QBUS 6840 Lecture 9

Seasonal ARIMA Models
Model Combination

Professor Junbin Gao

The University of Sydney Business School

Review of $ARMA(p, q)$ and $ARIMA(p, d, q)$ Processes

- $ARMA(p, q)$ Formulation with backshift operators

$$\left(1 - \sum_{i=1}^{p} \phi_i B^i\right) Y_t = c + \left(1 + \sum_{i=1}^{q} \theta_i B^i\right) \varepsilon_t,$$

- $ARIMA(p, d, q)$ Formulation with backshift operators

$$\left(1 - \sum_{i=1}^{p} \phi_i B^i\right) (1 - B)^d Y_t = c + \left(1 + \sum_{i=1}^{q} \theta_i B^i\right) \varepsilon_t$$

Review of $ARMA(p, q)$ and $ARIMA(p, d, q)$ Processes

- $ARMA(p, q)$ Formulation with backshift operators

$$\left(1 - \sum_{i=1}^{p} \phi_i B^i\right) Y_t = c + \left(1 + \sum_{i=1}^{q} \theta_i B^i\right) \varepsilon_t,$$

- $ARIMA(p, d, q)$ Formulation with backshift operators

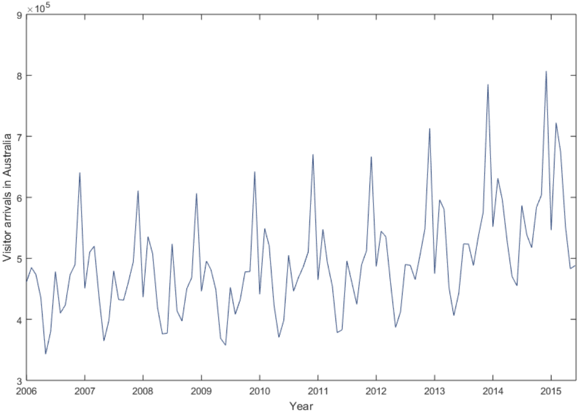
$$\left(1 - \sum_{i=1}^{p} \phi_i B^i\right) \quad Z_t \quad = c + \left(1 + \sum_{i=1}^{q} \theta_i B^i\right) \varepsilon_t,$$

- Let $Z_t = (1 - B)^d Y_t$, then Z_t is the d-order differencing of Y_t. Hence $ARMA(p, q)$ of Z_t is the $ARMA(p, d, q)$ of Y_t

Procedure to Estimate $ARMA(p,q)/ARIMA(p,d,q)$ processes: `Lecture08_Example04.py`

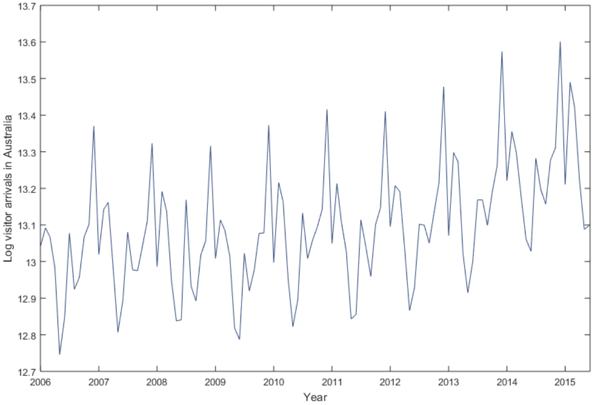
1. For the given time series $\{Y_t\}$, check its stationarity by looking at its Sample ACF and Sample PACF.
2. If ACF does not die down quickly, which means the given time series $\{Y_t\}$ is nonstationary, we seek for a transformation, e.g., log transformation $\{Z_t = log(Y_t)\}$, or the first order difference $\{Z_t = Y_t - Y_{t-1}\}$, or even the difference of log time series, or the difference of the first order difference, so that the transformed time series is stationary by checking its Sample ACF
3. When both Sample ACF and Sample PACF die down quickly, check the orders at which ACF or PACF die down. The order of ACF will be the lag q of the ARIMA and the order of PACF will be the lag p of the ARIMA, and the order of difference will be d.
4. Estimate the identified $ARIMA(p,d,q)$, or $ARMA(p,q)$ (if we did not do any difference transformation)
5. Make forecast with estimated $ARIMA(p,d,q)$, or $ARMA(p,q)$ model

Seasonal ARIMA models



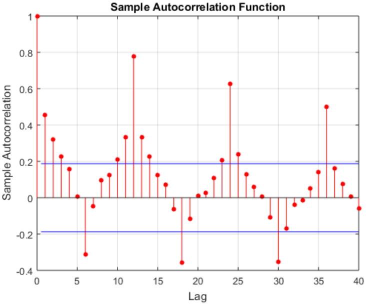
Seasonal ARIMA models

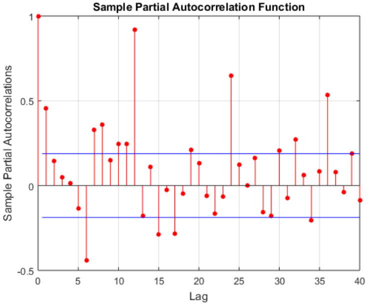
Variance stabilising transform



This is the Log transformed data.

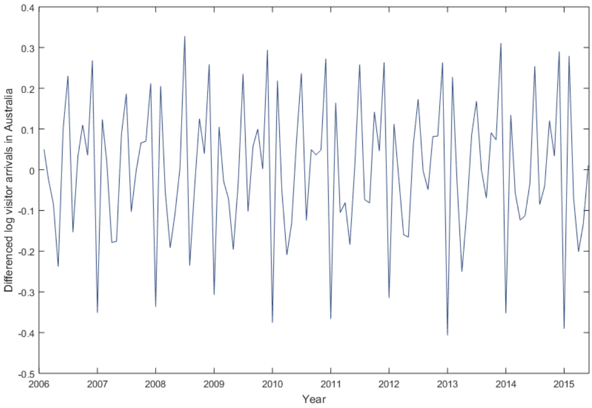
ACF and PACF for the log visitors series





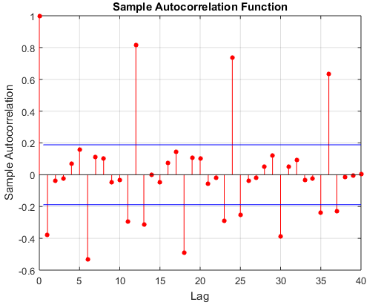
Seasonal ARIMA models

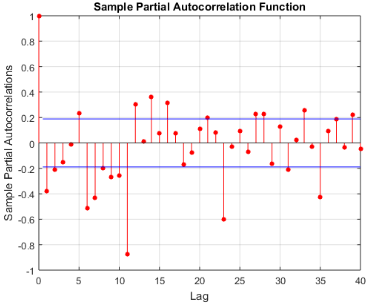
First differenced log visitors series



Seasonal ARIMA models

ACF and PACF for the first differenced log visitors series



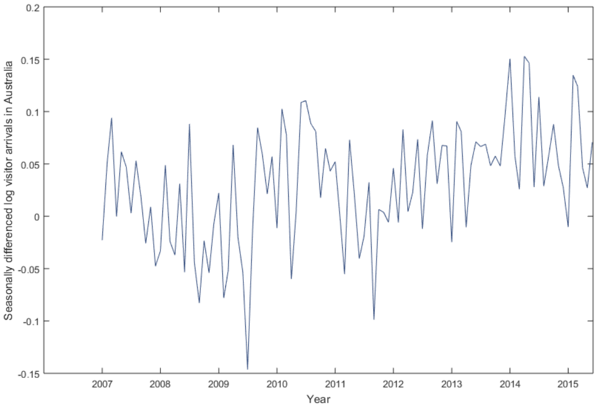


We can use seasonal differencing to remove the nonstationarity caused by the seasonality:

$$y_t - y_{t-12}$$

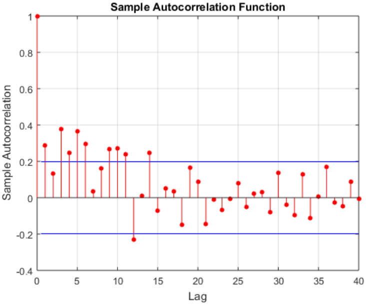
Seasonal ARIMA models

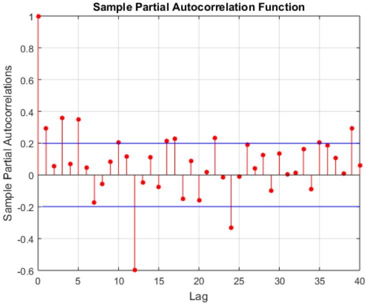
Seasonally differenced log visitors series



Seasonal ARIMA models

ACF and PACF for the seasonally differenced log visitors series





Seasonal AR model

$AR(1)$ specification

$$Y_t = c + \Phi_1 Y_{t-12} + \varepsilon_t$$

Seasonal AR model

$AR(1)$ specification

$$Y_t = c + \Phi_1 Y_{t-12} + \varepsilon_t$$

In the form of B operator

$$(1 - \Phi_1 B^{12}) Y_t = c + \varepsilon_t$$

Seasonal AR model

$AR(1)$ specification

$$Y_t = c + \Phi_1 Y_{t-12} + \varepsilon_t$$

In the form of B operator

$$(1 - \Phi_1 B^{12}) Y_t = c + \varepsilon_t$$

Considering the Seasonal Differencing series

$$Z_t = Y_t - Y_{t-12}$$

what is Z_t's Seasonal AR(1)?

$$Y_t = c + \Phi_1 Y_{t-12} + \varepsilon_t$$

In the form of B operator

$$(1 - \Phi_1 B^{12}) Y_t = c + \varepsilon_t$$

Considering the Seasonal Differencing series

$$Z_t = Y_t - Y_{t-12}$$

what is Z_t's Seasonal AR(1)?

$$Z_t = c + \Phi_1 Z_{t-12} + \varepsilon_t$$

i.e., for Y_t,

$$Y_t - Y_{t-12} = c + \Phi_1 (Y_{t-12} - Y_{t-24}) + \varepsilon_t$$

$$Y_t = c + \Phi_1 Y_{t-12} + \varepsilon_t$$

In the form of B operator

$$(1 - \Phi_1 B^{12}) Y_t = c + \varepsilon_t$$

Considering the Seasonal Differencing series

$$Z_t = Y_t - Y_{t-12}$$

what is Z_t's Seasonal AR(1)?

$$Z_t = c + \Phi_1 Z_{t-12} + \varepsilon_t$$

i.e., for Y_t,

$$Y_t - Y_{t-12} = c + \Phi_1 (Y_{t-12} - Y_{t-24}) + \varepsilon_t$$

In the form of B operator

$$(1 - \Phi_1 B^{12})(1 - B^{12}) Y_t = c + \varepsilon_t$$

Seasonal MA model

$MA(1)$ specification

$$Y_t = c + \Theta_1 \varepsilon_{t-12} + \varepsilon_t$$

Seasonal MA model

$MA(1)$ specification

$$Y_t = c + \Theta_1 \varepsilon_{t-12} + \varepsilon_t$$

In the form of B operator

$$Y_t = c + (1 + \Theta_1 B^{12})\varepsilon_t$$

Seasonal MA model

MA(1) specification

$$Y_t = c + \Theta_1 \varepsilon_{t-12} + \varepsilon_t$$

In the form of B operator

$$Y_t = c + (1 + \Theta_1 B^{12})\varepsilon_t$$

Considering seasonally differencing $Z_t = Y_t - Y_{t-12}$, its seasonal MA(1) is

$$Z_t = c + \Theta_1 \varepsilon_{t-12} + \varepsilon_t$$

Seasonal MA model

MA(1) specification

$$Y_t = c + \Theta_1 \varepsilon_{t-12} + \varepsilon_t$$

In the form of B operator

$$Y_t = c + (1 + \Theta_1 B^{12})\varepsilon_t$$

Considering seasonally differencing $Z_t = Y_t - Y_{t-12}$, its seasonal MA(1) is

$$Z_t = c + \Theta_1 \varepsilon_{t-12} + \varepsilon_t$$

Hence in terms of Y_t, it becomes

$$Y_t - Y_{t-12} = c + \Theta_1 \varepsilon_{t-12} + \varepsilon_t$$

Seasonal MA model

MA(1) specification

$$Y_t = c + \Theta_1 \varepsilon_{t-12} + \varepsilon_t$$

In the form of B operator

$$Y_t = c + (1 + \Theta_1 B^{12})\varepsilon_t$$

Considering seasonally differencing $Z_t = Y_t - Y_{t-12}$, its seasonal MA(1) is

$$Z_t = c + \Theta_1 \varepsilon_{t-12} + \varepsilon_t$$

Hence in terms of Y_t, it becomes

$$Y_t - Y_{t-12} = c + \Theta_1 \varepsilon_{t-12} + \varepsilon_t$$

In the form of B operator

$$(1 - B^{12})Y_t = c + (1 + \Theta_1 B^{12})\varepsilon_t$$

$$\text{ARIMA} \underbrace{(p,d,q)}_{\substack{\uparrow \\ \left(\begin{array}{l}\text{Non-seasonal part}\\ \text{of the model}\end{array}\right)}} \underbrace{(P,D,Q)_m}_{\substack{\uparrow \\ \left(\begin{array}{l}\text{Seasonal part}\\ \text{of the model}\end{array}\right)}}$$

where m = number of seasonal period (e.g. $m = 12$).

$$\begin{aligned}
&(1-\phi_1 B-\phi_2 B^2-\cdots-\phi_p B^p)(1-\Phi_1 B^{m}-\Phi_2 B^{2m}-\cdots-\Phi_P B^{Pm})(1-B)^d(1-B^{m})^D Y_t\\
=&c+(1+\theta_1 B+\theta_2 B^2+\cdots+\theta_q B^q)(1+\Theta_1 B^{m}+\Theta_2 B^{2m}+\cdots+\Theta_Q B^{Qm})\epsilon_t
\end{aligned}$$

Seasonal Box-Jenkins models: Example

$$\underbrace{(1-\phi_1 B)}_{\begin{pmatrix}\text{Non-seasonal}\\ \text{AR}(1)\end{pmatrix}}\ \underbrace{(1-\Phi_1 B^4)}_{\begin{pmatrix}\text{Seasonal}\\ \text{AR}(1)\end{pmatrix}}\ \underbrace{(1-B)}_{\begin{pmatrix}\text{Non-seasonal}\\ \text{difference}\end{pmatrix}}\ \underbrace{(1-B^4)}_{\begin{pmatrix}\text{Seasonal}\\ \text{difference}\end{pmatrix}}y_t \;=\; \underbrace{(1+\theta_1 B)}_{\begin{pmatrix}\text{Non-seasonal}\\ \text{MA}(1)\end{pmatrix}}\ \underbrace{(1+\Theta_1 B^4)}_{\begin{pmatrix}\text{Seasonal}\\ \text{MA}(1)\end{pmatrix}}e_t.$$

The above is an $ARIMA(1,1,1)(1,1,1)_4$ model (with $c = 0$)

$ARIMA(1,0,0)(0,1,1)_{12}$ for monthly data:

$$(1-\phi_1 B)(1-B^{12})Y_t = c + (1+\Theta_1 B^{12})\varepsilon_t$$

This is equivalent to

$$Y_t - Y_{t-12} = c + \phi_1(Y_{t-1} - Y_{t-13}) + \varepsilon_t + \Theta_1\varepsilon_{t-12}$$

Factored:

$$(1 - \phi_1 B)(1 - \Phi_1 B^{12}) Y_t = c + \varepsilon_t$$

Or write it out

$$Y_t = c + \phi_1 Y_{t-1} + \Phi_1 Y_{t-12} - \phi_1 \Phi_1 Y_{t-13} + \varepsilon_t$$

Factored:

$$(1-\phi_1 B)(1-\Phi_1 B^{12})Y_t = c + \varepsilon_t$$

Or write it out

$$Y_t = c + \phi_1 Y_{t-1} + \Phi_1 Y_{t-12} - \phi_1 \Phi_1 Y_{t-13} + \varepsilon_t$$

This is because informally

$$(1-\phi_1 B)(1-\Phi_1 B^{12}) = 1 - \phi_1 B - \Phi_1 B^{12} + \phi_1 \Phi_1 B^{13}$$

Hence

$$(1-\phi_1 B)(1-\Phi_1 B^{12})Y_t = Y_t - \phi_1 Y_{t-1} - \Phi_1 Y_{t-12} + \phi_1 \Phi_1 Y_{t-13}$$

$ARIMA(1, 1, 1)(1, 1, 0)_{12}$ model:

$$(1 - \phi_1 B)(1 - \Phi_1 B^{12})(1 - B)(1 - B^{12})Y_t = c + (1 + \theta_1 B)\varepsilon_t$$

ARIMA$(1, 1, 1)(1, 1, 0)_{12}$ model:

$$(1 - \phi_1 B)(1 - \Phi_1 B^{12})(1 - B)(1 - B^{12})Y_t = c + (1 + \theta_1 B)\varepsilon_t$$

First denote $Z_t = (1 - B)(1 - B^{12})Y_t$. For this new time series Z_t, the model is

$$(1 - \phi_1 B)(1 - \Phi_1 B^{12})Z_t = c + (1 + \theta_1 B)\varepsilon_t$$

$ARIMA(1,1,1)(1,1,0)_{12}$ model:

$$(1-\phi_1 B)(1-\Phi_1 B^{12})(1-B)(1-B^{12})Y_t = c + (1+\theta_1 B)\varepsilon_t$$

First denote $Z_t = (1-B)(1-B^{12})Y_t$. For this new time series Z_t, the model is

$$(1-\phi_1 B)(1-\Phi_1 B^{12})Z_t = c + (1+\theta_1 B)\varepsilon_t$$

or

$$(1-\phi_1 B - \Phi_1 B^{12} + \phi_1 \Phi_1 B^{13})Z_t = c + (1+\theta_1 B)\varepsilon_t$$

$ARIMA(1, 1, 1)(1, 1, 0)_{12}$ model:

$$(1 - \phi_1 B)(1 - \Phi_1 B^{12})(1 - B)(1 - B^{12})Y_t = c + (1 + \theta_1 B)\varepsilon_t$$

First denote $Z_t = (1 - B)(1 - B^{12})Y_t$. For this new time series Z_t, the model is

$$(1 - \phi_1 B)(1 - \Phi_1 B^{12})Z_t = c + (1 + \theta_1 B)\varepsilon_t$$

or

$$(1 - \phi_1 B - \Phi_1 B^{12} + \phi_1 \Phi_1 B^{13})Z_t = c + (1 + \theta_1 B)\varepsilon_t$$

Hence

$$Z_t = \phi_1 B Z_t + \Phi_1 B^{12} Z_t - \phi_1 \Phi_1 B^{13} Z_t + c + \varepsilon_t + \theta_1 B \varepsilon_t$$

or

$$Z_t = \phi_1 Z_{t-1} + \Phi_1 Z_{t-12} - \phi_1 \Phi_1 Z_{t-13} + c + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

ARIMA$(1, 1, 1)(1, 1, 0)_{12}$ model:

$$(1-\phi_1 B)(1-\Phi_1 B^{12})(1-B)(1-B^{12})Y_t = c + (1+\theta_1 B)\varepsilon_t$$

Finally from

$$Z_t = \phi_1 Z_{t-1} + \Phi_1 Z_{t-12} - \phi_1 \Phi_1 Z_{t-13} + c + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

$ARIMA(1,1,1)(1,1,0)_{12}$ model:

$$(1-\phi_1 B)(1-\Phi_1 B^{12})(1-B)(1-B^{12})Y_t = c + (1+\theta_1 B)\varepsilon_t$$

Finally from

$$Z_t = \phi_1 Z_{t-1} + \Phi_1 Z_{t-12} - \phi_1 \Phi_1 Z_{t-13} + c + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

and

$$\begin{aligned} Z_t =& (1-B)(1-B^{12})Y_t = Y_t - Y_{t-1} - Y_{t-12} + Y_{t-13} \\ =& (Y_t - Y_{t-1}) - (Y_{t-12} - Y_{t-13}) \end{aligned}$$

$ARIMA(1,1,1)(1,1,0)_{12}$ model:

$$(1-\phi_1 B)(1-\Phi_1 B^{12})(1-B)(1-B^{12})Y_t = c + (1+\theta_1 B)\varepsilon_t$$

Finally from

$$Z_t = \phi_1 Z_{t-1} + \Phi_1 Z_{t-12} - \phi_1 \Phi_1 Z_{t-13} + c + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

and

$$\begin{aligned} Z_t =& (1-B)(1-B^{12})Y_t = Y_t - Y_{t-1} - Y_{t-12} + Y_{t-13} \\ =& (Y_t - Y_{t-1}) - (Y_{t-12} - Y_{t-13}) \end{aligned}$$

we

$$\begin{aligned} (Y_t - Y_{t-1}) - (Y_{t-12} - Y_{t-13}) =& c + \phi_1 \left[(Y_{t-1} - Y_{t-2}) - (Y_{t-13} - Y_{t-14})\right] \\ & + \Phi_1 \left[(Y_{t-12} - Y_{t-13}) - (Y_{t-24} - Y_{t-25})\right] \\ & - \phi_1 \Phi_1 \left[(Y_{t-13} - Y_{t-14}) - (Y_{t-25} - Y_{t-26})\right] \\ & + \varepsilon_t + \theta_1 \varepsilon_{t-1} \end{aligned}$$

In our example for the log visitors series, we saw that seasonally differencing is not enough to make the series stationary. We can then consider the transform:

$$(1 - B^{12})(1 - B)Y_t = (Y_t - Y_{t-1}) - (Y_{t-12} - Y_{t-13})$$

In our example for the log visitors series, we saw that seasonally differencing is not enough to make the series stationary. We can then consider the transform:

$$(1 - B^{12})(1 - B)Y_t = (Y_t - Y_{t-1}) - (Y_{t-12} - Y_{t-13})$$

Note that

$$(Y_t - Y_{t-1}) - (Y_{t-12} - Y_{t-13}) = (Y_t - Y_{t-12}) - (Y_{t-1} - Y_{t-13})$$

In our example for the log visitors series, we saw that seasonally differencing is not enough to make the series stationary. We can then consider the transform:

$$(1-B^{12})(1-B)Y_t = (Y_t - Y_{t-1}) - (Y_{t-12} - Y_{t-13})$$

Note that

$$\begin{aligned}(Y_t - Y_{t-1}) - (Y_{t-12} - Y_{t-13}) &= (Y_t - Y_{t-12}) - (Y_{t-1} - Y_{t-13}) \\ &= (1-B^{12})Y_t - (1-B^{12})Y_{t-1}\end{aligned}$$

In our example for the log visitors series, we saw that seasonally differencing is not enough to make the series stationary. We can then consider the transform:

$$(1 - B^{12})(1 - B)Y_t = (Y_t - Y_{t-1}) - (Y_{t-12} - Y_{t-13})$$

Note that

$$\begin{aligned} (Y_t - Y_{t-1}) - (Y_{t-12} - Y_{t-13}) &= (Y_t - Y_{t-12}) - (Y_{t-1} - Y_{t-13}) \\ &= (1 - B^{12})Y_t - (1 - B^{12})Y_{t-1} \\ \text{denote } Z_t = (1 - B^{12})Y_t : \quad &= Z_t - Z_{t-1} = (1 - B)Z_t \end{aligned}$$

In our example for the log visitors series, we saw that seasonally differencing is not enough to make the series stationary. We can then consider the transform:

$$(1-B^{12})(1-B)Y_t = (Y_t - Y_{t-1}) - (Y_{t-12} - Y_{t-13})$$

Note that

$$\begin{aligned}
(Y_t - Y_{t-1}) - (Y_{t-12} - Y_{t-13}) &= (Y_t - Y_{t-12}) - (Y_{t-1} - Y_{t-13}) \\
&= (1-B^{12})Y_t - (1-B^{12})Y_{t-1} \\
\text{denote } Z_t = (1-B^{12})Y_t : \quad &= Z_t - Z_{t-1} = (1-B)Z_t \\
&= (1-B)(1-B^{12})Y_t
\end{aligned}$$

In our example for the log visitors series, we saw that seasonally differencing is not enough to make the series stationary. We can then consider the transform:

$$(1 - B^{12})(1 - B)Y_t = (Y_t - Y_{t-1}) - (Y_{t-12} - Y_{t-13})$$

Note that

$$\begin{aligned}
(Y_t - Y_{t-1}) - (Y_{t-12} - Y_{t-13}) &= (Y_t - Y_{t-12}) - (Y_{t-1} - Y_{t-13}) \\
&= (1 - B^{12})Y_t - (1 - B^{12})Y_{t-1} \\
\text{denote } Z_t = (1 - B^{12})Y_t : \quad &= Z_t - Z_{t-1} = (1 - B)Z_t \\
&= (1 - B)(1 - B^{12})Y_t
\end{aligned}$$

Hence

$$(1 - B^{12})(1 - B)Y_t = (1 - B)(1 - B^{12})Y_t$$

$ARIMA(0, 0, 0)(P, 0, 0)$

- Sample autocorrelations die down for lags m, $2m$, $3m$, etc.
- Sample partial autocorrelations cut off at lag Pm.

$ARIMA(0, 0, 0)(0, 0, Q)$

- Sample autocorrelations cuts off at lag Qm.
- Sample partial autocorrelations die down for lags m, $2m$, $3m$, etc.

$ARIMA(0, 0, 0)(0, 1, 0)$

- Sample autocorrelations and partial autocorrelations die down very slowly for lags m, $2m$, $3m$, etc.

Seasonal ARIMA models

Lecture09_Example01.py

Carefully read the scripts in Lecture09_Example01.py

Carefully read the scripts in `Lecture09_Example01.py`
The concluded model is

$$(1-\phi_1 B-\phi_2 B^2-\phi_3 B^3-\phi_4 B^4-\phi_5 B^5)(1-B^{12})Z_t=\varepsilon_t+\Theta_1 B^{12}\varepsilon_t$$

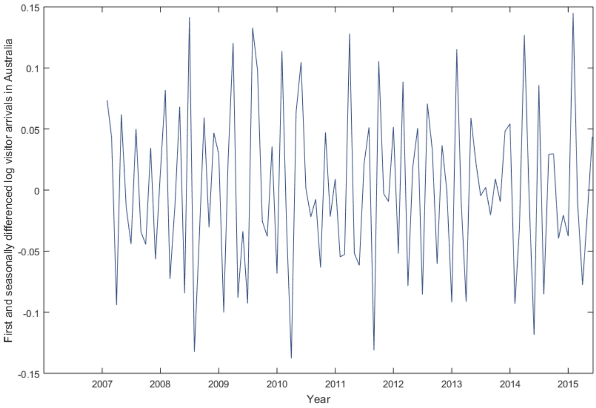
or

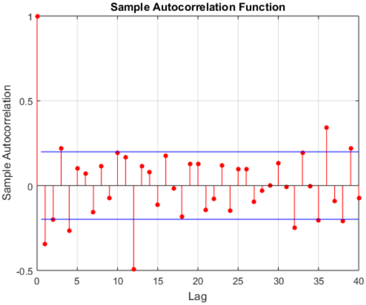
$$(1-\phi_1 B-\phi_3 B^3-\phi_5 B^5)(1-B^{12})Z_t=\varepsilon_t+\Theta_1 B^{12}\varepsilon_t$$

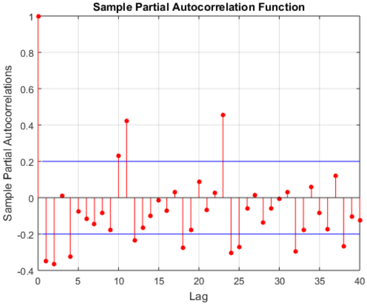
where Z_t is the quartic root data, i.e., $Z_t=Y_t^{1/4}$.

Seasonal ARIMA models

First and seasonal differencing




Sample Autocorrelation Function
Sample Autocorrelation
Lag
1
0.5
0
-0.5
0
5
10
15
20
25
30
35
40


Sample Partial Autocorrelation Function
Sample Partial Autocorrelations
Lag
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
0
5
10
15
20
25
30
35
40

$ARIMA(2,1,2)(0,1,1)_{12}$ model:

$$(1-\phi_1 B-\phi_2 B^2)(1-B)(1-B^{12})Y_t = c+(1+\theta_1 B+\theta_2 B^2)(1+\Theta_1 B^{12})\varepsilon_t$$

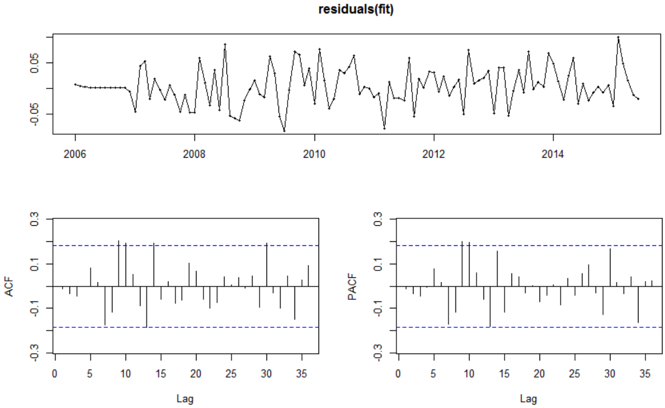
Estimated coefficients (using R):

	ar1	ar2	ma1	ma2	sma1
	-0.7817	-0.3154	-0.0300	-0.4007	-0.7471
s.e.	0.2212	0.1227	0.2213	0.1909	0.1073

log likelihood=178.99, AIC=-345.97, AICc=-345.08, BIC=-330.28.

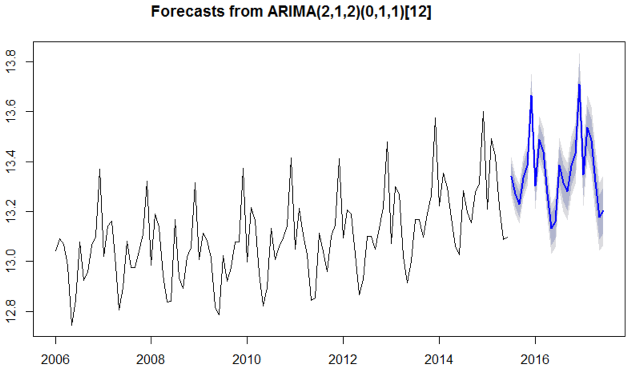
Seasonal ARIMA models

$ARIMA(2,1,2)(0,1,1)_{12}$ model: residuals



Seasonal ARIMA models

$ARIMA(2,1,2)(0,1,1)_{12}$ model: forecasts



- Intercept terms induce permanent trends.
- Only seasonally difference once.
- Usually only either one seasonal AR or MA term is needed.
- Seasonal AR terms are often used when the lag m sample autocorrelation terms are positive.
- Seasonal MA terms are often used when the lag m sample autocorrelation terms are negative.

Classical reference:

Bates, J. M., and C. W. J. Granger (1969). The combination of forecasts, *Operational Research Quarterly*, 20, 451–468.

They provide the following illustration:

TABLE 1. ERRORS IN FORECASTS (ACTUAL LESS ESTIMATED) OF PASSENGER MILES FLOWN, 1953

Month	Brown's exponential smoothing forecast errors	Box–Jenkins adaptive forecasting errors	Combined forecast ($\frac{1}{2}$ Brown + $\frac{1}{2}$ Box–Jenkins) errors
Jan	1	−3	−1
Feb.	6	−10	−2
March	18	24	21
April	18	22	20
May	3	−9	−3
June	−17	−22	−19·5
July	−24	10	−7
Aug.	−16	2	−7
Sept.	−12	−11	−11·5
Oct.	−9	−10	−9·5
Nov.	−12	−12	−12
Dec.	−13	−7	−10
Variance of errors	196	188	150

- It is possible to combine unbiased forecasts $\hat{y}^{(i)}_{T+1|T}$ from models $i = 1, \ldots, m$.
- The models can be various ARIMA type of models or a set of ARIMA models, HW exponential smoothing models and regression models for example.
- $\hat{y}^{(i)}_{T+1|T}$, $i = 1, \ldots, m$ could also be m expert forecasts.

Forecasting combinations

Weights

- The forecasts can be combined as follows

$$\hat{y}^{c}_{T+1|T} = \sum_{i=1}^{m} w_i \; \hat{y}^{(i)}_{T+1|T}$$

- The simplest way is to set $w_i = \frac{1}{m}$, then you are using a simple average.
- Simple averages often work surprisingly well.
- We often use convex combinations, that is $0 \leq w \leq 1$.
- The question is how to combine forecasts "optimally"?

Forecasting combinations

Variance reduction: example of two forecasts

If you have two unbiased forecasts $\hat{y}^{(1)}_{T+1|T}$ and $\hat{y}^{(2)}_{T+1|T}$ with the corresponding variances σ_1^2 and σ_2^2, then we can combine them linearly

$$\hat{y}^{c}_{T+1|T} = w\hat{y}^{(1)}_{T+1|T} + (1-w)\hat{y}^{(2)}_{T+1|T}.$$

The variance of the combined forecast will be

$$\sigma_c^2 = w^2\sigma_1^2 + (1-w)^2\sigma_2^2 + 2\rho w\sigma_1(1-w)\sigma_2.$$

It will have minimum at

$$w = \frac{\sigma_2^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}.$$

where w is the "optimal" value. In case where $\hat{y}^{(1)}_{T+1|T}$ and $\hat{y}^{(2)}_{T+1|T}$ are uncorrelated ($\rho = 0$), then $w = \sigma_2^2/(\sigma_1^2 + \sigma_2^2)$, which is no greater than the smaller of the two individual variances.

- Unfortunately we dont know the actual σ_1^2 and σ_2^2. Now we design a way to estimate them.
- One-step ahead forecasts over a sample $T+1$ observations.
- Dividing the $T+1$ observations into an initial estimation (regression) subsample (e.g., from time 1 to t_0) and a second evaluation (prediction) subsample (from t_0+1 to $T+1$).
- The first subsample enables you to estimate the parameters of each model.
- In the second subsample, the forecasting performance of each model can be evaluated. Each model's performance will differ from period to period:

$$e_{1,t_0+1},\, e_{1,t_0+2},\, \dots,\, e_{1,T}$$

$$e_{2,t_0+1},\, e_{2,t_0+2},\, \dots,\, e_{2,T}$$

- We are going to use the estimates (many possibilities)

$$\sigma_1^2 \approx \frac{1}{t_1 - t_0} \sum_{t=t_0+1}^{t_1} e_{1,t}^2; \quad \text{for } t_1 = t_0 + 1, \ldots, T$$

$$\sigma_2^2 \approx \frac{1}{t_1 - t_0} \sum_{t=t_0+1}^{t_1} e_{2,t}^2; \quad \text{for } t_1 = t_0 + 1, \ldots, T$$

- In general, we can define The simplest version of such adaptive weights is, for $t_1 = t_0 + 1, \ldots, T$,

$$w_{t_1+1} = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \approx \frac{\sum_{t=t_0+1}^{t_1} (e_{2,t})^2}{\sum_{t=t_0+1}^{t_1} (e_{1,t})^2 + \sum_{t=t_0}^{t_1} (e_{2,t})^2},$$

- At the time T, we can have the simplest version of such adaptive weights

$$w_{T+1} = \frac{\sum_{t=t_0+1}^{T}(e_{2,t})^2}{\sum_{t=t_0+1}^{T}(e_{1,t})^2 + \sum_{t=t_0+1}^{T}(e_{2,t})^2},$$

- Or updating as (a clever way)

$$\begin{aligned} w^*_{T+1} &= \alpha w^*_T + (1-\alpha)\frac{\sum_{t=t_0+1}^{T}(e_{2,t})^2}{\sum_{t=t_0+1}^{T}(e_{1,t})^2 + \sum_{t=t_0+1}^{T}(e_{2,t})^2}, \\ &= \alpha w^*_T + (1-\alpha) w_{T+1} \end{aligned}$$

Numerous other alternatives are available, see

Clemen, R. T. (1989). Combining forecasts: a review and annotated bibliography, *International Journal of Forecasting*, 5, 559–583

Hendry, D. F., and M. P. Clements (2002). Pooling of forecasts, *The Econometrics Journal*, 7, 1–31

Forecasting combinations

There is extensive empirical evidence in favour of combinations as a forecasting strategy.

- Forecasting combinations offer diversification gains that make them very useful compared to relying on a single model, as we have just seen.
- There may be structural breaks in the data, making it plausible that combining models with different levels adaptability will lead to better results than relying on a single model.

Forecasting combinations

- Even without structural breaks, individuals models may be subject to misspecification bias: it is implausible that a single model dominates all others at all time periods.

- An additional argument for combining forecasts is that predictions from different forecasters may have been constructed under different loss functions.

- Estimation errors are a serious problem for obtaining the combination weights. Simple averages are often better.
- Similarly, nonstationarities can cause instabilities on the weights.

Forecasting combinations

Additional references

If you are interested in this topic, you can also check:

- Stock and Watson (2004) Combination Forecasts of Output Growth in a Seven-Country Data Set, J. of Forecasting, 23, 405–430
- Timmermann (2006) Forecast combinations, Handbook of Economic Forecasting, vol 1, pp 135–196
- Vasnev A, Skirtun M and Pauwels L (2013) Forecasting Monetary Policy Decisions in Australia: A Forecast Combination Approach, Journal of Forecasting, 32, 151–156