TIMESER LAB NOTES

MANAY, Justin Gabrielle A.

02-21-2019

STAGES IN ARIMA MODELING:

- ***Model Identification***
  - Check if the series is stationary. From the plot/ACF/PACF, determine if it has a trend or exhibits seasonality. You could also use the Augmented Dickey-Fuller test for stationarity. If it does, apply differencing (if there is a trend) or deseasonalization (if there is seasonality).
  - Check if the series is white noise using the Ljung-Box Q-statistic for white noise. If not white noise, check ACF/PACF to determine the correct ARIMA model.
- ***Parameter Estimation***
- ***Model Diagnostics***
  - Check if residuals are white noise using the Ljung-Box Q-statistic for white noise. If not white noise, there are misspecfications in the fitted model.
- ***Forecast Verification and Reasonableness***

As an example, we can take the ***W1*** dataset.

Model Identification

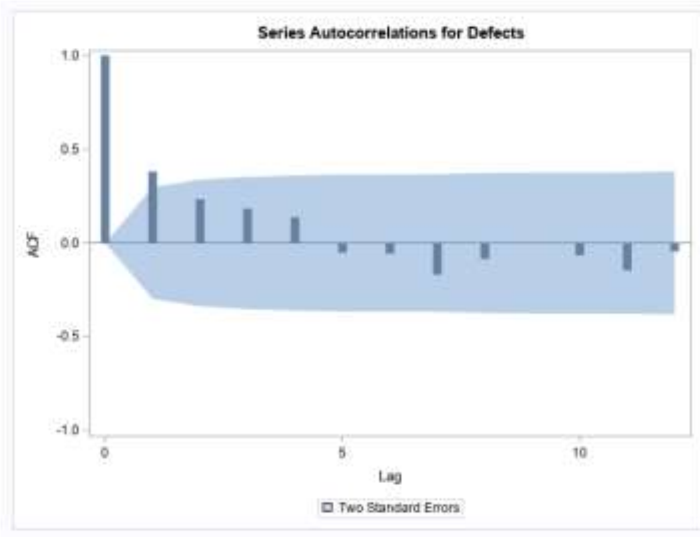
Warning: The value of NLAG is larger than 25% of the series length. The asymptotic approximations used for correlation based statistics and confidence intervals may be poor.

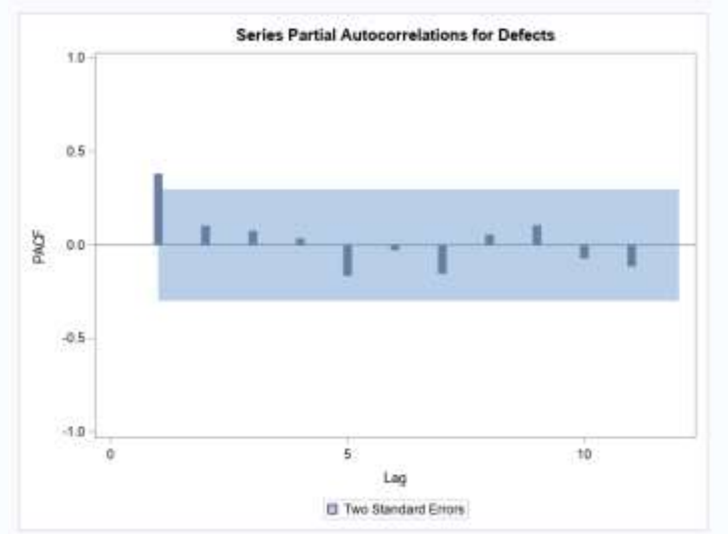
Name of Variable = Defects	
Mean of Working Series	1.766444
Standard Deviation	0.521012
Number of Observations	45

- The number of lags should be ¼ of the number of observations. In this case, the number of lags should be ¼(45) = 11.25. We round up and choose 12 lags, so SAS issues a warning since 12 > ¼(45).

Autocorrelation Check for White Noise									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	12.61	6	0.0496	0.381	0.233	0.183	0.137	-0.052	-0.057
12	16.36	12	0.1752	-0.169	-0.085	-0.002	-0.067	-0.147	-0.045

- Using the Ljung-Box Q-statistic for white noise,
  H_0: The series is white noise.
  H_a: The series is not white noise.
- We reject H_0 since at least one p-value < 0.05. Thus, *Defects* must not be white noise. Therefore, we proceed with fitting an ARIMA model.





- From the plots, ACF is exponentially decreasing, while PACF is insignificant after lag 1. Thus, we fit an AR(1) model.
- The shaded region determines the 95% confidence interval that the data is a white noise. Since there are values where the ACF/PACF exceeds the shaded region, the series *Defects* is not white noise.
- Note that ACF at lag 0 is always equal to 1 and that ACF at lag 1 is always equal to PACF at lag 1.
- You can alternatively use the IACF (Inverse Autocorrelation Function) plot if the trends in the ACF/PACF are not clear.

Parameter Estimation

Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	1.75530	0.11673	15.04	<.0001	0
AR1,1	0.38153	0.14103	2.71	0.0097	1

Constant Estimate	1.085603
Variance Estimate	0.242863
Std Error Estimate	0.492811
AIC	65.97205
SBC	69.58537
Number of Residuals	45

- Note that
  MU = μ
  AR1.1 = ϕ_1

Constant Estimate = θ_0

- In general, the resulting equation for an AR(1) model is:

$$(1 - \phi_1 B)Z_t = \theta_0 + a_t$$
$$Z_t = \theta_0 + \phi_1 Z_{t-1} + a_t$$

Taking the expectation of both sides,

$E[Z_t] = E[\theta_0 + \phi_1 Z_{t-1} + a_t]$

Thus,

$$\widehat{Z_t} = \widehat{\theta_0} + \widehat{\phi_1} Z_{t-1}$$

- For our model in particular, since all the p-values < 0.05,

$$\widehat{Z_t} = 1.0856 + 0.3815 Z_{t-1}$$

- Note that

$$(1 - \phi_1)\mu = \theta_0$$

should hold for an AR(1) model. Checking if it does,

$$(1 - 0.382)755 = 1.086$$

- The Akaike Information Criterion (AIC) and Schwarz/Bayesian Information Criterion (SBC) are useful for comparing different models (the lower, the better).

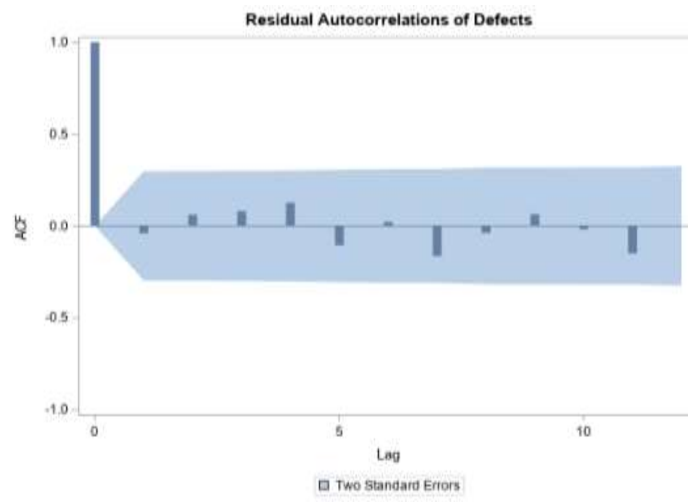
Model Diagnostics

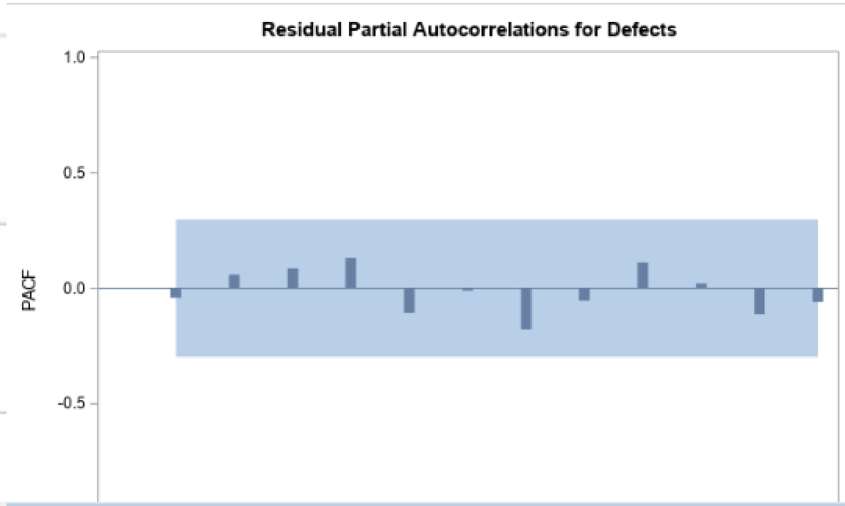
Correlations of Parameter Estimates		
Parameter	**MU**	**AR1,1**
MU	1.000	-0.017
AR1,1	-0.017	1.000

- The pairwise correlations are small (-0.017). Thus, there is no significant multicollinearity among the parameters of the AR model (i.e., there is no model redundancy).

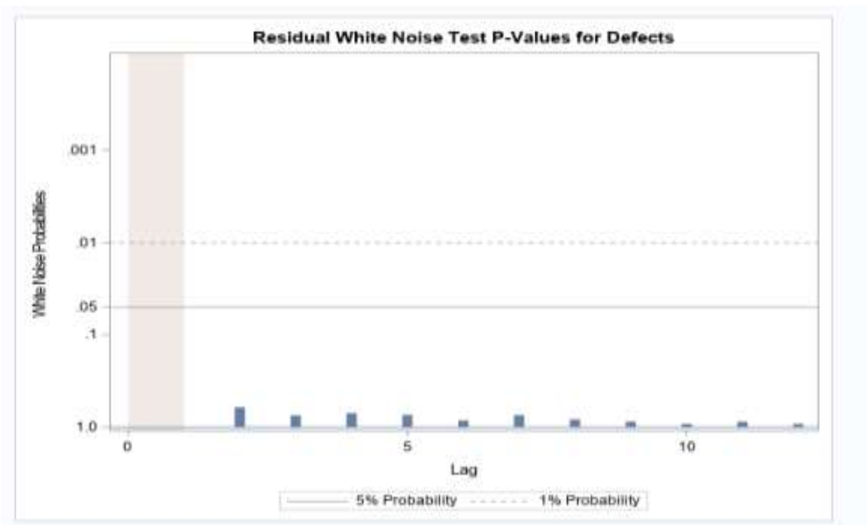
Autocorrelation Check of Residuals									
To Lag	**Chi-Square**	**DF**	**Pr > ChiSq**	**Autocorrelations**					
6	2.02	5	0.8462	-0.040	0.061	0.081	0.125	-0.106	0.022
12	5.28	11	0.9166	-0.165	-0.039	0.064	-0.020	-0.150	-0.002
18	11.22	17	0.8448	0.081	-0.067	-0.080	-0.096	0.091	-0.210
24	12.14	23	0.9683	-0.021	0.007	-0.068	-0.064	0.028	-0.013

- Autocorrelations are small between lags, which can potentially signal that the residuals are white noise.
- Using the Ljung-Box Q-statistic for white noise,
  H_0: The residuals are white noise.
  H_a: The residuals are not white noise.
- Since the p-values > 0.05, we do not reject H_0. Thus, the errors must be white noise. Thus, there is no information in the residuals which needs to be included in the model.

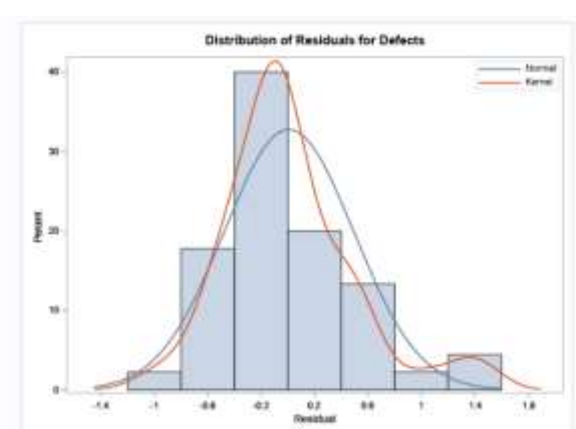


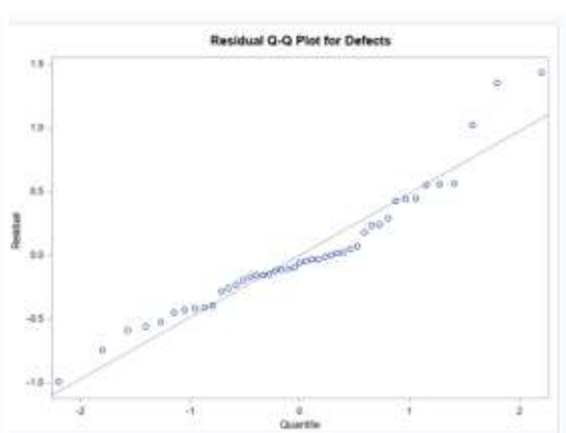


- The ACF is insignificant after lag 1, and the PACF is insignificant all throughout since the residuals are white noise.



- The residual white noise test p-values must be large to not reject H_0 and conclude that the residuals are white noise. This plot shows that the p-values are closer to 1, enabling us to not reject H_0 and conclude that the residuals are white noise.





- We want the residuals to be normally distributed. The QQ-plot closely follows the normal line, potentially indicating that the residuals are normal.

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.93041	Pr < W	0.0097
Kolmogorov-Smirnov	D	0.155872	Pr > D	<0.0100
Cramer-von Mises	W-Sq	0.177979	Pr > W-Sq	0.0096
Anderson-Darling	A-Sq	1.038387	Pr > A-Sq	0.0091

- We can also test for the normality of the residuals. Using the different normality tests,
  H_0: The residuals are normally distributed.
  H_a: The residuals are not normally distributed.
- Since p-value > 0.05, we reject H_0. Thus, the errors must not be normally distributed. REMEDY FOR NON-NORMALITY

- Note that any pattern in the residual ACF, residual PACF and residual plots will also tell you what components must be included that were omitted from the model.

Model for variable Defects	
Estimated Mean	1.755297

Autoregressive Factors	
Factor 1:	1 - 0.38153 B**(1)

- Note that the estimated mean is

$$\mu = 1.755297$$

and the autoregressive factors are

$$(1 - \phi_1 B) = 1 - 0.38153B$$

Forecast Verification and Reasonableness

Forecasts for variable Defects						
Obs	Forecast	Std Error	95% Confidence Limits		Actual	Residual
1	1.7553	0.4928	0.7894	2.7212	1.2000	-0.5553
2	1.5434	0.4928	0.5775	2.5093	1.5000	-0.0434
3	1.6579	0.4928	0.6920	2.6238	1.5400	-0.1179
4	1.6732	0.4928	0.7073	2.6390	2.7000	1.0268
5	2.1157	0.4928	1.1498	3.0816	1.9500	-0.1657

- The above table is a snippet of the forecast table outputted by SAS. From the forecast equation

$$\widehat{Z_t} = 1.0856 + 0.3815Z_{t-1}$$

At time 45,

$$\begin{aligned}\widehat{Z_{45}} &= 1.0856 + 0.3815(1.84) \\ &= 1.7876\end{aligned}$$

At time 47, there is no previous actual value. Thus, SAS uses the previous forecasted value.

$$\begin{aligned}\widehat{Z_{46}} &= 1.0856 + 0.3815(1.7876) \\ &= 1.7676\end{aligned}$$

- To test for overfitting, we can examine the AR(2) model.

Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	1.74991	0.12982	13.48	<.0001	0
AR1,1	0.34229	0.15352	2.23	0.0312	1
AR1,2	0.10422	0.15354	0.68	0.5010	2

- Note that the p-value for AR1,2 > 0.05. Thus, we drop AR1,2 and only an AR(1) model is needed.

Constant Estimate	0.96855
Variance Estimate	0.245958
Std Error Estimate	0.495942
AIC	67.48304
SBC	72.90302
Number of Residuals	45

- AIC for AR(2) model = 67.48. Compared to the AIC for AR(1) which is 65.97, AR(1) is the better model.
- We can also check if

$$(1 - \phi_1 - \phi_2)\mu = \theta_0$$

holds. Substituting,

$$(1 - 0.34229 - 0.10422)1.74991 = 0.96855$$

- Thus, the model is an AR(1) model, and for *Defects*, the value in the immediately succeeding period depends only on the value on the immediately previous period.

02-28-2019

NONSEASONAL ARIMA(p, d, q) MODEL

- The ARIMA(p, d, q) model is defined by the following equation:

$$\left(1-\phi_1 B-\cdots-\phi_p B^p\right)(1-B)^d Z_t=\theta_0+\left(1-\theta_1 B-\cdots-\theta_q B^q\right) a_t$$

Taking expectations on both sides,

$$\left(1-\phi_1 B-\cdots-\phi_p B^p\right)(1-B)^d E[Z_t]=\theta_0+\left(1-\theta_1 B-\cdots-\theta_q B^q\right) E[a_t]$$
$$\left(1-\phi_1 B-\cdots-\phi_p B^p\right)(1-B)^d \mu=\theta_0$$

To simplify, we examine ARMA(p, q). Note that ARMA(p,q) = ARIMA(p,0,q).

$$\left(1-\phi_1 B-\cdots-\phi_p B^p\right)\mu=\theta_0$$
$$(1-\phi_1-\cdots-\phi_p)\mu=\theta_0$$

Again, to simplify, we examine AR(1), AR(2),… and MA(1), MA(2), ….

AR(1): $1-\phi_1\mu=\theta_0$

AR(2): $(1-\phi_1-\phi_2)\mu=\theta_0$

…

AR(p): $\boldsymbol{(1-\phi_1-\phi_2-\ldots-\phi_p)\mu=\theta_0}$

Thus, for a pure AR process, the equations relate the AR coefficients with the constant term in the MA model. Note that for a pure AR process, $\mu=0 \Leftrightarrow \theta_0=0$.

MA(1): $\mu=\theta_0$

MA(2): $\mu=\theta_0$

…

MA(q): $\boldsymbol{\mu=\theta_0}$

Thus, for a pure MA process, the constant term is equal to the mean of the series μ.

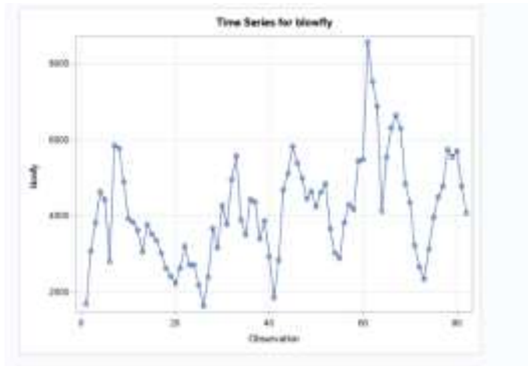
W3 Dataset

Autocorrelation Check for White Noise									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	80.85	6	<.0001	0.735	0.499	0.301	0.212	0.131	0.030
12	81.33	12	<.0001	-0.009	-0.047	-0.020	-0.032	-0.020	0.032
18	113.11	18	<.0001	0.103	0.191	0.257	0.284	0.267	0.199

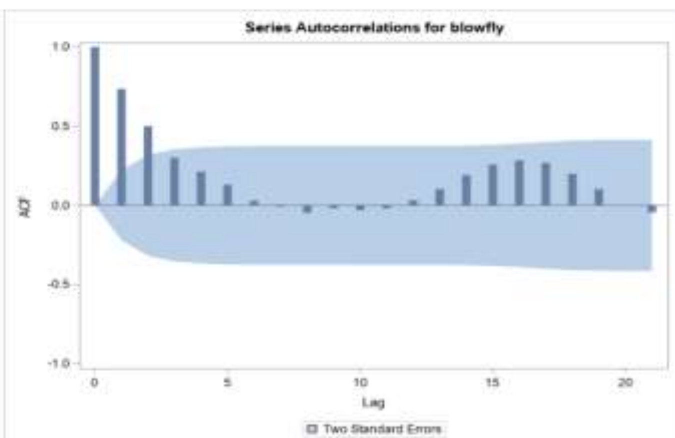
- Using the Ljung-Box Q-statistic for white noise,
  H_0: The series is white noise.
  H_a: The series is not white noise.
- Since all p-values < 0.05, we reject H_0. Thus, the series *Blowfly* must not be white noise. Therefore, we proceed with fitting an ARIMA model.
- If the series is a white noise,
  $$\widehat{Z}_t = \theta_0$$
  where $\theta_0 = \mu$. To confirm this, we use the Augmented Dickey-Fuller test.

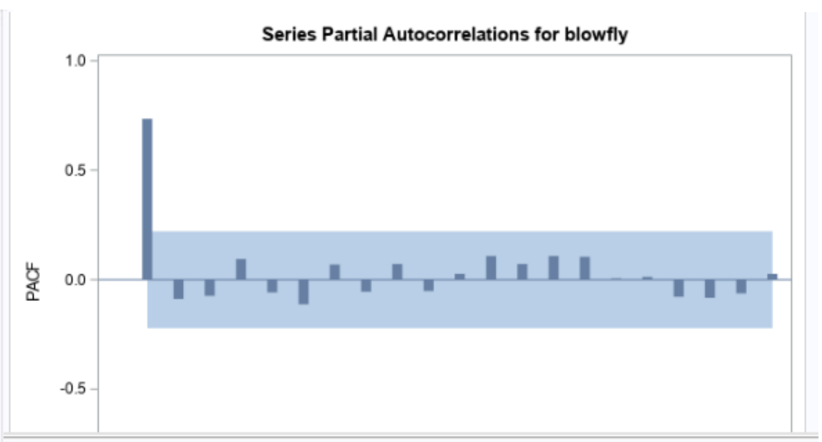
Augmented Dickey-Fuller Unit Root Tests							
Type	**Lags**	**Rho**	**Pr < Rho**	**Tau**	**Pr < Tau**	**F**	**Pr > F**
Zero Mean	0	-1.5687	0.3834	-0.80	0.3689		
	1	-1.6762	0.3690	-0.85	0.3439		
	2	-1.5710	0.3831	-0.84	0.3518		
Single Mean	0	-21.4770	0.0052	-3.64	0.0068	6.67	0.0016
	1	-25.2816	0.0017	-3.55	0.0091	6.30	0.0096
	2	-29.8645	0.0008	-3.48	0.0111	6.05	0.0148
Trend	0	-24.1958	0.0191	-3.74	0.0249	7.16	0.0324
	1	-30.8432	0.0033	-3.81	0.0209	7.29	0.0288
	2	-40.1224	0.0002	-3.86	0.0185	7.45	0.0249

- The Augmented Dickey-Fuller test is used to test for stationarity.
- Do NOT use the last column since the F-statistic assumes normality. Recall that an F-distributed variable is the ratio of two chi-squared distributed variables, and the chi-squared distribution is itself the square of the normal distribution.
- Thus, we use the tau-statistic.
- The Augmented Dickey-Fuller (ADF) test of stationarity is also called the ADF test of random walk (because as ϕ_1 approaches 1, the series becomes a random walk/non-stationary) or the ADF unit root test.
- Using the ADF test,
  H_0: A unit root is present (i.e., not stationary)
  H_a: There is no unit root (i.e., stationary).



- There are three options: zero mean (i.e., The series is moving up and down around a center of zero), single mean (similar to a random walk with a drift term) and trend (i.e., The series has a linear trend that is deterministic.)
- From the time plot, the movement of the series *Blowfly* is far from zero, so it does not have zero mean.
- All p-values < 0.05 so we reject H_0. Thus, the series is stationary and there is no need for differencing since there is no significant trend.





- From the plots, ACF is exponentially decreasing. PACF is insignificant after lag 1. Thus, we fit an AR(1) model.

Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	3824.7	386.83442	9.89	<.0001	0
AR1,1	0.75300	0.07520	10.01	<.0001	1

Constant Estimate	944.6945
Variance Estimate	853637.2
Std Error Estimate	923.9249
AIC	1354.577
SBC	1359.39
Number of Residuals	82

- The equation is:

$$\widehat{Z}_t = 944.69 + 0.7532Z_{t-1}$$

or

$$(1 - 0.7532B)Z_t = 944.69$$

Correlations of Parameter Estimates		
Parameter	MU	AR1,1
MU	1.000	-0.204
AR1,1	-0.204	1.000

- The pairwise correlations are small (-0.204). Thus, there is no significant multicollinearity among the parameters of the AR model.

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	2.73	5	0.7409	0.063	0.018	-0.131	0.025	0.050	-0.082
12	4.78	11	0.9411	0.002	-0.111	0.060	-0.039	-0.062	-0.023
18	10.12	17	0.8983	-0.005	0.072	0.106	0.113	0.127	0.075
24	12.52	23	0.9616	0.032	-0.085	-0.090	-0.015	0.064	0.022

- Using the Ljung-Box Q-statistic for white noise,

  H_0: The residuals are white noise.

H_a: The residuals are not white noise.

- All p-values > 0. Thus, we do not reject H_0. Thus, the errors must be white noise. Thus, there is no information in the residuals which needs to be included in the model.
- To check for overfitting, we examine the AR(2) model.

Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	3865.9	357.58585	10.81	<.0001	0
AR1,1	0.81366	0.11272	7.22	<.0001	1
AR1,2	-0.08601	0.11276	-0.76	0.4478	2

Constant Estimate	1052.866
Variance Estimate	858211.9
Std Error Estimate	926.3973
AIC	1355.983
SBC	1363.204
Number of Residuals	82

- AR1,2 is not significant. Thus, we drop it from the model. Thus, an AR(1) model is a better fit for the data.
- AIC is not significantly different from AR(1) but we prefer AR(1) by the principle of parsimony.

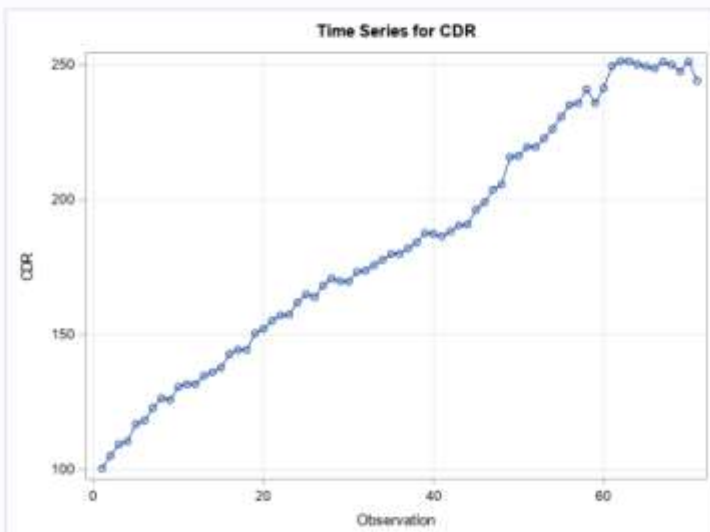
W5 dataset

Autocorrelation Check for White Noise									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	338.83	6	<.0001	0.960	0.918	0.878	0.835	0.793	0.751
12	521.51	12	<.0001	0.709	0.666	0.620	0.575	0.531	0.488
18	588.57	18	<.0001	0.446	0.402	0.359	0.318	0.279	0.239

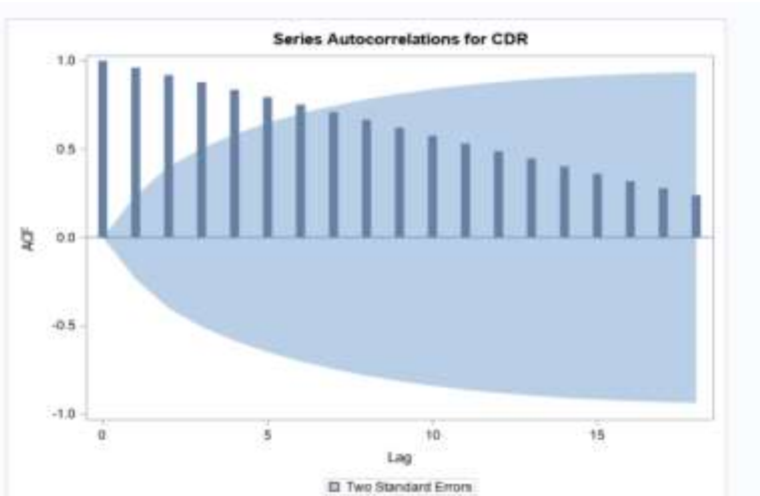
- Using the Ljung-Box Q-statistic for white noise,
  H_0: The series is white noise.
  H_a: The series is not white noise.
- All p-values < 0.05 so we reject H_0. Thus, the series is not white noise.

Augmented Dickey-Fuller Unit Root Tests							
Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Zero Mean	0	0.6858	0.8472	5.29	0.9999		
	1	0.6701	0.8436	4.36	0.9999		
	2	0.6350	0.8353	2.95	0.9991		
Single Mean	0	-0.9840	0.8847	-1.93	0.3170	21.76	0.0010
	1	-0.8345	0.8981	-1.92	0.3223	16.92	0.0010
	2	-0.7793	0.9027	-1.62	0.4677	8.76	0.0010
Trend	0	-6.4688	0.6915	-1.35	0.8666	2.53	0.6745
	1	-3.7759	0.8942	-0.88	0.9527	2.06	0.7676
	2	-4.9274	0.8170	-0.97	0.9412	1.63	0.8508

- H_0: A unit root is present (i.e., not stationary)
  H_a: There is no unit root (i.e., stationary)
- All p-values (for *trend*) > 0.05 so we do not reject H_0. Thus, the series is not stationary.



- The time plot shows that the series has a trend, so we use trend.



- Another indication of non-stationarity is an ACF plot which is slowly decreasing.

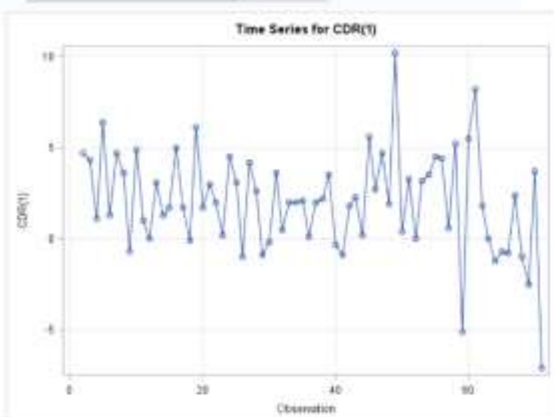
If we mistakenly fit an AR(1) model,

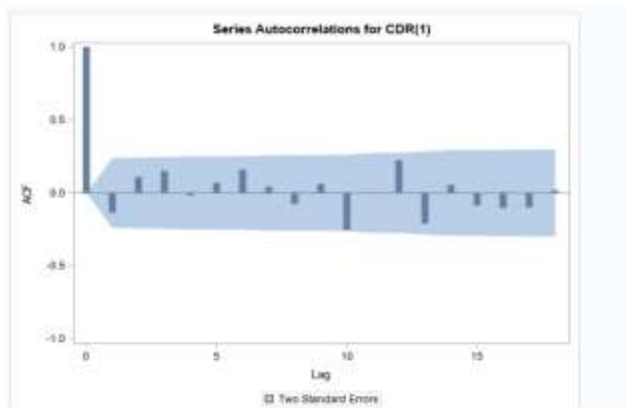
Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	107.76845	3.57563	30.14	<.0001	0
AR1,1	1.00000	0.0048789	204.96	<.0001	1

Constant Estimate	0.000011
Variance Estimate	12.78516
Std Error Estimate	3.575634
AIC	384.3888
SBC	388.9141
Number of Residuals	71

- AR1,1 = 1. However, |AR1,1| < 1 and when AR1,1 = 1, we have a random walk.

- Since the data is not white noise and is non-stationary, we employ differencing on the data; i.e., ARIMA(0, 1, 0)





- At first glance, the data now seems stationary. Also, all of the lags are within the blue band.

Autocorrelation Check for White Noise									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	6.35	6	0.3852	-0.135	0.110	0.150	-0.016	0.068	0.160
12	17.03	12	0.1486	0.042	-0.077	0.061	-0.253	0.005	0.224
18	23.83	18	0.1608	-0.209	0.056	-0.087	-0.103	-0.099	0.018

- From the Ljung-Box Q-statistic, all p-values > 0.05, so we do not reject H_0, signifying that the data is white noise.

Augmented Dickey-Fuller Unit Root Tests							
Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Zero Mean	0	-48.3738	<.0001	-5.94	<.0001		
	1	-21.5790	0.0006	-3.32	0.0012		
	2	-9.1685	0.0325	-2.08	0.0367		
Single Mean	0	-80.0712	0.0007	-8.86	0.0001	39.42	0.0010
	1	-64.1872	0.0007	-5.25	0.0001	13.90	0.0010
	2	-34.3883	0.0006	-3.22	0.0231	5.32	0.0322
Trend	0	-82.4550	0.0002	-9.19	<.0001	42.35	0.0010
	1	-70.6827	0.0002	-5.51	0.0001	15.21	0.0010
	2	-40.1851	0.0002	-3.48	0.0498	6.23	0.0626

- Furthermore, from the Dickey-Fuller tests, the p-values are < 0.05, so we reject H_0. Thus, the data is now stationary after differencing.
- Thus, after differencing once, the data becomes stationary and is a random walk.
- In general, for a random walk,

$$Z_t = \theta_0 + Z_{t-1} + a_t$$

- We just have to determine if $\theta_0 = 0$ (stochastic trend) or θ_0 is significantly different from zero (with drift)

Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	2.05429	0.33203	6.19	<.0001	0

Constant Estimate	2.054286
Variance Estimate	7.71701
Std Error Estimate	2.777951
AIC	342.6841
SBC	344.9326
Number of Residuals	70

- In this case, $\mu = \theta_0$. Thus, $\theta_0 = 0 \Longrightarrow \mu = 0$ for a random walk with a stochastic trend.
- The drift term is 2.05429, which is the slope of the random walk with a drift. Thus, the equation for this series is

$$\widehat{Z_t} = 2.05 + Z_{t-1}$$

03-05-2019

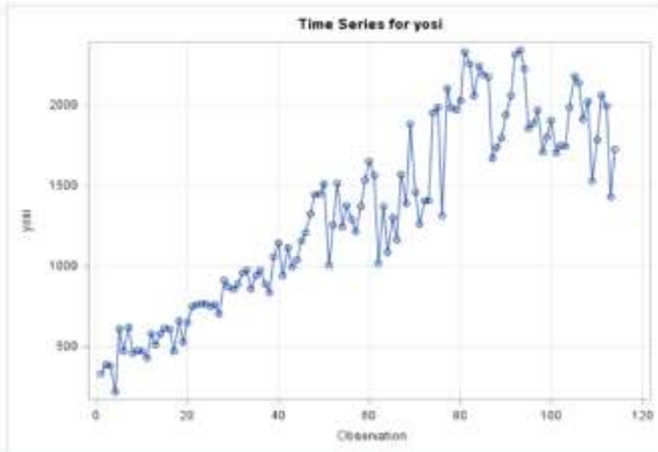
W6 dataset

Autocorrelation Check for White Noise									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	508.22	6	<.0001	0.909	0.880	0.849	0.818	0.798	0.790
12	894.80	12	<.0001	0.771	0.756	0.716	0.697	0.682	0.655
18	1152.37	18	<.0001	0.626	0.616	0.580	0.549	0.521	0.494
24	1303.77	24	<.0001	0.482	0.451	0.429	0.410	0.378	0.364
30	1378.35	30	<.0001	0.356	0.320	0.295	0.265	0.237	0.220

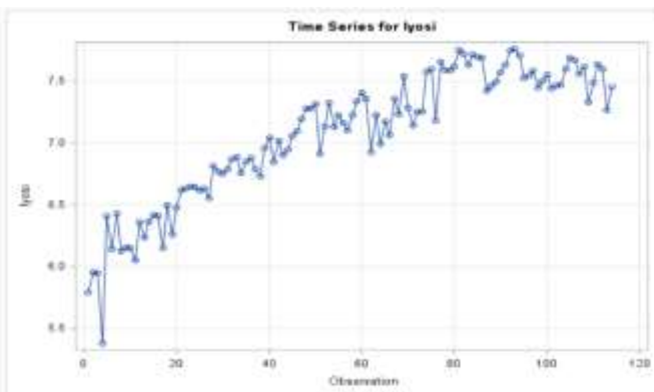
- Using the Ljung-Box Q-statistic for white noise,
  H_0: The series is white noise.
  H_a: The series is not white noise.
- All p-values < 0.05 so we reject H_0. Thus, the series is not white noise.

Augmented Dickey-Fuller Unit Root Tests							
Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Zero Mean	0	-0.6788	0.5310	-0.41	0.5339		
	1	0.0008	0.6814	0.00	0.6814		
	2	0.2337	0.7366	0.25	0.7580		
Single Mean	0	-9.7650	0.1342	-2.41	0.1424	3.07	0.2880
	1	-5.3875	0.3916	-1.87	0.3431	2.08	0.5412
	2	-4.0996	0.5219	-1.71	0.4238	1.96	0.5711
Trend	0	-49.9934	0.0004	-5.43	0.0001	14.88	0.0010
	1	-32.2927	0.0029	-3.60	0.0340	6.76	0.0385
	2	-25.1767	0.0172	-2.91	0.1619	4.60	0.2572

- H_0: A unit root is present (i.e., not stationary)
  H_a: There is no unit root (i.e., stationary)
- Most p-values (for *trend*) > 0.05 so we do not reject H_0. Thus, the series is not stationary.



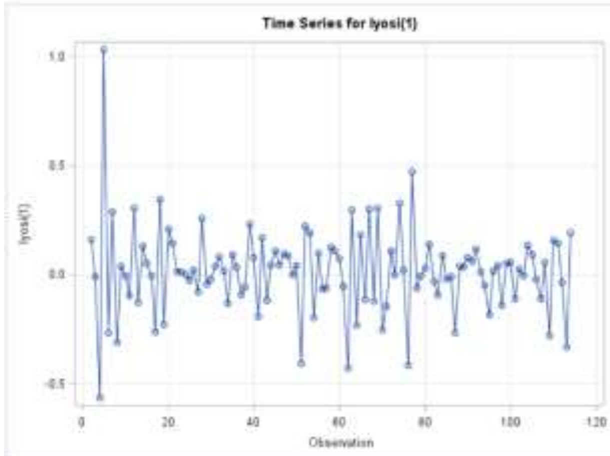
- From the time plot, we see that the series yosi is nonstationary in both mean and variance. Thus, we employ both (ordinary) differencing and log transformation. We do log tranformation before differencing because negative differences are undefined for the log function.
- After log transformation,



- The series lyosi is now stationary in the variance.

Augmented Dickey-Fuller Unit Root Tests							
Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Zero Mean	0	0.1753	0.7223	0.59	0.8416		
	1	0.1828	0.7241	1.07	0.9255		
	2	0.1855	0.7247	1.31	0.9519		
Single Mean	0	-10.7663	0.1041	-2.75	0.0693	4.11	0.0819
	1	-5.3289	0.3969	-2.26	0.1867	3.35	0.2180
	2	-4.4504	0.4837	-2.25	0.1888	3.67	0.1353
Trend	0	-46.5001	0.0004	-5.36	0.0001	14.72	0.0010
	1	-19.6727	0.0615	-2.95	0.1512	5.20	0.1388
	2	-14.5382	0.1816	-2.41	0.3737	4.05	0.3687

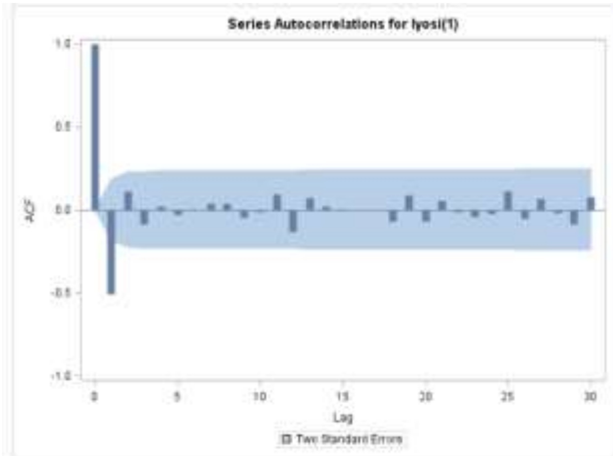
- However, from the Augmented Dickey-Fuller test, some p-values (for *trend*) > 0.05, so we also have to do (ordinary) differencing because the series is not stationary in the mean. The time plot for lyosi from earlier supports this.
- After ordinary differencing,

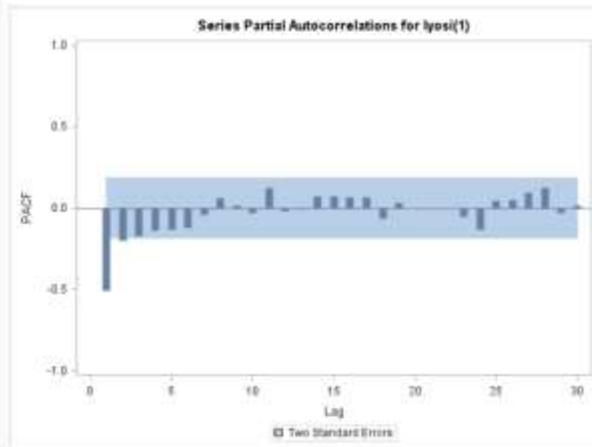


- The series is now stationary in both mean and variance.

Augmented Dickey-Fuller Unit Root Tests							
Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Zero Mean	0	-168.562	0.0001	-18.34	<.0001		
	1	-241.198	0.0001	-10.86	<.0001		
	2	-422.946	0.0001	-8.66	<.0001		
Single Mean	0	-169.315	0.0001	-18.43	<.0001	169.84	0.0010
	1	-251.498	0.0001	-11.02	<.0001	60.71	0.0010
	2	-566.501	0.0001	-9.06	<.0001	41.02	0.0010
Trend	0	-170.293	0.0001	-18.52	<.0001	171.52	0.0010
	1	-261.246	0.0001	-11.18	<.0001	62.48	0.0010
	2	-769.364	0.0001	-9.55	<.0001	45.86	0.0010

- From the Augmented Dickey-Fuller test, all p-values (for *zero mean*) < 0.05, so the series is now stationary.





- Since the ACF is only significant at lag 1 and the PACF is exponentially decreasing, we fit an MA(1) model; i.e., an ARIMA(0, 1, 1) model.

Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	0.01460	0.0049567	2.95	0.0039	0
MA1,1	0.68289	0.07063	9.67	<.0001	1

Constant Estimate	0.014602
Variance Estimate	0.026738
Std Error Estimate	0.163519
AIC	-86.585
SBC	-81.1302
Number of Residuals	113

- All parameters have a p-value < 0.05, meaning that they are all significant.
- Furthermore, note that for a pure MA model, $\mu = \theta_0$, which is true here.

Correlations of Parameter Estimates		
Parameter	MU	MA1,1
MU	1.000	0.057
MA1,1	0.057	1.000

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	1.49	5	0.9141	-0.019	0.066	-0.074	-0.028	-0.026	0.032
12	4.87	11	0.9373	0.093	0.084	0.002	0.028	0.082	-0.061
18	7.24	17	0.9802	0.076	0.074	0.040	0.004	-0.031	-0.064
24	8.18	23	0.9981	0.046	-0.038	0.035	-0.008	-0.028	0.033
30	14.62	29	0.9877	0.153	0.052	0.073	-0.036	-0.100	0.013

- The correlation between the parameters is low, indicating non-reduncancy. The residuals are also white noise based on the Ljung-Box Q-statistic (can also be checked through ACF/PACF plots of the residuals).
- To check, we employ overfitting and try to fit an MA(2) model.

Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	0.01458	0.0050807	2.87	0.0049	0
MA1,1	0.69863	0.09540	7.32	<.0001	1
MA1,2	-0.02193	0.09693	-0.23	0.8214	2

Constant Estimate	0.014583
Variance Estimate	0.026965
Std Error Estimate	0.16421
AIC	-84.6539
SBC	-76.4717
Number of Residuals	113

- MA1,2 has a p-value > 0.05, meaning that we should drop it from the model.
- The AIC and SBC for the MA(2) model are less than those for the MA(1) model. However, the differences are not significant (A rule of thumb is that the AIC and SBC differences are significant if they differ by more than 10) so by the rule of parsimony, we prefer the MA(1) model.
- Thus, the final model is given by:

$$\widehat{Z_t} = 0.01460 + Z_{t-1} + 0.31711 a_t$$

03-07-2019

- MULTIPLICATIVE

$$(1 - \phi_1 B)(1 - \Phi_{S=4} B^4)$$

$$AR(p = 1, d = 0, q = 0)AR(P = 1, D = 0, Q = 0)$$

ADDITIVE

$$(1 - \phi_1 B - \Phi_4 B^4)$$

- In general, an $ARIMA(p,d,q)(P,D,Q)_S$ model is given by:

$$\left(1 - \phi_1 B - \cdots - \phi_p B^p\right)\left(1 - \Phi_s B^s - \cdots - -\Phi_{ps} B^{ps}\right) = \theta_0 + \left(1 - \theta_1 B - \cdots - \theta_q B^q\right)\left(1 - \Theta_s B^s - \cdots - \Theta_{qs} B^{qs}\right) a_t$$

- Also,

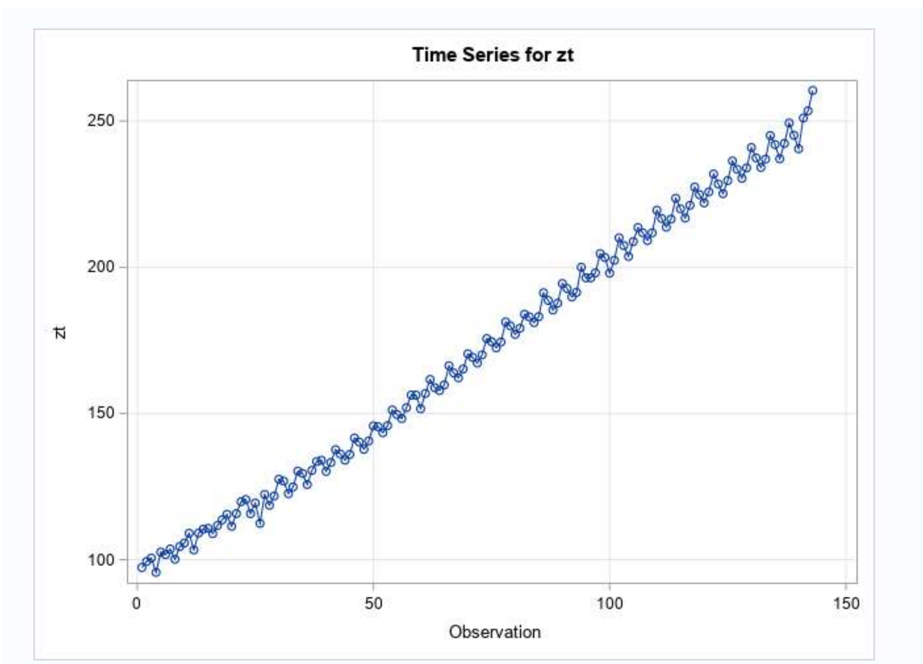
$$\theta_0 = (1 - \phi_1 - \cdots - \phi_p)(1 - \Phi_s - \cdots - \Phi_{ps})\mu$$

- ***Determining Seasonality***

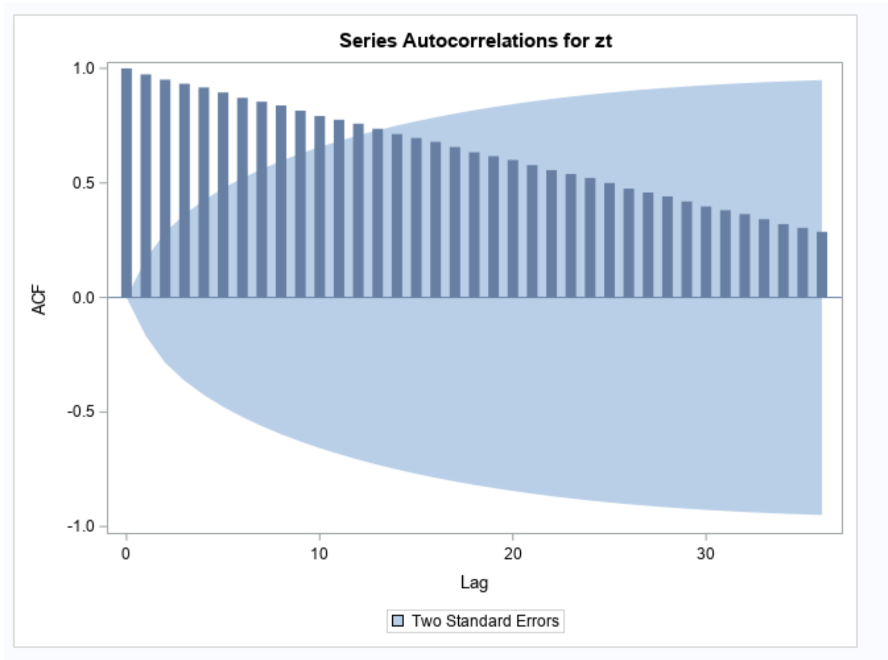
  As an example, we use the AR(1) model.

  There is quarterly seasonality if (1) the lags at s = 4 for the ACF plot are exponentially decreasing and (2) if only the first quarterly lag is significant.

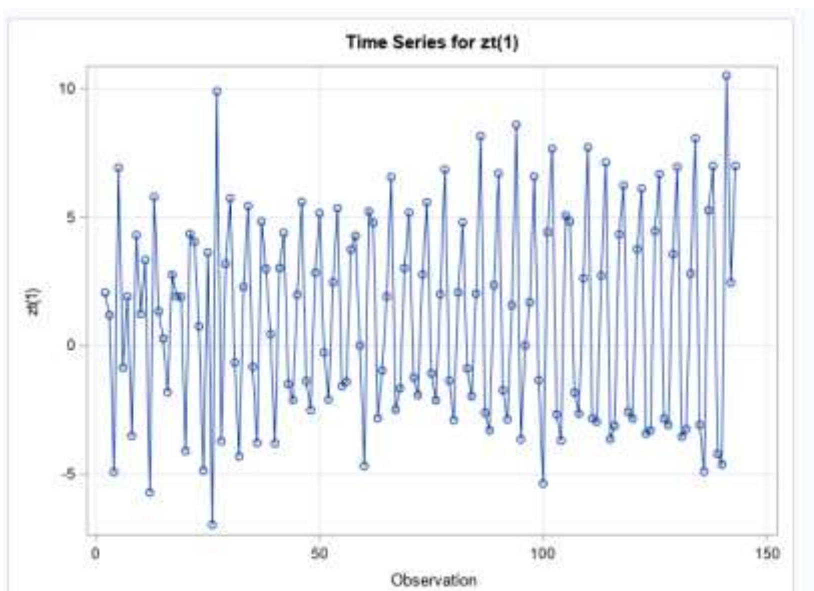
W8 Dataset



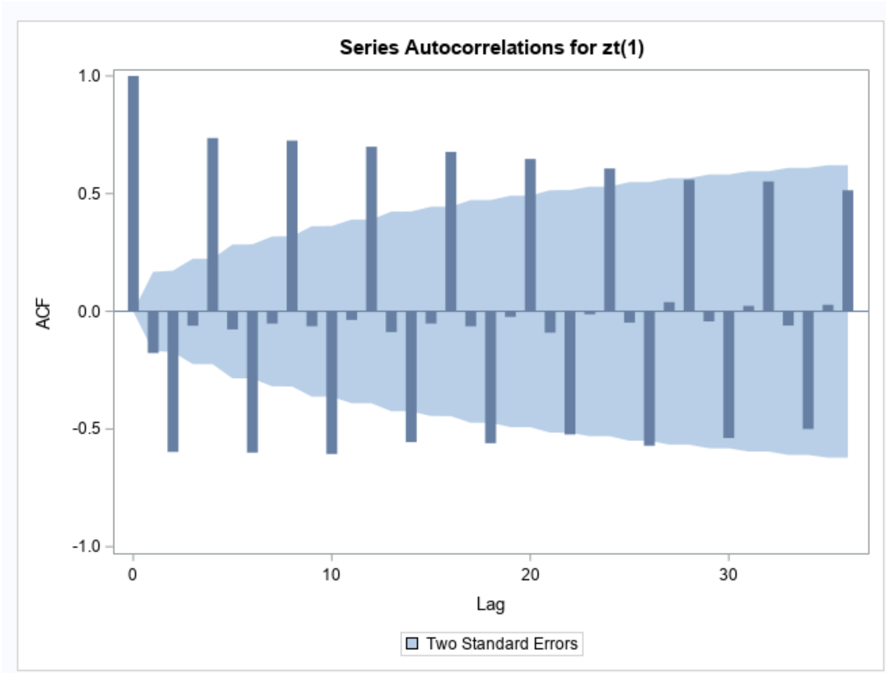
- From the time plot, the series is stationary in the variance but not in the mean.



- Since the ACF is slowly decreasing with increasing lags, we employ differencing.



- The data now looks stationary…



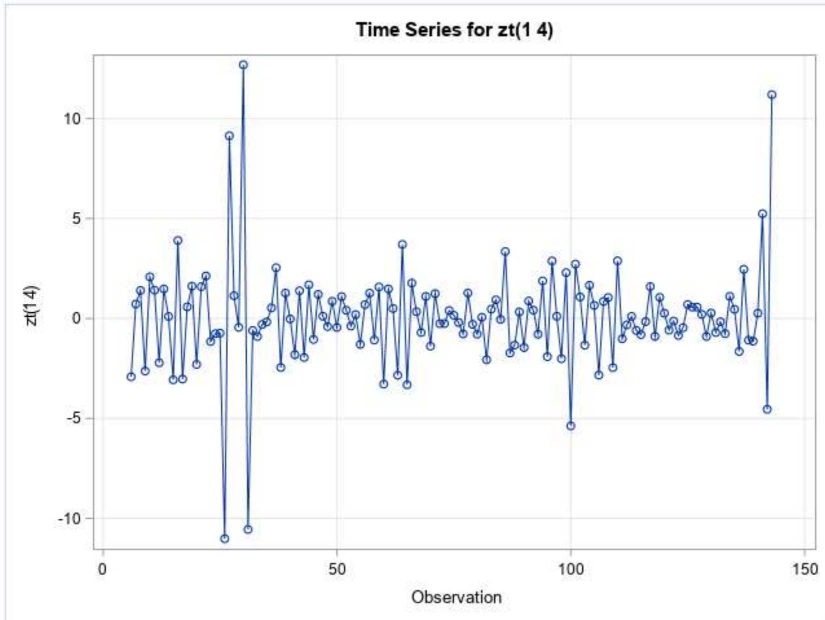
- However, the ACF is slowly decreasing for lags which are multiples of 4. Thus, we employ seasonal differencing. In total, we did differencing twice: one for ordinary differencing and another for seasonal differencing. Ergo,

$$(1-B)(1-B^4)Z_t \text{ or } Z_t(1\ 4)$$

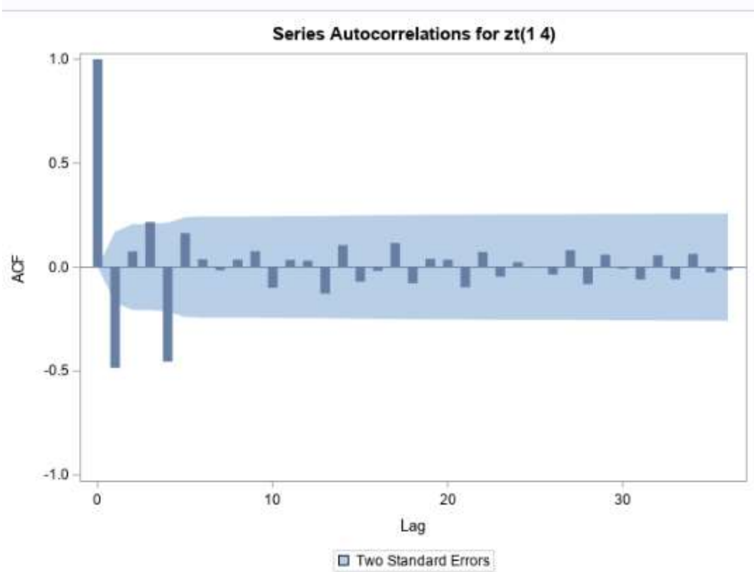
- Thus, we fit an ARIMA(-,1,-)(-,1,-) model.

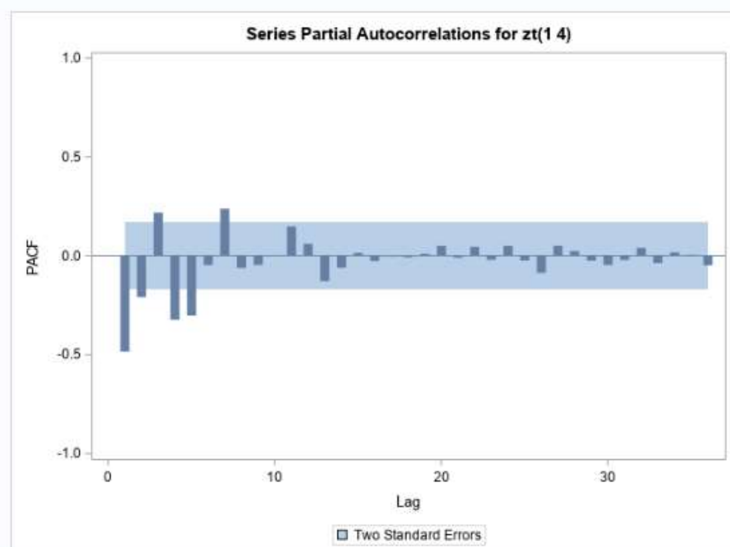
Autocorrelation Check for White Noise									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	74.63	6	<.0001	-0.485	0.075	0.217	-0.455	0.163	0.038
12	77.58	12	<.0001	-0.016	0.036	0.076	-0.100	0.035	0.031
18	85.78	18	<.0001	-0.127	0.105	-0.070	-0.019	0.116	-0.078
24	89.16	24	<.0001	0.040	0.036	-0.098	0.073	-0.046	0.023
30	92.35	30	<.0001	-0.003	-0.036	0.081	-0.082	0.059	-0.008
36	95.10	36	<.0001	-0.060	0.056	-0.059	0.063	-0.026	-0.013

- Since the p-values are very small, the series is not white noise.



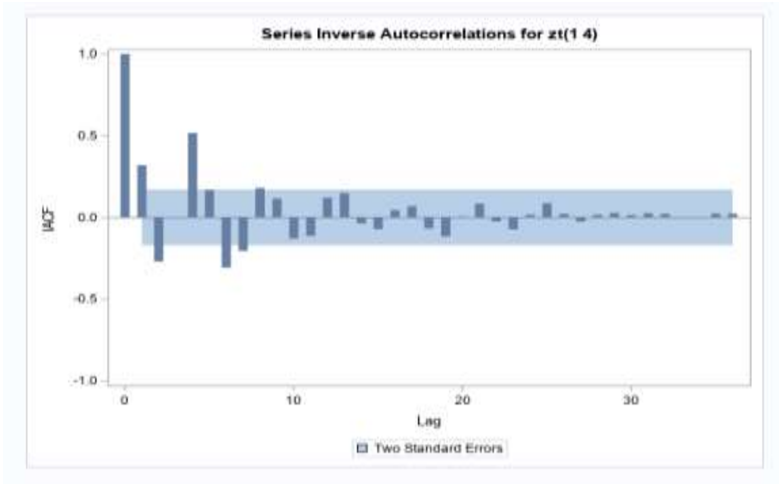
- The time plot is centered at zero and could have some outliers.





- In the ACF, lag 4 is significant.
- In the PACF, lags 4 and 7 are significant and exponentially decreasing. Lag 7 is significant since this lags "echoes" the succeeding lags (in this case, lag 8).
- Thus, we fit a seasonal MA(1) model.
- All lags below lag 4 are non-seasonal or ordinary lags. For the ACF, only lag 1 is significant while for the PACF, the non-seasonal lags are exponentially decreasing. Thus, we fit an MA(1) model.

- In total, we fit an ARIMA(0,1,1)(0,1,1)$_4$ model.
- Note that in the ACF/IACF plots, there is a plot for lag 0 while in the PACF plot, there is no such plot.



- This is even clearer in the IACF plot.

Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	0.0058592	0.01236	0.47	0.6363	0
MA1,1	0.73052	0.07435	9.83	<.0001	1
MA2,1	0.72365	0.07441	9.73	<.0001	4

Constant Estimate	0.005859
Variance Estimate	3.082198
Std Error Estimate	1.755619
AIC	549.9327
SBC	558.7145
Number of Residuals	138

* AIC and SBC do not include log determinant.

- MU $= \theta_0$
- MA1,1 $= \theta_1$
- MA2,1 $= \Theta_4$
- Since the p-value for the constant term > 0.05, we drop the constant term.
- Since this is a pure MA model, $\mu = \theta_0$.

Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MA1,1	0.73120	0.07387	9.90	<.0001	1
MA2,1	0.72409	0.07352	9.85	<.0001	4

Variance Estimate	3.064903
Std Error Estimate	1.750687
AIC	548.1746
SBC	554.0291
Number of Residuals	138

* AIC and SBC do not include log determinant.

Correlations of Parameter Estimates		
Parameter	MA1,1	MA2,1
MA1,1	1.000	-0.314
MA2,1	-0.314	1.000

Correlations of Parameter Estimates		
Parameter	MA1,1	MA2,1
MA1,1	1.000	-0.314
MA2,1	-0.314	1.000

- Correlation is relatively low (absolute value less than 0.9), indicating non-multicollinearity
- All parameters have p-value < 0.05, indicating that they are significant.

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	6.61	4	0.1580	-0.078	0.151	-0.020	-0.115	-0.029	0.054
12	9.06	10	0.5268	0.053	0.082	0.051	-0.046	0.020	0.041
18	11.60	16	0.7712	-0.088	0.031	-0.023	0.069	0.033	0.034
24	13.63	22	0.9142	0.021	0.060	-0.074	0.051	-0.011	0.006
30	15.15	28	0.9768	-0.068	-0.009	-0.012	-0.063	-0.008	-0.005
36	17.14	34	0.9929	-0.073	-0.004	-0.043	0.034	-0.048	-0.006

- p-value > 0.05 for all lags, indicating that the residuals are white noise
- Thus, the final model is

$$(1-B)(1-B^4)Z_t = (1-0.7312B)(1-0.72409B^4)a_t$$

- To check if this is the appropriate model, we employ overfitting and fit an ARIMA(0,1,3)(0,1,1)$_4$.

Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MA1,1	0.83804	0.09566	8.76	<.0001	1
MA1,2	-0.36123	0.12418	-2.91	0.0042	2
MA1,3	0.27801	0.11757	2.36	0.0195	3
MA2,1	0.66724	0.10032	6.65	<.0001	4

Variance Estimate	2.945113
Std Error Estimate	1.716133
AIC	544.6282
SBC	556.3372
Number of Residuals	138

- The additional parameters MA1,2 and MA1,3 are significant but AIC = 544, a difference of only 5 points from previous AIC = 549. By the principle of parsimony, we choose the simpler ARIMA(0,1,1)(0,1,1)$_4$ model.

<u>03-12-2019</u>

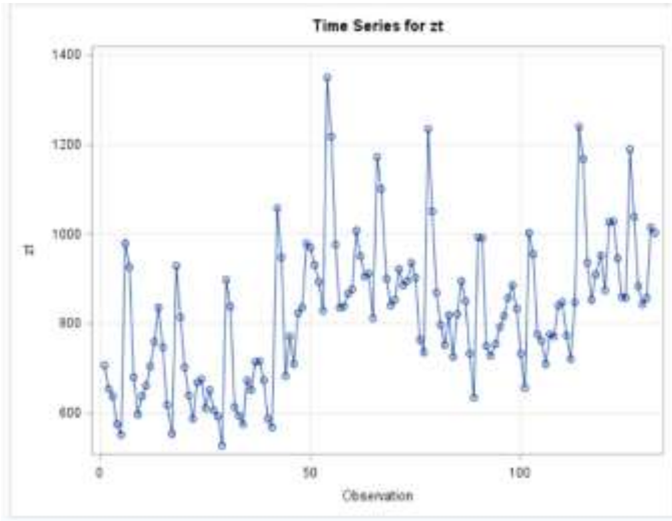
W9 Dataset

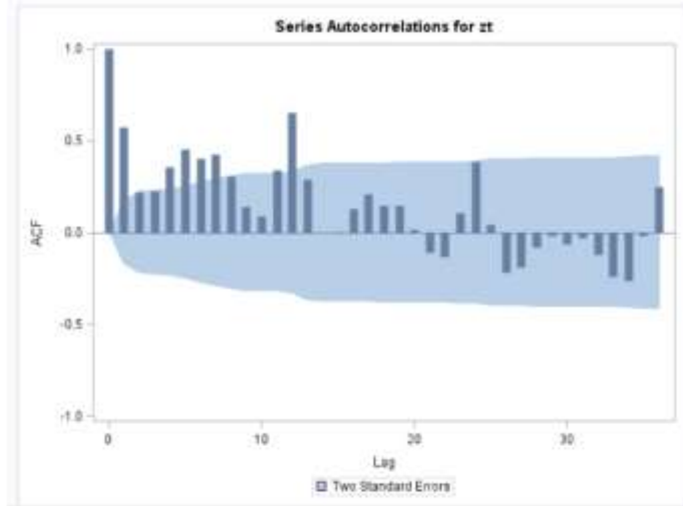
Monthly Employment (in 1000's)

n = 132 ⇒ nlag = 33 or 36 to observe 3 seasonal lags

Autocorrelation Check for White Noise									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	126.31	6	<.0001	0.571	0.218	0.224	0.358	0.454	0.400
12	248.32	12	<.0001	0.423	0.303	0.137	0.086	0.341	0.653
18	273.54	18	<.0001	0.291	-0.004	0.001	0.131	0.211	0.144
24	307.72	24	<.0001	0.144	0.016	-0.109	-0.132	0.105	0.386
30	323.63	30	<.0001	0.043	-0.216	-0.188	-0.081	-0.019	-0.063
36	360.88	36	<.0001	-0.032	-0.124	-0.240	-0.267	-0.021	0.246

- All the p-values < 0.05, so we reject H_0 and conclude that the series is not white noise and subsequently fit an ARIMA model.

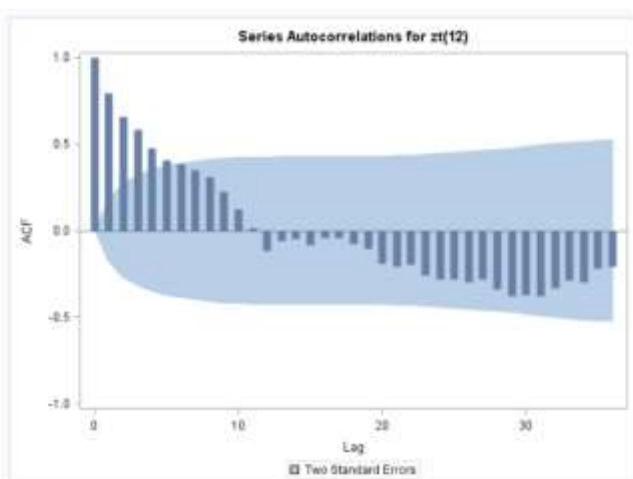


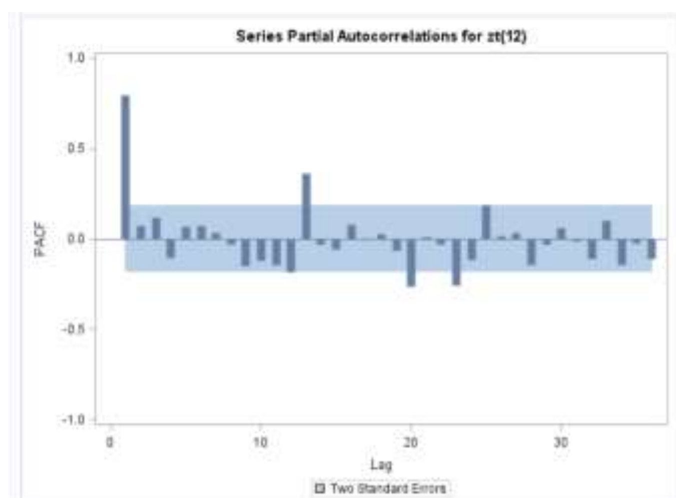


- From the time plot, the series is non-stationary in mean but not in variance, so we employ differencing. Notice that we need to do differencing because the ACF plots for the seasonal lags (lags 12, 24 and 36) are very slowly decreasing. Thus, we employ seasonal differencing: $Z_t(12)$ or $Z_t = (1 - B^{12})Z_t$
- After seasonal differencing,

Autocorrelation Check for White Noise									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	242.88	6	<.0001	0.794	0.657	0.586	0.473	0.410	0.386
12	282.04	12	<.0001	0.353	0.310	0.224	0.125	0.016	-0.114
18	285.16	18	<.0001	-0.058	-0.045	-0.084	-0.044	-0.044	-0.077
24	326.17	24	<.0001	-0.104	-0.191	-0.210	-0.195	-0.257	-0.280
30	429.37	30	<.0001	-0.283	-0.297	-0.280	-0.341	-0.381	-0.375
36	516.14	36	<.0001	-0.377	-0.331	-0.288	-0.299	-0.222	-0.207

- We reject H_0 since all p-values < 0.05, so the series is not white noise.





- The ACF plots of the nonseasonal lags are exponentially decreasing but their PACF plots are signficant only at lag 1.
- The PACF plots of the seasonal logs are exponentially decreasing.
- Thus, we can try to fit a nonseasonal AR(1) and a seasonal MA(1) or an ARIMA(1,0,0)(0,1,1)$_{12}$; i.e.,

$$(1-\phi_1 B)(1-B^{12})Z_t = \theta_0 + (1-\Theta_{12}B^{12})a_t$$

Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	28.22783	9.27539	3.04	0.0029	0
MA1,1	0.79816	0.06416	12.44	<.0001	12
AR1,1	0.84602	0.05012	16.88	<.0001	1

Constant Estimate	4.346632
Variance Estimate	3170.095
Std Error Estimate	56.3036
AIC	1310.889
SBC	1319.252
Number of Residuals	120

- All the parameters are signficant.

Correlations of Parameter Estimates			
Parameter	MU	MA1,1	AR1,1
MU	1.000	-0.006	0.121
MA1,1	-0.006	1.000	0.008
AR1,1	0.121	0.008	1.000

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	6.09	4	0.1927	-0.139	-0.043	0.137	-0.063	0.019	0.070
12	11.58	10	0.3144	0.147	-0.014	0.101	0.055	-0.081	0.018
18	14.19	16	0.5848	-0.002	0.074	-0.093	0.044	0.050	-0.002
24	24.59	22	0.3172	0.145	-0.180	0.005	0.061	-0.101	-0.060
30	28.33	28	0.4473	-0.035	-0.010	0.032	-0.042	-0.115	-0.077
36	39.16	34	0.2493	-0.086	-0.061	0.069	-0.205	0.073	-0.025

- There is no redundancy and the residuals are white noise.
- To check, we can employ overfitting and fit an ARIMA(2,0,0)(0,1,1)$_{12}$.

Conditional Least Squares Estimation					
Parameter	**Estimate**	**Standard Error**	**t Value**	**Approx Pr > |t|**	**Lag**
MU	30.24976	10.66325	2.84	0.0054	0
MA1,1	0.80833	0.06416	12.60	<.0001	12
AR1,1	0.70576	0.09197	7.67	<.0001	1
AR1,2	0.16821	0.09215	1.83	0.0705	2

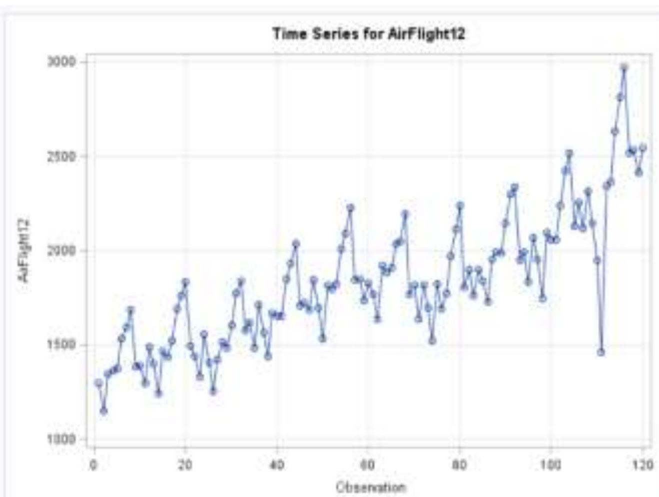
Constant Estimate	3.812425
Variance Estimate	3109.432
Std Error Estimate	55.76228
AIC	1309.541
SBC	1320.69
Number of Residuals	120

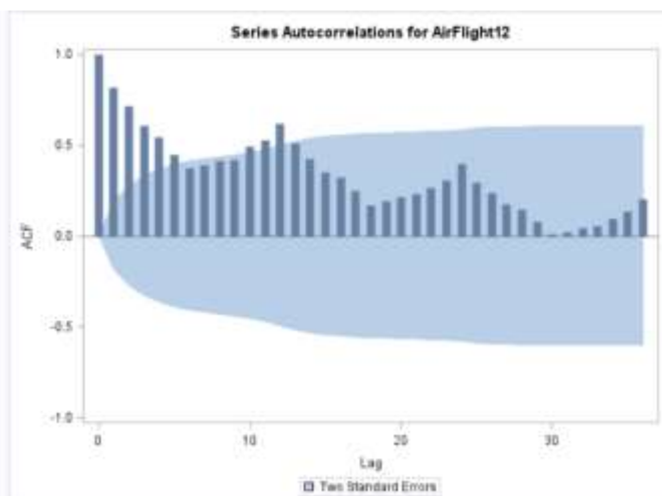
- AR1,2 is not significant and the difference in AIC and SBC are negligible compared to the first model. Thus, we prefer the first model.

Case 9 (Pankratz)

Monthly freight volume

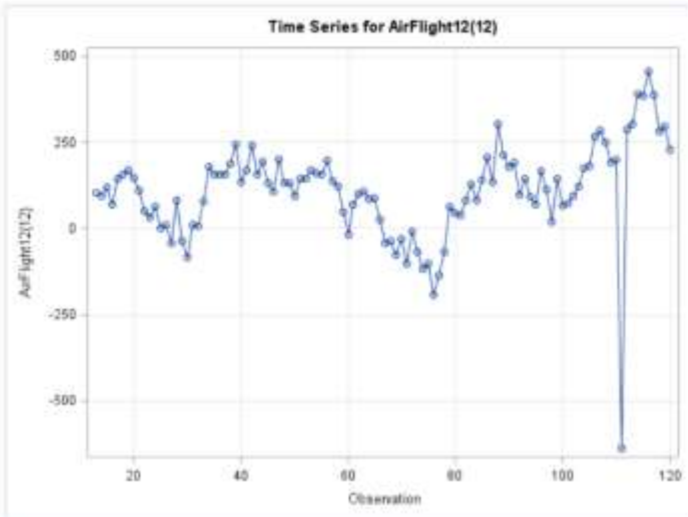
n = 120 ⇒ nlag = 30 or 36 to observe 3 seasonal lags

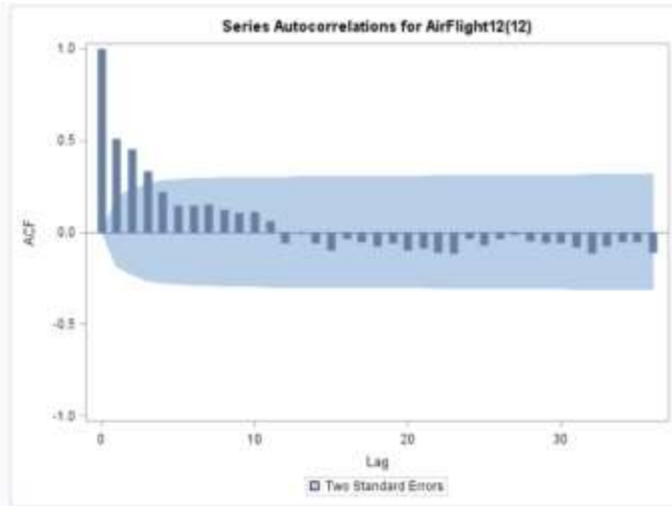




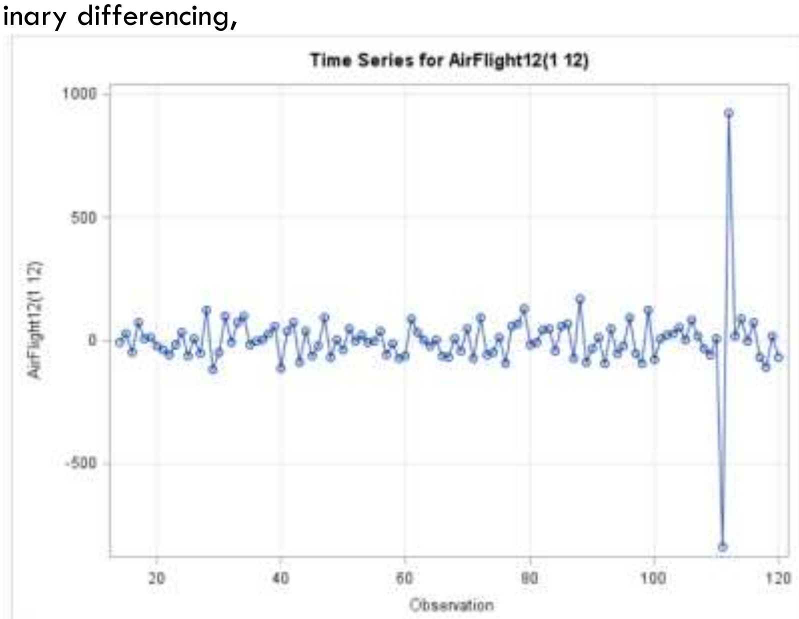
- The series is stationary in the variance but non-stationary in the mean. Thus, we employ differencing. From the ACF plot, the seasonal and nonseasonal lags are slowly decreasing. Thus we employ both, doing seasonal differencing first (arbitrary).

* After seasonal differencing,

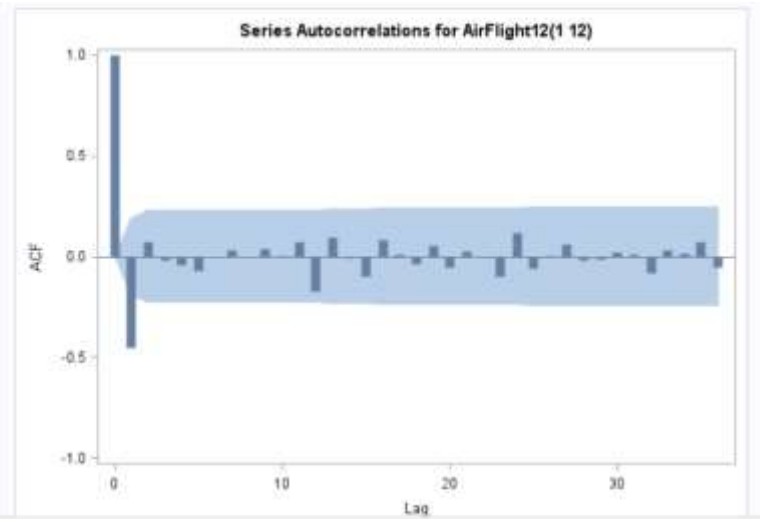


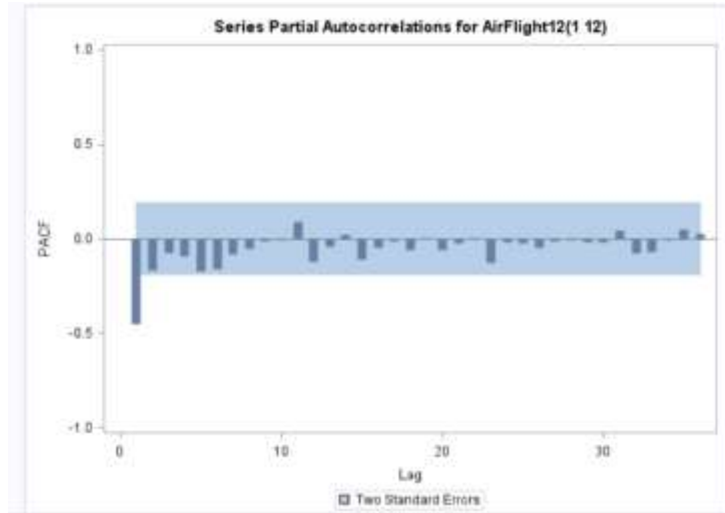


* The series is still nonstationary in the mean and the ACF plots for the nonseasonal lags are slowly decreasing. Thus, we employ ordinary differencing.
* After ordinary differencing,



* The series is now stationary in the mean.





- The ACF plot is significant only at lag 1 while the PACF plots are exponentially decreasing. Thus, we fit an MA(1) model; i.e. an ARIMA(0,1,1)(0,1,0)$_{12}$ model given by

$$(1-B)(1-B^{12})Z_t = \theta_0 + (1-\theta_1 B)a_t$$

Conditional Least Squares Estimation

Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	1.73424	4.31631	0.40	0.6887	0
MA1,1	0.62589	0.07627	8.21	<.0001	1

Constant Estimate	1.734243
Variance Estimate	13871.71
Std Error Estimate	117.7782
AIC	1326.158
SBC	1331.504
Number of Residuals	107

- θ_0 is not significant. Thus,

Conditional Least Squares Estimation

Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MA1,1	0.62113	0.07629	8.14	<.0001	1

Variance Estimate	13761.68
Std Error Estimate	117.3102
AIC	1324.32
SBC	1326.993
Number of Residuals	107

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	6.05	5	0.3015	0.040	0.061	-0.055	-0.136	-0.152	-0.060
12	11.10	11	0.4346	0.029	0.049	0.089	0.018	-0.001	-0.173
18	13.16	17	0.7253	-0.010	-0.056	-0.100	0.046	0.029	-0.003
24	15.61	23	0.8714	0.035	-0.040	-0.015	-0.048	-0.091	0.064
30	16.48	29	0.9697	-0.015	0.028	0.069	0.006	-0.012	0.006
36	19.83	35	0.9816	-0.025	-0.079	0.027	0.065	0.096	-0.004

- The residuals are white noise.
- Thus, the final model is given by:

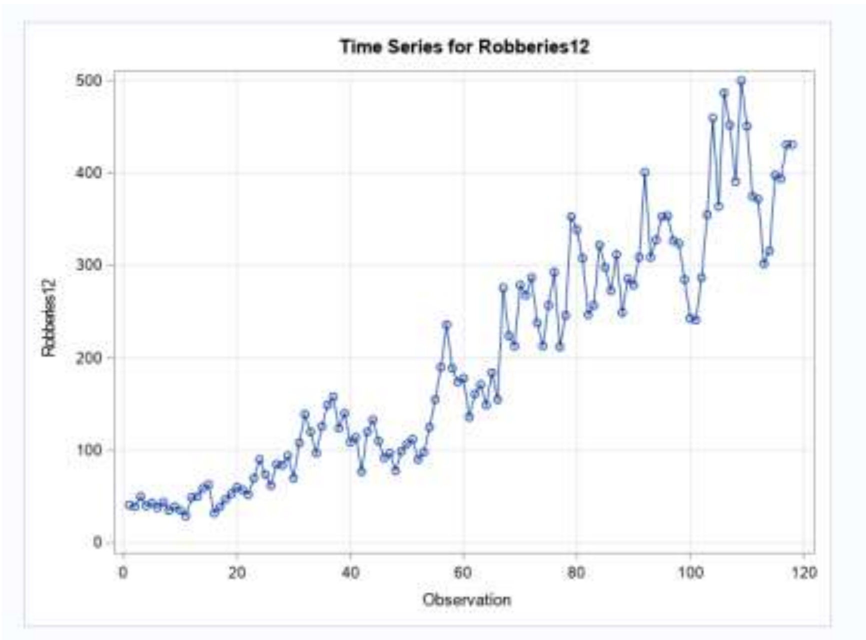
$$(1-B)(1-B^{12})Z_t = (1-0.62113B)a_t$$

03-14-2019

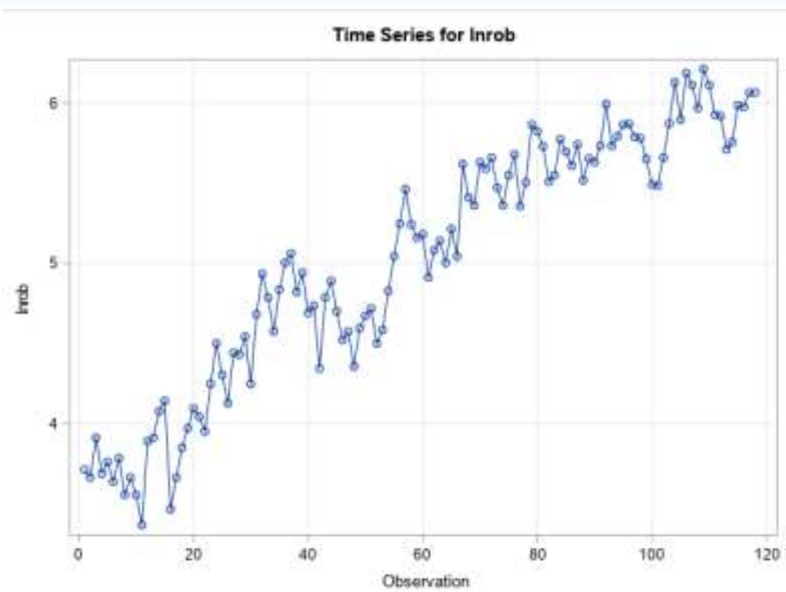
Case 11 (Pankratz)

Monthly crime robberies (Robberies12)

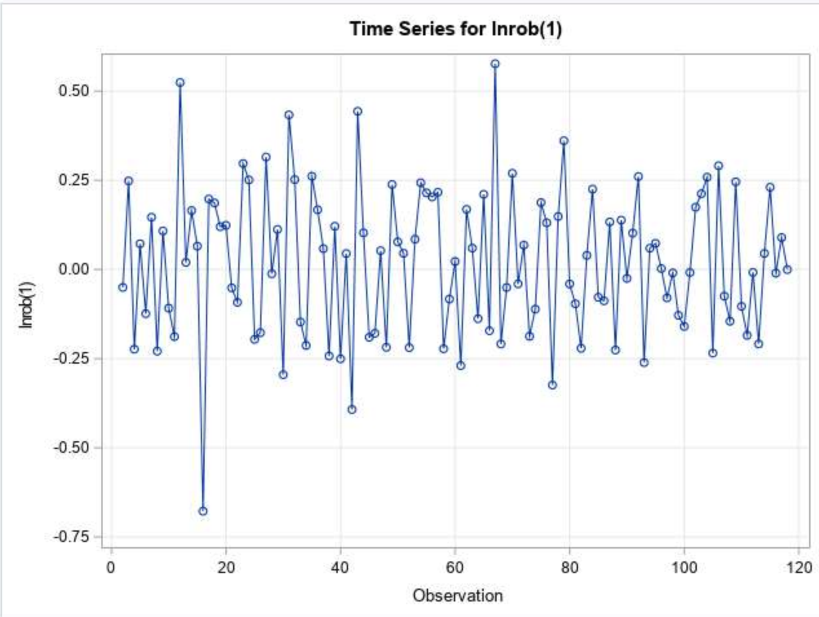
n = 118 ⇒ nlag = 29.5 or 36 to observe 3 seasonal lags



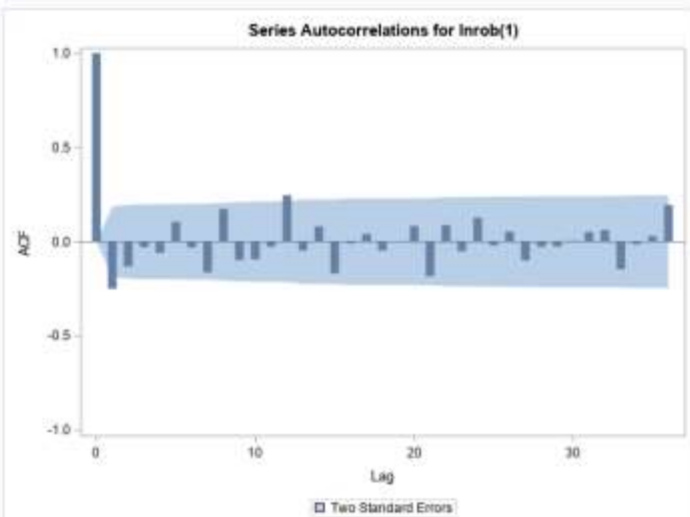
- Because the time plot exhibits non-constant mean and variance, we do both log transformation and differencing.

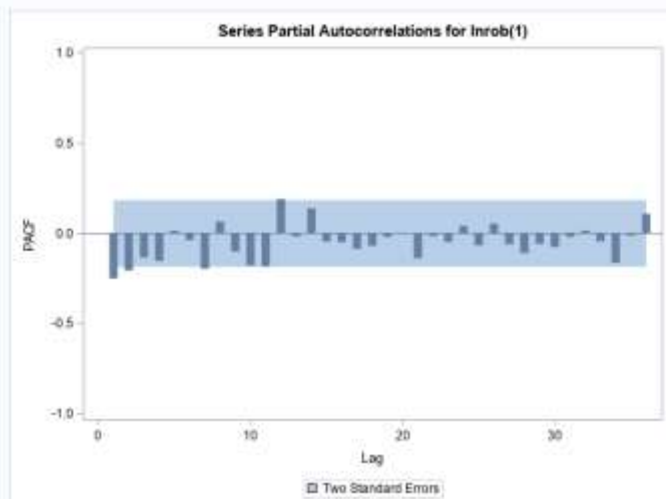


- Because the time plot still exhibits non-constant mean, we do differencing.



- After differencing, the series now looks stationary.
- Thus, we now have an ARIMA(-,1,-)(-,0,-)$_{12}$ model.





- The lags overall (ACF is significant at lag 1, PACF is exponentially decreasing) suggest an MA(1) model
- The seasonal lags (lags 12, 24 and 36) (ACF and PACF look alike) also suggest a seasonal ARMA(1,1) model.
- Thus, we try an ARIMA(0,1,1)(1,0,1)$_{12}$ model, i.e.

$$(1-\Phi_{12}B^{12})(1-B)Z_t = \theta_0 + (1-\theta_1 B)(1-\Theta_{12}B^{12})a_t$$

Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	0.02028	0.01606	1.26	0.2091	0
MA1,1	0.45086	0.08419	5.36	<.0001	1
MA2,1	0.74296	0.19589	3.79	0.0002	12
AR1,1	0.92801	0.13884	6.68	<.0001	12

Constant Estimate	0.00146
Variance Estimate	0.034214
Std Error Estimate	0.184971
AIC	-58.9259
SBC	-47.8772
Number of Residuals	117

- Only the constant term is non-significant.

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	3.95	3	0.2674	0.062	-0.118	-0.010	-0.080	0.070	-0.055
12	11.99	9	0.2140	-0.205	0.106	-0.015	-0.096	-0.012	0.021
18	16.62	15	0.3420	0.036	0.090	-0.124	-0.088	-0.011	-0.036
24	20.45	21	0.4930	-0.026	0.020	-0.069	0.106	-0.023	-0.092
30	25.32	27	0.5563	0.025	0.072	-0.069	-0.116	-0.077	0.035
36	29.17	33	0.6582	0.066	0.042	-0.048	0.023	0.097	0.068

- The residuals of the model are white noise.

Correlations of Parameter Estimates				
Parameter	MU	MA1,1	MA2,1	AR1,1
MU	1.000	-0.007	-0.004	-0.014
MA1,1	-0.007	1.000	0.066	0.071
MA2,1	-0.004	0.066	1.000	0.938
AR1,1	-0.014	0.071	0.938	1.000

- There is suspected redundancy since the correlation between MA2,1 and AR1,1 are highly correlated.

Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MA1,1	0.42787	0.08485	5.04	<.0001	1
MA2,1	0.72493	0.18776	3.86	0.0002	12
AR1,1	0.92328	0.13039	7.08	<.0001	12

Variance Estimate	0.034364
Std Error Estimate	0.185375
AIC	-59.3848
SBC	-51.0983
Number of Residuals	117

- The AIC and SBC improved.

- Alternatively, we can fit an ARIMA(2,1,0)(1,0,1)$_{12}$ model with no constant term, i.e.

$$(1-\phi_1 B-\phi_2 B^2)(1-\Phi_{12}B^{12})(1-B)Z_t=\theta_0+(1-\Theta_{12}B^{12})a_t$$

Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MA1,1	0.68768	0.18508	3.72	0.0003	12
AR1,1	-0.34144	0.09188	-3.72	0.0003	1
AR1,2	-0.24174	0.09175	-2.63	0.0096	2
AR2,1	0.90156	0.12896	6.99	<.0001	12

Variance Estimate	0.034331
Std Error Estimate	0.185287
AIC	-58.5269
SBC	-47.4782
Number of Residuals	117

- The AIC and SBC decreased (but not significantly). Also, the previous model is more parsimonious. Thus, we prefer the ARIMA(0,1,1)(1,0,1)$_{12}$ model.

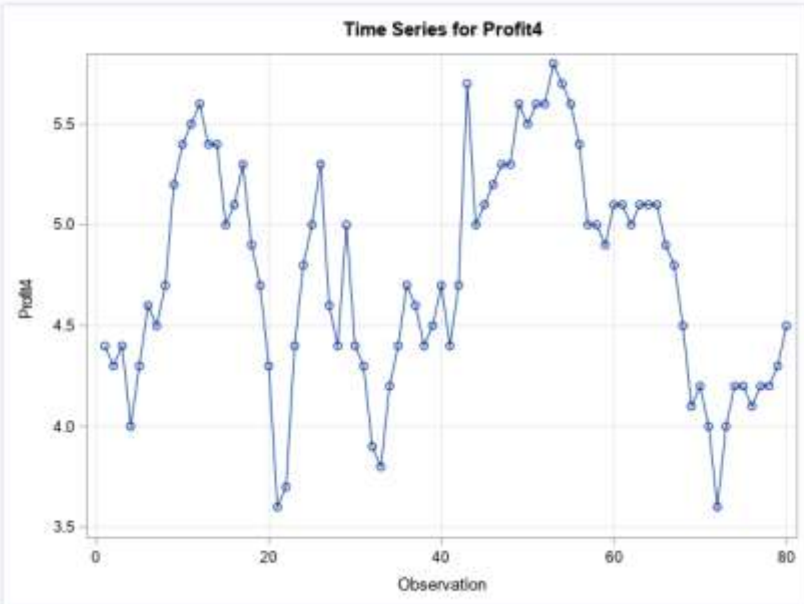
Correlations of Parameter Estimates				
Parameter	MA1,1	AR1,1	AR1,2	AR2,1
MA1,1	1.000	-0.074	-0.037	0.921
AR1,1	-0.074	1.000	0.277	-0.101
AR1,2	-0.037	0.277	1.000	-0.041
AR2,1	0.921	-0.101	-0.041	1.000

- However, the high correlation persists; this time between MA1,1 and AR2,1.

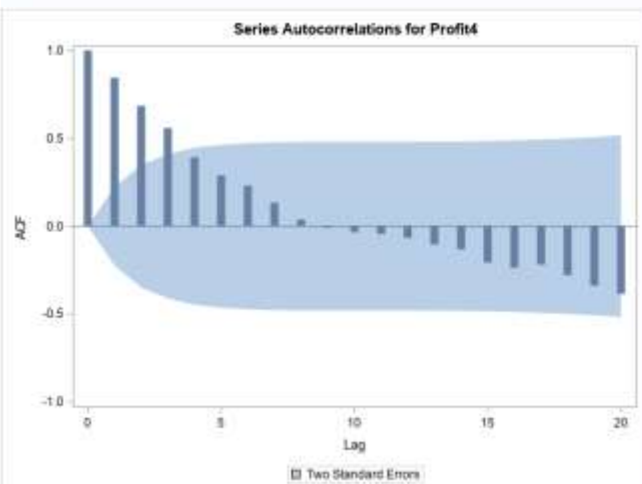
Case 10 (Pankratz)

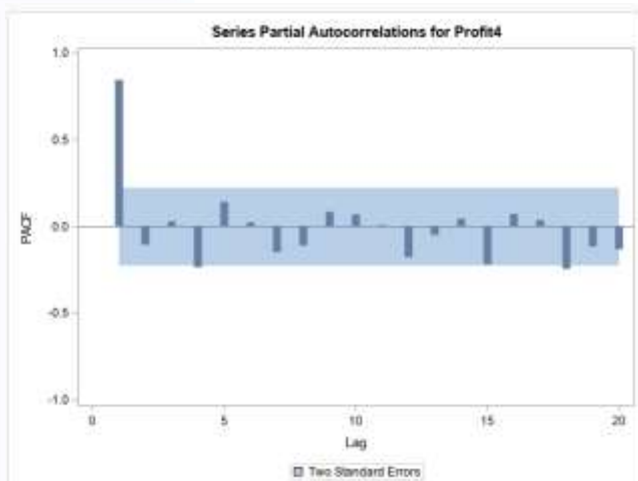
Quarterly profit margins (Profit4)

n = 80 ⇒ nlag = 20



- Because the time plot exhibits constant mean and variance, we do not do any transformation/differencing.





- The lags overall (ACF is exponentially decreasing, PACF is significant at lag 1) suggest an AR(1) model
- Thus, we fit an ARIMA(1,0,0) model, i.e.,

$$(1 - \phi_1 B)Z_t = \theta_0 + a_t$$

Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	4.62825	0.18536	24.97	<.0001	0
AR1,1	0.85587	0.05960	14.36	<.0001	1

Constant Estimate	0.667089
Variance Estimate	0.088076
Std Error Estimate	0.296777
AIC	34.64071
SBC	39.40477
Number of Residuals	80

- Both the constant term and AR1,1 are significant.
- The equation is:

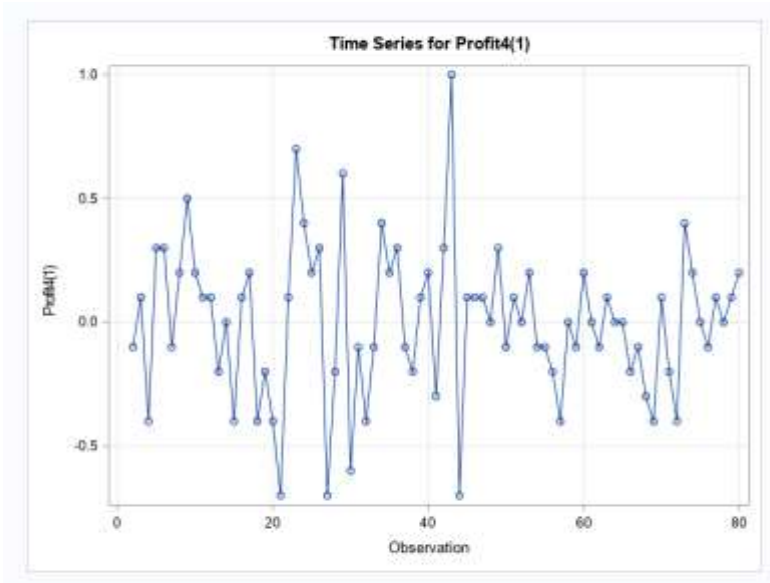
$$\widehat{Z}_t = 0.667 + 0.855Z_{t-1}$$

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	7.89	5	0.1623	0.077	-0.053	0.167	-0.158	-0.100	0.141
12	10.66	11	0.4720	0.010	-0.146	-0.035	-0.043	0.046	0.058
18	22.15	17	0.1790	-0.027	0.138	-0.141	-0.155	0.220	-0.025
24	38.32	23	0.0235	-0.064	0.107	-0.230	-0.274	-0.033	0.007

- Not all p-values < 0.05. However, for lag 24, the p-value is close enough to 0.05 that we can conclude that the residuals are probably white noise.

Augmented Dickey-Fuller Unit Root Tests							
Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Zero Mean	0	-0.1408	0.6485	-0.25	0.5943		
	1	-0.1237	0.6523	-0.21	0.6064		
	2	-0.1151	0.6542	-0.22	0.6033		
Single Mean	0	-12.0516	0.0716	-2.53	0.1116	3.21	0.2633
	1	-14.5681	0.0362	-2.67	0.0833	3.58	0.1710
	2	-13.3562	0.0502	-2.42	0.1391	2.93	0.3327
Trend	0	-12.1838	0.2766	-2.55	0.3061	3.32	0.5203
	1	-14.6958	0.1674	-2.69	0.2431	3.73	0.4399
	2	-13.4375	0.2159	-2.44	0.3559	3.09	0.5646

- p-values > 0.05. Thus, the series must have a unit root, and the series is not stationary. Thus, we conduct differencing.



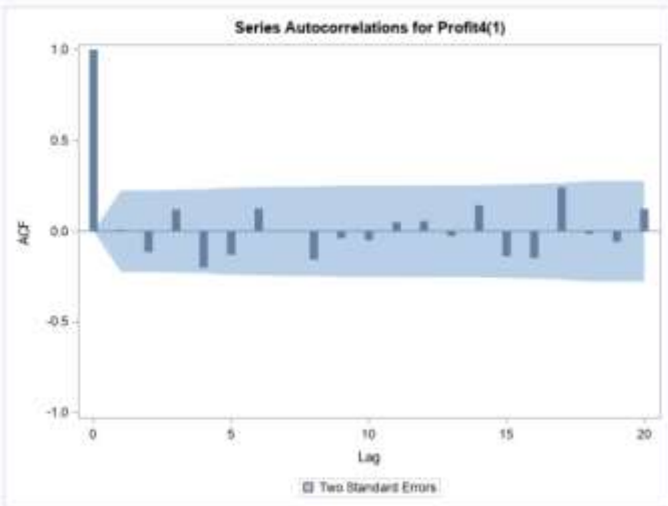
- Now, the series look more stationary.

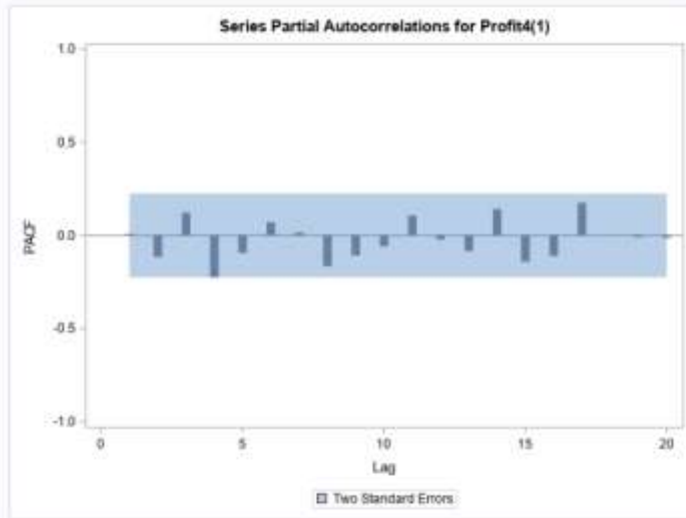
Autocorrelation Check for White Noise									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	8.60	6	0.1972	0.008	-0.116	0.118	-0.201	-0.131	0.125
12	11.69	12	0.4710	-0.001	-0.156	-0.039	-0.049	0.049	0.056
18	23.94	18	0.1571	-0.027	0.143	-0.139	-0.148	0.241	-0.015

- p-values < 0.05. Thus, the differenced series must be white noise.

Augmented Dickey-Fuller Unit Root Tests							
Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Zero Mean	0	-77.3598	<.0001	-8.69	<.0001		
	1	-96.4143	<.0001	-6.83	<.0001		
	2	-65.5498	<.0001	-4.72	<.0001		
Single Mean	0	-77.3570	0.0008	-8.63	<.0001	37.23	0.0010
	1	-96.4089	0.0007	-6.78	<.0001	22.99	0.0010
	2	-65.4939	0.0007	-4.68	0.0003	10.99	0.0010
Trend	0	-77.6028	0.0002	-8.60	<.0001	36.94	0.0010
	1	-97.4678	0.0002	-6.76	<.0001	22.85	0.0010
	2	-67.7683	0.0002	-4.71	0.0015	11.10	0.0010

- p-values < 0.05. Thus, the differenced series must be stationary.





- All lags are within the blue band. Thus, this confirms that the series is white noise.

Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	0.0012658	0.03454	0.04	0.9709	0

Constant Estimate	0.001266
Variance Estimate	0.094229
Std Error Estimate	0.306968
AIC	38.58587
SBC	40.95532
Number of Residuals	79

- The constant term is not significant. Thus, this model is a random walk with a stochastic trend, i.e.,

$$(1-B)Z_t = a_t$$

- Further note that AIC and SBC are slightly higher than before. Thus, in terms of AIC and SBC and in terms of intuitive appeal, the AR(1) model is preferred.
- The equation is:

$$\widehat{Z}_t = Z_{t-1}$$

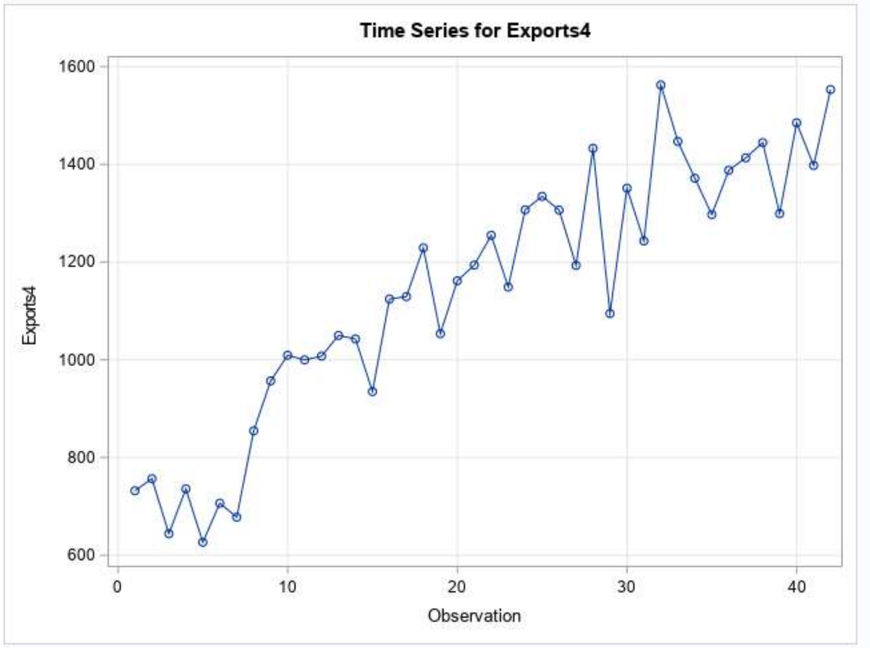
Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	8.60	6	0.1972	0.008	-0.116	0.118	-0.201	-0.131	0.125
12	11.69	12	0.4710	-0.001	-0.156	-0.039	-0.049	0.049	0.056
18	23.94	18	0.1571	-0.027	0.143	-0.139	-0.148	0.241	-0.015
24	37.70	24	0.0372	-0.059	0.122	-0.214	-0.241	0.008	0.042

- Further note that the residuals are white noise.

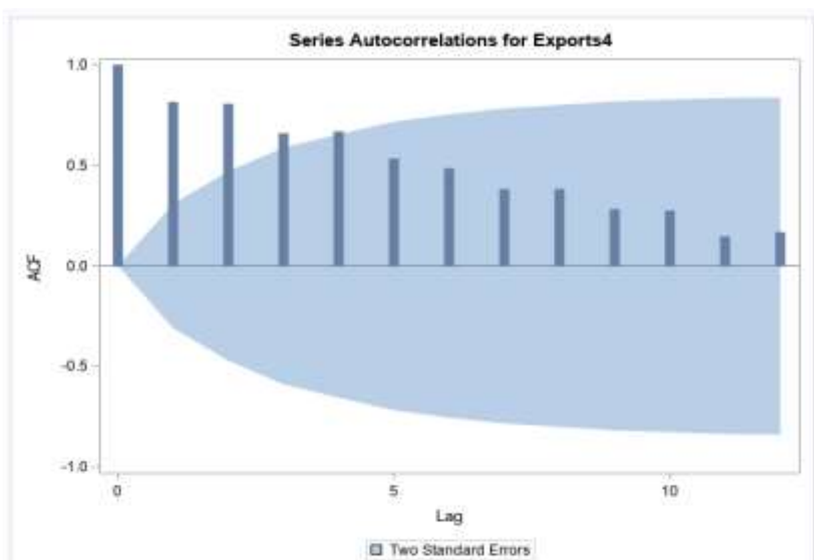
Case 15 (Pankratz)

Quarterly exports (Exports4)

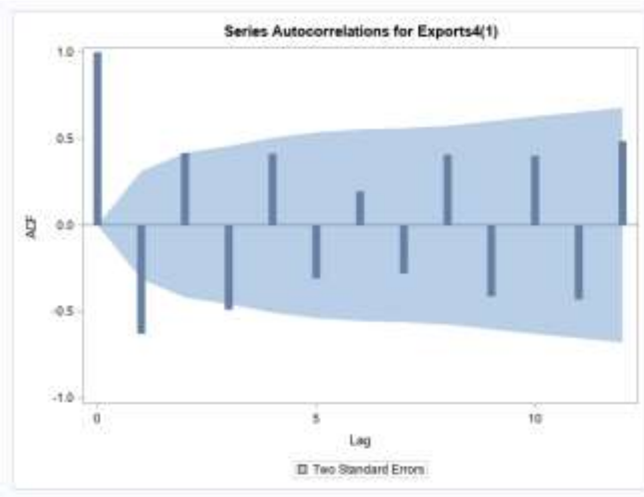
n = 42 ⇒ nlag = 12

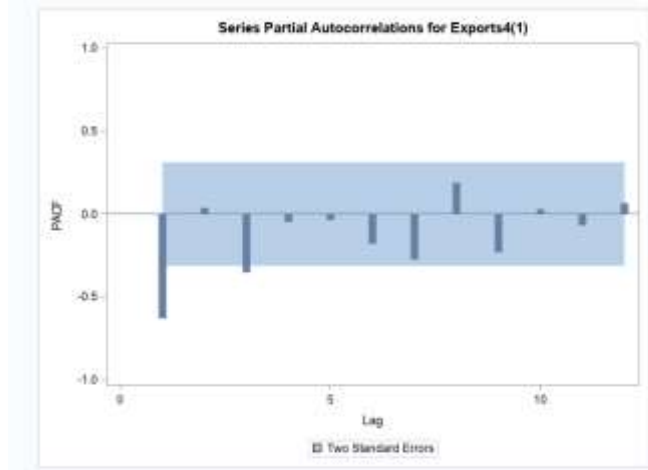


- Because the time plot exhibits non-constant mean, we do differencing.



- Because all ACF lags decrease slowly, we do ordinary differencing.





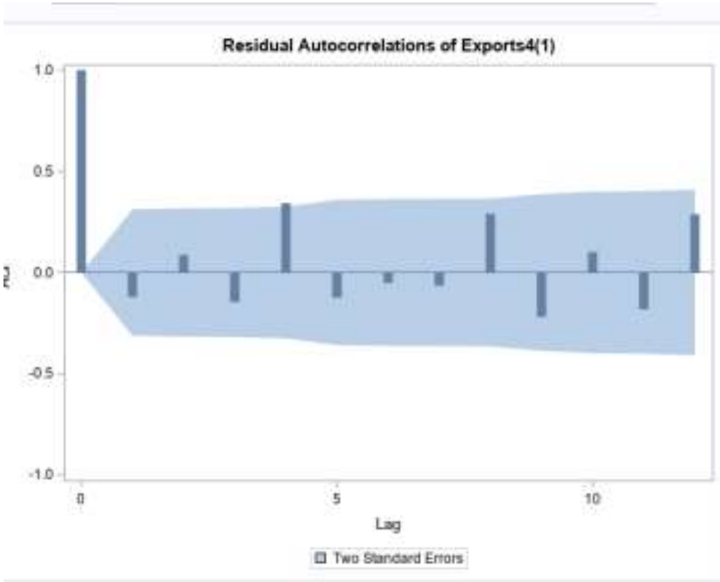
- For the non-seasonal lags, lags 1-3 are significant in the ACF but are exponentially decreasing in the PACF. Thus, we first fit an ARIMA(0, 1, 3)/MA(3) model.
- PACF seasonal lags are exponentially decreasing. ACF seasonal lag is significant only at lag = 4.

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	8.46	3	0.0374	-0.122	0.086	-0.147	0.342	-0.125	-0.053
12	23.41	9	0.0053	-0.067	0.290	-0.221	0.100	-0.183	0.287
18	37.31	15	0.0011	-0.344	0.031	-0.022	0.120	-0.238	-0.125
24	44.10	21	0.0023	-0.064	0.055	-0.189	0.015	-0.051	0.163

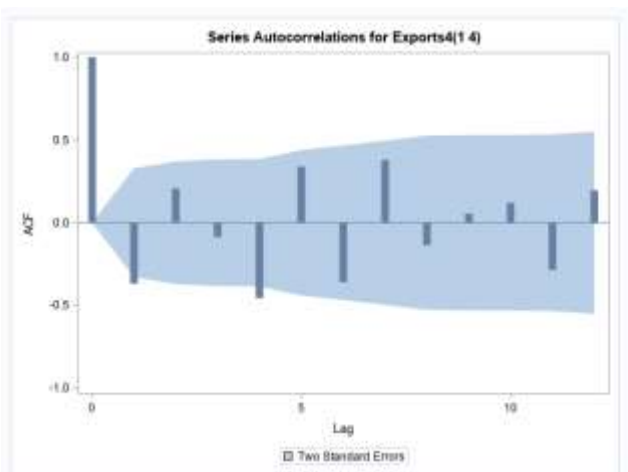
Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	19.99014	1.15700	17.28	<.0001	0
MA1,1	0.68397	0.16193	4.22	0.0002	1
MA1,2	-0.05032	0.19050	-0.26	0.7932	2
MA1,3	0.36635	0.16254	2.25	0.0302	3

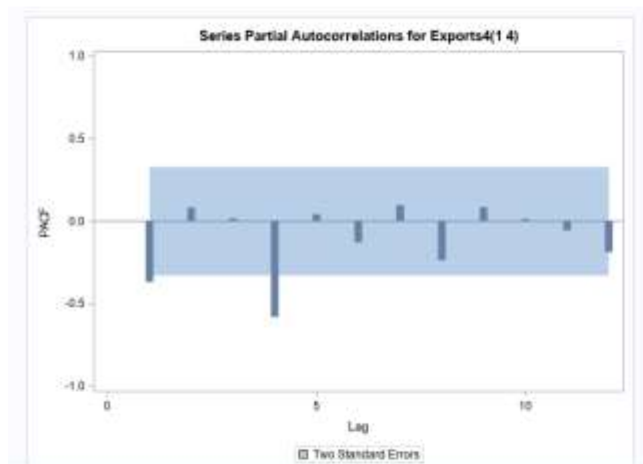
Constant Estimate	19.99014
Variance Estimate	9748.051
Std Error Estimate	98.73222
AIC	496.7219
SBC	503.5762
Number of Residuals	41

- MA1,2 is insignificant and p-values < 0.05, suggesting that the error is not white noise.



• The seasonal lags (lags = 4, 8, 12) in the residual ACF are slowly decreasing. Thus, we employ seasonal differencing.





- At the PACF, the seasonal lag at lag = 4 is significant while at the ACF, the seasonal lags are exponentially decreasing. At the PACF, the non-seasonal lag at lag = 1 is significant while at the ACF, the non-seasonal lags are exponentially decreasing. Thus, we fit an ARIMA(1,1,0)(0,1,1)$_{12}$,i.e.,

$$(1-\phi_1 B)(1-B^4)(1-B)Z_t = \theta_0 + (1-\Theta_4 B^4)a_t$$

Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	6.36142	3.36097	1.89	0.0669	0
MA1,1	0.98379	0.10128	9.71	<.0001	4
AR1,1	-0.29357	0.16937	-1.73	0.0921	1

Constant Estimate	8.228937
Variance Estimate	9851.402
Std Error Estimate	99.25423
AIC	448.1015
SBC	452.9342
Number of Residuals	37

- p-value > 0.05 for AR1,1 so we can drop it in the model

Correlations of Parameter Estimates			
Parameter	MU	MA1,1	AR1,1
MU	1.000	0.725	0.171
MA1,1	0.725	1.000	0.223
AR1,1	0.171	0.223	1.000

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	2.75	4	0.6004	0.035	0.107	-0.040	0.061	0.196	-0.075
12	6.97	10	0.7286	0.233	0.045	0.033	0.125	-0.032	0.099
18	14.46	16	0.5648	-0.157	0.108	0.167	-0.037	0.006	-0.210
24	17.17	22	0.7537	0.079	-0.143	0.020	0.002	0.016	0.061

- No redundancy and p-value < 0.05, indicating that the residuals are white noise.

03-19-2019

Case 6 (Pankratz)

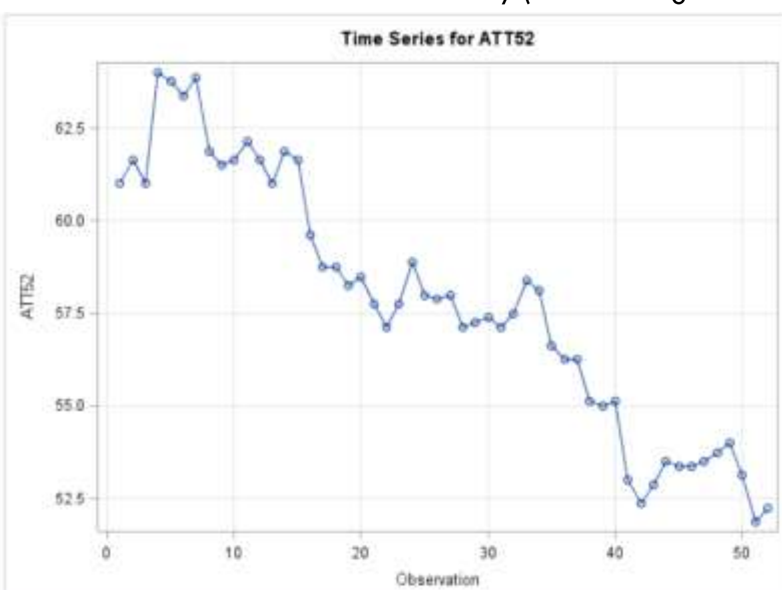
AT&T Stock Prices (att52)

n = 52 ⇒ nlag = 13

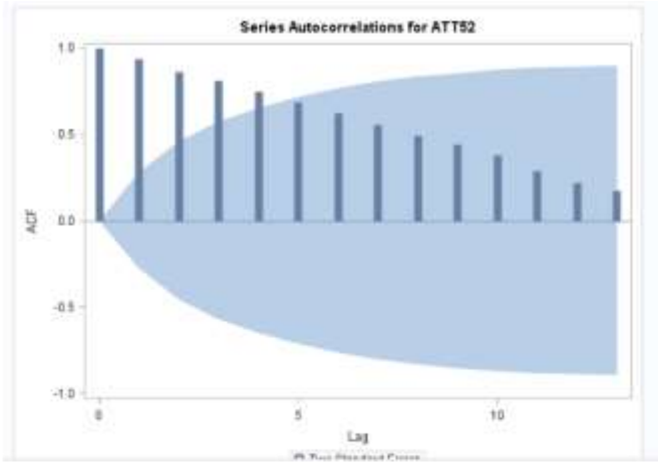
Autocorrelation Check for White Noise									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	212.15	6	<.0001	0.935	0.862	0.814	0.748	0.684	0.622
12	278.08	12	<.0001	0.553	0.493	0.442	0.377	0.289	0.217

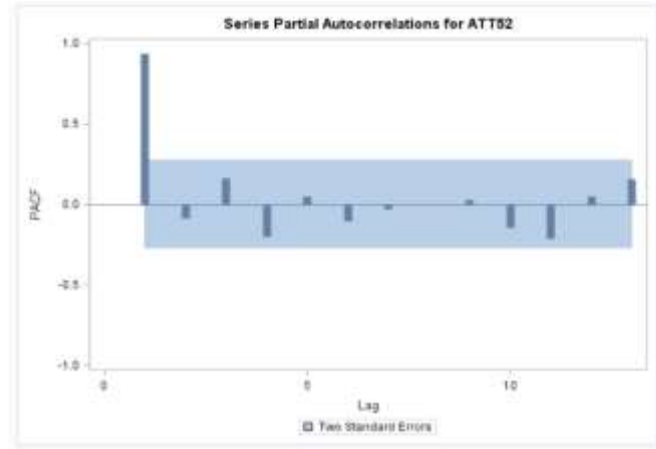
Augmented Dickey-Fuller Unit Root Tests							
Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Zero Mean	0	-0.1530	0.6443	-1.45	0.1357		
	1	-0.1640	0.6418	-1.58	0.1055		
	2	-0.1496	0.6449	-1.71	0.0818		
Single Mean	0	-0.7053	0.9078	-0.38	0.9041	1.08	0.7981
	1	-0.8199	0.8981	-0.45	0.8925	1.30	0.7448
	2	-0.1914	0.9437	-0.13	0.9406	1.44	0.7106
Trend	0	-19.4120	0.0496	-3.69	0.0322	7.08	0.0344
	1	-24.8558	0.0112	-3.65	0.0351	6.83	0.0476
	2	-29.2310	0.0029	-3.77	0.0267	7.47	0.0328

- The series att52 is not white noise and is non-stationary (refer to single mean and trend).

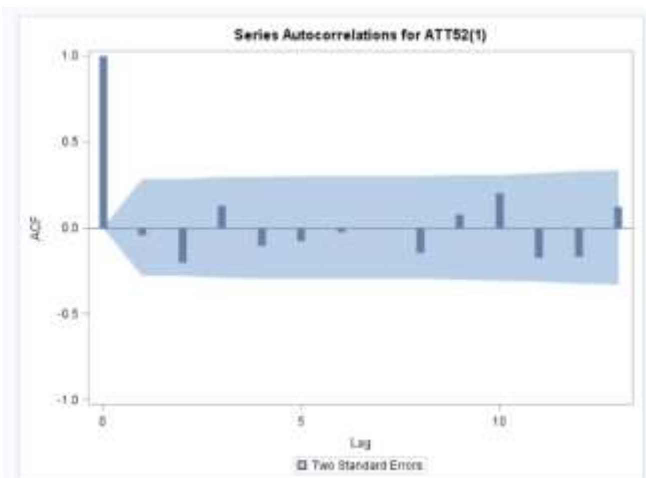


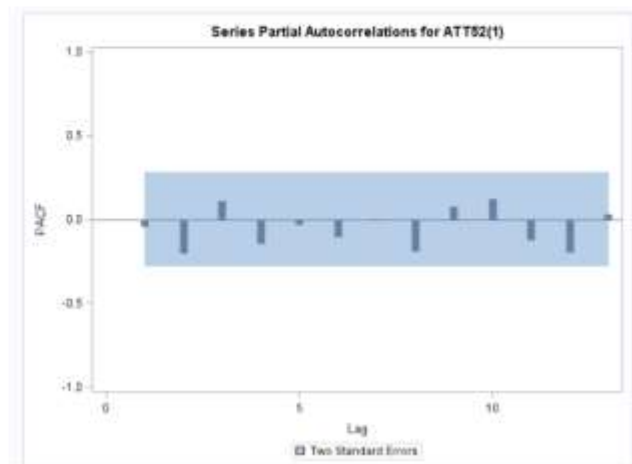
- From the time plot, the series is non-stationary in the mean but stationary in the variance. Thus, we employ (ordinary) differencing.





- The ACF plots are slowly decreasing, which again necessitates (ordinary) differencing.





- After differencing, the ACF and PACF plots suggest that the differenced series is now white noise.

Autocorrelation Check for White Noise

To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	4.25	6	0.6423	-0.043	-0.201	0.128	-0.106	-0.074	-0.026
12	12.73	12	0.3886	-0.002	-0.144	0.076	0.204	-0.174	-0.170

Augmented Dickey-Fuller Unit Root Tests

Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Zero Mean	0	-49.8566	<.0001	-7.00	<.0001		
	1	-67.2812	<.0001	-5.66	<.0001		
	2	-37.9037	<.0001	-4.07	0.0001		
Single Mean	0	-52.1954	0.0004	-7.27	0.0001	26.45	0.0010
	1	-77.6288	0.0004	-6.02	0.0001	18.12	0.0010
	2	-55.3893	0.0004	-4.90	0.0003	12.28	0.0010
Trend	0	-52.6337	<.0001	-7.22	<.0001	26.16	0.0010
	1	-79.7895	<.0001	-6.04	<.0001	18.23	0.0010
	2	-54.2835	<.0001	-4.75	0.0020	11.81	0.0010

- The Ljung-Box Q Statistic and the ADF (refer to zero mean, lag = 1) results support this.

Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	-0.17157	0.12010	-1.43	0.1593	0

Constant Estimate	-0.17157
Variance Estimate	0.7356
Std Error Estimate	0.857672
AIC	130.0613
SBC	131.9931
Number of Residuals	51

- After estimating a random walk model, the constant estimate turns out to be insignificant. Thus, the model for the series must be a random walk with a stochastic trend, given by $\widehat{Z}_t = Z_{t-1}$

Case 8 (Pankratz)

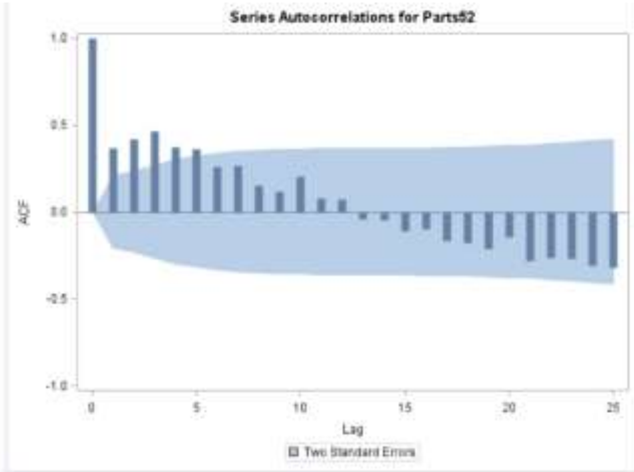
Parts Availability – Weekly (Parts52)

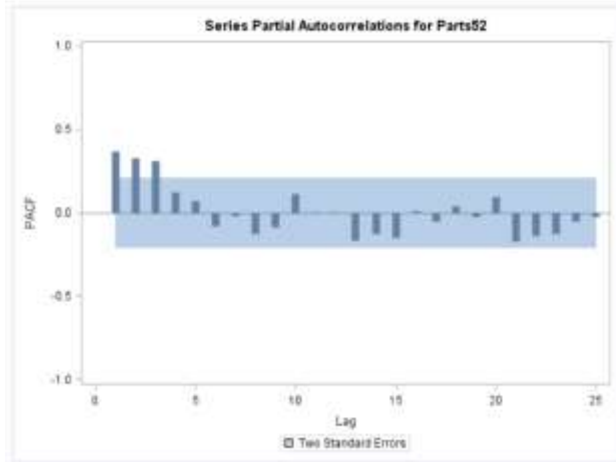
n = 90 ⇒ nlag = 23

Autocorrelation Check for White Noise									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	82.66	6	<.0001	0.370	0.421	0.466	0.374	0.359	0.258
12	98.71	12	<.0001	0.265	0.151	0.117	0.200	0.078	0.073
18	108.31	18	<.0001	-0.040	-0.049	-0.109	-0.100	-0.167	-0.177
24	155.07	24	<.0001	-0.211	-0.143	-0.282	-0.264	-0.272	-0.308

Augmented Dickey-Fuller Unit Root Tests							
Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Zero Mean	0	0.0633	0.6951	0.21	0.7461		
	1	0.0251	0.6863	0.15	0.7284		
	2	0.0237	0.6860	0.22	0.7486		
Single Mean	0	-52.6762	0.0009	-5.73	<.0001	16.50	0.0010
	1	-25.1223	0.0019	-3.34	0.0162	5.59	0.0245
	2	-11.4208	0.0857	-2.13	0.2347	2.30	0.4915
Trend	0	-58.0030	0.0003	-6.19	<.0001	19.25	0.0010
	1	-30.7007	0.0038	-3.78	0.0222	7.31	0.0283
	2	-15.1544	0.1550	-2.56	0.2978	3.53	0.4793

- The series Parts52 is not white noise and is stationary (refer to single mean and trend, lag = 0).





- Lags 1 to 3 are significant in the ACF plot. On the other hand, the PACF plots are exponentially decreasing. Thus, we fit an AR(3) model; i.e.,

$$(1-\phi_1 B-\phi_2 B^2-\phi_3 B^3)Z_t=\theta_0+a_t$$

Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	82.29065	0.71783	114.64	<.0001	0
AR1,1	0.15294	0.10848	1.41	0.1622	1
AR1,2	0.24234	0.11086	2.19	0.0315	2
AR1,3	0.34709	0.11094	3.13	0.0024	3

Constant Estimate	21.20031
Variance Estimate	3.987114
Std Error Estimate	1.996776
AIC	383.7934
SBC	393.7926
Number of Residuals	90

- All p-values are significant except for AR1,1. Thus, using the AR(3) model,

$$(1-\phi_1 B-\phi_2 B^2-\phi_3 B^3)Z_t=\theta_0+a_t$$

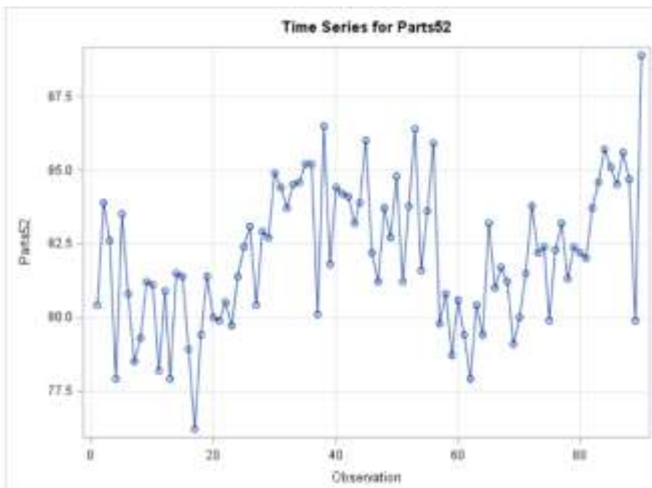
- Note that since this is a pure AR model,

$$(1-\phi_1-\phi_2-\phi_3)\mu=\theta_0$$

Correlations of Parameter Estimates				
Parameter	MU	AR1,1	AR1,2	AR1,3
MU	1.000	0.053	0.038	0.040
AR1,1	0.053	1.000	-0.304	-0.286
AR1,2	0.038	-0.304	1.000	-0.334
AR1,3	0.040	-0.286	-0.334	1.000

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	2.62	3	0.4532	-0.038	-0.061	-0.030	0.085	0.118	-0.002
12	8.35	9	0.4991	0.040	-0.076	-0.042	0.189	0.061	0.083
18	11.91	15	0.6856	-0.117	-0.020	-0.071	0.028	-0.101	-0.046
24	16.69	21	0.7295	-0.032	0.133	-0.080	-0.054	-0.068	-0.084

- Moreover, the parameter estimates are not significantly correlated and the residuals are white noise.

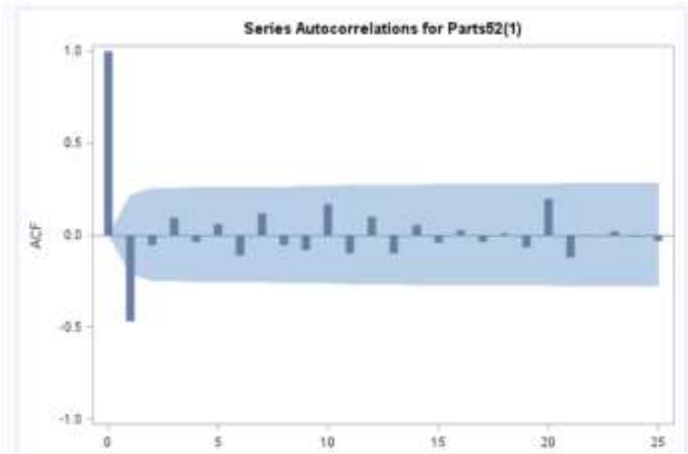


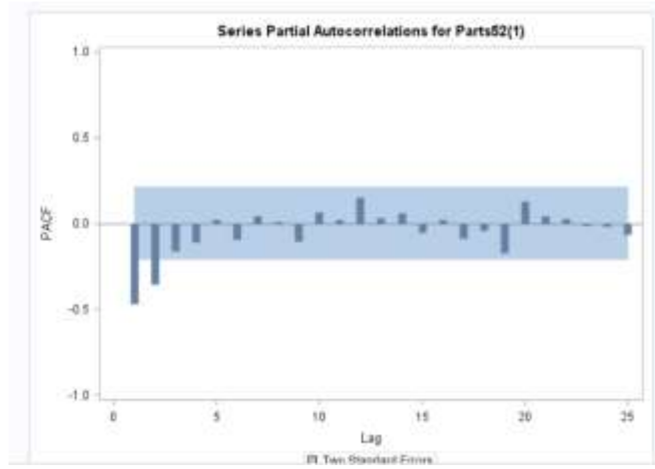
- However, from the time plot, the series is not stationary in the mean. Thus, we can employ (ordinary) differencing.

Autocorrelation Check for White Noise									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	23.15	6	0.0007	-0.471	-0.054	0.092	-0.038	0.060	-0.108
12	30.24	12	0.0026	0.118	-0.053	-0.082	0.165	-0.097	0.097
18	32.09	18	0.0214	-0.102	0.054	-0.040	0.028	-0.038	0.008
24	38.94	24	0.0277	-0.066	0.195	-0.123	-0.005	0.023	-0.009

Augmented Dickey-Fuller Unit Root Tests							
Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Zero Mean	0	-136.132	0.0001	-15.77	<.0001		
	1	-333.241	0.0001	-12.08	<.0001		
	2	-1933.57	0.0001	-8.70	<.0001		
Single Mean	0	-136.128	0.0001	-15.69	<.0001	123.07	0.0010
	1	-333.477	0.0001	-12.01	<.0001	72.16	0.0010
	2	-2020.60	0.0001	-8.67	<.0001	37.67	0.0010
Trend	0	-135.979	0.0001	-15.60	<.0001	122.17	0.0010
	1	-333.911	0.0001	-11.98	<.0001	71.89	0.0010
	2	-2161.00	0.0001	-8.63	<.0001	37.26	0.0010

- After differencing, we note that the data is not white noise and is stationary (refer to single mean and trend, lag = 1).





- The ACF plot is significant only at lag = 1 while the PACF plots are exponentially decreasing. Thus, we try and fit an MA(1) model.

Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	0.04642	0.05907	0.79	0.4341	0
MA1,1	0.73359	0.07405	9.91	<.0001	1

Constant Estimate	0.046417
Variance Estimate	4.138964
Std Error Estimate	2.034444
AIC	380.9679
SBC	385.9452
Number of Residuals	89

- All parameter estimates are significant except the constant estimate. Thus, we drop the constant term from the MA(1) model.
- Note that since this is a pure MA model,

$$\mu = \theta_0$$

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	3.01	5	0.6992	-0.094	-0.009	0.121	0.042	0.069	-0.040
12	6.96	11	0.8025	0.075	-0.049	-0.047	0.157	0.004	0.063
18	8.75	17	0.9477	-0.073	-0.021	-0.077	-0.016	-0.061	-0.024
24	13.11	23	0.9496	-0.034	0.134	-0.097	-0.049	-0.028	-0.071

- After fitting the MA(1) model with no constant estimate, the residuals are now white noise.

Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MA1,1	0.72530	0.07472	9.71	<.0001	1

Variance Estimate	4.120356
Std Error Estimate	2.029866
AIC	379.584
SBC	382.0726
Number of Residuals	89

- Thus, we are left to choose between the AR(3) and MA(1) model. Given the AIC and SBC, they do not significantly differ from each other, but MA(1) has a slightly lower values for both. By parsimony, MA(1) uses less parameters. Thus, we prefer the MA(1) model.

Case 13 (Pankratz)

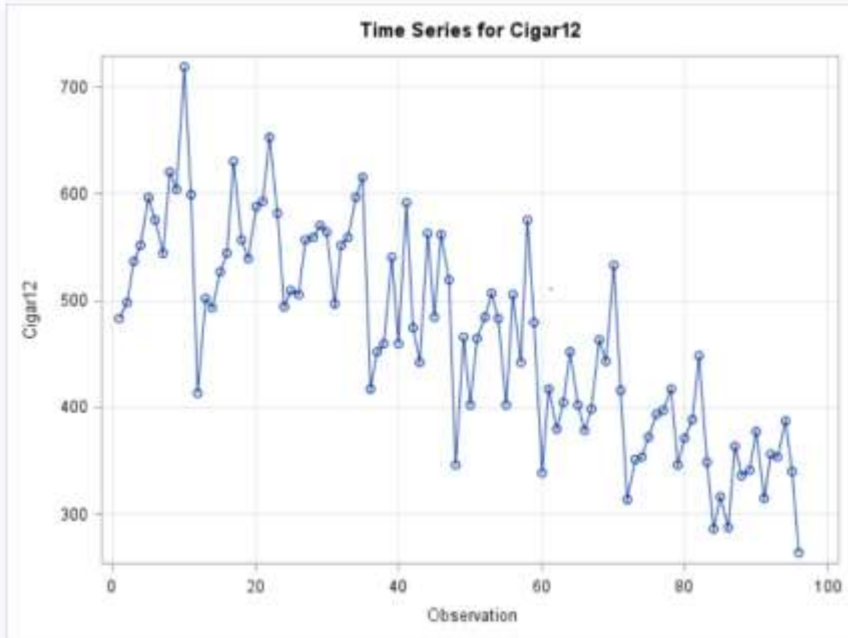
Cigarette Consumption – Monthly (Cigar12)

n = 96 ⇒ nlag = 24

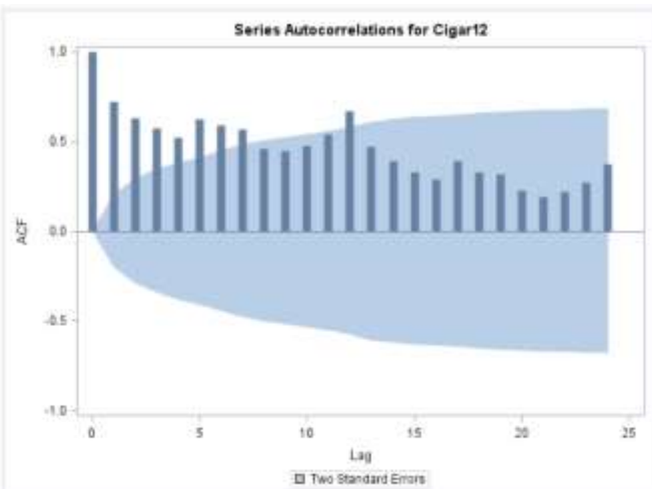
Autocorrelation Check for White Noise									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	228.77	6	<.0001	0.721	0.629	0.575	0.522	0.625	0.587
12	414.36	12	<.0001	0.568	0.461	0.448	0.476	0.537	0.671
18	510.79	18	<.0001	0.469	0.391	0.330	0.290	0.393	0.329
24	567.38	24	<.0001	0.316	0.224	0.193	0.218	0.270	0.373

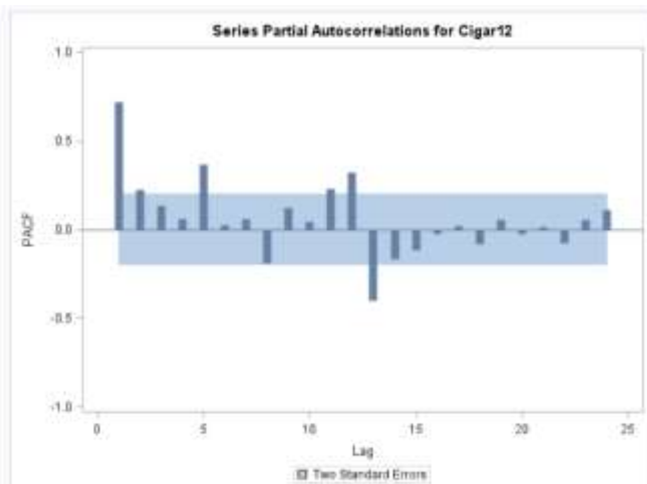
Augmented Dickey-Fuller Unit Root Tests							
Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Zero Mean	0	-1.3653	0.4127	-0.96	0.2977		
	1	-0.8640	0.4950	-0.85	0.3448		
	2	-0.7365	0.5191	-0.88	0.3313		
Single Mean	0	-23.1770	0.0035	-3.41	0.0131	5.86	0.0188
	1	-13.3518	0.0520	-2.33	0.1642	2.82	0.3615
	2	-9.5409	0.1400	-1.88	0.3390	1.93	0.5833
Trend	0	-71.4651	0.0003	-7.57	<.0001	28.84	0.0010
	1	-83.4495	0.0003	-6.56	<.0001	21.77	0.0010
	2	-135.556	0.0001	-6.38	<.0001	20.53	0.0010

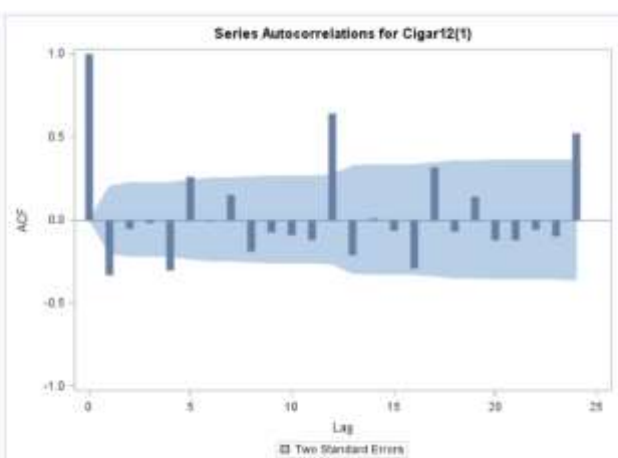
- The series Cigar12 is not white noise and

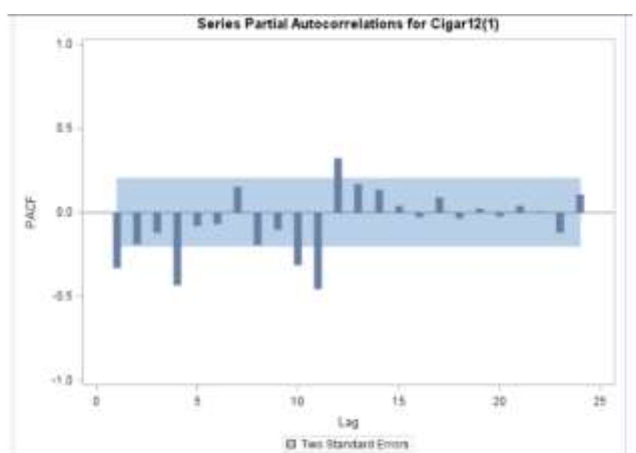


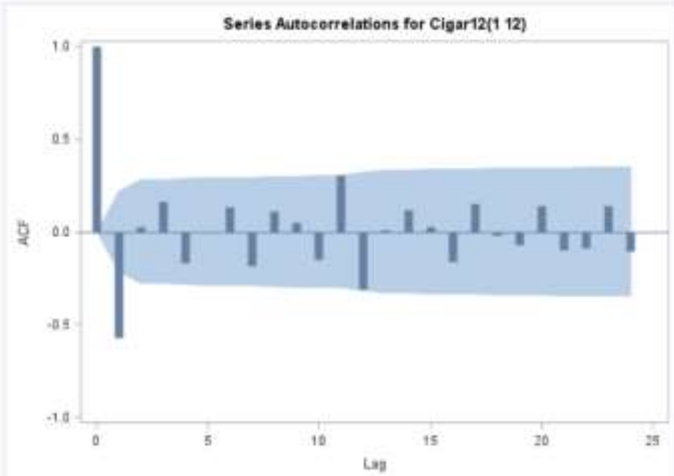
- The series is non-stationary in the mean. Thus, we employ (ordinary) differencing.

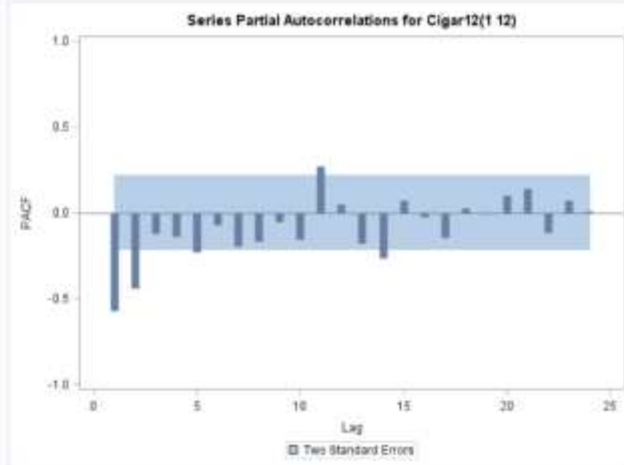












Conditional Least Squares Estimation

Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	-0.74690	0.32227	-2.32	0.0231	0
MA1,1	0.85473	0.06358	13.44	<.0001	1
MA2,1	0.72520	0.18557	3.91	0.0002	12
AR1,1	0.23239	0.22868	1.02	0.3126	12

Constant Estimate	-0.57333
Variance Estimate	1170.431
Std Error Estimate	34.21156
AIC	825.8498
SBC	835.5251
Number of Residuals	83

Conditional Least Squares Estimation

Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	-0.70872	0.32938	-2.15	0.0344	0
MA1,1	0.84780	0.06341	13.37	<.0001	1
MA2,1	0.56554	0.10634	5.32	<.0001	12

Constant Estimate	-0.70872
Variance Estimate	1176.092
Std Error Estimate	34.2942
AIC	825.2943
SBC	832.5508
Number of Residuals	83

Correlations of Parameter Estimates			
Parameter	MU	MA1,1	MA2,1
MU	1.000	-0.154	-0.218
MA1,1	-0.154	1.000	-0.147
MA2,1	-0.218	-0.147	1.000

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	9.54	4	0.0489	-0.127	0.018	0.190	-0.166	-0.040	0.159
12	20.47	10	0.0251	-0.198	0.089	0.119	-0.103	0.193	0.073
18	30.33	16	0.0163	-0.117	0.155	0.070	-0.118	0.182	0.067
24	41.69	22	0.0068	-0.082	0.154	-0.092	-0.122	0.198	-0.076

Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	-0.64159	0.41380	-1.55	0.1250	0
MA1,1	0.87391	0.05790	15.09	<.0001	1
AR1,1	-0.33672	0.10990	-3.06	0.0030	12

Constant Estimate	-0.85762
Variance Estimate	1259.252
Std Error Estimate	35.48594
AIC	830.9649
SBC	838.2214
Number of Residuals	83

Correlations of Parameter Estimates			
Parameter	MU	MA1,1	AR1,1
MU	1.000	-0.194	0.031
MA1,1	-0.194	1.000	0.127
AR1,1	0.031	0.127	1.000

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	10.87	4	0.0281	-0.162	0.040	0.206	-0.177	-0.025	0.143
12	20.89	10	0.0219	-0.180	0.107	0.099	-0.107	0.192	-0.052
18	33.72	16	0.0059	-0.143	0.171	0.071	-0.103	0.225	0.082
24	48.77	22	0.0009	-0.063	0.157	-0.102	-0.138	0.185	-0.192

OUTLIER DETECTION

W8 (with outliers)

- There are two potential additive outliers and one temporary level shift outlier (for around 6 time points).
- From the ACF plot, we need to employ ordinary differencing.
- After ordinary differencing, the PACF plots are exponentially decreasing while the ACF plot is only significant at lag 1. Thus, we can fit an MA(1) model/ARIMA(0,1,1); i.e., (1-B)Z_t = theta_0 + (1-theta_1B)a_t
- Since there are some significant lags still, we can try to do seasonal differencing. Since the plots aren't as good as when we only did ordinary differencing, we stick to that. The significant lags may be due to outliers.
- We fit an ARIMA(0,1,1)(1,0,1)$_4$ model. We also detect outliers.
- tc(5,6,7,8) – detects temporary change of 5, 6, 7 and 8 periods.
- We detected 3 outliers – one temporary change of 6 periods and 2 additive outliers.
- At time 7, magnitude came up by 73.638 more than the usual
- At time 98 and the 6 time points after, magnitude came up by an average of 81.959 compared to the usual
- However, model is still not adequate because the residuals are still not white noise. Thus, we check the ACF and PAF of the residuals.
- From the ACF and PACF, since the lag at time 16 is significant, we fit a non-hierarchical model. p = (1)(4 16) q = (1)(4 16)
- MA(16) and AR(16) did not help with the model.
- We again employ seasonal differencing.
- From the ACF and PACF plots, we employ an ARIMA(1,1,1)(0,1,1)$_4$ model.
- Doing backward elimination, we remove AR(1); i.e., an ARIMA(0,1,1)(0,1,1)$_4$ model
- Only the constant estimate is insignificant. Also, the residuals are now white noise.