Time Series Analysis - Homework 1

Tom Davidson

18/09/2018

Source code can be found here: "https://github.com/t-davidson/time-series-analysis-fall2018/tree/master/HW1"

Loading packages and data

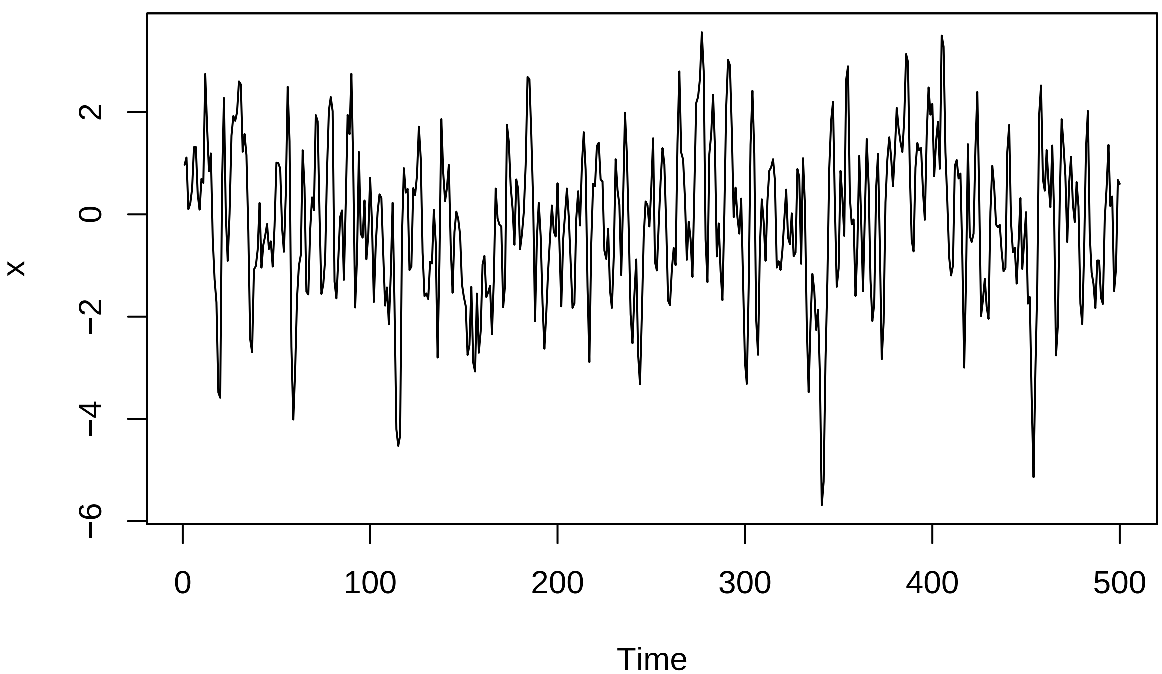
```
require("foreign")
require("tseries")
require("zoo")
require("urca")
require("knitr")

data <- read.dta('data/ts_hw1.dta')
x <- zoo(data$x, data$t)
z <- zoo(data$z[1:60], data$t[1:60]) # values after 60 are null
```

Analyzing series x

1.1. Plot time series x

```
plot(x, xlab = 'Time')
```



Looking at the series `x` it appears that the mean is around zero (actually value is -0.16). The variance looks relatively constant, although there are some large spikes later in the series suggesting that it may not be a white noise process. Overall this suggests that the series is stationary.

1.2. Show correlogram

It appears that there is no equivalent function in R to produce the same output as the `corrgram` command in Stata (I even wrote a post up on Stackoverflow but have not had any useful replies). I was able to manually construct the first 5 columns of the table but was not able to produce the final two columns. Since the last two columns show the same information as the following plots I hope that this is not too much of a problem.

I also wanted to compare my results to those produced by Stata to ensure that my logic was correct. It turns out that R's default `acf` function includes the first observation (where correlation between y_t and y_t is equal to 1) so I had to remove it. Otherwise the output exactly matches the results I obtained from Stata. The PACF, on the other hand, is estimated differently. R uses the Yule-Walker method while Stata relies upon a simple regression based approach; the former can be estimated in Stata using the command `pac x, yw` but there is not an equivalent to the latter in R. Nonetheless the estimates are almost identical so this shouldn't have any impact on our interpretations. Additionally, the Q statistic in R does not the Ljung-Box method by default, so it needed to be specified. These results exactly match those produced by Stata.

```r
max_lags = 40
acf_vec <- acf(x, plot=FALSE, lag.max=max_lags)$acf
pacf_vec <- acf(x, plot=FALSE, lag.max=max_lags, type = 'partial')$acf
Q_stats <- c()
Q_pvals <- c()
for (i in 1:max_lags) {
  Q = Box.test(x, lag=i, type="Ljung-Box")
  Q_stats[[i]] <- Q$statistic
  Q_pvals[[i]] <- Q$p.value
}
corrgram <- cbind(LAG=seq(1,max_lags), ACF=acf_vec[2:41], PAC=pacf_vec, Q=Q_stats, "Prob>Q"=Q_pvals)
kable(corrgram)
```

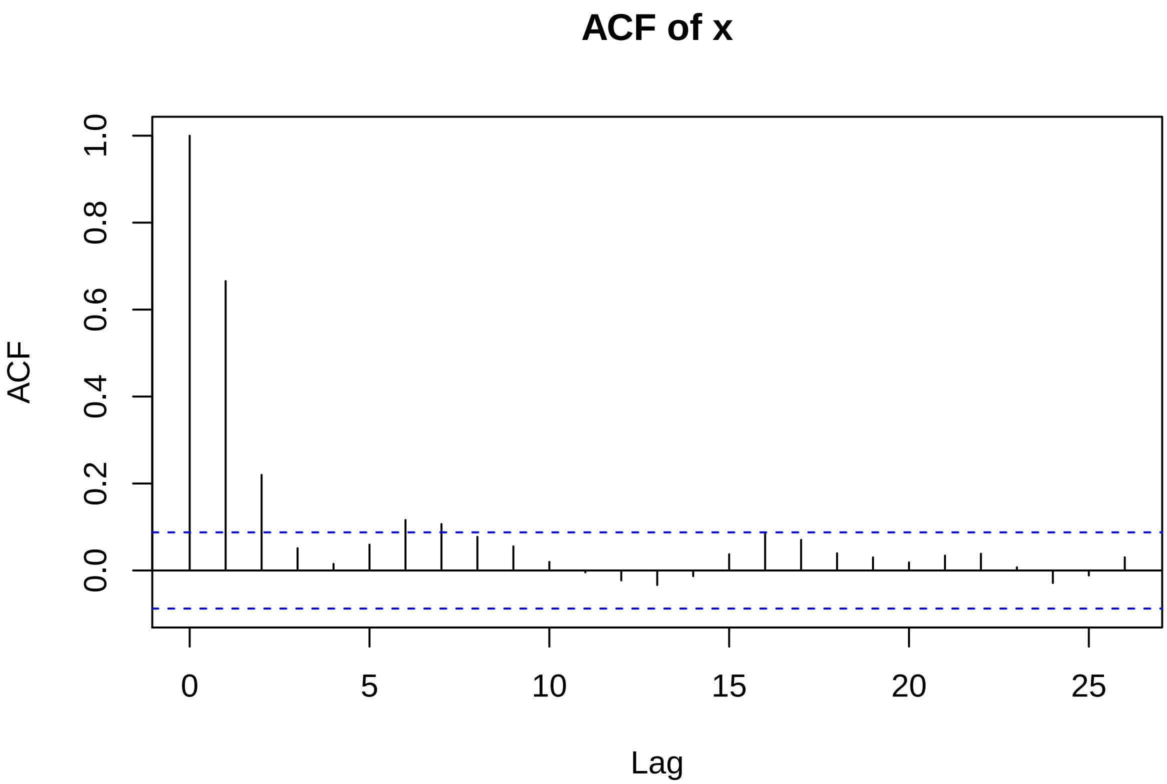
LAG	ACF	PAC	Q	Prob>Q
1	0.6657	0.6657	222.9	0
2	0.2201	-0.4006	247.3	0
3	0.0512	0.2411	248.6	0
4	0.0152	-0.1373	248.8	0
5	0.0594	0.2013	250.5	0
6	0.1162	-0.0488	257.4	0
7	0.1069	0.0365	263.2	0
8	0.0774	0.0187	266.3	0
9	0.0555	-0.0009	267.8	0
10	0.0198	-0.0271	268.0	0
11	-0.0044	0.0060	268.1	0
12	-0.0229	-0.0537	268.3	0
13	-0.0335	0.0187	268.9	0
14	-0.0132	0.0054	269.0	0
15	0.0374	0.0676	269.7	0
16	0.0852	0.0297	273.5	0
17	0.0705	-0.0444	276.1	0
18	0.0398	0.0671	276.9	0
19	0.0303	-0.0217	277.4	0
20	0.0187	0.0096	277.6	0

LAG	ACF	PAC	Q	Prob>Q
21	0.0343	0.0477	278.2	0
22	0.0387	-0.0642	279.0	0
23	0.0073	0.0116	279.0	0
24	-0.0290	-0.0639	279.4	0
25	-0.0116	0.0950	279.5	0
26	0.0304	-0.0254	280.0	0
27	0.0317	-0.0141	280.5	0
28	-0.0075	-0.0344	280.5	0
29	-0.0580	-0.0365	282.3	0
30	-0.0973	-0.0477	287.4	0
31	-0.1040	-0.0143	293.2	0
32	-0.0599	0.0252	295.1	0
33	-0.0052	0.0145	295.1	0
34	0.0168	-0.0069	295.3	0
35	0.0324	0.0593	295.8	0
36	0.0336	-0.0187	296.5	0
37	0.0371	0.0632	297.2	0
38	0.0380	-0.0296	298.0	0
39	0.0361	0.0596	298.7	0
40	0.0492	0.0114	300.0	0

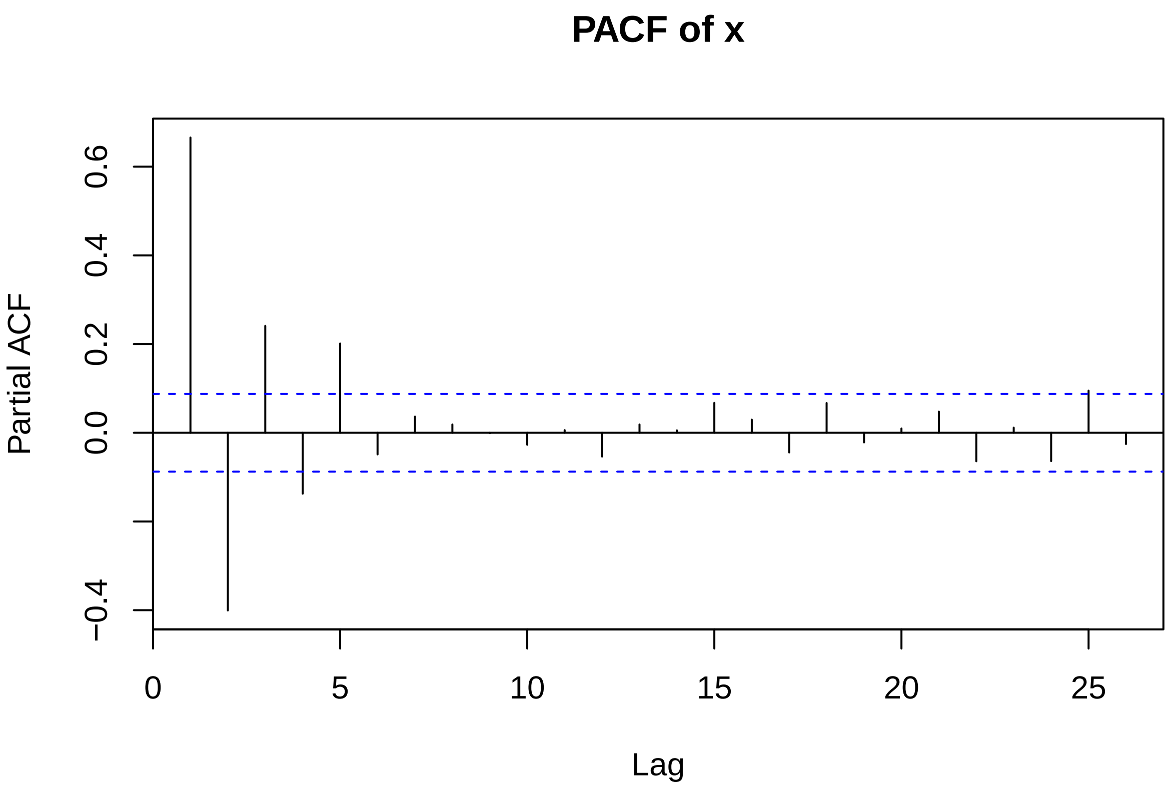
The Q statistic is significant at every lag until 40, providing strong evidence that `x` is not a white noise series. The ACF values appear to be quite large for the first couple of lags but then quickly decay. The PACF is the same as the ACF for the first lag then also rapidly decays. Based on this the series appears to be AR(1).

1.3. Plot ACF and PACF.

```
acf(x, main='ACF of x')
```




```r
pacf(x, main='PACF of x')
```



The graphical versions of the ACF and PACF provide a clearer picture of the results by allowing us to see which autocorrelations are statistically significant. In particular the relatively large a significant negative partial autocorrelation at lag 2 suggests that `x` may in fact be an MA(1) process. The second lag in the ACF plot is also larger than one might expect if we were looking at an AR(1) series.

1.4. Diagnosis of time series based on results

Based on these results I am still uncertain as to whether the series is AR(1) or MA(1). Since it shows characteristics of both AR(1) and MA(1) series it is possible that it is an ARIMA(1,0,1) series.

1.5. ADF and KPSS tests

```r
summary(ur.df(x, lags=1))
```

```
##
## ###############################################
## # Augmented Dickey-Fuller Test Unit Root Test #
## ###############################################
##
## Test regression none
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -3.498 -0.752 -0.037  0.596  3.381
##
## Coefficients:
##            Estimate Std. Error t value Pr(>|t|)
## z.lag.1     -0.4616     0.0335  -13.79   <2e-16 ***
## z.diff.lag   0.3972     0.0412    9.64   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.04 on 496 degrees of freedom
## Multiple R-squared:  0.297,  Adjusted R-squared:  0.294
## F-statistic:  105 on 2 and 496 DF,  p-value: <2e-16
##
##
## Value of test-statistic is: -13.79
##
## Critical values for test statistics:
##       1pct  5pct 10pct
## tau1 -2.58 -1.95 -1.62
```

```r
summary(ur.kpss(x))
```

```
##
## #######################
## # KPSS Unit Root Test #
## #######################
##
## Test is of type: mu with 5 lags.
##
```

```
## Value of test-statistic is: 0.1505
##
## Critical value for a significance level of:
##                 10pct  5pct 2.5pct  1pct
## critical values 0.347 0.463  0.574 0.739
```

The ADF test produces a test-statistic of -13.79. Based on this we can reject the null hypothesis at the $p < 0.01$ level since the critical value $\tau = -2.58$. This means we can reject the null hypothesis that a unit root is present in `x`.

For the KPSS test the absence of a unit root is the null hypothesis. In this case the test-statistic is below the critical value, even at the $p < 0.10$ level, so we cannot reject the null hypothesis that there is a unit root.

In sum, both of these tests are in agreement that there is not a unit root in series `x`. This allows us to restrict the order of integration in the ARIMA models to 0.

1.6. ARIMA

```r
m1 <- forecast::Arima(x, order = c(1,0,0))
m2 <- forecast::Arima(x, order = c(0,0,1))
m3 <- forecast::Arima(x, order = c(1,0,1))
print(m1)
```

```
## Series: x
## ARIMA(1,0,0) with non-zero mean
##
## Coefficients:
##         ar1    mean
##       0.665  -0.153
## s.e.  0.033   0.150
##
## sigma^2 estimated as 1.28:  log likelihood=-770.1
## AIC=1546   AICc=1546   BIC=1559
```

```r
print(Box.test(m1$residuals, lag=40, type="Ljung-Box"))
```

```
##
##  Box-Ljung test
##
## data:  m1$residuals
## X-squared = 120, df = 40, p-value = 3e-10
```

```r
print(m2)
```

```
## Series: x
## ARIMA(0,0,1) with non-zero mean
##
## Coefficients:
##         ma1    mean
##       0.845  -0.160
## s.e.  0.022   0.085
##
## sigma^2 estimated as 1.07:  log likelihood=-725.5
## AIC=1457   AICc=1457   BIC=1470
```

```r
print(Box.test(m2$residuals, lag=40, type="Ljung-Box"))
```

```
##
##  Box-Ljung test
##
## data:  m2$residuals
## X-squared = 88, df = 40, p-value = 2e-05
```

```r
print(m3)
```

```
## Series: x
## ARIMA(1,0,1) with non-zero mean
##
## Coefficients:
##         ar1    ma1    mean
##       0.372  0.721  -0.159
## s.e.  0.048  0.035   0.120
##
## sigma^2 estimated as 0.963:  log likelihood=-699.2
## AIC=1406   AICc=1406   BIC=1423
```

```r
print(Box.test(m3$residuals, lag=40, type="Ljung-Box"))
```

```
##
##  Box-Ljung test
##
## data:  m3$residuals
## X-squared = 33, df = 40, p-value = 0.8
```

```r
which.min(c(m1$aic, m2$aic, m3$aic))
```

```
## [1] 3
```

```r
which.min(c(m1$bic, m2$bic, m3$bic))
```

```
## [1] 3
```

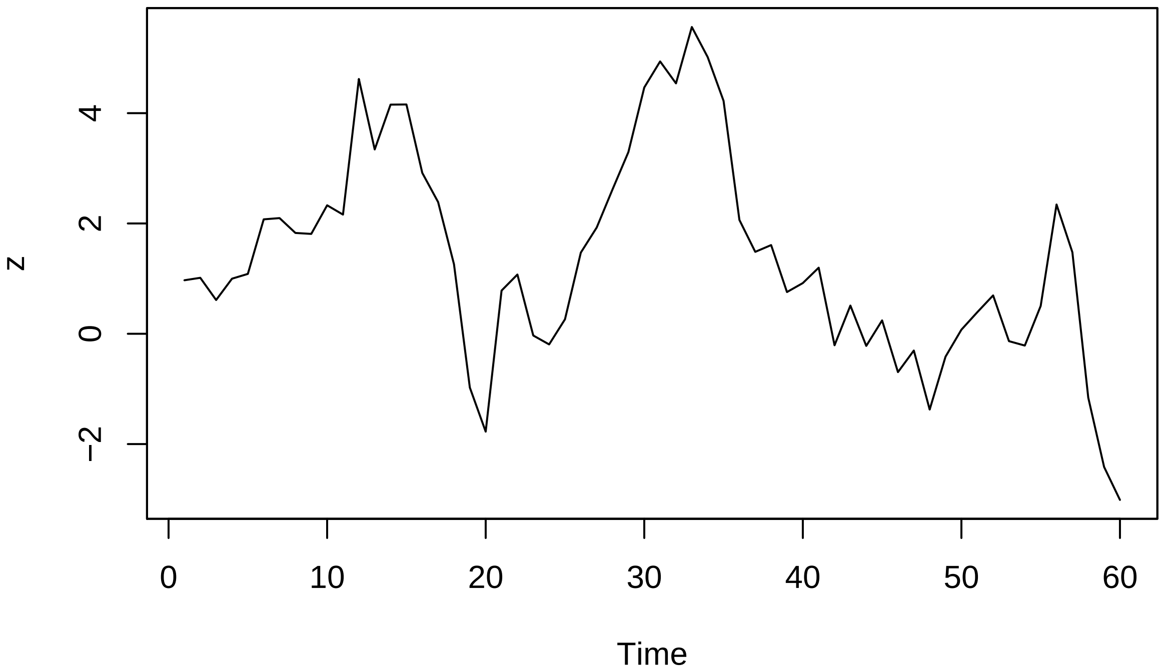
The ARIMA(1,0,1) model performers best in terms of both AIC and BIC, suggesting that it is an improvement over either the AR(1) or MA(1) models. Moreover, the Ljung-Box test Q statistic is not statistically significant, unlike the other two models. This means that the residuals can be considered white noise. In both the AR(1) and MA(1) models the Q statistic is statistically significant, suggesting that both models poorly fit the data.

It was unclear exactly what this process was based on inspection of the ACF and PACF plots, but after testing different model specifications I conclude that series `x` is likely to be an ARIMA(1,0,1) process.

Analyzing series z

I have omitted the code from the analysis of series z to avoid repetition.

2.1. Plot time series z



This series appears to show evidence of a unit root. We can see that there is not a constant mean and that that variance changes over time, for example between around 0 to 10 and 35 to 50 we see very low variance, while the variance is very high in other places, with large peaks and troughs. This suggests that shocks to the series persist for a long time. Based on this the series appears to be nonstationary. However, given that we only observe 60 points in this realization of the series it is possible, however, that the series is stationary and that it simply looks cointegrated due to the short time frame.

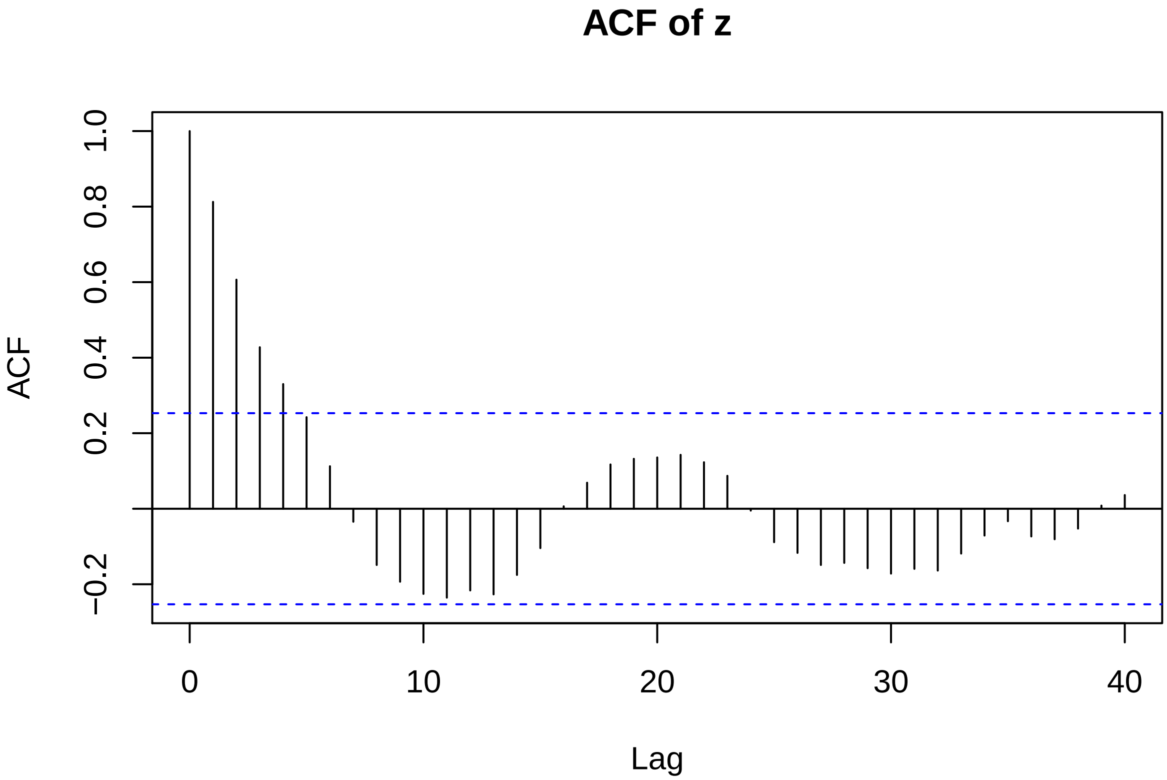
2.2. Show correlogram

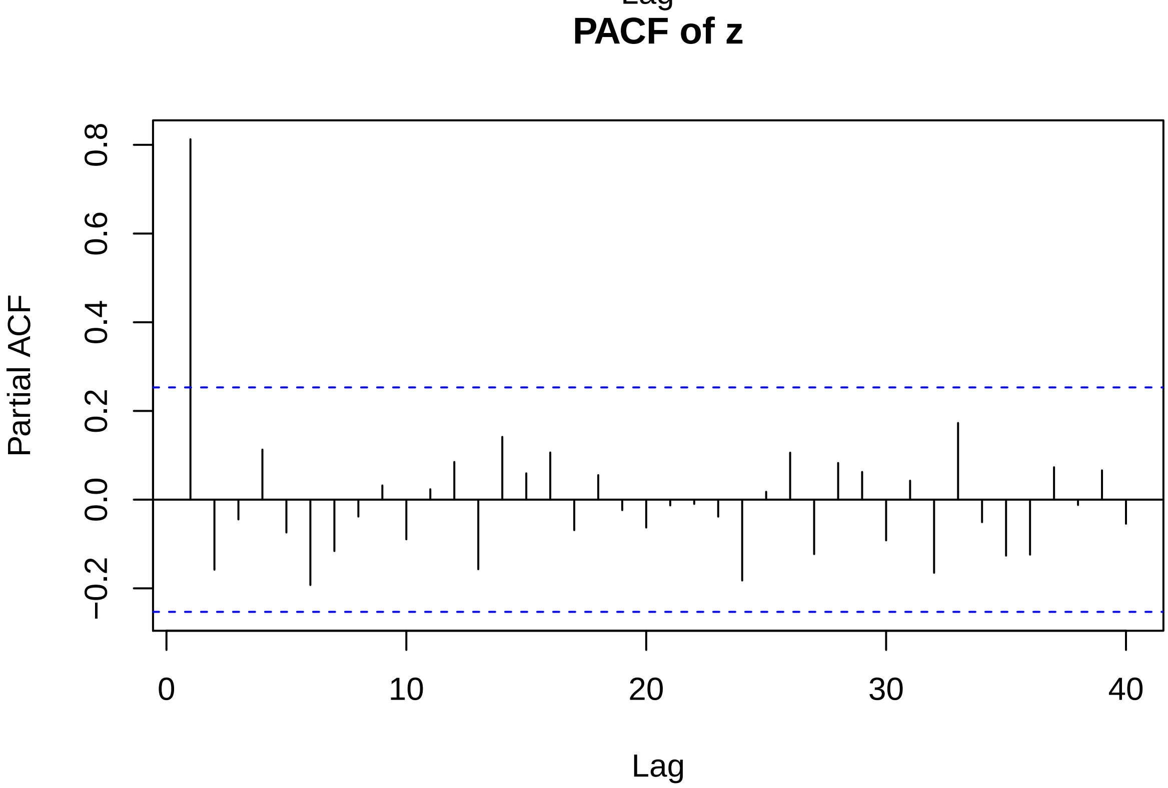
LAG	ACF	PAC	Q	Prob>Q
1	0.8126	0.8126	41.64	0
2	0.6067	-0.1581	65.24	0
3	0.4277	-0.0446	77.18	0
4	0.3301	0.1128	84.42	0
5	0.2424	-0.0741	88.40	0
6	0.1126	-0.1926	89.27	0
7	-0.0345	-0.1159	89.35	0
8	-0.1490	-0.0382	90.94	0
9	-0.1933	0.0320	93.66	0
10	-0.2256	-0.0894	97.45	0
11	-0.2355	0.0234	101.66	0
12	-0.2164	0.0849	105.29	0
13	-0.2268	-0.1568	109.36	0
14	-0.1753	0.1415	111.85	0
15	-0.1043	0.0593	112.75	0
16	0.0064	0.1063	112.75	0
17	0.0690	-0.0688	113.16	0
18	0.1173	0.0553	114.38	0
19	0.1322	-0.0236	115.97	0

LAG	ACF	PAC	Q	Prob>Q
20	0.1359	-0.0629	117.68	0
21	0.1428	-0.0130	119.63	0
22	0.1232	-0.0097	121.11	0
23	0.0872	-0.0384	121.88	0
24	-0.0050	-0.1823	121.88	0
25	-0.0886	0.0178	122.72	0
26	-0.1170	0.1060	124.21	0
27	-0.1489	-0.1228	126.71	0
28	-0.1435	0.0827	129.11	0
29	-0.1575	0.0624	132.08	0
30	-0.1718	-0.0919	135.75	0
31	-0.1592	0.0428	139.00	0
32	-0.1638	-0.1650	142.56	0
33	-0.1188	0.1729	144.51	0
34	-0.0711	-0.0508	145.23	0
35	-0.0331	-0.1263	145.39	0
36	-0.0734	-0.1241	146.23	0
37	-0.0808	0.0731	147.28	0
38	-0.0527	-0.0121	147.75	0
39	0.0083	0.0660	147.77	0
40	0.0363	-0.0543	148.01	0

The significant Q statistics allow us to reject the hypothesis that the series is white noise, although this was apparent from the plot above. The ACF values in particular seem to remain relatively large and decline somewhat linearly, a tell-tale sign of a unit root. The PACF shows a large correlation for the first lag and then low values, also typical of a unit root but also of an AR(1) process.

2.3. Plot ACF and PACF.





The ACF clearly shows that the autocorrelations decay almost linearly, indicating that the series has a unit root. The PACF results were clear just from the table above.

2.4. Diagnosis of time series based on results

Based on these results I expect that the series has a unit root. The PACF more closely resembles an AR than an MA process. I expect that the series may be an ARIMA(1,1,0) series.

2.5. ADF and KPSS tests

```
##
## ###############################################
## # Augmented Dickey-Fuller Test Unit Root Test #
## ###############################################
##
## Test regression none
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.3447 -0.6909  0.0961  0.6680  2.7044
##
## Coefficients:
##            Estimate Std. Error t value Pr(>|t|)
## z.lag.1     -0.1006     0.0585   -1.72    0.091 .
## z.diff.lag   0.1642     0.1336    1.23    0.224
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1 on 56 degrees of freedom
## Multiple R-squared:  0.0615, Adjusted R-squared:  0.0279
## F-statistic: 1.83 on 2 and 56 DF,  p-value: 0.169
##
##
## Value of test-statistic is: -1.718
##
## Critical values for test statistics:
##      1pct  5pct 10pct
## tau1 -2.6 -1.95 -1.61
```

```
##
## #######################
## # KPSS Unit Root Test #
## #######################
##
## Test is of type: mu with 3 lags.
##
## Value of test-statistic is: 0.4093
##
## Critical value for a significance level of:
##                 10pct  5pct 2.5pct  1pct
## critical values 0.347 0.463  0.574 0.739
```

Starting the the ADF test we see that we cannot reject the null hypothesis that there is a unit root at the conventional $p < 0.05$ level, although we can if we accept $p < 0.10$. Similarly, for the KPSS test we can reject the null hypothesis that there is not a unit root at the $p < 0.10$ level. Both tests provide some weak support the hypothesis that there is a unit root in `z`. The low statistical power of the tests is likely due to the fact that we only observe 60 points in the series.

2.6. ARIMA

```
## [1] "ARIMA(1,1,0):"
```

```
## Series: z
## ARIMA(1,1,0)
##
## Coefficients:
##         ar1
##       0.108
## s.e.  0.129
##
## sigma^2 estimated as 1.02:  log likelihood=-83.87
## AIC=171.7   AICc=171.9   BIC=175.9
```

```
##
##  Box-Ljung test
##
## data:  m1$residuals
## X-squared = 42, df = 40, p-value = 0.4
```

The Box-Ljung test shows that residuals of the series are white noise after fitting the ARIMA(1,1,0) model. This suggests that the model including the AR(1) component and an order of integration equal to 1 fits the data well.

While I know I was only supposed to present one model here, I was curious as to whether an ARIMA(0,1,1) model would fit better. The output below shows this model. We can see that the AIC and BIC and the p-value for the Box-Ljung test are almost identical.

```
## [1] "ARIMA(0,1,1):"
```

```
## Series: z
## ARIMA(0,1,1)
##
## Coefficients:
##         ma1
##       0.107
## s.e.  0.125
##
## sigma^2 estimated as 1.02:  log likelihood=-83.86
## AIC=171.7   AICc=171.9   BIC=175.9
```

```
##
##  Box-Ljung test
##
## data:  m2$residuals
## X-squared = 42, df = 40, p-value = 0.4
```