Forecasting the 10-Year US Treasury:

Machine Learning Methods vs. Market Expectations

Riley Conlon Shukrit Guha Kenneth Rios

[rc3608@nyu.edu](mailto:rc3608@nyu.edu) [sg5297@nyu.edu](mailto:sg5297@nyu.edu) [kr2152@nyu.edu](mailto:kr2152@nyu.edu)

**Business Understanding**

This paper will answer the Question:

**Can Machine Learning methods forecast the 10-year US Treasury more accurately than the market’s expectation?**

**We hypothesize the answer to be Yes.** Machine learning methods are recognized for their superior forecasting capabilities relative to traditional econometric techniques. We demonstrate the degree to which this is true in our research context by analyzing how well machine learning models forecast the 10-year Treasury relative to our basis.

The Value of this research ties to the importance and ubiquity of the 10-year Treasury in finance and financial econometrics. The 10-year Treasury note is the most widely tracked and studied government debt instrument in finance. Its corresponding yield curve is often used as a benchmark for other interest rates. Examples of such rates include mortgage rates, loan rates, and corporate bonds. The 10-year Treasury is also used as a barometer for the general state of the economy, with an inversion of the yield curve suggestive of an economic downturn in the future.

The Value our research purports to add is twofold. We aim to:

* use machine learning techniques to build a predictive model with superior forecasting performance relative to a basis, and to
* apply machine learning techniques not traditionally used in the literature to gain insights into the major determinants of the 10-year Treasury.

The basis, or baseline method for comparison, we use to forecast the 10-year Treasury is the market’s expectation of the rate itself, known as the *forward rate*. The forward rate is the market’s expectation of the future rate, also known as the future yield. Assuming efficient markets and a minimal-arbitrage condition, the 10-year Treasury yield can theoretically be constructed using the forward rates. The natural shortcoming of this theoretical framework is that forward rates are not perfect predictors of future rates. The spread between the predicted and realized rates represents the inherent risk in investing in the 10-year Treasury.

Forward rates for the 10-year Treasury are available several “years hence”, which signifies how far ahead the market’s forecast is made. Examples include the 10-year forward rate 1 year hence, 3 years hence, 5 years hence, etc. Under the efficient markets hypothesis, these forward rates can be backed out using the yield curve and spot rates for various Treasury maturities. Thus, the forward rate is the natural baseline for predicting the 10-year Treasury as it encapsulates what the market believes the future rate will be using all available information at the time the forecast is made.

Traditional methods for forecasting interest rates focus on the merits of standard econometric models, often linear in structure. These models, however, emphasize inferencing rather than forecasting. The Value Add of our analysis is to improve upon these models by exploiting the flexibility of particular machine learning methods in order to improve forecasting. This will aid financial market participants in preempting interest rate movements. In addition, the models we use reveal important relationships between the 10-year Treasury and several key macroeconomic indicators. These findings are of interest to applied economists seeking to understand which variables are most important to the determination of the 10-year Treasury.

We use Regression Tree methods and LASSO regression as our machine learning alternatives. Both of these machine learning methods promise to improve on forecasting accuracy relative to traditional econometric techniques. In the Model Understanding section, we describe the three specific types of Regression Tree models we choose to use and expand on the LASSO in more detail.

In order to gauge how well these models perform in forecasting the 10-year Treasury using macroeconomic data, we must decide on a Metric. We decide to use the root-mean-square error (RMSE) as our metric of forecasting accuracy. Both our machine learning models and our market baseline will be evaluated using this criterion. Models that give rise to RMSEs smaller than those from our forward rate baseline offer use cases where our alternative model performs better than the market’s expectation.

In addition to evaluating RMSE, we will observe how a *visual examination* of the forecast curves generated by our models may be used to preempt the directionality of the 10-year Treasury. Models which track short-term movements in the 10-year Treasury well provide another Value Add to market participants looking to profit based on whether the market will move up or down.

**Data Understanding**

The variables we decide to include in our forecasting models are summarized in the Variable List table on the following page. We choose the 18 independent variables listed since all of these variables, aside from TIGHT, FED\_FUNDS and CTI, are derived from indicator series that the Federal Reserve uses in their determination of monetary policy. Changes in monetary policy then influence movements in nominal interest rates. The 10-year Treasury is a crucial interest rate that is sufficiently long-run in maturity to take into account Fed policy. Thus, incorporating variables from indicator series used by the Fed in their determination of monetary policy represents a logical starting point for our own analysis of the 10-year Treasury.

**Data Understanding: Variable List**

|  |  |  |
| --- | --- | --- |
| **Variable Name** | **Type** | **Description** |
| UST10Y | Dependent Variable | 10-year Constant Maturity US Treasury |
| SPREAD | Independent Variable | 10-year US Treasury - Effective Fed Funds rate |
| FED\_FUNDS | Independent Variable | Fed Funds Rate |
| PCE | Independent Variable | Personal Consumption Expenditures excluding Food & Energy |
| TCU | Independent Variable | Total Industry Capacity Utilization |
| UR | Independent Variable | Civilian Unemployment Rate |
| M1 | Independent Variable | M1 US Money Stock |
| M3 | Independent Variable | M3 US Money Stock |
| INCOME | Independent Variable | Real Personal Income excluding current Transfer Receipts |
| CTI | Independent Variable | Percent Change in Total Outstanding Credit (owned and securitized) to INCOME |
| FISC\_BAL | Independent Variable | Federal Surplus/Deficit |
| MTS | Independent Variable | Real Manufacturing and Trade Industries Sales |
| INDPRO | Independent Variable | Industrial Production Index |
| LOANS | Independent Variable | Commercial and Industrial Loans |
| PRIME\_RATE | Independent Variable | Bank Prime Loan Rate |
| PERMITS | Independent Variable | Total New Privately Owned Housing Starts |
| CONS\_SENT | Independent Variable | Consumer Sentiment Index |
| SP500 | Independent Variable | S&P500 Index |
| TIGHT\* | Independent Variable | Dummy for tightening and loosening regimes |

\* Only included in specifications fitted within Moderation/ZIRP regimes

All variables in the Variable List table except for SP500 are generated using data we pull from FRED. We obtain S&P500 financial data from historical Yahoo Finance quotes. All variables represent a time series within a time horizon spanning from July 1990 to Dec 2015. The indicator series are split between the Leading, Coincident, and Lagging Indicator Indices used by the Fed in their determination of monetary policy. Although originally compiled and tracked by the US Department of Commerce, in 1995 the non-profit Conference Board assumed the role of reporting these indices. We pull the names and descriptions of the indicator series in these indices from Conference Board.

The Leading Index is meant to be an aggregation of variables that represent the current level of investment in the economy. As a result, changes in this Index tend to lead changes in the economy at large. The variables from this index that we include are SP500, CONS\_SENT, PERMITS, and SPREAD.

The Coincident Index is meant to be an aggregation of variables related to total income, or output. As a result, this Index provides a view of the current state of the economy. The variables from this index that we include are INDPRO, MTS, FISC\_BAL and INCOME.

The Lagging Index is meant to be an aggregation of variables that represent the level of debt in the economy. Changes in this Index tend to follow changes in the rest of the economy. The variables from this index that we include are PRIME\_RATE, LOANS, PCE, UR, and CTI. To capture inflationary pressures, we also include money supply measures M1 and M3.

We select all of the independent variables in the Variable List table for their prevalence in macroeconomic analyses, aside from their ubiquity in Federal Reserve communiques. We include the SPREAD and FED\_FUNDS variables since the relevant literature points to these two variables being significant predictors of the 10-year Treasury. We include the growth in the credit-to-income ratio (CTI) since this variable is a widely used and robust indicator of banking and currency crises in the country risk literature. For example, Comelli (2013) finds the ratio of private credit to GDP to be a significant and robust Early Warning Indicator for currency crises. We include SP500 in order to control for influences on the 10-year Treasury resulting from bear or bull equity markets.

The consumer sentiment index, CONS\_SENT, is compiled by University of Michigan. It is based on telephone surveys which gather information on consumer expectations about the aggregate economy. Since consumer spending accounts for roughly 70% of GDP, the UMich Consumer Sentiment Index is regarded as one of the many important economic indicators followed by businesses, policymakers, and investors.

TIGHT is a dummy variable we use to track tightening and loosening by the Federal Reserve as measured by movements in the target Fed Funds rate. We label periods during which the Fed continuously raised the Fed target rate as tightening regimes. We label periods during which the Fed continuously lowered the Fed target rate as loosening regimes. Tightening regimes receive a value of 1 in the data while loosening regimes receive a value of 0. We include this dummy in order to observe whether there is a direct effect of tightening and loosening within particular time periods (Moderation and ZIRP regimes). We later account for tightening and loosening by constructing data blocks corresponding to tightening and loosening regimes and then running our models separately on those blocks.

We lag all independent variables by one year, representing year-ahead forecasting from the practitioner's perspective. We roll forward our basis, the 3 year hence 10-year forward rate, by three years in order to compare the market’s expectation with the realized rate. In the Business Understanding, we mention that there are several candidate forward rates that could be used as the potential basis, such as the 1 year hence and 5 years hence forward rates.

In an earlier presentation, we discussed our research that sought to discern the best-performing forward rate over our entire time horizon of interest, according to RMSE. In it, we concluded that the 3 years hence 10-year forward rate performed the best among all commonly used forward rates in forecasting the 10-year US Treasury. Thus, for the rest of our analysis, we take the 3 years hence 10-year forward rate to be our baseline.

The time series data we initially obtain is recorded at the monthly level. We go from a collection of monthly time series to a collection of weekly time series via linear interpolation. We perform interpolation in order to increase our sample sizes which will lead to an improvement in the performance of our models. We try various other forms of interpolation, including moving-average interpolation and structural Kalman filtering, but find linear interpolation to work the best in terms of RMSE. We also find linear interpolation to be the simplest algorithm in our research context and most robust to changing specifications.

**Regimes**

We now discuss the two pairs of regimes we use to condition our training and test sets for estimation of our machine learning models. Incorporating regime switching is an important aspect of our statistical analysis. We find it crucial to condition the data for regime switching by partitioning it into distinct regimes for estimation. Doing so takes into account an evolving data generating process responsible for determining the 10-year Treasury over time. We can focus on particular policies within regimes and perform a more appropriate statistical comparison between our training and test sets. In our analysis, we employ two distinct pairs of regimes that look at two different facets of potential rate determination.

Our first pair of regimes corresponds to periods of differing monetary policy employed by the Federal Reserve. Gavin (2018), published by the Federal Reserve Bank of St. Louis, discusses determination of monetary policy. The paper provides a timetable of the various monetary policy regimes employed by the Federal Reserve from 1965 to 2015. These years span over our own time horizon of interest, which is from July 1990 to Dec 2015.

Gavin (2018) discusses the nature of these regimes and the justification for treating each time period separately from a policy standpoint. The Great Moderation was characterized by a period of stable economic output and inflation. From Oct 1982 to Dec 2008, the Fed rebuked money supply targeting in favor of indirect interest rate targeting. This was done in order to maintain the Fed’s credibility of low inflation that was achieved in the prior regime. Traditional reduced-form approximations of central bank policy, such as various types of Taylor-like rules, have been shown to model nominal interest rates well during this regime. All our time series data from Jan 1991 to Dec 2008 falls into the Great Moderation regime.

The Zero Interest Rate Policy regime was a period of roughly seven years from Dec 2008 to Dec 2015, characterized by the Fed’s policy of maintaining a low target Fed Funds rate. In this regime, the target Fed Funds rate oscillated between a peg of 0 and 0.25 basis points. The Fed abruptly undertook this expansionary monetary policy in response to the financial crisis in late 2008. During this regime, the Fed Funds target was kept at or near 0 as economic growth was low relative to the long run trend.

In addition to maintaining low rates, the Fed also engaged in Quantitative Easing in order to maintain low rates in money markets. Thus, due to the stable nature of the target rate, we expect the Fed Funds rate (FED\_FUNDS) and the spread itself (SPREAD) to be the most significant predictors of the 10-year Treasury during this time period. All our time series data from Jan 2009 to Dec 2015 falls into the Zero Interest Rate Policy (ZIRP) regime.

Our second pair of regimes, or blocks, consist of periods of tightening and loosening regimes. We determine regime boundaries according to stretches of continuously increasing or decreasing target Fed Funds rate set by the Federal Reserve. During tightening regimes, the Fed raises the rate in order to make borrowing more expensive and combat inflationary pressures, among other considerations. During loosening regimes, the Fed lowers the rate in order to make borrowing cheaper and stimulate the economy.

By partitioning the data into tightening and loosening regimes, we control for regime change and Fed policy objectives insofar as they influence the 10-year Treasury. Since tightening and loosening regimes are interspaced between each other, we pool regimes of each type into blocks. We then use these blocks of tightening and loosening regimes to generate our training and test sets. During our time horizon of interest from July 1990 to Dec 2015, tightening regimes occur from Feb 1994 to Jan 2001 and from June 2004 to Sep 2007. During the same time period, loosening regimes occur from July 1990 to Feb 1994, from Jan 2001 to June 2004, and from Sep 2007 to Dec 2015.

Due to the economic nature of these regimes and the behavior and concerns of the Fed during the corresponding time periods, we thus partition data for our entire time horizon into

* Great Moderation/ZIRP regimes, and
* Tight/Loose blocks.

We run our machine learning models separately on both pairs of regimes. Doing so will control for the various policies that are likely to influence the 10-year US Treasury over time and thus improve the performance of our forecasting models. We report RMSE estimates for each of these four regimes in the Model Performance section.

In addition, our machine learning models will also allow us to report the most important predictors of the 10-year US Treasury within each regime. Our Regression Tree methods report *Variable Importance plots* which rank the importance of the various independent variables according to their percentage reduction in RMSE. LASSO regression reports traditional estimated coefficients on those predictors selected by the model as being significant. We rank the estimated coefficients of the predictors according to their absolute value. The marginal effect of each predictor on the 10-year Treasury can then be interpreted.

**Sampling**

Due to the time series nature of our research and the temporal aspect of the regimes that partition our data, the method of sampling and model validation is critical to obtaining reliable estimates. In typical econometric analyses, establishing inference rather than generating forecasts is of interest to the econometrician. In such analyses, the entirety of the data available is pooled and then used to run various regression specifications. Within time series research, there is great variability in the types of regression models used. For example, Hamilton and Jorda (2001) refer to numerous papers which use various linear Vector Autoregression models to determine the directionality of interest rates.

Our research is not primarily concerned with inferencing, but rather with forecasting. In particular, part of our Value Add is to use models not traditionally used in the literature to improve forecasting performance of the 10-year Treasury. We draw upon machine learning methods that purport to improve forecasting accuracy relative to traditional econometric models.

In order to test the accuracy of these models, machine learning analyses typically divide the data into two groups. One group, the training set, is used to “fit” the model and provide estimates of the parameters of interest. The second group, the test set, is used to gauge how well the fitted model performs out-of-sample. This out-of-sample measurement of predictive accuracy is critical. A model which both trains and tests within the same sample is likely to suffer from overfitting. In this case, estimates of the model’s accuracy is likely to be unreliable as the ability of the fitted model to capture variations in new, unseen data is not well defined.

The construction of appropriate training and tests sets is an integral part of the machine learning analysis. In cross-sectional analysis, observations in the master data set are typically assigned randomly to either the training set or the test set according to some split ratio. The sampling procedure can be different in time series analyses. Since all our independent variables are lagged predictors, care must be taken in deciding the appropriate validation approach.

In our research context, assigning time series observations randomly to training and test sets would be an inappropriate approach. This is because of a well-established time series phenomenon known as “look-ahead bias”. If a random sampling approach is taken, then data corresponding to future rates would be used to train on current-period lags in the training set. Using future outcomes to train on current-period data is inappropriate in our time series estimation. The forecaster using such models, when deciding to generate a forecast of next-period rates, is not privy to outcomes in the future. Since our time series indicators are not time-independent and future values are dependent on current values, the distinction between current data and future data must be respected in the training and test sets.

We circumvent look-ahead bias by instead choosing a cutoff year within each regime that demarcates the boundary between training and test set. This approach is known as *year-forward validation*. To establish validation consistency across all regimes, we aim for an approximate 70/30 split of the data. In particular, we aim to include roughly 70% of the available data in each regime in our training set and place the remaining 30% in our test set. We experiment with various split ratios and find a 70/30 split ratio to be the best-performing split in terms of RMSE.

Splitting the data within Great Moderation and ZIRP into training and test sets proves to be straightforward as these regimes are contiguous, that is, there are no time gaps. A 70/30 split implies a cutoff year of 2003 for Great Moderation. Thus, the training set for Great Moderation consists of data from Jan 1991 to Dec 2003. The test set for Great Moderation consists of data from Jan 2004 to Dec 2008. A 70/30 split implies a cutoff year of 2013 for ZIRP. Thus, the training set for ZIRP consists of data from Jan 2009 to Dec 2013. The test set for ZIRP consists of data from Jan 2014 to Dec 2015.

The data within the Tight and Loose blocks, however, are not contiguous since tightening and loosening regimes follow each other over time and introduce gaps in the timeline. As a result, we take a subset of each block which allows us to have the most contiguous possible stretch of data, conditional on having a minimum of *five years* of training data. Maintaining a 70/30 split for training and test sets implies having roughly two years of data in the test set.

Of the years that span the Tight block, we choose a training set spanning from Jan 1997 to Jan 2001 and from June 2004 to June 2005 and a test set spanning from June 2005 to Sep 2007. Of the years that span the Loose block, we choose a training set spanning from Dec 2008 to Dec 2013 and a test set spanning from Dec 2013 to Dec 2015, which happen to be contiguous. We find that these training and test sets we construct are the ones which best reflect the considerations previously mentioned.

The table on the following page summarizes the years spanned for the various regimes. For Tight and Loose blocks, we include data strictly post-July 1990. We include the entire time span for Great Moderation and ZIRP mentioned in Gavin (2018) for sake of completeness.

**Data Understanding: Regime Switching**

|  |  |  |  |
| --- | --- | --- | --- |
| **Regime Name** | **Years Spanned** | **Training Set** | **Test Set** |
| Great Moderation | Oct 1982 - Dec 2008 | Jan 1991 - Dec 2003 | Jan 2004 - Dec 2008 |
| Zero Interest Rate Policy | Jan 2009 - Dec 2015 | Jan 2009 - Dec 2013 | Jan 2014 - Dec 2015 |
| Tight Block | Feb 1994 - Jan 2001, June 2004 - Sep 2007 | Jan 1997 - Jan 2001, Jun 2004 - Jun 2005 | Jun 2005 - Sep 2007 |
| Loose Block | July 1990 - Feb 1994, Jan 2001 - Jun 2004, Sep 2007 - Dec 2015 | Dec 2008 - Dec 2013 | Dec 2013 - Dec 2015 |

**Model Understanding**

The two distinct types of machine learning methods we use are:

1. LASSO (**L**east **A**bsolute **S**hrinkage **S**election **O**perator), and
2. Regression Trees methods

Both types are documented to achieve better forecasting accuracy relative to traditional econometric techniques. These machine learning methods also provide further insight into the independent variables that contribute most to accurately forecasting the 10-year Treasury.

**LASSO**

Equation (1) below describes an ordinary least squares regression model where *yt* is the value of the 10-year Treasury observed at time *t*, *xtj*is the value of the jth independent variable observed at time t, and is the disturbance term that follows a normal distribution with zero mean and unit variance.

(1)

(2)

Expression (2) is the Lagrangian which characterizes the LASSO’s constrained optimization problem. The difference between ordinary least squares regression and LASSO regression is the inclusion of the penalty term . This term represents the penalty placed on the absolute value of the coefficients. It allows the LASSO to shrink coefficients on variables that do not contribute significantly to forecasting all the way to zero. By doing so, LASSO selects and estimates marginal effects only for variables found to be significant for forecasting. is known as the *shrinkage parameter* and controls the extent to which the magnitude of coefficients is penalized. The optimal level of may be estimated using cross-validation.

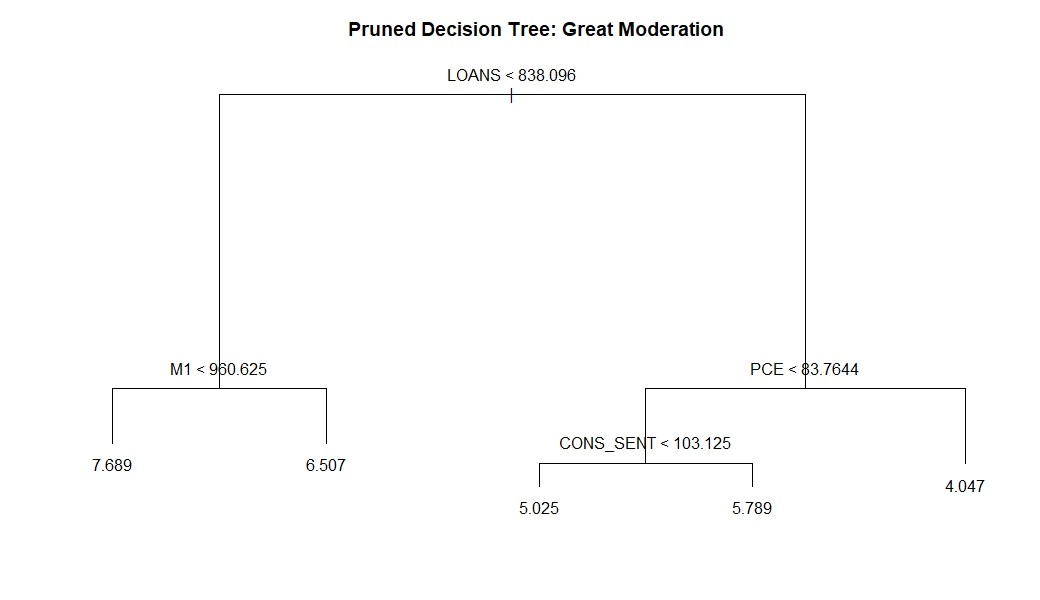
The primary benefits of the LASSO are its

* Inherent variable selection capabilities, and
* ability to model linear relationships with a high degree of interpretability.

We cross-validate for the optimal value of over a grid of possible values. The best is chosen by estimating the RMSE of the model’s forecasts for each in our grid via a 10-fold cross-validation of the corresponding training set.

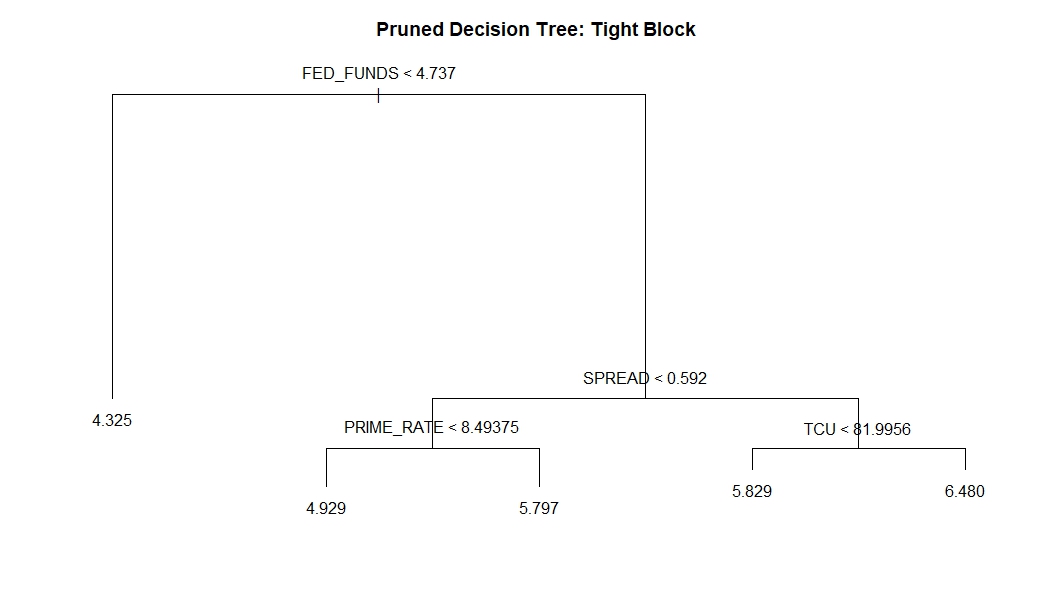
**Regression Tree Methods**

The Regression tree models we use in our analysis are ensembles of smaller decision trees fit on the training sets for each regime. These decision trees partition the data into mutually exclusive blocks using binary splits of the predictors. The average value realized by the dependent variable for each of these blocks is the forecast made by the final model when the predictors’ values fall within the partitions. Individual decision trees, however, cannot be used to estimate forecasting performance. These decision trees instead form the building blocks of the more sophisticated Regression Tree models that we use in our analysis. A visual representation of such a tree, after pruning[[1]](#footnote-1), is shown on the following page.



This tree is fit on the training set for the Great Moderation regime. The length of the vertical lines is proportional to the *Relative Importance* of the variables from which they stem from. The Relative Importance of a variable is determined by the extent to which forecasting accuracy is reduced when it is excluded it from the model. In the example above, the simple decision tree reveals that LOANS is the variable with the greatest predictive power, followed by M1, PCE and finally CONS\_SENT.

The decision tree first checks whether LOANS is less than or greater than 838.096. If it is less than 838.906, then the tree looks at M1. If it is greater than 838.096, then it looks at PCE to evaluate the next split. In the case that LOANS < 838.096, the model checks M1. If M1 < 960.625, the tree forecasts the 10-year Treasury to be 7.689. If M1 > 960.625 instead, it forecasts the 10-year Treasury to be 6.507. The decision tree continues to evaluate splits at different branches depending on the input values it is fed. It outputs a forecast when it arrives at a terminal node after the conditions in the preceding nodes are satisfied.



The decision tree above was similarly fit (and pruned) on the training set for the Tight Block. Here we see that the most important variable is FED\_FUNDS, followed by SPREAD, PRIME\_RATE and TCU. If the FED\_FUNDS rate is less than 4.737, the tree requires no further information and forecasts the 10-year Treasury to be 4.325. However, if FED\_FUNDS is greater than 4.737, it evaluates the next split by looking at the value of SPREAD. Depending on the value of SPREAD, the tree then evaluates either the PRIME\_RATE or TCU. In total, five possible realizations of the 10-year Treasury may be forecasted. These forecasts correspond to the values present in each of the five terminal nodes.

We use three specific Regression Tree methods that build upon the structure of decision trees as their fundamental unit of analysis. These methods may be characterized as *Regression Tree Ensembles*. They are:

1. *Bagged Trees* - An ensemble of deep regression trees that are fit on bootstrapped training samples and then averaged together to generate the final predictive model.
2. *Random Forest* - In Bagged Trees, many of the trees may be highly correlated due to a concentration of strong predictors among the trees. Averaging these highly correlated trees does not lead to a substantial reduction in variance. To deal with the correlation issue within Bagged Trees, Random Forests decorrelate the fitted trees by choosing a randomized subset of the predictors within each of the regression trees in the ensemble.
3. *Boosted Trees* - Each regression tree in the ensemble is grown sequentially using information from previously grown trees. Each tree is fit on a modified version of the original training set. Learning in Boosted Trees is slowly controlled via a learning parameter, η. The subsequent tree in the ensemble is then fit to the residuals of the current model. This process is iterated with updated residuals in each stage representing the variation in the data yet to be explained by the model. Generally, statistical learning approaches that *learn slowly* tend to perform well[[2]](#footnote-2).

**Value Add of Models**

If the relationship between the predictors and the dependent variable is well approximated by a linear model, then an approach such as LASSO regression will likely outperform Regression Tree methods that do not exploit this linear structure. If instead there is a highly non-linear and complex relationship between the predictors and the dependent variable, then Regression Tree methods may outperform traditional linear models.

In addition, Regression Tree methods tend to mimic human decision-making better than more traditional approaches. This would be useful, for instance, if we consider the Fed as being able to determine how the 10-year Treasury moves by looking at threshold values of key macroeconomic indicators. The ability of Regression Tree methods to provide well-defined threshold values for key predictors provides an additional layer of interpretation that linear models cannot capture.

LASSO is a form of regularized regression. It is more restrictive than typical ordinary least squares (OLS) regression as it shrinks the estimated coefficients according to By restraining the coefficients of variables that are found to be insignificant, LASSO selects only a subset of the independent variables to generate a predictive model. This process of variable selection reduces the variance of the estimates in exchange for a decrease in model flexibility. LASSO’s linear structure and its ability to generate simple and interpretable results is its Value Add.

Both LASSO and Regression Tree methods perform better than linear regression in the context of forecasting accuracy. The LASSO improves upon linear regression by reducing the variance of its estimator while introducing bias via in order to achieve a lower RMSE. In general, OLS regression fails to forecast as well as the LASSO precisely because the OLS estimator is an unbiased estimator by construction. As a result, the OLS estimator is suboptimal for forecasting due to its large variance.

Regression Tree methods tend to forecast better than traditional econometric techniques since no structure is imposed or assumed on the sampling distribution. Simple splitting rules allow complex interactions between the dependent and independent variables, and between the independent variables themselves, to be estimated. There is once again a tradeoff between bias and variance within Regression Tree methods that is balanced such that an optimal RMSE is reached.

As an example of this tradeoff within a Regression Tree method, consider the manner in which Random Forests attempt to improve upon the performance of Bagged Trees. Bagged Trees is a Regression Tree method which aggregates a large number of deeply fitted trees in order to reduce the variance of individual, deep trees. This method, however, suffers from high forecast variance when there are only a few significant predictors among the independent variables. With few significant predictors being used to fit a large number of trees, the trees themselves become highly correlated.

To resolve this, Random Forest fits deep trees on a randomized subset of the independent variables for each bootstrapped sample. Doing so decorrelates the individual trees as the same predictors no longer appear in each fitted tree. This decorrelation across trees allows for substantial reduction in the variance of the final forecasts when aggregated, thus improving the forecasting accuracy of the model.

The following table lists the baseline and the alternative machine learning models we used, covers their technical specifications, and expands on what each model’s Value Add is to accurately forecasting the 10-year Treasury.

**Model Understanding: Model Value-Adds**

|  |  |  |
| --- | --- | --- |
| **Model** | **Specifications** | **Value Add** |
| Baseline | 3 Years Hence 10-Year Forward Rate, rolled forward 3 years | * Incorporates all available current-period information in the market to forecast future 10-year Treasury rates * Represents market expectations of the 10-year Treasury and serves as our baseline model for comparison |
| Bagged Trees | **1.** De **1.** Deep trees fit on 500 bootstrapped samples  **2.** 18 **2.** 18 variables evaluated at each split | * Fits deep trees that increase the flexibility of the predictive model at the cost of high variance * To contain this high variance, Bagged Trees aggregate and average them to produce a final predictive model |
| Random Forest | **1.** Deep trees fit on 500 bootstrapped samples  **2.** 4 of 18 variables randomly evaluated at each split | * Because Bagged Trees evaluates all 18 variables at each split, several trees share similar structures at the top (determined by the strongest predictors). High correlation across trees increases the variance of forecasts * To decorrelate trees, Random Forest evaluates random subsets of predictors at each split, reducing the variance of forecasts and thereby increasing accuracy |
| Boosted Trees | **1.** Shallow trees (with maximum depth of 10), successively fit on residuals of previous trees  **2.** Learning parameter η set at 0.3 for Moderation/ZIRP and 0.2 Tight/Loose Blocks  **3.** Optimal number of iterations for Moderation and ZIRP estimated at 112 and 64, respectively. 71 and 67 iterations for Tight and Loose, respectively | * Trees are fit sequentially allowing the model to learn slowlyfrom residual information unexplained by previously fit trees * Trees are fit using current residuals at each learning stage * Each new decision tree iteratively updates the residuals * By fitting shallow trees, the model slowly improves the predictive model in areas where it does not perform well |
| LASSO | **1.** Optimal value of estimated to be 0.00104 and 0.00053 for Great Moderation and ZIRP, respectively. 0.00058 and 0.00059 for Tight and Loose Blocks, respectively.  **2.** Includes Interaction and higher-order terms | * Assumes a simple linear relationship between the dependent and independent variables * Will outperform Regression Tree approaches if true relationship is near linear as opposed to non-linear * Reports marginal effects which are straightforward to interpret * Allows for inherent variable selection by driving to zero marginal effects of insignificant predictors of the 10-year Treasury |

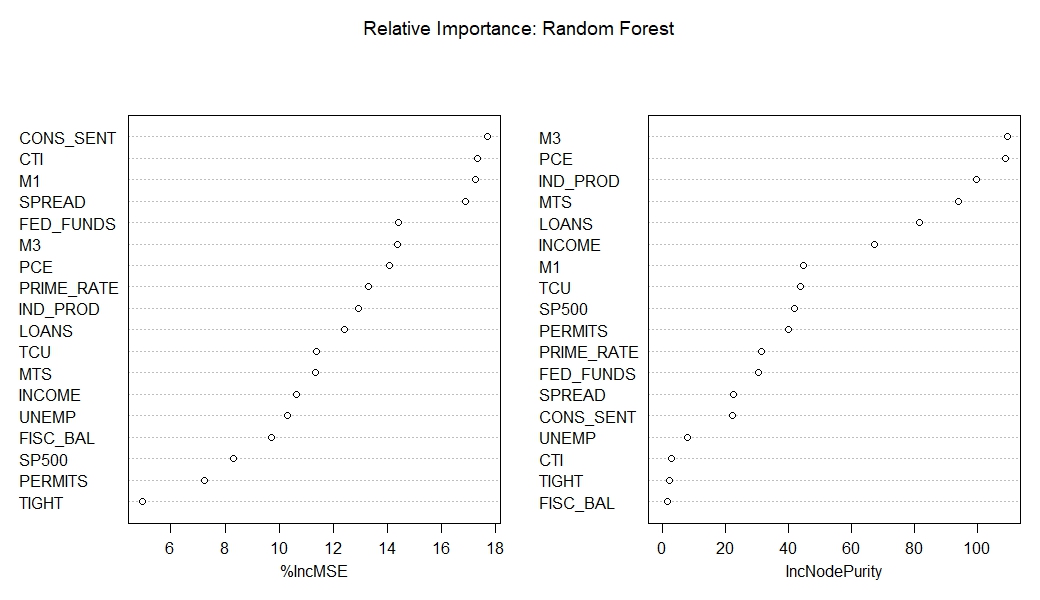
**Variable Relationships**

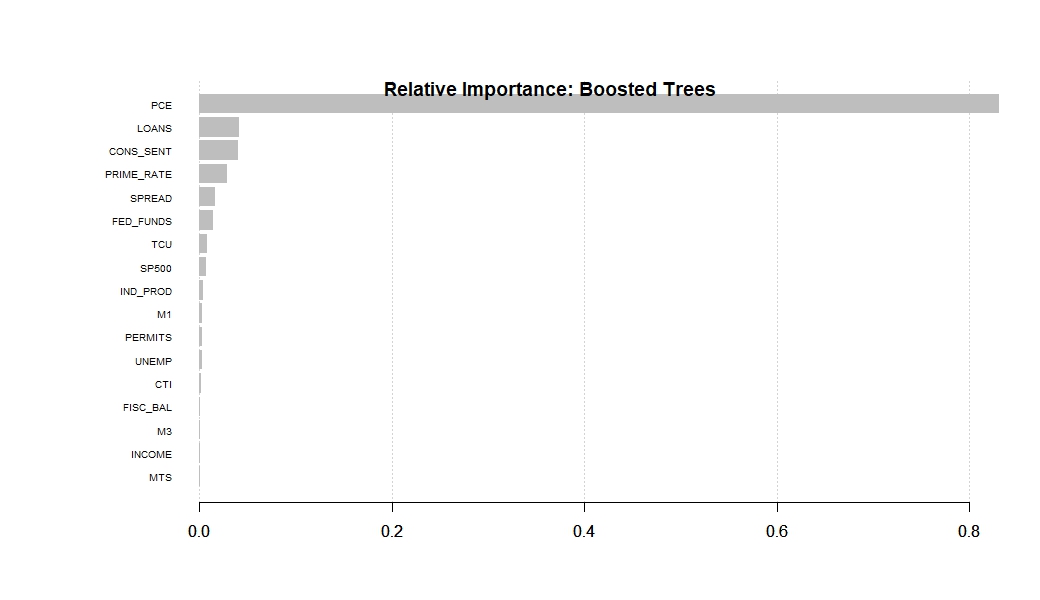
Both types of machine learning methods we use allow for an interpretation of the extent to which our various predictors influence the 10-year Treasury. In particular, for our best-performing Regression Tree and LASSO models within each regime, we report:

* Variable Importance plots generated by the Regression Tree models, and
* a table of estimated coefficients generated from LASSO regression. The variables in these tables are listed in descending order according to the absolute value of their estimated coefficients.

The relationships we observe correspond well with what is observed in the literature, such as the ubiquity and importance of SPREAD and FED\_FUNDS in determining the 10-year Treasury. We find less-established results as well, such as the importance of CONS\_SENT and CTI. Money supply measures M1 and M3 also tend to be an important predictor within most regimes.

**Great Moderation**

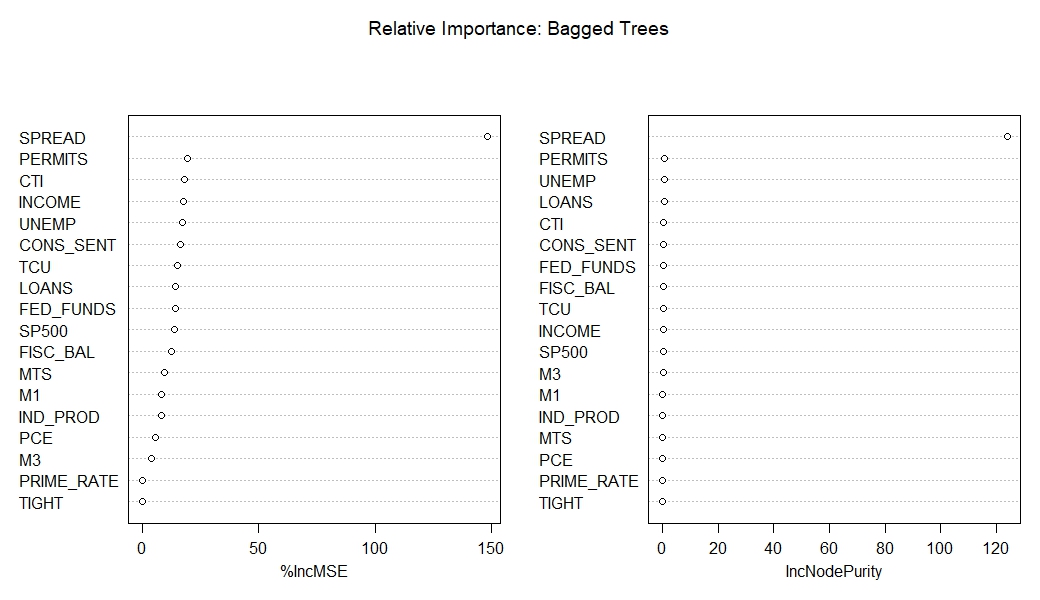
****

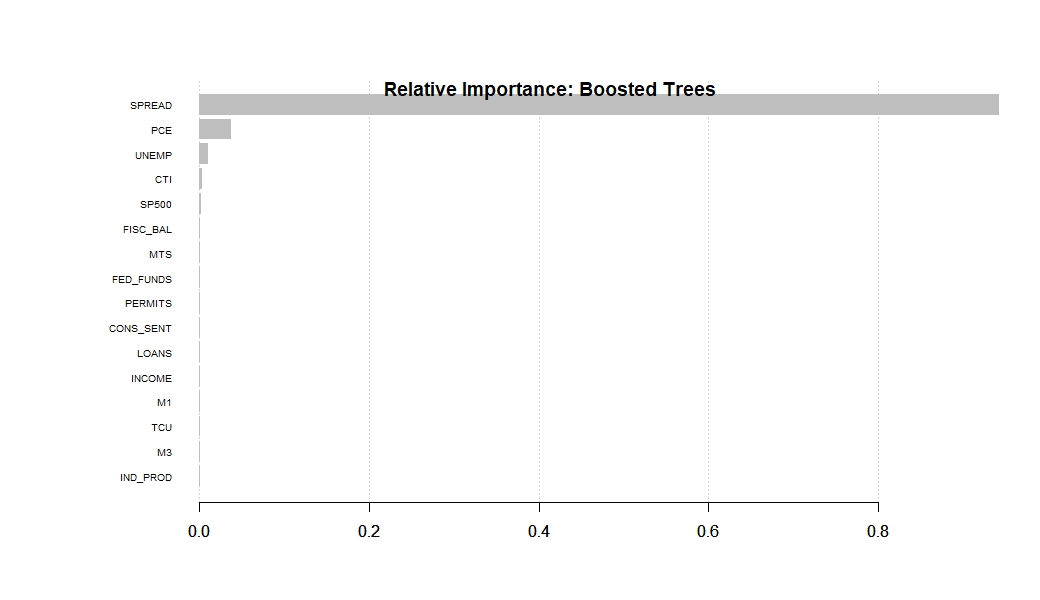
****

**Great Moderation: LASSO - Estimated Coefficients**

|  |  |
| --- | --- |
| **Variables** | **Coefficients** |
| SPREAD | 0.967470441105566 |
| FED\_FUNDS | 0.956894277506111 |
| TIGHT | -0.0588945123382847 |
| PRIME\_RATE | 0.0348769670298527 |
| IND\_PROD | 0.00355895599579273 |
| CONS\_SENT | 0.00215820491099017 |
| PCE | 0.00169572194003616 |
| TCU | 0.00040203479418237 |
| PERMITS | -0.000206656097311963 |
| SP500 | -0.000157710234305611 |
| M1 | -0.0000125751393898055 |
| INCOME | 0.000011069901001783 |
| UNEMP | -5.03832363285714E-06 |
| FISC\_BAL | -9.61314242130718E-08 |
| M3 | 0 |
| CTI | 0 |
| MTS | 0 |
| LOANS | 0 |

**ZIRP**

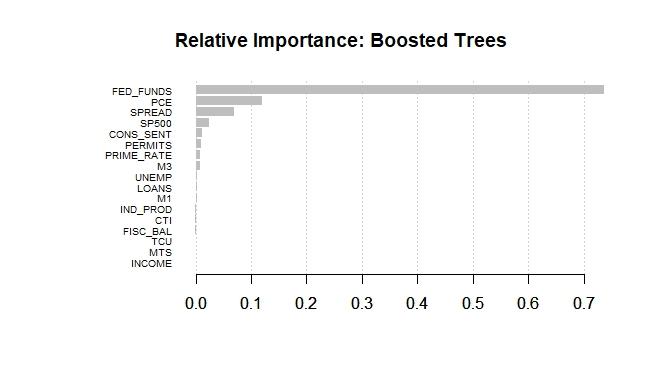
****

****

**ZIRP: LASSO - Estimated Coefficients**

|  |  |
| --- | --- |
| **Variables** | **Coefficients** |
| FED\_FUNDS | 2.01694226078864 |
| SPREAD | 1.08591129098169 |
| PCE | 0.139636623570869 |
| CTI | -0.0241834568647381 |
| IND\_PROD | 0.0191081563423715 |
| UNEMP | 0.0148625839249231 |
| CONS\_SENT | -0.0101840037596552 |
| TCU | 0.00161627841800156 |
| SP500 | -0.000370498048781341 |
| LOANS | -0.000308142004417766 |
| PERMITS | -0.0000449641949950017 |
| M1 | -0.0000198960304382723 |
| INCOME | -0.0000127132541251595 |
| MTS | -2.64655219796377E-06 |
| FISC\_BAL | 1.39742739812127E-07 |
| M3 | 0 |
| PRIME\_RATE | 0 |
| TIGHT | 0 |

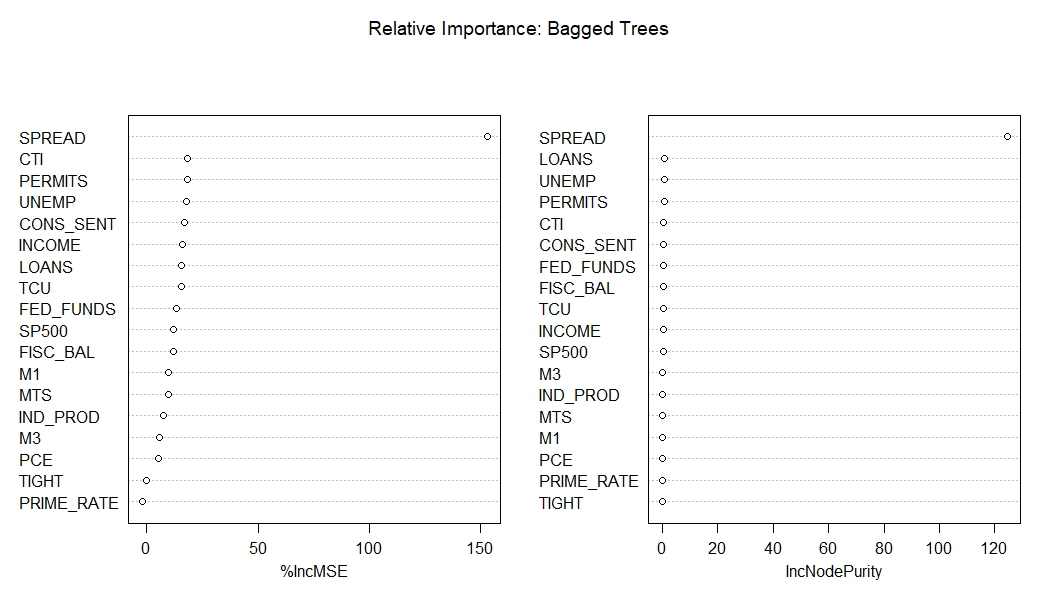
**Tight Block**

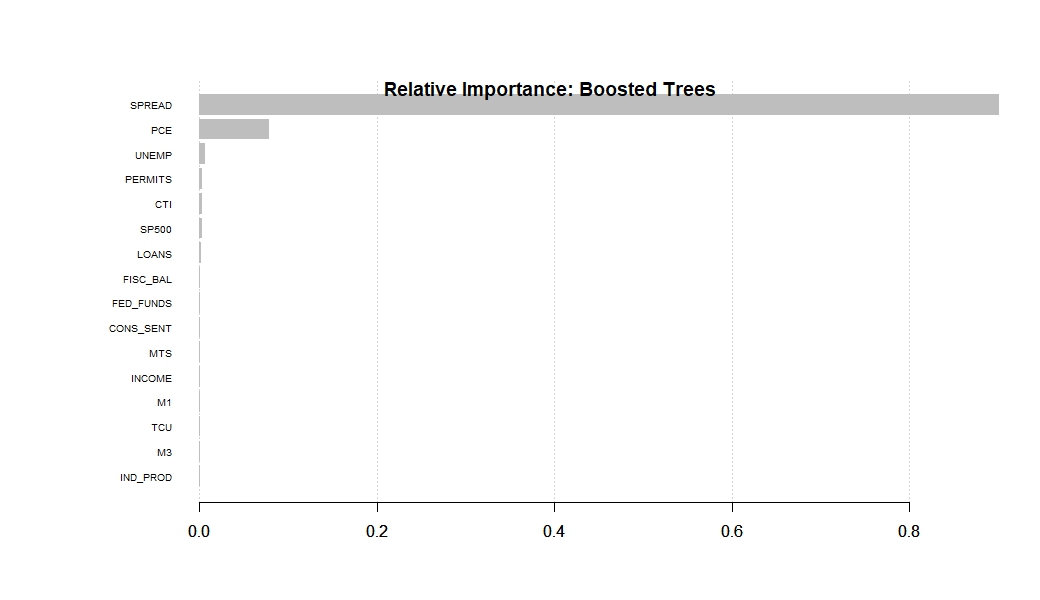


**Tight Block: LASSO - Estimated Coefficients**

|  |  |
| --- | --- |
| **Variables** | **Coefficients** |
| FED\_FUNDS | 1.06424002109514 |
| SPREAD | 1.00240650338564 |
| TCU | -0.0195416233408686 |
| UNEMP | -0.0135009614736049 |
| IND\_PROD | 0.00454905103366568 |
| LOANS | -0.000517268529494715 |
| M1 | -0.000475035857174911 |
| SP500 | -0.000431971116197268 |
| INCOME | 0.000150657723865595 |
| PERMITS | 0.000138632760797947 |
| CONS\_SENT | -0.0000582516540914046 |
| MTS | -1.45866291155197E-07 |
| FISC\_BAL | 1.01454460507888E-07 |
| PCE | 0 |
| M3 | 0 |
| PRIME\_RATE | 0 |

**Loose Block**

****

****

**Loose Block: LASSO - Estimated Coefficients**

|  |  |
| --- | --- |
| **Variables** | **Coefficients** |
| FED\_FUNDS | 2.05062173606407 |
| SPREAD | 1.08474249201339 |
| PCE | 0.148656109917791 |
| CTI | -0.0243064325275002 |
| UNEMP | 0.0155962521428006 |
| IND\_PROD | 0.0124553641155075 |
| CONS\_SENT | -0.0101977113874277 |
| TCU | 0.00876722993243563 |
| PRIME\_RATE | 0.00176080914525598 |
| SP500 | -0.00038529836458179 |
| LOANS | -0.0002840400351838152 |
| M1 | -0.0000334116787380296 |
| INCOME | -0.0000228880805082655 |
| PERMITS | -0.0000193493649094964 |
| MTS | -2.70834761891186E-06 |
| FISC\_BAL | 1.3643832114806E-072 |
| M3 | 0 |

We observe that SPREAD and FED\_FUNDS have the greatest predictive power among the various macroeconomic predictors used in our specifications. This is intuitive as the 10-year Treasury is sensitive to the Fed target rates and often tracks its movement. SPREAD is a measure of the difference between current and future yields. It contains relevant information on future movements of the 10-year Treasury and provides an outlook on the general state of the economy.

Both Regression Trees methods and the LASSO share similar results. FED\_FUNDS and SPREAD consistently have the largest estimated coefficients in absolute value from LASSO regression. In our LASSO specifications, we included interaction terms across all our independent variables and higher-order terms as well. In all our specifications, however, we find that *all* interaction and higher-order terms drop out via variable selection.

In relative terms, although SPREAD and FED\_FUNDS appear more important than the other variables, some variables consistently appear within the top five predictors. These macroeconomic variables include PCE, CTI, UNEMP, CONS\_SENT and IND\_PROD. PCE and UNEMP are measures of activity in the markets for goods, services, and labor. Their consistent inclusion within the top five predictors is expected given that changes in these variables induce changes in the Fed’s policy outlook. These changes in policy affect future movements in nominal interest rates.

The Industrial Production Index is a measure of the level of production within the economy. Since production, investment, and credit are closely linked, we find INDPRO’s strength as a predictor of the 10-year Treasury intuitive. CONS\_SENT is an indicator for the level of consumer confidence in the economy. Since consumption is a significant determinant of output, economic agents rely on this measure to gauge the general state of the economy. If conditions are conducive to investment, there is an increase in capacity to satisfy the growing demand for capital. This increase in capacity requires greater liquidity and access to credit. These indirect channels influence the demand for loanable funds in the economy and thus the market rate for the 10-year Treasury.

Several papers in the country risk literature find the credit to GDP ratio to be a significant and robust Early Warning Indicator of impending banking and currency crises. Large deviation of credit from the long-run trend represent credit booms which may prove to be unsustainable. Consequent busts in credit markets precipitate economic downturns. We note that during the ZIRP regime and the Loose Block, CTI is among the top three predictors according to the Bagged Trees Variable Importance plots. In the Boosted Trees Variable Importance plots, CTI is among the top five predictors within these regimes.

**Model Performance**

The table below displays the RMSEs for our baseline and each of our machine learning models within each of the four regimes we describe in the Data Understanding section.

**Model Performance: Model RMSEs**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Moderation | ZIRP | Tight Block | Loose Block |
| BASELINE | 0.295 | 0.413 | 0.266 | 0.422 |
| LASSO | 0.012\* | 0.059 | 0.043 | 0.059 |
| Bagged Tree | 0.428 | 0.053 | 0.007\* | 0.049\* |
| Random Forest | 0.166 | 0.256 | 0.066 | 0.201 |
| Boosted Tree | 0.267 | 0.007\* | 0.037 | 0.104 |
| \*Lowest RMSE within Regime | | | | |

We note that in all use cases except one (Bagged Trees within the Great Moderation), every machine learning approach outperforms the forward rate baseline. Furthermore, a *visual examination* of the forecast curves will demonstrate the extent to which the LASSO forecasts track the 10-year Treasury, most notably during periods of downturn such as the financial crisis in late 2008. The LASSO tracks the 10-year Treasury the best in all regimes except for Loose, where it falls just behind Bagged Trees in this regard. Thus, we argue that the LASSO provides additional Value Add beyond what alternative models with lower RMSEs within a particular regime may provide.

For traders that are concerned with movements of the rate in a short-run horizon, the LASSO provides a robust, parsimonious model that tracks the 10-year Treasury well. For investors that are concerned with whether the market will move up or down, the LASSO provides that Value Add to those financial market participants. This is most apparent in our forecasts for the Great Moderation regime and the Tight Block. From the point of view of long-term investors, an error measure such as RMSE over a longer-term horizon provides a more useful metric. For this audience, the Value Add from

* LASSO within Great Moderation,
* Bagged Trees within Tight/Loose blocks, and
* Boosted Trees for ZIRP,

lies in the models’ forecasting accuracy as measured by their RMSEs.

Combining our results, the table on the following page summarizes which regimes and use cases the various models we use work best for:

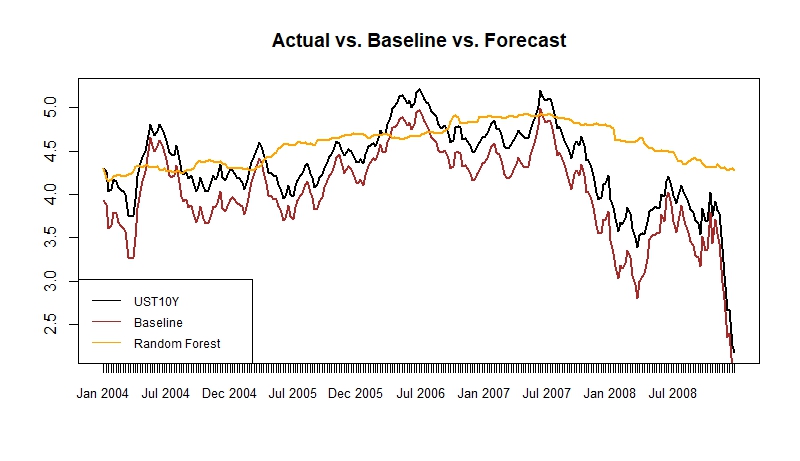
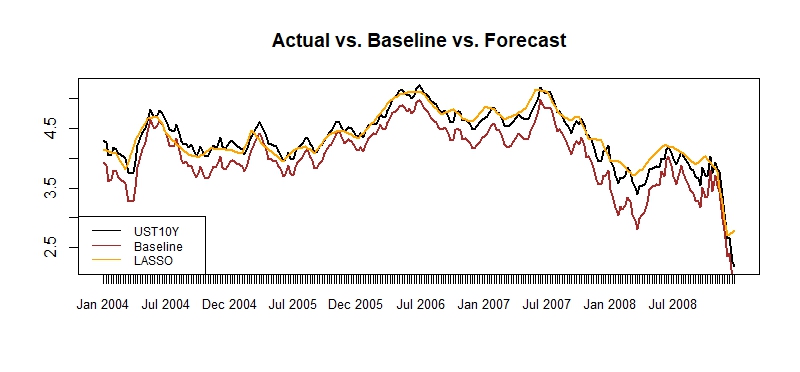
**Model Performance: Use Case Value-Adds**

|  |  |  |
| --- | --- | --- |
| **Model** | **Best Regimes** | **Value Add in Use Case** |
| Bagged Trees | 1. ZIRP  2. Tight Block  3. Loose Block | * Lowest RMSE within Tight/Loose Blocks * Best tracking within Loose Block |
| Random Forest | 1. Great Moderation | * Best tracking within Great Moderation among all Regression Tree models * Captures long-term trends the best within Great Moderation |
| Boosted Trees | 1. ZIRP  2. Tight Block  3. Loose Block | * Lowest RMSE for ZIRP * Comparable tracking for ZIRP and Loose Block * Overestimates rates in Tight regime between July 2006 and July 2007, but captures the financial crisis dip better than any other Regression Tree model |
| LASSO | 1. Great Moderation  2. ZIRP  3. Tight Block  4. Loose Block | * Lowest RMSE for Great Moderation * Tracks the 10-year Treasury better than all other models in all regimes except for Loose Block, where it falls just behind Bagged Trees * Substantially outperforms all other models in capturing the financial crisis dip in late 2008 |

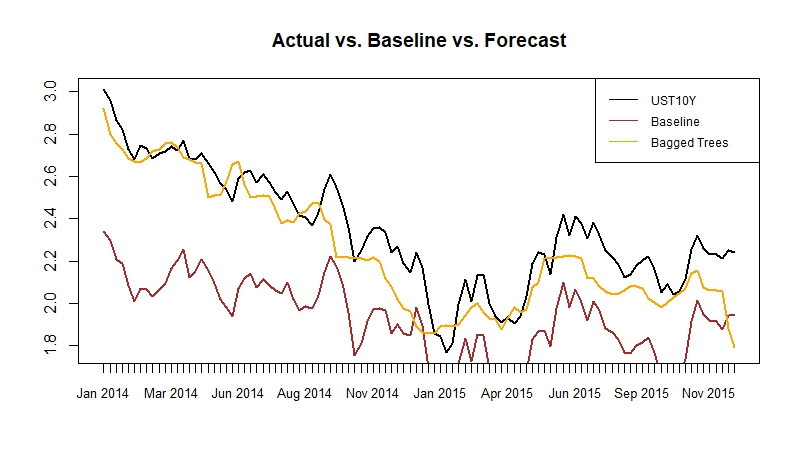
The following graphs depict the forecast curves within each regime generated by the best models for the regime according to the table above. As described in the table, we take “best” to mean the models that, within a given regime,

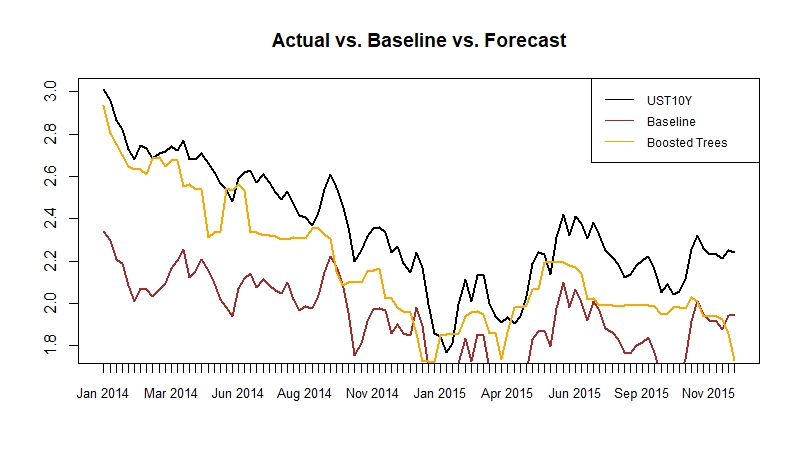
* generate the lowest RMSEs
* capture the directionality of the 10-year Treasury.

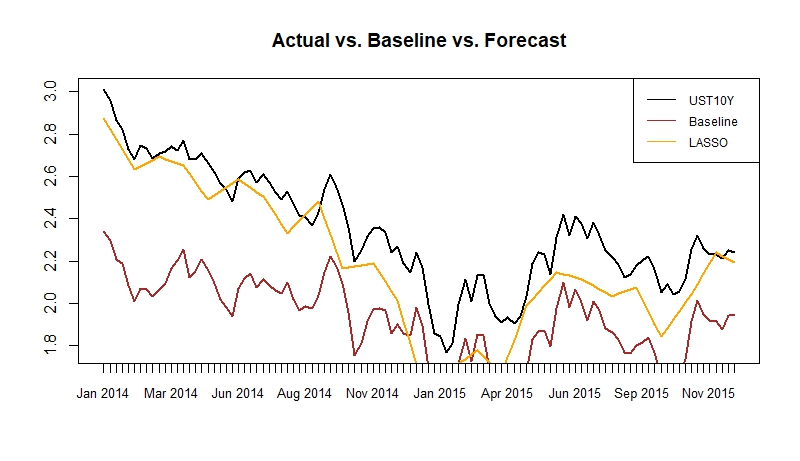
The forecast curve generated by the model is indicated by the yellow curve and spans the entirety of the corresponding test set indicated in our Regime Switching table. The observed 10-year Treasury is depicted by the black curve. Our baseline, the 3 year hence 10-year forward rate rolled forward three years, is depicted by the red curve.

**Great Moderation**

This set of graphs displays the forecast curves for Random Forest and LASSO within the Great Moderation regime. These two methods generate the lowest RMSEs within this regime, with the LASSO performing the best. The LASSO forecasts track closely with movements of the 10-year Treasury. In particular, the LASSO is able to capture the sharp dip in the 10-year Treasury during the downturn in late 2008. This signifies that lagged macroeconomic data, used by the Fed in determining monetary policy, is able to capture interest rate movements during the financial crisis. It is also suggestive of a linear relationship existing between the 10-year Treasury and macroeconomic indicators during this time period.

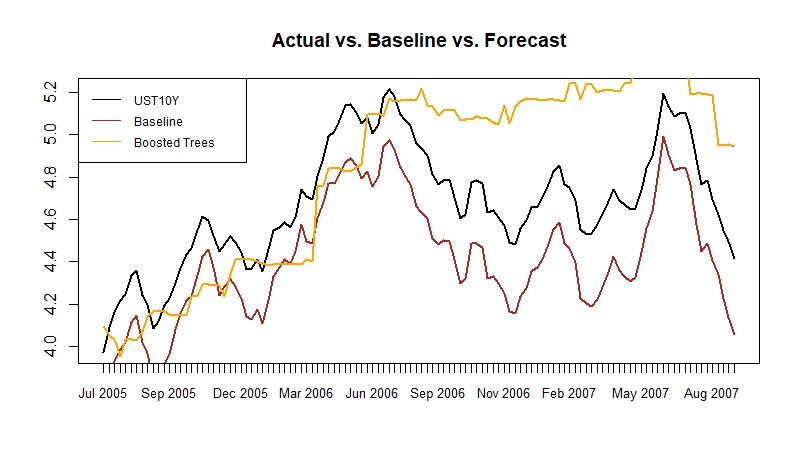
**ZIRP**

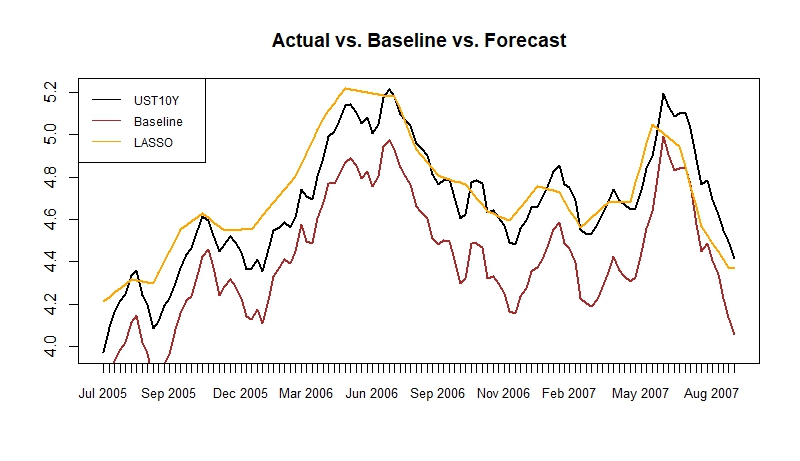




This set of graphs depicts the forecast curves for Bagged Trees, Boosted Trees, and LASSO within the ZIRP regime. While Boosted Trees provide the lowest RMSE among all models, LASSO and Bagged Trees are comparable for second-best fit. LASSO is more consistent in its tracking, however, than both Bagged Trees and Boosted Trees. Bagged Trees, in particular, does a poor job of tracking the 10-year Treasury during 2014 and experiences a sharp divergence from observed rates in Nov 2015. Boosted Trees does not track the 10-year Treasury well from May 2015 onwards.

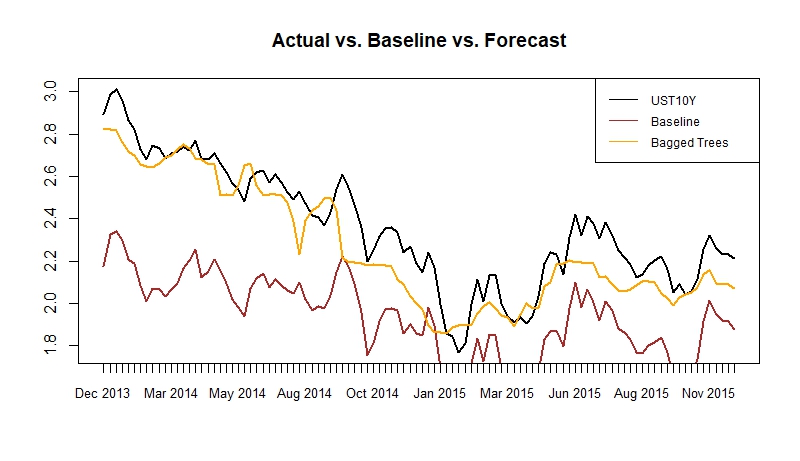
**Tight Block**

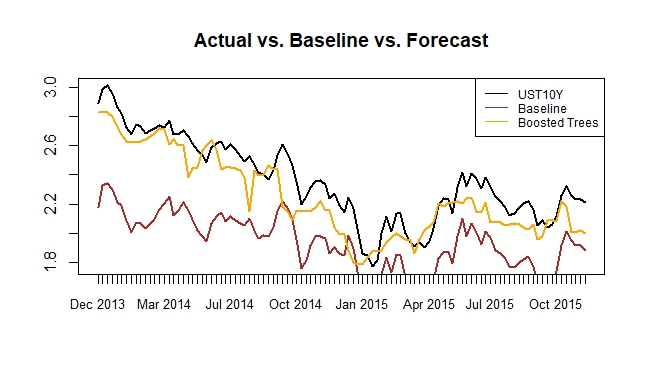


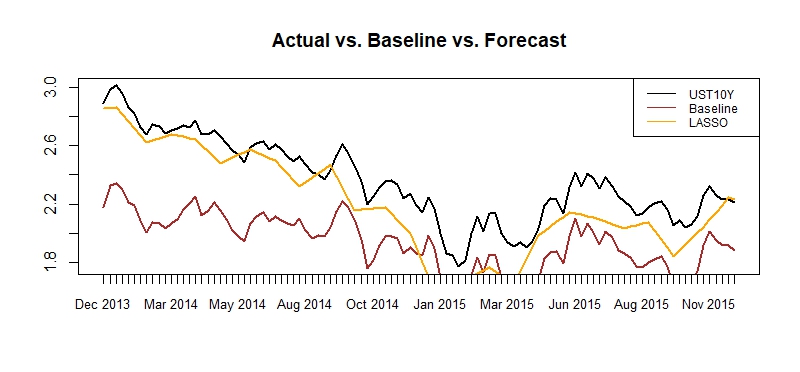


This set of graphs depicts the forecast curves for Boosted Trees and LASSO within the Tight Block. From the RMSE table, we glean that all three models performed well over this regime, with Bagged Trees performing the best. Bagged Trees, however, did not track the 10-year Treasury well. Of note, we observe the performance of Boosted Trees and LASSO around the two prominent rate peaks in July 2006 and July 2007. Boosted Trees perform the best among all Regression Tree methods in capturing the downward dip observed at the end of the test set. LASSO tracks the many undulations of the 10-year Treasury during this regime the best among all models.

**Loose Block**







This set of graphs depicts the forecast curves for Boosted Trees, Bagged Trees, and LASSO within the Loose Block. LASSO and Bagged Trees provide the lowest RMSEs among all models in this regime. Both of these models also have comparable RMSEs, similar to their performance within ZIRP. We find this intuitive as the ZIRP regime and the Loose Block possess a substantial amount of temporal overlap. Bagged Trees track the 10-year Treasury better than the LASSO. LASSO suffers from periods of poor directionality, particularly during Aug and Nov 2015.

**Do Our Results Support Our Hypothesis?**

The question our research sought to answer is whether machine learning methods can forecast the 10-year Treasury more accurately than the market’s expectation baseline. **Based on our results, we confirm our initial hypothesis of Yes.** Our machine learning models represent an improvement in forecasting accuracy over the baseline. Moreover, the models offer meaningful insight into the nature of the relationships between our independent variables and the 10-year Treasury.

Within the Great Moderation, our LASSO model reports back the lowest RMSE. We also note that the LASSO tracks the 10-year Treasury the best during this regime. Within ZIRP, Boosted Trees reports back the lowest RMSE. LASSO once again appears to perform the best in capturing the directionality of the 10-year Treasury during this regime. SPREAD provides the greatest relative importance to the model. PCE and UNEMP follow in importance, but at a rate of less than 10% than that of SPREAD.

For the Tight and Loose blocks, Bagged Trees outperform all other models. The discrepancy in RMSE is most apparent in the Loose Block, where Bagged Trees outperform the second-best model by an order of magnitude. Within the Tight Block, LASSO is the best tracker of the 10-year Treasury. Within the Loose Block, Bagged Trees is the best tracker of the 10-year Treasury. The most important predictors of the 10-year Treasury in both blocks are SPREAD and FED\_FUNDS. Depending on regime, different arrangements of PERMITS, CTI, UNEMP, and LOANS follow in importance, but are less important than SPREAD by at least an order of magnitude.

From our results, we surmise that SPREAD and FED\_FUNDS are predictors fundamental to the determination of the 10-year Treasury. This finding coincides with much of the financial econometrics literature that also finds the spread to be among the most significant predictors of interest rate movements. Bagged Trees and Boosted Trees report back the lowest RMSE in three of the four regimes we considered. LASSO provides the best or second-best fit in all regimes. The LASSO model also captures the directionality of the 10-year Treasury better than all other models except within the Loose Block, where it falls just behind Bagged Trees in this regard.

The LASSO’s strength intracking the 10-year Treasury represents a Value Add to market participants in the short-term looking to predict and profit from whether the market will move up or down. On the other hand, the ability of Regression Tree methods to provide lower RMSEs represents a Value Add to investors looking to profit over a longer-term horizon. The strength of our results also speak to the importance of taking into account regime switching in order to maintain a statistically valid comparison between training and test sets for time series.

In conclusion, our models’ improvement in forecasting accuracy with respect to the 3-year hence 10-year forward rate baseline reflects the ability of machine learning models to outperform traditional econometric techniques in terms of forecasting. Our results confirm our hypothesis and the presence of a Value Add for financial market participants. The Value Add also extends to applied economists studying the extent to which particular macroeconomic variables may influence the evolution of Treasury rates.

1. Pruning is the process of limiting the depth of fitted trees in order to curtail overfitting. The increased complexity of deeply fitted trees reduces forecasting performance due to high variance. A shallower tree allows for better interpretation and performance at the cost of introducing some bias. [↑](#footnote-ref-1)
2. James,G., Witten, D., Hastie, T., Tibshirani, R. (2013) *An Introduction to Statistical Learning with Applications in R. (Ch. 8, p. 322)* [↑](#footnote-ref-2)