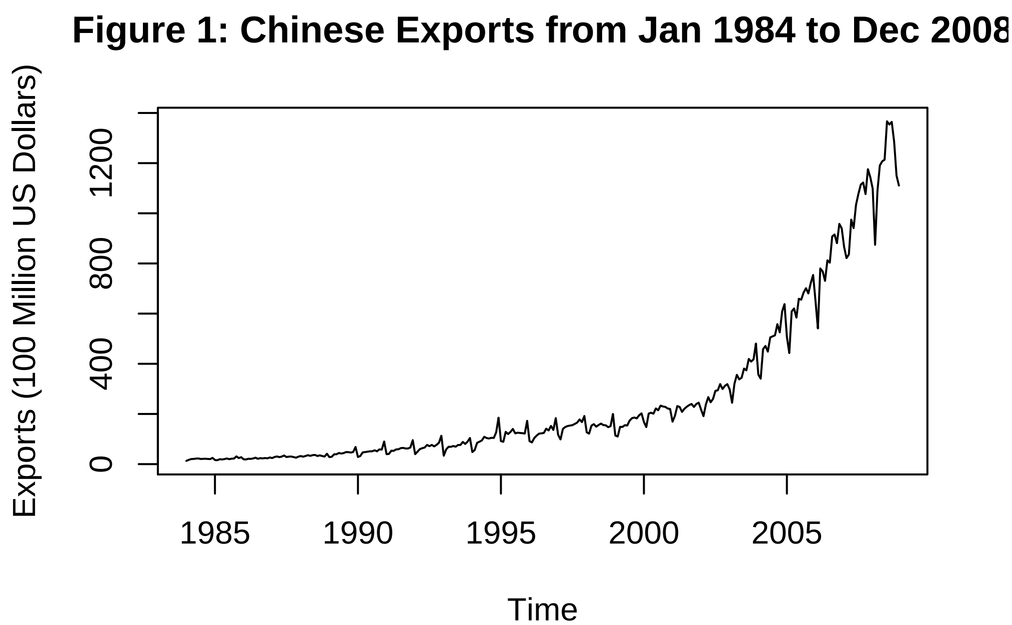
Homework 3

Francisco Calderon

December 3rd, 2016

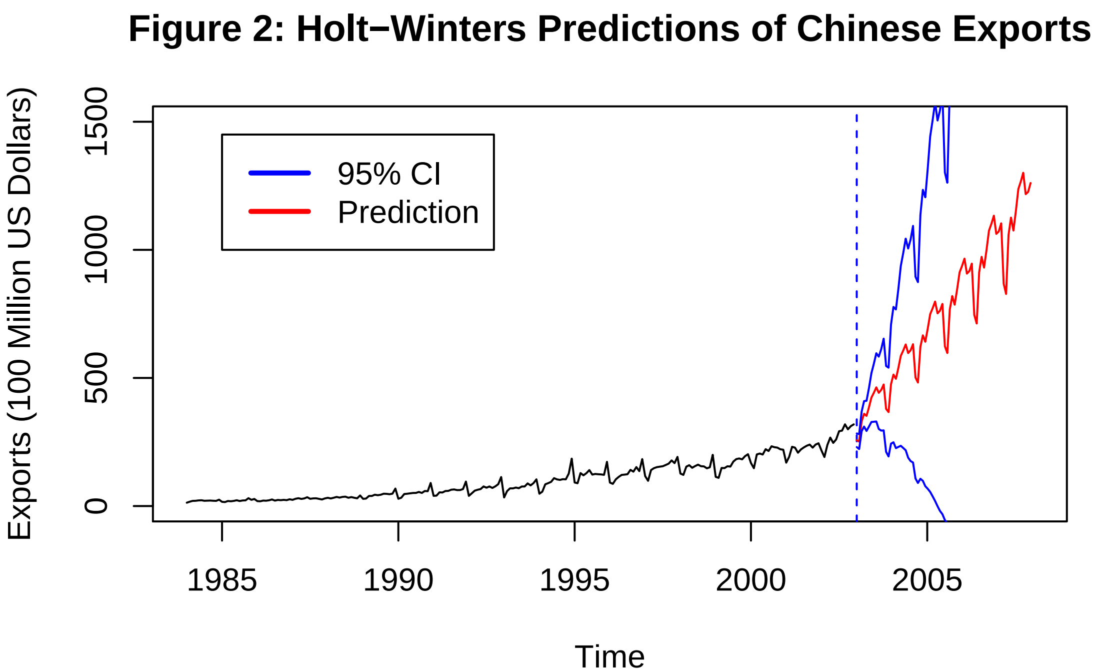
Intro

In this study, exports to China are modeled and forecasted using four time-series modeling approaches. These modeling approaches include HoltWinters, SARIMA, SARIMAX and VAR. For each modeling approach, we find the best performing model and report the root mean suqared error. Included with the exports to China are imports from China which will also be used to model and forecast exports in the SARIMAX and VAR approaches. In Figure 1 we show the time-series plots of Chinese exports. As one can see, the time series exhibits trend, seasonality, as well as changing variance throughout time. As we will see later on, the SARIMAX approach models and forecasts the data better than the other approaches.



Holt-Winters

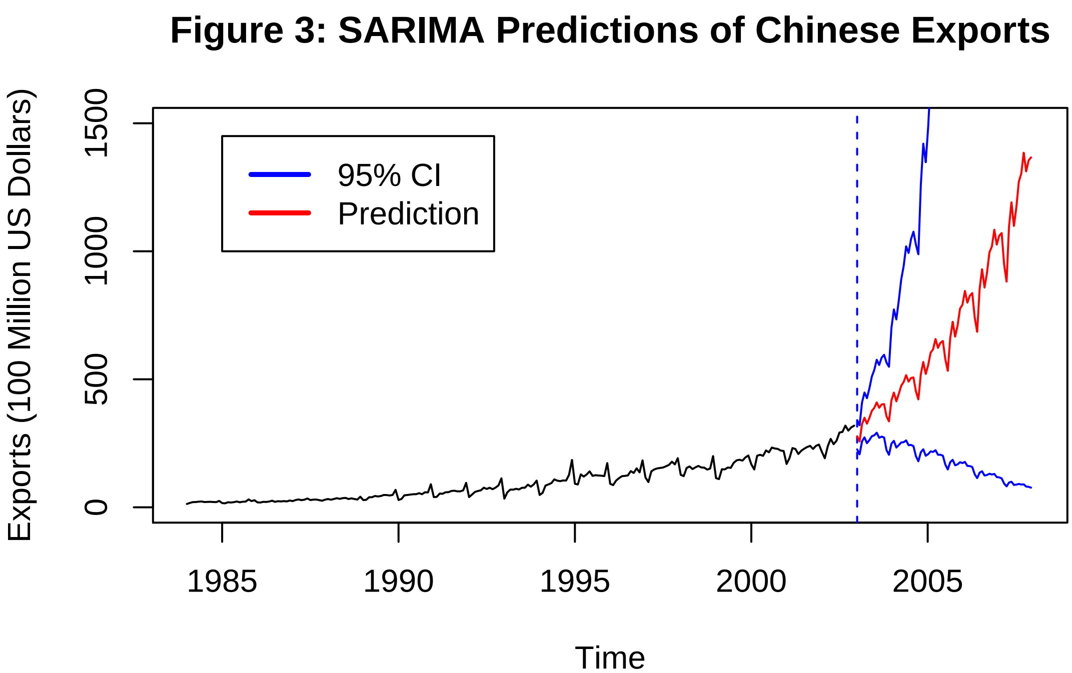
Since the plot of the series shows that trend and seasonality exist, we use the Triple Exponential Smooting method. The Holt-winters implementation in R automatically chooses the appropriate α, β, and γ to minimize the sum of square error. However, iterating through all possbile combinations of parameters with values from 0.01 to 0.99, we find the optimal hyper parameters are: $\alpha = 0.14$, $\beta = 0.87$, $\gamma = 0.47$. A plot of chinese exports forecasted is shown in Figure 2 which shows a 95% confidence interval for those predictons. The root mean squared error derived from this model is 65.19, which means that on average the model is off by 65.19 hundred-million dollars. The residuals do show a mean of zero but the assumption of constant variance is not met. Residuals also show high correlatedness throughout many lags. This can be seen in Holt-Winters Diagnostic Plots in the appendix.



Prediciton	Lower	Upper
256.8103	230.2401	283.3804
251.5613	222.3893	280.7334
330.0995	291.8315	368.3675
359.9560	310.3390	409.5731
351.9635	293.0649	410.8620
385.7271	310.3954	461.0589
423.3284	328.0956	518.5612
442.3370	329.0377	555.6363
463.2935	330.1455	596.4415
441.9418	300.7353	583.1484

SARIMA

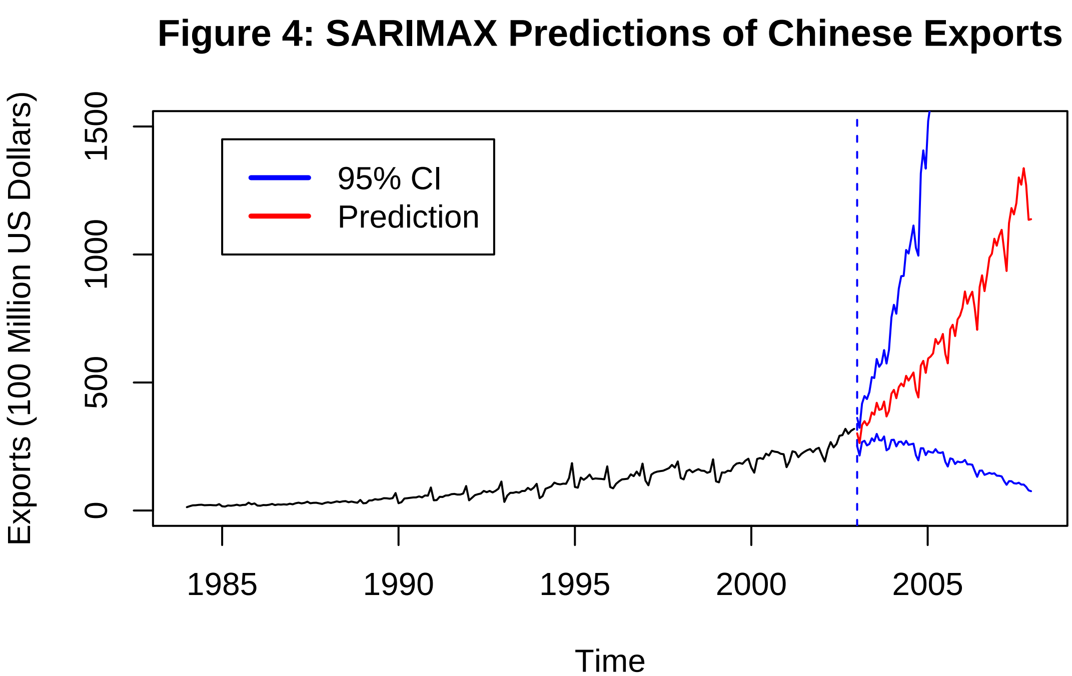
The original time series plot showed increasing varation with time, so after taking the log transformation, ACF and PACF plots showed that the order of the SARIMA model could have the following hyper paramters: $p = 1, 2, 3$ $q = 1, 2, 3$ $s = 12$ $d = 1$ $D = 1$ $Q = 1, 2, 3$ $P = 1, 2, 3$ The model SARIMA(2,1,2)x(2,1,2) with $s = 12$ performed the best with regards to RMSE. A plot of the residual time series showed that the residuals are homoskedastic and have mean of zero. The residuals did not meet the assumption of normality, but this is not a problem due to SARIMA being fit using the least squares method. Lastly, p-values for the Ljung-Box statistic show there is no presence of autocorrelation at any lags. These plots are displayed in the SARIMA Diagnostic plots in the appendix. This model derived a root mean squared error of 78.89



Prediciton	Lower	Upper
277.6162	226.3328	340.5194
257.2784	206.6303	320.3411
324.7767	257.6126	409.4517
350.0009	273.2553	448.3009
326.5797	250.3315	426.0523
349.0883	262.4696	464.2924
376.4624	277.6591	510.4240
388.1944	281.0113	536.2593
409.7780	291.3535	576.3378
388.4579	271.4799	555.8406

SARIMAX

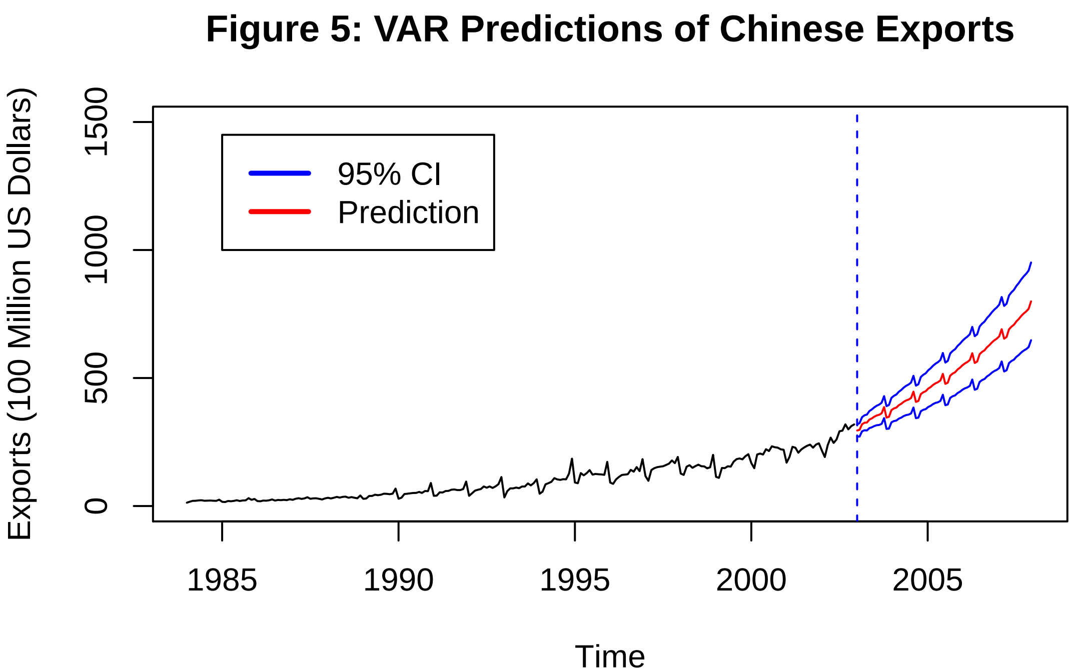
With the additional information on Chinese imports, we use the SARIMAX approach to use this new exongenous variable. Using the same potential values for p, q, P, Q, and s as the SARIMA approach before, we find that SARIMAX(1,1,3)x(3,1,3) with s = 12 performed the best with regards to root mean squared error. The residual plot showed there is a little bit of heteroskedasticity but not enough to be concerned; however, a mean of zero is very reasonable assumption from the residual plot. A QQ-plot of the residuals show they exhibit normality for the most part with exception of a few outliers at the tails. A plot of the p-values for the Ljung-Box tests show there is not autocorrelation at any lags. All of the these diagnostic plots are shown in the SARIMAX Diagnostic Plots section in the appendix. This model produces an rmse of 56.27.



Prediciton	Lower	Upper
301.5895	251.3812	361.8259
263.4072	214.6980	323.1672
334.0891	267.3904	417.4253
349.0005	272.3813	447.1723
332.7305	254.3126	435.3289
346.7487	259.8273	462.7483
383.3957	282.0382	521.1785
374.0943	270.4442	517.4692
420.8635	299.2690	591.8624
392.9604	275.0583	561.4006

VAR(P)

Now using Chinese exports and imports as endogenous variables, we use the Vector Autoregressive approach to forecast chinese exports. Using `R`'s `VARselect`, we find that an order of p=9 has the lowest AIC value. But after fitting VAR models with orders beween 1 and 10, we find that an order of 5 produced that best root mean squared error of 296.4. The residual plots shown in VAR Diagnostics Plots section in the appendix show that residuals have a mean of zero and are homoskedastic. However, the ACF and PACF plots show that there is autocorrelation at lag 12, but other than that there is no autocorrelation exhibited.



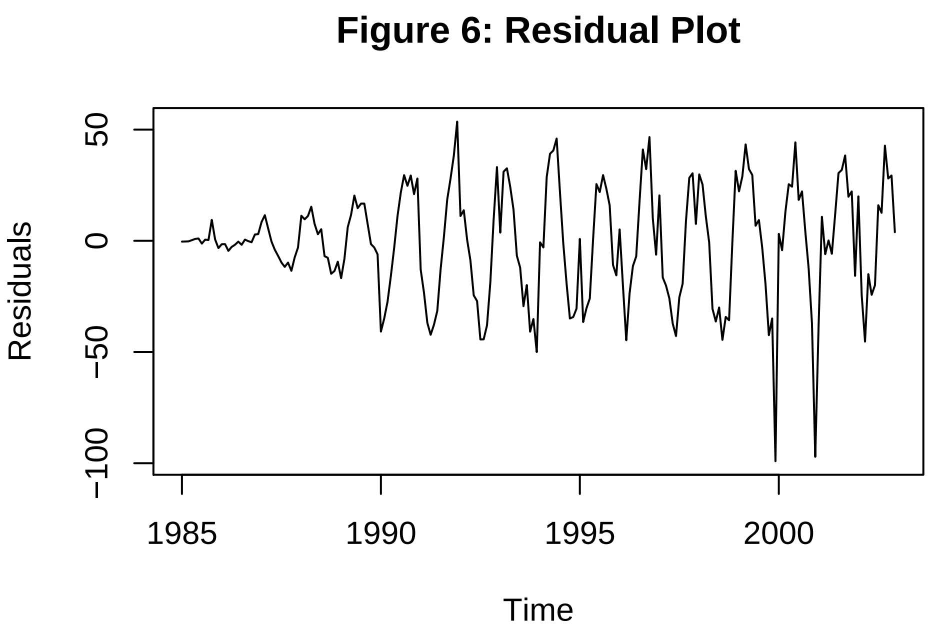
Prediciton	Lower	Upper
295.0997	271.7148	318.4846
297.5549	271.1135	323.9962
319.3281	291.7810	346.8751
325.2937	296.0578	354.5295
326.1213	295.0903	357.1523
337.7321	304.4646	370.9996
342.3380	307.3950	377.2810
348.9085	312.4254	385.3917

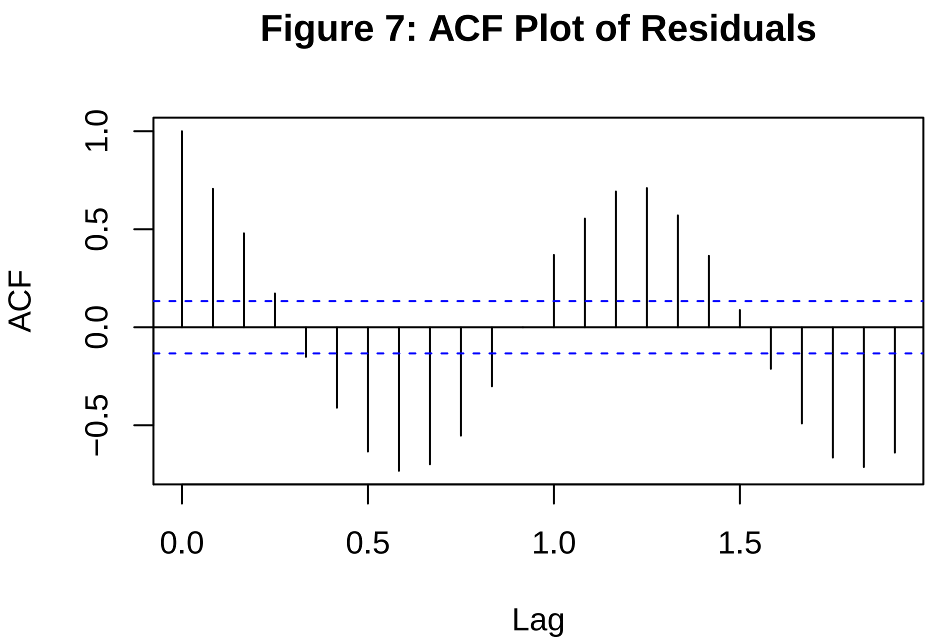
Conclusion

In forecasting Chinese exports, we find that the SARIMAX model performs the best with regards to rmse while also meeting all the residual assumptions. Even though the Holt-Winters method and SARIMA also peformed pretty well, Holt-winters did not meet the assumption of homoskedasticity or autocorrelation. Vector autorergression however, did not perform as one would like. Ultimately, this may mean that Chinese imports data were much more successful as an exogenous variable.

Appendix

HoltWinters Diagnostic Plots





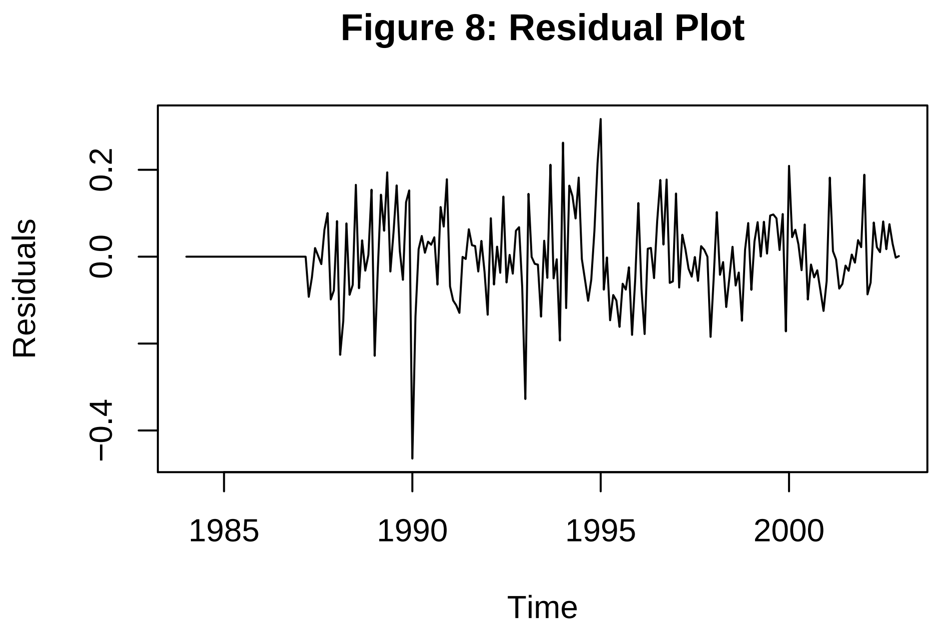


Figure 8: Residual Plot
Residuals
0.2
0.0
−0.4
1985
1990
1995
2000
Time

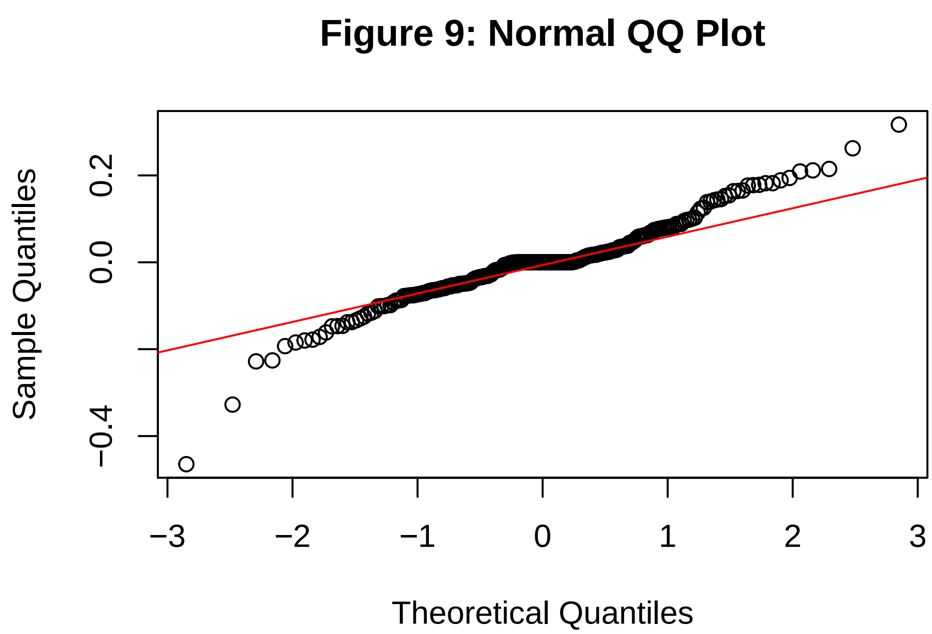
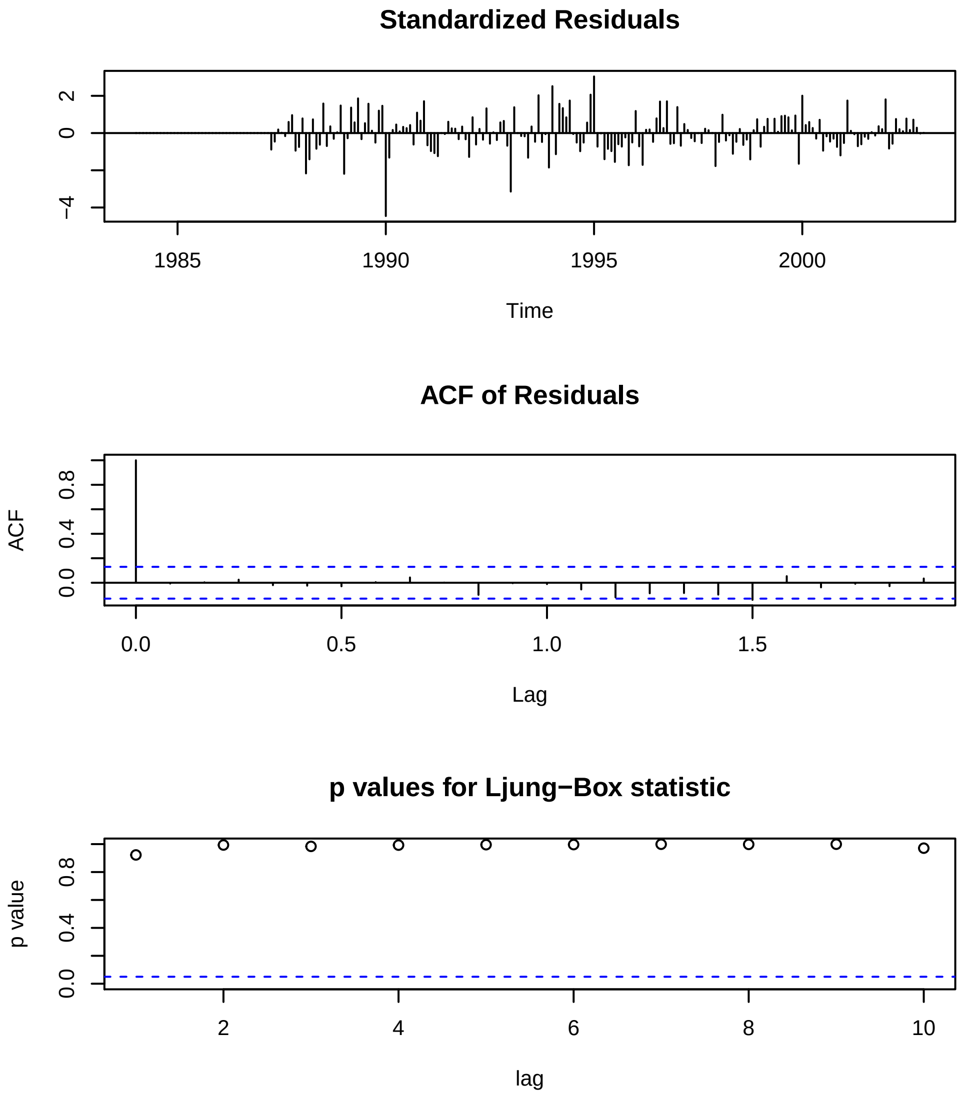


Figure 9: Normal QQ Plot
Sample Quantiles
0.2
0.0
−0.4
−3
−2
−1
0
1
2
3
Theoretical Quantiles


Standardized Residuals
2
0
−4
1985
1990
1995
2000
Time
ACF of Residuals
ACF
0.0
0.4
0.8
0.0
0.5
1.0
1.5
Lag
p values for Ljung−Box statistic
p value
0.0
0.4
0.8
2
4
6
8
10
lag

SARIMAX Diagnostic Plots

Figure 10: Residual Plot

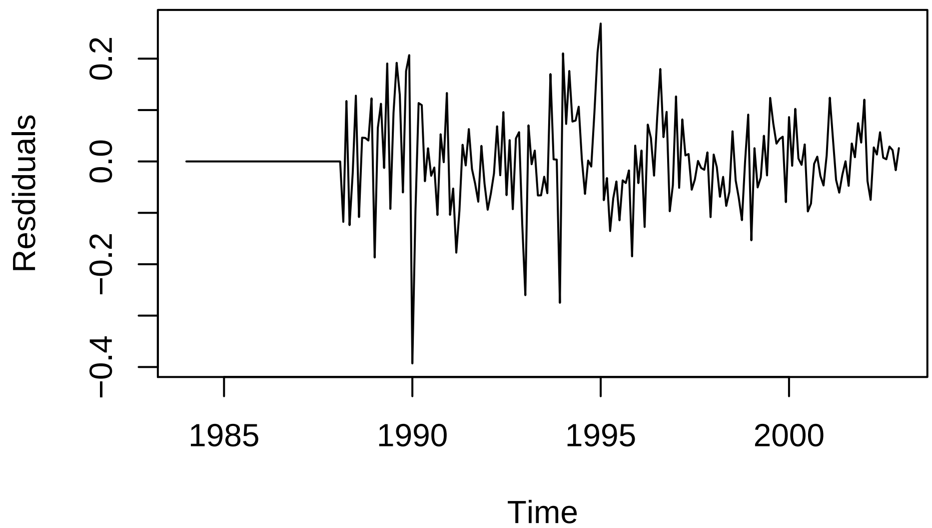
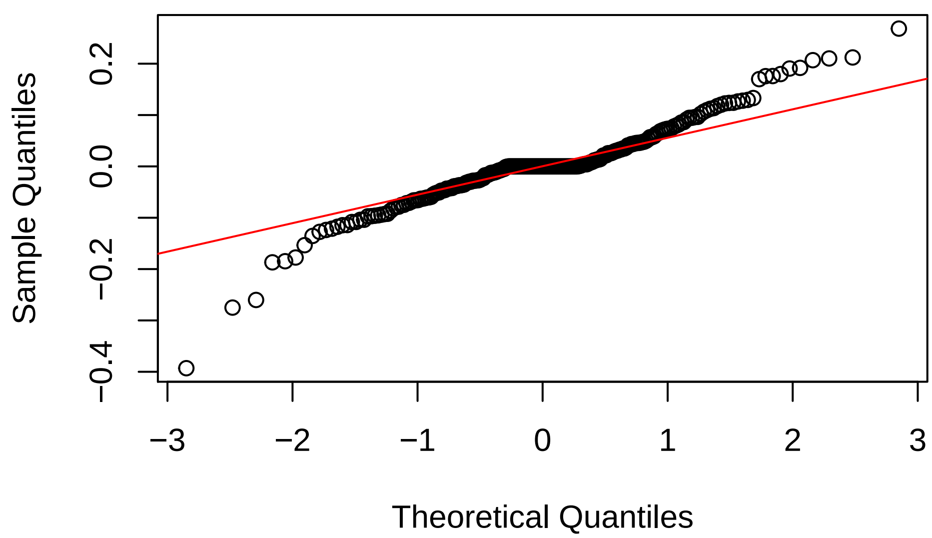
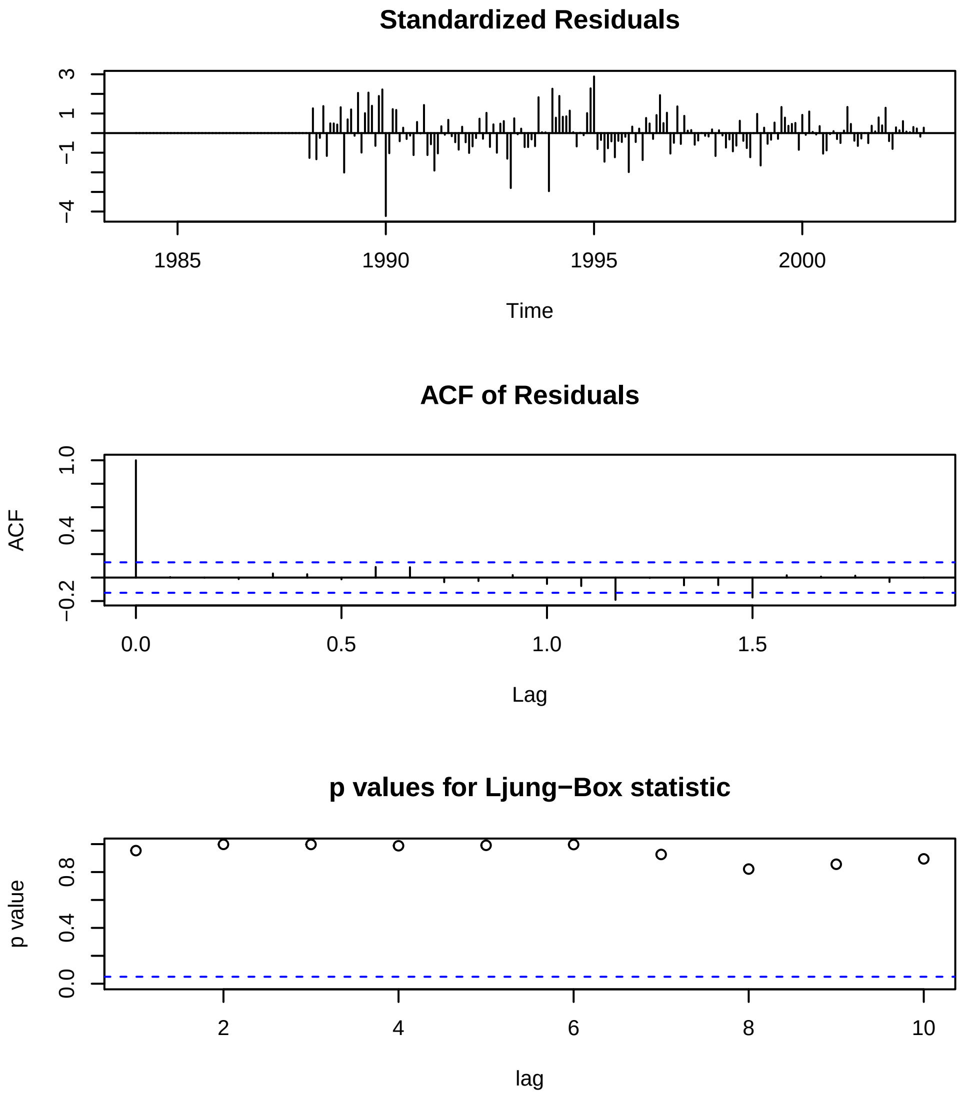


Figure 11: Normal QQ Plot




Standardized Residuals
3
1
−1
−4
1985
1990
1995
2000
Time
ACF of Residuals
ACF
1.0
0.4
−0.2
0.0
0.5
1.0
1.5
Lag
p values for Ljung−Box statistic
p value
0.8
0.4
0.0
2
4
6
8
10
lag

VAR Diagnostic Plots

Figure 12: Residual Plot

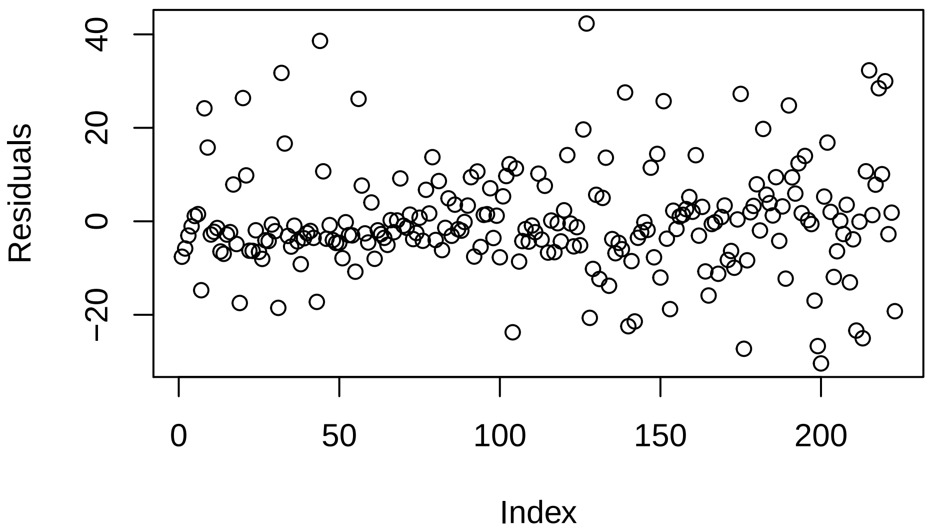


Figure 13: ACF Plot

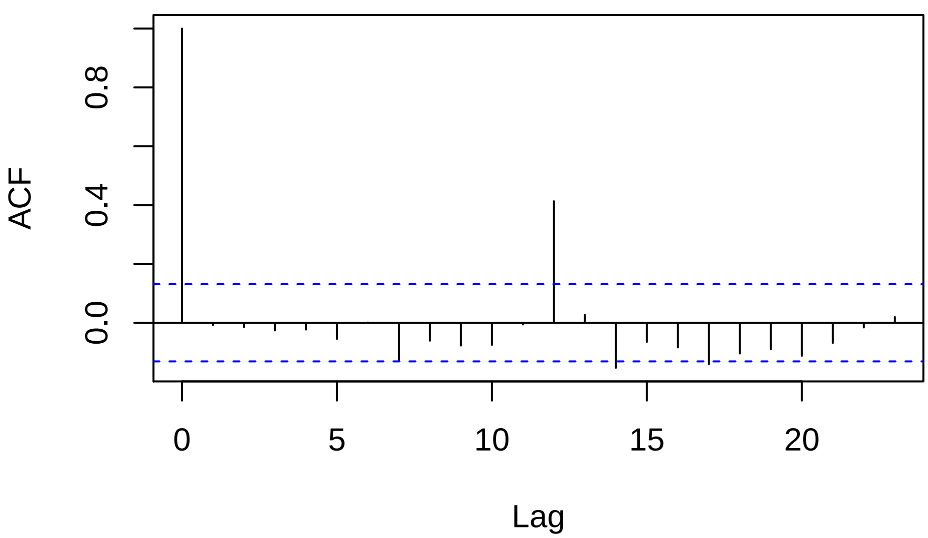


Figure 14: PACF plot