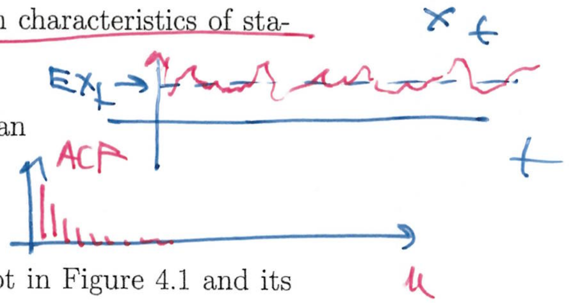
Lecture 5

Model fitting and diagnostic checking

We consider another example of fitting the model to the data and checking its fit. The same approach applies in practical work.

We fit ARMA model to stationary time series. Main characteristics of stationary time series are:

- observations are in equilibrium, around the mean
- autocorrelation function will die out quickly.



Example. We discuss temperature data X_t. Its plot in Figure 4.1 and its ACF indicate that the times series in non-stationary.

Poll 1

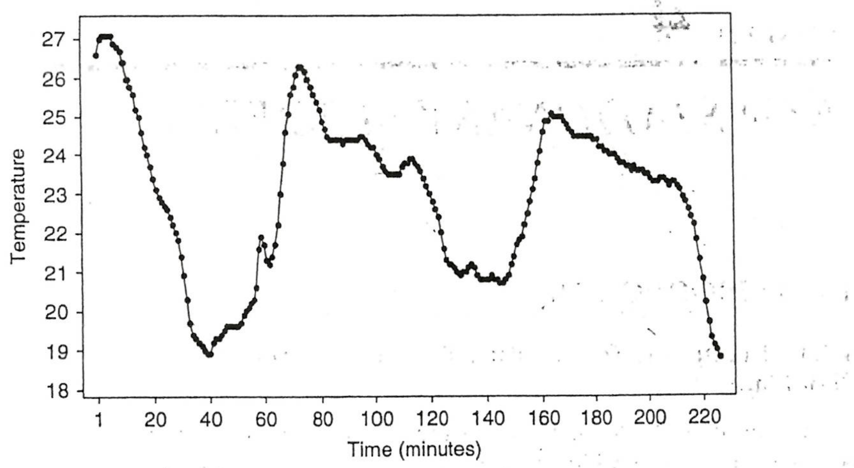
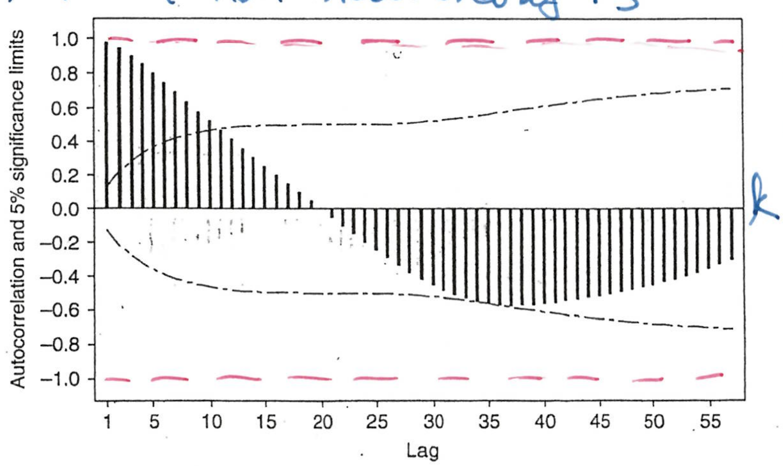


Figure 4.1 Time series plot of the temperature from a pilot plant observed every minute

ACF of non-stationary TS



ACF of stationary time series

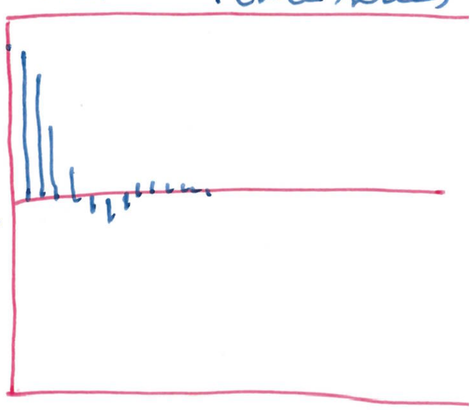


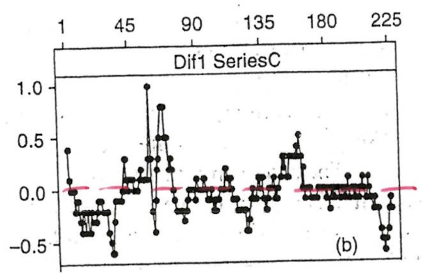
Figure 4.2 ACF for the temperature data.

How to transform to stationarity?

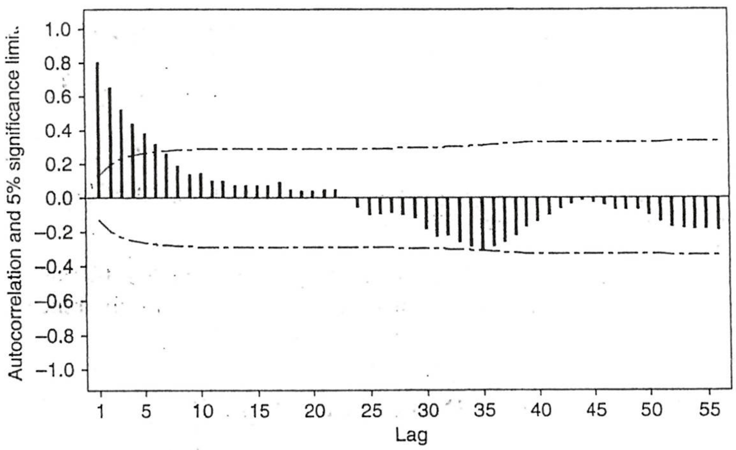
DIFFERENCING

We can try the first differences $\boxed{z_t = X_t - X_{t-1}.}$ We check for stationarity of z_t using the plot and the shape of ACF, see Figures 4.4 and 4.3.

The time series z_t looks as a stationary one.



Mean seems to be constant over time



ACF decays to 0 fast

Figure 4.3 ACF of the first difference of series C.

We would like to build a model to this data

Model building. How we proceed? In practice we never now the order p and q of AR or MA model we can fit, or of more general ARMA(p,q) model.

Poll 2

Model selection

For that we use ACF, PACF and information criterions (AIC, BIC).

We suggest the following iterative approach, and recall the patterns of AR, MA and ARMA models.

TABLE 4.1 Summary of Properties of Autoregressive (AR), Moving Average (MA), and Mixed Autoregressive Moving Average (ARMA) Processes

	AR(p)	MA(q)	ARMA(p,q)
Model	$w_t = \phi_1 w_{t-1} + \dots + \phi_p w_{t-p} + a_t$	$w_t = a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}$	$w_t = \phi_1 w_{t-1} + \dots + \phi_p w_{t-p} - \theta_1 a_{t-1} + \dots - \theta_q a_{t-q} + a_t$
Autocorrelation function (ACF)	Infinite; damped exponentials and/or damped sine waves; Tails off	Finite; cuts off after q lags	Infinite; damped exponentials and/or damped sine waves; Tails off
Partial autocorrelation function (PACF)	Finite; cuts off after p lags	Infinite; damped exponentials and/or damped sine waves; Tails off	Infinite; damped exponentials and/or damped sine waves; Tails off

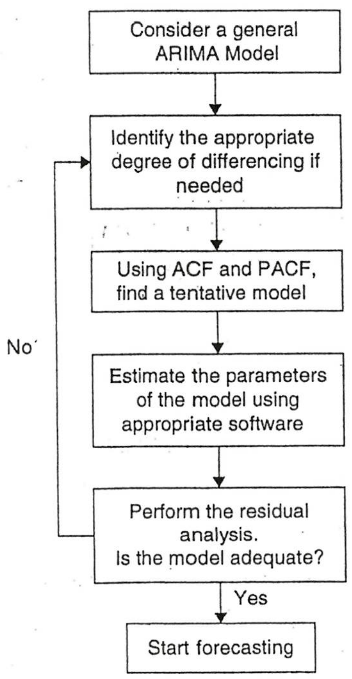


Figure 4.7 Stages of the time series model building process using ARIMA (Adapted from BJR, p. 18).

We now demonstrate how the iterative model building works.

PACF of z_t cuts off after lag 1. This suggest that z_t could be modelled by AR(1) model:

$$z_t = \phi z_{t-1} + \varepsilon_t.$$

AR(1) model

Substituting z_t into initial model, we obtain

$$X_t - X_{t-1} = \phi(X_{t-1} - X_{t-2}) + \varepsilon_t.$$

$z_t = X_t - X_{t-1}$

$z_1, \dots, z_n$

PACF of z_t

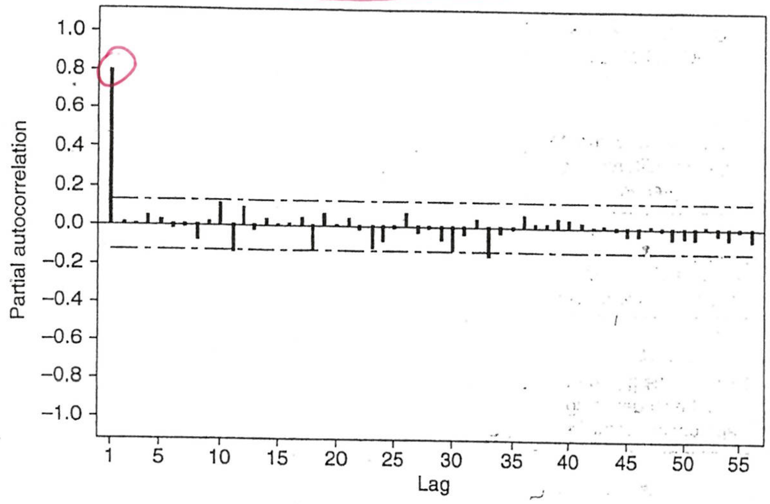


Figure 4.8 PACF for the differenced chemical process data.

Parameter estimation. Now when we identified the model, we estimate the parameters of the model. Depending on software package and method of estimation we may get slightly different results.

Table 4.2 shows estimation results of AR(1) model. Parameter $\hat{\phi}$ is significant, estimate variance of the noise is $\hat{\sigma}^2_\varepsilon = 0.018$.

TABLE 4.2 Estimation Summary from Fitting an AR(1) Model to the First Difference of the Chemical Process Data — Series C

Type	Coefficient	SE of coefficient	t	p
AR(1): $\hat{\phi}_1$	0.8239	0.0382	21.55	0.000

Poll 3

Model checking. Once the model is fitted we need to do diagnostic checks. If model fits well, the residuals should behave as a white noise, i.e. be uncorrelated.

Standard checks are to compute the ACF and PACF of residuals

$$\boxed{\hat{\varepsilon}_t = z_t - \hat{\phi} z_{t-1}.} \Leftarrow \quad z_t = \hat{\phi} z_{t-1} + \hat{\varepsilon}_t$$

Although there are a few autocorrelations and PACF's are larger than the 5% significance limits, there are no particularly alarming indications that the model does not fit well.

To do further checks we use Ljung -Box test to test for correlation in residuals, see Table 4.3. Since p values are larger than 0.05, we have no evidence that residuals are not a white noise.

Conclusion: AR(1) fits well.

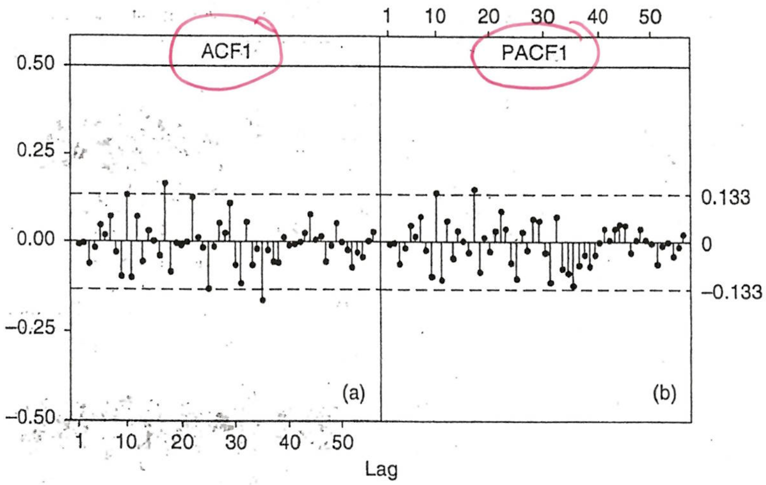


Figure 4.9 (a) ACF and (b) PACF of the residuals with 5% significance limits.

TABLE 4.3 Modified Ljung–Box–Pierce Chi-Square Statistic for the Residuals

Lag K	12	24	36	48
Chi square: $\tilde{Q}$	13.0	27.0	49.2	53.9
Degrees of freedom	11	23	35	47
p-values	0.292	0.254	0.056	0.229

For $m = 12$, we have $H_0: \rho_1 = \rho_2 = \dots = \rho_{12} = 0$

$H_1: \rho_j \neq 0$ for some $1 \leq j \leq 12$

Since $p = 0.292 > 0.05$, we do not reject H_0 (no correlation) at 5% significance level sample

$X_1, \ldots, X_t$

last observation

Forecasting out of sample

X_{t+k}

4 Forecasting

Forecasting is an important area of application of time series.

Once we fitted a model to a time series, the we can use it for forecasting about the future.

To forecast we need a model!

Assume that we have an AR(p) model

$$\boxed{X_t = \phi_0 + \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} + \varepsilon_t.}$$

unknown

Suppose we are at time t, and we have observations $X_1, \ldots, X_t$. — X_{t+k}

Our goal is to make forecast about X_{t+k}, k steps into the future.

X_t last observation

At this point, we assume that we know parameters $\phi_0, \phi_1, \ldots, \phi_p$. In applications these parameters will be replaced by estimated values $\hat{\phi}_0, \hat{\phi}_1, \ldots, \hat{\phi}_p$

Definition. The time t is called the forecast origin. The positive integer k is called the forecast step.

We denote such k-step ahead forecast by $X_t(k)$. → it is forecast of X_{t+k}

The error of such forecast is $\boxed{e_t(k) = X_{t+k} - \hat{X}_t(k).}$

Poll 4

Forecast criterion. In order to judge how good is the forecast we use the mean square error (MSE) $\boxed{Ee_t^2(k) = E(X_{t+k} - \hat{X}_t(k))^2.}$

- The best forecast is the one which minimizes the forecast error.
- It turns out the the forecast that minimizes MSE, is the conditional expectation of X_{t+k} at time t, that is

$$\boxed{\hat{X}_t(k) = E[X_{t+k}|F_t]}$$

FORMALA

where $F_t = \{X_t, X_{t-1}, \cdots, X_1\}$ denotes al information available at time t.

F_t = all past information until time t:

$$F_t = \left\{ \begin{matrix} X_1, \ldots, X_t \\ \varepsilon_1, \ldots, \varepsilon_t \end{matrix} \right\}$$

Question: How we actually compute that expectation or forecast?

Notation: We shall use notation $[X_{t+k}] = E[X_{t+k}|F_t]$.

Notation:

$$[X_{t+k}] = E[X_{t+k} | F_t]$$

RULES:

$[X_t] = E[X_t | \mathcal{F}_t] = X_t$

$[X_{t-1}] = E[X_{t-1} | \mathcal{F}_t] = X_{t-1}$

$[\varepsilon_t] = E[\varepsilon_t | \mathcal{F}_t] = \varepsilon_t$

$[\varepsilon_{t-2}] = E[\varepsilon_{t-2} | \mathcal{F}_t] = \varepsilon_{t-2}$

Poll 5

The calculation of the forecast is based on the following rules:

- The conditional expectation of the present and past values are actually those observations:

  $[X_t] = X_t$, $[X_{t-1}] = X_{t-1}$, $[X_{t-2}] = X_{t-2}$, and so on.

- The conditional expectation of the present and past shocks ε_j are actually those shocks:

  $[\varepsilon_t] = \varepsilon_t$, $[\varepsilon_{t-1}] = \varepsilon_{t-1}$, $[\varepsilon_{t-2}] = \varepsilon_{t-2}$, and so on.

- The conditional expectations of the future shocks are 0:

  $[\varepsilon_{t+1}] = E[\varepsilon_{t+1}|F_t] = 0$, $[\varepsilon_{t+2}] = E[\varepsilon_{t+2}|F_t] = 0$.

  Why? Because future shocks are independent of the past, so

  $[\varepsilon_{t+1}] = E\varepsilon_{t+1} = 0$.

  $[\varepsilon_{t+1}] = E[\varepsilon_{t+1} | \mathcal{F}_t] = E\varepsilon_t = 0$

- In computations, we replace the conditional expectations of the future observations by the forecasts:

  $[X_{t+1}] = \hat{X}_t(1)$, $[X_{t+2}] = \hat{X}_t(2)$, $\cdots$

  $[X_{t+1}] = \hat{X}_t(1) \sim X_{t+1}$ forecast of

4.1 Forecasting using AR(p) model

$\boxed{\textbf{1-Step Ahead Forecast } X_t(1).}$ By definition of AR(p) model,

$$X_{t+1} = \phi_0 + \phi_1 X_t + ... + \phi_p X_{t+1-p} + \varepsilon_{t+1}.$$

$\leftarrow$ Model

$X_1, \dots, X_t$

last observation

Forecast $\hat{X}_t(1)$ of X_{t+1} is the conditional expectation

$$\begin{aligned} \hat{X}_t(1) &= E[X_{t+1}|X_t, X_{t-1}, ..., X_1] \\ &= [\phi_0 + \phi_1 X_t + ... + \phi_p X_{t+1-p} + \varepsilon_{t+1}]. \end{aligned}$$

$\leftarrow$ Formula

$= E[\phi_0 + \phi_1 X_t + \dots + \phi_p X_{t+1-p} + \varepsilon_t | \mathcal{F}_t]$

Using above rules, we find that

$$\begin{aligned} \hat{X}_t(1) &= E[X_{t+1}|F_t] = [\phi_0 + \phi_1 X_t + ... + \phi_p X_{t+1-p} + \varepsilon_{t+1}] \\ &= [\phi_0] + \phi_1 [X_t] + ... + \phi_p [X_{t+1-p}] + [\varepsilon_{t+1}] \\ &= \phi_0 + \phi_1 X_t + ... + \phi_p X_{t+1-p}. \end{aligned}$$

$\leftarrow$ the forecast

$[\varepsilon_{t+1}] = 0$

$[X_t] = X_t$

Notice $[\phi X_{t+1}] = E[\phi X_{t+1} | \mathcal{F}_t]$

$= \phi E[X_{t+1} | \mathcal{F}_t]$

$= \phi \hat{X}_t(1)$

Poll 6

The forecast error:

$$\boxed{e_t(1) = \cancel{X}_{t+1} - \cancel{\hat{X}}_{t+1} = \varepsilon_{t+1}.}$$

The variance of the 1-step ahead forecast error:

$$Var(e_t(1)) = Var(\varepsilon_{t+1}) = \sigma_\varepsilon^2.$$

Forecast interval. If ε_t is normally distributed, then the 95%- confidence interval for 1-step ahead forecast of X_{t+1} is

$$\left[\hat{X}_t(1) - 1.96\sigma_\varepsilon, \quad \hat{X}_t(1) + 1.96\sigma_\varepsilon\right].$$

X_{t+1} with 95% probability

2-Step Ahead Forecast $X_t(2)$. Consider AR(2) model:

$$X_t = \phi_0 + \phi_1 X_{t-1} + ... + \phi_p X_{t-2} + \varepsilon_t.$$

From above, we know that 1-step ahead forecast is

$$X_t(1) = \phi_0 + \phi_1 X_t + \phi_2 X_{t-1}.$$

To compute $X_t(2)$, first we write equation for X_{t+2}:

Using AR(2) model write $X_{t+2} = \ldots$

$$X_{t+2} = \phi_0 + \phi_1 X_{t+1} + \phi_2 X_t + \varepsilon_{t+2}.$$

Using our rules, we find that

$$\begin{aligned}
\hat{X}_t(2) &= E[X_{t+2}|F_t] = [\phi_0 + \phi_1 X_{t+1} + \phi_2 X_t + \varepsilon_{t+2}] \\
&= \phi_0 + \phi_1 [X_{t+1}] + \phi_2 [X_t] + [\varepsilon_{t+2}] \\
&= \phi_0 + \phi_1 \hat{X}_t(1) + \phi_2 X_t \\
&= \phi_0 + \phi_1 \{\phi_0 + \phi_1 X_t + \phi_2 X_{t-1}\}) + \phi_2 X_t \\
&= \phi_0 + \phi_1 \phi_0 + (\phi_1^2 + \phi_2) X_t + \phi_2 X_{t-1}.
\end{aligned}$$

$[X_{t+1}] = \hat{X}_t(1)$

$[\varepsilon_{t+2}] = 0$

The forecast error is

$$\begin{aligned}
e_t(2) &= X_{t+2} - \hat{X}_{t+2} \\
&= \phi_0 + \phi_1 X_{t+1} + \phi_2 X_t + \varepsilon_{t+2} - \{\phi_0 + \phi_1 \hat{X}_t(1) + \phi_2 X_t\} \\
&= \phi_1 (\underbrace{X_{t+1} - \hat{X}_t(1)}_{= \varepsilon_{t+1}}) + \varepsilon_{t+2} = \phi_1 \varepsilon_{t+1} + \varepsilon_{t+2}.
\end{aligned}$$

$= Var(\phi_1 \varepsilon_{t+1}) + Var(\varepsilon_{t+2})$

The variance of the forecast error is

$$Var(e_t(2)) = Var(\phi_1 \varepsilon_{t+1} + \varepsilon_{t+2}) = \phi_1^2 \sigma_\varepsilon^2 + \sigma_\varepsilon^2 = (1 + \phi_1^2)\sigma_\varepsilon^2.$$

The 95% confidence interval for X_{t+2}: setting $\sigma_2^2 = Var(e_t(2))$,

$$\left[\hat{X}_t(2) - 1.96\sigma_2, \quad \hat{X}_t(2) + 1.96\sigma_2\right].$$

Note: Observe that

$$\boxed{Var(e_t(2)) \geq Var(e_t(1)).}$$

Comparing the 2-step ahead forecasting with 1-step ahead forecasting we see that

- when forecasting step k increases

  the uncertainty in forecasting increases.

Example. The researcher analyzed the sample $X_1, \cdots, X_t$ and found that it is not from a stationary time series. He checked the differenced series $z_t = X_t - X_{t-1}$ and found that fitting to it the AR(1) model gives uncorrelated residuals.

The fitted model was $z_t = 1 + 0.5z_{t-1} + \varepsilon_t$.

AR(1) model

He/she is interested in forecasting X_{t+1} and X_{t+2}. How to compute these forecasts?

Solution. First we compute the forecast of z_{t+1} and z_{t+2} as we did above.

To compute $\hat{z}_t(1)$, write $z_{t+1} = 1 + 0.5z_t + \varepsilon_{t+1}$. Then

$[\varepsilon_{t+1}] = 0$

$[z_t] = z_t$

$$\begin{aligned}\hat{z}_t(1) &= [1 + 0.5z_t + \varepsilon_{t+1}] \\ &= 1 + 0.5[z_t] + [\varepsilon_{t+1}] \\ &= 1 + 0.5z_t.\end{aligned}$$

To compute $\hat{z}_t(2)$, write $z_{t+2} = 1 + 0.5z_{t+1} + \varepsilon_{t+2}$. Then

$[z_{t+1}] = \hat{z}_t(1)$

$[\varepsilon_{t+2}] = 0$

$$\begin{aligned}\hat{z}_t(2) &= [1 + 0.5z_{t+1} + \varepsilon_{t+2}] \\ &= 1 + 0.5[z_{t+1}] + [\varepsilon_{t+2}] \\ &= 1 + 0.5\hat{z}_t(1) \\ &= 1 + 0.5(1 + 0.5z_t) = 1.5 + 0.25z_t.\end{aligned}$$

To compute $\hat{X}_t(1)$, we use equality $z_{t+1} = X_{t+1} - X_t$.

Then

$$\begin{aligned} \hat{z}_t(1) &= [z_{t+1}] = [X_{t+1} - X_t] \\ &= [X_{t+1}] - [X_t] = \hat{X}_t(1) - X_t. \end{aligned}$$

So,

$$\begin{aligned} \hat{X}_t(1) &= X_t + \hat{z}_t(1) \\ &= X_t + 1 + 0.5z_t \\ &= X_t + 1 + 0.5(X_t - X_{t-1}) \\ &= 1 + 1.5X_t - 0.5X_{t-1}. \end{aligned}$$

O_2

$X_{t+1} = X_t + Z_{t+1}$

$$\begin{aligned} \hat{X}_t(1) &= [X_{t+1}] = [X_t + Z_{t+1}] \\ &= [X_t] + [Z_{t+1}] \\ &= X_t + \hat{Z}_t(1) \end{aligned}$$

To compute $\hat{X}_t(2)$, we use equality $z_{t+2} = X_{t+2} - X_{t+1}$.

Then

$$\begin{aligned} \hat{z}_t(2) &= [z_{t+2}] = [X_{t+2} - X_{t+1}] \\ &= [X_{t+2}] - [X_{t+1}] \\ &= \hat{X}_t(2) - \hat{X}_t(1). \end{aligned}$$

O_2

$X_{t+2} = X_{t+1} + Z_{t+2}$

$$\begin{aligned} \hat{X}_t(2) &= [X_{t+2}] = [X_{t+1} + Z_{t+2}] \\ &= [X_{t+1}] + [Z_{t+2}] \\ &= \hat{X}_t(1) + \hat{Z}_t(2) \end{aligned}$$

So,

'' $\hat{Z}_t(2)$ (see p. 55)

$$\begin{aligned} \hat{X}_t(2) &= X_t(1) + \hat{z}_t(2) \\ &= \{X_t + 1 + 0.5z_t\} + \{1.5 + 0.25z_t\} \\ &= 2.5 + X_t + 0.75z_t = 2.5 + X_t + 0.75(X_t - X_{t-1}) \\ &= 2.5 + 1.75X_t - 0.75X_{t-1}. \end{aligned}$$

$z_t = x_t - x_{t-1}$

X_t – stationary sequence

Mean reversion. In AR(p) model,as the step k increases, the forecast $\hat{X}_t(k)$ has property:

$$\hat{X}_t(k) \to E[X_t], \quad as \quad k \to \infty.$$

$\hat{X}_t(k) = E[X_{t+k}|\mathcal{F}_t]$

That means: forecasts of the far future $\hat{X}_t(k)$ are close to unconditional mean. This property is called mean reversion of the forecast.

In addition, the variance of the forecast error

$$Var(e_t(k)) \to Var(X_t), \quad as \quad k \to \infty.$$

Poll 7

In finance, this property is called mean reversion.

Example. Table 2.2 below shows forecast of the value weighted monthly simple returns at the forecast origin 858. Forecasting is done using AR(3) model which parameters were estimated using the first 858 observations:

$$r_t = 0.0103 + 0.104 r_{t-1} - 0.010 r_{t-2} - 0.12 r_{t-3} + \varepsilon_t, \quad \hat{\sigma}_\varepsilon = 0.054.$$

The actual returns are also given. The sample mean is 0.0098.

Note: because of the weak serial correlation, the forecast and standard deviations of forecast errors converges to the sample mean and standard deviation of the data quickly.

Figure 2.7 shows the first 6-steps of forecast for these return data.

Finding. For this particular series, we observe that

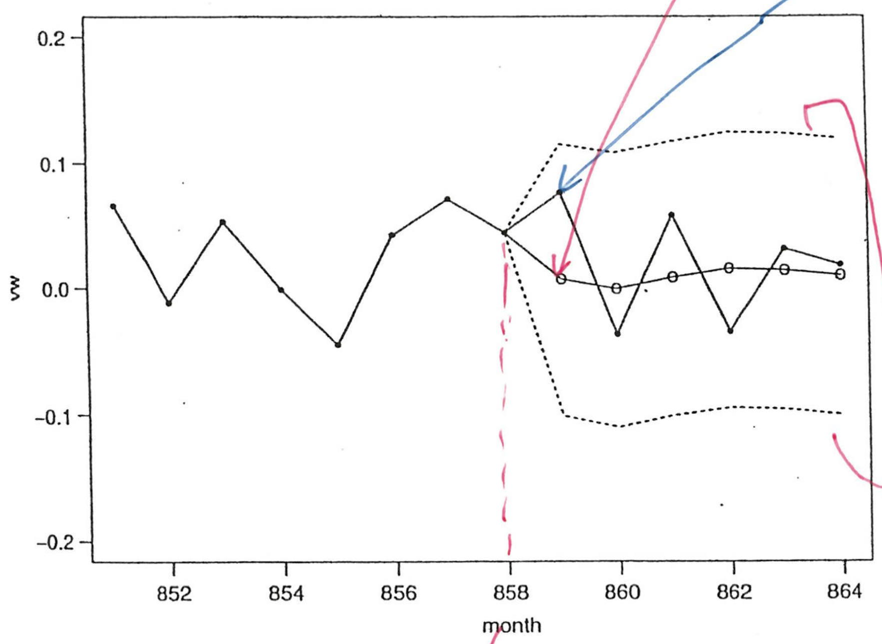
- the forecasts are close to the actual values
- actual values are within 95-percent confidence interval.

Table 2.2. **Multistep Ahead Forecasts of an AR(3) Model for the Monthly Simple Returns of CRSP Value-Weighted Index with Forecast Origin of 858**

Step	1	2	3	4	5	6
Forecast	0.0088	0.0020	0.0050	0.0097	0.0109	0.0106
Standard error	0.0542	0.0546	0.0546	0.0550	0.0550	0.0550
Actual	0.0762	−0.0365	0.0580	−0.0341	0.0311	0.0183

~ 0.0098
~ 0.054

forecasted value
true value



95% confidence interval for the forecast

Figure 2.7. Plot of 1-step to 6-step ahead out-of-sample forecasts for the monthly log returns of the CRSP value-weighted index. The forecast origin is $t = 858$. The forecasts are denoted by o and the actual observations by black dots. The two dashed lines denote two standard-error limits of the forecasts.

forecast origin
(last observation)
t=858

4.1.1 Forecasting using MA Model

Forecast of an MA model is rather easy. Let t be the forecast origin

- The forecast MA model goes fast to the mean of the model. !

1-step ahead forecast of an MA(1) model. To forecast X_{t+1} using $X_1, \cdots, X_t$, write

$$\boxed{X_{t+1} = c_0 + \varepsilon_{t+1} + \theta_1 \varepsilon_t.}$$

Model

We obtain the actual forecast taking the conditional expectation

$$\begin{aligned} \hat{X}_t(1) &= E[X_{t+1}|F_t] = [X_{t+1}] \\ &= [c_0 + \varepsilon_{t+1} + \theta_1 \varepsilon_t] \\ &= c_0 + [\varepsilon_{t+1}] + \theta_1 [\varepsilon_t] \\ &= c_0 + \theta_1 \varepsilon_t. \end{aligned}$$

$[\varepsilon_{t+1}] = E[\varepsilon_{t+1} | F_t] = 0$

$[\varepsilon_t] = E[\varepsilon_t | F_t] = \varepsilon_t$

The forecast error is

$$\boxed{e_t(1) = X_{t+1} - \hat{X}_t(1) = \varepsilon_{t+1}.}$$

The variance of the 1-step ahead forecast is

$$Var(e_t(1)) = Var(\varepsilon_{t+1}) = \sigma_\varepsilon^2.$$

Question: How to compute ε_t, which we need for forecasting?

We can compute it setting $\varepsilon_0 = 0$.

2-step ahead forecast of an MA(1) model. The model says that

$$\boxed{X_{t+2} = c_0 + \varepsilon_{t+2} + \theta_1 \varepsilon_{t+1}.}$$

Then

$$\begin{aligned} \hat{X}_t(2) &= [c_0 + \varepsilon_{t+2} + \theta_1 \varepsilon_{t+1}] \\ &= c_0 + \underbrace{[\varepsilon_{t+2}]}_{0} + \theta_1 \underbrace{[\varepsilon_{t+1}]}_{0} = c_0. \end{aligned}$$

$[\varepsilon_{t+2}] = E[\varepsilon_{t+2} | F_t]$
$= 0$

The forecast error is

$$\begin{aligned} e_t(2) = X_{t+2} - \hat{X}_t(2) &= \{\cancel{c_0} + \varepsilon_{t+2} + \theta_1 \varepsilon_{t+1}\} - \cancel{c_0} \\ &= \varepsilon_{t+2} + \theta_1 \varepsilon_{t+1}. \end{aligned}$$

The variance of the 2-step ahead forecast is $= Var(\varepsilon_{t+2}) + Var(\theta_1 \varepsilon_{t+1})$

$$Var(e_t(2)) = Var(\varepsilon_{t+2} + \theta_1 \varepsilon_{t+1}) = (1+\theta_1^2)\sigma_\varepsilon^2. \quad = \sigma_\varepsilon^2 + \theta_1^2 \sigma_\varepsilon^2$$

Note:

$$Var(e_t(2)) = (1+\theta_1^2)\sigma_\varepsilon^2 > Var(e_t(1)) = \sigma_\varepsilon^2. \quad !$$

Theory shows that 2-step ahead forecast for MA(1) model is its mean $EX_t = c_0$. More generally, the multi $k-$step forecast, $k \geq 2$, is

Poll 8

$$\hat{X}_t(k) = c_0 = E[X_t]. \quad \text{for } k = 2, 3, \ldots$$

Conclusion: for MA(1) model forecast mean reverting starts at time 1 period.

Forecast using an MA(2) model. For AR(2) model we have, that

$$X_{t+k} = c_0 + \varepsilon_{t+k} + \theta_1 \varepsilon_{t+k-1} + \theta_2 \varepsilon_{t+k-2}.$$

Taking conditional expectation we obtain the forecast

$$\begin{aligned} \hat{X}_t(1) &= [X_{t+1}] = c_0 + \theta_1 \varepsilon_t + \theta_2 \varepsilon_{t-1}. \\ \hat{X}_t(2) &= [X_{t+2}] = c_0 + \theta_2 \varepsilon_t. \\ \hat{X}_t(k) &= [X_{t+k}] = c_0, \quad for \quad k \geq 3. \end{aligned}$$

Conclusion: Forescasts of MA(2) model have property:

- reverse to the mean $E[X_t] = c_0$ after two steps. $\hat{X}_t(k) = c_0, \ k = 3, 4, \ldots$

|| **Property** of MA(q) model: the forecast $X_t(k)$ reverts to the mean EX_t when the step $k > q$.

MA(9)

Example. Table 2.3 contains forecasts of the ~~AR(9)~~ model for the monthly simple returns of the equal weighter index at forecast origin $t = 926$.

The sample mean and the standard error of the first 926 observations are 0.0126 and 0.0751.

We conclude that

- The 10-step ahead forecast is the sample mean;
- forecast error converges (increases) to the standard deviation when forecast step increases.

Table 2.3. Forecast Performance of a MA(9) Model for the Monthly Simple Returns of the CRSP Equal-Weighted Index[a]

Step	1	2	3	4	5
Forecast	0.0140	−0.0050	0.0158	−0.0008	0.0171
Standard error	0.0726	0.0737	0.0737	0.0743	0.0743
Actual	0.0097	0.0983	0.1330	0.0496	0.0617
Step	**6**	**7**	**8**	**9**	**10**
Forecast	0.0257	0.0009	0.0149	0.0099	0.0126
Standard error	0.0743	0.0743	0.0743	0.0743	0.0748
Actual	0.0475	0.0252	0.0810	0.0381	0.0391

[a] The forecast origin is February 2003 with $h = 926$. The model is estimated by the conditional maximum likelihood method.

4.2 Forecasting of ARMA(1,1) model

Forecast of ARMA(1,1) model has characteristics similar to those of an AR(1) model.

Denote the forecast origin by t. Using the available information $X_1, \cdots, X_t$, we compute the forecast $X_t(1)$.

Question: predict X_{t+1} !

The 1-step ahead forecast. By definition of ARMA(1,1) model,

$$\boxed{X_{t+1} = \phi_0 + \phi_1 X_t + \varepsilon_{t+1} + \theta_1 \varepsilon_t.}$$

The 1-step ahead forecast is obtained as follows:

X_t 0 ε_t

$$\begin{aligned}\hat{X}_t(1) &= E[X_{t+1}|F_t] = [\phi_0 + \phi_1 X_t + \varepsilon_{t+1} + \theta_1 \varepsilon_t] \\ &= \phi_0 + \phi_1[X_t] + [\varepsilon_{t+1}] + \theta_1[\varepsilon_t] \\ &= \phi_0 + \phi_1 X_t + \theta_1 \varepsilon_t.\end{aligned}$$

Forecast error is:

$$\begin{aligned}e_t(1) &= X_{t+1} - \hat{X}_t(1) \\ &= \{\phi_0 + \phi_1 X_t + \varepsilon_{t+1} + \theta_1 \varepsilon_t\} - \{\phi_0 + \phi_1 X_t + \theta_1 \varepsilon_t\} \\ &= \varepsilon_{t+1}.\end{aligned}$$

Variance of the error:

$$Var(e_t(1)) = Var(\varepsilon_{t+1}) = \sigma_\varepsilon^2.$$

To apply forecast $X_t(1)$, How to compute the forecast $\hat{X}_t(1)$

- we estimate the unknown parameters ϕ_0, ϕ_1, θ_1 from the data.
- X_t we know.
- ε_t can be estimated from data $X_1, \cdots, X_t$.