Econ 8011 Final Assignment DUE: Wednesday, April 28

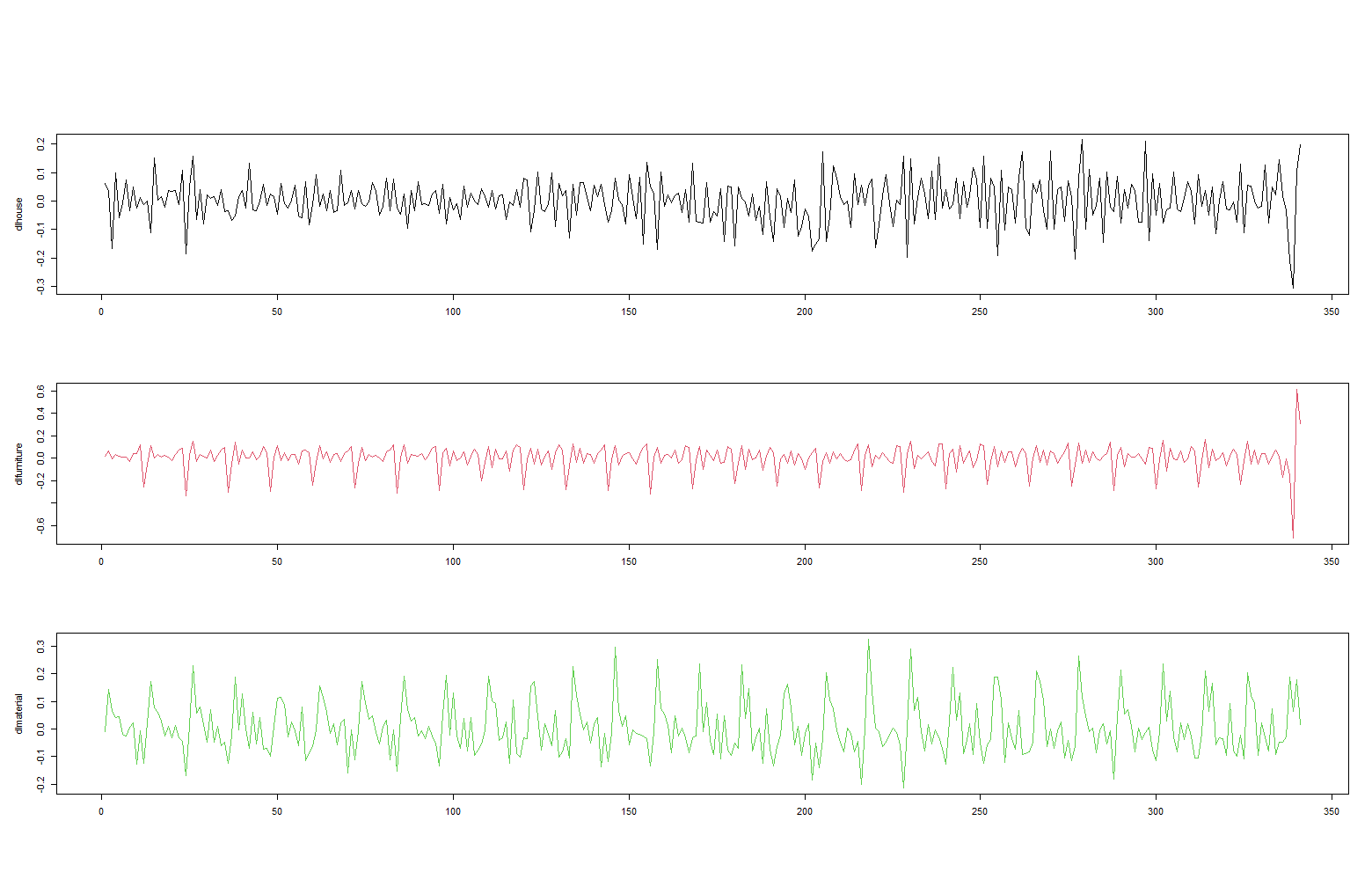
Zach Taylor

**1. Create a sub-sample of your data for model building and estimation purposes that removes the last 8 observations from the data set (for later use for out-of-sample forecasting). As your answer to this question, provide the source(s) of your data, the variable definitions, etc., and a table (after differencing if needed) with the mean, max, min, S.D and number of obs. for each variable (using only the sub-sample). Use the sub-sample to answer the following questions.**

To predict monthly spend at furniture and home furnishing stores, this paper will be models utilizing monthly spend data on building materials and monthly data on the number of new constructions of privately owned housing units. Below are the logs of the sub samples (all but final 8 observations

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Variable | Count  Count | SD | Min | 1QR | Median | Mean | 3QR | Max |
| lhouses.train | 336 | 0.07782 | -0.3065 | -0.04565 | -0.00171 | 0.000214 | 0.052 | 0.21478 |
| lfurniture.train | 336 | 0.11387 | -0.7125 | -0.03587 | 0.020497 | 0.002714 | 0.071 | 0.61319 |
| lmaterials.train | 336 | 0.09407 | -0.2132 | -0.05813 | -0.00921 | 0.004652 | 0.052 | 0.32387 |

Differenced once, logged, sub sample



U.S. Census Bureau, Retail Sales: Building Material and Supplies Dealers [MRTSSM4441USN], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/MRTSSM4441USN, April 22, 2021.

U.S. Census Bureau, Retail Sales: Furniture and Home Furnishings Stores [MRTSSM442USN], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/MRTSSM442USN, April 22, 2021.

U.S. Census Bureau and U.S. Department of Housing and Urban Development, New Privately-Owned Housing Units Started: Total Units [HOUST], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/HOUST, April 22, 2021.

**2. Determine the order of integration of each of the three variables.**

Each variable was analyzed on four characteristics to determine the order of integration. Houses and materials satisfy every condition for being integrated of order one, while furniture initially satisfies only 3 conditions. The ADF Test for furniture results in a value of -2.72 which is slightly more negative than the critical value of -2.57 and means it is technically in the rejection range and potentially integrated of order 0. However, as the other three conditions are satisfied for furniture being integrated of order one and the ADF test is only slightly more negative than the rejection point, furniture will be considered integrated of order 1 and also differenced once.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Variable | SD | SD of 1st diff | SD of 2nd Diff | ADF Test | ADF Test 1st diff | ACF 1st diff | Order of Integration |
| lhouses.train | 0.3585 | 0.0779 | 0.1263 | -1.2 | -16.65 | 2nd lag | I(1) |
| lfurniture.train | 0.2386 | 0.114 | 0.1773 | -2.72 | -6.54 | 3rd lag. Evidence of seasonality at 12 | I(1) |
| lmaterials.train | 0.3302 | 0.0942 | 0.1199 | -1.84 | -3.96 | 2nd lag. Evidence of sesaonality at 12 | I(1) |

**3. Determine whether any of the variables are cointegrated.**

To be cointegrated, variables must be integrated of the same order, and not stationary. All three variables satisfy these conditions as they are all I(1).

Both versions of the Johansen test show that there is at least one cointegrated relationship between the variables because the test statistics for r = 0 are larger than the critical values at 1%.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | Critical Value at significance level: | | |
| Johansen test type | Test Statistic when **r = 0** | 10% | 5% | 1% |
| Max eigenvalue | 64.43 | 19.77 | 22 | 26.81 |
| Trace | 73.54 | 32 | 34.91 | 41.07 |

Engle-Granger methodology was followed to determine which variables were cointegrated. After regressing all variables on each other, the resulting residuals were examined for stationarity. As shown in the table below, all three regressions resulted in stationary residuals indicating there is a cointegrative relationship and that an error correction term is necessary.

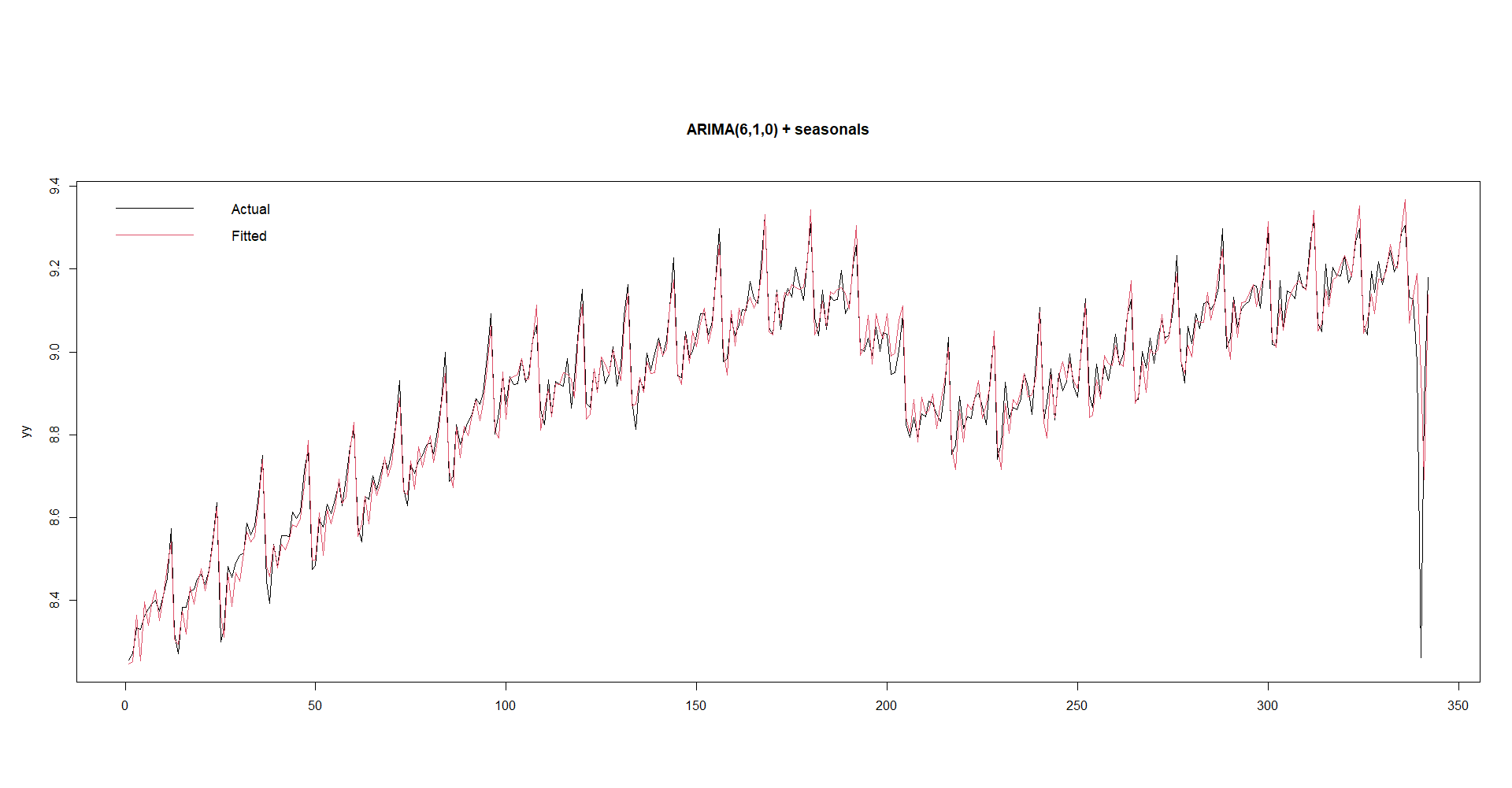
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Regression | SD of res | SD of 1st diff of res | SD of 2nd Diff of res | ADF Test | ADF Test 1st diff | ACF level | Order of Integration |
| dlhouse~dlfurniture + dlmaterial | 0.0768 | 0.126 | 0.2272 | -17.13 | -10.84 | 1st lag | I(0) |
| dlfurniture~dlhouse + dlmaterial | 0.1037 | 0.1573 | 0.2661 | -8.11 | -16.78 | 3rd lag | I(0) |
| dlmaterial~dlhouse + dlfurniture | 0.0866 | 0.1043 | 0.1698 | -4.29 | -18.48 | 1st lag | I(0) |

**4. Build an ARIMA model for your focus variable (only one of the three variables in your data set) choosing appropriate values for p, d and q. Be sure to examine AIC, BIC, serial correlation test results, potential seasonality, etc. to choose your ‘best’ model.**

ARIMA model was constructed by testing the number of lags to include of the furniture variable and optimizing the R2, AIC, and BIC results. The first table below shows these results. A model with the specification of (6,1,0), (or 6 lags of the variable, differenced once, and with no moving average term) minimizes the AIC and has almost the smallest BIC while maximizing the R2 value. This (6,1,0) ARIMA can explain 95.7% of the variation in dlfurniture.

The Breusch Godfrey and Box-Ljung tests were used to ensure no serial correlation was present in the error term. The Breusch Godfrey results (middle table below) show that there is no serial correlation for the first 4 lags and the Box-Ljung test does not produce significant values for any lags 1-20 and therefore serial correlation is not present in the chosen model for lfurniture.train, ARIMA (6,1,0).



**5. Build a dynamic regression model (or ECM if appropriate) for your focus variable using the other variables you have chosen to help improve the model of your focus variable.**



To test if the seasonal variables improve the model, a joint F test is performed which compares the best model to the restricted model (best model less the seasonal dummy variables) to see if there is a significant difference. This F test results in a small p value of 2.2e-16 which causes the rejection of the null hypothesis that the model with seasonal dummies produces a result no different than the model without seasonal dummies. Therefore, seasonality is important and will be included.

No evidence of Serial Correlation was shown in the Bresch Godfrey test results so we fail to reject the null hypothesis that there is no serial correlation in the error. This indicates that the seasonal dummies are useful in explaining variation as well.



**6. Build a VAR/VECM using the other variables you have chosen to help improve the model of your focus variable. Be sure to examine AIC, BIC, serial correlation test results, potential seasonality, etc. to choose your ‘best’ model.**



Following a similar process as when choosing the best dynamic regression, a VECM was built. The best model used 4 lags:

Δtdlfurniture𝑡=𝛼1+∑𝑝𝑖=1𝜙1𝑖 Δtdlfurniture𝑡 𝑡―𝑖+∑𝑝𝑖=1𝛽1𝑖 Δtdlhouse𝑡 𝑡―𝑖+𝜃1𝑒𝑡―1+∑𝑝𝑖=1δ1𝑖 Δtdlmaterial𝑡 𝑡―𝑖 𝑢1𝑡, 𝑢1𝑡~𝑁(0,𝜎21)

Δtdlhouse𝑡=𝛼2+∑𝑝𝑖=1𝜙2𝑖 Δtdlfurniture𝑡 𝑡―𝑖+∑𝑝𝑖=1𝛽2𝑖 Δtdlhouse𝑡 𝑡―𝑖+𝜃2𝑒𝑡―1+∑𝑝𝑖=1 δ 1𝑖 Δtdlmaterial𝑡 𝑡―𝑖 + 𝑢2𝑡, 𝑢2𝑡~𝑁(0,𝜎22)

Δtdlmaterial𝑡=𝛼2+∑𝑝𝑖=1𝜙2𝑖 Δtdlfurniture𝑡 𝑡―𝑖+∑𝑝𝑖=1𝛽2𝑖Δ𝑥𝑡―𝑖+𝜃2𝑒𝑡―1++∑𝑝𝑖=1 δ 1𝑖 Δtdlmaterial𝑡 𝑡―𝑖 + 𝑢2𝑡, 𝑢2𝑡~𝑁(0,𝜎22)

Error correction term: 𝑒𝑡―1=furniture―1―𝑎―𝑏house𝑡―cmaterial𝑡 ―1

Show the outputs of the 3 equations and describe how it minimizes AIC and BIC (shown above). The equations for dlfurniture and dlmaterial have high R2 values and all of the lags are significant. The equation for dlhouse has a low R2 value and most of the lags are not significant. However, this overall is the best when judging by AIC, BIC, and R2.

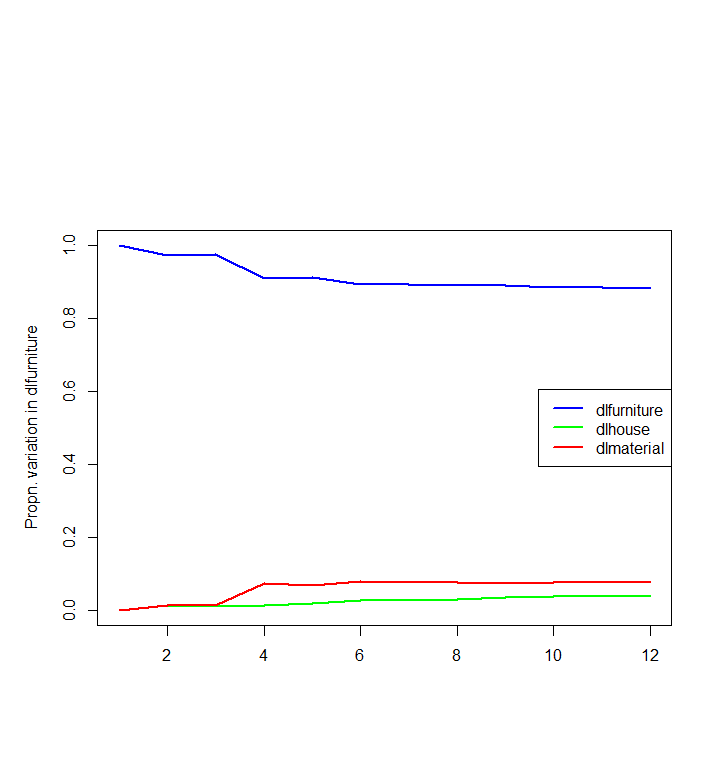
**7. a) Perform Granger causality testing**

To test for Granger causality, three F-test’s are performed to compare the best model with models missing lags of each variable. The resulting p-value indicates the missing lags in each equation are significant and therefore it can be said that each variable granger causes the dependent variable, dlfurniture.



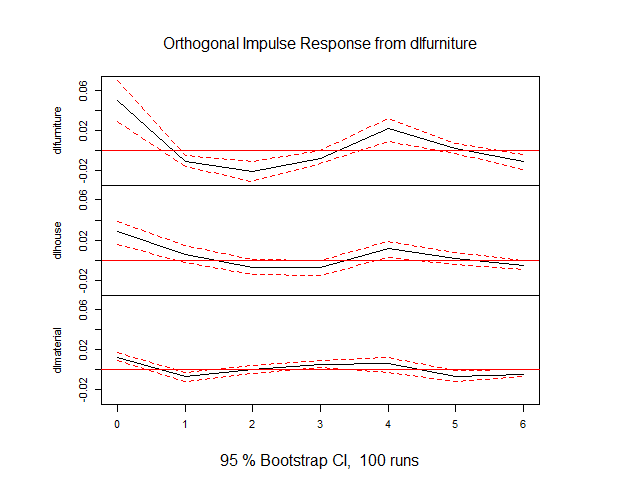
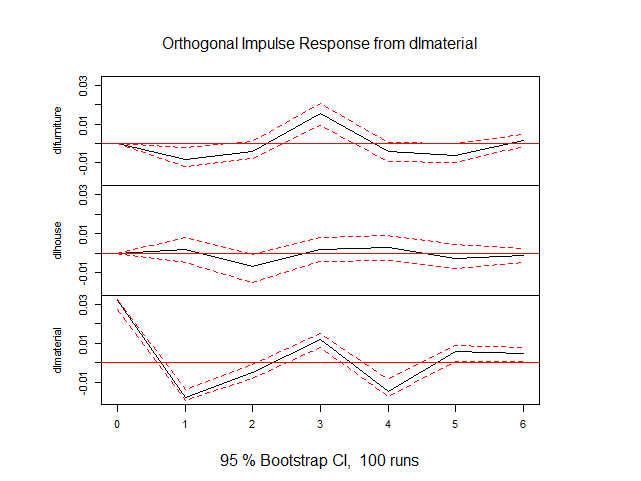
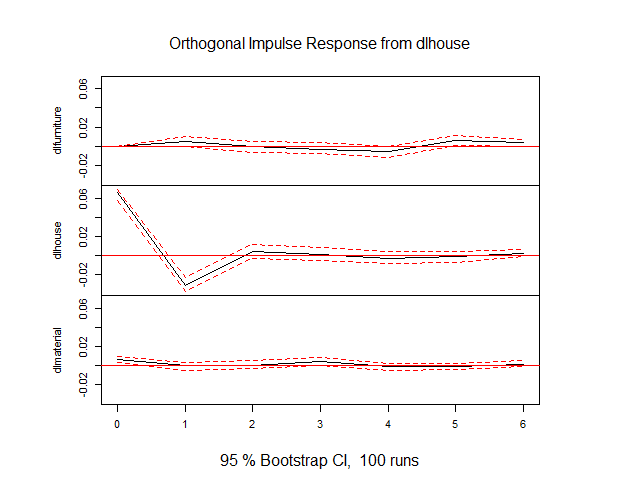
**b) generate variance decompositions and**

Decomposing the variance shows that lags of dlfurniture can explain approximately 89% of the variance in dlfurniture depending on the lag length. Dlmaterial is the second most important variable as it can explain about 8% of the variance in dlfurntiture, and dlhouse can explain an average of 3% of the variance in dlfurniture across 12 lags.



**c) impulse response functions to assess the effect of a change in each of the other variables in your model on your focus variable.**

Impulse response plots were generated for each variable with a 95% Bootstrap CI and 1000 runs.The log of each variable is utilized so interpretation of the impulse response function is straightforward. A 1% change in the impulse variable results in a x% change in the response variable. It is clear that all three variables receive about a 2% increase from a 1% increase in the other variables that dies out quickly, usually by the 4th lag. Each variables strongest impulse received is from its own lags, other than furniture which also receives a noticeable shock from dlmaterial. This is in line with the variance decomposition above which shows that dlmaterial and dlfurniture can account for virtually all of the variance in dlfurniture.



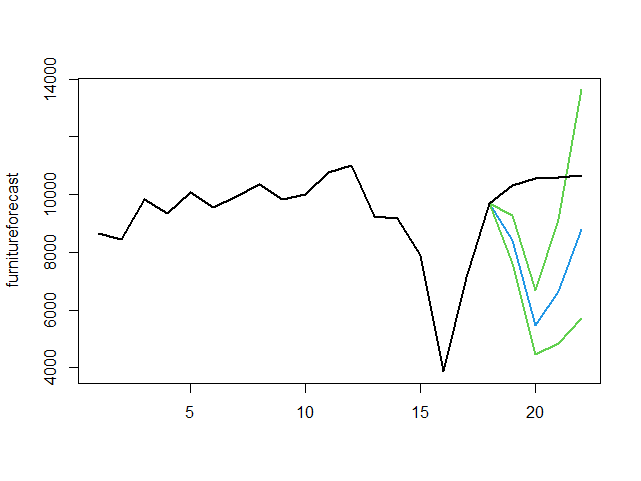
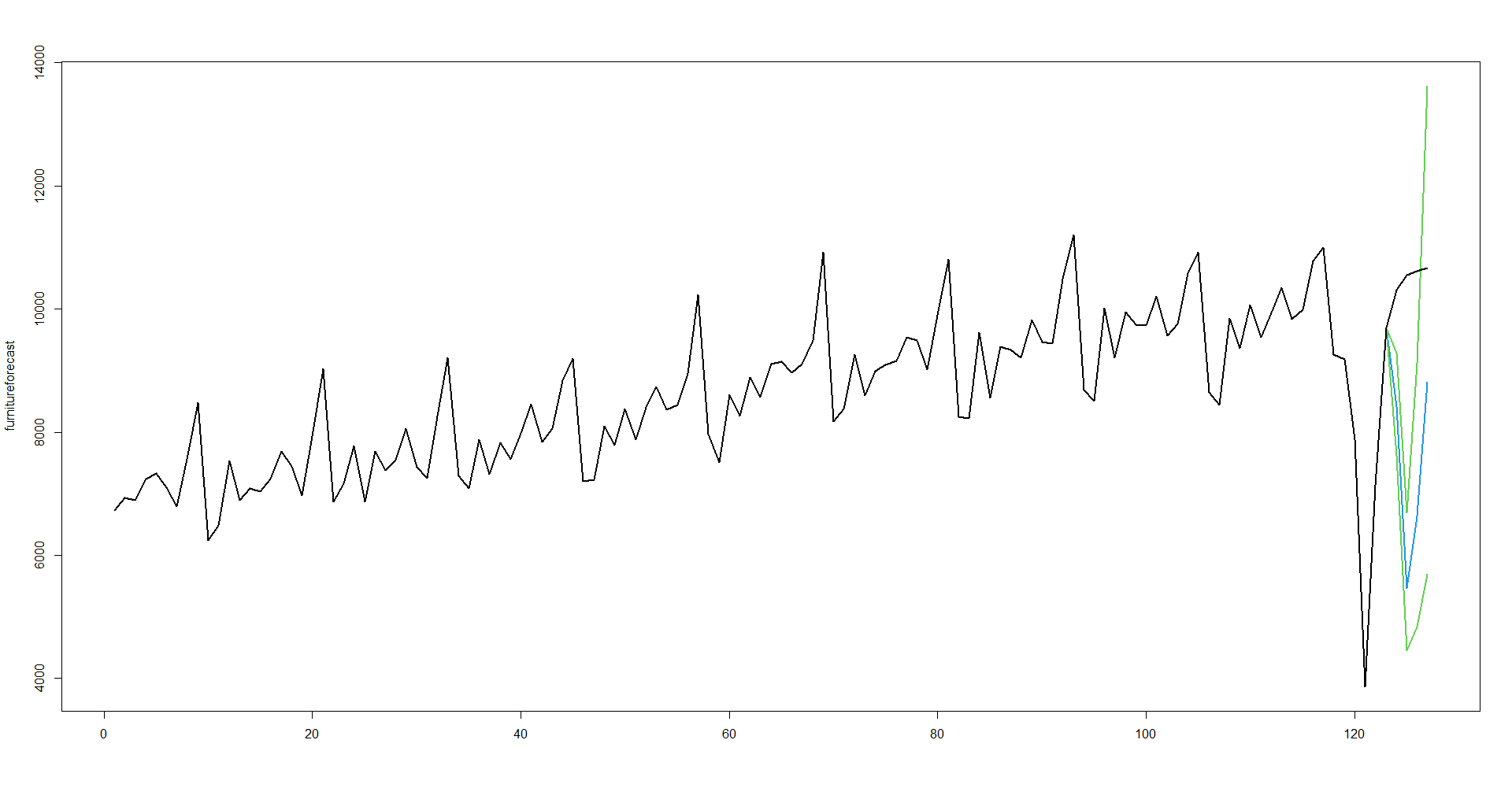
**8. Compare in-sample fit of your three (or more) ‘best’ models from questions 4, 5 and 6 to evaluate which is best (such as plots of actual and fitted values, in-sample RMSPEs, etc.).**

Final 8 observations. Actual vs predicted of ARIMA.

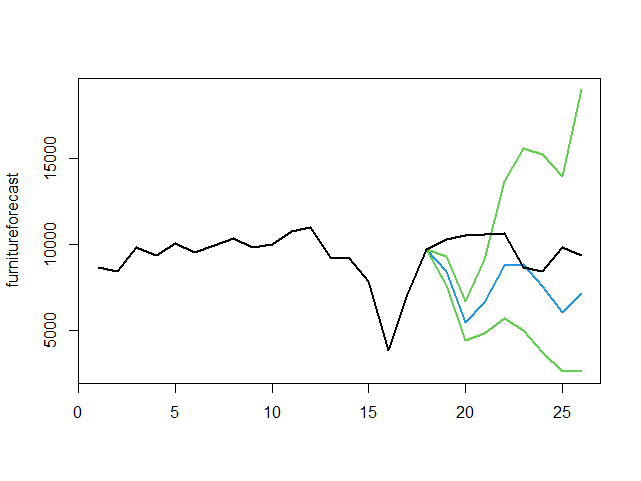
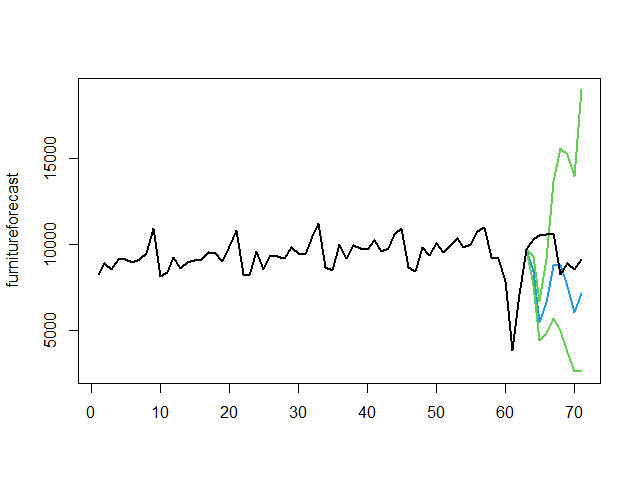


**9. For each of your three best models, generate out-of-sample forecasts for 4 periods ahead (1 through 4), then 8 periods ahead and compare with the actual values. For the dynamic regression model, you will have to either “cheat” and use actual future values of the other two variables, or use forecast values for those variables (say from a VAR or an ARIMA model for each of them).**

Using VECM to forecast last 4 and 8 observations. Green is Upper and Lower Confidence limits. Magnified view of final 20 observations is also below. The precipitous drop around observation 200 brought on by the recession caused by Covid-19 greatly affected the models ability to predict. However, in the magnified view, the quick recovery of the model is also apparent as by the end of the 8 predictions it is close to actuals once again.

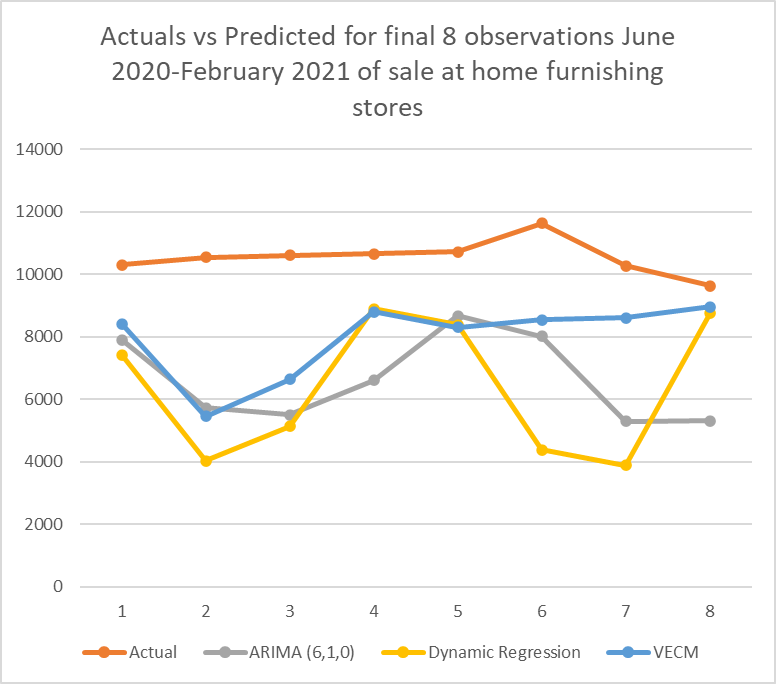


A similar story for the VECM model with every month being under forecasted. The magnified view of the same model is presented again on the left.



**10. Compare the out-of-sample forecasting performance of your three best models (using plots, out-of-sample RMSPE, etc.). Comment on which model performs best and why that might be.**

VECM performed the best of the three models. For this data, there is a fairly level long term trend with an extreme recent shock. It’s possible that the error correction term allowed the model to perform better by not overreacting to the shock. As seen in the actuals vs predicted values below, the ARIMA and dynamic regression react very quickly to the shock and do not return as quickly to the actual level. Another advantage that the VECM may have over the other models is that it allows for the effects of the right hand side variables on each other. This could result in a better prediction during a period of such extreme sensitivity.



**11. Attempt to obtain improved forecasts by combining the forecasts from at least two of you three best models**

A simple average of the three models was calculated to attempt to improve the forecast. However, since all three forecasts consistently under forecast, and the same forecast was always the second highest, an average of them ends up being worse than the existing best model, the VECM. Likewise, a weighted average of the three forecasts would be worse than VECM, but slightly better than the simple average.

