Fixed Income： Predicting bond excess return

Multi-dimensional Selection Factor Model

Team members ：

Xugan Chen, Yufeng Hu, Wencheng Chen, Jiayi Wang

Wutong Zhouxin, Chen Liu, Yinglu Shi, Yixuan Fang

Contents

Multiple Dimensions & Higher R^2

1 Multi-dimensional Factor

Construct our factors in **multiple classes**:

- Technical factors
- Macroeconomic & Financial factors
- Composite Leading factors
- Market factors
- Sentiment factors
- Stock model factors
- Classical model factors

2 Multi-selection Processing

Process data in **two dimensions**:

- Class-within PCA: **One factor** from each class ➡ contribute to reveal economic meanings
- Backward: remaining class-within factors filtered by **significance** ➡ process selection

3 Robustness and advantages

Robust in **different aspects:**

- Good empirical results for other frequencies
- Not fastidious about time to maturities
- Test for restricted model and unrestricted model then run better for the later
- Better results when ruling out crisis

Higher R^2 :

- Moderate factors combined with a higher R^2 than the classical models.

Choosing and Filtering Factors

1 Choosing Data

Factors	Data Source
Liquidity indicators	S&P monthly data
Fama-French 5	http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/f-f_5_factors_2x3.html
Financial indicators	Raw data given
Macroeconomic indicators	Raw data given
Technical indicators	Yahoo Finance
CP indicators	Raw data given
Composite leading indicators	http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/f-f_5_factors_2x3.html
Investors' sentiment factors	http://people.stern.nyu.edu/jwurgler/
Market jump risk measure	https://www.investing.com

2 PCA & Backward

- For **financial factor, technical factor and macroeconomic factor**, each contains several different classes and consists of more than 100 factors.
- **Choose only one factor** from one class by PCA to eliminate **multicollinearity problem.**

Regression model:

$$Y = X\beta + \varepsilon \qquad t_k = \frac{b_k}{\left[s^2 (X'X)^{-1}_{kk}\right]^{-\frac{1}{2}}}$$

$$S^2 = \frac{e'e}{n-k} \qquad p_k = 2P(t(n-k) > |t_k|)$$

Step1: do regression and calculate p-value

Step2: eliminate the factor which has the **biggest p-value**

Step3: repeat until $(P_i = 2P(t(n-k) > |t_i|)) \leq 0.1$ for each i

Step4: do regression with the remaining factors

Prediction: In-sample & Out-of-sample

1 In-sample analysis

- 2-5 year maturity, 1980.01-2002.12

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

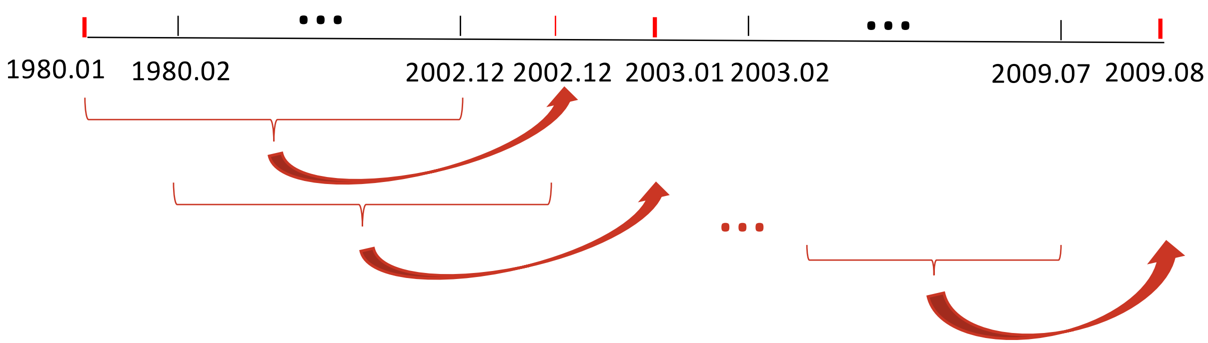
  → To select the best combinations of the factors

- Test: Wald statistics of large samples

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

2 Out-of-sample analysis

- **Rolling window**



Use the observations of a certain period of time to predict next month's bond risk premium.

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

- **Predictive ability**
  - R^2 (coefficient of determination)

    $$R^2 = \frac{\text{SSR}}{\text{SST}} = 1 - \frac{SSE}{SST} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2}$$

  - RMSE(root-mean-square error)

    $$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

  - R_{OOS}^2

    $$R_{OOS}^2 = 1 - \frac{\sum_{i=1}^{n}(y^{(n)} - \hat{y}^{(n)})^2}{\sum_{i=1}^{n}(y^{(n)} - \bar{y}^{(n)})^2}$$

→ The closer R^2 is to 1, the smaller RMSE is, the better predictive effect our model has.
A $R_{OOS}^2 > 0$ indicates a higher predictive accuracy.

Whole sample regression

1 Backward selection

- We use **Backward Selection** based on **p-value** to eliminate 22 less influential factors and **get 18 factors remained**.

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

Results of regression:

- 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).**
- As maturity increase, the explanatory abilities decrease which are consistent with the results in CP.

Some drawbacks:

1. excessive number of factors
2. have not go through the term structure model as CP does

2 Whole sample regression

	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

In-sample regression

1 In-sample regression

- 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.

$$y_{t+1}^{(n)} = \beta_{1t}x_{1t} + \beta_{2t}x_{2t} \dots + \beta_{18t}x_{18t} + \varepsilon_{t+1}$$

R-squared	2y	3y	4y	5y
CP	0.252	0.189	0.165	0.126
NG	0.529	0.502	0.490	0.466
CP+NG	0.592	0.553	0.532	0.497
CP+NG+Tech	0.638	0.593	0.570	0.530
CP+NG+SENT	0.619	0.581	0.557	0.519
CP+NG+Fama	0.654	0.622	0.611	0.578
CP+NG+BV	0.600	0.561	0.541	0.508
CP+NG+CLI	0.637	0.595	0.569	0.530
Te+SE+BV+CLI	0.484	0.450	0.430	0.397
Total	0.755	0.737	0.729	0.701

- We find that the model with the whole 18 variables shows the highest adjusted R-squared.

2 Wald Test of in-sample

χ^2	Fama	SENT	tech	macro	finan
2y	24.37 ***	16.53 ***	16.96 ***	34.67 ***	4.89 ***
3y	31.45 ***	17.77 ***	16.53 ***	35.88 ***	8.81 ***
4y	36.02 ***	16.56 ***	17.52 ***	37.45 ***	11.02 ***
5y	36.52 ***	14.43 ***	16.39 ***	36.01 ***	12.21 ***

- Compute factor model on Fama, sentiments, technology, micro and financial.
- Both two statistics convincingly **reject the H_0 hypothesis** which strengthen the feasibility of in-sample regression model.

3 Expectation Hypothesis

- R-squared for EH model approach zero
- It indicates EH **does not hold** in our research
- (main reason) We ignore term structure model

Out-of-sample regression

1 out of sample regression

- we use the **Rolling Window** to do the out-of-sample predicting regression with 18 variables

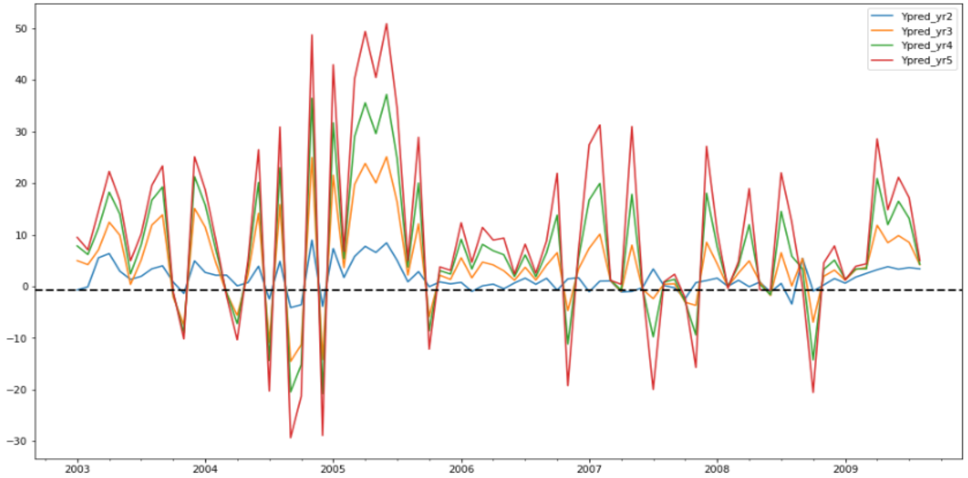
$$y_t^{(n)} = \beta_{1t-1}x_{1t-1} + \beta_{2t-1}x_{2t-1} \dots + \beta_{18t-1}x_{18t-1} + \varepsilon_t$$

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

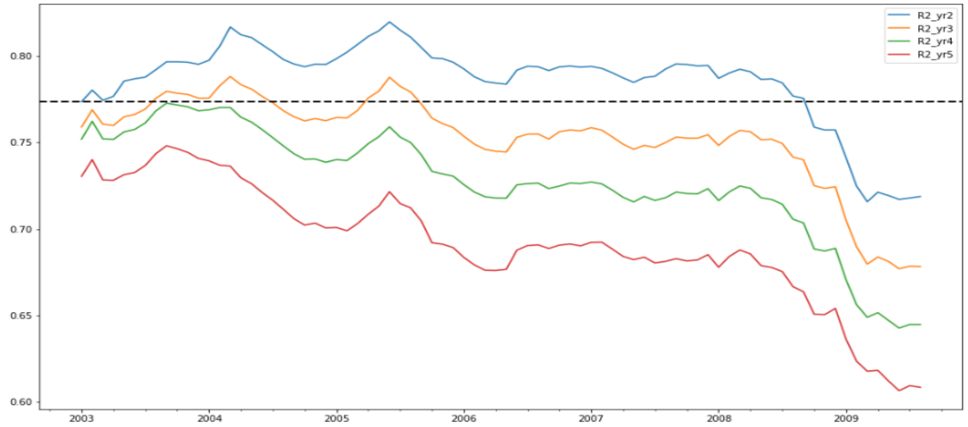
2 R^2_{OOS}、 $RMSE$ analysis

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

- 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.



The predicted excess return curve deviate more from true curve as maturity increases



The R-squared get smaller in longer-term bond

Four tests for our factor

1 Restricted Regression

- **Step 1:** $\overline{rx_t} = \beta X + \epsilon,$
  - Denote $\gamma = \beta X$ the new factor.
- **Step 2:** $rx_t^i = b_i \gamma + \epsilon$

Maturity	$\bar{R}^2$	Factor loading
2	0.688	0.4728
3	0.683	0.8788
4	0.669	1.2109
5	0.638	1.4375

- The results are slightly better. Proves the robustness of our factor

2 Three-month Frequency

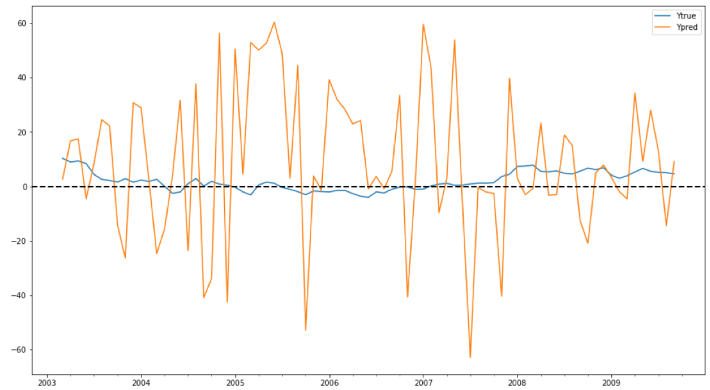
- Resample the given monthly data to three month frequency.
- The result are significantly better.
- Why? Lack of sample? Better cyclic behavior?
- Prediction: precisely caught trends.
  - Figure: Real (blue) and prediction (yellow) result

Maturity	2	3	4	5
R^2	0.826	0.826	0.813	0.790

Four tests for our factor

3 Longer forecast?

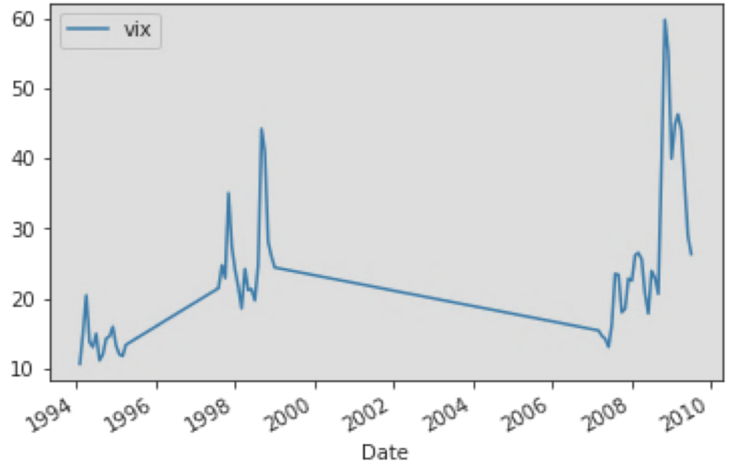
- Test the longer-run predictability of our factor.
- Shift y three months ahead.
- The results are bad. Market is stochastic.



4 Financial Crisis

- Three major financial crisis.
  - 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
- Figure: VIX index during financial crisis.
- We exclude financial crisis in our regression.
- As our factor is robust, the result is much better.
- Table: R2 statistics for comparison.

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



Thank you !

Team members：

Xugan Chen, Yufeng Hu, Wencheng Chen, Jiayi Wang

Wutong Zhouxin, Chen Liu, Yinglu Shi, Yixuan Fang