Is a forecast better than the competing forecasts? One criteria to answer this question is forecast encompassing:

> *A forecast is better than its competing forecast if the competing forecasts embody no useful information absent in the preferred forecasts.*

Consider the following regression

$$y_{t+h} = \beta_a y^a_{t+h,t} + \beta_b y^b_{t+h,t} + \epsilon_{t+h,t}$$

- Model A is said to forecast-encompasses model B if $(\beta_a, \beta_b) = (1, 0)$;
- Model B is said to forecast-encompasses model A if $(\beta_a, \beta_b) = (0, 1)$.
- We say neither model encompasses the other for other general values of (β_a, β_b).

To test forecast encompassing, we simply run the above regression and test the joint hypothesis $(\beta_a, \beta_b) = (1, 0)$, or $(\beta_a, \beta_b) = (0, 1)$.

What is the implication if neither model encompasses the other?

When neither model encompasses the other, we conclude that both forecasts have their merits or values in forecasting the object in question, i.e., y_{t+h}. Consequently, it is possible to combine the two forecasts to produce a better forecast.

9. Forecast combination

9.1. When forecasts are known to be unbiased.

Consider a simple weighted average of two unbiased forecasts to form a new forecast

$$y^c_{t+h,t} = \omega y^a_{t+h,t} + (1-\omega) y^b_{t+h,t}$$

It implies the forecast error of the new forecast is also a weight average of the two original forecast errors

$$e^c_{t+h,t} = \omega e^a_{t+h,t} + (1-\omega) e^b_{t+h,t}$$

Extreme cases

ω	$y^c_{t+h,t}$
0	$y^b_{t+h,t}$
1	$y^a_{t+h,t}$

It can be easily verified that the unbiasedness of $y^a_{t+h,t}$ and $y^b_{t+h,t}$ implies that $y^c_{t+h,t}$ is also unbiased.

$$\begin{aligned} E[e^c_{t+h,t} \mid \Omega_t] &= E[\omega e^a_{t+h,t} + (1-\omega)e^b_{t+h,t} \mid \Omega_t] \\ &= \omega E[e^a_{t+h,t} \mid \Omega_t] + (1-\omega)E[e^b_{t+h,t} \mid \Omega_t] \\ &= \omega \times 0 + (1-\omega) \times 0 = 0 \end{aligned}$$

By Law of Iterated Expectations,

$$E[e^c_{t+h,t} \mid \Omega_t] \Longrightarrow E[e^c_{t+h,t}] = 0$$

Given that the new forecast is unbiased, it is logical to choose the weight (i.e., ω) such that the variance of the new forecast is minimized.

$$\omega^* = \arg\min_{\omega} \sigma_c^2$$

$$\sigma_c^2 = \omega^2 \sigma_a^2 + (1-\omega)^2 \sigma_b^2 + 2\omega(1-\omega)\sigma_{ab}$$

where $\sigma_k^2 = Var(e_{t+h,t}^k)$, $k = a, b, c$, and $\sigma_{ab} = Cov(e_{t+h,t}^a, e_{t+h,t}^b)$.

Using calculus (using the first-order condition of the optimization), we find that

$$\omega^* = \frac{\sigma_b^2 - \sigma_{ab}}{\sigma_a^2 + \sigma_b^2 - 2\sigma_{ab}}$$

To obtain some intuition for this formula, it is useful to consider a special case of $\sigma_{ab} = 0$. *When $\sigma_{ab} = 0$*, the optimal weight reduces to

$$\omega^* = \frac{\sigma_b^2}{\sigma_a^2 + \sigma_b^2} = \frac{1}{(\sigma_a^2/\sigma_b^2) + 1}$$

The formula suggest that

- a *noisier* forecast will receive a *lighter* weight.
- a more accurate forecast will receive a heavier weight.

To see this, consider the extreme cases.

(1) If $y^a_{t+h,t}$ is a much noisier forecast than $y^b_{t+h,t}$, say, $(\sigma_a^2/\sigma_b^2) = \infty$. The formula suggests $\omega^* = 0$, i.e., we give all the weight to $y^b_{t+h,t}$.

(2) Conversely, if $y^a_{t+h,t}$ is a much more accurate forecast than $y^b_{t+h,t}$, say, $(\sigma_a^2/\sigma_b^2) = 0$, the formula will suggest $\omega^* = 1$, i.e., we give all the weight to $y^a_{t+h,t}$.

$$\omega^* = \frac{\sigma_b^2}{\sigma_a^2 + \sigma_b^2} = \frac{1}{(\sigma_a^2/\sigma_b^2) + 1}$$

Extreme cases

σ_a^2	σ_b^2	ω	$y_{t+h,t}^c$
∞	> 0	0	$y_{t+h,t}^b$
> 0	≈ 0	0	$y_{t+h,t}^b$
≈ 0	> 0	1	$y_{t+h,t}^a$
≥ 0	∞	1	$y_{t+h,t}^a$

Of course, in real applications, the variances and covariance have to be estimated as

$$\hat{\sigma}_a^2 = \frac{1}{T}\sum_{t=1}^{T}\left(e_{t+h,t}^{a}\right)^2$$

$$\hat{\sigma}_b^2 = \frac{1}{T}\sum_{t=1}^{T}\left(e_{t+h,t}^{b}\right)^2$$

$$\hat{\sigma}_{ab} = \frac{1}{T}\sum_{t=1}^{T} e_{t+h,t}^{a} e_{t+h,t}^{b}$$

and an estimate of the optimal weight is

$$\hat{\omega}^* = \frac{\hat{\sigma}_b^2 - \hat{\sigma}_{ab}}{\hat{\sigma}_a^2 + \hat{\sigma}_b^2 - 2\hat{\sigma}_{ab}}$$

9.2. **When forecasts need not be unbiased.**

More generally, the original two forecasts need not be unbiased and we may not want to restrict the weights to add up to unity. So, generally, we will estimate the following regression

$$y_{t+h} = \beta_0 + \beta_1 y^a_{t+h,t} + \beta_2 y^b_{t+h,t} + \epsilon_{t+h,t}$$

and use the estimated coefficients to form a new forecast

$$y^c_{t+h,t} = \hat{\beta}_0 + \hat{\beta}_1 y^a_{t+h,t} + \hat{\beta}_2 y^b_{t+h,t}$$

9.3. **Multiple period forecasts.**

We can also allow for different serial correlation structure in the errors in order to obtain a more precise estimate of the coefficients. For instance, as in the h-step-ahead forecast combinations, we may want to *allow the errors to follow an MA(h-1)*

$$
\begin{aligned}
y_{t+h} &= \beta_0 + \beta_1 y^a_{t+h,t} + \beta_2 y^b_{t+h,t} + \epsilon_{t+h,t} \\
& \epsilon_{t+h,t} \sim MA(h-1)
\end{aligned}
$$

We can also allow for the very general serial correlation structure in the errors and use the sample AIC and SIC to determine the *ARMA* orders.

$$
\begin{aligned}
y_{t+h} &= \beta_0 + \beta_1 y^a_{t+h,t} + \beta_2 y^b_{t+h,t} + \epsilon_{t+h,t} \\
& \epsilon_{t+h,t} \sim ARMA(p,q)
\end{aligned}
$$

9.4. **Time-varying weights.**

We can also allow for time-varying weights, such as,

$$y_{t+h} = (\beta_0^0 + \beta_0^1 TIME) + (\beta_1^0 + \beta_1^1 TIME) y^a_{t+h,t} + (\beta_2^0 + \beta_2^1 TIME) y^b_{t+h,t} + \epsilon_{t+h,t}$$

9.5. **Nonlinear combination.**

We can also allow for nonlinear combinations of the forecasts

$$y_{t+h} = \beta_0 + \beta_1 y^a_{t+h,t} + \beta_2 y^b_{t+h,t} + \beta_3 \left(y^a_{t+h,t}\right)^2 + \beta_4 \left(y^b_{t+h,t}\right)^2 + \epsilon_{t+h,t}$$

9.6. **Shrinkage of combining weights toward equality.**

Simple arithmetic averages of forecasts sometimes perform very well in out-of-sample forecast competitions, even relative to "optimal" combinations (say, based on regressions).

If we have forecasts from two models,

$$y^c_{t+h,t} = \frac{1}{2} y^{(1)}_{t+h,t} + \frac{1}{2} y^{(2)}_{t+h,t}$$

If we have forecasts from three models,

$$y^c_{t+h,t} = \frac{1}{3} y^{(1)}_{t+h,t} + \frac{1}{3} y^{(2)}_{t+h,t} + \frac{1}{3} y^{(3)}_{t+h,t}$$

Combination using simple arithmetic averages has the advantage that *weights are fixed* and *need not be estimated*. Consequently, the sampling error due to the estimation of weight is avoided. The drawback is that the weights are not optimal and hence a bias might be present.

10. Application: Shipping Volume

Oversea Shipping Volume on the Atlantic East Trade Lane

- To help guide fleet allocation decisions, each week OverSea makes forecasts of volume shipped over each of its major trade lanes, at horizons ranging from 1 week ahead through 16 weeks ahead.

- Two set of forecasts:
  - A quantitative forecast is produced using modern quantitative techniques (VOLQ)
  - A judgmental forecast is produced by soliciting the opinion of the sales representatives many of whom have years of valuable experience. (VOLJ)

FIGURE 10.1. *Two-week ahead* Shipping Volume Quantitative Forecast and Realization

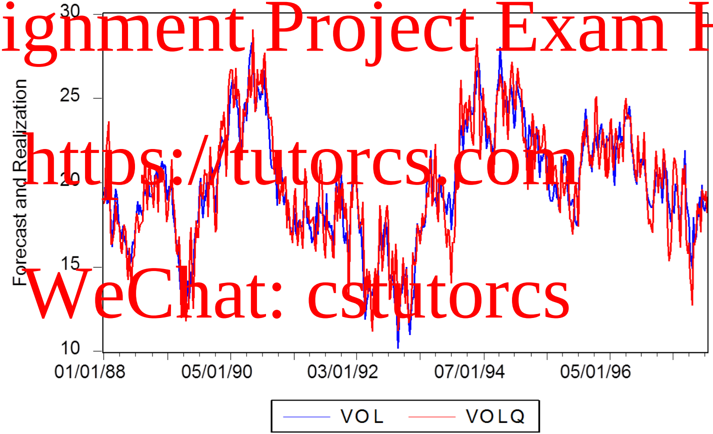


Figure 10.2. *Two-week ahead* Shipping Volume Judgmental Forecast and Realization

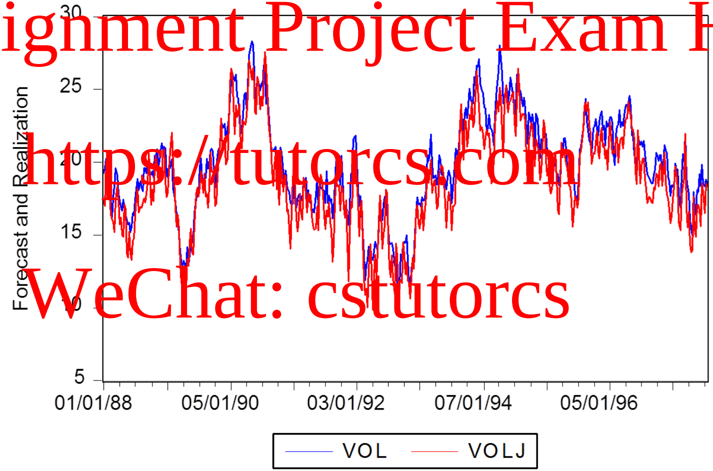


FIGURE 10.3. Quantitative forecast error

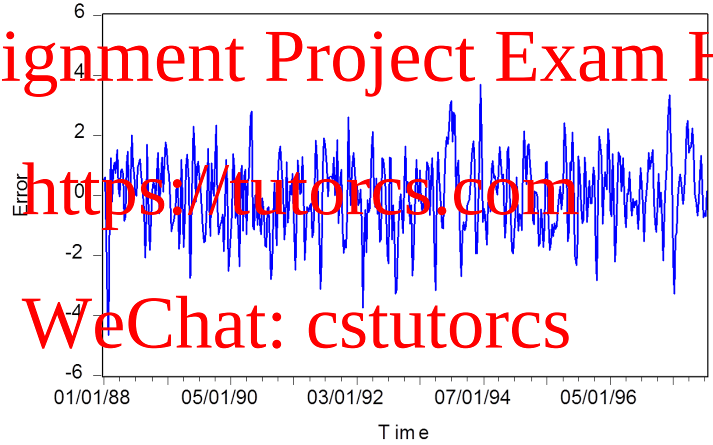


FIGURE 10.4. Judgmental forecast error

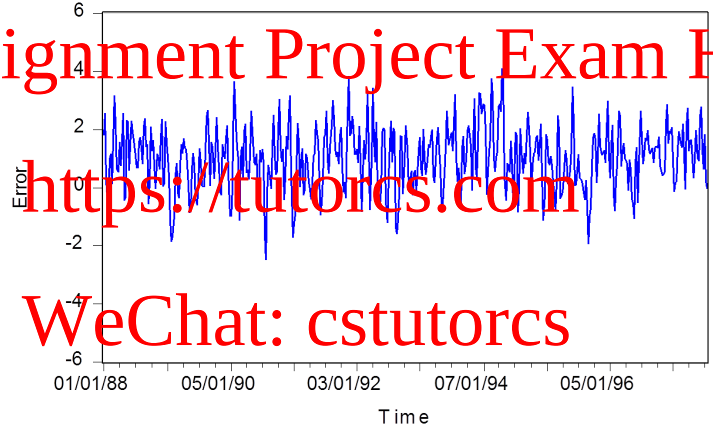


TABLE 10.1. Correlogram, quantitative forecast error

Sample: 1/01/1988 7/18/1997
Included observations: 499

	Acorr.	P. Acorr.	Std. Error	Ljung-Box	p-value
1	0.518	0.518	.045	134.62	0.000
2	0.010	−0.353	.045	134.67	0.000
3	−0.044	0.205	.045	135.65	0.000
4	−0.039	−0.172	.045	136.40	0.000
5	0.025	0.195	.045	136.73	0.000
6	0.057	−0.117	.045	138.36	0.000

TABLE 10.2. Correlogram, Judgmental forecast error

Sample: 1/01/1988 7/18/1997
Included observations: 499

	Acorr.	P. Acorr.	Std. Error	Ljung-Box	p-value
1	0.495	0.495	.045	122.90	0.000
2	−0.027	−0.360	.045	123.26	0.000
3	−0.045	0.229	.045	124.30	0.000
4	−0.056	−0.238	.045	125.87	0.000
5	−0.033	0.191	.045	126.41	0.000
6	0.087	−0.011	.045	130.22	0.000

FIGURE 10.5. Sample autocorrelation of quantitative forecast error

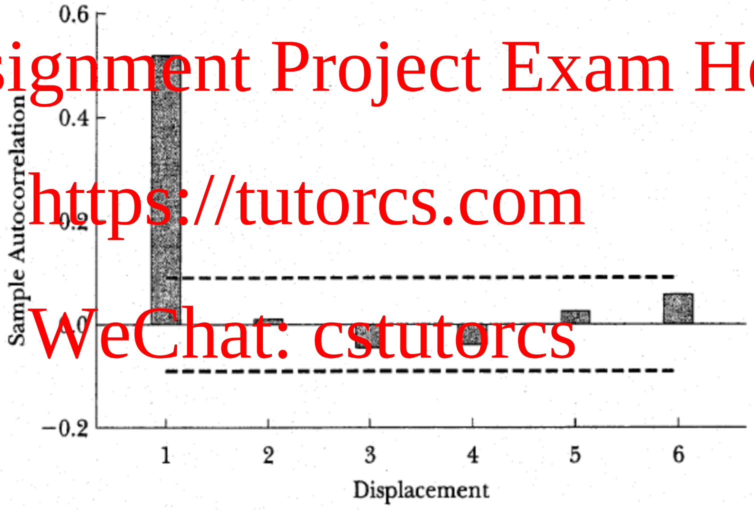


FIGURE 10.6. Sample partial autocorrelation of quantitative forecast error

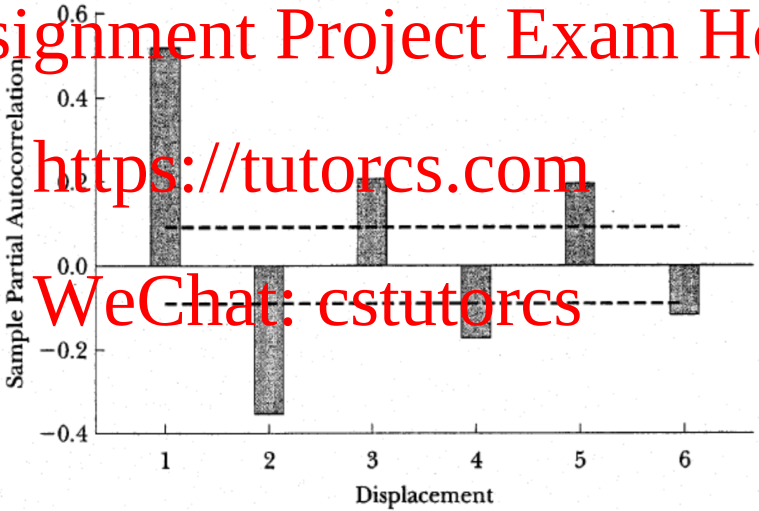


FIGURE 10.7. Sample autocorrelation of Judgmental forecast error

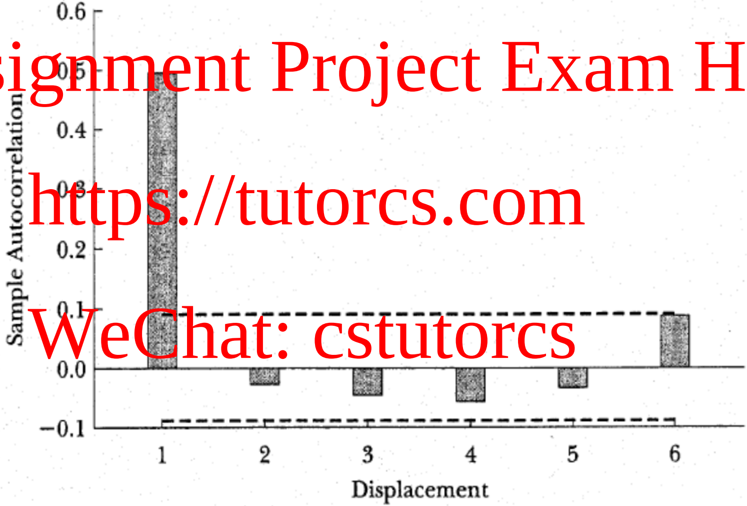


FIGURE 10.8. Sample partial autocorrelation of Judgmental forecast error

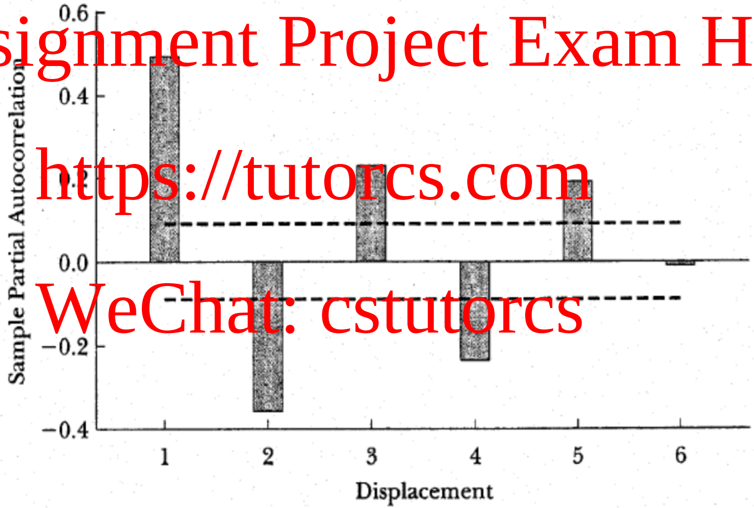


Table 10.3. Quantitative Forecast Error, Regression on Intercept, allowing MA(1) Disturbances

LS // Dependent variable is EQ.
Sample: 1/01/1988 7/18/1997
Included observations: 499
Convergence achieved after 6 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	−0.024770	0.079851	−0.310200	0.7565
MA(1)	0.935393	0.015850	59.01554	0.0000

R^2	0.468347	Mean dependent var.	−0.026572
Adjusted R^2	0.467277	SD dependent var.	1.262817
SE of regression	0.921708	Akaike info criterion	−0.159064
Sum squared resid.	422.2198	Schwarz criterion	−0.142180
Log likelihood	−666.3639	F-statistic	437.8201
Durbin-Watson stat.	1.988237	Prob(F-statistic)	0.000000
Inverted MA roots	−.94		

Table 10.4. Judgmental Forecast Error, Regression on Intercept, allowing MA(1) Disturbances

LS // Dependent variable is EJ.
Sample: 1/01/1988 7/18/1997
Included observations: 499
Convergence achieved after 7 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.026372	0.067191	15.27535	0.0000
MA(1)	0.961524	0.012470	77.10450	0.0000

R^2	0.483514	Mean dependent var.	1.023744
Adjusted R^2	0.482475	SD dependent var.	1.063681
SE of regression	0.765204	Akaike info criterion	−0.531226
Sum squared resid.	291.0118	Schwarz criterion	−0.514342
Log likelihood	−573.5094	F-statistic	465.2721
Durbin-Watson stat.	1.968750	Prob(F-statistic)	0.000000
Inverted MA roots	−.96		

TABLE 10.5. Mincer Zarnowitz Regression, Quantitative Forecast

LS // Dependent variable is VOL.
Sample: 1/01/1988 7/18/1997
Included observations: 499
Convergence achieved after 10 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.958191	0.341841	8.653696	0.0000
VOLQ	0.849559	0.016839	50.45317	0.0000
MA(1)	0.912559	0.018638	48.96181	0.0000

R^2	0.936972	Mean dependent var.	19.80609
Adjusted R^2	0.936718	SD dependent var.	3.403283
SE of regression	0.856125	Akaike info criterion	−0.304685
Sum squared resid.	363.5429	Schwarz criterion	−0.279358
Log likelihood	−619.0315	F-statistic	3686.790
Durbin-Watson stat.	1.815577	Prob(F-statistic)	0.000000

Inverted MA roots −.91

Wald test:

Null hypothesis: $C(1) = 0$ $C(2) = 1$ i.e., unforecastability

F-statistic	39.96862	Probability	0.000000
Chi-square	79.93723	Probability	0.000000

Table 10.6. Mincer Zarnowitz Regression, Judgmental Forecast

LS // Dependent variable is VOL.
Sample: 1/01/1988 7/18/1997
Included observations: 499
Convergence achieved after 11 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.592648	0.271740	9.540928	0.0000
VOLJ	0.916576	0.014058	65.20021	0.0000
MA(1)	0.949690	0.014621	64.95242	0.0000

R^2	0.952896	Mean dependent var.	19.80609
Adjusted R^2	0.952706	SD dependent var.	3.403283
SE of regression	0.740114	Akaike info criterion	−0.595907
Sum squared resid.	271.6936	Schwarz criterion	−0.570581
Log likelihood	−555.3715	F-statistic	5016.993
Durbin-Watson stat.	1.917179	Prob(F-statistic)	0.000000
Inverted MA roots	−.95		

Wald test:

Null hypothesis:	$C(1) = 0$	$C(2) = 1$	i.e., unforecastability
F-statistic	143.8323	Probability	0.000000
Chi-square	287.6647	Probability	0.000000

FIGURE 10.9. (Squared) Loss differential $(e^a_{t+h,t})^2 - (e^b_{t+h,t})^2$

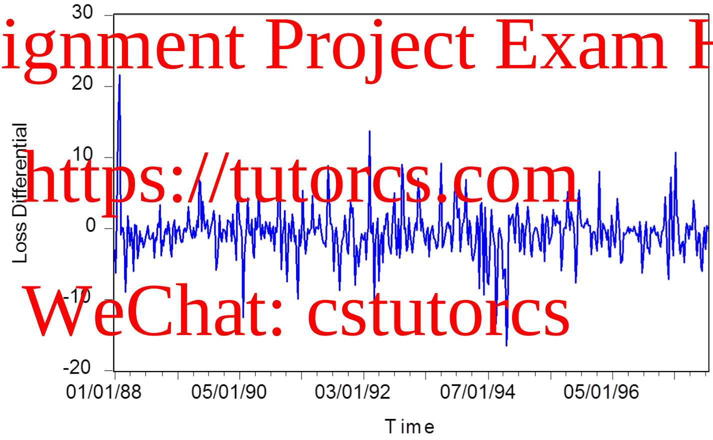


TABLE 10.7. Loss differential correlogram

Sample: 1/01/1988 7/18/1997
Included observations: 499

	Acorr.	P. Acorr.	Std. Error	Ljung-Box	p-value
1	0.357	0.357	.045	64.113	0.000
2	−0.069	−0.226	.045	66.519	0.000
3	−0.050	0.074	.045	67.761	0.000
4	−0.044	−0.080	.045	68.746	0.000
5	−0.078	−0.043	.045	71.840	0.000
6	0.017	0.070	.045	71.989	0.000

FIGURE 10.10. Sample autocorrelation of loss differential

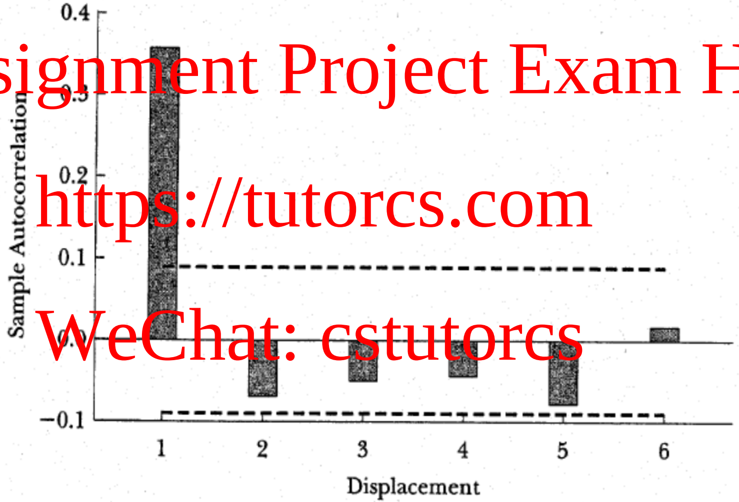


FIGURE 10.11. Sample autocorrelation of loss differential

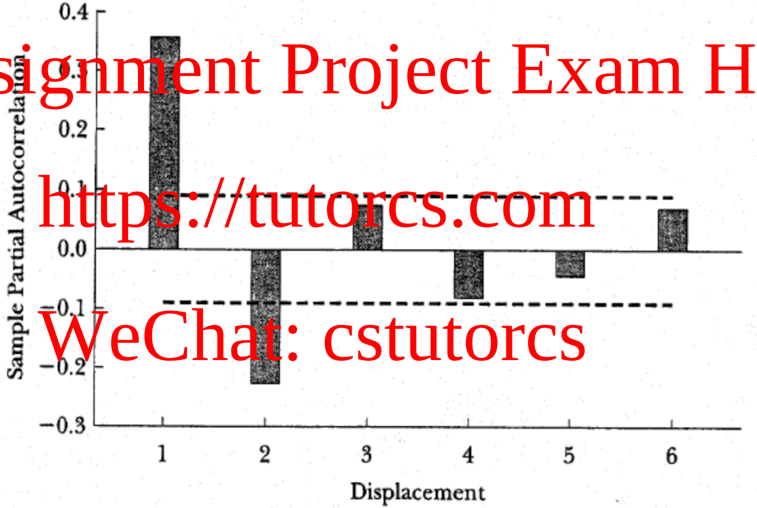


TABLE 10.8. Loss Differential, regression on intercept with $MA(1)$ disturbances

LS // Dependent variable is DD.
Sample: 1/01/1988 7/18/1997
Included observations: 499
Convergence achieved after 4 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	−0.585338	0.204737	−2.858945	0.0044
MA(1)	0.472901	0.039526	11.96433	0.0000

R^2	0.174750	Mean dependent var.	−0.584984
Adjusted R^2	0.173089	SD dependent var.	3.416190
SE of regression	3.106500	Akaike info criterion	2.270994
Sum squared resid.	4796.222	Schwarz criterion	2.287878
Log likelihood	−1272.663	F-statistic	105.2414
Durbin-Watson stat.	2.023606	Prob(F-statistic)	0.000000
Inverted MA roots	−.47		

TABLE 10.9. Shipping volume combining regression

LS // Dependent variable is VOL.
Sample: 1/01/1988 7/18/1997
Included observations: 499
Convergence achieved after 11 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.181977	0.259774	8.399524	0.0000
VOLQ	0.291577	0.038346	7.603919	0.0000
VOLJ	0.630551	0.039935	15.78944	0.0000
MA(1)	0.951107	0.014174	67.10327	0.0000

R^2	0.957823	Mean dependent var.	19.80609
Adjusted R^2	0.957567	SD dependent var.	3.403283
SE of regression	0.701049	Akaike info criterion	−0.702371
Sum squared resid.	243.2776	Schwarz criterion	−0.668603
Log likelihood	−528.8088	F-statistic	3747.077
Durbin-Watson stat.	1.925091	Prob(F-statistic)	0.000000
Inverted MA roots	−.95		