Bond Risk Premia

John H. Cochrane[*] and Monika Piazzesi[‡]

June 24, 2002

Abstract

This paper studies risk premia in the term structure. We start with regressions of annual holding period returns on forward rates. We find that a single factor, which is a tent-shaped function of forward rates, can predict one-year bond excess returns with an R^2 up to 40%. Though the return forecasting factor has a clear business cycle correlation, it does not forecast output, and business cycle variables do not forecast bond returns. The return forecasting factor does forecast stock returns, about as much as it would a 7 year duration bond. Its forecast power is retained in the presence of the dividend price ratio and the yield spread.

[*] University of Chicago and NBER. Graduate School of Business, University of Chicago, 1101 E. 58th St. Chicago IL 60637. john.cochrane@gsb.uchicago.edu. I gratefully acknowledge research support from the Graduate School of Business and from an NSF grant administered by the NBER.

[†] UCLA and NBER. Anderson School, 110 Westwood Plaza, Los Angeles CA 90095, piazzesi@ucla.edu

[‡] We thank Lars Hansen, Bob Hodrick, Pedro Santa-Clara, Martin Schneider and seminar participants at Arizona, Chicago, Colorado, Duke, Northwestern, the Stanford Institute for Theoretical Economics 2001, and the CIRANO, CIREQ and MITACS Conference on Forecasting for helpful comments.

1 Introduction

This paper studies risk premia in the term structure of interest rates for maturities between one and five years. We start by extending Fama and Bliss' (1987) classic regressions. Fama and Bliss found that the spread between forward rates and one-year rates can predict excess bond returns. Campbell and Shiller (1991) also find that the slope of the term structure forecasts bond returns. We find that one particular linear combination of forward rates predicts excess bond returns even better. It raises the R^2 in excess return forecasting regressions from about 17% to as much as 40%. Furthermore, the same linear combination of forward rates predicts bond returns at all maturities, where Fama and Bliss relate each bond's return to a separate forward-spot spread. In a horse race, our return-forecasting factor completely drives out the separate forward-spot spreads used by Fama and Bliss. It survives a number of subsample and real time robustness checks.

The return-forecasting factor is a tent-shaped linear combination of forward rates. It is not a "curvature factor" in yields, though, as our tent-shaped function of forward rates loads strongly on the 4 and 5 year yields. The return-forecasting factor has a clear business cycle correlation. However, it does not forecast output, and business cycle variables do not forecast bond returns. The return forecasting factor does forecast stock returns, about as much as it would a 7 year duration bond. This forecast power is retained in the presence of the dividend price ratio and the yield spread.

Our specification is similar to the "single index" or "latent variable" models used by Hansen and Hodrick (1983) and Gibbons and Ferson (1985) to capture time-varying expected returns. Stambaugh (1988) found a similar pattern in short maturity bond forecasts, but rejected our single factor specification.

2 Fama-Bliss and beyond

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

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

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

$$f_t^{(n-1 \to n)} = p_t^{(n-1)} - p_t^{(n)}$$

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

$$hpr_{t+1}^{(n)} = p_{t+1}^{(n-1)} - p_t^{(n)}.$$

We summarize the excess return by

$$hprx_{t+1}^{(n)} \equiv hpr_{t+1}^{(n)} - y_t^{(1)}.$$

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

$$\begin{aligned} y_t &= \begin{bmatrix} 1 & y_t^{(1)} & y_t^{(2)} & y_t^{(3)} & y_t^{(4)} & y_t^{(5)} \end{bmatrix}^\top, \\ f_t &= \begin{bmatrix} 1 & y_t^{(1)} & f_t^{(1\to 2)} & f_t^{(2\to 3)} & f_t^{(3\to 4)} & f_t^{(4\to 5)} \end{bmatrix}^\top. \end{aligned}$$

2.2 Fama-Bliss regressions

Fama and Bliss (1987) run a regression of one-year excess returns on long-term bonds against the forward-spot spread for the same maturity. The expectations hypothesis predicts a coefficient of zero – nothing should forecast bond excess returns. Table 1 updates Fama and Bliss' regressions to include more recent data. We see in the one-year return regressions in the left hand panel that the forward-spot spread moves essentially one-for-one with expected excess returns on long term bonds – the expectations hypothesis is exactly wrong at the one year horizon.

Table 1. Fama-Bliss regressions

Maturity		1 year excess returns			Change in $y^{(1)}$		
n		const.	β	R^2	const.	b	R^2
2		0.04	0.94	0.14	−0.04	0.06	0.00
	large T	(0.30)	(0.28)		(0.30)	(0.28)	
	small T	(0.15)	(0.34)	[0.01,0.35]			
	EH			[0, 0.14]			
3		−0.14	1.24	0.14	−0.17	0.32	0.02
	large T	(0.54)	(0.38)		(0.61)	(0.34)	
	small T	(0.31)	(0.44)	[0.01, 0.36]			
	EH			[0, 0.15]			
4		−0.41	1.50	0.15	−0.44	0.56	0.06
	large T	(0.76)	(0.50)		(0.69)	(0.21)	
	small T	(0.45)	(0.51)	[0.01, 0.38]			
	EH			[0, 0.15]			
5		−0.11	1.10	0.06	−0.58	0.76	0.11
	large T	(1.05)	(0.62)		(0.82)	(0.20)	
	small T	(0.59)	(0.70)	[0, 0.28]			
	EH			[0, 0.14]			

NOTE: The “1 year excess return” regression is

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

The “Change in $y^{(1)}$ regression” is

$$y_{t+n-1}^{(1)} - y_t^{(1)} = a + b \left(f_t^{(n-1 \to n)} - y_t^{(1)} \right) + \varepsilon_{t+n-1}^{(n)}.$$

‘Large T’ standard errors (first rows of brackets) use the Hansen-Hodrick GMM correction for overlap. ‘Small T’ standard errors (second rows of brackets) are based on 50,000 bootstrapped samples from a VAR with 12 lags for yields. Intervals below the R^2 indicate 95% confidence intervals computed from these bootstrapped samples. ‘EH’ imposes the expectations hypothesis on the bootstrap. Details are in the Appendix. Data sample 1964:1-2001:12.

Fama and Bliss also run a regression of multi-year changes in the one-year rate against forward-spot spreads. The expectations hypothesis predicts a coefficient of 1.0 – the forward rate should vary one for one with the expected future spot rate. Corresponding to the failure in the left hand panel, the right hand panel of Table 1 shows that the 1-year forward rate (from year one to year two, hence the $n = 2$ row) has essentially no power to forecast changes in the 1-year rate one year from now. However, moving down the rows in the right hand column, longer and longer forward rates correspond more and more to changes in spot rates, so that a 4-year forward rate is within one standard error of moving one-for-one with the expected change in the spot rates. The expectations hypothesis seems to work better over longer horizons. This success for the expectations hypothesis means that the 5-year forward-spot spread does not forecast the *four* year return on 5-year bonds, though it does forecast the *one*-year return on such bonds.[1]

Fama and Bliss’ regressions are driven by robust stylized facts in the data. When forward rates are higher than the 1-year rate, all rates often rise subsequently, as predicted by the expectations hypothesis. However, this rise may take 3 years or more to happen; there can be several years in which the forward rates are above the one-year rate before the interest rate rise takes place. During these years, holders of long-term bonds make money. The period since 1987 is a great out-of-sample success for Fama and Bliss. The regressions have held up well since publication, unlike many other anomalies. In particular, Figure 1 shows that forward-spot forecasts were high in 1990-1993, but interest rates declined, and so long-term bond holders made money. They lost money when interest rates rose in 1994, and the forecast was in the wrong direction in 2000 and 2001, but Fama-Bliss trading rule still made money on average in the post-publication sample.

While Fama-Bliss regressions seem to provide some evidence against the expectations hypothesis, the evidence has been questioned because of poor small sample properties

[1] Here and below, we use Fama and Bliss’ start date of 1964:01, and we do not use the more recently available 1952:6-1963:12 data. A visual inspection of the earlier data suggests a lot more measurement error, which is natural given the thinner selection of bonds and less interest rate movement. The results are quite different for this period – for example, the Fama-Bliss coefficients are all -1 rather than +1.

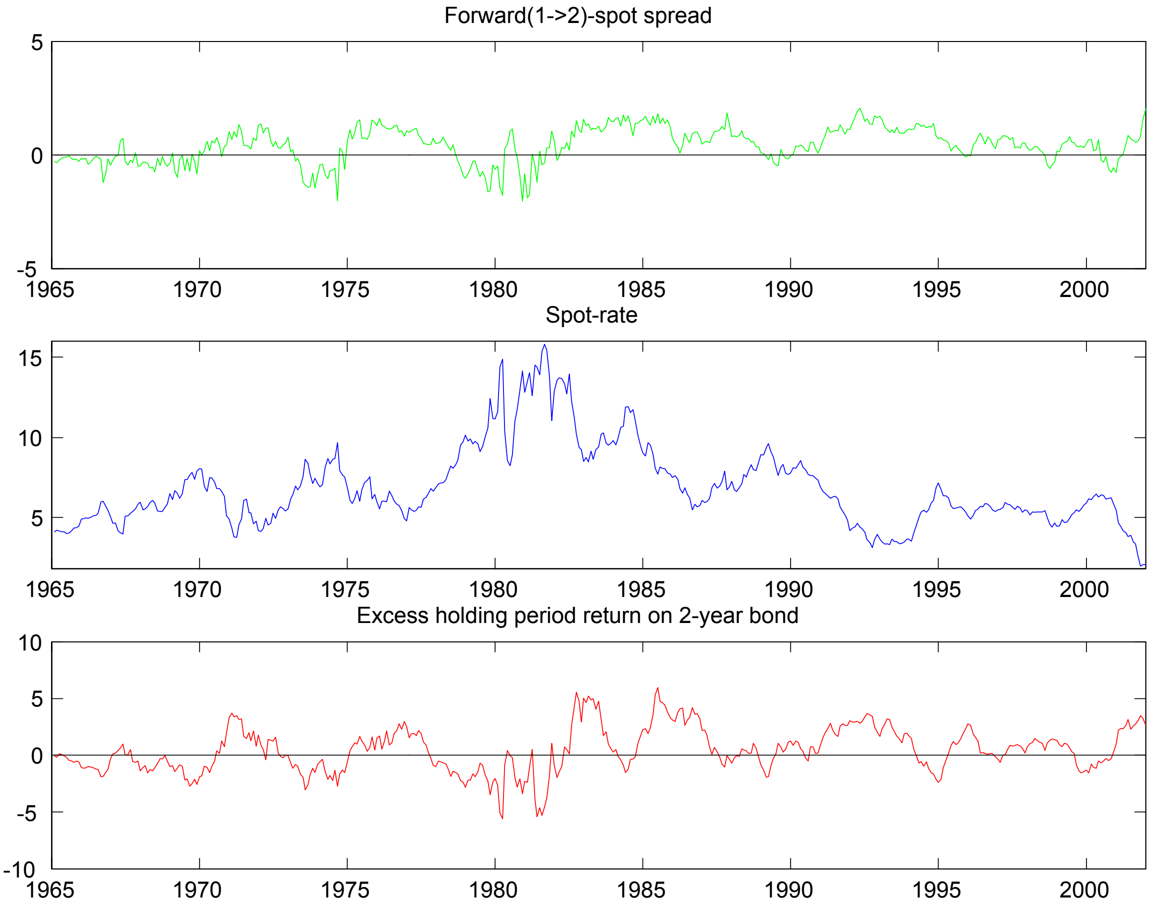

Figure 1:

of standard errors and R^2 (for example, see Bekaert, Marshall and Hodrick (1997)). We therefore compute small sample standard errors and confidence intervals for R^2 using bootstraps. Table 1 shows that small T standard errors are indeed larger than their asymptotic counterparts and that 95% confidence intervals for R^2 contain the estimated R^2 even when we impose the expectations hypothesis on the data-generating process from which the distributions are computed.

2.3 The return-forecasting factor emerges

While Fama and Bliss' specification is the most sensible for exploring the expectations hypothesis and its failures, we are more interested in characterizing expected excess bond returns. For this purpose, there is no reason why only the 4-year forward rate spread should be important for forecasting the expected returns on 4-year bonds. Other spreads may matter. Table 2 follows up on this thought by regressing the one-year return on long-term bonds on *all* of the forward rates separately.

Table 2. Regressions of one year excess returns on all forward rates

n		const.	$y^{(1)}$	$f^{(1\to2)}$	$f^{(2\to3)}$	$f^{(3\to4)}$	$f^{(4\to5)}$	R^2	$\bar{R}^2$	level R^2
2		-1.96	-0.94	0.74	1.15	0.24	-0.91	0.34	0.33	0.38
	large T	(0.64)	(0.18)	(0.43)	(0.30)	(0.27)	(0.18)			
	small T	(0.81)	(0.30)	(0.50)	(0.39)	(0.30)	(0.27)	$[0.20, 0.55]$		
	EH							$[0, 0.18]$		
3		-3.28	-1.66	0.74	2.96	0.29	-1.90	0.34	0.34	0.37
	large T	(1.21)	(0.32)	(0.72)	(0.49)	(0.50)	(0.32)			
	small T	(1.43)	(0.53)	(0.88)	(0.69)	(0.54)	(0.49)	$[0.22, 0.55]$		
	EH							$[0, 0.17]$		
4		-4.57	-2.40	1.11	3.46	1.18	-2.78	0.37	0.36	0.39
	large T	(1.68)	(0.46)	(0.94)	(0.62)	(0.67)	(0.42)			
	small T	(1.91)	(0.71)	(1.18)	(0.92)	(0.72)	(0.67)	$[0.24, 0.58]$		
	EH							$[0, 0.17]$		
5		-5.78	-2.98	1.48	3.93	1.14	-2.88	0.34	0.33	0.36
	large T	(2.13)	(0.58)	(1.12)	(0.73)	(0.80)	(0.53)			
	small T	(2.36)	(0.89)	(1.46)	(1.14)	(0.89)	(0.85)	$[0.21, 0.56]$		
	EH							$[0, 0.17]$		

NOTE: The regression equation is

$$hprx_{t+1}^{(n)} = \beta_0 + \beta_1 y_t^{(1)} + \beta_2 f_t^{(1\to2)} + ... + \beta_5 f_t^{(4\to5)} + \varepsilon_{t+1}^{(n)}$$

'Large T' standard errors (first rows of brackets) use the Hansen-Hodrick GMM correction for overlap. 'Small T' standard errors (second rows of brackets) are based on 50,000 bootstrapped samples from a VAR with 12 lags for yields. Intervals below the R^2 indicate 95% confidence intervals computed from these bootstrapped samples. 'EH' imposes the expectations hypothesis on the bootstrap. Details are in the Appendix. $\bar{R}^2$ 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, $e^{hpr_{t+1}^{(n)}} - e^{y_t^{(1)}}$. Data sample 1964:1-2001:12.

These regressions pick far more than the matched forward-spot spread as the best regressor for holding period returns. For example, the first line of Table 2 suggests that the $f^{(2\to3)} - f^{(4\to5)}$ spread is just as important as Fama and Bliss' variable, the $f^{(1\to2)} - y^{(1)}$ spread, for forecasting the one-year returns of two-year bonds. The top panel of Figure 2 graphs the regression coefficients as a function of the maturity on the right hand side – each row of Table 2 is a solid line of the graph. (For now, ignore the bottom panel and the dashed line in the top 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.

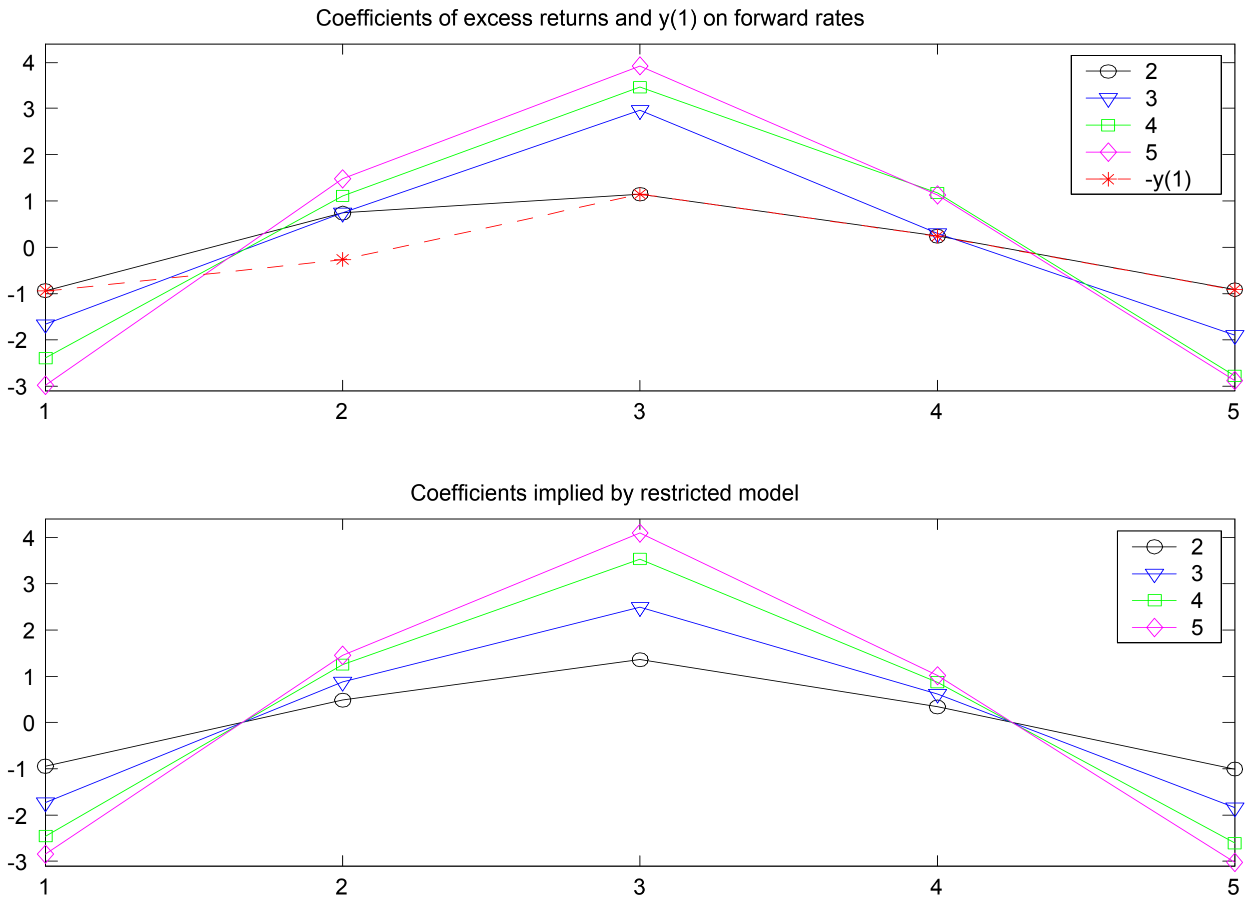

Figure 2: Coefficients in a regression of holding period excess returns on the one-year yield and 4 forward rates. The top panel presents unrestricted estimates from Table 2. The bottom panel presents restricted estimates from a single-factor model reported in Table 4. The legend (2, 3, 4, 5) refers 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. The dashed line in the top panel gives the negative of the regression coefficients of the one year yield on the same right hand variables.

These regressions more than double the R^2 from below 0.15 in Table 1 to 0.33-0.37 across all maturities. The 5-year rate R^2 is particularly dramatic, jumping from 0.06 in Table 1 to 0.34 in Table 2. One might worry that the rise in R^2 comes from the larger number of right hand variables. We report in Table 2 the conventional adjusted $\bar{R}^2$ as well, and it is nearly identical. Of course, that adjustment presumes i.i.d. data which is not valid in this case. We report correct test statistics below for the hypothesis that the extra forecastability is spurious. Moreover, we compute 95% confidence intervals for the R^2. The R^2 from the Fama-Bliss regressions do not lie within these confidence intervals. Also, our regressions make a stronger case against the expectations hypothesis than Fama-Bliss. When confidence intervals for R^2 are computed under the null of the expectations hypothesis, they do not contain the 0.33-0.37 point estimates from Table 2.

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

2.3.1 Short rate forecast

Fama and Bliss also run regressions of changes in short rates on forward-spot spreads. Such regressions are important, since the two ingredients of any term structure model are short rate forecasts plus risk premia. Table 3 presents regressions that forecast the 1-year rate using all the available forward rates.

Again, these results contrast strongly with the updated Fama-Bliss regressions in Table 1. The R^2 in Table 1 was essentially zero using the 2 year forward-spot spread as a right hand variable. (The remaining rows in the right half of Table 1 look at horizons longer than a year as well as using longer maturity forward rates as regressors.) Using all of the forward rates in Table 3, the R^2 jumps to a substantial 22%. Whereas it appeared that the one-year change in the one-year rate was completely unpredictable, it now appears that all the forward rates taken together have substantial power to predict one-year changes in one-year rates.

The coefficient of one-year rate changes on the lagged one-year rate is close to zero. There is a near-unit root in interest rates. Whether one runs the regression in levels or changes makes no difference, of course, except for the interpretation and value of R^2, and by a difference of 1.0 on the coefficient on $y_t^{(1)}$.

Table 3. Predicting the short rate with all forward rates

lhv	const.	$y_t^{(1)}$	$f_t^{(1\to2)}$	$f_t^{(2\to3)}$	$f_t^{(3\to4)}$	$f_t^{(4\to5)}$	R^2
$y_{t+1}^{(1)} - y_t^{(1)}$	1.96	−0.06	0.26	−1.15	−0.24	0.91	0.23
$y_{t+1}^{(1)}$	1.96	0.94	0.26	−1.15	−0.24	0.91	0.61
	(0.64)	(0.18)	(0.43)	(0.30)	(0.27)	(0.18)	

NOTE: The regression equation is

$$lhv_{t+1} = \beta_0 + \beta_1 y_t^{(1)} + \beta_2 f_t^{(1\to2)} + ... + \beta_5 f_t^{(4\to5)} + \varepsilon_{t+1}$$

where lhv is either the level or the change in the one-year rate $y_{t+1}^{(1)}$ as indicated. Standard errors in parentheses use the Hansen-Hodrick GMM correction for overlap and are the same for both regressions. Sample 1964:1-2001:12.

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

The one-year yield regression contains no information that is not already contained in the holding period return regressions. The holding period return of two year bonds, which are sold as one year bonds next year, contains a forecast of next year's one-year rate. Mechanically,

$$hprx_{t+1}^{(2)} = p_{t+1}^{(1)} - p_t^{(2)} - y_t^{(1)} = -y_{t+1}^{(1)} - p_t^{(2)} + p_t^{(1)} = -y_{t+1}^{(1)} + f_t^{(1\to 2)}. \tag{1}$$

Thus, the regression of the one-year yield on our variables should give exactly the negative of the coefficients of the two year holding period return on the same variables, with a 1.0 difference in the coefficient on $f^{(1\to 2)}$. We include in Figure 2 the negative of the one-year yield forecasting coefficients from the second row of Table 3, and you can see this pattern exactly.

More deeply, the identity (1) implies that the forward-spot spread equals the change in yield plus the holding period excess return, and hence, using any set of forecasting variables,

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

(Fama and Bliss use this identity as well.) In Fama and Bliss' regressions, the forward-spot spread corresponds almost one to one to changes in expected returns – both components on the right hand side vary, but they vary in equal amounts, so the one-year rate is unpredictable. Now we have variables that forecast the holding period returns beyond the forward-spot spread. Equation (2) implies that those variables *must* also forecast changes in the spot rate. In this way, the forecastability of the spot rate documented in Table 3 does not mean that the expectations hypothesis is working. It means that the spot rate must be predictable precisely because the expectations hypothesis is not working.

2.4 A single factor for expected bond returns

The pattern of coefficients in Figure 2 cries for us to describe expected excess returns of bonds on all maturities in terms of a single factor, as follows,

$$hprx_{t+1}^{(n)} = b_n\left(\gamma_0 + \gamma_1 y_t^{(1)} + \gamma_2 f_t^{(1\to 2)} + ... + \gamma_5 f_t^{(4\to 5)}\right) + \varepsilon_{t+1}^{(n)}. \tag{3}$$

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,

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

The specification (3) constrains the constants as well as the regression coefficients. We show below that this restriction holds closely as well.

With this normalization, we can fit (3) in two stages. First, we estimate γ by running the regression

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

The second equality introduces the notation γ, f_t for corresponding 6×1 vectors. Then, we can estimate b_n by running the four regressions

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

This procedure is consistent. While one can estimate the parameters with somewhat greater asymptotic efficiency (essentially, using the estimated 24×24 covariance matrix to find a weighted sum in (4)) we prefer the clarity and robustness of the two-stage OLS procedure.

This is a restricted model. We describe the [4 maturities $\times$ (5 right hand variables + 1 intercept)] = 24 unrestricted regression coefficients with (4 bs + 6 γs - 1 normalization) = 9 parameters. The essence of the restriction is that a *single* linear combination of forward rates $\gamma^\top f_t$ is the state variable for time-varying expected returns of *all* maturities.

Table 4 presents the estimated values of γ and b and standard errors. The $\gamma_1 - \gamma_5$ estimates are just about what one would expect from inspection of Figure 2. 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 4 are essentially the same as in Table 2. This fact indicates that the cross-equation restrictions implied by the model (3) – that bonds of each maturity are forecast by the *same* portfolio of forward rates – do no damage to the forecast ability.

Table 4. Estimates of the common return-factor

	γ_0	γ_1	γ_2	γ_3	γ_4	γ_5	R^2
	-3.90	-2.00	1.02	2.87	0.71	-2.12	0.35
large T	(1.41)	(0.38)	(0.79)	(0.53)	(0.56)	(0.36)	
small T	(1.62)	(0.60)	(1.00)	(0.79)	(0.61)	(0.57)	[0.22, 0.56]
EH							[0, 0.17]

		s. e.				
n	b_n	GMM	OLS	small T	R^2	small T
2	0.47	(0.05)	(0.06)	(0.02)	0.32	[0.19, 0.53]
3	0.87	(0.03)	(0.12)	(0.02)	0.34	[0.21, 0.55]
4	1.23	(0.02)	(0.18)	(0.02)	0.37	[0.24, 0.58]
5	1.43	(0.04)	(0.23)	(0.03)	0.34	[0.21, 0.55]

NOTE: The top panel regression is

$$\frac{1}{4}\sum_{n=2}^{5} hprx_{t+1}^{(n)} = \gamma_0 + \gamma_1 y_t^{(1)} + \gamma_2 f_t^{(1\to2)} + ... + \gamma_5 f_t^{(4\to5)} + \bar{\varepsilon}_{t+1}$$

γ_0 has units of annual percent return. 'Large T' standard errors (first rows of brackets) use the Hansen-Hodrick GMM correction for overlap. 'Small T' standard errors (second rows of brackets) are based on 50,000 bootstrapped samples from a VAR with 12 lags for yields. Intervals below the R^2 indicate 95% confidence intervals computed from these bootstrapped samples. 'EH' imposes the expectations hypothesis on the bootstrap. The lower panel reports each excess return's loading on the return-forecasting factor. The regression is

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

γ is the parameter estimate from the upper panel, and f denotes the vector of all forward rates. GMM standard errors 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. OLS standard errors gives conventional standard errors including the Hansen-Hodrick correction for overlap, i.e., treating γ as fixed. 'Small T' standard errors are based on the bootstrapped samples from the upper panel. Sample 1964:1-2001:12.

The standard errors that correct for the fact that γ is a generated regressor are much *smaller* than the "s.e. OLS" conventional (equation-by-equation) standard errors that treat γ as a fixed number. The second set of regressions, each holding period return on the common factor, cannot impose the restriction $1_4^\top b_n = 4$, where 1_4 denotes a 4×1 vector of ones. That restriction is imposed in sample by the first regression. Imposing that restriction in sample removes (places on γ) the largest, common, source of sample variation in b_n. Therefore, the correct standard errors for estimates that impose the restriction $1_4^\top b_n = 4$ in each sample are smaller than the standard errors that would occur if γ were known, in which case the restriction $1_4^\top b_n = 4$ would not hold in each sample. The OLS and GMM standard errors for the first regression are identical.

The bottom panel of Figure 2 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 one factor model almost exactly captures the unrestricted parameter estimates. Figure 3 plots the restricted and unrestricted estimates of the constant, and you can see similarly that the estimates are very close. Standard errors around estimates in Figures 2 and 3 can be obtained from standard errors for β in Table 2 or from those of (γ, b) in Table 4. (More precisely, the standard errors under the null are computed with the delta-method applied to the variance of (γ, b) underlying the standard-error calculations in Table 4.) Confidence intervals constructed as $\beta \pm 1 \times se$ contain both the unconstrained β and the constrained $b\gamma^\top$ estimates.

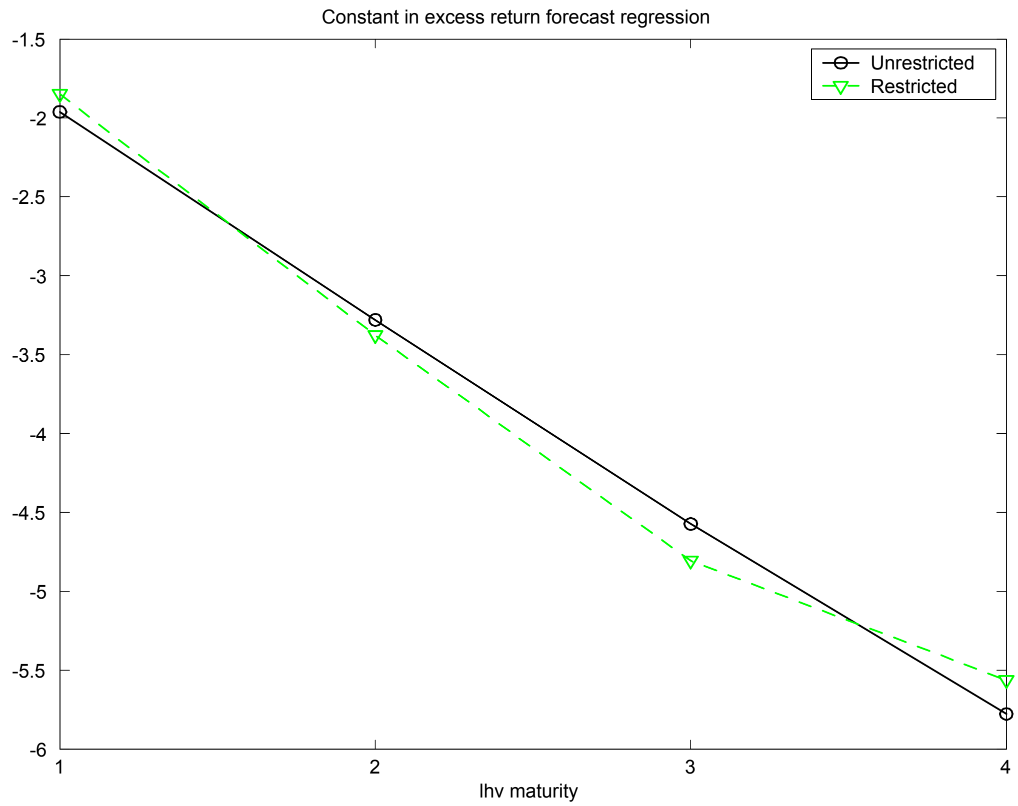

Figure 3: Restricted and unrestricted estimates of the constants. The unrestricted estimate are $\beta_0^{(n)}$ in the regressions $hprx_{t+1}^{(n)} = \beta_0{}^{(n)} + \beta_1 y_t^{(1)} + \ldots \beta_5 f_t^{(4\to 5)} + \varepsilon_{t+1}$. The restricted estimate is $b^{(n)}\gamma_0$ in $hprx_{t+1}^{(n)} = b^{(n)}\left(\gamma_0 + \gamma_1 y_t^{(1)} + \ldots + \gamma_5 f_t^{(4\to 5)}\right) + \varepsilon_{t+1}$.

We need a test of the one-factor model and a test of the constant restrictions. The underlying moments are the regression forecast errors multiplied by forward rates (right hand variables),

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

where ε_{t+1} denotes the 4×1 vector of holding period return regression residuals, and f_t denotes the 6×1 vector of a constant, the one-year yield, and four available forward rates. The unconstrained regression of Table 2 sets all of these moments to zero in each sample.

The single factor model sets only some combinations of these moments to zero

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

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

(We have indicated which parameter is identified by each moment before the colon.) We used the moments (6) and (7) to compute the second set of standard errors in Table 4.

For the restricted model, we can compute the χ^2 test that the remaining moments in (5) are zero, a J_T test. (Details are in the appendix.) We have not yet been able to produce a J_T test because the asymptotic covariance of the same mean of $g\left(\beta f_t\right)$ is not invertible when we compute it using Newey-West or other large-T methods, or when we compute it from the bootstrapped samples. From the confidence bounds implied by Tables 2 and 4 it should be clear, however, that the single factor model is not rejected by the data.

Stambaugh (1988) ran similar regressions of 2-6 month bond excess returns on 1-6 month forward rates. Stambaugh's coefficients are quite similar to the pattern in Figure 2. (See Stambaugh's Figure 2, p. 53.) In the basic regression, Stambaugh found that the matched-maturity forward-spot spread rate – the Fama-Bliss variable – remained the single strongest predictor for excess returns in this multiple regression. However, Stambaugh rightly suspected measurement error – if a bill has a bad price, then the spurious "spread" gives rise to a spurious "return" in the next period. Stambaugh then used a slightly different bill as predictor and predicted variable. This specification resulted in estimates that look a lot like Figure 2. Unlike us, Stambaugh soundly rejected a one or two factor representation of this forecast.

2.5 Checks, extensions and objections

2.5.1 Contest with Fama-Bliss spreads

If $\gamma^\top f_t$ really is the single factor for expected excess returns, it should drive out other forecasting variables, and the Fama-Bliss slope variables in particular. Table 5 presents a multiple regression. In the presence of the Fama-Bliss forward-spot spread, the coefficients and significance of the regression on the return-forecasting factor from Table 4 are unchanged. The R^2 is also unaffected, meaning that the addition of the Fama-Bliss forward-spot 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-spot spread.

Table 5. Horse race between $\gamma^\top f$ and Fama-Bliss

n	a_n	$\sigma\left(a_n\right)$	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 holding period returns on the return-forecasting factor and Fama-Bliss slope. The regression is

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

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

2.5.2 Historical performance

Figure 4 plots the forecast of the holding period excess returns on three year bonds implied by the Fama-Bliss regression of Table 1 (top), the forecast from the regression on the return-forecasting factor from Table 4 (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 the forecast with its outcome. For many episodes, the return-forecasting factor and the forward-spot spread agree. This pattern is particularly visible in the three swings from 1975 to 1982. The return-forecasting factor is correlated with the forward-spot spread. However, the figure shows the much better fit of the return-forecasting factor in the middle. In particular, the fit is much better through the turbulent early 1980s, end 1980s, and the mid 1990's. 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.

2.5.3 Additional Lags

We investigated whether additional lags of forward rates help to forecast bond returns. One additional monthly lag does enter with both statistical and economic significance. Table 6 reports the R^2 of this regression, in the rows labeled "$f_t,\ f_{t-1/12}$." The R^2 rise by about 0.05 to 0.38-0.43. (Adjusted R^2 are about 0.01 lower than unadjusted R^2, though again the conventional R^2 adjustment assumes i.i.d. data.) A χ^2 test overwhelmingly rejects the hypothesis that the coefficients on the additional lag of forward rates are zero. Figure 5 plots the coefficients from these regressions. You can see that the shape of the coefficients is roughly the same at the first and second lag.

Table 6. Additional lags of forward rates

	Maturity n of lhv			
rhv	2	3	4	5
f_t	0.34	0.34	0.37	0.34
$f_t, f_{t-1/12}$	0.39	0.41	0.43	0.41
$\frac{f_t+f_{t-1/12}}{2}$	0.39	0.40	0.43	0.41

NOTE: "f_t" reports the adjusted R^2 from the regression of excess returns on all forward rates in Table 2. "$f_t,\ f_{t-1/12}$" reports the adjusted R^2 from

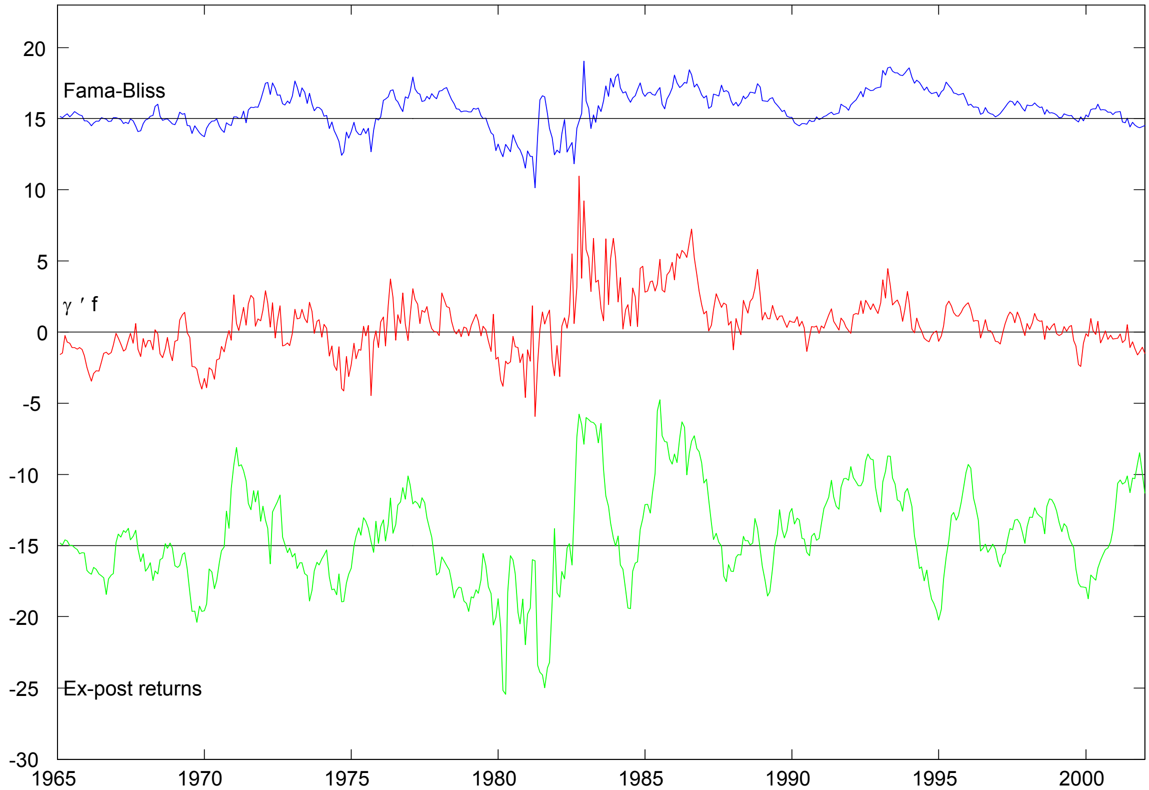

Figure 4: Fitted and actual holding excess returns of three year bonds. Top: Fitted value using Fama-Bliss regression, 3 year forward-spot spread. Middle: Fitted value using the restricted regression on all forward rates. Bottom: ex-post excess returns. The forecasts in the top two lines are graphed 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.

> a regression with an additional monthly lag of all right hand variables in the regression equation, "$(f_t + f_{t-1/12})/2$" reports the adjusted R^2 from a regression using a one-month moving average of right hand variables.

The data seem to want a one-month moving average of forward rates to predict bond returns. We ran a regression with this restricted specification, i.e. $hprx_{t+1}^{(n)}$ on $\left(y_t^{(1)} + y_{t-1/12}^{(1)}\right)/2$, $\left(f_t^{(1\to 2)} + f_{t-1/12}^{(1\to 2)}\right)/2$, etc. Figure 5 includes a plot of the coefficients, and Table 7 includes the R^2 in the row "$\left(f_t + f_{t-1/12}\right)/2$." The R^2 is the same, and the restriction is not rejected statistically, so this seems a good way to include the lagged information.

Following up on these unconstrained regressions, we run bond returns on additional lags of the state variable $\gamma^\top f_t$. Table 7 presents the results. Regression 1 repeats the regression of holding period excess returns on $\left(\gamma^\top f_t\right)$ from Table 4 for comparison. In the

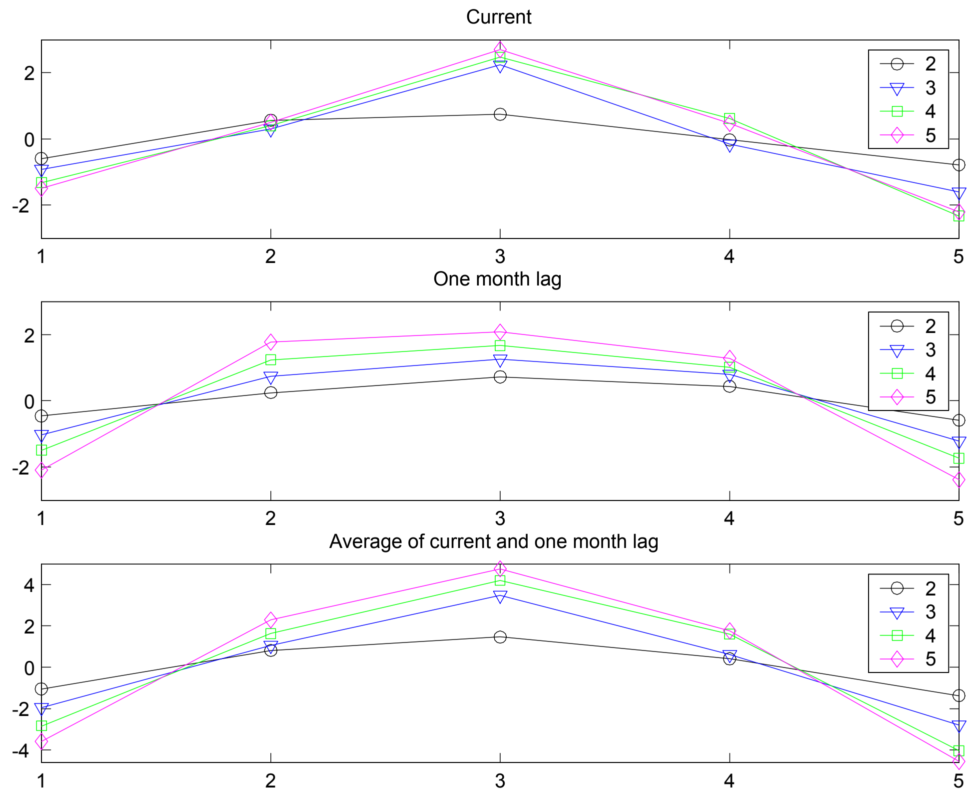

Figure 5: Coefficients in a regression of bond excess returns on the one year yield and 1 to 4 year forward rates, including an extra one-month lag. The top panel plots the coefficients of $hprx_{t+1}^{(n)}$ on forward rates at time t, while the middle panel plots the coefficients on forward rates at time $t-1/12$. The bottom panel presents the coefficients of $hprx_{t+1}^{(n)}$ on a one month moving average of forward rates at t and at $t-1/12$. Sample 1964-2001.

second regression, we add an additional lag $\left(\gamma^{\top} f_{t-1/12}\right)$. The adjusted R^2 now goes up to 0.38-0.41, nearly equal to the 0.38-0.42 values from the unconstrained two-lag regression in Table 4. Once again, the single factor seems to capture all of the information in all 5 forward rates. The coefficients in the second regression are about half of the coefficients in the first regression, and the new coefficients have the same pattern across maturities. The data again suggest $\gamma^{\top}\left(f_t + f_{t-1/12}\right)/2$ as a state variable, and the third regression checks this specification. The additional constraint on the coefficients makes no difference whatever to the R^2, and the coefficients themselves are very close to the value in the first regression. Adding a one-year lag (not reported) or another one-month lag (also not reported) does essentially nothing for the R^2 of the regression.

Table 7. Additional lags of the return-forecasting factor

	(1)		(2)			(3)	
	$\gamma^\top f_t$	R^2	$\gamma^\top f_t$	$\gamma^\top f_{t-1/12}$	R^2	$\frac{\gamma^\top(f_t+f_{t-1/12})}{2}$	R^2
$hprx_{t+1}^{(2)}$	0.46	0.33	0.26	0.27	0.38	0.53	0.38
$hprx_{t+1}^{(3)}$	0.86	0.34	0.50	0.47	0.39	0.97	0.39
$hprx_{t+1}^{(4)}$	1.23	0.37	0.74	0.66	0.41	1.40	0.41
$hprx_{t+1}^{(5)}$	1.45	0.34	0.74	0.94	0.40	1.68	0.40

NOTE: Estimate of each excess return's loading on the return-forecasting factor. The left hand variable is shown in each row heading and the right hand variables are shown in the column headings. γ are the estimates from Table 4. OLS on overlapping monthly data 1964-2001.

Additional lags are uncomfortable forecasting variables for bond yields. Term structure models are usually Markovian: period t's bond yields are sufficient statistics for the evolution of future bond yields. This occurs precisely because bond prices are period t expected values of future discount factors. Building a term structure model around a VAR representation for bond yields with many lags does not seem like a promising way to address the patterns in the data. Instead, the pattern of regression coefficients suggests an ARMA(1,1) model for monthly yields induced by Markovian prices contaminated with i.i.d. measurement error. In the interests of space, we leave the construction of a measurement error model and integration of the forecast patterns we have found here with a Markovian term structure model for future work, and despite the forecast power of the first monthly lag, we focus the rest of the analysis on forecasts using only the most current forward rates.

2.5.4 Subsamples

Table 8 reports a breakdown by subsamples of a regression of average holding period returns $\frac{1}{4}\sum_{n=2}^{5} hprx_{t+1}^{(n)}$ on yields and forwards. The first set of columns run the average return on the yields and forwards separately. The second set of columns runs the average return on $\gamma^\top f$ where γ are 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 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.

The second row shows the effect on the results of the last two disastrous years in the sample, in which $\gamma' f$ and the Fama-Bliss regression both forecast negative expected excess returns, but in fact long term bonds did well. Without these last two years, the R^2 rises to 0.4!

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 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 second set 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 fourth 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 90s, 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 to $\gamma' f$ reveals the terrible performance in these two years – the forecast was small, and the outcome was large.

Table 8. 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 regression

$$\frac{1}{4}\sum_{n=2}^{5} hprx_{t+1}^{(n)} = \gamma^\top f_t + \varepsilon_{t+1}$$

The second set of columns report the coefficient estimate b from

$$\frac{1}{4}\sum_{n=2}^{5} hprx_{t+1}^{(n)} = b\left(\gamma^\top f_t\right) + \varepsilon_{t+1}$$

using the γ parameter from the full sample regression as presented in the top row. Overlapping annual forecasts using monthly data 1964-2001.

2.5.5 Real time forecasts

A trader in, say, 1982, does not have access to our full sample to estimate the parameters of the return forecasting model, so he will not forecast as well. How well can one forecast bonds using real time data? Of course, the conventional rational-expectations answer to this question is that traders have access to historical information and rules of thumb that come from far longer time series than our data set, so their estimates of the coefficients will have converged long before ours. Still, it is an interesting robustness exercise to see how well a trader 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 the full sample estimate was required to see any forecast power.

Figure 6 contrasts the full sample and the real-time forecasts. The top line, marked "full sample" presents the fitted value of the regression $hprx_{t+1} = \gamma^{\top} f_t + \varepsilon_{t+1}$ 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.

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 7 compares them with a simple calculation. The figure calculates "trading rule returns" as

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

and then cumulates these returns so that the different calculations can be more easily compared. (If one follows a linear trading rule to invest $\$1 \times E_t(R_{t+1})$ in each end of a zero - cost portfolio with excess return R_{t+1}, then the profit from this strategy is $R_{t+1} \times E_t(R_{t+1})$. Our regression uses logs rather than levels, hence quotes around "trading rule." The calculation is also T times the covariance of the forecasted variable $hprx_{t+1}$ with the forecast, which is zero if there is no forecastability, so it has a purely statistical interpretation as well.) For the Fama - Bliss calculations, the figure calculates the expected excess return of each bond from its matched forward-spot 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 $E_t(hprx_{t+1})$.

The full sample line 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 real time forecast overall produces only about half of the cumulative profits as

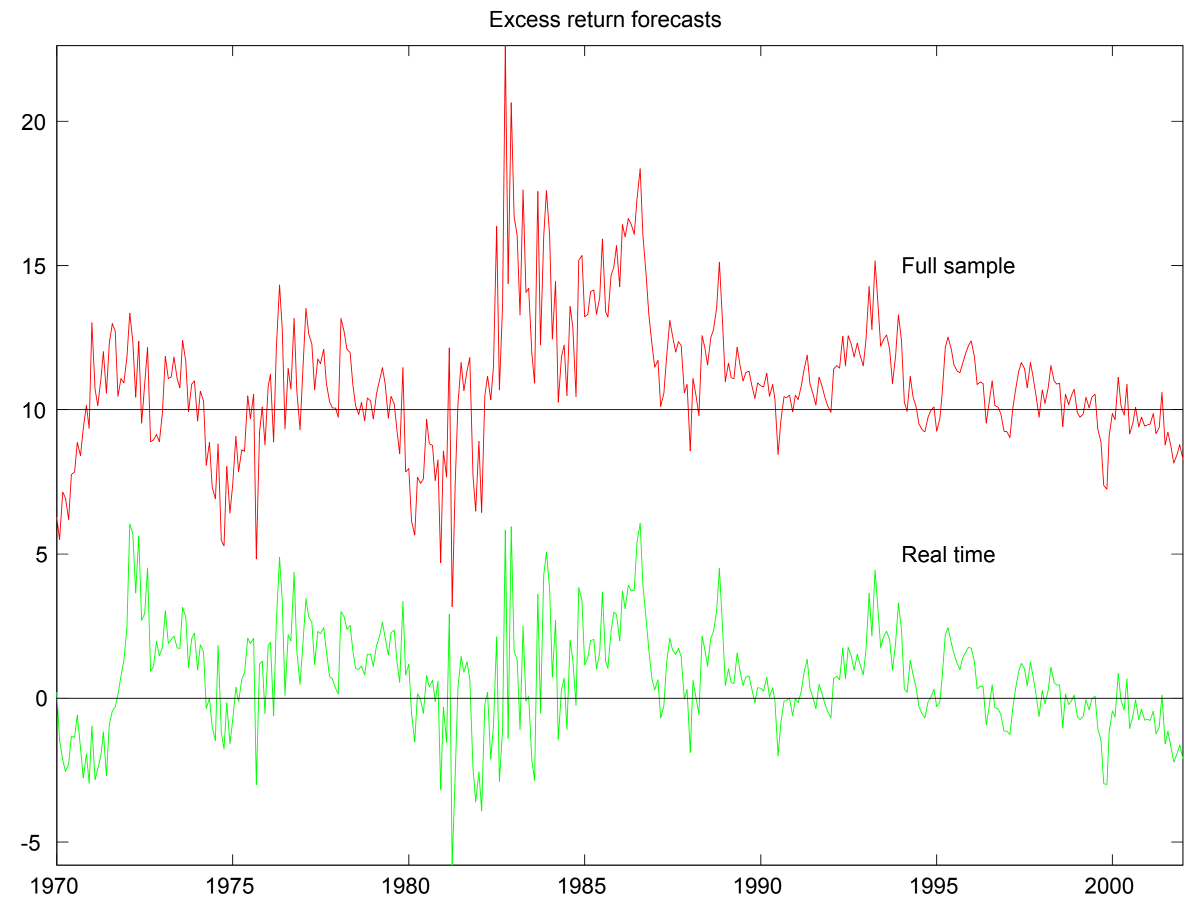

Figure 6: Comparison of full sample and real time forecasts of average (across bond maturities) one year excess returns. "Full sample" is the fitted value of the regression $hprx_{t+1} = \gamma' f_t + \varepsilon_{t+1}$ using 1964:1-2001:12 data to estimate the parameter γ. "Real time" uses data from 1964:1 to time t only to estimate the same regression.

does the full sample estimate. Inspecting the graph, however, 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 6. At this point, the regression has had only 19 years to estimate the 6 γ from collinear 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 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.

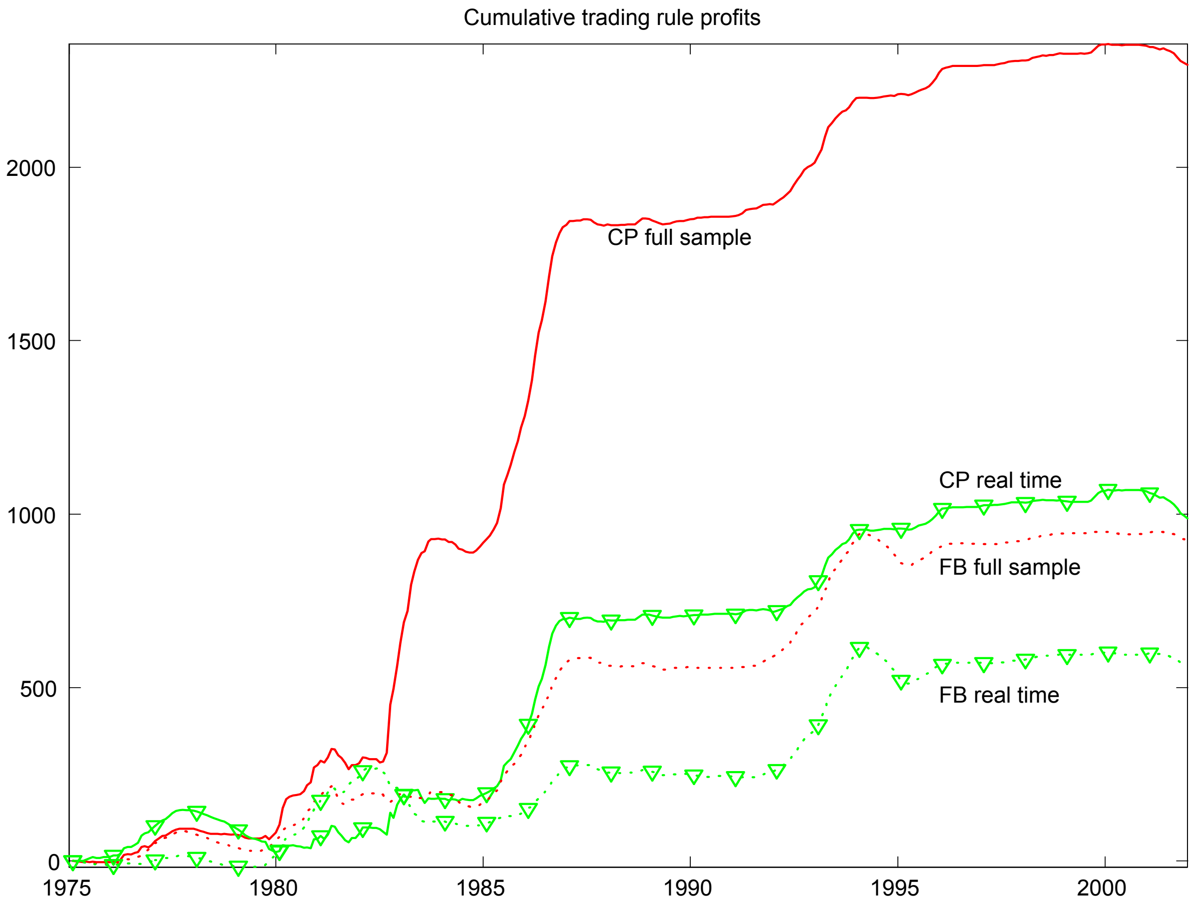

Figure 7: Cumulative profits from linear 'trading rules' using full sample and real time information. Each line plots the cumulative value of $hprx_{t+1} \times E_t(hprx_{t+1})$ where $hprx_{t+1}$ is the average (across maturities) one year log excess return on long term bonds. $E_t(hprx_{t+1})$ are formed from regression forecasts using the full 1964-2001 sample or data from 1964-t as marked. The CP lines use the forecast $hprx_{t+1} = \gamma^\top f_t$ where f_t consists of a constant, the one year rate and 1-5 year forward rates. The FB (Fama-Bliss) lines forecast each holding period excess return from the corresponding maturity forward-spot spread, and then average the forecasts across maturities.

2.5.6 Other data

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 9 shows the R^2s 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. Interestingly, the low R^2 for the excess holding period return of the 5-year bond may be an artifact of the Fama-Bliss data, as the McCulloch-Kwon data gives a 0.12 R^2 for the Fama-Bliss regression where the Fama-Bliss data only gives a 0.05 R^2. Looking at the lower panel of Table 9, the tent-shape of γ estimates is even more pronounced in McCulloch-Kwon data than in the Fama-Bliss data.

Table 9. Comparison with McCulloch-Kwon data

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

	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 used are McCulloch-Kwon and Fama-Bliss zero-coupon yields starting 1964:1 until the end of the McCulloch-Kwon dataset, 1991:12. McCulloch-Kwon data can be downloaded from http://www.econ.ohio-state.edu/jhm/ts/mcckwon/mccull.htm. The upper panel shows R^2 from running $hprx_{t+1}^{(n)}$ on the regressors indicated on top of the table: forward-spot spread, all forwards f_t, and $\gamma^\top f_t$. The lower panel shows the estimated γs using the two datasets.

2.5.7 Interpretation

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 one recovers from yields, say by an eigenvalue decomposition of their conditional or unconditional covariance matrices. However, our tent-shaped function is a function of *forward rates.* Forward rates and yields span the same bond prices of course, so we can express the forecasting factor $\gamma^\top f$ as a function of yields – the forecasts are exactly the same. The top of Figure 8 plots the results – equivalent to a regression of holding period excess returns on all yields rather than all forward rates. The bottom of Figure 8 plots the loadings of the first 3 principal components of yields. The third principal component is 'curvature'. The return-forecasting factor is clearly not a 'curvature' factor in yields. The correlation coefficient between changes in $\gamma^\top f$ and curvature is only 39.9%. The return forecasting factor is also not a 'level' or 'slope' factor. The corresponding correlation coefficients are 7.4% and 49.5%. At best, one can interpret it as long the 3-2 spread and short the 5-4 spread, but the interpretation in terms of forward rates seems cleaner.

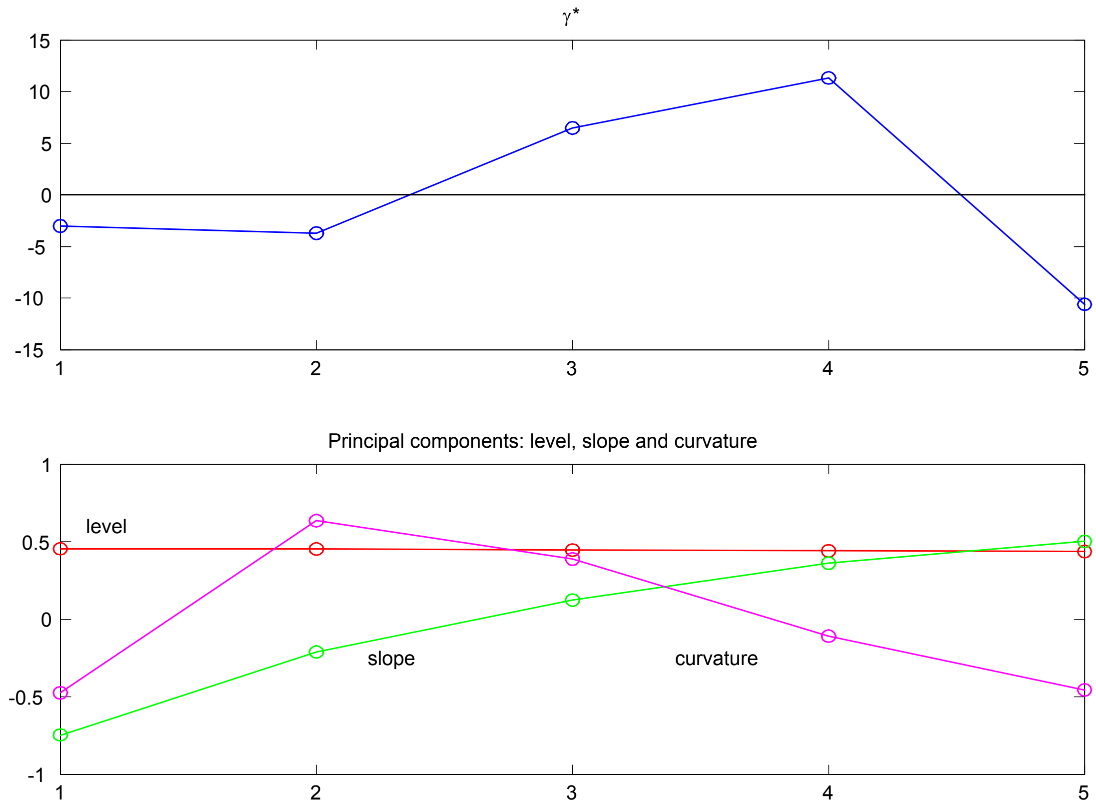

Figure 8: The top graph shows coefficients γ^* in a regression of average (across maturities) holding period returns on all yields, $\frac{1}{4}\sum_{n=2}^{5} hprx_{t+1}^{(n)} = \gamma_0^* + \gamma_1^* y_t^{(1)} + \gamma_2^* y_t^{(2)} + \ldots + \gamma_5^* y_t^{(5)}$. The bottom graph shows the loadings of the first three principal components of yields labelled 'level factor', 'slope factor' and 'curvature factor'. Sample 1964:1-2001:12.

3 Macroeconomics and bond return forecasts

Figure 4 already shows that the return-forecasting factor is highly correlated with the slope of the term structure, which is well known to be associated with recessions (Fama and French 1989) and to forecast output growth (Harvey 1989, Stock and Watson 1989, Estrella and Hardouvelis 1991, Hamilton and Kim 1999).

We discover a surprising difference between the return forecasting factor and the term structure slope. The return forecasting factor, like the slope, is highly correlated with business cycle measures. However, the forecasting relations are lost. Business cycle measures have no power alone, and even less in competition with the return forecasting factor, to forecast bond returns. Worse, the return forecasting factor loses the slope's ability to forecast output. Apparently, the component of the slope of the term structure that forecasts excess returns has nothing to do with the component that forecasts output.

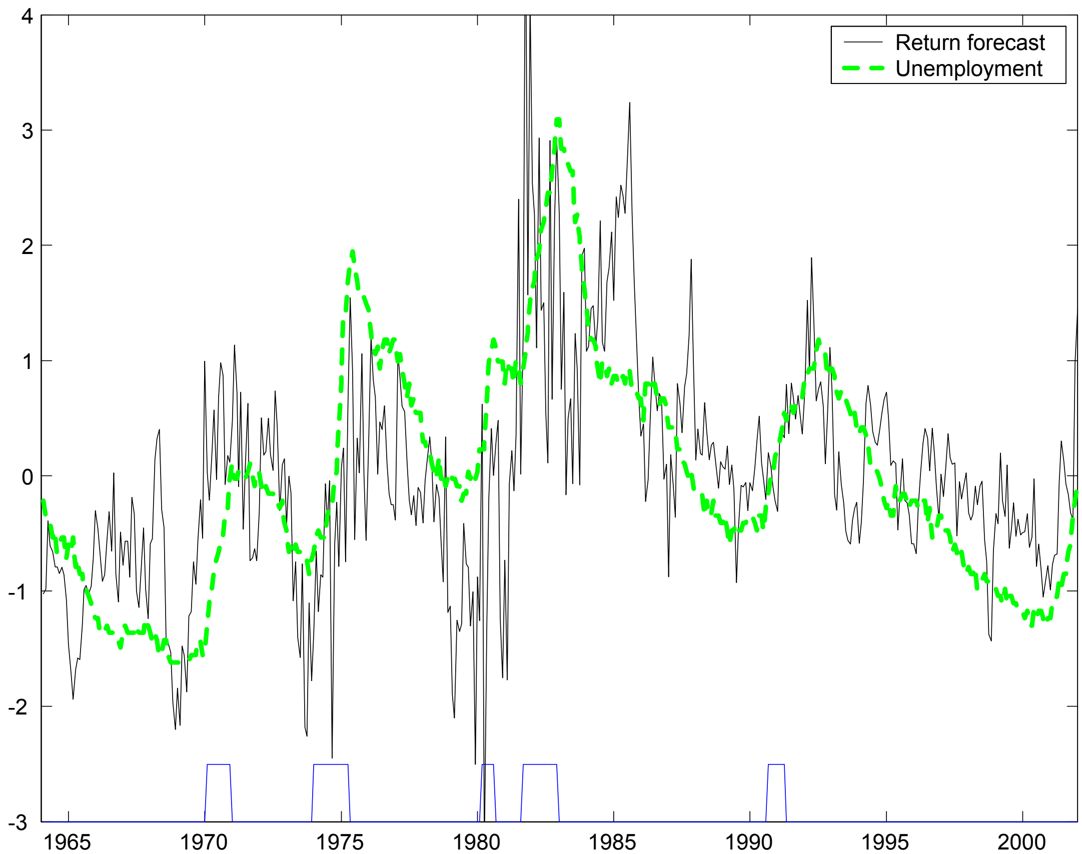

Figure 9: Return forecasting factor $\gamma' f_t$ and unemployment rate. Both series are transformed to $[x_t - E(x)]/\sigma(x)$ so that they fit on the same graph. The teeth at the bottom represent NBER business cycles.

3.1 Correlation between the return forecast and business cycles

Figure 9 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, high in bad times and low in good times. The graph shows the very nice correlation between the return forecasting factor and recessions. As Fama and French (1989) document for the yield curve slope, the time-varying expected return is clearly related to business cycles.

Interestingly, 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 does many measures of a decade long drop in productivity during that period. The bond return forecasting factor is a "level" variable rather than a "growth rate" 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*. 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.

3.2 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 excess returns on bonds. The answer is no, or at least "not among the variables we have tried so far."

This is an unfortunate result for economic interpretation. It would be much nicer if we could understand the return forecasting factor as a simple mirror of macroeconomic conditions. It appears instead that the bond market uses additional information beyond that available in macroeconomic aggregates to forecast bond returns. On the other hand, it is a fortunate result for our empirical analysis: it means we can stick to the model $E_t\left(hprx_{t+1}\right) = \gamma^\top f_t$ with great accuracy, even in VAR systems that include macroeconomic variables.

Table 10 contrasts regressions of the average one year bond excess return $\frac{1}{4}\sum_{n=2}^{5} hprx_{t+1}^{(n)}$ on the return forecasting factor $\gamma^\top f$, on the unemployment rate U and other macroeconomic variables. The first part of the table reminds us of the 0.35 and 0.40 R^2 when we forecast bond excess returns from $\gamma^\top f$. Despite its beautiful correlation with the return forecasting factor, unemployment forecasts bond excess returns with an R^2 of only 0.05. In a multiple regression it does not affect the size and significance of the $\gamma^\top f$ coefficient, and only raises the R^2 to 0.38.

The Stock-Watson (1989) leading index 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, which forecasts income growth and stock returns, does no better. Finally, CPI inflation is just as useless as the variables. A large variety of macroeconomic variables do no better.

Table 10. Macro forecasts of bond returns

$\gamma^\top f$	$\gamma^\top f_{-1}$	R^2	$\gamma^\top f$	U	R^2
1		0.35		0.54	0.05
(7.2)				(1.5)	
0.56	0.59	0.40	1.19	-0.50	0.38
(6.0)	(5.3)		(7.6)	(-1.6)	

$\gamma^\top f$	XLI	R^2	$\gamma^\top f$	cay	R^2	$\gamma^\top f$	cpi	R^2
	-0.11	0.01		0.44	0.02		-0.24	0.03
	(-0.6)			(1.03)			(-0.84)	
1.01	-0.14	0.36	1.01	-0.08	0.35	0.99	-0.20	0.37
(6.8)	(-1.2)		(7.8)	(-0.2)		(7.6)	(-1.0)	

Table 10. Forecasts of average bond returns $\frac{1}{4}\sum_{n=2}^{5} hprx_{t+1}^{(n)}$. U_t = the unemployment rate. XLI = Stock-Watson leading indicator. cay = the Lettau-Ludvigson consumption-wealth ratio using end of period wealth. cpi is inflation, the one-year growth in the CPI index. We estimate $\gamma^\top f$ by running the regression $\frac{1}{4}\sum_{n=2}^{5} hprx_{t+1}^{(n)} = \gamma^\top f_t + \varepsilon_{t+1}$ in a first stage. Overlapping annual forecasts, 1964:01-2001:12. Standard errors corrected for overlap and heteroskedasticity by GMM.

3.3 Term structure forecasts of output growth

The slope of the term structure slope forecasts output growth as well as bond returns. How does the return forecasting factor $\gamma^\top f$ forecast output growth? Table 11 presents regressions. The left hand panel forecasts industrial production, while the right hand panel forecasts growth in Stock and Watson's coincident index. The table verifies that the term structure slope $y^{(5)} - y^{(1)}$ forecasts both output growth measures, with statistical significance and R^2 of 0.14. The Stock-Watson leading index, which includes term structure variables as well as a variety of other macroeconomic variables, does even better, with stunning t statistics and R^2 of 0.35-0.41.

Surprisingly, though, the return forecasting factor is a miserable failure at forecasting output growth. The coefficients are tiny and insignificant, the R^2 almost vanish. The return factor is correlated with the yield spread, and the return factor forecasts bond returns much better, but it nonetheless loses any ability to forecast output growth. Apparently, the component of the yield spread that forecasts output growth is uncorrelated with the component that forecasts bond excess returns.

Table 11. Term structure forecasts of output growth

industrial production				coincident index			
$\gamma^\top f$	$y^{(5)} - y^{(1)}$	LI	R^2	$\gamma^\top f$	$y^{(5)} - y^{(1)}$	LI	R^2
0.11			0.01	0.09			0.004
(0.34)				(0.41)			
	-0.76		0.14		-0.54		0.14
	(-2.66)				(-2.3)		
		0.86	0.35			0.66	0.41
		(9.4)				(10.4)	
0.38	-0.64	0.70	0.41	0.27	-0.43	0.55	0.46
(1.9)	(-2.84)	(7.2)		(2.4)	(-2.8)	(6.9)	

Table 11. Regression forecasts of one-year industrial production growth and one-year growth in the Stock-Watson coincident index on the bond return forecasting factor $\gamma^\top f$, the term spread $y^{(5)} - y^{(1)}$, and the Stock-Watson leading index. Overlapping annual forecasts, 1964:01-2001:12. Standard errors corrected by GMM.

3.4 Forecasting stock returns

The slope of the term structure forecasts stock returns, as emphasized by Fama and French (1989). Table 12 evaluates how well our return forecasting factor forecasts stock returns.

The first 4 regressions remind us of 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%.

Table 12. Forecasts of excess stock returns

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

Table 12. Stock return forecasts. The left hand variable is the one-year

return on the value-weighted NYSE stock return, less the one 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 in parentheses. Standard errors are corrected for overlap.

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 4 had a coefficient of 1.43 on the return forecasting factor. Thus, the stock return coefficient is just about what would expect of a 6 or 7 year duration bond, which is perfectly sensible.

The sixth and seventh regressions compare the bond return forecasting factor with the term spread and d/p. The bond return 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 insignificant. It seems that the bond return forecasting factor subsumes most of the term spread and d/p's power to forecast stock returns.

Last, we ask whether a regression of stock returns on all forward rates produces a better fit than on the return forecasting factor, and whether such a regression recovers the tent-shaped pattern of coefficients all on its own. 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%. Figure 10 graphs the coefficients, along with the return forecasting factor coefficients γ, and two standard error bands. The stock return forecasting coefficients have the same general tent shape, though not exactly the same as those of the return forecasting factor. The 2-1 forward spread seems to enter more than it does for the return forecasting factor.

4 Conclusions

One-year expected excess returns in the Fama-Bliss (1987) data follow a one-factor structure almost exactly. The single factor is a tent-shaped function of forward rates, $\gamma^\top f_t$. Then, expected excess returns on bonds of maturity n are $E_t(hprx_{t+1}^{(n)}) = b_n(\gamma^\top f_t)$.

Regressions of excess returns on this common factor show a much improved R^2. In contrast to Fama and Bliss' R^2 of about 15%, the R^2 on the common factor is about 35%, and as much as 40% if we use a one-month moving average of the common factor $\gamma'(f_t + f_{t-1})$ to attenuate measurement error. (Ignoring the last two years these numbers jump to 40-45%).

The single factor $\gamma^\top f$ drives out the separate forward-spot spreads in predicting excess bond returns. The forecast works well across subsamples since 1964. It is somewhat

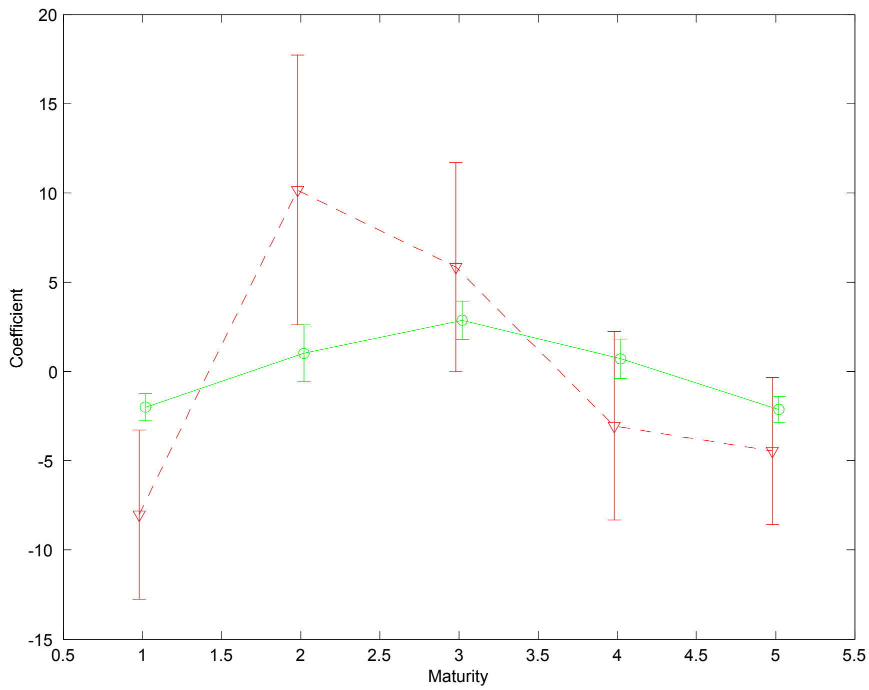

Figure 10: Coefficients in a regression of one-year value weighted NYSE stock excess returns on all forward rates (dashed line, triangles) and average bond excess returns on all forward rates (solid line, circles). Error bars are +/- two standard errors.

stronger in the latter part of the sample in which real interest rate movements dominate the term structure, than in the earlier part of the sample in which much interest rate movement reflects expected inflation.

The return forecasting factor $\gamma^\top f$ has a strong contemporaneous correlation with business cycle measures, especially the unemployment rate. However, macro variables do not forecast bond returns, either alone or in competition with the return forecasting factor. Curiously, the return forecasting factor does not forecast output, unlike the slope of the term structure. Apparently, the part of the slope that does forecast output is the part that does not forecast bond returns. The bond return forecasting factor does forecast stock returns, with a coefficient about what one would expect of a 7 year duration bond. Its forecast power is maintained in competition with a term spread and dividend price ratio.

References

[1] Bekaert, Geert, David Marshall, and Robert J. Hodrick, 1997, "On Biases in Tests of the Expectations Hypothesis of the Term Structure of Interest Rates," *Journal of Financial Economics*, 44, 309-348.

[2] 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.

[3] Fama, Eugene F. and Robert R. Bliss, 1987, "The Information in Long-Maturity Forward Rates," *American Economic Review* 77, 680-692.

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

[5] Ferson, Wayne, and Campbell R. Harvey, 1999, "Conditioning Variables and Cross-section of Stock Returns", *Journal of Finance* 54, 1325-1360

[6] Estrella, Arturo, and Gikas A. Hardouvelis, 1991, "The Term Structure as Predictor of Real Economic Activity," *Journal of Finance* 46, 1991, 555-576.

[7] Ferson, Wayne, and Michael R. Gibbons, 1985, "Testing Asset Pricing Models with Changing Expectations and an Unobservable Market Portfolio," *Journal of Financial Economics* 14, 216-236

[8] 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.*

[9] 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.

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

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

[12] Lettau, Martin, and Sydney Ludvigson, 2001 "Consumption, Aggregate Wealth and Expected Stock Returns," *Journal of Finance* LVI 815-849.

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

[14] 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.

5 Appendix

5.1 Small sample distributions

Data-generating process

The data-generating process is taken to be 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 3.

Imposing the expectations hypothesis

The data-generating process in this case is taken to be an AR(12) for the short rate

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

and long yields are computed as

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

5.2 GMM approach to the factor model for expected bond returns

Restricted and unrestricted model

The unrestricted regression is

$$hprx_{t+1} = \beta f_t + \varepsilon_{t+1}$$

where $hprx$ is the 4×1 vector of excess returns, $f_t = \begin{bmatrix} 1 & y_t^{(1)} & f_t^{(1\to2)} & \ldots & f_t^{(4\to5)} \end{bmatrix}'$ is the 6×1 vector of forward rates and β is a 4×6 matrix of unrestricted regression coefficients. The moment conditions of the unrestricted model are

$$g_T(\beta) = E(\varepsilon_{t+1} \otimes f_t) = 0. \tag{8}$$

The restricted model is $\beta = b\gamma^\top$ where b is a 4×1 vector and γ is a 1×6 vector of coefficients. Since b and γ are only separately identified up to a constant (double b, halve γ), we normalize to $b^\top 1_4 = 4$.

Efficient estimate and test

The efficient estimate of the restricted model is

$$J_T = \min_{\{\gamma,\beta:1_4'b=4\}} g_T(b\gamma^\top)' S^{-1} g_T(b\gamma^\top)$$

where S is the spectral density matrix corresponding to the moments (8). The standard errors of the efficient estimate are given by Hansen's (1982) theorem 3.1

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

where $d \equiv \partial g_T / \partial \left[b^\top \; \gamma^\top\right]$ is calculated below. We can test the restricted model with the usual J_T test statistic

$$T J_T \tilde{} \chi^2(15).$$

There are 15 degrees of freedom. There are $6 \times 4 = 24$ moments and $6(\gamma) + 4(b) - 1(\text{normalization}) = 9$ restrictions. Any inefficient estimate can only produce a larger value for the objective function and therefore for the J_T statistic, which means that we tend to overreject the null.

Inefficient 2-step estimate

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

$$\begin{aligned}
\frac{1}{4}\sum_{n=2}^{5} hprx_{t+1}^{(n)} &= \gamma^\top f_t + \bar{\varepsilon}_{t+1} \\
hprx_{t+1} &= b\left(\hat{\gamma}^\top f_t\right) + \varepsilon_{t+1}
\end{aligned}$$

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

We focus on the 2-step procedure in this case, as OLS is often better in small samples than GLS. We experienced a lot of difficulty with the 24×24 spectral density matrix, using 24 lags, and formed from the highly cross-correlated regression errors ε multiplied by the even more cross-correlated and highly serially correlated forward rates. It is often singular or nearly so, so using it to weight estimates is likely not to work well in small sample.

Standard errors for 2-step estimate and another test

To provide standard errors for the two-step estimate, we append the average excess return equation to the unrestricted moments

$$\widetilde{g}_T = \left(\begin{array}{c} E\left(\varepsilon_{t+1} \otimes f_t\right) \\ E\left(\overline{\varepsilon}_{t+1} \times f_t\right)\end{array}\right),$$

which gives us $4{\times}6 + 6 = 30$ equations. The two-step OLS regressions set some linear combinations of these moments equal to zero, so that we are left with only 10 equations for b and γ :

$$a_T \widetilde{g}_T = 0, \; a_T = \left(\begin{array}{cc} I_4 \otimes \gamma^\top & 0_{4\times 6} \\ 0_{6\times 24} & I_6 \end{array} \right).$$

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, $\gamma = E_T(ff^\top)^{-1}E_T(\overline{hprx}\; f)$. 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 $a_T \widetilde{g}_T\left(\gamma, b\right) = 0$.)

The GMM formula for the standard error of the estimates now is

$$\frac{1}{T}\left(a_T d\right)^{-1} a_T S a_T^\top \left(a_T d\right)^{-1\top},$$

with

$$d = -\left[\begin{array}{cc} I_4 \otimes E\left(f_t\left(\gamma^\top f_t\right)\right) & b \otimes E\left(f_t f_t^\top\right) \\ 0_{6\times 4} & E\left(f_t f_t^\top\right) \end{array} \right].$$

An alternative test of overidentifying restrictions can therefore be constructed from the covariance matrix of the sample moments

$$cov(g_t) = \left(I - d(ad)^{-1}a\right) S \left(1 - d\left(ad\right)^{-1} a\right)^\top.$$

The χ^2-distributed test statistic is

$$T\widetilde{J}_T = T\widetilde{g}_T^\top cov\left(\widetilde{g}_T\right)^{+} \widetilde{g}_T,$$

where $+$ refers to a pseudoinverse, since the covariance matrix is singular. The degrees of freedom is the rank of $cov\left(\widetilde{g}_T\right).$