Comparing Time Series Forecasting Models for Economic Indicators

SARIMA w/ Garch vs. Regression w/ Autocorrelated Errors vs. VAR

Dennis Li
ID: 704651739

November 30, 2018

Section 1. Introduction

The main question that we are trying to address in this paper is: **When are our forecasts ahead better, i.e., more accurate?**

(a) When the only information we use is the historical information of the time series we want to forecast, and nothing else, as when using SARIMA(p,d,q)(P,D,Q), with perhaps garch.

(b) When the variable we want to forecast is the dependent variable in a traditional regression model and we use other information in the form of other variables that play the role of exogenous variables. Such a model could also have dummy variables for seasonals and polynomial trends for trend. The residuals of such a model probably need to be modeled as AR or ARIMA to account for autocorrelation. We would further use gls to approach the regression with autocorrelated residuals.

(c) When we use information in other variables, but the variable we want to forecast (as well as the other variables) is both dependent and independent at the same time. This approach is VAR, or vector autoregression.

(d) When we average the forecasted values of (a), (b), and (c), to obtain a consensus forecast.

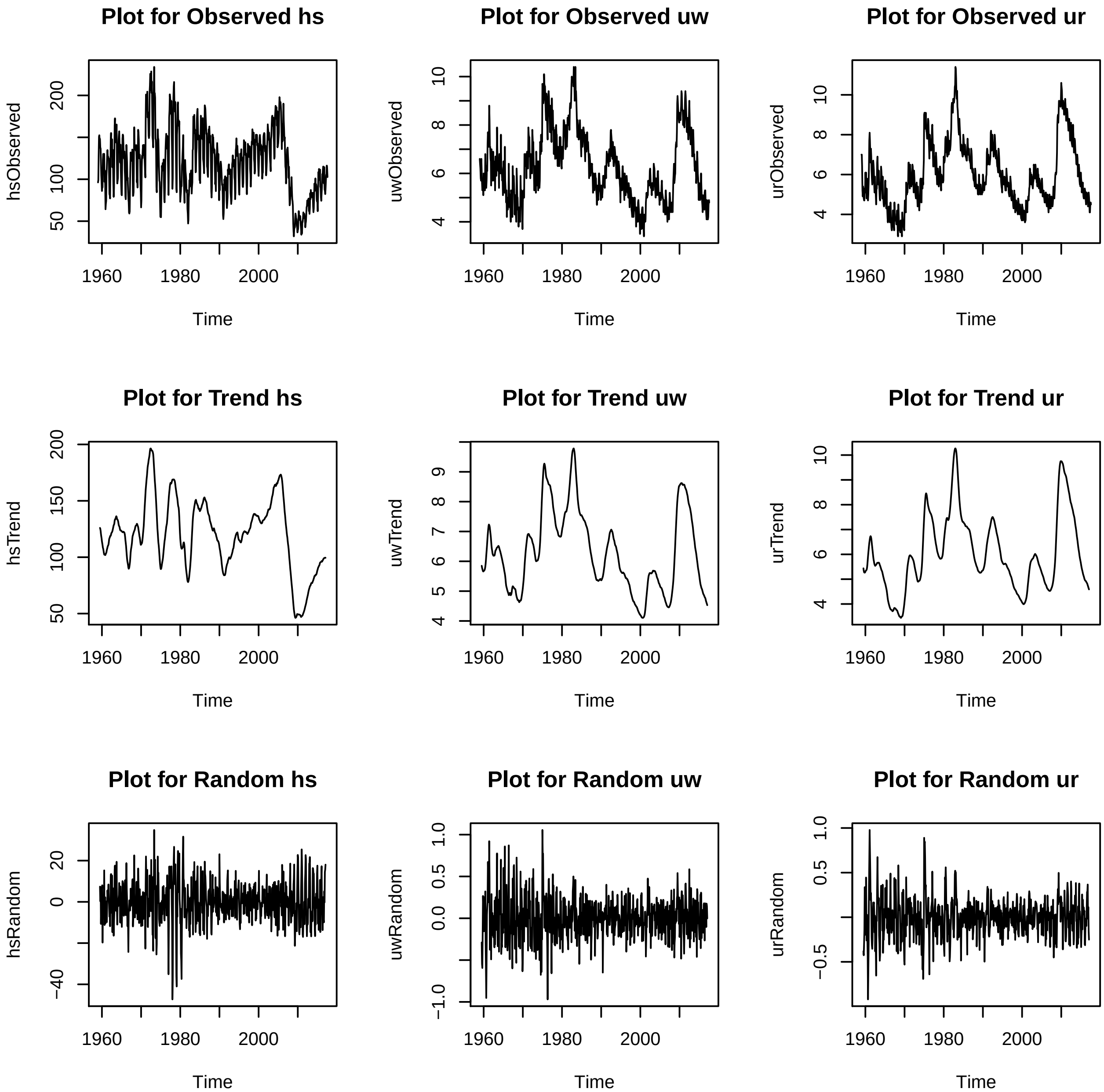
These questions always arise in practice in a number of sciences, among them economics and meterology. Nobody would ever use a single model to make a forecast for time t + 1. Almost all areas that predict the future fit several models, and then they average the forecasts obtained for each future time t + 1 from all the models to obtain the “consensus” forecast for time t. The Blue Chip consensus forecasts of economic indicators are an example of how the many forecasting companies out there combine their forecasts into a consensus forecast.

Section 2. The Data

Variable name	R name	Description and Source	Training Set	Test Set
HOUSTNSA	hs	Housing Starts: Total new privately owned housing units (in thousands). Monthly. Not seasonally adjusted. **Variable we want to forecast**	Jan 1, 1959 to Aug 1, 2017	Sept 1, 2019 to Aug 1, 2018
LNU04000002	uw	Unemployment rate: Women (percent). Monthly. Not seasonally adjusted	Jan 1, 1959 to Aug 1, 2017	
UNRATENSA	ur	Civilian Unemployment Rate (percent). Monthly. Not seasonally adjusted	Jan 1, 1959 to Aug 1, 2017	

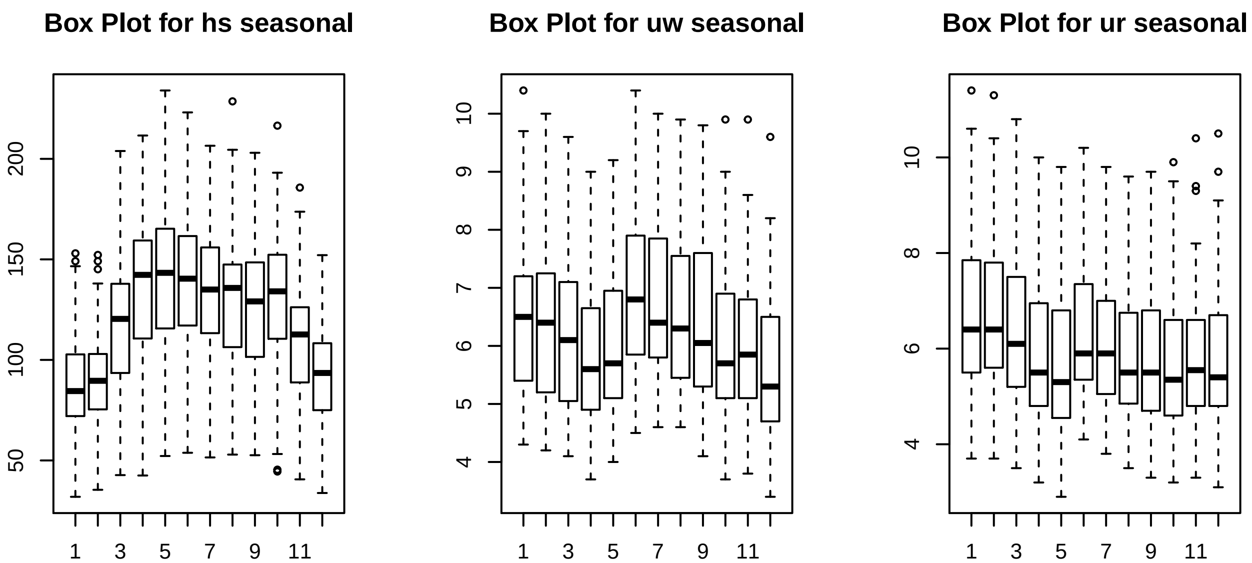
Section 3. Complete Data Description - Unit Roots Test, Volatility Checking & Cointegration

Observed and Decomposed Data



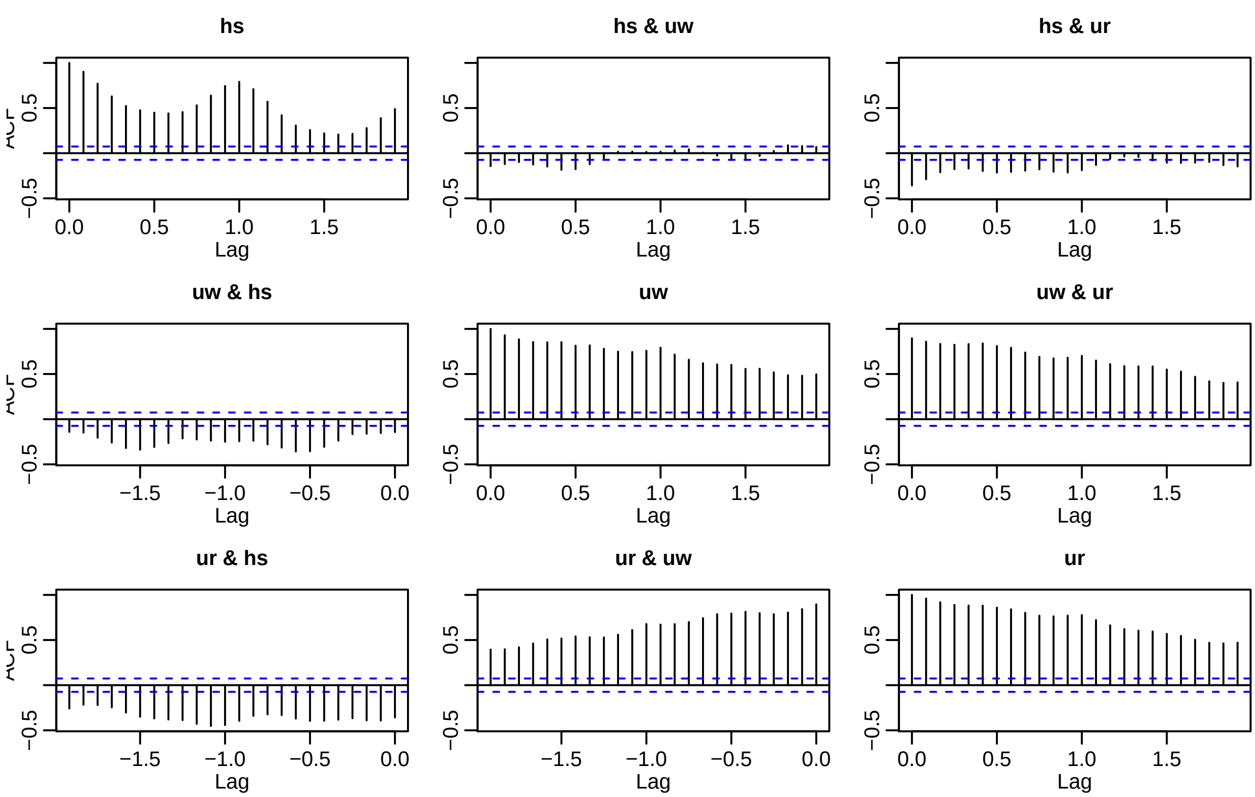
The above plots show the 3 series, decomposed by extracting the trend and random component. Based on the time plots, we can see some seasonality in the series. All 3 have trends, although not clear cut upwards or downwards trends. Although a bit hard to tell, it seems that the unemployment data and housing starts tend to move opposite from each other. In addition, the random components seem to be a bit non-stationary as the variance increases and decreases at certain points in the 3 series.

Seasonal Box Plots



The above plots show the seasonality of the 3 series using boxplots of the 12 months. We can see in the housing starts data that construction increases in the non-winter months, and is lowest in the months November-March. The unemployment rate for women (uw) time series shows some seasonality as well. Unemployment is high in January, and steadily decreases until June, where it spikes up, then steadily decreases again throughout the year. This seasonality pattern is also visible in the unemployment rate (ur) boxplots.

ACF of the Raw Data



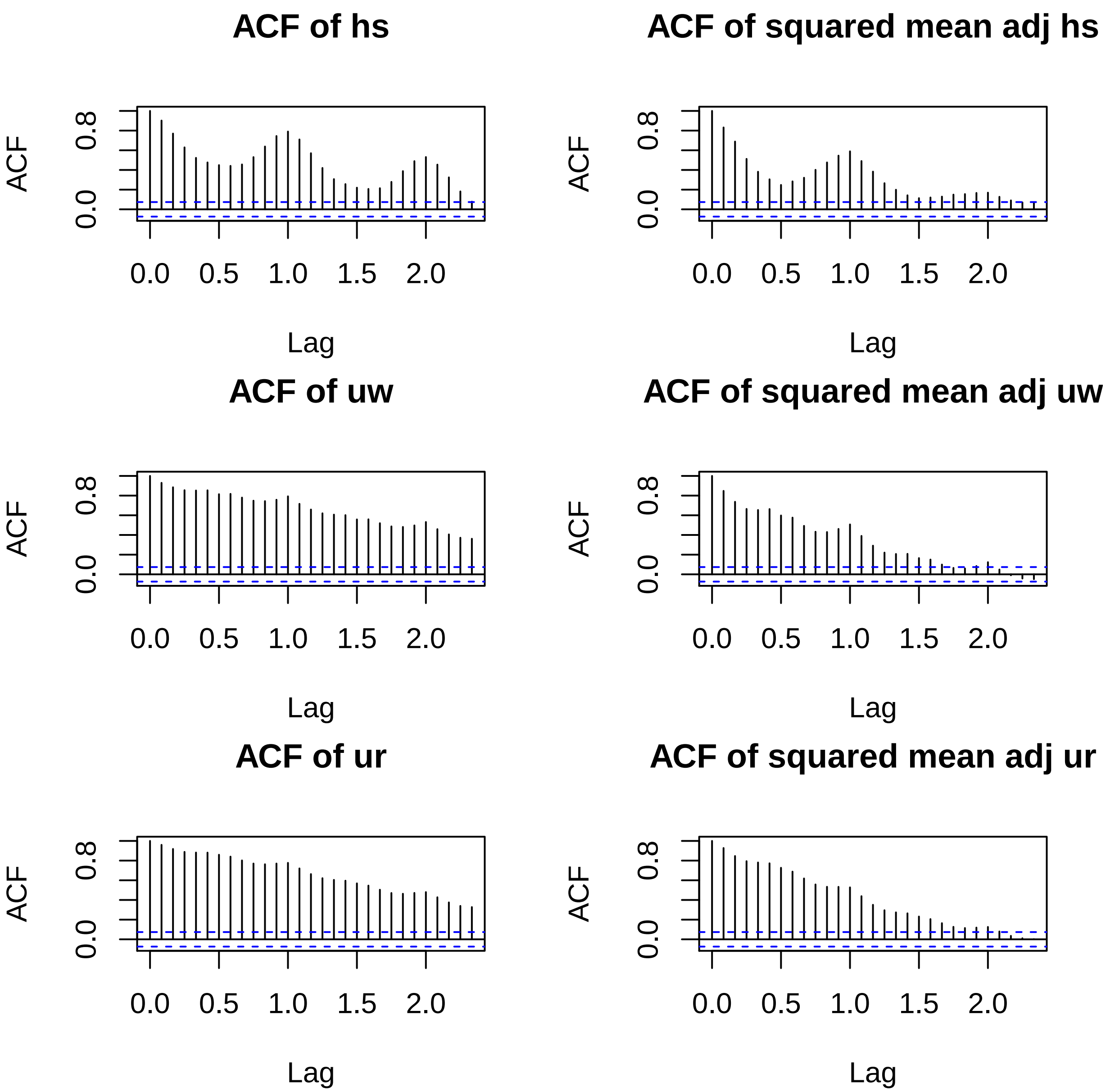
When we look at the acf and ccf of the 3 variables, we see for most of the plots that there are a lot of significant autocorrelations. This means that we must difference when analyzing the data so we obtain stationary time plots.

In addition, we conduct unit root tests to see if the data resemble a random walk. When running the Augmented Dickey-Fuller Test on the 3 time series, we obtain p-values of 0.2943 (for hs), 0.119 (for uw) and 0.09635 (for ur), so we fail to reject the null-hypothesis that the series are a random walk.

We then conduct a test to see if the 3 variables are cointegrated. We conduct the Phillips-Ouliaris Cointegration Test, and obtain a p-value lower tha 0.01, so we reject the null hypothesis that the series are not cointegrated.

Due to these 2 tests, we can be assured that any relationships from the VAR model are not spurious.

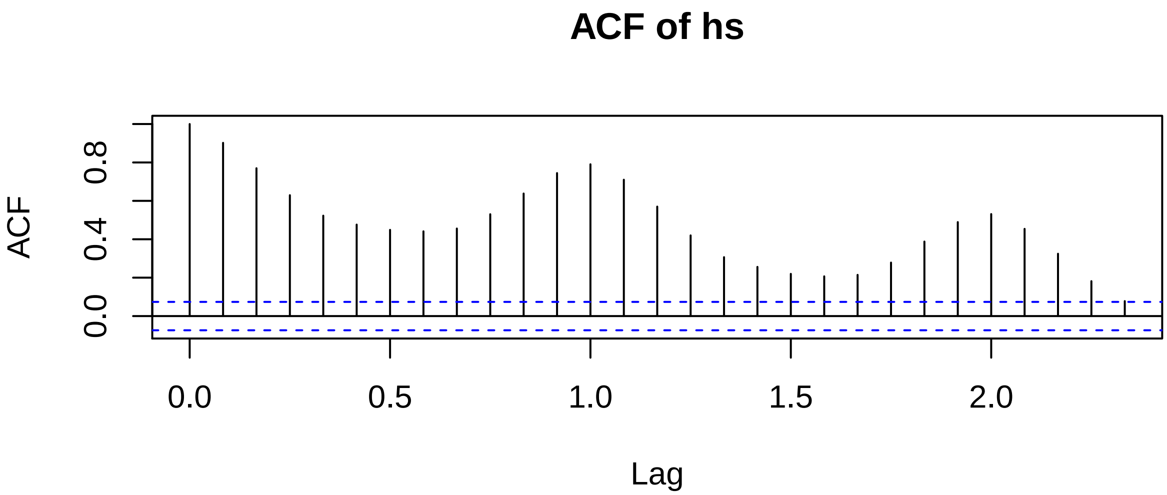
Volatility Check



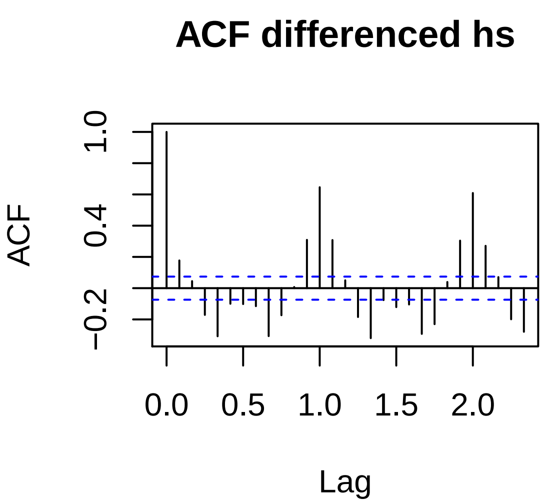
Finally, we do volatility checking to see if the data is volatile, or conditionally heteroskedastic. A time series is volatile if the ACF of the series is white noise and if the ACF of the squared series has significant autocorrelations. Here, the ACF of all 3 series are not white noise, so we do not have to worry about volatility in the time series.

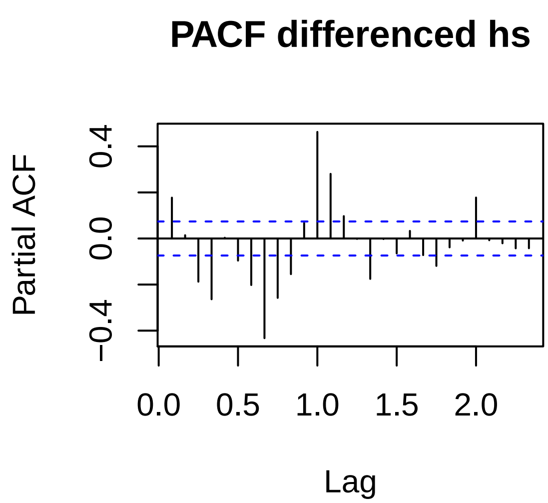
Section 4. SARIMA(p,d,q)(P,D,Q) w/ Garch

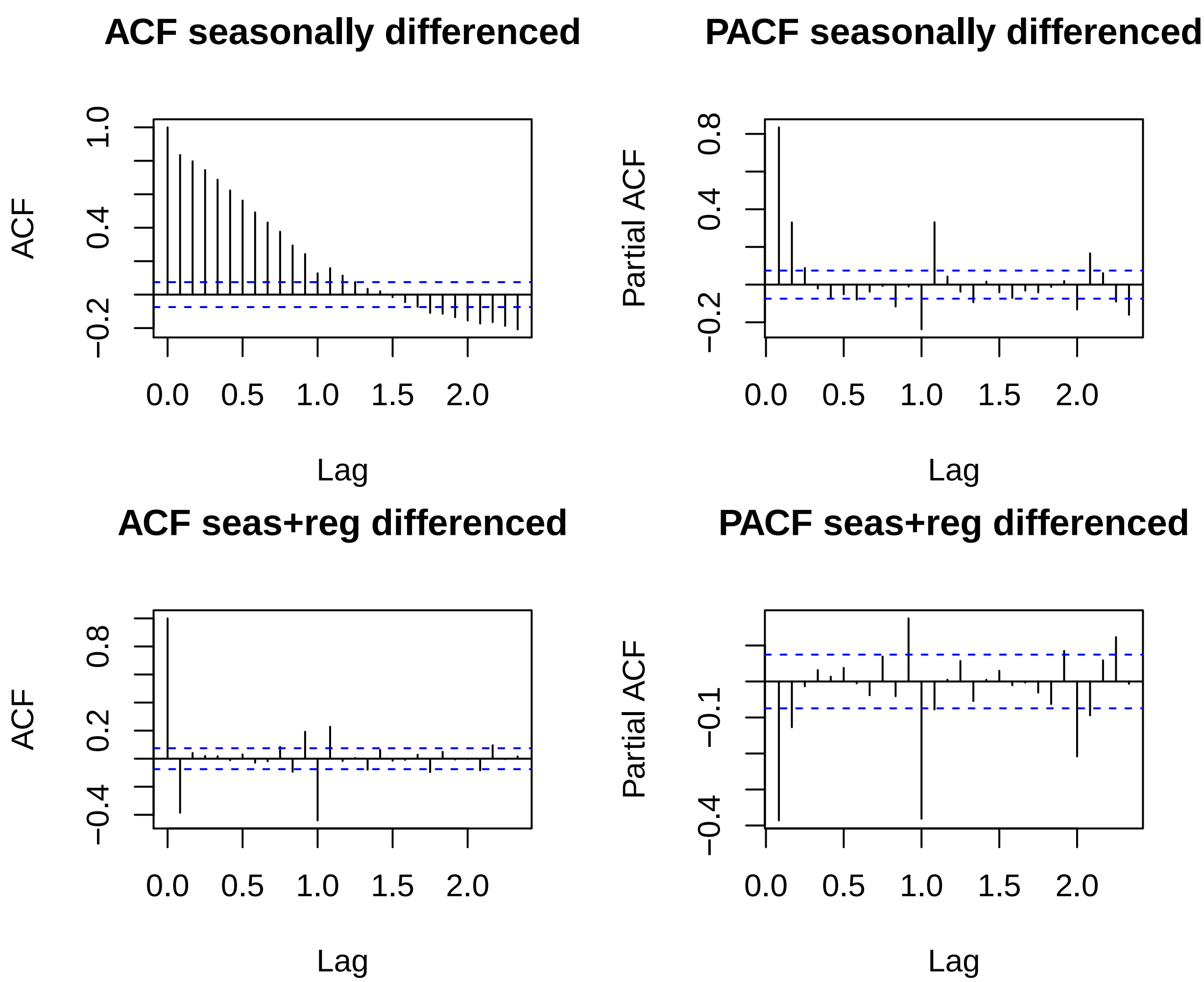
ACF of housing starts



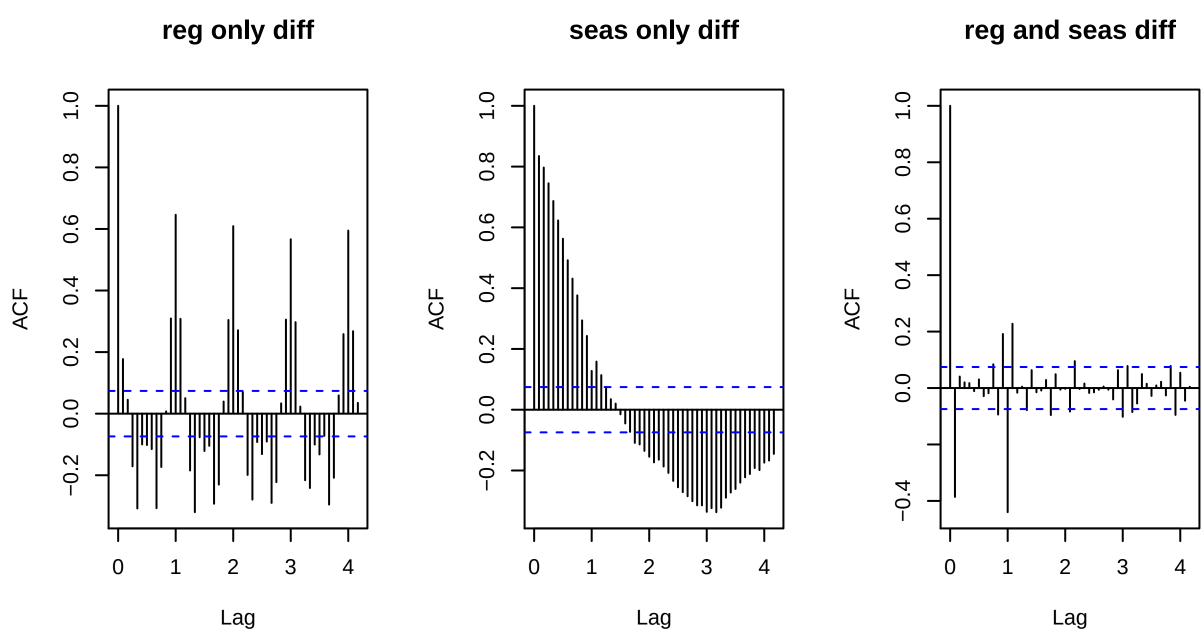
Based on the time series plot, we do not need a pretransformation such as a log or sqrt of the data because the variance is not increasing/decreasing with time. However, the acf shows very significiant autocorrelations. Due to this, we must first difference the data in order to identify a model. We try regular differencing, seasonal differencing, and both regular+seasonal differencing below.





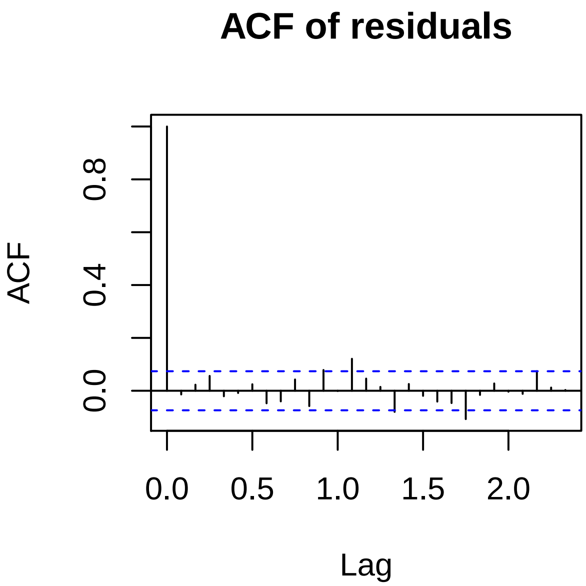


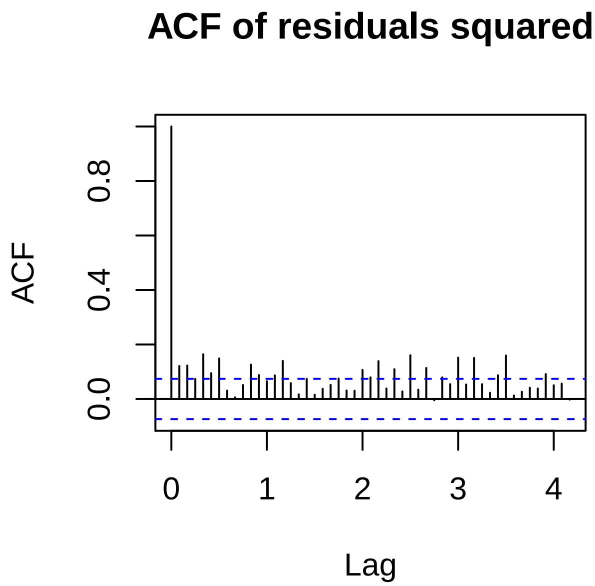
Putting Them Together



Based on the acf of the 3 differenced time series, we decide that regular+seasonal differencing is most valid because it makes the time series closest to a stationary series. Based on the acf and pacf of this differencing, we can identify a model. Based on the acf, it seems clear that an MA(1) model is appropriate for the non-seasonal part because the autocorrelation at lag = 1 is significant and it cuts off very quickly afterward. The early lags in the PACF are significant but don't cut off. It is also seen that there is a sudden seasonal spike in both the ACF and PACF, so we fit an ARMA(1,1) for the seasonal. Since we differenced regularly and seasonally, the model we will fit is a SARIMA$(0,1,1)(1,1,1)_{12}$

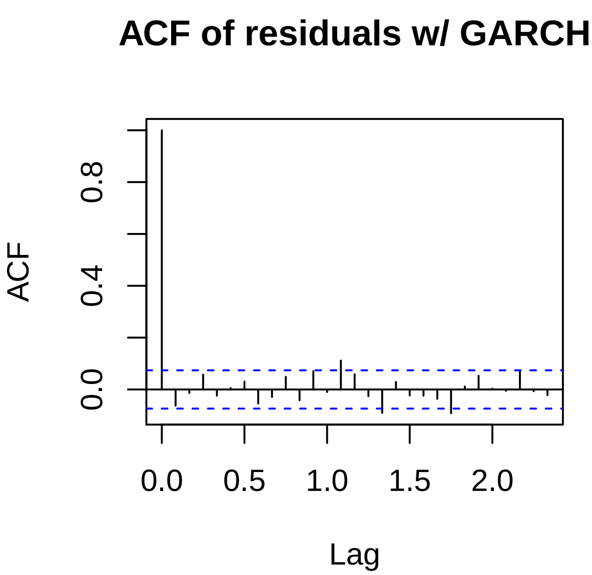
ACF of SARIMA Model Residuals

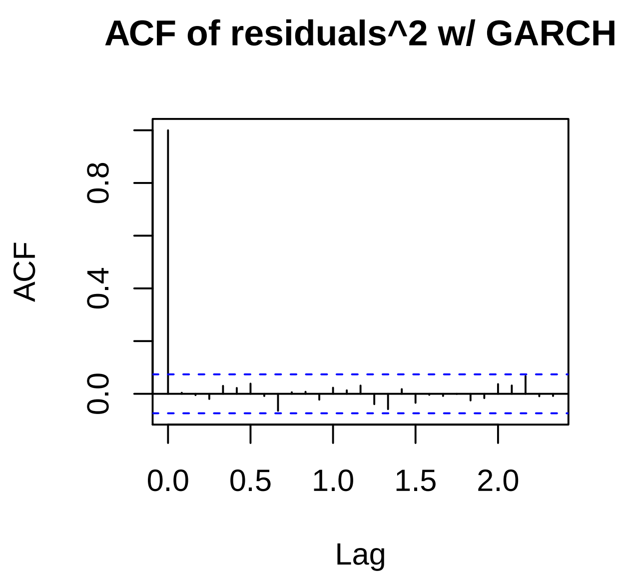




After fitting the SARIMA$(0,1,1)(1,1,1)_{12}$ model, we look at the residuals. Although the ACF of the residuals look like white noise, we further examine the ACF of the squared residuals and see that there is volatility in the residuals. We need to further capture this by fitting a garch model, which we will do next.

Fitting Garch to SARIMA Model Residuals





We choose a garch(1,1) model as this is usually adequate. After fitting this to the SARIMA model residuals, we get white noise in both the residuals and the squared residuals.

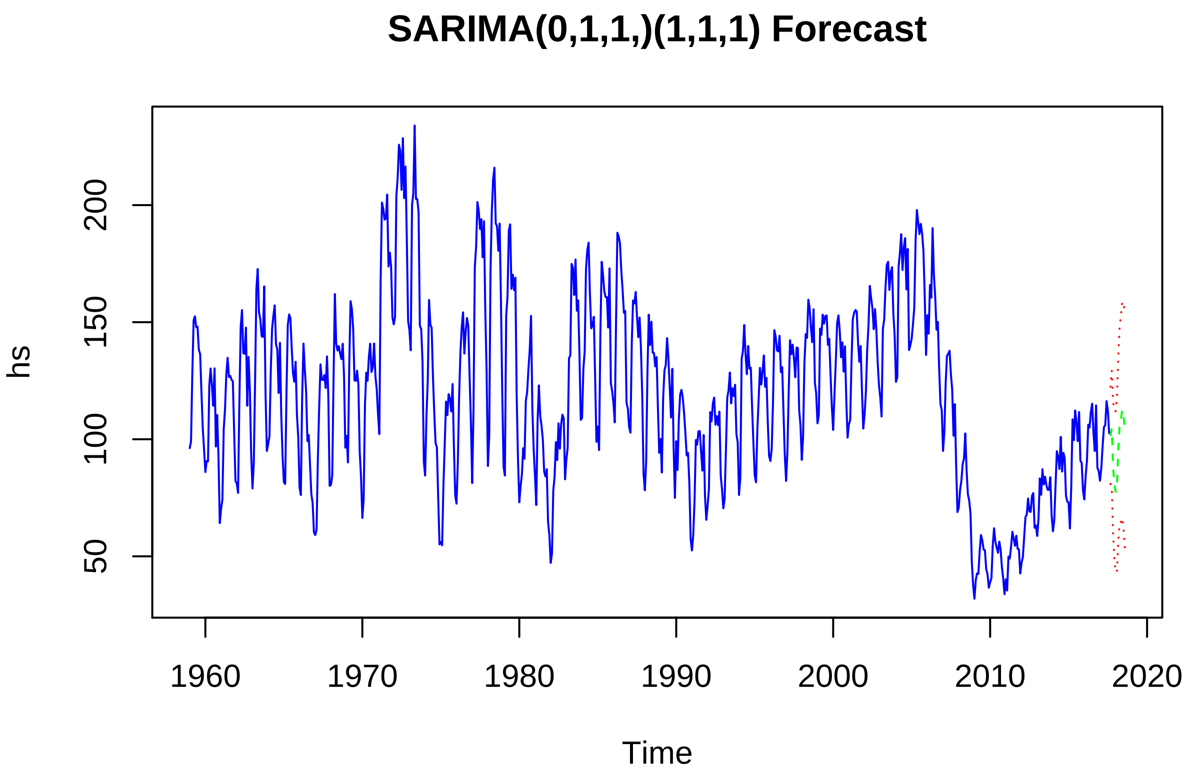
The equation of the fitted SARIMA model in polynomial form is:

$(1 - \alpha_{12}B^{12})(1 - B^{12})(1 - B)x_t = (1 - \beta_1 B)(1 - \beta_{12}B^{12})w_t$

which expands into:

$$\begin{aligned} x_t = 1.686x_{t-25} - 1.686x_{t-24} - 2.686x_{t-13} + 2.686x_{t-12} + x_{t-1} \\ + 3.58563w_{t-13} + 10.662w_{t-12} + 0.3363w_{t-1} + w_t \end{aligned} \quad (1)$$

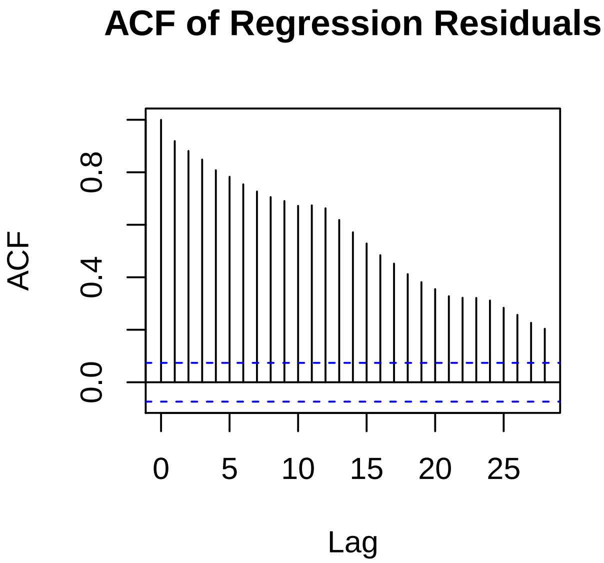
Forecast Plot for SARIMA

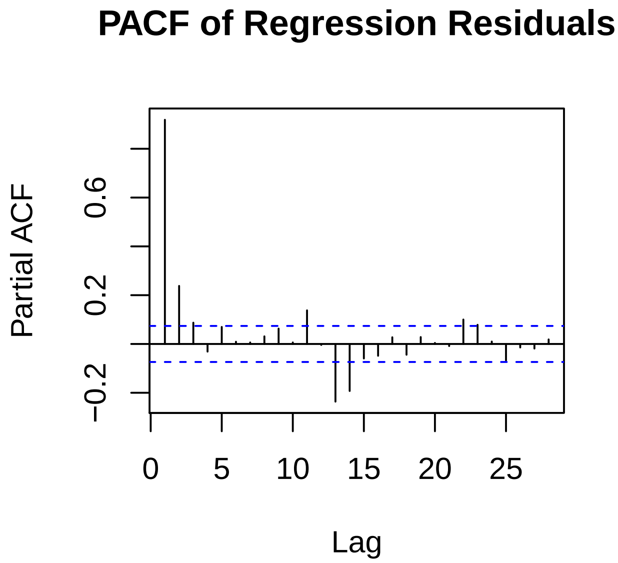


Here, we see the forecasted values for the next 12 months, with confidence intervals. The RMSE of the SARIMA model, using the test set data, is 8.958358.

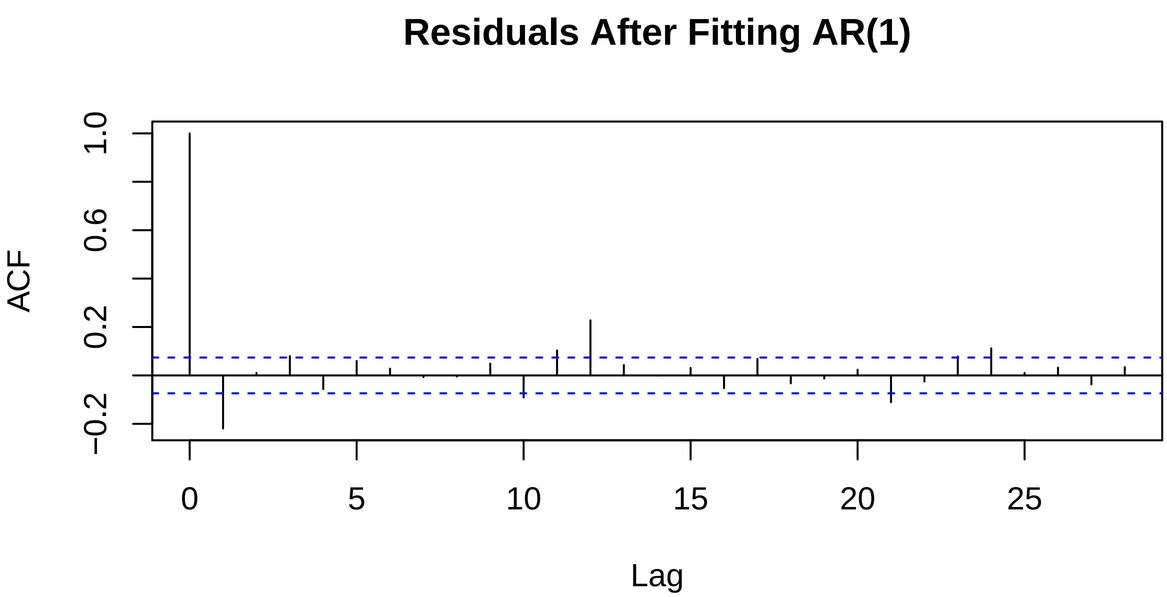
Section 5. Regression w/ Autocorrelated Errors

Using uw, ur, and seasonal to Predict



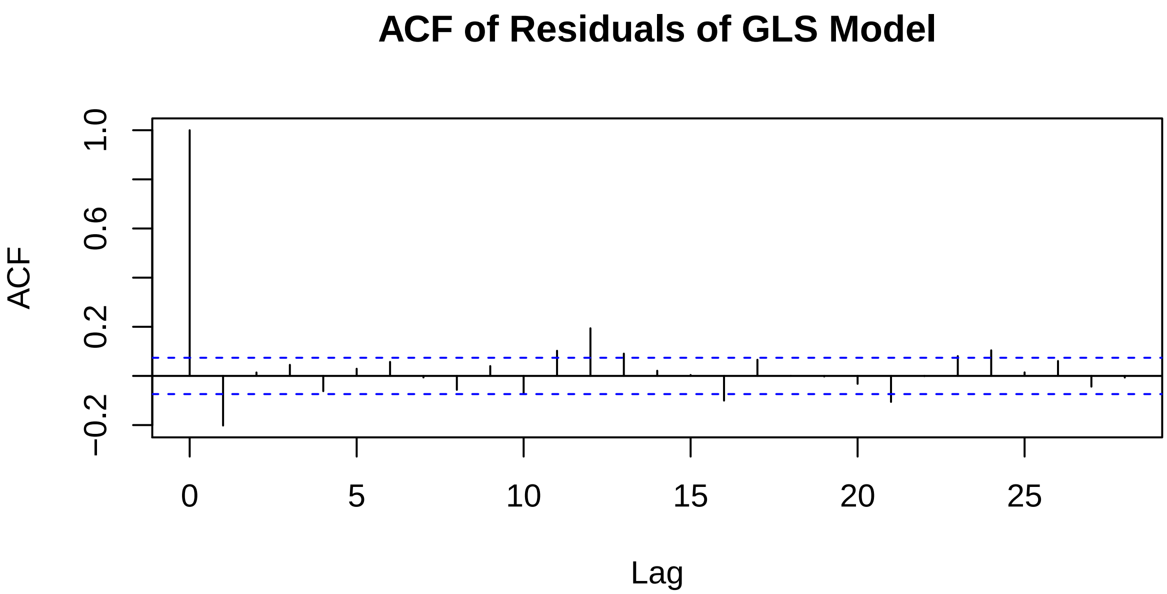


In this section, we fit a regression model using the 2 other times series variables (uw and ur) as predictor variables. We also add a seasonal term. When we examine the ACF of the residuals after fitting the regression, we see a very autocorrelated pattern resembling an autoregressive model. Therefore, we fit an AR(1) to the residuals. Below is the ACF of the residuals after doing so.



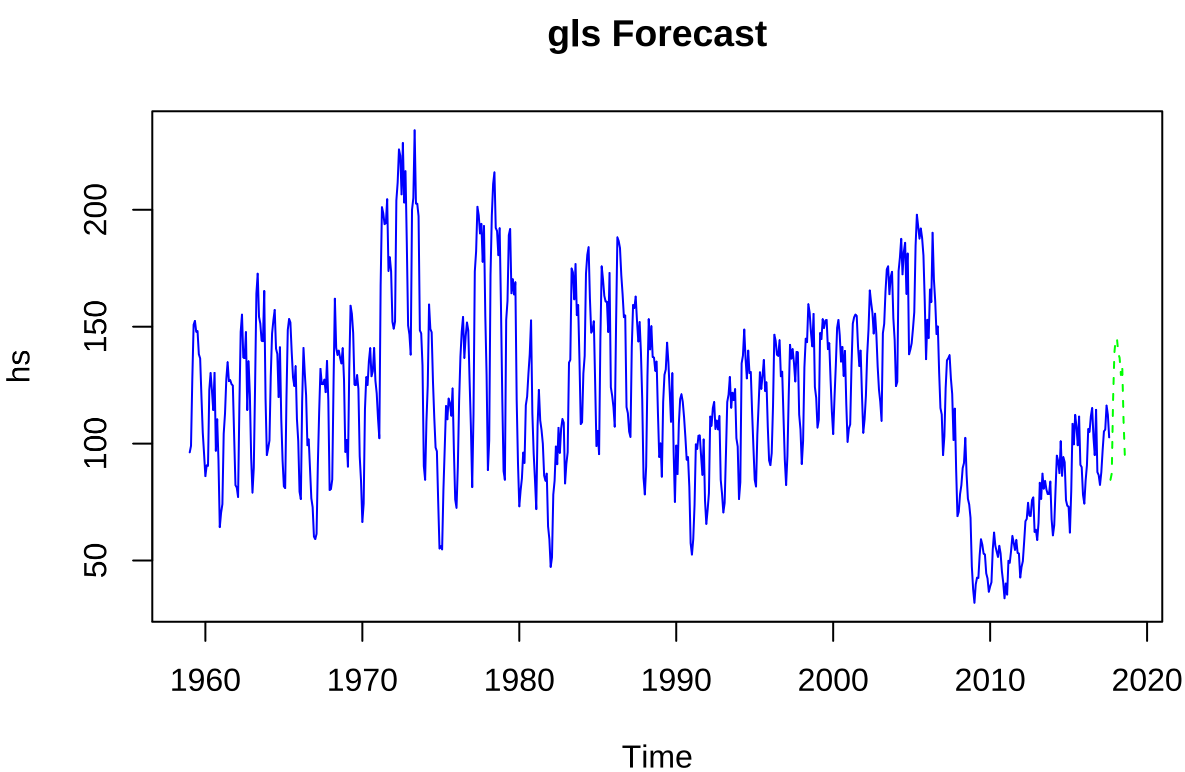
As seen above, the ACF of the residuals resemble white noise now. To capture this AR model on top of the regression, we use gls to account for the autocorrelation not explained by the original variables included (uw + ur + seasonal). Using the coefficient from the AR(1) model (0.9189), we fit the gls model and obtain the following acf of the residuals:

Fitting to GLS



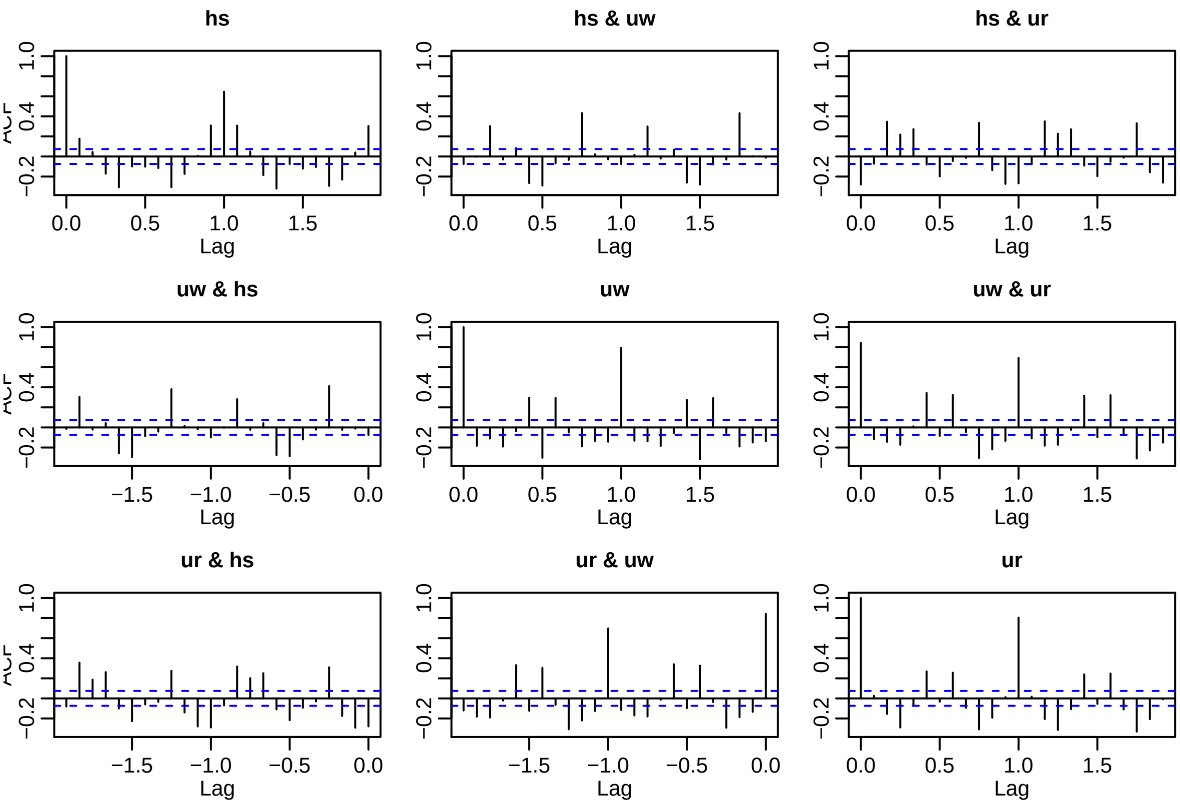
We can now forecast using the gls model, as the residuals are white noise. Below is the forecast, and we obtain an RMSE of 32.47106, which is worse than the RMSE from SARIMA (8.958358).

Forecast Plot for Regression w/ GLS



Section 6. VAR Model

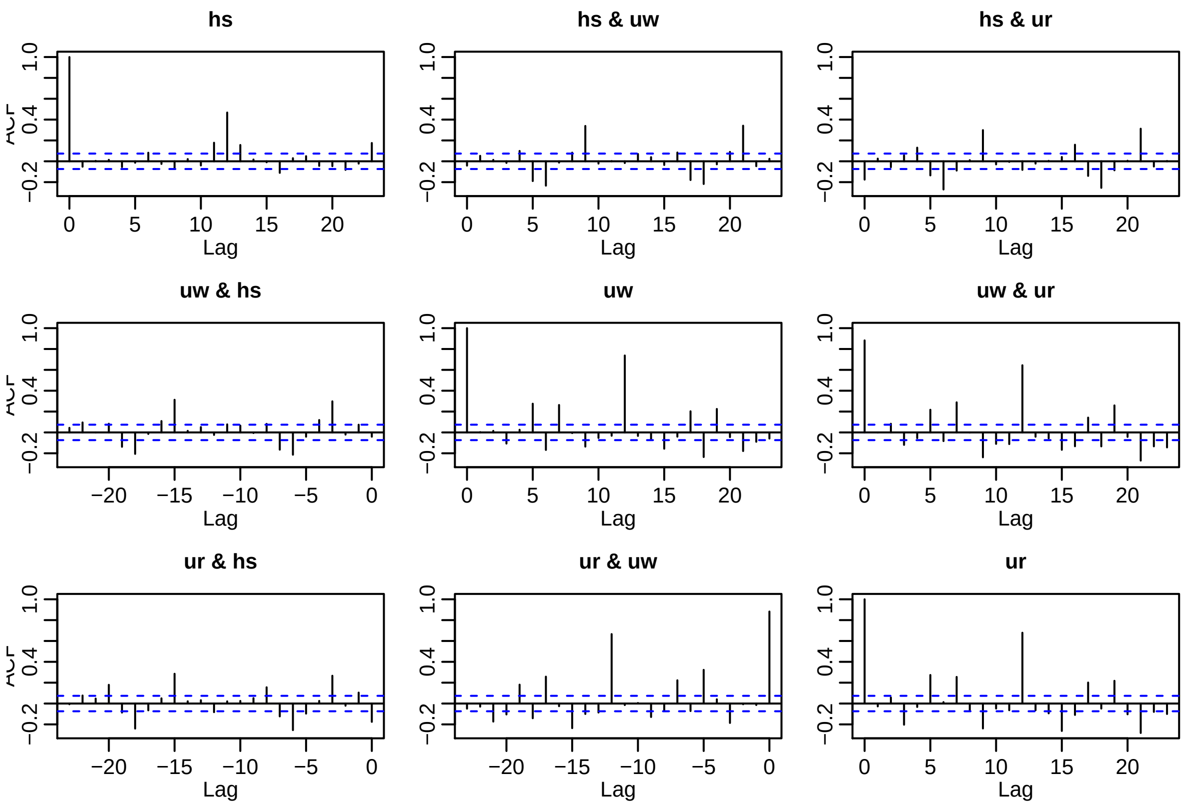
To Identify the VAR Model



To identify a VAR model, we examine the acf/ccf of the 3 variables. We already did this in Section 3, and we saw that there needed to be differencing done to identify a model. Therefore, we examine the acf/ccf of the differenced data. This is shown above. It looks better, and we can more clearly identify a model.

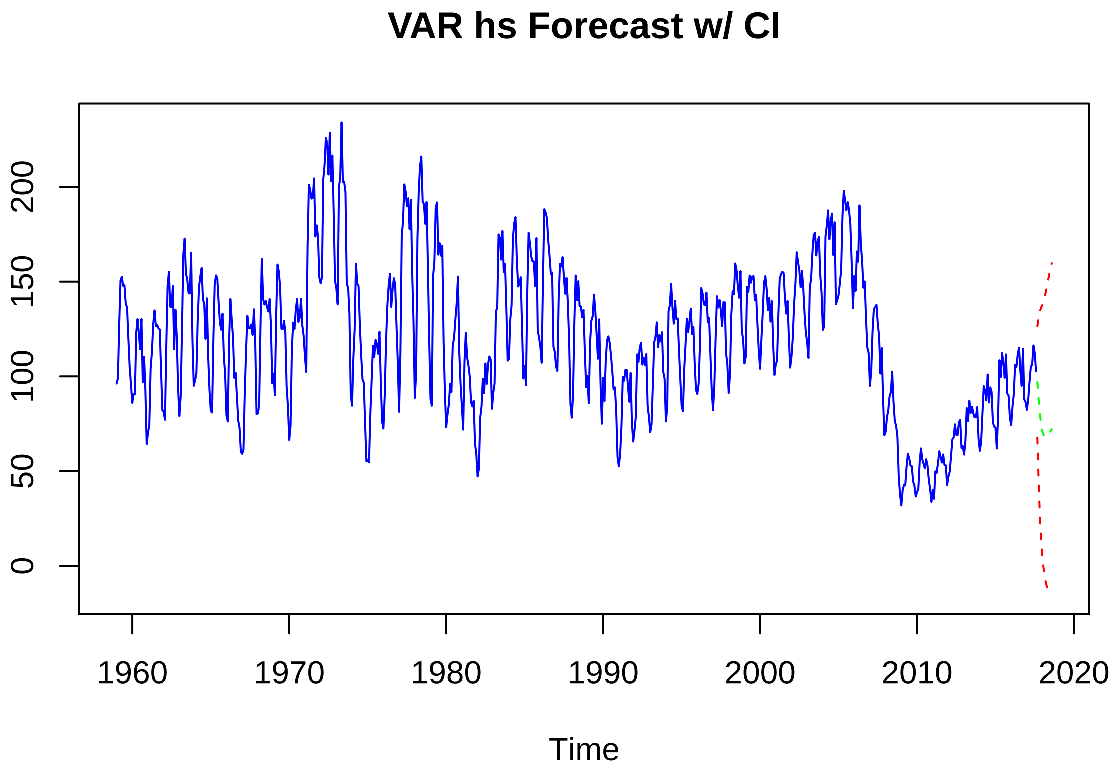
Based on the acfs (on the diagonal), we can see that the autocorrelations follow a sinusoidal pattern, so we can conclude that the 3 variables lag themselves by 2. When examining the cross-correlations, we see that the maximum number of significant lags until dying down is 3. Therefore, we will fit a VAR(3) model.

Residuals of the Model



After fitting the VAR(3) model, we obtained the above residuals. Admittedly, these residuals are not great. While some resemble white noise, others do not. However, when fitting other VAR(x) models, the residuals do not seem to improve enough to change the justification of the model. We will proceed to forecast.

Forecast Plot for VAR



We obtained a RMSE of 34.01208 for the VAR forecast, which is the worst of the three models.

For the equation where hs is the response, the following terms were significant (leading):

- hs_{t-1}, uw_{t-1}, uw_{t-2}, ur_{t-1}, ur_{t-2}, ur_{t-3}, trend

For the equation where uw is the response, the following terms were significant (leading):

- uw_{t-1}, uw_{t-2}, uw_{t-3}, $urt-1$, ur_{t-2}, ur_{t-3}, hs_{t-3}, trend

For the equation where ur is the response, the following terms were significant (leading):

- $urt-1$, ur_{t-2}, ur_{t-3}, uw_{t-1}, uw_{t-2}, hs_{t-1}, hs_{t-2}, hs_{t-3}

Here is the final model, with bold showing statistically significant coefficients:

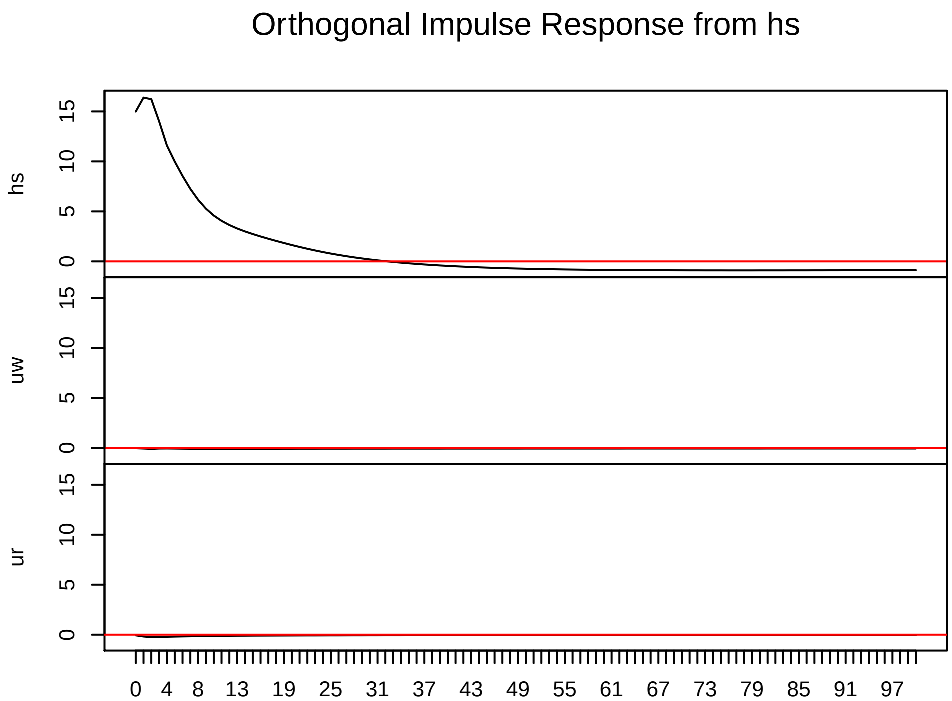
$$\begin{aligned} hs = \mathbf{1.06125655} hs_{t-1} + \mathbf{6.62886534} uw_{t-1} - \mathbf{8.76786611} ur_{t-1} - 0.04626321 hs_{t-2} \\ - \mathbf{11.59203367} uw_{t-2} + \mathbf{28.32152608} ur_{t-2} - 0.02203779 hs_{t-3} - 1.46655379 uw_{t-3} \\ - \mathbf{12.07923594} ur_{t-3} - \mathbf{0.01221720} trend \end{aligned} \tag{2}$$

$$\begin{aligned} uw = -\mathbf{0.0010722404} hs_{t-1} + \mathbf{0.4737445025} uw_{t-1} + 0.3719051484 ur_{t-1} \\ - 0.0002179218 hs_{t-2} + \mathbf{0.3944954511} uw_{t-2} - \mathbf{0.5198975127} ur_{t-2} \\ + \mathbf{0.0037529509} hs_{t-3} - \mathbf{0.1896093923} uw_{t-3} + \mathbf{0.4598845780} ur_{t-3} - \mathbf{0.0004911933} trend \end{aligned} \tag{3}$$

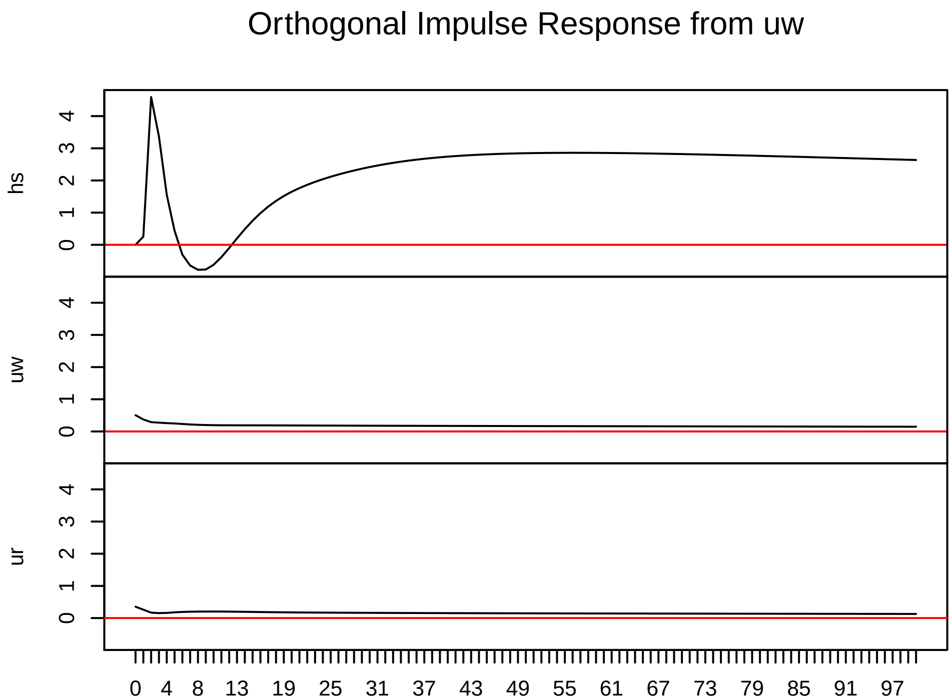
$$
\begin{aligned}
ur = &-\mathbf{0.0066941888}hs_{t-1} - \mathbf{0.4138842102}uw_{t-1} + \mathbf{1.3303632398}ur_{t-1} \\
&+ \mathbf{0.0035123354}hs_{t-2} + \mathbf{0.3019779741}uw_{t-2} - \mathbf{0.4960377890}ur_{t-2} \\
&+ \mathbf{0.0040686253}hs_{t-3} - 0.0031211458uw_{t-3} + \mathbf{0.2745496680}ur_{t-3} - 0.0001489535trend
\end{aligned} \tag{4}
$$

The cointegration and unit root tests done in Section 3 say that the relations are not spurious.

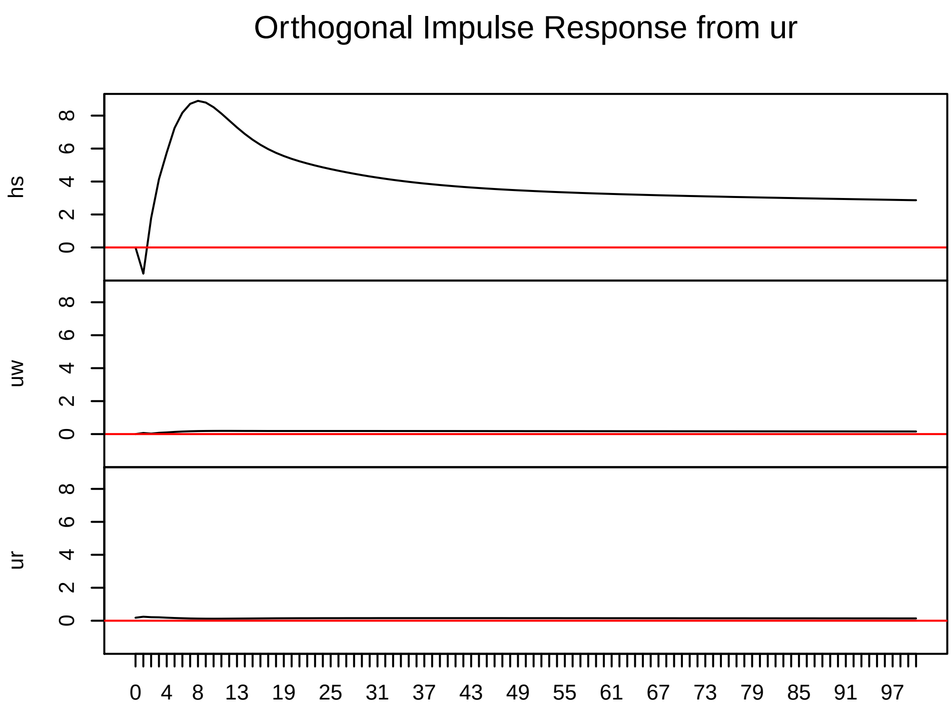
Impulse Response Analysis



We do impulse response analysis to show what a one-unit shock of one of the time series variables would do to the other variables. This can be a useful way of visualizing how one series may lead another, and to see the magnitude of such shocks. Above, we shock hs. What we see is that uw and ur are barely affected, and that hs rises up 15 units, before slowly dropping down back to a stable 0 at around time 50.



Here, we shock uw. Again, uw and ur are barely affected. They experience very minimal rises but quickly stabilize. hs, on the other hand, experiences a sharp rise and fall early on, but then gradually increases until stabilizing at around 3 units at approximately time 40.



Finally, we do a shock of ur. Here, uw and ur are barely affected. hs rises sharply to around 8 units early on, then gradually decreases to a stable 4 units by time 50. What we see from all 3 impulse response plots is that a shock to any of the variables causes a significant reaction in hs, but none of the shocks really affect uw or ur.

Section 7. Forecasts and Conclusions

Time Forecast	SARIMA	Regression w/ GLS	VAR	Consensus	Actual Data
Sept 1, 2017	101.69519	84.52104	97.10529	94.44051	104.4
Oct 1, 2017	104.44352	87.26864	87.49983	93.07066	109.6
Nov 1, 2017	88.12797	118.77163	79.66608	95.52189	97.9
Dec 1, 2017	80.42764	138.82954	74.30901	97.8554	81.4
Jan 1, 2018	77.56587	144.83975	70.63144	97.67902	91.6
Feb 1, 2018	81.02832	144.71710	68.68572	98.14371	89.7
Mar 1, 2018	96.02236	138.51507	67.96996	100.8358	107.2
Apr 1, 2018	106.16091	136.65931	68.12197	103.6474	117.5
May 1, 2018	108.74752	129.34525	68.84569	102.3128	123.7
Jun 1, 2018	112.49573	133.42196	69.86713	105.2616	112.0
Jul 1, 2018	110.25166	109.93937	70.99066	97.06056	111.9
Aug 1, 2018	103.62926	94.68026	72.08754	90.13235	112.6
RMSE	**8.958358**	**32.47106**	**34.01208**	**13.59325**	

The methodology of this analysis centered around a few key aspects of time-series analysis: identifying a model, gaining white-noise residuals (to show that all of the autocorrelation is explained by the model), and using RMSE on the out-of-sample test data to evaluate our models. Some models performed better than others.

As seen, the SARIMA model predicted the housing starts data the best. The RMSE was the lowest out of all of the methods, including the consensus forecast. I would recommend the SARIMA model, but it is always safe to use a consensus forecast. In cases where there is high risk or degree of uncertainty attached to the forecasts, a consensus forecast is appropriate. In our case, the consensus forecast also did well, as the RMSE (13.59325) was relatively low, while the RMSE for Regression w/ GLS and VAR were higher (32.47106 and 34.01208).

Our results show us that for forecasting housing starts, it is best to forecast based off past values of itself (historical). When we used other variables (regression w/ autocorrelated errors and VAR), it performed worse.

Based on these results, I would disagree with those economists who thought that housing starts is a leading indicator. While housing starts may coincide with the trend of the overall economy, unemployment rate seems to be the leading indicator. The results from our impulse response functions showed this, as a shock in housing starts did not lead to much chance in unemployment, but a shock in unemployment drastically affected housing starts.