US Real GDP Growth Prediction from Lagged GDP Growth & Spread Between 10-year Treasury Constant Maturity and 3-month Treasury Constant Maturity

John Balzani

1/20/2020

Abstract

Can the spread between long-term and short-term interest rates predict future economic growth? The purpose of this research is to explore whether or not the difference between the 10-year Treasury Constant Maturity yield and the 3-month Treasury Constant Maturity yield (10-year 3-month term spread) is significant in predicting US real GDP growth, above the past values of real GDP growth itself. In particular, one of the purposes of this research is to explore if these predictions can be made accurately for the 2008 recession and current post-2008 recession business cycle. Different lagged values of the 10-year 3-month term spread and real GDP growth are tested in an Autoregressive Distributed Lag (ADL) model, and this is compared to an Autoregressive (AR) model. Data from 1983 - April 2008 are used to estimate the model, and data after that are used to evaluate the model performance. It is found that the 1-quarter and 2-quarter lagged values of real GDP growth and the 2-quarter lagged value of the 10-year 3-month term spread are significant in predicting US real GDP growth from 1983-2008. However, these do not accurately predict US real GDP growth from July 2008 - July 2019 any better than a simple AR(2) model, with a RMSE of 2.36% annualized quarterly real growth and a MAE of 1.61% for the ADL model, compared to a RMSE of 2.33% and MAE of 1.62% for the AR model.

Introduction:

As yields are an indication of the market's projections of future economic growth, it would make sense that yields may predict future economic growth. The difference between long term and short term yields may predict future GDP growth, as it reflects market particpants' expectations of short term and long term economic growth. Past research has shown a link between the spread between long term and short term yields and future GDP growth. Laurent (1988) finds evidence that the spread between the 20-year US Treasury Constant Maturity rate and the federal funds rate can predict future GNP growth, and Estrella and Hardouvelis (1991) find that the term spread can predict future GDP growth. However, the relationship between term spread and US GDP growth has not been the same for all time periods. Haubrich and Dombrowski (1996) find that adding the term spread to a model of lagged GDP growth results in a worse fit for 1985-1995 data. In order to explore whether or not there is a relationship between term spread and real GDP growth with the most recent data, various ADL models were tested and compared to an AR model. Of particular interest was whether or not the term spread can predict real GDP growth for the 2008 recession and current post-2008 recession business cycle, so this period was used for the estimation sample.

Methodology:

Below is a short description of the data used in this report.
US Real GDP Growth:
Units: Percent Change from Previous Period, Seasonally Adjusted Annual Rate
Frequency: Quarterly
Source: https://fred.stlouisfed.org/series/A191RL1Q225SBEA

10-year Treasury Constant Maturity Rate Minus 3-month Treasury Constant Maturity Rate (T10Y3M):
Units: Percent, Not Seasonally Adjusted
Frequency: Daily
Source: https://fred.stlouisfed.org/series/T10Y3M
T10Y3M data is confined to the data on first day of the month in which data is available. Only quarterly data matching the data on real GDP growth is used.

After tests for a unit root and Granger causality, ADL and AR models are tested for the estimation sample (Jan 1893 - Apr 2008). These models are then tested and evaluated for the evaluation sample (Jul 2008 - Jul 2019). April 2008 was chosen to separate the estimation and evaluation samples because this was the latest data point before the serious recession of 2008.

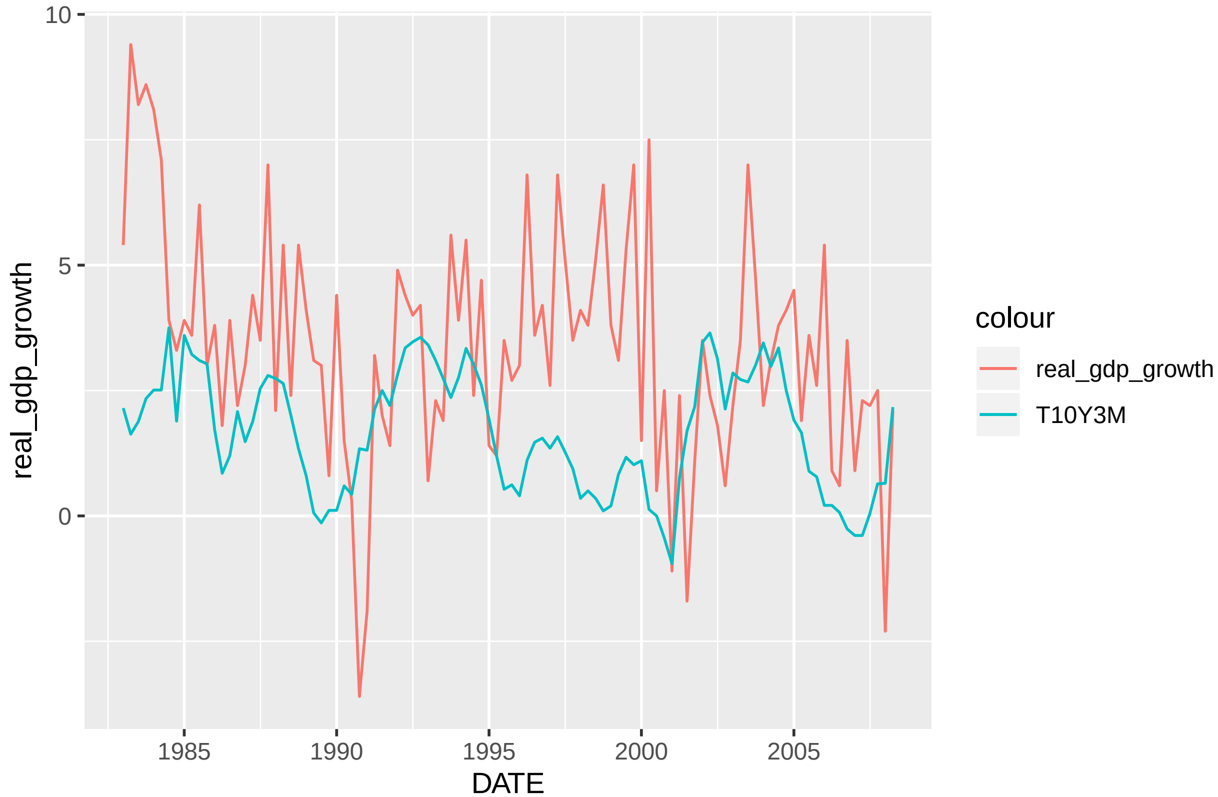
Unrestricted Model:
For the unrestricted model, the 4 most recent lagged values of T10Y3M and the 4 most recent lagged values of real GDP growth are tested. A 5% level of significance is used for all tests throughout the study.

The unrestricted model is shown below.
real_gdp_growth = alpha + beta1*T10Y3M_lag + beta2*T10Y3M_lag2 + beta3*T10Y3M_lag3 + beta4*T10Y3M_lag4 + gamma1*real_gdp_growth_lag + gamma2*real_gdp_growth_lag2 + gamma3*real_gdp_growth_lag3 + gamma4*real_gdp_growth_lag4 + epsilon, where epsilon is the error term

Implementation:

Below is a plot of the term spread and the real GDP growth rate from 1983 - April 2008:

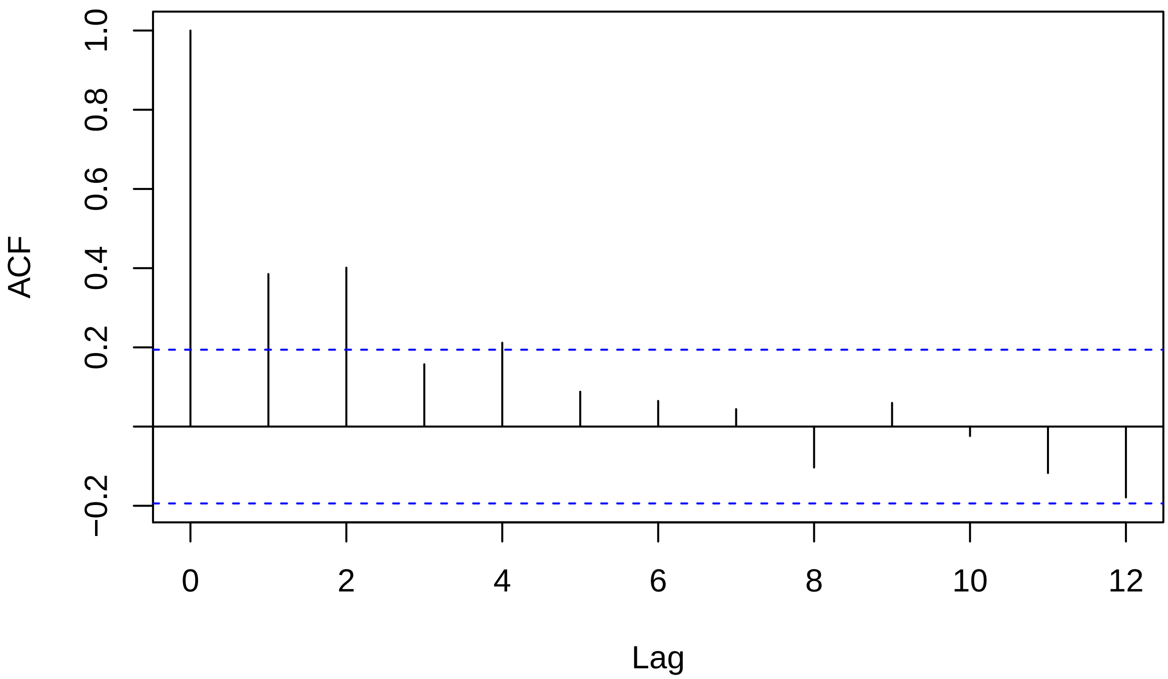
Figure 1: Real GDP Growth and T10Y3M 1983−Apr 2008



It appears that T10Y3M and real GDP growth are correlated. There does not appear to be seasonality or a long-term trend. The lack of seasonality matches past findings indicating that there is no seasonal trend in monthly interest rates (Wooldridge, 2009), and the real GDP growth data used is already seasonally adjusted.

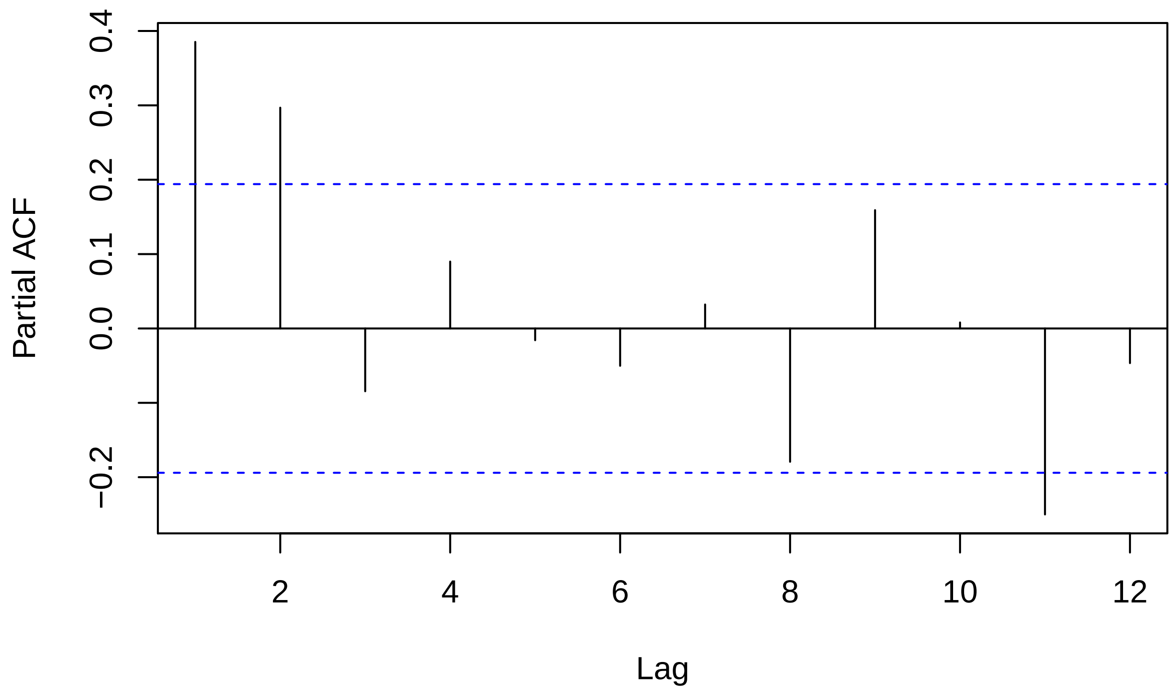
In order to explore the number of lags to test for real GDP growth, a correlogram of the ACFs was made (Figure 2). It can be seen that the 1-4 lagged values of real GDP growth are correlated with its current level.

Figure 2: ACFs of Lags of Real GDP Growth

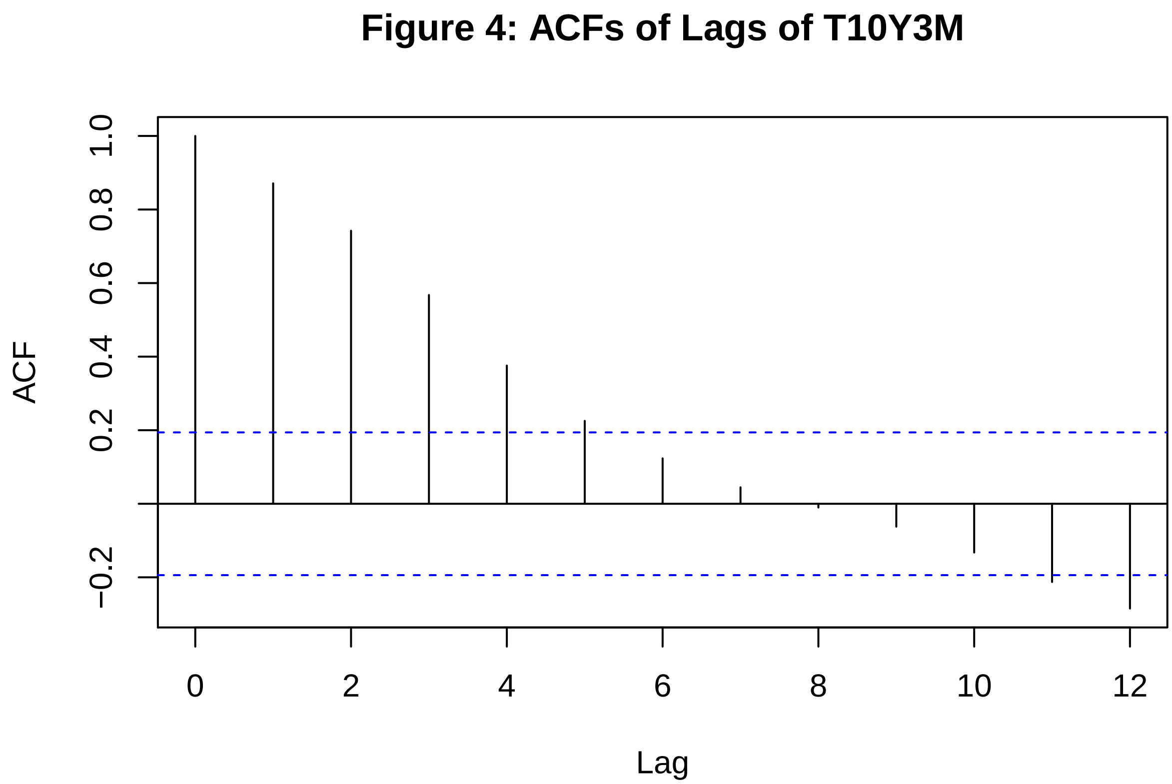


A correlogram of the PACF for real GDP growth was also generated (Figure 3). It can be seen that the first and second lags are correlated with its current level. The 11th lag is also significant, but this may be an artifact of the data rather than real causation because none of the other lags after lag 2 are significant.

Figure 3: PACFs of Lags of Real GDP Growth

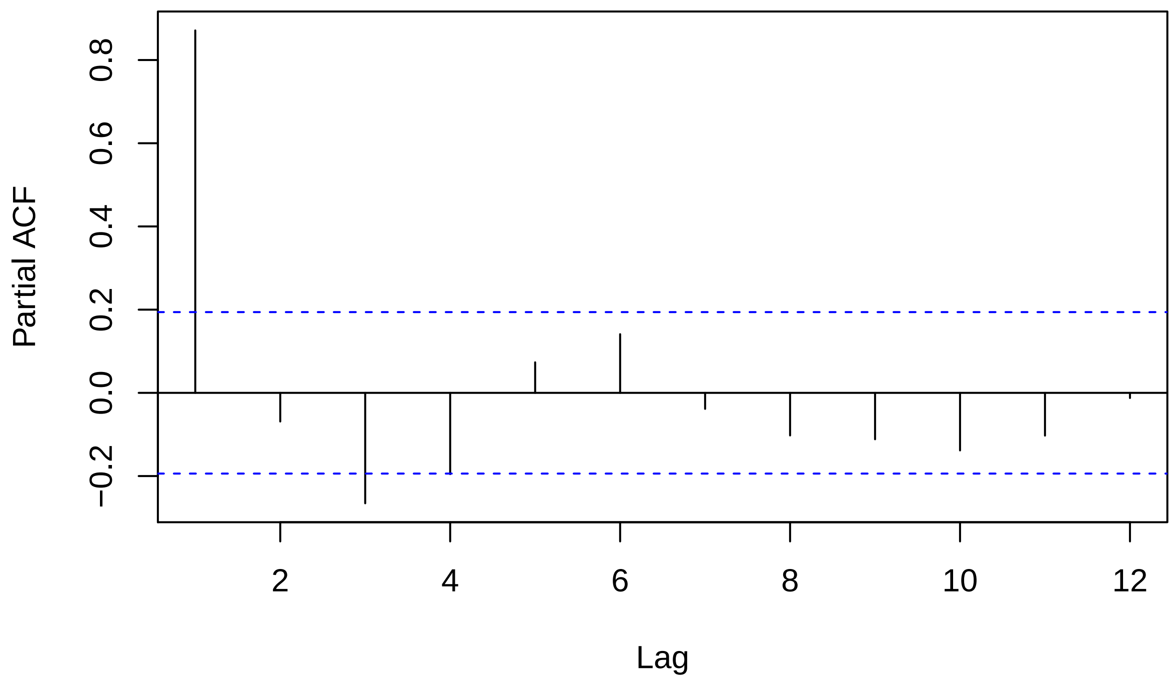


A correlogram of the ACFs of T10Y3M was also generated. From this, we can see a clear trend, with lags 1-6 being significant and also lags 11 and 12 (Figure 4).



However, when looking at a correlogram of the PACF, we can see that only lag 1, lag 3, and potentially lag 4 are significant (Figure 5).

Figure 5: PACF of Lags of T10Y3M



Test for Stationarity - Augmented Dickey-Fuller Test (ADF) for Real GDP Growth:
First an Augmented Dickey-Fuller test is performed to test whether or not real GDP growth is stationary. An ADF test for all 4 lags of delta_real_gdp_growth is used to start the ADF test, after reviewing the results of the ACFs and PACFs and after considering economic theory. It is reasonable to think that economic growth can be correlated to the future level of growth 1 year in the future. The model for the ADF test is shown below.
delta_real_gdp_growth = alpha_adf_1 + rho*real_gdp_growth_lag + gamma_adf_1*delta_real_gdp_growth_lag + gamma_adf_2*delta_real_gdp_growth_lag2 + gamma_adf_3*delta_real_gdp_growth_lag3 + gamma_adf_4*delta_real_gdp_growth_lag4 + epsilon_adf_1

Ng and Perron recommend the following process for determining the lag length for the ADF test (Ng and Perron, 2001). First one should decide on a maximum value for the lag length, and then perform the test with that lag length. If the longest lagged variable in the regression has a t value with an absolute value of less than 1.6, then reduce the lag length by one and perform the test again.

After performing an ADF test for lags 1-4 of delta_real_gdp_growth, it can be concluded that the ADF test should be repeated with lag length 3, as the absolute value of the t statistic of the last lagged value is 1.41, which is less than 1.6.

An ADF test with lags 1-3 of delta_real_gdp_growth is then performed. From this, it can be concluded that the ADF test should be repeated with lag length 2, as the absolute value of the t statistic of the last lagged value is 0.14, which is less than 1.6.

After performing an ADF test with lags 1 and 2 of delta_real_gdp_growth, it can be concluded that the ADF test should be repeated with lag length 1, as the absolute value of the t statistic of the last lagged value is 1.53, which is less than 1.6.

Finally, an ADF test with lag 1 of delta_real_gdp_growth is performed; it is seen that the t value of real_gdp_growth_lag is -4.46. This is below the critical value of -2.9, so we reject the null hypothesis of non-stationarity of real GDP growth. Real GDP growth is stationary.

Test for Stationarity - Augmented Dickey-Fuller Test for T10Y3M:
The ADF test is started with all 4 lags of delta_T10Y3M, based on the results of the PACF chart and economic theory. It is reasonable to think that the level of the T10Y3M yield will be correlated to its value in the near future, say, within the next year.

Model for ADF test: delta_T10Y3M = alpha_adf_2 + rho1*T10Y3M_lag + beta_adf_1*delta_T10Y3M_lag + beta_adf_2*delta_T10Y3M_lag_2 + beta_adf_3*delta_T10Y3M_lag3 + beta_adf_4*delta_T10Y3M_lag4 + epsilon_adf_2

After performing an ADF test for lags 1-4 of delta_T10Y3M, it can be concluded that the ADF test should be repeated with lag length 3, as the absolute value of the t statistic of the last lagged value is 0.48, which is less than 1.6.

From executing an ADF test for lags 1-3 of delta_T10Y3M, it is seen that the t value of the last lagged value is 2.4, which is above 1.6, so we next examine the t value of T10Y3M_lag. The t value of T10Y3M_lag is -3.925, which is below the critical value of -2.9, so we reject the null hypothesis of non-stationarity for T10Y3M. T10Y3M is stationary.

Cointegration: As the series are both stationary, no need to test for cointegration.

Granger Causality Test for T10Y3M on Real GDP Growth:
Next a test for Granger causality is performed to see if T10Y3M Granger-causes real GDP growth. The unrestricted model is shown below.
real_gdp_growth = alpha + beta1*T10Y3M_lag + beta2*T10Y3M_lag2 + beta3*T10Y3M_lag3 + beta4*T10Y3M_lag4 + gamma1*real_gdp_growth_lag + gamma2*real_gdp_growth_lag2 + gamma3*real_gdp_growth_lag3 + gamma4*real_gdp_growth_lag4 + epsilon

The restricted model is shown below.
real_gdp_growth = alpha_r + gamma_r1*real_gdp_growth_lag + gamma_r2*real_gdp_growth_lag2 + gamma_r3*real_gdp_growth_lag3 + gamma_r4*real_gdp_growth_lag4 + epsilon_r

The p value of the F statistic is 0.09. This is above 0.05, so we cannot reject the null hypothesis that T10Y3M does not Granger cause real GDP growth at a 5% level of significance. However, since much other literature indicates that T10Y3M can predict real GDP growth, I will explore its use as a predictor.

Models:

Unrestricted Model:
For the unrestricted model, the 4 most recent lagged values of T10Y3M and the 4 most recent lagged values of real GDP growth are tested. A 5% level of significance is used for all tests throughout the study. These lag values are indicated from the ACF/PACF charts and, more importantly, are reasonable based on economic theory.

The unrestricted model is shown below.
real_gdp_growth = alpha + beta1*T10Y3M_lag + beta2*T10Y3M_lag2 + beta3*T10Y3M_lag3 + beta4*T10Y3M_lag4 + gamma1*real_gdp_growth_lag + gamma2*real_gdp_growth_lag2 + gamma3*real_gdp_growth_lag3 + gamma4*real_gdp_growth_lag4 + epsilon, where epsilon is the error term

After testing the unrestricted model with lags 1-4 of T10Y3M and real_gdp_growth, it can be concluded that lags 1-4 of T10Y3M and lags 3 and 4 of real GDP growth are not significant in predicting US real GDP growth in this model. As lag 1 of T10Y3M has the highest p value (0.89), and since it is reasonable to think that the effect of T10Y3M may be longer-term than a 1 quarter lag, this variable is eliminated and the model is re-estimated.

A model of lags 2-4 of T10Y3M and lags 1-4 of real GDP growth is then tested. It can be seen that lag 4 of T10Y3M and lags 3 and 4 of real GDP growth are not significant in predicting US real GDP growth in this model. As lag 4 of real GDP growth has the highest p value out of the lag 4 values (0.76), it is eliminated and the model is re-estimated.

A model of lags 2-4 of T10Y3M and lags 1-3 of real GDP growth is tested. It can be concluded that lags 3 and 4 of T10Y3M and lag 3 of real GDP growth are not significant in predicting US real GDP growth in this model. As lag 4 of T10Y3M has the highest p value (0.13), it is eliminated and the model is re-estimated.

After testing a model of lags 2 and 3 of T10Y3M and lags 1-3 of real GDP growth, the result is that lag 3 T10Y3M and lag 3 of real GDP growth are not significant in predicting US real GDP growth in this model. As lag 3 of T10Y3M has the highest p value (0.20), it is eliminated and the model is re-estimated.

A model of lag 2 of T10Y3M and lags 1-3 of real GDP growth is tested next. The result is that lag 3 of real GDP growth is not significant in predicting US real GDP growth in this model, with a p value of 0.12. It is eliminated and the model is re-estimated.

After testing a model with lag 2 of T10Y3M and lags 1 and 2 of real GDP growth, it is seen that the twice lagged values of T10Y3M and once and twice lagged values of real GDP growth are jointly significant in predicting US real GDP growth at the 5% level. A summary of this model is shown below.

```
##
## ================================================
##                          Dependent variable:
##                      ---------------------------
##                            real_gdp_growth
## ------------------------------------------------
## T10Y3M_lag2                    0.354**
##                                (0.175)
##
## real_gdp_growth_lag            0.239**
##                                (0.098)
##
## real_gdp_growth_lag2           0.225**
##                                (0.097)
##
## Constant                       1.214**
##                                (0.464)
##
## ------------------------------------------------
## Observations                     102
## R2                              0.219
## Adjusted R2                     0.195
## Residual Std. Error        2.076 (df = 98)
## F Statistic             9.163*** (df = 3; 98)
## ================================================
## Note:                *p<0.1; **p<0.05; ***p<0.01
```

Autoregressive Model (AR Model): The ADL model is compared to an AR model of only GDP growth, in order to further explore if T10Y3M has predictive values above the past values of the real_gdp_growth time series. An AR(4) model consisting of the 4 last lagged values of real_gdp_growth is first tested. This model is equivalent to the restricted model tested earlier in the Granger causality test.

real_gdp_growth = alpha_ar + gamma_ar1*real_gdp_growth_lag + gamma_ar2*real_gdp_growth_lag2 + gamma_ar3*real_gdp_growth_lag3 + gamma_ar4*real_gdp_growth_lag4 + epsilon_ar

After testing an AR Model with lags 1-4 of real_gdp_growth, it can be seen that the 3 and 4 times lagged values of real GDP growth are not significant in predicting US real GDP growth in this model. As lag 4 is the longest lag and has the highest p value (0.89), it is eliminated and the model is re-estimated.

Next, an AR Model with lags 1-3 of real_gdp_growth is tested. The conclusion of this test is that the 3 times lagged values of real GDP growth is not significant in predicting US real GDP growth in this model, with a p value of 0.13. It is eliminated and the model is re-estimated.

Finally, testing an AR Model with lags 1-2 of real_gdp_growth yields the conclusion that lags 1 and 2 of real GDP growth are significant in predicting real GDP growth at the 5% level of significance. A summary of this model is shown below.

```
##
## ================================================
##                          Dependent variable:
##                      ---------------------------
##                            real_gdp_growth
## ------------------------------------------------
## real_gdp_growth_lag           0.276***
##                                (0.097)
##
## real_gdp_growth_lag2           0.241**
##                                (0.098)
##
## Constant                      1.639***
##                                (0.420)
##
## ------------------------------------------------
## Observations                     102
## R2                              0.187
## Adjusted R2                     0.170
## Residual Std. Error        2.108 (df = 99)
## F Statistic           11.352*** (df = 2; 99)
## ================================================
## Note:                *p<0.1; **p<0.05; ***p<0.01
```

Results:

Model Evaluation:
In order to determine which model is optimal between the ADL and AR models, Akaike Information Criteria and Bayes Information Criteria values are calculated. Below is the model with the lowest AIC value.

```
## real_gdp_growth ~ T10Y3M_lag2 + T10Y3M_lag3 + T10Y3M_lag4 + real_gdp_growth_lag +
##     real_gdp_growth_lag2 + real_gdp_growth_lag3
## <environment: 0x0000000011eda318>
```

The model with the lowest BIC value is also shown.

```
## real_gdp_growth ~ real_gdp_growth_lag + real_gdp_growth_lag2
## <environment: 0x000000000cf13a30>
```

A table showing the AIC and BIC values is also presented. From this table, the General to Specific ADL model found earlier with lag 2 of T10Y3M and lags 1 and 2 of real GDP growth seems to be optimal. The model with the lowest AIC value has a high BIC, and the model with the lowest BIC has a high AIC, but the General to Specific ADL model found earlier has a AIC close to the lowest AIC and a BIC close to the lowest BIC. This model has the lowest values overall across the information criteria.

Table 1: Model Comparison

Criteria	General.To.Specific.ADL.Model	Optimal.Model.AIC	Optimal.Model.BIC
Adjusted R Squared	0.195	0.224	0.187
AIC	444.427	443.543	446.586
BIC	457.552	464.543	457.086

Optimal Model:
The ADL(2, 1) model real_gdp_growth = alpha_final + beta_final*T10Y3M_lag2 + gamma_final_1*real_gdp_growth_lag + gamma_final_2*real_gdp_growth_lag2 + epsilon_final appears to be the optimal model.

Below is a table showing the test statistic and p value of the RESET test (p=1) and the Jarque-Bera test applied to the optimal model.

Table 2: Model Evaluation

Test	Test.Statistic	p.value
RESET (p=1)	1.981	0.122
Jarque-Bera	0.255	0.120

The additional parameter from the RESET test is not significant, indicating that the assumption that a linear model is correct cannot be rejected. The Jarque-Bera test indicates that the null hypothesis that the distribution of normality of the residuals cannot be rejected. These test results do not signal misspecification of the model.

Performing the Breusch-Godfrey test for serial correlation in the residuals gives the result below. As the p value is above 0.05, the null hypothesis of no serial correlation is not rejected. This test result does not signal any misspecfication of the model.

```
##
##  Breusch-Godfrey test for serial correlation of order up to 2
##
## data:  reg_lag_2_12
## LM test = 5.7033, df = 2, p-value = 0.05775
```

Performing the Breusch-Pagan test for heteroskedasticity gives the result below. As the p value is above 0.05, the null hypothesis of homoskedasticity is not rejected. This test result does not signal any misspecfication of the model.

```
##
##  studentized Breusch-Pagan test
##
## data:  reg_lag_2_12
## BP = 2.9491, df = 3, p-value = 0.3995
```

US Real GDP Growth Quarterly Predictions, Annualized Rate, for Quarters 1 - 45:
Evaluation Sample: 2008-07-01 to 2019-07-01:

Plot of Predicted vs. Actual Growth Rates July 2008 - July 2019:

Figure 6: ADL(2, 1) Model Predicted vs. Actual Real GDP Growth Rates

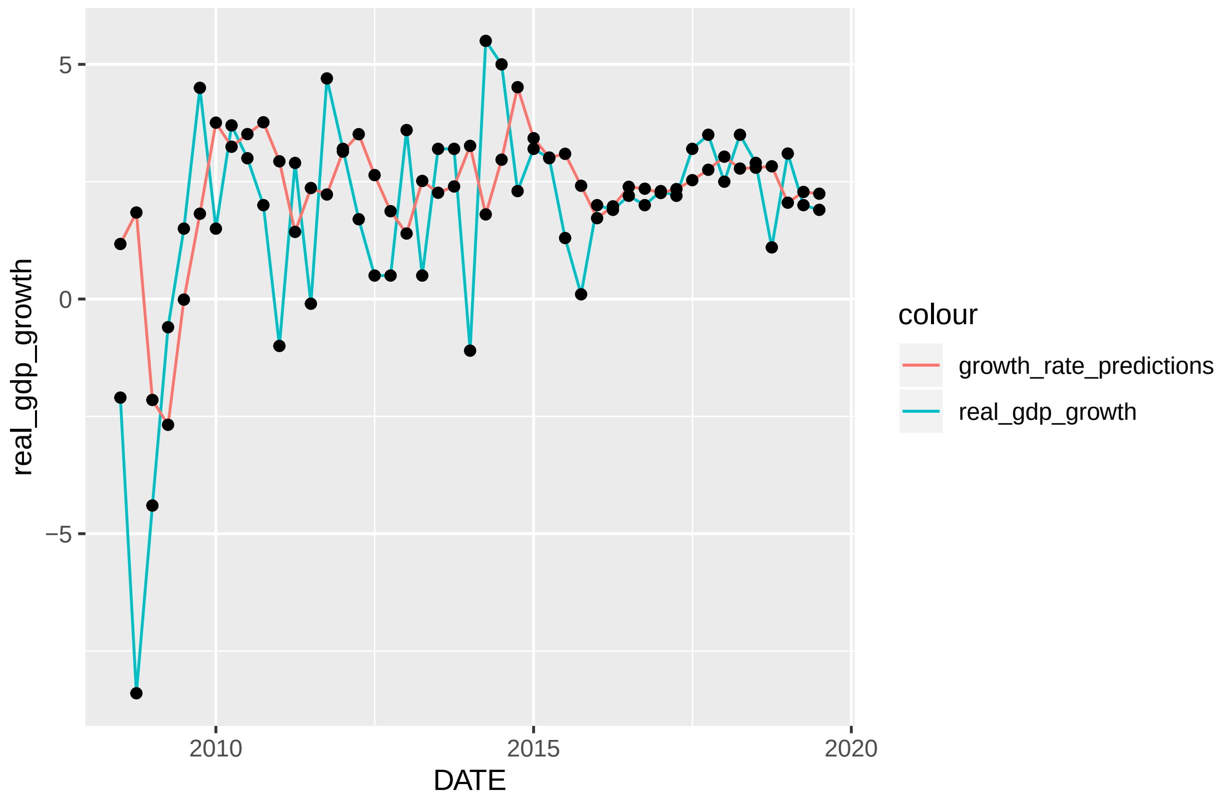
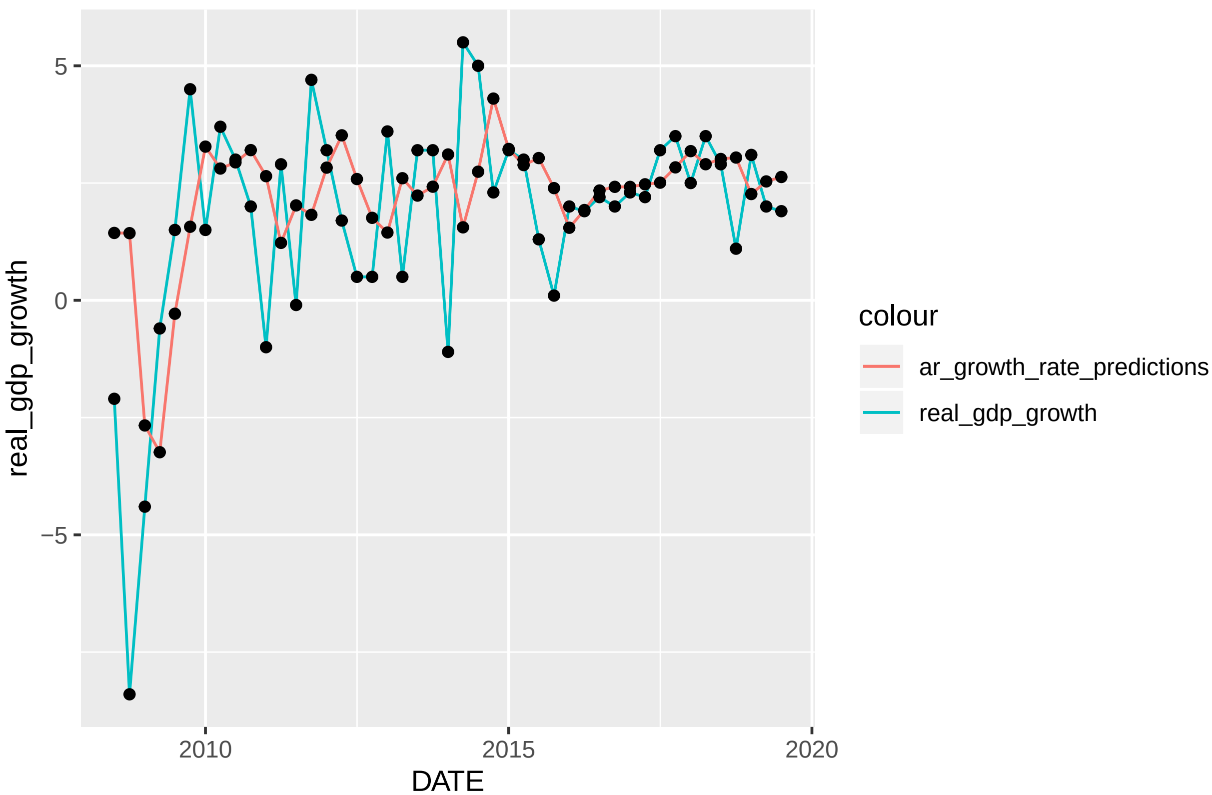


Figure 7: AR(2) Model Predicted vs. Actual Real GDP Growth



Conclusion

The training data indicates that T10Y3M does give predictive value above past values of real GDP growth itself. This was not seen in the evaluation sample, however. The RMSE and MAE of an AR(2) model were 2.33 and 1.62, respectively. By comparison, the RMSE and MAE of the optimal ADL model were 2.36 and 1.61. These are essentially equal, the RMSE indicating a forecasting error of about 2.4% on an annualized quarterly growth basis, and the MAE indicating a forecasting error of about 1.6%. These are high, so the models using 1983-2008 data do not accurately predict real GDP growth from July 2008 - July 2019. The models predict a recession in late 2008, after the recession was already under way. The models do not fully capture the variability in real GDP growth, but are more accurate in times of more stable economic growth.

One limitation of this research was the relatively small sample size, as data was only readily available from 1982. This research also does not use model averaging techniques to reduce variations in model estimates. Future work includes addressing these limitations as well as exploring more than just interest rates in order to better predict real GDP growth.

References:

Ben Hamner, and Michael Frasco. Metrics: Evaluation Metrics for Machine Learning (version 0.1.4), 2018. https://CRAN.R-project.org/package=Metrics.

Dijk, Dick van, Philip Hans Franses, and Christiaan Heij. “Lecture 6.3 on Time Series: Specification and Estimation.” n.d.

Estrella, Arturo, and Gikas A. Hardouvelis. “The Term Structure as a Predictor of Real Economic Activity.” The Journal of Finance 46, no. 2 (1991): 555–76. https://doi.org/10.2307/2328836.

Grolemund, Garrett, and Hadley Wickham. “Dates and Times Made Easy with Lubridate.” Journal of Open Source Software 40, no. 3 (2011): 1–25.

Haubrich, Joseph G., and Ann M. Dombrowski. "Predicting Real Growth Using the Yield Curve." Federal Reserve Bank of Cleveland Economic Review 32, no. 1 (Q1 1996): 26–35.

Heiss, Florian. Using R for Introductory Econometrics, 2016. http://www.URfIE.net.

Hlavac, Marek. Stargazer: Well-Formatted Regression and Summary Statistics Tables (version 5.2.1), 2018. https://CRAN.R-project.org/package=stargazer.

Komsta, Lukasz, and Frederick Novometsky. Moments: Moments, Cumulants, Skewness, Kurtosis and Related Tests (version 0.14), 2015. https://CRAN.R-project.org/package=moments.

Laurent, Robert G. "An Interest Rate-Based Indicator of Monetary Policy." Federal Reserve Bank of Chicago Economic Perspectives 12, no. 1 (January 1988).

Ng, Serena, and Pierre Perron. "Lag Length Selection and the Construction of Unit Root Tests with Good Size and Power." Econometrica 69, no. 6 (2001): 1519–54.

R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria, 2019. https://www.R-project.org/.

Robinson, David, and Alex Hayes. Broom: Convert Statistical Analysis Objects into Tidy Tibbles. (version 0.5.2), 2019. https://CRAN.R-project.org/package=broom.

Venables, W. N., and B. D. Ripley. Modern Applied Statistics with S. Fourth Edition. New York: Springer, 2002.

Wickham et al. "Welcome to the Tidyverse." Journal of Open Source Software 4, no. 43 (2019): 1686.

Wickham, Hadley, Jim Hester, and Romain Francois. Readr: Read Rectangular Text Data (version 1.3.1), 2018. https://CRAN.R-project.org/package=readr.

Wooldridge, Jeffrey M. Introductory Econometrics A Modern Approach. 4th ed. CENGAGE Learning, 2009.

Zeileis, A. Dynlm: Dynamic Linear Regression (version 0.3-6), 2019. https://CRAN.R-project.org/package=dynlm.

Zeileis, Achim, and Torsten Hothorn. "Diagnostic Checking in Regression Relationships." R News, 2002.