Essays on Return Predictability and Yield Factors

Xuyang Ma

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2014

Reading Committee:

Eric Zivot, Chair

Charles R. Nelson, Chair

Chang-Jin Kim

Program Authorized to Offer Degree:
Economics

University of Washington

Abstract

Essays on Return Predictability and Yield Factors

Xuyang Ma

Co-Chairs of the Supervisory Committee:
Professor Eric Zivot
Department of Economics

Professor Charles R. Nelson
Department of Economics

This dissertation includes three chapters in which the first two are on return predictability and the third is on yield curve and yield factors. The abstract of each of them is as follows:

- This paper proposes using capital gains instead of total returns in return predictability tests. Total return predictability can be inferred from capital gain predictability since total returns with dividends are highly correlated with returns based on capital gains only. An exact linear relationship exists among log dividend growth, log capital gain and log dividend price ratio. This exact linear relationship has similar implication as the Campbell-Shiller (1988) linear approximation but is more precise and easier for predictability tests. I verify the standard empirical findings on return predictability using capital gain predictability. Separation of price change and dividend change also leads to a new finding: shocks to dividend growth is shown to have significant positive correlation with shocks to dividend price ratio in the vector autoregressive regression (VAR) rather than close to zero as shown in previous literature.

- This paper tests the return predictability of the cyclical and trend components in the log dividend price ratio. The log dividend ratio is found to have a near-unit root trend factor if the expectation of the future discount factor is highly persistent. We

use Bayesian analysis and the Kalman filter to extract the strictly stationary and near-random-walk components in the log dividend price ratio. The extracted cyclical process can predict one-year ahead total returns during the post-war period and one-year ahead dividend growth rates during the pre-war and war period with notable R^2. We also demonstrate a reverse of predictability: returns become more predictable while dividend growth rates become more unpredictable.

- This paper examines the fourth principal component of the yields matrix, which is largely ignored in macro-finance forecasting applications, in the context of predicting excess bond returns. Using yields data from the Fama-Bliss and the Federal Reserve, we present the significant in-sample and out-of-sample predictive power of models including the fourth yield factor. Additionally, the "return-forecasting factor" in [7] is shown to be a restricted linear combination of all yield factors and to be highly correlated with the second and fourth factors. We interpret the fourth yield factor as a factor representing "$S - shape$" (the shape of a sigmoid curve) and demonstrate the connection between the $S - shape$ factor and the yield curve.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGMENTS

The author wishes to express sincere appreciation to Department of Economics, University of Washington, where she has had the opportunity to work on research and get help and support from the faculty members.

DEDICATION

To my dear mom and dad

Chapter 1

CAPITAL GAINS FOR RETURN PREDICTABILITY TESTS

1.1 Introduction

The theory of return predictability using dividend price ratio starts from the Campbell-Shiller (1988) log linearization of total returns. This linearization is just an approximation. It connects log dividend price ratio, log dividend growth and log total returns together and serves as one fundamental theoretical support for return predictability using log dividend price ratio. In this paper, I suggest using capital gains instead of total returns when dealing with return predictability tests. My approach allows separation of predictability tests on price change and dividend change. Testing these two changes separately, we can see where the total return predictability comes from: from dividend or price or both.

Total returns have high positive correlations with both capital gains and dividend growth rates, since total returns contain these two terms. If predictability of capital gains or dividend growth rates is found, it is likely that predictability of total returns can also be found because of the high correlations but we cannot infer where the predictability of total returns comes from. On the contrary, capital gains returns do not have a significant correlation with dividend growth rates. With this separation, the question whether dividend price ratio predicts price change or dividend change would become clear to test.

Cochrane (2008) runs a vector autoregressive regression (VAR) using log total returns, log dividend growth rates and log dividend price ratio as dependent variables to test the predictability using lagged log dividend price ratio as indicators. He derives an approximate linear relationship among coefficients of these three regression equations and makes this linear relationship the theoretical support for the total return predictability. In this paper, I run a similar VAR except that I change one dependent variable — log total returns into log capital gains. With this change, I am able to derive an exact linear relationship among coefficients of three regression equations. This exact linear relationship provides theoretical support for capital gains returns predictability. Since total returns and capital gains are highly correlated, total returns predictability is supported too.

With log capital gains serving as new dependent variables, I find shocks to log dividend growth rates have significant positive correlation with shocks to log dividend price ratio while Cochrane (2008) finds there is no correlation between these two shocks. As mentioned, total returns have high correlations with dividend growth rates while capital gains do not have significant correlations with dividend growth rates, the new finding on correlations among shocks is notable and worthy attention.

Using capital gains instead of total returns, I verify previous empirical findings on return predictability. One important empirical finding is the reversal of predictability in Chen (2009). He finds pre-war data predicts dividend growth rates but not total returns while post-war data predicts total returns but not dividend growth rates. Another important empirical finding is short horizon return predictability is only marginally significant while longer horizon regression shows stronger evidence of return predictability using lagged dividend price ratio, in the sense that the R-square increases as the horizon increases. I verify these findings using capital gains instead of total returns. Thus I prove that, empirically, capital gains tests can serve as well as total returns tests.

There are also papers questioning whether returns are predictable at all. Among others, Welch and Goyal (2008) find the out-of-sample predictabilities of a large amount of predictors including dividend price ratio, book value, corporate issuing activity, treasury bills, long term yield, inflation, investment to capital ratio and consumption wealth ratio are poor and then conclude that return is best predicted by history average but not financial ratios. This will not be the focus of this paper. I do not intend to argue whether return is predictable or not, neither do I intend to test which variable can better predict returns. This paper is just to emphasize that capital gains are good indicators for return predictability tests and are better indicators than total returns.

The rest of this paper is organized as follows. Section 2 presents theoretical analysis with a comparison with Campbell-Shiller (1988) linearization. Section 3 provides the empirical results with discussion. Section 4 concludes. Reference follows in the end.

1.2 Theory: Capital Gains

Capital gains returns are returns without dividend or returns on prices change only. They are widely analyzed in the literature as one type of returns. For example, Campbell and Shiller (1988) develop their linearization theory using total returns but consider four kinds of returns including capital gains returns; Fama and French (1988) research on permanent and temporary components of stock prices and denote the continuously compounded returns as capital gains; Cochrane (2008) states that It is quite robust fact that return variation is dominated by variation of prices or valuations with little change in cash flows, and more so at high frequencies.

Besides the fact that capital gains are widely used in previous literature, capital gains can be considered as good indicators or even equivalent variables with total returns when deriving the property of returns as they are highly correlated. In this section, I show reasons why using returns without dividend when running predictive regressions is reliable and even preferred.

1.2.1 High Correlation between Total Returns and Capital Gains

Log total returns are nearly perfectly correlated with log capital gains, which makes it safe to use capital gains predictability to infer total returns predictability. Consider the formula for total returns:

$$R_t = \frac{P_t + D_t}{P_{t-1}} \ .$$

Extract the capital gain on the right hand side, we have:

$$R_t = \frac{P_t}{P_{t-1}}(1 + \frac{D_t}{P_t}) \ .$$

Take the log of both sides:

$$r_t = \Delta p_t + ln(1 + \frac{D_t}{P_t}) \ ,$$

in which lowercase letters represent log values of corresponding capital letters, Δ represents the change, i.e. the growth rate. r_t is the log total return from time $t-1$ to t; Δp_t is the log capital gain return from time $t-1$ to t.

Since the value of D_t/P_t is close to zero, we can rewrite the above formula as following:

$$r_t \approx \Delta p_t + \frac{D_t}{P_t} .$$

INSERT TABLE. 1.1 NEAR HERE

The statistics of these three variables are shown in table 1. From Table 1, we can see that the log total returns and log capital gains are volatile, while dividend price ratios are much more stable. The correlation between log return and log capital gain is 0.997. Thus, we can imagine that most information contained in log total returns is contained in log capital gains. The co-movement of log total returns and log capital gains suggests a model as following:

$$r_t = \mu + \Delta p_t + \varepsilon_t ,$$

in which mu is a constant, ε_t is an error term with a zero mean.

INSERT FIG. 1.1 NEAR HERE

Figure 1 plots the annual log total returns and annual log capital gains plus a constant using Shiller's data. We can see that movements in log total returns and in log capital gains returns are nearly identical. Thus, to test return predictability, we can test capital gains predictability and infer total returns predictability.

We have shown that it is reliable to use capital gains instead of total returns when dealing with return predictability. Next I provide reasons why return predictability tests using capital gains are preferred.

1.2.2 An Exact Linear Relationship

The most widely used linear relationship when testing return predictability is Campbell-Shiller (1988) linearization. It starts with the identity:

$$R_{t+1} = \frac{P_{t+1} + D_{t+1}}{P_t} = \frac{(1 + P_{t+1}/D_{t+1})D_{t+1}/D_t}{P_t/D_t} .$$

After log-linearizing, it becomes

$$r_{t+1} = log[1 + exp(p_{t+1} - d_{t+1})] + \Delta d_{t+1} - (p_t - d_t) .$$

Approximately, we have

$$r_{t+1} \approx \; const. + \frac{P/D}{1+P/D}(p_{t+1} - d_{t+1}) + \Delta d_{t+1} - (p_t - d_t) \; .$$

where $log(P/D)$ is the point of linearization. After ignoring constant means and orders higher than one, and defining $\rho = (P/D)/(1 + P/D)$, we obtain the following equation:

$$\rho(d_{t+1} - p_{t+1}) \approx \; (d_t - p_t) + (\Delta d_{t+1} - r_{t+1}) \; . \tag{1.1}$$

Most research on return predictability starts from this approximate equation. For example, Cochrane (2008) runs a VAR(1) of log total returns r_{t+1}, log dividend price ratio $d_{t+1} - p_{t+1}$, and log dividend growth Δd_{t+1} to look for theoretical support of return predictability. The VAR(1) takes the following form:

$$\begin{cases} r_{t+1} & = a_r + b_r(d_t - p_t) + \varepsilon^r_{t+1} \\ \Delta d_{t+1} & = a_d + b_d(d_t - p_t) + \varepsilon^{\Delta d}_{t+1} \\ d_{t+1} - p_{t+1} & = a_{dp} + \varphi(d_t - p_t) + \varepsilon^{dp}_{t+1} \; . \end{cases}$$

Inserting those three equations into equation 1.2.3, we collect two identities from equality of the left side and the right side. One is from matching coefficients of dividend price ratio $d_t - p_t$:

$$b_r \approx \; 1 - \rho\varphi + b_d \; . \tag{1.2}$$

The other one is about the relationship among three error terms: ε^r_{t+1}, $\varepsilon^{\Delta d}_{t+1}$ and ε^{dp}_{t+1}:

$$\varepsilon^r_{t+1} \approx \; \varepsilon^{\Delta d}_{t+1} - \rho\varepsilon^{dp}_{t+1} \; .$$

When dealing with implied coefficients from these equations, research takes ρ in this equation as a constant, and assumes its value equals its mean, 0.96. For this approximation to work well, ρ needs to be stable. The approximation works well if ρ behaves like a constant.

Using post-war data, the coefficient b_r in the VAR is around 0.1 and the coefficient b_d is around 0.01, the implied coefficients in equation 1.2 are very sensitive to changes in the value of ρ. Figure 2 plots the value of ρ from 1871 to 2011 where the value varies from 0.9 to 1.0.

INSERT FIG. 1.2 NEAR HERE

The maximum value of ρ in our sample data is around 0.99 and the minimum value is around 0.91. There is a difference of around eight percent between these two possible values of ρ and the instability of ρ makes inferences which base on the assumption of constant ρ unreliable. However, if I take a slightly different approach, this approximation issue can be avoided.

My approach starts with an identity as follows:

$$\frac{D_{t+1}}{D_t} = \frac{D_{t+1}/P_{t+1}}{D_t/P_t} * \frac{P_{t+1}}{P_t} ,$$

then take the log of both sides and move the log dividend price ratio at time $t+1$ to the left side as the dependent variable:

$$(d_{t+1} - p_{t+1}) = (d_t - p_t) + (\Delta d_{t+1} - \Delta p_{t+1}) , \tag{1.3}$$

in which lowercase letters represent the log value of corresponding capital letters; Δ represents the change. So, d_t is the log dividend at time t, p_t is the log price at time t; Δd_{t+1} is log dividend growth from time t to $t+1$; Δp_{t+1} is the log capital gain from time t to $t+1$.

Comparing equation 1.2.3 with equation 1.2.3, we can see that equation 1.2.3 looks almost identical with equation 1.2.3 except that we no longer have to concern about the term ρ when using Δp_{t+1} instead of r_{t+1}. The fact that ρ is close to one also suggests that Δp_{t+1} behaves nearly identical with r_{t+1}.

If I build a VAR similar with Cochrane (2008) using log capital gains instead of log total returns, it takes the following form:

$$\begin{cases} \Delta p_{t+1} & = a_p + b_p(d_t - p_t) + \varepsilon_{t+1}^{\Delta p} \\ \Delta d_{t+1} & = a_d + b_d(d_t - p_t) + \varepsilon_{t+1}^{\Delta d} \\ d_{t+1} - p_{t+1} & = a_{dp} + \varphi(d_t - p_t) + \varepsilon_{t+1}^{dp} \ . \end{cases}$$

Inserting those three equations into equation 1.2.3, we collect two identities from equality of the left side and the right side. One is from matching coefficients of dividend price ratio $d_t - p_t$:

$$b_p = 1 - \varphi + b_d \ . \tag{1.4}$$

The other one is about the relationship among three error terms: $\varepsilon_{t+1}^{\Delta p}$, $\varepsilon_{t+1}^{\Delta d}$ and ε_{t+1}^{dp}:

$$\varepsilon_{t+1}^{\Delta p} = \varepsilon_{t+1}^{\Delta d} - \varepsilon_{t+1}^{dp} \ .$$

Equation 1.3.2 looks almost identical with equation 1.2 except it does not have the ρ term and thus avoids possible problems caused by the assumption of constant ρ, also it is an exact equation rather than an approximation like equation 1.2 because there is no first order approximation involved.

In later section of this paper, I will show that my VAR with gives similar empirical implications with the VAR in Cochrane (2008), and also supplies new implications among error terms.

1.2.3 Nearly Zero Correlation between Capital Gains and Dividend Growth

Another advantage of using log capital gains instead of log total returns is that it avoids confusion caused by high correlation between log total returns and log dividend growth rates. This high correlation makes it difficult to explain the origin of predictability that log dividend price ratio has for total returns and dividend growth rates. We cannot tell whether predictability comes from price or dividend if log total returns are predictable. For example, Chen (2009) shows that dividend growth rates with reinvestment are highly correlated with total returns: the correlation is 0.65 for annual data. Cochrane (2008) also points that the correlation between total returns and dividend growth rates is high, 0.66.

To see how this argument works mathematically, we iterate equation forward and get Campbell-Shiller present value identity:

$$dp_t \approx \sum_{j=1}^{\infty} \rho^{j-1} r_{t+j} - \sum_{j=1}^{\infty} \rho^{j-1} \Delta d_{t+j} \ , \tag{1.5}$$

Where lowercase letters are logarithms of corresponding capital letters, dp_t is log dividend price ratio at time t and is identical with $d_t - p_t$; $\rho = PD/(1 + PD)$ is the point of linearization, PD is the dividend price ratio at which one linearizes; r_t is the log total return from time $t-1$ to time t; Δd_t is the log dividend growth from time $t-1$ to time t.

From equation 1.5 we can see that current log dividend price ratio has information for future log total returns and log dividend growth rates. However, if right hand side variables in equation 1.5 were highly correlated, it would not be straightforward to separate the predictability of log dividend price ratio for future log total returns and future log dividend growth rates. If log dividend price ratio forecasts future log dividend growth rates, it should forecast future log total returns too because log total returns are correlated with log dividend growth rates.

To clear the confusion, it is critical to lower the correlation between log returns and log dividend growth rates. A few papers have come up with different dividends reinvestment strategies in order to lower the correlation between right hand variables. Among others, Binsbergen and Koijen (2010) suggest that two different reinvestment strategies for dividends should be considered when running predictive regressions: reinvestment in a 30-bill T-bill and reinvestment in the aggregate stock market. Returns without dividend being reinvested in the aggregate stock market are recommended to use when dealing with predictability on returns, because in this way returns then have a lower correlation with dividend growth rates than the other way around.

However, since total returns by definition contain information of dividends, these two variables are correlated despite of the reinvestment strategy. Even without dividend reinvestment, the correlation between dividend growth rates and total returns is 0.18 for annual data, which is not small enough to ignore. In contrast, the correlation between capital gains and dividend growth rates is lower than 0.05.

If we iterate equation forward, we get the following identity:

$$dp_t = \sum_{j=1}^{\infty} \Delta p_{t+j} - \sum_{j=1}^{\infty} \Delta d_{t+j} \ . \tag{1.6}$$

Comparing equation 1.6 with equation 1.5, we see that equation 1.6 looks very similar to equation 1.5 but equation 1.6 avoids the ρ term and is also a precise equation. Because the correlation between log capital gains and log dividend growth rates is close to zero, it is now direct to separate the predictability effect that log dividend price ratio has for the right hand side variables.

Since equation 1.6 works $ex - post$, we can take expectation of both sides based on information up to time t and make it work $ex - ante$:

$$dp_t = E_t(\sum_{j=1}^{\infty} \Delta p_{t+j}) - E_t(\sum_{j=1}^{\infty} \Delta d_{t+j}) \ . \tag{1.7}$$

Equation 1.7 shows that current log dividend price ratio includes information of expected future log capital gains and expected future log dividend growth rates. Since log capital gains co-move with log total returns, this equation also implies that current log dividend price ratio does predict expected future log total returns or, at least, it should predict the sum of expected future log total returns — log total returns of long horizon.

Coefficients in the long horizon linear regression are expected to be more significant according to equation 1.7. However, the predictability of log dividend price ratio on single holding period capital gain in the far future does not disappear as the horizon rises because equation 1.7 does not have a term ρ or a term ρ^j. Equation 1.5, however, implies that as the horizon moves into far future, the predictability dies out exponentially because ρ^j goes to zero eventually as j goes to infinity since ρ is smaller than one.

1.3 Empirical Results

This section presents empirical results of return predictability tests using capital gains. As mentioned, return predictability tests using capital gains verify two main empirical findings on return predictability using total returns: reversal of predictability and increasing R-square with horizon. Besides that, I also present new findings on correlations among error terms.

1.3.1 VAR

The Joint Hypothesis Tests

Using VAR, we show the reversal of predictability between pre-war period and post-war period. The VAR takes the following form:

$$\begin{bmatrix} \Delta p_{t+1} \\ \Delta d_{t+1} \\ dp_{t+1} \end{bmatrix} = \begin{bmatrix} a_p \\ a_d \\ a_{dp} \end{bmatrix} + \begin{bmatrix} - & - & b_p \\ - & - & b_d \\ - & - & \varphi \end{bmatrix} * \begin{bmatrix} \Delta p_t \\ \Delta d_t \\ dp_t \end{bmatrix} + \begin{bmatrix} \varepsilon^p_{t+1} \\ \varepsilon^d_{t+1} \\ \varepsilon^{dp}_{t+1} \end{bmatrix},$$

or

$$\tilde{y}_{t+1} = \tilde{\mu} + \phi \tilde{y}_t + \tilde{\epsilon}_{t+1} \ .$$

The main interest in return predictability tests will be to test the significance of some coefficients in matrix ϕ, especially b_p and b_d. If b_p and b_d are significant, then log dividend price ratio proves to have predictability on log capital gains and log dividend growth.

The significance of a coefficient is decided by its t-statistic. For regression variables that are strictly stationary, the critical value of t-statistics should be set to 1.96. However, because log dividend price ratio is a highly persistent process and displays a nearly unit-root property, simulation based critical values should be used. To decide the critical value for this series, I adopt criteria set in Ng and Perron (1995). Roughly, they suggest that if the absolute value of the t-statistics of a coefficient is greater than 1.6, then the coefficient is viewed as significant. Otherwise, it should be viewed as insignificant.

Table 2 presents estimation of coefficients in matrix ϕ. The whole sample data is from 1872 to 2012 and I also consider three sub-sample periods: pre-war, war and post-war to test the change in the predictability.

INSERT TABLE. 1.2 NEAR HERE

As shown in table 1.2, during the post-war period, the coefficient b_p is around 0.08 and the coefficient b_d is around 0.01. Cochrane (2008) use sample data from 1927 to 2004 and finds that b_r is around 0.09 and b_d is around 0.01. Our statistics verify what has been found in previous research, that is, during post-war period, the estimated coefficients that log dividend price ratio has on returns and dividend growth are small.

Table 1.2 also indicates that return predictability for either log total returns or log capital gains is only a post-war case, while during the pre-war period or the war period log dividend price ratio would predict log dividend growth only but not returns. During pre-war period, the t-statistic of b_d is -3.56, which is statistically significant at the 1% level; during the period of war, t-statistic of b_d is -2.96, which is still statistically significant at the 1% level although it is a bit weaker than pre-war period. It suggests that log dividend growth is predictable during both pre-war and war periods. On the contrary, during the post-war period, log dividend growth rates become unpredictable but log capital gains become predictable. During the post-war period, the t-statistic of b_p is 1.65, which is statistically significant at 10% level; Cochrane (2008) finds the t-statistics for b_r is 1.92, which is a similar number to mine. These marginally significant t-statistics indicate that return predictability during post-war periods is only marginal using log dividend price ratio.

The equation $b_p = 1 - \varphi + b_d$ which describes the relationship among the regression coefficients is verified by the empirical statistics in the VAR. Since φ is smaller than one, if b_d is zero or close to zero, then b_p must be different from zero, then we are able to conclude that return is predictable. The sum of b_p and b_d is $(1 - \varphi)$, which is larger than zero, so at least one of the numbers, b_p or b_d has to be different from zero, which means, at least one of the variables, log capital gains or log dividend growth has to be predictable using log dividend price ratio. Cochrane (2008) first uses this theory to back up return predictability using the formula: $b_r \approx 1 - \rho\varphi + b_d$. Here I am using capital gains instead of total returns, consequently I do not have to deal with first order approximation from Taylor expansion, nor do I need to worry about the changing value of ρ. The analysis of return predictability becomes more direct and clearer.

Impulse Response Analysis

The structure of dividend has changed during post-war period compared to pre-war period. During pre-war period, dividend growth did not correlate with previous dividend growth or previous capital gains but now it does. Previous research finds that dividend growth becomes much more smooth after war. If we use autoregressive moving average (ARMA) model to fit the series of dividend growth, the estimated coefficients in the ARMA model must have changed. There are many possible factors that caused the change. For example, it may be due to change in macroeconomics environment or change is companies policy.

The estimated coefficients matrix ϕ during the post-war period takes the following form:

$$\begin{bmatrix} 0 & 0 & A \\ B & C & 0 \\ 0 & 0 & D \end{bmatrix},$$

In which A, B, C and D represent non-zero coefficients in the matrix ϕ. A represents the same coefficient with b_p and will be our main interest for return predictability test.

Since VAR takes the form $\tilde{y}_{t+1} = \tilde{\mu} + \phi\tilde{y}_t + \tilde{\epsilon}_{t+1}$, the impact that a shock $\tilde{\epsilon}_t$ at time t on $\tilde{y}_{t+j}$ at time $t+j$ is: $\phi^j = \partial\tilde{y}_{t+j}/\partial\tilde{\epsilon}_t$. Since ϕ takes a special form in which a few coefficients inside the matrix are zero, ϕ^j takes the following form:

$$\begin{bmatrix} 0 & 0 & A^* \\ B^* & C^* & E^* \\ 0 & 0 & D^* \end{bmatrix},$$

In which A^*, B^*, C^*, D^* and E^* represent non-zero coefficients in the matrix ϕ^j.

The impact of shocks at time t for log capital gains at time $t+j$ is: $0*\varepsilon_t^p + 0*\varepsilon_t^d + 0*\varepsilon_t^{dp}$. A^* is equal to $AD^{(j-1)}$ and represents the impact that a log dividend price ratio shock at time t has on log capital gains at time $t + j$. D is estimated to be 0.93 during post-war and as j goes to infinity, this impact of shocks on future log capital gains will die away slowly.

Correlations Among Error Terms in VAR

This sub-section presents empirical findings on correlations among error terms in VAR. Previous research uses log total returns as dependent variables in predictive regressions and in this paper I use log capital gains returns. Since the correlation between log total returns and log dividend growth is significantly positive while the correlation between log capital gain and log dividend growth is nearly zero, correlations among errors and standard deviations of errors in my VAR regression are different from previous researchs findings.

INSERT TABLE. 1.3 NEAR HERE

Table 1.3 presents correlations among error terms of VAR using data from 1872 to 2012. We can see that the error correlation between log capital gains and log dividend price ratio is -0.80 during pre-war period and -0.94 during post-war period; the error correlation between log dividend growth and log dividend price ratio is 0.55 during pre-war period and 0.30 during post-war period. So the connection between the log dividend price ratio and log capital gain has become stronger while the connection between the log dividend price ratio and log dividend growth has become weaker. Cochrane (2008) uses data from 1927 to 2004 and finds that during this period correlations are -0.70 and 0.08 respectively.

The negative correlations -0.86, -0.80, -0.94 and -0.94 imply that shocks to log dividend price ratio impact log capital gains adversely; the positive correlations 0.44, 0.55, 0.28 and 0.30 imply that shocks to the log dividend price ratio impact log dividend growth rate in the same direction. These non-zero numbers suggest that log dividend price ratio does have predictability on log dividend growth and log capital gains.

Recall that we derived an equation $\varepsilon_{t+1}^{dp} = \varepsilon_{t+1}^{\Delta d} - \varepsilon_{t+1}^{\Delta p}$, which describes a linear relationship among error terms. This equation implies that the error correlation between log dividend price ratio and log dividend growth $cor(\varepsilon^{dp}, \varepsilon^{\Delta d})$ should be positive and the error correlation between log dividend price ratio and log capital gains $cor(\varepsilon^{dp}, \varepsilon^{\Delta p})$ should be negative. Table 1.3 verifies the implication of this equation.

INSERT TABLE. 1.4 NEAR HERE

Table 1.4 presents standard deviations of three error terms in the VAR. During post-war period, the standard deviations of these error terms are estimated to be 0.16, 0.06 and 0.17, respectively. The estimated standard deviation of log dividend price ratio is higher than the estimated standard deviation of log capital gain. This empirical result gets along with the equation $\varepsilon_{t+1}^{dp} = \varepsilon_{t+1}^{\Delta d} - \varepsilon_{t+1}^{\Delta p}$ from which we can derive that $var(\varepsilon^{dp}) = var(\varepsilon^{\Delta d}) + var(\varepsilon^{\Delta p})$ given that the log dividend growth is uncorrelated with the log capital gains. In Cochrane (2008), the standard deviations of these error terms are estimated to be 0.20, 0.14 and 0.15, respectively. The estimated standard derivation of log total return is higher than the estimated standard derivation of log dividend price ratio. The high estimated standard deviation of log total return in Cochrane (2008) is partly contributed by the fluctuation of dividends.

Comparing post-war period with pre-war period, we can see that the log capital gains are about same volatile; however, post-war dividend growth becomes smoother than that during pre-war: the standard deviation changes into almost half — from 0.11 to 0.06, which means the variance changes into about one quarter of its previous level. The standard deviation of the log dividend price ratio also declines because of the smoother dividend series.

1.3.2 Long-horizon Regressions

This section uses ordinary least square (OLS) to show another empirical finding on return predictability: the evidence of return predictability gets stronger as regression horizon increases. The in-sample long horizon regressions are as following:

$$\begin{cases} \sum_{j=1}^{k} \Delta p_{t+j} & = a_p + b_p^{(k)}(dp_t) + \epsilon_{t+k}^{\Delta p} \\ \sum_{j=1}^{k} \Delta d_{t+j} & = a_d + b_d^{(k)}(dp_t) + \epsilon_{t+k}^{\Delta d} \\ dp_{t+k} & = a_{dp} + b_{dp}^{(k)}(dp_t) + \epsilon_{t+k}^{\Delta dp} \end{cases} .$$

Combining the present value identity $dp_t = \sum_{j=1}^{k} \Delta p_{t+j} - \sum_{j=1}^{k} \Delta d_{t+j} + dp_{t+k}$ with the above three long horizon regressions, I get a linear relationship among coefficients:

$$b_p^{(k)} - b_d^{(k)} = 1 - b_{dp}^{(k)} \ .$$

If dp_t can be viewed as an autoregressive one (AR(1)) process, which takes the form $dp_{t+1} = a_{dp} + \varphi * (dp_t) + \varepsilon_{t+1}^{dp}$, then we have $b_{dp}^{(k)} = \varphi^k$. Under normal situations, $0 < \varphi < 1$, we have:

$$b_{dp}^{(k)} = \varphi^k \to 0, \ \text{as} \ k \to +\infty.$$

Thus,

$$b_p^{(k)} - b_d^{(k)} \to 1, \ \text{as} \ k \to +\infty, \tag{1.8}$$

which states that difference between the coefficient in long horizon capital gain regression and the coefficient in long horizon dividend growth regression goes to one as the horizon goes to infinity.

The equation $b_p = 1 - \varphi + b_d$ describes the relationship among one step ahead forecast coefficients. Combing equation and equation , I derive the implied long run coefficients:

$$b_p^{(\infty)} = \frac{b_p}{1-\varphi}$$

and

$$b_d^{(\infty)} = \frac{b_d}{1-\varphi}.$$

INSERT TABLE. 1.5 NEAR HERE

INSERT TABLE. 1.6 NEAR HERE

Table 1.5 and table 1.6 present empirical estimation of coefficients b_p and b_d of different horizons. Since the log dividend price ratio predicts the log dividend growth only during the pre-war and war periods and predicts the log capital gains only during the post-war period, I calculate implied coefficients of these periods separately. The theoretical implied long horizon coefficient for log capital gains is $0.07/(1-0.93) \approx 1.10$, which can be verified by the empirical estimations; but the theoretical implied long horizon coefficient for log dividend growth is $-0.41/(1-0.55) \approx -0.91$, which cannot be verified by the empirical estimations. According to the present value identity, the long horizon predictability for the log dividend growth using the log dividend price ratio should increase with forecasting horizon but empirically it is not.

From table 1.5 and table 1.6 we can see that the coefficients and the R-square in log capital gains regressions get more significant or larger as the horizon increases, during both pre-war and post-war period; however, coefficients in log dividend growth regressions do not become more significant as horizon increases — they either behave less significant or display the wrong sign. The R-square in log dividend growth regressions during post-war periods increases with horizon but is of wrong sign and should not be considered significant. Predictability shown for log dividend growth using log dividend price ratio is only significant in one-step ahead prediction during pre-war and war period.

1.4 Summary

To summarize, this paper presents a simpler way to test return predictability compared to previous literature. It proposes to test capital gains predictability and infer total return predictability from that. This method leads to an exact linear relationship among prediction coefficients. This exact linearization avoids the concern of change of the point of linearization when taking first order Taylor expansion in Campbell-Shiller (1988) log linearization.

Using capital gains, I verify empirical findings on return predictabilities found in previous literature. They are the reversal of predictability between the pre-war and the post-war period and the stronger evidence of return predictability in longer horizon regressions. I get a new present value identity by iterating linearization forward and compare it with Campbell-Shiller present value identity. Current dividend price ratio does include information of future dividend growth rates and future returns. Different from what Campbell-Shiller present value identity implies, there is no dying out effect on predictability when moving into far future since there is no term ^{j} involved which will go to zero eventually as j increases.

The new present value identity is easier to estimate and cleaner for analysis. Since the correlation between capital gains and dividend growth rates is small enough to neglect, problems in predictability test caused by high correlation between total returns and dividend growth rates can be avoided. New empirical findings on error terms show up when running a similar VAR with that in Cochrane (2008), that is, shocks to dividend price ratio have a significant positive correlation with shocks to dividend growth instead of close to zero. The correlation between shocks to dividend price ratio and shocks to capital gains are more negative than with the correlation between shocks to dividend price ratio and shocks to total returns.

1.5 References

Binsbergen, Jules H.Van, and Ralph S. Koijen, 2010, Predictive Regressions: A Present-Value Approach, The Journal of Finance 65, 1439-1471

Campbell, John Y., and Robert Shiller, 1988, The dividend-price ratio and expectations of future dividends and discount factors, Review of Financial Studies 1, 195-228

Campbell, John Y., 1991, A variance decomposition for stock returns, Economic Journal 101, 157-179

Chen, Long, 2009, On the reversal of return and dividend growth predictability: A tale of two periods, Journal of Financial Economics 92, 128-151

Cochrane, John H., 2008, The Dog That Did Not Bark: A Defense of Return Predictability, The Review of Financial Studies 21, 1533-1573

Cochrane, John H., 2011, Presidential Address: Discount Address, The Journal of Finance 66, 1047-1108 Fama, Eugene, and Kenneth French, 1988, Permanent and Temporary Components of Stock Prices, Journal of Political Economy 96, 246-273

Lanne, Markku, 2002, Testing the Predictability of Stock Returns, The Review of Economics and Statistics 84, 407-415

Lettau, Martin, and Sydney Ludvigson, 2001, Consumption, Aggregate Wealth, and Expected Stock Returns, The Journal of Finance 56, 815-849

Ng, Serena, and Pierre Perron, 1995, Unit Root Tests in ARMA Models with Data-Dependent Methods for the Selection of the Truncation Lag, Journal of the American Statistical Association 90, 268-281

Welch, Ivo and Amit Goyal, 2008, A Comprehensive Look at The Empirical Performance of Equity Premium Prediction, Review of Financial Studies 21, 1455-1508

Fig. 1.1. Total returns and capital gains.

This figure plots the annual log total returns r_t and annual log capital gains plus a constant $\mu + \Delta p_t$.

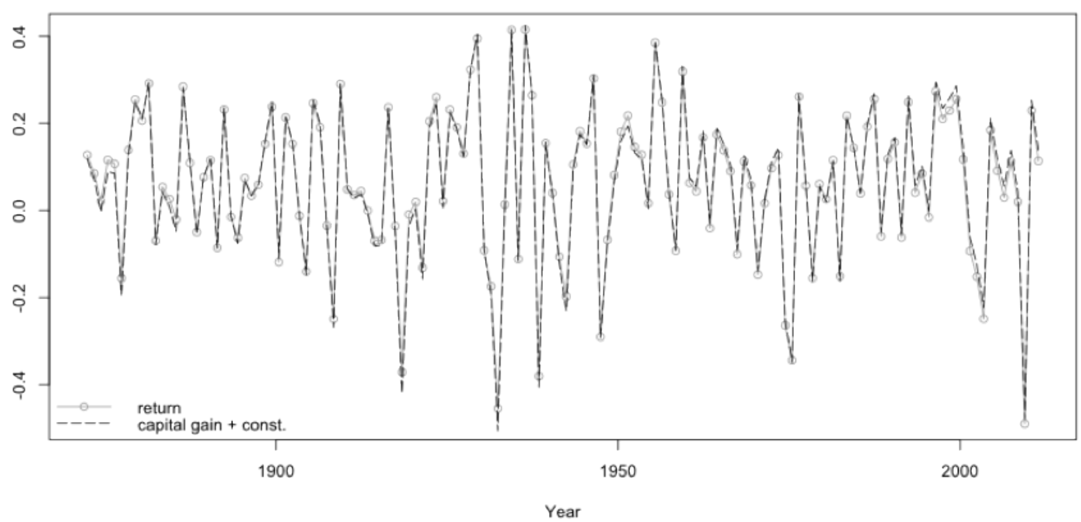


Figure 1.1: Total returns and capital gains.

Fig. 1.2. Value of the point of linearization ρ.

This figure plots the value of ρ, the point of linearization defined as $\frac{P/D}{1+P/D}$.

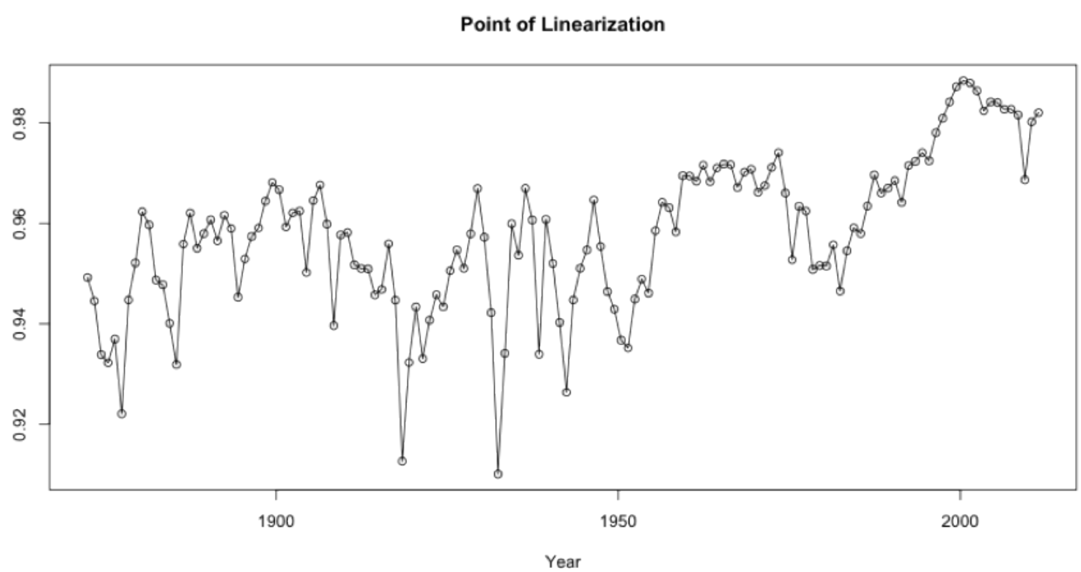


Figure 1.2: Value of ρ.

Table 1.1

Statistics of variables.

This table presents means and standard deviations of log total returns, log capital gains and dividend price ratios. Data is annual series from Robert Shillers website. Time spans from 1871 to 2011.

Table 1.1: Statistics of variables

$Variables$	r_t	Δp_t	D_t/P_t
Mean	0.06	0.02	0.04
SD	(0.17)	(0.18)	(0.02)

Table 1.2

Estimated coefficients matrix ϕ.

This table presents coefficients in matrix ϕ in the VAR and t-statistics of corresponding coefficients in parentheses. Data is from Robert Shillers website. The significant codes are "***" for the level of 0.01, "**" for 0.05, and "*" for 0.10. Numbers that are significant are in bold.

Table 1.2: Estimated coefficients matrix

Whole sample (1872-2012)			
Coef	0.07	**-0.24***	0.04
T-stat	(0.77)	(-1.83)	(1.10)
	0.30***	0.10	**-0.04***
	(5.77)	(1.35)	(-1.73)
	0.23**	**0.33****	**0.92****
	(2.28)	(2.34)	(21.50)
Pre-war (1872-1925)			
Coef	0.03	-0.17	0.05
T-stat	(0.16)	(-1.03)	(0.46)
	0.09	0.05	**-0.27*****
	(0.76)	(0.45)	(-3.56)
	0.07	0.22	**0.68*****
	(0.32)	(1.13)	(5.28)
War (1926-1945)			
Coef	0.35	-0.45	0.24
T-stat	(0.68)	(-1.30)	(0.48)
	0.13	-0.06	**-0.53*****
	(0.69)	(-0.48)	(-2.96)
	-0.22	0.39	0.23
	(-0.41)	(1.08)	(0.45)
Post-war (1946-2012)			
Coef	0.07	0.13	**0.08**
T-stat	(0.55)	(0.40)	(1.65)
	0.13***	**0.30*****	0.01
	(2.93)	(2.67)	(0.64)
	0.06	0.17	**0.93*****
	(0.45)	(0.51)	(18.31)

Table 1.3

Errors Correlations in VAR.

This table presents correlations among error terms in the VAR. The whole sample data is from 1872 to 2012 and three sub-periods are pre-war, war and post-war. Numbers in bold will be discussed in the paper.

Table 1.3: Errors Correlations

Whole sample (1872-2012)			
Corr	$\varepsilon^{\Delta p}$	$\varepsilon^{\Delta d}$	ε^{dp}
$\varepsilon^{\Delta p}$	1.00		
$\varepsilon^{\Delta d}$	0.08	1.00	
ε^{dp}	-0.86	0.44	1.00
Pre-war (1872-1925)			
	1.00		
	0.06	1.00	
	-0.80	0.55	1.00
War (1926-1945)			
	1.00		
	0.06	1.00	
	-0.94	0.28	1.00
Post-war (1946-2012)			
	1.00		
	0.04	1.00	
	-0.94	0.30	1.00

Table 1.4

Standard Deviations of Error Terms in VAR.

This table presents standard derivations of error terms in the VAR. The whole sample data is from 1872 to 2012 and three sub-periods are pre-war, war and post-war.

Table 1.4: S.D. of Errors Correlations

	Whole sample	Pre-war	War	Post-War
S.D.	1872-2012	1872-1925	1926-1945	1946-2012
$\varepsilon^{\Delta p}$	0.18	0.15	0.25	0.16
$\varepsilon^{\Delta d}$	0.10	0.11	0.09	0.06
ε^{dp}	0.20	0.18	0.26	0.17

Table 1.5

Predictability on Capital Gains using Log Dividend Price Ratio.

This table presents the linear in-sample predictive regressions for the log capital gains using one-period lagged log dividend price ratio. The regression is $\sum_{j=1}^{k} \Delta p_{t+j} = a_p + b_p^{(k)} * (dp_t) + \epsilon_{t+k,p}$ with k represents the number of years ahead. The significant codes are "***" for the level of 0.01, "**" for 0.05, and "*" for 0.10.

Table 1.5: Predictability on Capital Gains

Coef $b_p^{(k)}$ (T-stat)	whole sample (1872-2012)	Pre-war (1872-1925)	War (1926-1945)	Post-war (1946-2012)
K=1	0.03	0.04	0.05	0.07
	(0.94)	(0.45)	(1.11)	(1.54)
K=5	0.22***	0.60***	0.23**	0.36***
	(2.80)	(3.48)	(2.49)	(3.63)
K=10	0.38***	0.62***	0.45***	0.82***
	(3.25)	(2.98)	(3.09)	(5.00)
K=15	0.72***	1.07***	0.94***	1.68***
	(4.19)	(4.66)	(3.98)	(6.28)
K=20	0.90***	1.05***	1.29***	2.11***
	(4.74)	(4.93)	(5.03)	(7.41)
R-square				
K=1	0.6%	0.3%	1.4%	3.6%
K=5	5.6%	15.3%	7.3%	18.5%
K=10	7.6%	12.5%	11.5%	31.6%
K=15	12.5%	27.2%	18.7%	45.1%
K=20	16.0%	31.9%	28.4%	56.1%

Table 1.6

Predictability on Dividend Growth using Log Dividend Price Ratio.

This table presents the linear in-sample predictive regressions for the log dividend growth rates using one-period lagged log dividend price ratio. The regression is $\sum_{j=1}^{k} \Delta d_{t+j} = a_d + b_d^{(k)} * (dp_t) + \epsilon_{t+k,d}$ with k represents the number of years ahead. The significant codes are "***" for the level of 0.01, "**" for 0.05, and "*" for 0.10.

Table 1.6: Predictability on Dividend Growth

Coef $b_d^{(k)}$ (T-stat)	whole sample (1872-2012)	Pre-war (1872-1925)	War (1926-1945)	Post-war (1946-2012)
K=1	-0.08***	-0.41***	-0.05**	0.01
	(-3.39)	(-7.72)	(-2.02)	(0.39)
K=5	-0.11**	-0.37**	-0.06	0.02
	(-2.15)	(-2.61)	(-1.05)	(0.54)
K=10	-0.01	-0.09	0.01	0.12*
	(-0.12)	(-0.59)	(0.20)	(1.83)
K=15	-0.03	-0.24	0.15	0.50***
	(-0.29)	(-1.29)	(1.35)	(5.39)
K=20	0.1	-0.03	0.28**	0.56***
	(1.09)	(-0.19)	(2.31)	(4.61)
R-square				
K=1	7.7%	45.6%	4.6%	0.2%
K=5	3.3%	9.2%	1.4%	0.5%
K=10	0.0%	0.6%	0.1%	5.8%
K=15	0.0%	2.8%	2.6%	37.7%
K=20	1.0%	0.1%	7.7%	33.1%

Chapter 2

WILL DECOMPOSED LOG DIVIDEND PRICE RATIO PREDICT RETURNS BETTER?

2.1 Introduction

Log dividend price ratio is one of the most fundamental predictors for returns because it can be decomposed into the sum of the future returns minus the sum of the future dividend growth rates as shown in Campbell- shillers (1988) present value identity.

In this paper I try to answer a question: is log dividend price ratio stationary? The answer to this question is critical. If log dividend price ratio were not stationary, previous research based on the stationary assumption would be problematic. For example, without the assumption that log dividend price ratio is stationary, the vector autoregressive (VAR) regressions in Cochrane (2008) are not reliable.

For past thirty years or more, researchers have been trying to forecast future returns using log dividend price ratios. However, the predictability on one-year returns is not statistically significant. Long run forecast is getting more significant as the horizon gets longer. The well-known fact is log price ratio is very persistent: if I run an autoregressive order one (AR (1)) model on the price ratio itself, the coefficient can be as high as 0.94. This is, in my opinion, part of the reason why dividend price ratio forecasts long run returns with a higher R-square. Long run returns are much more persistent than one year ahead returns. Empirically, as the persistence of two time series gets closer, the prediction gets better.

One view about dividend price ratio is that it is stationary after taking into account structural breaks. For example, Lettau and Nieuwerburgh (2008) view dividend price ratio as having structural breaks in means. There is at least one structural break in around 1991; the second possible structure break is around 1954. They use demeaned price ratio to forecast returns and dividend growth and find that demeaned price ratio has a better prediction power than the original price ratio. The draw back of this view is that there is no solid economic explanation of structural breaks. For different finance ratios, estimated structural break dates are different. This looks more like a fitting data to a function than explaining what really goes on.

In this paper, I argue that log dividend price contains a unit root process if the expectation of future discount factor follows a random walk. The intuition for price ratio being non-stationary is straight forward: during the past thirty years, existing dividend paying firms become less likely to pay dividends; more and more firms move away from paying dividends toward share repurchases or simply stop paying dividends. The proportion of small firms with strong growth opportunities that never paid dividends is also increasing. Fama and French (2000) state that the proportion of firms paying cash dividends falls from 66.5% in year 1978 to 20.8% in year 1999. It is likely that dividend paying will continue to decrease in the future. However, price is not decreasing. Firms are never reluctant to increase their stock prices and stock prices keep going up. How can dividend price ratio be stationary if they move toward opposite directions?

My approach is analogous to Letttau and Ludvigson (2001). They use deviation from consumption wealth ratio (cay) to predict returns. In this paper, I construct a Bayesian framework and use the Kalman Filter to extract the stationary component in dividend price ratio. The extracted stationary process can predict one-year return in the post-war period with a R-square of 42.3%; using dividend price ratio directly, we have a R-square of 7.0%; Lettau and Nieuwerburgh (2008) use demeaned price ratio and have a R-square of 10.0% for one structural break assumption and 22.3% for two structural break assumptions for 1927 to 2004; while cay in Lettau and Ludvigson (2001) has a R-square of 9.0% for quarterly returns in fourth quarter of 1952 to third quarter of 1998.

The rest of this paper is organized as follows: section two uses a simple present value model to show that real price and real dividend are cointegrated with a timing varying cointegration vector, then moves on to explain where the persistence of log dividend price ratio comes from; section three first shows that log dividend price ratio is not empirically stationary by using unit root test and stationary test, then use the stationary component in it to predict returns and dividend growth. A little bit discussion on other literature follows after the prediction regressions. Section four includes a brief summary. Reference follows in the end.

2.2 Theoretical Analyses

Theoretical analyses include two parts. The first part is theoretical analysis on the cointegration relationship between real price and real dividend and the time-varying cointegration vector between them; the second part represents the non-stationary characters of log dividend price ratio.

2.2.1 Cointegration between real price and real dividend

I use a simple model here to present that there is a cointegration relationship between real price and real dividend and the cointegration vector depends on expectation of future discount factors.

Model

According to the efficient markets theory, the real price p_t of a share at the beginning of the time period t is given by:

$$P_t = E_t(\sum_{k=1}^{\infty} \gamma_{t+k} D_{t+k}),\ 0 < \gamma_{t+k} < 1,$$

in which $D_(t+1)$ is the real dividend paid at the end of time period t or at the beginning of the time period $t+1$. Similarly, D_{t+k} is the real dividend paid at the beginning of the time period $t+k$. E_t denotes mathematical expectation conditional on information available at t, γ_{t+k} represents the time varying real discount factor.

If we assume zero correlation between real discount factor and real dividend, we have a formula for real price as follows:

$$P_t = \sum_{k=1}^{\infty} E_t(\gamma_{t+k}) E_t(D_{t+k}),\ 0 < \gamma_{t+k} < 1.$$

INSERT FIGURE. 2.1 NEAR HERE

Figure 2.1 plots annual real dividend series from 1872 to 2011. It seems a non-stationary process and a quick computer test using R tells us it is integrated one ($I(1)$) process. Since all $I(1)$ processes can be decomposed into a non-stationary (random walk) term and a stationary term, we can decompose the real dividend series as follows:

$$D_t = x_t + z_t,$$

in which x_t follows a random walk:

$$x_t = x_{t-1} + \varepsilon_{x,t},\ x_0 \neq 0.$$

and z_t is a stationary term that has unconditional mean zero.

INSERT FIGURE. 2.2 NEAR HERE

Figure 2.2 plots annual real dividend growth from 1872 to 2011. It seems a stationary process and a simple stationary test using R tells that it is a stationary process. Since every stationary process has a Wold representation, we can write real dividend growth as follows:

$$\Delta D_{t+1} = cost. + a(L)\varepsilon_{D,t+1},$$

and,

$$\begin{cases} D_{t+1} = & D_t + a(L)\varepsilon_{D,t+1} \\ D_{t+2} = & D_{t+1} + a(L)\varepsilon_{D,t+2} \\ \ldots\ldots & \\ \ldots\ldots & \\ D_{t+k} = & D_{t+k-1} + a(L)\varepsilon_{D,t+k}. \end{cases}$$

Take expectation at time t, we have:

$$\begin{cases} E_t(D_{t+1}) = & D_t + e_{t,1} \\ E_t(D_{t+2}) = & D_t + e_{t,2} \\ \cdots\cdots & \\ \cdots\cdots & \\ E_t(D_{t+k}) = & D_t + e_{t,k}, \end{cases}$$

where e_t is realized error term at time t and is function of $\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}\cdots\cdots$.

If we take the decomposition of dividend series $D_t = x_t + z_t$ into the above formula, we have:

$$\begin{cases} E_t(D_{t+1}) = & x_t + z_t + e_{t,1} \\ E_t(D_{t+2}) = & x_t + z_t + e_{t,2} \\ \cdots\cdots & \\ \cdots\cdots & \\ E_t(D_{t+k}) = & x_t + z_t + e_{t,k}. \end{cases}$$

The first term in the right hand side variables x_t is a random walk component, the rest two terms are stationary components.

Now if we put $E_t(D_{t+k}) = x_t + z_t + e_{t,k}$ back into the present value identity $P_t = \sum_{k=1}^{\infty} E_t(\gamma_{t+k})E_t(D_{t+k})$, we get:

$$P_t = \sum_{k=1}^{\infty} E_t(\gamma_{t+k})(x_t + z_t + e_{t,k}),\ 0 < \gamma_{t+k} < 1.$$

It is clear now there a cointegration relationship between real dividend and real price since they share the same random walk component x_t. The cointegration vector is time-varying if the expected discount factor is time-varying.

2.2.2 Non-stationary component in log dividend price ratio

Now we have a model for real dividend and real price series:

$$\begin{cases} P_t = & \alpha_t * x_t + u_t \\ D_t = & x_t + z_t, \end{cases}$$

where $\alpha_t = \sum_{k=1}^{\infty} E_t(\gamma_{t+k})$ is the sum of expected future discount factor; u_t is a function of z_t and is stationary.

Take log of both sides, we have:

$$\begin{cases} p_t = & log\alpha_t + logx_t + w_t \\ d_t = & logx_t + v_t, \end{cases}$$

in which small letters represent log values of corresponding capital letters; w_t and v_t are left stationary process. We can formulate log dividend price ratio now:

$$dp_t = \; - \; log\alpha_t + v_t - w_t.$$

Since both v_t and w_t are stationary process, the non-stationary character of log dividend price ratio must come from $log\alpha_t$. If expected discount factor $_t$ is non-stationary, then the log of it — $log\alpha_t$ is non-stationary because the one-to-one relationship between log value and original value, then log dividend price ratio has a non-stationary component and is not stationary.

Now the question becomes, is $\alpha_t = \sum_{k=1}^{\infty} E_t(\gamma_{t+k})$ stationary or not? The persistence of peoples expectation has been researched in literature for a long time. Research shows that people do not seem to adjust their expectation accordingly even the reality tells them later that their expectations were wrong. See Buehler, Griffin and Ross (1994) for example. If I assume that expectation of future discount factor follows a random walk, the sum of expected future discount factor $_t$ follows a random walk. That is, we assume the expectation of discount factor at time t is equal to the expectation at time $t-1$, plus an expectation error or update; at time $t-1$, this expectation error or update has a value of zero:

$$\begin{cases} E_t(\gamma_{t+1}) = & E_{t-1}(\gamma_t) + \varepsilon^1_{\gamma,t} \\ E_t(\gamma_{t+2}) = & E_{t-1}(\gamma_{t+1}) + \varepsilon^3_{\gamma,t} \\ \dots\dots \\ \dots\dots \\ E_t(\gamma_{t+k}) = & E_{t-1}(\gamma_{t+k-1}) + \varepsilon^k_{\gamma,t}, \end{cases}$$

where $E_{t-1}(\varepsilon_\gamma, t) = 0$. Then we have the sum of expectation $_t$ follows a random walk

$$\alpha_t = \sum_{k=1}^{\infty} E_t(\gamma_{t+k}) = \alpha_{t-1} + \sum_{k=1}^{\infty} \varepsilon^k_{\gamma,t}.$$

Thus both α and $log\alpha_t$ are non-stationary. Since log dividend price ratio contains $log\alpha_t$, it must be non-stationary too. This analysis also tells us the persistence of log dividend price ratio comes from the persistence of expectation on future discount factors. Discount factors is constructed by one over one plus discount rates $\gamma = 1/(1+r)$; persistent expected discount factors also reflects that persistent expected discount rates, in other words, expected future returns are persistent. This persistence makes sense because it is saying peoples expectation of this years return is based on what previous years return was. Intuitively, that is how expectation is formed: it is formed according to the history.

2.3 Empirical Evidences

The empirical work includes two parts. The first part is unit root test on log dividend price ratio. Since I derive that there is a non-stationary component in log dividend price ratio, it is necessary to check whether existing tests agree with me. The second part is prediction on returns, capital gains and dividend growth using the stationary component in log dividend price ratio. Previous research shows that using original log dividend price ratio has little prediction power; I show here that using the stationary part in it improves the prediction significantly.

I use data from Robert Shiller website. Dividend data are computed from the $S\&P$ four-quarter totals for the quarter since 1926. Dividend before 1926 are from Cowles and associates (1939), interpolated from annual data. Stock price data are yearly averages of daily closing prices through 2012. After taking inflation into account, all prices and dividends are real and effectively in January 2000 dollars. More detailed data description can be seen from Robert Shillers website.

2.3.1 Unit Root Test and Stationary Test

I construct both unit root test and stationary test to test the character of log dividend price ratio.

Unit Root Test

For the unit root test, I use Zivot and Andrews unit root test (1992) (Z&A). The null hypothesis of Z&A test is that the series contains a unit root and the alternative hypothesis is that the series is stationary with a structural break. Z&A test has a similar regression with augmented Dickey fuller (ADF) test but allows a structural break and the time of structural break is no exogenous but is decided using the data. The reason why I chose this test over ADF test is there are quite a few papers discussing structural breaks in log dividend price ratio and the common view is that there is one structural break happening during 1990 or 1991 but the ratio is still stationary, see Lettau and Nieuwerburgh (2008) for example.

To use this test, I have to decide the model type, number of lagged endogenous differenced variables to be included and then generate critical value. Since the mean of log dividend price ratio is non-zero and the series itself does not have an obvious trend, I choose to regress it as "constant without trend".

Next I need to decide the number of lagged endogenous differenced variables to be included. Following Ng and Perron (1995) first I need to set the upper bound of this number, second I need to choose the optional number. The useful rule of thumb for determining the maximum number p_{max} to be considered, suggested by Schwert (1989), is

$$p_{max} = [12 \cdot (\frac{T}{100})^{\frac{1}{4}}],$$

where $[x]$ denotes the integer part of x and T is the number of observations. This choice allows p_{max} to grow with the sample. I have data from year 1871 to year 2011, in total around 140 yearly data, so the p_{max} is 12. After deciding the maximum lags, I need to decide the optional lag length p. I do this by estimating the regression with $p = p_{max}$. If the absolute value of the t-statistics for testing the significance of the lagged difference is greater than 1.6 then set $p = p_{max}$ and perform the unit root test. Otherwise, reduce the lag length by one and repeat the process. Thus I find the optional lag order equals to 10.

Finally I can generate the critical value at different quantiles and compare them with the test statistics. The simulated critical values are -5.34, -4.80 and -4.58 at quantiles 1%, 5% and 10% and the test statistic is -3.64. Since the test statistic is greater than critical values at all quantiles, the null hypothesis that there is a unit root component is not rejected. The test also suggests that the possible break date is during 1994.

INSERT FIGURE. 2.3 NEAR HERE

Figure 2.3 plots the Z&A test. Test summary is in appendix.

Stationary Test

Unit root tests has the null hypothesis that a time series is integrated of order one (I(1)). Stationary tests, on the other hand, has the null hypothesis that a time series is integrated of order zero (I(0)). The most commonly used stationarity test, the KPSS test, is due to Kwiatkowski, Phillips, Schmidt and Shin (1992). They derive their test by starting with the model:

$$y_t = c + \mu_t + u_t$$
$$\mu_t = \mu_{t-1} + \varepsilon_t,\ \varepsilon_t \sim WN(0, \sigma_\varepsilon^2),$$

where c is a constant, u_t is I(0) and may be heteroskedastic. μ_t is a pure random walk with innovation variance σ_ε^2. The null hypothesis that y_t is I(0) is formulated as $H_0 : \sigma_\varepsilon^2 = 0$, which implies that μ_t is a constant.

Under the null that y_t is I(0), KPSS converges to a function of standard Brownian motion as follows:

$$KPSS \xrightarrow{d} \int_0^1 V_1(r)dr.$$

The stationary test is a one-sided right-tailed test so that one rejects the null of stationarity at the $100\alpha\%$ level if the KPSS test statistic is greater than the $100(1-\alpha)\%$ quantile from the appropriate asymptotic distribution. The simulated critical values are 0.76, 0.45 and 0.35 at the quantiles of 0.01, 0.05 and 0.10. The test statistics is 2.29, which is greater than each of the simulated critical values, so the null hypothesis that the series is stationary is rejected. If I only use data up to year 1990 to avoid the impact of possible structural break, the test statistics generated is 0.91, which is still greater than each of the simulated critical values and the null hypothesis that the series is stationary is still rejected.

2.3.2 Prediction using the Stationary Component

I have shown above both Z&A unit root test and KPSS stationary test tend to support my point that log dividend price ratio is non-stationary. Now I want to predict returns, capital gains and dividend growth using the stationary component in price ratio. The theory I employ here is to view the random walk component as a latent variable in the price ratio and construct a Bayesian framework. The dynamic linear model that I want to estimate is simply:

$$\begin{cases} y_t = \theta_t + v_t, \ v_t \sim N(0, V) \\ \theta_t = \theta_{t-1} + w_t, \ w_t \sim N(0, W) \end{cases}$$

with a prior distribution for θ_0:

$$\theta_0 = N(m_0, C_0),$$

in which y_t is price ratio series; θ_t is the latent random walk component, the start point θ_0 has a mean of m_0 and a fairly big variance C_0 to show the uncertainty about the mean;

v_t and w_t are error terms with mean zero.

To estimate this Bayesian framework, I use Kalman filter to update the information of latent variable:

$$\theta_{t-1} \| y_{1:t-1} \sim N(m_{t-1}, C_{t-1}).$$

Then I get the stationary component by subtracting the random walk component from the price ratio, namely, I get $v_t = y_t - \theta_t$. Then I use v_t to predict.

INSERT FIGURE. 2.4 NEAR HERE

The maximum likelihood estimation results is: $W = 0.026$, $V = 0.009$. The innovation of the random walk has a variance of 0.026 and the stationary component has a variance of 0.009.

INSERT FIGURE. 2.5 NEAR HERE

INSERT FIGURE. 2.6 NEAR HERE

The random walk component in log dividend price ratio is plotted in figure 2.5. The left stationary component is plotted in figure 2.6.

The predictive regressions are the following three equations:

$$\begin{cases} \sum_{j=1}^{k} r_{t+j} = a_r^{(k)} + b_r^{(k)}(dp_t) + \varepsilon_{t+k,r} \\ \sum_{j=1}^{k} \Delta p_{t+j} = a_p^{(k)} + b_p^{(k)}(dp_t) + \varepsilon_{t+k,p} \\ \sum_{j=1}^{k} \Delta d_{t+j} = a_d^{(k)} + b_d^{(k)}(dp_t) + \varepsilon_{t+k,d}, \end{cases}$$

in which r_{t+j}, Δp_{t+j} and Δd_{t+j} represent total return, capital gain return and dividend growth rate at time $t+j$ respectively; dp_t is the log dividend price ratio at time t; other letters correspond to intercepts, coefficients and error terms in the linear regressions. To compare predictions at different horizons, I choose to regress those equations at $k =$ 1,5,10,15 and 20 and cut the whole sample into three sub-samples: pre-war and war period (1872-1945), war and post-war period (1926-2012) and post-war period (1946-2012). Results are summarized in tables 1.1, 1.2, 1.3 and 1.4 respectively.

INSERT TABLE. 2.1, 2.2, 2.3 and 2.4 NEAR HERE

All these four tables show us a clear message: stationary component in log dividend price ratio can predict short horizon returns better than long horizon returns; log dividend ratio itself can predict long horizon returns better than short horizon returns. Since we know that long horizon returns are much more persistent than short horizon returns, log dividend price ratio itself is much more persistent than the stationary component in it, we might ask: does prediction need to match the level of persistence?

Figure 2.7 plots a comparison of returns and price ratio, which gives us an impression that prediction needs to match the level of persistence.

INSERT FIGURE. 2.7 NEAR HERE

INSERT TABLE. 2.5 NEAR HERE

Results in table 2.5 also show us the reverse of predictability between return and capital growth. Chen (2009) first points out this phenomenon. He points out that, log dividend price ratio forecast one-year ahead dividend growth in the pre-war and war period. However, in the post-war period, the predictability for capital gains disappears but the price ratio starts to predict returns more. What he meant can be seen from the upper right corner of table 2.5. From 1872 to 1945, the prediction R^2 for dividend growth rates is 45.6% but the prediction R-square for returns is only 0.9%. From year 1946 to year 2012, the R-square for return increases to 7.0% but the R-square for dividend growth almost disappears, only 0.2%.

The reversal of short run predictability can also be seen from the prediction using stationary component in price ratio: it is shown in the left corner of table 2.5. The stationary component in price ratio always has predictive power on returns and dividend growth but it used to predict dividend growth much better than it does now: R-square drops from 55.3% to 11.9%. On the contrary, it starts to predict return more significantly: R-square increases from 7.1% to 42.3%.

The last two rows in table 2.5 tell us something about long-run prediction. Because long horizon returns and long horizon dividend growth are very persistent, the stationary component does not have much prediction power on them. Price ratio itself, however, predicts long run returns very well: the R-square is 42.4% for pre-war and war period and is 61.7% for post-war period. An abnormal phenomenon is numbers in the bottom right corner: R-square for 20-year-ahead dividend growth is basically 0 for pre-war and post-war, but is 33.1% for post-war period. Is long-run dividend growth becoming more predictable? This is a question worthy digging.

2.3.3 Discussion on Other Papers

Previous research generally takes log dividend price ratio as stationary process with structural break in mean. See Rapach and Wohar (2006) and Lettau and Nieuwerburgh (2008) among many others. If this theory stands, it implies a model as follows:

$$\begin{cases} p_t = \alpha_t + u_t \\ d_t = \alpha_t + w_t, \end{cases}$$

In which p_t is the log value of real price series; d_t is the log value of the real dividend series; α_t is a random walk: $\alpha_t = \alpha_{t-1} + \varepsilon_{\alpha,t}$; u_t and w_t are stationary processes. The assumption that log dividend price ratio is stationary with structural breaks implies a formula as follows:

$$dp_t = w_t - u_t = \mu_{s_t} + \epsilon_t,$$

in which μ_{s_t} is a state-dependent mean and ϵ_t is an error term. However, this specification implies that the cointegrating vector should always be [1, -1]. As shown in my analysis, this does not hold.

My forecast estimation results are better than Lettau and Nieuwerburgh (2008). They have a R-square of 10.0% for one structural break assumption and 22.3% for two structural break assumptions for year 1927 to year 2004; while my estimation for one-year ahead return has a R-square of 42.3% for year 1946 to year 2012. Also, as they say: it is difficult to estimate the mean after the structural break. But there is a more series difficulty:

how to give multiple structural breaks reasonable economic explanations? Without proper explanations, fitting price ratio with different structural breaks does not seem convincible.

Cochrane (2008) strengths stationarity of log dividend price ratios by using Campbell-Shiller present value identity:

$$dp_t \cong \sum_{j=1}^{\infty} \rho^j r_{t+j} - \sum_{j=1}^{\infty} \rho^j \Delta d_{t+j},$$

in which rho is the point of linearization which has a mean of 0.96 and maximum value of 0.99; r is the real total return and Δd is the log dividend growth. He argues that since both real return and log dividend growth are stationary, ρ^j goes to zero eventually, log dividend price ratio equals a finite sum of stationary process and should be stationary.

However, there are two non-neglectable points here weaken his argument. First is that, the above present value identity is correct but is just an approximation. It is from the first term in Tylor expansion while omitting higher order terms. We should be more conservative when deriving properties from an approximation than from an identity. If we start with a precise identity:

$$dp_t = \sum_{j=1}^{\infty} \rho^j \Delta p_{t+j} - \sum_{j=1}^{\infty} \rho^j \Delta d_{t+j},$$

in which Δp_{t+j} represents capital gains. we can not get a similar conclusion.

The second is that, the point of linearization ρ is not a constant and its maximum value can reach 0.99. When ρ gets infinitely close to one, the time that takes for ρ^j to go to zero goes to infinity too. Thus, it is not clear that we have finite sum of stationary terms in the right hand side.

Fama and French (2002) point out that if dividend price ratio, D_t/P_t , is stationary, in long sample period the compound rate of dividend growth approaches the compound rate of capital gain. That is: $(D_t - D_{t-1})/D_{t-1} = (P_t - P_{t-1})/P_{t-1}$. However, from 1872 to 2012, the mean of capital growth rate is 2% while the mean of dividend growth rate is 1%. The first one is double of the second one and there is no sign of increasing dividend growth rate. If dividend price ratio is not stationary, the log of it is unlikely to be stationary.

2.4 Conclusions

Dividends are disappearing but prices are keeping going up. It is likely that dividends in the future will have less connection with prices or profits. Thus, dividend price ratio may not be stationary. Dividend price ratio still contains information about returns, but to predict short horizon future returns, we need to use stationary component in it rather than to use price ratio itself.

The implication of non-stationarity of dividend price ratio means in the long run, dividend growth does not have to equal to capital gain. In fact, since fewer and fewer firms are paying dividends, dividend paying is no longer a reliable sigh whether a firm is doing well or not. On the contrary, capital gain is going to be the dominant part in total return. In the future, we might wan to focus on the capital gain changes to deduce the profitability and growth opportunity of a firm.

2.5 References

Buehler, Roger, Dale Griffin and Michael Ross, Exploring the Planning Fallacy: Why People Underestimate Their Task Completion Times, 1994, Journal of Personality and Social Psychology 67, 366-381

Chen, Long, 2009, On the reversal of return and dividend growth predictability: A tale of two periods, Journal of Financial Economics 92, 128-151

Cochrane, John H., 2008, The Dog That Did Not Bark: A Defense of Return Predictability, The Review of Financial Studies 21, 1533-1573

Henkel, Sam James, J. Spencer Martin and Federico Nardari, 2011, Time-varying short-horizon predictability, Journal of Financial Economics 99, 560-580

Lettau, Martin, and Sydney Ludvigson, 2001, Consumption, Aggregate Wealth, and Expected Stock Returns, The Journal of Finance 56, 815-849

Lettau, Martin, and Sydney Ludvigson, 2005, Expected returns and expected dividend growth, Journal of Financial Economics 76, 583-626

Lettau, Martin, and Stijn Van Nieuwerburgh, 2008, Reconciling the Return Predictability Evidence, The Review of Financial Studies 21, 1607-1652

Priestley, Richard, 2000, Time-Varying persistence in expected returns, Journal of Banking & Finance 25, 1271-1286

Welch, Ivo and Amit Goyal, 2008, A Comprehensive Look at The Empirical Performance of Equity Premium Prediction, Review of Financial Studies 21, 1455-1508

Rapach, David E. and Mark E. Wohar, 2006, Structural Breaks and Predictive Regression Models of Aggregate U.S. Stock Returns, Journal of Financial Econometrics 4, 238-274

Fig. 2.1. Real Dividend Series.

This figure plots the annual real dividend series from 1872 to 2011. Dividends are effectively in January 2000 dollars. More detailed information of data can be seen from Robert Shillers website.

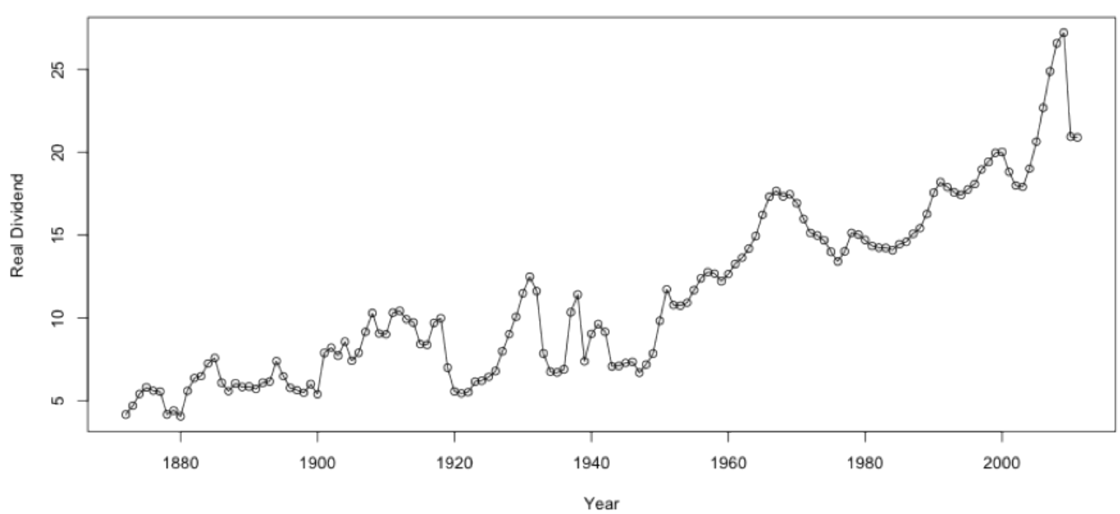


Figure 2.1: Real Dividend.

Fig. 2.2. Real Dividend Growth.

This figure plots the annual real dividend growth from 1872 to 2011. Dividends are effectively in January 2000 dollars. More detailed information of data can be seen from Robert Shillers website.

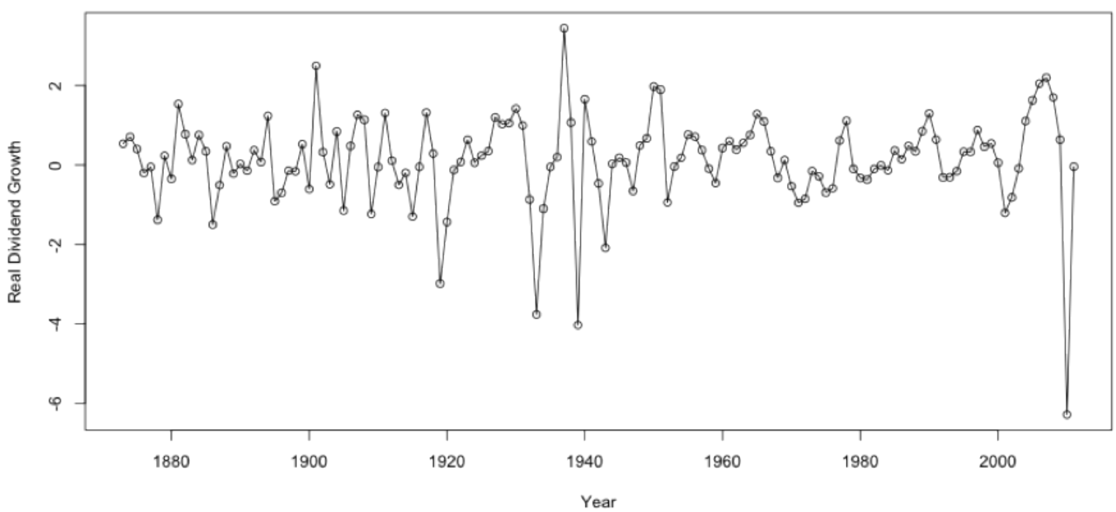


Figure 2.2: Real Dividend Growth.

Fig. 2.3. Zivot & Andrews Unit Root Test.

This figure plots the Zivot & Andrews unit root test on log dividend price ratio. The null hypothesis that there is a unit root is not rejected.

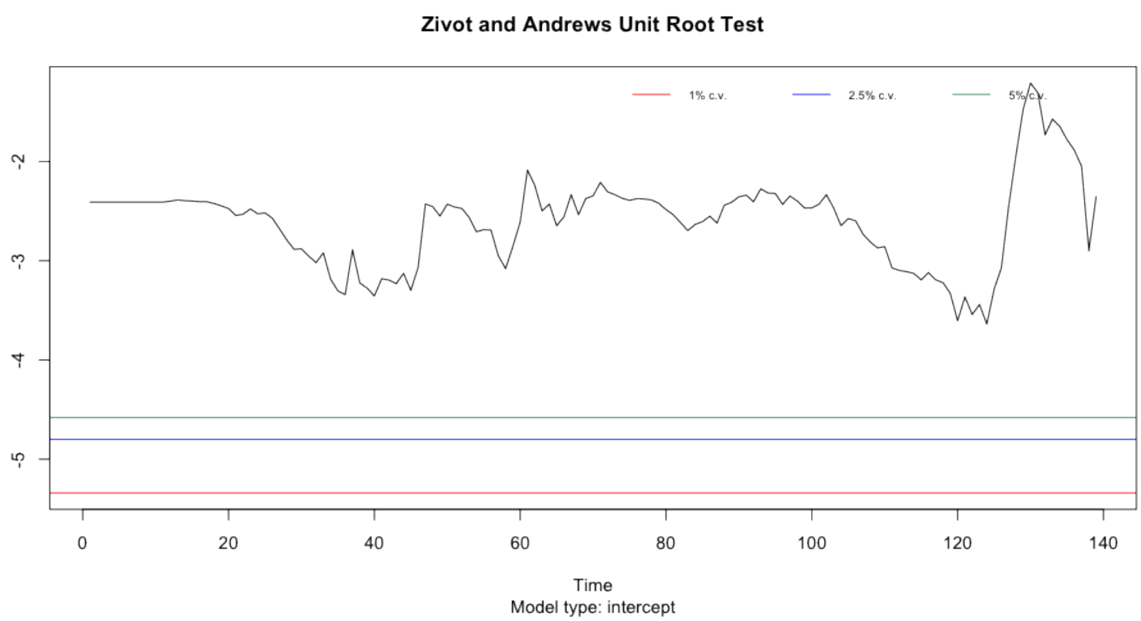


Figure 2.3: Unit Root Test.

Fig. 2.4. The Log Dividend Price Ratio.

This figure shows the plot of log dividend price ratio from 1871 to 2011. Prices and dividends are effectively in January 2000 dollars. More detailed information of data can be seen from Robert Shillers website.

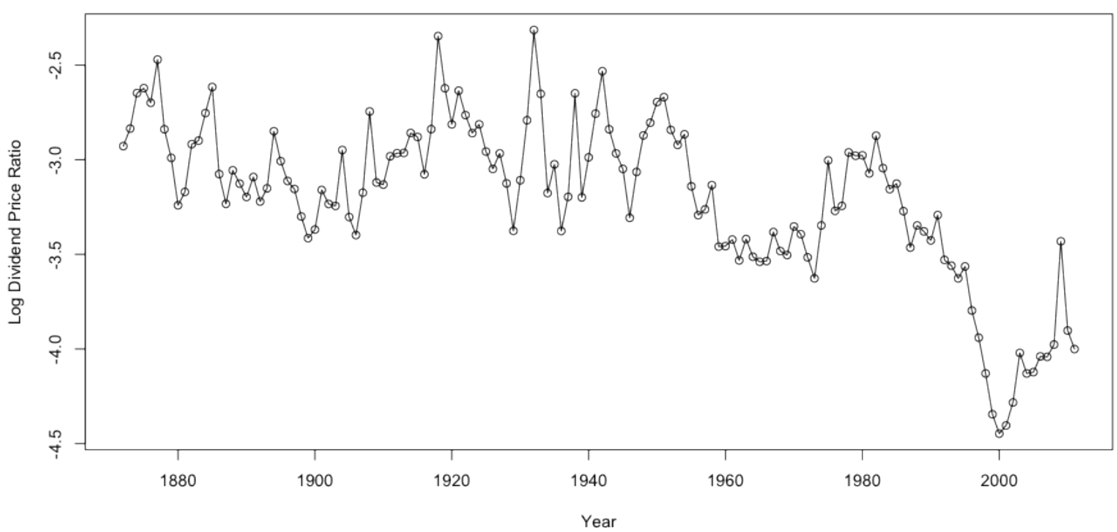


Figure 2.4: Log Dividend Price Ratio.

Fig. 2.5. The trend component in the log dividend price ratio.

This figure plots the trend component in log dividend price ratio and its 95% probability limits. I use Kalman filter to estimate the latent variable in the original data. Kalman filter is based on maximum likelihood method.

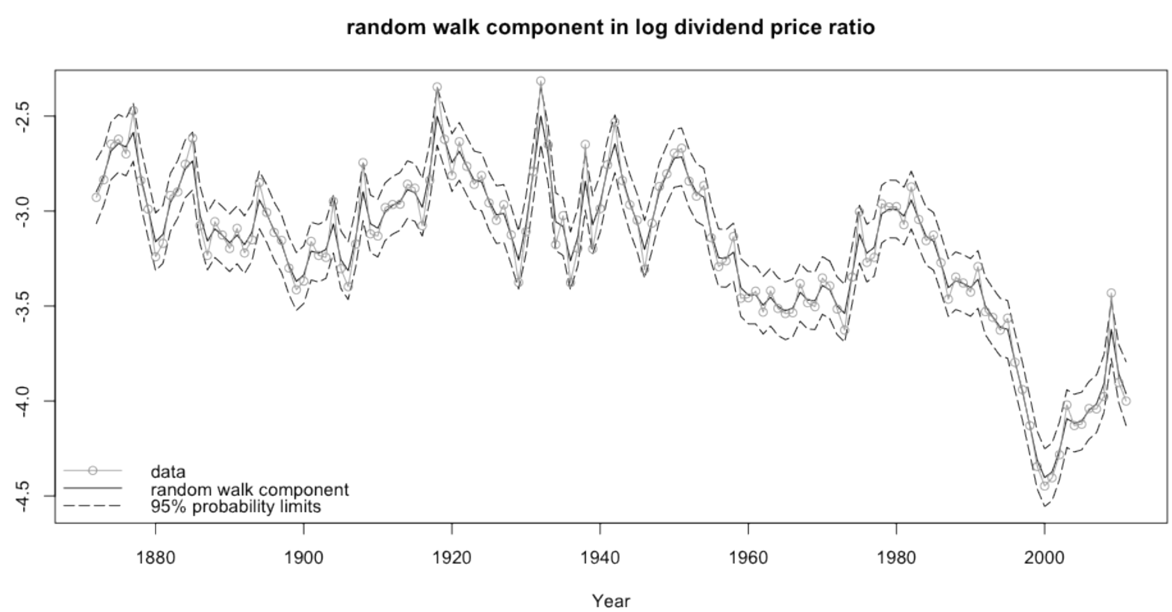


Figure 2.5: Trend Component.

Fig. 2.6. The cyclical component in the log dividend price ratio.

I get this cyclical component by subtracting the latent random walk variable from the original data. I use Kalman filter to estimate the latent variable in the original data. Kalman filter is based on maximum likelihood method.

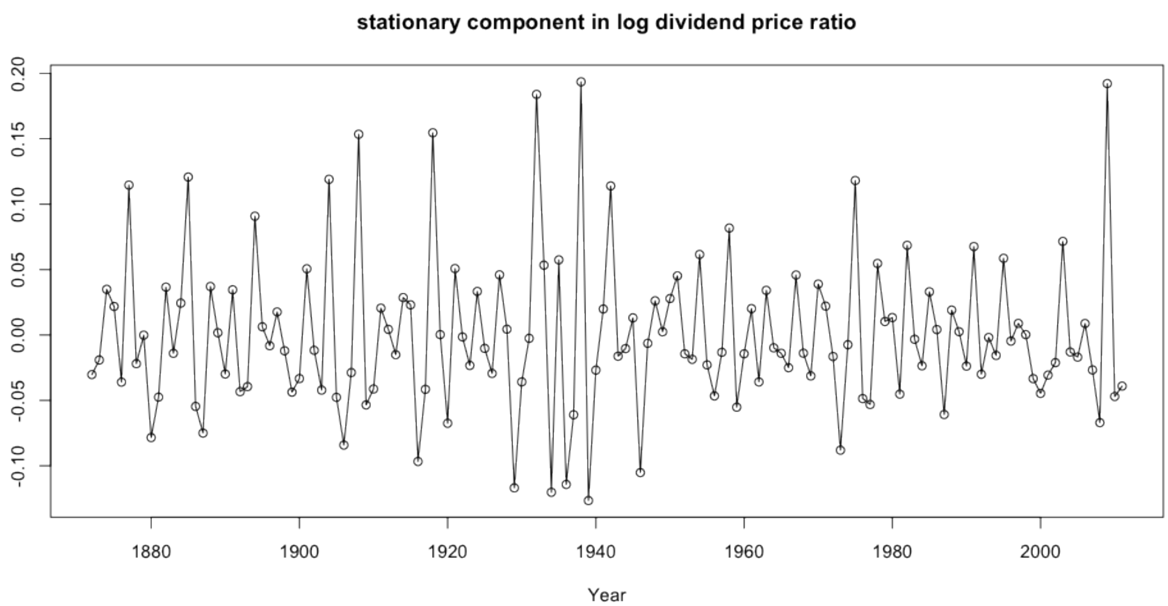
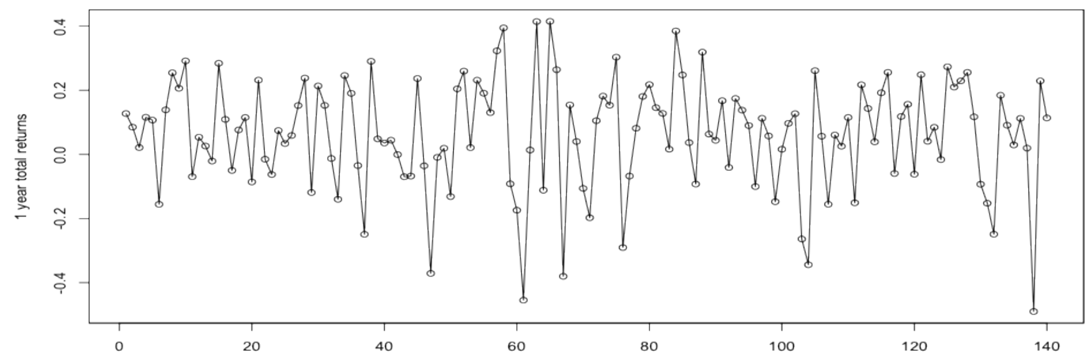


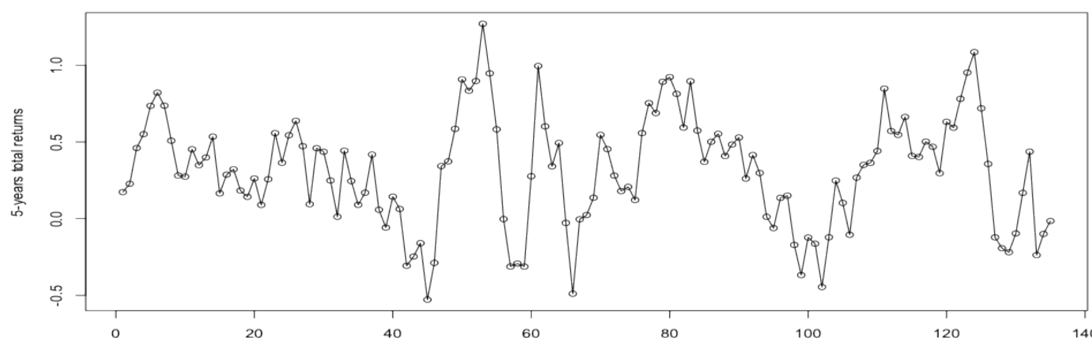
Figure 2.6: Cyclical Component.

Fig. 2.7. Aggregate returns in the future.

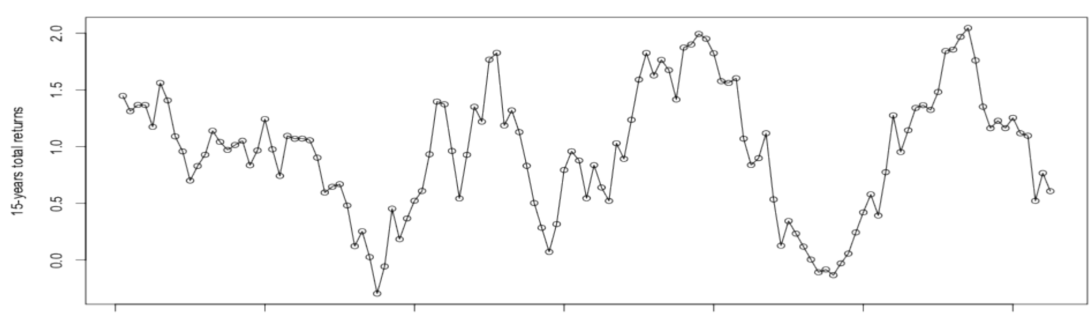
This figure shows the total returns in one year ahead, five years ahead and fifteen years ahead. As can been seen from the graph, the time series is getting more and more persistent as the forecast horizon increases.



(a) 1-year.



(b) 5-year.



(c) 15-year.

Figure 2.7: Aggregate returns in the future

Table 2.1

Predictability of whole sample (1872-2012).

This table presents prediction results for total return (r), capital gain return (Δp) and dividend growth (Δd) from 1872 to 2012. T-statistics are in the parentheses under the coefficients. R-squares are listed following coefficients and t-statistics. The first three columns present prediction results using the stationary component extracted from the log dividend price ratio. The last three columns present prediction results using the log dividend price ratio. The significant codes are "***" for the level of 0.01, "**" for 0.05, and "*" for 0.10.

Table 2.1: Statistics of variables

Coef & T-stat						
	Stationary Component			DP Ratio		
	$b_r^{(k)}$	$b_p^{(k)}$	$b_d^{(k)}$	$b_r^{(k)}$	$b_p^{(k)}$	$b_d^{(k)}$
K=1	1.33***	1.33***	-1.26***	0.06*	0.03	-0.08***
	(5.95)	(5.69)	(-9.45)	(1.86)	(0.94)	(-3.39)
K=5	2.35***	2.41***	-0.82**	0.34***	0.22***	-0.11**
	(4.62)	(4.56)	(-2.22)	(4.85)	(2.80)	(-2.15)
K=10	1.73**	1.73**	-0.45	0.61***	0.38***	-0.01
	(2.38)	(2.29)	(-1.09)	(5.90)	(3.25)	(-0.12)
K=15	2.05**	2.17**	-1.09**	1.01***	0.72***	-0.03
	(2.47)	(2.40)	(-2.43)	(7.10)	(4.19)	(-0.29)
K=20	1.53*	1.64*	-0.61	1.15***	0.90***	0.10
	(1.76)	(1.71)	(-1.41)	(7.48)	(4.74)	(1.09)
R^2						
	$b_r^{(k)}$	$b_p^{(k)}$	$b_d^{(k)}$	$b_r^{(k)}$	$b_p^{(k)}$	$b_d^{(k)}$
K=1	20.6%	19.1%	39.4%	2.5%	0.6%	7.7%
K=5	13.9%	13.5%	3.6%	15.0%	5.6%	3.3%
K=10	4.2%	3.9%	0.9%	21.4%	7.6%	0.0%
K=15	4.7%	4.5%	4.6%	29.1%	12.5%	0.0%
K=20	2.6%	2.4%	1.7%	32.2%	16.0%	1.0%

Table 2.2

Predictability till end of war (1872-1945)

This table presents prediction results for total return (r), capital gain return (Δp) and dividend growth (Δd) from 1872 to 1945. T-statistics are in the parentheses under the coefficients. R-squares are listed following coefficients and t-statistics. The first three columns present prediction results using the stationary component extracted from the log dividend price ratio. The last three columns present prediction results using the log dividend price ratio. The significant codes are "***" for the level of 0.01, "**" for 0.05, and "*" for 0.10.

Table 2.2: Statistics of variables

Coef & T-stat						
	Stationary Component			DP Ratio		
	$b_r^{(k)}$	$b_p^{(k)}$	$b_d^{(k)}$	$b_r^{(k)}$	$b_p^{(k)}$	$b_d^{(k)}$
K=1	0.91***	0.90***	-1.66***	0.07	0.04	-0.41***
	(3.10)	(2.92)	(-9.37)	(0.81)	(0.45)	(-7.72)
K=5	2.03***	2.12***	-0.86	0.65***	0.60***	-0.37**
	(3.37)	(3.31)	(-1.62)	(4.13)	(3.48)	(-2.61)
K=10	1.34	1.37	-0.36	0.72***	0.62***	-0.09
	(1.64)	(1.59)	(-0.57)	(3.80)	(2.98)	(-0.59)
K=15	2.05**	2.15*	-1.21	1.17***	1.07***	-0.24
	(2.02)	(1.97)	(-1.61)	(5.82)	(4.66)	(-1.29)
K=20	0.88	0.9	-0.71	1.12***	1.05***	-0.03
	(0.89)	(0.84)	(-1.07)	(6.19)	(4.93)	(-0.19)
R^2						
	$b_r^{(k)}$	$b_p^{(k)}$	$b_d^{(k)}$	$b_r^{(k)}$	$b_p^{(k)}$	$b_d^{(k)}$
K=1	11.9%	10.7%	55.3%	0.9%	0.3%	45.6%
K=5	14.5%	14.0%	3.8%	20.3%	15.3%	9.2%
K=10	4.2%	3.9%	0.5%	18.9%	12.5%	0.6%
K=15	6.5%	6.3%	4.3%	36.8%	27.2%	2.8%
K=20	1.5%	1.4%	2.2%	42.4%	31.9%	0.1%

Table 2.3

Predictability since 1926 (1926-2012)

This table presents prediction results for total return (r), capital gain return (Δp) and dividend growth (Δd) from 1926 to 2012. T-statistics are in the parentheses under the coefficients. R-squares are listed following coefficients and t-statistics. The first three columns present prediction results using the stationary component extracted from the log dividend price ratio. The last three columns present prediction results using the log dividend price ratio. The significant codes are "***" for the level of 0.01, "**" for 0.05, and "*" for 0.10.

Table 2.3: Statistics of variables

Coef & T-stat	Stationary Component			DP Ratio		
	$b_r^{(k)}$	$b_p^{(k)}$	$b_d^{(k)}$	$b_r^{(k)}$	$b_p^{(k)}$	$b_d^{(k)}$
K=1	1.48***	1.47***	-1.12***	0.08.	0.05	-0.05**
	(4.88)	(4.69)	(-7.19)	(1.80)	(1.11)	(-2.02)
K=5	2.52***	2.58***	-0.46	0.36***	0.23**	-0.06
	(3.72)	(3.69)	(-1.06)	(4.13)	(2.49)	(-1.05)
K=10	2.01*	2.02*	-0.39	0.67***	0.45***	0.01
	(1.98)	(1.93)	(-0.87)	(5.17)	(3.09)	(0.20)
K=15	2.29*	2.46*	-0.64	1.22***	0.94***	0.15
	(1.88)	(1.90)	(-1.16)	(6.19)	(3.98)	(1.35)
K=20	1.99	2.16	-0.55	1.50***	1.29***	0.28**
	(1.51)	(1.55)	(-0.95)	(6.90)	(5.03)	(2.31)
R^2						
	$b_r^{(k)}$	$b_p^{(k)}$	$b_d^{(k)}$	$b_r^{(k)}$	$b_p^{(k)}$	$b_d^{(k)}$
K=1	22.1%	20.8%	38.1%	3.7%	1.4%	4.6%
K=5	14.9%	14.7%	1.4%	17.8%	7.3%	1.4%
K=10	5.0%	4.8%	1.0%	26.6%	11.5%	0.1%
K=15	4.9%	5.0%	1.9%	35.7%	18.7%	2.6%
K=20	3.4%	3.6%	1.4%	42.6%	28.4%	7.7%

Table 2.4

Predictability since 1946 (1946-2012)

This table presents prediction results for total return (r), capital gain return (Δp) and dividend growth (Δd) from 1946 to 2012. T-statistics are in the parentheses under the coefficients. R-squares are listed following coefficients and t-statistics. The first three columns present prediction results using the stationary component extracted from the log dividend price ratio. The last three columns present prediction results using the log dividend price ratio. The significant codes are "***" for the level of 0.01, "**" for 0.05, and "*" for 0.10.

Table 2.4: Statistics of variables

Coef & T-stat	Stationary Component			DP Ratio		
	$b_r^{(k)}$	$b_p^{(k)}$	$b_d^{(k)}$	$b_r^{(k)}$	$b_p^{(k)}$	$b_d^{(k)}$
K=1	2.26***	2.25***	-0.36**	0.10**	0.07	0.01
	(6.80)	(6.67)	(-2.19)	(2.18)	(1.54)	(0.39)
K=5	3.60***	3.48***	0.03	0.48***	0.36***	0.02
	(3.18)	(3.13)	(0.06)	(5.14)	(3.63)	(0.54)
K=10	3.36*	3.42.	-0.38	1.00***	0.82***	0.12.
	(1.88)	(1.93)	(-0.62)	(6.80)	(5.00)	(1.83)
K=15	3.58	3.63	0.36	1.86***	1.68***	0.50***
	(1.57)	(1.54)	(0.46)	(8.27)	(6.28)	(5.39)
K=20	3.45	3.6	0.32	2.09***	2.11***	0.56***
	(1.42)	(1.40)	(0.35)	(8.32)	(7.41)	(4.61)
	(1.51)	(1.55)	(-0.95)	(6.90)	(5.03)	(2.31)
R^2						
K=1	42.3%	41.4%	7.1%	7.0%	3.6%	0.2%
K=5	14.8%	14.4%	0.0%	31.3%	18.5%	0.5%
K=10	6.1%	6.5%	0.7%	46.1%	31.6%	5.8%
K=15	4.9%	4.7%	0.4%	58.7%	45.1%	37.7%
K=20	4.5%	4.4%	0.3%	61.7%	56.1%	33.1%

Table 2.5

Predictability R^2 comparison

This table presents prediction R^2 using the stationary component in the log dividend price ratio and the log dividend price ratio itself for returns in the future. K represents years.

Table 2.5: R^2

	Stationary Component		DP Ratio	
K=1				
1872-1945	11.9%	55.3%	0.9%	45.6%
1946-2012	42.3%	7.1%	7.0%	0.2%
K=20				
1872-1945	1.5%	2.2%	42.4%	0.1%
1946-2012	4.5%	0.3%	61.7%	33.1%

Chapter 3

THE $S - SHAPE$ FACTOR AND BOND RISK PREMIA

3.1 Introduction

There is a perception in the current literature that the first three yield factors, namely the *level*, *slope* and *curvature* factors, are enough for predictive exercises in macroeconomic and finance area. The *level* factor is approximated by the short-rate or the risk-free rate, the *slope* factor is approximated by the difference between long-term and short-term rates, and the *curvature* factor is often approximated by a butterfly spread (a mid-maturity rate minus a short- and long-rate average). The first factor, *level*, contributes most to the variation of yields and is thus non-negligible; the second factor, *slope* or *yield spread*, is proven highly significant in predictive regressions for multiple economic variables (e.g. growth rate of GDP, consumption, inflation, etc.); the third factor, *curvature* factor, has smaller variation compared to the first two factors and is insignificant in most predictive regressions. Thus, it seems natural to conclude that higher order yield factors, which have even smaller variations compared to the third factor, will not be useful for forecasting exercises.

However, this paper finds the fourth yield factor does have significant predictive power for excess bond returns. The fourth factor has a shape of S and represents how much S-shape like the yield curve is. If the shape of the yield curve becomes more S-shape like, the change would be reflected by the $S - shape$ factor. In another word, the $S - shape$ factor represents the interest change rate under median maturities is different from interest change rates under short and long maturities.

To test the predictive power of the $S - shape$ factor, we fit two nested models to data: the benchmark model includes the first four yield factors — *level*, *slope*, *curvature*, and $S - shape$ — while the baseline model only includes the first three yield factors. Comparing these two models, we can see how much more predictability the $S - shape$ factor brings for predicting excess bond return. To prove that the empirical evidence of predictability is reliable and robust, we use two data sets from the Fama-Bliss and the Federal Reserve, five sample periods which are 1964–2007, 1964–1984, 1985–2007, 1964–2002, and 1985–2012 to forecast bonds starting with four different maturities, $n = 5, 4, 3, 2$.

The in-sample and out-of-sample statistics present consistent results that the $S - shape$ factor is a significant predictor for excess bond returns: the coefficients before the $S - shape$ factors are significant under most cases evaluated by t-statistics; the predictive power stands robust for different data sets, different sample periods and bonds with different maturities; the increase of R^2 because of the additional $S - shape$ factor is notable. For example, for bonds that bought with 3 years maturities and sold with 2 years maturities, during sample period 1964-2007, the in-sample R^2 increase from 15% to 28%, which is a 87% increase; for forecasting period 1985-2007, the out-of-sample R^2 increase from 13% to 26%, which is a 100% increase.

To see how economic meaningful the $S - shape$ factor is, this paper makes use of analysis in [5]. [5] calculate how much more expected return can be increased for a typical investor using forecasting variables. The typical investor is assumed to have a single-period investment horizon and mean-variance preference. Setting the risk-averse coefficient to 1, this paper finds that according to the Fama-Bliss data, the $S - shape$ factor can increase a typical investor's absolute return in the range of 5%–15%, which is 13%–48% proportionally; according to the Federal Reserve data, the $S - shape$ factor can increase the absolute return in the range of 9%–34%, which is 28%–131% proportionally.

As evidenced by empirical statistics, smaller variations of higher order yield factors do not necessarily mean that they are less useful for predictive purposes or less meaningful economically. Each yield factor contains different economic meaning, and the importance of economic implication that each factor conveys does not necessarily be proportional to its variance. The variance of the *slope* factor is close to 1% of that of the *level* factor but the *slope* factor or *yield spread* is found to be a much more significant predictor in macroeconomic forecasting applications than the *level* factor. The variation of the $S - shape$ factor is comparable to that of the *curvature* factor but this paper shows that the $S - shape$ factor has much stronger predictive power for bond risk premia than the *curvature* factor.

Yield factors are principal components of the yields matrix. Principal components of yields are derived from the variance-covariance matrix of yields through orthogonal transformation and are a set of linearly uncorrelated variables. The initial motivation of Principal Component Analysis (PCA) is to find the direction in the data with the most variation. Nowadays PCA has been applied in a variety of empirical settings. It has been developed as one popular latent variable approach or dynamic variable analysis. The benefits of PCA include that the common factors can be constructed without concerns on the structural instability of the data, and also that the extracted factors can be combined freely to construct other factors.

Research on prediction of economics variables using yields factors has been fruitful, especially using the second yield factor or the *yield spread*. For example, using the *yield spread*, [4], [7] and [14] forecast future short yields; [1] and [11] forecast real activity; [13] and [21] predict inflation; and [12] predict future recessions. More recently, [10] provide a macroeconomic interpretation of the yield factors by combining it with VAR dynamics for the macro-economy. There are also papers that investigate time-variation in the predictive power (see [2]), and for instability of it (see [25]).

A few other papers run predictive regressions or construct affine term structure models including the fourth yield factor. [8] regress average excess returns on all yield factors including the fourth factor. [19] point out that the fourth yield factor shows substantial correlation with the forward rate factor constructed by [7] and also with real economic growth as demonstrated in [16]. Both of these variables are known to have strong predictive power for excess bond returns. However, to the author's best knowledge, this paper is the first to focus on exploring the economic information in the fourth yield factor or the $S - shape$ factor and its predictive power for excess bond returns.

Another contribution of this paper is to connect yield factors with the "return-forecasting factor" in [7]. [7] regress excess bond returns on yield and forward rates and obtain a return-forecasting factor which is a linear combination of yields and forward rates. Their return-forecasting factor is influential in the bond return forecasting literature and serves as a benchmark factor in return forecasting exercises. [8] conclude that the return-forecasting factor is not spanned by the first three yield factors.

This paper derives theoretically that the return-forecasting factor is a restricted linear combination of all yield factors in the data. The connection is verified using empirical data. All yield factors derived from the Fama-Bliss data can explain 100% of the variation of the return-forecasting factor. This paper also finds that the return-forecasting factor has a high correlation with the *slope* factor and the $S - shape$ factor.

To compare the predictability of yield factors and the return-forecasting factors, this paper builds a third model which only includes the return-forecasting factor as a single predictor. As expected, both yield factors and the return-forecasting factor present excellent in-sample statistics for predicting excess bond returns. There are some advantages of using yield factors to predict economic variables instead of using the return-forecasting factor though. The first advantage is that it is straightforward to check which yield factor captures the most predictive information for excess bond returns and then research could focus more on these factors' economic implications than on others'. The second is that the estimation of loadings on yield factors is much less sensitive to the range of data used compared to the estimation of coefficients of the return-forecasting factor and is much more stable. The third is that the estimation of yield factors faces less econometric issues such as collinearity than the return-forecasting factor would face.

The rest of the paper is organized as follows. Section 2 gives a brief analysis on yield factors. Section 3 demonstrates the theoretical connection between yield factors and the return-forecasting factor and also supplies the empirical evidence on their correlation. Sections 4 and 5 present the in-sample and out-of-sample statistics for predictive models, aiming to show the significant predictive power of the $S - shape$ factor. Section 6 applies expected returns analysis used in [5] and calculates how much expected return increases for a typical investor because of the additional $S - shape$ factor. Section 7 concludes.

3.2 Functions of Yield Factors

As pointed out in the introduction, yield factors are principal components derived from the variance-covariance matrix of yields. One projects the data onto the directions of eigenvectors of the variance-covariance matrix to get principal components. The first eigenvector of the variance-covariance matrix of the data corresponds to the largest eigenvalue of it and the first principal component explains most variation of the data. The number of directions used depends on the specific situation. This section follows the notation in [24]. Let $\boldsymbol{Y}$ represent the $m * n$ matrix of yields with different maturities. The variance-covariance matrix of $\boldsymbol{Y}$ can be written as

$$var(\boldsymbol{Y}) = \Omega\Lambda\Omega^T \ ,$$

where Λ is a diagonal matrix of eigenvalues of the matrix $var(\boldsymbol{Y})$ and Ω is an orthogonal matrix whose columns are standardized eigenvectors. Principal components (PCs) can be defined by

$$PC = \Omega^T \boldsymbol{Y} \ , \tag{3.1}$$

Details on this method can be found in [17].

We use two data sets on yields to verify the robustness of our analysis. The first is the Fama-Bliss monthly data consisting of 1 through 5 year zero-coupon government bond prices from CRSP and the second is monthly observations of market yields on U.S. Treasury securities at 3-month, 6-month, 1-year, 3-year, 5-year and 10-year from the Federal Reserve. In this section, the only sample data range considered is 1964-2007. We will consider more subsamples in the section of forecasting excess bond returns.

3.2.1 Level, Slope, Curvature

The first three yield factors are analyzed intensively in literature. [20] are among the first to interpret the first three latent factors. They named the first three factors as "*level*", "*steepness*" and "*curvature*". These names deliver much of the intuition for what drive yields as shown in figure 3.1.

INSERT FIG. 3.1 NEAR HERE

Figure 3.1 plots the first four yield factors' loadings on yields since each yield factor can be denoted as a linear combination of yields. These loadings are just the eigenvectors of the variance-covariance matrix of yields— columns in Ω in equation 3.1. Each yield factor's loading is a vector of coefficients of yields of different maturities. We connect these points of loadings for each factor to make a line. Figure 3.1 shows us the shape of the fourth yield factor is stable S-shape like using different data sets.

INSERT FIG. 3.2 NEAR HERE

Figure 3.2 plots the time series of the first four yield factors during 1964–2007 using the data from the Federal Reserve. The *level* factor is the most persistent series with an autocorrelation of 0.99. The *slope* factor is the second persistent series with an autocorrelation of 0.96. The autocorrelation for the *curvature* and $S - shape$ factors are 0.87 and 0.84, respectively. The Fama-Bliss data also reveals that the first yield factor is the most persistent among all yield factors. The autocorrelations of the first four yield factors using the Fama-Bliss data are 0.99, 0.94, 0.60 and 0.43, respectively.

INSERT FIG. 3.3 NEAR HERE

As shown in figure 3.1, the loadings of the *level* factor is flat and thus, the *level* factor roughly represents an average of yields of all maturities. The change of the *level* factor represents a parallel shift of the yield curve as indicated in figure 3.3(a). The loadings of the *level* factor using the Fama-Bliss data are (0.44, 0.45, 0.45, 0.45, 0.45). They are (0.41, 0.41, 0.41, 0.41, 0.41, 0.40) using the Federal Reserve data. Suppose yields of all maturities go up by 0.20%, then according to equation 3.1, the value of the *level* factor would go up by (0.44+0.45+0.45+0.45+0.45)*0.20% or 0.45 according to the Fama-Bliss data and it would go up by (0.41+0.41+0.41+0.41+0.41+0.40)*0.20% or 0.49 according to the Federal Reserve data. The *level* factor goes up (down) if the overall level of yields goes up (down). Due to this fact, the *level* factor is often used as duration hedging in portfolio analysis as indicated in [20].

An important fact to note is that the parallel shifts in the yield curve do not cause other yield factors to change, which means this effect is uniquely captured by the *level* factor. Under the same assumption that yields of all maturity go up by 0.20%, values of the *slope*, *curvature* and $S - shape$ factors would barely change. In another word, the change of the overall level of yields is uniquely represented by the *level* factor.

The *slope* factor is widely used in macroeconomic forecasting exercises. One reason for its popularity is that it can be approximated by the difference between a long rate and a short rate or *yield spread* and *yield spread* proves to be very informative about the future economy. A higher long rate than the short rate is likely to indicate a positive future economy while a higher short rate than the long rate often forecasts a recession or economic slowdown.

The *slope* factor's loadings on yields are (0.74, 0.23, −0.10, −0.35, −0.52) according to the Fama-Bliss data and are (0.48, 0.41, 0.25, −0.17, −0.38, −0.60) according to the Federal Reserve data. Its loadings are positive on yields of short maturities and are negative on yields of long maturities. Also, the absolute values of loadings are relatively big at the tail and relatively small in the middle. As shown in figure 3.3(b), the *slope* factor reflects difference of yields under short maturities and yields under long maturities. In another word, it reflects changes in the slope of the yield curve. If the short-term rate increases while the long-term rate does not change, the value of the *slope* factor would increase and the yield curve would appear more flat; whereas if the long term rate increase while the short term rate does not change, the value of the *slope* factor would decrease and the yield curve would appear steeper.

As mentioned in the introduction, empirical research approximates the *level* factor by the short rate or the risk free rate, the *slope* factor by the difference between the long rate and the short rate, the *curvature* factor by the butterfly spread or a mid-maturity rate minus a short- and long-rate average. The first two latent factors have high correlation with their proxies while the third latent factor has a relatively low correlation with its proxy. For example, [1] find the *level* factor has a correlation of 0.96 with the short rate (three-month risk free rate) and we find that it is 0.98 using the Fama-Bliss data. The correlation between the *slope* factor and the *yield spread* is also close to 1. However, for the *curvature* factor the correlation between the latent factor and its empirical proxy is only around 0.5.

The *curvature* factor is also the least significant variable among the first three yield factors in predictive regressions. For example, [6] show that the first three yield factors can help predict exchange rate movements, with the *slope* factor being the most robust one, but movements in the *curvature* factor have a much smaller effect on exchange rates. [20]

interpret the *curvature* factor as representing how “hump-shaped” the yield curve is.

Figure 3.3(c) shows our interpretation on what the *curvature* factor represents. Its loadings on yields are (−0.48, 0.54, 0.46, −0.01, −0.52) using the Fama-Bliss data and are (−0.57, 0.02, 0.40, 0.50, 0.13, −0.50) using the Federal Reserve data. The factor has relatively large negative loadings at the tails and has relatively large positive loadings under median maturities. If yields at the short or long ends go up, the value of the *curvature* factor would decrease and the yield curve would become less hump-shaped. If yields under median maturities go up, the value of the *curvature* factor would increase and the yield curve would become more hump-shaped. Thus a higher value of the third principal component represents a more curvy yield curve as shown in figure 3.3(c) while other yield factors would not catch this effect as effectively. For example, under the same assumption, the value of the *slope* factor would not change much because the opposite signs of loadings at the short and long ends would offset the changes.

Current literature also builds connections between yield factors and macroeconomic variables. Since changes in the overall level of the yields or interest rates would change the value of the *level* factor, the *level* factor is often related to inflation or expected inflation. See [10], [27]. A higher level of inflation or expected inflation could be because the government is encouraging saving and investment and thus is increasing interest rates and a positive expected inflation rate or *level* factor is often a positive sign for future economy.

Meanwhile, literature connects the *slope* factor with real activity as its macroeconomic representation. See [12], [26] and [18]. One interpretation is that when central banks tighten monetary policy, the short rate increases and a recession often follows. Another interpretation is that lower long rate than short rate signals that peoples expectation on future short rates is low and the economy is likely to slow down. Hence, a flat or inverted yield curve is typically considered as a signal for an economic slowdown or a recession.

However, it is not as clear yet to see what it means for the economy when the yield curve becomes more hump-shaped. Current literature also finds it difficult to connect it with a specific macroeconomic variable. [20] find that changes in the *curvature* of the yield curve are associated with changes in yields volatility. However, as pointed out by [24], stochastic volatility behaves like a *curvature* factor in some estimated models but it turns out to be so persistent that it becomes the *level* factor in others. Also, results in [20] are difficult to replicate on more recent and non-U.S. data as pointed out by [9].

3.2.2 The $S - shape$ Factor

As seen in figure 3.1, the fourth yield factor has a shape of S which is the reason why we name it the $S - shape$ factor. Its loadings on yields are (0.15, −0.53, 0.23, 0.64, −0.49) according to the Fama-Bliss data and are (0.40, −0.20, −0.58, 0.45, 0.32, −0.39) according to the Federal Reserve data. If we separate different maturities of yields into four different regions: short, median short, median long and long, the value of the $S - shape$ factor would go up if yields under short or median long maturities increase, or if yields under median short maturities or long maturities decrease. Also notice that the sign of loadings in neighbor regions are opposite to each other. The loadings of short maturities are positive but the loadings of median short maturities are negative; the loadings of long maturities are negative but the loadings of median long maturities are positive. If the yield curve become more S-shape like, the change would be reflected in the $S - shape$ factor.

INSERT FIG. 3.4 NEAR HERE

Figure 3.3(d) shows our understanding of what $S-shape$ factor represents. The $S-shape$ factor measures yields change rate under "median short — median long" maturities are different from yields change rates under "short — median short" and "median long — long" maturities. If yield curve changes in the direction of figure 3.3(d), the curve would appear more S-shape like, also the change rate under median maturities would be much higher than that under both ends. For example, as show in historical yield curve of Nov. 2007 in figure 3.4(b), the yield curve is flat under short and long ends but has a clear upward trend under median maturities. The S-shape character of the yield curve is also documented in previous literatures including [22]. Meanwhile, it is important to notice that the S-shape of the yield curve can take another form as shown in figure 3.4(a).

The shape of the yield curve is often upward but takes more abnormal forms during recessions. It can be reverted, S-shaped, flat or displaying a mixed form. On the whole, the shape of the yield curve changes more frequently and more dramatically than what people would normally expect. As discussed in previous subsection, macroeconomic research ties the change of expected inflation to the *level* factor, the real output with business cycle to the *slope* factor, and nothing in particular to the *curvature* factor, it is still left as a question what macro variable should be tied to the $S-shape$ factor. Monetary policy seems a good candidate. However, monetary policy has no reason to just impact the $S-shape$ factor but not other yield factors. We leave this interesting question for further research.

3.3 Yield Factors and the "Return-Forecasting Factor"

This section demonstrates the theoretical connection between yield factors and the "return-forecasting factor" in [7] and verifies the correlation with empirical data. We build the theoretical connection using the Fama-Bliss data since that was the data used to create the return-forecasting factor in [7]. However, the theoretical relationship derived is in fact data-free. We use both the Fama-Bliss data and the Federal Reserve data to provide empirical evidence on the correlation.

3.3.1 Notation

The analysis here uses the same notation as [7]. The notation for log bond price is

$$p_t^{(n)} = \text{log price of } n\text{-year discount bond at time } t \text{ with face value 1.}$$

The parentheses are used to distinguish maturity from exponentiation in the superscript. The log yield is

$$y_t^{(n)} \equiv -\frac{1}{n} p_t^{(n)},$$

and the log forward rate at time t for loans between time $t+n-1$ and $t+n$ is

$$f_t^{(n)} \equiv p_t^{(n-1)} - p_t^{(n)},$$

and the log holding period return from buying an n-year bond at time t and selling it as an $(n-1)$-year bond at time $t+1$ is

$$r_{t+1}^{(n)} \equiv p_{t+1}^{(n-1)} - p_t^{(n)}.$$

The excess log return is denoted by

$$rx_{t+1}^{(n)} \equiv r_{t+1}^{(n)} - y_t^{(1)}.$$

The same letters without n index are used to denote vectors across maturity, e.g.,

$$\mathbf{rx}_{t+1} \equiv \left[rx_{t+1}^{(2)} \;\; rx_{t+1}^{(3)} \;\; rx_{t+1}^{(4)} \;\; rx_{t+1}^{(5)} \right]^T.$$

When used as right hand variables, these vectors include an intercept, e.g.,

$$\mathbf{y}_t \equiv \left[1 \;\; y_t^{(1)} \;\; y_t^{(2)} \;\; y_t^{(3)} \;\; y_t^{(4)} \;\; y_t^{(5)} \right]^T,$$

$$\mathbf{f}_t \equiv \left[1 \;\; y_t^{(1)} \;\; f_t^{(2)} \;\; f_t^{(3)} \;\; f_t^{(4)} \;\; f_t^{(5)} \right]^T.$$

Over bars are used to denote averages across maturity, e.g.,

$$\overline{rx}_{t+1} \equiv \frac{1}{4} \sum_{n=2}^{5} rx_{t+1}^{(n)}.$$

With this notation, [7] build the return-forecasting factor by running the regression

$$\frac{1}{4} \sum_{n=2}^{5} rx_{t+1}^{(n)} = \hat{\gamma}_0 + \hat{\gamma}_1 y_t^{(1)} + \hat{\gamma}_2 f_t^{(2)} + \cdots + \hat{\gamma}_5 f_t^{(5)} + \hat{\varepsilon}_{t+1},$$

or

$$\overline{\mathbf{rx}}_{t+1} = \hat{\boldsymbol{\gamma}}^T \mathbf{f}_t + \hat{\varepsilon}_{t+1}.$$

Where $\boldsymbol{\gamma}^T$ denotes a vector consisting of a constant and coefficients, $[\gamma_0 \;\; \gamma_1 \;\; \gamma_2 \;\; \gamma_3 \;\; \gamma_4 \;\; \gamma_5]$, and $\hat{\varepsilon}$ is the estimated regression residuals. The estimated return-forecasting factor is $\hat{\boldsymbol{\gamma}}^T \mathbf{f}_t$.

3.3.2 Theoretical Correlation Between PCs and CP

For the rest of the paper, we denote the return-forecasting factor as CP and the yield factors or principal components of the yields matrix as PCs. This section demonstrates the theoretical correlation between CP and PCs.

Since by definition,

$$CP_t \equiv \boldsymbol{\gamma}^T \boldsymbol{f}_t = \gamma_0 + \gamma_1 y_t^{(1)} + \gamma_2 f_t^{(2)} + \cdots + \gamma_5 f_t^{(5)},$$

and $y_t^{(1)} = -p_t^{(1)}$, $f_t^{(2)} = p_t^{(1)} - p_t^{(2)}$, $f_t^{(3)} = p_t^{(2)} - p_t^{(3)}$, $f_t^{(4)} = p_t^{(3)} - p_t^{(4)}$, $f_t^{(5)} = p_t^{(4)} - p_t^{(5)}$, one can show with substitution that CP can be denoted as a linear combination of log bond prices:

$$CP_t \equiv \gamma_0 + \gamma_1(-p_t^{(1)}) + \gamma_2(p_t^{(1)} - p_t^{(2)}) + \cdots + \gamma_5(p_t^{(4)} - p_t^{(5)}).$$

After collecting terms, it becomes:

$$CP_t \equiv \gamma_0 + (-\gamma_1 + \gamma_2)p_t^{(1)} + (-\gamma_2 + \gamma_3)p_t^{(2)} + \cdots + (-\gamma_4 + \gamma_5)p_t^{(4)} + (\gamma_5)p_t^{(5)}. \tag{3.2}$$

One can write the yield factors into the following equations according to equation 3.1:

$$\begin{aligned}
PC1_t &= \alpha_{1,0} + \alpha_{1,1}y_t^{(1)} + \alpha_{1,2}y_t^{(2)} + \alpha_{1,3}y_t^{(3)} + \alpha_{1,4}y_t^{(4)} + \alpha_{1,5}y_t^{(5)},\\
PC2_t &= \alpha_{2,0} + \alpha_{2,1}y_t^{(1)} + \alpha_{2,2}y_t^{(2)} + \alpha_{2,3}y_t^{(3)} + \alpha_{2,4}y_t^{(4)} + \alpha_{2,5}y_t^{(5)},\\
PC3_t &= \alpha_{3,0} + \alpha_{3,1}y_t^{(1)} + \alpha_{3,2}y_t^{(2)} + \alpha_{3,3}y_t^{(3)} + \alpha_{3,4}y_t^{(4)} + \alpha_{3,5}y_t^{(5)},\\
PC4_t &= \alpha_{4,0} + \alpha_{4,1}y_t^{(1)} + \alpha_{4,2}y_t^{(2)} + \alpha_{4,3}y_t^{(3)} + \alpha_{4,4}y_t^{(4)} + \alpha_{4,5}y_t^{(5)},\\
PC5_t &= \alpha_{5,0} + \alpha_{5,1}y_t^{(1)} + \alpha_{5,2}y_t^{(2)} + \alpha_{5,3}y_t^{(3)} + \alpha_{5,4}y_t^{(4)} + \alpha_{5,5}y_t^{(5)},
\end{aligned}$$

where the first, second, third, fourth and fifth yield factors are denoted as $PC1$, $PC2$, $PC3$, $PC4$ and $PC5$ respectively; α represents the loadings of yield factors on yields. Since $y_t^{(1)} = -p_t^{(1)}$, $y_t^{(2)} = -\frac{1}{2}p_t^{(2)}$, $y_t^{(3)} = -\frac{1}{3}p_t^{(3)}$, $y_t^{(4)} = -\frac{1}{4}p_t^{(4)}$ and $y_t^{(5)} = -\frac{1}{5}p_t^{(5)}$, yield factors can be rewritten into following equations:

$$PC1_t = \alpha_{1,0} - \alpha_{1,1}p_t^{(1)} - \frac{1}{2}\alpha_{1,2}p_t^{(2)} - \frac{1}{3}\alpha_{1,3}p_t^{(3)} - \frac{1}{4}\alpha_{1,4}p_t^{(4)} - \frac{1}{5}\alpha_{1,5}p_t^{(5)}, \tag{3.3}$$

$$PC2_t = \alpha_{2,0} - \alpha_{2,1}p_t^{(1)} - \frac{1}{2}\alpha_{2,2}p_t^{(2)} - \frac{1}{3}\alpha_{2,3}p_t^{(3)} - \frac{1}{4}\alpha_{2,4}p_t^{(4)} - \frac{1}{5}\alpha_{2,5}p_t^{(5)}, \tag{3.4}$$

$$PC3_t = \alpha_{3,0} - \alpha_{3,1}p_t^{(1)} - \frac{1}{2}\alpha_{3,2}p_t^{(2)} - \frac{1}{3}\alpha_{3,3}p_t^{(3)} - \frac{1}{4}\alpha_{3,4}p_t^{(4)} - \frac{1}{5}\alpha_{3,5}p_t^{(5)}, \tag{3.5}$$

$$PC4_t = \alpha_{4,0} - \alpha_{4,1}p_t^{(1)} - \frac{1}{2}\alpha_{4,2}p_t^{(2)} - \frac{1}{3}\alpha_{4,3}p_t^{(3)} - \frac{1}{4}\alpha_{4,4}p_t^{(4)} - \frac{1}{5}\alpha_{4,5}p_t^{(5)}, \tag{3.6}$$

$$PC5_t = \alpha_{5,0} - \alpha_{5,1}p_t^{(1)} - \frac{1}{2}\alpha_{5,2}p_t^{(2)} - \frac{1}{3}\alpha_{5,3}p_t^{(3)} - \frac{1}{4}\alpha_{5,4}p_t^{(4)} - \frac{1}{5}\alpha_{5,5}p_t^{(5)}, \tag{3.7}$$

Equations 3.2 — 3.7 clearly indicate that both CP and PCs can be expressed as linear combinations of log bond prices of different maturities. Therefore, the log bond prices serve as the foundation of the theoretical connection between the return-forecasting factor and principal components.

As shown in equations 3.3 — 3.7, we have five left-hand side variables and five right-hand side variables. Solving them as multiple equations, we can in fact denote log bond prices p_t as linear combinations of principal components PCs. Then back to equation 3.2, we can rewrite the return-forecasting factor as a linear combination of principal components:

$$CP_t = \beta_0 + \beta_1 PC1_t + \beta_2 PC2_t + \beta_3 PC3_t + \beta_4 PC4_t + \beta_5 PC5_t. \tag{3.8}$$

Thus, we have shown that the return-forecasting factor is a restricted linear combination of all yield factors in the data. Note that this analysis works with arbitrary data sets with any numbers of yield factors.

3.3.3 Empirical Evidence

This subsection supplies empirical evidence on the correlation between the return-forecasting factor and yield factors. To do this, we continue to use both the Fama-Bliss and the Federal Reserve data. The sample period in this subsection is still set to 1964-2007.

[8] conclude that the return-forecasting factor is not spanned by the standard *level*, *slope* and *curvature* factors. Their conclusion is reasonable since the return-predicting factor is not a linear conbination of those three yield factors and those three factors do not explain all the variation of the return-forecasting factor.

However, as shown in the previous subsection, the return-forecasting factor is in fact a linear combination of all yield factors in the data if the return-forecasting factor and yield factors are derived from the same data set. Moreover, the return-forecasting factor is shown to be highly correlated with the *slope* factor and the $S - shape$ factor. According to the Fama-Bliss data, the return-forecasting factor has a correlation of -0.76 with the *slope* factor and a correlation of 0.52 with the $S - shape$ factor. Meanwhile, the correlation with the *level* factor ranks the third — 0.30, and the correlations with the *curvature* factor and the fifth factor are 0.25 and -0.05, respectively. According to the Federal Reserve data, the return-forecasting factor has a correlation of -0.74 with the *slope* factor and a correlation of 0.28 with the $S - shape$ factor. The correlation with the *level* factor is 0.24 and the correlations with the *curvature* factor, the fifth factor and the sixth factor are 0.06, 0.02 and -0.03, respectively.

INSERT TABLE. 3.1 NEAR HERE

Table 3.1 presents statistics of five univariate regressions and two multivariate regressions using the Fama-Bliss data. We regress the return-forecasting factor onto yield factors to verify the derived correlation. The regressed coefficients of yield factors are stable: the coefficients of the *level* factor are around 0.003, the coefficients of the *slope* factor are around -0.06, the coefficients of the *curvature* factor are around 0.13, and the coefficients of the $S-shape$ factor are around 0.39. $\overline{R}^2$ of the second multivariate regression which includes the first four yield factors as right-hand variables can be rounded to 100%, which means the first four yield factors explain almost all the variation of the return-forecasting factor. Univariate regressions also show that the variation of the return-forecasting factor can be explained mostly by the *slope* factor and the $S-factor$ with $\overline{R}^2$ of 57% and 27%, respectively. The *level* factor and and the *curvature* factor can explain 9% and 6% of the variation of the yield curve, respectively. The fifth principal components is insignificant in the univariate regression. Comparing $\overline{R}^2$ of two multivariate regressions in table 3.1 , we see that the $S-shape$ factor plays an important role in explaining the variation of the return-forecasting factor by adding almost 30% $\overline{R}^2$.

INSERT TABLE. 3.2 NEAR HERE

Table 3.2 presents statistics with similar implications with table 3.1 by using yield factors derived from the Federal Reserve treasury yields. The return-forecasting factor is still derived from the Fama-Bliss data. Since the return-forecasting factor and yield factors are now derived from different data sets and contain different sets of information, the $\overline{R}^2$ in the multivariate regressions can not reach 100%. However, table 3.2 verifies that the *slope* factor has the highest correlaton with the return-forecasting factor, -0.74, the $S-shape$ factor ranks the second with a correlation of 0.28, and the *level* factor ranks the third with a correlation of 0.24. Other factors do not have notable correlations with the return-forecasting factor. It is 0.06, 0.02 and -0.03 for the third, fifth and sixth factors respectively.

Given the high correlation between the return-forecasting factor and yield factors, if the return-forecasting factor has predictive power for excess bond returns, it is reasonable to infer that yield factors have predictive power too. In fact, the inference is correct and yield factors do have significant predictive power for excess bond returns. The difference lies in that the estimation of the return-forecasting factor is much less stable than the estimation of yield factors and is very data dependent. Recall that the definition of the return-forecasting factor is a linear combination of yield and forward rates. The coefficients of these rates are estimated by running a linear regression of average excess bond returns on these rates. For different sample periods, the estimated coefficients may change substantially and in fact they do. Yield factors, on the other hand, are estimated through decomposing the variance-covariance matrix of the yields and can be taken as non-stochastic. The estimated loadings of yield factors stay the same during different sample periods.

The difficulty of estimating the return-forecasting factor becomes more serious when working with large datasets that have ten or more yields with different maturities. The econometric problem of multicollinearity will show up when we include a large number of yields and forward rates on the right hand side. The estimation of coefficients would be very unstable because the right hand side variables are highly correlated. Yield factors, on the other hand, do not have to face this problem. We will also show in the following section that yield factors are robust predictors for excess bond returns across different datasets and the $S - shape$ factor adds on to the predictability significantly.

3.4 Bond Returns Forecast

This section presents the in-sample and out-of-sample statistics of predictive regressions for excess bond returns. The main purpose of this section is to verify the predictive power of the $S - shape$ factor and also to compare the predictive ability between the return-forecasting factor and yield factors.

3.4.1 In-Sample Forecast

In order to compare their predictabilities, we make use of three regression models. $Model\,1$ is a baseline model that include the first three yield factors as predictors. $Model\,2$ is the benchmark model which includes predictive factors in $Model\,1$ plus the $S - shape$ factor. We can check the marginal contribution for prediction of the $S - shape$ factor by comparing these two models. $Model\,3$ just contains a single predictor — the return-forecasting factor. Section 3 has shown that using the Fama-Bliss data, the first four yield factors can explain almost all the variation in the return-forecasting factor, and by comparing $Model\,2$ and $Model\,3$, we can see how yield factors and the return-forecasting factor perform relatively in predictive regressions.

The mathematic expressions for the three predictive regressions are:

$$\begin{aligned} Model\,1: \; rx_{t+1}^{(n)} =& \delta_{1,0} + \delta_{1,1}PC1_t + \delta_{1,2}PC2_t + \delta_{1,3}PC3_t + \varepsilon_{t+1}^{(n)}, \\ Model\,2: \; rx_{t+1}^{(n)} =& \delta_{2,0} + \delta_{2,1}PC1_t + \delta_{2,2}PC2_t + \delta_{2,3}PC3_t + \delta_{2,4}PC4_t + \varepsilon_{t+1}^{(n)}, \\ Model\,3: \; rx_{t+1}^{(n)} =& \delta_{3,0} + \delta_{3,1}CP_t + \varepsilon_{t+1}^{(n)}, \end{aligned}$$

Following [7], this section presents regression results of excess log bond returns with four different maturities: $n = 2, 3, 4, 5$. $N = 2$ represents the case of buying a 2-year bond and selling it as a 1-year bond, and $n = 3$ represents the case of buying a 3-year bond and selling it as a 2-year bond. Similar meanings apply to the cases for $n = 4$ and $n = 5$. Figure 3.5 plots average excess bond returns $\overline{rx}$ and the lagged $S - shape$ factor derived from the Fama-Bliss data. These two time series have a correlation of 0.28. Figure 3.6 plots average excess bond returns $\overline{rx}$ and the lagged $S - shape$ factor derived from the Federal Reserve data. These two time series have a correlation of 0.34. Interestingly, although average excess bond returns $\overline{rx}$ is derived from the Fama-Bliss data, it has a higher correlation with the fourth yield factor derived from the Federal Reserve data than with the fourth yield factor derived from the Fama-Bliss data.

INSERT FIG. 3.5 NEAR HERE

INSERT FIG. 3.6 NEAR HERE

It is customary to check for the stability of regressors, so we include different sample periods into regressions. [7] use the sample data of 1964–2002. We consider this sample period for comparison and also divide the whole sample period from 1964 into some other subsample periods for different considerations. The great recession post–2007 may have changed many economic variables' usual meanings. Many financial ratios such as dividend price ratio lose their usual predictive power during this financial crisis period. The benchmark sample period is set to 1964–2007 to avoid the abnormal impact of financial crisis. Another potential interesting subsample is 1985–2007 since the period of 1985–2007 is called "Great Moderation". The subsample period pre–1985 is also considered. In addition, we present statistics during the period of 1985–2012 to show the impact of the financial crisis. We follow previous research's conventions to assume that the return-forecasting factor, excess returns and yield factors are stationary.

For regressions including yield factors as regressors, we apply the Ordinary Least Squares (OLS) method to estimate the coefficients and the method in [23] to fix the standard error estimation problem caused by heteroskedasticity and autocorrelation among errors. [23] use bartlett kernel and automaticly choose the bandwidth considering correlations among errors. For regressions including the return-forecasting factor as a singular variable, we use the two-step OLS to estimate the coefficients and the method in [15] to fix the standard error estimation problem caused by generated regressors.

INSERT TABLE 3.3 NEAR HERE

The stability of estimated coefficients of predictors reflects predictors' predictive ability. The $S-shape$ factor and the return-forecasting factor have relatively more stable estimated coefficients than the *curvature* factor. From table 3.3 we can see that there is a sign change for the coefficients of the *curvature* factor from 1964–1984 to 1985–2007. For bonds with different maturities, the *curvature* and the $S-shape$ factor have more predictive power for excess bond returns during pre-"Great Moderation" period than during "Great Moderation" period. The return-forecasting factor appears to be a significant predictor during post-1985 periods for bond with different maturities. During pre–1985 period, the return-forecasting factor is a significant predictor for bonds with maturities $n = 4$ and $n = 5$, but not for bonds with maturities $n = 2$ and $n = 3$.

Across different data sets, the $S - shape$ factor proves to be a much stronger predictor than the *curvature* factor. Yield factors derived from the Fama-Bliss data appear more significant in predictive regressions than yield factors derived from the Federal Reserve data, which is within the expectation, since the dependent variables in the regressions —

excess bond returns are derived from the Fama-Bliss data.

INSERT TABLE 3.4 NEAR HERE

Table 3.4 compares performances of models in term of the in-sample $\overline{R}^2$ and shows what is the marginal gain in the $\overline{R}^2$ with an additional predictor: the $S-shape$ factor. $Model\,2$ with the $S-shape$ factor performs better than $model\,1$ in all cases: the $S-shape$ factor significantly increases the $\overline{R}^2$. For example, for bonds with $n=3$, during period 1964–2007, according to the Fama-Bliss data, the $S-shape$ factor increases the in-sample $\overline{R}^2$ from 18% to 27%, which is a 50% increase ; according to the Federal Reserve data, the increase is from 15% to 28%, which is a 87% increase. For bonds with all maturities, comparing to the Fama-Bliss data, the Federal Reserve data indicates a larger amount of increase in the $\overline{R}^2$ because of the $S-shape$ factor.

$Model\,2$ using the Fama-Bliss data is not supposed to outperform $model\,3$ since we have shown in previous section that the return-forecasting factor contains the information of all yield factors in the Fama-Bliss data. However, as we can see from table 3.4, the third column and the six column have similar values, which means, $model\,2$ performs as good as $model\,3$ without considering the fifth factor.

Comparing $model\,2$'s performances under two datasets, surprisingly we find that the in-sample $\overline{R}^2$ of $model\,2$ using the Federal Reserve data is even bigger than the in-sample $\overline{R}^2$ using the Fama-Bliss data, even given that the Fama-Bliss data is where excess bond returns are derived from. Also, $Model\,2$ using the Federal Reserve data outperforms $model\,3$ during almost all subsamples except the subsample period covering the financial crisis.

Statistics of sample period 1964–2002 give similar implications with those of period 1964–2007. [7] use data of 1964–2002 and forecast average excess bond returns with an R^2 around 35%. We are able to verify their finding in table 3.4. Also, the in-sample $\overline{R}^2$ of each model during 1964–2002 is bigger than the in-sample $\overline{R}^2$ during 1964–2007. This is consistent with the fact that all three models perform better during the earlier period 1964-1984 than during the later period 1985–2007. There is a significant decrease in the $\overline{R}^2$ comparing pre- and post-1985. For example, for bonds with $n=2$, comparing sample period 1964–1984 with 1985–2007, according to the Federal Reserve data, the in-sample R^2

drops from 30% to 17% for *model* 1, it drops from 52% to 29% for *model* 2; for *model* 3, the in-sample R^2 of drops from 36% to 22%. On the whole, according to the Federal Reserve data, the in-sample R^2 during post–1985 are around one half of the R^2 during post–1985.

To summarize, bonds with different maturities give similar implications. First, both the $S - shape$ factor and the return-forecasting factor prove to be strong predictors for predicting bond risk premia. Second, predictors are more useful during the pre–1985 period than during the "Great Modification" or financial crisis period. Third, the $S - shape$ factor adds significant predictive power to the regressions and the first four yield factors together can outperform the return-forecasting factor.

3.4.2 Out-of-Sample Forecast

For out-of-sample performances, we focus on comparing *model* 1 with *model* 2 to emphasize the additional predictive power brought by the $S - shape$ factor. Among different methods to measure out-of-sample performances, we choose to calculate the Root Mean Squared Error (RMSE) and the out-of-sample R^2. RMSE is a standard method widely used in measuring the out-of-sample performances (see [3]). Meanwhile, calculation of the out-of-sample R^2 makes it convenient to compare with the in-sample R^2.

Our way to calculate the out-of-sample RMSE and R^2 is standard: regress the true excess bond returns on fitted excess bond returns and extract the residuals and the R^2 of fitting. The fitted excess return at time t is estimated by multiplying coefficients estimated through time $t - 1$ with predictors at time t. We use recursive regressions starting from 20 years. Since the sample starts from 1964, the forecast period starts from 1985.

INSERT TABLE 3.5 NEAR HERE

Table 3.5 presents the RMSE of *model* 1 and *model* 2. The RMSE of *model* 2 is smaller than the RMSE of *model* 1 in all cases. Table 3.5 also reports the ratio of RMSE of these two models. All ratios are smaller than one, which indicates that *model* 2 creates smaller estimates errors and predicts more accurately compared to *model* 1. Data from the Federal Reserve provides even smaller RMSE ratios than data from the Fama-Bliss and even stronger evidence that the $S - shape$ factor is a useful predictor.

INSERT TABLE 3.6 NEAR HERE

Table 3.6 presents the out-of-sample R^2 of *model* 1 and *model* 2. The out-of-sample R^2 for *model* 1 are around 20% using the Fama-Bliss data and are around 16% using the Federal Reserve data; for *model* 2, the out-of-sample R^2 are around 25% using the Fama-Bliss data and are around 30% using the Federal Reserve data. According to the Fama-Bliss data, there are around 5% increases in the out-of-sample R^2 comparing *model* 2 with *model* 1 , and according to the Federal Reserve data, the increases are around 14%. The Federal Reserve data provides stronger evidence that the $S-shape$ factor is a useful predictor compared to the Fama-Bliss data. For example, for bonds with maturities $n = 3$, according to the Fama-Bliss data, the out-of-sample R^2 increases from 18% to 24% because of the $S-shape$ factor, which is a 33% increase; according to the Federal Reserve data, the out-of-sample R^2 increases from 13% to 26%, which is a 100% increase.

3.5 Utility Analysis

This section applies the expected return analysis in [5]. They develope the calculation because the out-of-sample R^2 is very small in their paper: lower than 1%. Because of this, they use a utility function to check whether the prediction is economically meaningful. It turns out that a statistically insignificant number can still be economically significant. Predictors in their paper proved to be useful economically. Similarly, we employ the expected return calculation in this section to compare the predictive ability of *model* 2 and *model* 1 and to see how much the additional $S-shape$ factor increases a typical investor's expected return.

To do this, this section considers an investor with a single-period investment horizon and mean-variance preference. The investor's objective function is:

$$E_t(r^W_{t+1}) - \frac{\gamma}{2} var_t(r^W_{t+1}),$$

where r^W_{t+1} is the return on wealth or the expected portfolio return and γ is the coefficient of relative risk aversion. The return on wealth follows

$$r^W_{t+1} = r^f_t + \alpha'_t r^e_{t+1},$$

where r_t^f is the return on risk-free asset, $\alpha_t^{'}$ is the weight invested on risky asset, and r_{t+1}^e is the excess return on long bonds with mean $E_t(r_{t+1}^e)$ and variance Σ_t. Suppose:

$$r_{t+1}^e = \mu + x_t + \varepsilon_{t+1},$$

where μ is the unconditional average excess return, x_t is a predictor variable with mean zero, and constant variance σ_x^2, and ε_{t+1} is a random shock with mean zero and constant variance σ_ε^2. Then the investor should set the weight on risky asset in the optimal portfolio without constraints as:

$$\alpha_t = \frac{1}{\gamma}\Sigma_t^{-1}E_t(r_{t+1}^e),$$

which equals to $(\frac{1}{\gamma})(\frac{\mu}{\sigma_x^2+\sigma_\varepsilon^2})$ if the investor does not observe x_t, and equals to $(\frac{1}{\gamma})(\frac{\mu+x_t}{\sigma_\varepsilon^2})$ if the investor does observe x_t. The Euler equation implies excess return on long bonds is

$$E_t(r_{t+1}^e) = \gamma\Sigma_t\alpha_t = \gamma cov(r_{t+1}^e, r_{t+1}^W),$$

which equals to $(\frac{1}{\gamma})(\frac{\mu^2}{\sigma_x^2+\sigma_\varepsilon^2}) = \frac{S^2}{\gamma}$ if the investor does not observe x_t, and equals to $(\frac{1}{\gamma})(\frac{\mu^2+\sigma_x^2}{\sigma_\varepsilon^2}) = (\frac{1}{\gamma})(\frac{S^2+R^2}{1-R^2})$ if the investor does observe x_t. Note here the S represents the unconditional Sharpe Ratio of the risky asset and the R^2 represents the R^2 statistics for regressions of excess return on the predictor variable x_t.

The difference between two expected returns represents how much better expected return gets because of observing x_t: portfolio with larger portion of risky assets has higher expected excess return. The absolute difference between two expected returns is:

$$(\frac{1}{\gamma})(\frac{R^2}{1-R^2})(1+S^2).$$

The proportional increase of expected return because of observing x_t is:

$$(\frac{R^2}{1-R^2})(\frac{1+S^2}{S^2}).$$

Comparing $model\,1$ with $model\,2$, the absolute increases on expected return because of the $S-shape$ factor is:

$$(\frac{1}{\gamma})(\frac{R_2^2}{1-R_2^2} - \frac{R_1^2}{1-R_1^2})(1+S^2),$$

in which R_2^2 represents the R^2 from $model\,2$ and R_1^2 represents the R^2 from $model\,1$. The proportional increase of expected return because of the $S-shape$ factor is:

$$(\frac{R_2^2}{1-R_2^2})/(\frac{R_1^2}{1-R_1^2})-1.$$

INSERT TABLE 3.7 NEAR HERE

Table 3.7 shows what the absolute and proportional increases on expected return are for a typical investor if she uses $model\,2$ instead of $model\,1$. The calculation makes use of the out-of-sample R^2 in table 3.6 and set the risk adverse coefficient γ to 1.

According to the Fama-Bliss data, the additional $S-shape$ factor can increase the absolute expected returns around 5%–15% more, which are around 13%–48% proportionally. According to the Federal Reserve data, the additional $S-shape$ factor can increase the absolute expected returns around 9%–34% more, which are around 28%–131% proportionally. Also, the statistics of 1985–2012 indicate that the $S-shape$ factor becomes less helpful during periods later than 2008, probably due to the impact of the financial crisis period starting 2008.

3.6 Conclusion

In this paper, we propose a new factor to predict excess bond returns: the $S-shape$ factor. It is the fourth principal component of the yields matrix. Similar to the *level*, *slope* and *curvature* factors whose names deliver the intuition of their implications, we name the fourth factor $S-shape$ according to its shape of loadings on yields and the $S-shape$ factor represents how much S-shape like the yield curve is.

To test the predictive power of the $S-shape$ factor for excess bond returns, we fit two nested models to two datasets, the Fama-Bliss data and the Federal Reserve data. The benchmark model includes the first four yield factors (*level*, *slope*, *curvature*, and $S-shape$) while the baseline model only includes the first three factors. The in-sample and out-of-sample statistics present consistent results that the $S-shape$ factor is a significant predictor for excess bond returns.

The S-shape character is an important feature of the yield curve. Historically, market yields on U.S. Treasury securities have displayed the S-shape multiple times. The $S-shape$ factor captured the S-shape of the yield curve. It represents yields change rate under median maturities are different from yields change rates under short and long maturities.

In this paper we also demonstrate that the return-forecasting factor in [7] is a linear combination of all yield factors in the data. The return-forecasting factor has high correlations with the second and fourth yield factors, the *slope* and $S-shape$ factors. The advantages of using yield factors to predict economic variables instead of using the return-forecasting factor cover three aspects. First, it is straightforward to check which yield factor captures the most predictive information for excess bond returns, namely the second and the fourth one. We can focus on these yield factors' economic implications instead of others' and do research beyond the predicting exercises. Second, the estimation of loadings of yield factors is much less sensitive to the data used compared to the estimation of coefficients of the return-forecasting factor and is much more stable over time. Third, the estimation of yield factors faces less econometric issues (such as collinearity) than what the return-forecasting factor would face.

BIBLIOGRAPHY

[1] Andrew Ang, Monika Piazzesi, and Min Wei. What does the yield curve tell us about gdp growth? *Journal of Econometrics*, 131(1):359–403, 2006.

[2] Luca Benati and Charles Goodhart. Investigating time-variation in the marginal predictive power of the yield spread. *Journal of Economic Dynamics and Control*, 32(4):1236–1272, 2008.

[3] Michael D Bordo and Joseph G Haubrich. The yield curve as a predictor of growth: Long-run evidence, 1875-1997. *The Review of Economics and Statistics*, 90(1):182–185, 2008.

[4] John Y Campbell and Robert J Shiller. Yield spreads and interest rate movements: A bird's eye view. *The Review of Economic Studies*, 58(3):495–514, 1991.

[5] John Y Campbell and Samuel B Thompson. Predicting excess stock returns out of sample: Can anything beat the historical average? *Review of Financial Studies*, 21(4):1509–1531, 2008.

[6] Yuqin Chen and Kwok Ping Tsang. What does the yield curve tell us about exchange rate predictability? *Review of Economics and Statistics*, 95(1):185–205, 2013.

[7] John H Cochrane and Monika Piazzesi. Bond risk premia. *The American economic review*, 95(1):138–160, 2005.

[8] John H Cochrane and Monika Piazzesi. Decomposing the yield curve. *Graduate School of Business, University of Chicago, Working Paper*, 2008.

[9] Francis X Diebold and Glenn D Rudebusch. *Yield Curve Modeling and Forecasting?* Princeton University Press, 2012.

[10] Francis X Diebold, Glenn D Rudebusch, and Boragan S Aruoba. The macroeconomy and the yield curve: a dynamic latent factor approach. *Journal of econometrics*, 131(1):309–338, 2006.

[11] Arturo Estrella and Gikas A Hardouvelis. The term structure as a predictor of real economic activity. *The Journal of Finance*, 46(2):555–576, 1991.

[12] Arturo Estrella and Frederic S Mishkin. Predicting us recessions: financial variables as leading indicators. *Review of Economics and Statistics*, 80(1):45–61, 1998.

[13] Eugene F Fama. Term-structure forecasts of interest rates, inflation and real returns. *Journal of Monetary Economics*, 25(1):59–76, 1990.

[14] Eugene F Fama and Robert R Bliss. The information in long-maturity forward rates. *The American Economic Review*, pages 680–692, 1987.

[15] Lars Peter Hansen. Large sample properties of generalized method of moments estimators. *Econometrica: Journal of the Econometric Society*, pages 1029–1054, 1982.

[16] Scott Joslin, Marcel Priebsch, and Kenneth J. Singleton. Risk premiums in dynamic term structure models with unspanned macro risks. *Journal of Finance*, page forthcoming, 2013.

[17] JT Kent, JM Bibby, and KV Mardia. Multivariate analysis. *London, Aca*, 1979.

[18] André Kurmann and Christopher Otrok. News shocks and the slope of the term structure of interest rates. *Federal Reserve Bank of St. Louis Working Paper No*, 2012.

[19] Anh Le and Kenneth J Singleton. The structure of risks in equilibrium affine models of bond yields. Technical report, Working Paper, UNC, 2013.

[20] Robert B Litterman and Jose Scheinkman. Common factors affecting bond returns. *The Journal of Fixed Income*, 1(1):54–61, 1991.

[21] Frederic S Mishkin. What does the term structure tell us about future inflation? *Journal of monetary economics*, 25(1):77–95, 1990.

[22] Charles R Nelson and Andrew F Siegel. Parsimonious modeling of yield curves. *Journal of business*, pages 473–489, 1987.

[23] Whitney K Newey and Kenneth D West. Automatic lag selection in covariance matrix estimation. *The Review of Economic Studies*, 61(4):631–653, 1994.

[24] Monika Piazzesi. Affine term structure models. *Handbook of financial econometrics*, 1:691–766, 2010.

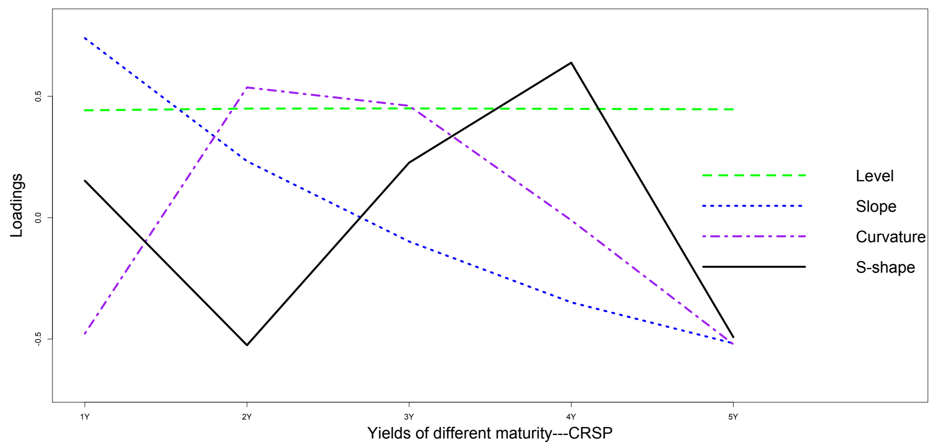
[25] James H Stock and Mark W Watson. Forecasting output and inflation: The role of asset prices. *Journal of Economic Literature*, 41:788–829, 2003.

[26] James H Stock and Mark W Watson. *Introduction to econometrics*, volume 104. Addison Wesley Boston, 2003.

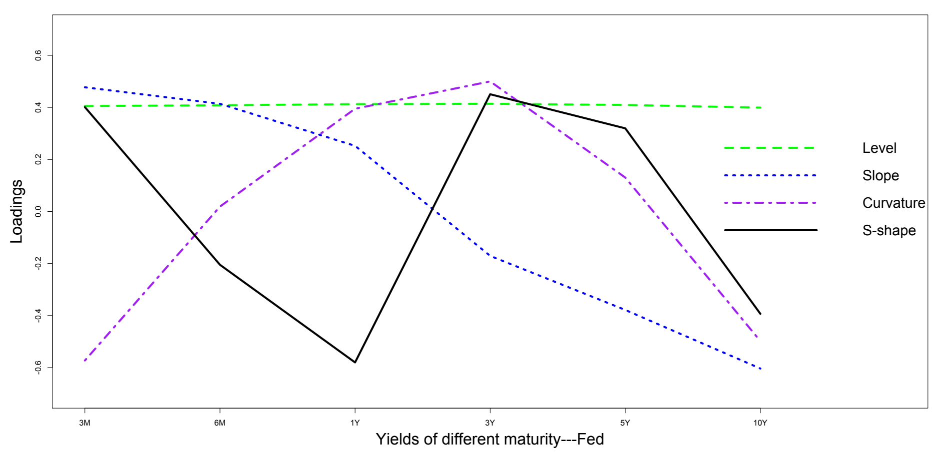
[27] Dick Van Dijk, Siem Jan Koopman, Michel Van der Wel, and Jonathan Wright. Forecasting interest rates with shifting endpoints. 2012.

Fig. 3.1. The first four yield factors' loadings.

This graph uses the Fama-Bliss monthly data consisting of 1 through 5 year zero coupon bond prices and the Federal Reserve data containing market yields on U.S. treasury securies at 3-month, 6-month, 1-year, 3-year, 5-year and 10-year. Since yield factors can be denoted as linear conbinations of yields, for each factor, we connect yield factors' loadings on yields to draw a line. The first three factors are well-known as "*level*", "*slope*" and "*curvature*", and we name the fourth factor "$S-shape$" according to its shape of loadings. Data range is 1964–2007.



(a) Loadings using the Fama-Bliss data.

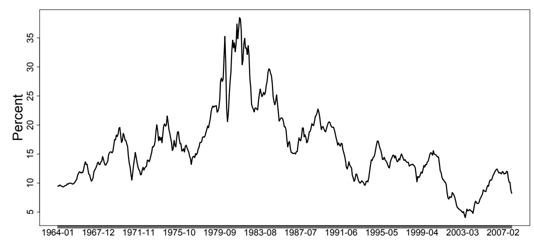


(b) Loadings using the Federal Reserve data.

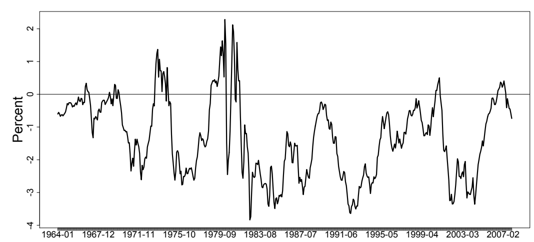
Figure 3.1: Loadings of yield factors.

Fig. 3.2. Time series plots of yield factors.

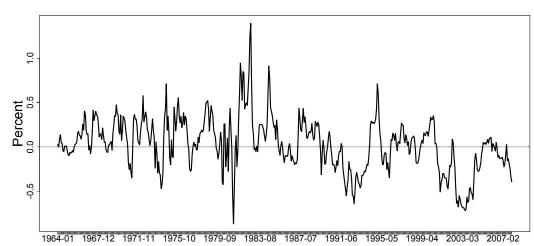
This figure plots the time series of the first four yield factors using monthly U.S. Treasury data from the Federal Reserve. Data range is 1964–2007.



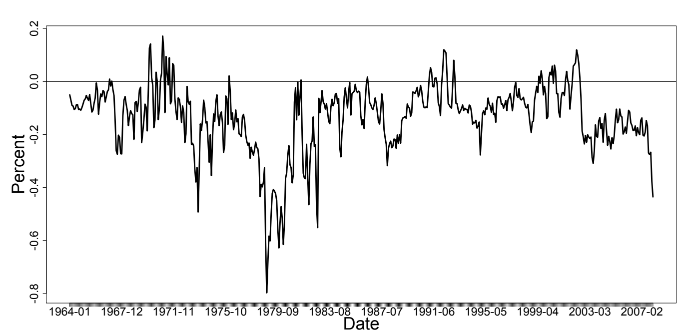
(a) Level.



(b) Slope.



(c) Curvature.

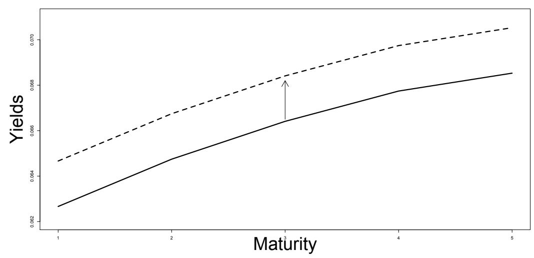


(d) S-shape.

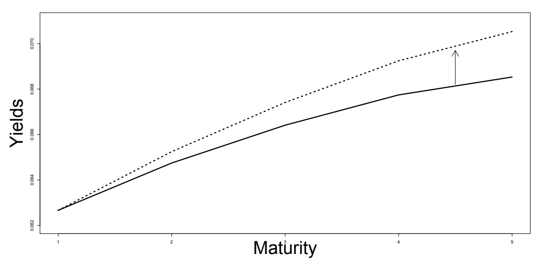
Figure 3.2: Time series plots of yield factors.

Fig. 3.3. The first four yield factors' effect.

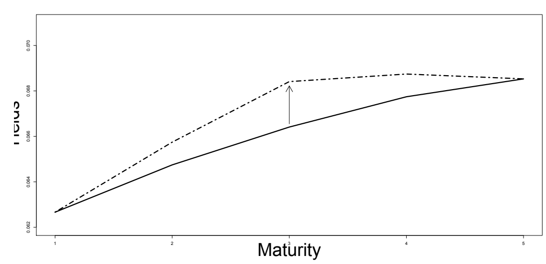
$PC1$ represents the first yield factor, $PC2$ represents the second yield factor, etc. The solid lines represent yield curves at a certain time. The dashed lines represent yield factors' effects on the shape of the yield curve.



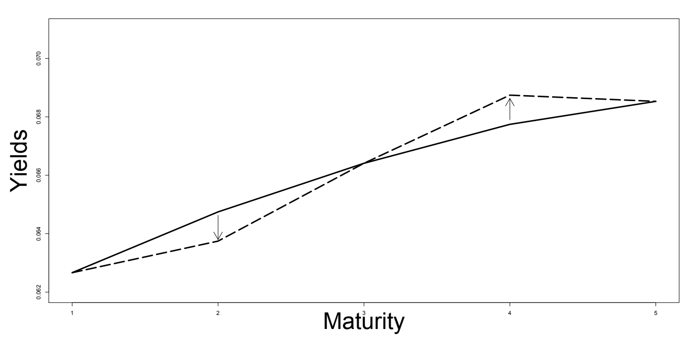
(a) PC1: "Level" effect.



(b) PC2: "Slope" effect.



(c) PC3: "Curvature" effect.

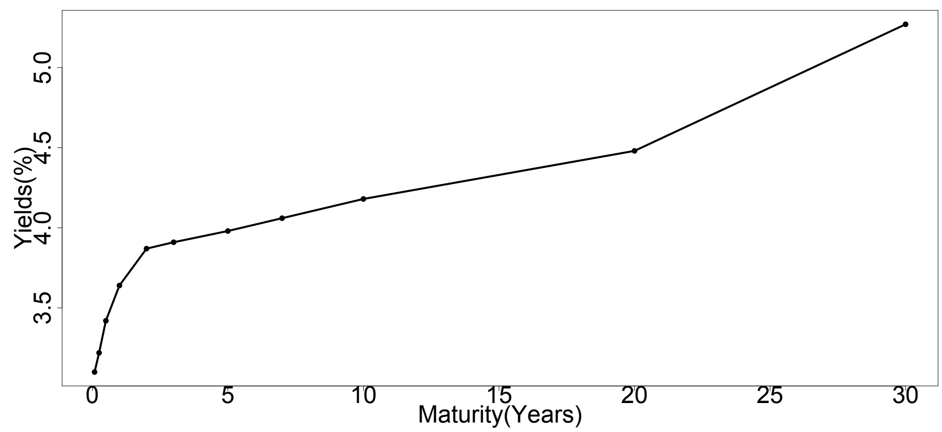


(d) PC4: "S-factor" effect.

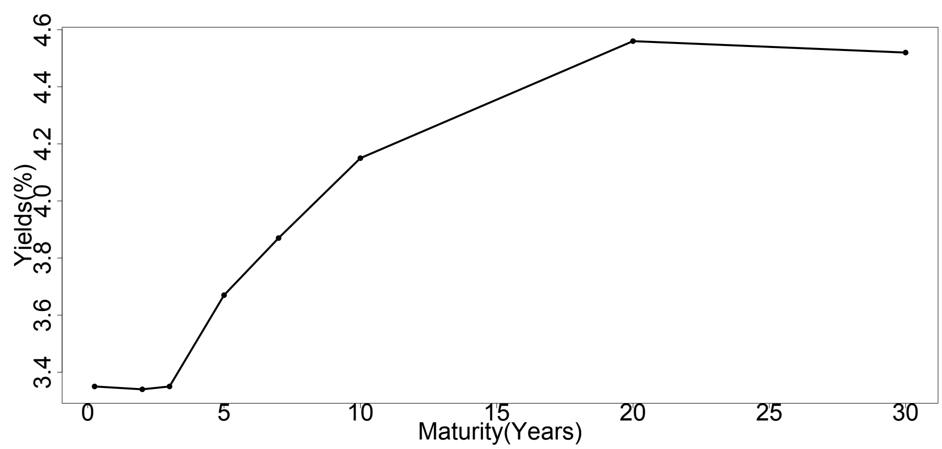
Figure 3.3: Function of yield factors.

Fig. 3.4. Historical yield curve.

This figure plots historical yield curves at two different times, July 2005 and Nov. 2007. The purpose of this figure is to show that yield curve did appear S-shape in history. Data is from the Federal Reserve.



(a) July 2005.



(b) Nov. 2007.

Figure 3.4: Historical yield curves.

Fig. 3.5. Average excess bond returns and lagged $S-shape$ factor using the Fama-Bliss data. This figure plots the average excess bond returns $\overline{rx}$ of four bonds with different maturities, $n = 5, 4, 3, 2$, and the lagged fourth yield factor derived from the Fama-Bliss data. The data consists of 1 through 5 year zero coupon government bond prices. The sample period is set to 1964-2007. The correlation between these two time series is 0.28.

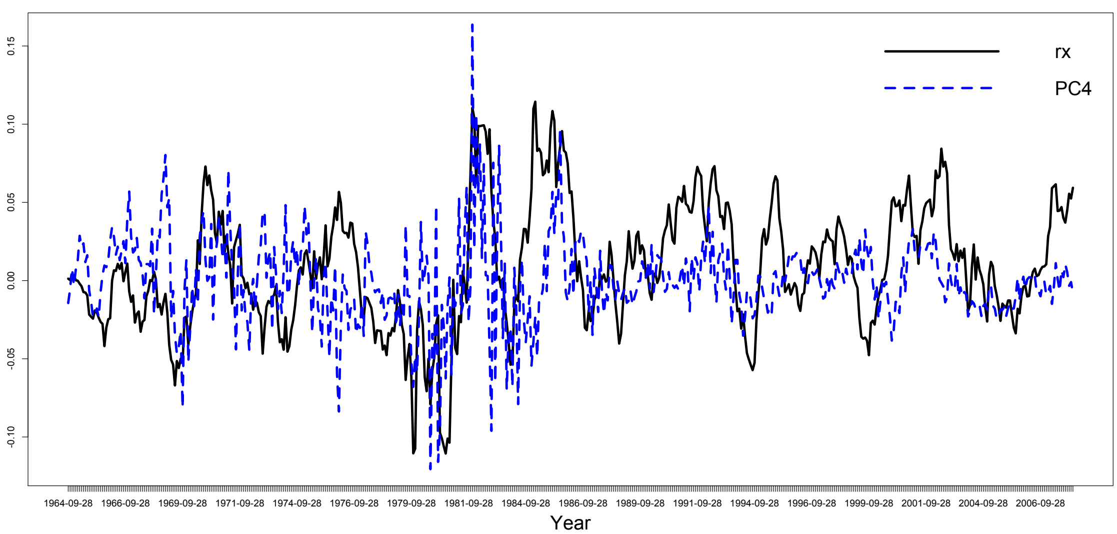


Figure 3.5: Excess bond returns and lagged $S - shape$ factor, Fama-Bliss.

Fig. 3.6. Average excess bond returns and lagged $S - shape$ factor using the Federal Reserve data.

This figure plots the average excess bond returns $\overline{rx}$ of bonds with four different maturities, $n = 5, 4, 3, 2$, and the lagged fourth yield factor derived from the Federal Reserve data. The data consists of monthly observations of market yields on U.S. treasury securities at 3-month, 6-month, 1-year, 3-year, 5-year and 10-year. The sample period is set to 1964-2007. The correlation between these two time series is 0.34.

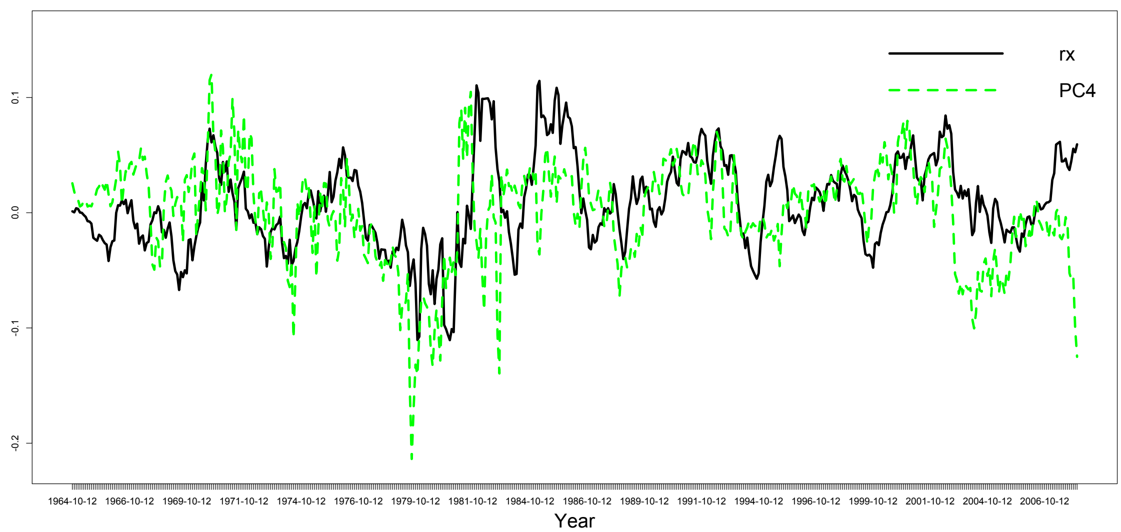


Figure 3.6: Excess bond returns and lagged $S - shape$ factor, Fed.

Table 3.1

Regression results of the return-forecasting factor on yield factors derived from the Fama-Bliss data.

Yield factors are principal components derived from 1 through 5 year zero coupon bond prices from CRSP. The sample period is set to 1964-2007. This table presents statistics of univariate and multivariate regressions with the return-forecasting factor (CP) in [7] as the dependent variable and principal components (PCs) of yields as regressors. The regression equation for the last line of this table is: $CP_t = \beta_0 + \beta_1 PC1_t + \beta_2 PC2_t + \beta_3 PC3_t + \beta_4 PC4_t + \varepsilon_t$. We use OLS to estimate the coefficients and method in [23] to fix the standard error estimation problem due to heteroskedasticity and autocorrelation among errors. $\overline{R}^2$ reports adjusted R^2. Standard errors are in parentheses. The significant codes are "***" for the level of 0.01, "**" for 0.05, and "*" for 0.10.

Table 3.1: Regressions of the return-forecasting factor on principal components

$PC1$	$PC2$	$PC3$	$PC4$	$PC5$	$\overline{R}^2$
0.003**					9%
(NW)(0.00)					
	−0.06***				57%
	(0.01)				
		0.13***			6%
		(0.03)			
			0.39***		27%
			(0.07)		
				−0.04	0%
				(0.08)	
0.003***	−0.06***	0.13***			72%
(0.00)	(0.00)	(0.02)			
0.003***	−0.06***	0.13***	0.39***		100%
(0.00)	(0.00)	(0.00)	(0.00)		

Table 3.2

Regression results of the return-forecasting factor on yield factors derived from the Federal Reserve data.

Yield factors are principal components derived from monthly data of U.S. Treasury securities at 3-month, 6-month, 1-year, 3-year, 5-year and 10-year. The sample period is set to 1964-2007. This table presents statistics of univariate and multivariate regressions with the return-forecasting factor (CP) in [7] as the dependent variable and principal components (PCs) of yields as regressors. The regression equation for the last line of this table is: $CP_t = \beta_0 + \beta_1 PC1_t + \beta_2 PC2_t + \beta_3 PC3_t + \beta_4 PC4_t + \beta_5 PC5_t + \beta_6 PC6_t + \varepsilon_t$. We use OLS to estimate the coefficients and the method in [23] to fix the standard error estimation problem due to heteroskedasticity and autocorrelation among errors. $\overline{R}^2$ reports adjusted R^2. Standard errors are in parentheses. The significant codes are "***" for the level of 0.01, "**" for 0.05, and "*" for 0.10.

Table 3.2: Regressions of the return-forecasting factor on principal components

$PC1$	$PC2$	$PC3$	$PC4$	$PC5$	$PC6$	$\overline{R}^2$
0.002*						5%
(NW)(0.00)						
	−0.03***					54%
	(0.00)					
		0.01				0%
		(0.02)				
			0.13***			8%
			(0.04)			
				0.02		0%
				(0.09)		
					−0.04	0%
					(0.09)	
0.002***	−0.03***	0.01				60%
(0.00)	(0.00)	(0.01)				
0.002***	−0.03***	0.01	0.13***			68%
(0.00)	(0.00)	(0.01)	(0.01)			
0.002***	−0.03***	0.01	0.13***	0.02	−0.04	68%
(0.00)	(0.00)	(0.01)	(0.01)	(0.03)	(0.04)	

Table 3.3

In-sample regression results of predictive models.

The regression equation of *model* 1 is: $rx_{t+1}^{(n)} = \delta_{1,0} + \delta_{1,1}PC1_t + \delta_{1,2}PC2_t + \delta_{1,3}PC3_t + \varepsilon_{t+1}^{(n)}$, in which $PC1$ stands for the first yield factor (principal component), etc. The regression equations of *model* 2 is *model* 1 plus an additional right-hand-side variable: $PC4$. *Model* 3 is a regression with a single predictor: CP—the return-forecasting factor in [7]. This table records the coefficients, standard errors and significant codes of the last right-hand-side variable in each model during different sample periods. 1964-2002 is the sample period used in [7]. The first two columns use yield factors derived from the Fama-Bliss data. The third and fourth columns use yield factors derived from the Federal Reserve data. For the first four columns, we use OLS to estimate the coefficients and the method in [23] to fix the standard error estimation problem due to heteroskedasticity and autocorrelation among errors. For the last column, we use two step OLS to estimate the coefficients and the method in [15] to fix the standard error estimation problem due to generated regressors. The significant codes are "***" for the level of 0.01, "**" for 0.05, and "*" for 0.10.

Table 3.3: Forecasts of 1-year excess returns using 3 models: coefficients

Panel A:

	PCs (F-B)		PCs (Fed)		CP (F-B)	
$n = 2$	*Model* 1	*Model* 2	*Model* 1	*Model* 2	*Model* 3	
Periods	$PC3$	$PC4$	$PC3$	$PC4$	CP	
1964-2007	0.07*	0.16***	0.01	0.15**	0.45***	
(NW)	(0.03)	(0.05)	(0.02)	(0.07)	(0.13)	(2 OLS)
t-stat	1.91	3.16	0.37	2.27	3.45	t-stat
1964-2002	0.08***	0.15***	0.005	0.19***	0.45***	
	(0.03)	(0.04)	(0.02)	(0.05)	(0.16)	
	2.81	3.44	0.27	4.90	2.83	
1964-1984	0.14***	0.15***	−0.05	0.23***	0.49	
	(0.04)	(0.05)	(0.03)	(0.05)	(0.80)	
	3.70	2.72	−1.64	4.91	0.62	
1985-2007	−0.02	0.16**	0.02	0.11	0.45***	
	(0.05)	(0.08)	(0.02)	(0.20)	(0.07)	
	−0.43	2.00	1.02	0.55	> 5.00	
1985-2012	−0.04	0.13*	0.02	0.07	0.43***	
	(0.05)	(0.08)	(0.02)	(0.05)	(0.07)	
	−0.77	1.75	0.95	1.29	> 5.00	

Panel B:

	PCs (F-B)		PCs (Fed)		CP (F-B)	
$n = 3$	$Model\,1$	$Model\,2$	$Model\,1$	$Model\,2$	$Model\,3$	
Periods	$PC3$	$PC4$	$PC3$	$PC4$	CP	
1964-2007	0.12**	0.36***	0.03	0.28*	0.86***	
(NW)	(0.06)	(0.09)	(0.03)	(0.16)	(0.06)	(2 OLS)
t-stat	1.98	4.12	0.82	1.77	> 5.00	t-stat
1964-2002	0.14***	0.36***	0.02	0.34***	0.85***	
	(0.05)	(0.08)	(0.04)	(0.09)	(0.06)	
	2.71	4.39	0.65	3.86	> 5.00	
1964-1984	0.25***	0.37***	-0.11*	0.42***	0.89	
	(0.07)	(0.10)	(0.06)	(0.08)	(0.77)	
	3.41	3.74	-1.84	4.96	1.26	
1985-2007	-0.03	0.33**	0.03	0.20	0.84***	
	(0.10)	(0.15)	(0.04)	(0.46)	(0.04)	
	-0.29	2.27	0.69	0.43	> 5.00	
1985-2012	-0.07	0.25*	0.03	0.11	0.80***	
	(0.09)	(0.14)	(0.05)	(0.10)	(0.04)	
	-0.73	1.79	0.71	1.04	> 5.00	

Panel C:

	PCs (F-B)		PCs (Fed)		CP (F-B)	
$n = 4$	$Model\,1$	$Model\,2$	$Model\,1$	$Model\,2$	$Model\,3$	
Periods	$PC3$	$PC4$	$PC3$	$PC4$	CP	
1964-2007	0.16*	0.52***	0.04	0.35**	1.24***	
(NW)	(0.08)	(0.12)	(0.05)	(0.18)	(0.03)	(2 OLS)
t-stat	1.91	4.46	0.85	1.99	> 5.00	t-stat
1964-2002	0.18**	0.52***	0.03	0.44***	1.24***	
	(0.07)	(0.11)	(0.05)	(0.10)	(0.04)	
	2.44	4.78	0.67	4.22	> 5.00	
1964-1984	0.32***	0.52***	-0.15**	0.53***	1.21***	
	(0.10)	(0.13)	(0.07)	(0.10)	(0.38)	
	3.17	4.06	-2.08	> 5.00	3.23	
1985-2007	-0.03	0.42**	0.03	0.25	1.26***	
	(0.16)	(0.19)	(0.06)	(0.71)	(0.03)	
	-0.21	2.22	0.56	0.35	> 5.00	
1985-2012	-0.11	0.23*	0.05	0.12	1.28***	
	(0.12)	(0.19)	(0.06)	(0.14)	(0.02)	
	-0.89	1.75	0.78	0.87	> 5.00	

Panel D:

	PCs (F-B)		PCs (Fed)		CP (F-B)	
$n = 5$	$Model\,1$	$Model\,2$	$Model\,1$	$Model\,2$	$Model\,3$	
Periods	$PC3$	$PC4$	$PC3$	$PC4$	CP	
1964-2007	0.18*	0.54***	0.06	0.40	1.45***	
(NW)	(0.10)	(0.15)	(0.05)	(0.27)	(0.03)	(2 OLS)
t-stat	1.76	3.66	1.02	1.50	> 5.00	t-stat
1964-2002	0.21**	0.54***	0.05	0.52***	1.46***	
	(0.09)	(0.13)	(0.06)	(0.13)	(0.03)	
	2.48	4.03	0.88	4.08	> 5.00	
1964-1984	0.34***	0.55***	-0.18*	0.63***	1.40***	
	(0.12)	(0.16)	(0.10)	(0.12)	(0.09)	
	2.92	3.44	-1.93	> 5.00	> 5.00	
1985-2007	-0.03	0.39*	0.02	0.26	1.46***	
	(0.25)	(0.24)	0.07	0.62	(0.02)	
	-0.12	1.64	0.27	0.42	> 5.00	
1985-2012	0.16	0.23	0.05	0.09	1.49***	
	(0.15)	(0.24)	(0.08)	(0.17)	(0.02)	
	-1.06	0.97	0.65	0.50	> 5.00	

Table 3.4

In-sample adjusted R^2 of predictive models.

The regression equation of $model\,1$ is: $rx_{t+1}^{(n)} = \delta_{1,0} + \delta_{1,1}PC1_t + \delta_{1,2}PC2_t + \delta_{1,3}PC3_t + \varepsilon_{t+1}^{(n)}$, in which $PC1$ stands for the first yield factor (principal component), etc. The regression equations of $model\,2$ is $model\,1$ plus an additional right-hand-side variable, $PC4$. $Model\,3$ is a regression with a single predictor, CP: the return-forecasting factor in [7]. This table reports the in-sample adjusted R^2 of each model during different sample periods. 1964-2002 is the sample period used in [7]. The first two columns use yield factors derived from the Fama-Bliss data. The third and fourth columns use yield factors derived from the Federal Reserve data.

Table 3.4: Forecasts of 1-year excess returns: in-sample $\overline{R}^2$

	PCs (F-B)		PCs (Fed)		CP (F-B)
Periods/ In-sample $\overline{R}^2$	$Model\,1$	$Model\,2$	$Model\,1$	$Model\,2$	$Model\,3$
$n = 2$					
1964-2007	21%	26%	17%	30%	25%
1964-2002	26%	31%	21%	39%	31%
1964-1984	32%	36%	30%	52%	36%
1985-2007	16%	21%	17%	29%	22%
1985-2012	12%	16%	12%	17%	15%
$n = 3$					
1964-2007	18%	27%	15%	28%	27%
1964-2002	24%	33%	20%	38%	34%
1964-1984	32%	36%	27%	51%	36%
1985-2007	14%	20%	13%	23%	21%
1985-2012	11%	14%	10%	13%	14%
$n = 4$					
1964-2007	20%	29%	17%	28%	30%
1964-2002	26%	37%	23%	39%	37%
1964-1984	32%	36%	28%	49%	36%
1985-2007	17%	22%	15%	23%	24%
1985-2012	13%	16%	12%	13%	18%
$n = 5$					
1964-2007	20%	27%	19%	28%	27%
1964-2002	27%	34%	25%	39%	34%
1964-1984	27%	34%	29%	49%	34%
1985-2007	17%	19%	15%	20%	21%
1985-2012	13%	14%	12%	12%	16%

Table 3.5

Out-of-sample Root Mean Squared Error (RMSE) of predictive regressions.

The calculation method is to regress the true excess bond returns on fitted excess bond returns. The fitted excess return at time t is estimated by multiplying coefficients estimated through time $t-1$ with predictors at time t. The out-of-sample Root Mean Squared Error (RMSE) is estimated by taking the square root of the mean of the squared errors. The regression equation of $model\,1$ is: $rx_{t+1}^{(n)} = \delta_{1,0} + \delta_{1,1}PC1_t + \delta_{1,2}PC2_t + \delta_{1,3}PC3_t + \varepsilon_{t+1}^{(n)}$, in which $PC1$ stands for the first yield factor (principal component), etc. The regression equations of $model\,2$ is $model\,1$ plus an additional right-hand-side variable: $PC4$. This table reports the RMSE of each model during period 1964-2007 and 1964-2012 using recursive method starting from 20 years. The forecast samples are 1985-2007 and 1985-2012 and data is from the Fama-Bliss and the Federal Reserve.

Table 3.5: Out-of-sample RMSE using recursive regressions

RMSE	PCs (F-B)			PCs (Fed)		
Forecast sample	$Model\,1$	$Model\,2$	$Model\,2/Model\,1$	$Model\,1$	$Model\,2$	$Model\,2/Model\,1$
n=2						
1985-2007	0.014	0.013	0.97	0.014	0.013	0.91
1985-2012	0.013	0.013	0.98	0.013	0.013	0.94
n=3						
1985-2007	0.027	0.026	0.96	0.027	0.025	0.92
1985-2012	0.025	0.024	0.97	0.026	0.024	0.95
n=4						
1985-2007	0.037	0.036	0.96	0.038	0.036	0.93
1985-2012	0.035	0.034	0.97	0.036	0.034	0.96
n=5						
1985-2007	0.045	0.044	0.97	0.046	0.044	0.95
1985-2012	0.043	0.043	0.99	0.044	0.043	0.97

Table 3.6

Out-of-sample R^2 of predictive regressions.

The calculation method is to regress the true excess bond returns on fitted excess bond returns and extract the R^2 of fitting. The fitted excess return at time t is estimated by multiplying coefficients estimated through time $t-1$ with predictors at time t. The regression equation of *model* 1 is: $rx_{t+1}^{(n)} = \delta_{1,0} + \delta_{1,1}PC1_t + \delta_{1,2}PC2_t + \delta_{1,3}PC3_t + \varepsilon_{t+1}^{(n)}$, in which $PC1$ stands for the first yield factor (principal component), etc. The regression equations of *model* 2 is *model* 1 plus an additional right-hand-side variable: $PC4$. This table reports the out-of-sample R^2 of each model during period 1964-2007 and 1964-2012 using recursive method starting from 20 years. The forecast samples are 1985-2007 and 1985-2012 and data is from the Fama-Bliss and the Federal Reserve.

Table 3.6: Out-of-sample R^2 using recursive regressions

R^2	PCs (F-B)			PCs (Fed)		
Forecast sample	*Model* 1	*Model* 2	Increase	*Model* 1	*Model* 2	Increase
n=2						
1985-2007	22%	26%	18%	17%	31%	82%
1985-2012	21%	24%	14%	16%	27%	69%
n=3						
1985-2007	18%	24%	33%	13%	26%	100%
1985-2012	17%	22%	29%	13%	22%	69%
n=4						
1985-2007	20%	26%	30%	15%	26%	73%
1985-2012	19%	24%	26%	16%	22%	38%
n=5						
1985-2007	21%	25%	19%	17%	26%	53%
1985-2012	20%	22%	10%	18%	22%	22%

Table 3.7

The absolute/proportional increases on expected return because of the additional $S - shape$ **factor.**

This table calculates the absolute/proportional increases on expected returns for an investor with a single-period horizon and mean-variance preference because of the additional $S - shape$ factor. The calculation makes use of the method in [5], which makes use of the out-of-sample R^2. The risk adverse coefficient γ is set to 1. The regression equation of $model\,1$ is: $rx_{t+1}^{(n)} = \delta_{1,0} + \delta_{1,1}PC1_t + \delta_{1,2}PC2_t + \delta_{1,3}PC3_t + \varepsilon_{t+1}^{(n)}$, in which $PC1$ stands for the first yield factor (principal component), etc. The regression equations of $model\,2$ is $model\,1$ plus an additional right-hand-side variable: $PC4$. Data is from the Fama-Bliss and the Federal Reserve.

Table 3.7: Absolute/Proportional increases on expected return: comparing $model\,2$ with $model\,1$

	Absolute increases: E(r)		Proportional increases: E(r)	
Forecast sample	PCs (F-B)	PCs (Fed)	PCs (F-B)	PCs (Fed)
n=2				
1985-2007	0.11	0.34	30%	122%
1985-2012	0.08	0.24	22%	85%
n=3				
1985-2007	0.14	0.27	48%	131%
1985-2012	0.10	0.18	33%	83%
n=4				
1985-2007	0.15	0.24	45%	102%
1985-2012	0.10	0.14	30%	54%
n=5				
1985-2007	0.09	0.19	28%	70%
1985-2012	0.05	0.09	13%	28%