TS7

Dylan Hayashi

MSDS 413, Fall 2022, Section 55

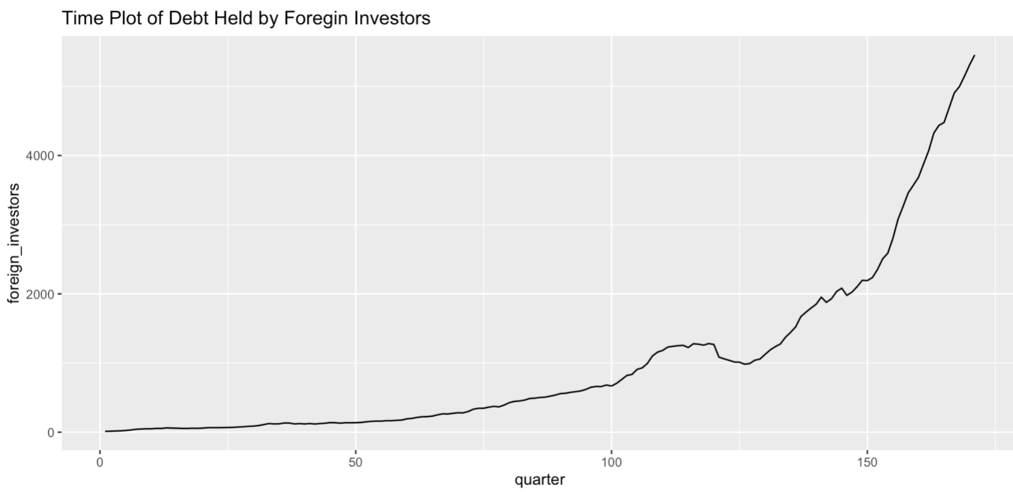
Northwestern University, Time-Series Analysis & Forecasting

October 31, 2022

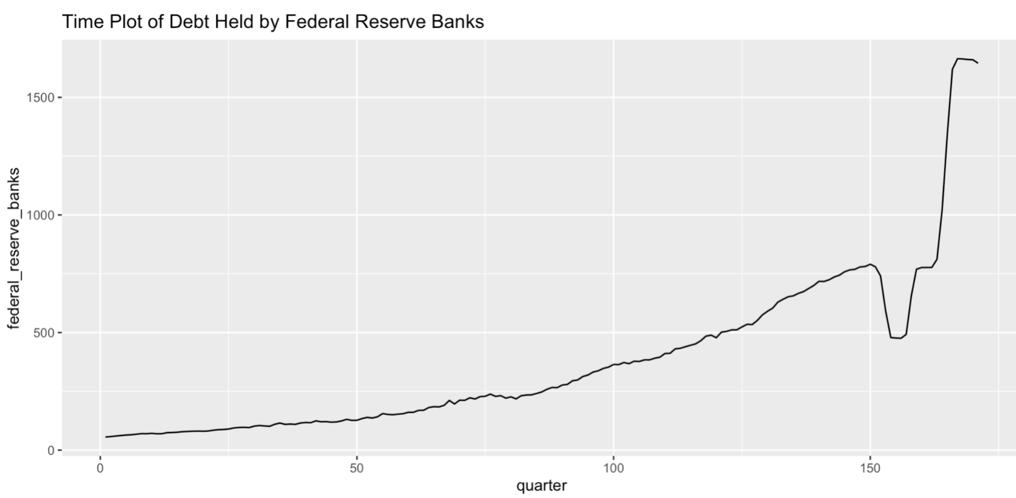
1. Debt

1.1 EDA

Below are time plots of the level of debt (in billions) held by (1) foreign investors and (2) Federal Reserve banks, as well as their summary statistics. I have converted the year and month variables of the dataset into a quarter variable, which is shown on the x-axes of the plots.




```
              X..data.hbfin
nobs           171.000000
NAs              0.000000
Minimum         12.400000
Maximum       5455.000000
1. Quartile    131.750000
3. Quartile   1263.750000
Mean          1005.757895
Median         502.000000
Sum         171984.600000
SE Mean         96.709153
LCL Mean       814.852410
UCL Mean      1196.663379
Variance   1599304.919393
Stdev         1264.636279
Skewness         1.822921
Kurtosis         2.771397
```



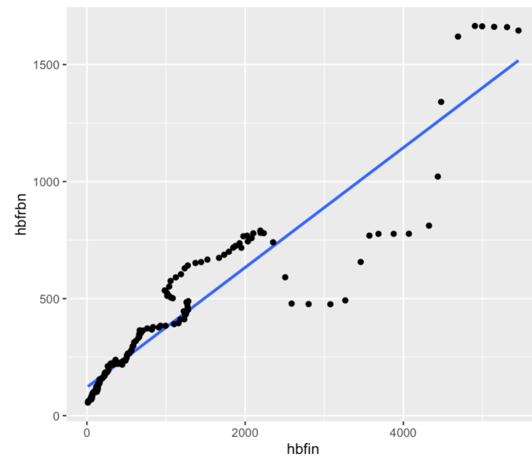

```
              X..data.hbfrbn
nobs            171.000000
NAs               0.000000
Minimum          55.800000
Maximum        1664.700000
1. Quartile     120.350000
3. Quartile     517.800000
Mean            378.573099
Median          247.500000
Sum           64736.000000
SE Mean          26.468296
LCL Mean        326.324240
UCL Mean        430.821959
Variance     119797.591390
Stdev           346.117887
Skewness          1.893926
Kurtosis          4.213803
```

Prior to conducting analysis, confirmation that these data are time series data is required. In order to fit the definition, the data must be time ordered sequences of observations of a stochastic variable over constant time intervals.

To test that this is a time ordered sequence, we can confirm that the data are indexed by unique time periods and that there are observations of each variable for each time period. The vectors quarter, unique(quarter), debt held by foreign investors, and debt held by federal reserve banks all have lengths of 171, confirming that this is the case.

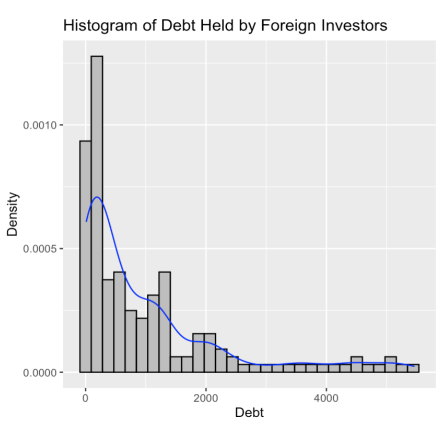
To test for constant intervals between time periods, we can sum the difference between each pair of successive quarters, and this should equal one less than the total number of observations. The sum of these differences is 170, one less than 171, confirming that the data have constant time intervals.

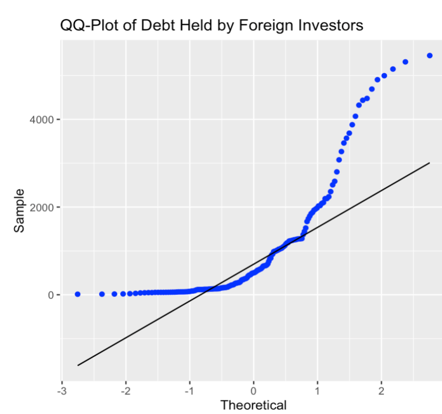
The data summaries on the previous page show that both of the variables for debt held by foregin investors and federal reserve banks have variance, indicating that they are stochastic. Thus, we have confirmed each part of the definition of time series for both variables.

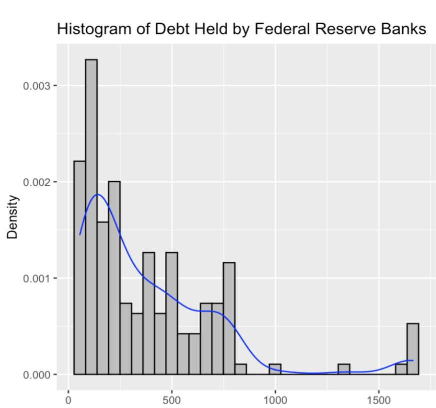


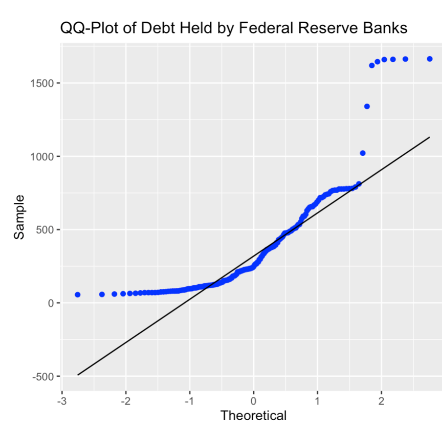
To the right is a scatter plot of the two variables. The plot demonstrates a strong positive relationship between the two variables. Given the similarity of the time plots, this is reasonable.

Neither of the variables are normally distributed. As the histograms below show, they are both right skewed. Their QQ plots also demonstrate how the tails of both variables deviate strongly from the normal line. Additionally, neither the 95% confidence intervals for skew nor kurtosis include zero for either variable. For debt held by foreign investors, they were (1.44, 2.21) and (0.686, 4.722). For debt held by Federal Reserve banks, they were (1.592, 2.322) and (2.435, 6.157). Given the shape of the variables' distributions and their confidence intervals for skew and excess kurtosis, it can be concluded that neither variable is normally distributed.





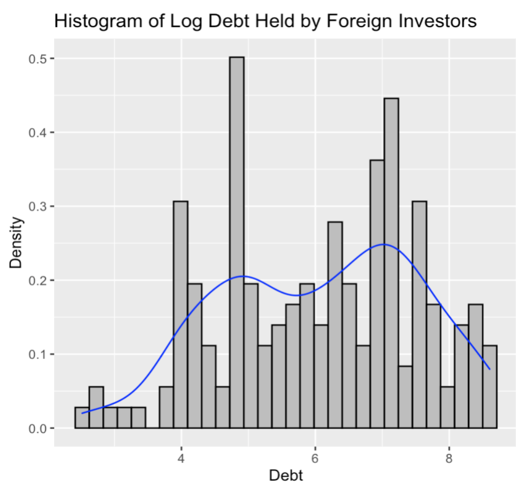


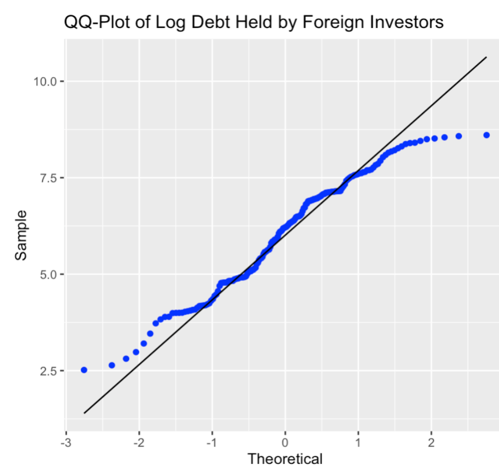


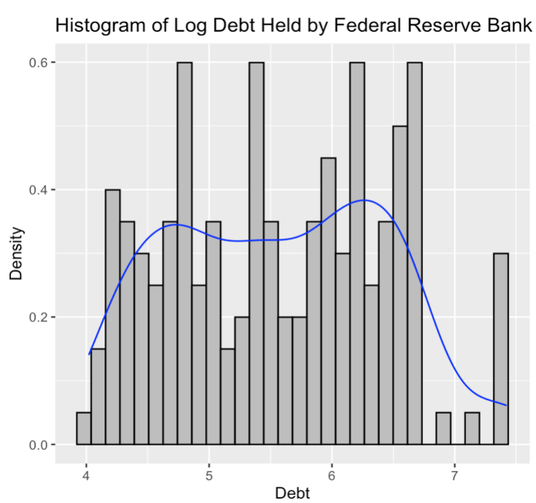
Lack of normality in both the individual variables and the multivariate time series are asserted by the MVN function, which displayed the following output. The results for multivariate normal skew, kurtosis, and overall distribution, as well as for univariate normality were 'NO.' Thus, I will apply log transformation.

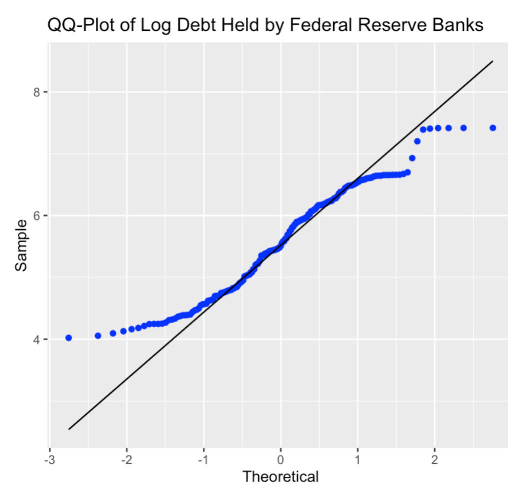
```
$multivariateNormality
             Test        Statistic
1 Mardia Skewness 342.991869755552
2 Mardia Kurtosis 18.9345131762861
3             MVN             <NA>
                                                                                    p value Result
1 0.000000000000000000000000000000000000000000000000000000000000000000000000571532167102754     NO
2                                                                                         0     NO
3                                                                                      <NA>     NO

$univariateNormality
             Test  Variable Statistic   p value Normality
1 Anderson-Darling   hbfin     14.433  <0.001      NO
2 Anderson-Darling  hbfrbn      8.806  <0.001      NO
```

Log transformation of the variables produces the histograms and QQ plots below:







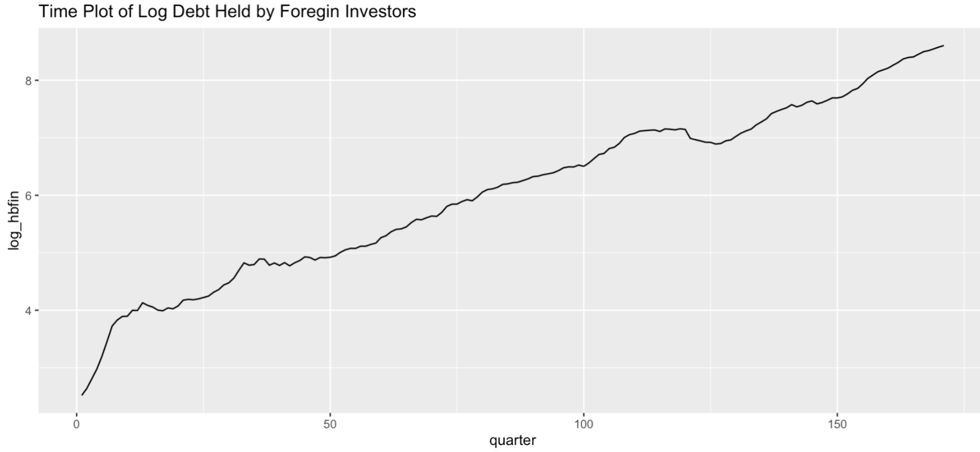


The histograms show that both variables are now less skewed. Additionally, the QQ plots indicate that more of the observations fall on the normal line. However, compression has caused their distributions to appear more bimodal than before. The new 95% confidence intervals for skew and excess kurtosis of the foregin investor variable are now (-0.445, 0.049) and (-1.193, -0.582). For the Federal Reserve bank variable, they are now (-0.133, 0.329) and (-1.248, -0.766). So, while the distributions are not normally distributed as a result of their shape, their statistics for skew and kurtosis have improved.

Lack of normality is substantiated by the MVN function which, when applied to the log transformed variables, produced the output below. It indicates that the individual variables are still not normally distributed, and that the multivariate time series is not either. However, the p-values and test statistics have improved, corroborating the improvement I described in the previous paragraph. On the bottom of the page are the time plots of the log data.

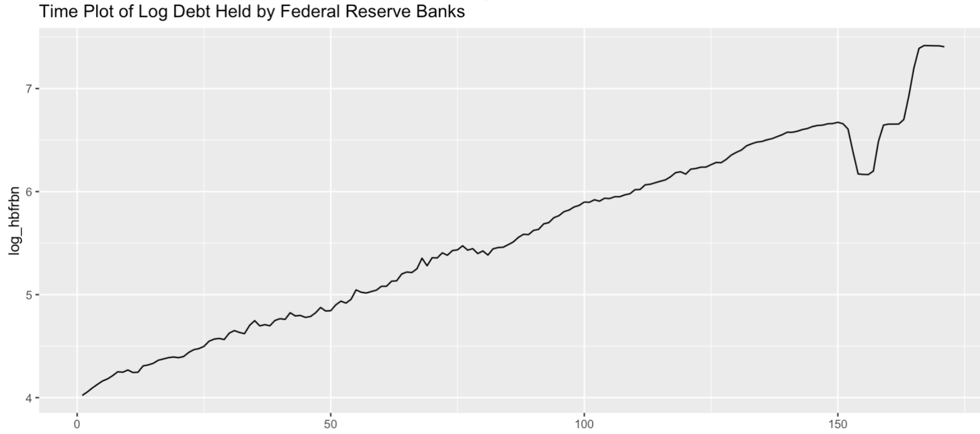
```
$multivariateNormality
             Test        Statistic                p value Result
1 Mardia Skewness 30.2209003783539  0.00000441288429266606     NO
2 Mardia Kurtosis 5.20346263375788 0.000000195609098030403     NO
3             MVN             <NA>                   <NA>     NO

$univariateNormality
              Test   Variable Statistic   p value Normality
1 Anderson-Darling log_hbfin     1.5901    0.0004    NO
2 Anderson-Darling log_hbfrbn    1.8336    0.0001    NO
```

Time Plot of Log Debt Held by Foregin Investors



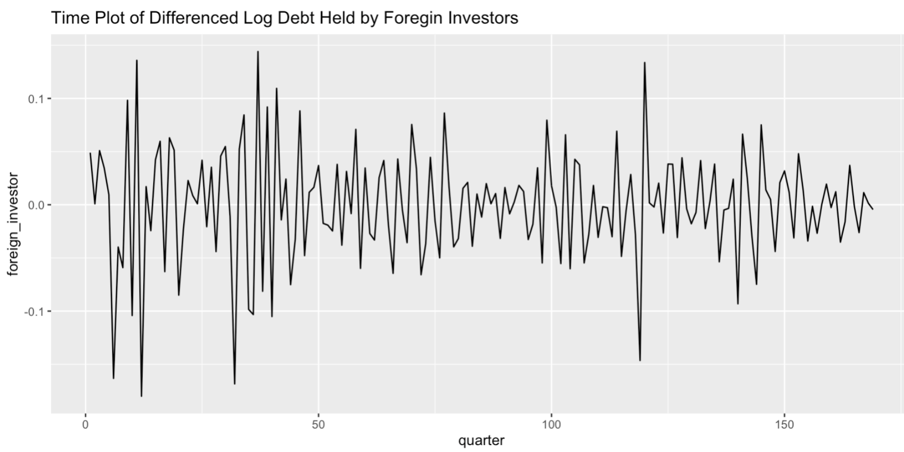

```
nobs              171.00000
NAs                 0.00000
Minimum             2.51770
Maximum             8.60429
1. Quartile         4.88087
3. Quartile         7.14183
Mean                6.05061
Median              6.21860
Sum              1034.65416
SE Mean             0.11173
LCL Mean            5.83005
UCL Mean            6.27117
Variance            2.13478
Stdev               1.46109
Skewness           -0.19717
Kurtosis           -0.88972
```

Time Plot of Log Debt Held by Federal Reserve Banks

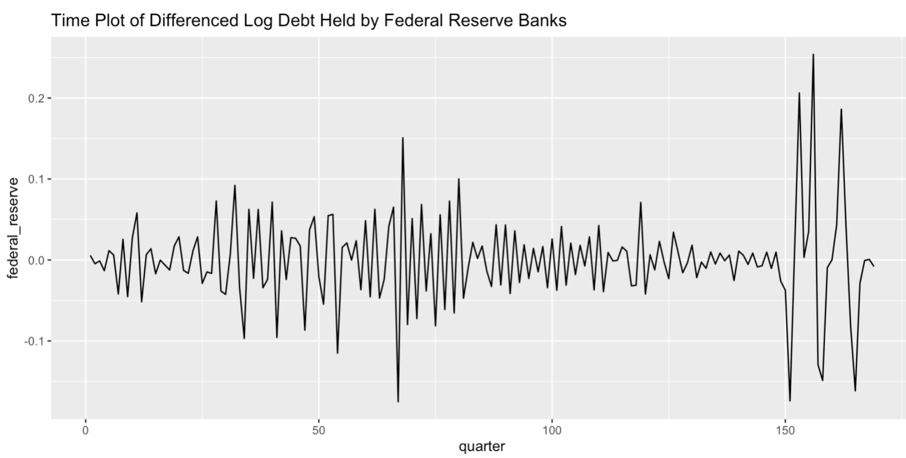



```
nobs              171.000000
NAs                 0.000000
Minimum             4.021774
Maximum             7.417400
1. Quartile         4.790400
3. Quartile         6.249520
Mean                5.566256
Median              5.511411
Sum               951.829786
SE Mean             0.066981
LCL Mean            5.434034
UCL Mean            5.698478
Variance            0.767184
Stdev               0.875891
Skewness            0.088245
Kurtosis           -0.999028
```

The plots of both variables appear to have trend. T-Tests of both variables returned 95% confidence intervals of the means that did not include zero. They were (5.83, 6.271) and (5.434, 5.698) for debt held by foreign investors and Federal Reserve banks respectively. Lack of mean zero in the variables indicates that differencing is appropriate. Twice differencing was required to get data with mean zero. The new time plots are shown below:



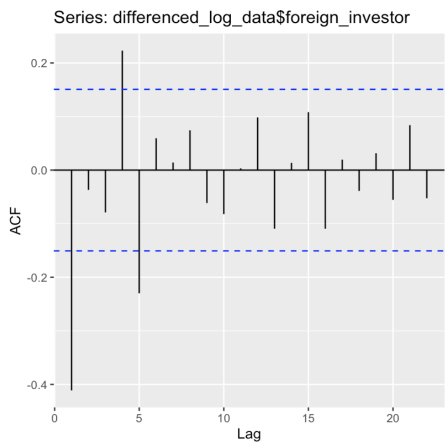
nobs	169.000000
NAs	0.000000
Minimum	-0.180052
Maximum	0.144105
1. Quartile	-0.030840
3. Quartile	0.034592
Mean	-0.000560
Median	0.000871
Sum	-0.094608
SE Mean	0.004092
LCL Mean	-0.008638
UCL Mean	0.007518
Variance	0.002830
Stdev	0.053194
Skewness	-0.381554
Kurtosis	1.252030

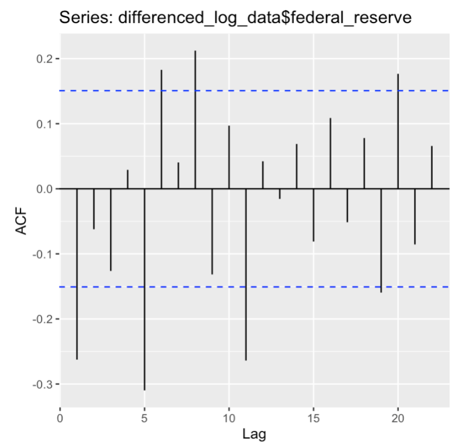


nobs	169.000000
NAs	0.000000
Minimum	-0.175179
Maximum	0.253954
1. Quartile	-0.027831
3. Quartile	0.025515
Mean	-0.000250
Median	-0.000414
Sum	-0.042318
SE Mean	0.004311
LCL Mean	-0.008762
UCL Mean	0.008261
Variance	0.003141
Stdev	0.056047
Skewness	0.456543
Kurtosis	4.287645

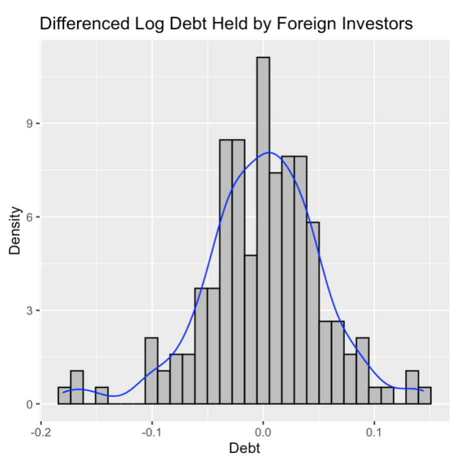
As stated, the twice differenced data are mean zero. The 95% confidence intervals for the means of debt held by foreign investors and debt held by Federal Reserve banks are (-0.009, 0.008) and (-0.009, 0.008) respectively. Linear-trend stationarity is asserted for both variables by ADF and KPSS tests, which each returned p-values/test-statistics beneath those required for stationarity. For debt held by foreign investors, they were ADF p-value: $0.01 < 0.05$, KPSS test statistic: $0.0178 < 0.146$. For debt held by Federal Reserve banks, they were ADF p-value: $0.01 < 0.05$, KPSS test statistic: $0.0136 < 0.146$. McLeod-Li tests of both variables returned sets with all tested lags, indicating rejections of their null hypotheses that the variables do feature constant variance. Thus, the variables are not strictly stationary. A March test of the multivariate time series produced p-values of all less than one, indicating that multivariate ARCH effects are present.

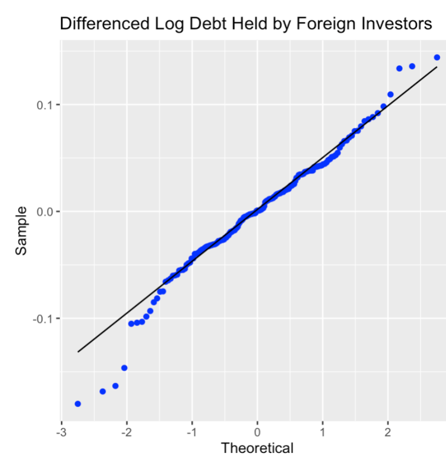
Ljung-Box tests of both variables returned p-values less than 0.01, indicating rejections of their null hypotheses that the variables do not feature auto-correlation. Auto-correlation can also be demonstrated in their ACF plots below, which each show multiple lags with statistically significant auto-correlation.

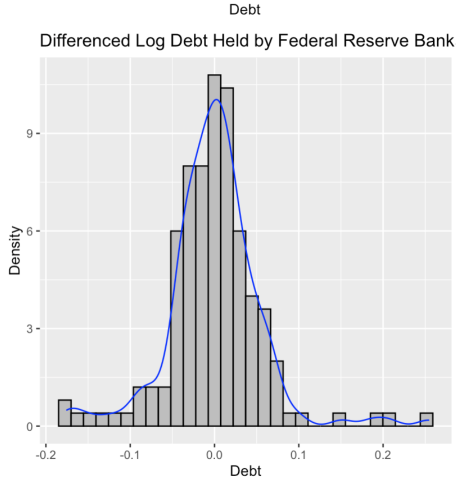


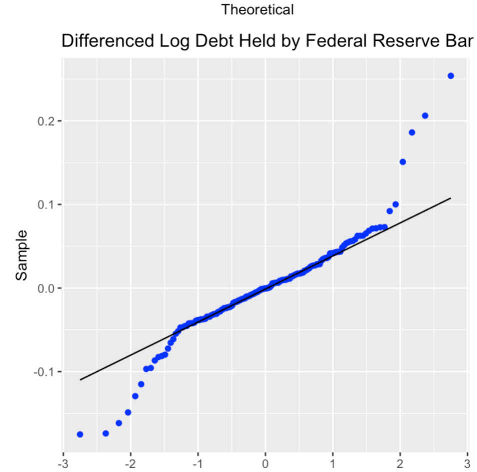


Below are histograms and QQ plots of the differenced data. They demonstrate that the twice differencing also improved their distributions, making them appear more normal. They are now unimodal, however their tails appear to deviate more from the normal line than the undifferenced data for debt held by foreign investors.









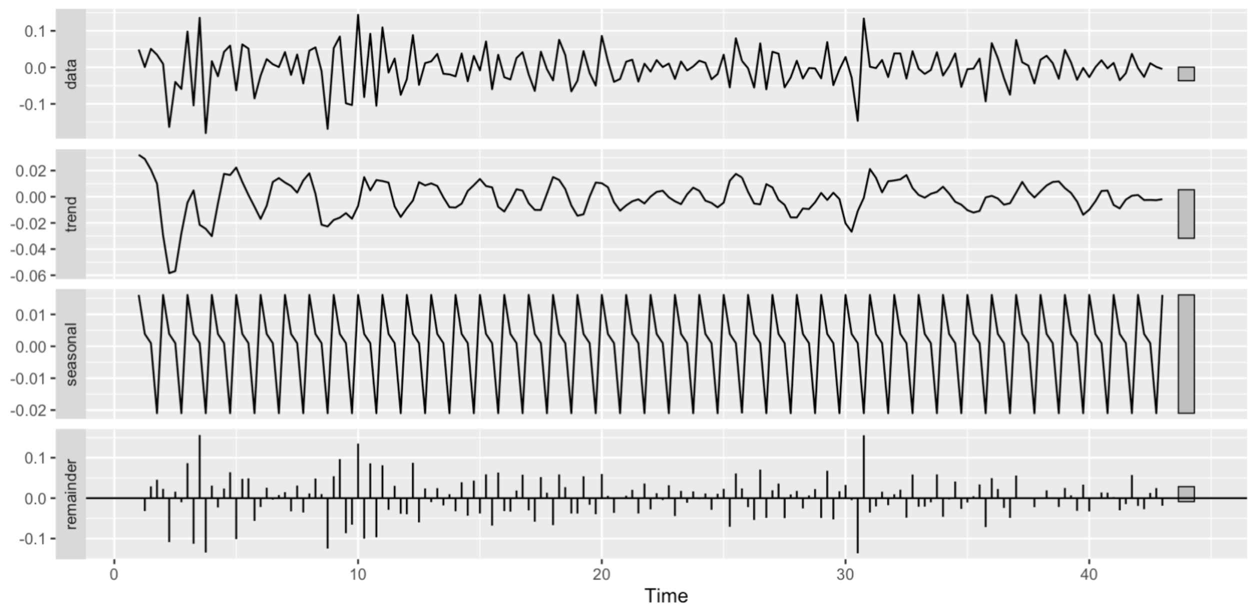
The plots on the previous page suggest the variables are now approximately normal, although with high kurtosis. The 95% confidence intervals for both of the variables still include zero. They are (-0.923, 0.133) for debt held by foreign investors and (-0.56, 1.659) for debt held by Federal Reserve banks. As expected, the 95% confidence intervals for excess kurtosis were slightly high for both variables, they were (0.42, 2.233) and (2.241, 6.93).

The MVN function indicated that neither of the variables were normally distributed, nor are they when combined in a multivariate time series. The results of the function are shown below. Overall, however, the variables are now much closer to normal than when they were undifferenced and not log-transformed; they are not unimodal, with skew zero, and have less deviation from the normal line in their QQ plots.

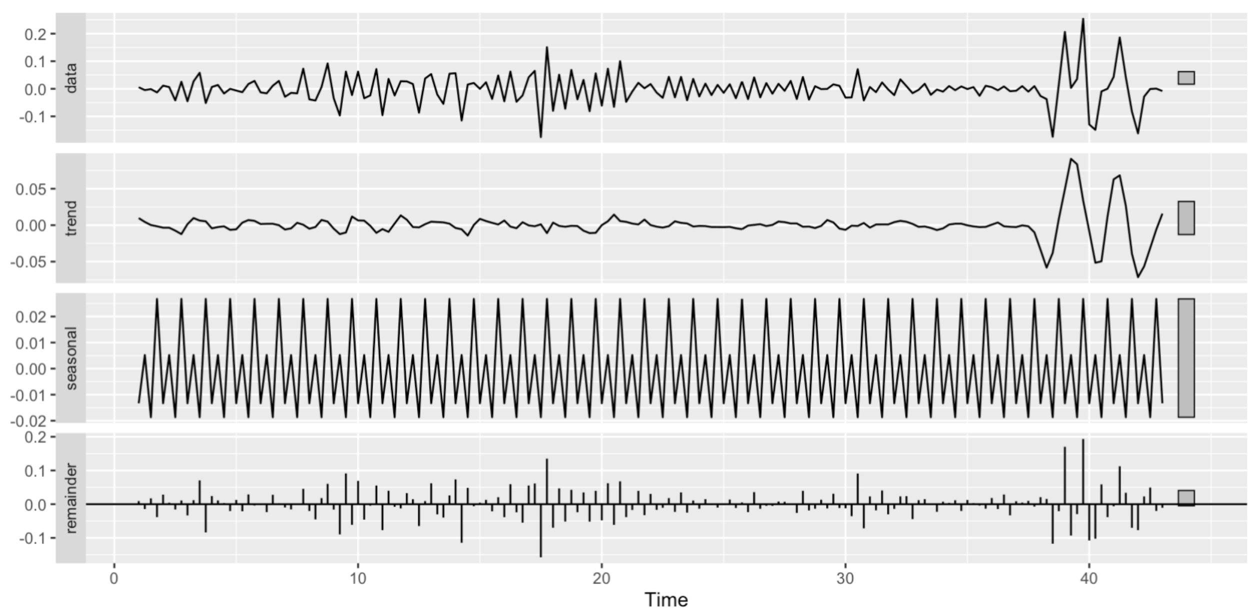
```
$multivariateNormality
             Test        Statistic            p value Result
1 Mardia Skewness 12.2741082772358 0.0154251801058141     NO
2 Mardia Kurtosis 9.78143977386677                  0     NO
3             MVN             <NA>               <NA>     NO

$univariateNormality
              Test         Variable Statistic   p value Normality
1 Anderson-Darling foreign_investor    0.8853  0.0231      NO
2 Anderson-Darling federal_reserve     3.4063  <0.001      NO
```

On the top of the following page are decompositions of the two variables. For both decompositions, the trend lines are somewhat smoothed versions of the line plot. While mostly smooth, they demonstrate that the decomposition attributed a large amount of variance to the trends when it occurred. Both of the decompositions show well defined seasonal patterns, although they are different. Notably, the seasonal pattern for debt held by foreign investors features less frequent spiking in comparison to that of debt held by Federal Reserve banks. Overall, however, their ranges are small and how unsmooth the trend line is in combination with how closely the residuals resemble the time plot indicate that decompositions were not able to adequately explain much of the variance while keeping a smooth trend line.

Debt Held by Foreign Investors

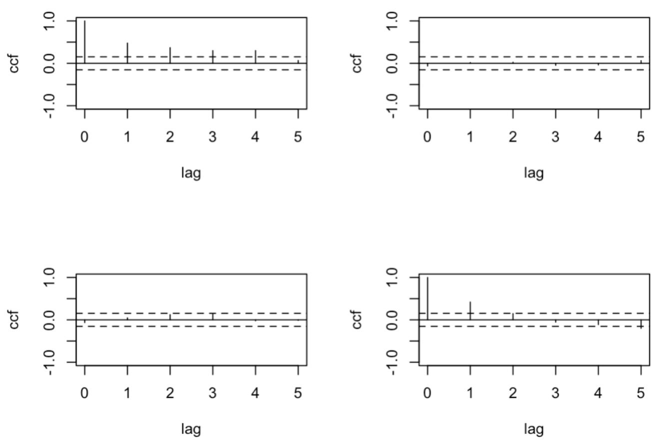


Debt Held by Federal Reserve Banks



1.2 Cross-Correlation Matrices

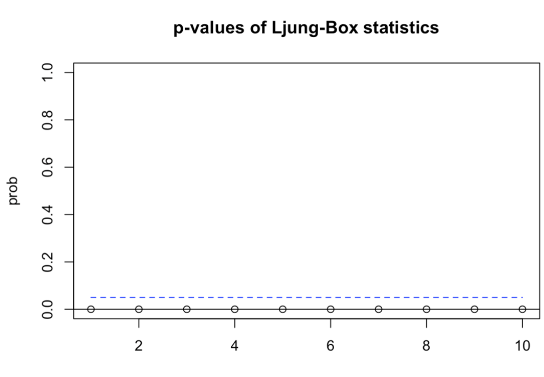
Below are the outputs of the CCM function. They collectively demonstrate that there is not significant cross-correlation between the two time-series, but that the individual time series do feature auto-correlation. The dots in the off-diagonals of the matrices indicate that the variables do not feature cross-correlation, while the positive and negative signs in the diagonals indicate that the individual variables feature auto-correlation. The ACF and CCF plots show the same lack of cross-correlation in the plots in the off-diagonals and significant auto-correlation in the diagonals.




```
CCM at lag:  1
+ .
. +
CCM at lag:  2
+ .
. .
CCM at lag:  3
+ .
. .
CCM at lag:  4
+ .
. .
CCM at lag:  5
. .
. -
```

1.3 Cross-Correlation Hypothesis Testing

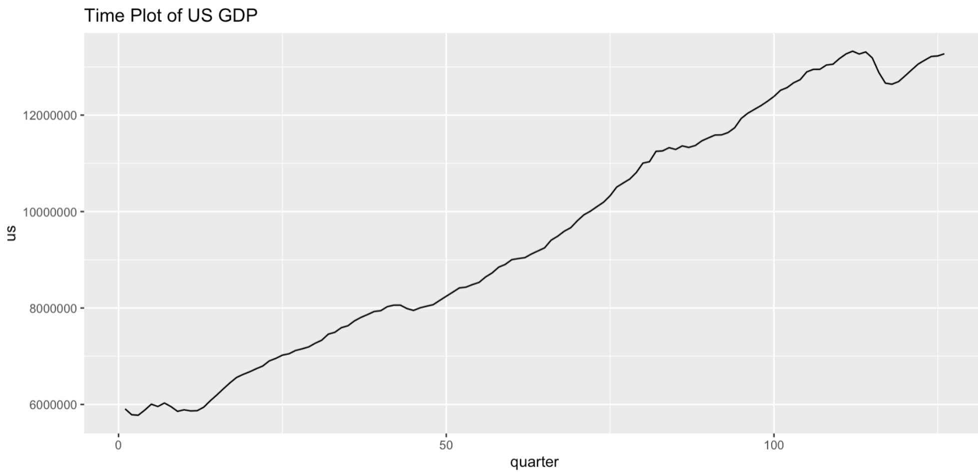
The MQ function indicates that there is statistically significant correlation across all lags up to 10. The Ljung-Box tests in the function assume the null hypothesis that the correlation at each lag is equal to zero. As the plot below shows, the p-value for all lags are beneath 0.05, indicating a rejection of this null hypothesis. Thus, there is significant correlation at all tested lags.



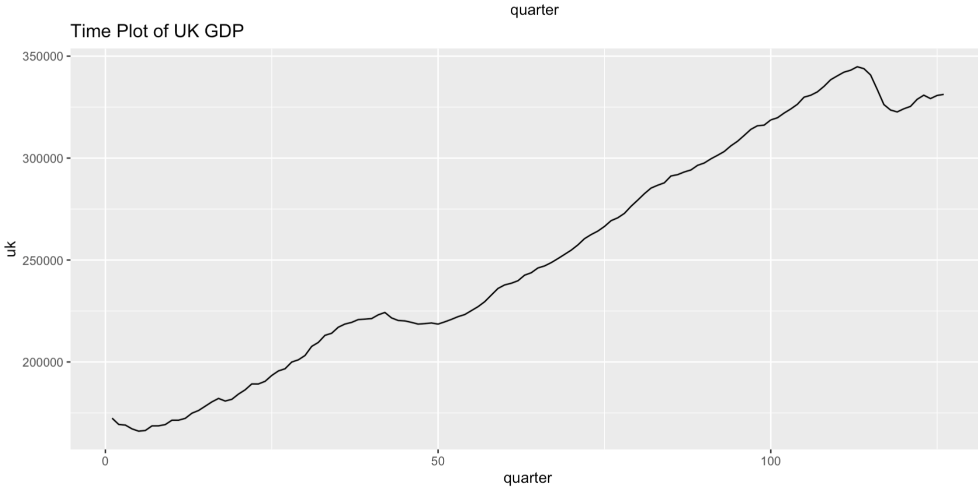
2. GDP

2.1 EDA

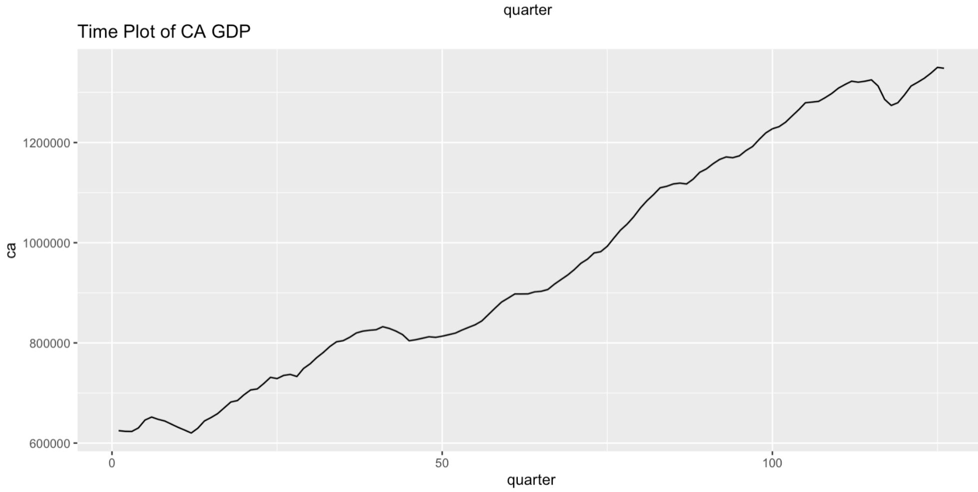
Below are time plots of the US, UK, and Canada GDP, as well as their summary statistics.



nobs	126.000000
NAs	0.000000
Minimum	5776600.000000
Maximum	13326000.000000
1. Quartile	7467650.000000
3. Quartile	11882400.000000
Mean	9551488.095238
Median	9152500.000000
Sum	1203487500.000000
SE Mean	222536.937714
LCL Mean	9111059.876571
UCL Mean	9991916.313905
Variance	6239858769537.142578
Stdev	2497970.930483
Skewness	0.074972
Kurtosis	-1.411987



nobs	126.00000
NAs	0.00000
Minimum	166052.00000
Maximum	344809.00000
1. Quartile	210459.00000
3. Quartile	307689.25000
Mean	252570.88095
Median	243169.00000
Sum	31823931.00000
SE Mean	5068.50468
LCL Mean	242539.68115
UCL Mean	262602.08076
Variance	3236907199.12971
Stdev	56893.82391
Skewness	0.11363
Kurtosis	-1.36261



nobs	126.00000
NAs	0.00000
Minimum	620116.00000
Maximum	1350078.00000
1. Quartile	783929.75000
3. Quartile	1172881.00000
Mean	959827.03175
Median	899888.00000
Sum	120938206.00000
SE Mean	20989.77997
LCL Mean	918285.65217
UCL Mean	1001368.41133
Variance	55511928733.61497
Stdev	235609.69575
Skewness	0.22542
Kurtosis	-1.35943

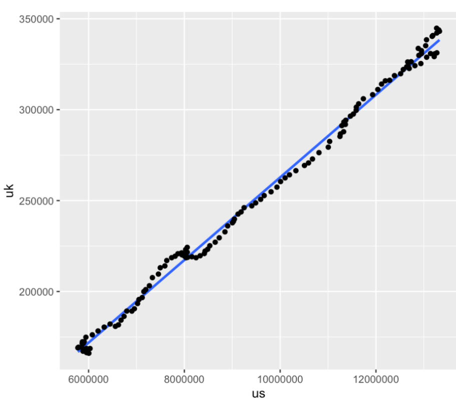
Prior to analysis, the data must be confirmed to be time series data. This requires that the data be time ordered sequences of observations of a stochastic variable over constant time intervals.

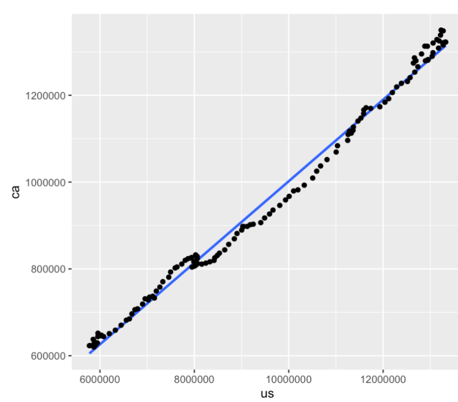
To test that the data are time ordered sequences, the data must be indexed by unique time periods with observations of each variable for each time period. The vectors quarter, unique(quarter), US, UK, and CA all have lengths of 126, confirming that this is the case.

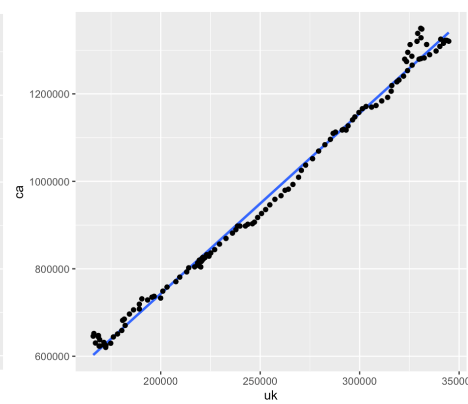
To test for constant intervals between time periods, we can sum the difference between each pair of successive quarters, and this should equal one less than the total number of observations. The sum of these differences is 125, one less than 126, confirming that the data have constant time intervals.

The data summaries on the previous page show that all three variables have positive variance, indicating that they are stochastic. Thus, we have confirmed each part of the definition of time series for both variables.

Below are scatter plots of each pair-wise combination of the variables. They all demonstrate a very strong positive relationship with one another. Given the similarity in the behavior of the variables shown in the time plots, it is intuitive that they feature high, positive correlations.

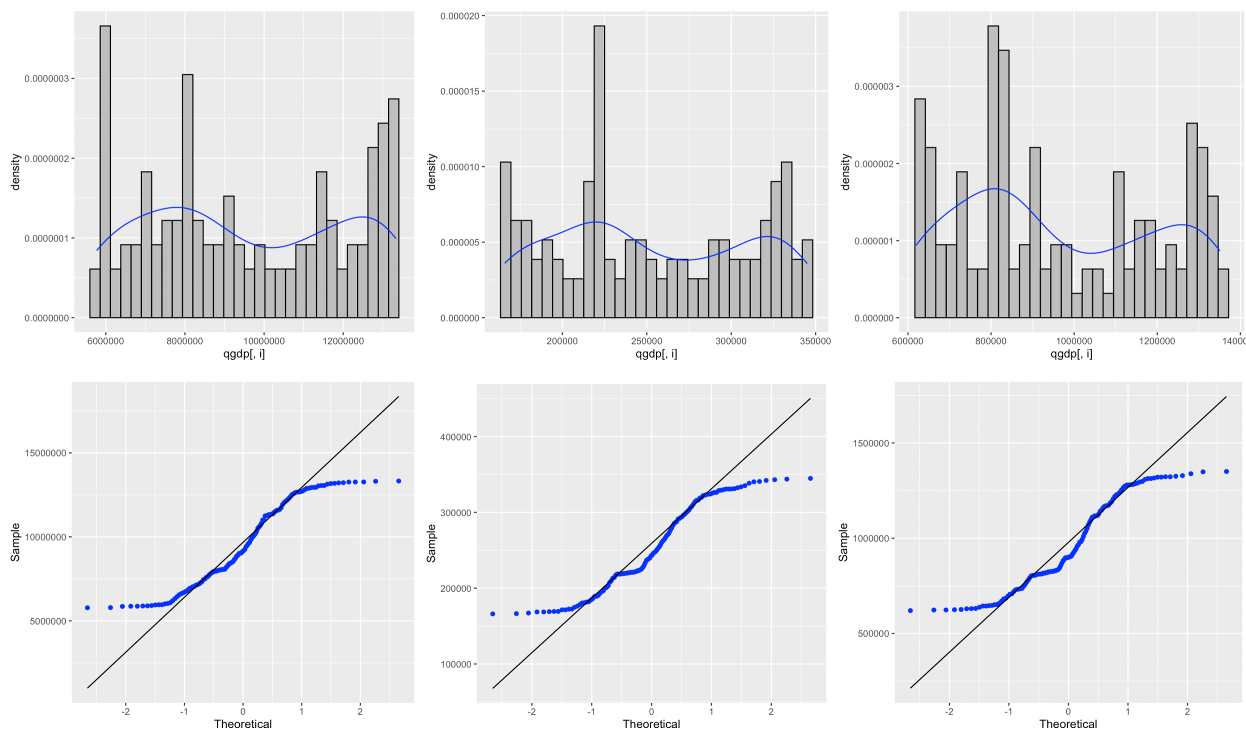






The line plots on the previous page demonstrates that the countries of the GDP all have had historically similar behavior, although at different levels. Unsurprisingly, their distributions appear similar as well. As the histograms and QQ plots on the top of the next page show, they are all bimodal and have tails that deviate from the normal line. It is apparent from these diagrams that the distributions are not normally distributed.

US UK CA



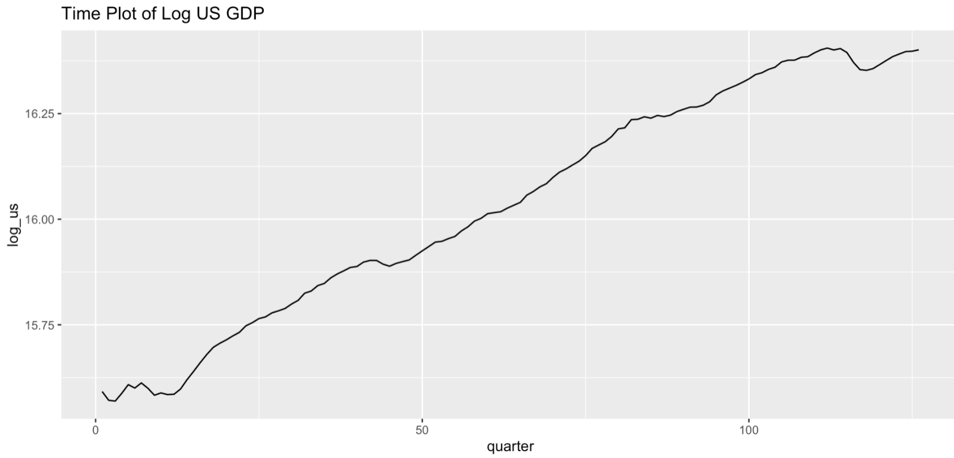
The 95% confidence intervals for skew of each variable do contain zero, however, that is not enough to assert that the variables are approximately normally distributed. The plots on the previous page demonstrate as much given their bimodality. Additionally, none of the variables have 95% confidence intervals for excess kurtosis that contain zero. The results are shown in the table below:

Variable	Skew LL	Skew	Skew UL	Kurt LL	Kurt	Kurt UL
US	0.075	-0.187	0.337	-1.412	-1.626	-1.256
UK	0.114	-0.159	0.368	-1.363	-1.601	-1.189
CA	0.225	-0.038	0.500	-1.359	-1.629	-1.137

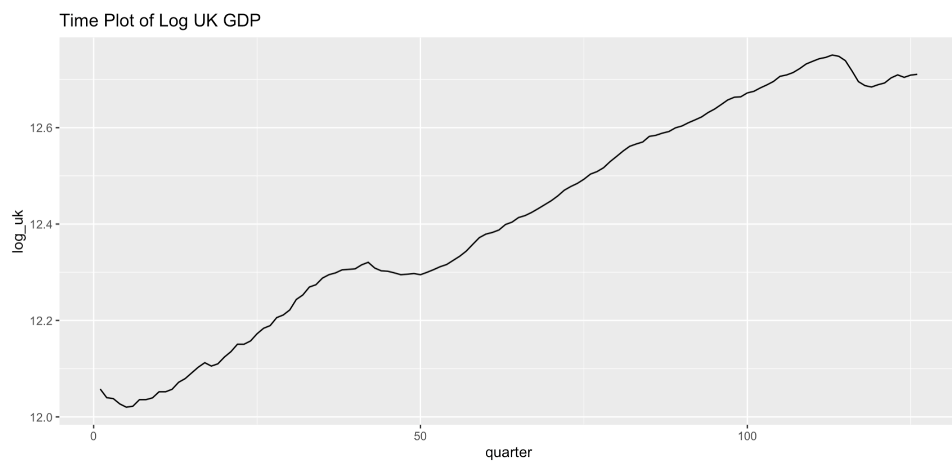
Lack of normality for each variable is also suggested by the MVN function, which produced the output below. As shown, each individual variable is indicated as not normal, and the multivariable timeseries containing all three variables is also not normal. Thus, log transformation is appropriate.

```
$multivariateNormality
             Test          Statistic                              p value Result
1 Mardia Skewness    85.7791046443066 0.0000000000000366378607316769     NO
2 Mardia Kurtosis -0.780008077168047              0.435386120884077    YES
3             MVN                <NA>                            <NA>     NO

$univariateNormality
              Test  Variable Statistic   p value Normality
1 Anderson-Darling     us        2.9283  <0.001      NO
2 Anderson-Darling     ca        3.3264  <0.001      NO
3 Anderson-Darling     uk        2.8067  <0.001      NO
```

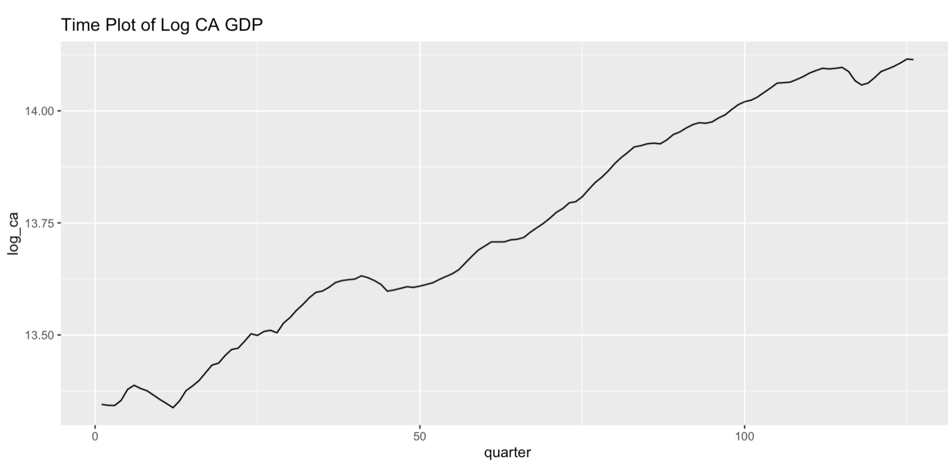
Applying log transformation produces variables that take the following form as line plots:



nobs	126.000000
NAs	0.000000
Minimum	15.569326
Maximum	16.405228
1. Quartile	15.826088
3. Quartile	16.290544
Mean	16.036740
Median	16.029532
Sum	2020.629242
SE Mean	0.024114
LCL Mean	15.989015
UCL Mean	16.084465
Variance	0.073268
Stdev	0.270681
Skewness	-0.184800
Kurtosis	-1.301279



nobs	126.000000
NAs	0.000000
Minimum	12.020056
Maximum	12.750746
1. Quartile	12.257021
3. Quartile	12.636841
Mean	12.413620
Median	12.401509
Sum	1564.116082
SE Mean	0.020479
LCL Mean	12.373090
UCL Mean	12.454150
Variance	0.052842
Stdev	0.229874
Skewness	-0.120582
Kurtosis	-1.286579



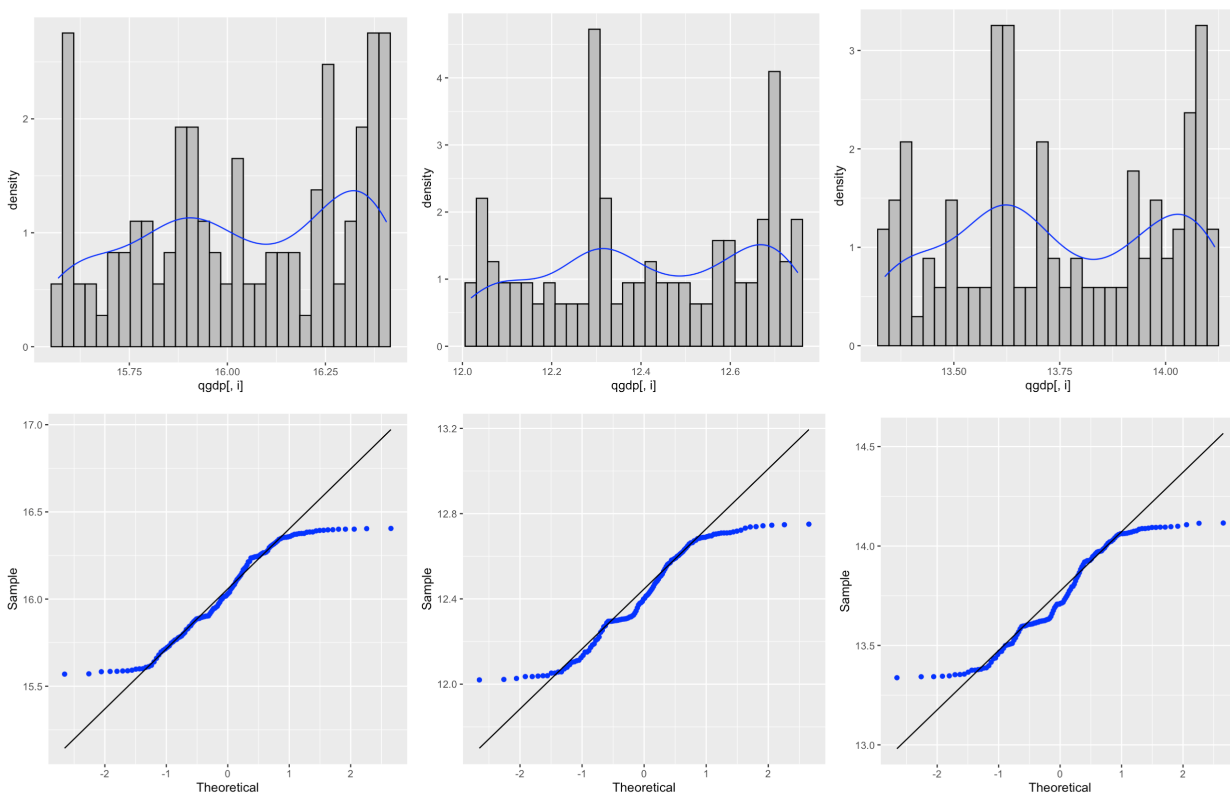
nobs	126.000000
NAs	0.000000
Minimum	13.337662
Maximum	14.115673
1. Quartile	13.572053
3. Quartile	13.974973
Mean	13.744251
Median	13.710023
Sum	1731.775625
SE Mean	0.022090
LCL Mean	13.700532
UCL Mean	13.787970
Variance	0.061484
Stdev	0.247960
Skewness	-0.016351
Kurtosis	-1.323562

Log transformation did not make these variables normal. As the histograms and QQ plots below show, they still have roughly bimodal distributions and have tails that deviate from the normal line. However, it does appear that more of the observations lie on the line. The confidence intervals for skew and excess kurtosis of each variable have shrunk and moved towards zero, but this is expected as log transformation reduces the range of the variables. The output of the MVN function indicates that none of the variables are normally distributed, nor is the multivariate time series that contains all three.

Variable	Skew LL	Skew	Skew UL	Kurt LL	Kurt	Kurt UL
US	-0.185	-0.445	0.071	-1.301	-1.522	-1.127
UK	-0.121	-0.369	0.134	-1.515	-1.515	-1.112
CA	-0.016	-0.271	0.226	-1.324	-1.526	-1.154

```
$multivariateNormality
             Test         Statistic                       p value Result
1 Mardia Skewness  76.5570045369632 0.0000000000236597731948479     NO
2 Mardia Kurtosis -1.11562962570727            0.264580705440418    YES
3             MVN              <NA>                         <NA>     NO

$univariateNormality
              Test  Variable Statistic   p value Normality
1 Anderson-Darling  log_us      2.6392  <0.001      NO
2 Anderson-Darling  log_uk      2.4506  <0.001      NO
3 Anderson-Darling  log_ca      2.5981  <0.001      NO
```

US CA UK

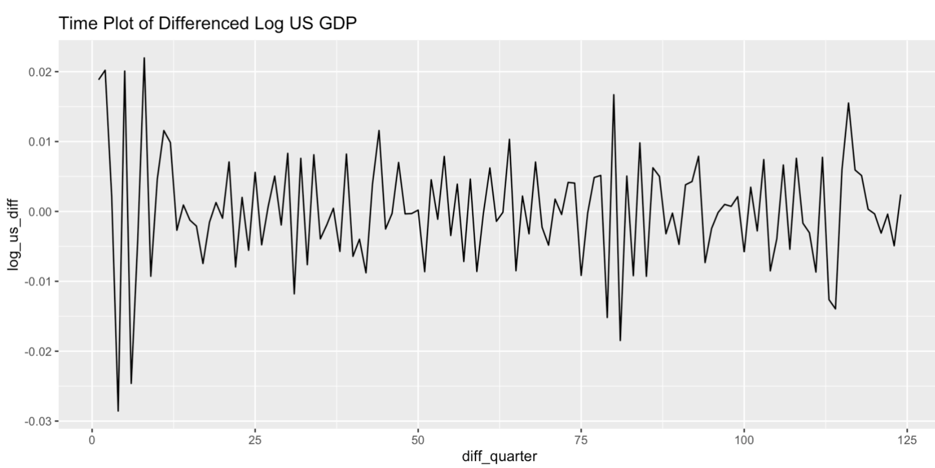


The log transformed variables are not mean zero. The 95% confidence intervals of their t-tests are shown below. These values indicate that differencing is appropriate.

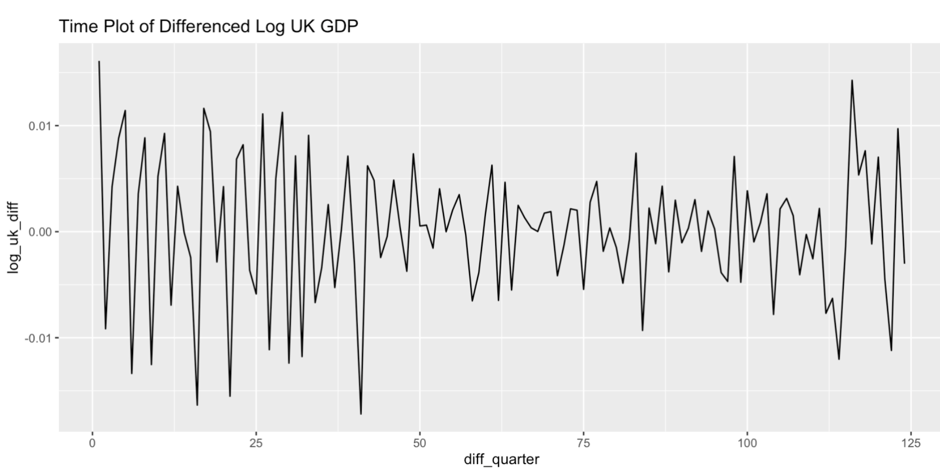
Variable	Mean LL	Mean	Mean UL
US	15.989	16.037	16.084
UK	12.373	12.414	12.454
CA	13.701	13.744	13.788

Below are time plots of the variables after twice differencing, which was required to reach mean zero. The line plots now show greater distinction between one another, which is notable. The 95% confidence intervals for the means of the variables now include zero, as shown in the table below:

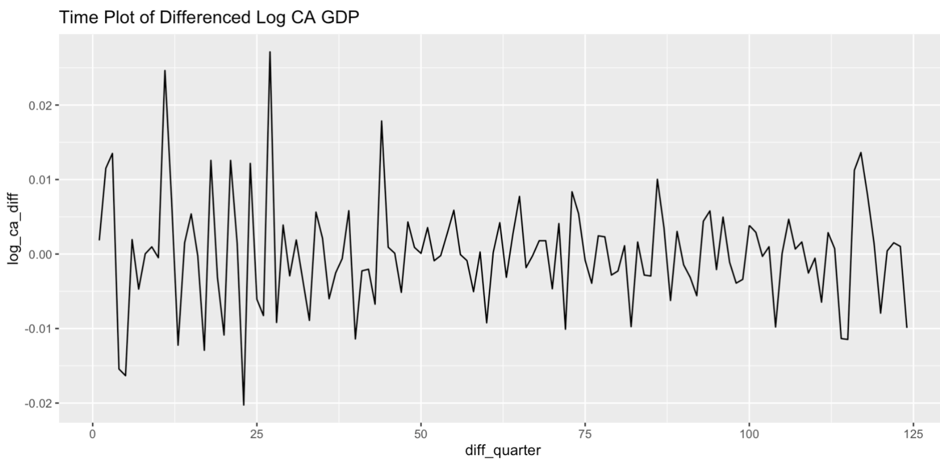
Variable	Mean LL	Mean	Mean UL
US	-0.00124	0.00019	0.00163
UK	-0.00100	0.00016	0.00131
CA	-0.00129	0.00001	0.00131




```
nobs         124.000000
NAs            0.000000
Minimum       -0.028565
Maximum        0.021974
1. Quartile   -0.004199
3. Quartile    0.005080
Mean           0.000194
Median        -0.000233
Sum            0.024022
SE Mean        0.000724
LCL Mean      -0.001240
UCL Mean       0.001628
Variance       0.000065
Stdev          0.008068
Skewness      -0.208047
Kurtosis       1.433915
```




```
nobs         124.000000
NAs            0.000000
Minimum       -0.017197
Maximum        0.016108
1. Quartile   -0.003807
3. Quartile    0.004286
Mean           0.000158
Median         0.000345
Sum            0.019634
SE Mean        0.000583
LCL Mean      -0.000995
UCL Mean       0.001312
Variance       0.000042
Stdev          0.006490
Skewness      -0.306125
Kurtosis       0.038083
```



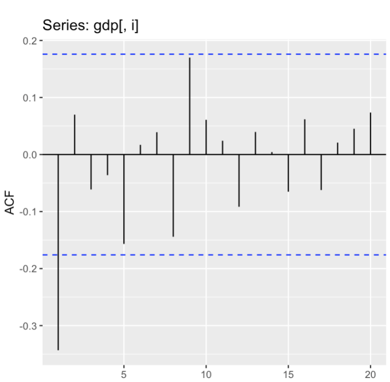

```
nobs          124.000000
NAs             0.000000
Minimum        -0.020268
Maximum         0.027137
1. Quartile    -0.003430
3. Quartile     0.003162
Mean            0.000007
Median          0.000104
Sum             0.000894
SE Mean         0.000656
LCL Mean       -0.001291
UCL Mean        0.001305
Variance        0.000053
Stdev           0.007302
Skewness        0.488532
Kurtosis        1.855989
```

Linear-trend stationarity is indicated for all variables by ADF and KPSS tests, each of which return p-values/test-statistics beneath that required to assert stationarity. The ADF for each variable returned a p-value of 0.01 (< 0.05). The KPSS tests for US, UK, and CA were 0.0395, 0.0484, and 0.0197 respectively, which all fall beneath the critical value of 0.146. McLeod-Li tests of the US and UK variables returned sets with all tested lags, indicating that the variables feature non-constant variance. A McLeod-Li test of CA returned a set with lags 6 through 20, also indicating non-constant variance. Thus, each of the variables are linear-trend stationary, but not strictly stationary. Expectedly, multivariate ARCH effects were detected in a March test of the variables, which produced p-values of less than 0.05, indicating a rejection of the null hypothesis that multivariate ARCH effects are not present.

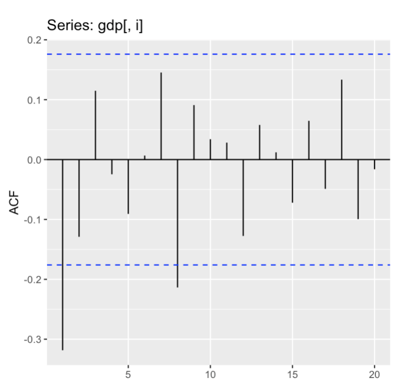
Ljung-Box tests of the null hypotheses that the variables US, UK, and CA do not feature auto-correlation returned p-values of 0.0026, 0.0017, and 0.012 respectively. Thus, the hypotheses of non-auto-correlation are rejected. Auto-correlation is also demonstrated in each of the ACF plots below, although the variables only have one to two lags with significant auto-correlation. To the right is a cross-correlation matrix of the three variables over the first five lags. The matrices indicate that there is statistically significant cross-correlation between variables through the fourth lag. This indicates that a VAR(4) model would be appropriate.

```
CCM at lag:  1
+ + +
+ + +
+ + +
CCM at lag:  2
+ + +
+ + +
. + +
CCM at lag:  3
+ . +
. + .
. . .
CCM at lag:  4
+ . .
+ . .
+ . .
CCM at lag:  5
. . .
. . .
. . .
```

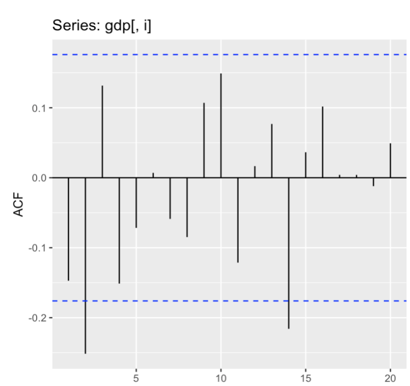
US



UK



CA

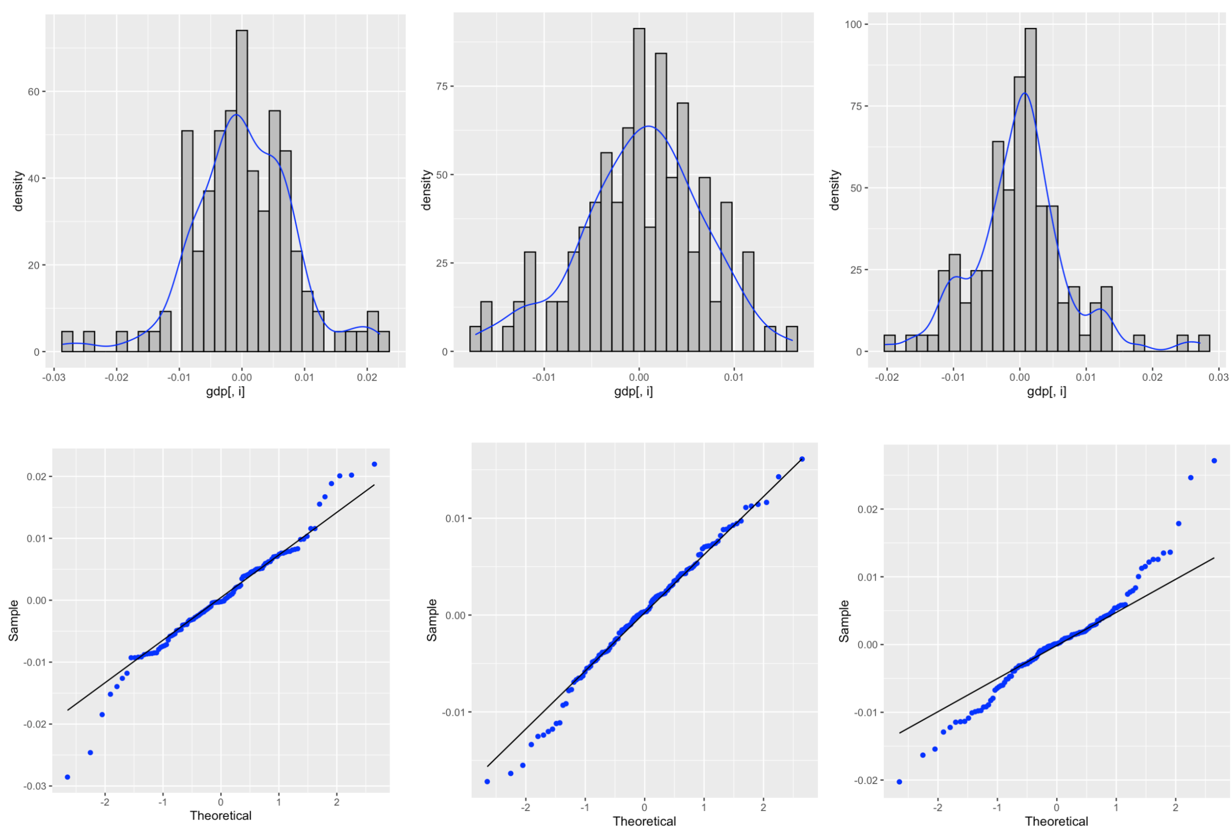


As the histograms and QQ plots below demonstrate, differencing results in the distributions of the variables more closely approximate a normal distribution. The variables now show unimodal distributions with fewer observations deviating from the normal line. The 95% confidence intervals for the skew and excess kurtosis after differencing are shown below. Again, the confidence intervals have shrunk, which is expected given the differencing. Notably, the 95% confidence interval for excess kurtosis for UK now contains zero. The MVN function now indicates that the variable is normally distributed, as shown in the output below the table. However, US and CA are still not normally distributed, and neither is the multivariate time series that contains all three variables.

Variable	Skew LL	Skew	Skew UL	Kurt LL	Kurt	Kurt UL
US	-0.907	-0.208	0.462	0.333	1.434	2.858
UK	-0.653	-0.306	0.033	-0.538	0.038	0.671
CA	-0.167	0.489	1.274	0.440	1.856	3.676

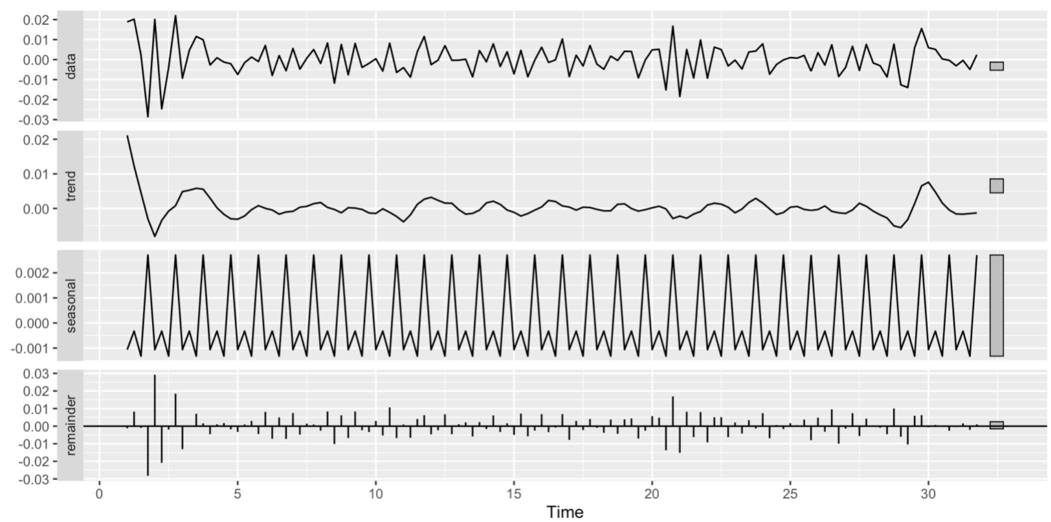
```
$multivariateNormality
             Test         Statistic                    p value Result
1 Mardia Skewness 16.3928276206614         0.0889259417883717    YES
2 Mardia Kurtosis 6.31718408972876 0.000000000266372257584635     NO
3             MVN             <NA>                       <NA>     NO

$univariateNormality
              Test    Variable Statistic   p value Normality
1 Anderson-Darling log_us_diff    0.7579    0.0474    NO
2 Anderson-Darling log_uk_diff    0.3849    0.3882    YES
3 Anderson-Darling log_ca_diff    1.4725    0.0008    NO
```

US UK CA

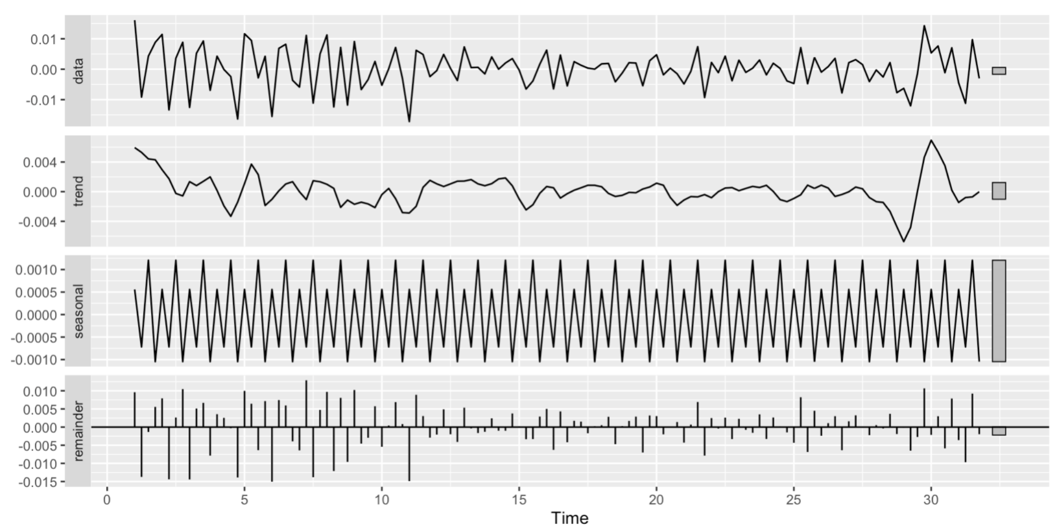


Below are decompositions of each variable. The line plots demonstrate how the twice-differenced, log-transformed data show greater differences in variance than was apparent in the original line plots. As expected, the trend lines for each look like moderately smoothed versions of the line plots. Similarly, the residual plots feature greater variance when the line plots have greater variance. These indicate that the decomposition did not explain much of the variance in the time series. It is notable, however, that the variables feature distinct seasonal components. The US and UK both feature seasonal patterns with one large spike and one small spike per period, however, the smaller spike in the UK's seasonal pattern is larger. The seasonal pattern of CA is unlike either of the other two, featuring a single spike per period with a small step at the top of the spike.

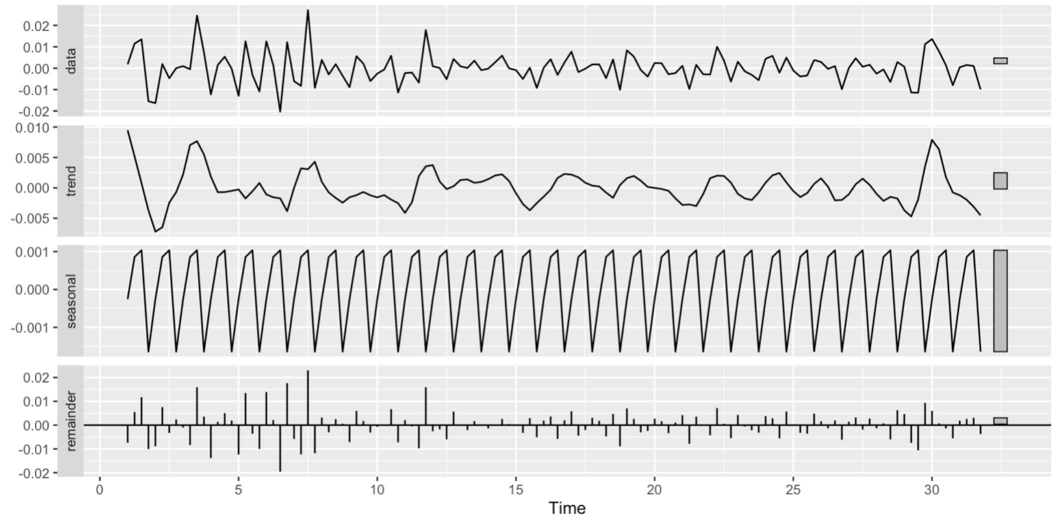
US



UK



CA



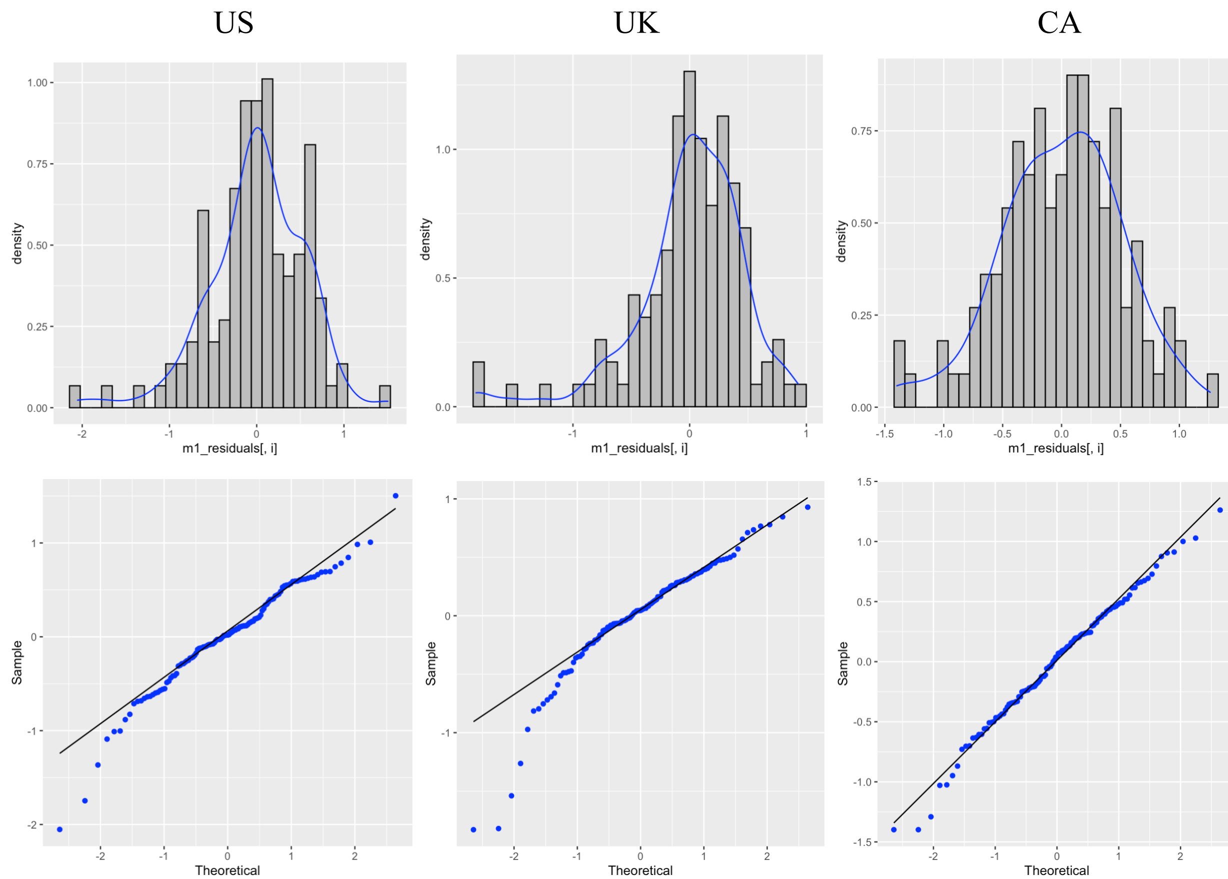
2.2 VAR(4) Model

As stated in the previous section, the cross-correlation matrices for the three variables indicate that a VAR(4) model is appropriate. The precise values for the coefficients of each equation are shown in the model output below. The variables 1, 2, and 3 are in the order UK, CA, US. The first output represents the constants of each equations, followed by the coefficients. This model has 36 degrees of freedom.

```
$Ph0
[1] 0.147957 0.077574 0.238677

$Phi
        [,1]     [,2]     [,3]     [,4]     [,5]      [,6]     [,7]      [,8]     [,9]    [,10]     [,11]     [,12]
[1,] 0.51569 0.071863 0.063904 -0.05037  0.16004 -0.001984 0.052363 -0.278830 0.141145 0.040130  0.261731 -0.246485
[2,] 0.37782 0.316026 0.409626 -0.17399 -0.25407  0.062947 0.096153  0.120348 0.013661 0.074697 -0.090308 -0.097807
[3,] 0.51905 0.173045 0.150389 -0.21784 -0.15931  0.225606 0.047828 -0.078562 0.073845 0.154090 -0.151819 -0.053496
```

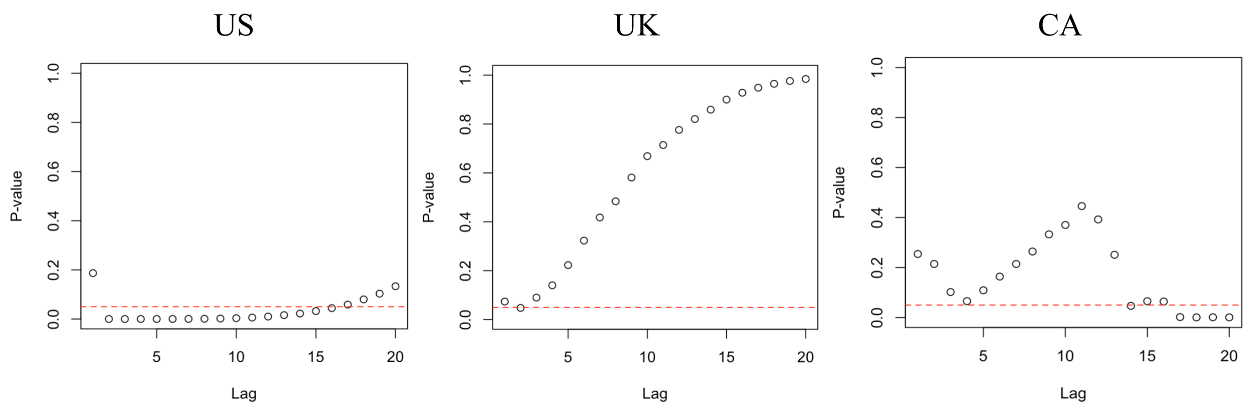
The residuals of this model indicate that it is a decent fit of the data. The individual variables' are linear-trend stationary and do not feature auto-correlation nor cross-correlation. Whether or not they are individually approximately normally distributed is debatable; they have visually desirable shapes but do not pass statistical tests. They are also not normally distributed as a multivariate time series, they do not feature strict stationarity, and they do have business cycles.

The histograms on the following page show that each of the variables' residuals appear approximately normal; they are unimodal with decent symmetry. While the tails do deviate from the normal line, the deviation is relatively small. However, only the residuals for CA had a 95% confidence interval for skew that included zero, as shown on the next page. None of the residuals had 95% confidence intervals that contained zero. The MVN function, which produced the plot after the table, agrees that only the CA residuals are normal and that the multivariate time series with all residuals are not normally distributed.



Residuals	Skew LL	Skew	Skew UL	Kurt LL	Kurt	Kurt UL
US	-1.382	-0.658	-0.053	0.101	1.512	3.336
UK	-1.976	-1.306	-0.793	0.090	3.010	5.441
CA	-0.709	-0.274	0.109	-0.472	0.123	0.814

```
$multivariateNormality
             Test        Statistic                       p value Result
1 Mardia Skewness 60.4965795051728    0.0000000291881188013298     NO
2 Mardia Kurtosis 7.12700857984278 0.0000000000102584607475364     NO
3             MVN             <NA>                          <NA>     NO

$univariateNormality
              Test          Variable Statistic   p value Normality
1 Anderson-Darling m1.residuals...3.    0.9035  0.0206      NO
2 Anderson-Darling m1.residuals...1.    2.3484  <0.001      NO
3 Anderson-Darling m1.residuals...2.    0.2531  0.7292      YES
```



As the table below shows, the residuals all have a mean within 3 decimal places of zero with t-test 95% confidence intervals that contain zero. Linear-trend stationarity is suggested for all of the residuals by ADF and KPSS tests, which all returned p-values/test-statistics beneath those required to assert stationarity. The values of each test are shown in the second table. However, the residuals are not strictly stationary. The McLeod-Li above show that all residuals feature lags with non-constant variance. This is corroborated by a March test of the residuals with the null hypothesis that multivariate ARCH effects are not present, which returns p-values of less than 0.05, indicating multivariate ARCH effects are present. Thus, they are linear-trend stationary, but not strictly stationary.

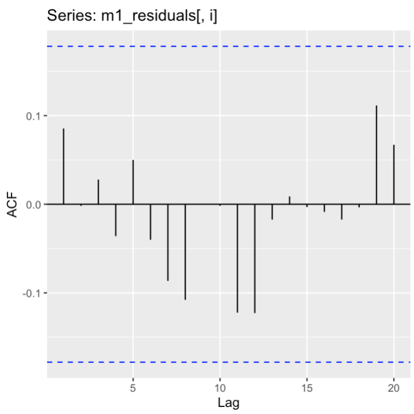
Residuals	Mean LL	Mean	Mean UL
US	-0.098	0.000	0.098
UK	-0.086	0.000	0.086
CA	-0.092	0.000	0.092

Residuals	ADF p-value	ADF Crit Val	KPSS statistic	KPSS Crit Val	Ljung-Box p-value
US	0.01	0.05	0.0706	0.146	0.35
UK	0.01	0.05	0.0656	0.146	0.69
CA	0.01	0.05	0.049	0.146	0.69

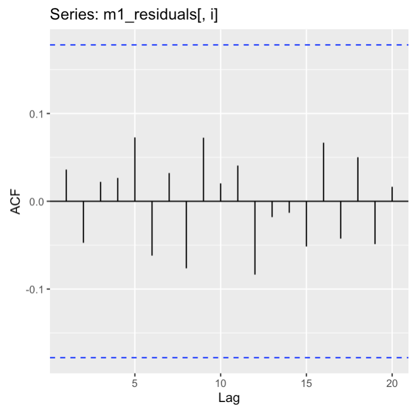
The table above shows that Ljung-Box tests of the residuals produced p-values above 0.05, indicating acceptance of their null hypotheses that the residuals do not feature auto-correlation. This is agreed with by the ACF plots on the top of the next page, which show that none of the residuals have lags with statistically significant auto-correlation. According to the CCM output on the right, the residuals do not feature cross-correlation. This is corroborated by the MQ function, which produced the plot below the ACF plots, which shows p-values of 1 for all lags, indicating acceptance of the null hypothesis that the residuals do not feature cross-correlation.

```
CCM at lag:  1
. . .
. . .
. . .
CCM at lag:  2
. . .
. . .
. . .
CCM at lag:  3
. . .
. . .
. . .
CCM at lag:  4
. . .
. . .
. . .
CCM at lag:  5
. . .
. . .
. . .
```

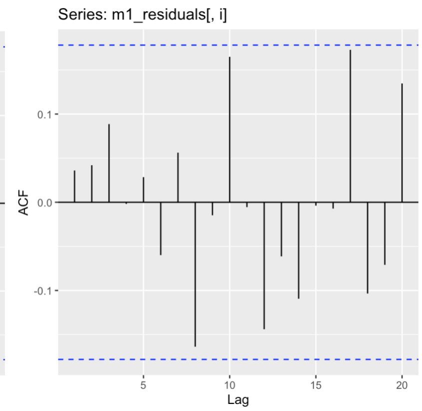
US

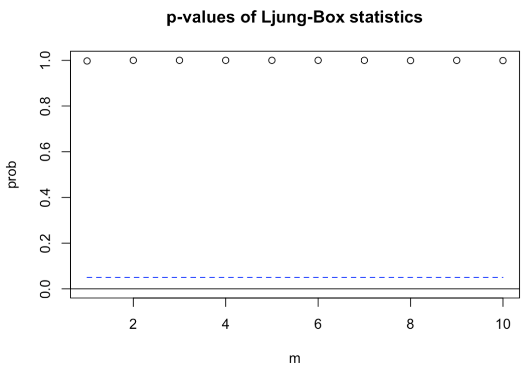


UK



CA





As previously mentioned, the residuals do have business cycles. Their respective frequencies are shown in the table below. They indicate that the model failed to account for some patterns of variance for each of the time series.

Residuals	Freq 1	Freq 2	Freq 3	Freq 4	Freq 5
US	10.478	2.18	3.554	5.227	2.635
UK	5.826	2.898	3.71	2.404	11.562
CA	10.229	2.206	3.667	2.804	4.77

2.3 Simplified VAR(4) Model

A simplified version of the previous model created with an alpha of 0.05 produced a model that removed the coefficients where zeros are present in the output shown on the right. The precise values of the remaining constants and coefficients are shown in the outputs below.

```
      [,1] [,2] [,3]
 [1,]    1    0    1
 [2,]    1    1    1
 [3,]    0    1    1
 [4,]    0    1    0
 [5,]    0    0    0
 [6,]    1    1    0
 [7,]    0    0    0
 [8,]    0    0    0
 [9,]    1    0    0
[10,]    0    0    0
[11,]    0    0    0
[12,]    1    0    1
[13,]    1    0    0
```

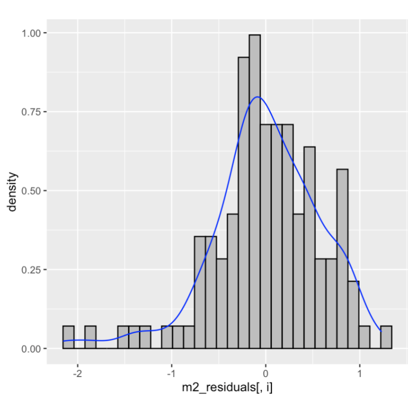
```
$Ph0
[1] 0.20443 0.00000 0.32321
```

```
$Phi
        [,1]    [,2]    [,3] [,4]     [,5] [,6] [,7]     [,8] [,9] [,10]    [,11]    [,12]
[1,] 0.56509 0.00000 0.00000    0  0.28370    0    0 -0.24891    0     0  0.29639 -0.24285
[2,] 0.37639 0.29890 0.40798    0 -0.15298    0    0  0.00000    0     0  0.00000  0.00000
[3,] 0.50816 0.26786 0.00000    0  0.00000    0    0  0.00000    0     0 -0.19901  0.00000
```

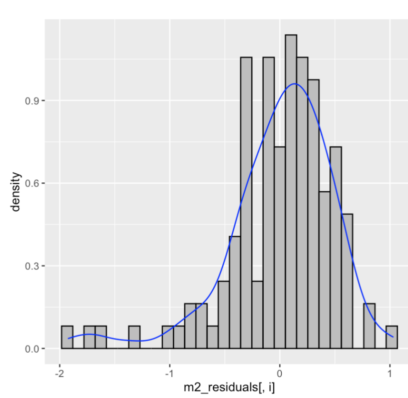
Overall, the residuals of this model indicate slightly worse fit than that of the previous model. For the most part, the residuals feature similar characteristics. The residuals of US and UK are not normally distributed, while they are for CA. The residuals are not multivariate normal and they do not feature cross-correlation. They are linear-trend stationary but not strictly stationary. Where the residuals for this model differ are in auto-correlation, which is featured in the US residuals.

The histograms below show that each of the variables' residuals appear approximately normal. The QQ plots on the following page show the extent to which the tails deviate from normal; it is small but still present. Only the residuals for CA had a 95% confidence interval for skew that included zero, as shown on the next page. None of the residuals had 95% confidence intervals that contained zero. The MVN function, which produced the plot after the table, agrees that only the CA residuals are normal and that the multivariate time series with all residuals are not normally distributed.

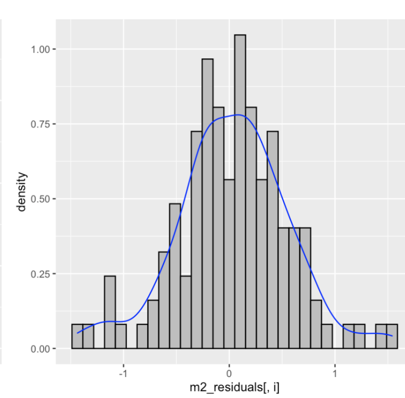
US

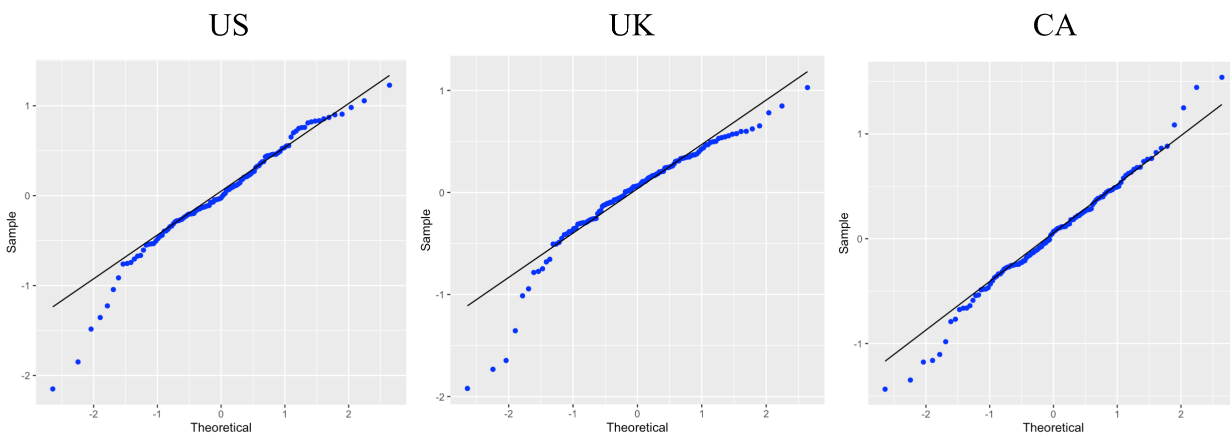


UK



CA





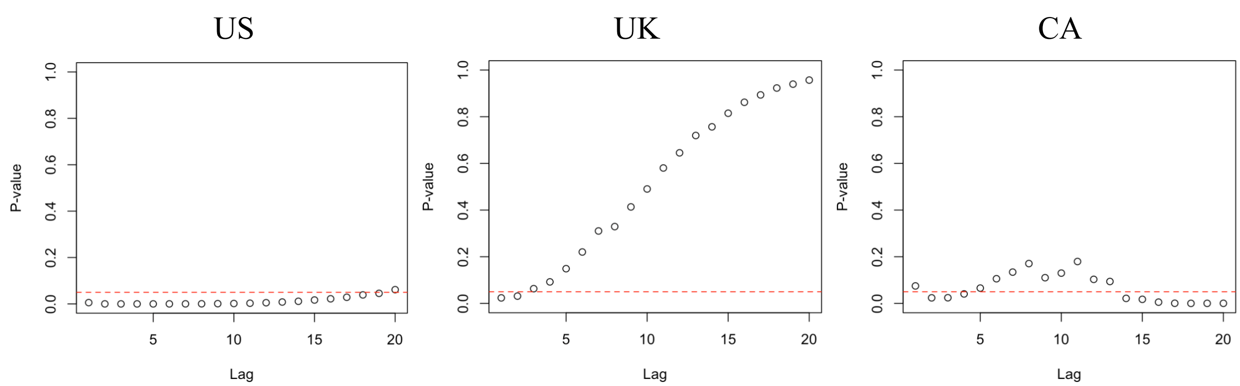
Residuals	Skew LL	Skew	Skew UL	Kurt LL	Kurt	Kurt UL
US	-1.338	-0.721	-0.220	0.112	1.453	3.208
UK	-1.974	-1.314	-0.807	1.167	2.923	5.109
CA	-0.581	-0.069	0.414	-0.016	0.619	1.376

```
$multivariateNormality
             Test        Statistic                        p value Result
1 Mardia Skewness 62.7655182105568     0.000000010821017851442 5     NO
2 Mardia Kurtosis 7.26630498009094 0.000000000000369482222595252     NO
3             MVN             <NA>                           <NA>     NO

$univariateNormality
              Test          Variable Statistic   p value Normality
1 Anderson-Darling m2.residuals...3.    0.7832    0.041      NO
2 Anderson-Darling m2.residuals...1.    2.1878   <0.001      NO
3 Anderson-Darling m2.residuals...2.    0.4183   0.3236      YES
```

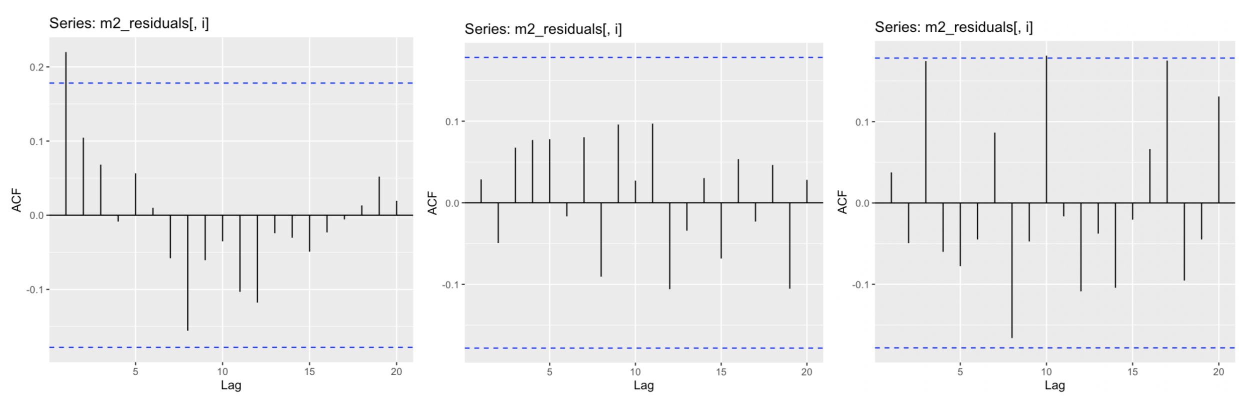
As the table below shows, all of the residuals are mean zero. They are also linear-trend stationary, as shown by the results of ADF and KPSS tests, which all returned p-values/test-statistics beneath those required to assert stationarity. The values of each test are shown in the table on the top of the next page. However, the residuals are not strictly stationary. The McLeod-Li outputs on the following page show that all residuals feature lags with non-constant variance. This is corroborated by a March test of the residuals with the null hypothesis that multivariate ARCH effects are not present, which returns a p-value of less than 0.05, indicating multivariate ARCH effects are present. Thus, they are linear-trend stationary, but not strictly stationary.

Residuals	Mean LL	Mean	Mean UL
US	-0.104	0.000	0.104
UK	-0.088	0.000	0.088
CA	-0.064	0.031	0.126

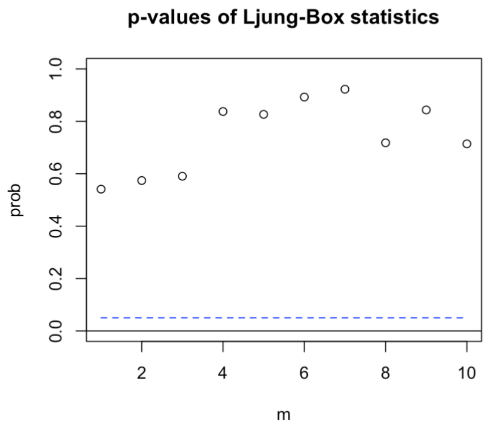
Residuals	ADF p-value	ADF Crit Val	KPSS statistic	KPSS Crit Val	Ljung-Box p-value
US	0.01	0.05	0.0849	0.146	0.016
UK	0.01	0.05	0.0832	0.146	0.75
CA	0.01	0.05	0.037	0.146	0.68



The table above shows the results of Ljung-Box tests of the residuals, producing p-values above 0.05 for the UK and US residuals, and a p-value beneath 0.05 for US residuals. The null hypotheses of these tests are that the residuals do not feature auto-correlation, and thus, from the p-values, it is demonstrated that the US residuals do feature auto-correlation, while the UK and CA residuals do not. This can also be seen visually in the ACF plots below. However, the plots show that the CA residuals have multiple lags with auto-correlation just barely below the significance line.



According to the CCMs on the top of the next page, the residuals do not feature cross-correlation, as there are no positive or negative signs in the off diagonals. This is corroborated by the MQ function, which produced the plot on the next page, which shows p-values of above 0.05 for all lags, indicating acceptance of the null hypothesis that the residuals do not feature cross-correlation.




```
CCM at lag:  1
+ . .
. . .
. . .
CCM at lag:  2
. . .
. . .
. . .
CCM at lag:  3
. . .
. . .
. . .
CCM at lag:  4
. . .
. . .
. . .
CCM at lag:  5
. . .
. . .
. . .
```

As previously mentioned, the residuals do have business cycles. Their respective frequencies are shown in the table below. They indicate that the model failed to account for some patterns of variance for each of the time series.

Residuals	Freq 1	Freq 2	Freq 3	Freq 4	Freq 5	Freq 6
US	8.978	2.656	68.048	4.874	2.185	3.395
UK	6.112	3.778	2.381	2.901	86.072	11.005
CA	2.201	3.776				

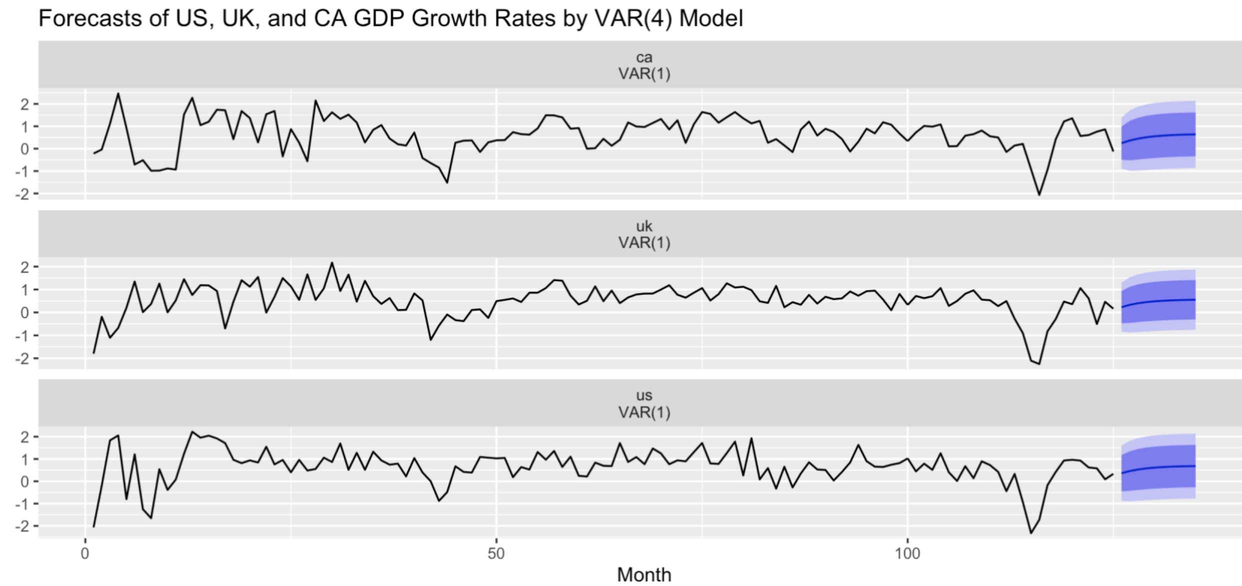
2.4 Model Comparison

I prefer the first model to the second. As I described in the previous section, the residuals of the simplified model indicate a slightly worse fit than the first. The largest difference is that the residuals of the second model's residuals of the US variable features auto-correlation, while the first model's do not. There are also numerous spikes in the ACF plot of the second model's residuals for the CA variable that are essentially at the significance line. The distribution of the first model's residuals are also more ideal than those of the second. While it is true that the distributions of the residuals in both model were identical in the sense that those of the US and UK variables were not normally distributed while those of the CA variable were, the residuals of the first model have more observations that fall on the normal line in their QQ plots.

There is an argument to be made that the second model is preferable to the first because of its simplicity. While this is true, there was not a distinguishable difference in compute time, making the difference in complexity negligible.

2.5 Forecasts

Below are my forecasts of the growth rates of the US, UK, and Canada produced by the unsimplified VAR(4) model.



3. Report

I have created a model that predicts the growth rates of the GDPs of the US, UK, and Canada. These predictions are based on each country's recent GDP growth, as well as on the relationship that exists among the GDPs of these three countries. The predictions are shown above. The dark blue line represents the expected growth rates, with the shaded regions indicating probabilities of other possible growth rates. The darker shaded region indicates where growth rates are expected to fall within 95% probability, and the lighter shaded regions are the 2.5% most extreme situations.

The predictions are that over the short-term, GDP growth rates of each country are expected to increase and stay positive. They do indicate the possibility of GDP growth rates becoming negative, but with a less than 50% probability.

There are several features of the GDPs of these three countries that are notable. The first is that they behave similarly. Historically, the GDPs of these countries have very high correlations. As the plots show, their growth rates have behaved similarly as well. This synchronicity in level and growth rate is why the predictions look similar. It is also notable that my analysis did not show a lead/lag relationship among the GDPs, thus it is not the case that any one or two GDPs will mirror in the near-term what the other GDP(s) have done in the near-past. Rather, they move together.