A. Initial Independent Variable Selection

The estimated model should explain real long-term interest rates as a function of low-frequency and high frequency determinants. Indeed, as per the replicated paper, "low-frequency determinants can be thought of as the fundamentals that influence saving and investment trends, while high-frequency determinants are those which proxy the movements in expectations about these fundamental factors". And so, we attempted to go beyond the determinants chosen in the paper, while staying true to the framework therein defined and developed.

We expanded the state space of possible low-frequency determinants with a plethora of functions on the raw data collected: long-term moving averages, polynomial combinations of data columns, etc… Also, first differences and short term expectations/predictions of the low-frequency determinants constituted the state space of possible high-frequency determinants. We must note that the paper makes use of the Hodrick-Prescott filter to generate high-frequency determinants. However, research regarding HP filters has revealed that despite it's popularity, the method inflicts patterns on time-series that do not originate from the data-generating process. Indeed, HP filters are especially inadequate when the observed variables follow a random walk; all selected determinants follow a random walk (https://voxeu.org/article/why-you-should-never-use-hodrick-prescott-filter).

B. Model Development Principles

Real long-term interest rates drive long-term saving and investment trends, and fundamentally influence business cycles, and the government's macroeconomic policies. And so, we aimed to develop a model with three main characteristics:

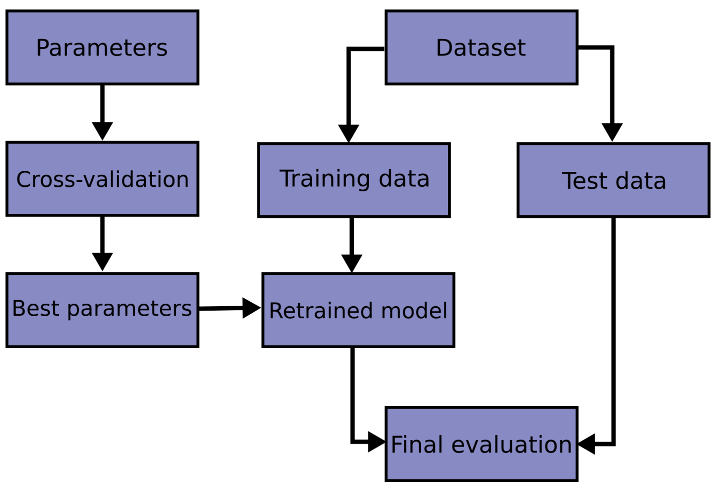
- Statistically significant, i.e. a regression model with satisfactory R^2
- Great generalization ability, i.e. a regression model with low variance
- Easily interpretable by humans, hence expanding the theoretical framework on the determinants of real interest-rates, i.e. a sparse model

C. Regression Method Selection

As in the paper, we expect real interest-rates to be a linear combination of the features. Linear models aim to minimize a certain measure of prediction error. In order to satisfy our model development principles, we chose to explore linear regression with regularization. Regularization are techniques used to avoid overfitting. The main regularization methods are L_1 and L_2. The first prefers models with fewer non-zero coefficients, leading to sparser models. The second penalizes large coefficients, leading to more robustness to collinearity. Consequently, we chose to explore Lasso regression (L_1), and Elastic Net Regression (weighted combination of L_1 and L_2).

D. Model Training

Model training is a two step process: hyper-parameter tuning and model fitting.



Hyper-parameters are parameters not directly learnt within estimators. We start by defining a hyper-parameter space, and search the space for the optimal hyper-parameter using cross-validation.

- Lasso Regression
  - Hyper-parameter: alpha
  - Hyper-parameter Space: 0.001 to 100
  - Cross-Validation: 5fold cross validation
  - Optimal Hyper-parameter Value: 0.126

- Elastic Net Regression
  - Hyper-parameters: alpha, lambda
  - Hyper-parameter space: 0.001 to 100, 0.001 to 1
  - Cross-Validation: 5x 10fold cross validation
  - Optimal Hyper-parameter Values: 0.163, 0.0185

Then, we randomly split the dataset into a test and train set, where the train set constitutes 70% of the dataset, and fit the above models to the training set, using the optimal hyper-parameters. Both models repeatedly deemed all transformed features as insignificant, and reduced the high-frequency determinant space the expected inflation and short-term interest rates.

We obtained the following model coefficients:

- Lasso Regression

```
Lasso Model
19 x 1 sparse Matrix of class "dgCMatrix"
                                                                                          1
(Intercept)                                                                     3.577673810
`Govt. Deficit as % of GDP`                                                     .
`Short-Term interest rates (%)`                                                 0.778051947
`Inflation, Consumer Prices, US (%)`                                            .
GDPIPD                                                                          .
`10 year average of %YoY (Ksen calc)`                                          -0.643625971
`5-Year, 5-Year Forward Inflation Expectation Rate`                            -0.118829697
`Inflation Expectations`                                                        0.325168451
`Capital Stock at Constant National Prices for United States`                   0.002694966
`Net capital stock, volume`                                                     .
`Net Operating Surplus, financial corporations, percentage of net value added`  .
`Net Operating Surplus, non financial corporations, percentage of net value added` .
`10 Year Tresury ex post Bond Yield`                                            .
`S&P Value`                                                                     .
`S&P 500 Returns`                                                               0.183738530
`Baa Corporate Bonds Return`                                                    0.568238069
`Market value of Bonds in Circulation`                                         -0.045899846
`US Fixed Income Securities Outstanding`                                        .
`GDP Change`                                                                    .
```

- Elastic Net Regression

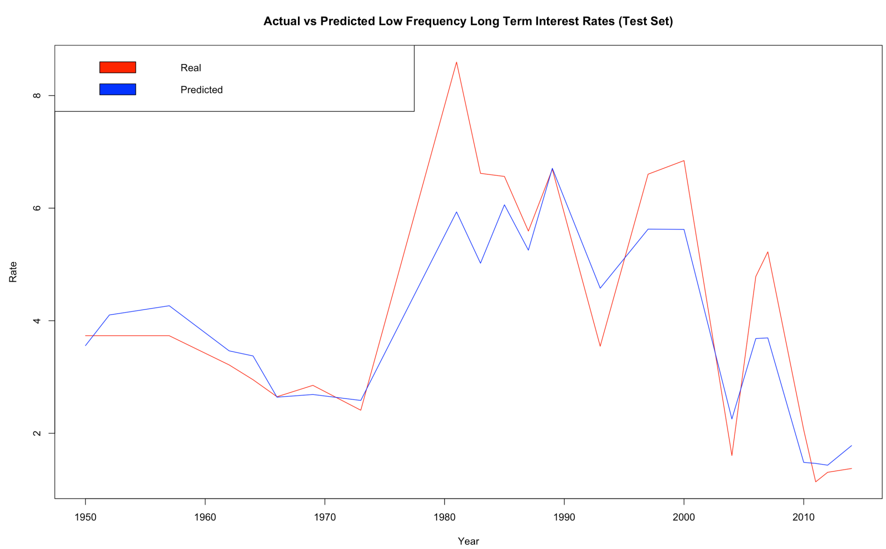
```
Elastic Net Model
19 x 1 sparse Matrix of class "dgCMatrix"
                                                                                          1
(Intercept)                                                                      3.57767381
`Govt. Deficit as % of GDP`                                                      0.15990209
`Short-Term interest rates (%)`                                                  1.04029181
`Inflation, Consumer Prices, US (%)`                                            -0.13875013
GDPIPD                                                                          -0.23690456
`10 year average of %YoY (Ksen calc)`                                           -0.62914721
`5-Year, 5-Year Forward Inflation Expectation Rate`                             -0.27814677
`Inflation Expectations`                                                         0.33739091
`Capital Stock at Constant National Prices for United States`                    0.25876983
`Net capital stock, volume`                                                      0.23726676
`Net Operating Surplus, financial corporations, percentage of net value added`  -0.02485816
`Net Operating Surplus, non financial corporations, percentage of net value added` .
`10 Year Tresury ex post Bond Yield`                                             0.37543815
`S&P Value`                                                                      0.36907309
`S&P 500 Returns`                                                                0.14922280
`Baa Corporate Bonds Return`                                                     0.45138239
`Market value of Bonds in Circulation`                                          -0.15064397
`US Fixed Income Securities Outstanding`                                        -0.30677536
`GDP Change`                                                                    -0.01527127
```

As expected, the Lasso Model is drastically sparser, consisting of 8 significant independent variables. The Elastic Net Model is more complex, potentially “capturing” more predictive value from the independent variables. The Lasso model seems to best adhere to our principles. However, a comparison of their performance on the test set is necessary to select the best model to predict long term interest-rates, as a linear combination of low and high frequency determinants.

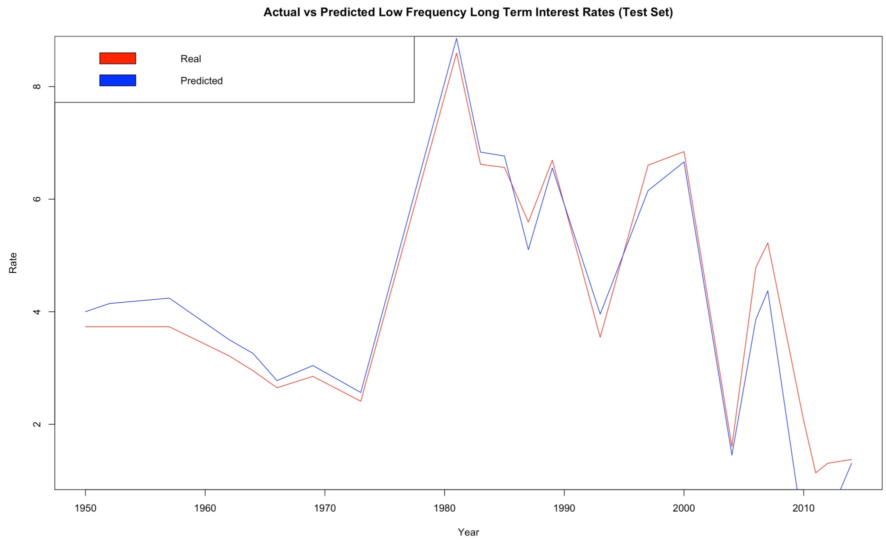
E. Model Testing

We then predict the real interest-rates of the test set observations, and obtain the R^2 of the regressions:

- Lasso Regression
  - Train Set R^2: 0.836466
  - Test Set R^2: 0.812906



- Elastic Net Regression
  - Train Set R^2: 0.9485969
  - Test Set R^2: 0.905927



Clearly, Elastic Net Regression performs better than Lasso Regression, as it most accurately predicts real interest-rates in the test dataset. Nevertheless, the choice of model depends on how one weighs the three model development principles.