

1 Appendices

A Research to Include Later

“The estimation of unobserved components: trend-cycle, seasonal and irregular component was made with SEATS program based on ARIMA models. Seasonally adjusted series were obtained by removing the seasonal component from the original data. Trend was obtained by removing the irregular component from the seasonally adjusted series” (?).

“A possible leading indicator variable for the unemployment rate is the number of initial claims of unemployment” (?).

B commentary moved to the end

Figure 1: Forecasting with ARIMA and VAR models

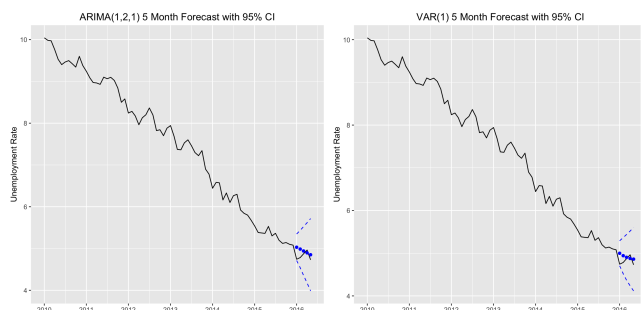
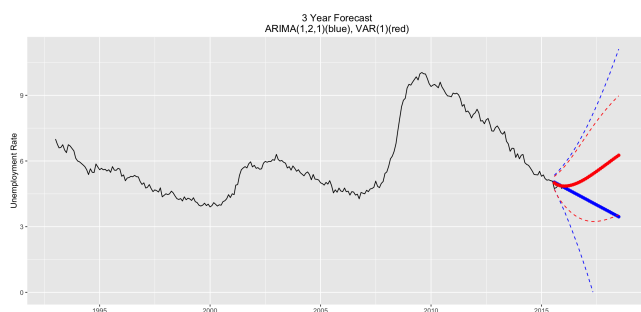


Figure 2: 3 year forecasts



The ARIMA(1, 2, 1) predicts that unemployment will continue to decrease indefinitely, which we know can't be true. The VAR(1) model shows a much more accurate picture in the long run.

I am glad to start doing some forecasting. I did some with the ARIMA(1, 2, 1) seasonally adjusted, no predictors. It's in the RScript "forecasting."

What other potential models are we considering? My only concern is that if we choose a model with predictors, we will have to forecast those predictors before we forecast the unemployment rate.

In case we go with the ARIMA(1, 2, 1) model for the seasonally adjusted data with no predictors, here are some forecast plots. I uploaded them in the Plots folder, too.

The graphs are for the $h = 5, 12$, and 24 step ahead forecasts. The first three were generated by `sarima()`, and the last three by `Arima()`. Personally, I think the last three look better. I think it's good to have a picture of the forecast in the context of all the data. I will play around with `sarima()` to see if I can adjust the default axes to accommodate all past data.

Figure 3: Plots described above

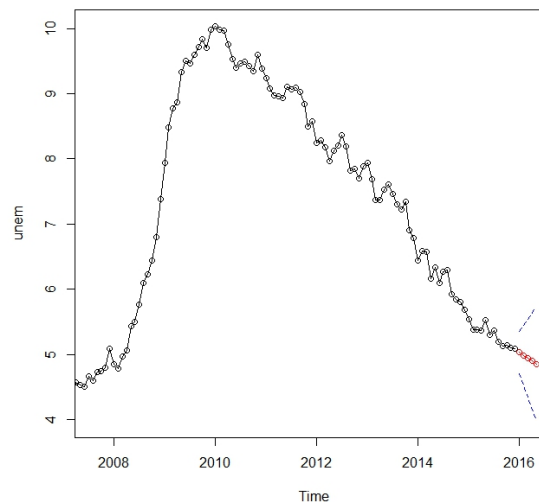


Figure 4: Plots described above

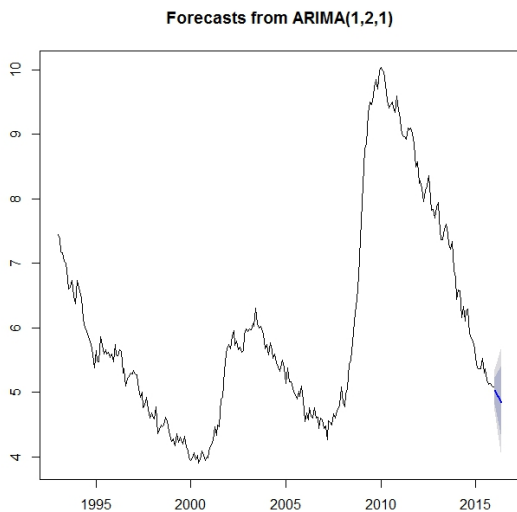
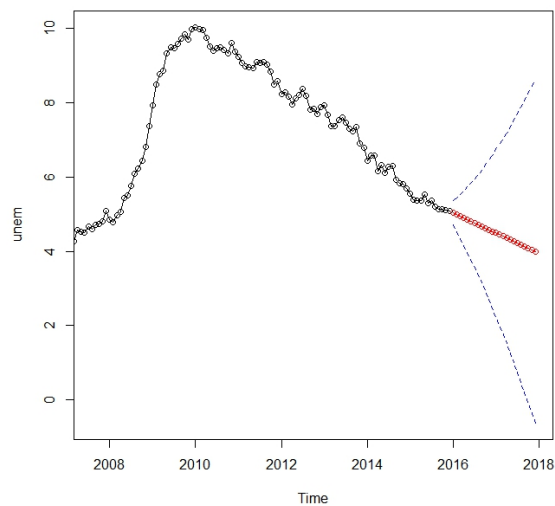
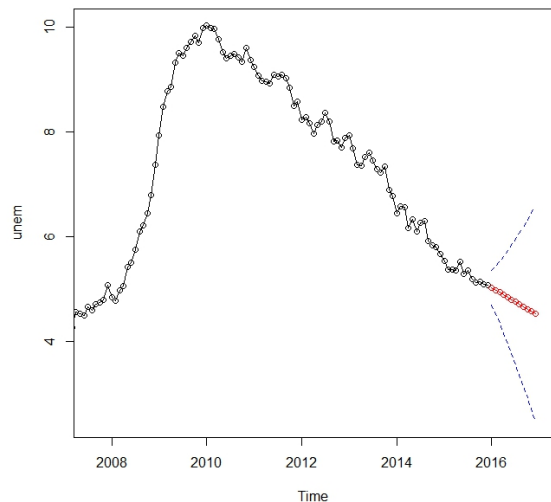


Figure 5: Plots described above

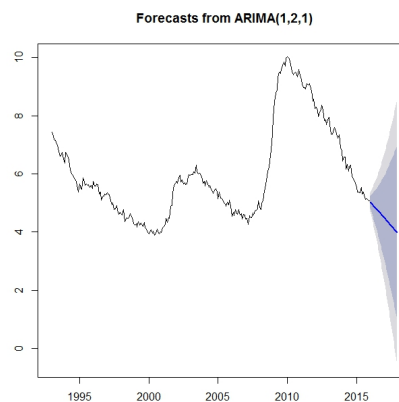
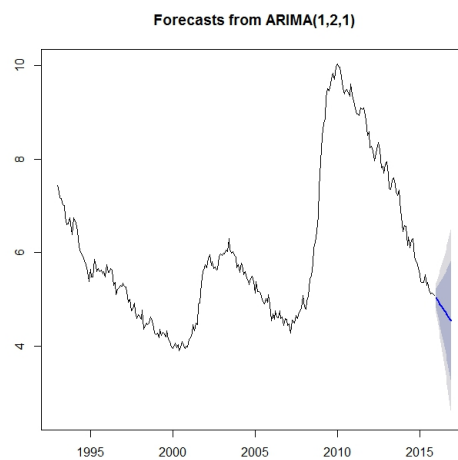
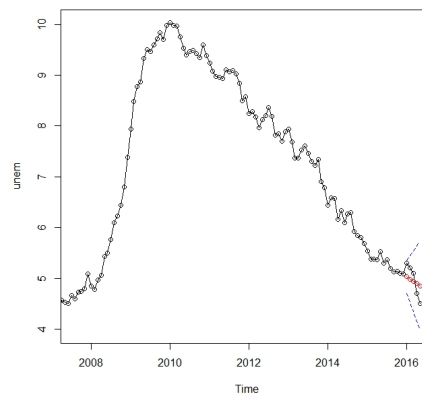


Figure 6: Plot described below



And here is a plot of the first five forecasted values (red) along with the actual observed values (black) from 2016.

I looked at the FRED website where we got our data,

and it looks like the unemployment for June 2016 has been posted at 5.1%. We could compare that to our predictor for June 2016 as well.

Here is a plot from Arima() that shows the predicted values through June 2016 (blue) and the observed values (black).

I put all the code for my plots in the RScript folder and named it “forecasting plots.”

Of the models we have discussed so far, I think the ARIMA(1, 2, 1) is best. It had the best diagnostics and the lowest AIC.

I added some predictors to the ARIMA(1, 2, 1), and only retail seemed significant. However, its coefficient is so small that I argue we don’t need it.

I then did some forecasting for the ARIMA(1, 2, 1) as well as two ARIMA(1, 2, 1) models with predictors. I then compared our predicted values for 2016 unemployment with the actual values:

Jan 2016: actual 5.3 , predicted = 5.0

Feb 2016: actual 5.2 , predicted = 5.0

Mar 2016: actual 5.1 , predicted = 4.9

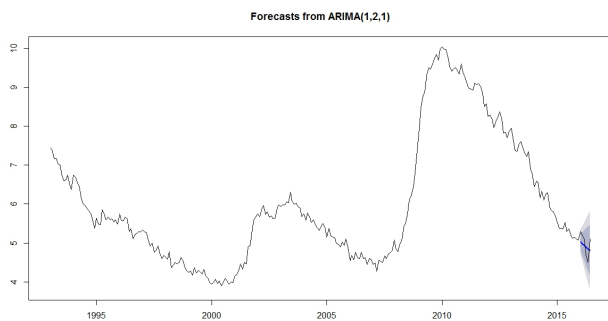
Apr 2016: actual 4.7 , predicted = 4.9

May 2016: actual 4.5 , predicted = 4.9

Overall, I think the ARIMA(1, 2, 1) is very good.

I uploaded all of my code as "forecasting 7_21_16"

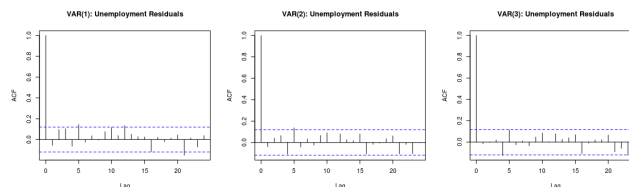
Figure 7: with June 2016



The professor seems to like the idea of splitting the data into training and validation sets. We didn’t split the data but luckily we have the new 5 months data as a validation set. From looking at the plots, it seems hard to distinguish the performance of two models. I computed the mean squared error of forecasting of the two best models. 0.01505823 for ARIMA(1,2,1) and 0.009663836 for VAR(1). This quantitative measure also

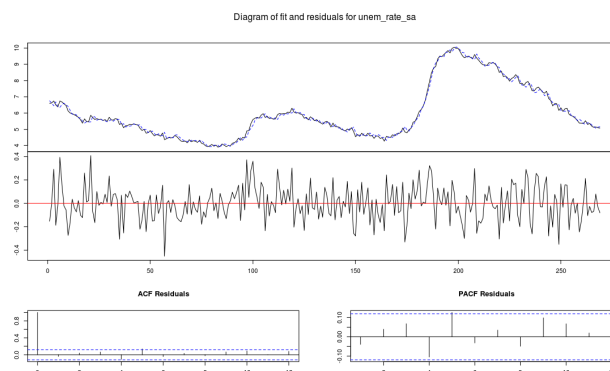
supports this VAR(1) model. Hope this would help a bit when we are comparing the two models.

Figure 8: with June 2016



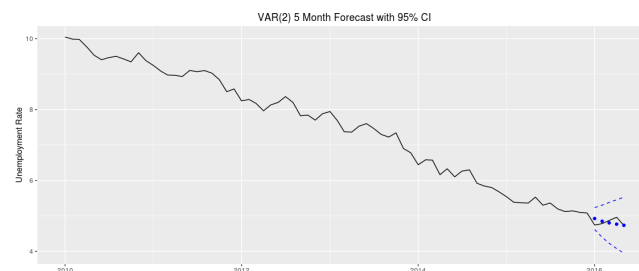
Here is a plot of the unemployment series in the best performing model by AIC: Var(2) with lagged xregs.

Figure 9: fit and residuals



There is also forecasting functionality in the package which is nice because in the case of an ARIMA model with xregs, you don't have to forecast the xregs. Vars will do that for you since all of them are essentially AR(p) models that only use lagged values to forecast.

Figure 10: Var(2) Forecast 5 mo



I also built a few VAR models. By VARselect, BIC suggests VAR(1) HQ suggest VAR(2). The VAR(1) results only show the

retail_sales_sa.l1 and recession_ind.l1

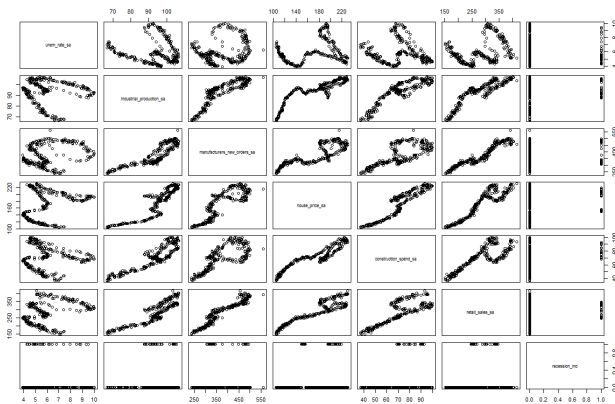
besides

unem_rate_sa.11

were significant predictors. I checked the correlation among these predictors and found that variables

industrial_production, manufacturers_new_orders, house_price_sa, construction_spend, and retail_sales are highly correlated.

Figure 11: Scatterplot matrix



It might be reasonable to leave out some highly correlated variables. Thus, I then fitted two models with only

unem_rate, retail_sales, and recession_ind

. Here are the AICs and BICs.

```
AIC(M1$varresult$unem_rate_sa) # -253.317
AIC(M2$varresult$unem_rate_sa) # -252.6457
AIC(M3$varresult$unem_rate_sa) # -247.1147
AIC(M4$varresult$unem_rate_sa) # -251.6351
```

```
BIC(M1$varresult$unem_rate_sa) # -217.1493
BIC(M2$varresult$unem_rate_sa) # -191.2225
BIC(M3$varresult$unem_rate_sa) # -225.414
BIC(M4$varresult$unem_rate_sa) # -219.117
```

AICs suggest the original VAR(1) model.

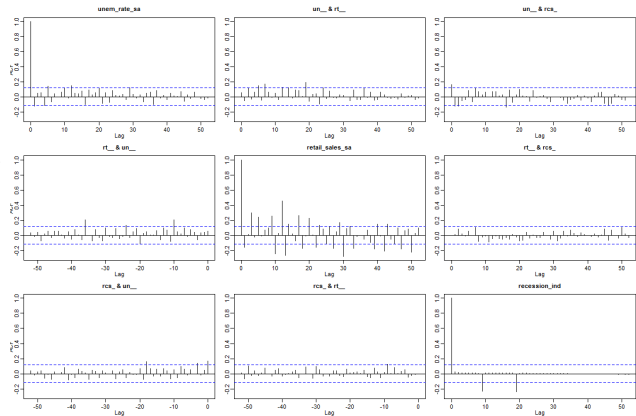
The BICs suggest the VAR(1) with only three variables.

These tables are included earlier.

5 Month Forecasts for the 2 best Models

Since we decomposed and adjusted the seasonal data ourselves, it differs slightly from what you would see on the BLS website so I applied the same seasonal adjustment to the first 5 months of unemployment that came with the original data set. Overall the two plots are very similar.

Figure 12: Plots of above

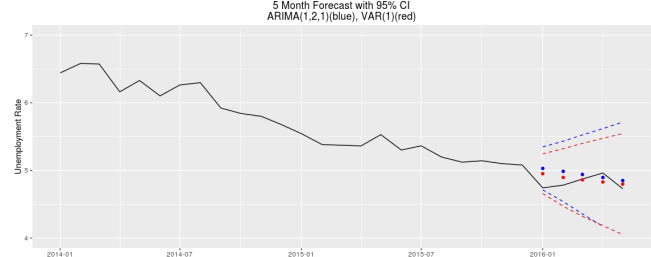


It also looks like the VAR model produced a slightly better forecast over this period, however the confidence intervals of the models overlap substantially.

The forecasts start to look significantly different when you look at the longer term forecasts. This plot shows a 36 month forecast for the two best models. We can see how the confidence interval of the ARIMA model quickly explodes, perhaps indicating that it is not a good choice for long term forecasts.

All of the previously mentioned plots are already included earlier except for:

Figure 13: Other plot



Updated the VAR to not include the insignificant variables I mentioned. The plots in *All_Final_Models.r* will reflect this... here are the updated tables now that those variables have been dropped. This matches the VAR(1) model I posted yesterday.

The professor seems to like the idea of splitting the data into training and validation sets. We didn't split the data but luckily we have the new 5 months data as a validation set. From looking at the plots, it seems hard to distinguish the performance of two models. I computed the mean squared error of forecasting of the two best models. 0.01505823 for ARIMA(1,2,1) and 0.009663836 for VAR(1). This quantitative measure also supports this VAR(1) model. Hope this would help a bit

when we are comparing the two models.

	unemployment	arma.pred	arma.pred.lwr	arma.pred.upr	var.pred	var.pred.lwr	var.pred.upr	dt
73	4.74				5.00	4.71	5.29	Jan
74	4.78				4.94	4.52	5.36	Feb
75	4.87				4.90	4.37	5.44	Mar
76	4.96				4.88	4.24	5.52	April
77	4.73				4.86	4.12	5.60	May

	unemployment	arma.pred	arma.pred.lwr	arma.pred.upr	var.pred	var.pred.lwr	var.pred.upr	dt
73	4.74	5.03	4.71	5.35	5.00	4.71	5.29	Jan-16
74	4.78	4.99	4.54	5.43	4.94	4.52	5.36	Feb-16
75	4.87	4.94	4.36	5.53	4.90	4.37	5.44	Mar-16
76	4.96	4.90	4.17	5.62	4.88	4.24	5.52	Apr-16
77	4.73	4.85	3.99	5.72	4.86	4.12	5.60	May-16

Table 1: 2016 Predicted Values

	unemployment	arma.pred	arma.pred.lwr	arma.pred.upr	var.pred	var.pred.lwr	var.pred.upr	dt
73	4.74	5.03	4.71	5.35	5.00	4.71	5.29	Jan-16
74	4.78	4.99	4.54	5.43	4.94	4.52	5.36	Feb-16
75	4.87	4.94	4.36	5.53	4.90	4.37	5.44	Mar-16
76	4.96	4.90	4.17	5.62	4.88	4.24	5.52	Apr-16
77	4.73	4.85	3.99	5.72	4.86	4.12	5.60	May-16

Date	Actual Unemployment	Prediction	Lower CI	Upper
Jan-16	4.74	5.03	4.71	5.35
Feb-16	4.78	4.99	4.54	5.43
Mar-16	4.87	4.94	4.36	5.53
Apr-16	4.96	4.90	4.17	5.62
May-16	4.73	4.85	3.99	5.72