Multi-dimensional Selection Factor Model

Xugan Chen, Yufeng Hu, Wencheng Chen, Jiayi Wang

Wutong Zhouxin, Chen Liu, Yinglu Shi, Yixuan Fang

Course Fixed Income Analytics and Models

December 11th, 2018

Abstract

We study time-varying risk premia in Treasury bonds. There are two innovations of the project. First, we set up a model using multi-dimensional factor, including technical factors, macroeconomic and financial factors, composite leading factors, market factors, sentiment Factors, stock model factors and other classical model factors. Then we use multi-selection algorithm to process the data and select factors, including using class-within PCA to get one factor from each class and using backward selection to remaining class-within factors filtered by significance. Besides, we do the robust test in different aspects and find that our model get better results when ruling out crisis. Finally, we conclude that our model with moderate factors performs a higher R^2 than the classical models.

Content

1. Introduction

We study time-varying risk premia in Treasury bonds. One strand of our research relates such variations to excess return*(Cochrane and Piazzesi,2005)*, financial variables*(Ludvigson and Ng ,2007)*,macroeconomics variables*(Ludvigson and Ng ,2009),*Fama French 5 factor, whereas more and more recent studies within this strand has uncovered a wide array of factors including jump risk*(Wright J H, Zhou H,2009)*, liquidity*(Lin H, Wang J, Wu C,2011; Amihud Y,2002)*,BwSent sentiment*(Baker and Wurgler,2006,2007)*,technical varibles and composite leading indicators.

In this paper, we process our data by machine learning. Every factor selected from these variations is filter by “Principal Component Analysis”(PCA).Then, we used the “Backward Selction” to screen these factors to get more significant ones. We can avoid multiple collinearity problems to a great extent by this way. The more details we will dicuss in the subsequent sections.

2. Literature Review

By meticulously reading the literatures given in GroupProjectReading file, we analyzed the models in these literatures and refine some factors of them. Next we classify them into groups and applies into our model. However, we not only use the factors and data in these literatures but also refer to other articles and make some innovations. Specifically, our liquidity index are consulted from “Lin H, Wang J, Wu C. Liquidity risk and expected corporate bond returns[J]. Journal of Financial Economics, 2011, 99(3): 628-650.” For CLI indictors, using CLI factors from eight countries, the first principal component is the CLI_all factor. We have more comprehensive technical indicators. The literature Goh et al 2013 only has 63 technical indicators, but we have 10 categories with 200 technical indicators which come from Yahoo Finance. Besides, CP and log_excess_return indicator are fully according to the article. For Macro_raw_data of LN and Financial data of LN, we do PCA for each category of indicators. In general, our model is constructed by comprehensive consideration and easier to analyze. The indicators of the model are more integrated and precise.

3. Data and Methodology

3.1 Indicators and Data Source

Nine categories of indicators have been considered in the model. Detailed information and data sources will be given as follows.

[Table 1 around here]

（1）Liquidity indicators

Recent studies have further suggested liquidity as a good candidate for a priced state variable, because liquidity is often viewed as an important feature of the investment environment. Investors should require higher returns on assets whose returns have greater sensitivities to market liquidity. Therefore, liquidity indicators are included in this report to predict bonds' access return.

1) Amihud Liquidity indicator*

$$L_t = \frac{V_t}{|\ln p_t - \ln p_(t-1)|}$$

V_t: daily trade volume in money: $V_t = \frac{hightest\_price + lowest\_price}{2} \times v_t$

P_t: daily market-close price

2) Pastor and Stambaugh Liquidity indicator

(2) Fama-French five factors

The Fama-French 5 factors are constructed using the 6 value-weight portfolios formed on size and book-to-market, the 6 value-weight portfolios formed on size and operating profitability, and the 6 value-weight portfolios formed on size and investment.

(3) Financial indicators

147 financial market variables, updated from Ludvigson and Ng (2007) "The Empirical Risk-Return Relation: A Factor Analysis Appraoch,". They can be divided into four classes, including 'PYD(price, yield and dividends)', RiskFactors, Industries and Size/BM. For detailed information, please refer to the appendix.

（4） Macroeconomic factors

132 macroeconomic variables, updated from Ludvigson and Ng(2009) “Macro Factors in Bond Risk Premia”. The 132 variables ca be divided into eight classes, including "output_and_income, labor_market, housing, consumption, money, bond_ex, price and stock". For detailed information, please refer to the appendix.

（5） Technical indicators

Technical indicators are composed of together 200 factors, which can be divided into ten classes by tradition. For detailed information, please refer to the appendix.

（6） CP_t indicators

Based on the (Cochrane and Piazzesi 2005), calculate the CP_t factors, which predict the access return of bond by forward rate of different maturities.

（7） Composite leading indicators

The composite leading indicator (CLI) is designed to provide early signals of turning points in business cycles showing fluctuation of the economic activity around its long-term potential level. CLIs show short-term economic movements in qualitative rather than quantitative terms.

（8） BWSENT sentiment indicators

Baker and Wurgler,"Investor Sentiment and the Cross-Section of Stock Returns" concluded previous work on studying investors' sentiment. It suggests a number of proxies for sentiment, including CFFD, TURN, NIPO, RIPO, S and $P^{D\text{-}ND}$:

1） CFFD(the closed-end fund discount) : average difference between the net asset values of closed-end stock fund shares and their market prices. Prior work suggests that CFFD is inversely related to sentiment (Zweig 1997 and Lee et al.1991).

2）TURN: the natural log of the raw turnover ratio, detrended by the 5-year moving average. Baker and Stein(2004) suggest that turnover can serve as a sentiment index.

3） NIPO,RIPO. The IPO market is often viewed as sensitive to sentiment. So the number of IPOs, NIPO and the average first-day returns RIPO can be viewed as indicators of the market sentiment.

4） S(the share of equity issues in total equity and debt issues) is another measure of financing activity that may capture sentiment.

5） $P^{D\text{-}ND}$ (dividend premium): log difference of the average market-to-book ratios of payers and nonpayers, regarded as the sixth proxy of market sentiment.

Therefore, SENTIMENT is defined as the first principal component of the correlation matrix of the above variables:

$$SENTIMENT_t = -0.241CEFD_t + 0.242TURN_{t-1} + 0.253NIPO_t + 0.257RIPO_{t-1} + 0.112S_t - 0.283P_{t-1}^{D-ND}$$

To form a better index, Baker and Wurgler use the orthogonalized proxies following the same procedure as before and get another indicator:

$$SENTIMENT_t^{\perp} = -0.198SCEFD_t^{\perp} + 0.225TURN_{t-1}^{\perp} + 0.234NIPO_t^{\perp} + 0.263RIPO_{t-1}^{\perp} + 0.211S_t^{\perp} - 0.234P_{t-1}^{D-ND,\perp}$$

（7）Market jump risk measures

In Wright J H, Zhou H. Bond risk premia and realized jump risk[J], Market jump risk measures is potentially constitute unspanned risk factors for predicting excess returns.

$$ds_t = \mu_t dt + \sigma_t dW_t + J_t dq_t$$

$$r^s_{t,j} \equiv s_{t,j\cdot\Delta} - s_{t,(j-1)\cdot\Delta}$$

$$\begin{aligned} RV_t &\equiv \sum_{j=1}^{m} |r^s_{t,j}|^2 \rightarrow \int_{t-1}^{t} \sigma_u^2 du + \int_{t-1}^{t} J_u^2 dq_u \\ BV_t &\equiv \frac{\pi}{2}\frac{m}{m-1}\sum_{j=2}^{m} |r^s_{t,j}||r^s_{t,j-1}| \rightarrow \int_{t-1}^{t} \sigma_u^2 du. \end{aligned}$$

$$RJ_t \equiv \frac{RV_t - BV_t}{RV_t}$$

μ_t and σ_t are the instantaneous drift and diffusion functions that are completely general and may be stochastic (subject to the regularity conditions), W_t is a standard Brownian motion, dq_t is a Poisson jump process with intensity JI_t, and J_t refers to the corresponding (log) jump size distributed as Normal (JMt,JVt2). Note that the jump intensity, mean and volatility are all allowed to be time-varying in a completely unrestricted way. s_t is the log form of the price of an asset.

3.2 Methodology

3.2.1 Filter factors

Principal components analysis (PCA)

For financial factor, technical factor and macroeconomic factor, each contains several different classes. In this project, we choose only one factor from each class by PCA to eliminate the problem of multicollinearity and reduce the number of factors to an acceptable number.

Backward method to reduce the number of factors

After the PCA step, we have reduced the number of factors to a large extent, the backward method is to eliminate factors with less important influence, thus reducing the number of factors further.

Supposing the multi-regression model is:

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_K x_{iK} + \varepsilon_i \text{, which is } y = X\beta + \varepsilon$$

$$t|X = \frac{b_k}{[s^2 (X'X)^{-1}_{kk}]^{1/2}} \text{ , in which } s^2 = \frac{e'e}{n-K}$$

1) Step1: do regression and calculate the p-value of different factors: $P_i = 2P(t(n-k) > |t_i|)$
2) Step2: eliminate the factor which has the biggest p-value
3) Step3: repeat the above steps until $(P_i = 2P(t(n-k) > |t_i|)) \leq 0.1$ for each i
4) Step4:do multi-regression with the remaining factors and get the final result

After we filtering principal components within classes, we follow the methodology described by Cochrane and Piazzesi (2005) and construct single factors by linear combination. The restricted model is a reasonable substitute for the unrestricted one. Our results are reported in Table 1 to Table 8 in appendix.

3.2.2 Prediction

In-sample analysis

After filtering the factors, we want to select the best combinations of them. So we construct different factor groups, run the regression respectively and find the highest R^2 among them:

$$y_{t+1}^{(n)} = X_t \beta_t + \varepsilon_t, \quad \text{for } n = 2,3,4,5,$$

The data we use are bonds of 2-5 year maturity during the period January 1980 to December 2002.

Next, we construct Wald statistics of large samples to test the in-sample regression model:

Under the constraint condition of $R\beta = q$, we have

$$W = (R\hat{\beta} - q)'[R(s^2(X'X)^{-1}R']^{-1}(R\hat{\beta} - q) \xrightarrow{d} \chi^2(J),$$

where J represents the number of conditions.

Out-of-sample analysis

Getting the best factor combination, we use the model to do out-of-sample analysis. The method we use in this part is rolling window regression:

1) Step1: Select an initial period and use its consecutive observations to run the regression:

$$y_t^{(n)} = X_{t\text{-}1}\beta_{t\text{-}1} + \varepsilon_t, \quad \text{for } n = 2,3,4,5,$$

then get the estimation of $\beta_{t\text{-}1}$, which is $\hat{\beta}_{t-1}$.

2) Step2: Predict the bond risk premium of the following next month after the rolling window in step1 using $\hat{\beta}_{t-1}$ and X_t:

$$\hat{y}_{t+1}^{(n)} = X_t\hat{\beta}_{t\text{-}1} + \varepsilon_{t+1}, \quad \text{for } n = 2,3,4,5,$$

3) Step3: Move the rolling window one month later and keep the length of window constant, use the subsamples of the new time period and similar method in step1 to get $\hat{\beta}_t$.

4) Step4: repeat the above steps until we get the result of the last period.

The first rolling window here is January 1980 to December 2002, and we predict the bond risk premium month by month from January 2003 to August 2009.

3.2.3 Predictive ability

We study the out-of-sample predictive ability based on three indicators: R^2, $RMSE$ and R_{OOS}^2.

R^2 (coefficient of determination): $R^2 = \frac{\text{SSR}}{\text{SST}} = 1 - \frac{SSE}{SST} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2}$, it measures how well our regression predictions approximate the real data points. The closer R^2 is to 1, the better the regression predictions fit the data.

$RMSE$ (root-mean-square error): $RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$, similar to R^2, $RMSE$ also measures how well our model is and how precise we predict, the smaller $RMSE$ is , the better predictive effect our model has.

$R_{OOS}^2 = 1 - \frac{\sum_{i=1}^{n}(y^{(n)} - \hat{y}^{(n)})^2}{\sum_{i=1}^{n}(y^{(n)} - \bar{y}^{(n)})^2}$, where $y^{(n)}$ is the true excess return we observed, $\hat{y}^{(n)}$ is the forecast using our own model, and $\bar{y}^{(n)}$ is the mean value of all true data we observed. This indicator evaluates the reduction in mean square prediction error (MSPE) for our regression model relative to the historical average forecast benchmark. A $R_{OOS}^2 > 0$ implies that the MSPE of our model is lower than that of the historical benchmark, indicating a higher predictive accuracy.

4. Empirical Findings

4.1 Backward Selection

We start from the whole sample regression of excess bond returns. We use the sample of bond with 2 year maturities. The sample period is January 1980 to December 2009. Firstly, we use Backward Selection based on p-value to eliminate 22 less influential factors and get 18 factors remained. The table 10 below shows the factor-removing process.

After Backward Selection, we then use the remaining 18 factors as dependent variables to regress excess returns on 2-year maturity bonds to examine the explanatory ability. The model is constructed as below.

y_t denotes log excess return at time t, $x_{it}\ (i = 1,2,3 \dots 18)$ represent 18 variables respectively, β_{it} is the coefficient of variable i at time t.

$$y_t = \beta_{1t}x_{1t} + \beta_{2t}x_{2t} \dots + \beta_{18t}x_{18t} + \varepsilon_t$$

The table 11 displays the regression results. The Adjusted R-squared reaches 0.697, indicating that the model has a strong explanatory power.

4.2 Whole Sample Regression (2-5-year maturity bond)

In this step, we apply the model above into further analysis on 2-5 year maturity bonds. The table 12 shows the result of whole sample regression.

Our research extends the previous work. Eugene Fama and Robert Bliss's (1987) report that the spread between the n-year forward rate and 1-year yield can be used to predicts the 1-year excess return of the n-year bond. It's R^2 is about 0.18. Cochrane and Piazzesi's (2005) find that a single tent-shaped linear combination of forward rates, predicts excess returns on 1-5-year maturity bond with R^2 up to 0.44. Ludvigson and Ng (2009) show that macroeconomic fundamentals have important predictive power for excess returns on U.S. government bonds which has R^2 of 0.45.

Based mainly on the research in CP, we do further research by combining the factors of FB、CP and Ng, together with the new ones concerning technology, investors sentiments, bipower variation and China land factor.

Our R^2 in whole sample regression reach to 0.712, 0.687, 0.669 and 0.645 respectively for 2-5-year bond, much higher than 0.44(CP). Besides, as maturity increase, the explanatory abilities decrease which are consistent with the results in CP.

Though our multifactor model is more comprehensive and has a better explanatory ability, it has some drawbacks.

Firstly, as the results show, when we eliminate the influence of the number of variables, the adjusted R-squared drop down by approximately 2%-3%. It indicate that further study should be done to address the problem of the excessive number of factors.

Secondly, we have not go through the term structure model as CP does. CP find that the return-forecasting factor is a symmetric, tent-shaped linear combination of forward rates. The shape implies that an important part of its forecast power is unrelated to the standard ”level”, ”slope” and “curvature”. In our research, we ignore the impact of the term structure of bond with different maturities.

4.3 In-sample regression

The table presents the in-sample regression of excess return using monthly data of 2-5-year maturity bonds with a sample period from January 1980 to December 2002.

In this step, we lag each of the excess return by one month. We further use different combination of 18 variables, trying to find out the best choice.

The model is as follows. y_{t+1} denotes log excess return at time t+1, x_{it} $(i = 1,2,3 \dots 18)$ represent 18 variables respectively, β_{it} is the coefficient of variable i at time t.

$$y_{t+1}^{(n)} = \beta_{1t}x_{1t} + \beta_{2t}x_{2t} \dots + \beta_{18t}x_{18t} + \varepsilon_{t+1} \textit{ for } n=2,3,4,5$$

The table 13-16 shows the regression results. Comparing 2-11 columns, we find that the model with the whole 18 variables shows the highest adjusted R-squared. However, as shown in the first column of each table, the expectation hypothesis display R-squared approximately to zero which indicate that the EH does not hold in our research. The main reason accounting for this may be that we have not concern about the term structure model, a drawback mentioned above.

We then compute a Wald test on factors in terms of Fama, investors' sentiments, technology, micro and financial. The table 17 shows the results. Calculation give relatively large □^{2} of 5 in-sample models with 2-5-year maturity bond excess return. The p-value approach to zero, indicating that the factors are significant under the confidence level of 1%. Both two statistics convincingly reject the H0 hypothesis which strengthen the feasibility of in-sample regression model.

4.3 Out-of-sample regression

The results of in-sample regression suggests that the combination with whole 18 variables has the best explanatory power. Therefore, we use the Rolling Window approach to do the out-of-sample predicting regression based on this model.

The model is as follows. y_t denotes log excess return at time t, $y_{pred,t+1}$ is predicted excess bond return at time t+1, $x_{it-1}\ (i = 1,2,3 \ldots 18)$ represent 18 variables respectively, β_{it-1} is the coefficient of variable i at time t-1, $\hat{\beta}_{it}$ is the estimated β_{it-1}.

$$y_t^{(n)} = \beta_{1t-1}x_{1t-1} + \beta_{2t-1}x_{2t-1} \ldots + \beta_{18t-1}x_{18t-1} + \varepsilon_t$$

$$\hat{y}_{pred,t+1}{}^{(n)} = \hat{\beta}_{1t}x_{1t} + \hat{\beta}_{2t}x_{2t} \ldots + \hat{\beta}_{18t}x_{18t} \quad (n=2,3,4,5)$$

The forecast period is from January 2003 to August 2009. The length of the historical period is 22 years, starting from January 1980 to December 2002.

The figure 1 and figure 2 contain 8 pictures, depicting the fitting ability of the prediction in two aspects.

The figure 1 shows that as the bond maturity increases, the curves of predicted bond excess return deviate away from the true bond excess return in much larger degree.

The figure 2 shows that the volatility of the predicted bond excess return increases when maturity get longer which is in accordance with the trend of the true excess return.

Besides, the R-squared curves of longer maturity bond lie below that of the shorter maturity.

Afterward, we use the statistics of R^2_{OOS}、RMSE to analyze the predicting result, The table 18 reports R^2_{OOS}、RMSE of 2-5 year maturity bonds.

The result shows that as bond maturity increases, RMSE gets larger and R^2_{OOS} gets smaller. It strengthens the evidence that the prediction ability become worse in longer maturity bond. Therefore, we can draw a conclusion that the out-of-sample predictability of excess returns is stronger in shorter maturity bond.

5. Robustness Check

5.1 Restricted vs Unrestricted Regression

First we employ comparison test between restricted and unrestricted regression similar to Cochrane and Piazessi (2005). The process is briefly described as follows.

The unrestricted regression is the vanilla regression described in Section 3. We regress all 18 factors to 4 types of log returns, which is of 2, 3, 4, 5 years' maturity.

The restricted regression, in contrary, uses average of returns as dependent variable. Denote $\overline{rx_t} = \frac{1}{4}\sum_{i=2}^{5} rx_t^i$ the average return of all maturities.

Step 1: Regress $\overline{rx_t}$ with respect to all factors.

$$\overline{rx_t} = \beta X + \epsilon,$$

Where X is matrix of vanilla factors and β their loadings. Denote $\gamma = \beta X$ the new factor.

Step 2: Regress log returns of different yields with the new factor.

$$rx_t^i = b_i\gamma + \epsilon$$

The following table19 shows results from Step 1, with 0.69 adjusted R^2

[table 19 around here]

The result resembles the result of unrestricted regression. Notice as Cochrane pointed out, as the maturity increases, the factor loadings magnify (whichever it's sign). We refer the reader to Section 3 to show that the loadings are in the middle of results from maturity year 2 and 3.

The result of Step 2 is briefly shown as table 20.

[table 20 around here]

In general, this test reproduces similar phenomenon to Cochrane and Piazessi (2005): as maturity increases from 2 to 5, the factor loadings magnify, but the predictability decreases. From unrestricted to restricted regression, there are minor differences of R^2 and even higher with longer maturities. We thus prove that our factor is time independent and robust for independently predict returns of all maturities.

5.2 Quarter Return Forecast

The original Cochrane and Piazessi (2005) and other articles based on it generally uses monthly return data from CRSP. We test predictability of our factor for longer duration. The frequency we choose is 3 months. All other conditions are the same.The in-sample regression results are assembled in the table 21. The first row is R^2 for each maturities, followed by factors and their standard error.

The out-sample prediction results are shown in the figures 3-4.

We can see from the result that the in-sample regression shows greater R^2. The improved predictability may result from more regular periodic behavior, or just because of smaller sample size. But the out-sample prediction shows promising results. Although it’s not precise, it successfully inspects up- and downward trends, proving the robustness of our factor.

5.3 Effect of Financial Crisis

Our regression timespan is chosen to include three recent major financial crisis, which we list as follows:

Mexico Financial Crisis, from Dec. 1994 to Mar. 1995

Asian Financial Crisis, from Jul. 1997 to Dec. 1998

American Subprime Mortgage Crisis, from Feb. 2007 to Jul. 2009

The VIX is an important indicator for volatility, and also an indispensable factor for our regression and prediction. The time-series plot(Figure 5-6) shows VIX peeks during financial crisis.

The financial crisis strongly affects financial market, creating abnormal fluctuations in bond return. In the last part of robustness check we exclude these times, and test our factor. Hypothesis is that excluding crisis will improve predictability.

We briefly assemble the result as table 22 shows.

Crisis and volatility are our enemies. Our factor shows better predictability excluding financial crisis times. It also proved our hypothesis: as financial crisis is unpredictable (in our model), excluding financial crisis should show better predictability, which, in turn, proves the robustness of our factor.

5.4 Further Research

To sum up, we chose factors that may affect bond return and tested their predictability. We also put up with our factor in recent time periods, and the result is promising.

An analysis of correlations within 18 factors might further prove the robustness of our factor. The factor seems to exaggerate volatility, maybe due to potential asymmetric units of measure. In general, our factor and model lacks ability to take volatility into account.

6. Conclusion

Through multi-dimensional factor selecting and data processing, we find a series of variables from different classes to interpret and predict excess return.

7. Appendix

Table 1 Data Source

Indicators	Data Source
Liquidity indicator	S&P monthly data (1950-2018)
Fama-French five factors	http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/f-f_5_factors_2x3.html
Financial indicator	Raw data given
Macroeconomic indicator	Raw data given
Technical indicator	Yahoo Finance
CP indicator	Raw data given
Composite leading indicator	https://data.oecd.org/leadind/composite-leading-indicator-cli.htm
Investors' sentiment indicator	http://people.stern.nyu.edu/jwurgler/
Market Jump Risk measure	https://www.investing.com

Table 2 Restricted Model For Liquidity

n	coef.	σ	$p-value$	[0.025	0.975]	R^2
2	0.54	0.06	0.00	0.42	0.66	0.18
3	0.91	0.12	0.00	0.68	1.13	0.15
4	1.21	0.16	0.00	0.90	1.53	0.14
5	1.34	0.20	0.00	0.95	1.73	0.11

Table 3 Restricted Model For Composite Leading Indicators

n	coef.	σ	$p-value$	[0.025	0.975]	R^2
2	0.51	0.04	0.00	0.43	0.60	0.28
3	0.90	0.08	0.00	0.73	1.06	0.25
4	1.21	0.12	0.00	0.98	1.44	0.23
5	1.38	0.14	0.00	1.10	1.66	0.21

Table 4 Restricted Model For Fama-French Five Factors

n	coef.	σ	$p-value$	*[0.025*	*0.975]*	R^2
2	0.46	0.04	0.00	0.38	0.54	0.28
3	0.88	0.07	0.00	0.74	1.02	0.29
4	1.22	0.10	0.00	1.03	1.42	0.29
5	1.44	0.12	0.00	1.20	1.68	0.28

Table 5 Restricted Model For Market Sentiment

n	coef.	σ	$p-value$	*[0.025*	*0.975]*	R^2
2	0.58	0.06	0.00	0.46	0.70	0.21
3	0.94	0.12	0.00	0.71	1.16	0.15
4	1.19	0.16	0.00	0.87	1.51	0.13
5	1.29	0.20	0.00	0.89	1.69	0.10

Table 6 Restricted Model For Market Jump Risk

n	coef.	σ	$p-value$	*[0.025*	*0.975]*	R^2
2	0.50	0.06	0.00	0.39	0.62	0.16
3	0.89	0.11	0.00	0.66	1.11	0.14
4	1.22	0.16	0.00	0.90	1.53	0.14
5	1.39	0.20	0.00	1.00	1.77	0.12

Table 7 Restricted Model For Technical Factors

n	coef.	σ	$p-value$	*[0.025*	*0.975]*	R^2
2	0.51	0.05	0.00	0.40	0.61	0.20
3	0.90	0.10	0.00	0.70	1.10	0.18
4	1.21	0.14	0.00	0.93	1.49	0.17
5	1.38	0.17	0.00	1.04	1.73	0.15

Table 8 Restricted Model For Macroeconomic Factors

n	coef.	σ	$p-value$	*[0.025*	*0.975]*	R^2
2	0.49	0.02	0.00	0.45	0.54	0.57
3	0.89	0.04	0.00	0.80	0.97	0.53
4	1.20	0.06	0.00	1.08	1.33	0.51
5	1.41	0.08	0.00	1.26	1.57	0.47

Table 9 Restricted Model For Financial Factors

n	coef.	σ	$p-value$	*[0.025*	*0.975]*	R^2
2	0.52	0.06	0.00	0.40	0.64	0.16
3	0.90	0.12	0.00	0.67	1.13	0.14
4	1.22	0.16	0.00	0.89	1.54	0.13
5	1.36	0.20	0.00	0.96	1.76	0.11

Table 10 Factor-Removing Process Base On Backward Selection

	feature	Adj-R2	p-values
step1	tech_class9	0.686249	0.996148
step2	RV	0.687226	0.974965
step3	AggLiq	0.688193	0.948885
step4	tech_class2	0.689151	0.930645
step5	output_and_income	0.690096	0.903777
step6	money	0.691038	0.909795
step7	CMA	0.691946	0.838128
step8	tech_class1	0.692819	0.7876
step9	RiF	0.693677	0.771447
step10	tech_class0	0.694479	0.709583
step11	SizeBM	0.695246	0.679265
step12	bond_ex	0.695894	0.586928
step13	tech_class3	0.696403	0.50519
step14	CLI_allPca	0.696981	0.54609
step15	InnovLiq	0.697512	0.519664
step16	tech_class7	0.697966	0.481291
step17	RJ	0.698367	0.457401
step18	tech_class8	0.698756	0.45261
step19	TradedLiq	0.698996	0.393281
step20	AmihudLiq	0.699026	0.326168
step21	SMB	0.698027	0.145485
step22	RMW	0.697123	0.156101

Table 11 Whole Sample Regression With 18 Factors (2-year maturity)

	coef	Std err	P>|t|
CP	4.784	0.686	0.000
CLI	-0.929	0.561	0.099
Stock	-3.591	0.842	0.000
BWSENTcoth	-7.094	1.417	0.000
Industries	3.275	1.240	0.009
housing	-3.212	0.576	0.000
consumption	-1.709	0.535	0.002
tech_class5	-2.854	1.310	0.030
MktRF	2.845	0.833	0.001
labor_market	-0.971	0.505	0.055
PYD	-1.360	0.764	0.076
tech_class6	6.776	0.983	0.000
BV	-1.147	0.507	0.024
Price	3.964	0.400	0.000
RF	-3.449	0.614	0.000
HML	1.071	0.507	0.035
BWSENT	6.179	1.262	0.000
tech_class4	-2.612	0.973	0.008
R-squared		0.712	
Adj. R-squared		0.697	
Obs		359	

Table 12 Whole Sample Regression with 18 Factors (2-5year maturity)

	2 year	3 year	4 year	5 year
CLI	-0.929*	-2.065*	-2.724*	-3.846*
MktRF	2.845***	7.029***	11.966***	17.043***
HML	1.071**	2.605***	4.345***	5.695***
RF	-3.449***	-7.094***	-11.044***	-13.632***
BWSENTcoth	-7.094***	-13.704***	-18.124***	-21.576***
BWSENT	6.179***	11.604***	15.406***	18.929***
BV	-1.147**	-2.520**	-3.751***	-4.827***
CP	4.784***	6.876***	6.165***	3.996*
tech_class4	-2.612***	-5.502***	-9.091***	-13.040***
tech_class5	-2.854**	-4.903*	-5.873	-4.743
tech_class6	6.776***	13.243***	18.419***	22.225***
labor_market	-0.971*	-1.752*	-2.745*	-3.246*
housing	-3.212***	-6.925***	-10.554***	-14.088***
consumption	-1.709***	-3.754***	-5.310***	-6.780***
price	3.964***	7.823***	11.434***	14.338***
stock	-3.591***	-6.485***	-8.987***	-11.316***
PYD	-1.360*	-2.722*	-3.469	-4.292
Industries	3.275***	7.925***	13.528***	19.299***
R-squared	0.712	0.687	0.669	0.645
Adj. R-squared	0.697	0.671	0.652	0.626
Obs	359	359	359	359

Table 13 In-Sample Regression (2-year maturity)

	EH	CP	NG	CP+NG	CP+NG+Tech	CP+NG+SENT	CP+NG+Fama	CP+NG+BV	CP+NG+CLI	技术+投资者情绪+Fama+波动率+CLI	全部
β0	0.8697										
	(0.100)										
CLI									-3.5949	-2.7722	-3.3065
									(0.626)	(0.629)	(0.88)
MktRF			`				-1.6978			5.0289	5.2926
							(0.707)			(0.61)	(1.054)
HML							1.2958			3.8703	2.0905
							(0.665)			(0.713)	(0.603)
RF							-3.7425			-5.5909	-4.909
							(0.673)			(0.719)	(0.728)
BWSENTcoth						-7.6968				-9.3643	-8.4536
						(1.788)				(2.102)	(1.519)
BWSENT						6.0036				9.6732	6.3412
						(1.559)				(1.831)	(1.356)
BV								-1.4062		-1.7098	-1.5804
								(0.604)		(0.739)	(0.544)
CP		1.7291		4.7077	4.9411	4.7387	3.403	4.5712	5.9582		2.2457
		(0.18)		(0.737)	(0.716)	(0.761)	(0.816)	(0.733)	(0.73)		(0.839)
tech_class4					-5.6085					0.9552	1.4109
					(1.917)					(2.388)	(2.065)
tech_class5					-1.8814					-7.4435	-16.506
					(2.868)					(3.81)	(3.012)
tech_class6					13.9012					5.8043	9.2763
					(4.127)					(5.339)	(4.588)
labor_market			-0.6018	-1.531	-0.858	-1.5634	-1.022	-1.6007	-0.1313		-0.8388
			(0.674)	(0.646)	(0.629)	(0.627)	(0.625)	(0.641)	(0.657)		(0.56)
housing			-3.946	-3.6336	-2.6666	-3.7265	-2.8401	-3.5209	-5.4671		-3.411
			(0.759)	(0.709)	(0.706)	(0.695)	(0.671)	(0.705)	(0.743)		(0.863)
consumption			-1.8297	-2.8254	-1.1644	-2.4821	-1.7465	-2.7553	-1.4257		-1.3906
			(0.748)	(0.716)	(0.747)	(0.698)	(0.744)	(0.71)	(0.719)		(0.642)
price			7.0535	5.4527	6.3856	5.5851	6.168	5.3384	4.844		5.6606
			(0.499)	(0.528)	(0.527)	(0.514)	(0.508)	(0.526)	(0.51)		(0.548)
stock			-1.4141	-3.4082	-1.1105	-2.3713	-0.6664	-3.2912	-1.2123		0.9057
			(1.044)	(1.023)	(1.078)	(1.024)	(1.094)	(1.016)	(1.04)		(0.963)
PYD			1.4527	0.4191	-3.4127	-0.1654	2.5843	0.799	0.4339		-2.6085
			(0.43)	(0.433)	(0.86)	(0.441)	(0.538)	(0.459)	(0.409)		(1.081)
Industries			-0.8172	-0.6601	-0.2274	-0.8259	-3.3686	-0.4833	-0.7157		4.5663
			(1.202)	(1.122)	(1.07)	(1.089)	(1.354)	(1.116)	(1.06)		(1.562)
R-squared	0.000	0.252	0.529	0.592	0.638	0.619	0.654	0.6	0.637	0.484	0.755
Adj. R-squared	0.000	0.25	0.517	0.579	0.622	0.605	0.64	0.586	0.625	0.464	0.737

Table 14 In-Sample Regression (3-year maturity)

	EH	CP	NG	CP+NG	CP+NG+Tech	CP+NG+SENT	CP+NG+Fama	CP+NG+BV	CP+NG+CLI	技术+投资者情绪+Fama+波动率+CLI	全部
β0	1.4751										
	(0.190)										
CLI									-6.3528	-5.3094	-6.784
									(1.224)	(1.201)	(1.688)
MktRF							-1.6556			9.5765	12.5104
							(1.367)			(1.166)	(2.021)
HML							2.9614			7.5981	4.7862
							(1.286)			(1.361)	(1.155)
RF							-8.0318			-11.2454	-10.0901
							(1.301)			(1.372)	(1.396)
BWSENTcoth						-13.8271				-17.4235	-16.3459
						(3.469)				(4.014)	(2.913)
BWSENT						10.1646				17.1643	11.742
						(3.025)				(3.497)	(2.6)
BV								-2.465		-2.8707	-2.7348
								(1.171)		(1.411)	(1.043)
CP		2.7725		7.9201	8.3195	8.2972	4.4467	7.6808	10.1299		2.6064
		(0.347)		(1.427)	(1.404)	(1.476)	(1.579)	(1.422)	(1.427)		(1.609)
tech_class4					-10.6123					3.6677	2.3321
					(3.758)					(4.56)	(3.959)
tech_class5					-1.5522					-16.0455	-32.5069
					(5.621)					(7.276)	(5.775)
tech_class6					24.2458					8.5579	17.3224
					(8.089)					(10.195)	(8.798)
labor_market			-0.8449	-2.4081	-1.2692	-2.5084	-1.8449	-2.5303	0.0655		-1.5213
			(1.283)	(1.25)	(1.233)	(1.216)	(1.21)	(1.243)	(1.285)		(1.074)
housing			-7.5934	-7.0678	-5.4823	-7.362	-5.6071	-6.8701	-10.3079		-7.072
			(1.444)	(1.373)	(1.384)	(1.349)	(1.298)	(1.368)	(1.452)		(1.654)
consumption			-3.4278	-5.1029	-2.1986	-4.4606	-3.7607	-4.98	-2.6293		-2.9763
			(1.425)	(1.385)	(1.464)	(1.355)	(1.439)	(1.377)	(1.405)		(1.231)
price			13.3366	10.6434	12.2392	10.92	11.7913	10.4431	9.5678		10.7909
			(0.949)	(1.023)	(1.032)	(0.997)	(0.983)	(1.021)	(0.998)		(1.051)
stock			-3.602	-6.9567	-2.9672	-4.9367	-2.5864	-6.7517	-3.0761		0.6721
			(1.988)	(1.98)	(2.114)	(1.986)	(2.115)	(1.97)	(2.033)		(1.847)
PYD			2.1593	0.4205	-6.2016	-0.6724	4.9506	1.0864	0.4466		-6.1427
			(0.819)	(0.837)	(1.685)	(0.855)	(1.04)	(0.89)	(0.8)		(2.074)
Industries			-0.599	-0.3348	0.4642	-0.6043	-3.9068	-0.0248	-0.433		12.128
			(2.289)	(2.172)	(2.097)	(2.113)	(2.619)	(2.163)	(2.074)		(2.996)
R-squared	0.000	0.189	0.502	0.553	0.593	0.581	0.622	0.561	0.595	0.45	0.737
Adj. R-squared	0.000	0.186	0.489	0.54	0.577	0.566	0.607	0.546	0.581	0.43	0.718

Table 15 In-Sample Regression (4-year maturity)

	EH	CP	NG	CP+NG	CP+NG+Tech	CP+NG+SENT	CP+NG+Fama	CP+NG+BV	CP+NG+CLI	技术+投资者情绪+Fama+波动率+CLI	全部
β0	1.9774										
	(0.266)										
CLI									-8.4085	-7.2848	-10.4598
									(1.755)	(1.702)	(2.382)
MktRF							-1.4113			13.3907	19.5893
							(1.931)			(1.652)	(2.853)
HML							4.8929			11.1187	7.682
							(1.817)			(1.929)	(1.631)
RF							-12.2493			-16.3036	-14.5959
							(1.838)			(1.945)	(1.97)
BWSENTcoth						-16.872				-22.1914	-21.4344
						(4.967)				(5.688)	(4.11)
BWSENT						11.5246				21.4465	14.6518
						(4.331)				(4.956)	(3.669)
BV								-3.7401		-3.8995	-3.8228
								(1.665)		(2)	(1.472)
CP		3.5962		9.964	10.8053	10.8791	4.23	9.6009	12.8888		2.59
		(0.489)		(2.032)	(2.01)	(2.113)	(2.231)	(2.023)	(2.046)		(2.27)
tech_class4					-13.7906					7.3475	4.3512
					(5.379)					(6.463)	(5.587)
tech_class5					-1.7658					-25.216	-47.9334
					(8.047)					(10.311)	(8.15)
tech_class6					25.7138					5.5559	16.7567
					(11.58)					(14.448)	(12.416)
labor_market			-12.2818	-2.859	-1.5314	-3.0413	-2.3102	-3.0445	0.4148		-1.9873
			(1.807)	(1.78)	(1.764)	(1.742)	(1.709)	(1.768)	(1.842)		(1.516)
housing			-11.2939	-10.6327	-8.6475	-11.1719	-8.5649	-10.3328	-14.9212		-11.3893
			(2.033)	(1.956)	(1.981)	(1.932)	(1.833)	(1.946)	(2.082)		(2.334)
consumption			-4.9966	-7.104	-2.9909	-6.284	-5.638	-6.9176	-3.83		-4.3339
			(2.006)	(1.973)	(2.095)	(1.94)	(2.033)	(1.96)	(2.015)		(1.738)
price			18.5716	15.1835	17.3134	15.5756	16.7187	14.8795	13.7598		15.0526
			(1.336)	(1.457)	(1.478)	(1.428)	(1.388)	(1.452)	(1.431)		(1.483)
stock			-5.1533	-9.3738	-3.5371	-6.6866	-3.6089	-9.0627	-4.2375		1.3679
			(2.799)	(2.82)	(3.026)	(2.844)	(2.988)	(2.802)	(2.915)		(2.607)
PYD			3.078	0.8904	-8.6526	-0.5036	7.7414	1.9008	0.9249		-9.6129
			(1.153)	(1.193)	(2.412)	(1.224)	(1.469)	(1.266)	(1.147)		(2.926)
Industries			-0.7268	-0.3944	0.7897	-0.6832	-4.7217	0.0759	-0.5243		19.1873
			(3.223)	(3.093)	(3.002)	(3.026)	(3.699)	(3.077)	(2.974)		(4.228)
R-squared	0.000	0.165	0.49	0.532	0.57	0.557	0.611	0.541	0.569	0.43	0.729
Adj. R-squared	0.000	0.162	0.477	0.518	0.552	0.54	0.595	0.525	0.555	0.408	0.71

Table 16 In-Sample Regression (5-year maturity)

	EH	CP	NG	CP+NG	CP+NG+Tech	CP+NG+SENT	CP+NG+Fama	CP+NG+BV	CP+NG+CLI	技术+投资者情绪+Fama+波动率+CLI	全部
β0	2.1831										
	(0.328)										
CLI									-9.4884	-8.997	-13.4549
									(2.229)	(2.127)	(3.039)
MktRF							-0.3722			16.1149	26.1995
							(2.445)			(2.064)	(3.64)
HML							6.7906			13.4896	10.1276
							(2.301)			(2.41)	(2.081)
RF							-15.2468			-19.62	-17.9552
							(2.327)			(2.43)	(2.514)
BWSENTcoth						-19.0454				-25.6606	-25.6997
						(6.291)				(7.108)	(5.244)
BWSENT						13.0282				24.6309	17.708
						(5.486)				(6.193)	(4.682)
BV								-5.0268		-5.2704	-5.2391
								(2.095)		(2.499)	(1.878)
CP		3.83		10.5083	11.5664	11.5314	2.7782	10.0203	13.8087		0.8576
		(0.608)		(2.56)	(2.553)	(2.676)	(2.825)	(2.545)	(2.6)		(2.896)
tech_class4					-14.7444					11.0979	5.8143
					(6.831)					(8.076)	(7.128)
tech_class5					-3.2233					-33.8255	-60.1157
					(10.218)					(12.884)	(10.399)
tech_class6					26.6545					3.3701	17.6712
					(14.704)					(18.054)	(15.842)
labor_market			-0.9371	-3.0111	-1.5561	-3.2156	-2.7249	-3.2604	0.6832		-2.3478
			(2.248)	(2.242)	(2.24)	(2.206)	(2.164)	(2.225)	(2.34)		(1.935)
housing			-14.2963	-13.5989	-11.3316	-14.2036	-11.2444	-13.1958	-18.4382		-14.9812
			(2.529)	(2.464)	(2.515)	(2.447)	(2.321)	(2.448)	(2.645)		(2.979)
consumption			-6.5976	-8.8202	-4.1169	-7.8953	-7.7479	-8.5696	-5.1257		-6.0599
			(2.496)	(2.485)	(2.661)	(2.457)	(2.575)	(2.465)	(2.56)		(2.217)
price			22.5866	19.0134	21.4342	19.4548	20.6577	18.6048	17.4068		18.3555
			(1.663)	(1.835)	(1.877)	(1.809)	(1.758)	(1.827)	(1.818)		(1.892)
stock			-6.6471	-11.0981	-4.3416	-8.0696	-5.0857	-10.68	-5.3021		1.1236
			(3.481)	(3.552)	(3.842)	(3.602)	(3.784)	(3.525)	(3.703)		(3.326)
PYD			3.825	1.5178	-9.5081	-0.0544	9.98	2.8759	1.5569		-11.9692
			(1.434)	(1.502)	(3.063)	(1.55)	(1.861)	(1.593)	(1.457)		(3.734)
Industries			-0.5153	-0.1647	1.1736	-0.4916	-4.2722	0.4674	-0.3113		26.1787
			(4.009)	(3.897)	(3.812)	(3.832)	(4.685)	(3.871)	(3.778)		(5.394)
R-squared	0.000	0.126	0.466	0.497	0.53	0.519	0.578	0.508	0.53	0.397	0.701
Adj. R-squared	0.000	0.123	0.452	0.482	0.511	0.5	0.56	0.491	0.514	0.375	0.68

Table 17 Wald Tests of The In-Sample Regression Model

	Fama		InvestorSENT		tech		macro		financial	
	χ^2	p-values	χ^2	p-values	χ^2	p-values	χ^2	p-values	χ^2	p-values
2 year	24.37	0.00	16.53	0.00	16.96	0.00	34.67	0.00	4.892	0.01
3 year	31.45	0.00	17.77	0.00	16.53	0.00	35.88	0.00	8.808	0.00
4 year	36.02	0.00	16.56	0.00	17.52	0.00	37.45	0.00	11.02	0.00
5 year	36.52	0.00	14.43	0.00	16.39	0.00	36.01	0.00	12.21	0.00

Table 18 Out-of-Sample Forecasting Results For 2-5 Year Maturity Bonds, 2003.1-2009.8

	2-year maturity	3-year maturity	4-year maturity	5-year maturity
RMSE	3.0953	9.149	13.864	19.233
R^2_{OOS}	-3.312	-10.354	-13.156	-18.530

Table 19

FACTOR	COEF	STDERR
CLI	-2.391	1.284
MKTRF	9.7207	1.908
HML	3.429	1.161
RF	-8.8048	1.406
BWSENTCOTH	-15.1244	3.244
BWSENT	13.0295	2.89
BV	-3.0613	1.161
CP	5.455	1.572
TECH_CLASS4	-7.5612	2.226
TECH_CLASS5	-4.5933	2.998
TECH_CLASS6	15.1657	2.251
LABOR_MARKET	-2.1784	1.155
HOUSING	-8.6948	1.318
CONSUMPTION	-4.3879	1.225
PRICE	9.3894	0.917
STOCK	-7.5944	1.927
PYD	-2.9606	1.748
INDUSTRIES	11.0065	2.838

Table 20

Maturity	R^2 （Adjusted）	Factor Loading
2	0.688	0.4728
3	0.683	0.8788
4	0.669	1.2109
5	0.638	1.4375

Table 21

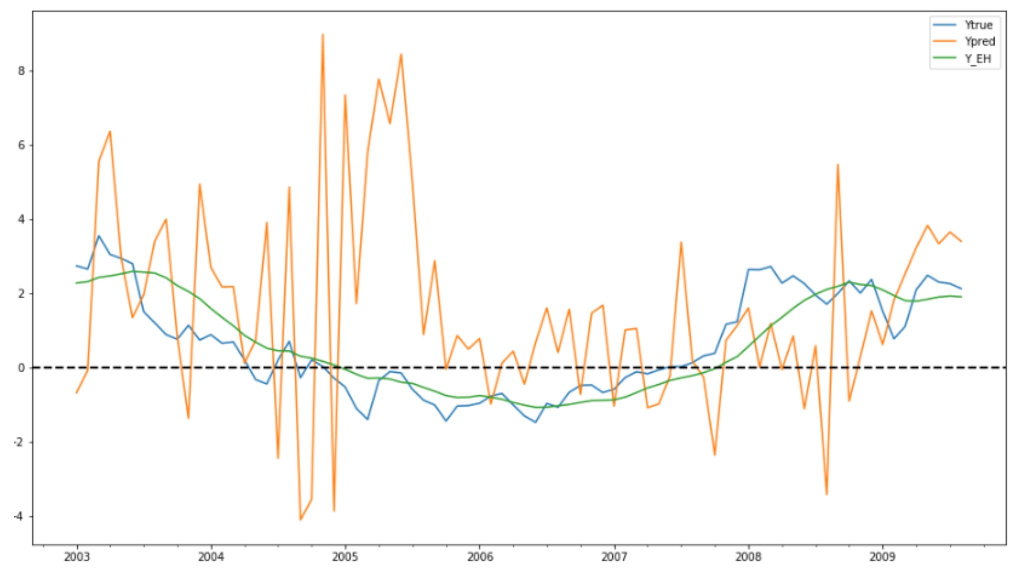
Maturity	2		3		4		5	
R^2	0.826		0.826		0.813		0.790	
COlumn	COEF	STDERR	COEF	STDERR	COEF	STDERR	COEF	STDERR
CLI	-3.5746	1.467	-7.1013	2.714	-11.3735	3.875	-15.4052	5.06
MktRF	8.323	1.889	19.1128	3.496	29.3672	4.991	39.0912	6.517
HML	1.0676	0.805	2.3994	1.49	3.6019	2.127	4.5609	2.777
RF	-3.9103	1.174	-8.7971	2.174	-12.627	3.103	-15.3317	4.052
BWSENTcoth	-6.8329	2.115	-13.5194	3.914	-17.4342	5.588	-23.6155	7.295
BWSENT	4.9236	1.878	9.4366	3.475	11.5911	4.962	16.2967	6.479
BV	-1.511	0.756	-3.2926	1.4	-4.8685	1.999	-5.7054	2.61
CP	2.4251	1.001	2.7611	1.852	1.9266	2.645	0.7505	3.453
tech_class4	1.6635	3.576	0.7089	6.617	0.5339	9.448	-0.9998	12.335
tech_class5	-19.7813	5.091	-41.2517	9.421	-60.4872	13.451	-76.4787	17.562
tech_class6	8.9072	5.941	20.3631	10.995	25.5772	15.698	30.6845	20.496
labor_market	-1.5532	0.943	-3.0317	1.745	-3.9782	2.491	-3.8927	3.253
housing	-3.9015	1.419	-7.7978	2.626	-11.9063	3.749	-15.7552	4.895
consumption	-0.1595	0.608	-0.3408	1.124	-0.1195	1.605	0.0788	2.096
price	4.269	0.674	8.6094	1.247	12.4856	1.781	15.6162	2.325
stock	-2.4295	1.9	-3.4409	3.516	-3.9249	5.02	-4.656	6.554
PYD	-3.9424	2.115	-10.5276	3.914	-16.4301	5.588	-22.8342	7.296
Industries	10.3259	2.915	23.263	5.394	35.2183	7.701	46.0612	10.055

Table 22

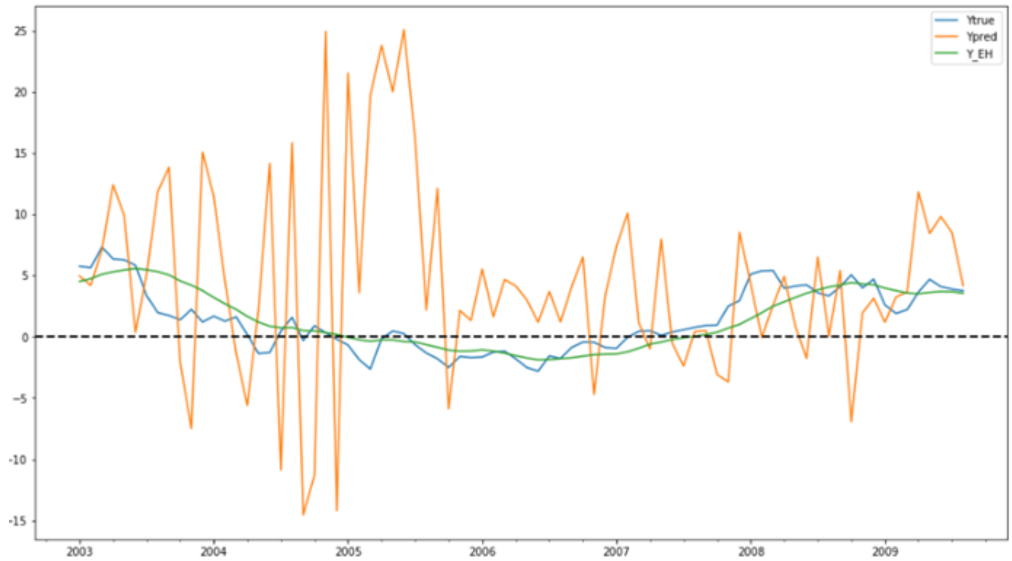
Maturity	With Crisis	Without crisis
2	0.697	0.763
3	0.671	0.746
4	0.652	0.730
5	0.626	0.695

Figure 1

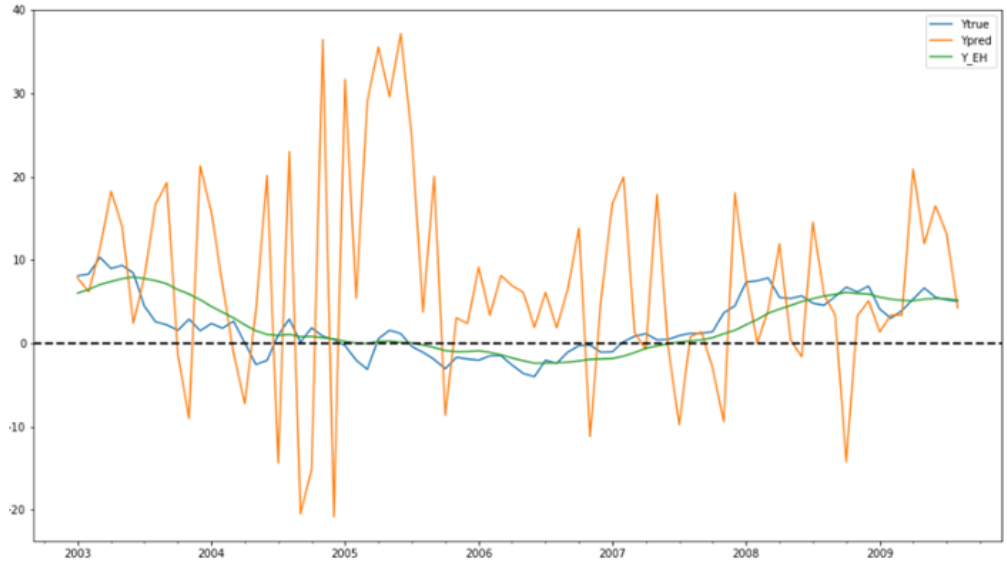
2-year maturity



3-year maturity



4-year maturity



5-year maturity

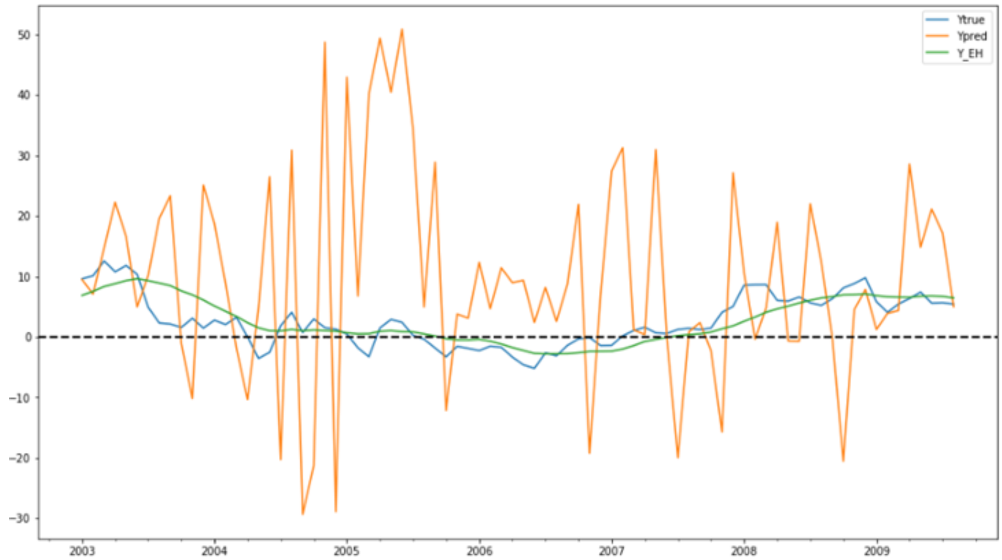
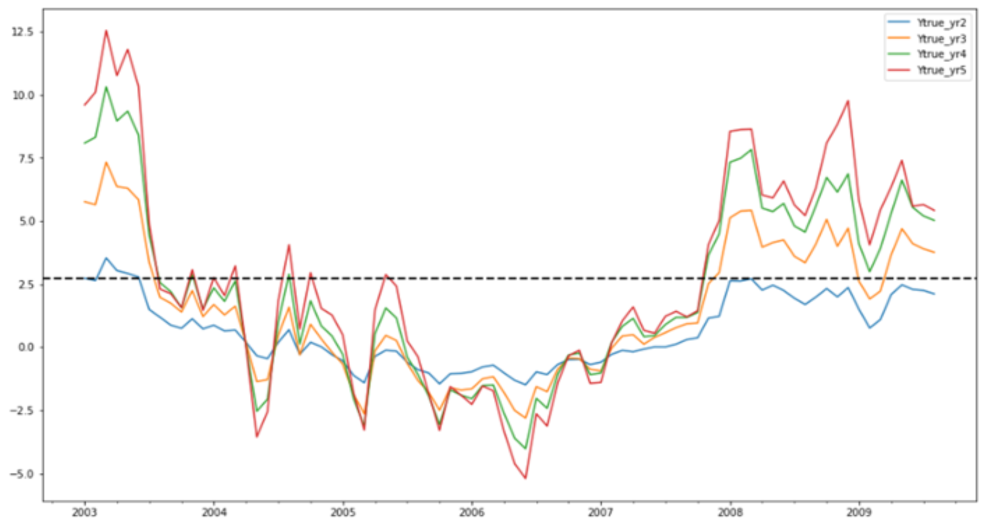
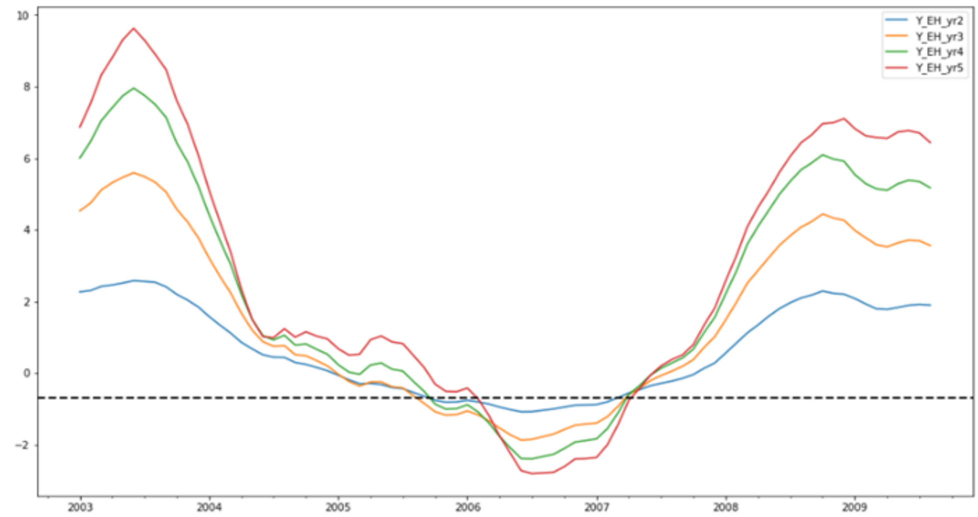


Figure 2

2-5year maturity




Y_EH_yr2
Y_EH_yr3
Y_EH_yr4
Y_EH_yr5
10
8
6
4
2
0
−2
2003
2004
2005
2006
2007
2008
2009


Ypred_yr2
Ypred_yr3
Ypred_yr4
Ypred_yr5
50
40
30
20
10
0
−10
−20
−30
2003
2004
2005
2006
2007
2008
2009


R2_yr2
R2_yr3
R2_yr4
R2_yr5
0.80
0.75
0.70
0.65
0.60
2003
2004
2005
2006
2007
2008
2009

Figure 3 Maturity = 2, R^2_{OOS} = -1.32

Figure 4 Maturity = 3, R^2_{OOS} = -2.37

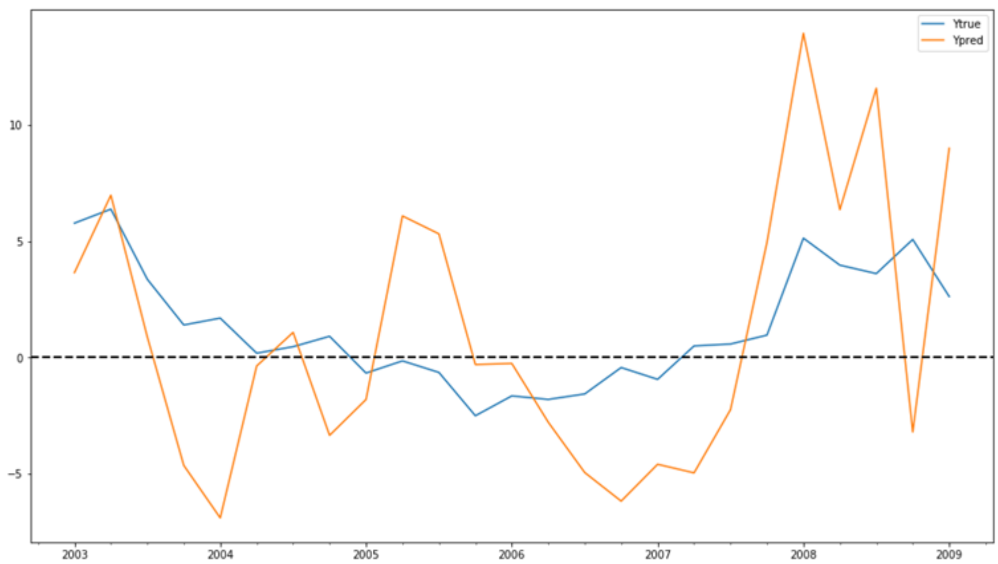


Figure 5 VIX indicator

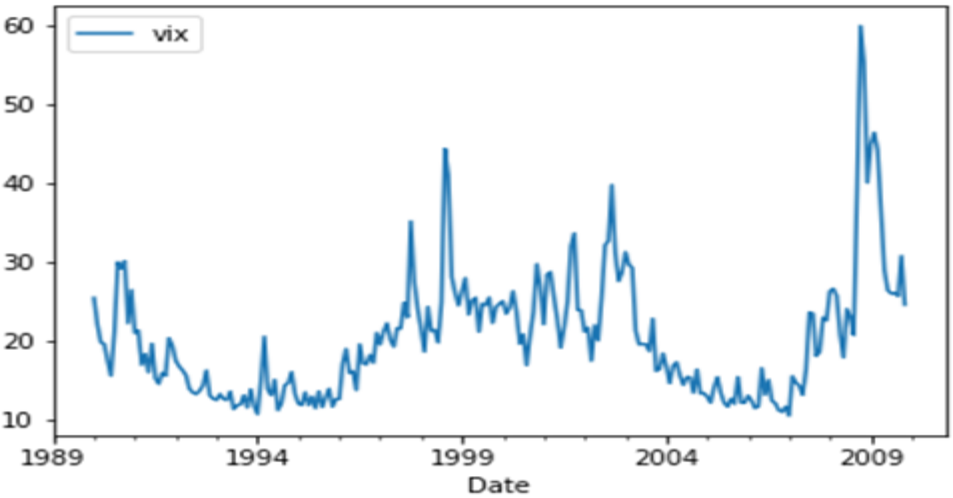
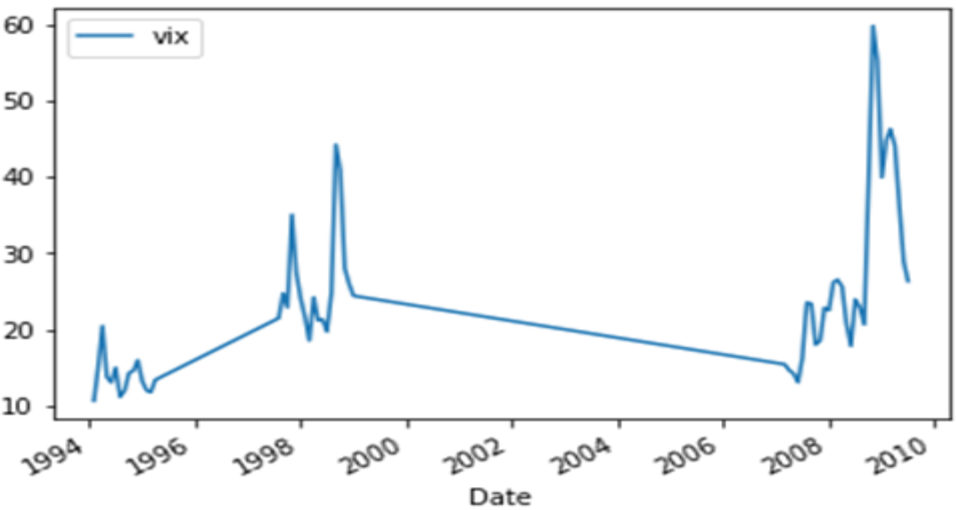


Figure 6 VIX in financial crisis



Reference

John h. Cochrane and Monika Piazzesi.,"Bond Risk Premia", 2005

Eugene F. Fama and Robert R. Bliss ,"The Information in Long-Maturity Forward Rates", 1987

Sydney C. Ludvigson and Serena Ng,"Macro Factors in Bond Risk Premia", 2009

Jeremy Goh, Fuwei Jiang, Jun Tu and Guofu Zhou, "Forecasting Government Bond Risk Premia Using Technical Indicators".2013

Hao Zhou, "Variance Risk Premia, Asset Predictability Puzzles, and Macroeconomic Uncertainty",2012

Lin H, Wang J, Wu C ,"Liquidity risk and expected corporate bond returns",Journal of Financial Economics,2011

Amihud Y. ,"Illiquidity and stock returns: cross-section and time-series effects", Journal of financial markets, 2002

Pástor Ľ, Stambaugh R F., "Liquidity risk and expected stock returns", Journal of Political economy, 2003

Wright J H, Zhou H., "Bond risk premia and realized jump risk", Journal of Banking & Finance, 2009