Chap 7: Seasonal ARIMA Models

Ani Shabri

Department of Mathematical Sciences,
Faculty of Science, Universiti Teknologi Malaysia,
81310 UTM Johor Bahru, Malaysia
ani@utm.my

Jun 8, 2014

Chap 7: Seasonal ARIMA models

Outline:

- Introduction to SARIMA models
- Seasonality and ARIMA models
- Box-Jenkins methodology
- Stationary
- Model identification222
- Parameter estimation
- Diagnostic checking
- Forecasting

Introduction to SARIMA models

- Common in economic, agricultural and geophysical time series have cycle components within a specific time period.
- The smallest time period for this repetitive phenomenon is called a seasonal period (s). For example,

  - unemployment temperatures have a 24-hour cycle, s = 24.

  - hourly temperature have a 12-month cycle, s = 12.

  - monthly temperature have a 12-month cycle, s = 12.

  - the quarterly ice cream sales have a 4-quarterly cycle, s = 4.
- It may be useful to use a s-fold difference operator $(1-B^s)$ with s = 4 to remove the cycle component from quarterly data, s = 12 to remove annual fluctuations from monthly data.

Seasonality and ARIMA models

- The ARIMA models can be extended to handle seasonal components of a data series.
- The multiplicative seasonal autoregressive moving average model, SARIMA (p, d, q)(P, D, Q)$_s$ is given by

$$\phi(B)\Phi_P(B^s)(1-B)^d(1-B^s)^D y_t = \delta + \theta(B)\Theta_Q(B^s)\varepsilon_t$$

where $\{\varepsilon_t\}$ is Gaussian white noise. $\phi(B)$ is ordinary autoregressive and $\theta(B)$ moving average components; $\Theta_Q(B^s)$ and $\Phi_P(B^s)$ are seasonal autoregressive and moving average components, respectively, and $(1-B)^d$ are $(1-B^s)^D$ the ordinary and seasonal difference component of order *d* and *D*.

Box-Jenkins methodology

- ARIMA models for seasonal time series (SARIMA) are built the SAME ITERATIVE PROCEDURES used for non-seasonal data.
- The Box –Jenkins approach uses an iterative model-building strategy that consist of
  - Stationary
  - Selecting an initial model (model identification)
  - Estimating the model coefficients (parameter estimation)
  - Analyzing the residuals (model checking)
  - Forecasting

Stationary

- The General Formula to transform not stationary to stationary series is given by

$$W_t = (1-B)^d (1-B^s)^D y_t$$

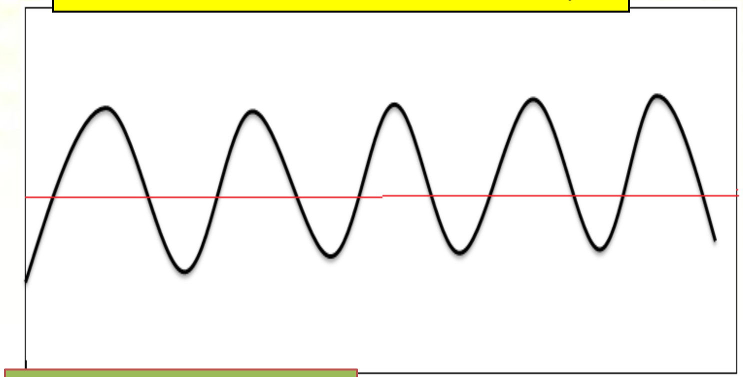
- With seasonal data which is not stationary, it is appropriate to take seasonal differences and check the time plot, ACF and PACF. We take a seasonal difference

$$W_t = y_t - y_{t-s} = (1-B^S) y_t$$

- If the seasonally differenced data appears to be non-stationary (the plots are not shown), so we difference the data again.

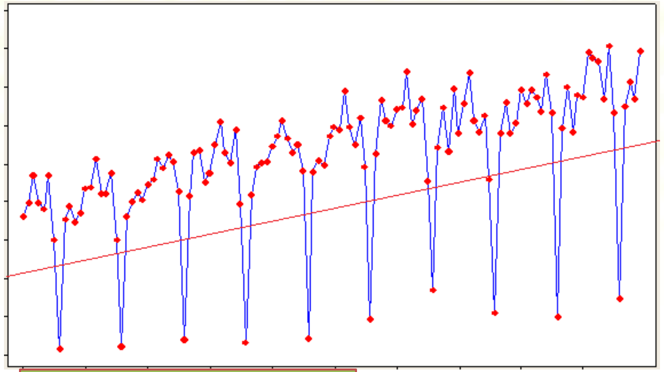
Seasonally stationary process

No trend and additive seasonal variability



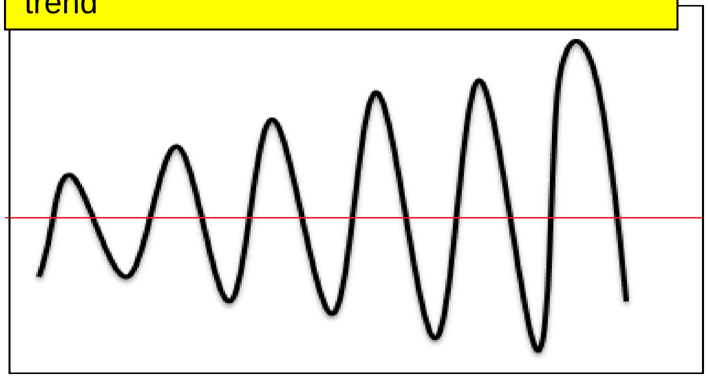
Take d = 0 and D = 1 $W_t = (1 - B^s) y_t$

Additive seasonal variability with an additive trend



Take d = 1 and D = 1 $W_t = (1 - B)(1 - B^s) y_t$

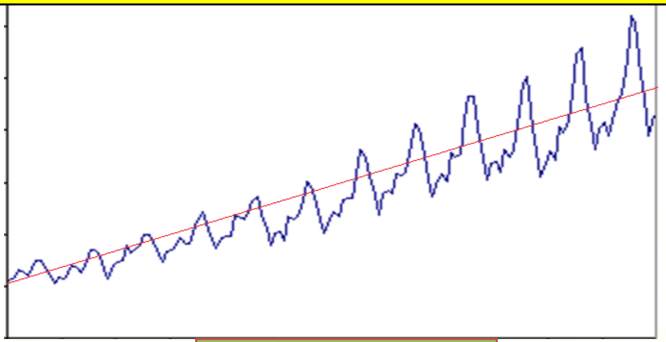
Multiplicative seasonal variability with no trend



Take d = 0 and D = 1

$x_t = \log y_t$ and $W_t = (1 - B^s) x_t$

Multiplicative seasonal variability with an additive trend

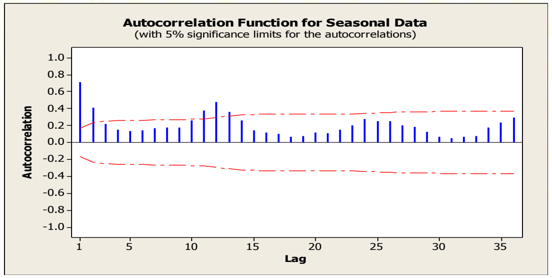


Take d = 1 and D = 1

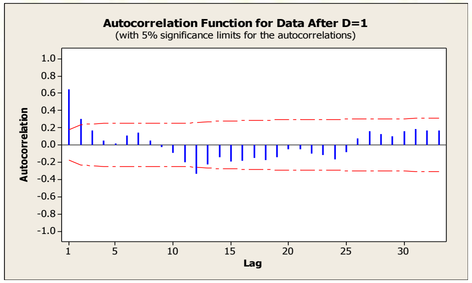
$x_t = \log y_t$ and $W_t = (1 - B)(1 - B^s) x_t$

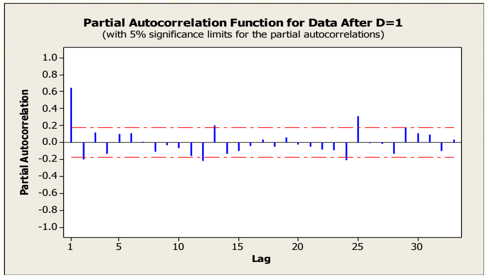
ACF for non-stationary seasonal data

ACF die down slowly at multiples of 12

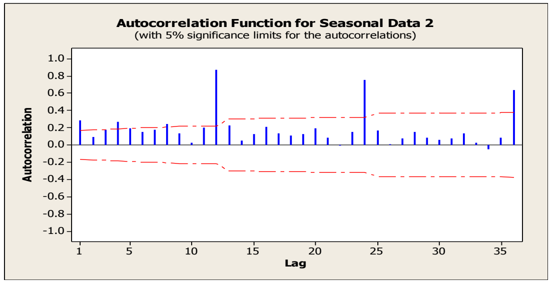


ACF and PACF for first seasonal differencing (D=1)

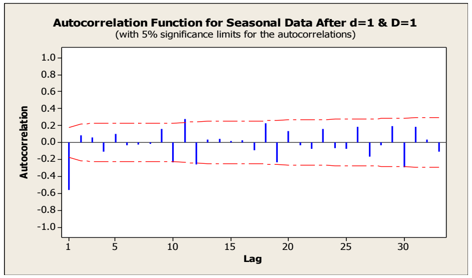


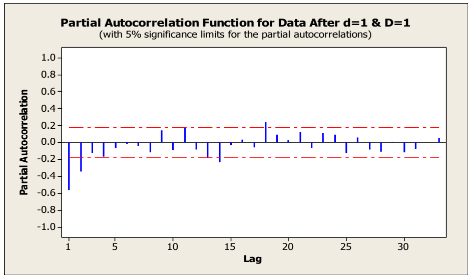


ACF die down slowly before 12 and at multiples of 12

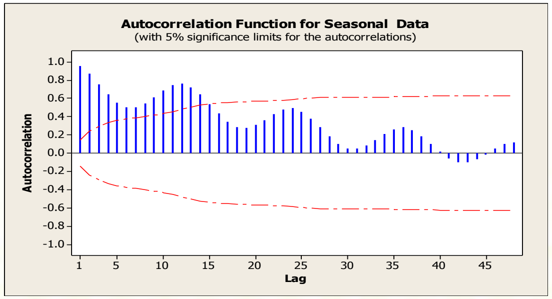


ACF and PACF for first ordinary (d=1)and first seasonal differencing (D=1)

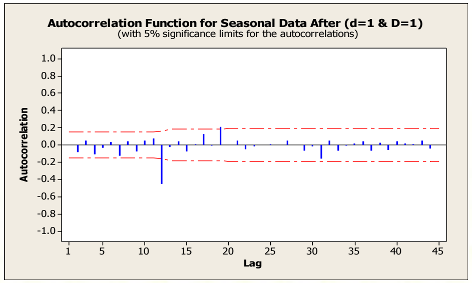


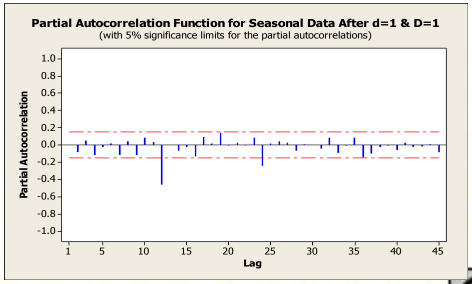


ACF die down slowly and has a cycle pattern at multiples of 12



ACF and PACF for first ordinary (d=1)and first seasonal differencing (D=1)





Model identification

In a purely seasonal ARIMA model (no non-seasonal components), the ACF and PACF have similar forms as for the previous non-seasonal ARIMA models. The difference is now only lags h=s, 2s, 3s,… are examined. The remainder of the lags have ACF and PACF equal to 0.

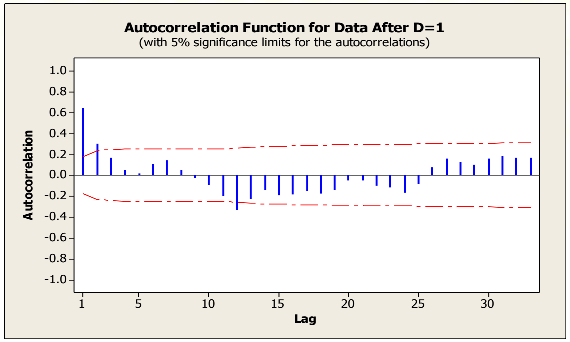
Examples of Identification

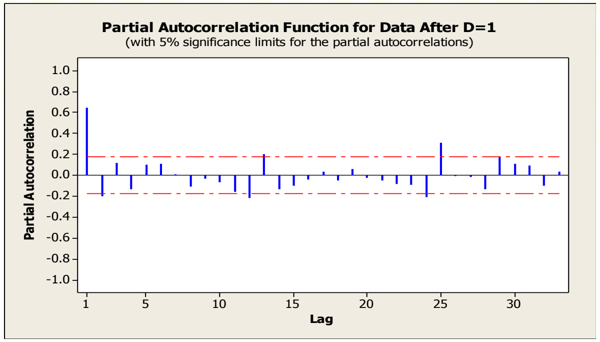
	$SAR(P)_s$	**$SMA(Q)_s$**	**$SARMA(P,Q)_s$**
ACF	Tails off to 0 at lags ks for k=1,2,…	Cuts off to 0 after lag Qs	Tails off to 0 after lag Qs
PACF	Cuts off to 0 after lag Ps	Tails off to 0 at lags ks for k=1,2,…	Tails off to 0 after lag Ps

Note: The values of the ACF and PACF are zero at lags other than s, 2s, 3s, ….

Model identification for Example 1

ACF and PACF for first seasonal differencing (D=1)





Tentative SARIMA MODELS

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

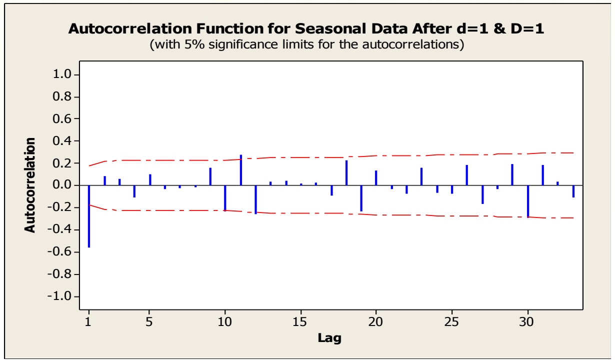
- ACF shows a spike at lag 1, 2 and 12 but no other significant spikes.
- The PACF tails off

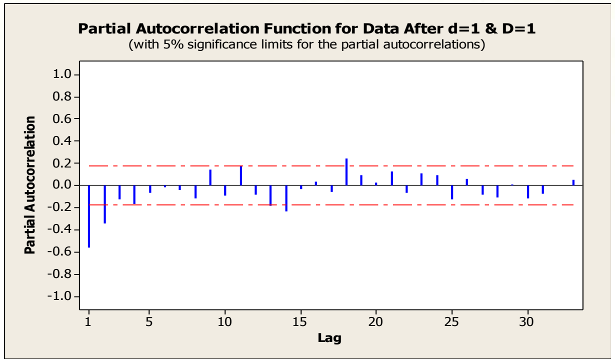
$ARIMA(2,0,0)(2,1,0)_{12}$

- ACF tails off.
- The PACF spike at lag 1, 2,12, 13, 24 and 25 but no other significant spikes.

Model identification for Example 2

ACF and PACF for first ordinary (d=1)and first seasonal differencing (D=1)





Tentative SARIMA MODELS

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

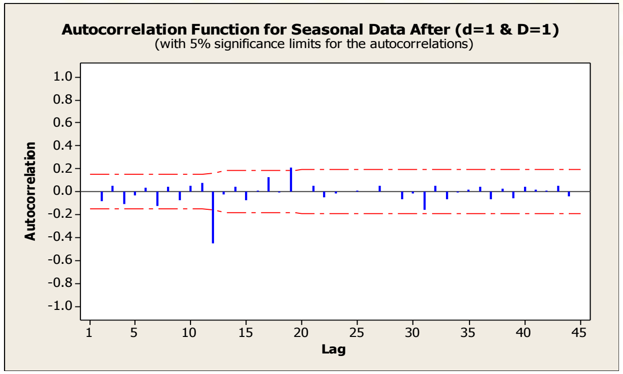
- ACF shows a spike at lag 1, 11 and 12 but no other significant spikes.
- ACF tails off.

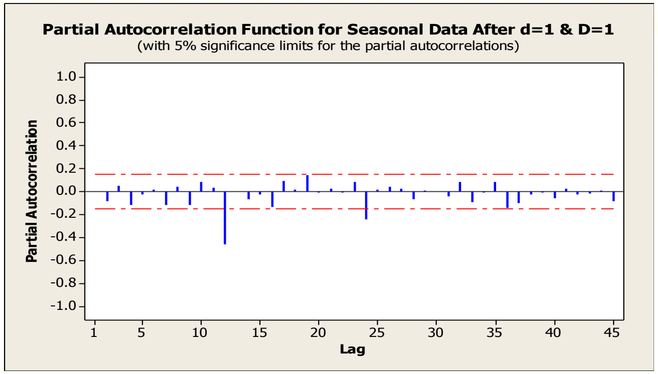
$ARIMA(2,1,0)(1,1,0)_{12}$

- ACF tails off.
- The PACF spike at lag 1, 2, 14 and 18 but no other significant spikes.

Model identification for Example 3

ACF and PACF for first ordinary (d=1)and first seasonal differencing (D=1)





Tentative SARIMA MODELS

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

- ACF shows a spike at lag 12 but no other significant spikes.
- The PACF dying down quickly at seasonal level.

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

- ACF tails off.
- The PACF spike at lag 12 and 24 but no other significant spikes.

Parameter estimation

- Once a tentative model has been identified, the estimates for constant and the coefficients of the parameter SARIMA models must be obtained.
- The model should be parsimonious (simplest form)
- All parameters and constant estimated should be significantly different from zero. Significance of parameters is tested using standard t-test

$$t_{stat} = \frac{\text{point estimate of parameter}}{\text{standard error of estimate}}$$

- The parameters model are significances if

$$\left|t_{stat}\right| > 2 \text{ for } \alpha = 0.05.$$

Diagnostic checking

SARIMA(p,d,q)(P,D,Q)s models are adequate if the residuals nearly the properties white noise process, i.e. the errors

- constant on variances
- Independent
- normally distributed with zero means and variance σ^2

Diagnostics checking for Example 1

Final Estimates of Parameters

Type	Coef	SE Coef	T	P
MA 1	-0.9540	0.0850	-11.23	0.000
MA 2	-0.3076	0.0835	-3.68	0.000
SMA 12	0.8716	0.0659	13.23	0.000

The t statistics are significant at α = 5%

Differencing: 0 regular, 1 seasonal of order 12
Number of observations: Original series 144, after differencing 132
Residuals: SS = 1826299 (backforecasts excluded)
MS = 14157 DF = 129

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	12	24	36	48
Chi-Square	44.0	68.2	88.3	99.4
DF	9	21	33	45
P-Value	0.000	0.000	0.000	0.000

Final Estimates of Parameters

Type	Coef	SE Coef	T	P
AR 1	0.8031	0.0880	9.13	0.000
AR 2	-0.1669	0.0882	-1.89	0.061
SAR 12	-0.6279	0.0751	-8.36	0.000
SAR 24	-0.6053	0.0751	-8.06	0.000

The t statistics are significant at α = 10%

Differencing: 0 regular, 1 seasonal of order 12
Number of observations: Original series 144, after differencing 132
Residuals: SS = 1890332 (backforecasts excluded)
MS = 14768 DF = 128

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	12	24	36	48
Chi-Square	28.4	65.7	112.0	141.8
DF	8	20	32	44
P-Value	0.000	0.000	0.000	0.000

The LBQ statistics are significant as indicated by the small p-values for either model.

Conclusion: Both models are not Adequate models. Try using the other tentative models.

Diagnostics checking for Example 2

Final Estimates of Parameters

Type		Coef	SE Coef	T	P
MA	1	0.7805	0.0564	13.83	0.000
SMA	12	0.4914	0.0804	6.11	0.000

The t statistics are significant at α = 5%

Differencing: 1 regular, 1 seasonal of order 12
Number of observations: Original series 144, after differencing 131
Residuals: SS = 9271841 (backforecasts excluded)
MS = 71875 DF = 129

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	12	24	36	48
Chi-Square	16.0	33.3	62.8	84.7
DF	10	22	34	46
P-Value	0.101	0.058	0.002	0.000

The LBQ statistics are not significant at lag 12 and 24 .

Final Estimates of Parameters

Type		Coef	SE Coef	T	P
AR	1	-0.8119	0.0796	-10.20	0.000
AR	2	-0.4402	0.0800	-5.50	0.000
SAR	12	-0.4086	0.0877	-4.66	0.000

The t statistics are significant at α = 5%

Differencing: 1 regular, 1 seasonal of order 12
Number of observations: Original series 144, after differencing 131
Residuals: SS = 9748163 (backforecasts excluded)
MS = 76158 DF = 128

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	12	24	36	48
Chi-Square	11.2	26.8	43.0	56.5
DF	9	21	33	45
P-Value	0.265	0.179	0.114	0.117

The LBQ statistics are not significant as indicated by the p-values greater than 0.05.

Conclusion: Both models are Adequate models.

Diagnostics checking for Example 3

Final Estimates of Parameters

Type		Coef	SE Coef	T	P
SMA	12	0.6776	0.0580	11.67	0.000

The t statistics are significant at α = 5%

Differencing: 1 regular, 1 seasonal of order 12
Number of observations: Original series 192, after differencing 179
Residuals: SS = 1169.30 (backforecasts excluded)
MS = 6.57 DF = 178

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	12	24	36	48
Chi-Square	19.2	32.9	41.1	49.8
DF	11	23	35	47
P-Value	0.058	0.083	0.221	0.362

Final Estimates of Parameters

Type		Coef	SE Coef	T	P
SMA	12	0.5854	0.0756	7.75	0.000
SMA	24	0.0957	0.0781	1.23	0.222

The t statistics is not significant at α = 5%. Consider SARIMA(0,1,0)(1,1,0)

Differencing: 1 regular, 1 seasonal of order 12
Number of observations: Original series 192, after differencing 179
Residuals: SS = 1184.15 (backforecasts excluded)
MS = 6.69 DF = 177

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	12	24	36	48
Chi-Square	19.3	33.3	41.9	50.6
DF	10	22	34	46
P-Value	0.037	0.058	0.164	0.296

The LBQ statistics are not significant as indicated by the p-values greater than 0.05, except at lag 12

Conclusion: Both models are Adequate models.

Forecasting

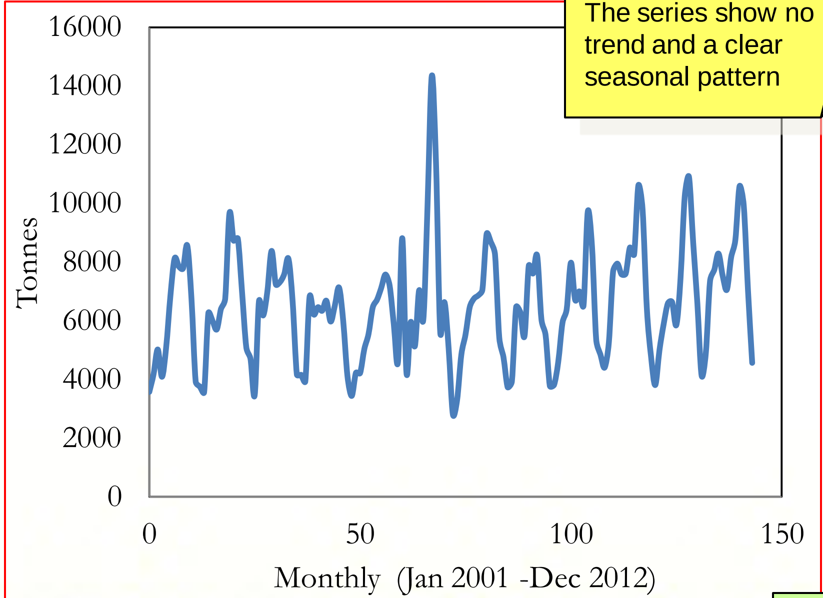
- Once the fitted model has been selected, it can be used to generate forecasts for future time periods.
- The forecast values of h-period ahead for SARMA(p,d,q)(P,D,Q) model is given by

$$\phi(B)\Phi_P(B^s)(1-B)^d(1-B^s)^D\hat{y}_{t+h} = \delta + \theta(B)\Theta_Q(B^s)\varepsilon_t$$

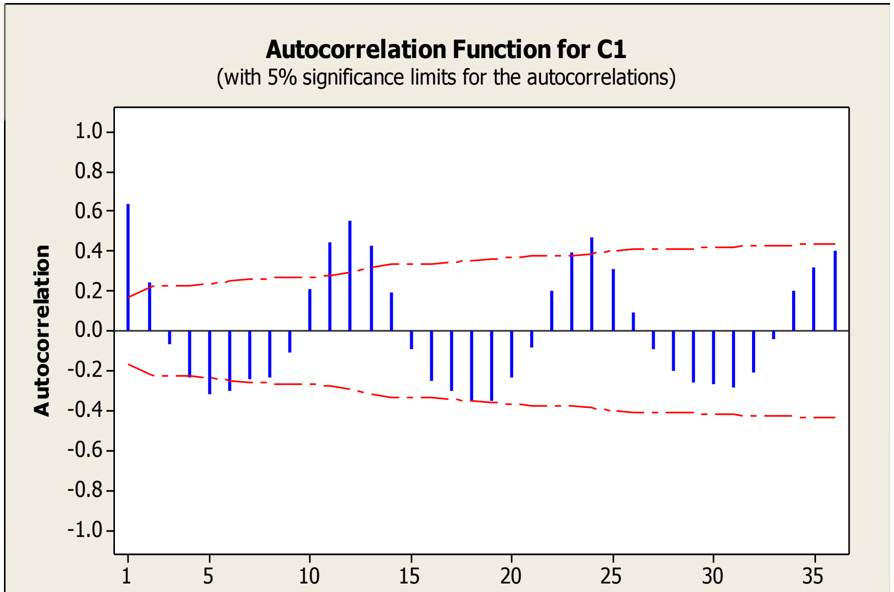
where the forecast values of the SARIMA model may be found by replaced by their estimates when the actual values are not available.

Example

The monthly data of fishery landing in East Johor, covering the period from January 2001 to December 2012 with a total of 144 observations are used, as shown in Fig. 4. The time plot shows a clear seasonal pattern and the series fluctuate around a constant mean. The ACF cuts of to zero quickly and significance in the large values at the seasonal lags 12, 24 and 36.



The series show no trend and a clear seasonal pattern



The ACF trailed off to zero rather quickly→ no trend exists and ACF at lags 12, 24, 36 are significantly and slowly decreasing (seasonality exits)

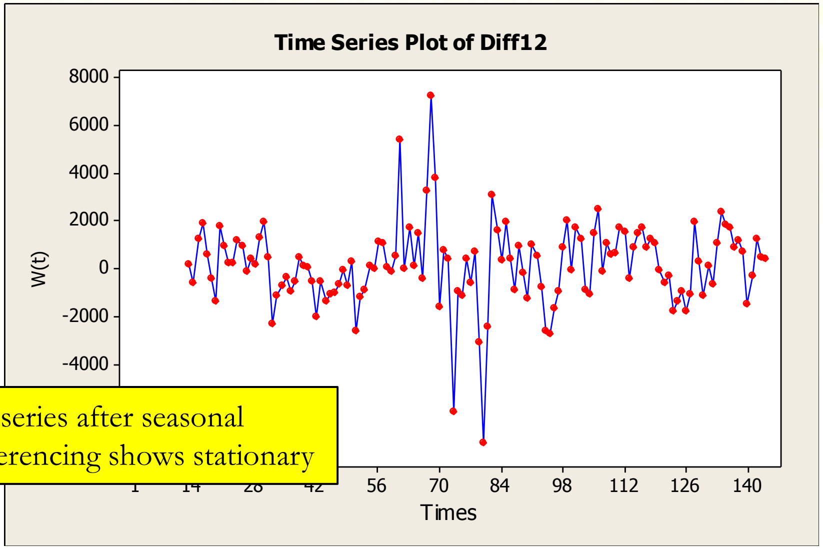
Seasonal differencing

- ACF at the seasonal lags 12, 24 and 36 were large and failed to die out quickly. This suggested the series is non-stationary seasonality data . Thus with non-stationary seasonal data, we have to difference the observations using

$$w_t = y_t - y_{t-s} = (1 - B^s)y_t$$

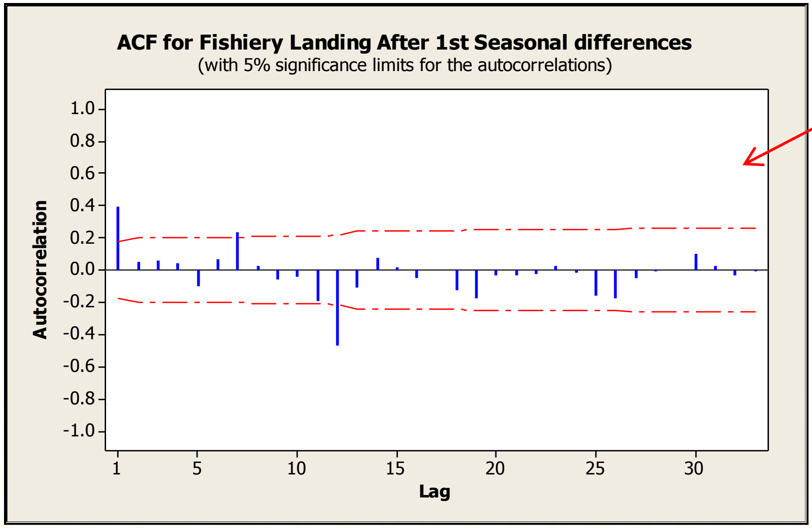
where s = 12

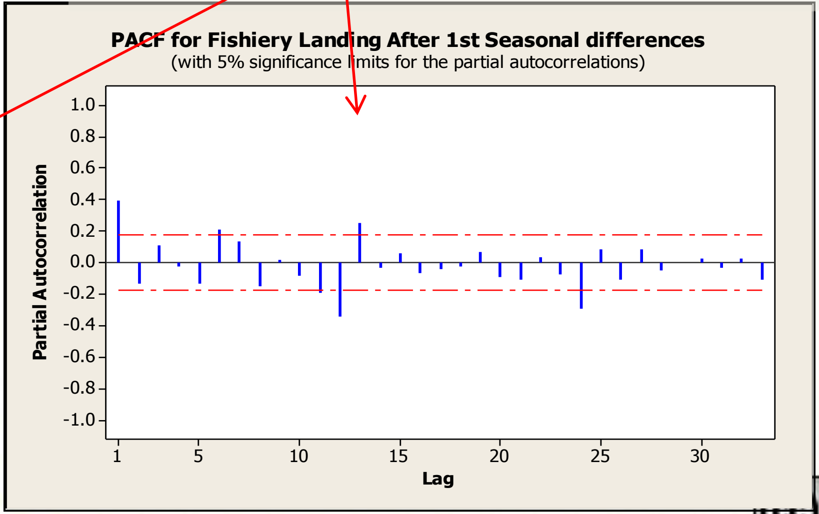
Seasonal differencing



The series after seasonal differencing shows stationary

Comparing the ACF with their error limits, the significant ACF are at lag 1, 7 and 12, , indicating MA(1) and SMA(1) behavior. The PACF appears to cut off after lag 1, 6, 12, 13, 24 indicating AR(1) and SAR(2) behavior → we will try: SARIMA(1,0,0)(2,1,0)12 and SARIMA(0,0,1)(0,1,1)12





ARIMA with MINITAB

SARIMA(1,0,0)(2,1,0) 12

The t statistics are significant at α = 5%

SARIMA(0,0,1)(0,1,1) 12

SARIMA(1,0,0)(2,1,0) 12

Final Estimates of Parameters

Type	Coef	SE Coef	T	P
AR 1	0.4455	0.0790	5.64	0.000
SAR 12	-0.7078	0.0842	-8.41	0.000
SAR 24	-0.3252	0.0862	-3.77	0.000

Differencing: 0 regular, 1 seasonal of order 12
Number of observations: Original series 144, after differencing 132
Residuals: SS = 210115255 (backforecasts excluded)
MS = 1628800 DF = 129

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	12	24	36	48
Chi-Square	15.3	22.7	42.5	51.2
DF	9	21	33	45
P-Value	0.082	0.360	0.125	0.243

SARIMA(0,0,1)(0,1,1) 12

Final Estimates of Parameters

Type	Coef	SE Coef	T	P
MA 1	-0.4458	0.0782	-5.70	0.000
SMA 12	0.8981	0.0643	13.96	0.000

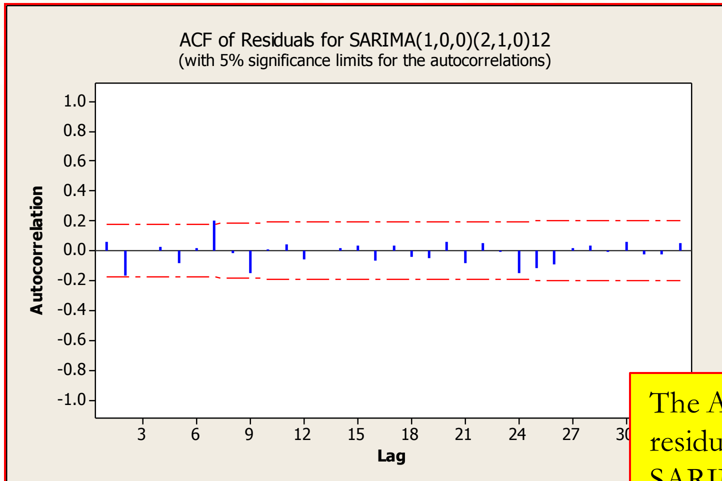
Differencing: 0 regular, 1 seasonal of order 12
Number of observations: Original series 144, after differencing 132
Residuals: SS = 175940657 (backforecasts excluded)
MS = 1353390 DF = 130

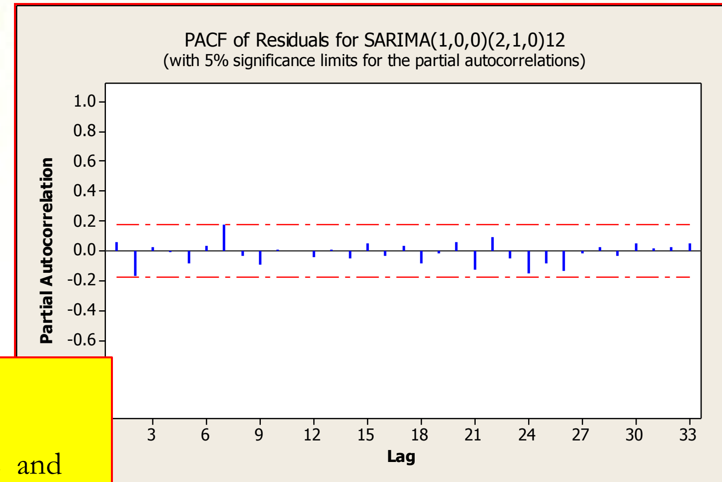
Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	12	24	36	48
Chi-Square	12.2	16.6	25.7	31.2
DF	10	22	34	46
P-Value	0.275	0.787	0.847	0.954

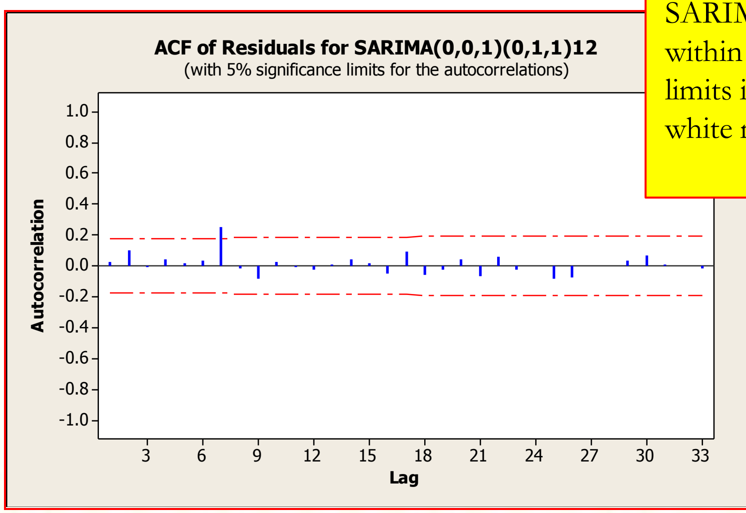
The LBQ statistics are not significant as indicated by the large p-values for either model.

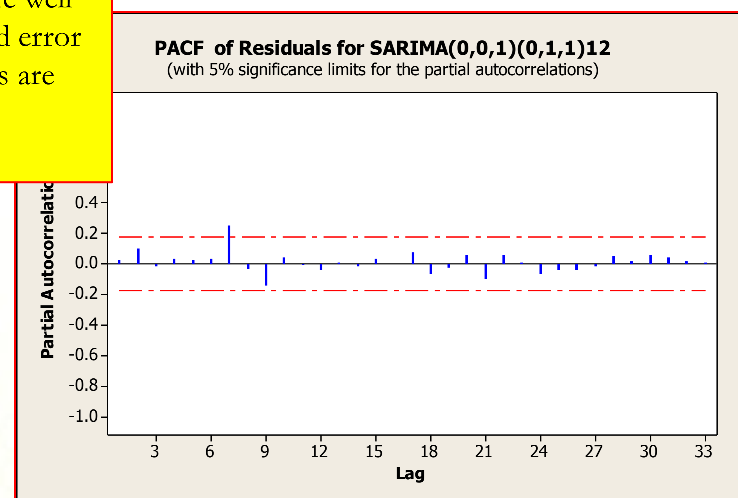
ACF and PACF of residuals of SARIMA Model





The ACF and PACF of residuals for SARIMA(1,0,0)(2,1,0)12 and SARIMA(0,0,1)(0,1,1) are well within their two standard error limits indicating residuals are white noise.





Model selection criteria

	t-test	Q-test	AIC	BIC
$SARIMA(1,0,0)(2,1,0)_{12}$	√	√	14.235	14.297
$SARIMA(0,0,1)(0,1,1)_{12}$	√	√	14.044	14.085

Judging these results, it appears that the estimated $SARIMA(0,0,1)(0,1,1)_{12}$ model best fits the data.

Comparison of actual and forecasted Values for SARIMA(0,0,1)(0,1,1)12 model