NBER WORKING PAPER SERIES

BOND RISK PREMIA

John H. Cochrane
Monika Piazzesi

Working Paper 9178
http://www.nber.org/papers/w9178

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2002

We thank Geert Bekaert, Michael Brandt, Lars Hansen, Bob Hodrick, Narayana Kocherlakota, Pedro Santa-Clara, Martin Schneider and many seminar participants for helpful comments. Cochrane acknowledges research support from the Graduate School of Business and from an NSF grant administered by the NBER. Updated drafts of this paper are maintained on the authors' web pages, including color graphs. The views expressed herein are those of the authors and not necessarily those of the National Bureau of Economic Research.

Bond Risk Premia
John H. Cochrane and Monika Piazzesi
NBER Working Paper No. 9178
September 2002
JEL No. G1, E4

ABSTRACT

This paper studies time variation in expected excess bond returns. We run regressions of annual excess returns on forward rates. We find that a single factor predicts 1-year excess returns on 1-5 year maturity bonds with an R^2 up to 43%. The single factor is a tent-shaped linear function of forward rates. The return forecasting factor has a clear business cycle correlation: Expected returns are high in bad times, and low in good times, and the return-forecasting factor forecasts long-run output growth. The return-forecasting factor also forecasts stock returns, suggesting a common time-varying premium for real interest rate risk. The return forecasting factor is poorly related to level, slope, and curvature movements in bond yields. Therefore, it represents a source of yield curve movement not captured by most term structure models. Though the return-forecasting factor accounts for more than 99% of the time-variation in expected excess bond returns, we find additional, very small factors that forecast equally small differences between long term bond returns, and hence statistically reject a one-factor model for expected returns.

John H. Cochrane
Graduate School of Business
University of Chicago
1101 E. 58th St.
Chicago, IL 60637
and NBER
john.cochrane@gsb.uchicago.edu

Monika Piazzesi
Anderson Graduate School of Management
UCLA
110 Westwood Plaza
Los Angeles, CA 90095
and NBER
piazzesi@ucla.edu

1 Introduction

This paper studies time-varying risk premia in the term structure of interest rates. We run regressions of one year excess returns – borrow at the one year rate, buy a long term bond, and sell it in one year – on all forward rates available at the beginning of the period. We find R^2 values as high as 43%. The forecasts are statistically significant for all maturities, even taking into account the small sample properties of test statistics. The forecasts survive in subsamples, in real time, and across two datasets. Most importantly, the pattern of regression coefficients is the same for all maturities. A *single* "return-forecasting factor," a single linear combination of forward rates (or yields), describes time-variation in the expected return of *all* bonds. A rise in the return-forecasting factor implies steadily greater expected excess returns for longer maturity bonds.

This work extends Fama and Bliss's (1987) and Campbell and Shiller's (1991) classic regressions. Fama and Bliss found that the spread between the n-year forward rate and the 1-year yield predicts the 1-year excess return of the n-year bond, with R^2 about 15%. Campbell and Shiller found similar results forecasting yield changes with yield spreads. (Campbell 1995 is an excellent summary.) We more than double the R^2. We find that the *same* linear combination of forward rates predicts bond returns at all maturities, where Fama and Bliss relate each bond's expected excess return to a different forward spread. Our return-forecasting factor completely drives out Fama and Bliss's forward spread in a multiple regression. Our regressions improve on Fama and Bliss's statistical evidence for forecastable returns, especially in small samples.

We explore extensively the macroeconomic and financial interpretation and implications of our bond return forecasts, both for their own interest and to make the forecasts more concrete and believable. Fama and Bliss (1989), Harvey (1989) and many others document that the term spreads which forecast bond returns are correlated with economic conditions and forecast output and stock returns. It is natural to see how the return-forecasting factor behaves in these roles. We find that the return-forecasting factor has a clear business cycle correlation. Expected excess returns are high in bad times, and low in good times. However, business cycle variables do not forecast bond returns. The return-forecasting factor apparently reflects additional or better-measured information. The return-forecasting factor forecasts GDP growth, but only at horizons over a year, in contrast to the yield spread which forecasts output over short horizons. Thus, we overturn the link between bond excess return forecasts and short term (monetary?) output fluctuations suggested by term spreads. The return-forecasting factor also forecasts stock returns, about as much as it would forecast the return of an 8-year duration bond. The return-forecasting factor is a bit stronger in the 1990s, when interest rate changes were largely real, than it is in the 1970s when inflation had a strong effect on interest rates, and the return-forecasting factor is poorly correlated with inflation. All of this evidence points to a time-varying, business cycle related premium for holding real interest rate risk.

The return-forecasting factor also forecasts changes in short-term interest rates. The classic expectations hypothesis predicts that a forward rate higher than the short rate should forecast a rise in the short rate. Fama and Bliss's (1987) rejection of the expectations hypothesis showed that this is not true – there is no tendency for the short rate to rise in the first year after we observe a high forward rate. This phenomenon is the flip side of Fama and Bliss's return forecasts: since the interest rate does not rise, long term bond prices do not fall to offset their higher initial yields, so long-term bond holders make money. Our return-forecasting factor *does* forecast changes in 1-year rates. However, our forecast is roughly speaking in the "wrong direction." When the return-forecasting factor signals high returns on long term bonds, it forecasts a *decline* in the short rate that will raise long term bond prices, so long term bond holders make even more money than by Fama and Bliss's mechanism.

We also relate our findings to the term structure literature in finance, especially the "multifactor affine model" literature in which all bond yields are driven by a few factors. The return-forecasting factor is a tent-shaped linear combination of forward rates. It is not a "curvature factor" in yields, though. Expressed as a function of yields rather than forward rates, it is curved through the 4-5 year maturity rather than curved at the short end. As a result, the return-forecasting factor is poorly related to the level, slope, and curvature factors that describe the vast bulk of moments in bond yields and that form the basis of most term structure models. Much of the variation in the return-forecasting factor is related to the small additional yield curve factors that are conventionally ignored. In addition, most term structure models are specified and estimated at monthly or even weekly frequency. We find a moving average structure in the monthly yield data, possibly induced by measurement error, so a monthly AR(1) yield representation raised to the 12th power misses much of the one-year bond return predictability and completely misses the single-factor representation. To see the return forecasts, you must look directly at the one year horizon, or more complex time series models.

These two facts may explain why the return-forecasting factor has gone unrecognized for so long in this well-studied data, and these facts carry important implications for term structure modeling. If you first posit a factor model for yields, estimate it on monthly data, and then look at one year expected returns, you will completely miss excess return forecastability. To incorporate our evidence on risk premia, a yield curve model must include something like our return-forecasting factor in addition to traditional factors such as "level," "slope," and "curvature," even though the return-forecasting factor does little improve the model's fit for yields, and it must reconcile the difference between our direct annual forecasts and those implied by short horizon regressions.

The return-forecasting factor accounts for more than 99% of the time-variation of bond expected excess returns. However, additional very small factors forecast very small spreads between long-term bond returns with statistical significance. These additional factors cause a statistical rejection of a single factor model for expected returns. The additional factors correspond to small, idiosyncratic, transitory movements in bond prices or yields. These movements may represent small liquidity premia, exploitable by extreme long-short positions (such as the famous 29.5 - 30 year LTCM bond spread), or they may reflect measurement error in prices.

The expectations hypothesis remains a workhorse in applied work in macroeconomics and finance. For example, central banks care a lot about interest rates – how to forecast interest rates, how monetary policy affects long term rates through its dynamic effect on short term rates, or how to read market interest rate and inflation expectations from the yield curve. Yet central bank researchers routinely impose the expectations hypothesis in these exercises (Clews 2002, Scholtes 2002, Söderlind and Svensson, 1997). The European Central Bank "Monthly Bulletin" explicitly uses the expectations hypothesis to compute expected future short rates from futures (see, for example, page 13 of the August 2002 issue). The monetary VAR literature, when it does not simply ignore the information in the term structure, routinely imposes the expectations hypothesis to distinguish expected from unexpected movements in interest rates, for example Rudebusch (1998), Krueger and Kuttner (1996). Our strong evidence against the expectations hypothesis casts doubt on these and related applications.

Our single factor model is similar to the "single index" or "latent variable" models used by Hansen and Hodrick (1983) and Ferson and Gibbons (1985) to capture time-varying expected returns. Stambaugh (1988) ran regressions similar to ours of 2-6 month bond excess returns on 1-6 month forward rates. After correcting for measurement error, Stambaugh found a pattern of coefficients similar to ours (Figure 2, page 53). Stambaugh rejected a one or two factor representation of this forecast. Ilmanen (1995) ran regressions of monthly excess returns on bonds in different countries on a term spread, the real short rate, stock returns, and bond return betas. Ilmanen did not reject a one factor representation for expected international excess returns consisting of a linear combination of these variables.

2 Bond return regressions

2.1 Notation

We use the following notation for log bond prices:

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

We use parentheses to distinguish maturity from exponentiation in the superscript. The log yield is

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

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

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

and we write 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 $t+1$ as

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

We denote excess log returns by

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

We use the same letters without n index to denote vectors across maturity, e.g.

$$rx_{t+1} = \left[\begin{array}{cccc} rx_t^{(2)} & rx_t^{(3)} & rx_t^{(4)} & rx_t^{(5)} \end{array} \right]^\top.$$

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

$$\begin{aligned} y_t &= \left[\begin{array}{cccccc} 1 & y_t^{(1)} & y_t^{(2)} & y_t^{(3)} & y_t^{(4)} & y_t^{(5)} \end{array} \right]^\top, \\ f_t &= \left[\begin{array}{cccccc} 1 & y_t^{(1)} & f_t^{(1\to2)} & f_t^{(2\to3)} & f_t^{(3\to4)} & f_t^{(4\to5)} \end{array} \right]^\top. \end{aligned}$$

We use overbars to denote averages across maturity, e.g.

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

2.2 Excess return forecasts

Our objective is to understand time-variation in expected excess bond returns. Hence, we run regressions of bond excess returns at time $t+1$ on variables at time t. The natural right hand variables are time t bond prices, yields or forward rates. Prices, yields and forward rates are linear functions of each other, so the forecasts are the same. We find that forward rates produce more elegant and interpretable results. Section 3 considers the addition of macroeconomic variables to forecast bond returns, and finds that they do not help. We focus on an annual horizon. We use the Fama-Bliss data (available from CRSP) of 1 through 5 year zero coupon bond prices, so we can compute annual returns directly.

Table 1 presents regressions of excess returns on all forward rates. The top panel of Figure 1 graphs the regression coefficients as a function of the maturity on the right hand side – each row of Table 1 is a line of the graph. (For now, ignore the bottom panel.) The plot makes the pattern clear – *the same function of forward rates forecasts holding period returns at all maturities. Longer maturities just have greater loadings on this same function.* The pattern of coefficients suggests a common return-forecasting factor that is a tent-shaped linear combination of forward rates.

Table 1. Regressions of 1-year excess returns on all forward rates

n		const.	$y^{(1)}$	$f^{(1\to 2)}$	$f^{(2\to 3)}$	$f^{(3\to 4)}$	$f^{(4\to 5)}$	R^2	$\bar{R}^2$	Level R^2	$\chi^2(5)$
2		−1.96	−0.94	0.74	1.15	0.24	-0.91	0.34	0.33	0.38	143.0
	Large T	(0.64)	(0.18)	(0.43)	(0.30)	(0.27)	(0.18)				⟨0.00⟩
	Small T	(0.81)	(0.30)	(0.50)	(0.40)	(0.30)	(0.27)	[0.20, 0.55]			
	EH							[0.00, 0.17]			⟨0.00⟩
3		−3.28	−1.66	0.74	2.96	0.29	−1.90	0.34	0.34	0.37	107.7
	Large T	(1.21)	(0.32)	(0.72)	(0.49)	(0.50)	(0.32)				⟨0.00⟩
	Small T	(1.45)	(0.53)	(0.88)	(0.70)	(0.54)	(0.49)	[0.22, 0.56]			
	EH							[0.00, 0.17]			⟨0.00⟩
4		−4.57	−2.40	1.11	3.46	1.18	−2.78	0.37	0.36	0.39	104.0
	Large T	(1.68)	(0.46)	(0.94)	(0.62)	(0.67)	(0.42)				⟨0.00⟩
	Small T	(1.92)	(0.71)	(1.18)	(0.93)	(0.72)	(0.67)	[0.24, 0.58]			
	EH							[0.00, 0.17]			⟨0.00⟩
5		−5.78	−2.98	1.48	3.93	1.14	−2.88	0.34	0.33	0.36	77.8
	Large T	(2.13)	(0.58)	(1.12)	(0.73)	(0.80)	(0.53)				⟨0.00⟩
	Small T	(2.38)	(0.89)	(1.46)	(1.14)	(0.89)	(0.85)	[0.21, 0.56]			
	EH							[0.00, 0.17]			⟨0.00⟩

NOTE: The regression equation is

$$rx_{t+1}^{(n)} = \beta_0 + \beta_1 y_t^{(1)} + \beta_2 f_t^{(1\to 2)} + ... + \beta_5 f_t^{(4\to 5)} + \varepsilon_{t+1}^{(n)}.$$

$\bar{R}^2$ reports adjusted R^2. "Level R^2" reports the R^2 from a regression using the level, not log, excess return on the left hand side, $e^{r_{t+1}^{(n)}} - e^{y_t^{(1)}}$. Standard errors are in parentheses. "Large T"standard errors use the Hansen-Hodrick GMM correction for overlap. "Small T" standard errors are based on 50,000 bootstrapped samples from an unconstrained 12 lag yield VAR. Square brackets "[]" are 95% bootstrap confidence intervals for R^2. "EH" imposes the expectations hypothesis on the bootstrap: We run a 12 lag autoregression for the 1-year rate and calculate other yields as expected values of the 1-year rate. Details are in the Appendix. "χ^2" is the Wald statistic that tests whether the slope coefficient is zero. All χ^2 statistics are computed with 18 Newey-West lags to ensure a positive definite covariance matrix. Pointed brackets "<>" report asymptotic and bootstrapped p-values. Data source CRSP, sample 1964:1-2001:12.

The expectations hypothesis predicts coefficients of zero – nothing should forecast bond excess returns. The regressions in Table 1 provide strong evidence against the expectations hypothesis. The p-values for the χ^2 test that all coefficients are zero are below 1%, for all maturities. We also compute small sample distributions[1] for our test

[1] Bekaert, Hodrick and Marshall (1997) and others have questioned the small sample properties of expectations hypothesis tests.

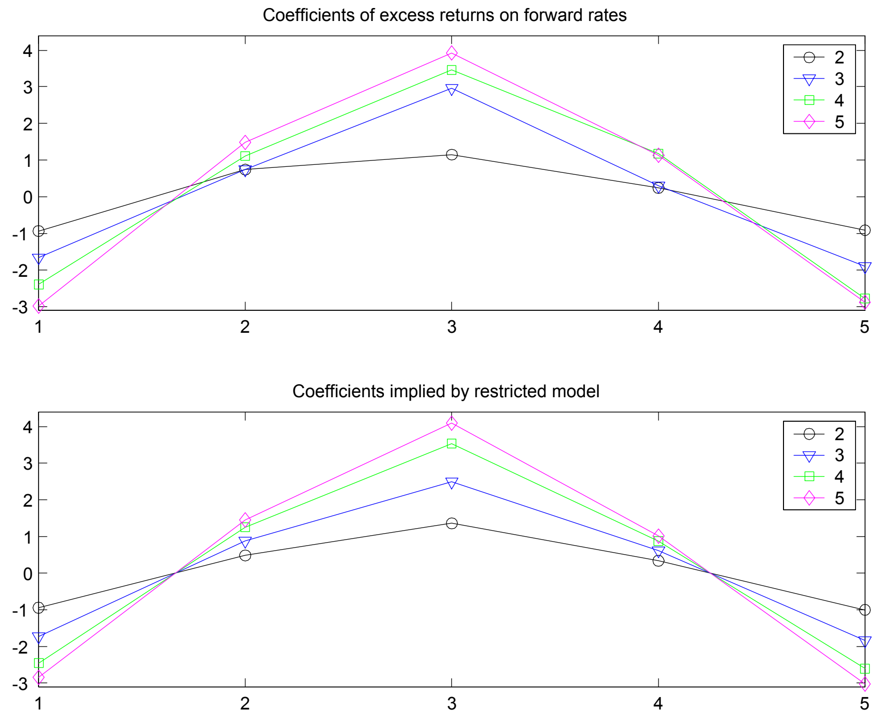

Figure 1: Coefficients in a regression of holding period excess returns on the one-year yield and 4 forward rates. The top panel presents unrestricted estimates from Table 1. The bottom panel presents restricted estimates from a single-factor model reported in Table 2. The legend (2, 3, 4, 5) refers to the maturity of the bond whose excess return is forecast. The x axis gives the maturity of the forward rate on the right hand side.

statistics. To construct standard errors, we run an unrestricted 12 monthly lag vector autoregression of all 5 yields, and bootstrap the residuals. To test the expectations hypothesis ("EH" in the table), we run an unrestricted 12 monthly lag autoregression of the one year yield, bootstrap the residuals, and calculate other yields according to the expectations hypothesis. (See the Appendix for details.) Table 1 shows that the small-T standard errors are indeed slightly larger than their asymptotic counterparts. However, inferences are not strongly affected. The small T p-values for the χ^2 test still overwhelmingly reject the expectations hypothesis.

The regressions give an impressive R^2 (for excess return forecasts) of 0.34-0.37, suggesting that the forecastability is economically as well as statistically important. The R^2 is consistent across maturities. One might worry that the R^2 comes from the large number of right hand variables. For reassurance, we report in Table 1 the conventional adjusted $\bar{R}^2$, and it is nearly identical. However, that adjustment presumes i.i.d. data which is not valid in this case. The point of adjusted $\bar{R}^2$ is to test whether the forecastability is spurious, and the χ^2 test that the coefficients are jointly zero addresses that issue properly. To assess sampling error and overfitting bias in R^2 directly, we also compute small-sample 95% confidence intervals for the unadjusted R^2. Our 0.34-0.37 R^2

lie well above the 0.17 upper end of the 95% R^2 confidence interval calculated under the expectations hypothesis.

One might worry about logs versus levels; that actual excess returns are not forecastable, so the coefficients in Table 1 only reflect $1/2\sigma^2$ terms and conditional heteroskedasticity.[2] We repeated the regressions using actual excess returns, $e^{r_{t+1}^{(n)}} - e^{y_t^{(1)}}$ on the left hand side. The coefficients are nearly identical. The penultimate column of Table 1 reports the R^2 from these regressions, labeled "Level R^2," and they are in fact slightly *higher* than the R^2 for the regression in logs.

2.3 A single factor for expected bond returns

The beautiful pattern of coefficients in Figure 1 cries for us to describe expected excess returns of bonds on all maturities in terms of a single factor, as follows:

$$rx_{t+1}^{(n)} = b_n \left(\gamma_0 + \gamma_1 y_t^{(1)} + \gamma_2 f_t^{(1\to 2)} + ... + \gamma_5 f_t^{(4\to 5)} \right) + \varepsilon_{t+1}^{(n)}. \tag{1}$$

b_n and γ_n are not separately identified by this specification, since you can double all the bs and halve all the γs. We normalize the coefficients by imposing that the average value of b_n is one,

$$\frac{1}{4}\sum_{n=2}^{5} b_n = 1.$$

With this normalization, we estimate (1) in two steps. First, we estimate the γ by running the regression

$$\frac{1}{4}\sum_{n=2}^{5} rx_{t+1}^{(n)} = \gamma_0 + \gamma_1 y_t^{(1)} + \gamma_2 f_t^{(1\to 2)} + ... + \gamma_5 f_t^{(4\to 5)} + \bar{\varepsilon}_{t+1}. \tag{2}$$
$$\overline{rx}_{t+1} = \gamma^\top f_t + \bar{\varepsilon}_{t+1}.$$

The second equality reminds us of the notation γ, f for corresponding 6×1 vectors and the notation $\overline{rx}$ for the average (across maturities) excess return. Then, we estimate b_n by running the four regressions

$$rx_{t+1}^{(n)} = b_n \left(\gamma^\top f_t \right) + \varepsilon_{t+1}^{(n)},\ n = 2,\ 3,\ 4,\ 5.$$

(We explain in Section 6 why we use this two-step procedure rather than efficient GMM.)

Table 2 presents the estimated values of γ and b and standard errors. The γ estimates are just about what one would expect from inspection of Figure 1. The loadings b_n of expected returns on the common return-forecasting factor $\gamma^\top f$ increase smoothly with maturity. The R^2 in Table 2 are essentially the same as in Table 1. This fact indicates that

[2] We thank Ron Gallant for raising this important question.

the cross-equation restrictions implied by the model (1) – that bonds of each maturity are forecast by the *same* portfolio of forward rates – do little damage to the forecast ability.

Table 2. Estimates of the return-forecasting factor.

	γ_0	γ_1	γ_2	γ_3	γ_4	γ_5	R^2	$\chi^2(5)$
	-3.90	-2.00	1.02	2.87	0.71	-2.12	0.35	97.2
Large T	(1.41)	(0.38)	(0.79)	(0.53)	(0.56)	(0.36)		$\langle 0.00 \rangle$
Small T	(1.62)	(0.60)	(1.00)	(0.79)	(0.61)	(0.57)	[0.22, 0.56]	
EH							[0.00, 0.17]	$\langle 0.00 \rangle$

		Standard Errors			
n	b_n	Large T	Small T	R^2	Small T
2	0.47	(0.05)	(0.02)	0.32	[0.19, 0.54]
3	0.87	(0.03)	(0.02)	0.34	[0.21, 0.55]
4	1.23	(0.02)	(0.02)	0.37	[0.24, 0.58]
5	1.43	(0.04)	(0.03)	0.34	[0.21, 0.55]

NOTE: The top panel regression is

$$\overline{rx}_{t+1} = \gamma_0 + \gamma_1 y_t^{(1)} + \gamma_2 f_t^{(1\to 2)} + ... + \gamma_5 f_t^{(4\to 5)} + \bar{\varepsilon}_{t+1}$$

where $\overline{rx}_{t+1}$ denotes the average (across maturities) excess log return. The lower panel regression is

$$rx_{t+1}^{(n)} = b_n \left(\gamma^\top f_t\right) + \varepsilon_{t+1}^{(n)}.$$

γ is the parameter estimate from the upper panel, and f denotes the vector of all forward rates. "Large T" standard errors in the lower panel correct for the fact that γ is estimated, by considering this estimate together with the regression in the top panel as a single GMM estimation. "χ^2" tests whether all slope coefficients are jointly zero. Standard errors are in parentheses, bootstrap 95% confidence intervals in square brackets "[]" and p-values angled brackets "<>". See notes to Table 1 for details.

The single factor model (1) is a restricted model. If we write the 4 unrestricted regressions of excess returns on all forward rates as

$$rx_{t+1} = \beta f_t + \varepsilon_{t+1}, \tag{3}$$

where β is a 4×6 matrix of regression coefficients, the single factor model amounts to the restriction

$$\beta = b\gamma^\top.$$

A *single* linear combination of forward rates $\gamma^\top f_t$ is the state variable for time-varying expected returns of *all* maturities.

The bottom panel of Figure 1 plots the coefficients of expected returns on each of the forward rates implied by the restricted model, i.e. for each n, it presents $\left[\begin{array}{ccc} b_n\gamma_1 & \cdots & b_n\gamma_5 \end{array}\right]$. Comparing this plot with the unrestricted estimates of the top panel, you can see that the single-factor model almost exactly captures the unrestricted parameter estimates. The specification (1) constrains the constants ($b_n\gamma_0$) as well as the regression coefficients plotted in Figure 1. Figure 2 plots the restricted and unrestricted estimates of the constant, and you can see similarly that the estimates are very close.

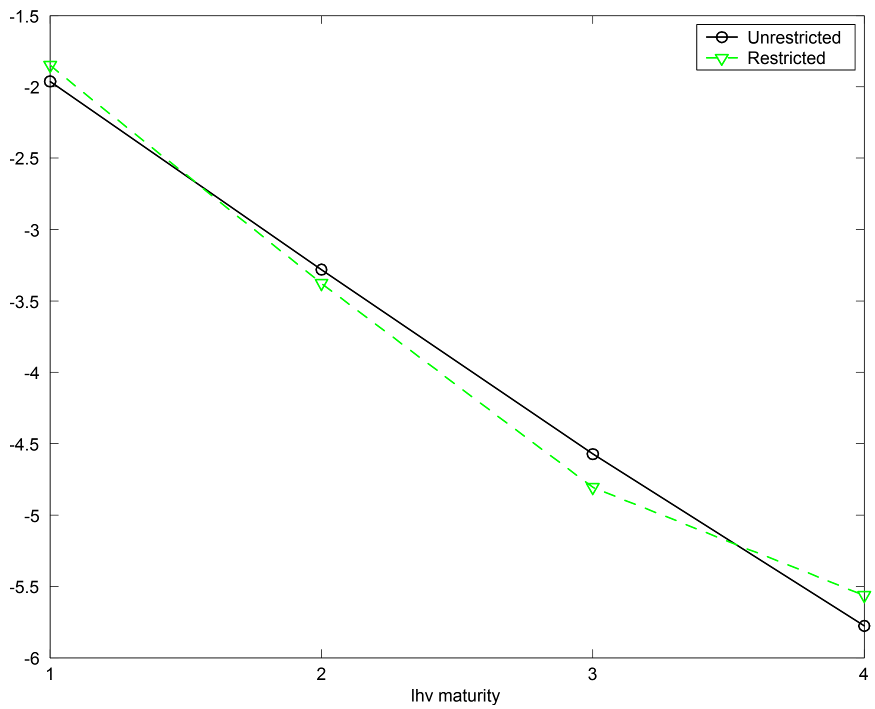

Figure 2: Restricted and unrestricted estimates of the constants. Restricted estimates are from Table 2, and unrestricted estimates are from Table 1. The x axis represents the maturity of the bond whose excess return is forecast.

The individual restricted and unrestricted estimates are close statistically as well as economically. We could not fit standard error bounds into Figures 1 and 2, but we computed t-statistics for the hypothesis that each parameter is individually equal to its restricted value.[3] The largest t-statistic is 0.9 and most of them are around 0.2. (Section 6 considers whether the restricted and unrestricted coefficients are *jointly* equal.)

[3]The test statistic is $vec(b\gamma^\top) - vec(\beta)$ divided by the GMM standard error of unrestricted parameter estimates $diag(cov(vec(\beta))^{1/2}$.

2.4 Fama-Bliss regressions

Fama and Bliss (1987) regressed each excess return against the same maturity forward spread. This is the classic regression that provided first evidence against the expectations hypothesis in this data set. Table 3 updates Fama and Bliss's regressions to include more recent data.

Table 3. Fama-Bliss excess return regressions

Maturity n		α	β	R^2	$\chi^2(1)$
2		0.04	0.94	0.14	14.2
	Large T	(0.30)	(0.28)		⟨0.00⟩
	Small T	(0.15)	(0.34)	[0.01, 0.35]	
	EH			[0.00, 0.13]	⟨0.02⟩
3		−0.14	1.24	0.14	13.5
	Large T	(0.54)	(0.38)		⟨0.00⟩
	Small T	(0.31)	(0.44)	[0.01, 0.36]	
	EH			[0.00, 0.14]	⟨0.02⟩
4		−0.41	1.50	0.15	11.6
	Large T	(0.76)	(0.50)		⟨0.00⟩
	Small T	(0.45)	(0.51)	[0.01, 0.38]	
	EH			[0.00, 0.14]	⟨0.03⟩
5		−0.11	1.10	0.06	3.7
	Large T	(1.05)	(0.62)		⟨0.11⟩
	Small T	(0.59)	(0.70)	[0.00, 0.28]	
	EH			[0.00, 0.14]	⟨0.20⟩

NOTE: The regressions are

$$rx_{t+1}^{(n)} = \alpha + \beta\left(f_t^{(n-1\rightarrow n)} - y_t^{(1)}\right) + \varepsilon_{t+1}^{(n)}.$$

Standard errors are in parentheses, bootstrap 95% confidence intervals in square brackets "[]" and p-values angled brackets "<>". See notes to Table 1 for details.

The expectations hypothesis predicts a coefficient of zero. Instead, Table 3 shows that the forward spread moves essentially one-for-one with expected excess returns on long term bonds. Fama and Bliss's regressions have held up well since publication, unlike many other anomalies.

The multiple regressions in Table 1 and the single factor model in Table 2 provide stronger evidence against the expectations hypothesis than do the updated Fama-Bliss regressions in Table 3 in many respects. For 3 of the 4 bonds, the Fama-Bliss regressions forecast excess returns with statistical significance, whether using asymptotic or bootstrap confidence intervals and p-values.[4] Tables 1 and 2 show stronger χ^2 rejections, and do so for all maturities. They more than double Fama and Bliss's R^2 from below 0.15 in Table 3 to 0.34-0.37 in Tables 1 and 2. The 5-year rate R^2 is particularly dramatic, jumping from 0.06 in Table 3 to 0.34 in Table 1. In the Fama-Bliss regressions, 3 of the 4 bonds show R^2 just above the 95% confidence interval, mirroring the narrow, but still significant statistical rejections. The R^2 in Tables 1and 2 are nearly double the upper end of their expectations hypothesis confidence intervals. The 0.20-0.24 *lower* end of the R^2 confidence intervals in Table 1 lie above the 0.15 R^2 of the Fama- Bliss regression.

If the return-forecasting factor really is an improvement, it should drive out other variables, and the Fama-Bliss spread in particular. The individual coefficient standard errors in Table 1 already suggest that the additional right hand variables are important. Table 4 presents multiple regressions contrasting the single factor model of Table 2 with the Fama-Bliss regressions.

Table 4. Contest between $\gamma^\top f$ and Fama-Bliss

n	a_n	$\sigma(a_n)$	b_n	$\sigma(b_n)$	c_n	$\sigma(c_n)$	R^2
2	0.13	(0.25)	0.47	(0.03)	−0.05	(0.19)	0.33
3	0.13	(0.52)	0.88	(0.10)	−0.07	(0.37)	0.34
4	−0.03	(0.67)	1.22	(0.15)	0.05	(0.46)	0.37
5	−0.31	(0.75)	1.42	(0.17)	0.15	(0.35)	0.34

NOTE: Multiple regression of excess holding period returns on the return-forecasting factor and Fama-Bliss slope. The regression is

$$rx_{t+1}^{(n)} = a_n + b_n \left(\gamma^\top f_t\right) + c_n \left(f_t^{(n-1\to n)} - y_t^{(1)}\right) + \varepsilon_{t+1}^{(n)}.$$

Standard errors in parentheses use the Hansen-Hodrick GMM correction for overlap.

In the presence of the Fama-Bliss forward spread, the coefficients and significance of the return-forecasting factor from Table 2 are unchanged in Table 4. The R^2 is also unaffected, meaning that the addition of the Fama-Bliss forward spread does not help to forecast bond returns. On the other hand, in the presence of the return-forecasting factor, the Fama-Bliss slope is destroyed as a forecasting variable. The coefficients decline from 1 or even more to almost exactly zero, and are insignificant. Clearly, the return-forecasting factor subsumes all the predictability of bond returns captured by the Fama-Bliss forward spread.

[4] We use χ^2 rather than a t test as we report χ^2 tests that multiple parameters are jointly zero in other tables. The χ^2 values are not squared t statistics, because we use Hansen-Hodrick weights to compute individual standard errors, but Newey-West weights with more lags to compute χ^2 statistics. Newey-West weights are necessary to ensure positive definite covariance matrices in tests of multiple parameters.

The most important comparison is economic, not statistical. The single factor model of Table 2 describes expected excess returns on *all* bonds with a *single* state variable, $\gamma^\top f_t$. The Fama-Bliss regressions describe the expected return of *each* bond with a *separate* state variable, its own forward spread. It is a much more useful specification to think of a common element to expected returns, at least across a range of maturities.

This comparison is not a criticism of Fama and Bliss. The Fama-Bliss specification is exactly right for Fama and Bliss's purpose. They wanted to explore the expectations hypothesis and, eventually, to reject it. The individual forward spreads have important interpretations in the expectations hypothesis, so they are natural right hand variables if one is guided by that null. Our purpose is different. We want to characterize expected excess returns, knowing the expectations hypothesis is false. The multiple regression is a more natural specification when we are guided by that null, and it's not surprising in retrospect that it works better for its purpose.

2.5 Short rate forecast

Fama and Bliss also regressed changes in the 1-year rate on forward spreads. Short rate forecasts and excess return forecasts are mechanically linked, but seeing the same phenomenon as a short rate forecast provides a useful complementary intuition and suggests many additional implications.

Here, the expectations hypothesis predicts a coefficient of 1.0 – if the forward rate increases one percentage point over the short rate, we should see the short rate rise one percentage point on average. Corresponding to the expectations hypothesis failure in Table 4, the Fama Bliss regression in Table 5 shows that the 2-year forward rate has no power to forecast a 1-year change in the 1-year rate. The coefficient is essentially zero, and one fourth of a standard error away from zero. Under the expectations hypothesis null the R^2 confidence interval is from 7 to 32% – short rates should be predictable, given there is variation in the forward spread. Yet the R^2 is zero, and the test for a zero coefficient passes at a 98% probability value.

The unconstrained regression in Table 5 again contrasts strongly with the Fama-Bliss regression. All the forward rates taken together have substantial power to predict one-year changes in the short rate. The R^2 for short rate changes jumps to a substantial 23%. The χ^2 test strongly rejects the null that the parameters are jointly zero.

This forecastability of the short rate does not revive the expectations hypothesis. Instead, the short rate becomes forecastable precisely because the expectations hypothesis is even more wrong than Fama and Bliss suspected. To understand this claim, suppose the forward spread rises one percentage point, so long term bond yields are higher than the 1-year rate. According to the expectations hypothesis, the 1-year rate is expected to rise one percentage point. This rise will lower prices of long term bonds, offsetting their now higher yields and generating no change in expected return. In Fama and Bliss's regressions, the 1-year rate moves sluggishly. When the forward spread is high, the offsetting rise in 1-year rates does not materialize for a few years, so long term bond holders make an extra return from the higher yield of long term bonds. In the regressions of Tables 1 and 2, we see much larger variation in the expected excess return. Fama and Bliss's mechanism has been exhausted. To generate larger expected returns, we must be able to forecast short rate movements, roughly speaking in the "wrong" direction. If our regression forecasts a 2% excess return on long term bonds, it must predict a *decline* in short rates, which will raise long term bond prices.

Table 5. Predicting short rate changes

	const.	$f_t^{(1\to 2)} - y_t^{(1)}$	$y_t^{(1)}$	$f_t^{(1\to 2)}$	$f_t^{(2\to 3)}$	$f_t^{(3\to 4)}$	$f_t^{(4\to 5)}$	R^2	χ^2
	Fama-Bliss								
	-0.04	0.06						0.00	0.1
Large T	(0.30)	(0.28)							<0.98>
EH								[0.07,0.32]	<1.00>
	Unconstrained								
	1.96		−0.06	0.26	−1.15	−0.24	0.91	0.23	118.5
Large T	(0.64)		(0.18)	(0.43)	(0.30)	(0.27)	(0.18)		
Small T	(0.82)		(0.30)	(0.50)	(0.40)	(0.30)	(0.27)	[0.16, 0.41]	$\langle 0.00 \rangle$
EH								[0.08, 0.35]	$\langle 0.00 \rangle$

NOTE: The Fama-Bliss regression is

$$y_{t+1}^{(1)} - y_t^{(1)} = \beta_0 + \beta_1 \left(f_t^{(1\to 2)} - y_t^{(1)} \right) + \varepsilon_{t+1}.$$

The unconstrained regression equation is

$$y_{t+1}^{(1)} - y_t^{(1)} = \beta_0 + \beta_1 y_t^{(1)} + \beta_2 f_t^{(1\to 2)} + ... + \beta_5 f_t^{(4\to 5)} + \varepsilon_{t+1}.$$

χ^2 tests whether all slope coefficients are jointly zero (5 degrees of freedom unconstrained, one degree of freedom for Fama-Bliss). Standard errors are in parentheses, bootstrap 95% confidence intervals in square brackets "[]" and p-values angled brackets "<>". See notes to Table 1 for details.

To see the same points more formally, start with the definition,

$$rx_{t+1}^{(2)} = p_{t+1}^{(1)} - p_t^{(2)} - y_t^{(1)} = -y_{t+1}^{(1)} - p_t^{(2)} + p_t^{(1)} = -y_{t+1}^{(1)} + f_t^{(1\to 2)}$$
$$rx_{t+1}^{(2)} = \left(y_{t+1}^{(1)} - y_t^{(1)} \right) + \left(f_t^{(1\to 2)} - y_t^{(1)} \right),$$

and hence, using any set of forecasting variables,

$$E_t \left(rx_{t+1}^{(2)} \right) = E_t \left(y_{t+1}^{(1)} - y_t^{(1)} \right) + \left(f_t^{(1\to 2)} - y_t^{(1)} \right). \tag{4}$$

Under the expectations hypothesis, expected excess returns are constant, so any movement in the forward spread must be matched by movements in the expected 1-year rate change. In Fama and Bliss's regressions, the expected yield change term is constant, so changes in the forward spread must move one for one with changes in the expected excess return. In our regressions, expected returns move more than changes in the forward spread. The only way to generate such changes is if the 1-year rate becomes forecastable as well.

Equation (4) also means that the regression coefficients which forecast the 1-year rate change in Table 5 are exactly equal to our return-forecasting factor $b_2\gamma^\top f_t$ which forecasts $E_t\left(rx_{t+1}^{(2)}\right)$ minus a coefficient of one on the 2-year forward spread. The same factor that forecasts excess returns is also the state variable that forecasts the short rate.

(Fama and Bliss found that the expectations hypothesis works better over longer horizons. Though the 2 year forward rate has little power to forecast the one year change in the one year rate, the 5-year forward rate moves nearly one-for-one with the expected four-year change in the 1-year rate. This means that the 5-year forward spread does not forecast the *four* year return on 5-year bonds, though it does forecast the *one*-year return on 5-year bonds. They relate this pattern to a slow moving AR(1) for the one year rate. The extension of our work to longer horizons is not straightforward, so in the interest of length we do not pursue it in this paper.)

3 Macroeconomic interpretation

What is the intuition and economic significance of the return-forecasting factor? We start by relating the return-forecasting factor to macroeconomic variables and to stock returns. The slope of the term structure is correlated with recessions and forecasts stock returns (Fama and French 1989) and it forecasts output growth (Harvey 1989, Stock and Watson 1989, Estrella and Hardouvelis 1991, Hamilton and Kim 1999). Since we substitute the return-forecasting factor for the term structure slope in forecasting bond excess returns, we naturally want to see how it performs in these other roles.

Figure 3 presents the return-forecasting factor together with the unemployment rate and the NBER peaks and troughs. The return-forecasting factor is closely associated with business cycles. Expected returns are high in bad times and low in good times. The return-forecasting factor is a "level" variable rather than a "growth" variable. It is high when the *level* of unemployment is high, or the level of income is low, rather than being high during recessions defined as periods of poor GDP *growth*. Campbell and Cochrane (1999) model this kind of business cycle related risk premium.

The correlation between the return-forecasting factor and unemployment is also evident at lower frequencies than usual business cycles. The return-forecasting factor increases throughout the 70s and decreases throughout the 80s, mirroring the unemployment rate as it mirrors many measures of a two-decade-long productivity dip.

The return-forecasting factor is correlated with many other recession indicators as well, including industrial production growth, Lettau and Ludvigson's (1999) consumption/wealth ratio, the investment/GDP ratio, and so on. It is much less correlated with inflation. We present the graph for unemployment as it has the highest correlation among the cyclical indicators we examined.

3.1 Macroeconomic forecasts of bond returns

Given the high correlation between the return factor and the unemployment rate, a natural question is whether we can use unemployment or other macro variables to forecast bond excess returns, either alone or in addition to the return-forecasting factor. The answer is no, or at least not among the variables we have tried. The fact that macro variables by themselves do not forecast bond excess returns is an unfortunate result for economic interpretation. (This is equally true for Fama and Bliss's 1987 or Campbell and Shiller's 1991 specifications). It would be much nicer if we could understand bond expected excess returns as a simple mirror of macroeconomic conditions. It appears instead that bond market prices reflect information beyond that available in macroeconomic aggregates, and this information is crucial to forecasting bond excess returns. On the other hand, the fact that macro variables add nothing to the return-forecasting factor

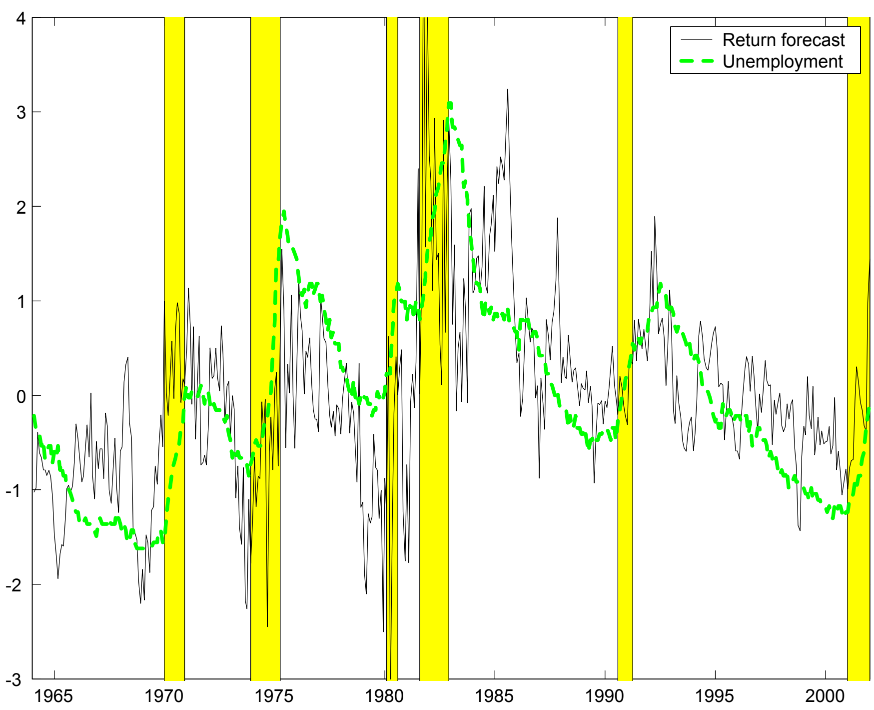

Figure 3: Return forecasting factor $\gamma' f_t$ and unemployment rate. Both series are transformed to $[x_t - E(x)]/\sigma(x)$ so that they fit on the same graph. Shaded areas are NBER recessions.

in multiple regressions is a fortunate result for empirical analysis: it means we can stick to the model $E_t(\overline{rx}_{t+1}) = \gamma^\top f_t$ with great accuracy, even in VAR systems that include macroeconomic variables.

Table 6 contrasts forecasting regressions of the average (across maturities) 1-year bond excess return $\overline{rx}_{t+1}$ on the return-forecasting factor $\gamma^\top f_t$, the unemployment rate U_t, and other macroeconomic variables. The first row reminds us of the 0.35 R^2 and high t statistic when we forecast bond excess returns from $\gamma^\top f_t$. Despite its beautiful correlation with the return-forecasting factor, unemployment U forecasts bond excess returns only 0.05 R^2. In a multiple regression, unemployment does not affect the size and significance of the $\gamma^\top f_t$ coefficient, and only raises the R^2 to 0.38. The Stock-Watson (1989) leading index XLI is designed to forecast output growth at a 6 month horizon. Alas, it forecasts bond excess returns with an even lower R^2 of 0.01 and has no effect in a multiple regression. Lettau and Ludvigson's (2001) consumption-wealth ratio *cay*, which forecasts income growth and stock returns, does no better. Finally, CPI inflation is just as useless as the others.

Table 6. Macro forecasts of bond excess returns

$\gamma^\top f$	(t)	U	(t)	XLI	(t)	*cay*	(t)	*cpi*	(t)	R^2
1	(7.2)									0.35
		0.54	(1.5)							0.05
1.19	(7.6)	-0.50	(-1.6)							0.38
				-0.11	(-0.6)					0.01
1.01	(6.8)			-0.14	(-1.2)					0.36
						0.44	(1.03)			0.02
1.01	(7.8)					−0.08	(−0.2)			0.35
								−0.24	(−0.84)	0.03
0.99	(7.6)							-0.5	(1.6)	0.38

NOTE: Forecasts of average (across maturities) bond returns $\overline{rx}_{t+1}$. U = the unemployment rate. XLI = Stock-Watson leading indicator. *cay* = the Lettau-Ludvigson consumption-wealth ratio using end of period wealth. *cpi* = inflation, the one-year growth in the CPI index. Overlapping annual forecasts, 1964:01-2001:12. Standard errors in parentheses are corrected for overlap and heteroskedasticity by GMM.

It is tempting to conclude from the fact that unemployment does not forecast output that the "level" movement of the return-forecasting factor shown in Figure 3 is irrelevant; that only the high frequency blips matter. This would be a mistake. The return-forecasting factor could easily have chosen not to load on the level of interest rates, producing a much flatter graph. It did not; the regression coefficients γ sum to about one (a level of interest rate) rather than zero (just spreads).

3.2 Term structure forecasts of output growth

The term structure slope forecasts output growth as well as bond returns. Does the return-forecasting factor $\gamma^\top f$ forecast output growth? Figure 4 contrasts regressions of GDP growth using the return forecasting factor $\gamma^\top f$ and the yield spread $y^{(5)} - y^{(1)}$. We look at GDP growth multiple periods ahead. For example, the regression using the return-forecasting factor is

$$\ln (GDP)_{t+x} - \ln(GDP)_{t+x-1/4} = \alpha + \beta \left(\gamma^\top f_t\right) + \varepsilon_{t+x},$$

where x is the number of years ahead, plotted on the x axis of Figure 4. For horizons out to a year, the term structure slope forecasts GDP growth well. However, the return-forecasting factor has absolutely no power to forecast GDP growth for the first year.

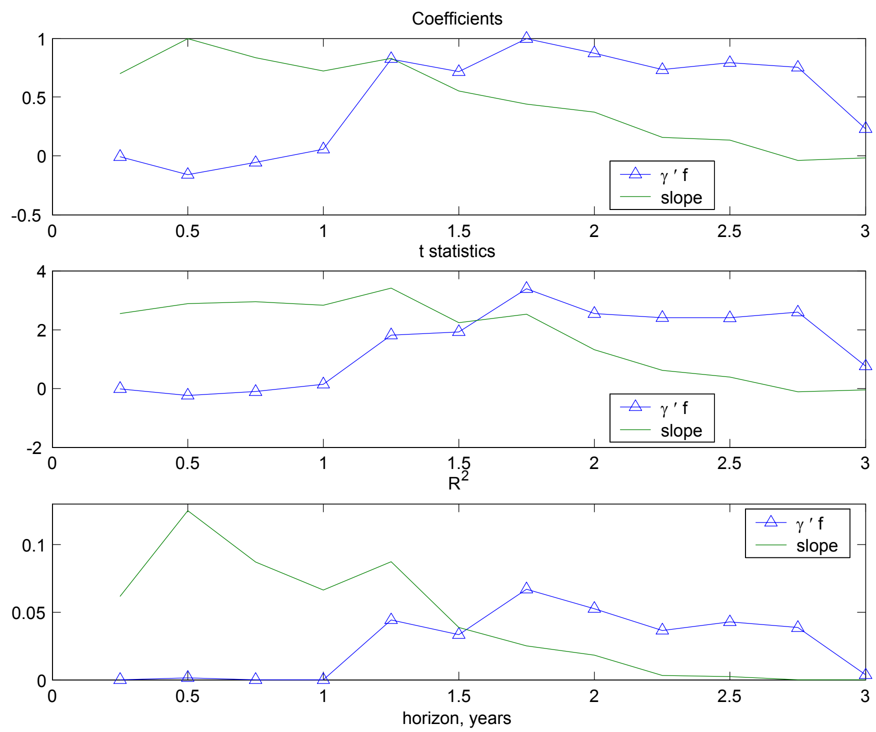

Figure 4: Coefficients, t-statistics and R^2 in long-run GDP growth forecasts. The underlying regressions forecast one-quarter GDP growth x years ahead using the return-forecasting variable $\gamma^\top f$ and the term structure slope $y^{(5)} - y^{(1)}$. Sample is quarterly, 1964-2001.

The fact that the slope forecasts output and bond returns has long pointed to an attractive common element to these forecasts. Our regressions overturn this conclusion. The component of the yield spread that forecasts near-term output growth is uncorrelated with the component that forecasts bond excess returns, and conversely, the component (correlated with the return-forecasting factor) that forecasts bond returns does not forecast near-term output growth. .

For horizons beyond a year, however, the pattern changes. Now the return-forecasting factor forecasts GDP growth, while the slope has less and less power to do so. We conclude again that bond expected excess returns are "level" variable like the unemployment rate, the consumption-income ratio (Cochrane 1994) or the consumption-wealth ratio (Lettau and Ludvigson 2001) that forecasts long-term output growth, rather than a variable like the slope that forecasts near-term growth, possibly due to short-term financial or monetary factors. (Multiple regressions show the same pattern of coefficients and t-statistics. A wide variety of other output forecast variables, including the consumption-wealth ratio, the Stock-Watson leading index, etc., and other definitions of output, including industrial production and the coincident index, lead to similar results.)

3.3 Forecasting stock returns

The slope of the term structure forecasts stock returns, as emphasized by Fama and French (1989). This is important evidence that the forecast corresponds to a real risk premium and not to a bond-market fad or measurement error in bond prices. We can view a stock as a long term bond plus risk; unless time-varying stock market risk premia move exactly opposite to time-varying bond market risk premia, a bond return forecasting factor should forecast stock returns much as it would a long-term bond. Table 7 evaluates how well our return-forecasting factor forecasts stock returns.

Table 7. Forecasts of excess stock returns

Regression	d/p	(t)	$y^{(5)} - y^{(1)}$	(t)	$\gamma^\top f$	(t)	R^2
1 (Full sample)	2.51	(1.18)					0.03
2 (1964 -1989)	7.08	(2.43)					0.14
3			4.16	(1.68)			0.05
4	2.50	(1.25)	4.15	(1.84)			0.08
5					2.10	(3.00)	0.10
6			1.00	(0.38)	1.89	(2.54)	0.10
7	1.09	(0.56)			1.94	(2.86)	0.11
8 (all f)							0.13

NOTE: The left hand variable is the one-year return on the value-weighted NYSE stock return, less the 1-year bond yield. The right hand variables are as indicated in the column headings. Overlapping monthly observations of annual returns, 1964-2001, except regression 2 from 1964-1989. The dividend price ratio is based on the return with and without dividends for the preceding year. T-statistics are in parentheses. Standard errors are corrected for overlap.

The first 4 regressions remind us of stock return forecastability from the dividend price ratio and term spread. Regressions 1 and 2 study the dividend price ratio. Until the 1990s, the dividend price ratio was a strong return forecaster, with a 14% R^2. The long boom of the 1990s cut down this forecastability dramatically, especially in our rather short sample (for these purposes) starting only in 1964. Of course, one good crash will restore the d/p forecastability. The term spread in the third regression forecasts the VW stock return with a 4.2 coefficient – one percentage point term spread corresponds to 4.2 percentage point increase in stock return. The R^2 is only 5% however. The fourth regression shows that the term spread and dividend price ratio forecast different components of returns, since the coefficients are unchanged in multiple regressions and the R^2 increases, though to a still low 8%.

The fifth regression introduces the return-forecasting factor. It is significant, which neither d/p (in this sample) nor the term spread are, and at 10%, its R^2 is slightly higher than that of the term spread and d/p combined. The coefficient is 2.10. The return-forecasting factor is the average expected return across 2-5 year bonds. The 5-year bond in Table 2 had a coefficient of 1.43 on the return-forecasting factor, and the coefficients rose 0.2-0.4 per year of maturity. Thus, the stock return coefficient of 2.10 is what would expect of a bond with about 8-year duration, which is sensible. (8 year duration is also the result of fitting a linear plus square root function of maturity to the coefficients in Table 2.)

The sixth and seventh regressions compare the bond return-forecasting factor with the term spread and d/p. The bond return forecasting factor's coefficient and significance are hardly affected in this multiple regression, while the d/p and term coefficients are cut more than in half and rendered even less significant.

Last, we consider an unrestricted regression of stock excess returns on all forward rates. Of course, this estimate will be noisy, since stock returns are more volatile than bond returns. All forward rates together produce an R^2 of 13%, only slightly more than the $\gamma^\top f$ R^2 of 10%. The stock return forecasting coefficients (not reported) recover a similar tent shape pattern, though not exactly the same as those of the return-forecasting factor.

4 Finance interpretation

We now relate the return-forecasting factor to term structure models in finance. These models decompose yield curve movements into linear combinations of yields, or "factors," that explain the majority of the variance of yield changes.[5] Most such decompositions find "level," "slope," and "curvature" factors that move the yield curve in corresponding shapes. It is tempting to look at our tent-shaped function of forward rates, and to conclude that the return-forecasting factor is a "curvature" factor in the yield curve.

[5] For example, Dai and Singleton 2002 and Duffee 2002 construct affine multifactor models consistent with Fama and Bliss's 1987 regressions. Much effort in these models, which we do not address here, goes in to specifying short rate dynamics and market prices of risk to *derive* a linear factor representation for bond yields. We simply characterize the results, i.e. we examine directly the linear factor representations which affine models would spend a lot of time deriving.

This temptation is misleading, however, because our tent-shaped function is a function of *forward rates*, not yields. Forward rates and yields span the same bond prices of course, so we can also express the forecasting factor as a function of yields, $\gamma^{*\top} y_t = \gamma^\top f_t$. The forecasts are exactly the same, but the weights γ^* are different than the weights γ.

The top of Figure 5 plots the return-forecasting factor as a function of yields. The return-forecasting factor is curved at the *long* end, not the short end of the yield curve. It loads most strongly on the 4-5 year spread, not short spreads. To make an explicit comparison, the bottom of Figure 5 plots the first 3 principal components of yield changes. We calculate these principal components from an eigenvalue decomposition of the covariance matrix of yield changes. (A similar factor analysis of the covariance matrix of yield levels produces very similar results.) We label the principal components "level," "slope," and "curvature" based on the shape of these loadings. As the figure shows, the curvature factor is curved at the short end of the yield curve. The return-forecasting factor is clearly not this "curvature" factor in yields, or any other of the first three principal components.

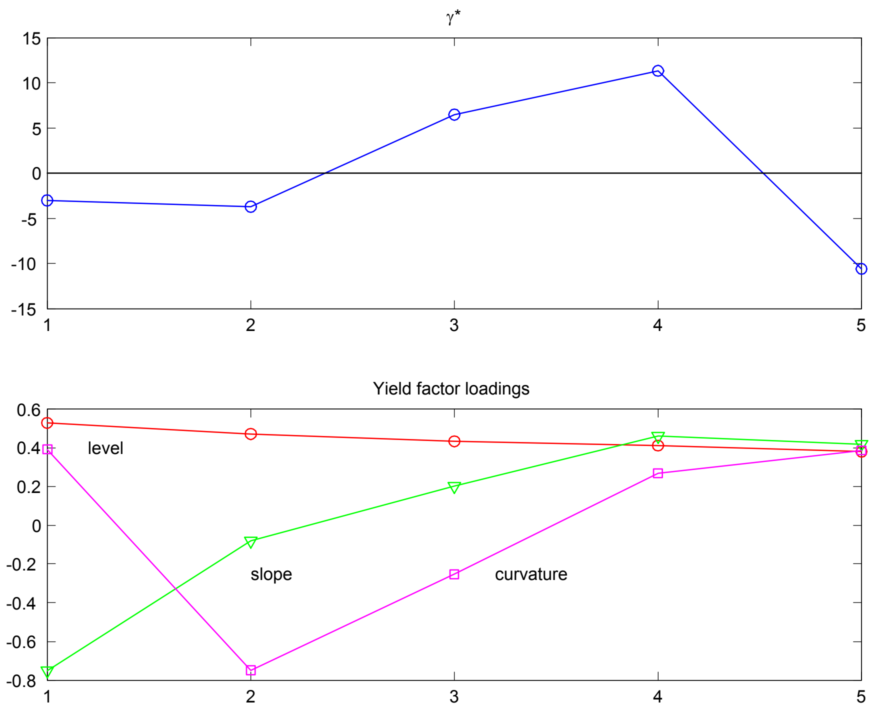

Figure 5: The top graph shows coefficients γ^* in a regression of average (across maturities) holding period returns on all yields, $\overline{rx}_{t+1} = \gamma^{*\top} y_t + \varepsilon_{t+1}$. The bottom graph shows the loadings of the first three principal components of yield changes labeled "level," "slope," and "curvature."

Table 8 collects facts about these principal components and comparisons with the return-forecasting factor. We dub the fourth yield factor "spreads" since it loads on the spread between 3 and 2,5 year yields, and we dub the fifth yield factor "zigzag" for a zigzag shape of its loadings. (We don't plot these loadings in Figure 5 for clarity.) The first row of Table 8 shows that as usual in such decompositions, the first two or three factors explain the vast majority of the variance of yield changes. The last column shows how $\gamma^\top f$ does in the factor model's job – explaining variance of yield changes. $\gamma^\top f$ explains only 0.3% of the variance of yield changes. Not a very useful factor.

Table 8. Yield factors

	Level	Slope	Curvature	Spreads	Zigzag	$\gamma^\top f$
% of $var(\Delta y)$ explained by this factor	93.7	4.4	0.8	0.6	0.6	0.3
% of $var(\gamma^\top f)$ explained by this factor	0.5	37	0.4	42	21	100
R^2 forecasting rx_{t+1} from this factor (%)	2.6	5.2	7.3	8.9	9.1	35.1
R^2 forecasting rx_{t+1} from up to this (%)	2.6	22.6	26.5	26.7	35.1	

NOTE: The yield curve factors x_t are formed from an eigenvalue decomposition of the covariance matrix of yield changes $var(\Delta y) = Q\Lambda Q^\top$. The first row is the fraction yield change variance due to the kth factor, $\Lambda(k,k)/\sum_k \Lambda(k,k)$. In the $\gamma^\top f$ column of the first row, we first run a regression $\Delta y_t^{(n)} = a + b\gamma^\top f_t + \varepsilon_t$, and then calculate $100 \times trace(cov(b\gamma^\top f))/trace(cov(\Delta y))$. The second row decomposes the variance of $\gamma^\top f$ into components due to each factor. We find α in $\gamma^\top f_t = \alpha^\top x_t$ and then calculate $100 \times \alpha(k)^2 \Lambda(k,k)^2 / var(\gamma^\top f)$. The third row presents R^2 from forecasting regressions of the average (across maturities) excess return on the factors, $\overline{rx}_{t+1} = \alpha + \beta x_t^{(k)} + \varepsilon_{t+1}$. The fourth row presents R^2 from corresponding multiple regressions $\overline{rx}_{t+1} = \alpha + \beta_1 x_t^{(1)} + ... + \beta_k x_t^{(k)} + \varepsilon_{t+1}$. Sample 1964-2001.

In the second row, we ask how $\gamma^\top f$ is related to the yield curve factors. The loadings (regression coefficients of $\gamma^\top f_t$ on the yield curve factors at time t) are not that interesting, so we calculate the fraction of the variance of $\gamma^\top f$ due to each of the (orthogonal) factors in turn. $\gamma^\top f$ is correlated with the slope factor, so the slope factor explains the largest fraction, 37% of $\gamma^\top f$ variance. Curvature accounts for only 0.4% of $\gamma^\top f$ variance - the return-forecasting factor $\gamma^\top f$ really does have nothing to do with the curvature factor. The spreads factor, essentially insignificant for yields turns out to be quite significant at 42% for explaining expected returns through $\gamma^\top f$. The zigzag factor is also important, explaining 21%.

The third row of Table 8 asks how well we can forecast returns using yield curve factors in place of $\gamma^\top f$. The yield curve factors individually do a poor job of forecasting returns with R^2 under 10%. Worse, the order is reversed — the "level" factor has the worst forecast performance, while the tiny "zigzag" factor has the best forecast performance.

The fourth row forecasts returns using the factors up to and including the factor listed in the column. For example, in the "slope" column, we forecast returns using level and slope factors. By the time we reach the zigzag column, of course, the actual $\gamma^\top f$ is in the span of the included factors, so we recover the 35% R^2. What's interesting about this row is how deep into the small factors we have to go to forecast returns. Level and slope together only produce a 23% R^3, little better than the single Fama-Bliss slope. Curvature and spreads add little. Even though 99.4% of yield change variance has already been explained, we have to include the tiny zigzag factor to reproduce the yield curve patterns that forecast excess returns.

Summary and implications

Our return-forecasting factor – the linear combination of yields that captures variation over time in bond *expected returns* – turns out to have little to do with yield curve factors – the linear combinations of yields that capture variation over time in bond *yields*. The return-forecasting factor does nothing to explain variation in yields, and the important yield curve factors turn out to have little forecast power for excess returns. Most variation in $\gamma^\top f$ and hence in expected excess returns is due to linear combinations of yields that are orthogonal to traditional yield curve factors, are tiny sources of yield curve movement, and are typically ignored in explicit term structure models.

If yields or forward rates followed an *exact* factor structure then all state variables including $\gamma^\top f$ would be functions of that exact factor structure. However, when as in fact yields do not follow an exact factor structure, an important state variable like $\gamma^\top f$ can be hidden in the small idiosyncratic factors that are often dismissed as minor specification or measurement errors.

This suggests a reason why the return forecast factor $\gamma^\top f$ has not been noticed before. Most studies *first* reduce yield data to a small number of factors and *then* look at expected returns. To see expected returns, it's important *first* to look at expected returns and *then* investigate reduced factor structures. A reduced factor representation for yields that is to capture the expected return facts in this data must include the return-forecasting factor $\gamma^\top f$ *as well as* yield curve factors such as the level and slope, even though inclusion of the former will do almost nothing to fit yields, i.e. to reduce pricing errors.

5 Checks and extensions

5.1 Historical and subsample performance

Figure 6 plots the forecast of the holding period excess returns on 3-year bonds implied by the Fama-Bliss regression of Table 3 (top), the forecast from the regression on the return-forecasting factor from Table 2 (middle, i.e. $b_3\left(\gamma^\top f_t\right)$) and the actual holding period returns (bottom). The forecast made at time $t-1$ for time t is plotted at time t, so you can directly compare each forecast with its outcome.

For many episodes, the return-forecasting factor and the forward spread agree. This pattern is particularly visible in the three swings from 1975 to 1982. The return-forecasting factor is correlated with the forward spread. However, the figure shows the much better fit of the return-forecasting factor in many episodes, including the late 1960s,

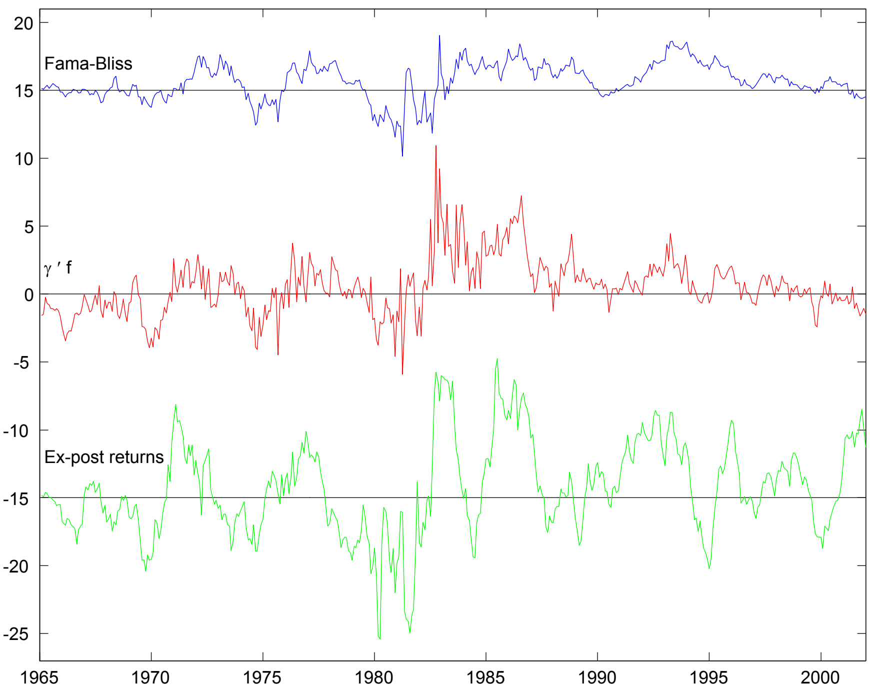

Figure 6: Forecast and actual excess returns of 3-year bonds. Top: Fitted value using Fama-Bliss regression, 3-year forward spread. Middle: Fitted value using the restricted regression on all forward rates, $b_3\gamma^\top f$. Bottom: ex-post excess returns. The forecasts in the top two lines are graphed at the date of the return; the forecast made at $t-1$ is graphed at year t to line up with the ex-post return at year t. The top and bottom graphs are shifted up and down 15% for clarity.

the turbulent early 1980s, the late 1980s, and the mid 1990's. (Campbell 1995 highlights the latter as a particularly challenging episode for yield curve models.) The improved R^2 is not driven by spurious forecasting of one or two unusual data points. Both the return-forecasting factor and the Fama-Bliss regression badly miss the last two years of the sample – they predict slightly negative returns where instead bond returns have been strongly positive as interest rates declined.

Table 9 reports a breakdown by subsamples of a regression of average (across maturity) excess returns $\overline{rx}_{t+1}$ on forward rates. The first set of columns run the average return on the forward rates separately. The second set of columns runs the average return on the return-forecasting factor $\gamma^\top f$ where γ is estimated from the full sample. This regression moderates the tendency to find spurious forecastability with 5 right hand variables in short time periods.

The first row of Table 9 reminds us of the full sample result – the pretty tent-shaped coefficients and the 0.35 R^2. Of course, if you run a regression on its own fitted value you get a coefficient of 1.0 and the same R^2, as shown in the two right hand columns of the first row.

The second row shows the effect of the last two years in the sample, in which $\gamma^\top f$ and the Fama-Bliss regression both forecast slightly negative expected excess returns, but in fact long term bonds did well. Without these last two years, the R^2 rises to 0.40.

Table 9. Subsample analysis

	γ_0	γ_1	γ_2	γ_3	γ_4	γ_5	R^2	$\gamma^\top f$	R^2
1964:01-2001:12	−3.9	−2.0	1.0	2.9	0.7	−2.1	0.35	1.00	0.35
1964:01-1999:12	-4.4	−2.0	0.9	2.9	0.8	-2.1	0.40	1.05	0.40
1964:01-1979:08	−5.4	−1.3	1.3	2.5	-0.1	−1.7	0.32	0.78	0.28
1979:08-1982:10	−32.6	0.8	0.5	1.2	0.6	−0.7	0.78	0.84	0.29
1982:10-2001:12	−3.5	−1.0	1.1	1	1.7	−2.1	0.27	0.88	0.23
1964:01-1969:12	0.6	−1.3	0.2	2.0	0.5	−1.9	0.31	0.71	0.24
1970:01-1979:12	−9.7	−1.4	0.5	2.4	0.3	−0.6	0.22	0.71	0.17
1980:01-1989:12	−11.9	-2.2	1.5	2.6	1.0	−1.8	0.42	1.15	0.37
1990:01-1999:12	−13.8	-1.6	0.5	4.3	1.5	−2.5	0.71	1.83	0.51
2000:01-2001:12								0.09	0.005

NOTE: Subsample analysis of average return-forecasting regressions. For each subsample, the first set of columns present the regression

$$\overline{rx}_{t+1} = \gamma^\top f_t + \overline{\varepsilon}_{t+1}.$$

The second set of columns report the coefficient estimate b and R^2 from

$$\overline{rx}_{t+1} = b\left(\gamma^\top f_t\right) + \varepsilon_{t+1}$$

using the γ parameter from the full sample regression. Overlapping annual forecasts using monthly data 1964-2001.

The third set of rows examine the period before, during, and after the momentous period 1979:8-1982:10, when the Fed changed operating procedures, interest rates were very volatile, and inflation declined and became much less volatile. The broad pattern of coefficients is the same before and after. The 0.73 R^2 looks dramatic in the experiment, but this period really only has three data points and 5 right hand variables. When we constrain the pattern of the coefficients in the right hand pair of columns, the R^2 is the same as the earlier period. It is comforting that the forecasts are so similar in the vastly different regimes of the pre and post experiment periods.

The last set of rows break down the regression by decades. Again, we see the pattern of the coefficients is quite stable. The R^2 is worst in the 70s, a decade dominated by inflation. This suggests that the forecast power derives from changes in the real rather than nominal term structure. The R^2 rises to a dramatic 0.71 in the 1990s, and still 0.51 when we constrain the coefficients γ to their full sample values. The first two years of the 2000 decade are too little to say anything meaningful about the unconstrained regression, but the regression on $\gamma^\top f$ shows again the low R^2 in these two years – the forecast was small, and the outcome was large.

5.2 Real time forecasts

Investors in, say, 1982, do not have access to our full sample to estimate the parameters of the return-forecasting model, so they will not forecast as well. How well can one forecast bond excess returns using real time data? Of course, the conventional rational-expectations answer to this question is that investors have historical information and evolved rules of thumb that summarize far longer time series than our data set, so their expectations will have converged long before ours. Still, it is an interesting robustness exercise to see how well an investor could do who has to estimate the forecasting rule based only on our data from 1964 up to the time the forecast must be made, and it would be discomforting if we could only see forecast power in sample.

Figure 7 contrasts the full sample and the real time forecasts. The top line, marked "full sample" presents the fitted value of the regression $rx_{t+1} = \gamma^{\top} f_t + \varepsilon_{t+1}$ using the full sample 1964:1-2001:12 to estimate the parameters γ. The bottom line presents the same fitted values, but at each time t, the regression is reestimated using data from 1964:1 to time t only.

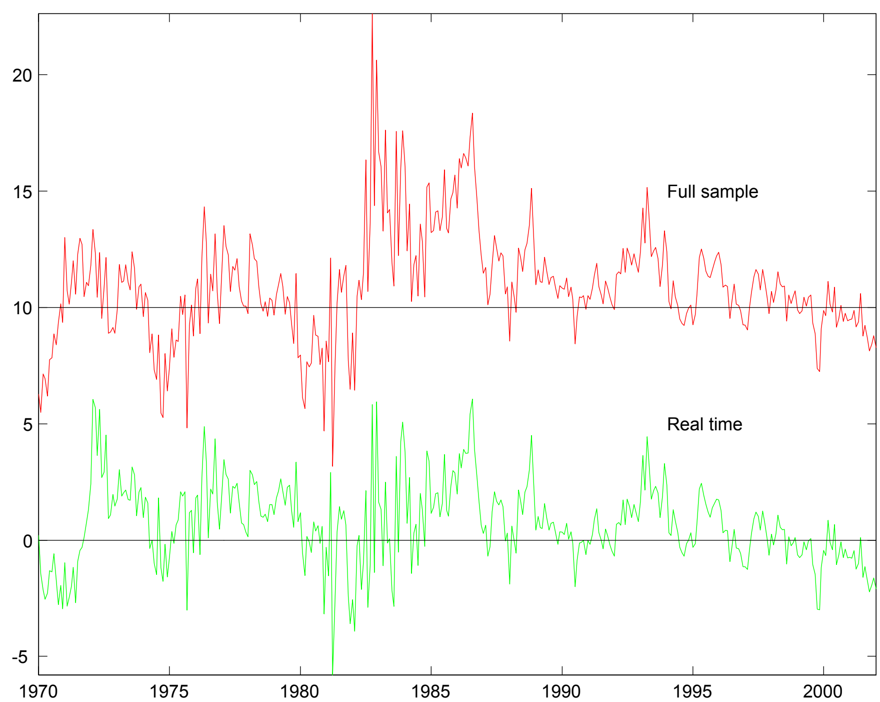

Figure 7: Comparison of full sample and real time forecasts of average (across bond maturities) one year excess returns. "Full sample" is the fitted value of the regression $\overline{rx}_{t+1} = \gamma^{\top} f_t + \varepsilon_{t+1}$ using 1964:1-2001:12 data to estimate the parameter γ. "Real time" uses data from 1964:1 to time t only to estimate the same regression.

The full sample and real time forecasts are quite similar. Even though the regression only starts the 1970s with 6 years of data, it still captures the same pattern of bond expected returns. By the big forecasts of 1987, the full sample and real time forecasts are essentially identical. The only significant discrepancy is in the 1983-1984 period. Here, the real time forecast is a good deal lower than the full sample forecast.

The forecasts are similar, but are they similarly successful? Figure 8 compares them with a simple calculation. We calculate "trading rule returns" as

$$\overline{rx}_{t+1} \times E_t(\overline{rx}_{t+1}) = \overline{rx}_{t+1} \times \left(\gamma^\top f_t\right),$$

and then we cumulate these returns so that the different cases can be more easily compared. (If one follows a linear trading rule to invest $\$1 \times E_t(R^e_{t+1})$ in each end of a zero - cost portfolio with excess return R^e_{t+1}, then the profit from this strategy is $R^e_{t+1} \times E_t(R^e_{t+1})$. We use logs rather than levels, hence quotes around "trading rule." The calculation is also T times the covariance of the forecasted variable $\overline{rx}_{t+1}$ with the forecast $E_t \overline{rx}_{t+1}$, i.e. the numerator of the forecast regression coefficient, so it has a purely statistical interpretation as well.) For the Fama - Bliss regressions, the figure calculates the expected excess return of each bond from its matched forward spread, and then finds the average expected excess return across maturities. The full sample lines use full sample estimates of the regressions. The real time lines use regression estimates only up to time t to calculate $E_t(\overline{rx}_{t+1})$. We start in 1975, with 10 years of data to estimate the return forecasts.

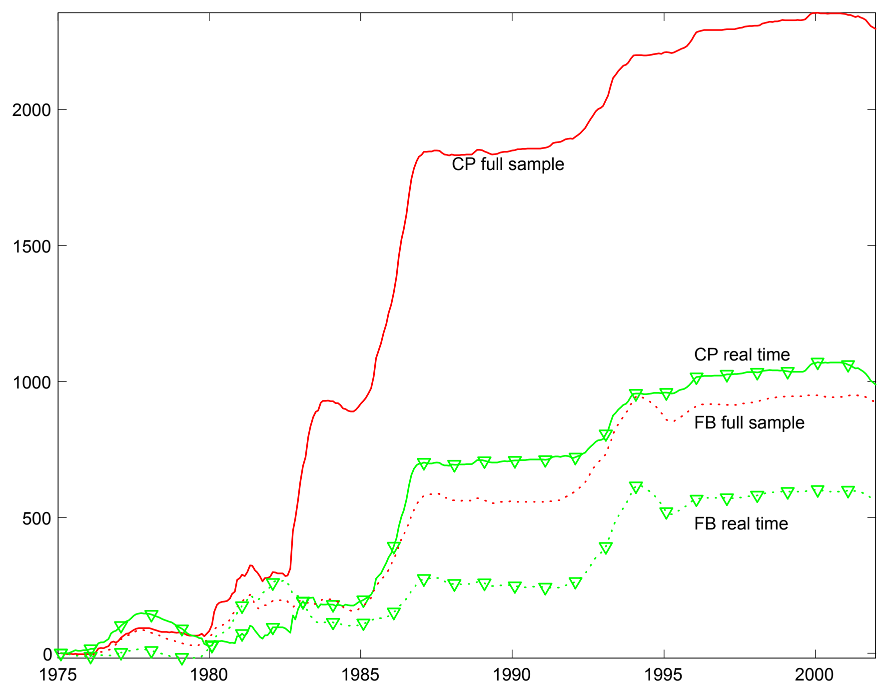

Figure 8: Cumulative profits fromr 'trading rules' using full sample and real time information. Each line plots the cumulative value of $\overline{rx}_{t+1} \times E_t(\overline{rx}_{t+1})$. $E_t(\overline{rx}_{t+1})$ are formed from the full 1964-2001 sample or "real time" data from 1964-t as marked. The CP lines use the forecast $\overline{rx}_{t+1} = \gamma^\top f_t$. The FB (Fama-Bliss) lines forecast each excess return from the corresponding forward spread, and then average the forecasts across maturities.

The full sample line of Figure 8 shows vividly the character of this return-forecasting exercise: it produces occasional spectacular gains, as in 1983, 1987, and 1994, while producing nearly nothing (and recommending small positions) for long periods. The last two years of the sample lost a little money, as the forecast was for slightly negative bond returns, while in fact long term bonds made money as interest rates declined. The loss of R^2 was primarily due to a large residual, rather than a forecast of a wrong sign.

The real time forecast overall produces only about half of the cumulative profits as does the full sample estimate. This underperformance essentially all comes from the 1983 period. The real time forecast had not quite settled on the coefficients that would let it forecast the spectacular return obtained by the full sample estimate in this period. This finding mirrors the difference in forecast for 1983 shown in Figure 7. At this point, the regression has had only 19 years to estimate the 6 γ from colinear forward rates. However, the real time forecast captures almost all of the impressive gains of the 1987 and 1994 episodes. Interestingly, neither the Fama-Bliss full sample or real time estimates capture this 1983 episode either. In fact, they lose money here.

Overall, we conclude that while the forecasts do degrade somewhat using real-time data (and given the limitations of our particular data set), the overall pattern remains. It does not seem to be the case that the forecast power, or the improvement over the Fama-Bliss forecasts, requires the use of ex-post data.

5.3 Other data

The Fama-Bliss data are interpolated zero-coupon yields. To check whether the predictability results are generated by the interpolation scheme, we run the regressions with McCulloch-Kwon data, which use a different interpolation scheme to derive zero-coupon yields from treasury data.

Table 10 compares the R^2 and γ estimates using McCulloch-Kwon and Fama-Bliss data over the McCulloch-Kwon sample (1964:1-1991:2). The R^2 are very similar across the two datasets. The tent-shape of γ estimates is even more pronounced in McCulloch-Kwon data than in the Fama-Bliss data. (Interestingly, the low 0.05 R^2 for the Fama-Bliss 5-year bond regression is raised to 0.12, comparable to the other maturities, in the McCulloch-Kwon data.)

5.4 Additional Lags

Do additional lags of forward rates help to forecast bond returns? We start with unrestricted regressions. Table 11 reports the R^2 in the columns labeled "*all* f." Specification (1) repeats the baseline regression of excess returns on one lag of forward rates from Table 1 for comparison. Specification (2) presents the R^2 with additional one-month lagged forward rates, $f_{t-1/12}$. The R^2 rise by about 0.05 to 0.39-0.43. A χ^2 test overwhelmingly rejects the hypothesis that the coefficients on the additional lag of forward rates are zero. The extra lag helps.

Table 10. Comparison with McCulloch-Kwon data

R^2

	All f_t		$\gamma^\top f_t$		$f_t^{(n-1\to n)} - y_t^{(1)}$	
n	M-K	F-B	M-K	F-B	M-K	F-B
2	0.39	0.39	0.39	0.38	0.16	0.15
3	0.37	0.39	0.37	0.40	0.15	0.16
4	0.36	0.41	0.36	0.42	0.13	0.17
5	0.35	0.37	0.35	0.38	0.12	0.05

Coefficients

	γ_0	γ_1	γ_2	γ_3	γ_4	γ_5
McCulloch-Kwon	-5.11	-2.52	1.78	3.19	1.94	-3.82
Fama-Bliss	-4.73	-1.84	0.95	2.98	0.52	-2.10

NOTE: The data are McCulloch-Kwon and Fama-Bliss CRSP zero-coupon yields starting 1964:1 until the end of the McCulloch-Kwon dataset, 1991:12. The upper panel shows R^2 from the regressions corresponding to Tables 1-3. The regressions run excess log returns $rx_{t+1}^{(n)}$ on the regressors indicated on top of the table: forward spread $f_t^{(n-1\to n)} - y_t^{(1)}$, all forwards f_t, and the return-forecasting factor $\gamma^\top f_t$. The lower panel shows the estimated γ coefficients in the regression of average returns on forward rates $\overline{rx}_{t+1} = \gamma^\top f_t + \bar{\varepsilon}_{t+1}$. McCulloch-Kwon data are downloaded from
http://www.econ.ohio-state.edu/jhm/ts/mcckwon/mccull.htm.

The coefficients in these regressions (not reported) have the familiar tent shape, and are roughly the same for the first and second lag. These fact suggests that a one-month moving average of forward rates predicts bond returns. Specification (3) presents the R^2 with a moving average $\left(f_t + f_{t-1/12}\right)/2$ regressor. The R^2 is nearly the same, and the restriction is not rejected statistically, so this moving average seems a good way to include the lagged information.

Since the unrestricted coefficients have the usual tent shape, Table 11 specification (2) investigates a regression of excess returns on an additional lag of the return-forecasting factor $\gamma^\top f_t$. The R^2 goes up from the corresponding specification (1) to 0.38-0.41, nearly equal to the 0.39-0.43 values from the unconstrained two-lag regression in the *all f* column. Once again, the single factor seems to capture all of the information in all 5 forward rates. The coefficients again suggest a moving average $\gamma^\top \left(f_t + f_{t-1/12}\right)/2$ as a state variable. Specification (3) examines this regression. The additional constraint on the coefficients makes little difference to the R^2, and the coefficients themselves are very close to the value in the first regression.

Table 11. Regressions of excess returns on additional lags of forward rates

	(1) Baseline			(2) One extra lag				(3) Moving average		
	all f	$\gamma^\top f$		*all f*	$\gamma^\top f$			*all f*	$\gamma^\top f$	
n	R^2	R^2	$\gamma^\top f_t$	R^2	$\gamma^\top f_t$	$\gamma^\top f_{t-1/12}$	R^2	R^2	R^2	$\frac{\gamma^\top(f_t+f_{t-1/12})}{2}$
2	0.34	0.33	0.46	0.39	0.26	0.27	0.38	0.39	0.38	0.53
3	0.34	0.34	0.86	0.41	0.50	0.47	0.39	0.40	0.39	0.97
4	0.37	0.37	1.23	0.43	0.74	0.66	0.41	0.43	0.41	1.40
5	0.34	0.34	1.45	0.41	0.74	0.94	0.40	0.41	0.40	1.68

NOTE: Specification (1) forecasts excess returns with one lag of forward rates. The "all f" column reports the R^2 from unrestricted regressions, $rx^{(n)}_{t+1} = \beta^\top f_t + \varepsilon^{(n)}_{t+1}$. The $\gamma^\top f$ columns report slope coefficients b_n and the R^2 from restricted regressions $rx^{(n)}_{t+1} = b_n\left(\gamma^\top f_t\right) + \varepsilon^{(n)}_{t+1}$. Specification (2) forecasts excess returns with one extra lag of forward rates. The "*all f*" column reports the R^2 from the unrestricted regression $rx^{(n)}_{t+1} = \beta_1^\top f_t + \beta_2^\top f_{t-1/12} + \varepsilon^{(n)}_{t+1}$. The $\gamma^\top f$ columns report slope coefficients b_1, b_2 and the R^2 from regressions $rx^{(n)}_{t+1} = b_{n,1}\left(\gamma^\top f_t\right) + b_{n,2}\left(\gamma^\top f_{t-1/12}\right) + \varepsilon^{(n)}_{t+1}$. Specification (3) forecasts excess returns with a moving average of forward rates. The "*all f*" column reports the R^2 from unrestricted regressions $rx^{(n)}_{t+1} = \beta_n\left(f_t + f_{t-1/12}\right)/2 + \varepsilon^{(n)}_{t+1}$. The $\gamma^\top f$ columns report slope coefficients b_n and R^2 from restricted regressions $rx^{(n)}_{t+1} = b_n\gamma^\top\left(f_t + f_{t-1/12}\right)/2 + \varepsilon^{(n)}_{t+1}$. OLS on overlapping monthly data 1964-2001.

In sum, one additional lag enters with statistical and economic significance, and a moving average of the return-forecasting factor offers an excellent summary of the extra information. Additional monthly or annual lags are not much help.

The extra lag means that an VAR(1) *monthly* representation, of the type specified by nearly every explicit term structure model, does not capture the patterns we see in *annual* return forecasts. To see the annual return forecastability, one must either look directly at annual horizons, or adopt more complex time-series models. To quantify this point, we fit an unconstrained monthly yield VAR, $y_{t+1/12} = Ay_t + \varepsilon_t$, and found the implied annual VAR representation $y_{t+1} = A^{12}y_t + u_t$. Yields, forward rates, and returns are all linear functions of each other, so we can calculate return forecasts and R^2 directly from an annual VAR representation, and the return forecasts implied by a directly estimated annual VAR $y_{t+1} = By_t + v_{t+1}$ are precisely the same as we found above using forward rates. The implied annual return forecasts from a monthly VAR hide the tent-shaped pattern of coefficients. Most importantly, they hide the single-factor structure: The yield on n-year bonds stands out as a much more important forecaster of the n-year bond return, so no single combination of yields or forward rates summarizes bond return forecastability. The forecast R^2 are cut to 0.21-0.26 instead of 0.34-0.39.

What do we do with these facts? This data summary must be a prelude to the construction of an economic model of the term structure. But additional lags are awkward forecasting variables for bond yields. In any model, bond prices are period t expected values of future discount factors, so a full set of time t bond yields drive out time $t-1/12$ yields in forecasting anything. To integrate the lagged yields into the analysis, and to reconcile VAR representations at different horizons, it may be more fruitful to specify a model for monthly yields in which date t yields are truly sufficient state variables, but they are contaminated with i.i.d. measurement error so lagged yields help to reveal the true date-t yield. This effort also requires an extension to additional maturities and integration with an explicit term structure model. In the interest of space, we leave this large project for future work.

6 Multiple return-forecasting factors

6.1 Testing the single-factor model

The parameters of the unrestricted ($rx_{t+1} = \beta f_t + \varepsilon_{t+1}$) return forecasting regressions and those of the restricted single-factor model ($rx_{t+1} = b\left(\gamma^\top f_t\right) + \varepsilon_{t+1}$) are *individually* indistinguishable (Figure 1 and 2), but are they *jointly* equal? Does an overall test of the single-factor model's restrictions reject?

The moments underlying the unrestricted regressions (3) are the regression forecast errors multiplied by forward rates (right hand variables),

$$g_T(\beta) = E\left(\varepsilon_{t+1} \otimes f_t\right) = 0. \tag{5}$$

By contrast, our two-step estimate of the single factor model sets to zero the moments

$$E\left[\left(1_4^\top \varepsilon_{t+1}\right) \otimes f_t\right] = 0, \tag{6}$$
$$E\left[\varepsilon_{t+1} \otimes \left(\gamma^\top f_t\right)\right] = 0. \tag{7}$$

We use these moments to compute the GMM standard errors in Table 2. The restricted model $\beta = b\gamma^\top$ does not set all the moments (5) to zero, $g_T(b\gamma^\top) \neq 0$. We can compute the "J_T" χ^2 test that the remaining moments are not too large. To do this, we express the moments (6)-(7) that define the estimate of the one-factor model as linear combinations of the moments (5) that are set to zero, $a_T g_T = 0$. Then we apply Hansen's (1982) Theorem 3.1 (Details are in the Appendix). We also compute a Wald test of the joint parameter restrictions $\beta = b\gamma^\top$. We find the standard GMM distribution $cov(vec(\beta))$, and then compute the χ^2 statistic $\left[vec(b\gamma^\top) - vec(\beta)\right]^\top cov(vec(\beta))^{-1}\left[vec(b\gamma^\top) - vec(\beta)\right]$. (*vec* since β is a matrix of coefficients).

The tests all reject the single factor model. The precise χ^2 values are sensitive to the number of lags, weighting scheme, and the use of restricted or unrestricted moments in the S matrix calculation, but the tests all reject with well below 1% probability values for all reasonable choices of these specifications. We have reproduced both J_T and Wald tests in finite samples by simulation, and the rejection is confirmed.

To understand the rejection, consider forecasting a linear combination of bond excess returns, $\delta^\top rx_{t+1}$, where the weights are orthogonal to the b loadings, $\delta^\top b = 0$. The single factor model predicts that all such linear combinations of excess returns are not forecastable: it predicts $E_t\left(\delta^\top rx_{t+1}\right) = \delta^\top b\gamma^\top f_t = 0$. If the single factor model is rejected, it must mean that such linear combinations *are* forecastable. It means there are *multiple* factors in expected excess bond returns–additional linear combinations of forward rates that forecast these bond portfolios. Clearly, we should understand these extra factors.

6.2 A multiple-factor model

To understand and interpret the additional forecastability of bond returns that rejects the single factor model, we express the regression coefficients β from the unconstrained regressions

$$E_t\left(rx_{t+1}\right) = \beta f_t$$

into an exactly-identified multi-factor model,

$$E_t(rx_{t+1}) = B\Gamma^\top f_t = \beta f_t. \tag{8}$$

Γ is a 6×4 matrix and $\Gamma^\top f_t$ is a 4×1 vector of factors that drive expected returns. B is a 4×4 matrix of loadings, that describe how each expected return is affected by a movement in each factor. Something like the b and γ we have found above will be the first columns of B and Γ. The remaining columns will help us to interpret the additional linear combinations of forward rates $\Gamma^\top f_t$ that evidently can forecast additional linear combinations of returns. Obviously, this rewriting adds nothing to the unrestricted regression, and there are an infinite number of ways to break up β in this way; we present a decomposition that gives economic intuition and verifies the statistical significance of all the remaining factors' forecasting ability.

We can interpret a factor decomposition of the form (8) in terms of forecasting regressions. Consider the expected return of a portfolio δ of bonds,

$$E_t\left(\delta^\top rx_{t+1}\right) = \delta^\top \beta f_t = \delta^\top B\Gamma^\top f_t.$$

We will choose the B matrix to have orthonormal columns. Then, the portfolio choice $\delta_i = B(:,i)$ will reveal the ith factor, since $\delta_i^\top B(:,i) = 1$, $\delta_i^\top B(:,j \neq i) = 0$. (We use MATLAB notation $B(:,i)$ to denote the ith column of the matrix B.) Once we have identified an interesting B, we can estimate Γ and evaluate the economic and statistical significance of multiple factors by running forecasting regressions of $B(:,i)^\top rx_{t+1}$ on forward rates,

$$B(:,i)^\top rx_{t+1} = \Gamma(:,i)^\top f_t + B(:,i)^\top \varepsilon_{t+1}, \tag{9}$$

a simple generalization of our regression of average (across maturities) excess returns on forward rates to estimate γ above. The $B(:,i)$ tell us which portfolio takes maximum advantage of factor i's forecast power, and the regression tests and R^2 tell us how great that power is. (Of course we can also calculate the Γ directly from (8) by $\Gamma^\top = B^{-1}\beta$.)

Now, what is an interesting identification of B and hence Γ? We start with an eigenvalue decomposition of the expected return covariance matrix, in the same way that we often define factor models for yields by an eigenvalue decomposition of the yield covariance matrix,

$$Q\Lambda Q^\top = cov\left[E_t(rx_{t+1}), E_t(rx_{t+1})^\top\right] = \beta cov(f_t, f_t^\top)\beta^\top.$$

Q is an orthogonal matrix of eigenvectors $Q^\top Q = I$ and Λ is a diagonal matrix of eigenvalues. We identify $B = Q$, and, inverting (8), we can calculate $\Gamma^\top = Q^\top\beta$. As usual, the eigenvalue decomposition produces orthogonal factors that each explain the largest possible fraction of expected return variance in turn. (For example, see Mardia, Kent and Bibby 1979, p.215-218.)

The first factor computed by this eigenvalue decomposition turns out to be almost exactly our return-forecasting factor, and it captures 99.4% of the variance of expected returns. As often happens in eigenvalue decompositions, however, the B and Γ loadings of factors past the first don't tell much of a story. Their relative variances are similar at 0.28%, 0.16% and 0.11%. They are selected to explain fractions of variance; when as here the remaining factors all have about the same and tiny variance, the eigenvalue decomposition is not very good at sorting them out.

Therefore, we find a recursive identification of the remaining factors more insightful. We identify the second factor to affect only 4 and 5 year bond returns and be orthogonal to the first,

$$\begin{aligned} B(:,2)^\top B(:,1) &= 0, \\ B(:,2)^\top B(:,2) &= 1, \\ B(1{:}2,2) &= 0. \end{aligned}$$

We identify the third factor to affect 3, 4, and 5 year returns while remaining orthogonal to the first two, and so forth.

Results

The result of this exact identification procedure is

$$B = \left[\begin{array}{ll|rrrr} & & \multicolumn{4}{c}{\text{Factor:}} \\ & & 1 & 2 & 3 & 4 \\ \hline & rx_{t+1}^{(2)} & 0.22 & 0 & 0 & -0.97 \\ \text{Return:} & rx_{t+1}^{(3)} & 0.41 & 0 & -0.91 & 0.09 \\ & rx_{t+1}^{(4)} & 0.58 & -0.76 & 0.27 & 0.13 \\ & rx_{t+1}^{(5)} & 0.67 & 0.65 & 0.32 & 0.15 \end{array}\right].$$

We calculate $\Gamma^\top = B^\top\beta$ from (8) (B is orthonormal so $B^{-1} = B^\top$), and Figure 9 plots the corresponding columns of Γ, i.e. how factors are formed from forward rates. Table 12 summarizes the forecast power of the factors by regressions of the form (9).

Table 12. Multi-factor excess return forecasts.

	Factor 1	2	3	4
$\sigma\left[\Gamma(:,i)^\top f_t\right]$	5.13	0.19	0.26	0.14
Variance %	99.4	0.13	0.26	0.14
R^2	0.35	0.18	0.27	0.16
EH upper 95%	0.14	0.12	0.12	0.12
$\chi^2(5)$	92.4	93.8	107.4	31.7
EH p-value	0.00	0.00	0.00	0.01

NOTE: Regressions of linear combinations of excess returns on forward rates,

$$B(:,i)^\top rx_{t+1} = \Gamma(:,i)^\top f_t + \varepsilon_{t+j}^{(i)}.$$

$\sigma\left[\Gamma(:,i)^\top f_t\right]$ is the standard deviation of the right hand variable in percent. "Variance %" gives the same information as variance divided by total (sum across factors) variance. The χ^2 statistic is the test for joint insignificance of all right hand variables except the constant. "EH upper 95%" are the upper bounds of 95% confidence intervals computed from the bootstrap that imposes the expectations hypothesis. "EH p-value" reports the p-values of Wald tests for joint significance of the right hand variables from the same bootstrap. Coefficient estimates are plotted in Figure 9.

The first factor is almost exactly the return-forecasting factor $\gamma^\top f_t$ we found above. The first column of B shows that the first factor moves all expected excess returns in the same direction, and long maturity returns more than short maturity returns. This column is almost exactly (up to normalization $\sum_i b_i^2 = 1$ rather than $\sum_i b_i = 4$) the b loading that we found in Table 2. We omitted the first column of Γ from Figure 9 for clarity; it is the familiar tent-shaped function of forward rates, visually indistinguishable from Figure 1. The first two rows of Table 12 show that this factor accounts for the overwhelming majority of the variance of return forecasts. Thus, just about any linear combination of returns regressed on forward rates will give almost exactly this factor; the fact that above we regressed $\frac{1}{4} \times 1_4^\top rx_{t+1}$ on forward rates and now we regress $B(:,1)^\top rx_{t+1}$ on forward rates makes almost no difference whatever in the estimate of $\Gamma(:,1)$. The other factors are so much smaller (first and second rows, Table 12) that you have very carefully to define portfolio weights orthogonal to $B(:,1)$ in order to see them.

The second column of B shows that the second factor moves the 5-year return up and the 4-year return down, by almost the same amounts. The corresponding factor $\Gamma(:,2)$ in Figure 9 – the function of yields that signals a high return to the long 5, short 4 portfolio – is basically the 5-4 year yield spread. The second column of Table 12 verifies that although this factor is tiny – normalizing its effect on expected excess returns so that $B(:,2)^\top B(:,2) = 1$, the standard deviation of $\Gamma(:,2)^\top f_t$ is only 0.19 % – it does significantly forecast its corresponding bond return portfolio (long 5, short 4). The R^2 is

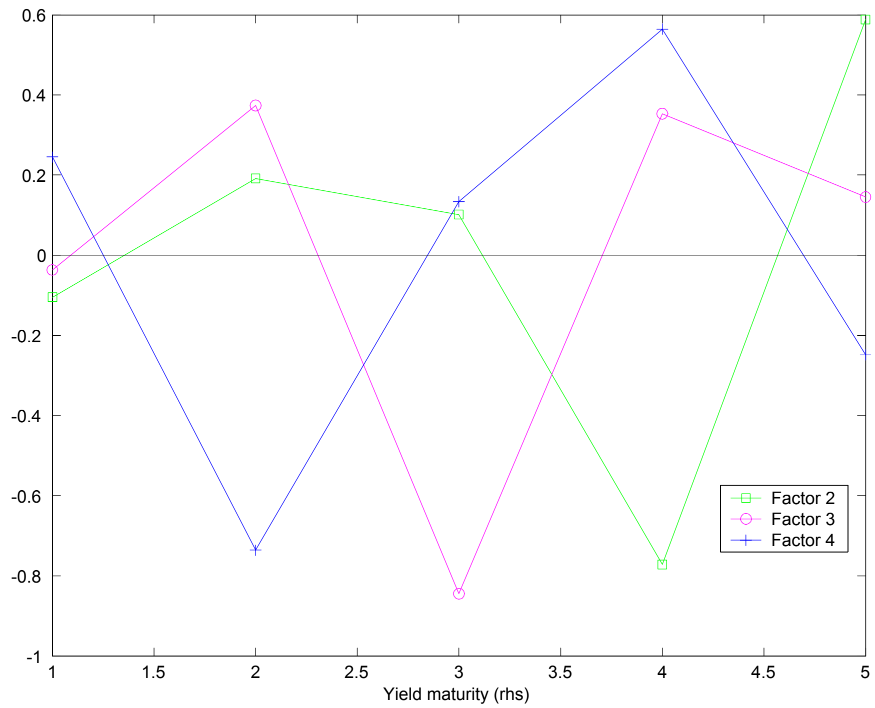

Figure 9: Expected excess return factors. The lines are the second through fourth columns of the matrix Γ in the representation $E_t\left(rx_{t+1}\right) = B\Gamma^\top f_t$. The lines are scaled by their sum of squares to fit on the same graph.

0.18, about the size of the Fama-Bliss R^2, and the GMM and small sample tests soundly reject lack of forecastability.

These results paint an intuitive picture. When the 5-4 yield spread is high − when the 5 year price "low" and the 4 year price is "high"− a portfolio that buys 5 and shorts 4 does well. This is the classic "spread trade" or "convergence trade."

The remaining factors work the same way. Looking at the B matrix, the third factor lowers the 3-year bond return, balanced by raising the 4 and 5 year returns, while the fourth factor lowers the 2-year bond return, raising all the others. Looking at Figure 9, the signal for the third factor is primarily a high 3-year bond yield, and the signal for the fourth factor is primarily a high 2-year bond yield. Like the second factor, these are small idiosyncratic movements in the yield curve that are reversed, giving a profit to those who buy low and sell high, hedging movements in the overall level of the yield curve. In Table 12, these factors again are tiny (first two rows), but significantly forecast the tiny returns on their corresponding bond portfolios.

Interpretation of additional factors

What should we make of these additional factors? They are very small. There are only small idiosyncratic movements of the individual bond yields, and they forecast small returns to the corresponding portfolios. But the forecasted returns form a nearly zero-cost portfolio (columns of B), so a huge short position matched by a huge long position can make money, as in the famous LTCM 29.5 - 30 year trade. Furthermore, since most bond return risk is "level" risk that moves all long-term bond prices together, these portfolios have small variance and potentially large Sharpe ratios. This low variance – equivalently, the respectable forecast R^2 values – means that the small forecastable movements in expected returns are well-measured and hence statistically significant, accounting for the statistical rejection of the appealing single factor model.

They could be real. The 29.5-30 year spread trade was real (among others, see Krishnamurthy 2002). Much of what bond traders do is precisely to pick bonds that are a little over or underpriced relative to a yield curve, leverage like mad, and wait for the small price fluctuations to melt away. On the other hand, measurement error also induces small transitory variation in prices. Our CRSP data are interpolated by Fama and Bliss from Federal Reserve surveys; surely not the kind of data on which one would recommend highly leveraged spread trades. The analysis in Section 5.4 also suggests i.i.d. measurement error in our data.

In any case, our main focus is to characterize the *economically* interesting variation in expected bond returns. For that purpose, the single factor model that explains 99.4% of the variance of bond expected returns is obviously the model on which we want to focus, even though our data show the kind of tiny fluctuations that may give bond traders profitable spread trades as well, and for this reason the single factor model is statistically rejected.

Two step vs. efficient estimates

At this point, we can finally answer the question, "Why estimate the model in Table 2 with an ad-hoc two step procedure, rather than use efficient GMM?" Under the null that the single factor model is true, efficient GMM ($\min_{\{b,\gamma\}} g_T(b\gamma^\top)^\top S^{-1} g_T(b\gamma^\top)$) produces asymptotically more efficient estimates. However, our model is statistically rejected. We want good estimates of a rejected model, not efficient estimates of a true single-factor model. Efficient GMM can do a poor job of that task, even in samples in which one can trust estimation and inversion of a 24×24 spectral density estimate with 12 lags.

The crucial question is, what moments will GMM use to choose γ, the linear combination of forward rates that forms the single factor? Once the single-factor parameter γ is estimated, even efficient GMM estimates the remaining b coefficients by regressions of each return on $\gamma^\top f_t$. In turn, taking linear combinations of moments is the same thing as forming a portfolio, so the crucial question becomes, "which portfolio of excess returns $\delta^\top rx_{t+1}$ will efficient GMM regress on all forward rates to estimate γ?" The answer is that efficient GMM pays attention to well-measured linear combinations of moments, guided by S, not "large" or "economically interesting" moments. For example, suppose that one of the tiny additional factors (say, $\Gamma(:,3)$) forecasts its corresponding tiny linear combination of returns ($B(:,3)^\top rx_{t+1}$) with 100% R^2. No regression error means that this moment is exactly measured, so efficient GMM will estimate parameters γ to make the moments of this regression $E(\varepsilon_{t+1}(b,\gamma) \otimes f_t)$ equal to zero exactly. It will proudly return the estimate $\gamma = \Gamma(:,3)$. Efficient GMM will *completely miss* the economically interesting factor $\Gamma(:,1)$ that describes 99.4% of the variance of expected returns. It will leave extremely large values for the corresponding moments, and it will produce a factor that explains almost *none* of the variance of expected returns.

In our data, the R^2 for all factors are roughly comparable. That means that efficient GMM pays about equal attention to all the regressions of $\Gamma(:,i)$ on forward rates, producing a factor that is a combination of the Γ coefficients rather than an estimate of $\Gamma(:,1)$ or any of the other factors identified above. Still, the resulting single factor model explains very little of the variance of expected excess returns.

We want a GMM estimate of the approximate single factor that explains most of the variance of expected returns, not the one that minimizes the best measured, even if tiny, moments. For that purpose, we want to force GMM to pay attention to a portfolio such as $B(:,1)^\top rx_{t+1}$ or $1^\top rx_{t+1}$ as we did with the two step procedure. The return-forecasting factor is such an overwhelming factor for the variance of expected returns that it doesn't really matter which linear combination we choose, so long as we keep GMM from paying attention to the special linear combinations $B(:,2{:}4)$ that produce our very small, but well measured additional factors.

7 Concluding remarks

This analysis is admittedly incomplete in many respects. We have confined ourselves to a one year horizon. Understanding how expected returns vary across investment horizon is of course a very important issue. We also confined ourselves to 1-5 year maturity bonds, and linking the analysis to shorter maturity bonds is important and challenging. However, these extensions require some subtle time-series analysis, probably including an explicit treatment of the measurement error we seem to see in our analysis of multiple lags. They also require extending our data beyond zero coupon bonds at one year spaced maturities and hence integration with an explicit term structure model since all returns cannot be measured directly. We have not mentioned variances or covariances, and hence we have not mentioned Sharpe ratios or optimal portfolios. Modeling time-varying second moments is likely to require as much effort as modeling the first moments in this paper. Finally, we have not tied the time-varying premia to macroeconomic or monetary fundamentals in more than a suggestive way. All these and more are important extensions, but we have strained space and the reader's patience enough for now.

References

[1] Bekaert, Geert, Robert J. Hodrick, and David Marshall, 1997. "On Biases in Tests of the Expectations Hypothesis of the Term Structure of Interest Rates." *Journal of Financial Economics* 44, 309-48.

[2] Campbell, John Y., 1995. "Some Lessons from the Yield Curve," *Journal of Economic Perspectives* 9, 129-52.

[3] Campbell, John Y. and John H. Cochrane, 1999. "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior," *Journal of Political Economy* 107, 205-51.

[4] Campbell, John Y. and Robert J. Shiller, 1991. "Yield Spreads and Interest Rate Movements: A Bird's Eye View." *Review of Economic Studies* 58, 495-514.

[5] Clews, Roger, 2002. "Asset prices and Inflation." *Quarterly Bulletin*, Bank of England, Summer, 178-85.

[6] Cochrane, John H., 1994. "Permanent and Transitory Components of GNP and Stock Prices," *Quarterly Journal of Economics* CIX, 241-66.

[7] Dai, Qiang and Kenneth J. Singleton, 2002. "Expectation Puzzles, Time-varying Risk Premia, and Affine Models of the Term Structure," *Journal of Financial Economics* 63, 415-41.

[8] Duffee, Gregory, 2002, "Term Premia and Interest Rate Forecasts in Affine Models," *Journal of Finance* 57, 405-43.

[9] Estrella, Arturo, and Gikas A. Hardouvelis, 1991. "The Term Structure as Predictor of Real Economic Activity." *Journal of Finance* 46, 1991, 555-76.

[10] Fama, Eugene F. and Robert R. Bliss, 1987. "The Information in Long-Maturity Forward Rates." *American Economic Review* 77, 680-92.

[11] Fama, Eugene F. and Kenneth R. French, 1989. "Business Conditions and Expected Returns on Bonds and Stocks." *Journal of Financial Economics* 25, 23-49.

[12] Ferson, Wayne, and Campbell R. Harvey, 1999. "Conditioning Variables and Cross-section of Stock Returns." *Journal of Finance* 54, 1325-60.

[13] Ferson, Wayne, and Michael R. Gibbons, 1985. "Testing Asset Pricing Models with Changing Expectations and an Unobservable Market Portfolio." *Journal of Financial Economics* 14, 216-36.

[14] Hamilton, James D. and Kim, Dong H., 1999. "A Re-Examination of the Predictability of the Yield Spread for Real Economic Activity." Forthcoming, *Journal of Money, Credit, and Banking.*

[15] Hansen, Lars Peter and Robert J. Hodrick, 1983. "Risk Aversion Speculation in the Forward Foreign Exchange Market: An Econometric Analysis of Linear Models." In Jacob Frenkel, ed., *Exchange Rates and International Macroeconomics.* Chicago, IL: University of Chicago Press.

[16] Harvey, Campbell R., 1989. "Forecasts of Economic Growth from the Bond and Stock Markets." *Financial Analysts Journal* 45, 38-45.

[17] Ilmanen, Antti, 1995. "Time Varying Expected Bond Returns in International Bond Markets." *Journal of Finance* 50, 481-506.

[18] Krishnamurthy, Arvind, 2002. "The Bond-Old Bond Spread." Forthcoming, *Journal of Financial Economics.*

[19] Krueger, Joel T. and Kenneth N. Kuttner, 1996. "The Fed Funds Futures Rate as a Predictor of Federal Reserve Policy," *Journal of Futures Markets* 16, 965-979.

[20] Lettau, Martin, and Sydney Ludvigson, 2001 "Consumption, Aggregate Wealth and Expected Stock Returns." *Journal of Finance* 56, 815-49.

[21] Mardia, K. V., J. T. Kent, and J. M Bibby, 1979. *Multivariate Analysis*. San Diego, CA: Academic Press.

[22] Rudebusch, Glenn D., 1998. "Do measures of monetary policy in a VAR make sense?" *International Economic Review* 39 (November), 907-31.

[23] Scholtes, Cedric, 2002. "On market-based measures of inflation expectations." *Quarterly Bulletin*, Bank of England, Spring, 67-77.

[24] Söderlind, Paul and Lars E.O. Svensson, 1997. "New Techniques to Extract Market Expectations from Financial Instruments," *Journal of Monetary Economics* 40, 373-429.

[25] Stambaugh, Robert F., 1988. "The information in Forward Rates: Implications for Models of the Term Structure." *Journal of Financial Economics* 22, 3-25.

[26] Stock, James H. and Mark W. Watson, 1989. "New Indexes of Coincident and Leading Indicators." In Olivier Blanchard and Stanley Fischer, Eds., *1989 NBER Macroeconomics Annual.* Cambridge MA: MIT Press.

Appendix

A. Small sample distributions

The data-generating process is a vector autoregression with 12 lags for the vector of yields

$$y_t = A_0 + A_1 y_{t-1/12} + \ldots + A_{12} y_{t-1} + \varepsilon_t.$$

Vector autoregressions based on fewer lags (such as one or two) are unable to replicate the long-horizon forecastability of the short rate documented in Table 10.

To impose the expectations hypothesis we start with an AR(12) for the short rate

$$y_t^{(1)} = a_0 + a_1 y_{t-1/12}^{(1)} + \ldots + a_{12} y_{t-1}^{(1)} + \varepsilon_t.$$

We then compute long yields as

$$y_t^{(n)} = \frac{1}{n} E_t \left(\sum_{i=1}^{n} y_{t+i-1}^{(n)} \right), n = 2, \ldots, 5.$$

To compute the expected value in this last expression, we expand the state space to rewrite the dynamics of the short rate as a vector autoregression with 1 lag. The 12-dimensional vector $x_t = \left[y_t^{(1)} \; y_{t-1/11}^{(1)} \; \cdots \; y_{t-11/12}^{(1)} \right]^\top$ follows

$$x_t = B_0 + B_1 x_{t-1/12} + \Sigma u_t$$

for

$$B_0 = \begin{pmatrix} a_0 \\ 0 \end{pmatrix}, B_1 = \begin{pmatrix} a_1 \cdots & a_{12} \\ I_{11\times 11} & 0_{11\times 1} \end{pmatrix}, \Sigma = \begin{pmatrix} 1 & 0_{1\times 11} \\ 0_{11\times 1} & 0_{11\times 11} \end{pmatrix}.$$

The vector $e_1 = [1 \;\; 0_{1\times 11}]$ picks the first element in x_t, which gives $y_t^{(1)} = e_1 x_t$. Longer yields can then be easily computed recursively as

$$y_t^{(n)} = \frac{n-1}{n} y_t^{(n-1)} + \frac{1}{n} e_1 \left(\left(\sum_{i=0}^{n-1} B_1^i \right) B_0 + B_1^n x_t \right), n = 2, \ldots, 5,$$

with the understanding that B_1^i for $i = 0$ is the 12×12 identity matrix.

B. GMM estimates and tests

The unrestricted regression is

$$rx_{t+1} = \beta f_t + \varepsilon_{t+1},$$

The moment conditions of the unrestricted model are

$$g_T(\beta) = E(\varepsilon_{t+1} \otimes f_t) = 0. \tag{10}$$

The restricted model is $\beta = b\gamma^\top$, with the normalization $b^\top 1_4 = 4$.

We focus on a 2-step OLS estimate of the restricted model – first estimate average (across maturities) returns on f, then run each return on $\hat{\gamma}^\top f$:

$$\overline{rx}_{t+1} = \gamma^\top f_t + \bar{\varepsilon}_{t+1}, \tag{11}$$
$$rx_{t+1} = b\left(\hat{\gamma}^\top f_t\right) + \varepsilon_{t+1}. \tag{12}$$

The estimates satisfy $1_4^\top b = 4$ automatically.

To provide standard errors for the two-step estimate, we use the moments corresponding to the two OLS regressions (11) and (12),

$$\tilde{g}_T(b,\gamma) = \left[\begin{array}{c} E\left(\bar{\varepsilon}_{t+1}(b,\gamma) \times f_t\right) \\ E\left[\varepsilon_{t+1}(b,\gamma) \times \gamma^\top f_t\right] \end{array}\right] = 0.$$

Since the estimate is exactly identified from these moments ($a = I$) Hansen's (1982) Theorem 3.1 gives the standard error,

$$var\left(\begin{array}{c} \hat{\gamma} \\ \hat{b} \end{array}\right) = \frac{1}{T}\tilde{d}^{-1}\tilde{S}\tilde{d}^{-1\top}$$

where

$$\begin{aligned}\tilde{d} = \frac{\partial \tilde{g}_T}{\partial\left[\gamma^\top\ b^\top\right]} &= \frac{\partial}{\partial\left[\gamma^\top\ b^\top\right]}\left[\begin{array}{c} E\left[f_t\left(\overline{rx}_{t+1} - f_t^\top\gamma\right)\right] \\ E\left[\left(rx_{t+1} - b\gamma^\top f_t\right)\left(f_t^\top\gamma\right)\right] \end{array}\right] \\ &= \left[\begin{array}{cc} -E\left(f_t f_t^\top\right) & 0_{6\times 4} \\ E(rx_{t+1}f_t^\top) - 2b\gamma^\top E\left(f_t f_t^\top\right) & -\left[\gamma^\top E\left(f_t f_t^\top\right)\gamma\right] I_4 \end{array}\right].\end{aligned}$$

Since the upper right block is zero, the upper left block of d^{-1} is $E\left(f_t f_t^\top\right)^{-1}$. Therefore, the variance of γ is not affected by the b estimate, and is equal to the usual GMM formula for a regression standard error, $var(\hat{\gamma}) = E(ff^\top)^{-1}\tilde{S}(1:6,1:6)E(ff^\top)^{-1}/T$. The variance of $\hat{b}$ is affected by the generated regressor γ, via the off diagonal term in $\tilde{d}^{-1}$.

To test the (inefficient) two step estimate, we apply Hansen's Lemma 4.1 – the counterpart to the J_T test that handles inefficient as well as efficient estimates. To do this, we must first express the restricted estimate as a GMM estimate based on the unrestricted moment conditions (10). The two step OLS estimate of the restricted model sets to zero a linear combination of the unrestricted moments:

$$a_T g_T = 0, \tag{13}$$

where

$$a_T = \begin{bmatrix} I_6 & I_6 & I_6 & I_6 \\ \gamma^\top & 0_{1\times 6} & 0_{1\times 6} & 0_{1\times 6} \\ 0_{1\times 6} & \gamma^\top & 0_{1\times 6} & 0_{1\times 6} \\ 0_{1\times 6} & 0_{1\times 6} & \gamma^\top & 0_{1\times 6} \end{bmatrix} = \begin{bmatrix} 1_4^\top \otimes I_6 & \\ I_3 \otimes \gamma^\top & 0_{3\times 6} \end{bmatrix}.$$

The first row of identity matrices in a_T sums across return maturities to do the regression of average returns on all forward rates. The last three rows sum across forward rates at a given return maturity to do the regression of each return on $\gamma^\top f$. An additional row of a_T of the form $\begin{bmatrix} 0_{1\times 6} & 0_{1\times 6} & 0_{1\times 6} & \gamma^\top \end{bmatrix}$ to estimate the last element of b would be redundant – the b_4 regression is implied by the first three regressions. The estimate is the same whether one runs that regression or just estimates $b_4 = 1 - b_1 - b_2 - b_3$. We follow the latter convention since the GMM distribution theory is written for full rank a matrices. It is initially troubling to see a parameter in the a matrix. Since we use the OLS γ estimate in the second stage regression, however, we can interpret γ in a_T as its OLS estimate, $\gamma = E_T(ff^\top)^{-1}E_T(\overline{rx}\ f)$. Then a_T is a random matrix that converges to a matrix a as it should in the GMM distribution theory. (I.e. we do not choose the γ in a_T to set $a_T(\gamma)g_T(\gamma, b) = 0$.)

We need the d matrix,

$$d \equiv \frac{\partial g_T}{\partial \begin{bmatrix} b^\top & \gamma^\top \end{bmatrix}}.$$

Recalling $b_4 = 4 - b_1 - b_2 - b_3$, the result is

$$d = \begin{bmatrix} \begin{bmatrix} -1 & & \\ & -1 & \\ & & -1 \\ 1 & 1 & 1 \end{bmatrix} \otimes E(ff^\top)\gamma & -b \otimes E(ff^\top) \end{bmatrix}.$$

Now, we can invoke Hansen's Lemma 4.1, and write the covariance matrix of the moments under the restricted estimate,

$$cov(g_T) = \frac{1}{T}(I - d(ad)^{-1}a)S(I - d(ad)^{-1}a)^\top.$$

The test statistic is

$$g_T^\top cov(g_T)^+ g_T \tilde{}\ \chi^2_{15}.$$

There are $4 \times 6 = 24$ moments and $6(\gamma) + 3(b)$ parameters, so there are 15 degrees of freedom. The $cov(g_T)$ matrix is singular, so the $+$ operator represents pseudoinversion. We use an eigenvalue decomposition for $cov(g_T)$ and then retain only the largest 15 eigenvalues, i.e. write $cov(g_T) = Q\Lambda Q^\top$ where Q is orthogonal and Λ is diagonal and then $cov(g_T)^+ = Q\Lambda^+ Q^\top$ where Λ^+ inverts the 15 largest diagonal elements of Λ and sets the remainder to zero.

To conduct the corresponding Wald test, we first find the GMM covariance matrix of the unrestricted parameters. We form a vector of those parameters, $vec(\beta)$, and then

$$cov(vec(\beta)) = \frac{1}{T} d^{-1} S\,(d)^{-1\top},$$

$$d = \frac{\partial g_T}{\partial (vec(\beta)^\top)} = I_4 \otimes E(ff^\top).$$

C. Recursive identification of the multifactor model

The first row of B is the eigenvector corresponding to the largest eigenvalue of the covariance matrix of expected returns, i.e. with

$$Q\Lambda Q^\top = cov\left[E_t rx_{t+1}, E_t rx_{t+1}^\top\right] = \beta cov(f, f^\top)\beta^\top,$$

we identify

$$B(:,1) = Q(:,1).$$

We identify the rest of B recursively. The zero restrictions are

$$B = \begin{bmatrix} Q(1,1) & 0 & 0 & f \\ Q(2,1) & 0 & c & g \\ Q(3,1) & a & d & h \\ Q(4,1) & b & e & i \end{bmatrix}.$$

We need the coefficients $a - i$. The columns must be orthonormal.

We simplify the algebra by postponing the normalization. We pick b, e, i arbitrarily at 1. Then we pick the other coefficients ones so that the columns are orthogonal. Finally, we rescale the columns numerically, i.e. choosing b, e, i, so that the sum of squares is one. The second column:

$$\begin{aligned} Q(3,1)a + Q(4,1)b &= 0 \\ a &= -\frac{bQ(4,1)}{Q(3,1)}. \end{aligned}$$

The third column:

$$\begin{aligned} Q(2,1)c + Q(3,1)d + Q(4,1)e &= 0 \\ ad + be &= 0 \end{aligned}$$

$$
\begin{aligned}
d &= -\frac{eb}{a} \\
c &= -\frac{Q(3,1)d + Q(4,1)e}{Q(2,1)}.
\end{aligned}
$$

The fourth column:

$$
\begin{aligned}
Q(1,1)f + Q(2,1)g + Q(3,1)h + Q(4,1)i &= 0 \\
ah + bi &= 0 \\
cg + dh + ei &= 0
\end{aligned}
$$

$$
\begin{aligned}
h &= -\frac{bi}{a} \\
c &= -\frac{dh + ei}{g} \\
f &= -\frac{Q(2,1)g + Q(3,1)h + Q(4,1)i}{Q(1,1)}.
\end{aligned}
$$