Stochastic Process and Forecasting Assignment

Nick Head

15 April 2015

Question 1
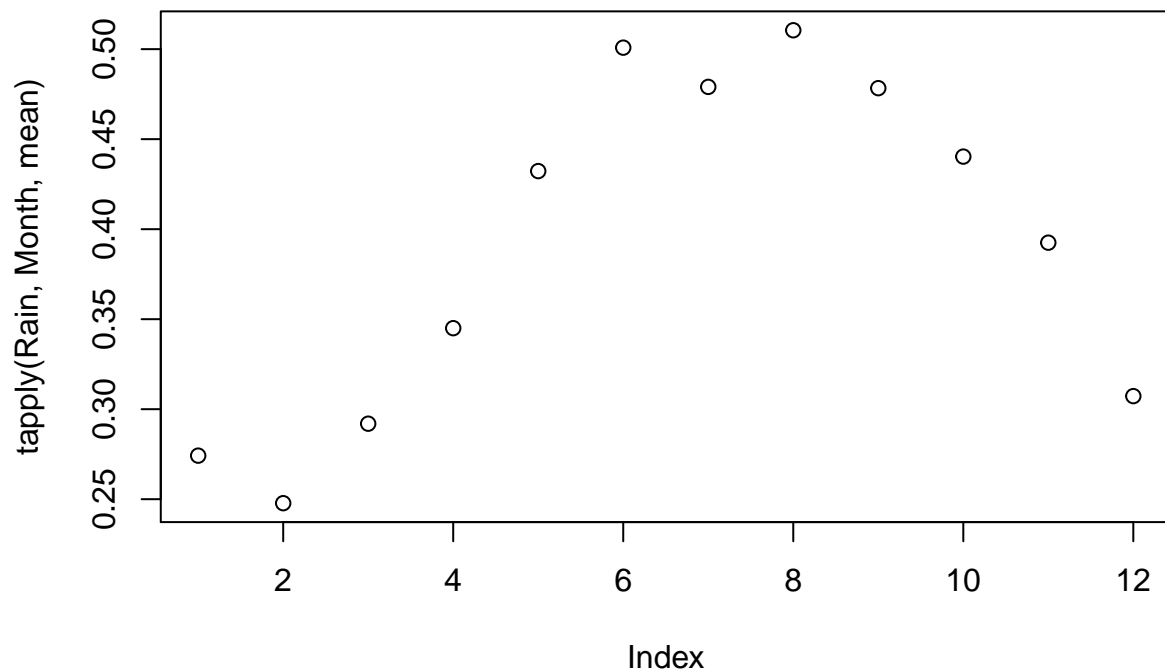
i)

The overall proportion of rainy days to sunny days is 0.392. The proportion of rainy days seems to increase over the (Southern Hemisphere) winter months and decreases over the summer months.

```
mean(Rain)
```

```
## [1] 0.392
```

```
plot(tapply(Rain,Month,mean))
```


ii)

For $m = 1$:

```
m <- 1
pi0 <- c(mean(Rain))
gamma0 <- matrix(c(1),nrow=m)
Melbourne.mle1 <- binary.HMM.mle(Rain,m,pi0,gamma0)
```

The fitted parameter values are:

$$\boldsymbol{\pi} = (\pi_1) = (0.392)$$

and

$$\boldsymbol{\Gamma} = (1)$$

For $m = 2$:

```
m <- 2
pi0 <- c(0.001,0.9)
gamma0 <- matrix(c(0.8,0.2,0.4,0.6),nrow=m,byrow=TRUE)
Melbourne.mle2 <- binary.HMM.mle(Rain,m,pi0,gamma0)
```

The fitted parameter values are:

$$\boldsymbol{\pi} = (\pi_1, \pi_2) = (0, 0.848)$$

and

$$\boldsymbol{\Gamma} = \begin{pmatrix} 0.702 & 0.298 \\ 0.346 & 0.654 \end{pmatrix}$$

For $m = 3$:

```
m <- 3
pi0 <- c(0.01,0.5,0.99)
gamma0 <-matrix(c(0.8,0.1,0.1,0.1,0.8,0.1,0.2,0.2,0.6),nrow=m,byrow=TRUE)
Melbourne.mle3 <- binary.HMM.mle(Rain,m,pi0,gamma0)
```

The fitted parameter values are:

$$\boldsymbol{\pi} = (\pi_1, \pi_2\, \pi_3) = \left(2.022 \times 10^{-15}, 0.581, 1\right)$$

and

$$\boldsymbol{\Gamma} = \begin{pmatrix} 0.747 & 1.112 \times 10^{-39} & 0.253 \\ 3.815 \times 10^{-75} & 0.967 & 0.033 \\ 0.457 & 0.017 & 0.526 \end{pmatrix}$$

iii)

	mllk	AIC	BIC
1	9785.50	19572.99	19580.58
2	9254.95	18517.89	18548.25
3	9208.61	18435.22	18503.52

As can be seen in the table, $m = 3$ minimises both the AIC and the BIC and therefore should be chosen as the most appropriate.

As $\pi_1 \approx 0$ and two of the γ_{ij} are also close to zero, three of the natural parameters are close to their boundary values and this could cause convergence issues. A modificied HMM could be used where $\pi_1 = 0$.

Question 2

i)

The following SAS code was used to identify and estimate appropriate univariate ARIMA models for each of the `avg` and `end` exchange rate time series:

```
proc arima data=LatDol;
identify var=avg(1) nlag=18;
estimate q=1 noint;
identify var=end nlag=18;
estimate p=1;
run;
```

The diagnostics plots indicates that the first difference of `avg` has an ACF cut-off at lag 1, thereby leading us to believe the data follows an Arima(0,1,1) model with the following model equation:

$$\Delta Y_t = -0.36347\epsilon_{t-1} + \epsilon_t$$
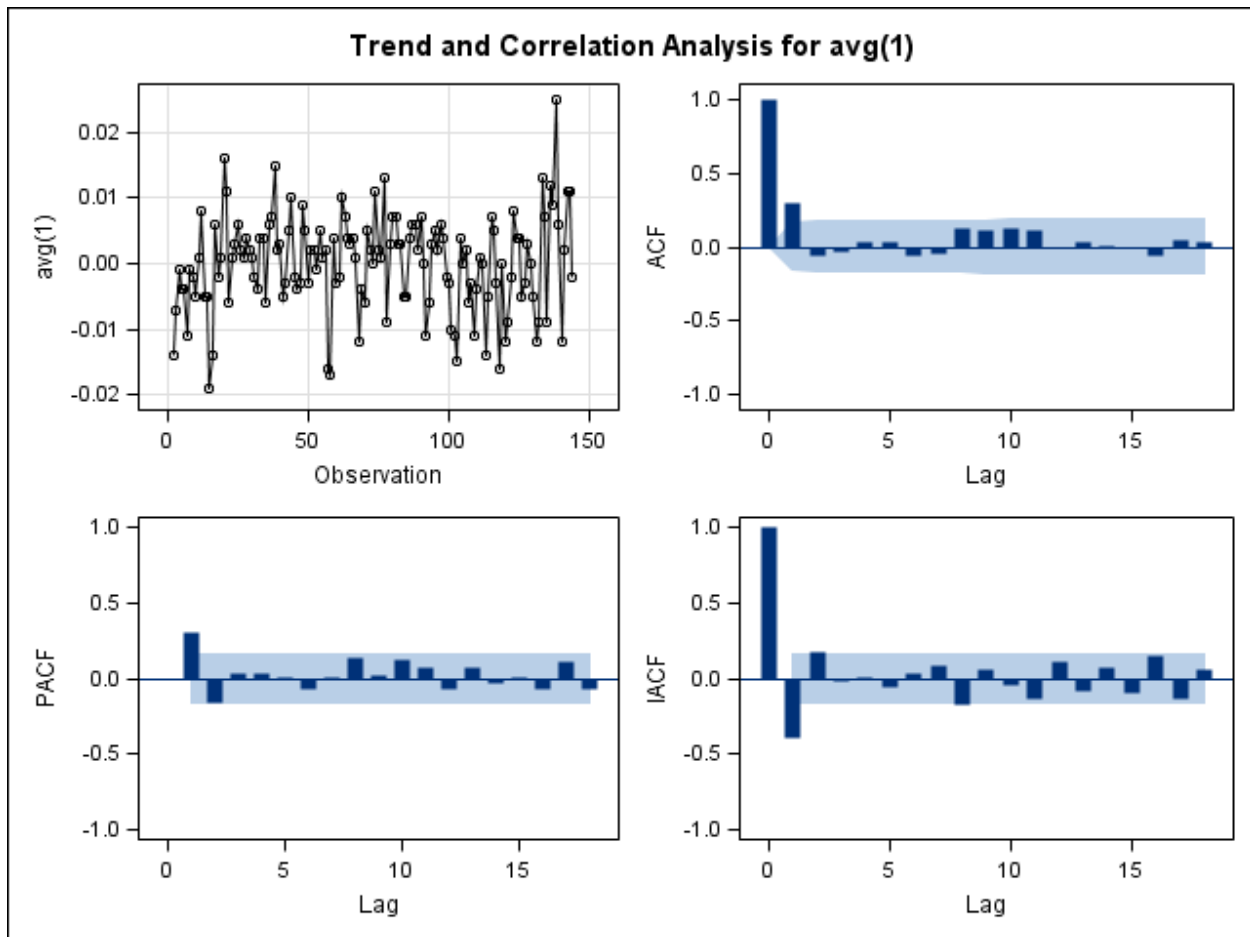
In addition the Ljung-Box chi-square statistics do not have any significant p-values, giving further weight to the chosen model:
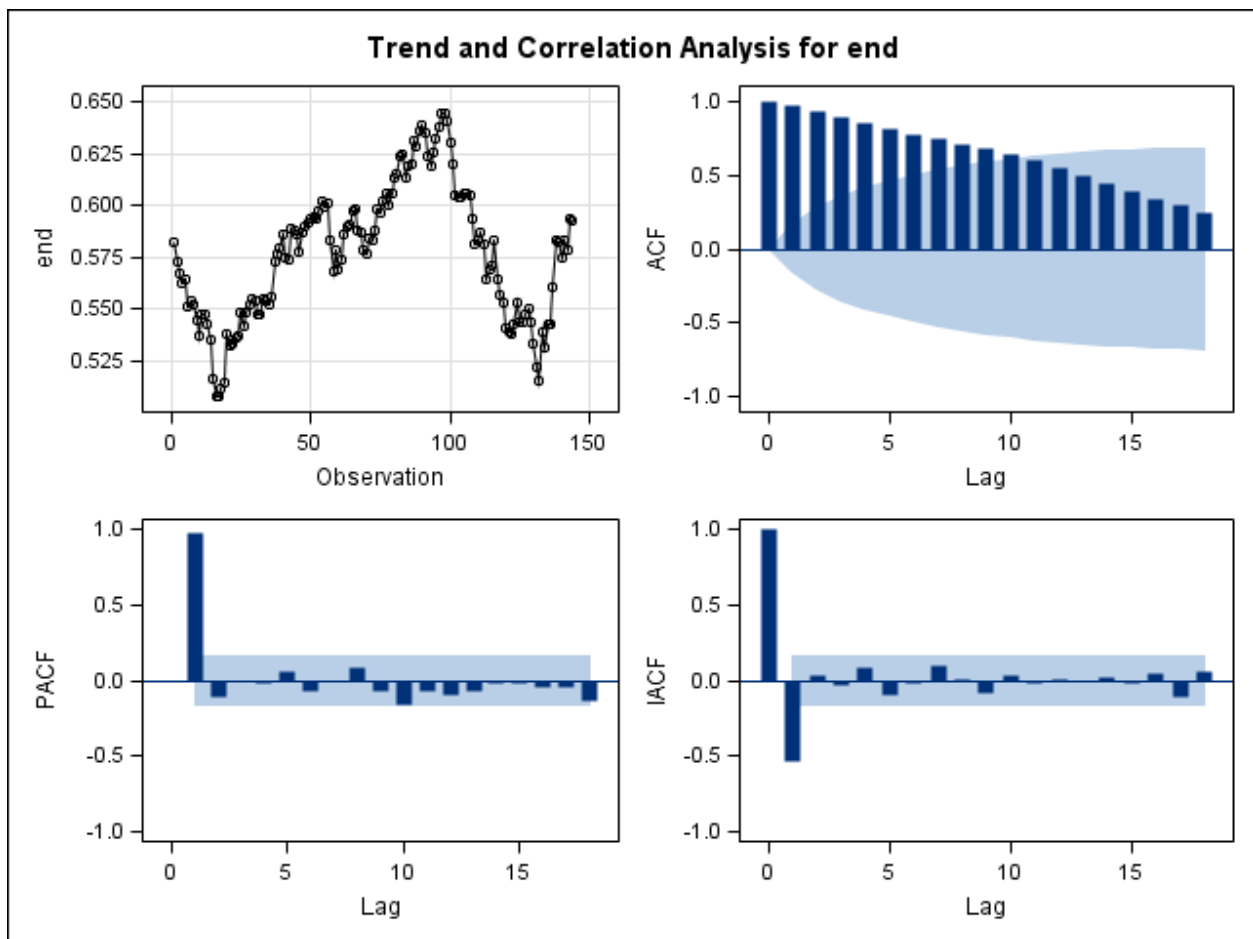
Autocorrelation Check of Residuals

To Lag	Chi-Square	DF	Pr > ChiSq			Autocorrelations			
6	1.30	5	0.9346	-0.010	-0.041	-0.034	0.032	0.046	-0.051
12	8.50	11	0.6681	-0.071	0.133	0.046	0.079	0.104	-0.066
18	11.49	17	0.8299	0.059	-0.022	0.030	-0.087	0.075	0.015
24	16.13	23	0.8497	-0.014	-0.121	-0.044	-0.074	-0.025	-0.066

end is found to follow an AR(1) model with the following model equation:

$$Y_t = 0.58167 + 0.96945Y_{t-1} + \epsilon_t$$

This model was identified by the lag 1 cut-off in the PACF as shown in the diagnostics plots below:

In addition the Ljung-Box chi-square statistics do not have any significant p-values, giving further weight to the chosen model:

Autocorrelation Check of Residuals

To Lag	Chi-Square	DF	Pr > ChiSq		Autocorrelations	

4

6	2.84	5	0.7239	0.116	0.043	0.007	-0.048	0.040	0.006
12	11.19	11	0.4279	-0.050	0.081	0.153	0.092	0.081	0.076
18	13.25	17	0.7192	-0.035	-0.019	-0.019	-0.022	0.095	-0.033
24	18.91	23	0.7064	-0.024	-0.076	-0.109	-0.044	-0.028	-0.108

ii)

The following SAS code identifies and fits a VAR(p) model to the bivariate data:

```
proc varmax data=LatDol;
model end avg /
minic = (type=sbc p=(0:10) q=0) noint dif=(end(1) avg(1)) PRINT=DIAGNOSE;
output lead=3;
run;
```

Which gives the following truncated output:

<div align="center">

Model Parameter Estimates

</div>

Equation	Parameter	Estimate	Standard Error	t Value	Pr > |t|	Variable
end	AR1_1_1	0.03600	0.11185	0.32	0.7480	end(t-1)
	AR1_1_2	0.11050	0.12356	0.89	0.3727	avg(t-1)
avg	AR1_2_1	0.70995	0.07553	9.40	0.0001	end(t-1)
	AR1_2_2	-0.22236	0.08344	-2.66	0.0086	avg(t-1)

<div align="center">

Covariances of Innovations

</div>

Variable	end	avg
end	0.00007	0.00004
avg	0.00004	0.00003

Based off which the fitted model is as follows:

$$\begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} .03600 & 0.11050 \\ .70995 & 0.22236 \end{pmatrix} \begin{pmatrix} x_{t-1} \\ y_{t-1} \end{pmatrix} + \epsilon_t$$

The following Portmanteau Statistics shows significant p-values at all lags indicating that this VAR(1) model does not fit adequatly:

<div align="center">

Portmanteau Test for Cross
Correlations of Residuals

</div>

Up To Lag	DF	Chi-Square	Pr > ChiSq
2	4	23.29	0.0001
3	8	23.46	0.0028
4	12	26.37	0.0095
5	16	34.09	0.0053
6	20	35.06	0.0198
7	24	43.30	0.0092
8	28	44.79	0.0232
9	32	51.56	0.0157
10	36	58.37	0.0106
11	40	63.56	0.0103
12	44	64.87	0.0219

The following outputs show the forecasts of the average exchange rate for the next three months of 2006 are: 0.59273, 0.59261, 0.59261

```
                                Forecasts

                                    Standard
    Variable        Obs      Forecast      Error       95% Confidence Limits

    end             145       0.59274     0.00810     0.57686        0.60863
                    146       0.59270     0.01204     0.56911        0.61630
                    147       0.59269     0.01530     0.56270        0.62268
    avg             145       0.59273     0.00547     0.58201        0.60346
                    146       0.59261     0.01102     0.57101        0.61421
                    147       0.59261     0.01435     0.56448        0.62074
```

iii)

The following SAS code was used to identify and estimate an appropriate transfer function model where `avg` was specified as the output variable and `end` the input variable:

```
proc arima data=LatDol;
identify var=end(1) nlag=18;
estimate q=1 noint;
identify var=avg(1) crosscor=end(1) nlag=18;
estimate q=1 input=(1 $ (1) /  end) noint;
forecast lead=3 out=results;
run;
```

The following diagnostics plots show from the lag 1 cutoff that an ARIMA(0,1,1) model is appropriate for the first difference of `end`:
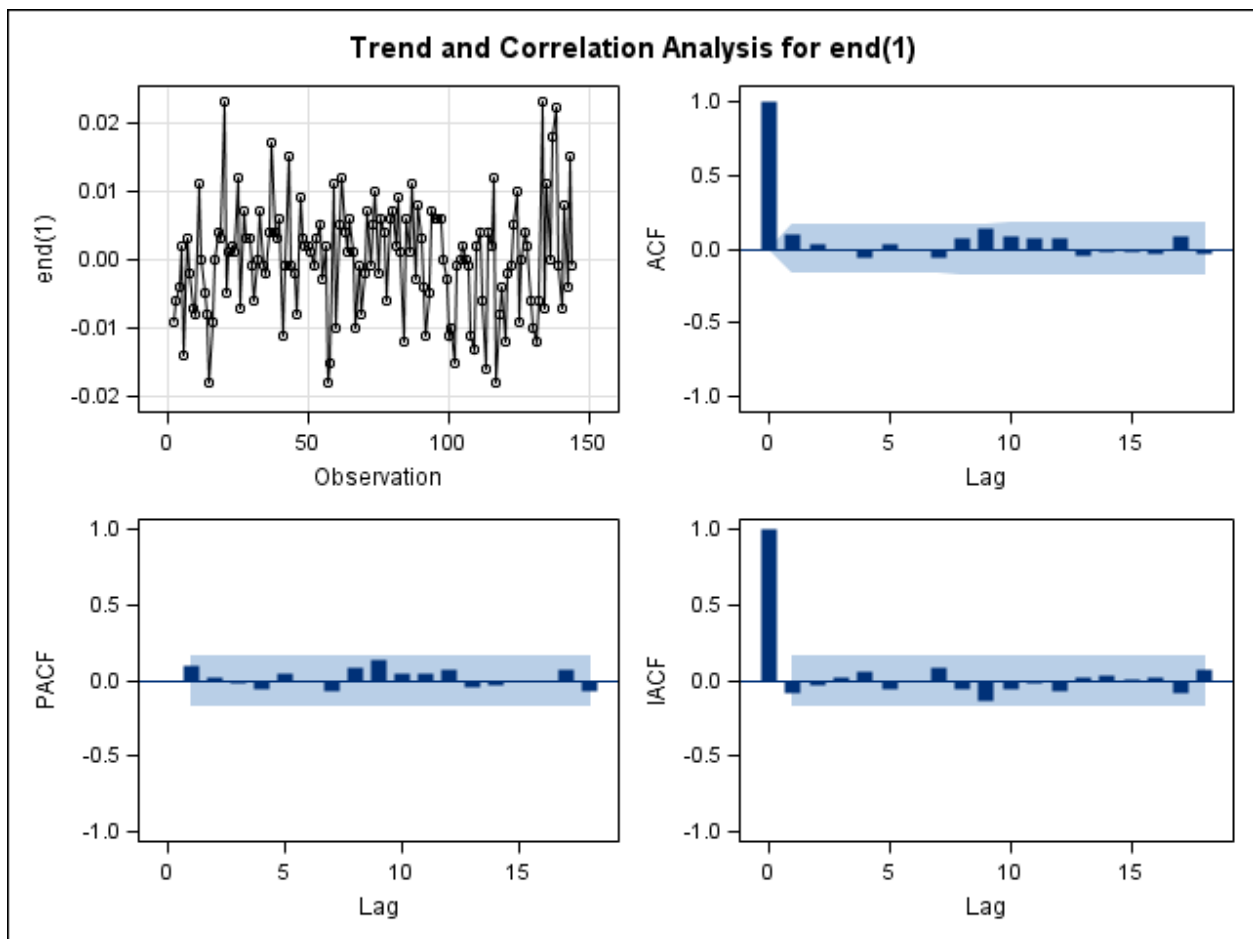
Hence the estimation procedure shows the following model for `end`:

$$\Delta X_t = \eta_t + 0.09801\eta_{t-1}$$

The following output shows the p-values from the portmanteau statistics indicate the model fits well:

```
                    Autocorrelation Check of Residuals
```

To Lag	Chi-Square	DF	Pr > ChiSq	-------------------Autocorrelations-------------------					
6	0.95	5	0.9668	0.003	0.029	-0.003	-0.065	0.035	-0.003
12	7.14	11	0.7880	-0.069	0.066	0.134	0.066	0.062	0.069
18	9.75	17	0.9136	-0.045	-0.017	-0.018	-0.034	0.101	-0.044
24	14.42	23	0.9143	-0.015	-0.067	-0.102	-0.033	-0.013	-0.105

The next statement shows the output of the second `identify` statement which shows the following cross-correlation function:
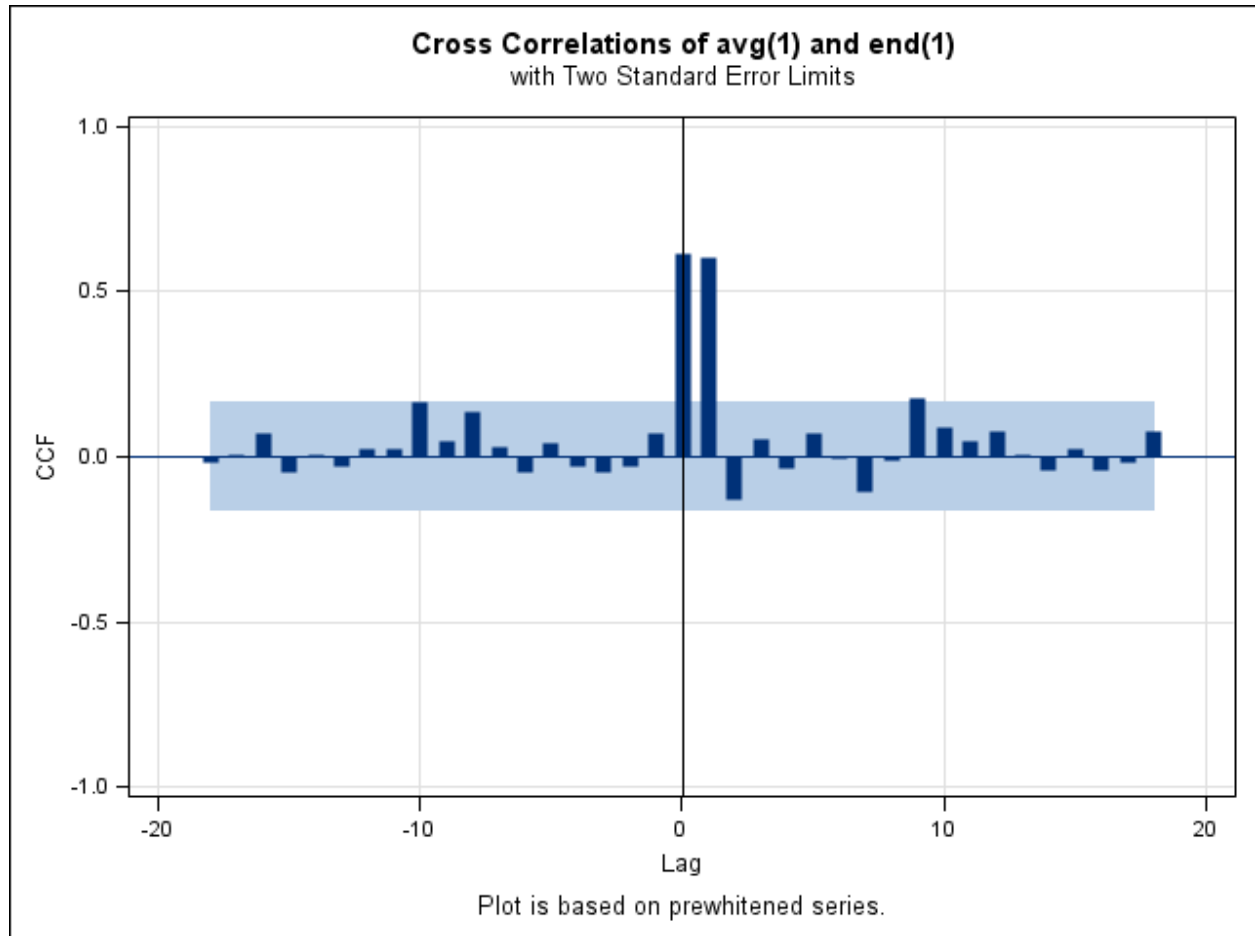
This indicates that the transfer function is in the form of a lag-1 pulse function. There is a sharp (not decayed) response after lag 1 which indicates a transfer function of the form:

$$Y_t = \omega X_t + U_t$$

NB: The lack of intercept is due to the first differencing, and the lack of denominator terms due to the pulse intervention

With the subsequent estimation run, SAS provides us the estimated model of:

$$Y_t = (1.1118 - 0.14899L)X_{t-1} + U_t$$

Where
$$U_t = (1 - 0.95978L)\epsilon_t$$

The ARIMA Procedure

Moving Average Factors

Factor 1: 1 - 0.95978 B**(1)

Input Number 1

Input Variable   end
Shift   1
Period(s) of Differencing   1

Numerator Factors

Factor 1: 1.1118 - 0.14899 B**(1)

The forecast values for avg are given as:

Forecasts for variable avg

Obs	Forecast	Std Error	95% Confidence Limits	
145	0.5924	0.0051	0.5824	0.6024
146	0.5922	0.0104	0.5719	0.6125
147	0.5923	0.0135	0.5658	0.6188