Chapter 5: Autoregressive Integrated Moving Average(ARIMA) Models

Rong Fan

South Dakota State University

rong.fan@sdstate.edu

5.7 ARIMA Model Building

Three step iterative procedure is used to build an ARIMA model.

1. Identify the model through analysis of historical data.
   AR(p), or MA(q), or ARMA(p,q), or ARIMA(p,d,q)...

2. Estimate the parameters of the model.
   Method of Moments (MOM), Maximum Likelihood (ML) and Least Squares (LS) estimates are commonly used methods.

3. Perform the diagnostic checks to determine the adequacy of the model, or indicate potential improvement.
   Residual analysis

5.7.1 Model Identification I

Plot the data and do a visual inspection.

Address issues of stationary, need for differencing or dth differencing

Other transformations.

For assumed stationary series after differencing etc, examine sample ACF and PACF up to lag 20-25 to suggest possible ARMA models.
Recall the properties of ACF and PACF for AR, MA, and ARMA models.

Model	ACF $\rho(k)$	PACF ϕ_{kk}
MA(q)	0 for $k > q$	exp decay ($q = 1$) mixture ($q > 1$)
AR(p)	exp decay ($p = 1$) mixture ($p > 1$)	0 for $k > p$
ARMA(p, q)	mixture	mixture

5.7.1 Model Identification II

1. For AR(p), it is a common practice to use $\pm 2/\sqrt{T}$ as limits to assess the significiance of PACF ϕ_{kk} in terms of deviation from 0 for $k > p$.
2. For MA(q), we can use $\pm 2/\sqrt{T}$ or $\pm 2\sqrt{1 + 2\sum_{i=1}^{q} \hat{\rho}(i)^2}/\sqrt{T}$ as limits to assess the significance of $\rho(k)$ in terms of deviation from 0 for $k > q$.
3. ARMA models would require more, as both ACF and PACF will exhibit a mixture of expoential decay and damped sinusiod behaviour.

Some remarks on ARIMA model identification:

- Because we are dealing with estimates, it will not always be clear whether the sample ACF or PACF is tailing off or cutting off.
- Two (or more) models that are seemingly different can actually be very similar, so we should not worry about being so percise at this stage.

5.7.1 Model Identification III

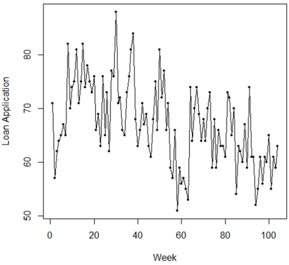
- A few perliminary values of p,d,and q could be at hand, and we can start estimating the parameters.
- If the sum of AR coefficients is almost equal to 1 (in this case, there is a unit root in the AR part of the model.), you should reduce p by one and increase d by one.
- If the sum of MA coefficients is almost equal to 1 (in this case, there is a unit root in the MA part of the model.), you should reduce q by one and reduce d by one.

5.7.1 Model Identification - Example 1 I

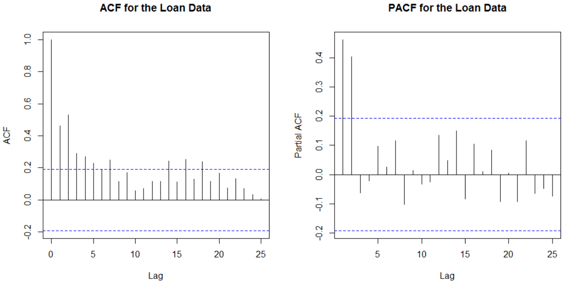
Example 1: The data set contains weekly total number of loan applications in a local bank for last 2 years.

TABLE 5.5 Weekly Total Number of Loan Applications for the Last 2 Years

Week	Applications	Week	Applications	Week	Applications	Week	Applications
1	71	27	62	53	66	79	63
2	57	28	77	54	71	80	61
3	62	29	76	55	59	81	73
4	64	30	88	56	57	82	72
5	65	31	71	57	66	83	65
6	67	32	72	58	51	84	70
7	65	33	66	59	59	85	54
8	82	34	65	60	56	86	63
9	70	35	73	61	57	87	62
10	74	36	76	62	55	88	60
11	75	37	81	63	53	89	67
12	81	38	84	64	74	90	59
13	71	39	68	65	64	91	74
14	75	40	63	66	70	92	61
15	82	41	66	67	74	93	61
16	74	42	71	68	69	94	52
17	78	43	67	69	64	95	55
18	75	44	69	70	68	96	61
19	73	45	63	71	64	97	56
20	76	46	61	72	70	98	61
21	66	47	68	73	73	99	60
22	69	48	75	74	59	100	65
23	63	49	66	75	68	101	55
24	76	50	81	76	59	102	61
25	65	51	72	77	66	103	59
26	73	52	77	78	63	104	63



The sample ACF and PACF plots are shown below.



The sample ACF plot suggests MA(2), MA(3) or even MA(4) models. Alternatively, it can be viewed an exponential decay pattern suggesting AR(p) models. Then we check sample PACF plot and conclude it cuts off after 2 suggesting AR(2) model.

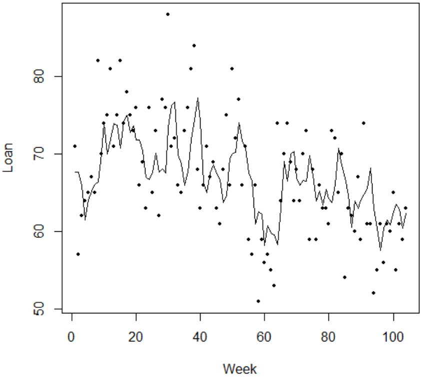
```
> acf(loan.data[,2], lag.max=25, type="partial", main="PACF for the Loan Data")
> loan.ar2 <- arima(loan.data[,2], order=c(2,0,0))
> loan.ar2

Call:
arima(x = loan.data[, 2], order = c(2, 0, 0))

Coefficients:
         ar1     ar2  intercept
      0.2659  0.4130    66.8538
s.e.  0.0890  0.0901     1.8334

sigma^2 estimated as 38.32:  log likelihood = -337.46,  aic = 682.92
```

Here the intercept is the mean, therefore the constant $\delta = 66.8538 * (1 - 0.2659 - 0.4130) = 21.467$. So the fitted model is

$$y_t = 21.467 + 0.2659y_{t-1} + 0.4130y_{t-2} + \epsilon_t$$

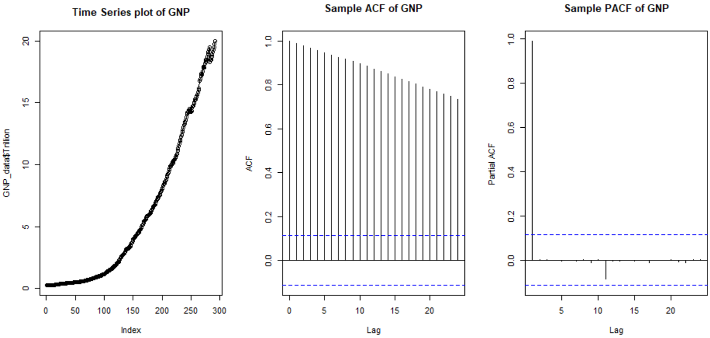
Here is a plot of the fitted values.

5.7.1 Model Identification - Example 1 IV



The analysis of the residuals is discussed later

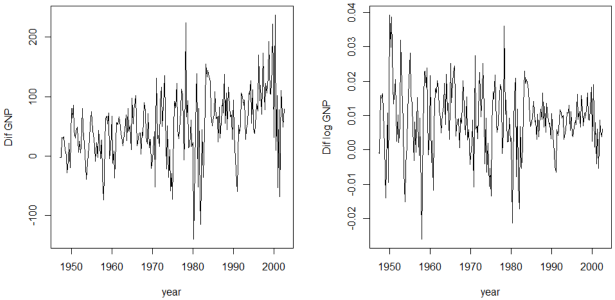
Example 2: The data set contains quarterly US GNP from 1947(1) to 2002(3).



5.7.1 Model Identification-Example 2 II

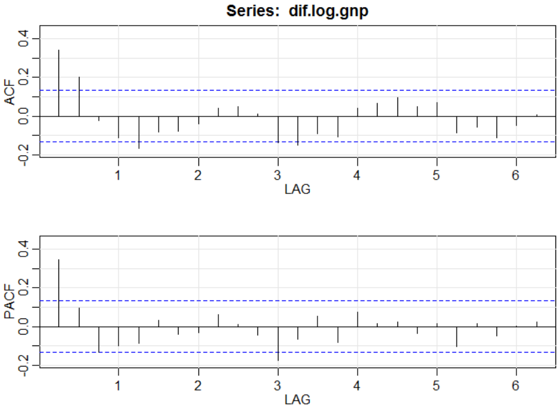
The plot shows a trend but no seasonality. Both Time series plot and ACF show that the process is nonstationary. Even though PACF shows a cut off at lag 1, we can't use a AR(q) model. Since for a finite q, MA(q) process is always stationary.Also notice that the ϕ_{11} in PACF is very close to 1 which implies a non-invertible process. The time series plot and the ACF both suggest that a difference is needed.
The plot of the difference shows an increasing variance over time. The plot of the differenced log looks a little better.

5.7.1 Model Identification-Example 2 III



Notice we take log before taking difference.

The ACF and PACF plots for diff log data are



5.7.1 Model Identification-Example 2 V

Models MA(2) and AR(1) are promising. They are fitted as

$$w_t = 0.0085(1 - 0.3429) + 0.3429 w_{t-1} + \epsilon_t \tag{1}$$

$$w_t = 0.0085 + 0.2998\epsilon_{t-1} + 0.2225\epsilon_{t-2} + \epsilon_t \tag{2}$$

In fact, the fitted models are nearly the same. One can see this from the plots of the fitted values. One can also find infinite MA form for the AR(1) model and compare it with the fitted MA(2) model. (or find infinite AR form for the MA(2) model and compare it with the fitted AR(1) model.)

5.7.2 Parameter Estimation I

Software packages can perform the estimation by using different methods. In particular, R uses Maximum Likelihood estimates.

For demonstration, we briefly go over the Method of Moments for some easy cases.

1. Review: For AR(1), $y_t = \phi y_{t-1} + \epsilon_t$, we have $\rho(k) = \phi^k$.This implies the following two equations:

$$\gamma(0) = \phi^2\gamma(0) + \sigma^2 \tag{3}$$

$$\rho(1) = \phi\rho(0) = \phi \tag{4}$$

   Thus the MOM (Methods of Moment) estimates for AR(1) are:

$$\hat{\phi} = \hat{\rho}(1) \tag{5}$$

$$\hat{\sigma}^2 = \hat{\gamma}(0)[1 - \hat{\phi}\hat{\rho}(1)] \tag{6}$$

5.7.2 Parameter Estimation II

2. Review: For AR(2), $y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t$, we have

$$\gamma(0) = \phi_1\gamma(1) + \phi_2\gamma(2) + \sigma^2 \tag{7}$$
$$\gamma(k) = \phi_1\gamma(k-1) + \phi_2\gamma(k-2), k = 1, 2, \cdots \tag{8}$$
$$\rho(k) = \phi_1\rho(k-1) + \phi_2\rho(k-2), k = 1, 2, \cdots \tag{9}$$

we use the following relationships

$$\gamma(0) = \phi_1\gamma(1) + \phi_2\gamma(2) + \sigma^2 \tag{10}$$
$$\rho(1) = \phi_1\rho(0) + \phi_2\rho(1), \tag{11}$$
$$\rho(k) = \phi_1\rho(k-1) + \phi_2\rho(k-2), \quad k = 2, 3, 4... \tag{12}$$

where (12) is the Yule-Walker equation (5.40, page 344). From (11), we have

$$\sigma^2 = \gamma(0)(1 - \phi_1\rho(1) - \phi_2\rho(2)) \tag{13}$$

5.7.2 Parameter Estimation III

To solve for $\hat{\sigma}^2$, $\hat{\phi}_1$ and $\hat{\phi}_2$, we just need $\hat{\gamma}(0)$, $\hat{\rho}(1)$, $\hat{\rho}(2)$.
For example, from a sample of AR(2) process we obtain $\hat{\gamma}(0) = 5.6$, $\hat{\rho}(1) = 0.48$, $\hat{\rho}(2) = -0.1$, then the MOM estimates are computed by solving

$$\begin{cases} \hat{\sigma}^2 = 5.6 * (1 - \hat{\phi}_1 * 0.48 - \hat{\phi}_2 * (-0.1)) \\ 0.48 = \hat{\phi}_1 * 1 + \hat{\phi}_2 * 0.48 \\ -0.1 = \hat{\phi}_1 * 0.48 + \hat{\phi}_2 * 1 \end{cases} \tag{14}$$

Thus $\hat{\phi}_1 = 0.686$ $\hat{\phi}_2 = -0.429$ and $\hat{\sigma} = 3.515$ Therefore the AR(2) model is

$$y_t = 0.686 y_{t-1} - 0.429 y_{t-2} + \epsilon_t. \tag{15}$$

5.7.2 Parameter Estimation IV

Exercise: How to estimate the δ in the following AR(2)?

$$y_t = \delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t \tag{16}$$

Solution: Since the mean of y is

$$\mu_y = \frac{\delta}{1 - \phi_1 - \phi_2}, \tag{17}$$

Replace μ_y by $\bar{y}$ in (17), then solve for δ,

$$\hat{\delta} = \bar{y}(1 - \hat{\phi}_1 - \hat{\phi}_2). \tag{18}$$

5.7.2 Parameter Estimation V

3. For MA(1) $y_t = \epsilon_t - \theta\epsilon_{t-1}$ we have 2 equations (See (5.12) & (5.13) on page 335)

$$\gamma(0) = \sigma^2(1+\theta^2) \tag{19}$$

$$\rho(1) = -\frac{\theta}{1+\theta^2} \tag{20}$$

Thus the MOM estimates for MA(1) are solved from

$$\hat{\theta}^2 + \frac{1}{\hat{\rho}(1)}\hat{\theta} + 1 = 0 \tag{21}$$

$$\hat{\sigma}^2 = \frac{\hat{\gamma}(0)}{1+\hat{\theta}^2} \tag{22}$$

5.7.2 Parameter Estimation VI

For example, from an MA(1) process sample we obtain $\hat{\gamma}(0) = 4.6$, and $\hat{\rho}(1) = 0.488$, so we have equations

$$\hat{\sigma}^2 = \frac{4.6}{1+\hat{\theta}^2} \qquad \hat{\theta}^2 + \frac{1}{0.488}\hat{\theta} + 1 = 0$$

We solve for $\hat{\theta} = 0.8$ and $\hat{\theta} = -1.25$, which one? Thus $\hat{\sigma}^2 =$?

5.7.3 Diagnostic Checking

The main idea is to examine residuals

$$\hat{\epsilon}_t = y_t - (\hat{\delta} + \sum_{i=1}^{p} \hat{\phi}_i y_{t-i} - \sum_{i=1}^{q} \hat{\theta}_i \hat{\epsilon}_{t-i}) \tag{23}$$

in various ways.

If the fitted model is the appropriate form, the residuals should behave similar to a white noise process. Then its sample ACF should not differ from zero significant for all lags greater than one. This is called the autocorrelation check.

We also want to check normality.

5.7.3 Autocorrelation Checking

1. To check the sample ACF values individually, we simply use $\pm 2/\sqrt{T}$ as the limits to judge significance.
2. To check multiple sample ACF values together, say the ACF values $\hat{\rho}_\epsilon(i)$ for lags $i = 1, 2, ...K$. The test statistics is

$$Q_k = (T - d) \sum_{i=1}^{K} \hat{\rho}_\epsilon(i)^2 \tag{24}$$

which is approximately a chi-square distribution with degrees of freedom $K - p - q$, i.e.

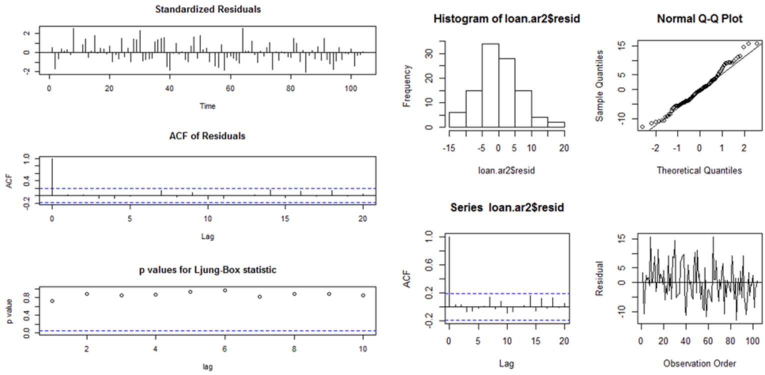
$$Q_k \sim \chi^2_{K-p-q} \tag{25}$$

if the model is appropriate.

Thus at α level, we would reject the hypothesis of model adequacy if Q exceeds the $1 - \alpha$ upper critical value of χ^2_{K-p-q}. This test is called **Ljung-Box-Pierce** test.

5.7.3 Analysis of Residuals for Loan Example

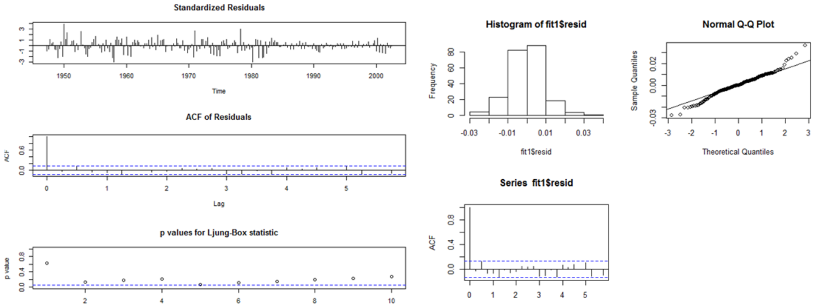
The analysis of the residual is conducted as follows:



The plots of analysis of residuals do not show any problem.

5.7.3 Analysis of Residuals for GNP Example

We only analysis one of them, say the AR(1) model. The analysis of the residual is conducted as follows:



Everything seems fine except for the fact that a distribution with heavier tails than the normal distribution should be employed.

For an ARIMA(p,d,q) process, at τ periods in the future, i.e. time $T+\tau$

$$y_{T+\tau} = \delta + \sum_{i=1}^{p+d} \phi_i y_{T+\tau-i} + \epsilon_{T+\tau} - \sum_{i=1}^{q} \theta_i \epsilon_{T+\tau-i}. \tag{26}$$

So for prediction of $y_{T+\tau}$, $\hat{y}_{T+\tau}(T)$, it requires the "best guess" for some past y's, and current and future y's, as well as the "best guess" for some past ϵ's, and the current and future ϵ's.

In particular,

1. $i \geq \tau$, then $y_{T+\tau-i}$ is a past or current observation, so the "best guess" should be the observations themselves;
2. $i < \tau$, then $y_{T+\tau-i}$ is a future observation, so the "best guess" should be the forecasts for them.

5.8.Forecast II

1. $i \geq \tau$, then $\epsilon_{T+\tau-i}$ is a past or current error, so the "best guess" should be the observed forecast errors;
2. $i < \tau$, then $\epsilon_{T+\tau-i}$ is a future error, so the "best guess" should be 0.

5.8.Forecast III

Therefore, the computation of forecasts should be as follows:

- For $\tau > q$,

$$\hat{y}_{T+\tau}(T) = \delta + \sum_{i=1}^{p+d} \phi_i \hat{y}_{T+\tau-i}(T). \tag{27}$$

- For $\tau = 1, 2, 3...q$,

$$\hat{y}_{T+\tau}(T) = \delta + \sum_{i=1}^{p+d} \phi_i \hat{y}_{T+\tau-i}(T) - \sum_{i=\tau}^{q} \theta_i \epsilon_{T+\tau-i} \tag{28}$$

where for both cases,

$$\hat{y}_{T+\tau-i}(T) = \begin{cases} \hat{y}_{T+\tau-i}(T) \ i.e. \text{ forecast,} & \text{if } i < \tau \\ y_{T+\tau-i} \ i.e. \text{ observation,} & \text{if } i \geq \tau \end{cases} \tag{29}$$

To get the prediction interval, we need standard deviation the τ-step-ahead forecast error. Suppose the infinite MA representation of the model is

$$y_{T+\tau} = \mu + \sum_{i=0}^{\infty} \Psi_i \epsilon_{T+\tau-i} \tag{30}$$

$$\hat{y}_{T+\tau} = \mu + \sum_{i=\tau}^{\infty} \Psi_i \epsilon_{T+\tau-i} \tag{31}$$

$$e_T(\tau) = y_{T+\tau} - \hat{y}_{T+\tau} = \sum_{i=0}^{\tau-1} \Psi_i \epsilon_{T+\tau-i} \tag{32}$$

then

$$Var[e_T(\tau)] = \sigma^2 \sum_{i=0}^{\tau-1} \Psi_i^2 \tag{33}$$

which gets bigger as lead time τ increases. This intuitively makes sense as we should expect wider(more uncertainty) Prediction Intervals in our forecast further into the future.

5.8 Forecasting ARIMA Processes I

Infinite MA representation of $y_{T+\tau}$ can be written as

$$
\begin{aligned}
y_{T+\tau} &= \sum_{i=0}^{\infty} \Psi_i \epsilon_{T+\tau-i} && (34)\\
&= \Psi_0 \epsilon_{T+\tau} + \Psi_1 \epsilon_{T+\tau-1} + \Psi_2 \epsilon_{T+\tau-2} + \cdots && (35)\\
&= (\Psi_0 + \Psi_1 B + \Psi_2 B^2 + \Psi_3 B^3 + \cdots)\epsilon_{T+\tau} && (36)
\end{aligned}
$$

An ARIMA (p,d,q) can be written as

$$\Phi(B)(1-B)^d y_{T+\tau} = \Theta(B)\epsilon_{T+\tau} \tag{37}$$

Plugging (36) into (37), we have

$$
\begin{aligned}
&(\Psi_0 + \Psi_1 B + \Psi_2 B^2 + \cdots)(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p)(1-B)^d \\
&= (1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q)
\end{aligned} \tag{38}
$$

Then the Ψ weights can be obtained by equating like powers of B in (38)

$$(\Psi_0 + \Psi_1 B + \Psi_2 B^2 + \cdots)(1 - \phi_1 B) = (1 - \theta B) \tag{39}$$

Equating like power of B, we find that

$$\begin{cases} B^0 : \Psi_0 = 1, \\ B^1 : \Psi_1 - \phi\Psi_0 = -\theta, \quad \text{or} \quad \Psi_1 = \phi - \theta \\ B^2 : \Psi_2 - \phi\Psi_1 = 0, \quad \text{or} \quad \Psi_2 = \phi\Psi_1 = \phi(\phi - \theta) \\ B^3 : \Psi_3 - \phi\Psi_2 = 0, \quad \text{or} \quad \Psi_3 = \phi\Psi_2 = \phi^2(\phi - \theta) \\ \vdots \end{cases}$$

In general, we can show for the ARMA(1,1) model that

$$\Psi_j = \phi^{j-1}(\phi - \theta). \tag{40}$$

5.8 Example: AR(2)

For the AR(2) model, the product of the required polynomials is

$$(\Psi_0 + \Psi_1 B + \Psi_2 B^2 + \Psi_3 B^3 \cdots)(1 - \phi_1 B - \phi_2 B^2) = 1 \qquad (41)$$

Equating like power of B, we find that

$$\begin{cases} B^0 : \Psi_0 = 1, \\ B^1 : \Psi_1 - \phi_1 \Psi_0 = 0, \quad \text{or} \quad \Psi_1 = \phi_1 \Psi_0 = \phi_1 \\ B^2 : \Psi_2 - \phi_1 \Psi_1 - \phi_2 \Psi_0 = 0, \quad \text{or} \quad \Psi_2 = \phi_1 \Psi_1 + \phi_2 \Psi_0 = \phi_1^2 + \phi_2 \\ B^3 : \Psi_3 - \phi_1 \Psi_2 - \phi_2 \Psi_1 = 0, \quad \text{or} \quad \Psi_3 = \phi_1 \Psi_2 + \phi_2 \Psi_1 \\ \vdots \end{cases}$$

In general, we can show for the AR(2) model that

$$\Psi_j = \phi_1 \Psi^{j-1} + \phi_2 \Psi_{j-2}. \qquad (42)$$

5.8 Example ARIMA(0,1,1) or IMA(1,1) I

For the ARIMA(0,1,1) model, the product of the required polynomials is

$$(\Psi_0 + \Psi_1 B + \Psi_2 B^2 + \Psi_3 B^3 \cdots)(1 - B) = (1 - \theta B) \tag{43}$$

Equating like power of B, we find that

$$\begin{cases} B^0 : \Psi_0 = 1, \\ B^1 : \Psi_1 - \Psi_0 = -\theta \quad \text{or} \quad \Psi_1 = \Psi_0 - \theta = 1 - \theta \\ B^2 : \Psi_2 - \Psi_1 = 0, \quad \text{or} \quad \Psi_2 = \Psi_1 = 1 - \theta \\ B^3 : \Psi_3 - \Psi_2 = 0, \quad \text{or} \quad \Psi_3 = \Psi_2 = 1 - \theta \\ \vdots \end{cases}$$

In general, we can show for the ARIMA(0,1,1) or IMA(1,1) model that

$$\Psi_j = \Psi_{j-1} = 1 - \theta, j = 2, 3, \cdots . \tag{44}$$

5.8.Exercise: AR(1)

What are Ψ_i weights in AR(1) model?

5.8.Forecast AR(1) I

Consider the AR(1) model

$$y_t = \phi y_{t-1} + \delta + \epsilon \tag{45}$$

Suppose we are currently at time T, then

$$\hat{y}_{T+1}(T) = \phi y_T + \delta \tag{46}$$
$$\hat{y}_{T+2}(T) = \phi \hat{y}_{T+1}(T) + \delta \tag{47}$$
$$\vdots \tag{48}$$
$$\hat{y}_{T+\tau}(T) = \phi \hat{y}_{T+\tau-1}(T) + \delta \tag{49}$$

We also know that the infinite MA representation coefficients $\Psi_i = \phi^i$, thus

$$Var[e_T(\tau)] = \sigma^2 \sum_{i=0}^{\tau-1} \Psi_i^2 = \sigma^2 \sum_{i=0}^{\tau-1} \phi^{2i} = \sigma^2 \frac{1-\phi^{2\tau}}{1-\phi^2} \tag{50}$$

5.8.Forecast AR(1) II

Thus the $100(1-\alpha)$ PI for $y_{T+\tau}$ is

$$\hat{y}_{T+\tau}(T) \pm Z_{\alpha/2} * \sqrt{Var[e_T(\tau)]} \tag{51}$$

$$\hat{y}_{T+\tau}(T) \pm Z_{\alpha/2} * \sigma \sqrt{\frac{1-\phi^{2\tau}}{1-\phi^2}} \tag{52}$$

5.8 Example: Prediction Interval of ARIMA(0,1,1) or IMA(1,1) I

Suppose we are currently at time T, then

$$\begin{aligned}
y_{T+\tau}(T) &= y_{T+\tau-1}(T) + \delta + \epsilon_{T+\tau} - \theta\epsilon_{T+\tau-1} \\
\hat{y}_{T+1}(T) &= y_T + \delta - \theta\epsilon_T \\
\hat{y}_{T+2}(T) &= \hat{y}_{T+1}(T) + \delta \\
&\vdots \\
\hat{y}_{T+\tau}(T) &= \hat{y}_{T+\tau-1}(T) + \delta
\end{aligned} \tag{53}$$

For ARIMA(0,1,1) or IMA(1,1), we know that $\Psi_0 = 1$ and $\Psi_i = 1 - \theta$, thus

$$Var[e_T(\tau)] = \sigma^2 \sum_{i=0}^{\tau-1} \Psi_i^2 = \sigma^2[1 + (\tau-1)(1-\theta)^2] \tag{54}$$

5.8 Example: Prediction Interval of ARIMA(0,1,1) or IMA(1,1) II

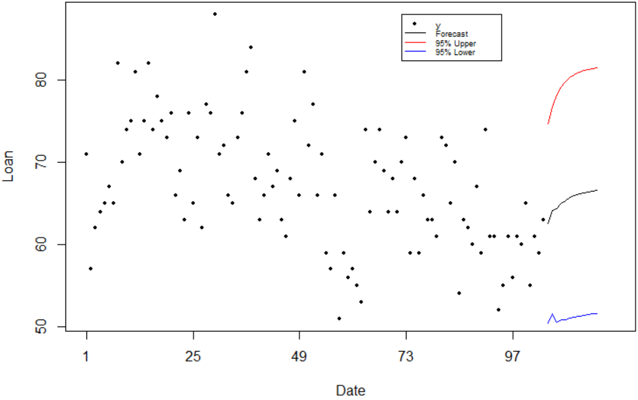
Thus the $100(1-\alpha)$ PI for $y_{T+\tau}$ is

$$\hat{y}_{T+\tau}(T) \pm Z_{\alpha/2} * \sqrt{Var[e_T(\tau)]} \tag{55}$$

$$\hat{y}_{T+\tau}(T) \pm Z_{\alpha/2} * \sigma\sqrt{1+(\tau-1)(1-\theta)^2} \tag{56}$$

5.8.Forecast Example: loan applications

Consider the forecast for the loan application example. Assume that we want to make forecasts for next 12 weeks using AR(2) model. The forecasts with the 95% prediction limits are shown in picture.



5.9 Seasonal ARIMA Models (SARIMA) I

Some time series exhibit seasonal behaviors with a period of s. As we discussed before, a first approach is to apply a seasonal difference

$$y_t - y_{t-s} = (1 - \mathbf{B}^s)y_t = w_t \tag{57}$$

so that w_t forms a stationary series. Sometimes, you may need seasonal differences more than once to get stationarity. Then it is denoted by $(1 - \mathbf{B}^s)^D$.

Although w_t is stationary, it may not eliminate all seasonal features in the profess. That is, it would likely have special seasonal features in its ACF structure. In other words, the seasonally differenced data may still show strong autocorrelation at lags s, 2s, $\cdots$ So we introduce the seasonal ARMA model.

5.9 Seasonal ARIMA Models (SARIMA) II

The "pure" seasonal ARMA model, which takes the form

$$\Phi(\mathbf{B}^s)w_t = \delta + \Theta(\mathbf{B}^s)\epsilon_t. \tag{58}$$

The operators

$$\Phi(\mathbf{B}^s) = 1 - \Phi_1\mathbf{B}^s - \Phi_2\mathbf{B}^{2s} - ... - \Phi_P\mathbf{B}^{Ps} \tag{59}$$

and

$$\Theta(\mathbf{B}^s) = 1 + \Theta_1\mathbf{B}^s + \Theta_2\mathbf{B}^{2s} + ... + \Theta_Q\mathbf{B}^{Qs} \tag{60}$$

are called the seasonal autoregressive operator and seasonal moving average operator of orders P and Q, respectively.

5.9 Seasonal ARIMA Models (SARIMA) III

Similar to regular ARMA models, we have

Model	ACF	PACF
Seasonal AR(P)	Decay at lags ks k=1,2,3...	Cut off after lag Ps
Seasonal MA(Q)	Cut off after lag Qs	Decay at lags ks, k=1,2,3...
Seasonal ARMA(P,Q)	Decay at lags ks k=1,2,3...	Decay at lags ks k=1,2,3...

Note the all values at nonseasonal lags are zero.

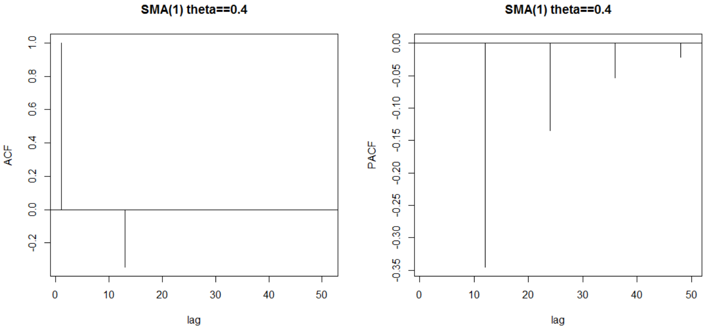
5.9 Example–Seasonal MA(1)

The seasonal MA(1) model with season length s=12 has the form:

$$w_t = \epsilon_t - \theta\epsilon_{t-12} = (1 - \theta \mathbf{B}^{12})\epsilon_t \tag{61}$$

The only nonzero ACF value, aside from lag zero is (See page 335,(5.13)),

$$\rho(12) = \frac{-\theta}{1+\theta^2} \tag{62}$$

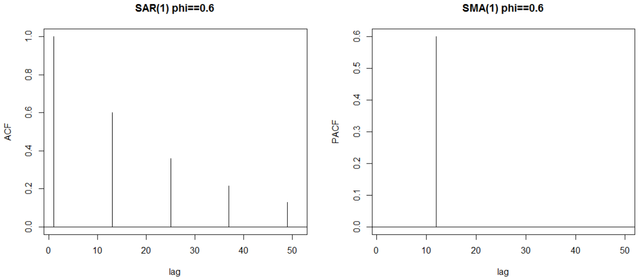


The seasonal AR(1) model with s=12 has the form:

$$w_t = \phi w_{t-s} + \epsilon_t \quad or \quad (1 - \phi \mathbf{B}^s) w_t = \epsilon_t \tag{63}$$

Then non-zero ACF values are at lag$= 12k$, (See page 340, (5.26))

$$\rho(12k) = \phi^k \tag{64}$$



5.9 ULTIMATE Form I

Recall w_t was obtained by seasonal differencing y_t if needed.
$y_t - y_{t-s} = (1 - \mathbf{B}^s)^D y_t = w_t$, so we have the pure SARIMA model

$$\Phi(\mathbf{B}^s)(1 - \mathbf{B}^s)^D y_t = \delta + \Theta(\mathbf{B}^s)\epsilon_t. \tag{65}$$

which is called a pure SARIMA (P,D,Q) model.

It is likely that the series has both seasonal behaviors and non-seasonal behaviors at the same time, to the ULTIMATE model is

$$\phi(\mathbf{B})\Phi(\mathbf{B}^s)(1 - \mathbf{B})^d(1 - \mathbf{B}^s)^D y_t = \delta + \theta(\mathbf{B})\Theta(\mathbf{B}^s)\epsilon_t. \tag{66}$$

or

5.9 ULTIMATE Form II

$$(1-\phi_1\mathbf{B} - ... - \phi_p\mathbf{B}^p)(1-\Phi_1\mathbf{B}^s - ... - \Phi_P\mathbf{B}^{Ps})(1-\mathbf{B})^d(1-\mathbf{B}^s)^D y_t \tag{67}$$

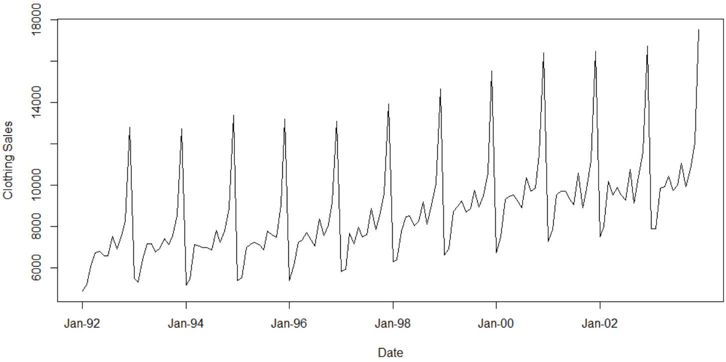
$$=\delta + (1-\theta_1\mathbf{B} - ... - \theta_q\mathbf{B}^q)(1-\Theta_1\mathbf{B}^s - ... - \Theta_Q\mathbf{B}^{Qs})\epsilon_t \tag{68}$$

which is called a seasonal ARIMA model of orders $(p, d, q) \times (P, D, Q)$.

In practice, although it is case specific, it is not expected to have P, D and Q greater than 1.

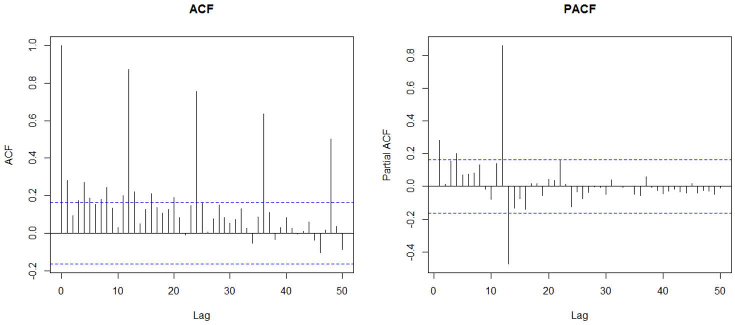
The forecast by using SARIMA$(p, d, q) \times (P, D, Q)$ is similar to the case of a nonseasonal ARIMA model. The estimate for the variance of the forecast errors as well as the prediction intervals follow the same idea, but is more complicated.

Reconsider the clothing sales example in chapter 4. The data exhibit seasonality with $s = 12$.



5.9 Example–SARIMA$(0, 1, 1) \times (0, 1, 1)_{12}$ II

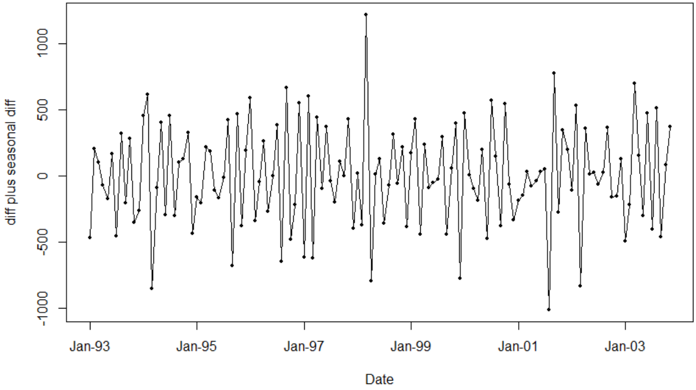
Sample ACF and PACF are plotted too.



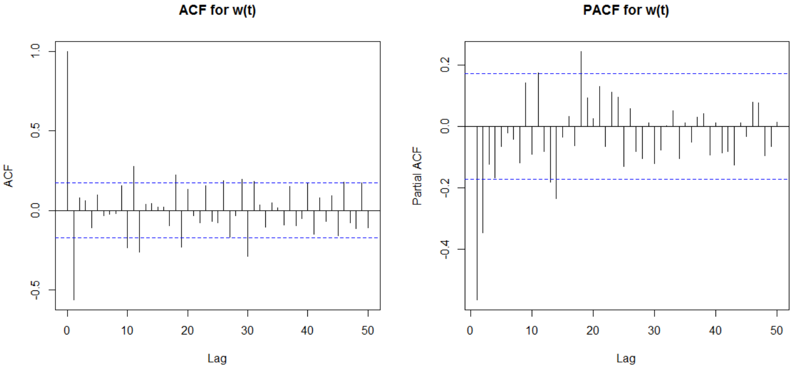
ACF values at lags 12, 24, 36 are significant and slowly decreasing. Other ACF values are also slowly decreasing. PACF value at 12 is huge. These suggest a seasonal difference and a difference may be considered. That is $w_t = (1 - \mathbf{B})(1 - \mathbf{B}_{12})y_t$.

5.9 Example–SARIMA$(0, 1, 1) \times (0, 1, 1)_{12}$ III

The plot of w_t along with its ACF and PACF are shown below.



5.9 Example–SARIMA$(0, 1, 1) \times (0, 1, 1)_{12}$ IV



5.9 Example–SARIMA$(0,1,1) \times (0,1,1)_{12}$ V

We can interpret the ACF plot as cutting off after 1, and PACF plot as exponential decay. So we propose MA(1) model for w_t. At this time, we know the model looks like $(P, 1, Q) \times (0, 1, 1)$.

The remaining seasonlity is a bit difficult to interpret. Again check the w_t's ACF and PACF at lags being multiples of $s = 12$: we interpret them as

- ACF at lag 12 seems to be significant.
- PACF at lag 12, 24,36 seems to be alternating in sign.

MA(1) model for seasonality is proposed. Hence $P = 0$, and $Q = 1$. So finally, the model is SARIMA $(0,1,1) \times (0,1,1)$.

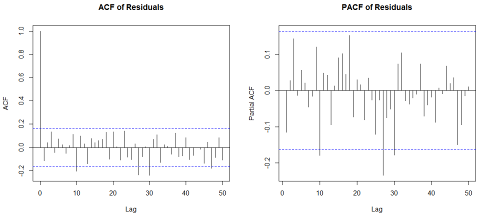
The estimates of the coefficients by R are

```
Coefficients:
          ma1     sma1
      -0.7745  -0.4319
s.e.   0.0517   0.0914
```

So the fitted model is

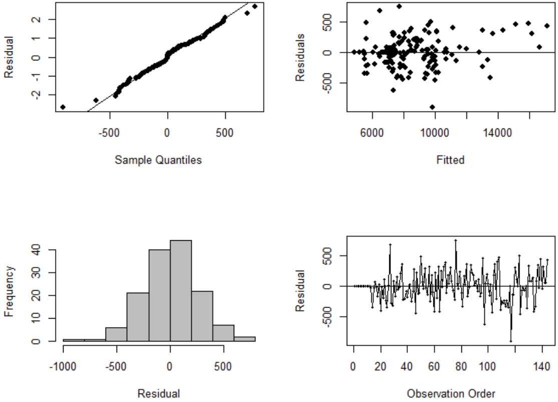
$$(1 - \mathbf{B})(1 - \mathbf{B}^{12})y_t = (1 - 0.7745\mathbf{B})(1 - 0.4319\mathbf{B}^{12})\epsilon_t \tag{69}$$

The ACF and PACF plots for residuals are



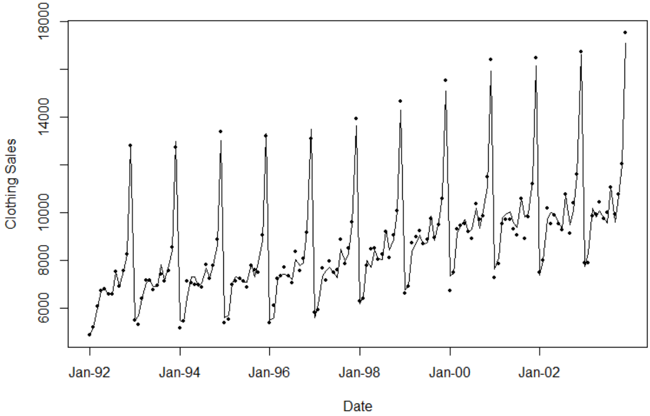
which confirm a white noise process.

Analysis of residuals does not show any problem.



5.9 Example–SARIMA$(0,1,1) \times (0,1,1)_{12}$ VIII

The plot of the fitted values shows a satisfactory fit.



The plot of the forecast for next year is