TIME SERIES

Nonseasonal Box-Jenkins Models

- A time series is stationary if: $E\left(Y_t\right) = \mu, Var\left(Y_t\right) = \sigma^2 \text{ for all } t$

In other words, if $y_1, y_2, \ldots, y_n$ values of the time series fluctuate around a constant mean with constant variation, it is reasonable to assume that the time series is stationary (Figure 1(b), next slide)

If the n values do not seem to fluctuate around a constant mean or do not fluctuate with constant variation around a constant mean, then it is non-stationary (Figure 1(a), next slide)

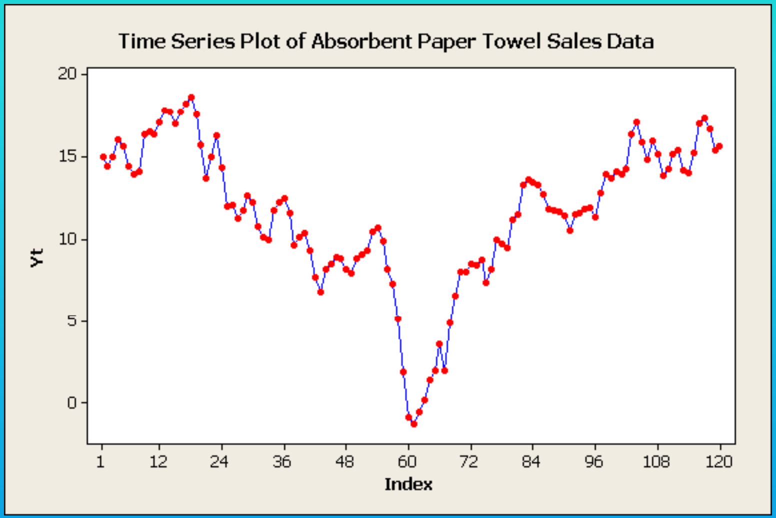


Figure 1(b): non-stationary time series

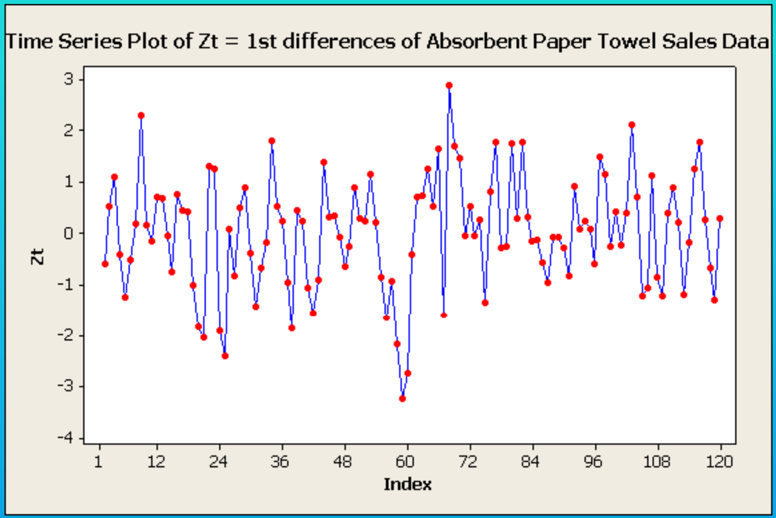


Figure 1(a): stationary time series

- A non-stationary time series can usually be transformed into a stationary one by computing
- the 1st differences as follows:

Original	First differences
y_1	-------------
y_2	$z_2 = y_2 - y_1$
y_3	$z_3 = y_3 - y_3$
…	…
y_n	$z_n = y_n - y_{n-1}$

Or by 2nd differences = 1st differences calculated for z_i

Original	1st differences	2nd differences
y_1	-------------	
y_2	$z_2 = y_2 - y_1$	-------------
y_3	$z_3 = y_3 - y_2$	$z_3 = (y_3 - y_2) - (y_2 - y_1) = y_3 - 2y_2 + y_1$
...	...	...
y_n	$z_n = y_n - y_{n-1}$	$z_n = y_n - 2y_{n-1} + y_{n-2}$

Identification of a Box-Jenkins Model is done by looking at the SAC and SPAC of the values of a stationary time series

$z_b, z_{b+1}, \ldots, z_n$ = working time series of 1st or 2nd differences

Sample Autocorrelation Function (SAC) at lag K is defined as:

$$r_k = Corr\left(z_b, z_{b+k}\right)$$

$$= \frac{\sum_{t=b}^{n-k}\left(z_t - \overline{z}\right)\left(z_{t+k} - \overline{z}\right)}{\sum_{t=b}^{n}\left(z_t - \overline{z}\right)^2}$$

$$where\ \overline{z} = \frac{\sum_{t=b}^{n} z_t}{n-b+1}$$

The standard error of r_k is

$$s_{r_k} = \begin{cases} \sqrt{\dfrac{1}{n-b+1}} & \text{if } k=1 \\ \sqrt{\dfrac{1+2\sum_{j=1}^{k-1} r_j^{\,2}}{n-b+1}} & \text{if } k=2,3,\ldots \end{cases}$$

The $t-value$ is given by $t_{r_k} = \dfrac{r_k}{s_{r_k}}$

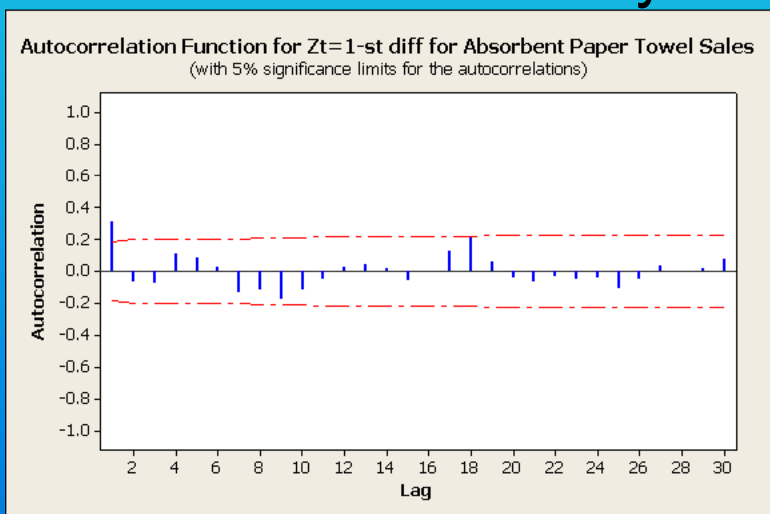
A spike exists at lag k if $|t_{r_k}| > 2$.

SAC cuts off after lag k if there are no spikes at lags $> k$.

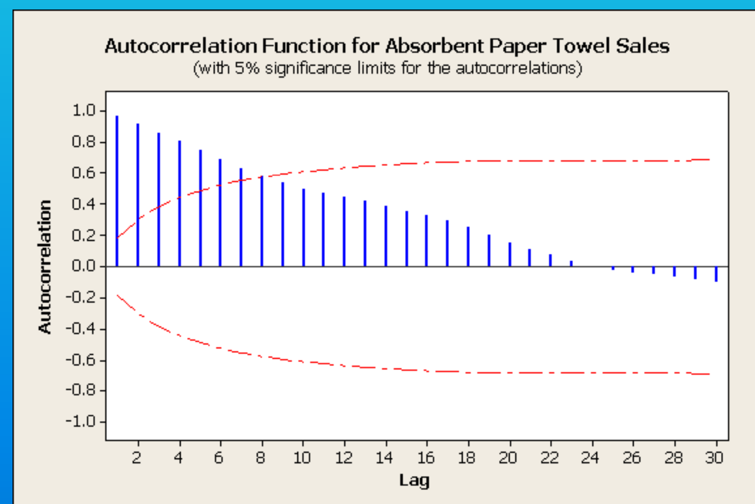
The SAC function is a graph of $r_k \ vs. \ k(lag)$

NOTE: If SAC of $z_b, z_{b+1}, \cdots, z_n$ either cuts off or dies down very quickly, then the time series is stationary.

If the SAC dies down extremely slowly, then the time series is non-stationary.



SAC cuts off quickly (after lag k = 1) as r_k is not significantly different from 0 for k > 1; working series is stationary.



SAC cuts off very slowly (after lag k > 7) as r_k is significantly different from 0 for k ≤ 7; time series is non-stationary.

The Sample Partial Autocorrelation (SPAC) Function is defined as:

$$r_{KK} = \begin{cases} r_1 & K = 1 \\ \dfrac{r_K - \sum_{j=1}^{K-1} r_{K-1,j}\, r_{K-j}}{1 - \sum_{j=1}^{K-1} r_{K-1,j}\, r_j} \end{cases}$$

$$where\ r_{Kj} = r_{K-1,j} - r_{KK} r_{K-1,K-j},\ j = 1, 2, .., K-1.$$

It's standard error is:

$$s_{r_{KK}} = \frac{1}{\sqrt{n - b + 1}}$$

and the t-statistic is:

$$t_{r_{KK}} = \frac{r_{KK}}{s_{r_{KK}}}$$

The SPAC function is a graph of $r_{KK}\ vs.\ K.$

$r_{KK} =$ SAC at lag K with the effect of intervening observations eliminated.

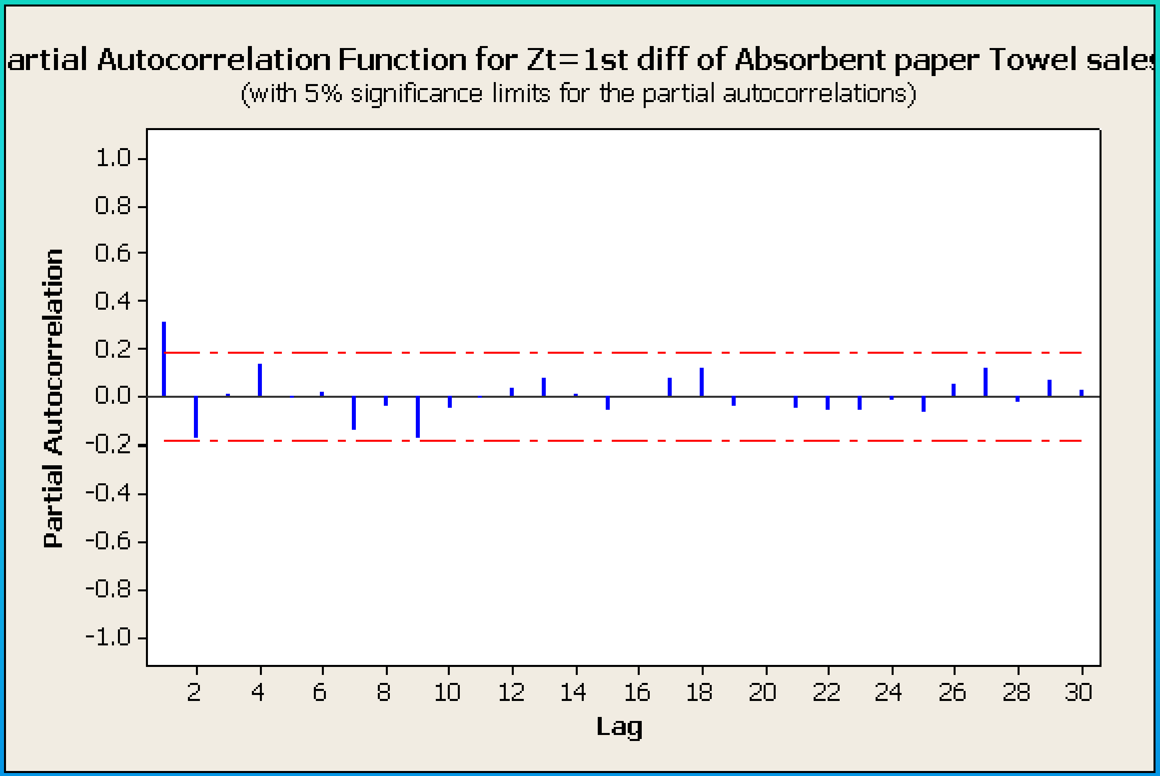
If $\left| t_{r_{KK}} \right| > 2$ we say that a spike at lag K exists in the time series.

If there are no spikes in the time series at lag>K in SPAC, then we say that SPAC cuts off after lag K.

If SPAC does not cut off but decreases steadily, we say that SPAC dies down.

NOTE: 1. For nonseasonal data, if SPAC cuts off it will do so (typically) for $K \leq 2$

2. The behavior of SPAC helps us to identify Box-Jenkins models.



SPAC oscillates and dies down.

Nonseasonal Box-Jenkins Models and Their Tentative Identification

$\{y_1, y_2, ..., y_n\} \xrightarrow{differencing} \{z_b, z_{b+1}, ..., z_n\}$, working series

We look at the SAC and SPAC of the working series to identify a Box-Jenkins model.

Two commonly used Box-Jenkins mod els are

1) Non-seasonal autoregressive (AR) model of order 1

$z_t = \phi_1 z_{t-1} + a_t$ where $a_t =$ random shock~N$(0, \sigma^2)$, independent

2) Non-seasonal moving average (MA) model of order 1

$z_t = a_t - \theta_1 a_{t-1}$

Theoretical Autocorrelation Function (TAC) is

$$\rho_k = Corr\left(Z_b, Z_{b+k}\right)$$

$$= \frac{E(Z_b - \mu)(Z_{b+k} - \mu)}{\sigma^2}$$

where $E(Z_b) = \mu, \quad Var(Z_b) = Var(Z_{b+k}) = \sigma^2$

The Theoretical Partial Autocorrelation (TPAC) Function is defined as:

$$\rho_{kk} = \begin{cases} \rho_1 & k = 1 \\ \dfrac{\rho_k - \sum_{j=1}^{k-1} \rho_{k-1,j}\, \rho_{k-j}}{1 - \sum_{j=1}^{K-1} \rho_{k-1,j}\, \rho_j} \end{cases}$$

$$where\ \rho_{Kj} = \rho_{k-1,j} - \rho_{kk}\rho_{k-1,k-j},\ j = 1, 2, .., k-1.$$

For the non-seasonal moving average (MA) model of order 1

$z_t = a_t - \theta_1 a_{t-1},$ we can show that

$$\rho_1 = \frac{-\theta_1}{1+\theta_1^2} \neq 0$$

$$\rho_k = 0 \text{ , } k \geq 2$$

PROOF : (you can skip the proof if you want)

$z_t = a_t - \theta_1 a_{t-1},\ a_t \sim N(0,\sigma^2),$ independently

$$E(z_t)=0, Var(z_t)=Var(a_t-\theta_1 a_{t-1})=\sigma^2(1+\theta_1^2)$$

$$z_{t-1} = a_{t-1} - \theta_1 a_{t-2}$$

$$\rho_1 = Corr(z_t, z_{t-1}) = \frac{E(z_t z_{t-1}) - E(z_t)E(z_{t-1})}{\sigma_{z_t}\sigma_{z_{t-1}}} = \frac{E(z_t z_{t-1})}{\sigma_{z_t}\sigma_{z_{t-1}}}$$

$$E(z_t z_{t-1}) = E[(a_t - \theta_1 a_{t-1})(a_{t-1} - \theta_1 a_{t-2})]$$

$$= E[a_t a_{t-1} - \theta_1 a_t a_{t-2} - \theta_1 a_{t-1}^2 + \theta_1^2 a_{t-1} a_{t-2}]$$

$$= -\theta_1 \sigma^2$$

$$\rho_1 = \frac{E(z_t z_{t-1})}{\sigma_{z_t}\sigma_{z_{t-1}}} = \frac{-\theta_1\sigma^2}{\sigma^2(1+\theta_1^2)} = \frac{-\theta_1}{(1+\theta_1^2)} \neq 0$$

$PROOF:$ (continued)

$z_t = a_t - \theta_1 a_{t-1},\ a_t \sim N(0,\sigma^2)$, independently

$E(z_t) = 0, Var(z_t) = Var(a_t - \theta_1 a_{t-1}) = \sigma^2(1+\theta_1^2)$

$z_{t-1} = a_{t-1} - \theta_1 a_{t-2}$

$$\rho_2 = Corr(z_t, z_{t-2}) = \frac{E(z_t z_{t-2}) - E(z_t)E(z_{t-2})}{\sigma_{z_t}\sigma_{z_{t-2}}} = \frac{E(z_t z_{t-2})}{\sigma_{z_t}\sigma_{z_{t-2}}}$$

$$E(z_t z_{t-2}) = E[(a_t - \theta_1 a_{t-1})(a_{t-2} - \theta_1 a_{t-3})]$$

$$= E[a_t a_{t-2} - \theta_1 a_t a_{t-3} - \theta_1 a_{t-1} a_{t-2} + \theta_1^2 a_{t-1} a_{t-3}]$$

$$=0$$

$$\rho_2 = \frac{E(z_t z_{t-1})}{\sigma_{z_t}\sigma_{z_{t-1}}} = 0$$

Similarly, $\rho_k = 0,\ k > 2$

We can also show that the TPAC of an MA-1 model dies down slowly:

$$\rho_{kk} = \begin{cases} \rho_1 & k=1 \\ \dfrac{\rho_k - \sum_{j=1}^{k-1} \rho_{k-1,j}\, \rho_{k-j}}{1 - \sum_{j=1}^{k-1} \rho_{k-1,j}\, \rho_j} \end{cases}$$

$$\text{where } \rho_{1j} = \rho_{k-1,j} - \rho_{kk}\rho_{k-1,k-j},\ j = 1, 2, .., k-1.$$

$$\rho_{11} = \rho_1 = \frac{-\theta_1}{1+\theta_1^2}$$

$$\rho_{22} = \frac{\rho_2 - \rho_{11}\rho_1}{1-\rho_{11}\rho_1} = \frac{-\rho_1^2}{1-\rho_1^2} = \frac{-\theta_1^2}{1+2\theta_1^2}$$

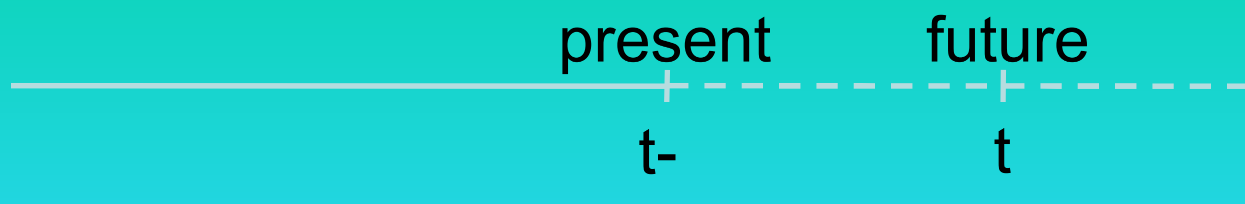
It can be verified that $|\rho_{11}| > |\rho_{22}| > |\rho_{33}| > \ldots$

The above two results for TAC/TPAC show that, if

1) The SAC has a spike at lag k = 1, cuts off after lag k=1, and
2) The SPAC dies down in a damped exponential decay

then we should try a moving average of order 1 (MA-1)
model for the working series.

The Least Squares Estimation of θ_1 for MA-1 model



$$z_t = a_t - \theta_1 a_{t-1} \Rightarrow y_t - y_{t-1} = a_t - \theta_1 a_{t-1}$$

$$\Rightarrow y_t = y_{t-1} + a_t - \theta_1 a_{t-1}$$

$$\hat{y}_t = y_{t-1} + \hat{a}_t - \theta_1 \hat{a}_{t-1}$$

$$\hat{a}_t = 0 \text{ for future shock (value of mean)}$$

$$\hat{a}_{t-1} = \begin{cases} y_{t-1} - \hat{y}_{t-1} \text{ if } \hat{y}_{t-1} \text{ can be computed} \\ 0 \text{ if } \hat{y}_{t-1} \text{ cannot be computed} \end{cases}$$

$$\hat{y}_1 = y_0 + \hat{a}_1 - \theta_1 \hat{a}_0 = 0 \text{ (no } y_0\text{)}$$

$$\hat{y}_2 = y_1 + \hat{a}_2 - \theta_1 \hat{a}_1$$

Following table shows calculations for a_t_hat, Y_t_hat for estimating θ_1 by minimizing Sum of Squared Residuals (SSE).

			ORDER OF CALCULATIONS		
			3	1	2
	t	Yt	at_hat	Yt_hat	Resid
Line 1	1	Y1	0	Y1hat=0	Y1-Y1hat = Y1
Line 2	2	Y2	Y1-Y1hat	Y1+0-theta1*a1hat	Y2-Y2that
Line 3	3	Y3	Y2-Y2hat	Y2+0-theta1*a2hat	Y3-Y3that
…	…	…	…	…	…

Initial table is obtained using θ_1 = 0.2

theta1	SSE
0.2	392.3257

		ORDER OF CALCULATIONS			
		3	1	2	
t	Yt	at_hat	Yt_hat	Resid	Resid_sqr
0					
1	15	0	0	15	225
2	14.4064	-0.5936	15	-0.5936	0.352361
3	14.9383	0.41318	14.52512	0.41318	0.170718
4	16.0374	1.181736	14.855664	1.181736	1.3965
5	15.632	-0.16905	15.801053	-0.16905	0.028579

Following optimal table is obtained using EXCEL SOLVER; there are no constraints for the parameter θ_1 .

theta1	SSE
-0.35339	352.4788

		ORDER OF CALCULATIONS			
		3	1	2	
t	Yt	at_hat	Yt_hat	Resid	Resid_sqr
0					
1	15	0	0	15	225
2	14.4064	-0.5936	15	-0.5936	0.352361
3	14.9383	0.74167	14.19663	0.74167	0.550075
4	16.0374	0.837004	15.200396	0.837004	0.700575
5	15.632	-0.70119	16.333186	-0.70119	0.491662

Non-seasonal moving average (MA) model of order q

$$z_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \ldots - \theta_q a_{t-q}$$

The term Moving Average refers to the inclusion the current random shock a_t as well as the past random shocks $a_{t-1}, a_{t-2}, \ldots, a_{t-q}$ in the model.

For $q = 1$,

1) TAC has a spike at lag 1 and cuts off after lag 1.

$$\rho_1 = \frac{-\theta_1}{1+\theta_1^2} \neq 0$$

$$\rho_k = 0 \ , k \geq 2$$

2) TPAC dies down $(\rho_{11}, \rho_{22}, \ldots \downarrow k)$, dominated by exponential decay.

For $q = 2$

1) TAC has spikes at lags 1 and 2, cuts off after lag 2.

$$\rho_1 = \frac{-\theta_1(1-\theta_1)}{1+\theta_1^{\ 2}+\theta_2^{\ 2}}$$

$$\rho_2 = \frac{-\theta_2}{1+\theta_1^{\ 2}+\theta_2^{\ 2}}$$

$$\rho_k = 0,\ k > 2$$

2) TPAC dies down $(\rho_{11}, \rho_{22}, \ldots \downarrow k),$ dominated by exponential decay.

In general,

1) TAC has spikes at lags 1, 2,..., q and cuts off after lag q.

$$\rho_k = \begin{cases} \neq 0, & k = 1, 2, \ldots, q \\ 0, & k > q \end{cases}$$

2) TPAC dies down $(\rho_{11}, \rho_{22}, \ldots \downarrow k)$, dominated by exponential decay.

Therefore if the SAC has spikes at lags 1, 2,..., q and cuts off after lag q, and the SPAC dies down, then we should try a moving average model of order q.

Nonseasonal Autoregressive (AR) Model of order p

$$z_t = \delta + \phi_1 z_{t-1} + \phi_2 z_{t-2} + ... + \phi_p z_{t-p}$$

Taking expected values of both sides and using the fact that $E(z_t) = \mu$, we get

$$\mu = \delta + \phi_1 \mu + \phi_2 \mu + ... + \phi_p \mu$$

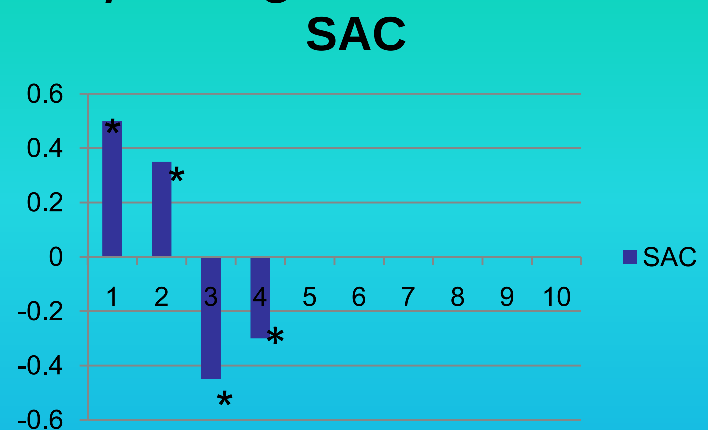
$$\delta = \mu(1 - \phi_1 - \phi_2 - ... - \phi_p)$$

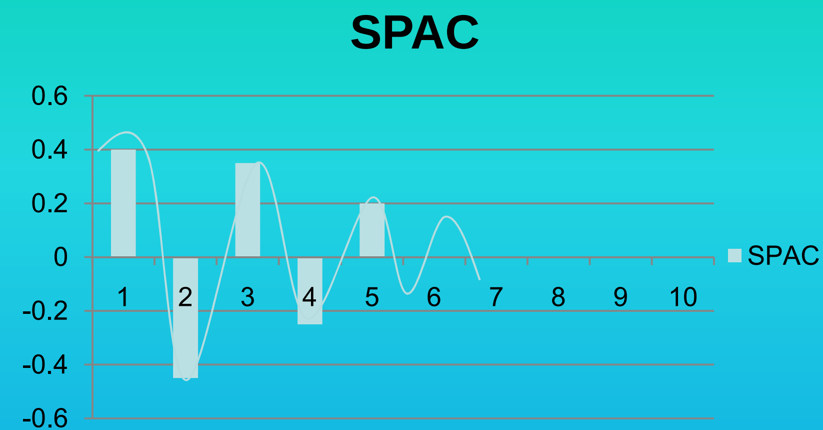
It can be shown that for this model,

1) The TAC dies down : $\rho_k \downarrow k$

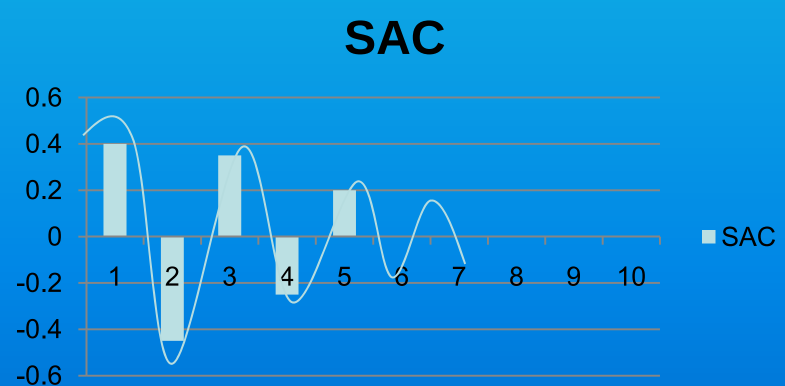
2) The TPAC has spikes at lags $1, 2, ..., p$ and cuts off after lag p, *i.e.*, $\rho_{kk} \neq 0,\ k = 1, 2, ..., p$ and $\rho_{kk} = 0, k > p$

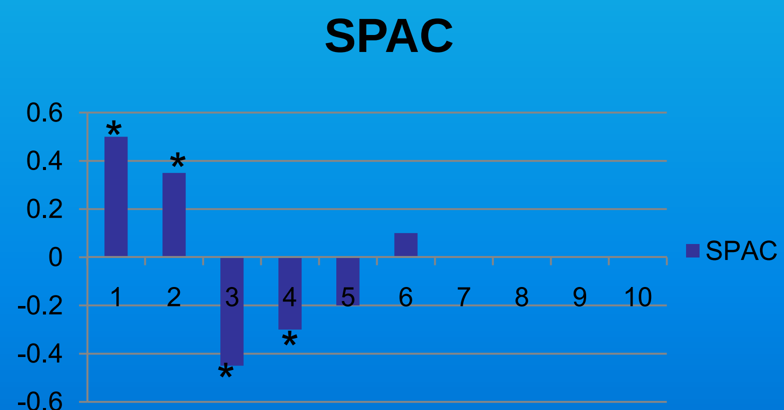
Nonseasonal MA Model of order *q=4: SAC has spikes upto lag 4, cuts off after k = 4; SPAC dies down*





Nonseasonal AR Model of order *p=4: SAC dies down; SPAC has spikes upto lag 4, cuts off after k = 4.*





Nonseasonal Mixed Autoregressive Moving Average (ARMA) Model of order (p,q)

$$z_t = \delta + \phi_1 z_{t-1} + \phi_2 z_{t-2} + \ldots + \phi_p z_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \ldots - \theta_q a_{t-q}$$

Taking expected values of both sides and using the fact that $\mathrm{E}(z_t) = \mu$, we get

$$\mu = \delta + \phi_1 \mu + \phi_2 \mu + \ldots + \phi_p \mu$$

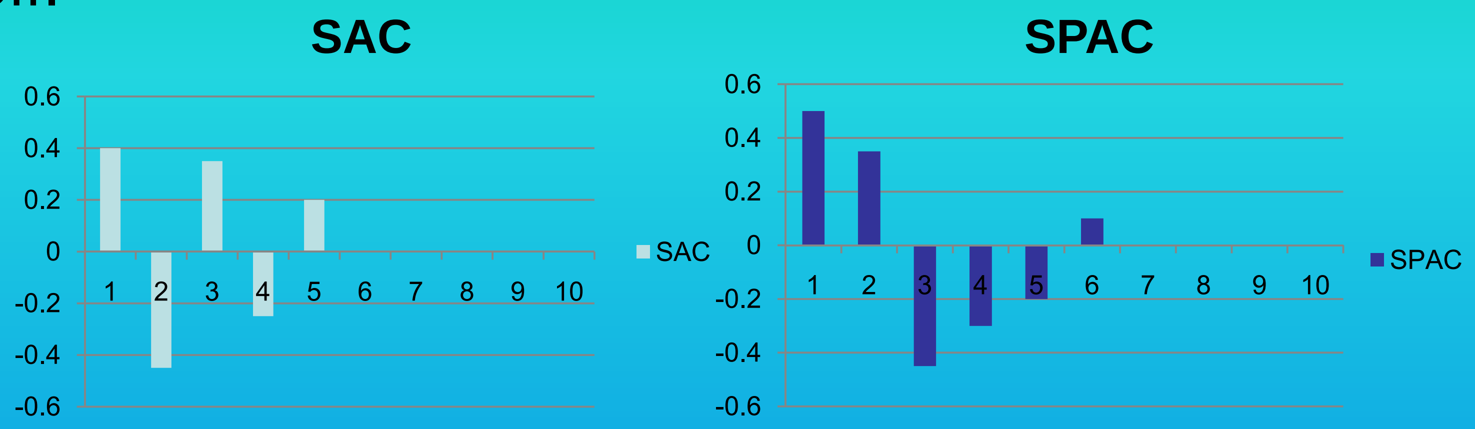
$$\delta = \mu(1 - \phi_1 - \phi_2 - \ldots - \phi_p)$$

It can be shown that for this model,

1) The TAC dies down : $\rho_k \downarrow k$

2) The TPAC also dies down: $\rho_{kk} \downarrow k$

NOTE: There is no theoretical Box-Jenkins model for which both SAC and SPAC have spikes and then cut off.



If data suggests that both SAC and SPAC have spikes and then cut off, we should try to see which of the SAC and SPAC cuts off more abruptly.

If SAC is cutting off more abruptly, then we should try *MA-q* model.

If SPAC is cutting off more abruptly, then we should try *AR-p* model.

If no choice can be made, then try both of these models.

Estimation, Diagnostics, and Forecasting using Nonseasonal Box-Jenkins Models

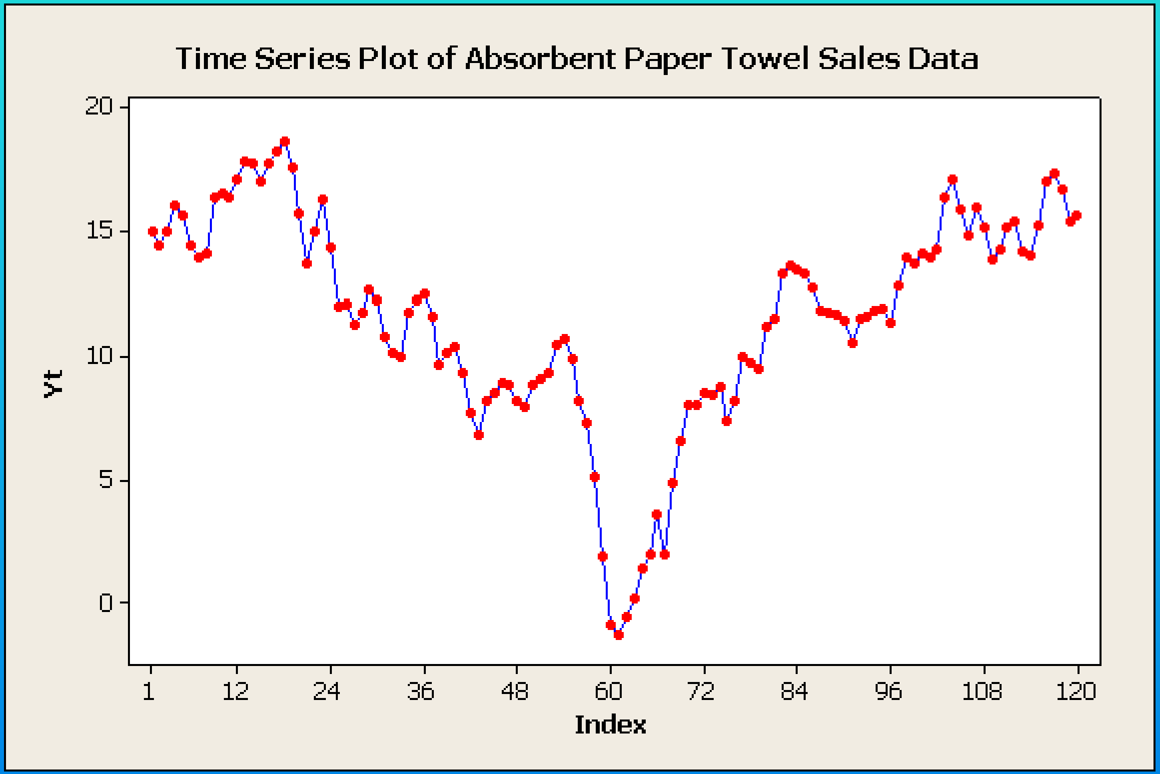
- Box-Jenkins models must be both **stationary** and **invertible**.
- **Stationarity** means the mean and variance of *working series* remain constant over time.
- **Invertibility** means that when a Box-Jenkins model is put in a form that relates z_t *to* past values $z_{t-1}, z_{t-2}, \ldots$ *then the weights decline as we move further into the past.*

Stationarity and Invertibility Conditions for Some Specific Box-Jenkins Models

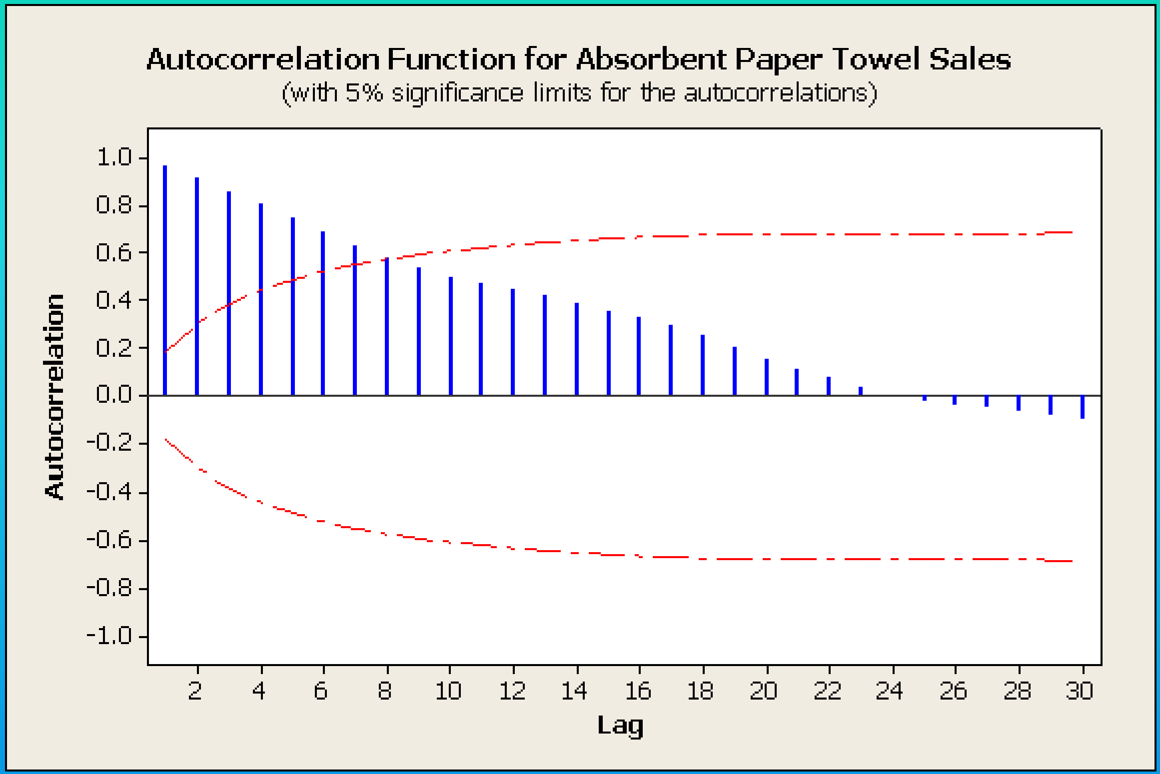
Model	Stationarity condition	Invertibility condition
MA-1 $z_t = \delta + a_t - \theta_1 a_{t-1}$	None	$\lvert\theta_1\rvert < 1$
MA-2 $z_t = \delta + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$	None	$\theta_1 + \theta_2 < 1$ $\theta_2 - \theta_2 < 1, \lvert\theta_2\rvert < 1$
AR-1 $z_t = \delta + \phi_1 z_{t-1} + a_t$	$\lvert\phi_1\rvert < 1$	None
AR-2 $z_t = \delta + \phi_1 z_{t-1} + \phi_2 z_{t-2} + a_t$	$\phi_1 + \phi_2 < 1$ $\phi_2 - \phi_2 < 1, \lvert\phi_2\rvert < 1$	None
ARMA (1,1) $z_t = \delta + \phi_1 z_{t-1} + a_t - \theta_1 a_{t-1}$	$\lvert\phi_1\rvert < 1$	$\lvert\theta_1\rvert < 1$

Example 9.2/p. 407, Bowerman, O'Connell, Koehler, 4-th Edition – Forecasting, Time Series, and Regression

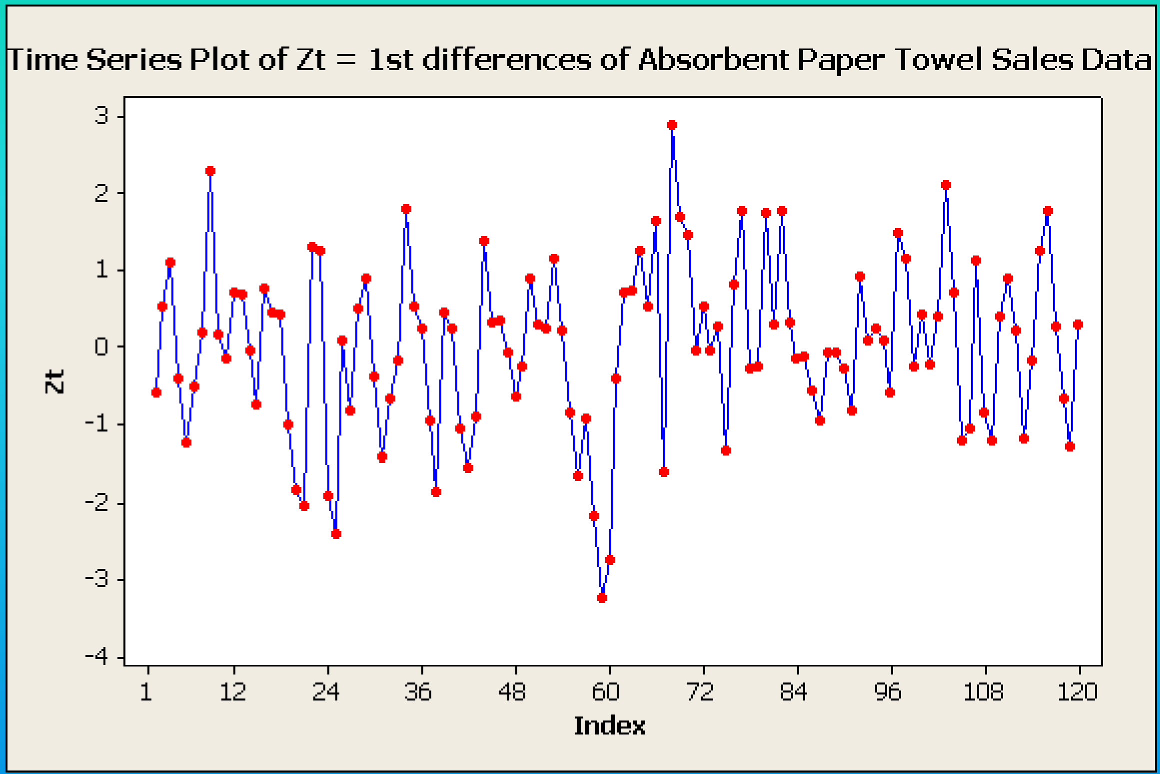
Absorbent Paper Towel Sales Data



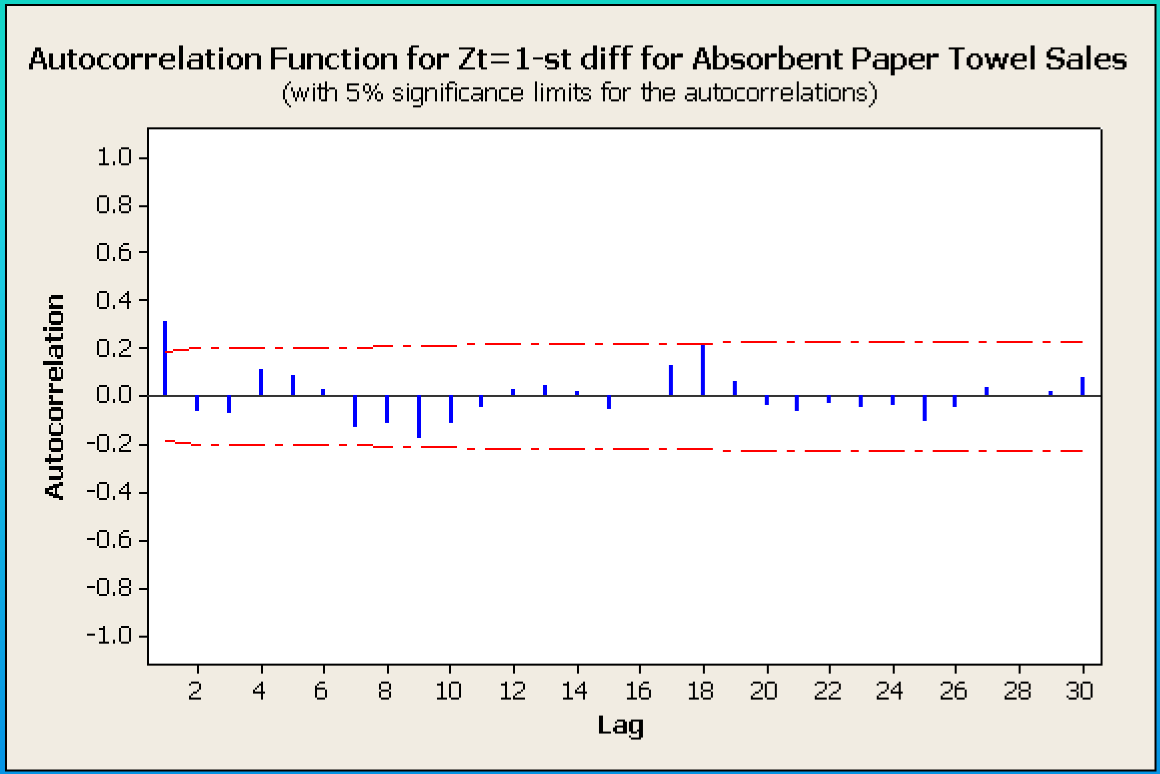
Yt do not seem to fluctuate around a constant mean so original time series does not appear to be stationary.



SAC of original time series dies down very slowly, another indication that original time series is not stationary.



Zt seems to fluctuate around a constant mean so time series of 1-st differences does appear to be stationary.

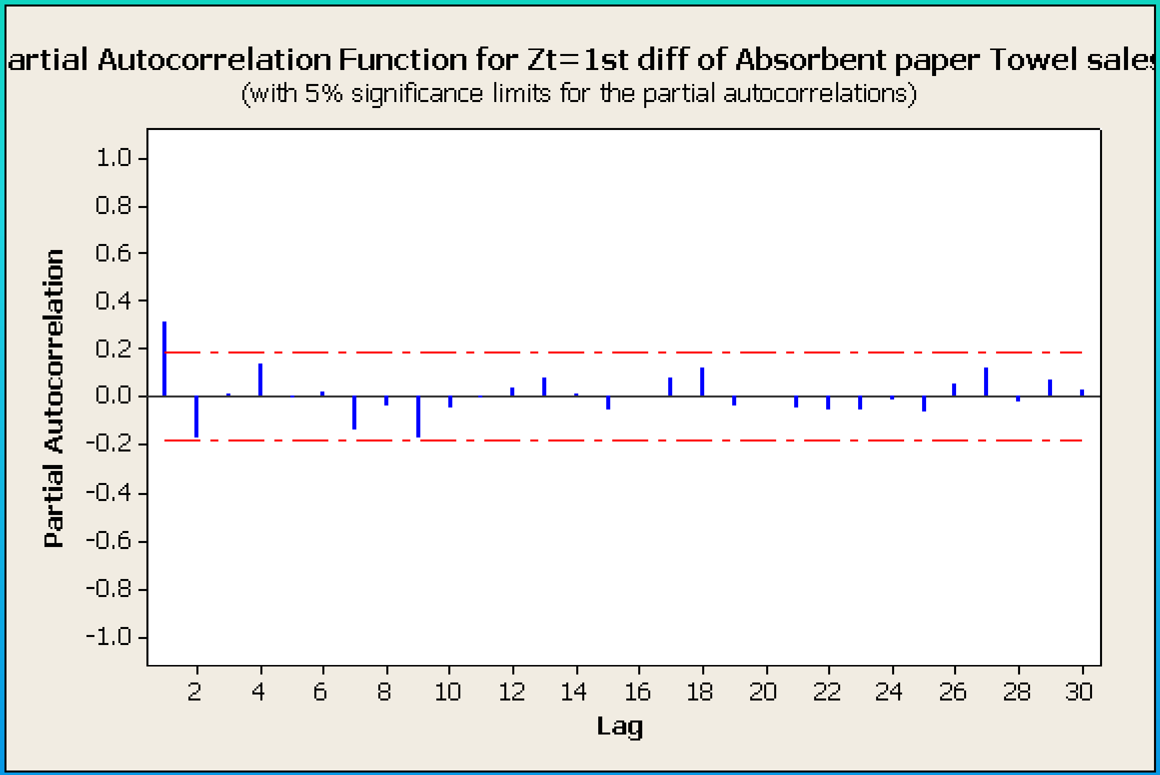


There is a spike at lag 1 (t = 3.34 > 2) and cuts off after lag 1.

SAC for 1-st differences of the Absorbent Paper Towel Sales time series

lag k	ACF2	TSTA2
1	0.306655	3.34521
2	-0.06474	-0.6479
3	-0.07166	-0.71468
4	0.104567	1.0384
5	0.084132	0.828
6	0.022841	0.22351
7	-0.13261	-1.29714
8	-0.11905	-1.1483
9	-0.17384	-1.65856
10	-0.11823	-1.10283

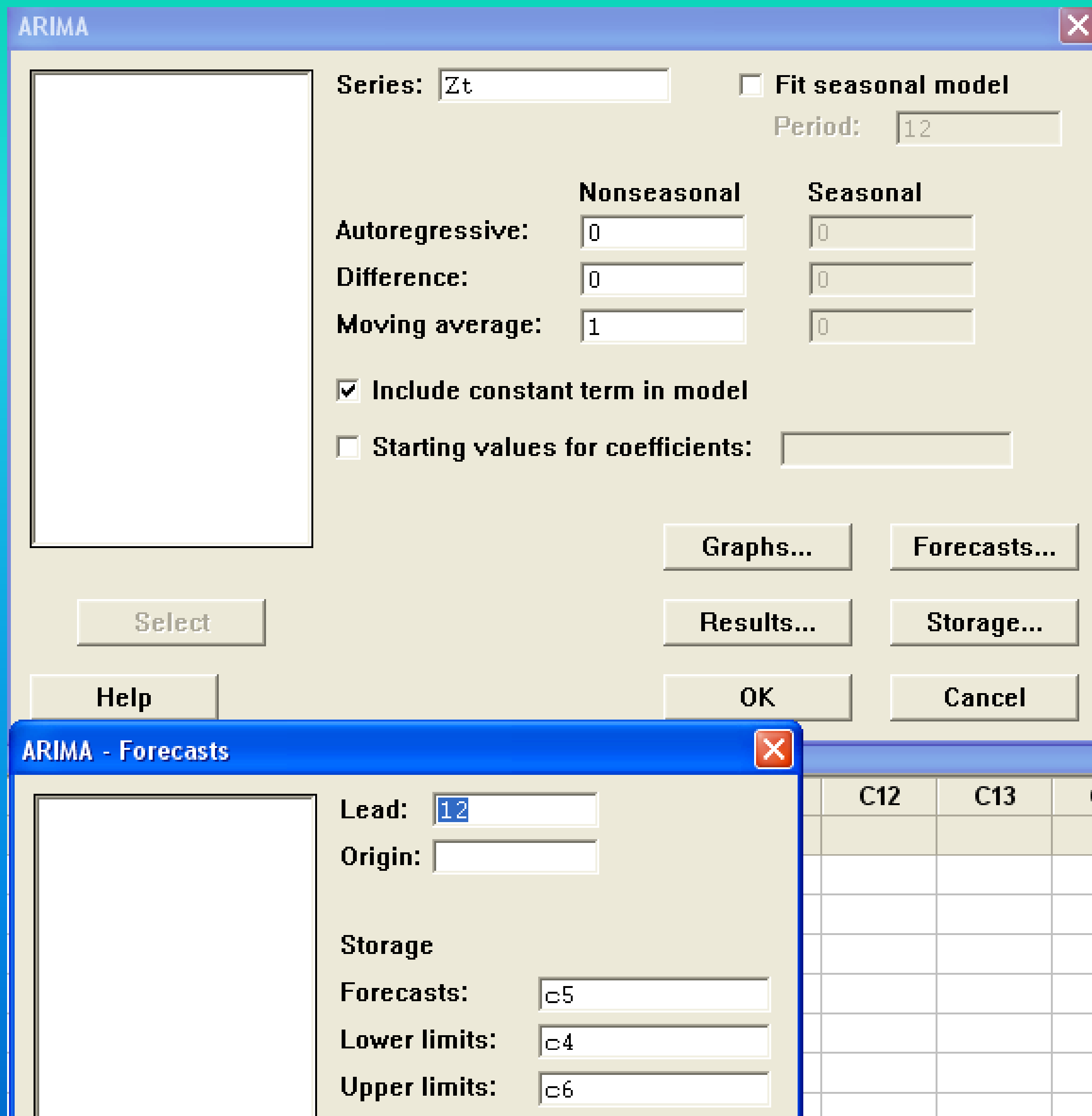
There is a spike at lag 1 (t = 3.34 > 2) and cuts off after lag 1.



PACF3	TSTA3
0.306655	3.34521
-0.175255	-1.91180
0.006163	0.06723
0.133346	1.45463
-0.009526	-0.10392
0.019915	0.21725
-0.139654	-1.52345
-0.040763	-0.44467
-0.177555	-1.93690
-0.052701	-0.57490
-0.010918	-0.11910

SPAC of Zt=1st diff of Absorbent Paper Towel Sales data oscillates and dies down

Hence a nonseasonal moving average (MA) model of order 1 should be used for this data.


ARIMA
Series: Zt
Fit seasonal model
Period: 12
Nonseasonal
Seasonal
Autoregressive: 0 0
Difference: 0 0
Moving average: 1 0
Include constant term in model
Starting values for coefficients:
Graphs...
Forecasts...
Select
Results...
Storage...
Help
OK
Cancel
ARIMA - Forecasts
Lead: 12
Origin:
Storage
Forecasts: c5
Lower limits: c4
Upper limits: c6
C12
C13

ARIMA Model: zt = 1st difference of Paper Towel Sales Data

Estimates at each iteration

Iteration	SSE	Parameters
0	153.997	0.100
1	139.734	-0.050
2	130.863	-0.200
3	127.423	-0.350
4	127.419	-0.355
5	127.419	-0.354
6	127.419	-0.354

Relative change in each estimate less than 0.0010

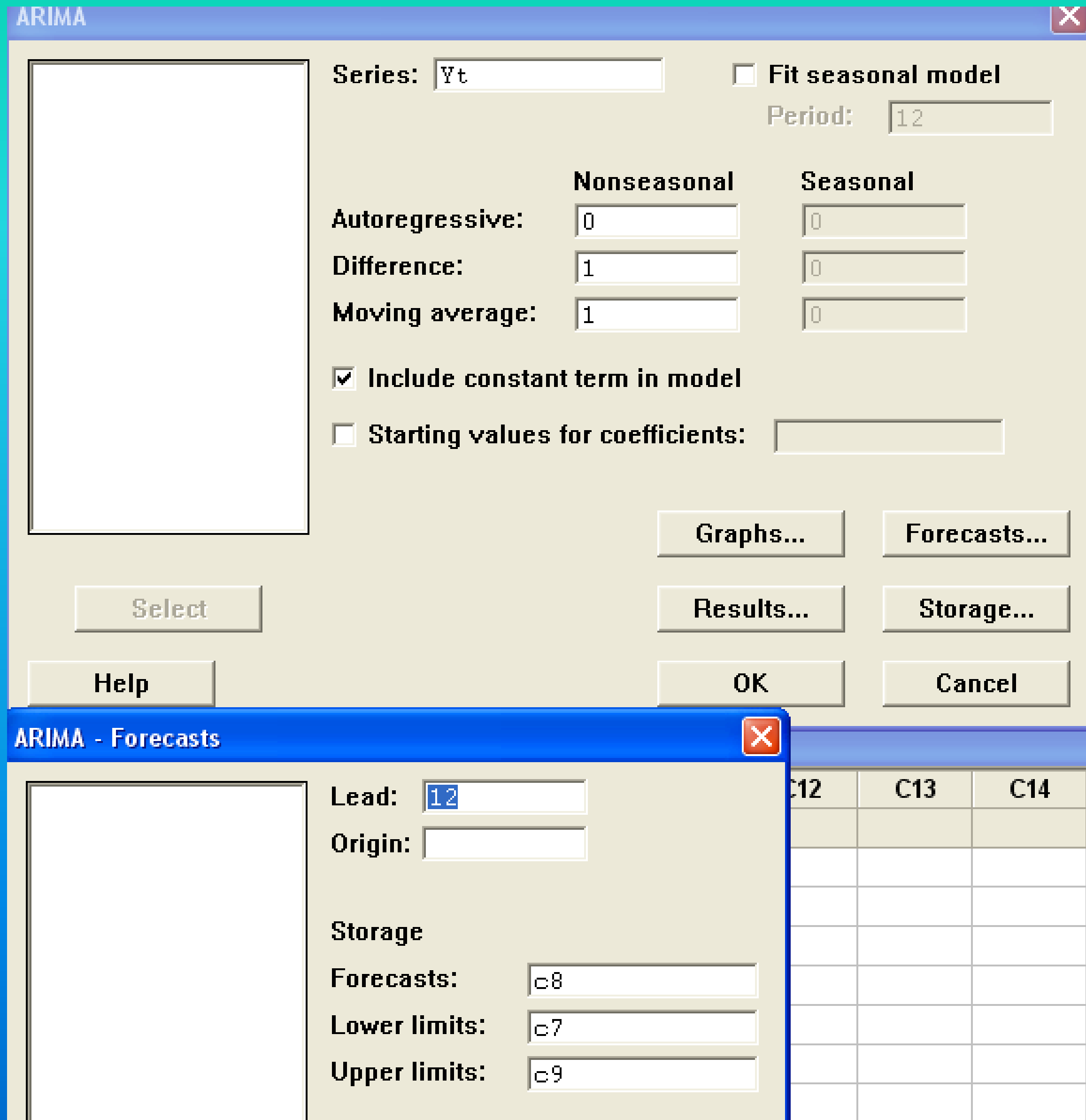
Final Estimates of Parameters

Type		Coef	SE Coef	T	P
MA	1	-0.3544	0.0864	-4.10	0.000

Number of observations: 119
Residuals: SS = 127.367 (backforecasts excluded)
MS = 1.079 DF = 118

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

Lag	12	24	36	48
Chi-Square	10.3	18.6	27.5	41.2
DF	11	23	35	47
P-Value	0.500	0.725	0.815	0.710


ARIMA
Series: Yt
Fit seasonal model
Period: 12
Nonseasonal
Seasonal
Autoregressive: 0 0
Difference: 1 0
Moving average: 1 0
Include constant term in model
Starting values for coefficients:
Graphs...
Forecasts...
Select
Results...
Storage...
Help
OK
Cancel
ARIMA - Forecasts
Lead: 12
Origin:
Storage
Forecasts: c8
Lower limits: c7
Upper limits: c9
C12
C13
C14

ARIMA Model: Yt

Estimates at each iteration

Iteration	SSE	Parameters
0	153.997	0.100
1	139.734	-0.050
2	130.863	-0.200
3	127.423	-0.350
4	127.419	-0.355
5	127.419	-0.354
6	127.419	-0.354

Relative change in each estimate less than 0.0010

Final Estimates of Parameters

Type		Coef	SE Coef	T	P
MA	1	-0.3544	0.0864	-4.10	0.000 (0, significant)

Analysis of Residuals from Box-Jenkins Model

- Compute the Ljung-Box Statistics

$$Q^* = n'(n'+2)\sum_{l=1}^{K}\frac{r_l^2(\hat{a})}{(n'-l)}$$

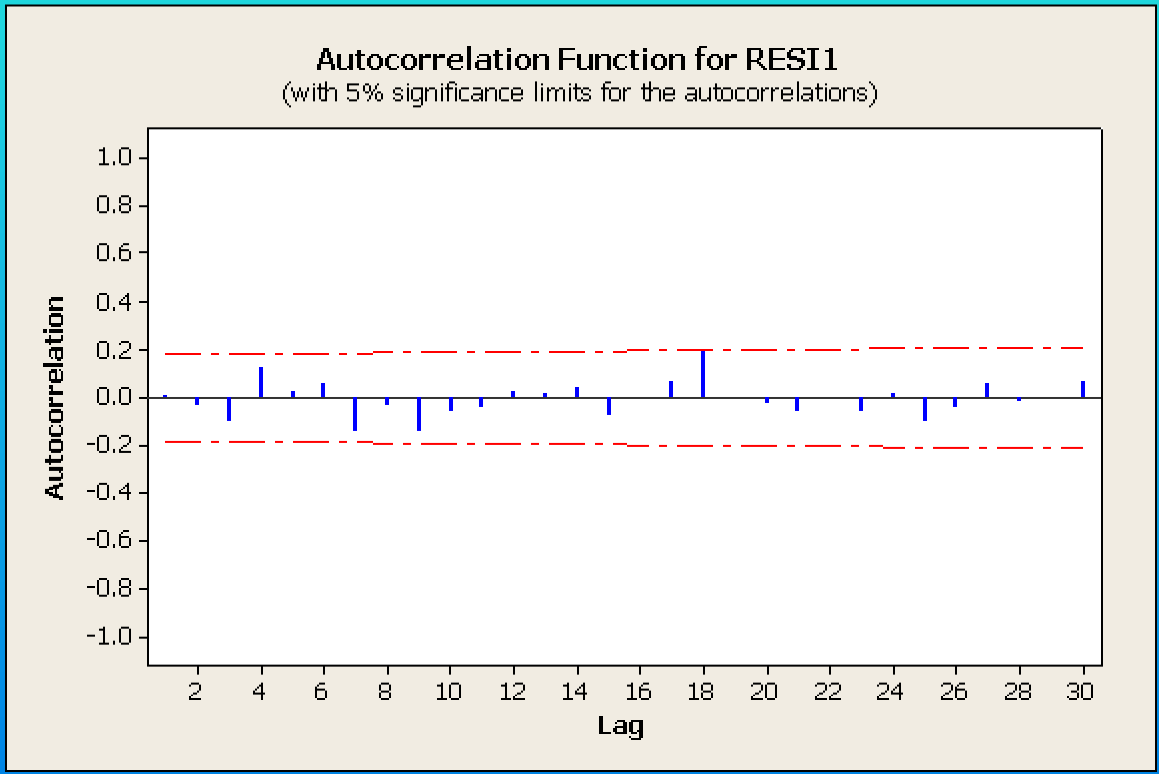
$n' = n - d, d =$ order of differencing to get a stationary series

$r_l^2(\hat{a}) =$ squared sample AC of residual at lag l

Reject H_0 : 1-st K $TAC = 0$, vs. H_1 : H_0 is false if Q^* is large $(>\chi^2_{K,\alpha})$, or $p = P(\chi^2_K > Q^*) < 0.05$

Notes: 1) some software packages use n instead of n'

2) typical choices for K are 6, 12, 18, 24.



Autocorrelation Function for RESI1
(with 5% significance limits for the autocorrelations)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
Autocorrelation
2
4
6
8
10
12
14
16
18
20
22
24
26
28
30
Lag

lag K	ACF1	TSTA1	LBQ1	p-value_from_R
1	0.004894	0.05339	0.0029	0.957053
2	-0.03586	-0.39122	0.1612	0.922563
3	-0.10234	-1.11495	1.4613	0.691231
4	0.128205	1.38236	3.5193	0.47495
5	0.027512	0.29199	3.6149	0.606078
6	0.061049	0.64747	4.0898	0.664525
7	-0.1416	-1.49655	6.6677	0.464279
8	-0.03073	-0.3188	6.7902	0.559429
9	-0.14622	-1.51572	9.5887	0.384796
10	-0.05787	-0.58868	10.0312	0.43776
11	-0.0405	-0.41075	10.2499	0.508059
12	0.025576	0.25903	10.3379	0.586343
13	0.019338	0.19574	10.3887	0.661882
14	0.038975	0.39439	10.597	0.717364
15	-0.07386	-0.74645	11.3524	0.727237

Computed in R as MINITAB 14 does not provide p-values

lag K	ACF1	TSTA1	LBQ1	p-value_from_R
16	-0.0023	-0.02309	11.3531	0.787184
17	0.067783	0.6818	12.0017	0.800034
18	0.196097	1.96482	17.4839	0.490106
19	0.002275	0.02209	17.4846	0.557073
20	-0.02703	-0.26246	17.5909	0.614338
21	-0.05983	-0.58059	18.1169	0.641595
22	-0.00062	-0.00602	18.1169	0.699035
23	-0.05497	-0.53195	18.5702	0.725869
24	0.012688	0.12248	18.5946	0.77325
25	-0.09794	-0.94534	20.0639	0.743465
26	-0.03979	-0.38122	20.309	0.776691
27	0.055507	0.53115	20.7913	0.795965
28	-0.01651	-0.1576	20.8344	0.832185
29	0.002756	0.0263	20.8356	0.865068
30	0.063576	0.60679	21.4895	0.87196

Computed in R as MINITAB 14 does not provide p-values

MINITAB 14 computes SAC, t-value, and Ljung-Box Q statistics, but does not output the corresponding P-values. The last column of the previous 2 slides (p-values) were computed in programming language R. You can calculate these in MINITAB, one line at a time. As an example, for K = 6 (slide 19), Ljung-Box Q* = LBQ = 4.0898.

In MINITAB:

Calc/probability distributions/chi-square

Pick cumulative probability, enter degrees of freedom = K = 6, input constant = 4.0898, and the click on OK to get

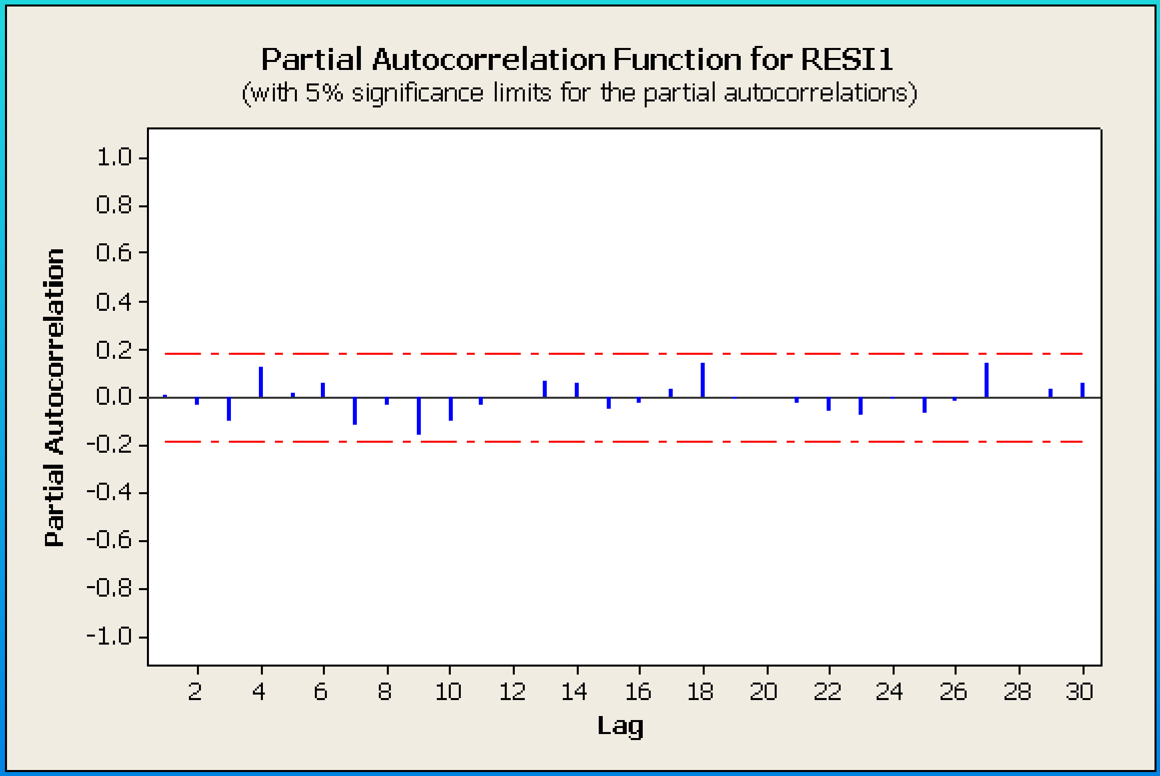
Chi-Square with 6 DF

x	P(X <= x)
4.0898	0.335475

from which p-value can be calculated : p = 1 - 0.335475 = 0.664525.

Since all p-values > .05, we conclude that there are no spikes in TAC and hence the model is adequate.

SPAC plot of RESIDUALS also shows the adequacy of the Box-Jenkins model chosen to analyze the Absorbent paper Towel Sales data.



Since the MA-1 model is found to be adequate for Paper Towel Sales data, we can use it to forecast future observations.

$100(1-\alpha)\%$ prediction interval for $y_{n+\tau}$ computed at time $t = n$ is $\hat{y}_{n+\tau} \pm t_{n-n_p,\alpha/2} SE_{n+\tau}(n)$

Function of *$s=SSE/(n-n_p)$*

Calculated by MINITAB

t*	lZthat95	Zthat	uZthat95	lYthat95	Ythat	uYthat95
121	-1.79666	0.248727	2.29411	13.8486	15.894	17.9394
122	-2.16451	0.005538	2.17559	12.456	15.8996	19.3431
123	-2.16451	0.005538	2.17559	11.4855	15.9051	20.3247
124	-2.16451	0.005538	2.17559	10.6946	15.9106	21.1267
125	-2.16451	0.005538	2.17559	10.0101	15.9162	21.8223
126	-2.16451	0.005538	2.17559	9.3982	15.9217	22.4453
127	-2.16451	0.005538	2.17559	8.8398	15.9273	23.0147
128	-2.16451	0.005538	2.17559	8.3232	15.9328	23.5424
129	-2.16451	0.005538	2.17559	7.8401	15.9383	24.0365
130	-2.16451	0.005538	2.17559	7.3849	15.9439	24.5028
131	-2.16451	0.005538	2.17559	6.9533	15.9494	24.9455
132	-2.16451	0.005538	2.17559	6.542	15.9549	25.3679

2) Double Exponential Smoothing is equivalent to forecasting with the Box-Jenkins MA-2 model

$$z_t = a_t - \theta_1 a_{t-1} \text{-} \theta_2 a_{t-2} \text{ where}$$

$$z_t = y_t - 2y_{t-1} + y_t - 2 \text{ (2-nd difference)}$$

$$\text{with } \theta_1 = 2 - \alpha - \gamma,\ \theta_2 = \alpha - 1$$

$$(\alpha, \gamma = \text{smoothing constants})$$