3RD DECEMBER 2020

THE GEORGE WASHINGTON UNIVERSITY

HAILEE KIM

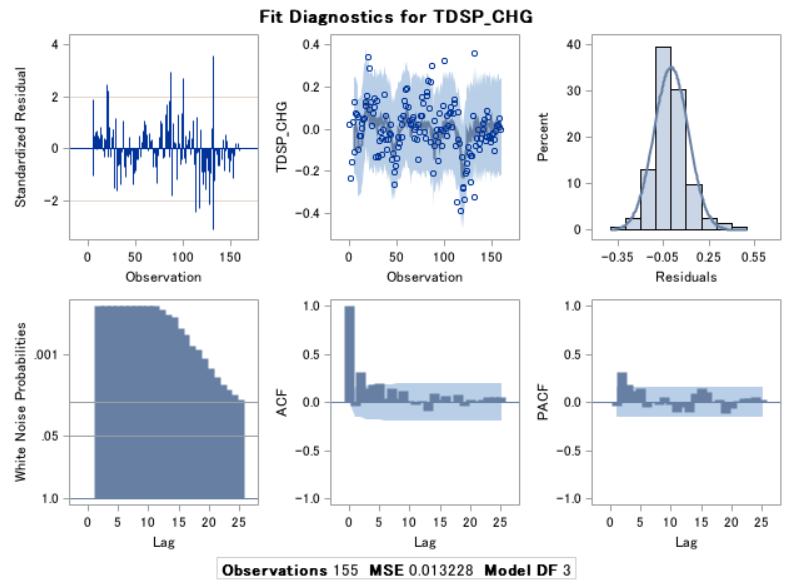
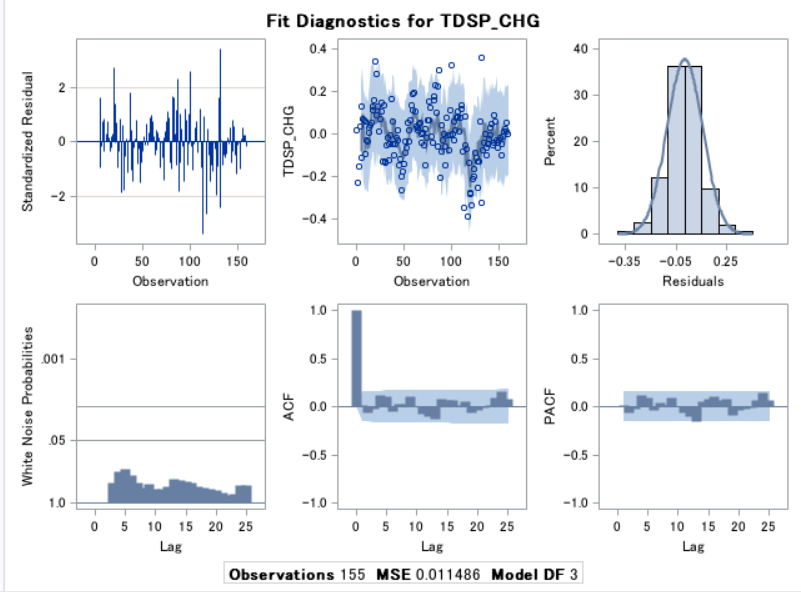
TIFFANY TIONO

PROFESSOR MEJIA

ASSIGNMENT 3

FINA 6271 FINANCIAL MODELLING/ECONOMETRICS

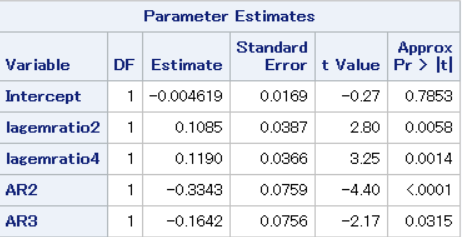
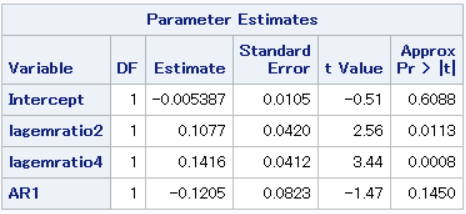
**1. Regression with Autocorrelated Errors**

**Use proc AUTOREG, as illustrated in class, to run a regression where the household debt variable is a function of lag 2 and lag 4 of the employment ratio variable. Show the SAS results. Describe observations related to the regression and the error, and indicate whether setting “nlag= (1)” is sufficient to make the final equation residuals white noise.**

The Second Model

The Original Series

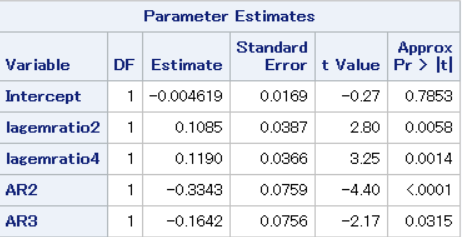
First of all, we generate the proc autoreg to run a regression in the lag 2 and lag 4. We put “nlag = (1)” for the original model. As a result, it is not sufficient enough to make the final equation residuals are white noise. As we can see in the original bar graph white noise probabilities are all low (high bars) indicating that the obtained residuals are not white noise. So, we need to add a couple more lags in addition to the nlag to make the final equation residuals are white noise. Therefore, we added lag 2 and lag 4 in the “nlag” since both ARs in their p-values are statistically significant. As a result, in the second model, we made the final equation residuals are white noise. As we can see that the white noise probabilities in the second model, the bar graph white noise probabilities are all high (zero and low bars) indicating that the obtained residuals are white noise.

**If your conclusion is that the final equation residuals are white noise after setting “nlag=(1)”, explain why. If your conclusion is that the final equation residuals are not white noise after setting “nlag=(1)”, explain why and perform any necessary changes to the “nlag” setting to make them white noise. Show the SAS outputs as needed to support your explanation.**

The Second Model

The Original Series

Achieving white noise is important for time series forecasting because it makes all the variables of a time series have the same variance and each value has a zero correlation with all other values in the series. In other words, white noise prevents the series from happening the sudden changes such as shocks (unnecessary residuals) and making it stable from the independent variable that are identically distributed with a mean of zero. As the white noise time series is random, so it makes us not be able to reasonably model it and make predictions. All of the signal information in the time series should be harnessed by the model to make predictions in order to make forecast errors white noise. That is, all that is left in the forecast errors are the random fluctuations that cannot be modeled. If there is any indication that possibly needs to be further improved, therefore, then that is going to affect the series not to be white noise. Since we need to make a final equation residual are white noise, we need to change the nlag for AR to become significant to create a white noise residual. For example, in the original model, we can see that AR 1 is not significant. Then, we try the possibilities from AR 1 to AR 20 and we throw away the one that is not significant. Therefore, we got the result in the second model that AR 2 and AR 3 are the only AR that are statistically significant. The white noise possibilities are higher than the previous original model.

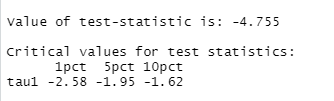
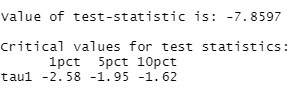
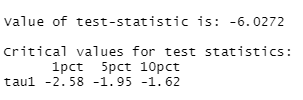
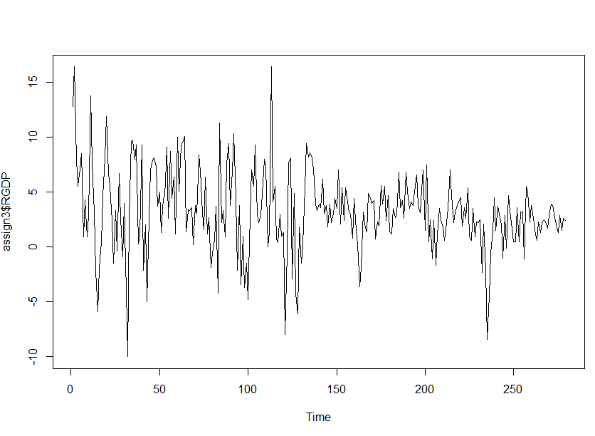
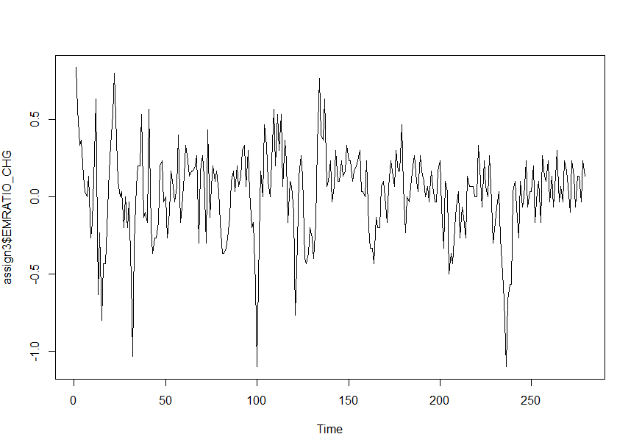
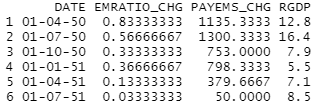
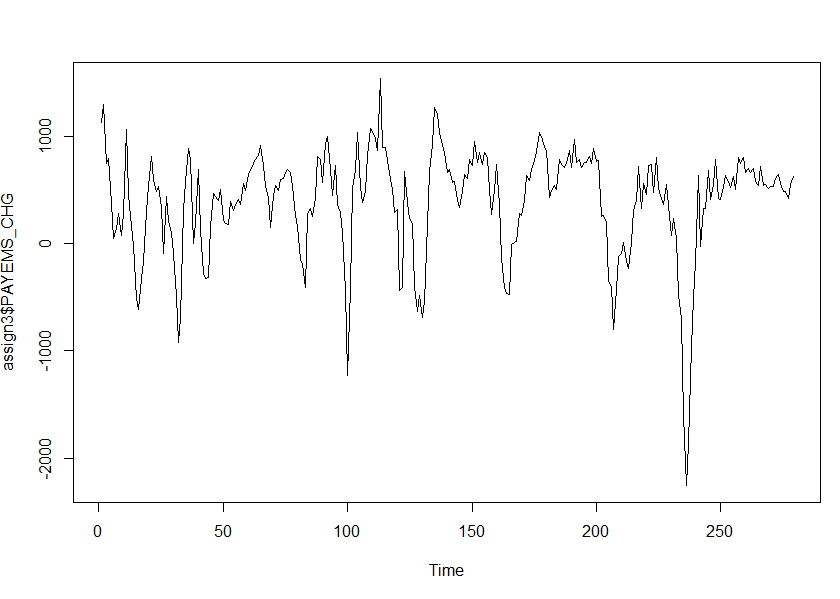
**Describe your final “nlag=” settings. Comment on the final model results, including comments on the explanatory variable coefficients and significance, after modeling the ovides you basically with the same information. Shouldn’t we consider that option?”**

We should consider the option that the lagemratio2 and lagemratio4 should have positive coefficients, which is the employment ratio variables positive impacts on debt. As we can see, both of lagemratio 2 and lagemratio4 p-values are less than 0.05, so we can conclude that both models are statistically significant.

**As you are presenting the results, your supervisor makes a statement and raises a question: “I don’t understand the results, especially those about the so-called ‘autocorrelated’ errors. I think it would be easier to build an ESM or ARIMA model on the household debt series. It’s simpler and provides it reasonable or not? Why? How would you respond? (max. 200)**

It should be reasonable to build an ARIMA model on the household debt series since ARIMA is more general and allows certain non-stationary time series and even stationary series that cannot be fit by low order autoregressive models. Moreover, the autoregressive integrated moving average (ARIMA) is consisted of the two terms of the autoregressive model (AR) and the moving average (MA), which all use the same class of model like the same dataset, time analysis, and historical context to predict future values. Furthermore, ARIMA requires p lags and q lags if applicable, but autoreg can choose depending on what we are going to find. In this case where we use the autoreg function instead, we only need to apply q lags. Autoreg function does not require us to indicate whether it has stationarity, while ARIMA does. Finally, as the autoregressive model refers to modeling of the series and the moving average refers to modeling of the lags of the error, we not only need the AR but also the MA components in order to find out the significant variables and achieve white noise which is the final goal of the modeling.

**2. Autoregressive Distributed Lags (ARDL)**

******Use the “head” function to show the first six lines of your dataset. Then, use the “ts.plot” function to produce plots for the three variables. Lastly, apply the “summary (ur.df (…)” step with “lag=1” to perform a unit root test (i.e., to see if you can reject the non-stationarity hypothesis). Show and explain the ADF results.**

EMRATIO T-Statistic

RGDP T-Statistic

PAYEMS Time series

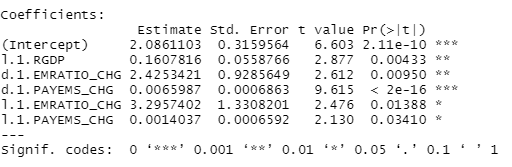
PAYEMS T-Statistic

EMRATIO Time series

First Six Data Results

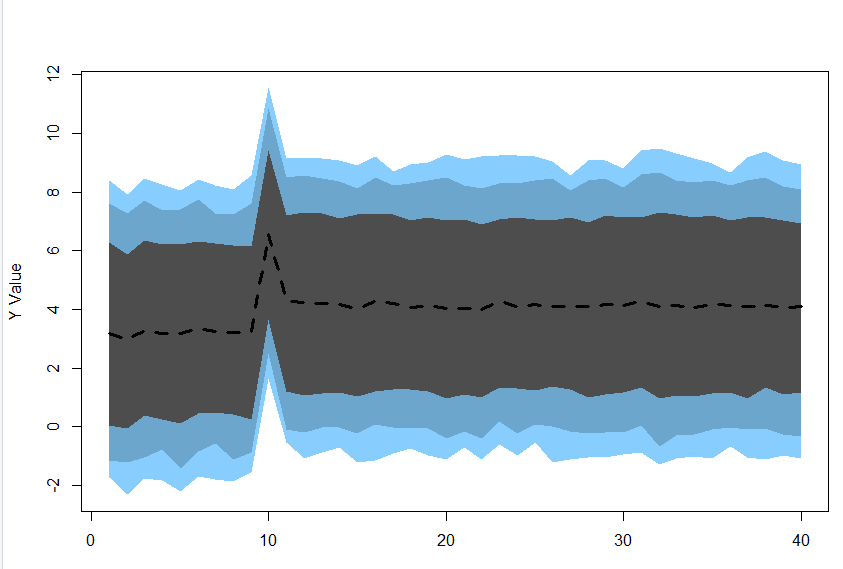
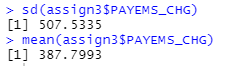
RGDP Time series

For the ADF test, we conclude that we can reject the null hypothesis of non-stationarity since all the T-statistics are less than the critical values.

**Show the R results. Discuss the estimates, pointing to the explanatory variables in the model, the corresponding coefficients, and their significance. Also, as background for the discussion the expectation is that all coefficients are positive (and hopefully significant) since the employment ratio and employment levels are presumed to have a positive effect on GDP growth. Use this assumption as you review the results.**

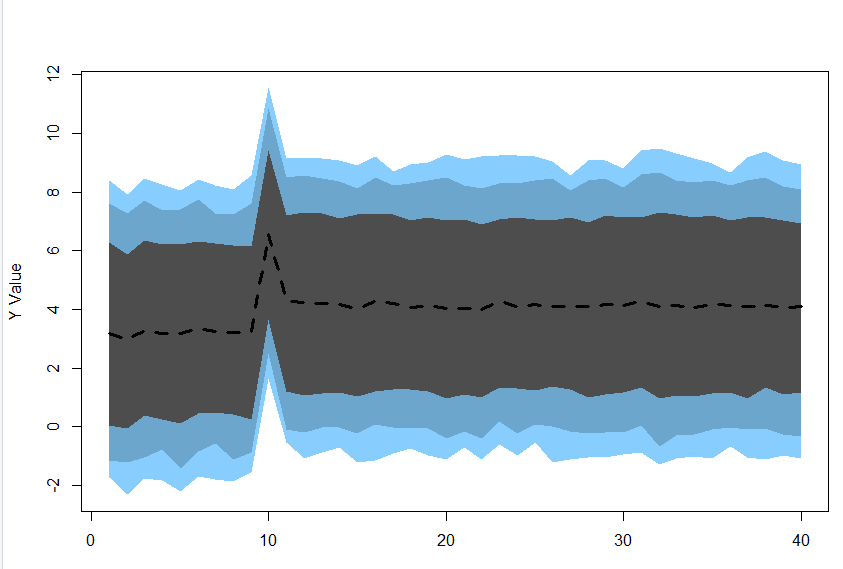
According to our result, for one period lag, all the variables (employment rate and non-farm employment) are statistically significant to the dependent variable (RGDP) since the all p-values are less than 0.05. Moreover, in the one-period differencing (d.1), the employment ratio and non-farm employment are also statistically significant since all p-values are less than 0.05. Therefore, we can conclude that the employment ratio and non-farm employment levels have a positive effect on the GDP growth rate.

**Since the default setting for the shock variable is one standard deviation, you may find it useful to estimate the value of one standard deviation in the non-farm employment variable. You can do this in Excel or in R using the “sd” function. You do not need the standard deviation information to complete the analysis, but you may find it useful to have it to interpret the response function (i.e., to understand the magnitude of the impulse response to whatever change in the shock variable that one standard deviation represents).**



Forecast with Shock in R Result

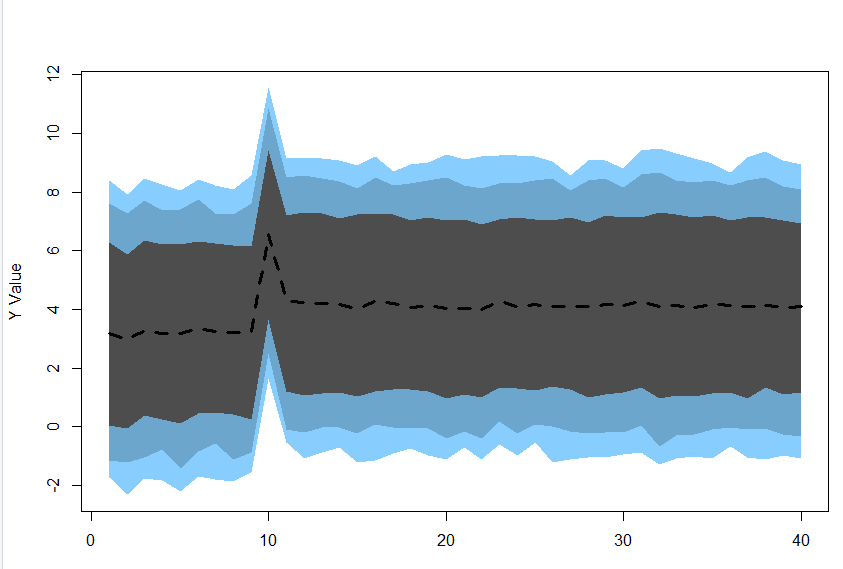
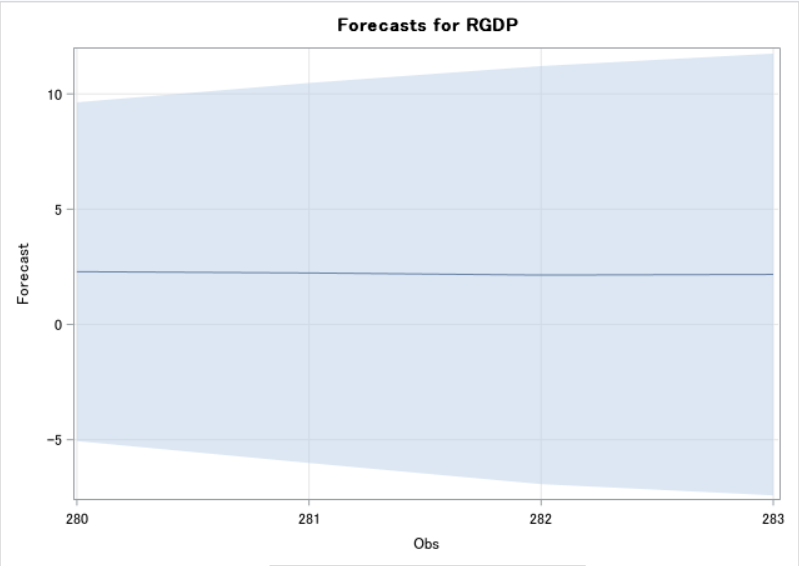
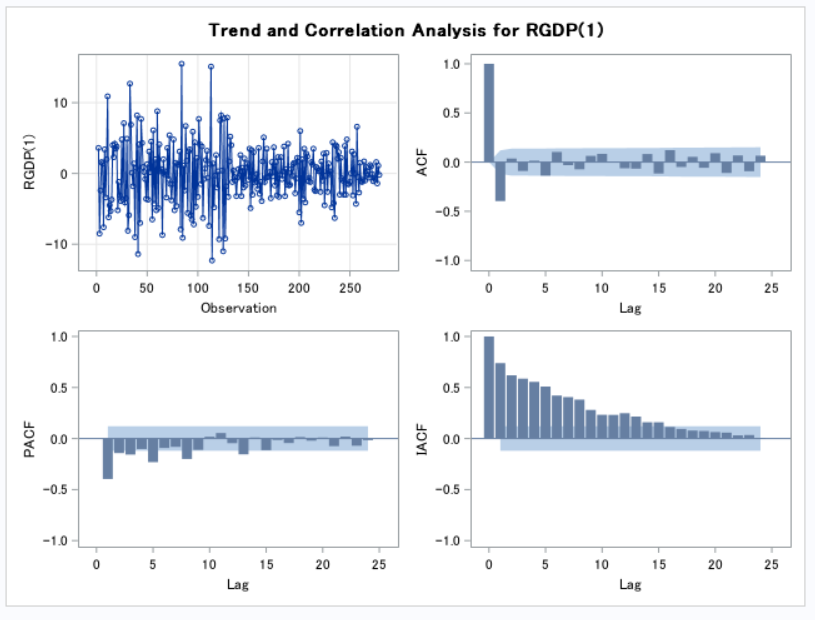
As calculated by the standard deviation divided by the mean, we got 1.31, which indicates that it has a high variance. The higher the coefficient of variation gets, the greater the level of dispersion around the mean gets. In other words, the greater the value of the coefficient of variation, the more precise the estimate. Moreover, with the “dynardl” function, this graph shows a shock of 1 standard deviation at the time period of 10. Furthermore, a change in the non-farm employment variable that is one standard deviation different from the mean change is also explained by the amount of change in dependent variable.

**Based on the chart, roughly state the expected reaction in GDP percentage growth to a shock in one standard deviation in non-farm employment. Point to the timing of the response function, indicating whether the response lasts for more than a few periods or goes away quickly.**

Forecast with Shock in R Result

Following the chart above, the shock goes up and down so quickly so that we can say it does not last for long, but goes away quickly. The shock just starts going up in approximately a time period of 9.7 in the Y value of approximately 3. Then, it goes up to the peak of the shock in the time period 10 and Y value of approximately 6.5. Finally, the shock goes down again in approximately a time period of 10.3 and Y value roughly 4.3. So, the simulation of the chart shows the effect of the non-farm employment shock.

**As with the last analysis, you have a last-minute comment and question. You receive an email from a colleague saying: “This type of model is too complex. It’d be easier to run an ARIMA model with intervention events. You can first run a classic ARIMA model on the GDP variable. Then you can run an ARIMA model with intervention events and shock the GDP variable. And then you can compare the two forecasts and see what the impact of the shock is. That way you don’t need to deal with any explanatory variable” Address this statement. How would you respond?**



Forecast with Proc ARIMA Intervation Result

Forecast with Shock in R Result

In the R programming, there was a shock in period 10 that caused the roughly unexpected growth in GDP because of non-farm employment making standard deviation further away from the mean. As a result, driven by R programming, we confirmed that all of non-farm employment, GDP and non-farm employment that all the variables are stationary. In Proc ARIMA with Intervention, in order to test stationarity, we look at the ACF and PACF graphs and see that there are some possibilities of stationarity in those variables. On the other hand, the IACF graph shows over-differencing in the analysis of time series, which results in not rejecting the null hypothesis of non-stationarity. Moreover, in PROC ARIMA, the forecast result doesn’t necessarily show a dramatic change overtime. Following the rules for number of lags by Schwert 1989, we look at ADF as a tool to test short-term stationarity behavior, and we can reject the null hypothesis of non-stationarity once there are no dramatic change in between short-term autocorrelation behavior.