
ANALYZING U.S. GDP TRENDS USING STOCK MARKET IDIOSYNCRASY

Stat 479 –Professor Douglas Wiens

Bohdan Horak

University of Alberta – School of Business

4/9/2015

CONTENTS

Motivation	3
Introduction	3
Measure of Stock Return Idiosyncrasy	4
Measure of Economic Production	4
Overview of Data	5
Overview of Data and Summary Statistics	5
Real GPD Series Analysis	7
Transformations	7
Fitting an SARIMA model	11
Residual Analysis	12
MODEL 1: $1,1,0 \times 1,0,02$	12
MODEL 2: $1,1,0 \times 0,0,02$	13
MODEL 3: $2,1,0 \times 0,0,02$	14
Forecasting	15
Test Set Evaluation	15
Predicting the Future	16
Idiosyncrasy Series Analysis	16
Transformations	16
Residual Analysis	22
MODEL 1: $1,1,2$	22
MODEL 2: $0,1,3$	23
MODEL 1: $0,1,2$	24
Forecasting	25
Test Set Evaluation	25
Predicting the Future	26

Frequency Domain Analysis.....	27
Spectra	27
Periodograms.....	27
Coherence.....	28
Optimal Filtering Real GDP Series.....	29
Optimal Filtering Idiosyncrasy Series	31
Impulse Response Model.....	33
Conclusion	36
Bibliography.....	37
Appendix: R-Code	38

MOTIVATION

In his work "Capitalism, Socialism and Democracy" Joseph Schumpeter's coined the term Creative Destruction to denote a "process of industrial mutation that incessantly revolutionizes the economic structure from within, incessantly destroying the old one, incessantly creating a new one." (Schumpeter, 1947) Thus, creative destruction predicts economic upheaval. New companies, with new ideas will force their way into the market and upend the older established companies. These new companies bring new better products and services into the market thereby increasing the productivity and as a result drive economic growth in the economy.

INTRODUCTION

In finance, and more specifically portfolio theory, it is often useful to split the volatility of the returns associated with any particular stock into two portions – idiosyncratic and systematic. Idiosyncratic volatility, sometimes referred to as firm-specific risk, is the risk associated with the day-to-day running of the company and the business and investment decisions that managers make. The systematic risk is the risk associated with the market as a whole. For example, if the oil price drops and the stocks of oil producers drop as a result, this would represent a systematic event, thereby contributing chiefly to the systematic volatility. However, if one of the oil companies is able to implement a new technology and cut its costs, thereby pushing its stock higher, would represent an event which increases contributes to idiosyncratic volatility.

In a sense the idiosyncratic volatility is a proxy for creative destruction and therefore innovation. The winner of creative destruction will outpace the losers, thereby creating a gap in the fundamentals of firms. This rise in firm specific fundamentals volatility elevates firm-specific or idiosyncratic risk (Morck, 2013). Thus, in a healthy, growing economy we expect to see a lot of idiosyncratic risk, while in a stagnant economy we expect to see a lot systematic risk. The purpose of this paper is to examine the opposite. Can we predict economic output, i.e. GDP, with a measure of market wide idiosyncratic risk?

How do we determine which stock price swings are results of firm specific events, and which are results of market wide swings? The answer is quite technical and there is some debate in

the finance literature as to how specifically it should be calculated, thus we will only go through one of the ways to calculate it and only briefly.

MEASURE OF STOCK RETURN IDIOSYNCRASY

The measure uses regression and is outlined in more detail and in addition to an alternative method in (Zhang, 2014).

$$R_{js} = \alpha_{jt} + \beta_{m,jt}R_{ms} + \beta_{i,jt}R_{is} + \epsilon_s$$

Where R_{js} is the stock return of stock j on day s and R_{ms} and R_{is} are value weighted market and industry returns on day s . We exclude return of stock j , in the calculation of both the market and industry return. The stocks systematic return volatility is the sum of squares of variation explained by the regression, SSM and the firm-specific volatility is defined as the squared variation attributed to the regression errors, SSE .

Stock j 's idiosyncrasy ratio, is defined as the log ratio of its firm-specific volatility over its systematic volatility, we denote this measure as $\phi_{j,t} = \log\left(\frac{SSE}{SSM}\right)$. To obtain market wide measure, which we will use in this analysis, we take value-weighted averages of the firm measures¹. The resulting time series is an aggregate measure of stock market idiosyncrasy with a quarterly frequency.

MEASURE OF ECONOMIC PRODUCTION

As we want to try and predict economic output, we will need its measure. We will use GDP, this series is updated quarterly by the U.S. Department of Commerce: Bureau of Economic Analysis (US. Bureau of Economic Analysis, 2015). We will use real GDP as opposed to nominal, as inflation might adversely affect our results and the prediction of inflation already has quite a few economists occupied so we will leave that to them and focus on real GDP. The base year of

¹ Big thanks to Lu Zhang for providing me with the already calculated and aggregated idiosyncrasy measures. For those looking to replicate the results, if asked, I will provide the STATA code used (also provided by Lu). STATA is used because R does not work well in this case as the New York Exchange features a huge number of stocks and therefore the data set is very large. So large in fact that R will not be able to handle it without some tinkering.

our GDP series is 2008. In other words, the GDP series stemming back to 1948 will be in 2008 dollars.

OVERVIEW OF DATA

OVERVIEW OF DATA AND SUMMARY STATISTICS

Measure	<i>Idiosyncrasy</i>	Real GDP
Mean	2.35	7702.64
Standard Error	0.08	267.44
Median	2.10	6507.05
Standard Deviation	1.25	4378.23
Sample Variance	1.57	19168901.15
Kurtosis	4.88	-1.12
Skewness	1.75	0.46
Range	8.08	14305.20
Minimum	0.39	1989.50
Maximum	8.47	16294.70
Count	268.00	268.00
Start	1/1/1948	1/2/1948
End	10/1/2014	10/2/2014

Figure 1: Summary Statistics of the two series. To get a look at what the data looks like.

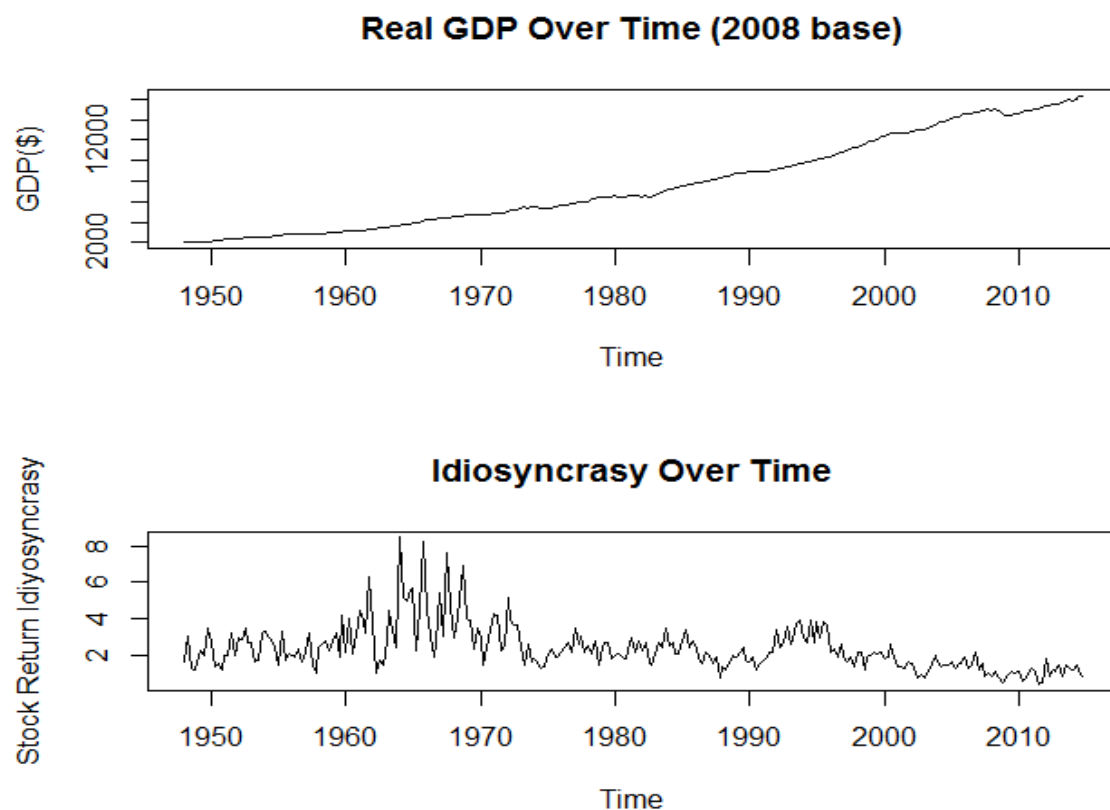


Figure 2: Time series plots of both series. We notice that there is a linear trend in the Real GDP series. We also notice a big spike in the variance in the Idiosyncrasy series. This leads us to believe that the series are not stationary.

The data is clearly not stationary. We should try transforming the series to get them to satisfy stationarity requirements we require for the rest of our analysis. There seems to be a linear trend in the GDP series so we could try de-trending it. For idiosyncrasy we will need to try to stabilize the variance somehow, we could try taking a square root of the data. De-trending in the idiosyncrasy case is unlikely to help, as we don't have a linear trend, we have what looks like a relatively constant mean.

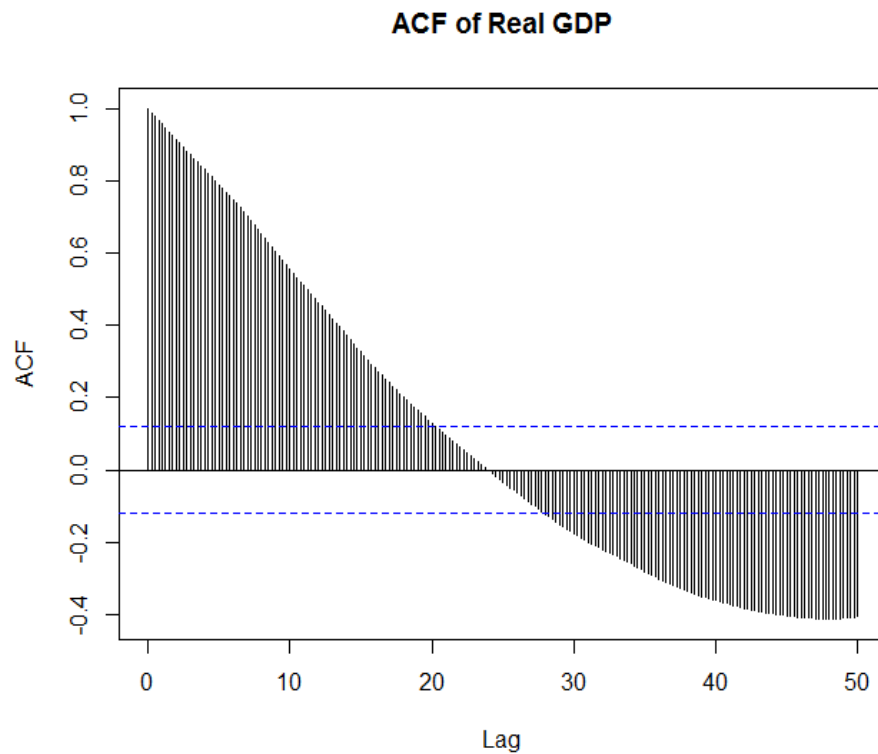


Figure 3: The ACF of real GDP clearly shows signs of non-stationary. We have highly significant autocorrelations at very high lags, we also have most values outside the confidence band denoted by the dotted blue lines. In addition, we see what seems like a sinusoidal pattern.

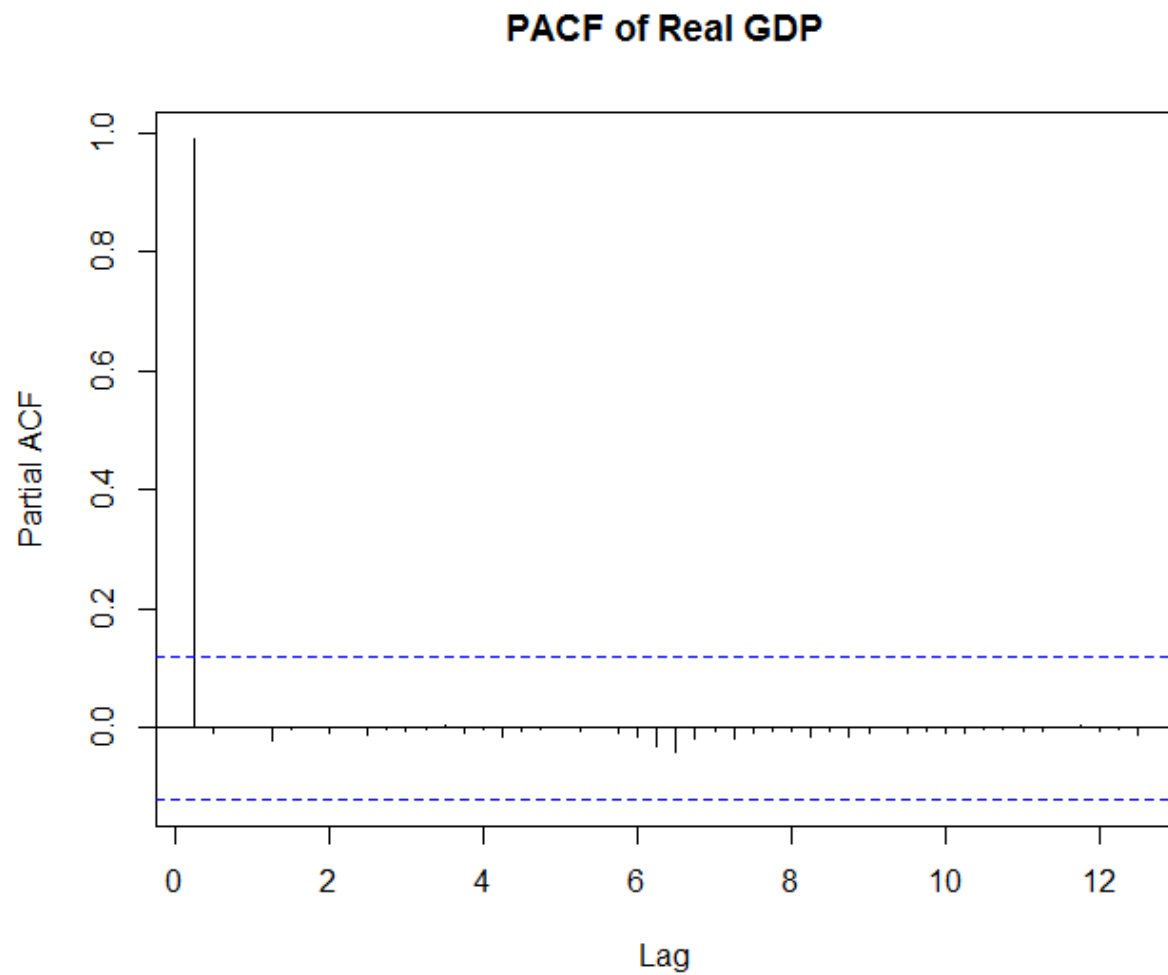


Figure 4: We see a very large PACF value at lag 1, which might indicate non-stationarity. The Real GDP series will need to be transformed in order to progress through our analysis.

First we try de-trending. That is, we fit a linear regression model on its own times then do the analysis on the resulting residuals. We saw what looks like a linear trend in Figure 2, thus we try this first.

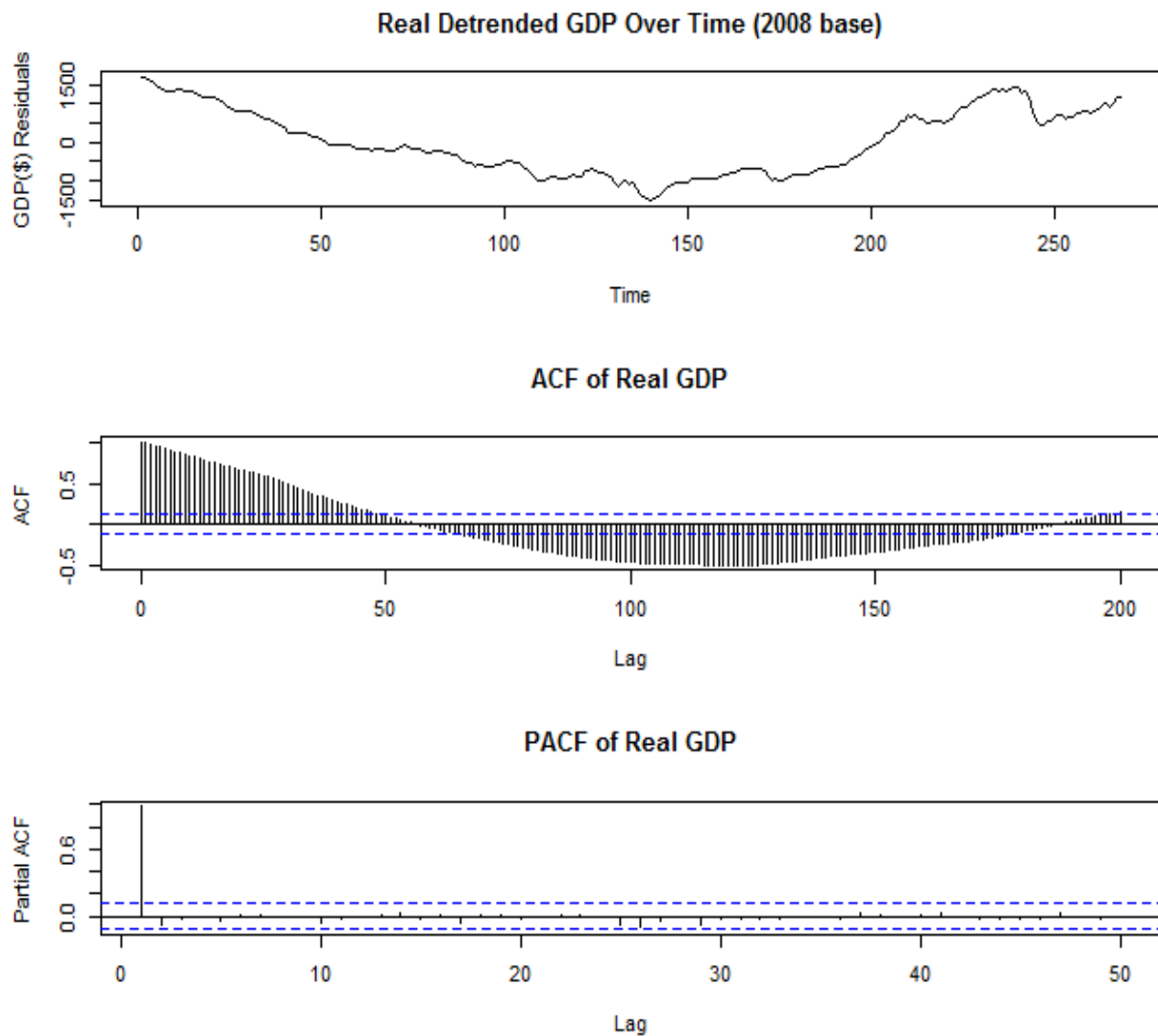


Figure 5: We see that the de-trending did not make things much better. We see variation in the mean (top plot), highly significant ACF's at high lags(plot 2).

As de-trending clearly did not help, we will move on to differencing or log transformations.

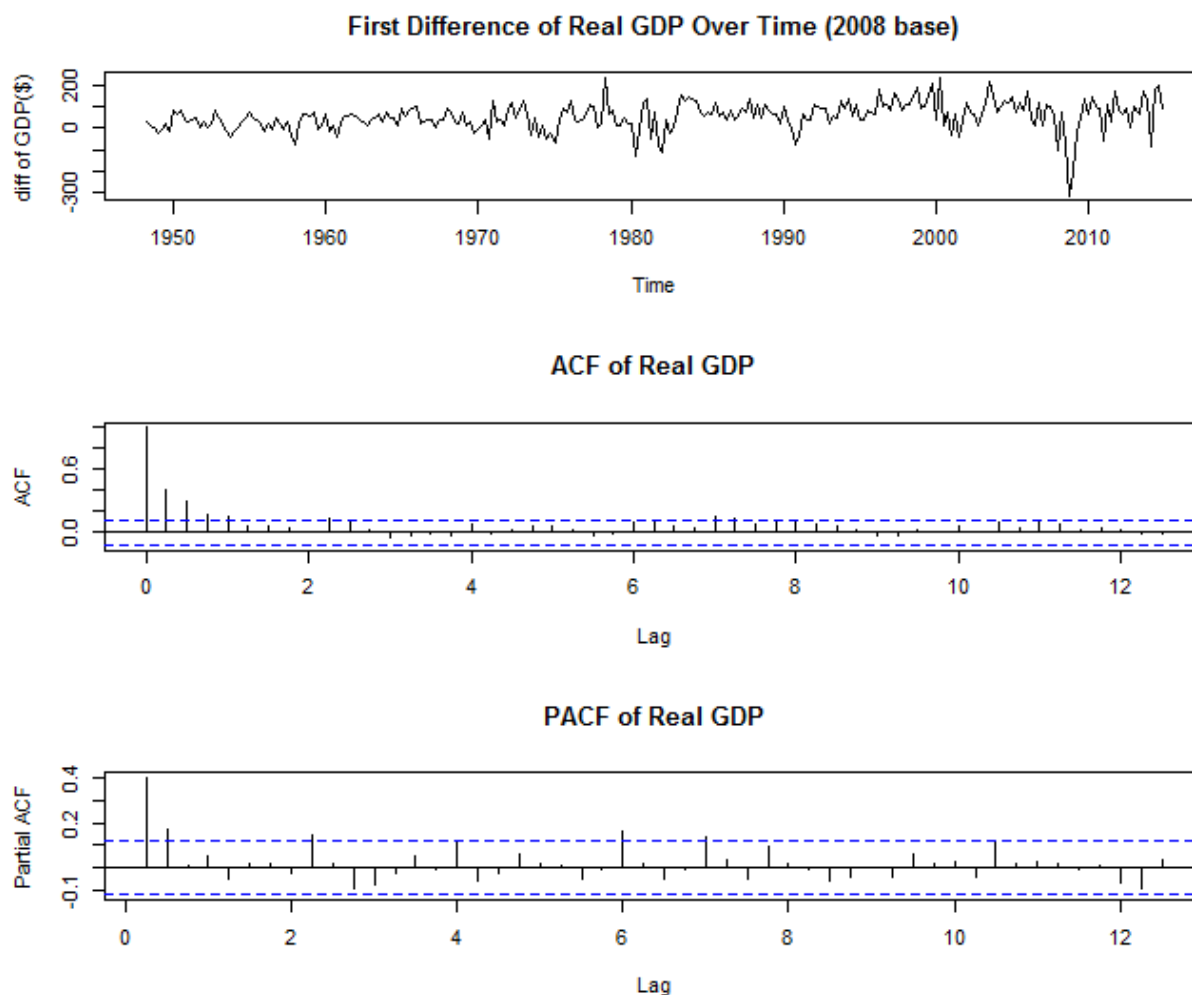


Figure 6: Plot one exhibits a seemingly constant mean. Economic side note: There is a significant dip in GDP during the 2007-2008 period when the housing market collapse caused a dramatic slowing of housing related production. In addition, the drop in property values also caused a negative wealth effect in which homeowners suddenly felt 10-20% poorer and thus cut back on consumption. The ACF tapers down slowly indicating a possible AR model. From the PACF plot we note that there is a significant value at lag=0.25 (1 quarter) and a regular repetition of significant PACF values every 1.75-2 years or 7-8 quarters, indicating a possibility of a seasonal AR component. Upon closer inspection of the ACF we also note some regularity, even though not always significant it does corroborate what the PACF is telling us. As none of these are significant we can assume no seasonal MA component.

FITTING AN SARIMA MODEL

We first guess that the model has some sort of autoregressive behavior from Figure 6 ACF plot. Furthermore we also think there is an autoregressive seasonality component with $s \approx 2$ (or 8 quarters) in the model from the PACF plot. We also know that this model needs be differenced so $d=1$. Using this knowledge our first guess for a model would be $(1,1,0) \times (1,0,0)_2$. With this in mind we run a batch process to test similar models and look for stand out performers based on AIC and BIC criteria. After doing this analysis we find that our original model, while good may not be the best.

p	d	q	P	D	Q	S	AIC	AICc	BIC
1	1	0	0	0	0	2	9.2277	9.2355	8.2545
1	1	0	1	0	0	2	9.2108	9.2188	8.2510
1	1	0	2	0	0	2	9.2104	9.2187	8.2640
1	1	1	0	0	0	2	9.2058	9.2139	8.2460
1	1	1	1	0	0	2	9.2120	9.2203	8.2656
1	1	1	2	0	0	2	9.2154	9.2241	8.2824
1	1	2	0	0	0	2	9.2122	9.2205	8.2658
1	1	2	1	0	0	2	9.2170	9.2256	8.2840
1	1	2	2	0	0	2	9.2230	9.2321	8.3034
2	1	0	0	0	0	2	9.2049	9.2130	8.2451
2	1	0	1	0	0	2	9.2108	9.2191	8.2644
2	1	0	2	0	0	2	9.2155	9.2241	8.2825
2	1	1	0	0	0	2	9.2119	9.2202	8.2655
2	1	1	1	0	0	2	9.2159	9.2246	8.2829
2	1	1	2	0	0	2	9.2228	9.2319	8.3032
2	1	2	0	0	0	2	9.2173	9.2259	8.2843
2	1	2	1	0	0	2	9.2233	9.2324	8.3037
2	1	2	2	0	0	2	9.2280	9.2375	8.3218

Figure 7: We see that our original model is slightly worse than some others (highlighted in green).

While useful a useful tool the criteria are not perfect, we will need to look at the residuals analysis to see which of our three models is viable.

RESIDUAL ANALYSIS

We will do a basic residual analysis to pick the best model. We will then use this model to try and forecast GDP.

MODEL 1: $(1,1,0) \times (1,0,0)_2$

Model Coefficients Are

Model 1	ar1	sar1	constant
<i>Coeff.</i>	0.3402	0.1688	53.6674
<i>s.e.</i>	0.0628	0.0662	6.6718

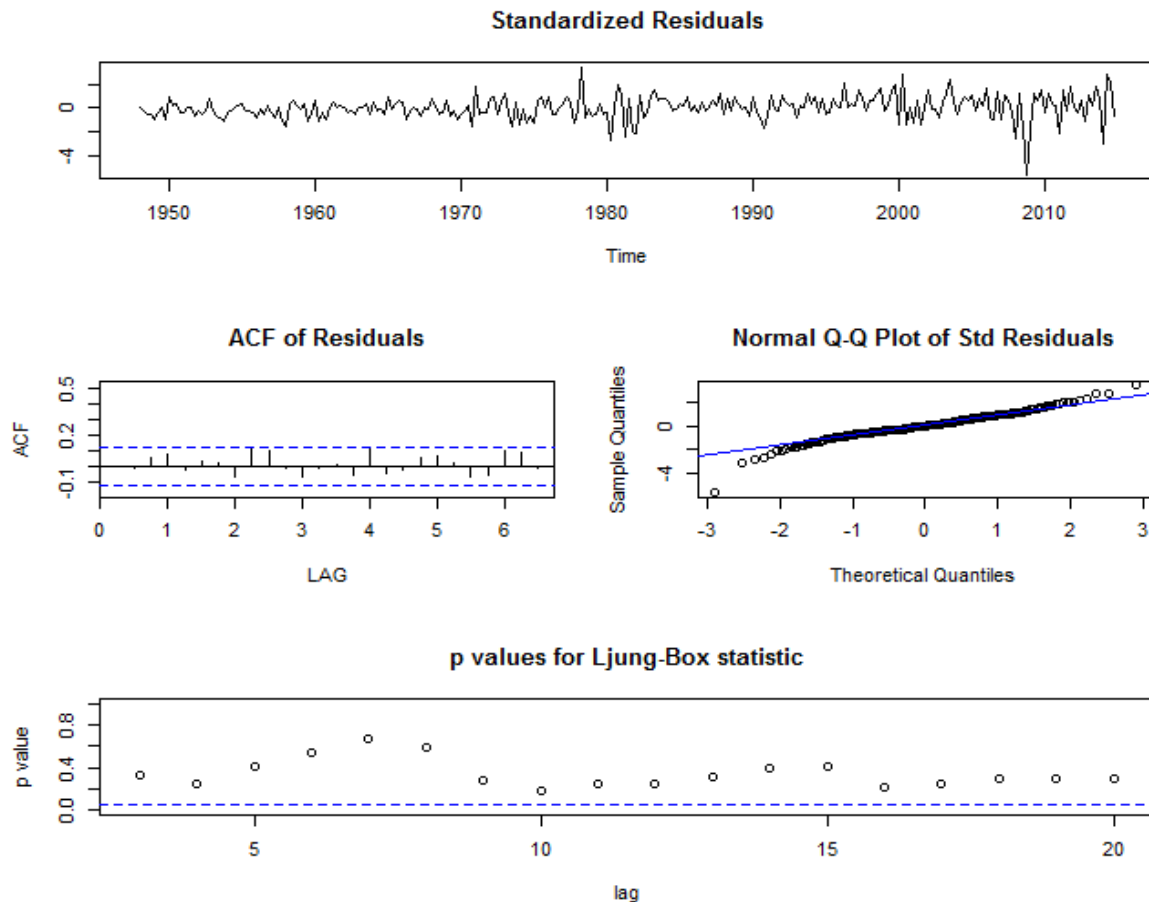


Figure 8 & 9: Displayed are the coefficients of the first model and the residual tables generated by R.

We see that the standardized residuals do indeed look like white noise, aside from the blip in 2008. The ACF also looks pretty good, no statistically significant values at any lag. The Normal Q-Q plot also looks good, it seems to follow the linear line fairly well, and we typically see the first and last few residuals deviating from the linear line. The p values for the Ljung-Box statistic are all significant. From this analysis this is not a bad model.

MODEL 2: $(1,1,0) \times (0,0,0)_2$

Second Model is

Model 2	ar1	constant
Coeff.	0.4019	53.6113
s.e.	0.056	6.1985

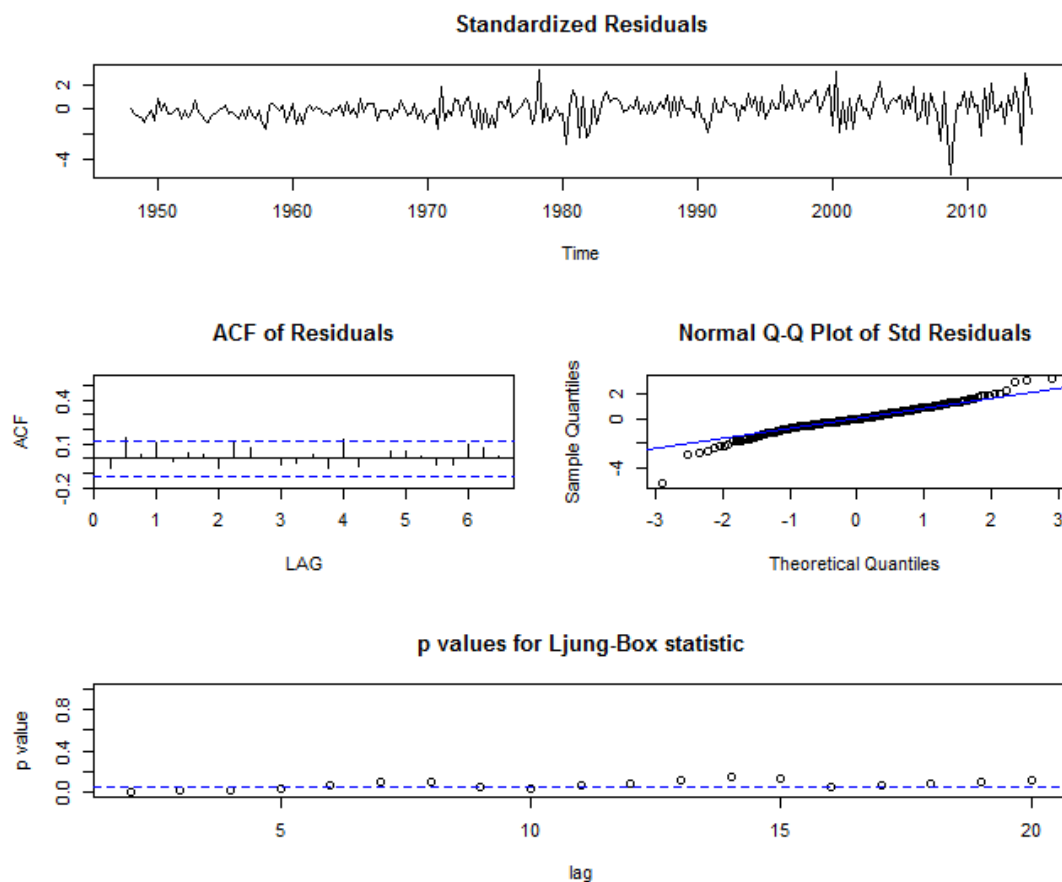


Figure 10 & 11: Model 2 and its residual analysis

While the white noise looks similar to model 1, the other indicators are significantly worse than in model 1. The ACF has some significant values, and the p-values for Ljung-Box statistic are far worse, they are barely significant at all. Thus, we discard this model in favor of the first.

MODEL 3: $(2,1,0) \times (0,0,0)_2$

Model 3	ar1	ar2	Constant
<i>Coeff.</i>	0.3327	0.173	53.7596
<i>s.e.</i>	0.0602	0.0607	7.3692

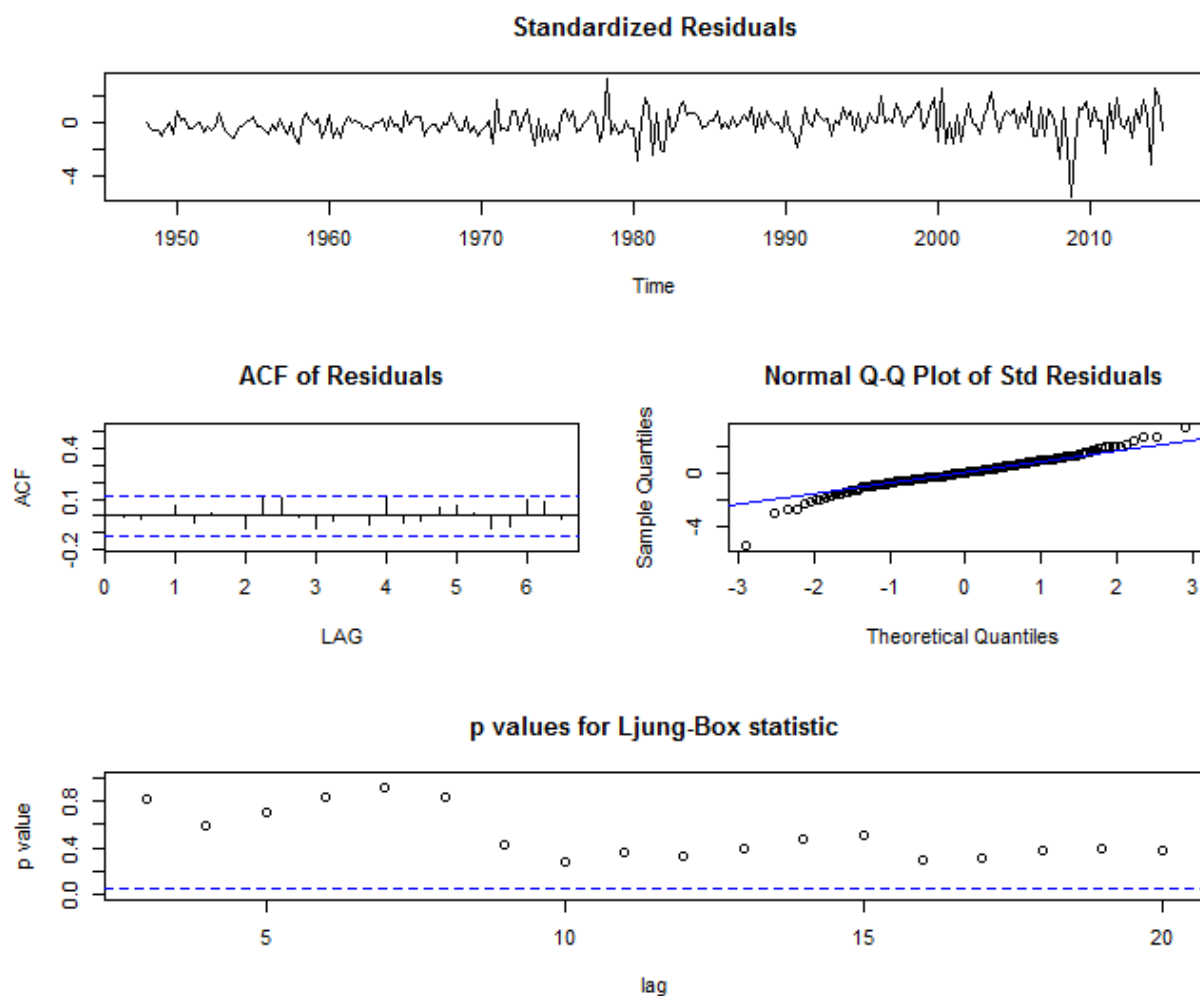


Figure 12 & 13: Model 3 and its residual analysis

We see that this model is also very good. It might be slightly better than the first but not by much. However, due to the Law of Parsimony we are inclined to pick the simpler model, which

in this case is the model with no seasonal component. However interesting the seasonality of GDP is, according to our metrics this model is superior.

FORECASTING

The final model is $(2,1,0) \times (0,0,0)_2$.

It looks like this

$$(1 - \phi_1 B - \phi_2 B^2)(1 - B)x_t = \omega_t$$

Where the B's are the Backshift operators. Plugging in our estimates

$$(1 - 0.3327B - 0.173B^2)(1 - B)x_t = \omega_t$$

TEST SET EVALUATION

In order to better judge the prediction quality of our model we split off the last four of the values in the GDP series and predicted them using our model. These are our results

Time	Actual	Predicted	% Difference	s.e.
1/1/2014	15916.2	16013.94608	-0.61%	58.25781
4/1/2014	15831.7	16088.10039	-1.62%	97.78644
7/1/2014	16010.4	16144.86087	-0.84%	136.606
10/1/2014	16205.6	16198.02133	0.05%	169.7691

As we can see over the last 4 quarters we were not off by very much at all. This just goes to show how consistent the U.S. is economically. We now try to predict the future.

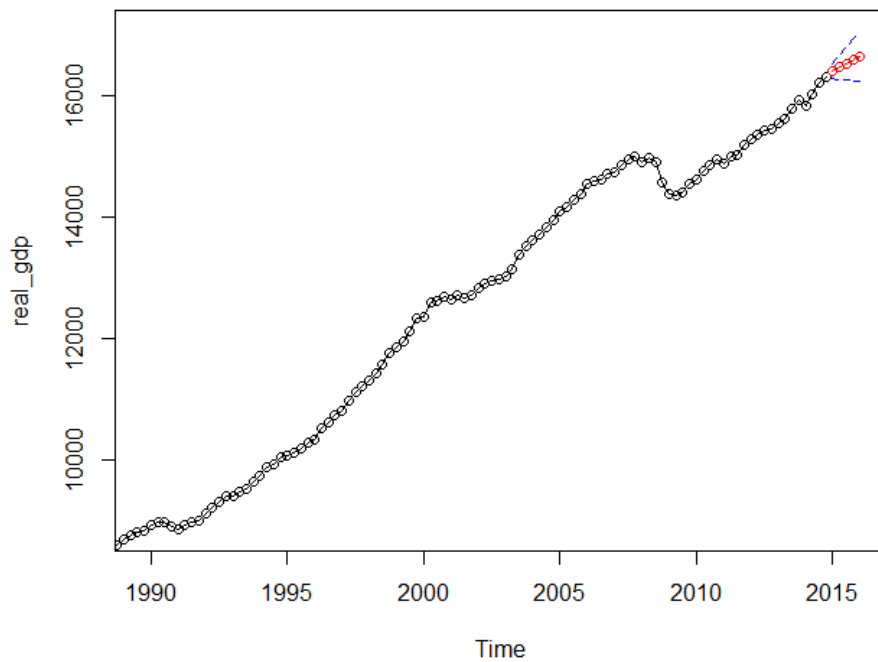


Figure 14: This is a 5-quarter or 1.25 year forecast of real GDP. It looks reasonable, from this we can expect the GDP to increase, but at a rate lower than what we have been seeing in the last few years.

IDIYOSYNCRASY SERIES ANALYSIS

TRANSFORMATIONS

From **Figure 2:** We notice the non-constant variability. We need to try and fix this before we start our analysis. However, let us first look at the ACF and PACF.

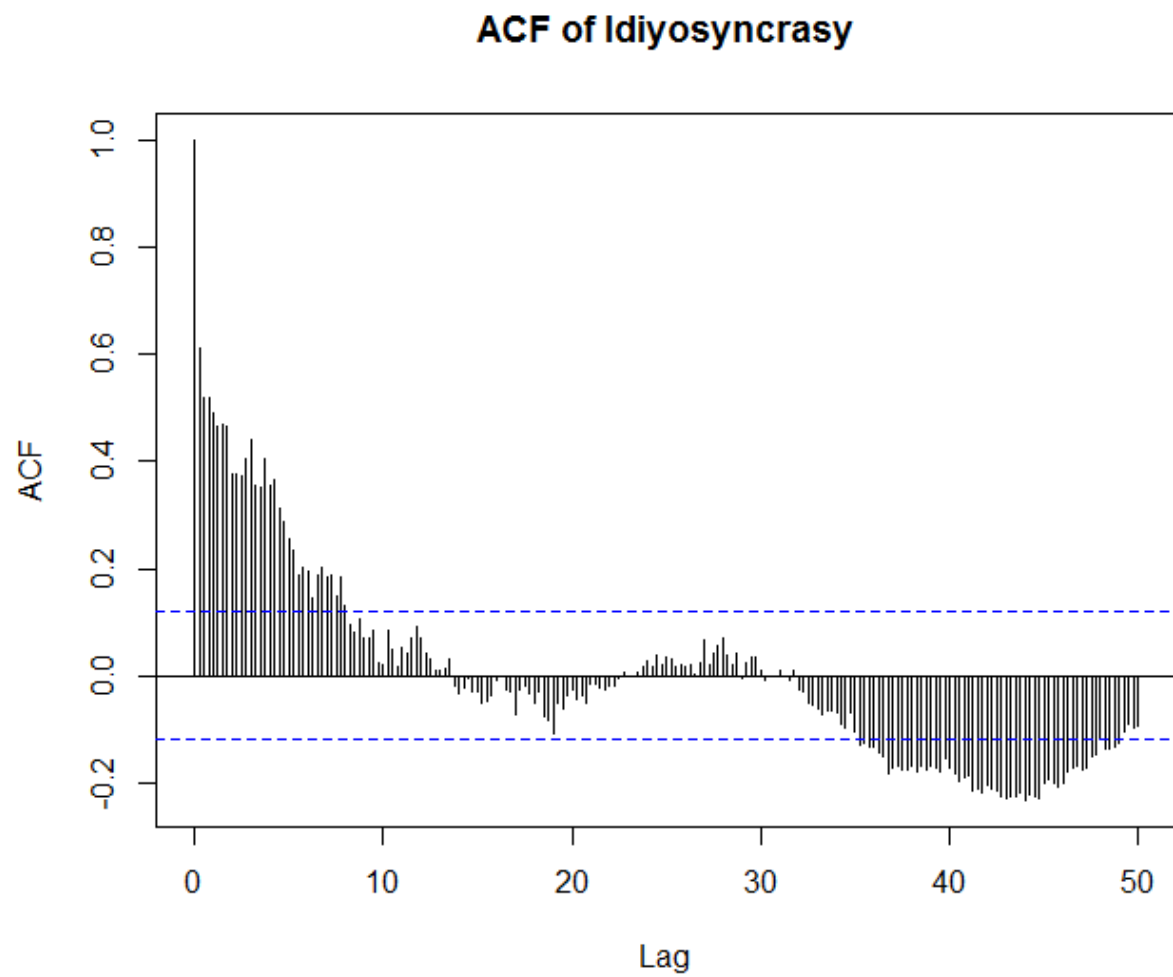


Figure 15: We notice significant ACF's at higher lags; this is an indicator of non-stationarity. ACF shows some periodicity, but not that significant and not nearly as pronounced as the GDP series.

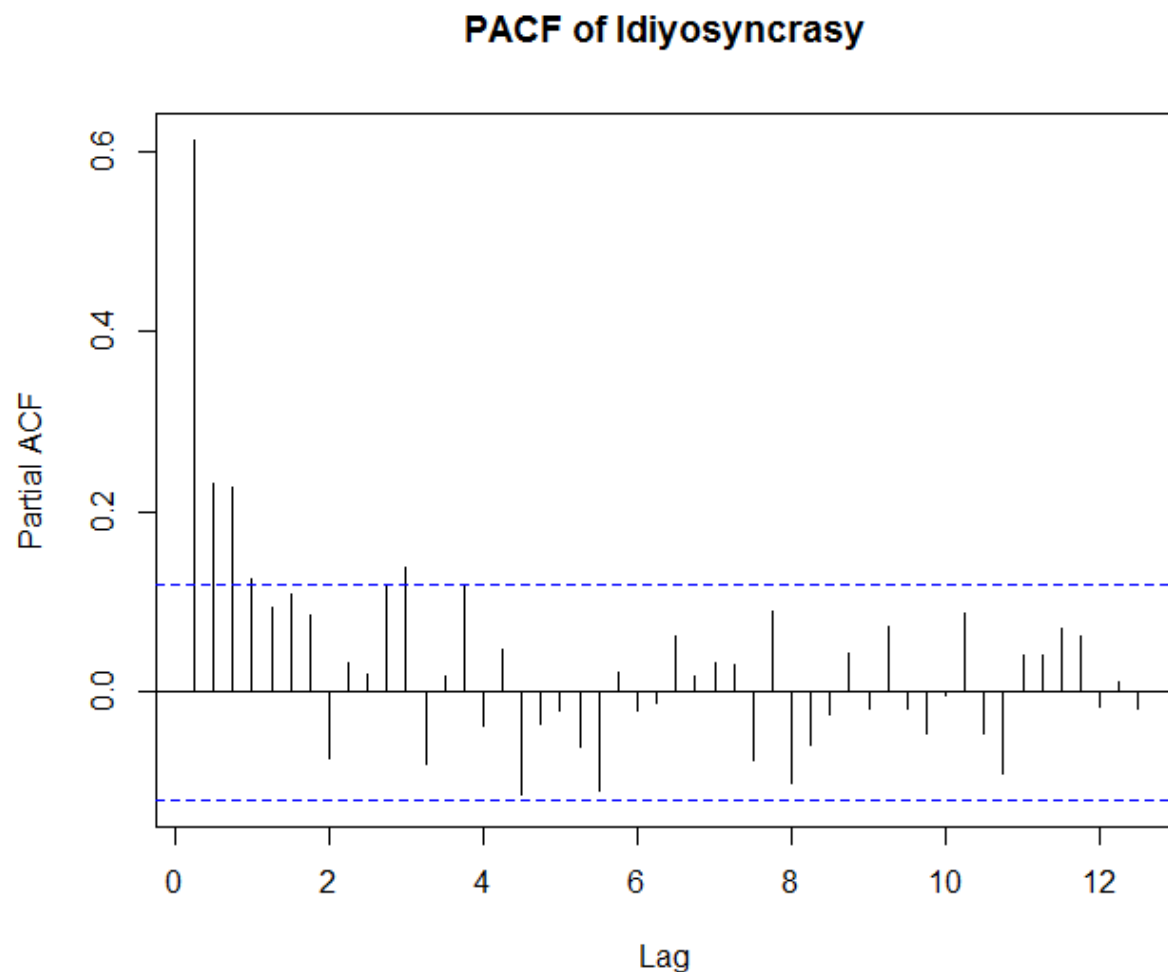


Figure 16: The PACF does not seem to have high values at higher lags, we do however, see a high PACF at lag 1 which is usually an indicator of non-stationarity.

From **Figure 2** and **Figure 15** we notice that the series is not stationary. Furthermore, from Figure 2 we note that there is no linear trend, as such, we do not even need to attempt to de-trend. What we need is a variance stabilizing transformation. We implement a basic square root transformation of the data to see if that helps with variance problem.

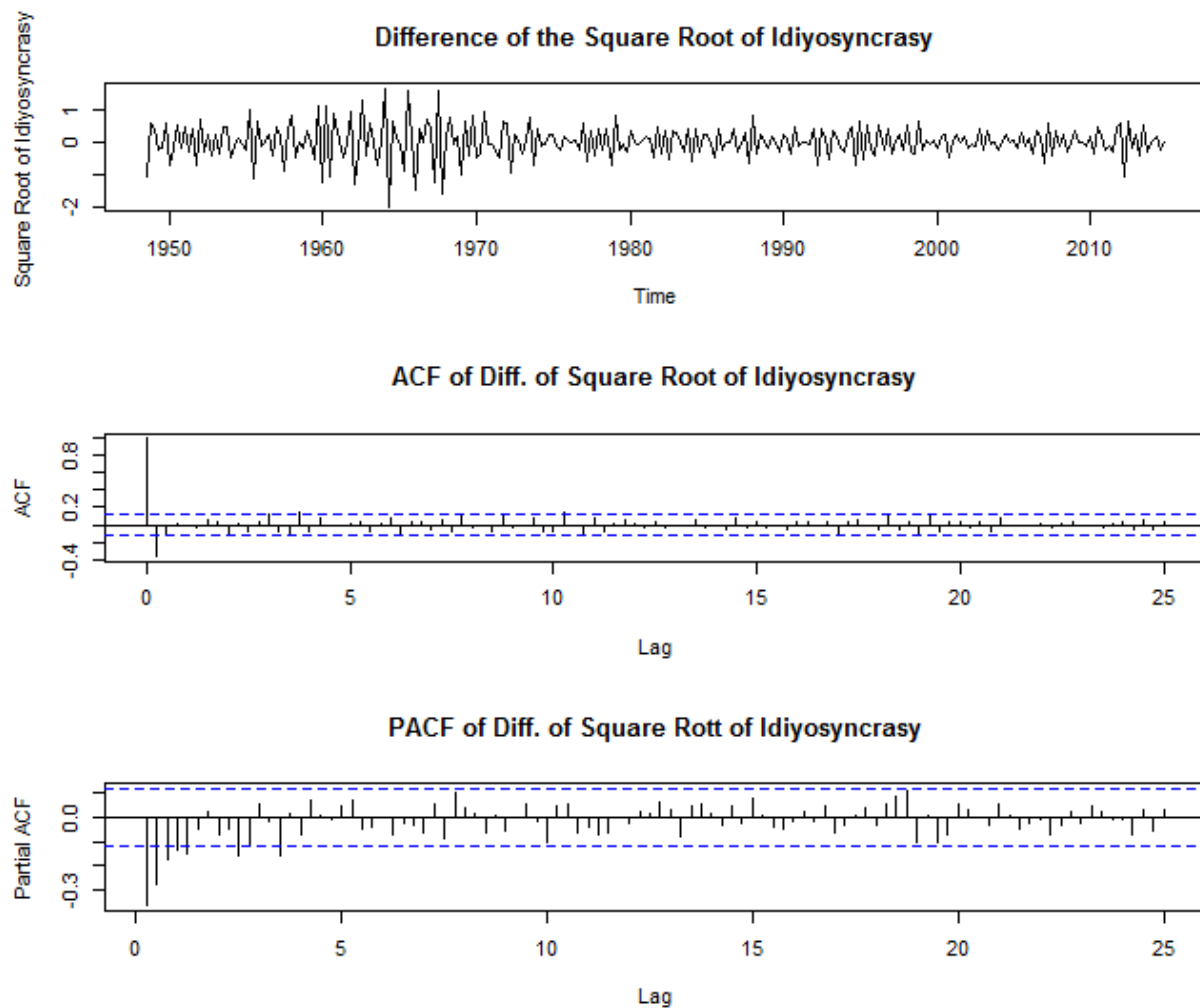


Figure 17: Plot 1 shows a better, more stable variance, however it is still not perfect. Plot 2 shows that there are still significant ACF values at high lags. Plot 3: the PACF does not look much different. We still have some issues with the uneven variance, but it is not too bad. What we can't see from the above plot, due to scaling, is that the mean of idiosyncrasy is no longer approximately constant. Thus we will need to take a difference to mitigate this, as de-trending is unlikely to help us.

Thus, we will take a difference to see if the mean problem can be resolved.

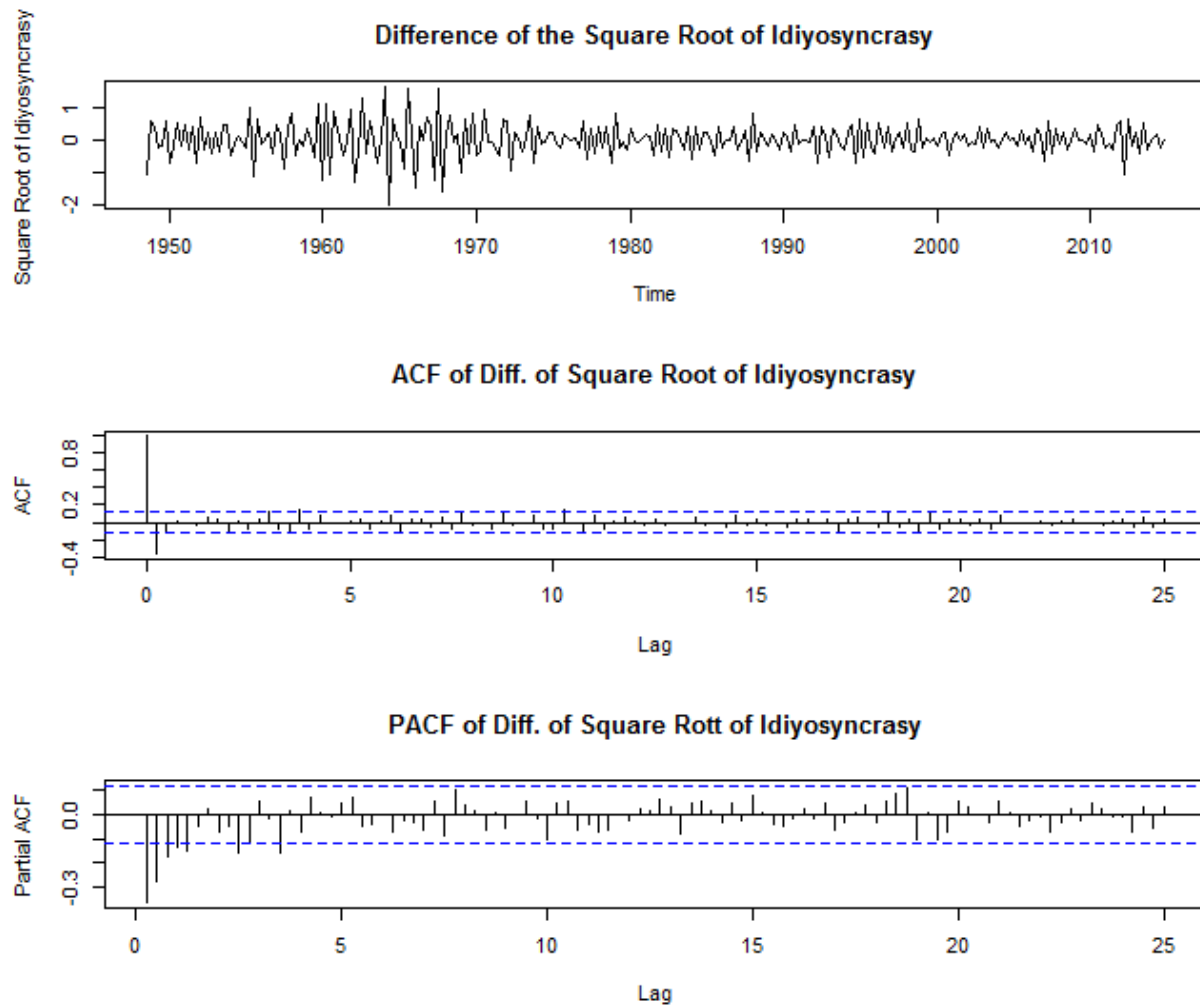


Figure 18: The first plot looks slightly better, there is still a slight issue with uneven variance but it is as good as we will likely ever make it. The second ACF is much better and the lags do tend to 0 or insignificance fairly quickly. The plot also shows clear evidence of an MA(2 or 3) model with no seasonality. The PACF plot shows clear evidence of an MA component as the values taper off, again no significant season trends were detected. We will re-rerun the same algorithm used before this time with no seasonality, and both p and q ranging from 0 to 3.

p	d	q	P	D	Q	S	AIC	AICc	BIC
---	---	---	---	---	---	---	-----	------	-----

0	1	0	0	0	0	0	-0.350	-0.342	-1.337
0	1	1	0	0	0	0	-1.344	-1.336	-2.317
0	1	2	0	0	0	0	-1.626	-1.618	-2.585
0	1	3	0	0	0	0	-1.646	-1.638	-2.593
1	1	0	0	0	0	0	-0.777	-0.770	-1.751
1	1	1	0	0	0	0	-1.482	-1.474	-2.441
1	1	2	0	0	0	0	-1.648	-1.640	-2.595
1	1	3	0	0	0	0	-1.627	-1.618	-2.560
2	1	0	0	0	0	0	-1.018	-1.010	-1.978
2	1	1	0	0	0	0	-1.556	-1.548	-2.502
2	1	2	0	0	0	0	-1.641	-1.632	-2.574
2	1	3	0	0	0	0	-1.633	-1.624	-2.553
3	1	0	0	0	0	0	-1.151	-1.143	-2.097
3	1	1	0	0	0	0	-1.580	-1.571	-2.513
3	1	2	0	0	0	0	-1.637	-1.628	-2.556
3	1	3	0	0	0	0	-1.627	-1.618	-2.533

Figure 19: AIC and BIC results. Three models to be tested are highlighted.

We will test these three models to see which is the best. As we can see on the basis of the information criteria these models are nearly identical, and as we have predicted before each has a large MA component.

RESIDUAL ANALYSIS

MODEL 1: (1,1,2)

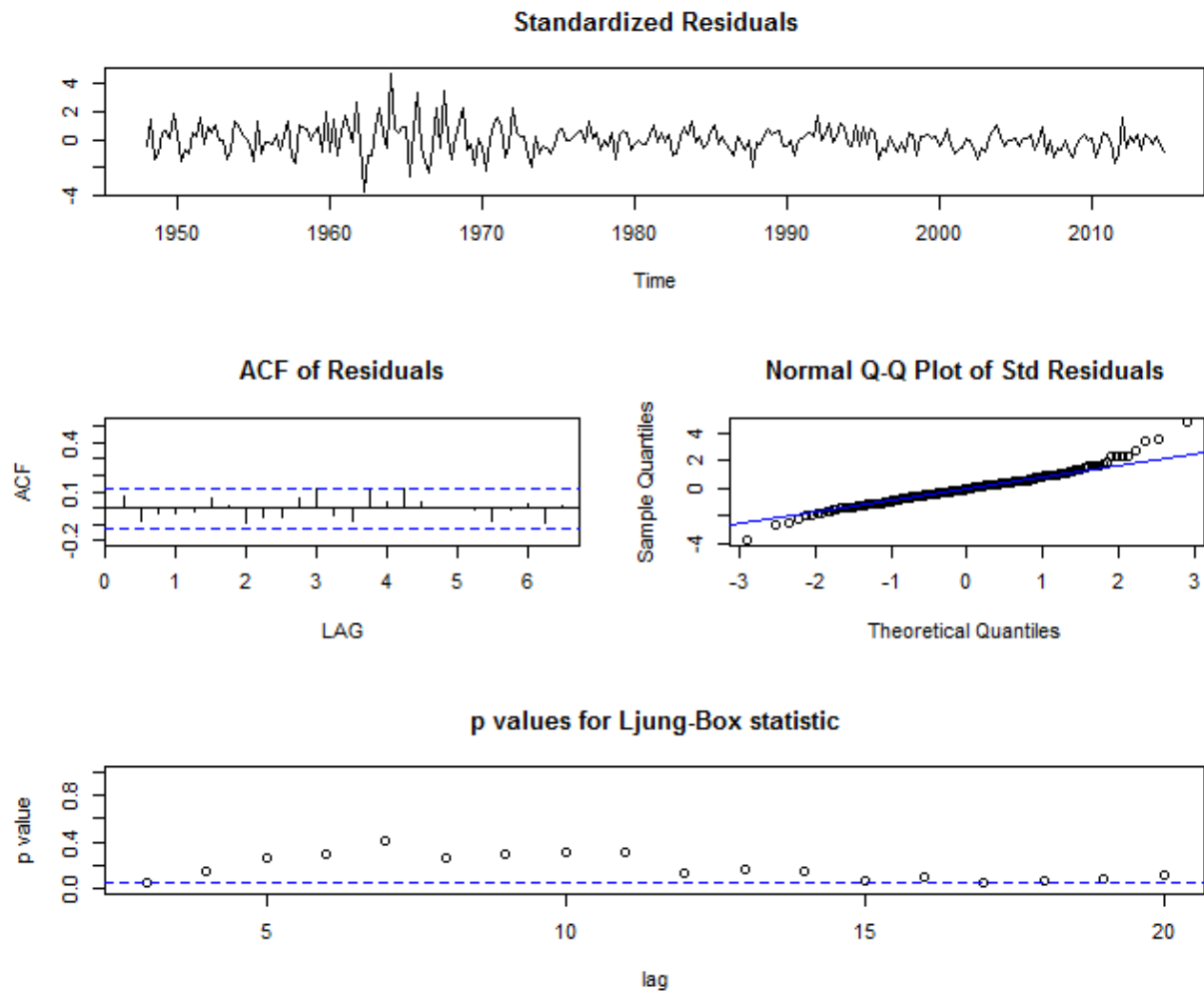


Figure 20: The standardized residuals do not necessarily look like white noise, due to the spike in the variance in the 1960's to 70's. However the ACF of the residuals looks good with few significant values. The normal Q-Q plot isn't great but we can forgive a little diverge from the linear in the beginning and ends of the interval. P values in the Ljung-Box statistics are all seem to be significant or on the verge of being significant.

This model is usable but not great, we will see if the other two are better.

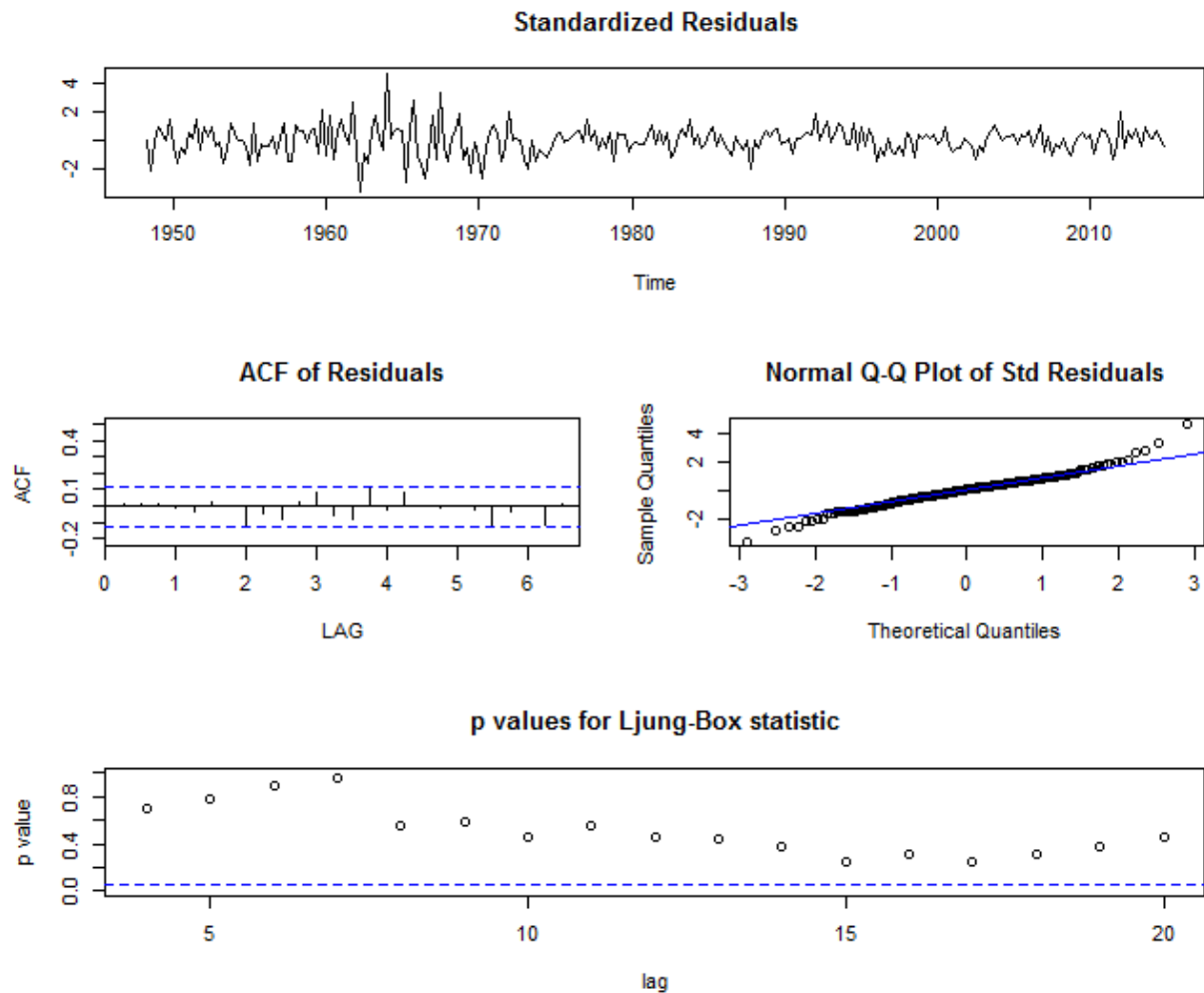


Figure 21: This model is better than the first. While the residual seem to have some non-constant variance just as in the first model. However, the normal Q-Q plot does seem to be slightly better indicating that the standardized residuals follow a normal distribution. The clear differentiator between models one and two are the p-values for the Ljung-Box statistic, all are significant and way above the blue dotted line.

If we had to make a decision between model 1 and 2, model 2 would be the clear winner.

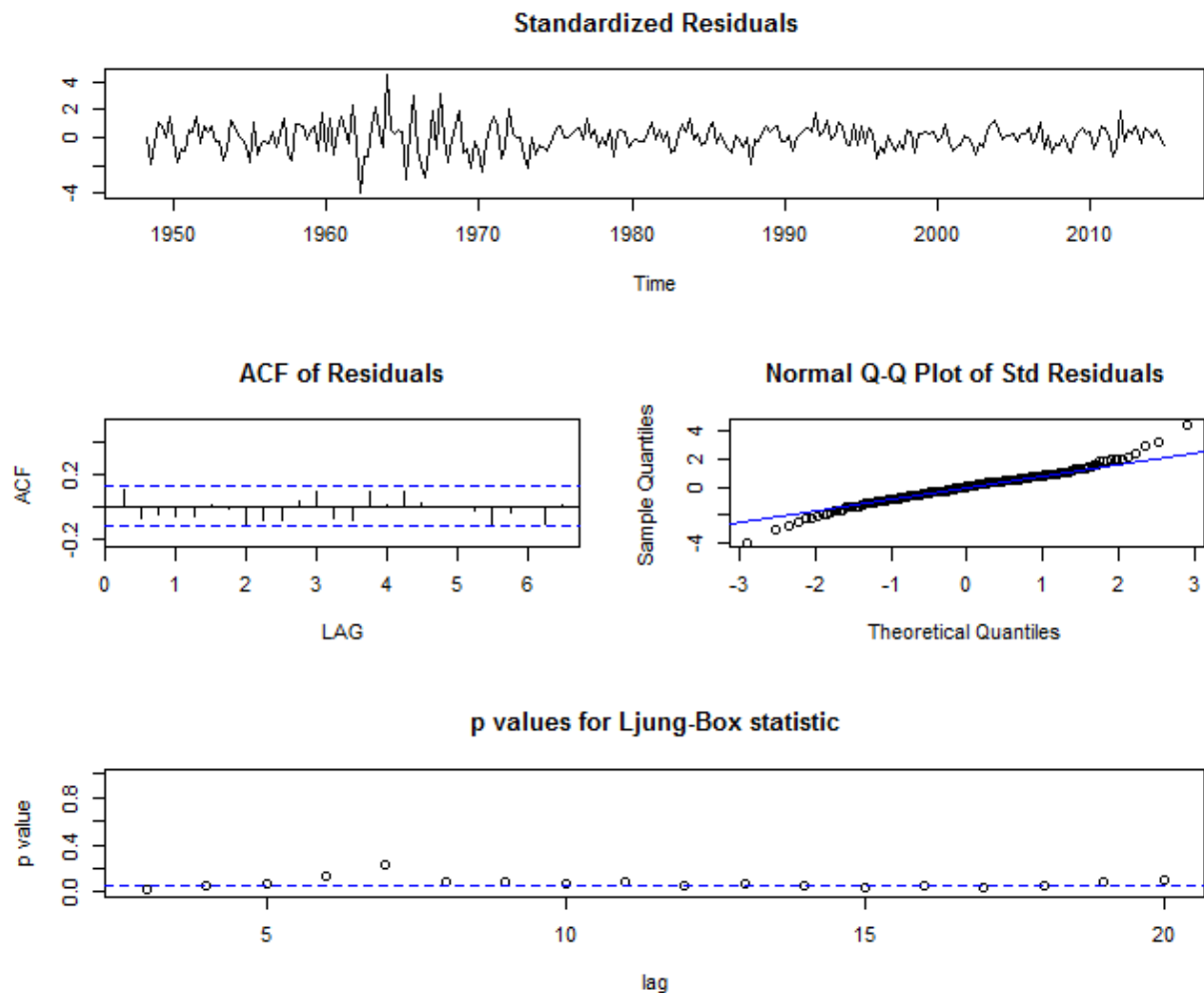


Figure 22: This model is strictly worse than model 2, the current favorite. The points of contention are the slightly less linear trends in the Normal Q-Q plots and largely insignificant p-values for the Ljung-Box Statistic.

Thus, model two is the clear favorite, as we saw from the ACF and PACF the model is (0,1,3). Even though the step with checking through multiple models and looking at the information criteria is not a strict requirement it is nice to know that it agrees with our eyeball analysis. We will now go into model detail of model 2 and use it to forecast.

FORECASTING

Model 2	ma1	ma2	ma3	constant
<i>Coeff.</i>	-1.591	0.4227	0.1683	0
<i>s.e.</i>	0.0617	0.111	0.0632	0

Note we use ψ as our market wide idiosyncrasy measure that we discussed in beginning. We also note that as we had to a square root of the data our model has to reflect that as well.

$$\sqrt{\psi_t}(1 - B) = (1 + \theta_1 B + \theta_2 B^2 + \theta_3 B^3)\omega_t$$

And plugging in our estimates

$$\sqrt{\psi_t}(1 - B) = (1 - 1.591B + 0.4227B^2 + 0.1683B^3)\omega_t$$

Gives us our finished model.

TEST SET EVALUATION

In order to better evaluate our model, I censored the last four quarter's worth of data and predicted them based on the incomplete data set and our model. These are the results.

Date	Actual	Predicted	%Difference	s.e.
1/1/2014	1.13783	1.06948	6.01%	0.26388
4/1/2014	1.07548	1.05274	2.11%	0.28515
7/1/2014	1.19204	1.04974	11.94%	0.29302
10/1/2014	1.06750	1.04843	1.79%	0.29922

As we can see the model did decently well. We now move onto predicting the future. Even with the problems we have had with stationarity being only 1-12% off is not too bad.

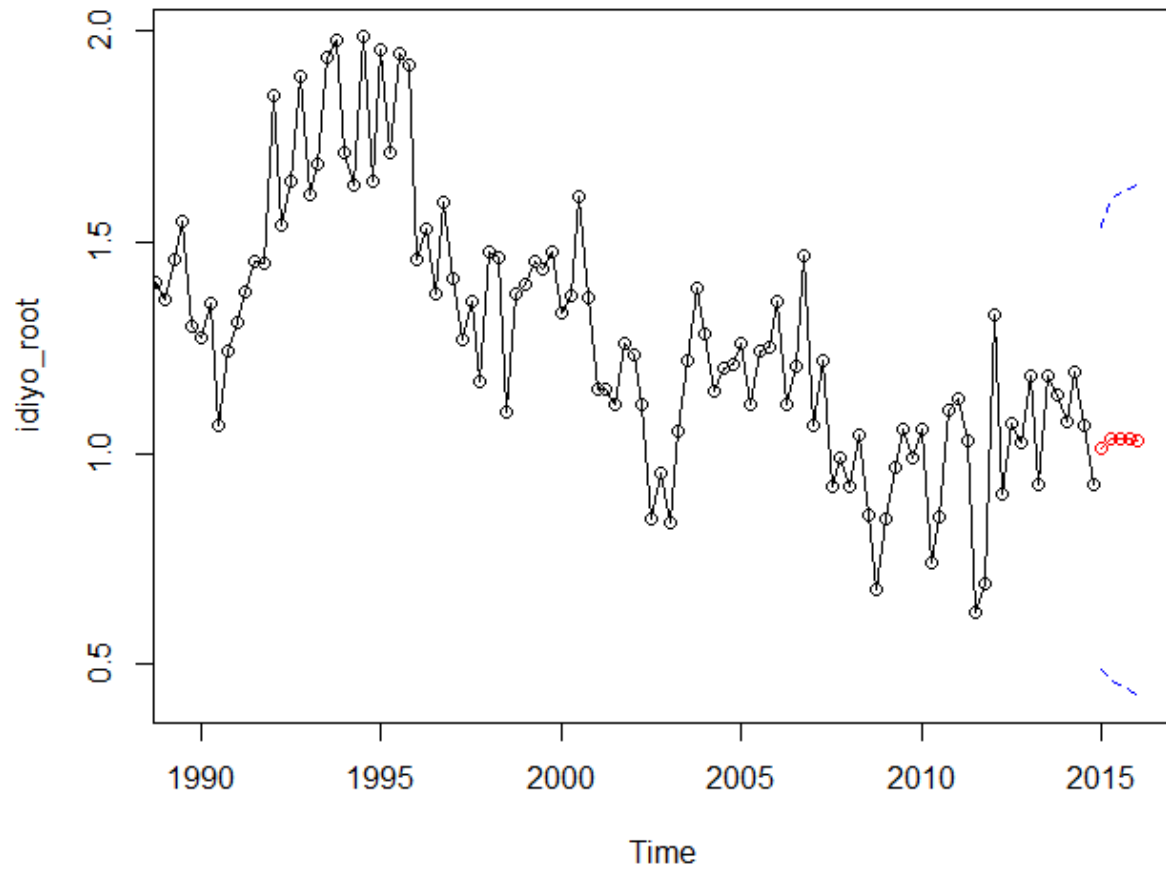


Figure 23: The prediction seems reasonable somewhat reasonable. There seems to be an issue with our prediction intervals, as they seem to be quite wide in comparison with the data.

We now move on to Frequency domain analysis to study inter-relationships the two series might have.

PERIODOGRAMS

We first look at the smoothed periodograms.

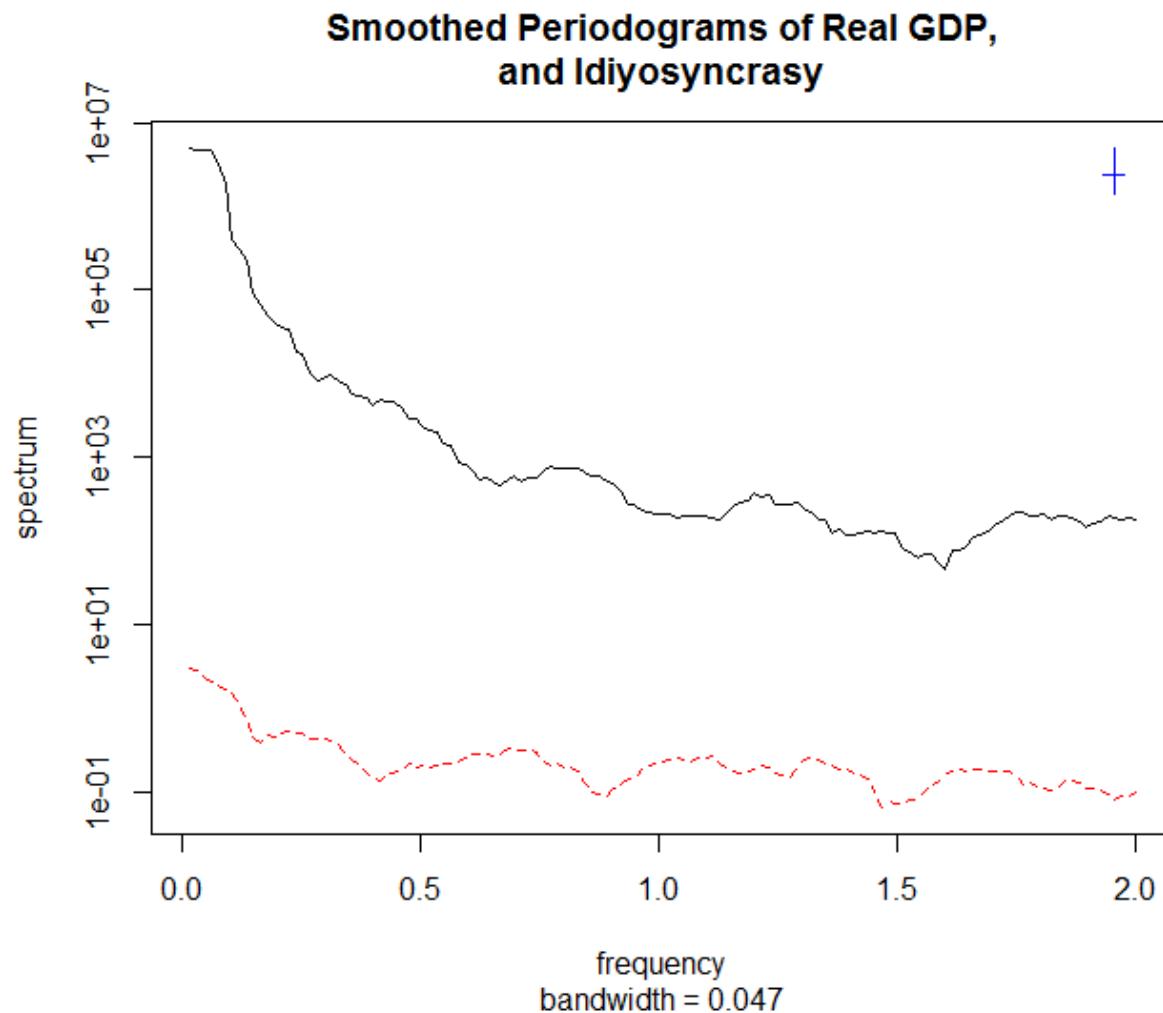


Figure 24; Real GDP – Black, Idiosyncrasy –Red. Smoothing was done with a factor of 11.

We see some interesting things, there seems to be a pattern that the two series follow. Particularly at lower frequencies and at frequency of around 1.6 we see a small dip in the powers of both series. We will look at the coherence plot to get a better idea.

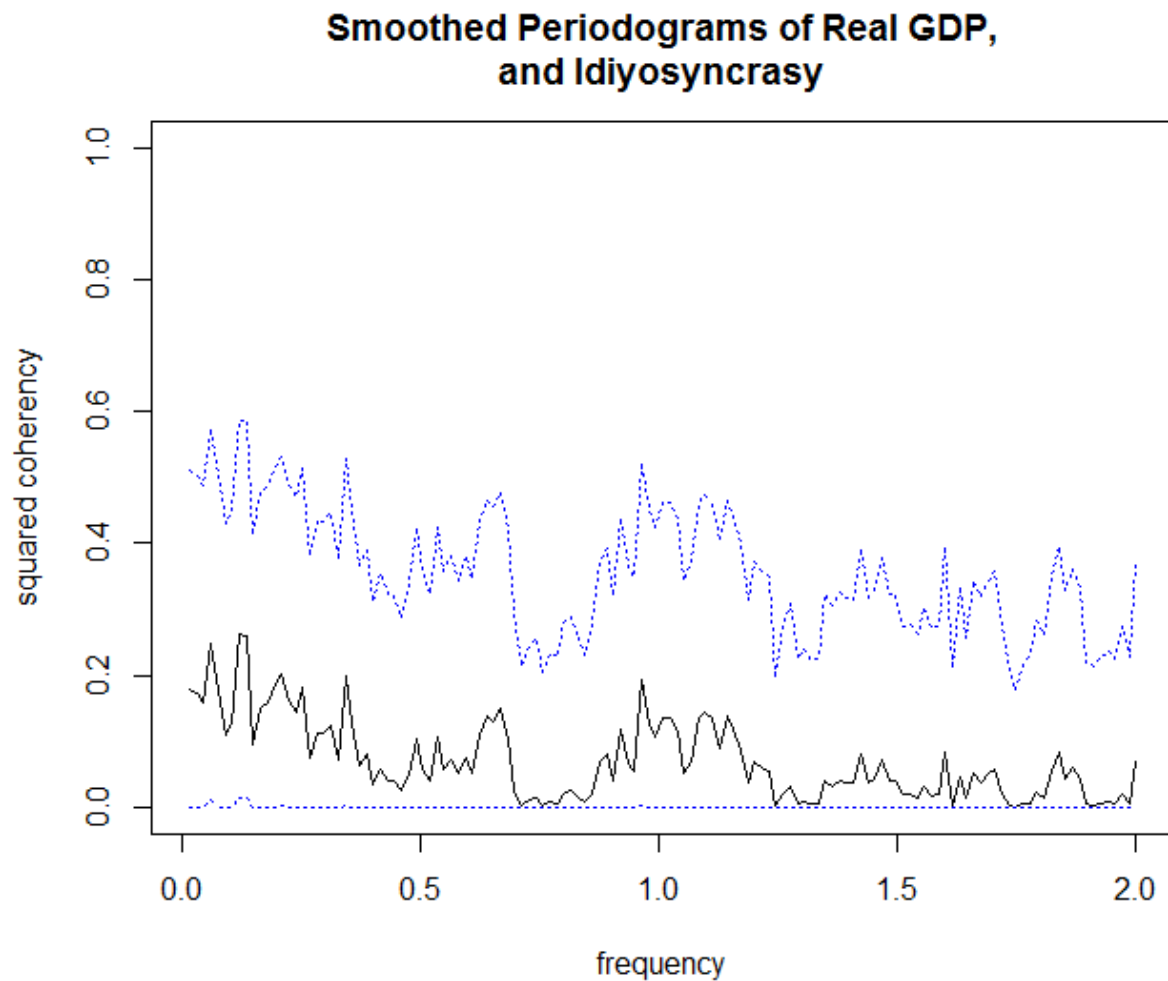


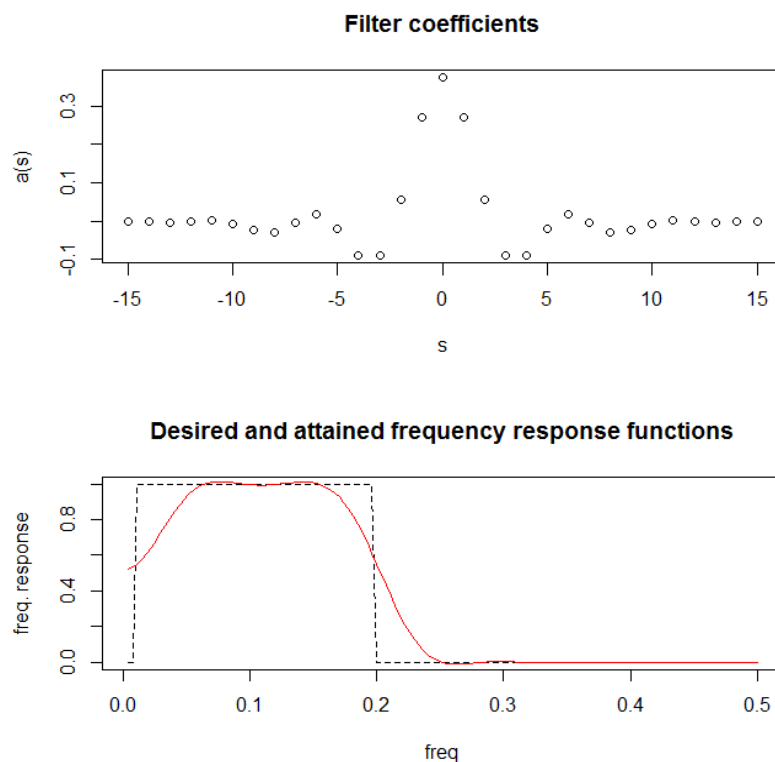
Figure 25: Squared Coherence plot. Smoothing factor of 11 was used.

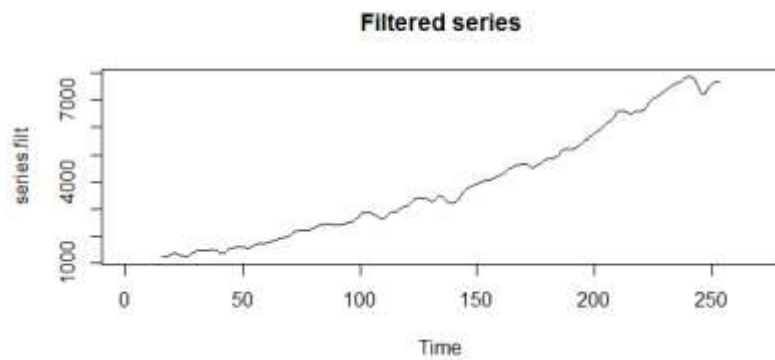
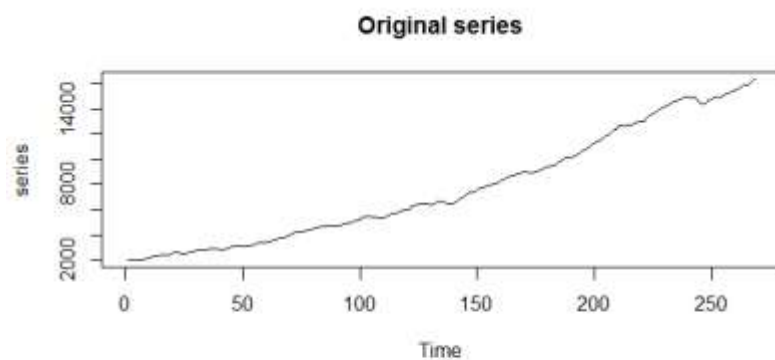
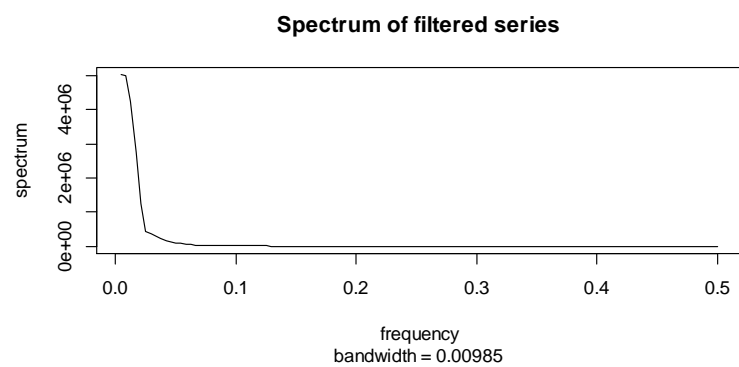
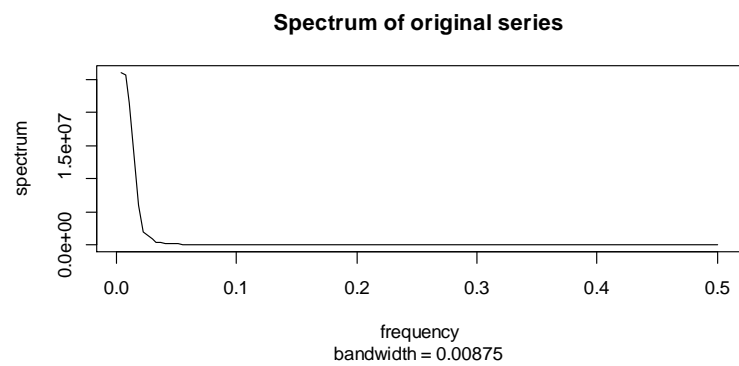
Coherence is strongest at the lower frequencies, it also has an uptick at the mid frequency and then plateaus again. This might suggest that there is a relationship between Real GDP and Idiosyncrasy. However, as the confidence interval do include zero for a fairly large amount of values, which can make whatever trend we think we are seeing, illusory.

For further analysis we might want to isolate the most important frequencies for each series and continue our analysis from there.

OPTIMAL FILTERING REAL GDP SERIES

From the plot in Figure 25, we can see the power of this series is highest for very low frequencies. This makes sense as with low frequencies correspond to very infrequent events, in the order of years for this data series. From what we know of economics, is that short-term growth is not variable while long-term growth outcomes are much more so, obviously recessions do not occur every third quarter. Thus we want to restrict our analysis to long-term trends. This is what the data is telling us to do, and what is most interesting in terms of what we understand of economics. The frequency band we are most interested in the one that is less than 0.2. This corresponds to a period of five years. We use R to extract our filter. In our analysis we found that using $L=9$ and $M=32$ has worked well.



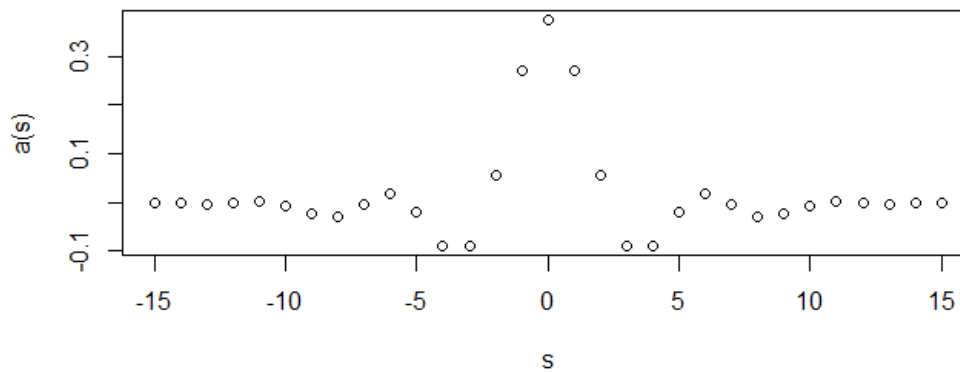


Figures 26-29: Are the printouts of the newly filtered series.

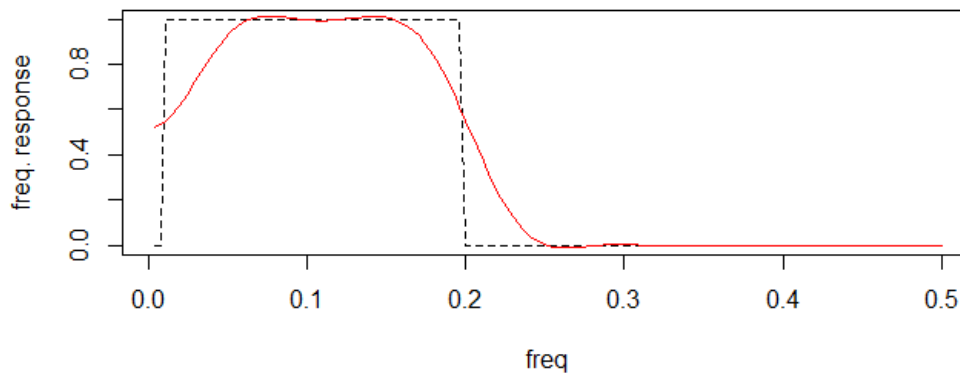
OPTIMAL FILTERING IDIOSYNCRASY SERIES

We will do the same to the idiosyncrasy series as we did to the Real GDP series but we will use a max frequency of 0.5 instead of 0.2. We do this because the periodogram of the idiosyncrasy series does not decrease as quickly after hitting its max as the Real GDP series. A frequency of 0.5 corresponds to a period of 2 years.

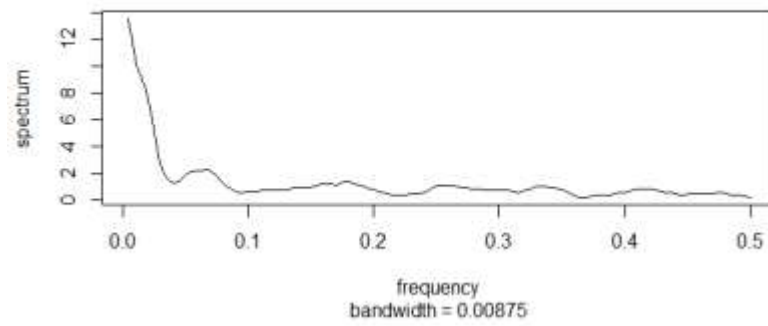
Filter coefficients



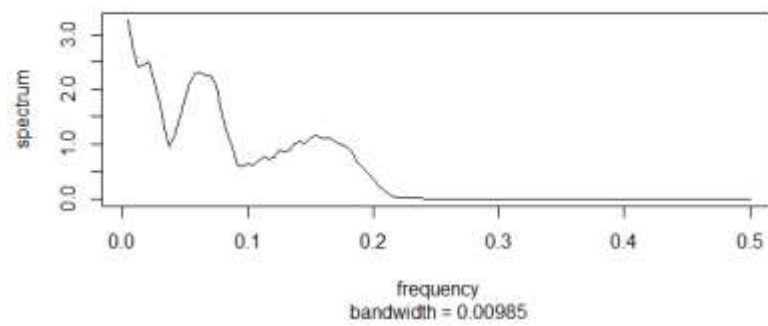
Desired and attained frequency response functions



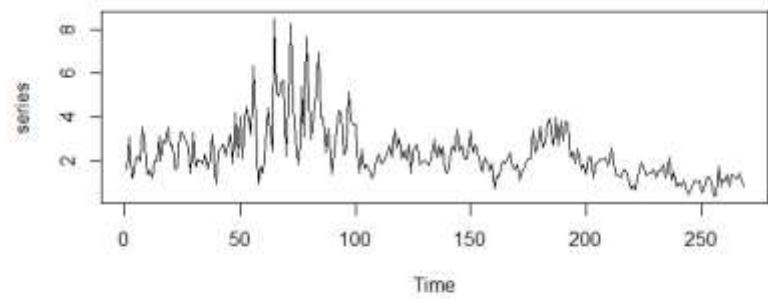
Spectrum of original series



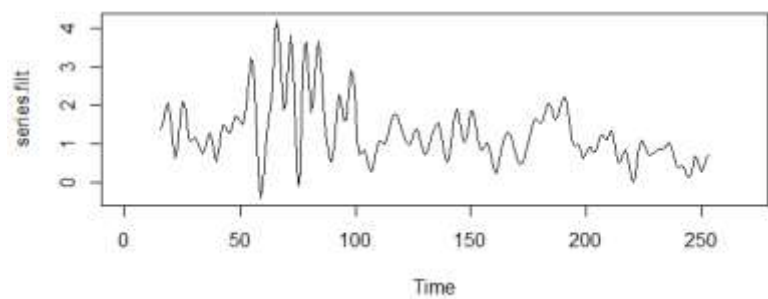
Spectrum of filtered series



Original series



Filtered series



Figures 30-31: Are the printouts of the newly filtered series. We note we have eventuated the powers at the longer frequencies and the resulting plot of the Filtered Series looks much smoother.

We see that both series have a lot of their power coming from low frequencies and thus long-term trends (long periods). This suggests that the two series might be in some way related. In order to figure out what that relationship is we do a lagged regression, which sometimes referred to as an impulse response analysis.

IMPULSE RESPONSE MODEL

As we have stated in the opening paragraphs we want to predict GDP with idiosyncrasy. Thus we will use GDP as output and Idiosyncrasy as input. This method of regression use the least squares and combines it with spectral analysis to simplify the calculations. We will use the transformed stationary variables or real GDP and idiosyncrasy.

Fitting the Model,

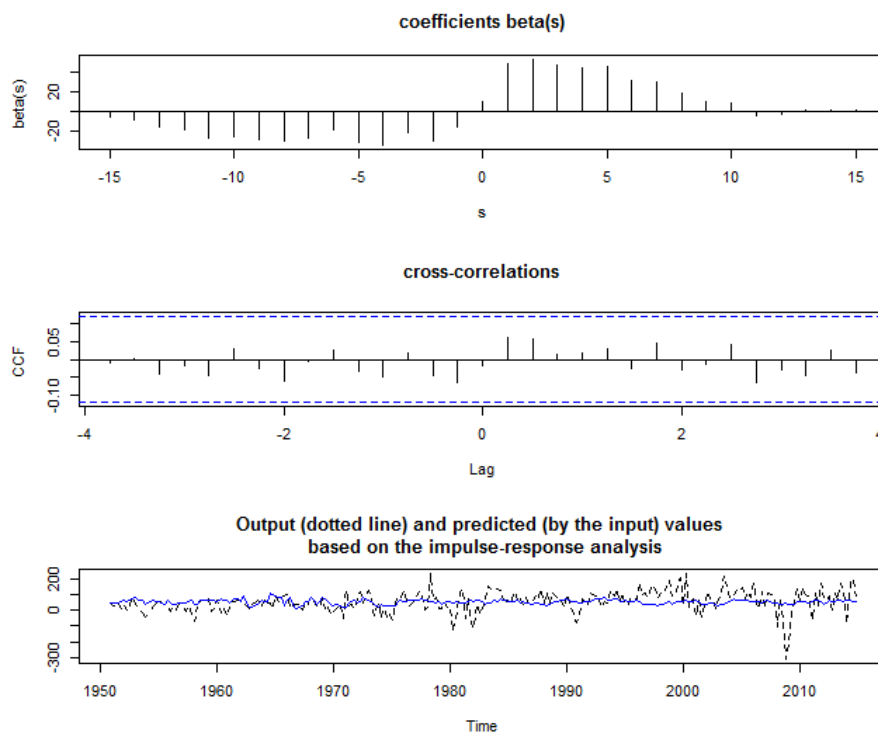


Figure 32: While seemingly significant we are not able to accurately predict the swings in real GDP

lag -s	Coefficient- β_s
1	47.4248
2	52.5029
3	47.1957
4	43.6501
5	44.8797
6	30.9124
7	29.5742
8	18.2792

Figure 33: Here a table of the coefficients that were provided by the LagReg function. L=15, M=32, threshold=9. Also, $\alpha = 53.98753$ and $MSE = 4492.055$

The values of L, M and threshold were chosen by trial and error while considering the MSE. Using M=64 would decrease the MSE to around 4444.91, but we felt that omitting so much data for a paltry decrease in the MSE was not worth it.

The prediction equation is

$$realgdp(t) - realgdp(t-1) = \alpha + \sum_{s=0}^8 \beta_s * \left(\sqrt{idiosyncrasy(t-s)} - \sqrt{idiosyncrasy(t-s-1)} \right)$$

It is important to note that we are not modelling as real GDP directly but rather its difference. We are also not using idiosyncrasy but the difference of its root.

The most likely culprit as to why idiosyncrasy was not a great predictor is that there are likely factors that we have omitted that influence GDP greatly, government spending or the interest rate for instance. In order to answer our original question fully we would need to control for all those other variables and then repeat our analysis. This would likely require some sort of multi-variable time series method.

IMPULSE RESPNSE ON FILTERED VARIABLES

In the Optimal Filtering sections we isolated the long term interesting frequencies and found filters which we can use to find series which only exhibit these long term trends. As we are most interested in these long term trends we will use the filtered series and try performing another lagged regression. Like before, we need to take differences of both series to make them stationary.

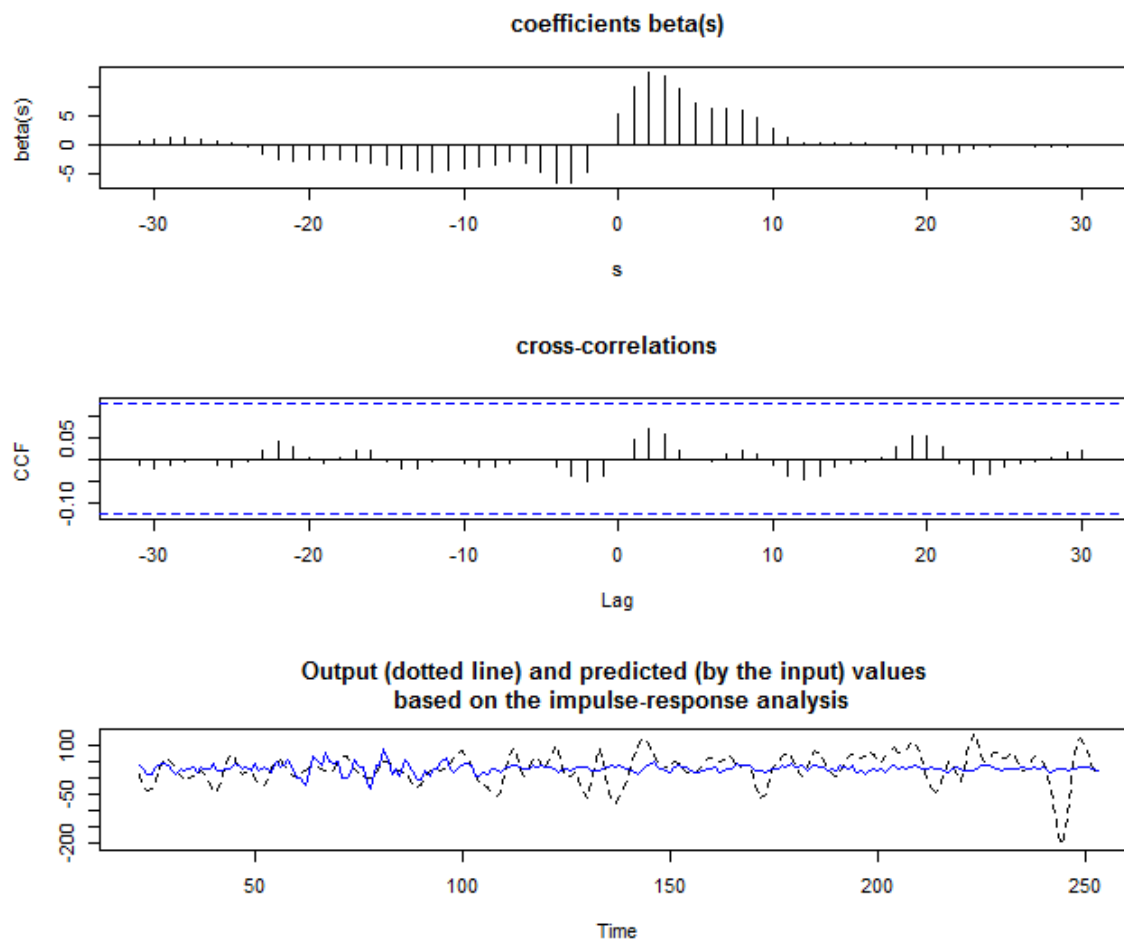


Figure 34: Seems better but far from perfect. The fit is decent in the earlier time periods but from 1980's and onwards the fit grows significantly worse.

lag -s	Coefficient-
1	10.0831
2	12.47746
3	11.94638
4	9.533105

Figure 35: A table of coefficients that were provided by the LagReg function. $L=15$, $M=32$, threshold=9. Also, $\alpha = 27.08332$ and $MSE = 2219.417$

The fit clearly improves for the smoothed series but the fit is far from perfect. If we tried using this model in 2007, just before the financial crisis, we would fail miserably. As the recessions

tend to be the most interesting phenomena in economics we consider this model somewhat unsatisfactory. As stated before the issue is likely with missing variables.

CONCLUSION

We began with the two series Real GDP and Idiosyncrasy and we wanted to answer the question whether we can predict Real GDP with idiosyncrasy. We first noticed that neither series was stationary. We had to take a difference of the Real GDP series in order to stabilize the mean. We then fit multiple SARIMA models on the GDP series, and to determine which is best, we took into account the information criterions as well as the residual analysis. In the end, the best model for Real GDP ended up being a non-seasonal ARMA, $(1 - 0.3327B - 0.173B^2)(1 - B)x_t = \omega_t$. We then took a look at the idiosyncrasy series and noticed that variance was not constant. In order to stabilize the variance, we took a square root transformation of the data. We then looked at this new series and noticed that it did not have a constant mean, in order to remedy this, we took a difference, and this resulting series looked approximately stationary. Just like for the Real GDP series we fit multiple ARIMA models to determine which is best. We decided on the model $\sqrt{\psi_t}(1 - B) = (1 - 1.591B + 0.4227B^2)\omega_t$.

We then moved on to the Frequency Domain. We looked at the periodogram of the two series as well as their coherence. Each showed that a lot of the power is concentrated at lower frequencies, which corresponds to higher periods. We noted that this makes sense economically and proceeded with our analysis.

We fit a lagged regression to determine the relationship between the two series. After some tinkering with the values of L, M and threshold we found a reasonable model. However, according to plot 3 of **Figure 32**, we noticed that our model is a poor predictor of Real GDP. We also performed a lagged regression on the filtered version of our variables and ended up with a similar, unsatisfactory result. This is probably because we are missing very significant drivers of GDP such as the interest rate and government spending among others.

BIBLIOGRAPHY

Morck, R. (2013). R2 and The Economy. *The Annual Review of Financial Economics* , 1.

Schumpeter, J. (1947). *Capitalism, Socialism and Democracy*.

Schumway, R. H. *Timer Series Analysis and Its Applications*.

US. Bureau of Economic Analysis. (2015, April 1). *Gross Domestic Product [GDP]*. Retrieved from FRED, Federal Reserve Bank of St. Louis: <https://research.stloiusfed.org/fred2/series/GDP/>

Wiens, D. (2015, 2 1). *Stat 479 Lecture Notes*. Retrieved from Stat 479 Course Website:
<http://www.mathstat.ualberta.ca/~wiens/stat479/stat479.html>

Zhang, L. (2014). Stock Market Idiyosyncrasy. *Job Market Paper* , 4.

APPENDIX: R-CODE

(Schumway)

(Wiens, 2015)

```
<<>>=
```

```
## initialize R and Data
```

```
library(astsa)
```

```
setwd("C:/Users/Bohdan/OneDrive/Stat 479 Project")
```

```
data<-read.csv("data_export_2.csv")
```

```
data_2<-ts(data,start=c(1948,1),frequency=4)
```

```
real_gdp<-data_2[,5]
```

```
idiyo<-data_2[,2]
```

```
@
```

```
<<>>=
```

```
# introductory plots
```

```
par(mfrow=c(2,1))
```

```
ts.plot(real_gdp,main="Real GDP Over Time (2008 base)", ylab="GDP($)",xlab="Time")
```

```
ts.plot(idiyo,main="Idiosyncrasy Over Time",ylab="Stock Return Idiyosyncrasy",xlab="Time")
```

```
@
```

```
<<>>=
```

```
#ACF and PACF of Real GDP
```

```
acf(real_gdp,lag.max=200,main="ACF of Real GDP")
```

```
pacf(real_gdp,lag.max=50,main="PACF of Real GDP")
```

```
@
```

```
<<>>=
```

```
# detrending
```

```
fit<-lm(real_gdp~time(real_gdp))
```

```
real_gdp_detrend<-fit$resid
```

```
par(mfrow=c(3,1))
```

```
ts.plot(real_gdp_detrend,main="Real Detrended GDP Over Time (2008 base)",
```

```
ylab="GDP($) Residuals",xlab="Time")
```

```
acf(real_gdp_detrend,lag.max=200,main="ACF of Real GDP")
```

```
pacf(real_gdp_detrend,lag.max=50,main="PACF of Real GDP")
```

```
@
```

```
<<>>=
#differencing real GDP
real_gdp_diff<-diff(real_gdp)
par(mfrow=c(3,1))
ts.plot(real_gdp_diff,main="First Difference of Real GDP Over Time (2008 base)",
        ylab="diff of GDP($)",xlab="Time")
acf(real_gdp_diff,lag.max=50,main="ACF of Real GDP")
pacf(real_gdp_diff,lag.max=50,main="PACF of Real GDP")
@
```

```
<<>>=
#fitting SARMA model
out<-matrix(ncol=10)
d=1
D=0
S=2
Q=0
for (p in 1:2) {for (q in 0:2) {for (P in 0:2){
  fits<-sarima(real_gdp,p,d,q,P,D,Q,S,details=F)
  out<-rbind(out,c(p,d,q,P,D,Q,S,fits$AIC,fits$AICc,fits$BIC))
}}}
out<-out[-1,]
colnames(out)<-c("p","d","q","P","D","Q","S","AIC","AICc","BIC")
write.csv(out,file="Real_GDP_Model.csv")
fit1<-sarima(real_gdp,1,1,0,1,0,0,2,details=F)
fit2<-sarima(real_gdp,1,1,0,0,0,0,2,details=F)
fit3<-sarima(real_gdp,2,1,0,0,0,0,0,details=F)
@
```

```
<<>>=
#forecasting
real_gdp_actual<-ts(real_gdp[264:268],start=c(2014,1),frequency=4)
real_gdp_test<-ts(real_gdp[1:264],start=c(1948,1),frequency=4)
forecast_test<-sarima.for(real_gdp_test,4,0,1,3)
```



```

prediction_summary<-cbind(real_gdp_actual,forecast_test$pred,forecast_test$sse)
prediction_summary<-prediction_summary[-5,]
write.csv(prediction_summary,file="real_gdp_pred_summary.csv")
forecast<-sarima.for(real_gdp,5,2,1,0)
@

```

Idiyosyncrasy

```
<<>>=
```

#ACF and PACF of Real GDP

```

acf(idiyo,lag.max=200,main="ACF of Idiyosyncrasy")
pacf(idiyo,lag.max=50,main="PACF of Idiyosyncrasy")
@

```

```
<<>>=
```

#square root transform idiyo

```

idiyo_root<-sqrt(idiyo)
par(mfrow=c(3,1))
ts.plot(diff(idiyo_root),main="Square Root of Idiyosyncrasy",
        ylab="Square Root of Idiyosyncrasy",xlab="Time")
acf(diff(idiyo_root),lag.max=100,main="ACF of Square Root of Idiyosyncrasy")
pacf(idiyo_root,lag.max=100,main="PACF of Square Rott of Idiyosyncrasy")
@

```

```
<<>>=
```

```

idiyo_root_diff<-diff(sqrt(idiyo))
par(mfrow=c(3,1))
ts.plot(diff(idiyo_root_diff),main="Difference of the Square Root of Idiyosyncrasy",
        ylab="Square Root of Idiyosyncrasy",xlab="Time")
acf(idiyo_root_diff,lag.max=100,main="ACF of Diff. of Square Root of Idiyosyncrasy")
pacf(idiyo_root_diff,lag.max=100,main="PACF of Diff. of Square Rott of Idiyosyncrasy")
@

```

```
<<>>=
```

```
#fitting the idiosyncrasy ARIMA model
```

```
out<-matrix(ncol=10)
```

```
d=.25
```

```
D=0
```

```
S=0
```

```
Q=0
```

```
for (p in 0:3) {for (q in 0:3) {for (P in 0:0){
```

```
  fits<-sarima(idiyo_root,p,d,q,P,D,Q,S,details=F)
```

```
  out<-rbind(out,c(p,d,q,P,D,Q,S,fits$AIC,fits$AICc,fits$BIC))
```

```
}}}
```

```
out<-out[-1,]
```

```
colnames(out)<-c("p","d","q","P","D","Q","S","AIC","AICc","BIC")
```

```
out
```

```
write.csv(out,file="Idiyo.csv")
```

```
fit1<-sarima(idiyo_root,1,1,2,0,0,0,details=F)
```

```
fit2<-sarima(idiyo_root,0,1,3,0,0,0,details=F)
```

```
fit3<-sarima(idiyo_root,0,1,2,0,0,0,details=F)
```

```
@
```

```
<<>>=
```

```
# Forecasting
```

```
idiyo_root_actual<-ts(idiyo_root[264:268],start=c(2014,1),frequency=4)
```

```
idiyo_root_test<-ts(idiyo_root[1:264],start=c(1948,1),frequency=4)
```

```
forecast_test<-sarima.for(idiyo_root_test,4,0,1,3)
```

```
prediction_summary<-cbind(idiyo_root_actual,forecast_test$pred,forecast_test$se)
```

```
prediction_summary<-prediction_summary[-5,]
```

```
write.csv(prediction_summary,file="Idiyo_predict_summary.csv")
```

```
forecast<-sarima.for(idiyo_root,5,0,1,3)
```

```
@
```

```

<<>>=
# spectra
data.ts<-cbind(real_gdp,idiyo)
L=11
spec.per=spec.pgram(data.ts,kernel("daniell",(L-1)/2),main="Smoothed Periodograms of Real GDP,
and Idiyosyncrasy")
spec.coh=spec.pgram(data.ts,kernel("daniell",(L-1)/2),plot.type="coh", main="Smoothed Periodograms of Real
GDP,
and Idiyosyncrasy")
x<-cbind(spec.coh$freq,spec.coh$coh,spec.coh$spec)
write.csv(x,file="coherence_values.csv")
@

```

```

<<>>=
# Optimal Filter Real GDP
out<-SigExtract(real_gdp,L=9,M=32,max.freq=0.2)
out<-SigExtract(idiyo,L=9,M=32,max.freq=0.5)
@

```

```

<<>>=
#impulse response on stationary variables
coef<-LagReg(input=idiyo_root_diff,output=real_gdp_diff,L=15,M=64,threshold=9)
@

```

```

<<>>=
# impulse response on filtered variables-getting filtered
out_gdp<-na.omit(out_gdp)
out_idiyo<-na.omit(out_idiyo)
out_gdp_diff<-diff(out_gdp)
out_idiyo_diff<-diff(out_idiyo)
@

```

```

<<>>=
#lagged regression on filtered variables
coef<-LagReg(input=out_idiyo_diff,output=out_gdp_diff,L=15,M=64,threshold=8)
@

```