

Chris Montgomery TS Assignment 2

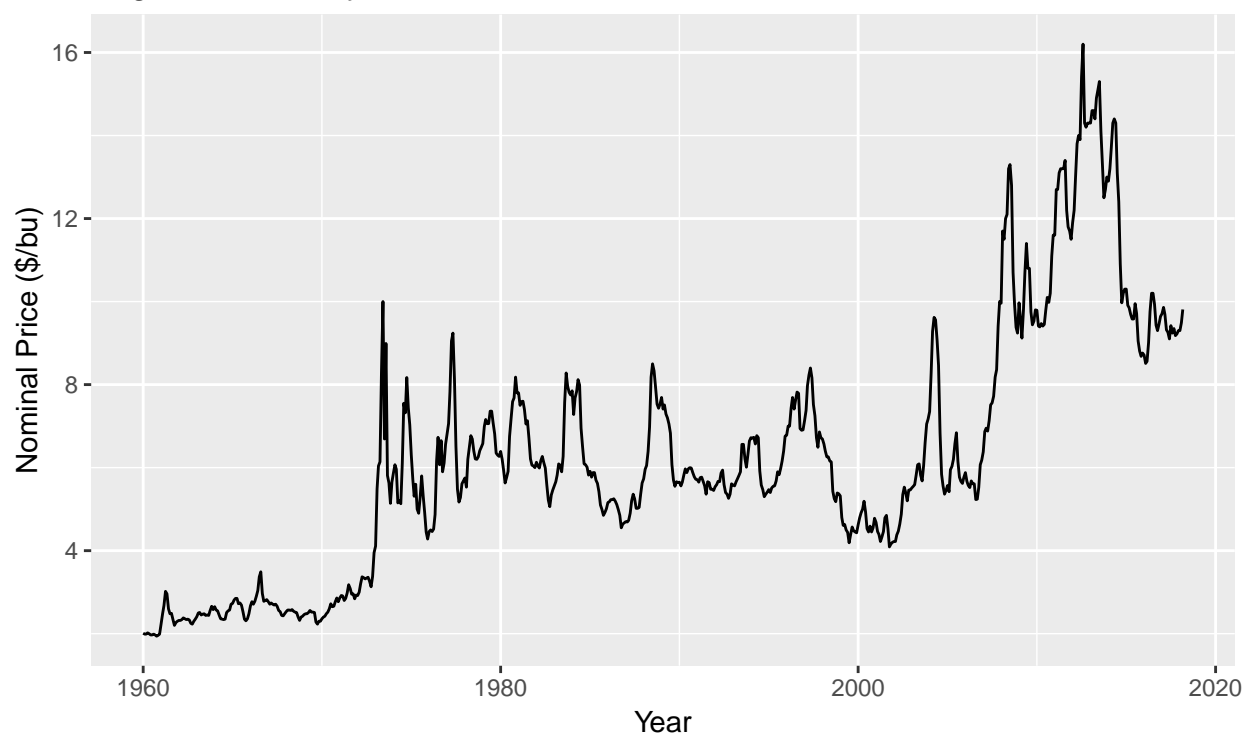
Introduction

For the purposes of part one of this assignment, the following section will analyze “soybean prices received” data. This data, collected by the USDA National Agriculture Statistics Service, depicts the average monthly price received per bushel by domestic soybean farmers. Price averages are collected by the monthly Agriculture Resource Management Survey (ARMS). This time series contains 699 observations covering the period beginning January, 1960 to March 2018. Prices are not adjusted for inflation.

Below is a plot of the time series. Upon visual inspection, it would appear that a positive drift parameter is present in the series. Conceptually this makes sense, as one would typically expect general inflation to drive prices up over time. Additional tests could more formally elucidate the presence of a drift parameter.

```
#plot (ts)
library(ggplot2)
library(ggfortify)
p <- autoplot(ts, xlab = "Year", ylab = "Nominal Price ($/bu)", main = " Figure 1: US Soybean Producer Prices Received")
p + labs(caption = "Source: USDA National Agriculture Statistics Service \n https://www.nass.usda.gov/Charts_and_Maps/Agricultural_Prices/pricesb.php")
```

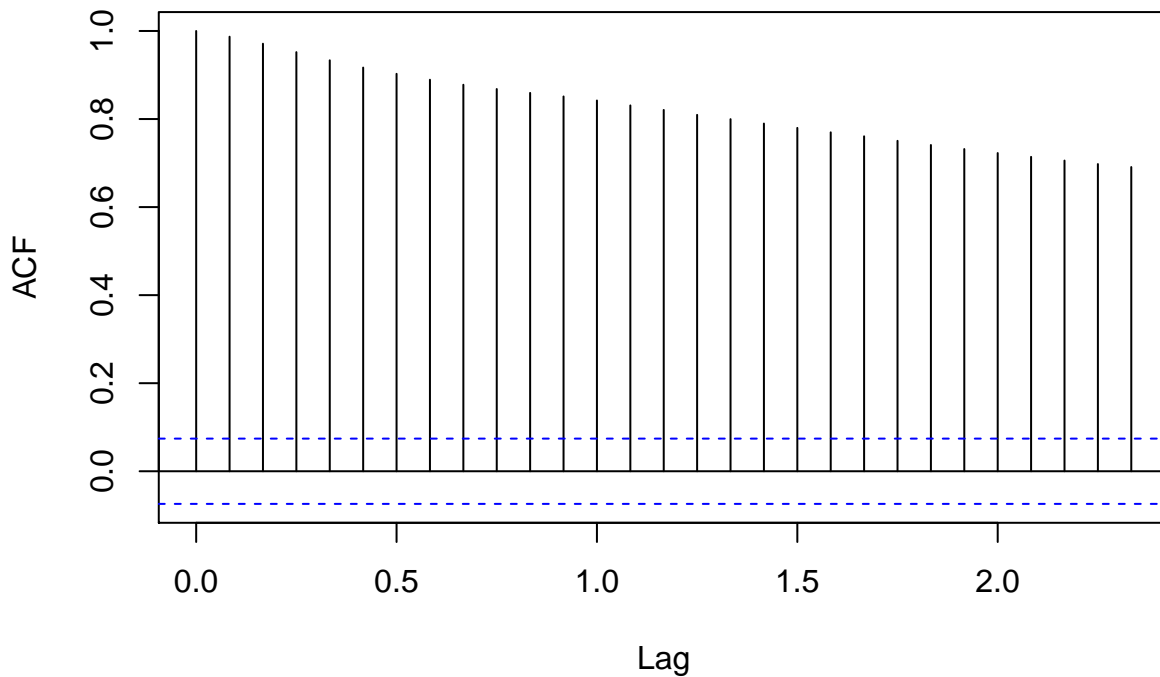
Figure 1: US Soybean Producer Prices Received



Source: USDA National Agriculture Statistics Service
https://www.nass.usda.gov/Charts_and_Maps/Agricultural_Prices/pricesb.php

Visual inspection of the correlogram shows some decay in correlation coefficients over time. This could be suggestive of the presence of a near-unit root. This plot could suggest that the time series does not follow a random walk.

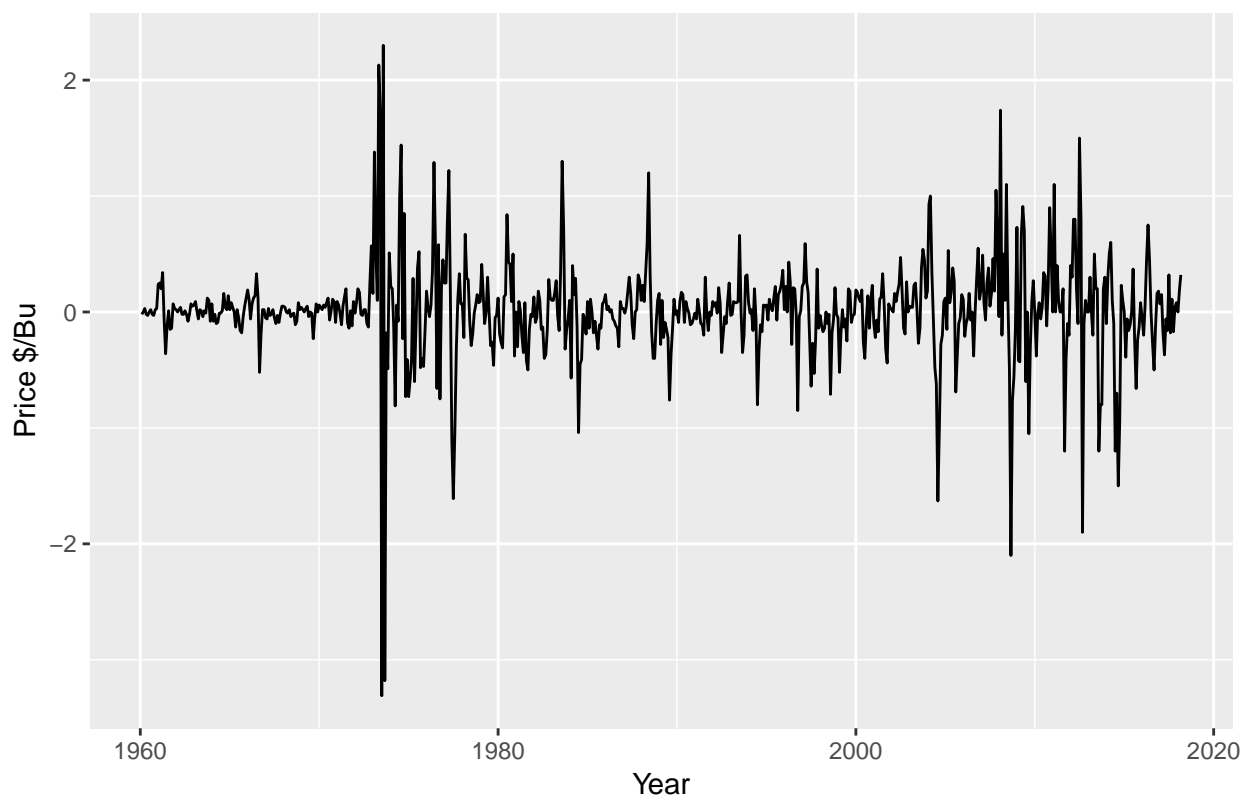
```
acf <- acf(ts, main = "")
```



Below is a plot of the first difference of the time series as well as its acf. The ACF plot shows that the first and second lags of the first-differenced time series are positive and statistically significantly correlated with the first difference. This suggests that taking the first difference failed to remove all autocorrelation in the series. In addition, the spike at the 12th lag appears to be significant at the 95% confidence level. This suggests that some seasonality may exist in the data

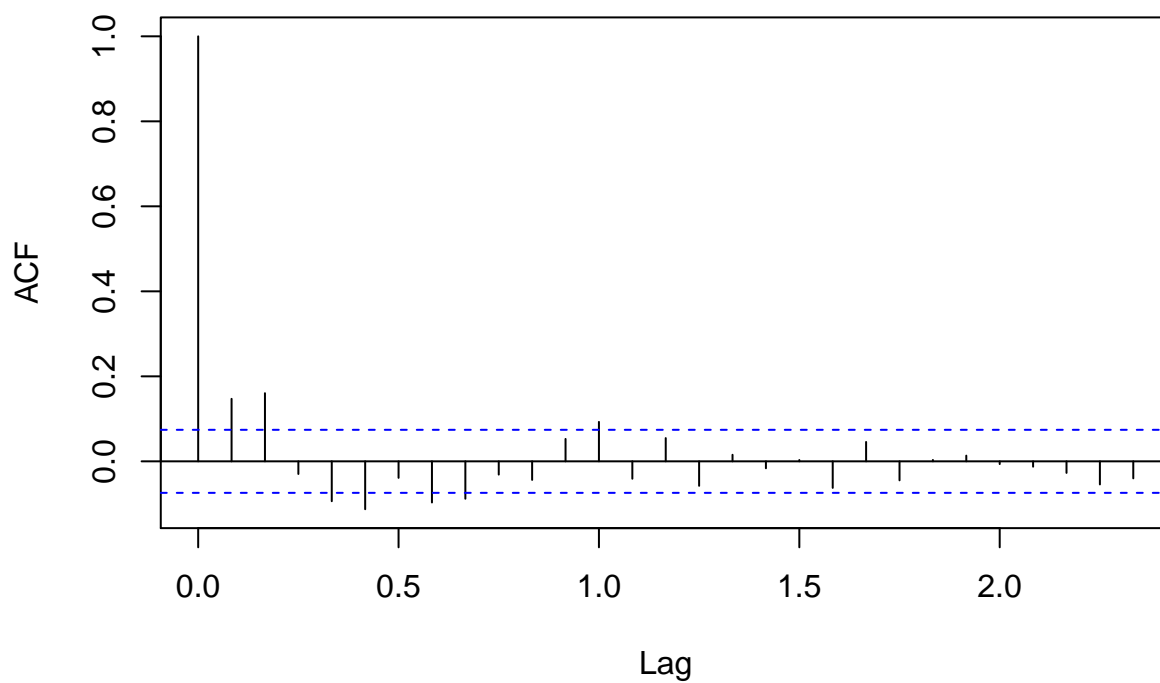
```
autoplot(diff(ts), ylab = "Price $/Bu", xlab = "Year", main = "First Difference, Prices Received")
```

First Difference, Prices Received



```
acf(diff(ts), main = "ACF of First Difference")
```

ACF of First Difference



Comparing the below descriptive statistics for the entire time series with a subset containing observations

between 2000 and 2018 indicate non-stationarity. Here, the mean of the post 2000 subset is \$2.60/bushel higher than the mean of the entire series. Again, this is fairly intuitive as we would expect to see a positive drift parameter associated with general inflation over time.

```
library("tseries")
summary(ts)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1.940   4.455   5.780   6.219   7.530   16.200
```

```
summary(window(ts, start = c(2000, 1)))
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    4.090   5.680   9.300   8.883  10.850   16.200
```

Testing a Random Walk Hypothesis

Looking at the mean of our first difference does not suggest the presence of a drift parameter. The mean is not significantly different from zero. However, the mean of the differenced series is the correct sign for the presence of a positive drift parameter

```
diff_ts <- diff(ts)
mean(diff_ts)
```

```
## [1] 0.01117479
```

```
mean(diff_ts)+c(-1.96,1.96)*sd(diff_ts)/sqrt(length(diff_ts))
```

```
## [1] -0.02079911  0.04314868
```

Fitting a random walk model suggests that the drift parameter (intercept term) is positive and statistically significant at the 95% confidence level. At the same time, the coefficient of the first lag is statistically different from one at the 95% confidence level. This supports the presence of a near unit root, as opposed to a random walk.

```
#Fitting a Random Walk model
library(dyn)
```

```
## Loading required package: zoo
```

```
##
```

```
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
##      as.Date, as.Date.numeric
```

```
reg <- dyn$lm(ts ~ lag(ts,-1))
```

```
confint(reg, level = 0.95)
```

```
##              2.5 %    97.5 %
## (Intercept) 0.004784613 0.1529455
## lag(ts, -1) 0.978351194 0.9998617
```

Subsetting the series beginning with observations in 2000 changes the significance levels of both the lag coefficient and drift parameter. In this case, the lag coefficient is no longer significant from one, while the drift parameter is no longer significantly different from zero. This could be explained in at least in part due to

increases in the standard errors of the regression coefficients from running the regression against a smaller sample.

```
ts_2k <- window (ts, start= c(2000,1))
reg <- dyn$lm(ts_2k ~ lag(ts_2k,-1))
summary(reg)

##
## Call:
## lm(formula = dyn(ts_2k ~ lag(ts_2k, -1)))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.06243 -0.17935  0.01603  0.21012  1.73312
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.16278    0.09772   1.666  0.0972 .
## lag(ts_2k, -1)  0.98435    0.01036  95.011  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4869 on 216 degrees of freedom
## (2 observations deleted due to missingness)
## Multiple R-squared:  0.9766, Adjusted R-squared:  0.9765
## F-statistic: 9027 on 1 and 216 DF, p-value: < 2.2e-16
confint(reg, level = 0.95)

##              2.5 %    97.5 %
## (Intercept)  -0.02982765 0.3553899
## lag(ts_2k, -1)  0.96392726 1.0047681
```

Applying an Augmented Dickey Fuller test fails to reject the null hypothesis of non-stationarity. Thus, according to the ADF test we also cannot reject the possibility of a unit root being present.

```
library(tseries)
adf.test(ts)

##
## Augmented Dickey-Fuller Test
##
## data:  ts
## Dickey-Fuller = -2.8691, Lag order = 8, p-value = 0.2104
## alternative hypothesis: stationary
adf.test(ts, k = 0)

##
## Augmented Dickey-Fuller Test
##
## data:  ts
## Dickey-Fuller = -3.03, Lag order = 0, p-value = 0.1423
## alternative hypothesis: stationary
```

Part II

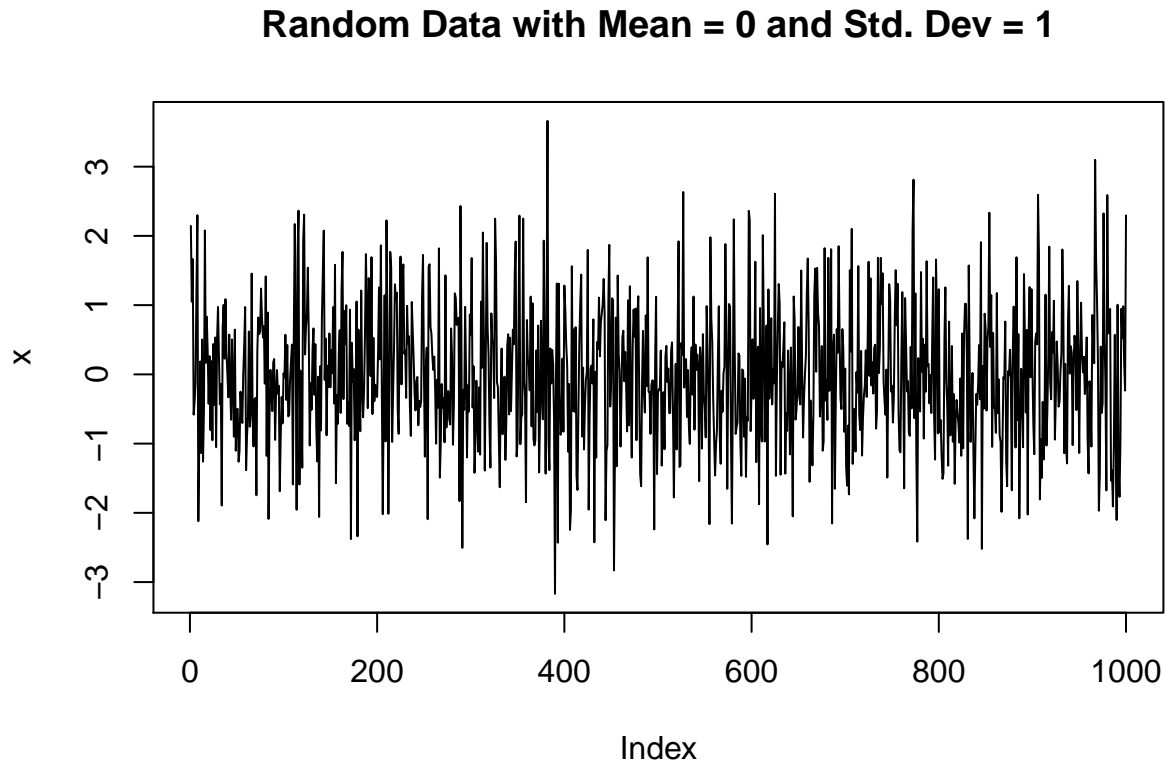
The below code creates a random walk series of 1000 observations. Knowing the data generation process

instills confidence that the series is stationary with a mean zero.

```
x <- rnorm(1000)
layout(1:1)
mean(x)
```

```
## [1] -0.009998258
```

```
plot(x,type="l", main = "Random Data with Mean = 0 and Std. Dev = 1")
```



Fitting an ARIMA (1,1,1) to the original random series proves to be analogous to fitting an ARMA (1,1) to the first difference of the series. This should be fairly intuitive, as the integrated term in an arima model specifies a difference transformation. In this case, the ARIMA (1,1,1) model specifies an ARMA (1,1) applied to the first difference of the time series. However, both AR and MA coefficients very slightly. This can be explained by the presence of an intercept term in the ARMA (1,1) model. By default, the ARIMA (1,1,1) model does not include a drift parameter.

*#Coefficients and standard errors are pretty darn close. But they're still off a bit.
#Weird...*

```
arima <- arima(x, order = c(1,1,1))
arma <- arima(diff(x), order = c(1,0,1))
arima$coef
```

```
##          ar1          ma1
## 0.03218244 -0.99999950
```

```
arma$coef
```

```
##          ar1          ma1      intercept
## 0.0312748229 -0.9999997088 -0.0001061521
```

It's important to note that the intercept term represents a drift parameter. As such, its presence (or absence) has significant implications for the fit of a model. Thus, one should always consider whether a model is better

fit by the presence of a drift parameter, and if so, write pick the right model in R.

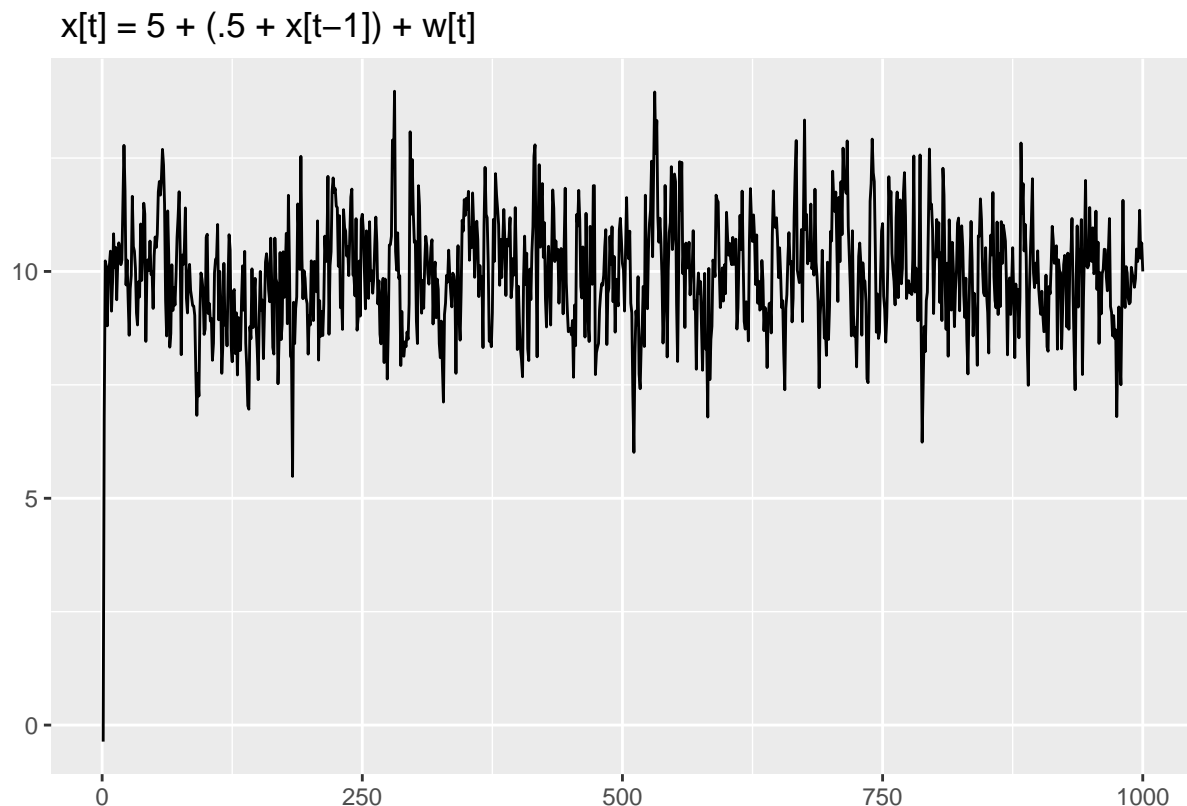
Part III $X_t = 5 + .5X_{t-1} + W_t$

The above equation represents a mean stationary time series. Since the coefficient of the lagged term is less than one, the series converges to a stationary mean as time progresses. A simulation shows that as t approaches infinity, the mean of the time series produced by the above equation approaches 10.

```
x <- w <- rnorm(1000)
for (t in 2:1000) x[t] <- 5 + (.5 * x[t-1]) + w[t]

x <- ts(x)
adf.test(x)

## Warning in adf.test(x): p-value smaller than printed p-value
##
## Augmented Dickey-Fuller Test
##
## data: x
## Dickey-Fuller = -8.7893, Lag order = 9, p-value = 0.01
## alternative hypothesis: stationary
autoplot(x, type="l", main = " x[t] = 5 + (.5 + x[t-1]) + w[t] ")
```

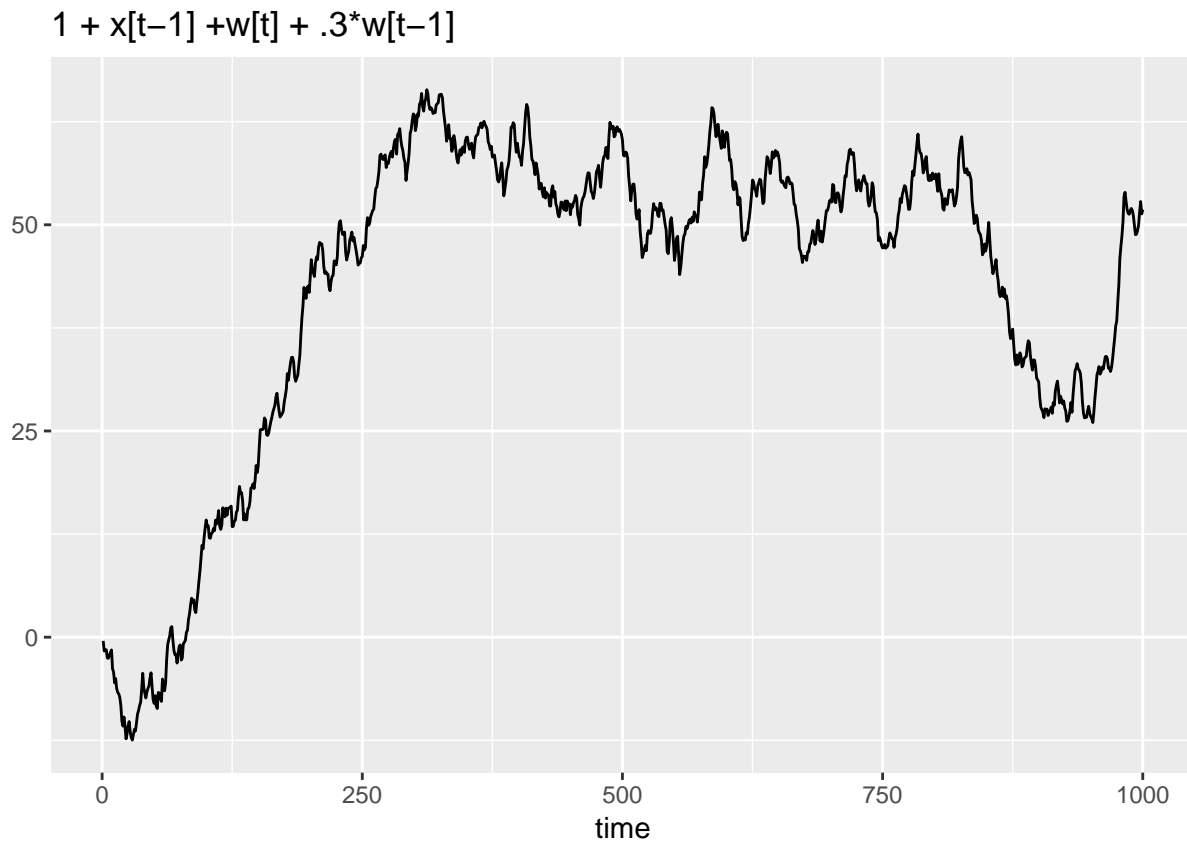


$X_t = 1 + X_{t-1} + W_t + .3W_t$

The above equation is not mean stationary. In this case, the alpha parameter value of one and the presence of a positive drift parameter follows a random walk with drift. Simulating the above equation validates the non-stationarity and positive drift of the time series.

```
#To better demonstrate the random walk, the drift parameter was reduced by a factor of 10. This produces
x <- w <- rnorm(1000)
for (t in 2:1000) x[t] <- .1 + x[t-1] +w[t] + .3*w[t-1]

x <- ts(x)
autoplot(x, type="l", main = "1 + x[t-1] +w[t] + .3*w[t-1]", xlab = "time")
```



```
adf.test(x)
```

```
##
## Augmented Dickey-Fuller Test
##
## data: x
## Dickey-Fuller = -1.976, Lag order = 9, p-value = 0.5885
## alternative hypothesis: stationary
```

One rather simple way to transform the above time series involves taking its first difference. In this case, the transformed equation can be shown as follows:

$$x[t] - x[t-1] = 1 + w[t] + .3w[t-1]$$

By taking the first difference time series now takes on a conditional mean of 1 (the new intercept) with each observation being a function of the current and previous time period's white noise term. The presence of a MA(1) term does not affect the stationarity of the series, as the conditional expectation of each white noise term is zero. The below first difference plot and ADF test support this explanation.

```
diff_x <- diff(x)
autoplot(diff_x, type = 'l', main = "Plot of First Difference")
```


Plot of First Difference

