Stat 153 Final Project

Hongzhe Li

Prof. Brillinger

Dec 15 2014

Time Series Analysis of Gambling Industry

Introduction:

Gambling industry has always been a controversial issue in people's daily life, because it could easily change people's mood or even life. I am, personally, fond of going to Las Vegas during my holidays. The scientific question motivating my work is because when I am in the casino, I always thought that Gambling industry has really brought enormous capitals to the U.S government, and are people working in casinos like their jobs. Therefore I decided to see if there exists certain patterns in the data of employment populations during several years. **The question I am addressing is that does the population of employment in Gambling Industry of Las Vegas follows any particular model, and what is the best-fit model or models for that data.**

Data:

The data used in this paper is the monthly population of employees (units are in thousands) in Las Vegas – Paradise from Jan 1990 to Sep 2014. I extracted the data from the website of Federal Reserve Bank of St. Louis, U.S. Department of Labor: Bureau of Labor Statistics, and the R-library of rdatamarket. Because, eventually, I am hoping to make predictions in order to examine how well my models work, I will only use the first 276 data for my modeling (From Jan 1990 to Sep 2014). After selecting models, I will compare the rest of data (From Jan 2013 to Sep 2014) with my prediction. Figure 1 on the next page is the plot of the raw data.

Methodology:

Figure 1

By conducting the original data, it does not look quite stationary, since the means and variances are changing over times or not constant. The next step of analyzing is to try to make the plot approaching stationary. As a result, I tried taking the square root and log of the data, which did not help a lot, because the shape of the plot still remained the same. Therefore, I decided to use Loess model in order to take away the curve, and also smoothing the original plot, which would be benefit for my future analysis.

Loess Model:

Figure 2

Figure 2 is the loess curve added on the original data. I chose span to be 0.5, because under this value, the residuals are the most similar one that close to white noise, and simultaneously the curve will not be over fitted.

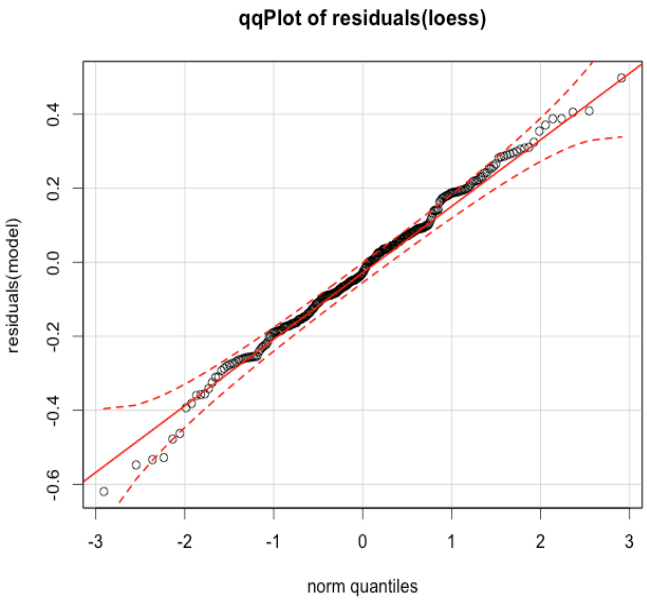
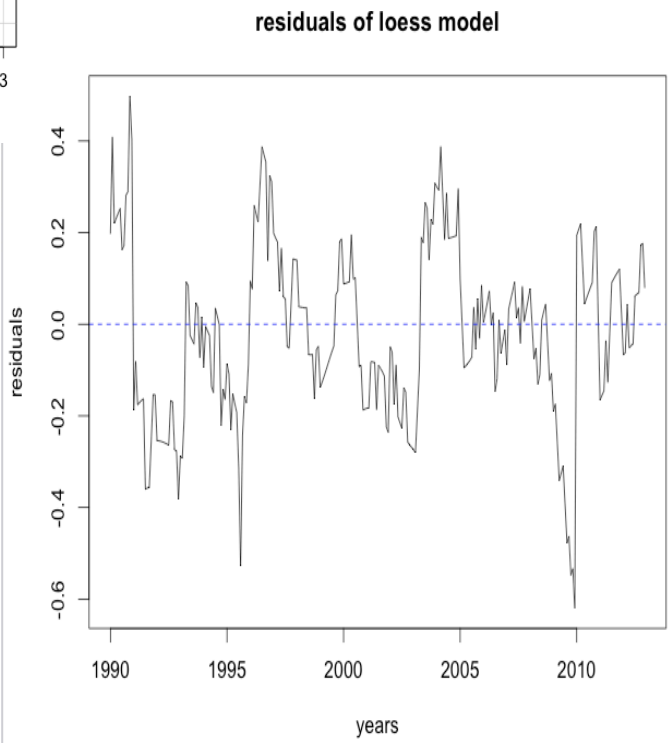


Figure 3

The graph on the left is the qqPlot(Figure 3) of residuals, based on this plot, there are only few outliers in the beginning and ending.

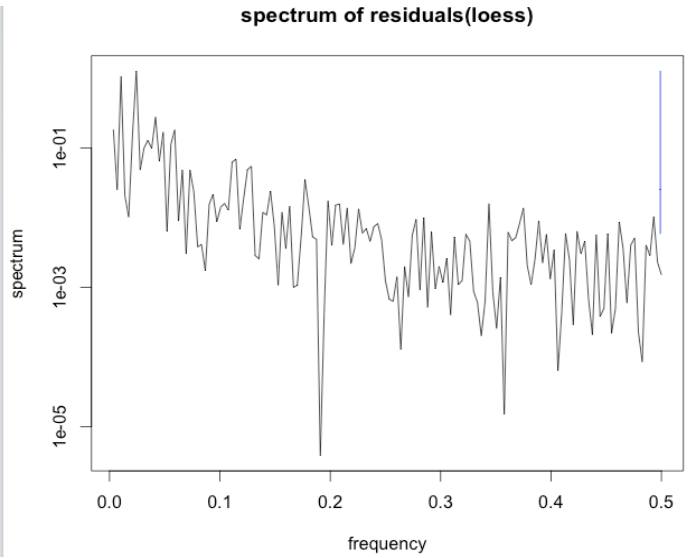
The next step is to find a suitable model for the residual after fitting loess curve.

Figure 4



I started from inspecting the plot of the residuals (Figure 4) and expected to see the evidence of stationary or whether there exists any trends. Unfortunately, the plot of residuals does not look stationary and I could not see any trend hiding behind it.

Figure 5



Then, I used the spectrum (Figure 5) and discovered that I could probably fit an ARMA model since the spectrum looks really random. In order to clarify my assumption, I also checked ACF (Figure 6) and PACF (Figure 7) of residuals, which indeed strengthened my early assumption.

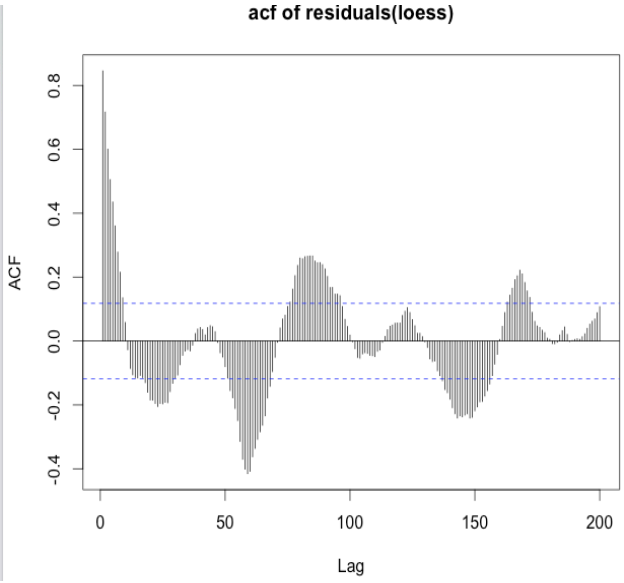


Figure 6

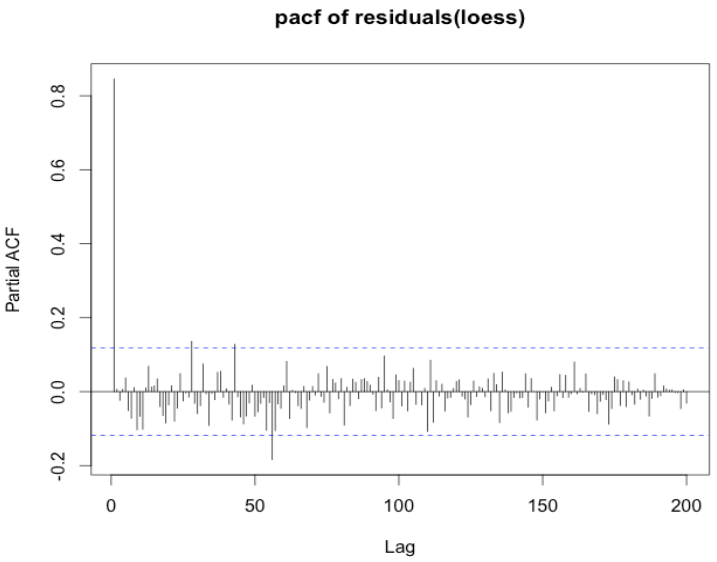
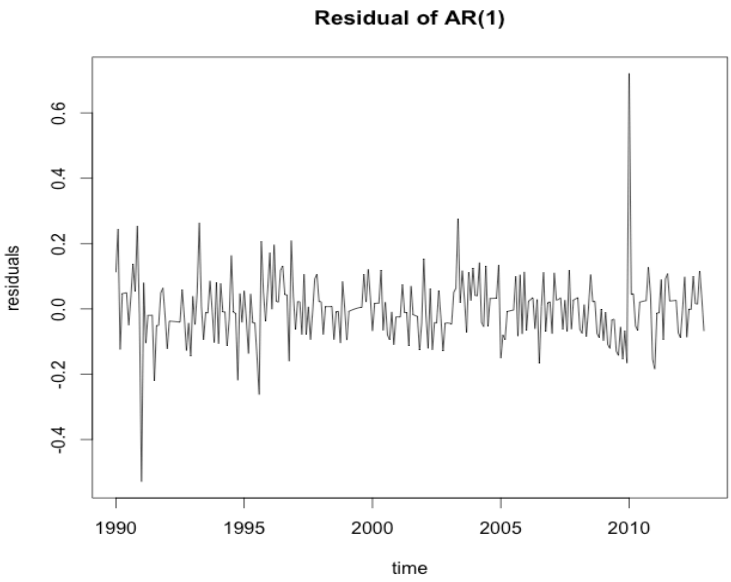


Figure 7

According to the ACF, the figure is tailing off, and by analyzing PACF, I discovered that there is a clear cut off after lag 1 (I omitted few small cut offs such as the negative one at Lag 52). Thereafter, I fitted an AR(1) model to the residuals of the Loess Model, and as I did earlier in the report, I will check how well my model fits by analyzing the plot of residuals.

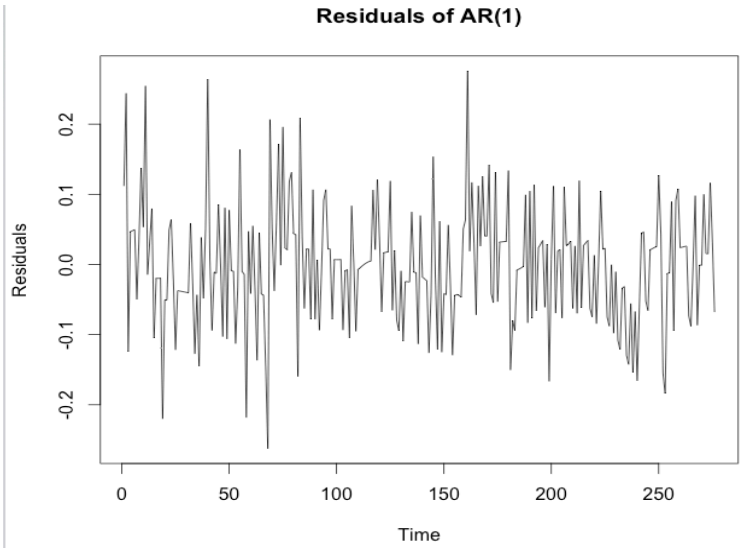
AR(1) Model:

Figure 8



I plotted the residuals (Figure 8) in order to check how well my model fits, the residuals looked much better than before and quite approach to white noise except for two outliers at year 1991 and 2010. What caused those two outliers really drew my attention, because my data is employment population, as known to all, it should be changing gradually over times. As a result, I did some researches online and finally figured it out. (I will discuss it in the discuss section at the end of the report).

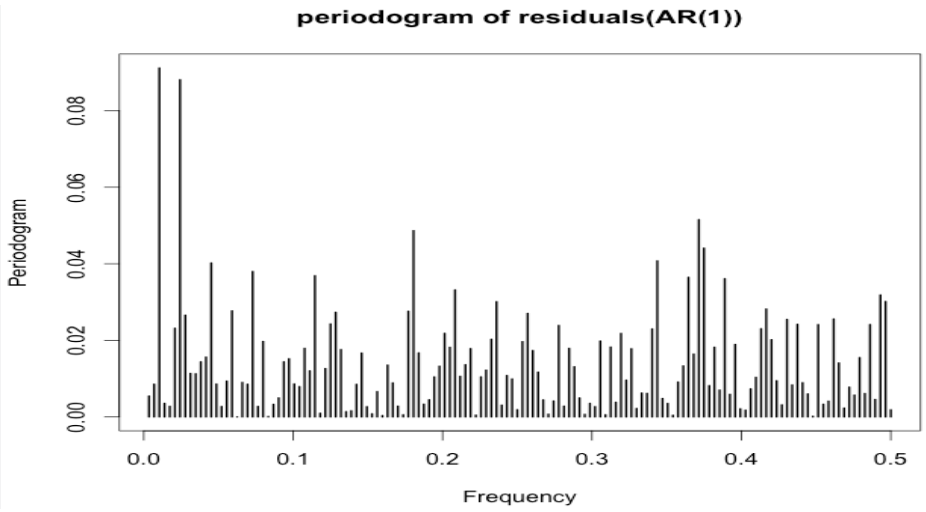
Figure 9



Then, I decided to ignore two outliers; as a result, I used the function tsoutlier to find two suitable substitutions for outliers. Figure 9 is the modified plot of residuals (Notice I did not use the same scale of y-axis because it will be benefit for my analyzing by enlarging the plot). Now, the plot of residuals looks quite similar to a white noise, but I looked thoroughly and noticed that there were several "flat floors" appeared around residuals equal to zero. In other

words, the plot sometimes goes up or down to residual of zero and keeps the same value for several lags. From my perspective, this happens due to my data are in thousands and are not very accurate (For example, the same value of 6.8 from Jul 2004 to Nov 2004), in which there are several months that have equal values. For dealing with the "flat floors", I suspected that there could be certain harmonic trends hiding behind the residuals. Hence, I looked at the periodogram of the residuals (Figure 10).

Figure 10



The periodogram looks quite messy but there are 4 apparent spikes. I would ignore the first two spikes, and used the frequency around 0.18 and 0.38, since the first two spikes were having very low frequencies. Then, I fitted 2 sine and cosine curves with frequency 0.3715278 and 0.1805556

Harmonic:

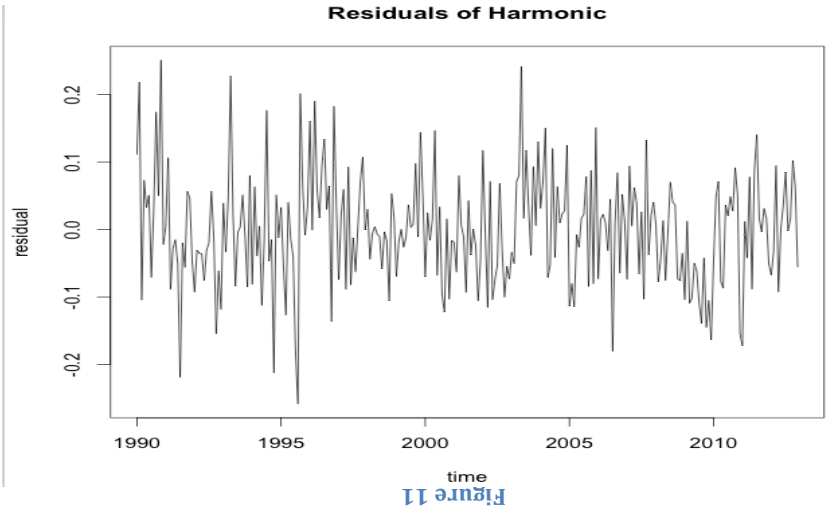


Figure 11

After fitting the harmonic model, the residuals (Figure 11) are the best I have seen so far with constant means and variances, which almost like a white noise. I looked at the

histogram (Figure 12) which looks very normal, to make sure the normality of residuals I also looked at qqPlot (Figure 13), which turned out to be really good. Additionally Sharipo-Wilk normality test gives me 0.2169 which forbid me from rejecting normality.

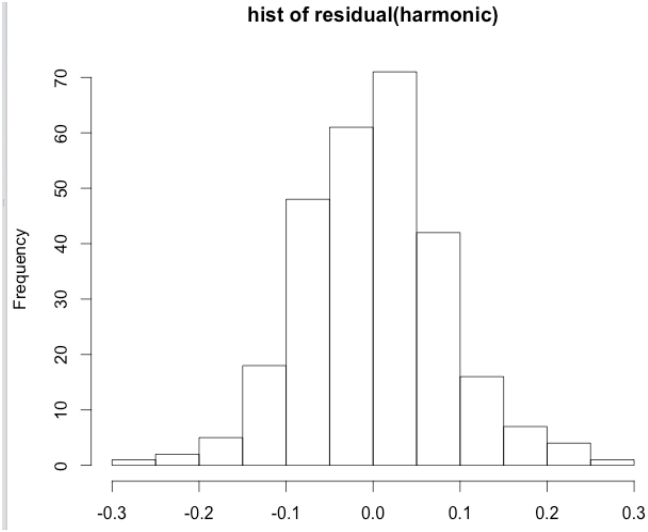


Figure 12

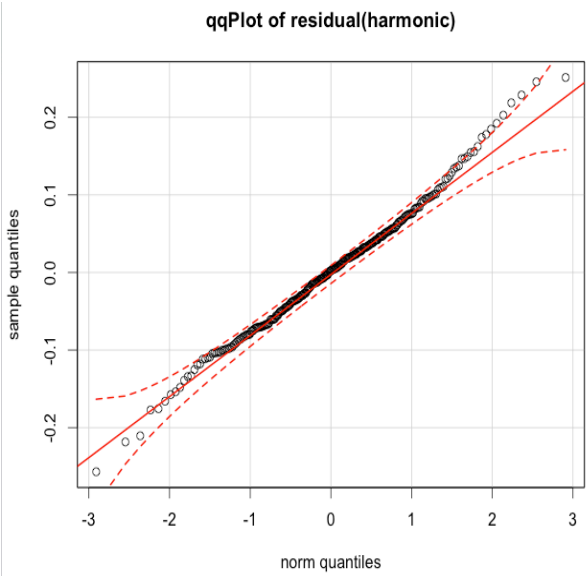


Figure 13

```
        Shapiro-Wilk normality test

data:  freq_lm$residuals
W = 0.9929, p-value = 0.2149
```

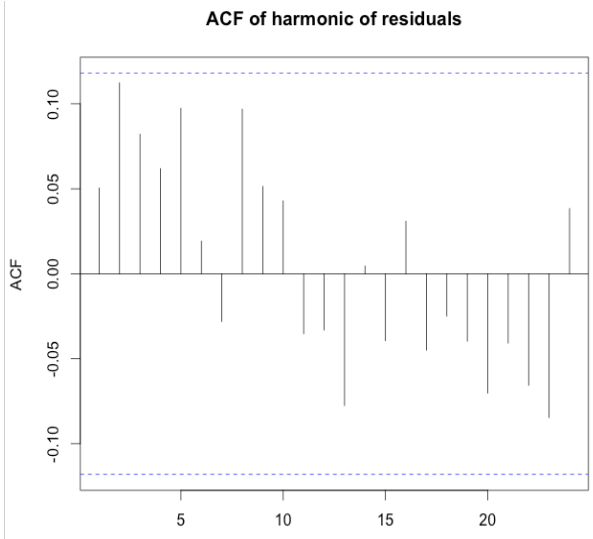


Figure 14

According to ACF (figure 14), there is no cut offs, which reveals that the residuals are independent.

Up to this point, I have already reached the most white noise likely step, and this finishes my modeling.

Prediction:

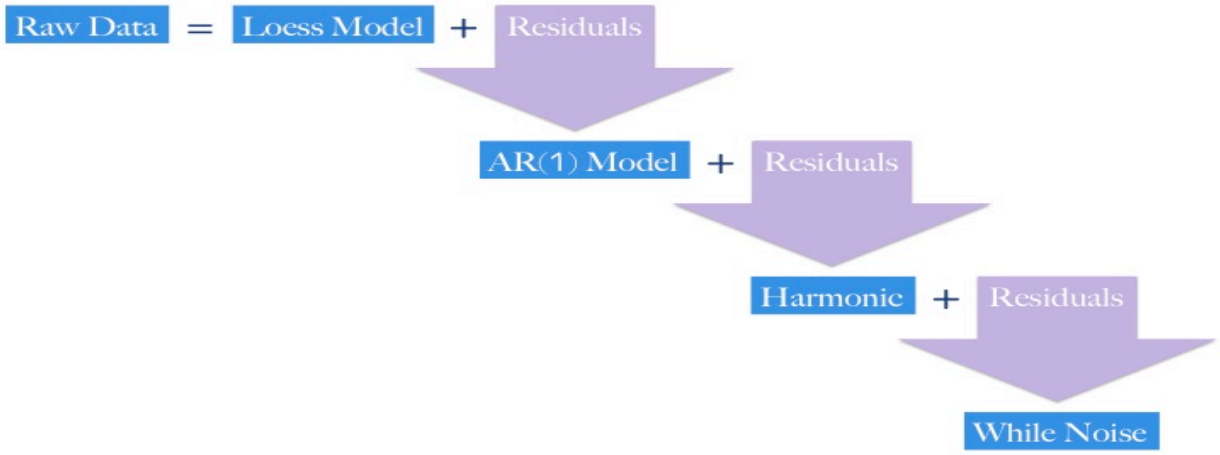
As I said in the beginning of the report, I only used the first 276 data points for selecting models, there are 21 data left to compare with my prediction. In order to make prediction I have to work backwards, starting from the harmonic model and go back to loess. I used the predict function in R to generate the prediction and finally conduct it (figure 14).

Figure 14

Even though the prediction does not look very close to the original data, probably because the reason I mentioned earlier about the accuracy of my data, but at least the actual data lies within the 95 percent confidence interval. And I will try to find better data for the gambling industry in order to predict well.

Conclusion:

The schema below summarizes my whole procedure for modeling the gambling data. Basically, I started from the raw data, and used Loess to take away the curve, then by fitting AR(1) model on the residuals of Loess, and two sine and cosine on the residuals of AR(1) which finally guide me to led the residuals to white noise.



<u>Throughout the stated analysis above, the answer to my question proposed in the beginning of this report is that the best-fit models for the population of employment in gambling industry of Las Vegas - Paradise is *Loess* + *AR(1)* + *Harmonic* which are my final models.</u>

Discussion

As I pointed out earlier, there are two outliers. One is from Dec 1990 to Jan 1991, which goes down from 6.1 to 5.5; the other one is from Dec 2009 to Jan 2010, which jumps from 4.9 to 5.7(thousands of people). Based on my research, the decline from 1990 to 1991 is probably because Chinatown was built in 1990 and many Asian people working for gambling industry

might probably give up their current jobs in casinos or hotels to work in Chinatown. The other time period that has a huge jump from 2009 to 2010 is probably due to the reason of the economic booming in the last half year of 2009. As a result, lots of Hotels and casinos were built in 2009 such as Hard Rock. Additionally, CityCenter starts accepting applications for more than 12,000 open jobs.

Reference:

Federal Reserve Bank of St. Louis, U.S. Department of Labor: Bureau of Labor Statistics

http://www.lvcva.com/stats-and-facts/history-of-las-vegas/