Michael Gleyzer, Mathew Witek and Harish Pillai
December 12,2019
960:486:01 - Comp Graph & Applied Stats

Objective/Purpose

The purpose of this project was to study the application of Bayesian methods to the forecasting of economic data. In this project we decided upon Bayesian Structural Time Series as our method of interest. To test the accuracy of Bayesian Structural Time Series we decided to find out whether this method could successfully predict the rate of loan delinquencies during the Great Recession.

Methodology

1. First we took delinquency rate data from the years **2000-2007** (32 quarters) for all loans(commercial,residential...etc.) and fitted a **Bayesian structural time series model**(including both trend and seasonal components)
2. Then we predicted data for the next 8 quarters with a **95% credible interval** for the forecast
3. Next we graphed the actual data from the recession to visually compare the prediction accuracy of the forecast
4. We then repeated the same procedure but only for residential loan delinquency data

Tools

We used the R programming language to manipulate financial time series data on the market crash of 2007. Rather than just doing regular time series data analysis, we employed the **BSTS** package within the R library to construct a Bayesian model of the time series data.

Datasets

While preparing for this project, we found a plethora of datasets exploring the situation centered around the mortgage crisis in 2007. Eventually, we decided to use the data on delinquency rates on loans and leases from the Federal Reserve. The reason for choosing this data was that it broke up the Years into quarters, and also the loans into residential, commercial, and farmland points, which we felt provided a clearer picture of the data.

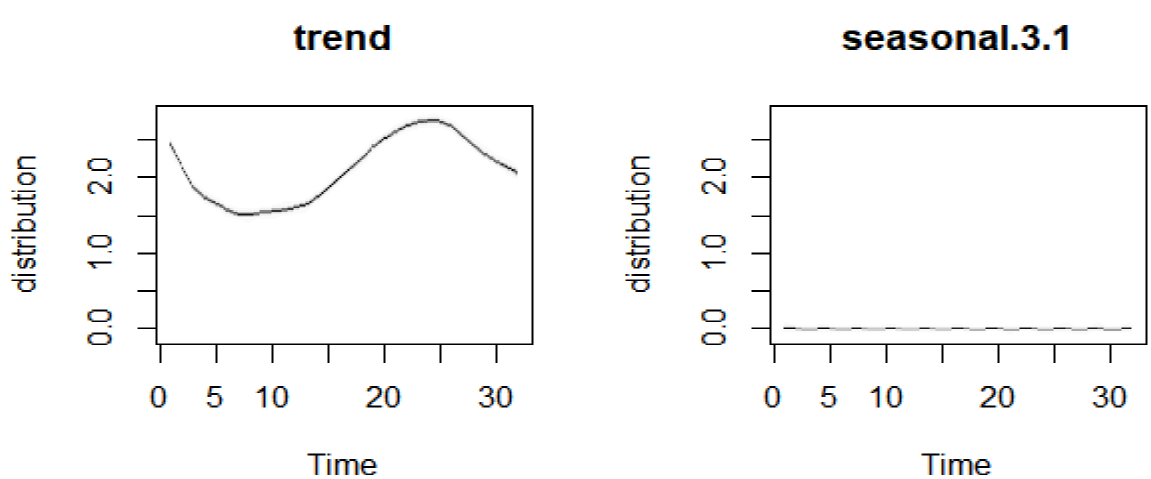
LINK TO: Federal Reserve data

CODE

```
#install & load bsts package
```

```
install.packages(“bsts”)
library(bsts)

#read in necessary data stored in delinquencyData variable
delinquencydata<-read.csv(“~/desktop/delinquencydata.csv”)

Deliquency <- as.numeric(deliquencydata$`Total loans and leases`)
dd <- AddLocalLinearTrend(list(), Deliquency)
ss <- AddSeasonal(dd,Deliquency, nseasons = 3)
#store a BSTS on the total loans and leases into model 1
model1 <- bsts(deliquencydata$`Total loans and leases`,
            state.specification = ss,
            niter = 10000)

plot(model1)
plot(model1, "components")
```

All Loans Delinquency Rate Seasonal and Trend Components



```
#plot(model1, "help")
summary.bsts(model1)
```

$residual.sd
[1] 0.0234888

$prediction.sd
[1] 0.1193062

$rsquare
[1] 0.9971038

$relative.gof
[1] -0.004642704

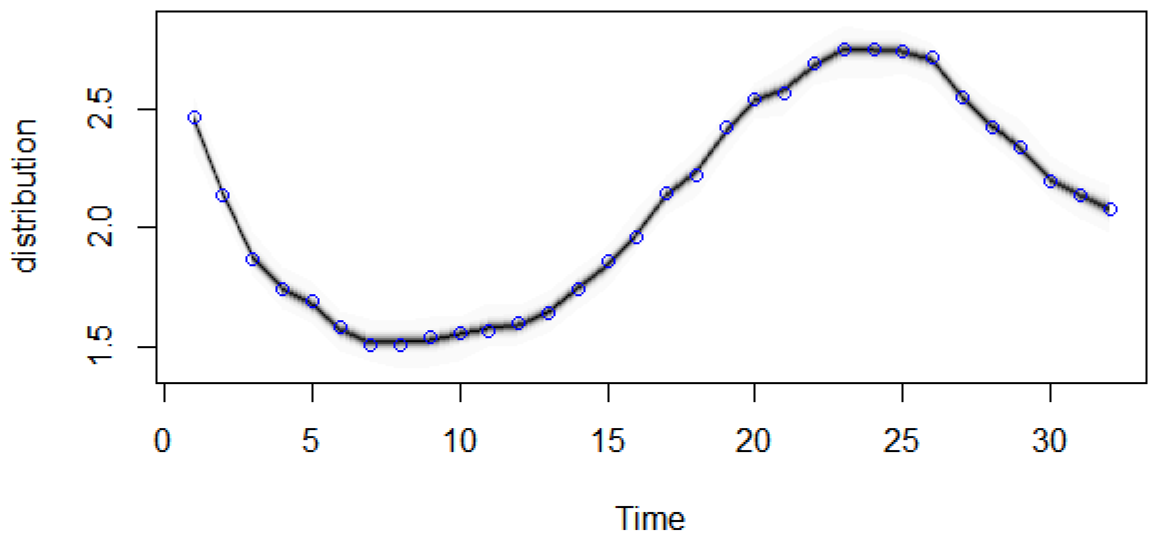
```
# Did the prediction with a burn
pred1 <- predict(model1, horizon = 8)
plot(pred1, plot.original = 32)
```

(1st graph)

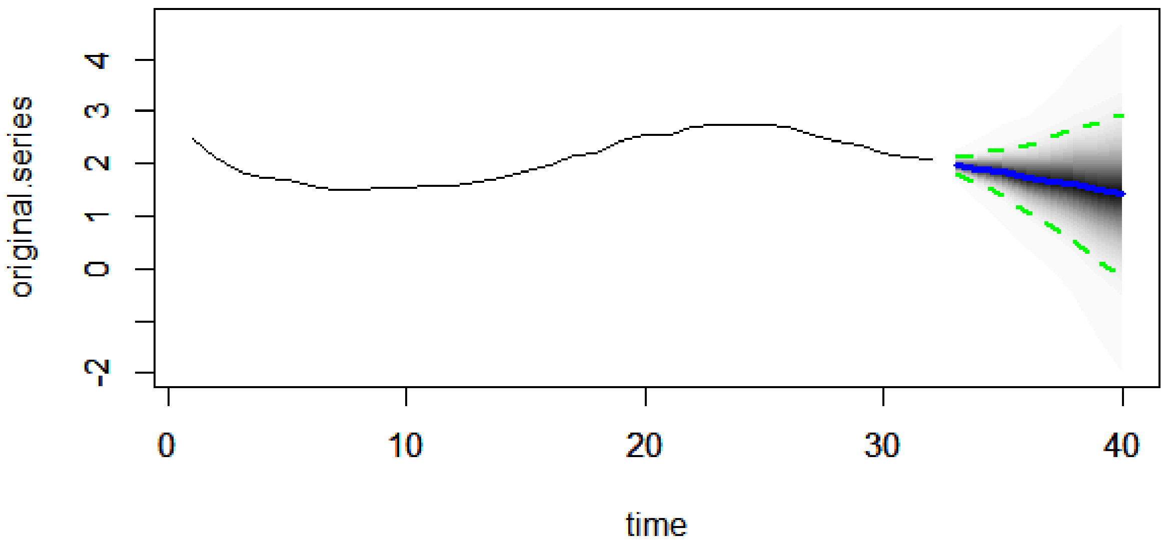
```
#Did prediction without a burn
burn <- SuggestBurn(0.1, model1)
pred <- predict(model1, horizon = 8, burn = burn, quantiles = c(.025, .975))
plot(pred, plot.original = 32)
```

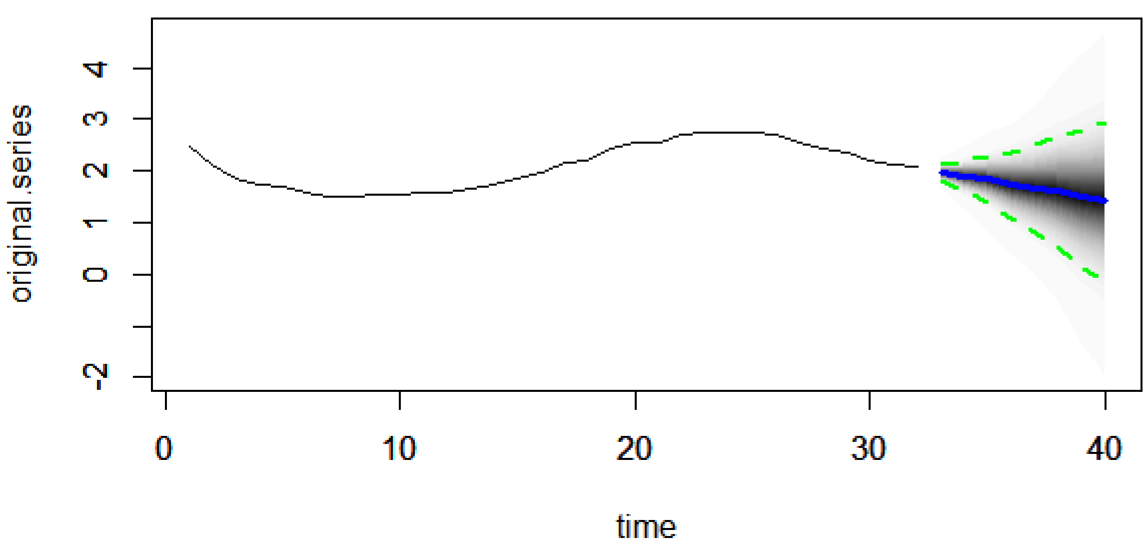
(2nd graph)

Fitted Bayesian Model for All Loan Delinquency Rates

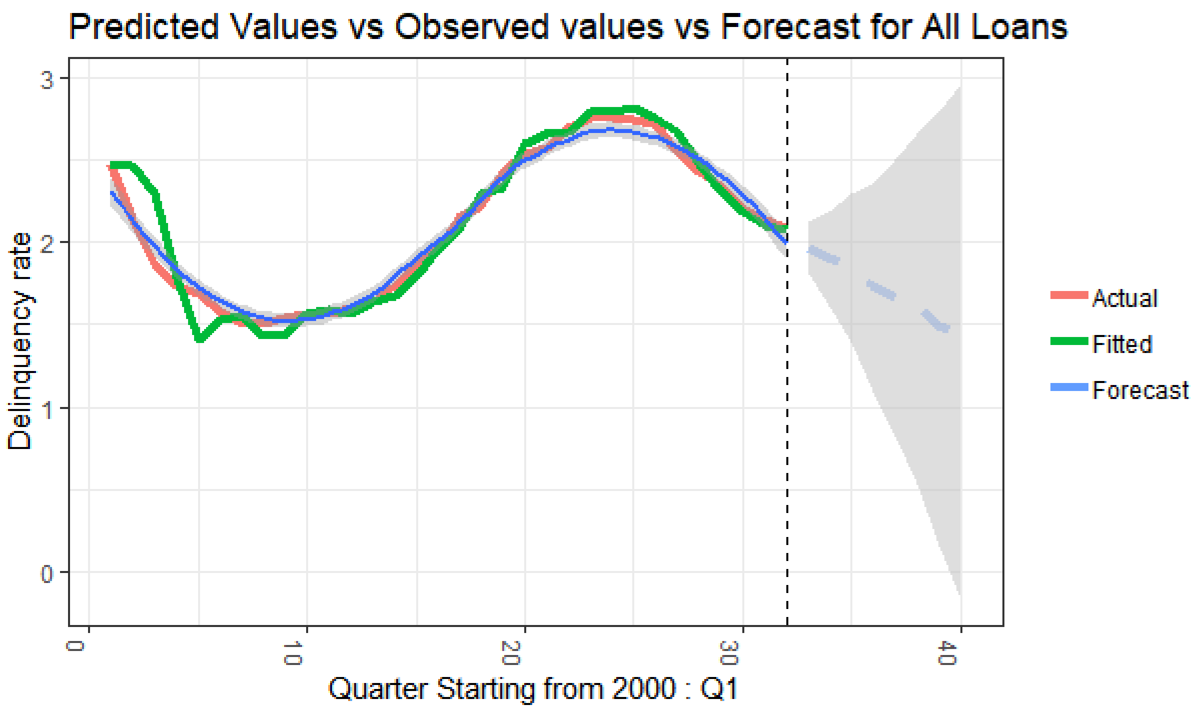


Prediction Forecasts for All Loans Delinquency Rates

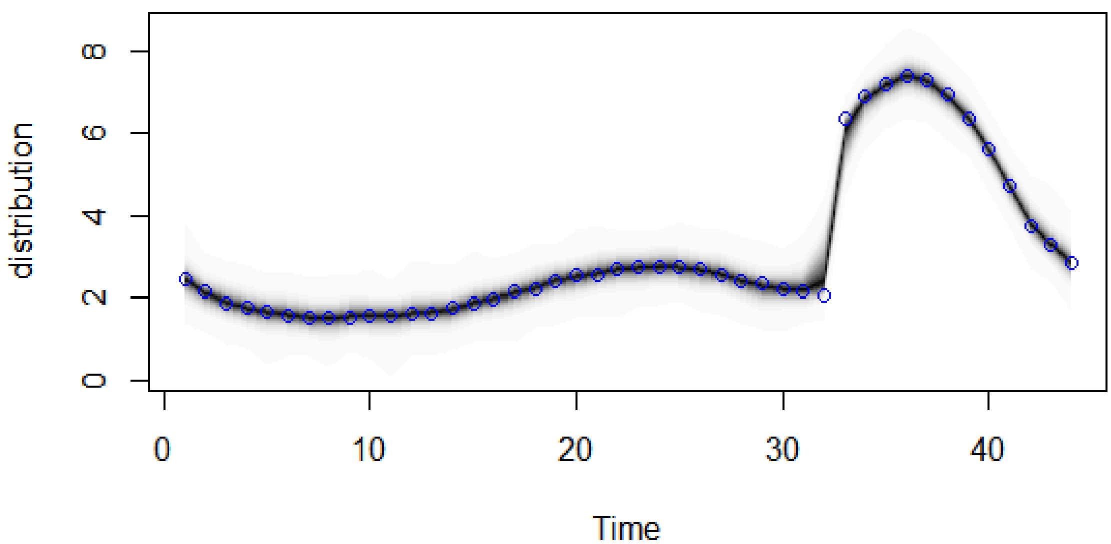




```
# data frame storing data from the prediction

d2=data.frame(t(pred$interval),Actual=pred$mean,pred$median,Index = 33:40)
names(d2) <- c("lower95", "upper95", "Actual","median","Index")

# data frame storing the actual time series and data and the Bayesian fit

d3=data.frame(Actual=pred$original.series,Fitted=-colMeans(model1$one.step.prediction.errors
[-(1:burn),])+ deliquencydata$`Residential`,Index=1:length(deliquencydata$`Residential`))

# Graph the forecast, actual data and the fit

ggplot(data=d3, aes = (x=d3$Index)) +
  geom_line(aes(x=d3$Index, y=d3$Actual, colour = "Actual"), size=1.5, linetype=1) +
  geom_line(aes(x=d3$Index, y=d3$Fitted, colour = "Fitted"), size=1.5, linetype=1) +
  theme_bw() + theme(legend.title = element_blank()) +
  geom_line(aes(x= Index, y = Actual, colour="Forecast"), data =d2, alpha = 0.5, size=1.5,
linetype=2) +
geom_ribbon(aes(x=Index,ymin=lower95, ymax=upper95), fill="grey", alpha=0.5,data=d2) +
  theme(axis.text.x=element_text(angle = -90, hjust = 0))+
  geom_vline(xintercept=32, linetype=2) +
  geom_smooth(aes(x = d3$Index, y = d3$Actual), data =d3,
            method = 'loess')  +
  labs(title = "Predicted Values vs Observed values vs Forecast")
```

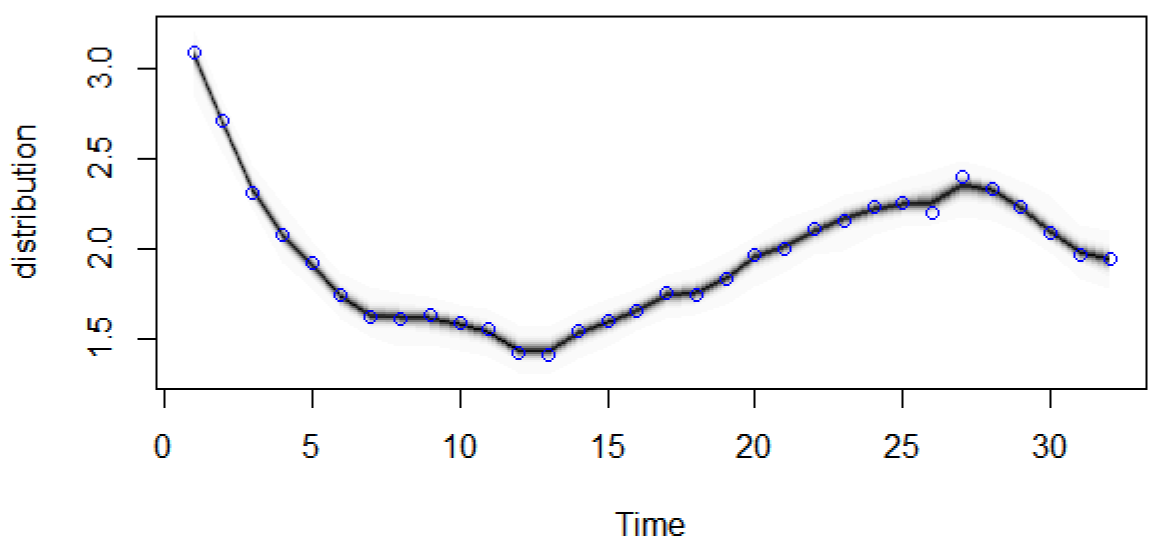

Predicted Values vs Observed values vs Forecast for All Loans
Delinquency rate
Quarter Starting from 2000 : Q1
0
1
2
3
0
10
20
30
40
Actual
Fitted
Forecast

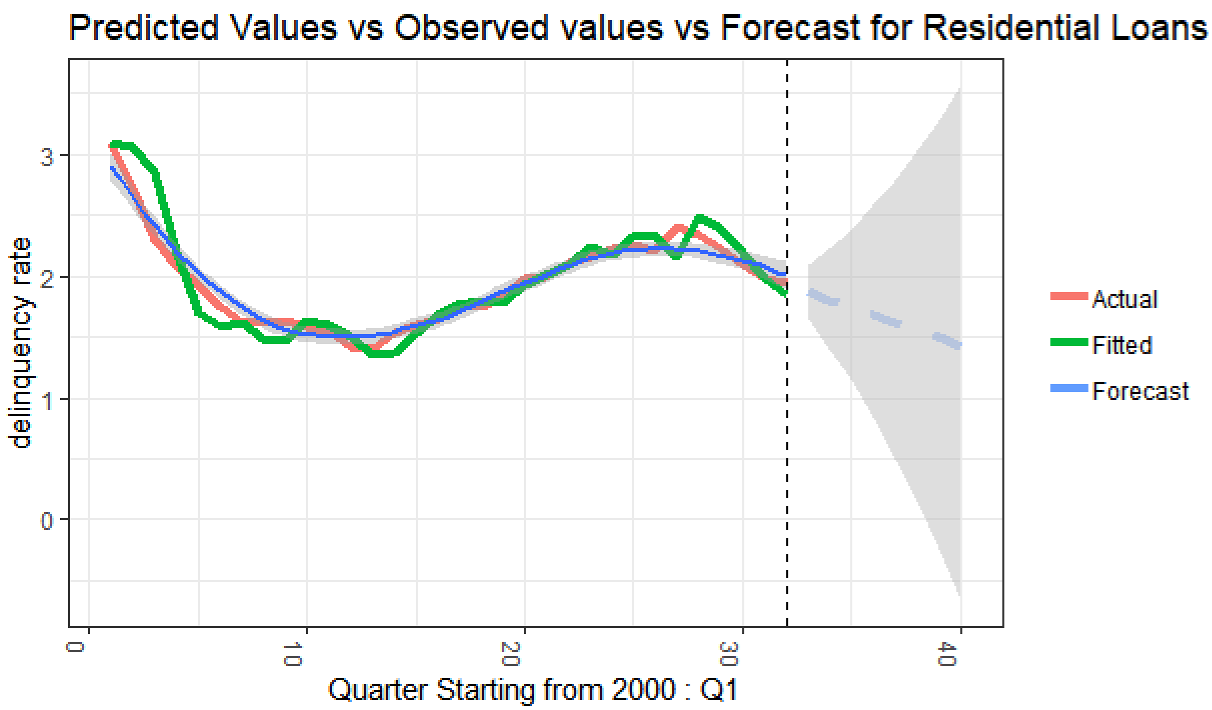
Fitted Bayesian Model for All Loans including 2008-2010 data



Note: All code for the residential loan delinquency rates is equivalent to the all loans code except the variable **<u>deliquencydata$ ‘Total Loans and Leases’</u>** was replaced by **<u>deliquencydata$`Residential`</u>**.

Fitted Bayesian Model for Residential Loans before Recession





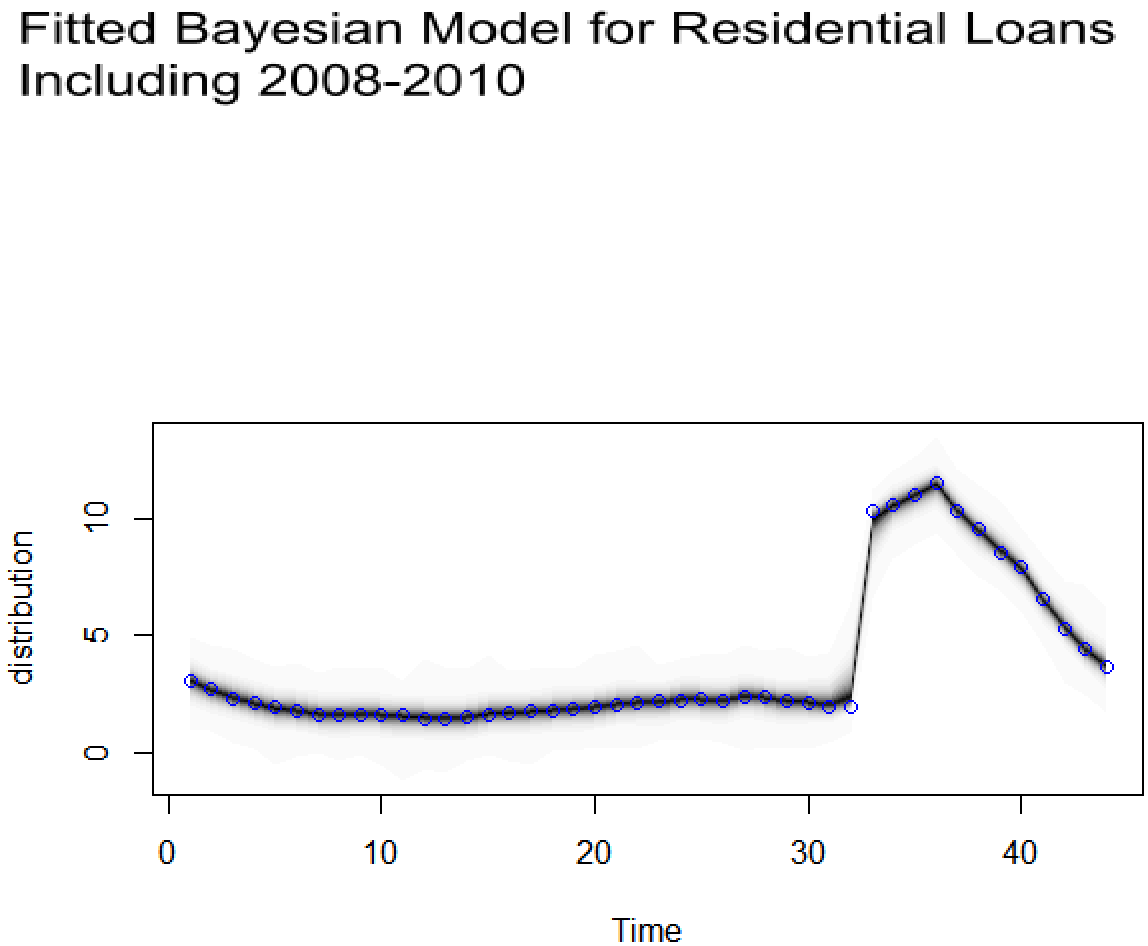
Residential Loan Diagnostic Statistics for the BSTS

$residual.sd
[1] 0.03538668

$prediction.sd
[1] 0.1555824

$rsquare
[1] 0.9913592

$relative.gof
[1] -0.2889068



<u>Analysis of the Results</u>

The first graph of the seasonal and trend components of the fitted model for all loan delinquency rates, shows a <u>somewhat cyclical</u> structure in the data. The fitted model seems to be pretty good given that the mean of the standard deviation’s posterior is pretty low and the R-squared value is pretty high. The slightly negative Harvey’s goodness of fit statistic suggests that a random walk model is only **slightly better** than the fitted bayesian model. The relatively low prediction standard deviation may suggest that the model has good predictive power for the given pre-recession data. Overall though it seems like a Bayesian structural time series model is a <u>good fit</u> of the data.

Next we did a preliminary forecast based on the obtained prediction results and generally the forecast predicted a local <u>negative</u> trend in loan delinquency rates with an expectation that loan delinquency rates would actually go down in 2008-2009. The actual data from the recession on the contrary shows a large spike in loan delinquency rates. Hence, a Bayesian Structural time series model although generally a good forecasting method for financial data was unable to accurately predict the significant rise in loan delinquency rates during the financial crisis. For the pre-recession data though, the forecast was generally very good and accurate. The forecasting for delinquencies on residential loans generally showed similar results: a good Bayesian model fit and a good prediction for pre-recession rates. The prediction for the recession though was not very good.

Overall I think the fact that the forecasts were good for the pre-recession data while not being good for the recession data, can be explained by the general nature of economic data. The values of many economic factors tend to be correlated with the business cycle. In our case, delinquency rates tend to have an inverse relationship with the business cycle(as the economy does well delinquency rates fall, as the economy does poorly delinquency rates rise). The Great Recession was a significant deviation from the classical business cycle which almost resulted in the collapse of the United States financial system. Thus, it would probably be difficult to accurately predict any economic metric in this period based on data from regular economic cycles.

Furthermore, based on this project Bayesian methods are a fairly good method for prediction but like all statistical modeling methods they are prone to serious error for very rare and significant events like the financial crisis.