Interval Forecasts for MA and AR Models

Ka-fu WONG

University of Hong Kong

January 2, 2020

Contents

1. Unbiasedness of point forecast

Assume that y_t is a covariance stationary time series (possibly with non-zero mean).
Let $\hat{y}_{T+h,T}$ denote the h-step ahead forecast of y formed at time T, i.e.,

$$\hat{y}_{T+h,T} = E(y_{T+h} \mid y_T, y_{T-1}, \ldots)$$

The h-step ahead forecast error is then:

$$e_{T+h,T} = y_{T+h} - \hat{y}_{T+h,T}$$

Given that y_t has a Wold representation, in terms of linear combintation of past innovations, $e_{T+h,T}$ *will be a linear combination of innovations too*.

$$y_t = \epsilon_t + b_1\epsilon_{t-1} + \ldots$$

$$y_{T+h} = \epsilon_{T+h} + b_1\epsilon_{T+h-1} + \ldots + b_h\epsilon_{T+1} + b_h\epsilon_{T+h-h} + b_{h+1}\epsilon_{T-1} + \ldots$$

$$\hat{y}_{T+h,T} = E(y_{T+h} \mid y_T, y_{T-1}, \ldots) = b_h\epsilon_T + b_{h+1}\epsilon_{T-1} + \ldots$$

$$e_{T+h,T} = y_{T+h} - \hat{y}_{T+h,T} = \epsilon_{T+h} + b_1\epsilon_{T+h-1} + \ldots + b_h\epsilon_{T+1}$$

It is easy to see that $E(e_{T+h,T}) = 0$, that is, the h-step ahead forecasts we have constructed are *unbiased*.

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

2. Forecast interval (with the assumption of normality)

Let σ_h^2 denote the variance of the h-step ahead forecast error, i.e.,

$$\sigma_h^2 = E(e_{T+h,T}^2)$$

Assume that the ϵ's (and, therefore, the y's) are *normally* distributed:

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

Then $e_{T+h,T}$, which is *a linear combination of innovations*, is also normally distributed, i.e.,

$$e_{T+h,T} \sim N(0, \sigma_h^2)$$

Note

$$e_{T+h,T} = y_{T+h} - \hat{y}_{T+h,T} \quad \Leftrightarrow \quad y_{T+h} = \hat{y}_{T+h,T} + e_{T+h,T}$$

and

$$\begin{aligned} e_{T+h,T} \sim N(0, \sigma_h^2) \quad &\Leftrightarrow \quad \hat{y}_{T+h,T} + e_{T+h,T} \sim N(\hat{y}_{T+h,T}, \sigma_h^2) \\ &\Leftrightarrow \quad y_{T+h} \sim N(\hat{y}_{T+h,T}, \sigma_h^2) \end{aligned}$$

Therefore, an approximate 95% forecast interval for y_{T+h} is

$$\hat{y}_{T+h,T} \pm 1.96\sigma_h$$

More generally, $(1 - 2\alpha) \times 100$ percent forecast interval for y_{T+h} is

$$\hat{y}_{T+h,T} \pm Z_{1-\alpha}\sigma_h$$

where $Z_{1-\alpha}$ is the $(1 - \alpha) \times 100$ percentile of the $N(0, 1)$ distribution.

3. Forecast Interval of $y_t \sim WN$

If $y_t = \epsilon_t$, $\epsilon_t \sim NWN(0, \sigma^2)$, then

(1) the h-step ahead forecast, for all $h > 0$, is

$$\hat{y}_{T+h,T} = E(y_{T+h} \mid y_T, y_{T-1}, \ldots) = E(\epsilon_{T+h} \mid \epsilon_T, \epsilon_{T-1}, \ldots) = 0$$

(2) the h-step ahead forecast error is

$$e_{T+h,T} = \epsilon_{T+h}$$

(3) the variance of the h-step ahead forecast error is

$$\sigma_h^2 = E(e_{T+h,T}^2) = E(\epsilon_{T+h}^2) = \sigma^2$$

So, the 95% forecast interval is

$$\hat{y}_{T+h,T} \pm 1.96\sigma_h$$

or

$$[-1.96\sigma, 1.96\sigma]$$

Obviously, σ can be estimated by

$$\hat{\sigma} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\epsilon_t^2}$$

or

$$\hat{\sigma} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}y_t^2}$$

The case of a white noise y_t is admittedly not very interesting. However, it does serve as a very important benchmark.

4. Forecast Interval of $y_t \sim MA(1)$

Consider the MA(1) process,

$$y_t = \epsilon_t + \theta\epsilon_{t-1}$$
$$\epsilon_t \sim NWN(0, \sigma^2)$$

Unlike the white noise process, the point forecast and the forecast interval of this MA(1) process will change with horizons of forecast, h.

4.1. $h = 1$.

(1) the 1-step ahead forecast is

$$\begin{aligned}\hat{y}_{T+1,T} &= E(y_{T+1} \mid y_T, y_{T-1}, \ldots) \\ &= E(\epsilon_{T+1} + \theta\epsilon_T \mid \epsilon_T, \epsilon_{T-1}, \ldots) \\ &= \theta\epsilon_T\end{aligned}$$

(2) the 1-step ahead forecast error is

$$\begin{aligned}e_{T+1,T} &= y_{T+1} - \hat{y}_{T+1,T} \\ &= \epsilon_{T+1} + \theta\epsilon_T - \theta\epsilon_T \\ &= \epsilon_{T+1}\end{aligned}$$

(3) the variance of the 1-step ahead forecast error is

$$\sigma_1^2 = E(e_{T+1,T}^2) = E(\epsilon_{T+1}^2) = \sigma^2$$

(4) the 95% forecast interval for y_{T+1} is

$$\hat{y}_{T+1,T} \pm 1.96\sigma_1 \qquad \text{or} \qquad \theta\epsilon_T \pm 1.96\sigma$$

How to estimate σ?

Recall that for MA(1) model, $Var(y_t) = (1+\theta^2)\sigma^2$. This results suggest that we can first estimate the variance of $Var(y_t)$ and obtain an estimate of σ^2 using the relationship

$$\begin{aligned} \hat{\sigma}^2 &= \widehat{Var(y_t)}/(1+\theta^2) \\ &= \left[\frac{1}{T}\sum_{t=1}^{T} y_t^2\right]/(1+\theta^2) \end{aligned}$$

In real forecast exercises, we will have to estimate θ too. Here, to reduce the complexity of discussion, we will assume that θ is known. And, we ignore the additional uncertainty due to the estimation of σ^2.

4.2. $h = 2$.

(1) the 2-step ahead forecast is

$$\begin{aligned}\hat{y}_{T+2,T} &= E(y_{T+2} \mid y_T, y_{T-1}, \dots) \\ &= E(\epsilon_{T+2} + \theta\epsilon_{T+1} \mid \epsilon_T, \epsilon_{T-1}, \dots) \\ &= 0\end{aligned}$$

(2) the 2-step ahead forecast error is

$$\begin{aligned}e_{T+2,T} &= y_{T+2} - \hat{y}_{T+2,T} \\ &= \epsilon_{T+2} + \theta\epsilon_{T+1} - 0 \\ &= \epsilon_{T+2} + \theta\epsilon_{T+1}\end{aligned}$$

(3) the variance of the 2-step ahead forecast error is

$$
\begin{aligned}
\sigma_2^2 &= E(e_{T+2,T}^2) \\
&= E\left[(\epsilon_{T+2} + \theta\epsilon_{T+1})^2\right] \\
&= E(\epsilon_{T+2}^2 + \theta^2\epsilon_{T+1}^2 + 2\theta\epsilon_{T+2}\epsilon_{T+1}) \\
&= E(\epsilon_{T+2}^2) + \theta^2 E(\epsilon_{T+1}^2) + 2\theta E(\epsilon_{T+2}\epsilon_{T+1}) \\
&= (1+\theta^2)\sigma^2
\end{aligned}
$$

(4) the 95% forecast interval for y_{T+1} is

$$\hat{y}_{T+1,T} \pm 1.96\sigma_2$$

or

$$0 \pm 1.96\sqrt{(1+\theta^2)\sigma^2}$$

4.3. $h \geq 2$.

(1) the h-step ahead forecast is

$$\begin{aligned}\hat{y}_{T+h,T} &= E(y_{T+h} \mid y_T, y_{T-1}, ...) \\ &= E(\epsilon_{T+h} + \theta\epsilon_{T+h-1} \mid \epsilon_T, \epsilon_{T-1}, ...) \\ &= 0\end{aligned}$$

(2) the h-step ahead forecast error is

$$\begin{aligned}e_{T+h,T} &= y_{T+h} - \hat{y}_{T+h,T} \\ &= \epsilon_{T+h} + \theta\epsilon_{T+h-1} - 0 \\ &= \epsilon_{T+h} + \theta\epsilon_{T+h-1}\end{aligned}$$

(3) the variance of the h-step ahead forecast error is

$$\begin{aligned}
\sigma_h^2 &= E(e_{T+h,T}^2) \\
&= E\left[(\epsilon_{T+h} + \theta\epsilon_{T+h-1})^2\right] \\
&= E(\epsilon_{T+h}^2 + \theta^2\epsilon_{T+h-1}^2 + 2\theta\epsilon_{T+h}\epsilon_{T+h-1}) \\
&= E(\epsilon_{T+h}^2) + \theta^2 E(\epsilon_{T+h-1}^2) + 2\theta E(\epsilon_{T+h}\epsilon_{T+h-1}) \\
&= (1+\theta^2)\sigma^2
\end{aligned}$$

(4) the 95% forecast interval for y_{T+1} is

$$\hat{y}_{T+1,T} \pm 1.96\sigma_h$$

or

$$0 \pm 1.96\sqrt{(1+\theta^2)\sigma^2}$$

5. Forecast Interval of $y_t \sim MA(q)$

Consider the MA(q) process,

$$\begin{aligned} y_t &= \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \ldots + \theta_q \epsilon_{t-q} \\ \epsilon_t &\sim NWN(0, \sigma^2) \end{aligned}$$

From our earlier discusison of MA(1), we have learned that different forecast horizons may result in different point forecasts and forecast intervals. We should consider $h \leq q$ and $h > q$ *separately*.

5.1. $h = 1$.

(1) the 1-step ahead forecast is

$$\begin{aligned}
\hat{y}_{T+1,T} &= E(y_{T+h} \mid y_T, y_{T-1}, \dots) \\
&= E(\epsilon_{T+1} + \theta_1\epsilon_T + \theta_2\epsilon_{T-1} + \dots + \theta_q\epsilon_{T+1-q} \mid \epsilon_T, \epsilon_{T-1}, \dots) \\
&= \theta_1\epsilon_T + \theta_2\epsilon_{T-1} + \dots + \theta_q\epsilon_{T+1-q}
\end{aligned}$$

(2) the 1-step ahead forecast error is

$$\begin{aligned}
e_{T+1,T} &= y_{T+1} - \hat{y}_{T+1,T} \\
&= \epsilon_{T+1} + \theta_1\epsilon_T + \theta_2\epsilon_{T-1} + \dots + \theta_q\epsilon_{T+1-q} \\
&\quad -[\theta_1\epsilon_T + \theta_2\epsilon_{T-1} + \dots + \theta_q\epsilon_{T+1-q}] \\
&= \epsilon_{T+1}
\end{aligned}$$

(3) the variance of the 1-step ahead forecast error is

$$\sigma_1^2 = E(e_{T+1,T}^2) = E(\epsilon_{T+1}^2) = \sigma^2$$

(4) the 95% forecast interval for y_{T+1} is

$$\hat{y}_{T+1,T} \pm 1.96\sigma_1$$

or

$$(\theta_1\epsilon_T + \theta_2\epsilon_{T-1} + ... + \theta_q\epsilon_{T+1-q}) \pm 1.96\sigma$$

How to estimate σ?

Recall that for MA(1) model, $Var(y_t) = (1 + \theta_1^2 + \theta_2^2 + ... + \theta_q^2)\sigma^2$. This results suggest that we can first estimate the variance of $Var(y_t)$ and obtain an estimate of σ^2 using the relationship

$$\begin{aligned}\hat{\sigma}^2 &= \widehat{Var(y_t)}/(1 + \theta_1^2 + \theta_2^2 + ... + \theta_q^2) \\ &= \left[\frac{1}{T}\sum_{t=1}^{T} y_t^2\right]/(1 + \theta_1^2 + \theta_2^2 + ... + \theta_q^2)\end{aligned}$$

Again, in actual forecast exercises, we will have to estimate θ too. Here, to reduce the complexity of discussion, we will assume that θ is known. And, we ignore the additional uncertainty due to the estimation of σ^2.

5.2. $h = 2$.

(1) the 2-step ahead forecast is

$$\begin{aligned}\hat{y}_{T+2,T} &= E(y_{T+2} \mid y_T, y_{T-1}, \dots) \\ &= E(\epsilon_{T+2} + \theta_1\epsilon_{T+1} + \theta_2\epsilon_T + \dots + \theta_q\epsilon_{T+2-q} \mid \epsilon_T, \epsilon_{T-1}, \dots) \\ &= \theta_2\epsilon_T + \theta_3\epsilon_{T-1} + \dots + \theta_q\epsilon_{T+2-q}\end{aligned}$$

(2) the 2-step ahead forecast error is

$$\begin{aligned}e_{T+2,T} &= y_{T+2} - \hat{y}_{T+2,T} \\ &= \epsilon_{T+2} + \theta_1\epsilon_{T+1} + \theta_2\epsilon_T + \dots + \theta_q\epsilon_{T+2-q} \\ &\quad -(\theta_2\epsilon_T + \theta_3\epsilon_{T-1} + \dots + \theta_q\epsilon_{T+2-q}) \\ &= \epsilon_{T+2} + \theta_1\epsilon_{T+1}\end{aligned}$$

(3) the variance of the 2-step ahead forecast error is

$$\begin{aligned}
\sigma_2^2 &= E(e_{T+2,T}^2) \\
&= E\left[(\epsilon_{T+2} + \theta_1 \epsilon_{T+1})^2\right] \\
&= E(\epsilon_{T+2}^2 + \theta_1^2 \epsilon_{T+1}^2 + 2\theta_1 \epsilon_{T+2}\epsilon_{T+1}) \\
&= E(\epsilon_{T+2}^2) + \theta_1^2 E(\epsilon_{T+1}^2) + 2\theta_1 E(\epsilon_{T+2}\epsilon_{T+1}) \\
&= (1+\theta_1^2)\sigma^2
\end{aligned}$$

(4) the 95% forecast interval for y_{T+2} is

$$\hat{y}_{T+2,T} \pm 1.96\sigma_2$$

or

$$(\theta_2 \epsilon_T + \theta_3 \epsilon_{T-1} + \ldots + \theta_q \epsilon_{T+2-q}) \pm 1.96\sqrt{(1+\theta_1^2)\sigma^2}$$

5.3. $h \le q$.

(1) the h-step ahead forecast is

$$
\begin{aligned}
\hat{y}_{T+h,T} &= E(y_{T+h} \mid y_T, y_{T-1}, \ldots) \\
&= E(\epsilon_{T+h} + \theta_1 \epsilon_{T+h-1} + \theta_2 \epsilon_{T+h-2} + \ldots + \theta_q \epsilon_{T+h-q} \mid \epsilon_T, \epsilon_{T-1}, \ldots) \\
&= \theta_h \epsilon_T + \theta_3 \epsilon_{T-1} + \ldots + \theta_q \epsilon_{T+h-q}
\end{aligned}
$$

(2) the h-step ahead forecast error is

$$
\begin{aligned}
e_{T+h,T} &= y_{T+h} - \hat{y}_{T+h,T} \\
&= \epsilon_{T+h} + \theta_1 \epsilon_{T+h-1} + \theta_2 \epsilon_{T+h-2} + \ldots + \theta_q \epsilon_{T+h-q} \\
&\quad -(\theta_h \epsilon_T + \theta_3 \epsilon_{T-1} + \ldots + \theta_q \epsilon_{T+h-q}) \\
&= \epsilon_{T+h} + \theta_1 \epsilon_{T+h-1} + \ldots + \theta_{h-1} \epsilon_{T+1}
\end{aligned}
$$

(3) the variance of the h-step ahead forecast error is

$$\begin{aligned}
\sigma_h^2 &= E(e_{T+h,T}^2) \\
&= E\left[(\epsilon_{T+h} + \theta_1 \epsilon_{T+h-1} + \ldots + \theta_{h-1}\epsilon_{T+1})^2\right] \\
&= (1 + \theta_1^2 + \ldots + \theta_{h-1}^2)\sigma^2
\end{aligned}$$

(4) the 95% forecast interval for y_{T+h} is

$$\hat{y}_{T+h,T} \pm 1.96\sigma_h$$

or

$$(\theta_h \epsilon_T + \theta_3 \epsilon_{T-1} + \ldots + \theta_q \epsilon_{T+h-q}) \pm 1.96\sqrt{(1 + \theta_1^2 + \ldots + \theta_{h-1}^2)\sigma^2}$$

5.4. $h > q$.

(1) the h-step ahead forecast is

$$\begin{aligned}\hat{y}_{T+h,T} &= E(y_{T+h} \mid y_T, y_{T-1}, \ldots) \\ &= E(\epsilon_{T+h} + \theta_1\epsilon_{T+h-1} + \theta_2\epsilon_{T+h-2} + \ldots + \theta_q\epsilon_{T+h-q} \mid \epsilon_T, \epsilon_{T-1}, \ldots) \\ &= 0\end{aligned}$$

(2) the h-step ahead forecast error is

$$\begin{aligned}e_{T+h,T} &= y_{T+h} - \hat{y}_{T+h,T} \\ &= \epsilon_{T+h} + \theta_1\epsilon_{T+h-1} + \theta_2\epsilon_{T+h-2} + \ldots + \theta_q\epsilon_{T+h-q} - 0 \\ &= \epsilon_{T+h} + \theta_1\epsilon_{T+h-1} + \theta_2\epsilon_{T+h-2} + \ldots + \theta_q\epsilon_{T+h-q}\end{aligned}$$

(3) the variance of the h-step ahead forecast error is

$$\begin{aligned}
\sigma_h^2 &= E(e_{T+h,T}^2) \\
&= E\left[(\epsilon_{T+h} + \theta_1\epsilon_{T+h-1} + \theta_2\epsilon_{T+h-2} + \ldots + \theta_q\epsilon_{T+h-q})^2\right] \\
&= (1+\theta_1^2 + \ldots + \theta_q^2)\sigma^2
\end{aligned}$$

(4) the 95% forecast interval for y_{T+h} is

$$\hat{y}_{T+h,T} \pm 1.96\sigma_h$$

or

$$0 \pm 1.96\sqrt{(1+\theta_1^2 + \ldots + \theta_q^2)\sigma^2}$$

6. Forecast Interval of $y_t \sim AR(1)$

Consider the AR(1) process,

$$
\begin{aligned}
y_t &= \phi y_{t-1} + \epsilon_t \\
\epsilon_t &\sim WN(0, \sigma^2)
\end{aligned}
$$

6.1. $h = 1$.

(1) the 1-step ahead forecast is

$$
\begin{aligned}
\hat{y}_{T+1,T} &= E(y_{T+1} \mid y_T, y_{T-1}, \ldots) \\
&= E(\phi y_T + \epsilon_{T+1} \mid y_T, y_{T-1}, \ldots) \\
&= \phi y_T
\end{aligned}
$$

(2) the 1-step ahead forecast error is

$$
\begin{aligned}
e_{T+1,T} &= y_{T+1} - \hat{y}_{T+1,T} \\
&= \phi y_T + \epsilon_{T+1} - \phi y_T \\
&= \epsilon_{T+1}
\end{aligned}
$$

(3) the variance of the 1-step ahead forecast error is

$$\sigma_1^2 = E(e_{T+1,T}^2) = E(\epsilon_{T+1}^2) = \sigma^2$$

(4) the 95% forecast interval for y_{T+1} is

$$\hat{y}_{T+1,T} \pm 1.96\sigma_1$$

or

$$\phi y_T \pm 1.96\sigma$$

How to estimate σ?

Recall that for AR(1) model, $Var(y_t) = \phi^2 Var(y_{t-1}) + \sigma^2$ and hence $Var(y_t) = \phi^2 Var(y_{t-1}) + \sigma^2$. This results suggest that we can first estimate the variance of $Var(y_t)$ and obtain an estimate of σ^2 using the relationship

$$\begin{aligned} \hat{\sigma}^2 &= (1-\phi^2)\widehat{Var(y_t)} \\ &= (1-\phi^2)\left[\frac{1}{T}\sum_{t=1}^{T} y_t^2\right] \end{aligned}$$

Equivalently, we can run an regression to estimate the AR coefficients and hence the residuals. Then the standard error of the regression or the mean squared residuals will be an estimate of σ^2.

In real forecast exercises, we will have to estimate ϕ too. Here, to reduce the complexity of discussion, we will assume that ϕ is known. And, we ignore the additional uncertainty due to the estimation of σ^2.

6.2. $h = 2$.

(1) the 2-step ahead forecast is

$$\begin{aligned}
\hat{y}_{T+2,T} &= E(y_{T+2} \mid y_T, y_{T-1}, \ldots) \\
&= E(\phi y_{T+1} + \epsilon_{T+2} \mid y_T, y_{T-1}, \ldots) \\
&= E[\phi(\phi y_T + \epsilon_{T+1}) + \epsilon_{T+2} \mid y_T, y_{T-1}, \ldots] \\
&= E[\phi^2 y_T + \phi \epsilon_{T+1} + \epsilon_{T+2} \mid y_T, y_{T-1}, \ldots] \\
&= \phi^2 y_T
\end{aligned}$$

(2) the 2-step ahead forecast error is

$$\begin{aligned}
e_{T+2,T} &= y_{T+2} - \hat{y}_{T+2,T} \\
&= \phi^2 y_T + \phi \epsilon_{T+1} + \epsilon_{T+2} - \phi^2 y_T \\
&= \phi \epsilon_{T+1} + \epsilon_{T+2}
\end{aligned}$$

(3) the variance of the 2-step ahead forecast error is

$$
\begin{aligned}
\sigma_2^2 &= E(e^2_{T+2,T}) \\
&= E\left[(\phi\epsilon_{T+1} + \epsilon_{T+2})^2\right] \\
&= E(\epsilon^2_{T+2} + \phi^2\epsilon^2_{T+1} + 2\phi\epsilon_{T+2}\epsilon_{T+1}) \\
&= E(\epsilon^2_{T+2}) + \phi^2 E(\epsilon^2_{T+1}) + 2\phi E(\epsilon_{T+2}\epsilon_{T+1}) \\
&= (1+\phi^2)\sigma^2
\end{aligned}
$$

(4) the 95% forecast interval for y_{T+2} is

$$\hat{y}_{T+2,T} \pm 1.96\sigma_2$$

or

$$\phi^2 y_T \pm 1.96\sqrt{(1+\phi^2)\sigma^2}$$

6.3. **Any** $h \geq 1$.

(1) the h-step ahead forecast is

$$\begin{aligned}
\hat{y}_{T+h,T} &= E(y_{T+h} \mid y_T, y_{T-1}, \ldots) \\
&= E(\phi^h y_T + \phi^{h-1}\epsilon_{T+1} + \phi^{h-2}\epsilon_{T+2} + \ldots + \epsilon_{T+h} \mid y_T, y_{T-1}, \ldots) \\
&= \phi^h y_T
\end{aligned}$$

(2) the h-step ahead forecast error is

$$\begin{aligned}
e_{T+h,T} &= y_{T+h} - \hat{y}_{T+h,T} \\
&= \phi^{h-1}\epsilon_{T+1} + \phi^{h-2}\epsilon_{T+2} + \ldots + \phi\epsilon_{T+h-1} + \epsilon_{T+h}
\end{aligned}$$

(3) the variance of the h-step ahead forecast error is

$$\begin{aligned}\sigma_h^2 &= E(e_{T+h,T}^2) \\ &= E\left[(\phi^{h-1}\epsilon_{T+1} + \phi^{h-2}\epsilon_{T+2} + \ldots + \phi\epsilon_{T+h-1} + \epsilon_{T+h})^2\right] \\ &= \left[\phi^{2(h-1)} + \phi^{2(h-2)} + \ldots + \phi^2 + 1\right]\sigma^2\end{aligned}$$

(4) the 95% forecast interval for y_{T+h} is

$$\hat{y}_{T+h,T} \pm 1.96\sigma_h$$

or

$$\phi^h y_T \pm 1.96\sqrt{\left[\phi^{2(h-1)} + \phi^{2(h-2)} + \ldots + \phi^2 + 1\right]\sigma^2}$$

7. A comparison of forecast intervals

h	MA(1)	AR(1)
1	$\theta\epsilon_T \pm 1.96\sigma$	$\phi y_T \pm 1.96\sigma$
2	$0 \pm 1.96\sqrt{(1+\theta^2)\sigma^2}$	$\phi^2 y_T \pm 1.96\sqrt{(1+\phi^2)\sigma^2}$
3	$0 \pm 1.96\sqrt{(1+\theta^2)\sigma^2}$	$\phi^3 y_T \pm 1.96\sqrt{(1+\phi^2+\phi^4)\sigma^2}$
$h \geq 2$	$0 \pm 1.96\sqrt{(1+\theta^2)\sigma^2}$	$\phi^h y_T \pm 1.96\sqrt{(1+\phi^2+\ldots+\phi^{2(h-1)})\sigma^2}$

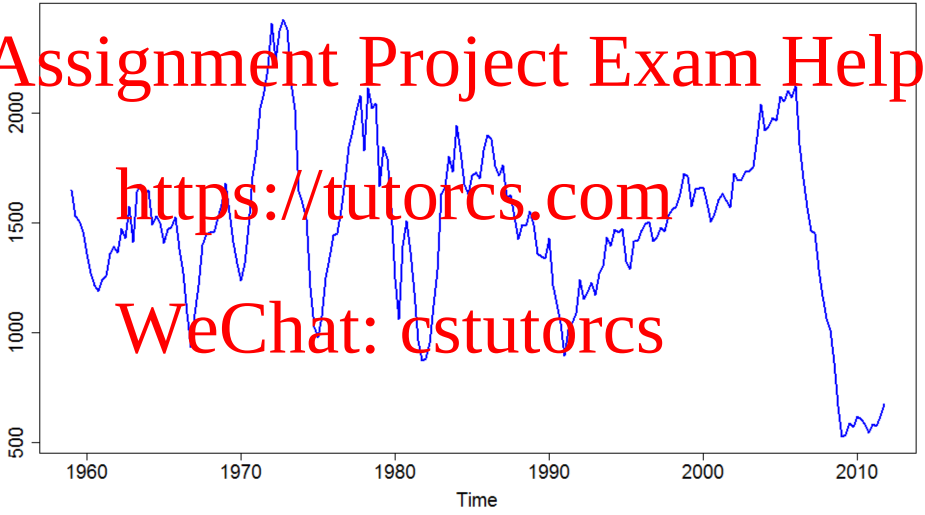
- The forecast interval for MA(1) process increases but stabilized after one period.
- The forecast interval for AR(1) continues to increase with forecast horizons. The forecast interval for AR(1) continues to increase with forecast horizon means that there is an increasing uncertainty in our forecast.

8. Example: US Housing Starts (Quarterly Seasonally Adjusted)

- Quarterly seasonally adjusted data from 1959:01 to 2013:04 (a total of 220 observations), downloaded from FRED[1].
- 8 observations were saved for out-of-sample comparison.
- That is, model selection and estimation are based on data from 1959:01 to 2011:04 (a total of 212 observations).
- We use seasonally adjusted data in this example because we want to illustrate with a model *without modelling seasonality*.

[1] Federal Reserve Economic Data (FRED, https://fred.stlouisfed.org/)

Estimation Sample



AIC

	q = 0	q = 1	q = 2	q = 3	q = 4	q = 5	q = 6	q = 7	q=8
p = 0	3128	2918	2780	2713	2684	2663	2648	2630	2631
p = 1	2633	2626	2614	2616	2616	2614	2616	2617	2618
p = 2	2621	2612	2616	2617	2611	2613	2613	2614	2616
p = 3	2613	2614	2615	2615	2617	2611	2613	2611	2613
p = 4	2615	2616	2612	2606	2608	2613	2614	2611	2610
p = 5	2614	2613	2614	2615	2610	2614	2615	2609	2609
p = 6	2610	2611	2613	2613	2612	2612	2608	2609	2613
p = 7	2612	2613	2612	2614	2613	2612	2610	2609	2611
p=8	2613	2615	2611	2613	2615	2616	2614	2615	2617

SIC

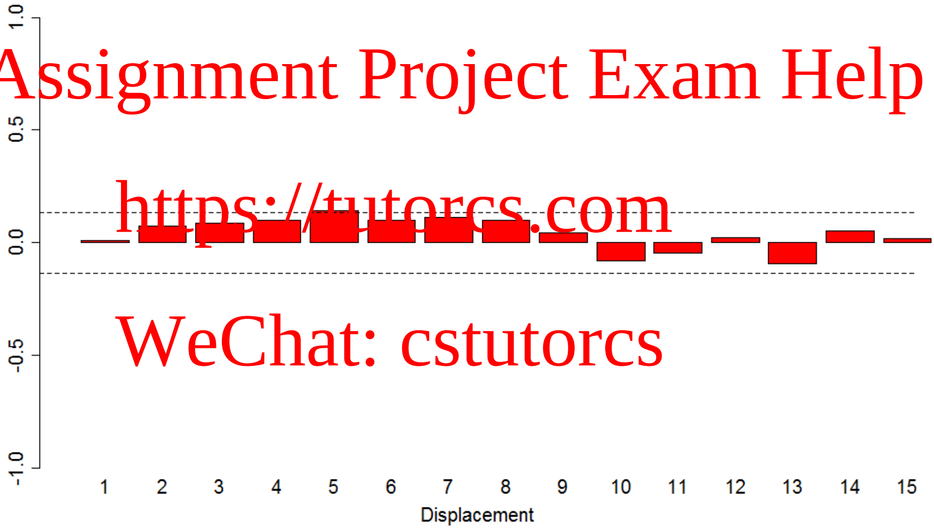
	q = 0	q = 1	q = 2	q = 3	q = 4	q = 5	q = 6	q = 7	q=8
p = 0	3135	2928	2794	2730	2704	2687	2675	2660	2665
p = 1	2643	2639	2631	2636	2639	2641	2646	2650	2655
p = 2	2634	2629	2636	2641	2638	2643	2646	2651	2657
p = 3	2630	2635	2638	2642	2647	2644	2649	2652	2657
p = 4	2635	2640	2638	2637	2641	2649	2655	2655	2657
p = 5	2637	2639	2645	2649	2647	2655	2658	2656	2659
p = 6	2637	2641	2646	2650	2652	2655	2655	2659	2667
p = 7	2642	2646	2649	2654	2657	2659	2660	2662	2668
p=8	2646	2652	2652	2657	2662	2666	2667	2673	2678

p-value of Box Test

	q = 0	q = 1	q = 2	q = 3	q = 4	q = 5	q = 6	q = 7	q=8
p = 0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.105	0.112
p = 1	0.000	0.004	0.360	0.370	0.465	0.879	0.867	0.920	0.920
p = 2	0.020	0.463	0.368	0.369	0.896	0.916	0.945	0.920	0.926
p = 3	0.396	0.398	0.498	0.451	0.473	0.967	0.963	0.998	0.999
p = 4	0.398	0.399	0.642	0.968	0.984	0.958	0.955	0.999	1.000
p = 5	0.467	0.729	0.773	0.489	0.982	0.953	0.905	0.997	0.983
p = 6	0.966	0.994	0.995	0.917	0.986	0.995	0.999	0.987	1.000
p = 7	0.985	0.995	0.996	0.977	0.984	0.863	0.999	1.000	0.997
p=8	0.992	0.995	1.000	1.000	1.000	0.899	1.000	0.970	1.000

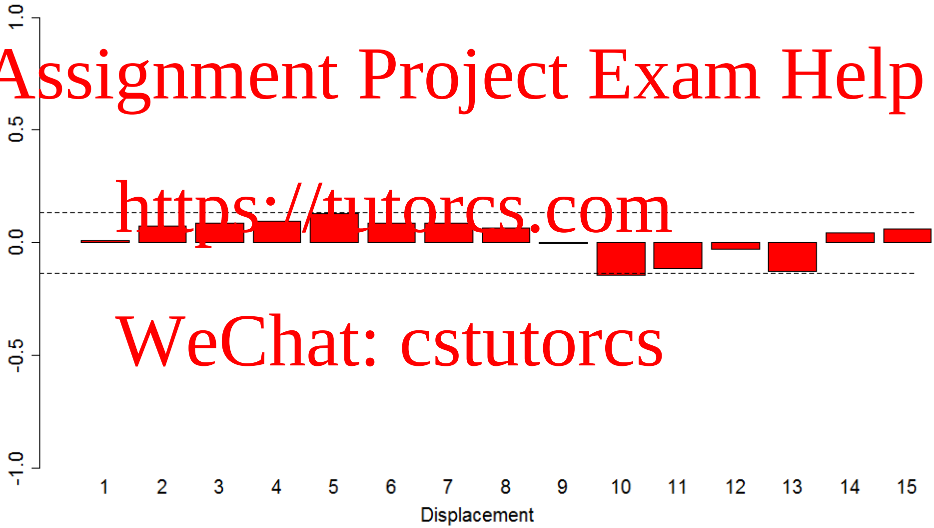
- The information (AIC, SIC and the white noise test) suggests *ARMA(2,1)*. But we also want to consider the pure *AR(3)* and pure *MA(8)* models.

ACF of ARMA(0,8)



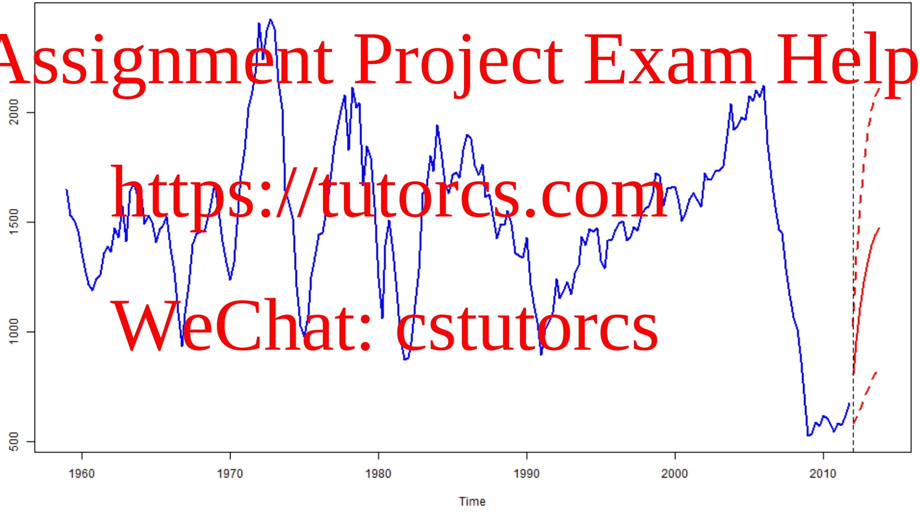
Example: US Housing Starts (Quarterly Seasonally Adjusted)

PACF of ARMA(0,8)



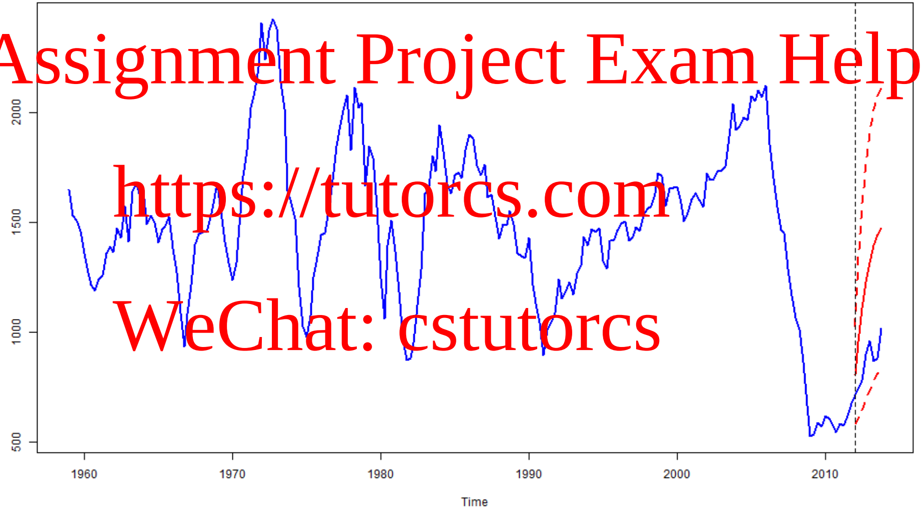
Example: US Housing Starts (Quarterly Seasonally Adjusted)

Forecast of ARMA(0,8)



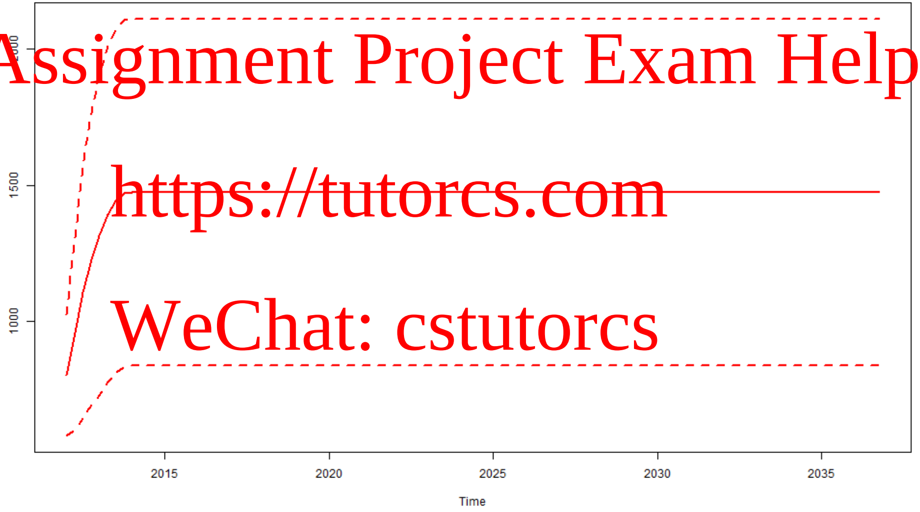
Example: US Housing Starts (Quarterly Seasonally Adjusted)

Forecast Comparison of ARMA(0,8)

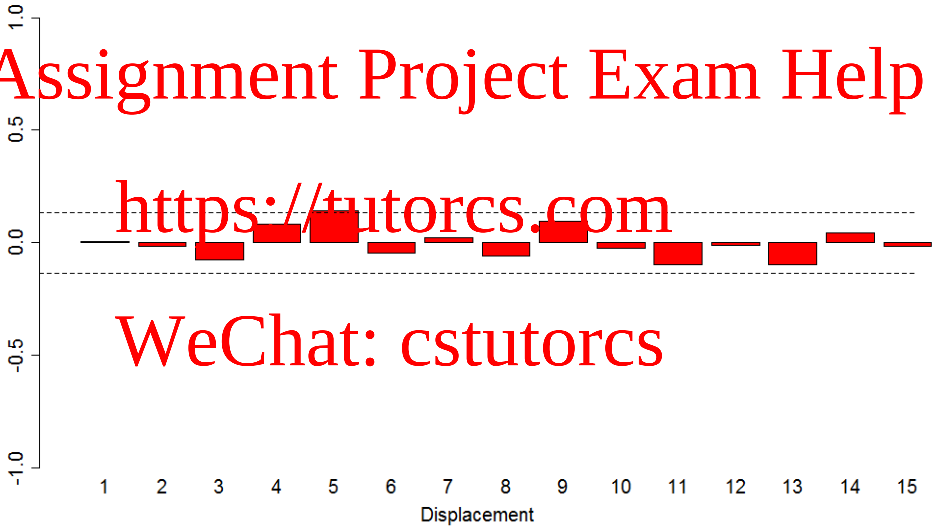


Example: US Housing Starts (Quarterly Seasonally Adjusted)

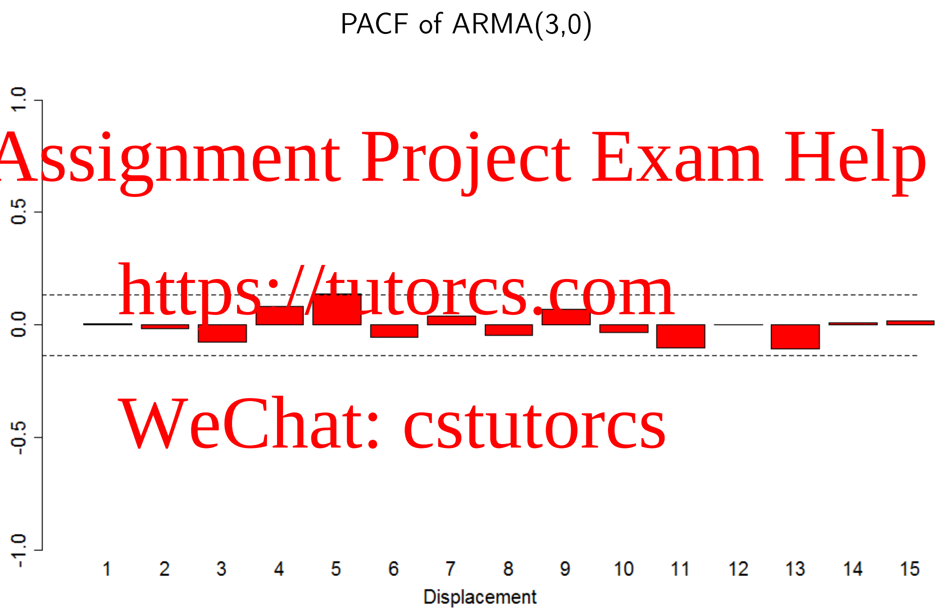
Long-horizon Forecast of ARMA(0,8)



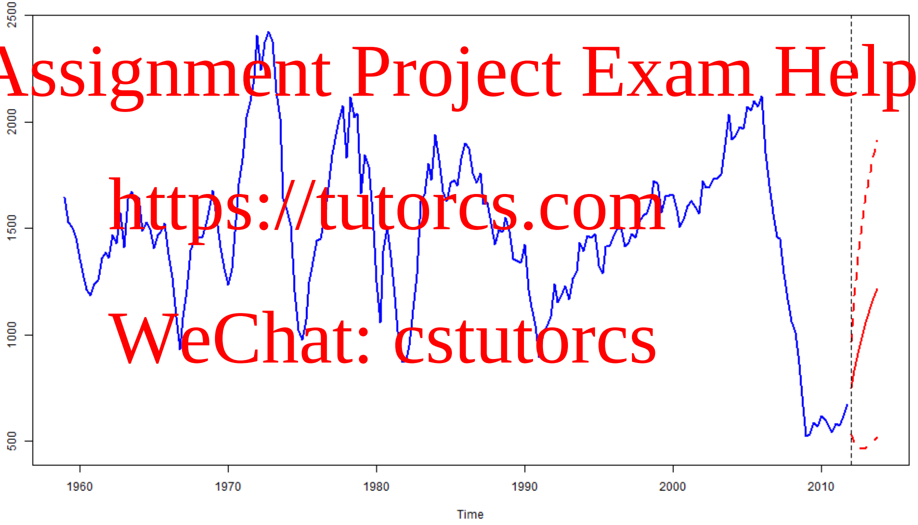
ACF of ARMA(3,0)



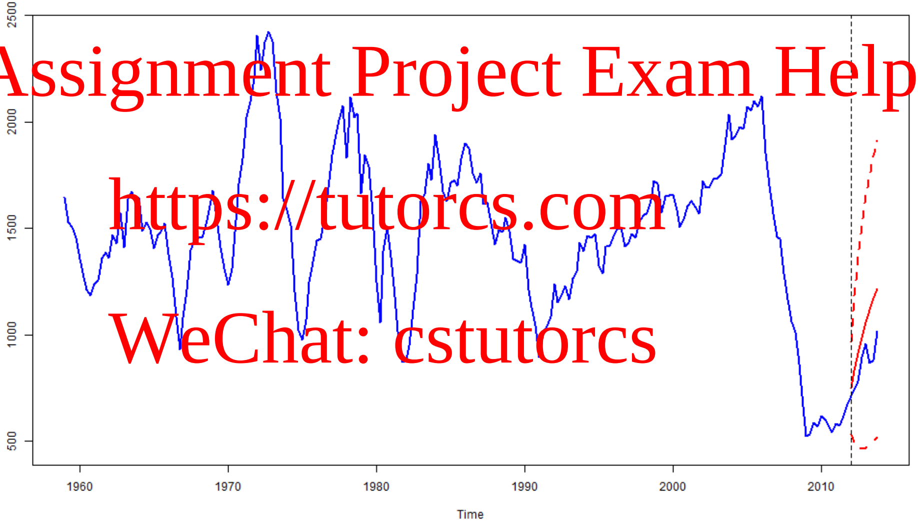
Example: US Housing Starts (Quarterly Seasonally Adjusted)


PACF of ARMA(3,0)
1.0
0.5
0.0
-0.5
-1.0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
Displacement

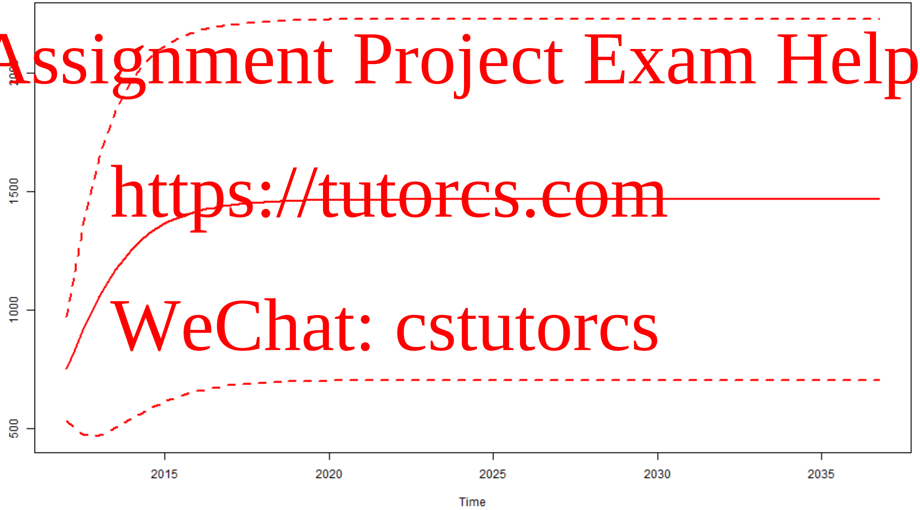
Forecast of ARMA(3,0)



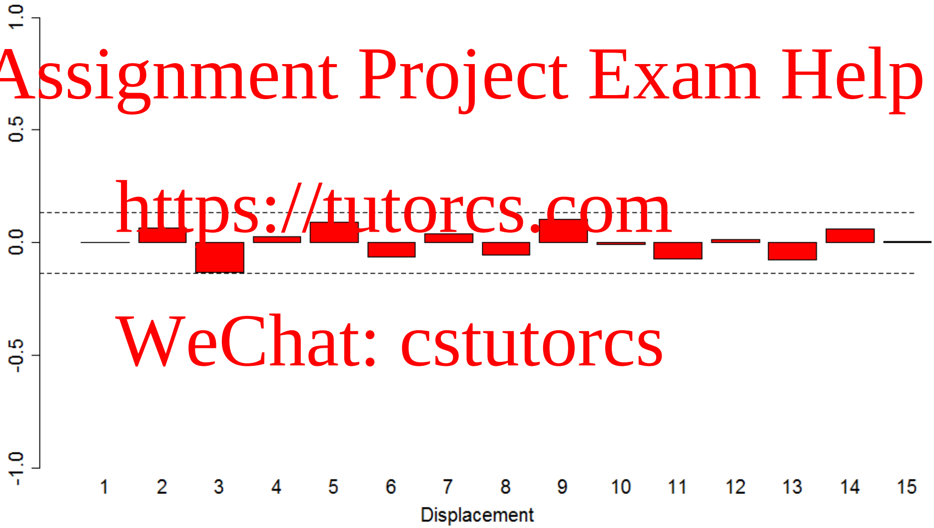
Forecast Comparison of ARMA(3,0)



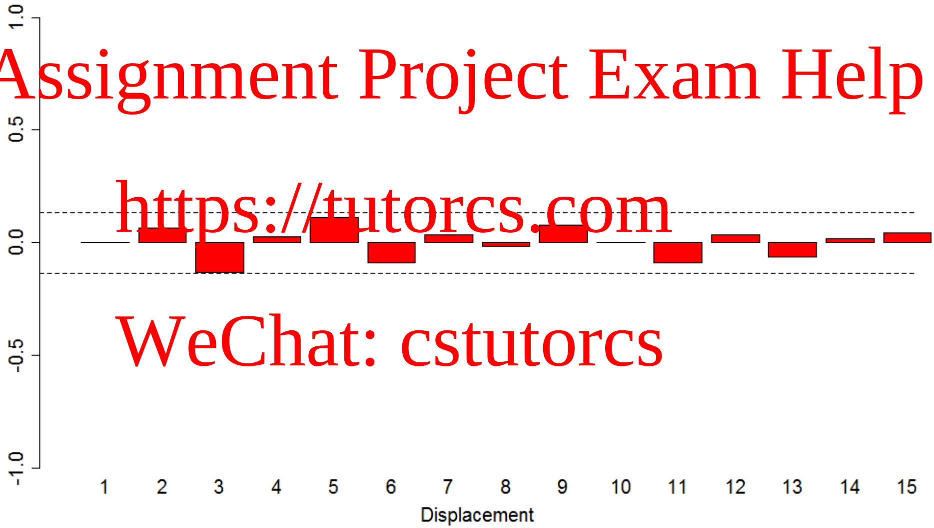
Long-horizon Forecast of ARMA(3,0)



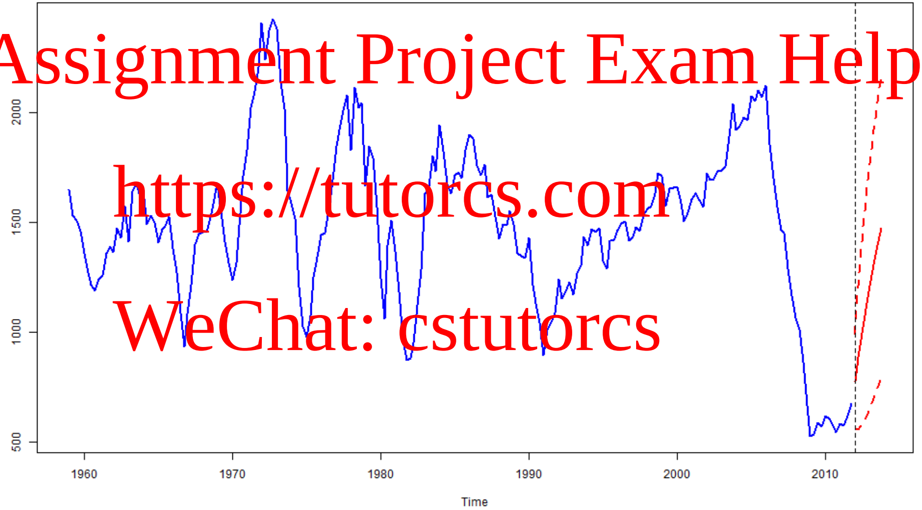
ACF of ARMA(2,1)



PACF of ARMA(2,1)

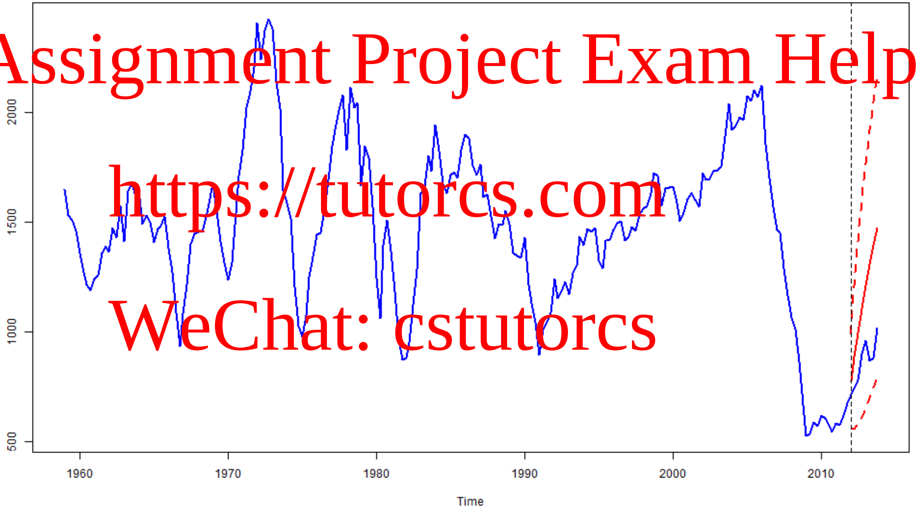


Forecast of ARMA(2,1)



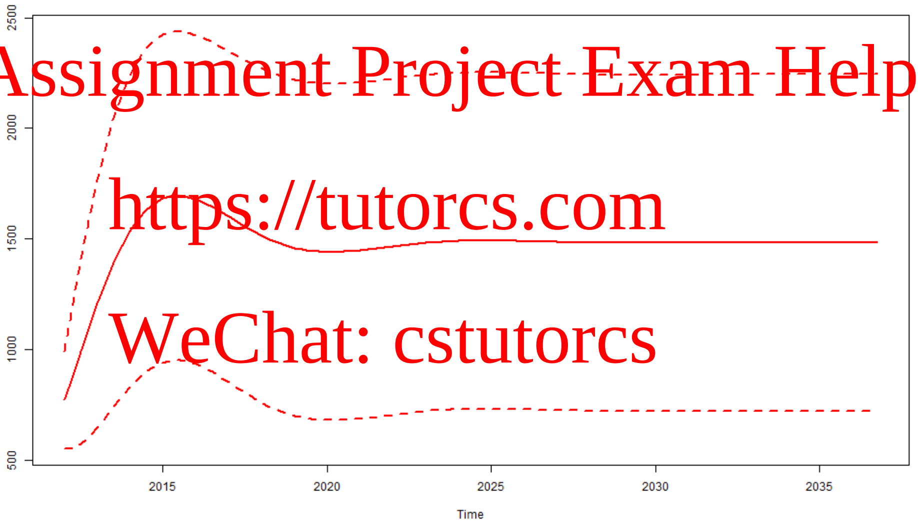
Example: US Housing Starts (Quarterly Seasonally Adjusted)

Forecast Comparison of ARMA(2,1)



Example: US Housing Starts (Quarterly Seasonally Adjusted)

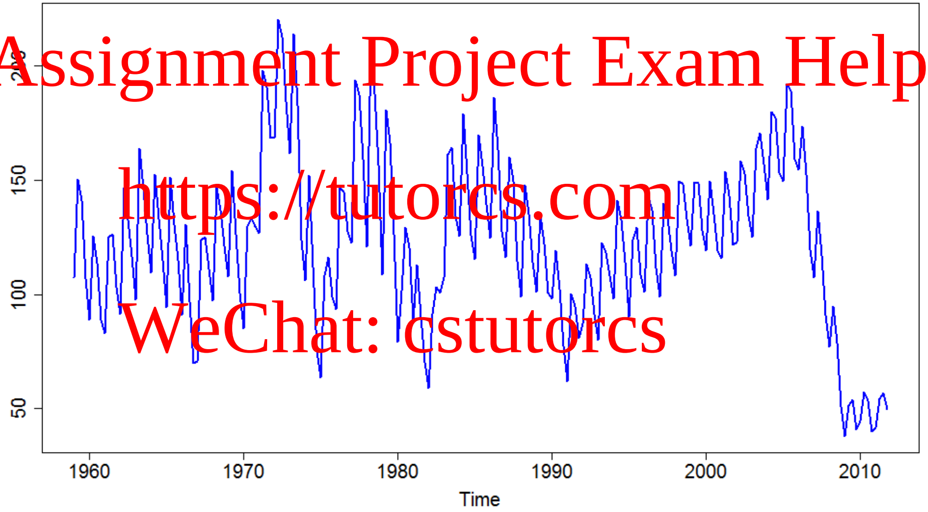
Long-horizon Forecast of ARMA(2,1)



9. Example: US Housing Starts (Quarterly Non-seasonally Adjusted)

- Quarterly non-seasonally adjusted data from 1959:01 to 2013:04 (a total of 220 observations), downloaded from FRED.
- 8 observations were saved for out-of-sample comparison.
- That is, model selection and estimation are based on data from 1959:01 to 2011:04 (a total of 212 observations).

Estimation Sample



AIC

	q = 0	q = 1	q = 2	q = 3	q = 4	q = 5	q = 6	q = 7	q=8
p = 0	2130	2031	1921	1919	1868	1842	1817	1814	1797
p = 1	1987	1988	1917	1920	1840	1843	1815	1816	1777
p = 2	1989	1989	1917	1897	1836	1832	1748	1739	1701
p = 3	1935	1923	1792	1773	1741	1741	1709	1702	1700
p = 4	1880	1772	1742	1662	1661	1659	1661	1650	1652
p = 5	1676	1675	1671	1660	1651	1651	1653	1652	1654
p = 6	1675	1669	1672	1654	1651	1653	1655	1653	1651
p = 7	1674	1670	1653	1655	1653	1654	1652	1656	1654
p=8	1676	1664	1660	1644	1646	1647	1647	1648	—

SIC

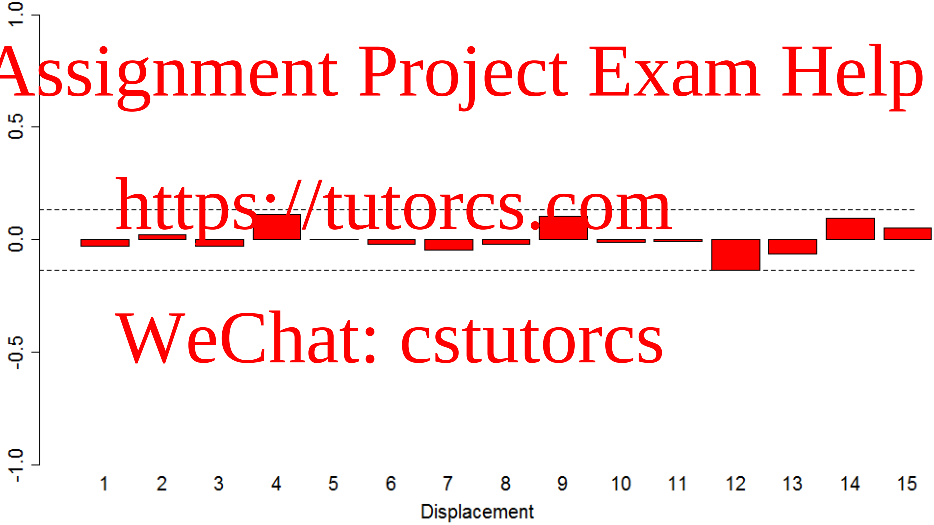
	q = 0	q = 1	q = 2	q = 3	q = 4	q = 5	q = 6	q = 7	q=8
p = 0	2137	2041	1934	1936	1888	1865	1844	1844	1831
p = 1	1997	2001	1934	1940	1863	1869	1846	1850	1814
p = 2	2002	2005	1937	1920	1863	1862	1781	1776	1741
p = 3	1952	1943	1815	1800	1771	1774	1746	1742	1743
p = 4	1901	1795	1768	1692	1694	1696	1701	1694	1699
p = 5	1699	1702	1702	1693	1687	1692	1697	1699	1704
p = 6	1702	1699	1706	1691	1692	1697	1702	1703	1705
p = 7	1704	1704	1690	1695	1697	1701	1702	1710	1711
p=8	1709	1701	1700	1688	1693	1698	1700	1705	—

p-value of Box Test

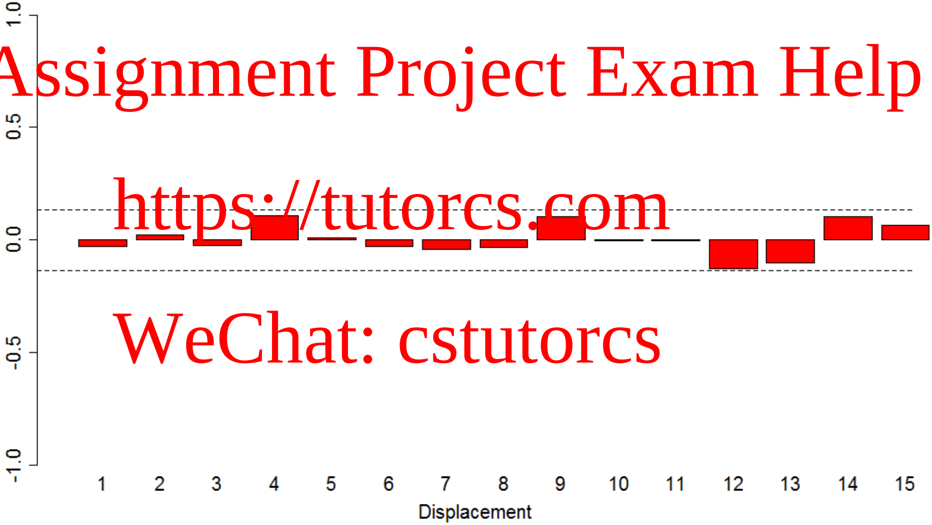
	q = 0	q = 1	q = 2	q = 3	q = 4	q = 5	q = 6	q = 7	q=8
p = 0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
p = 1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
p = 2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
p = 3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
p = 4	0.000	0.000	0.000	0.005	0.050	0.113	0.162	0.590	0.596
p = 5	0.014	0.020	0.009	0.101	0.541	0.614	0.621	0.587	0.598
p = 6	0.017	0.088	0.010	0.231	0.600	0.647	0.639	0.759	0.704
p = 7	0.013	0.085	0.280	0.350	0.583	0.589	0.630	0.723	0.812
p=8	0.020	0.112	0.224	0.766	0.788	0.780	0.917	0.651	—

- The information (AIC, SIC and the white noise test) suggests ARMA(5,4).

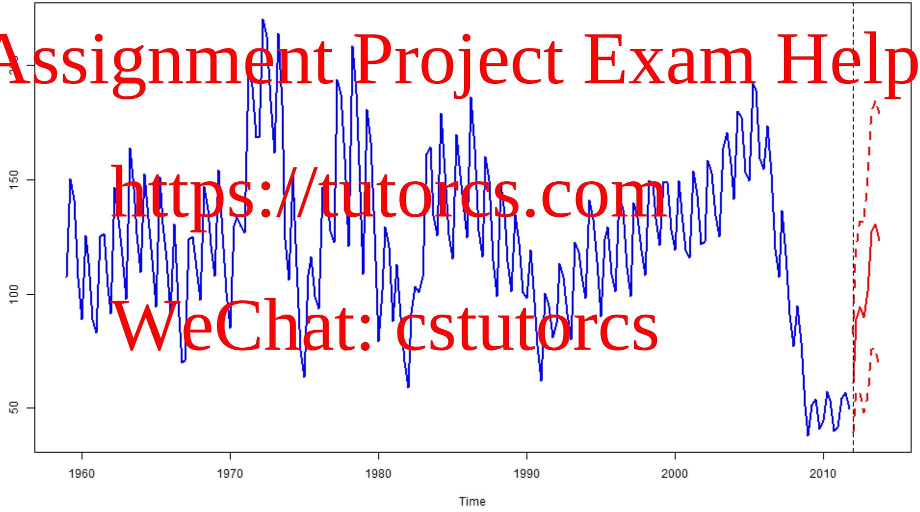
ACF of ARMA(5,4)



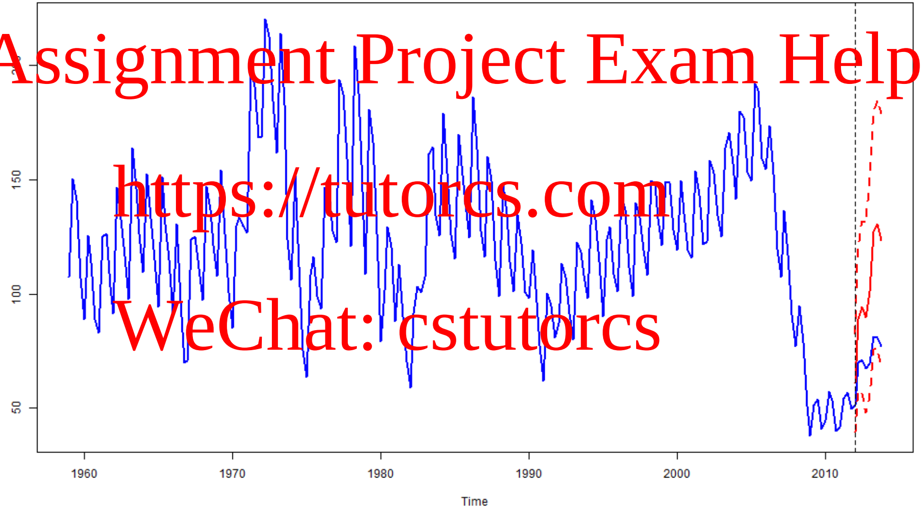
PACF of ARMA(5,4)



Forecast Comparison of ARMA(5,4)

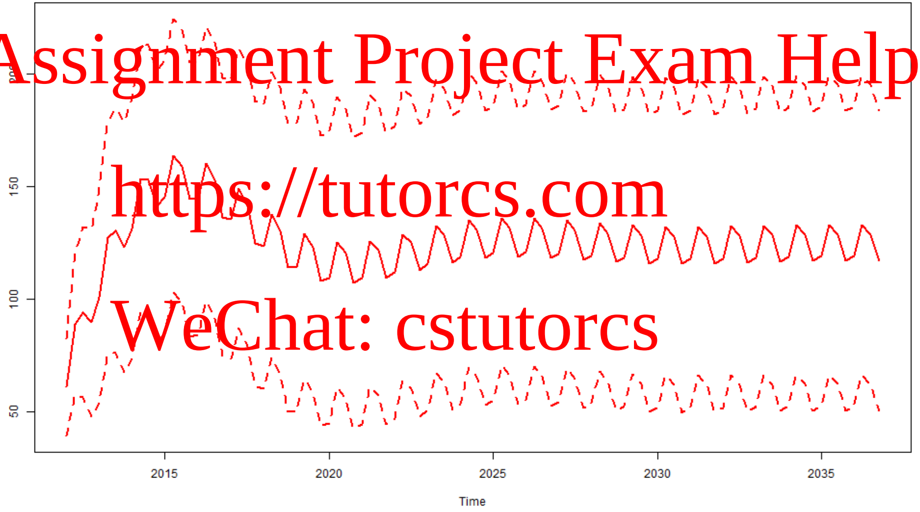


Forecast Comparison of ARMA(5,4)

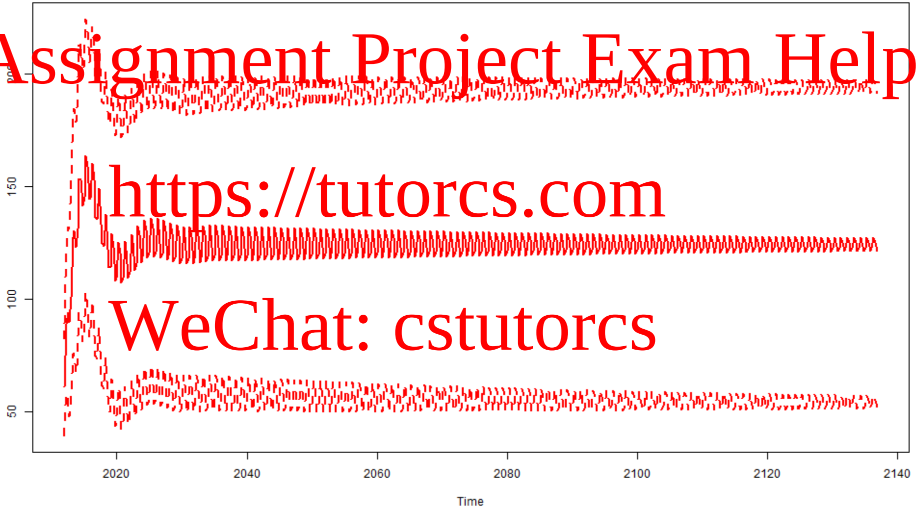


Example: US Housing Starts (Quarterly Non-seasonally Adjusted)

Long-horizon Forecast of ARMA(5,4)



Very Long-horizon Forecast of ARMA(5,4)



Remarks:

ARMA models will pick up *seasonality*.

The assumption of AR and MA components in the model building will lead to *different patterns* of the forecast (point and interval).

As forecast horizon (h) increases, the forecast will *converge to* the unconditional mean.

Estimation of pure AR models is based on OLS. The estimation of ARMA models with non-trivial MA components will be based on non-linear least squares or maximum likelihood, and hence requires numerical optimization. Because numerical optimization may fail some of the times, we may prefer *pure AR models* when we have relatively long series.

Additional Exercise:

Simulate 396 observations monthly data (33 years) with deterministic seasonality (using 12 monthly dummy variables).

- Use the first 360 observations (30 years) for estimation and remaining for out-of-sample comparison.
- Use the two approaches to estimate the model and produce one-step-ahead forecast recursively.
  - Model 1: model as deterministic seasonality using dummy variables.
  - Model 2: model as stochastic seasonality using ARMA models.

Which model yields a smaller mean squared prediction error?