MATH 474 - Time Series

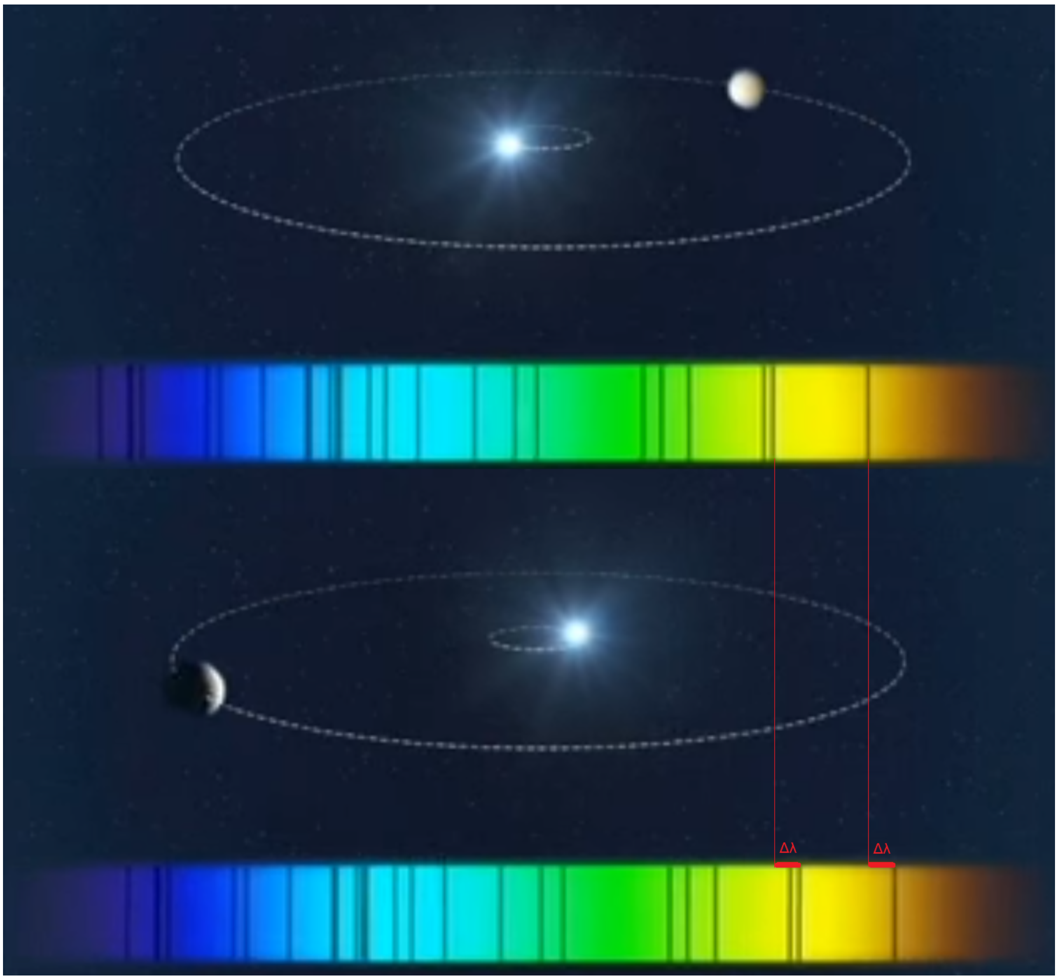
An Exercise in Time Series Analysis: Radial-Velocity Exoplanet Detection applied to HD 75767

Bryan Arnold

Fall 2015

Contents

1 Introduction



1.1 Remarks on Doppler Spectography

Doppler spectroscopy (also known as the radial-velocity method, or colloquially, the wobble method) is an indirect method for finding extrasolar planets (exoplanets) and brown dwarf stars from radial-velocity measurements via observation of Doppler shifts in the spectrum of the planet's parent star.

A star with an orbiting planet will move in its own small orbit in response to the planet's gravity. This leads to variations in the speed with which the star moves toward or away from Earth, i.e. the variations are in the radial velocity of the star with respect to Earth. The radial velocity can be deduced from the displacement in the parent star's spectral lines due to the Doppler effect.

A spectrograph, like a prism, splits the light from the star into its component colors producing a spectrum. Some of the starlight gets absorbed as it passes through the star's atmosphere, and this produces small, dark gaps or lines in the spectrum (spectral lines). As the star moves closer to us, these spectral lines shift toward the blue-end of the spectrum (blue-shift). As the star moves away, the lines shift back toward the red-end of the spectrum (red-shift). Astronomers can detect orbiting planets by looking for these back-and-forth motions of the lines in a star's spectrum. As of October 2012 about half of the extrasolar planets known were discovered using Doppler spectroscopy.

1.2 Bellerophon

51 Pegasi b (abbreviated 51 Peg b), sometimes unofficially named Bellerophon, is an extrasolar planet approximately 50 light-years away in the constellation of Pegasus. 51 Pegasi b was the first planet to be discovered orbiting a main-sequence star, the Sun-like 51 Pegasi, and marked a breakthrough in astronomical research. (The first exoplanet discovery was made by Aleksander Wolszczan in 1992, around pulsar PSR 1257.) It is the prototype for a class of

planets called hot Jupiters.

They used the radial velocity method with the ELODIE spectrograph on the Observatoire de Haute-Provence telescope in France and making world headlines with their announcement.

1.3 Motivation

Sometimes Doppler spectrography produces false signals which is more common in multi-planet and multi-star systems. Magnetic field and certain types of stellar activity can also give false signals. When the host star has multiple planets, false signals can also arise from having insufficient data where multiple solutions can fit with gathered data as stars are not generally observed continuously. Some of the false signals can be eliminated by analyzing the stability of the planetary system, conducting photometry analysis on the host star and knowing its rotation period and stellar activity cycle periods.

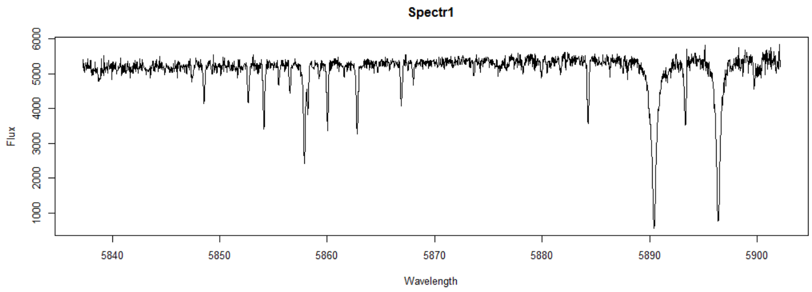
Astronomers can sometimes detect multiple planets via the spectral data by utilizing further statistical analysis and modeling. The idea is simple: More orbiting objects create more fluctuations in the spectral lines (more noise). Modeling allows further analysis to be done to detect these individual contributors. This is the motivation for this project –we would like to practice some of these techniques.

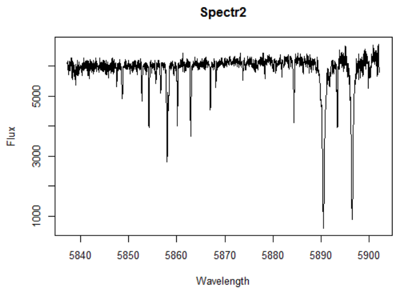
In this project we analyze a known binary star system: HD 75767. Detecting a dwarf star that emits almost no visible light or a gas giant (like Jupiter) is easier than Earth-like planets due to the magnitude of the mass (less noise due to relatively smaller masses in the system). This is why we have chosen this example for our analysis.

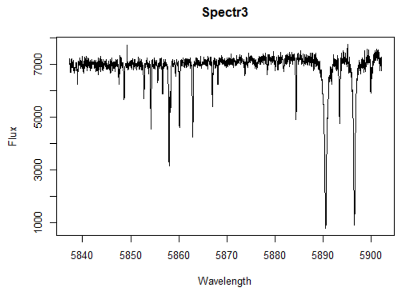
2 Data

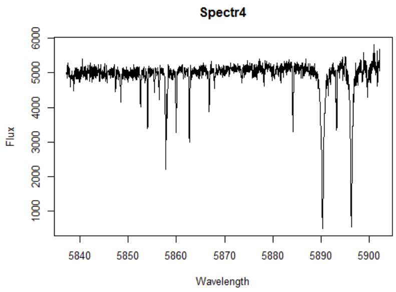
We shall visualize eleven spectra of a binary star on sequential days; we study the part of spectrum around Na (Sodium) lines: we choose Na lines because they are easy to find.

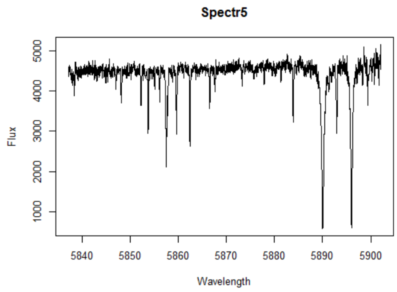
The data was retrieved from the NASA Exoplanet Archive. HD 75767 has been surveyed with the CORALIE spectrograph installed at the Swiss telescope at the ESO-La Silla observatory.

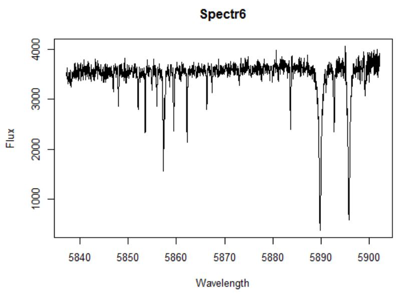

Spectr1
Flux
1000 2000 3000 4000 5000 6000
5840 5850 5860 5870 5880 5890 5900
Wavelength

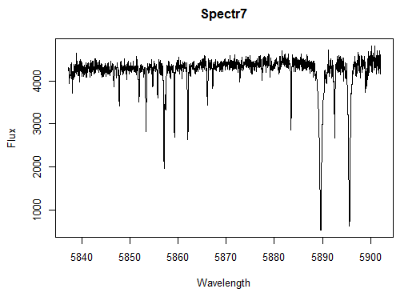

Spectr2
Flux
1000 3000 5000
5840 5850 5860 5870 5880 5890 5900
Wavelength

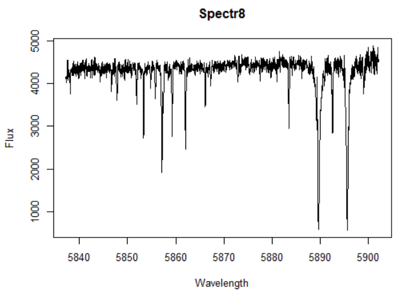

Spectr3
Flux
1000 3000 5000 7000
5840 5850 5860 5870 5880 5890 5900
Wavelength

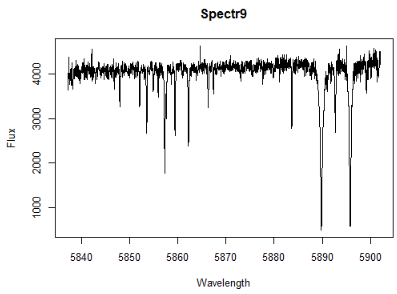

Spectr4
Flux
1000 2000 3000 4000 5000 6000
5840 5850 5860 5870 5880 5890 5900
Wavelength

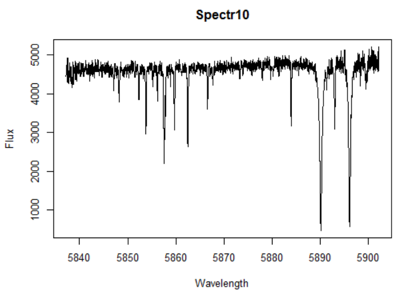

Spectr5
Flux
1000 2000 3000 4000 5000
5840 5850 5860 5870 5880 5890 5900
Wavelength

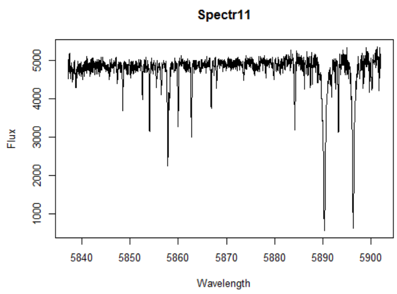

Spectr6
Flux
1000 2000 3000 4000
5840 5850 5860 5870 5880 5890 5900
Wavelength


Spectr7
Flux
1000 2000 3000 4000
5840 5850 5860 5870 5880 5890 5900
Wavelength

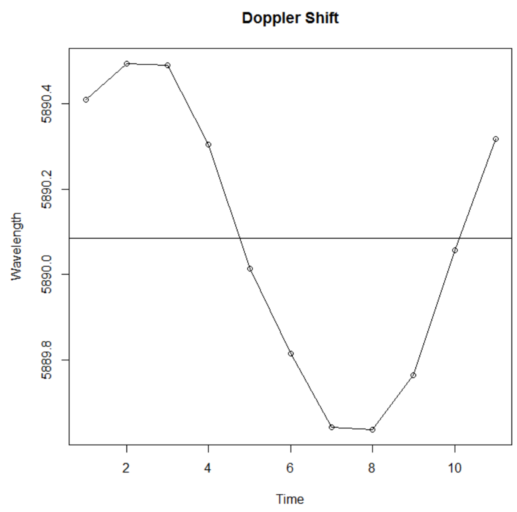

Spectr8
Flux
1000 2000 3000 4000 5000
5840 5850 5860 5870 5880 5890 5900
Wavelength

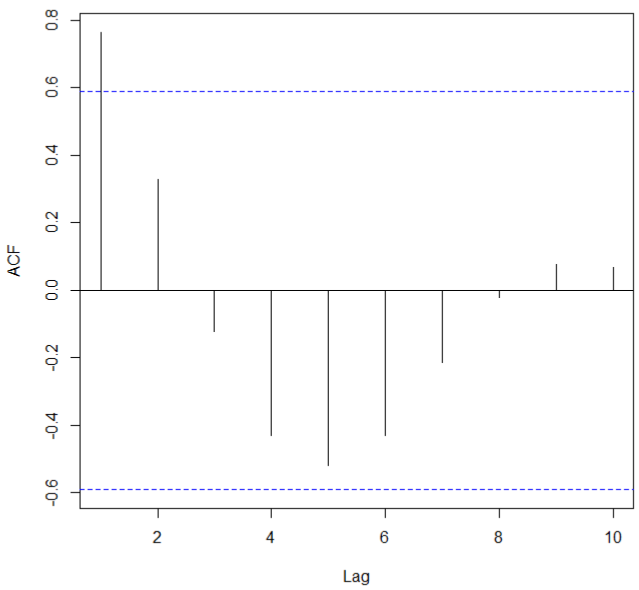

Spectr9
Flux
1000 2000 3000 4000
5840 5850 5860 5870 5880 5890 5900
Wavelength

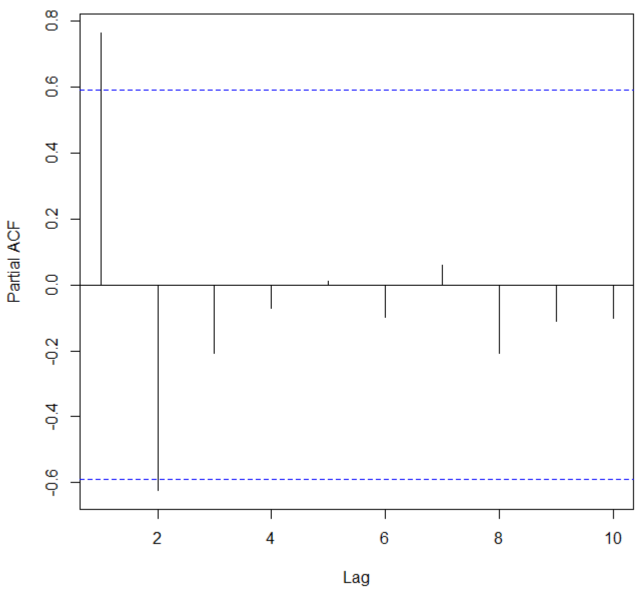

Spectr10
Flux
1000 2000 3000 4000 5000
5840 5850 5860 5870 5880 5890 5900
Wavelength


Spectr11
Flux
1000 2000 3000 4000 5000
5840 5850 5860 5870 5880 5890 5900
Wavelength

2.1 Analysis


Doppler Shift
Wavelength
5889.8
5890.0
5890.2
5890.4
2
4
6
8
10
Time


ACF
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
2
4
6
8
10
Lag


Partial ACF
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
2
4
6
8
10
Lag

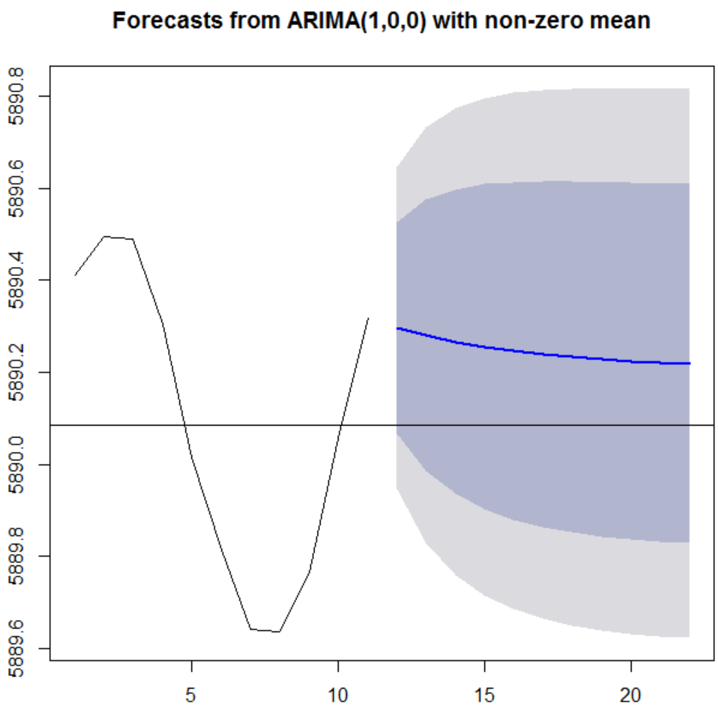
2.2 R-Code

```r
mydata1 = read.table("spectr1.data")
mydata2 = read.table("spectr2.data")
mydata3 = read.table("spectr3.data")
mydata4 = read.table("spectr4.data")
mydata5 = read.table("spectr5.data")
mydata6 = read.table("spectr6.data")
mydata7 = read.table("spectr7.data")
mydata8 = read.table("spectr8.data")
mydata9 = read.table("spectr9.data")
mydata10 = read.table("spectr10.data")
mydata11 = read.table("spectr11.data")

layout(matrix(c(1,1,2,3), 2, 2, byrow = TRUE))
plot(mydata1, type='l',main="Spectr1",xlab='Wavelength',ylab='Flux')
plot(mydata2, type='l',main="Spectr2",xlab='Wavelength',ylab='Flux')
plot(mydata3, type='l',main="Spectr3",xlab='Wavelength',ylab='Flux')

par(mfrow=c(2,2))
plot(mydata4, type='l',main="Spectr4",xlab='Wavelength',ylab='Flux')
plot(mydata5, type='l',main="Spectr5",xlab='Wavelength',ylab='Flux')
plot(mydata6, type='l',main="Spectr6",xlab='Wavelength',ylab='Flux')
plot(mydata7, type='l',main="Spectr7",xlab='Wavelength',ylab='Flux')

par(mfrow=c(2,2))
plot(mydata8, type='l',main="Spectr8",xlab='Wavelength',ylab='Flux')
plot(mydata9, type='l',main="Spectr9",xlab='Wavelength',ylab='Flux')
plot(mydata10, type='l',main="Spectr10",xlab='Wavelength',ylab='Flux')
plot(mydata11, type='l',main="Spectr11",xlab='Wavelength',ylab='Flux')

library(TSA)
library(forecast)

mydata = read.table("Book1.txt")
mydata = mydata[[1]]

plot(mydata, type='o', main='Doppler Shift',xlab='Time',ylab='Wavelength')
mean(mydata)
abline(h=5890.086)

acf(mydata)

pacf(mydata)

mod1 = Arima(mydata, order=c(1,0,0)); mod1
mod2 = Arima(mydata, order=c(0,0,1)); mod2
mod3 = Arima(mydata, order=c(1,0,1)); mod3
mod4 = Arima(mydata, order=c(1,0,2)); mod4

fc.mod1 <- forecast(mod1, h=11)
fc.mod2 <- forecast(mod2, h=11)
fc.mod3 <- forecast(mod3, h=11)
fc.mod4 <- forecast(mod4, h=11)

plot(fc.mod1)
abline(h=5890.086)

plot(fc.mod2)
abline(h=5890.086)

```

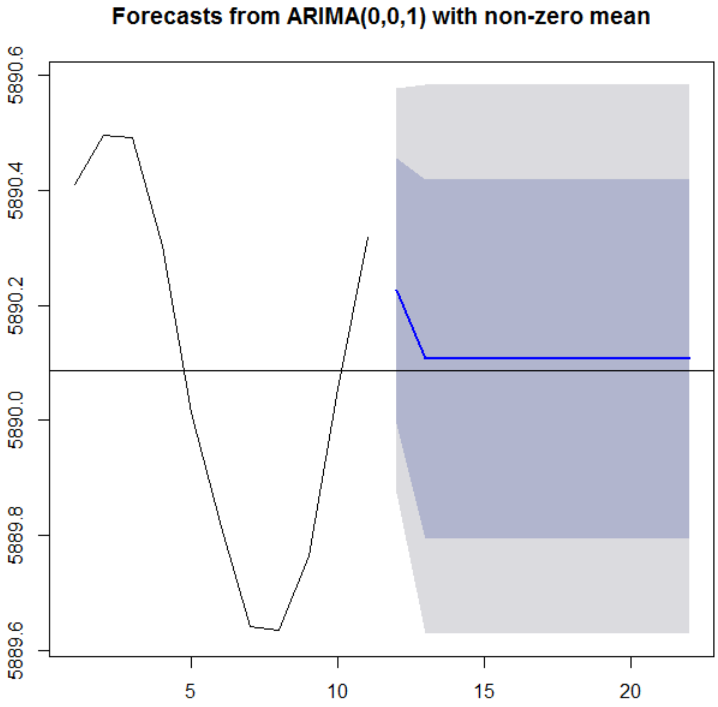
```
plot(fc.mod3)
abline(h=5890.086)

plot(fc.mod4)
abline(h=5890.086)

mod1 = arima(mydata, order=c(1,0,0)); res1=rstandard(mod1); mod1
mod2 = arima(mydata, order=c(0,0,1)); res2=rstandard(mod2); mod2
mod3 = arima(mydata, order=c(1,0,1)); res3=rstandard(mod3); mod3
mod4 = arima(mydata, order=c(1,0,2)); res4=rstandard(mod4); mod4

shapiro.test(residuals(mod1))
shapiro.test(residuals(mod2))
shapiro.test(residuals(mod3))
shapiro.test(residuals(mod4))

mod5=arima(mydata,order=c(0,0,1), seasonal=list(order=c(1,0,0),period=10)); mod5

shapiro.test(residuals(mod5))

mod5=Arima(mydata,order=c(0,0,1), seasonal=list(order=c(1,0,0),period=10));
fc.mod5 <- forecast(mod5, h=11)

plot(fc.mod5)
abline(h=5890.086)
```

3 Modeling

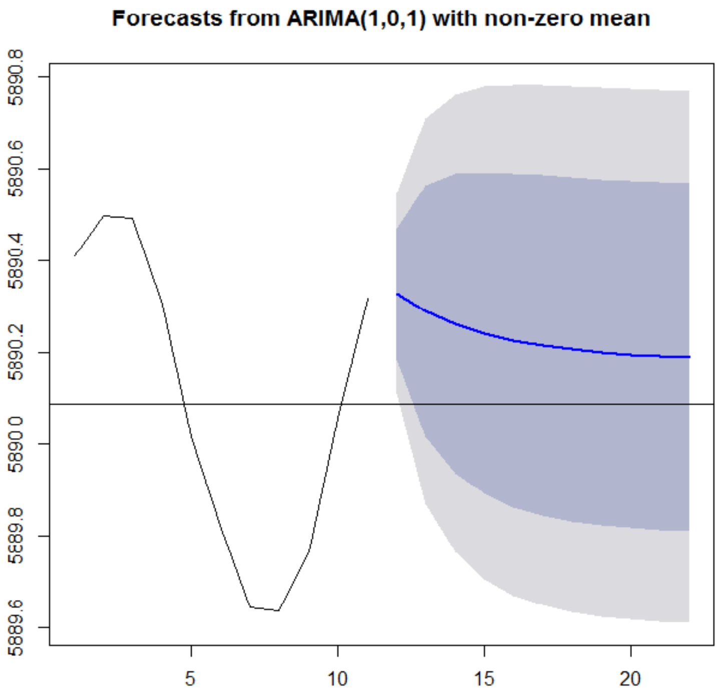
3.1 AR(1)



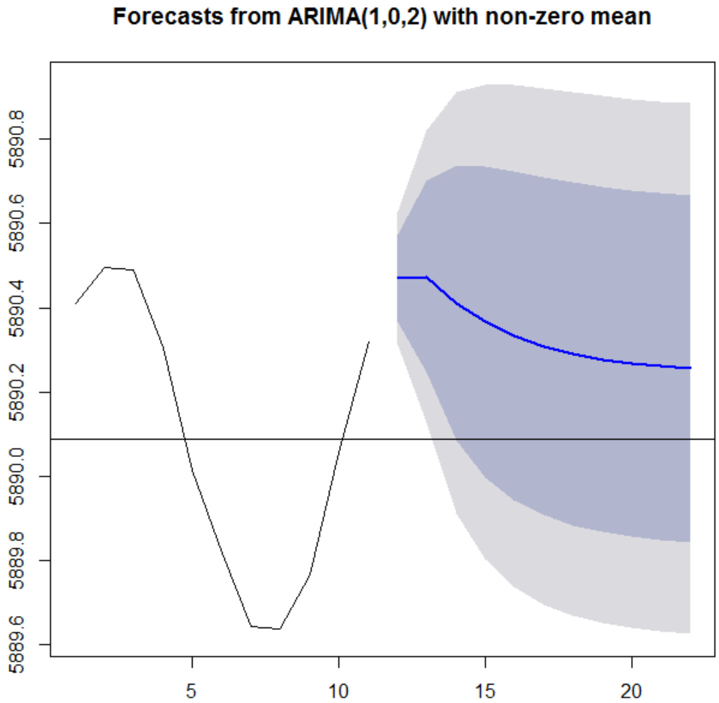
3.2 MA(1)



3.3 ARMA(1, 1)



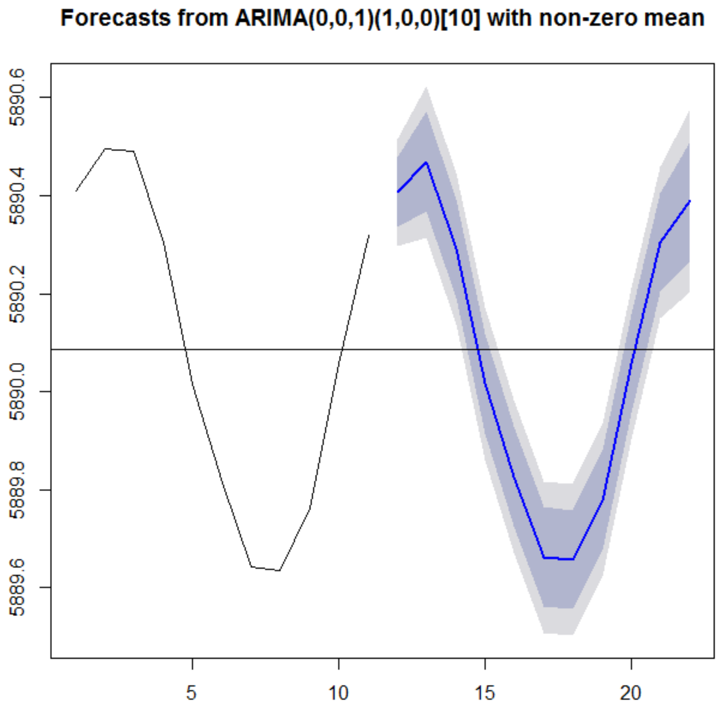
3.4 ARMA$(1, 2)$



3.5 Seasonal ARIMA$(1, 0, 1) \times (1, 0, 0)_{10}$

Seasonal Theoretical Box-Jenkins Model Identification:
STEP 1: Look at the ACF and PACF at the non-seasonal level to identify a non-seasonal model.
STEP 2: Look at the behaviors (spikes) of the ACF and PACF at the seasonal level to identify a seasonal model.
STEP 3: Combine models from STEP 1 and STEP 2 to identify an overall tentatively model.

Here we do not have enough data to do anything on the seasonal level. However, a priori, we know that our model should be somewhat periodic –in this case nearly sinusoidal. It seems from inspection of the data that the half-period is about 5. So we will use a seasonal lag of 10.


Forecasts from ARIMA(0,0,1)(1,0,0)[10] with non-zero mean
5890.6
5890.4
5890.2
5890.0
5889.8
5889.6
5
10
15
20

4 Comparison and Selection

```
> mod1 = Arima(mydata, order=c(1,0,0)); mod1
Series: mydata
ARIMA(1,0,0) with non-zero mean

Coefficients:
         ar1  intercept
      0.8114  5890.2082
s.e.  0.1451     0.2236

sigma^2 estimated as 0.03193:  log likelihood=2.8
AIC=0.4   AICc=3.83   BIC=1.6
> mod2 = Arima(mydata, order=c(0,0,1)); mod2
Series: mydata
ARIMA(0,0,1) with non-zero mean

Coefficients:
        ma1  intercept
      1.000  5890.1066
s.e.  0.224     0.0997

sigma^2 estimated as 0.0298:  log likelihood=2.47
AIC=1.06   AICc=4.48   BIC=2.25
> mod3 = Arima(mydata, order=c(1,0,1)); mod3
Series: mydata
ARIMA(1,0,1) with non-zero mean

Coefficients:
         ar1     ma1  intercept
      0.7441  0.9425  5890.1812
s.e.  0.1811  0.8118     0.2057

sigma^2 estimated as 0.01189:  log likelihood=6.89
AIC=-5.78   AICc=0.89   BIC=-4.19
> mod4 = Arima(mydata, order=c(1,0,2)); mod4
Series: mydata
ARIMA(1,0,2) with non-zero mean

Coefficients:
         ar1     ma1     ma2  intercept
      0.7315  1.3366  0.9997  5890.2413
s.e.  0.1854  0.5586  0.6338     0.2254

sigma^2 estimated as 0.005521:  log likelihood=9.58
AIC=-9.17   AICc=2.83   BIC=-7.18
```

```
> shapiro.test(residuals(mod1))

        Shapiro-Wilk normality test

data:  residuals(mod1)
W = 0.91597, p-value = 0.2865

> shapiro.test(residuals(mod2))

        Shapiro-Wilk normality test

data:  residuals(mod2)
W = 0.90349, p-value = 0.2038

> shapiro.test(residuals(mod3))

        Shapiro-Wilk normality test

data:  residuals(mod3)
W = 0.95873, p-value = 0.7557

> shapiro.test(residuals(mod4))

        Shapiro-Wilk normality test

data:  residuals(mod4)
W = 0.89752, p-value = 0.1724
```

```
> mod5=arima(mydata,order=c(0,0,1), seasonal=list(order=c(1,0,0),period=10)); mod5

Call:
arima(x = mydata, order = c(0, 0, 1), seasonal = list(order = c(1, 0, 0), period = 10))

Coefficients:
         ma1    sar1  intercept
      1.0000  0.9494   5890.057
s.e.  0.1076  0.0557      0.112

sigma^2 estimated as 0.00311:  log likelihood = 4.57,  aic = -3.13
>
> shapiro.test(residuals(mod5))

        Shapiro-Wilk normality test

data:  residuals(mod5)
W = 0.89194, p-value = 0.147
```

5 Calculating the Mass

We use the original data and forecasted values from the fifth model in what precedes:

1	5890.4106
2	5890.4956
3	5890.4912
4	5890.3047
5	5890.0142
6	5889.8154

7	5889.6421
8	5889.6357
9	5889.7637
10	5890.0562
11	5890.3179
12	5890.406
13	5890.469
14	5890.292
15	5890.016
16	5889.828
17	5889.663
18	5889.657
19	5889.779
20	5890.056
21	5890.305
22	5890.389

Using some basic physics allows to estimate the mass of this orbiting object.
First: We calcualte the radial velocity from the Doppler shift via

$$\frac{\Delta\lambda}{\lambda_{Na}} = \frac{v_{rad}}{c}.$$

Second: In order to obtain the radial velocity of the observed star, we will fit the measured data through a cosine.

$$v_{rad} = v_0 + \frac{A}{2}cos(2\pi t/T + \phi).$$

Kepler's Third Law states

$$\frac{GT^2}{(SN)^3} = \frac{4\pi^2}{M_S + m_N}.$$

Assuming the orbit is fairly circular, $M_S \approx M_S + m_N$, and applying the baryocenter law we get:

$$m_N = v_0 \left(\frac{TM_S}{2\pi G}\right)^{1/3}.$$

After all is said and done we get 300 Jovian Masses $\approx m_N \approx$ 0.3 Solar Masses. This only has an error of $4.41\% \pm 2 \cdot 0.27\%$. Since this supposed object (directly-unobserved object) is so much larger than Jupiter, the unknown companion must be a star. In fact, it is a brown dwarf star that emits almost no visible light.

6 Conclusions

Judging purely on the physical basis –the seasonal ARMA model was the best model. I attempted to fit cosine curves given data using the other models and carry out calculating the mass of the dwarf star. I couldn't get good fits and thus I couldn't accurately calculate the mass.

From the statistical perspective –there is too little data to speculate. The models selected, for the most part, had significant residuals (θ_2 in ARMA(1,2) was not –neither was θ_1 in ARMA(1,1)). However, there seemed to be problems with stationarity. Of course, this is due to the lack of data.

I'll leave the reader with a few fitting fun facts.

Fun Facts:
On average there are 1-3 Earth-sized planets per star system.
The majority of star systems are binary star systems.

6.1 Physical Implications and Follow Up Research

Given more data, more accurate data (our data comes from the 1990s –the technology is obsolete already), and utilizing more of the spectral absorption lines could allow us to fit a better model and help us get a more accurate calculation of the orbiting mass. If we could also ascertain information about its orbit: eccentricity, inclination, etc., then we could "delete " the gravitation effects of the dwarf star from our original data. Our transformed wavelength data would then reflect the effects of other bodies in the system.

Often times this is impossible to do –usually due to the difference in the relative masses. However, we can also get data and detect exoplanets via other means. I had originally intended to try to use other data to do this. However, due to the lack of spectral data I could find for this system, it seemed ridiculous to attempt this.

References

[1] Schneider, Jean (26 October 2012). "Interactive Extra-solar Planets Catalog". The Extrasolar Planets Encyclopedia. Retrieved 2012-10-26.

[2] http://arxiv.org/abs/1310.6907

[3] Griffin, R. F. 1991. An improved spectroscopic orbit for HD 75767. Astronomical Society of India. BASI, Vol. 19, No. 3 & 4, pp. 183-192

[4] Fuhrmann, Klaus; Guenther, Eike; König, Brigitte; Bernkopf, Jan. 2005. The case and fate of HD 75767 - neutron star or supernova? Monthly Notices of the Royal Astronomical Society, Volume 361, Issue 3, pp. 803-808.

[5] http://exoplanetarchive.ipac.caltech.edu/