Bond Risk Premia

John H. Cochrane*and Monika Piazzesi‡

November 6, 2001

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 45%.

Though the return forecasting factor has a clear business cycle correlation, it does not forecast output, and macro 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.

We relate these time-varying expected returns to covariances with various shocks, which is the same as finding the market prices of risk that justify a yield VAR as an affine term structure model. The time-varying expected return can be entirely accounted for by a time-varying risk premium for level-shocks in yields, and almost entirely accounted for by a time-varying risk premium for monetary policy shocks.

How could such an important factor have been missed? The return forecasting factor does very little for understanding yields. Conventional two or three factor models provide an excellent approximation for yields, but a poor approximation for expected returns. Also, bond yields do not follow a monthly AR(1), with a pattern that suggests measurement error. Hence, if you follow the usual approach in term structure analysis, starting with a monthly k-factor model chosen to minimize pricing errors, and then finding implied annual return forecasts, you completely miss the forecastability of annual returns.

*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 and seminar participants at the Stanford Institute for Theoretical Economics 2001 for helpful comments.

1 Introduction

This paper studies risk premia in the term structure of interest rates. 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 about 45%. 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. This finding paves the way for a simple representation of the term structure of interest rates, in which we can use a small number of linear combinations of yields as state variables, rather than requiring each yield or forward rate as a separate state variable, in order to forecast that maturity's return. Our return-forecasting factor is a tent-shaped linear combination of forward rates. In a horse race, our return-forecasting factor completely drives out the separate forward-spot spreads used by Fama and Bliss.

Expected returns should be related to covariances multiplied by risk premia. We find that covariance with a "level" shock to yields, multiplied by a time-varying risk premium proportional to the return-forecasting factor, describes the time-variation in expected bond returns. We also study shocks to expected bond returns, inflation shocks and monetary policy shocks, in an attempt to find a fundamental shock to explain the time-variation in returns. We find that shocks to expected returns and inflation do not help, but a time-varying premium for exposure to monetary policy shocks can explain the bulk of the time-varying expected bond return.

These results correspond with intuition. The unconditional mean excess return rises with maturity: long bonds return a bit more, on average, than short bonds. The much larger time-varying component of expected returns also varies systematically with maturity. Long bonds' expected excess returns load more heavily on our return-forecasting factor than do short bonds' expected excess returns. If we wish to explain these facts with a factor risk premium (the same for all bonds) multiplied by a covariance of each bond with a shock, then we must find a shock that affects all bond returns in the same direction, and affects long bond returns more than short bond returns. That is exactly the feature of a level shock; if the yield curve shifts up, bonds of successively longer maturity (duration) are successively more affected.

The monetary policy shock works precisely because it produces a similar "level" effect in bond returns. A monetary policy shock moves all yields up or down together; it thus produces a larger change in long-term bond returns than short-term bond returns. This larger covariance with long-term bond returns can, when multiplied by a factor risk premium, produce a larger expected return for long term bond returns. (*Why* monetary policy shocks move long term bonds so much is an interesting puzzle, but one we do not address here.)

The interesting part of this result is the strong time-variation in bond expected excess returns, and hence in the factor risk premium. Rises in yields mechanically produce declines in returns, so the sign of the covariance between returns and the shock cannot change. Hence, given that expected excess returns are sometimes positive and sometimes negative, the factor risk premium must change sign. Bonds whose prices will decline when there is a monetary policy shock sometimes earn a positive expected return, and thus have lower prices than predicted by expectations logic; those same bonds, subject to the same price decline when there is a monetary policy shock, will at other times earn a negative expected return, and thus have higher prices than predicted by expectations logic. It all depends on the state of the economy, as reflected in our tent-shaped linear combination of forward rates.

Our time-varying risk premium specification results in an affine term structure model based on a VAR representation for bond yields. Thus, we have constructed an affine model that can completely capture the predictability of bond returns, and it exactly reproduces the prices of bonds (or their linear combinations) used as state variables. Whether or not bond return predictability is consistent with affine models has been a contentious point in the literature. (See Fisher 1998, Duarte 1999, Duffie 2001 and Dai and Singleton 2001.) We show that, almost trivially, one can construct an affine model to mirror complex patterns of bond return predictability including those we find in the data.

Why have extensive investigations of the term structure of interest rates missed this return forecasting factor? Most term structure analyses are performed by first fitting an approximate $k-$factor model in high frequency data, and then (if at all) looking at implied one-year forecastability. Two interesting features of the data imply that this procedure will miss return forecastability.

First, a monthly autoregression raised to the 12th power completely misses the forecastability of returns at the one year horizon. The monthly bond data do not follow an AR(1). Monthly yields are closely approximated by an ARMA(1,1), which suggest an underlying AR(1) plus i.i.d. measurement error. Though the deviation from an AR(1) is small, the 12th power magnifies small misspecifications.

Second, the return forecasting factor is almost completely unimportant for describing prices or yields. "Level" and "slope" factors unrelated to the return forecasting factor can fit yields with a high degree of accuracy. (Analyses that include maturities less than a year often find a third "curvature" factor, in order to reconcile maturities longer than a year with those less than a year. This factor does not have the same shape as our return-forecasting factor, whose weights are symmetric around the 3 year forward rate.) For this reason, the return-forecasting factor is not recovered by traditional factor analysis or maximum likelihood estimation of term structure models. You have to look at excess return forecasts to see it.

Reduced factor representations are still interesting of course. We find that a three factor representation, using a 'level' and a 'slope' factor, deriving ultimately from the VAR shock covariance matrix, and a return factor deriving from the expected return regressions, does an excellent job of representing all this information in the term structure.

Our investigation is a little unusual in that we examine conditionally homoskedastic discrete time models, rather than continuous time models with heteroskedastic shocks as is common in the term structure literature. Our specification is closer 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. This fact has an important implication: though many affine models use conditionally heteroskedastic shocks to produce curved patterns in the term structure and time-varying expected returns, one does not *have* to use heteroskedastic shocks to obtain these results. For the expected return - beta questions we address in this paper, conditional homoskedasticity is not likely to have a major effect on the results. The covariances of bond returns with yield curve shocks are really driven by the arithmetic of duration, and do not change sign. Thus, we will have to understand expected returns that change sign with a risk premium that changes sign, rather than a covariance that changes sign multiplied by a constant risk premium. Thus, while time-varying covariances with yield shocks (really, time-varying variances of the yield shocks) can be part of the story, they must be a secondary part, and time-varying risk premia must be the most important part of the story. Of course, since $cov(r, \eta) [\lambda_0(1 + \lambda_1 x_t)] = cov [r, \eta(1 + \lambda_1 x_t)] \lambda_0$, where r = return, η = shock, x = state variable, and λ_0, λ_1 = parameters, much of what one can express with constant covariances and a time varying risk premium can be expressed as a changing covariance and a constant risk premium, with different shocks. Whether modeling conditional heteroskedasticity is really important, in the end, will have to be judged by constructing such a model and seeing whether it gives an importantly different characterization of expected returns.

2 Fama-Bliss and beyond

2.1 Notation

We use the following notation. Denote the log price of a n year discount bond at time t by $p_t^{(n)}$

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

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, e.g.

$$\begin{aligned}
y_t &= \left[\begin{array}{ccccc} y_t^{(1)} & y_t^{(2)} & y_t^{(3)} & y_t^{(4)} & y_t^{(5)} \end{array} \right]' \\
hprx_{t+1} &= \left[\begin{array}{cccc} hprx_{t+1}^{(2)} & hprx_{t+1}^{(3)} & hprx_{t+1}^{(4)} & hprx_{t+1}^{(5)} \end{array} \right]' \\
f_t &= \left[\begin{array}{ccccc} y_t^{(1)} & f_t^{(1\to 2)} & f_t^{(2\to 3)} & f_t^{(3\to 4)} & f_t^{(4\to 5)} \end{array} \right]'
\end{aligned}$$

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. The first panel of Table 1 updates Fama and Bliss' regressions to include more recent data. We see in the one-year return regression 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.

	1 year returns			Change in $y^{(1)}$		
maturity	b	$\sigma(b)$	R^2	b	$\sigma(b)$	R^2
2	1.02	0.27	0.17	-0.02	0.27	0.001
3	1.33	0.36	0.17	0.39	0.34	0.020
4	1.61	0.48	0.18	0.55	0.21	0.063
5	1.18	0.62	0.07	0.75	0.21	0.11

Table 1. Fama-Bliss regressions The "1 year returns" regression is

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

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

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

Standard errors use the Hansen-Hodrick GMM correction for overlap. Sample 1964:1-1999:12.

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 be equal to the expected future spot rate (plus a Jensen's inequality term). 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 one-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. 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 one-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, forward-spot spreads 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, but Fama-Bliss trading rule still made money on average in the post-publication sample.

[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. Also, the results are quite different for this period – for example, the Fama-Bliss coefficients are all -1 rather than +1 – so at a minimum one needs to think of a time-varying model to include the period.

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.

	Coefficients						R^2			
n	a	$y^{(1)}$	$f^{(1\to 2)}$	$f^{(2\to 3)}$	$f^{(3\to 4)}$	$f^{(4\to 5)}$	f_t	$f_t,\ f_{t-1/12}$	$\frac{f_t+f_{t-1/12}}{2}$	level
2	-2.24	-0.97	0.71	1.13	0.30	-0.88	0.38	0.44	0.44	0.39
3	-3.77	-1.72	0.67	2.97	0.41	-1.86	0.39	0.46	0.45	0.40
4	-5.20	-2.46	0.99	3.49	1.33	-2.75	0.41	0.48	0.48	0.43
5	-6.54	-3.04	1.30	4.00	1.33	-2.85	0.38	0.46	0.45	0.39
	Standard errors									
2	0.59	0.19	0.47	0.29	0.27	0.18				
3	1.11	0.35	0.79	0.46	0.50	0.31				
4	1.57	0.50	1.03	0.58	0.66	0.41				
5	2.01	0.62	1.24	0.67	0.79	0.51				

Table 2. Regression of one year holding period returns on forward rates, 1964:01-1999:12. The regression equation is

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

Standard errors correct for overlapping data. In the R^2 column, " f_t" reports R^2 from this regression. "$f_t,\ f_{t-\frac{1}{12}}$" reports R^2 from a regression with an additional monthly lag of all right hand variables. "$(f_t+f_{t-1/12})/2$" reports R^2 from a regression using a one-month moving average of right hand variables. "level" reports the R^2 from a regression using the level, not log, excess return on the left hand side, $e^{hpr_{t+1}^{(n)}} - y_t^{(1)}$.

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 $f^{(2\to 3)} - f^{(4\to 5)}$ spread is just as important as Fama and Bliss' variable, the $f^{(1\to 2)} - y^{(1)}$ spread, for forecasting the one-year returns of two-year bonds. These regressions more than double the R^2 from below 0.18 in Table 1 to 0.38-0.41 across all maturities. The 5 year rate R^2 is particularly dramatic, jumping from 0.07 in Table 1 to 0.38 in Table 2.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 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*

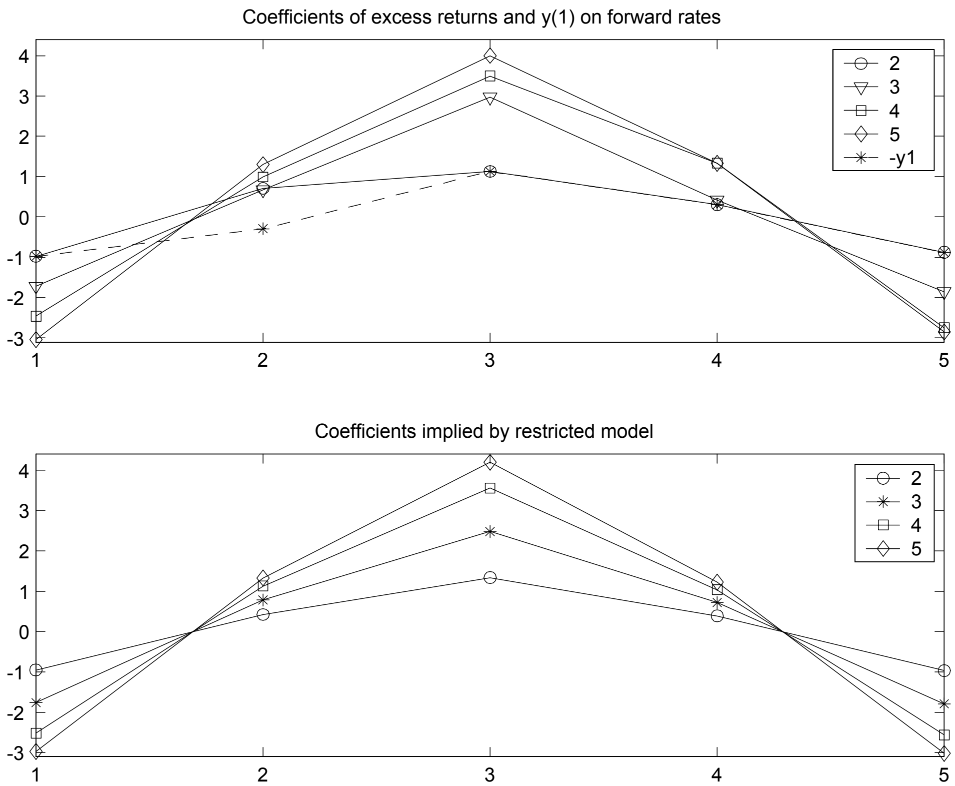

Figure 1: Coefficients in a regression of holding period excess returns on the one-year yield and 4 forward rates. The top panel presents unrestricted estimates from Table 2. The bottom panel presents restricted estimates from a single-factor model, from Tables 4 and 5. The legend (2, 3, 4, 5) refers 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.

maturities just have greater loadings on this same function. The pattern of coefficients suggests a tent-shaped factor.

We can, of course, run excess returns on bond yields rather than forward rates. The fitted values of the regression are exactly the same, since forward rates are linear functions of yields. The pattern of regression coefficients is less pretty.

One might worry about logs vs. levels; that actual excess returns are not forecastable, but that the coefficients in Table 2 only reflect $1/2\sigma^2$ terms and conditional heteroskedasticity.[2] We repeated the regressions using actual excess returns, $e^{hpr_{t+1}^{(n)}} - y_t^{(1)}$ 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]We thank Ron Gallant for raising this important question.

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 one-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 0.001%, 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 26%. 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 still 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 $y_t^{(1)}$.

lhv	$y_t^{(1)}$	$f_t^{(1\to2)}$	$f_t^{(2\to3)}$	$f_t^{(3\to4)}$	$f_t^{(4\to5)}$	R^2
$y_{t+1}^{(1)} - y_t^{(1)}$	-0.026	0.29	-1.13	-0.30	0.88	0.26
$y_{t+1}^{(1)}$	0.974	0.29	-1.13	-0.30	0.88	0.63
Standard errors (same for both regressions):						
	0.19	0.47	0.29	0.27	0.18	

Table 3. Regression of one year yields on forward rates, 1964:01-1999:12. The regression equation is

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

where lhv is either the level or the change in the one-year rate $y_{t+1}^{(1)}$ as indicated. Standard errors correct for overlapping data.

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,

$$hprx_{t+1}^{(2)} = p_{t+1}^{(1)} - p_t^{(2)} - y_t^{(1)} = -y_{t+1}^{(1)} - p_t^{(2)} + p_t^{(1)} = -y_{t+1}^{(1)} + f_t^{(1\to2)}. \tag{1}$$

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 $f^{(1\to2)}$. We include in Figure 1 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,

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

(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. (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.

2.3.2 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. In Table 2, we report the R^2 of this regression, in the column labeled "$f_t,\ f_{t-1/12}$." The R^2 rise by about 0.05 to 0.44-0.48. Rather than add them to a table, Figure 2 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. 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 $\left(y_t^{(1)}+y_{t-1/12}^{(1)}\right)/2$, $\left(f_t^{(1\rightarrow 2)}+f_{t-1/12}^{(1\rightarrow 2)}\right)/2$, etc. Figure 2 includes a plot of the coefficients, and Table 2 includes the R^2 in the column "$\left(f_t+f_{t-1/12}\right)/2$." The R^2 is lowered by no more than 0.01 by this additional restriction, and it is not rejected statistically, so this seems the best way to include the lagged information.

This finding suggests measurement error in the forward rates, so that the "true" forward rate is better recovered by the moving average. Additional monthly lags or a one year lag add little to the regression.

Despite the small increase in forecast power available from an additional lag, we focus our attention on specifications that use only the current variables f_t, as this drastically simplifies the analysis. Then we return to a treatment of the extra lags, while reconciling these annual horizon regressions with a monthly VAR representation for bond yields.

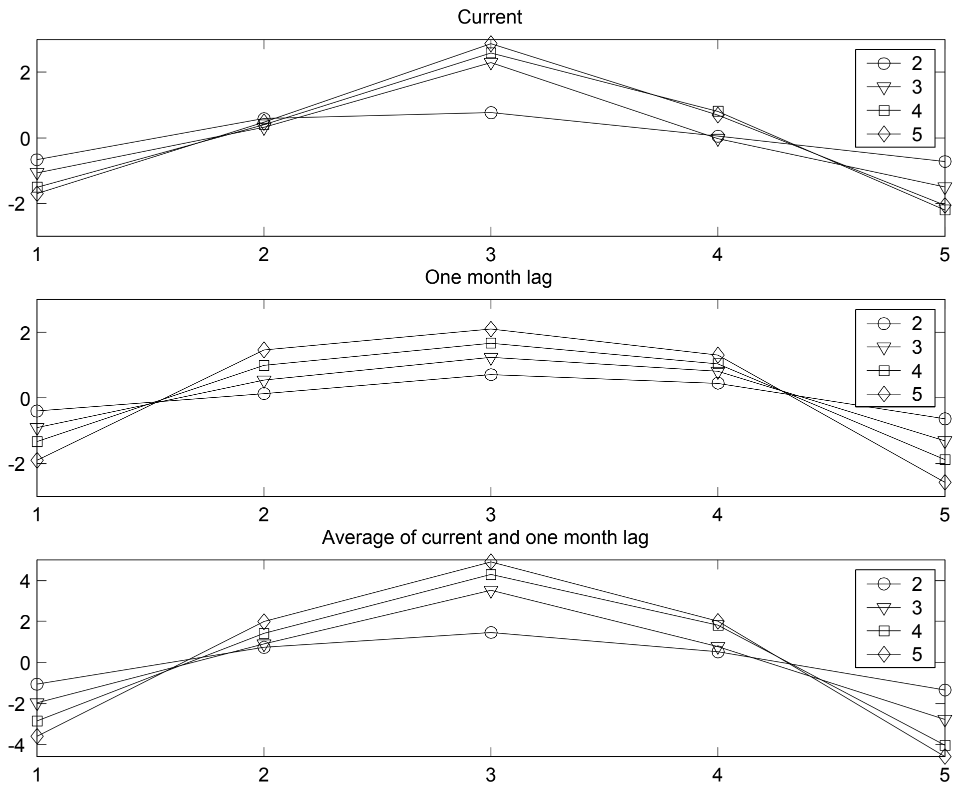

Figure 2: Coefficients in a regression of bond excess returns on the one year yield and 1 to 4 year forward rates, including an extra one-month lag. The top panel plots the coefficients of $hprx_{t+12}^{(n)}$ on forward rates at time t, while the middle panel plots the coefficients on forward rates at 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 1964-1999.

2.4 A single factor for bond expected returns

The 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.

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

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, and the average value of a_n are zero

$$\frac{1}{4}\sum_{n=2}^{5} b_n = 1;\ \sum_{n=2}^{5} a_n = 0$$

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

$$\frac{1}{4}\sum_{n=2}^{5} hprx_{t+1}^{(n)} = \gamma_0 + \gamma_1 y_t^{(1)} + \gamma_2 f_t^{(1\to 2)} + ... + \gamma_5 f_t^{(4\to 5)} + \bar{\varepsilon}_{t+1} \quad (4)$$
$$= \gamma_0 + \gamma' f_t + \bar{\varepsilon}_{t+1}.$$

The second equality introduces the notation γ, f_t for corresponding 4×1 vectors. Then, we can estimate the a_n, b_n by running the four regressions

$$hprx_{t+1}^{(n)} = a_n + b_n(\gamma_0 + \gamma' f_t) + \varepsilon_{t+1}^{(n)},\ n = 2,\ 3,\ 4,\ 5.$$

We use GMM standard errors to correct for the fact that $\gamma' f_t$ is a generated regressor, along with serial correlation due to overlap. We consider the additional restriction $a_n = 0$ that the intercepts as well as the slope coefficients follow the single-factor model. This procedure is consistent. While one can estimate the parameters with somewhat greater asymptotic efficiency (essentially, using the estimated 30×30 covariance matrix to find a weighted sum in (4)) we prefer the clarity of the two-stage OLS procedure.

This is a restricted model. We describe the (4 maturities $\times$ (5 right hand variables + 5 intercepts) = 25 unrestricted regression coefficients with (4 as + 4 bs + 6 γs - 2 normalizations) = 12 parameters, or, if $a_n = 0$ with 9 parameters. The essence of the restriction is that a *single* linear combination of forward rates $\gamma_0 + \gamma' f_t$ is the state variable for time-varying expected returns of *all* maturities.

Tables 4 and 5 presents the estimated values of γ, a and b.

	γ_0	γ_1	γ_2	γ_3	γ_4	γ_5	R^2
Estimate	-4.44	-2.05	0.91	2.90	0.84	-2.08	0.40
Std. error	1.31	0.41	0.88	0.50	0.55	0.35	

Table 4. Estimated common factor in bond expected returns. The regression is

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

γ_0 has units of annual percent return

The $\gamma_1 - \gamma_5$ estimates in Table 4 are just about what one would expect from inspection of Figure 1. The loadings b_n of expected returns on the common factor presented in Table 5 increase smoothly with maturity. The R^2 in Table 5 are the same as in Table 3 to two significant digits. 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.

Maturity n	a_n	s.e.	$a_n + \frac{1}{2}\sigma_n^2$	b_n	s.e.	s.e. OLS	R^2
2	0.11	0.62	0.13	0.47	0.02	0.05	0.37
3	0.09	1.15	0.13	0.86	0.02	0.11	0.39
4	0.01	1.60	0.08	1.23	0.01	0.15	0.41
5	-0.21	1.90	-0.10	1.43	0.03	0.20	0.38

Table 5. Estimate of each excess return's loading on the return-forecasting factor. The regression is

$$hprx_{t+1}^{(n)} = a_n + b_n \left(\gamma_0 + \gamma' f_t\right) + \varepsilon_{t+1}^{(n)}$$

where $hprx$ denotes bond return less one year rate, γ are the estimates from Table 4, and f denotes the vector of all forward rates. a_n has units of percent annual return. The "s.e." columns are GMM standard errors. They correct for the fact that γ is estimated, by considering this estimate together with the regression of Table 4 as a single GMM estimation. The "s.e. OLS" column gives conventional standard errors including the Hansen-Hodrick correction for overlap.

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 $\sum_n b_n = 4$. 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 $\sum_n b_n = 4$ in each sample are smaller than the standard errors that would occur if γ were known, in which case the restriction $\sum_n b_n = 4$ would not hold in each sample.

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 $\left[\begin{array}{cccc} b_n\gamma_1 & b_n\gamma_2 & b_n\gamma_3 & b_n\gamma_4 \end{array}\right]$. 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.

Table 5 suggests that we eliminate the constants in the individual regressions as well, i.e. that the intercept in each bond return regression is well modeled as the slope coefficient times the intercept in the average return regression $b_n\gamma_0$. This leaves a truly one-factor model,

$$hprx_{t+1}^{(n)} = b_n \left(\gamma_0 + \gamma' f_t\right) + \varepsilon_{t+1}^{(n)}.$$

The coefficients a_n in Table 5 are tiny. They are an order of magnitude below the standard errors, so that they are individually significant. Figure 3 plots the intercepts from the unrestricted regressions and their standard error bars along with the intercepts $b_n\gamma_0$ from the restricted regression and you can see the excellent fit.

Following this hunch, we repeated the two step estimation of Table 4 and Table 5 with $a_n = 0$. The γ estimates and standard errors are of course exactly the same, since

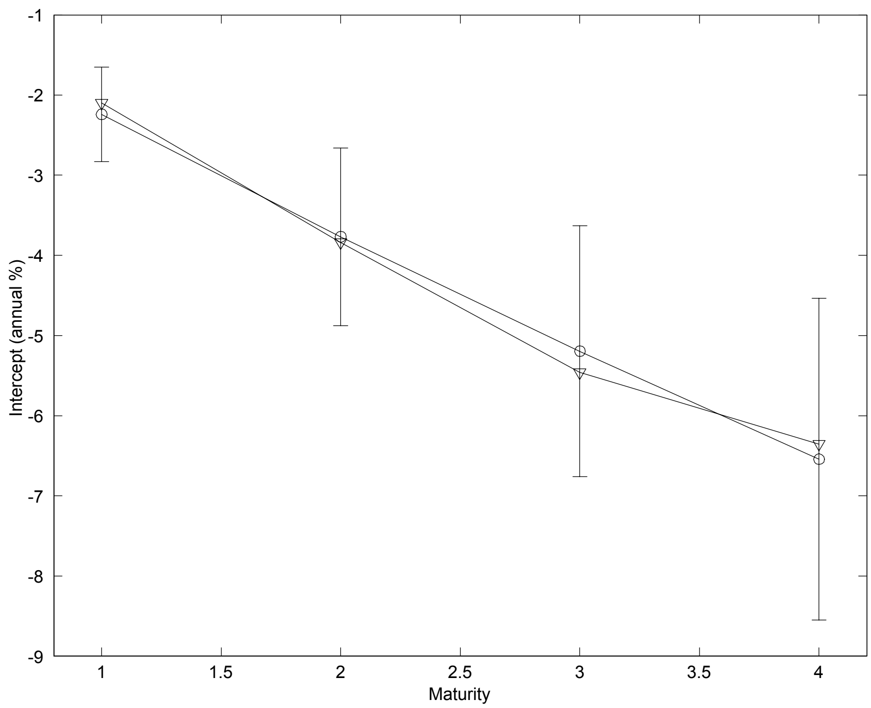

Figure 3: Restricted and unrestricted intercepts. The unrestricted intercepts are from the regressions $hprx_{t+1}^{(n)} = a_n + \beta'_n f_t + \varepsilon_{t+1}^{(n)}$. The restricted intercepts are $b_n\gamma_0$ from the regressions $hprx_{t+1}^{(n)} = b_n(\gamma_0 + \gamma' f_t) + \varepsilon_{t+1}^{(n)}$, where γ_0 and γ are estimated from $\frac{1}{4}\sum_{n=2}^{5} hprx_{t+1}^{(n)} = \gamma_0 + \gamma' f_t + \varepsilon_{t+1}$. Error bars are +/- 2 standard errors from the unrestricted regression.

they are estimated in the first step. The b_n coefficients, standard errors, and R^2 are the same to the decimals indicated in Table 4 and 5. However, overidentifying restrictions tests presented below reject this specification, so we keep the intercepts a_n.

If this 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 6 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 5 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.

maturity n	b_n	$\sigma(b_n)$	c_n	$\sigma(c_n)$	R^2
2	0.47	0.03	-0.04	0.19	0.37
3	0.87	0.09	-0.06	0.35	0.39
4	1.21	0.15	0.05	0.44	0.41
5	1.42	0.15	0.14	0.32	0.38

Table 6. Multiple regression of holding period returns on the return-forecasting factor and Fama-Bliss slope. The regression is

$$hprx_{t+1}^{(n)} = a_n + b_n\left(\gamma_0 + \gamma' f_t\right) + c_n\left(f_t^{(n-1\to n)} - y_t^{(1)}\right) + \varepsilon_{t+1}^{(n)}.$$

Standard errors corrected by GMM for overlap.

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 3 (middle, i.e. $a_3 + b_3\left(\gamma_0 + \gamma' f_t\right)$) and the actual holding period returns (bottom). For many episodes, you can see that 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, you can also see the much better fit of the regression using the return-forecasting factor in the middle. In particular, the fit is much better through the turbulent early 1980s and the mid 1990's. The improved R^2 is not driven by spurious forecasting of one or two unusual data points.

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 1. (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 1. Stambaugh soundly rejected a one or two factor representation of this forecast.

2.4.1 Tests

This section is very preliminary – this is the method, but we don't trust the numbers

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),

$$E\left(\left[\begin{array}{c} \varepsilon_{t+1} \\ \varepsilon_{t+1} \otimes f_t \end{array}\right]\right) = 0 \tag{5}$$

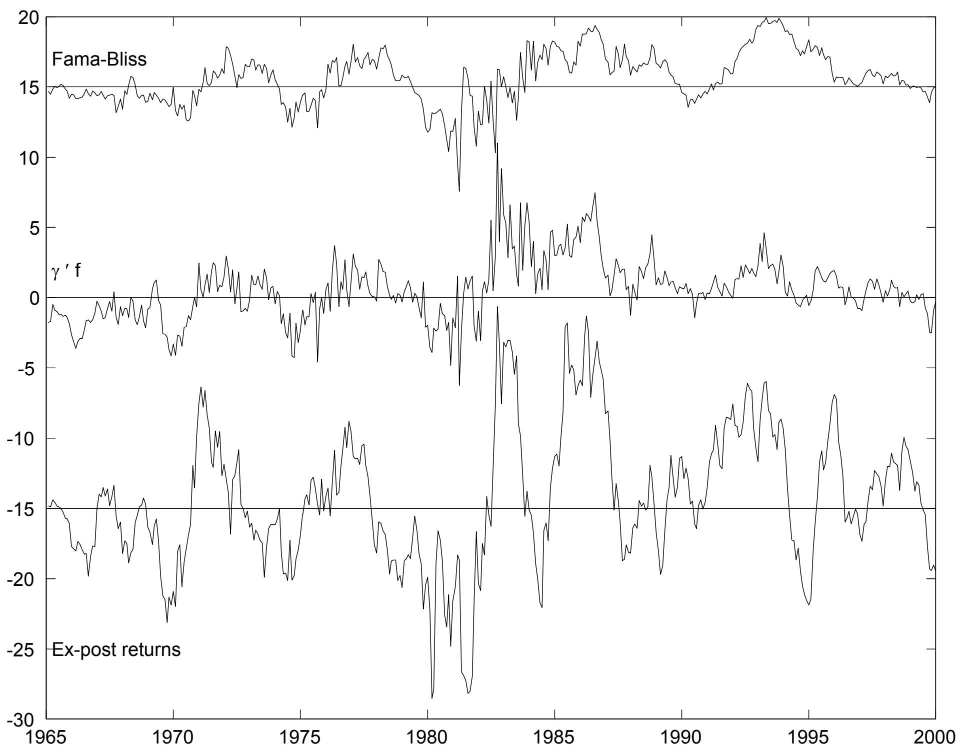

Figure 4: Fitted and actual holding excess returns of three year bonds. Top: Fitted value using Fama-Bliss regression, 3 year forward-spot spread. Middle: Fitted value using the restricted regression on all forward rates. Bottom: ex-post excess returns. The forecasts in the top two lines are graphed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

where ε_{t+1} denotes the 4×1 vector of holding period return regression residuals, and f_t denotes the 5×1 vector of 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 with constants ($a_n \neq 0$) sets only certain linear combinations of these moments to zero

$$\begin{aligned} \gamma_0 &: \; E\left[1_4'\varepsilon_{t+1}\right] = 0 && (6) \\ \gamma &: \; E\left[(1_4'\varepsilon_{t+1}) \otimes f_t\right] = 0 && \\ a &: \; E\left[\varepsilon_{t+1}\right] = 0 && (7) \\ b &: \; E\left[\varepsilon_{t+1} \otimes (\gamma' f_t)\right] = 0 && (8) \end{aligned}$$

where 1_4 denotes a 4×1 vector of ones.(We have indicated which parameter is identified by each moment before the colon.) We used the moments (6) to compute the second set of standard errors in Table 5.

The single factor model with no additional constants ($a = 0$) sets

$$\begin{aligned} \gamma_0 &: \quad E\left[1_4' \varepsilon_{t+1}\right] = 0 && (9) \\ \gamma &: \quad E\left[(1_4' \varepsilon_{t+1}) \otimes f_t\right] = 0 && (10) \\ b &: \quad E\left[\varepsilon_{t+1} \otimes (\gamma_0 + \gamma' f_t)\right] = 0 && (11) \end{aligned}$$

For both restricted models, we can compute the χ^2 test that the remaining moments in (5) are zero, which we denote the J_T test in Table 7. Denoting the sample moments by g_T the test is $g_T' cov(g_T)^+ g_T \sim \chi^2_{rank(cov(g_T))}$ where + denotes a pseudo-inverse. (Details in the appendix.) We can also use the variance covariance matrix of estimated parameters from less restricted models to test the parameter restrictions of more restricted models, which we label a Wald test in Table 7.

Table 7 collects our test results. The single factor model with free intercepts seems a great success, with a χ^2 value of 5 and 16 degrees of freedom. However, the Wald test of its parameter restrictions is decisively rejected with an enormous p value.

The single factor model with restricted intercepts fails its overidentifying restrictions test with a 77 χ^2 value and 16 degrees of freedom. The Wald test of the same parameter restrictions, based on the unconstrained parameter variance-covariance matrix, also dramatically rejects with a 478 χ^2 value. Puzzlingly, the additional restriction of this model, that the intercepts a are zero, is decisively not rejected, with a χ^2 value of only 0.014, exactly as Figure 3 suggests. But if this, its only extra restriction, is not rejected, why does the model with $a = 0$ fare so much worse than the model with $a \neq 0$? The next row suggests the answer: when you constrain the intercept, it affects the slope coefficients; these have much smaller standard errors than the intercept, so the *slope* restrictions are violated when we constrain the *intercept.*

Wald tests based on the parameter variance covariance matrix from the single factor model with $a \neq 0$ rather than the completely unrestricted regressions paint a different picture. Here, the $a = 0$ restriction is not rejected, (finally!) at a sensible p-value of 54%. This time the slope restrictions are also not rejected, as we might have expected given the good vidual indication of the fit. However, the *joint* test that the intercepts are zero and the small changes in the slope coefficients that result when the intercept is restricted to zero now rejects.

Test type	Hypothesis	χ^2	dof	%p
1. Single factor model with $a \neq 0$				
J_T		5	16	98
Wald from unrestricted	$\beta = b\gamma'$	1,1074	20	0
2. Single factor model with $a = 0$				
J_T		77.6	16	0.00
Wald from unrestricted,	$a = 0$ and $\beta = b\gamma'$	478	24	0
Wald from unrestricted,	$\alpha = b\gamma_0$ (i.e. $a = 0$)	0.014	4	99.9
Wald from unrestricted,	$\beta = b\gamma'$	1,117	20	0
Wald from $a \neq 0$ model,	$a = 0$	3.08	4	54
Wald from $a \neq 0$ model,	$b_{unr} = b_{restr}$	0.58	4	96
Wald from $a \neq 0$ model,	$a = 0$ and $b_{unr} = b_{restr}$	27.3	8	0.06

Table 7. Model tests. The single factor model is $hprx_{t+1} = a + b(\gamma_0 + \gamma' f_t) + \varepsilon_{t+1}$ In the first panel it is estimated imposing $a = 0$ in the second panel it is estimated allowing $a \neq 0$ (which also affects the b estimates, leading to the difference between b_{unr} and b_{restr}). The unrestricted model is $hprx_{t+1} = \alpha + \beta f_t + \varepsilon_{t+1}$. J_T tests are tests that the moments not set to zero in estimation are in fact zero after accounting for sampling errors. Wald tests use the variance covariance matrix of parameter estimates in less restricted models to test the restrictions of more restricted models.

In summary, these tests do not paint a clear picture. Wald and overidentifying restrictions tests do not agree, and Wald tests from different models do not agree. We suspect that the asymptotic statistics – based on a 30×30 moment matrix and 12 monthly lags in the spectral density matrix – are simply not reliable in our sample.

2.4.2 Additional Lags

Following up on the unconstrained regressions with additional monthly lags in Figure 2, we run bond returns on additional lags of the state variable $\gamma' f_t$. Table 8 presents the results.

Regression 1 repeats the regression of holding period excess returns on $(\gamma' f_t)$ from Table 5 for comparison. In the second regression, we add an additional lag $\left(\gamma' f_{t-1/12}\right)$. The R^2 now jumps up to 0.43-0.46, nearly equal to the 0.44-0.48 values from the unconstrained two-lag regression in Table 2. 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 $\gamma' \left(f_t + f_{t-1/12}\right)/2$ 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.

The fourth regression investigates an additional lag. The pattern of coefficients seems similar, and the coefficients seem to be dying off. Though the additional coefficients are statistically significant (not shown), adding a second monthly lag raises the R^2 by no more than 0.02. Adding a one-year lag (not reported) does absolutely nothing for the R^2 of the regression. We conclude that the moving average of the first two lags is a good robust specification, though one may want to consider additional lags with an autoregressive pattern. We argue below that this pattern suggests an ARMA(1,1) model for monthly yields induced by i.i.d. measurement error.

	(1)		(2)			(3)		(4)			
	$\gamma' f_t$	R^2	$\gamma' f_t$	$\gamma' f_{t-1/12}$	R^2	$\frac{\gamma'(f_t+f_{t-1/12})}{2}$	R^2	$\gamma' f_t$	$\gamma' f_{t-1/12}$	$\gamma' f_{t-1/12}$	R^2
$hprx_{t+1}^{(2)}$	0.46	0.37	0.26	0.27	0.43	0.53	0.43	0.21	0.21	0.12	0.44
$hprx_{t+1}^{(3)}$	0.86	0.39	0.50	0.47	0.44	0.97	0.43	0.42	0.34	0.24	0.45
$hprx_{t+1}^{(4)}$	1.23	0.41	0.74	0.65	0.46	1.38	0.46	0.64	0.50	0.29	0.47
$hprx_{t+1}^{(5)}$	1.45	0.38	0.74	0.92	0.44	1.66	0.44	0.60	0.71	0.42	0.46

Table 8. 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 1964-1999.

2.4.3 Subsamples

Table 9 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 $\gamma' f$ 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 0.40 R^2. Of course, if you run a regression on its own fitted value you get a coefficient of 1.0.

The second set of rows break down the regression into 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 R^2 is a little lower in the inflationary period. This suggests that *real* holding period excess returns are better forecast by yield curve movements in an environment such as after the great monetary experiment, in which *real* interest rate movements are more important than inflation in driving the term structure. The 0.77 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.

The third 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. It is a dramatic 0.70 in the 90s, and even 0.51 when we constrain the coefficients γ to their full sample values. Again, this suggests that the forecasts have greatest power when return shocks are real rather than nominal.

	$y^{(1)}$	$f^{(1\to2)}$	$f^{(2\to3)}$	$f^{(3\to4)}$	$f^{(4\to5)}$	R^2	$\gamma' f$	R^2
1964:01-1999:12	-2.0	0.9	2.9	0.8	-2.1	0.40	1	0.40
1964:01-1979:08	-1.3	1.3	2.5	-0.1	-1.7	0.31	0.74	0.28
1979:08-1982:10	0.8	0.5	1.2	0.6	-0.7	0.77	0.80	0.28
1982:10-1999:12	-1.7	1.6	1.2	1.7	-2.3	0.35	1.01	0.33
1964:01-1969:12	-1.3	0.2	2.0	0.5	-1.9	0.30	0.70	0.24
1970:01-1979:12	-1.4	0.5	2.4	0.3	-0.6	0.22	0.69	0.17
1980:01-1989:12	-2.2	1.5	2.6	1.0	-1.8	0.42	1.11	0.37
1990:01-1999:12	-1.6	0.5	4.3	1.5	-2.5	0.70	1.69	0.51

Table 9. Subsample analysis of average return forecasting regressions. The first set of columns present regression

$$\frac{1}{4}\sum_{n=2}^{5} hprx_{t+1}^{(n)} = \gamma_0 + \gamma' f_t + \varepsilon_{t+1}$$

The second set of columns report a regression $\frac{1}{4}\sum_{n=2}^{5} hprx_{t+1}^{(n)} = a + b\left(\gamma' f_t\right) + \varepsilon_{t+1}$ using the γ parameter from the full sample regression, as presented in the top row. Overlapping annual forecasts using monthly data.

2.5 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.

2.5.1 Correlation between the return forecast and business cycles

Figure 5 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.

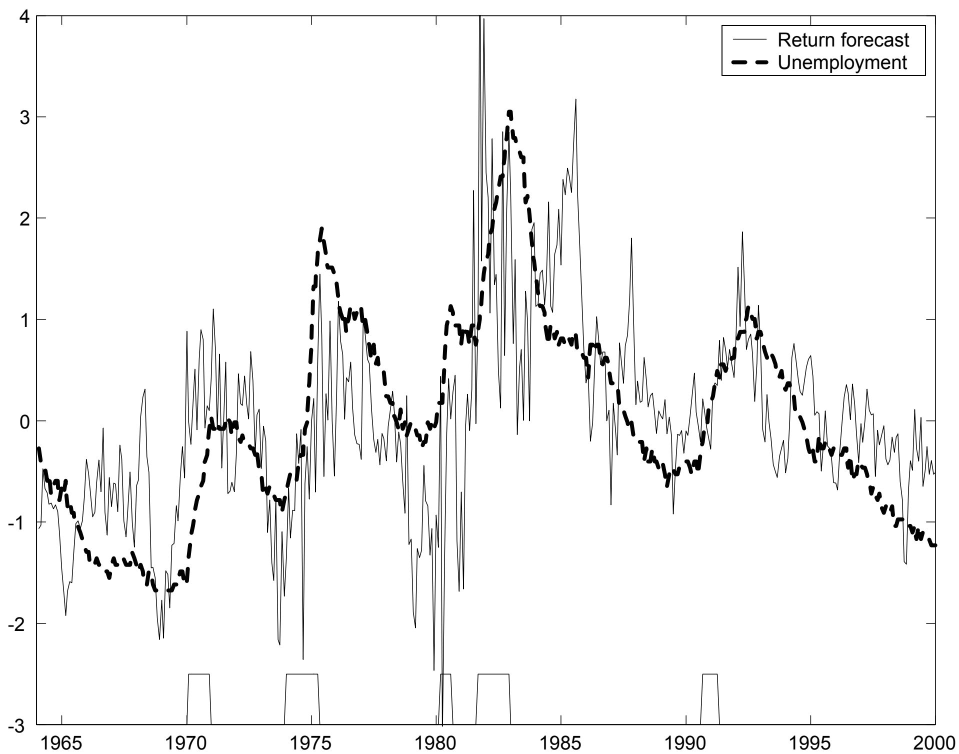

Figure 5: Return forecasting factor $\gamma' f_t$ and unemployment rate. Both series are transformed to $[x_t - E(x)]/\sigma(x)$ so that they fit on the same graph. The teeth at the bottom represent NBER business cycles.

Interestingly, the correlation 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.

2.5.2 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 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 $E_t\left(hprx_{t+1}\right) = a + b\gamma' f_t$ 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 $\gamma' f$, on the unemployment rate U and other macroeconomic variables. The first part of the table reminds us of the 0.40 and 0.45 R^2 when we forecast bond excess returns from $\gamma' f$. Despite its beautiful correlation with the return forecasting factor, unemployment forecasts bond excess returns with an R^2 of only 0.06. In a multiple regression it does not affect the size and significance of the $\gamma' f$ coefficient, and only raises the R^2 to 0.42.

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.

$\gamma' f$	$\gamma' f_{-1}$	R^2	$\gamma' f$	U	R^2
1		0.40		0.66	0.06
(8.0)				(1.9)	
0.56	0.58	0.45	1.17	-0.48	0.42
(6.4)	(5.3)		(7.7)	(-1.5)	

$\gamma' f$	XLI	R^2	$\gamma' f$	cay	R^2	$\gamma' f$	cpi	R^2
	-0.12	0.01		0.79	0.05		-0.24	0.03
	(-0.6)			(2.3)				
1.01	-0.16	0.41	0.97	0.30	0.40	0.99	-0.18	0.41
(7.4)	(-1.4)		(7.3)	(1.1)		(8.5)	(-0.9)	

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 $\gamma' f$ by running the regression $\frac{1}{4}\sum_{n=2}^{5} hprx_{t+1}^{(n)} = a + \gamma' f_t + \varepsilon_{t+1}$ in a first stage. Overlapping

annual forecasts, 1964:01-1999:12 Standard errors corrected for overlap and heteroskedasticity by GMM.

2.5.3 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 $\gamma' f$ 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 0.16-0.17. 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 0.39-0.44.

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.

industrial production				coincident index			
$\gamma' f$	$y^{(5)} - y^{(1)}$	LI	R^2	$\gamma' f$	$y^{(5)} - y^{(1)}$	LI	R^2
0.060			0.01	0.063			0.002
(0.19)				(0.29)			
	-0.83		0.17		-0.58		0.16
	(-3.0)				(-2.5)		
		0.88	0.39			0.67	0.44
		(9.3)				(10.5)	
0.32	0.70	-0.67	0.46	0.23	-0.44	0.55	0.50
(1.7)	(7.2)	(-3.0)		(2.2)	(-3.0)	(7.1)	

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 $\gamma' f$, the term spread $y^{(5)} - y^{(1)}$, and the Stock-Watson leading index. overlapping annual forecasts, 1964:01-1999:12 Standard errors corrected by GMM.

2.5.4 Forecasting stock returns

The slope of the term structure forecasts stock returns, as emphasized by Fama and French (1989). Table11.1 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 has 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 4.6 coefficient – one percentage point term spread corresponds to .4.6 percentage point increase in stock return. The R^2 is only 6.2% 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 8.8%.

Regression #	d/p	$y^{(5)} - y^{(1)}$	$\gamma' f$	R^2(%)
1	2.5			2.0
	(0.81)			
2: 1964-1989	6.97			14.4
	(2.43)			
3		4.6		6.2
		(1.8)		
4	2.8	4.8		8.8
	(0.97)	(2.0)		
5			1.87	8.9
			(2.7)	
6		2.0	1.46	9.7
		(0.69)	(1.82)	
7	(1.00)		1.76	9.2
	(0.37)		(2.79)	
8	$y_t^{(1)}, f_t^{(1\to2)}, f_t^{(2\to3)}, f_t^{(3\to4)}, f_t^{(4\to5)}$			13.7

Table 11.1. 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, 1964-1999. 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 8.9%, its R^2 is slightly higher than that of the term spread and d/p combined. The coefficient is 1.87. The return forecasting factor is the average expected return across 2-5 year bonds. The 5 year bond in Table 5 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 in half and rendered very 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 13.7%. Figure 6 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.

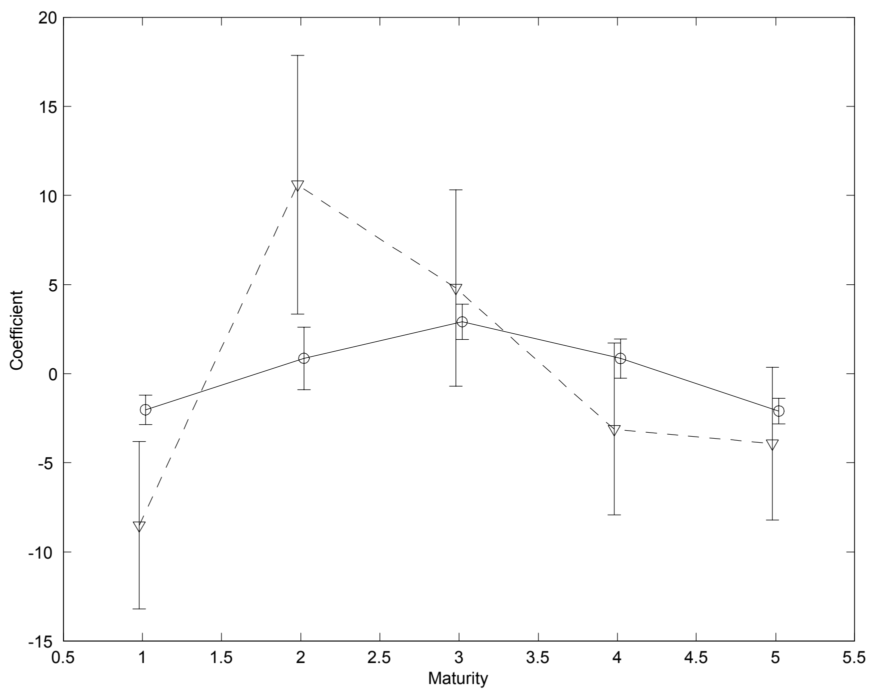

Figure 6: Coefficients in a regression of one-year value weighted NYSE stock excess returns on all forward rates (dashed line, triangles) and average bond excess returns on all forward rates (solid line, circles). Error bars are +/- two standard errors.

3 Risk premia

Seeing a pattern in expected returns, we naturally want to relate that pattern to covariances. The central parable of finance explains expected returns by their covariance with shocks to factors, or, in logs,

$$E_t(hprx_{t+1}) = cov_t(hprx_{t+1}, \varepsilon'_{t+1})\lambda_t - \frac{1}{2}\sigma_t^2(hprx_{t+1}) \tag{12}$$

where $hprx_{t+1}$ denote the (4×1) vector of excess log returns, ε_{t+1} is a vector of orthogonalized shocks to pricing factors (proxies for marginal utility), and λ_t is a vector of factor risk premia or market prices of risk.

3.1 Calculating market prices of risk

Let C denote the covariance matrix of returns with shocks,

$$C = cov_t(hprx_{t+1}, \varepsilon'_{t+1}).$$

In an exactly identified models with as many shocks as returns, we can solve for λ_t by simply premultiplying (12) by C^{-1}. Of course, we will look for elements of λ_t that seem close to zero, suggesting the traditional specification with only a few factors. We also examine underidentified situations with more shocks than returns. For example, in a VAR using yields on 1-5 year bonds, there are potentially 5 shocks and 4 returns. Including macroeconomic variables increases the number of shocks. In these cases, C is not invertible. CC' still is invertible so we can find a solution

$$\lambda_t = C'(CC')^{-1}\left[E_t(hprx_{t+1}) + \frac{1}{2}\sigma_t^2(hprx_{t+1})\right]. \tag{13}$$

This solution is not unique. With more shocks than returns, many different λ_t can exactly relate expected returns with covariances. We characterize the properties of this identification in Proposition 3 below.

The market prices of risk λ_t also depend on how shocks are orthogonalized, as the C matrix will reflect the orthogonalization. When underlying shocks are correlated, the order of orthogonalization will determine how the market price of their common component is assigned to one or the other shock. We orthogonalize our shocks with the most interesting ones first, in order to let as much of the market price of risk reflect the interesting shocks as possible.

We have a model for expected returns,

$$E_t\left(hprx_{t+1}\right) = a + b\left(\gamma_0 + \gamma' f_t\right). \tag{14}$$

We start by studying VAR type specifications with constant conditional shock variances and covariances. Plugging (14) in (13), we can calculate the risk premia λ_t for these specifications by

$$\lambda_t = C'(CC')^{-1}\left[a + b\left(\gamma_0 + \gamma' f_t\right)\right] + C'(CC')^{-1}\frac{1}{2}\sigma_t^2(hprx_{t+1})$$

or, compactly,

$$\begin{aligned} \lambda_t &= \lambda_0 + \delta\left(\gamma_0 + \gamma' f_t\right) && (15) \\ \lambda_0 &= C'\left(CC'\right)^{-1}\left(a + \frac{1}{2}\sigma_t^2(hprx_{t+1})\right) && (16) \\ \delta &= C'\left(CC'\right)^{-1} b. && (17) \end{aligned}$$

If we use all the forward rates as state variables in forecasting returns,

$$E_t\left(hprx_{t+1}\right) = a + Bf_t,$$

rather than the restricted model (14), λ_t will depend on all of the forward rates f_t, e.g. $\lambda_t = \lambda_0 + \lambda_1 f_t$ where λ_1 is a 4×5 matrix. The one-factor structure we have found for expected excess returns carries over and gives a nice one-factor structure to the time-varying risk premium.

3.2 Discount factors and affine models

The fundamental parable of finance is often stated in terms of a stochastic discount factor rather expected returns. Equation (12) is equivalent to a stochastic discount factor of the form

$$m_{t+1} = e^{-y_t^{(1)} - \frac{1}{2}\lambda_t' E_t(\varepsilon_{t+1}\varepsilon_{t+1}')^{-1}\lambda_t - \lambda_t'\varepsilon_{t+1}}. \tag{18}$$

With this discount factor and conditionally normal log returns and shocks, $1 = E_t\left(m_{t+1}R_{t+1}\right)$ where R denotes the level (not log) return is equivalent to (12). Thus, in calculating λ, we are also characterizing a stochastic discount factor that captures bond returns.

Thinking about bonds in terms of one-period expected return-beta representations, or even one-period discount factors, has a decidedly old-fashioned flair. Most of the vast term structure literature studies bond prices and yields rather than returns, in the context of explicit and typically affine models. An affine model is lurking here however, and by calculating λ we are also implicitly finding the market prices or risk and risk-neutral probabilities that define an affine model.

Proposition 1. Let X_t denote a vector of state variables that follows $X_t = \mu + \phi X_{t-1} + \Sigma\varepsilon_t$, with i.i.d. normally distributed shocks ε_t and $E(\varepsilon_t\varepsilon_t') = I$. Let the short rate $y_t^{(1)}$ be included in the state vector X_t, $y_t^{(1)} = e'X_t$. Let m_{t+1} be as given by (18), where λ is a linear function of X_t, e.g. $\lambda_t = \lambda_0 + \lambda_1' X_t$. Then bond prices, generated by $e^{p_t^{(n)}} = E_t(m_{t+1}m_{t+2}...m_{t+n})$ are linear functions of the state variables, i.e. we can find A_n and B_n such that

$$p_t^{(n)} = A_n + B_n' X_t. \tag{19}$$

The affine model is equivalent to risk-neutral pricing with distorted probabilities

$$\phi^* \equiv \phi - \Sigma\lambda_1 \tag{20}$$
$$\mu^* \equiv \mu - \Sigma\lambda_0. \tag{21}$$

Yields and forward rates are of course also linear functions of the state variables. The proof consists simply of algebra; grind out the conditional expectation that defines bond prices and derive the linear form. We present it in the Appendix.

Intuitively, the conditional heteroskedasticity of the discount factor results in time-varying risk premia that are linear functions of the state variables as in the continuous-time setups of Fisher (1998) and Dai and Singleton (2001). Ang and Piazzesi (2001) study a similar discrete-time affine model.

Often the state variable includes bond prices or yields; in fact often the state variable consists only of prices or yields. Now we have a tricky identification problem to solve: we have to make sure that the bond prices $p_t^{(n)}$ that come *out* of the model are the same as the state variables $p_t^{(n)}$ that go *in* the model. For example if the price $p_t^{(n)}$ is included in the state variable X_t, then we must choose λ_t so that $A_n = 0$ and B_n is the vector that recovers $p_t^{(n)}$ from X_t. In our case, however, this problem is solved by the identification (13):

Proposition 2. Suppose X_t contains a full set of prices, i.e. suppose that we can recover prices of 1 through N period bonds from X_t by $[\ p_t^{(1)} \quad p_t^{(2)} \quad .. \quad p_t^{(N)}] = PX_t$. Then, λ_t calculated by (13) form a self-consistent affine model; the predicted bond prices by (19) are the same as those recovered directly by PX_t.

Formula (13) requires a full set of returns, but if you have a full set of prices you have a full set of returns as well as yields and forward rates. Again, the proof is in the Appendix. It is possible to find a set of $\lambda_t = \lambda_0 + \lambda_1 X_t$ that form a self-consistent affine model even when X does not include a full set of returns, but the procedure is a bit more complicated than our simple formula (13).

3.2.1 Implications

This connection to affine models has important implications for our calculations, beyond just showing that there is such a connection and we are not hopelessly out of date in studying expected return-beta models.

First, many papers have been written on the subject whether one can construct affine models that are consistent with bond return predictability, and the Fama-Bliss (1987) regressions in particular. Examples include Dai and Singleton (2001), Duffee (2001) and Duarte (1999). If Fama and Bliss' 0.15 R^2 poses problems, one might think our 0.45 R^2 are fatal. Proposition 2 and our formulas for λ show that nothing of the sort is true. We can *construct* market prices of risk λ_t that capture Fama and Bliss' regression evidence, our much stronger regression evidence, or much more complex return forecasting regressions.

Furthermore, the model, by construction, exactly reproduces the bond prices, yields, or forward rates that are used as state variables. One may choose to examine specifications with a restricted number of factors, but this is not necessary. We can construct a model with zero pricing errors.

Second, not only is the term structure model affine, it is homoskedastic. Many affine models also include conditional heteroskedasticity of the shocks. In those models, expected return variation and curvature of the yield curve are tied to changes in conditional volatility. It's tempting to conclude that expected return variation and curvature *must* come from changing conditional volatility, but the propositions show that is not the case. (Conditionally heteroskedastic models may be important in fitting the data of course, and some time-varying risk premium may in fact be time-varying risk. On the other hand such models are often more complex, especially in discrete time. Our point here is that we do not *have* to study conditionally heteroskedastic models in order to fit yields and expected returns.)

In fact, the market prices of risk λ to make all this happen are *underidentified*. Lots of other choices would work as well. The choice we pursue has the following properties, familiar from Hansen and Jagannathan (1991) and Cochrane and Saá-Requejo (1999), and proved in the appendix:

Proposition 3. Among all market prices of risk λ_t *that price the available bond returns, or (equivalently) that produce a self-consistent affine model with yields as state variables, the market prices of risk defined by (*13*) produce the minimum size* $\lambda_t'\lambda_t$*, the discount factor with minimum volatility, and the minimum value of the maximum Sharpe ratio. They set to zero the prices of risk* λ_t *of any shock uncorrelated with bond returns.*

3.3 Yield curve shocks: an exact identification

Equations (15)-(17) tell us how to calculate market prices of risk λ. All that remains is to choose an interesting set of shocks ε_t. Eventually, we want to tie the shocks ε_t to macroeconomic risks. However, it is an interesting first step to describe expected returns in terms of covariances with factor-mimicking portfolios; i.e. a simple and interpretable set of portfolios such as the market return, or the size and book-to-market portfolios that describe stock returns. A natural set of bond shocks are the "level," "slope," and perhaps "curvature" factors that describe much of the variation in the term structure. Thus we ask questions such as, "Are expected returns compensation for holding "level" risk or "slope" risk?"

We follow the latent variable tradition started by Hansen and Hodrick (1983) to identify yield curve shocks that *exactly* generate our time-varying expected returns. If you have a vector of expected returns, you can always construct a single index model; you can find a single (ex-post mean-variance efficient) portfolio such that the expected return of each asset is a linear function of the covariance of that asset's return with the chosen portfolio. The portfolio has N weights, and there are N expected returns to match. Hansen and Hodrick noticed that the same point holds with a single factor model of time-varying expected returns.

$$E_t(hprx_{t+1}) = b\left(\gamma_0 + \gamma' f_t\right)$$

where $hprx$ and b are 4×1 vectors. Given this model for expected returns, we can find a single combination of return shocks so that conditional mean returns line up perfectly against covariances of returns with that linear combination of shocks, multiplied by a time-varying factor risk premium linear in $\gamma' f$. We just have to find a portfolio with shocks η such that $cov(hprx_{t+1}\eta) = b$.

In this way, we can *construct* a single factor–a single combination of yield curve shocks–that exactly explains our time-varying expected returns, and then interpret it, rather than try various shocks to see which ones produce covariances that explain time-varying expected returns. We can foresee that this procedure will lead to a "level" shock. The b coefficients rise steadily with maturity. Thus, our "factor" must produce covariances with returns that rise steadily with maturity. An upward shift in the entire yield curve is precisely such a factor.

We extend this idea to allow for the conditional variance term and a constant that may not exactly follow the single index model. Our model of time-varying expected returns is

$$E_t(hprx_{t+1}) + \frac{1}{2}\sigma_t^2(hprx_{t+1}) = \tilde{a} + b\left(\gamma_0 + \gamma' f_t\right)$$

We show below how to construct two orthogonal yield curve shocks $\eta_{t+1}^i = w_i'\left(y_{t+1} - E_t y_{t+1}\right)$ and two factor risk premia, one time varying and the other constant

$$\begin{aligned} \lambda_t^1 &= \lambda_0^1 + \delta\left(\gamma_0 + \gamma' f_t\right) \\ \lambda_t^2 &= \lambda_0^2 \end{aligned}$$

that exactly capture our model of bond expected returns, i.e. such that

$$E_t(hprx_{t+1}) + \frac{1}{2}\sigma_t^2(hprx_{t+1}) = cov(hprx_{t+1}\eta_{t+1}^1)\lambda_t^1 + cov(hprx_{t+1}\eta_{t+1}^2)\lambda_0^2. \tag{22}$$

If $\tilde{a} = 0$ and expected returns (plus the variance term) follow a true single-index model, our representation collapses to

$$E_t(hprx_{t+1}) + \frac{1}{2}\sigma_t^2(hprx_{t+1}) = cov(hprx_{t+1}\eta_{t+1}^{level})\,\delta\left(\gamma_0 + \gamma' f_t\right).$$

The second shock and factor risk premium are only there to capture deviations from the exact single-index model, $\tilde{a} = 0$.

Figure 7 shows how yields move in response to each of our two orthogonal yield curve shocks. Obviously, we can label the first shock a "level" shock and the second shock a "slope" shock, and we use these labels rather than "1" and "2" below. The level shock

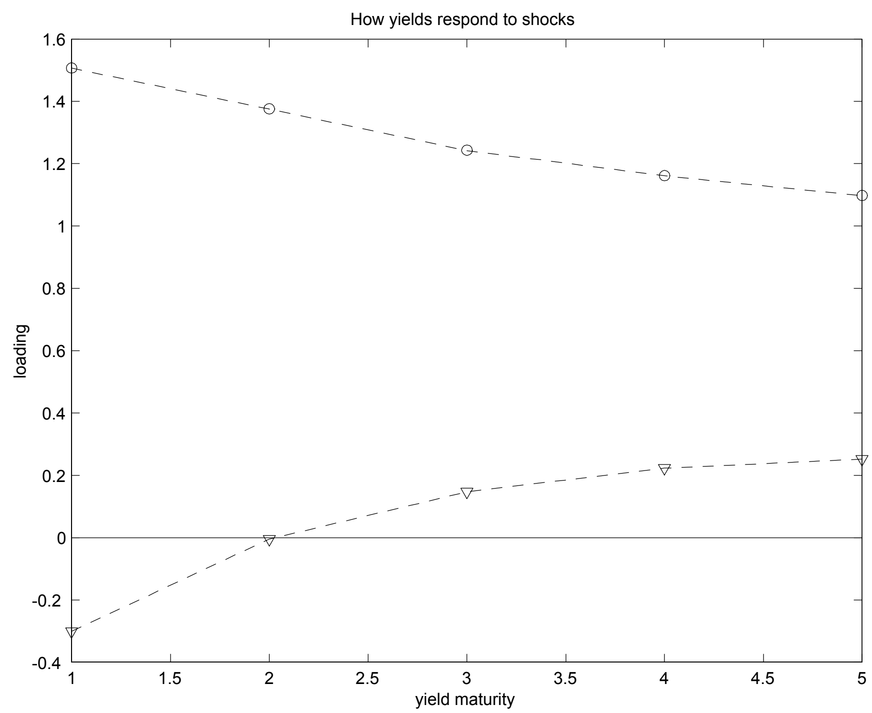

Figure 7: How yields move in response to the two yield curve shocks.

is not exactly level, which is fortunate since parallel shifts in the term structure violate arbitrage.

Table 12 presents how bond returns (rather than yields) are affected by each shock. These are also the covariances and betas, since the shocks are orthogonal and have unit variances. An upward level shock produces a negative return, and more and more so for longer maturities. This pattern of betas is of course exactly what we need to explain the strong rise in conditional mean returns across maturities when $\gamma' f$ is high, and the smaller rise in unconditional mean returns across maturities as well. The slope factor does exactly what a slope factor should do, raising the returns of short term bonds and lowering those of long-term bonds

	Maturity			
Shock	2	3	4	5
level	-1.5	-2.8	-3.7	-4.6
slope	0.3	0.08	-0.44	-0.89

Table 12. Effect of level and slope shocks on bond excess returns.

We estimate $\lambda_0^{level} = 0.0028$, $\delta = 2.22$, $\lambda_0^{slope} = 0.0004$ (and, by assumption $\delta^{slope} = 0$.) The units are annual percent excess returns. The numbers suggest that the single index model with $\lambda_0^{level} = 0$ and no premium for the slope shock will be an excellent approximation. The sign also makes sense. A shock that *raises* yields, *lowers* returns, and you get a *positive* premium for holding such shocks.

Table 13 shows how these covariances and factor risk premia add up to explain expected returns. The table starts at the sample mean value of $\gamma' f$. The first row shows the expected return to be explained by covariances. The numbers remind us of the slight upward slope in bond returns. Half of the slope (39 basis points) comes right away in the two year bond excess return, and then average returns increase slowly out to 72 basis points for 5 year bonds. Rows 2-4 break up this expected return into components. $\tilde{\alpha}$ gives the deviation from the exact single factor model, and $1/2\sigma^2$ gives the variance term. $b\left[\gamma_0 + \gamma' E(f)\right]$ gives the contribution of expected log returns in the exact single factor model. This breakdown shows that the $\tilde{a}$ term is quite small, so the single factor model is an excellent approximation. The $1/2\sigma^2$ terms are also small, comforting us that the phenomenon is really about expected returns not about variances.

		Maturity 2	3	4	5
	1. At $\gamma' f = E(\gamma' f)$				
	a. To be explained:				
1	$E(hprx) + \frac{1}{2}\sigma^2 = \tilde{a} + b\left[\gamma_0 + \gamma' E(f)\right]$	0.39	0.61	0.78	0.72
2	$\tilde{a}$	0.13	0.13	0.07	-0.10
3	$\frac{1}{2}\sigma^2$	0.02	0.04	0.07	0.11
4	$b\left[\gamma_0 + \gamma' E(f)\right]$	0.26	0.49	0.70	0.82
	b. Explanation				
5	level: $cov(hprx, \eta^{level})\left[\lambda_0^{level} + \delta(\gamma_0 + \gamma' E(f))\right]$	0.32	0.59	0.85	1.00
6	slope: $cov(hprx, \eta^{slope})\lambda_0^{slope}$	0.06	0.02	-0.08	-0.28
7	Total explained	0.39	0.61	0.78	0.72
	2. Effect of a 1σ increase in $\gamma' f$:				
	a. To be explained:				
8	$E_t(hprx_{t+1}) - E(hprx) = b\gamma' \left[\sigma(f)\right]$	1.19	2.20	3.16	3.72
	b. Explanation:				
9	level: $cov(hprx, \eta^{level})\left[\delta^{level}\ \gamma' \sigma(f)\right]$	1.19	2.20	3.16	3.72

Table 13. Mean excess returns and their explanation by covariance of returns with level and slope shocks to yields. Effect on expected returns of a one standard deviation increase in the return forecasting factor, and its explanation by covariance times time-varying factor risk premium. Units are all percentage annual returns.

Rows 5 and 6 show our estimates of covariances times factor risk premia evaluated when $\gamma' f$ is at its sample mean. Clearly, the level shock is doing almost all the work. The main effect of the slope shock is to reduce by 28 basis points the 1% risk premium on 5 year bond returns predicted by the level shock.

Next, we turn to the far more interesting issue: time-varying expected returns. Row 8 shows how much expected returns rise when $\gamma' f$ rises by one standard error above its mean. These numbers are much larger than the average returns; from 1.2% for 2 year bonds up to 3.7% expected excess returns on 5 year bonds. Row 9 shows that covariance with the level shock multiplied by the time-varying risk premium exactly captures this pattern in time-varying bond expected returns.

To understand the level shock more clearly, we compare it with a factor decomposition of the yield shock variance covariance matrix. Let

$$v_{t+1}^{y} = y_{t+1} - E_t(y_{t+1})$$

If we perform an eigenvalue decomposition of the yield shock covariance matrix,

$$E(v_{t+1}^{y} v_{t+1}^{y\prime}) = Q\Lambda Q',$$

where Q satisfies $Q' = Q^{-1}$and Λ is diagonal, and then let

$$\Sigma_y = Q\Lambda^{\frac{1}{2}},$$

then we can write

$$v_{t+1}^{y} = \Sigma_y \eta_{t+1}$$

where $E(\eta^y \eta^{y\prime}) = I$. The eigenvalue decomposition produces a factor analysis, with the orthogonal components that explain in order as much of yield innovation variance as possible.

Figure 15 presents the eigenvalue decomposition of the yield shock covariance matrix, the columns of Σ_y. The lines represents how much each yield rises in response to one of the orthogonalized unit variance shocks η_{t+1}. There is one strong "level" shock. There is a second "slope" shock that lowers short term yields and raises long term yields. Then, there are three idiosyncratic shocks with strong peaks on the 2, 3, and 4 year yields. (The largest of these, the two year yield shock, can also be interpreted as a "curvature" factor. We do not see stronger curvature factors because we do not use yields less than a year. In most analyses the curvature factor mainly accounts for the difference between bonds with less and more than a year maturity.) The idiosyncratic shocks are multiplied by small eigenvalues to produce small loadings in the figure. This leaves us with an approximate two-factor structure for yield shocks. [3]

The comparison between Figure 7 and Figure 15 is striking. The two factors whose shocks are constructed to explain bond *expected* returns turn out to be almost exactly the level and slope factors that dominate the yield covariance matrix. A market price of risk calculation using the eigenvalue factors of Figure 15 gives almost exactly the same result as in Table 12.

[3] A similar exercise for forward rates also produces level and slope factors. The idiosyncratics are a bit stronger for forward rates. This may be a simple result of smoothing, because yields are a cumulative average of forward rates. The conditional variance-covariance matrix of excess holding period returns also produces level and slope factors.

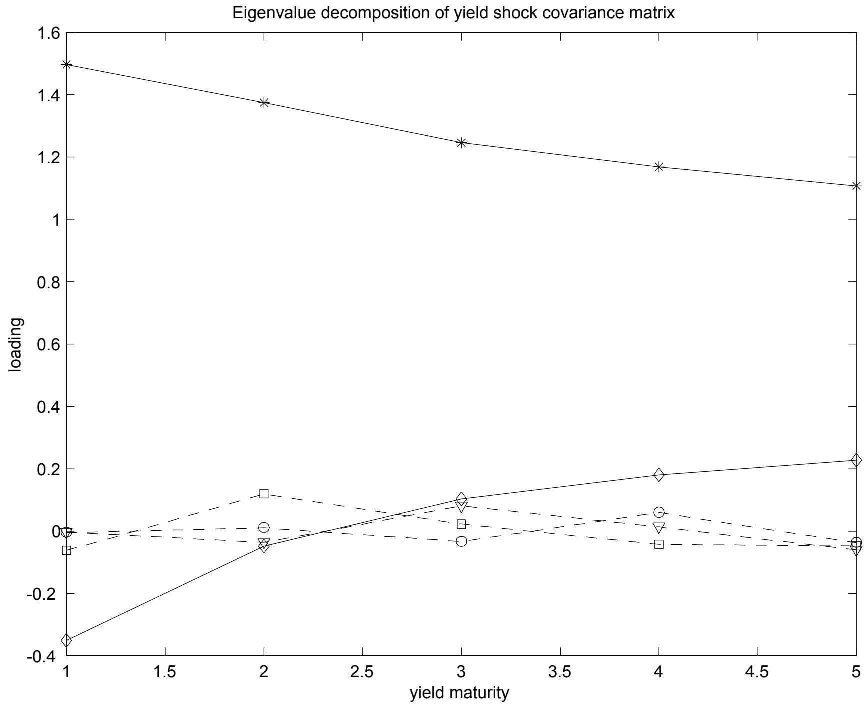

Figure 8: Eigenvalue decomposition of the restricted yield VAR innovation variance covariance matrix. The VAR is $y_{t+1} = \mu_y + \phi_y y_t + v_{t+1}^y$.The graph presents the columns of $Q\Lambda^{\frac{1}{2}}$ from the eigenvalue decomposition $E(v^y v^{y\prime}) = Q\Lambda Q'$. Each line represents how yields of various maturities are affected by a unit standard deviation movement in each orthogonalized shock.

3.3.1 Details of the calculation

Start with a VAR representation for bond yields

$$y_{t+1} = \mu_y + \phi_y y_t + v_{t+1}^y$$

where y_t contains all Fama-Bliss yields $y_t = \left[y_t^{(1)}, y_t^{(2)}, y_t^{(3)}, y_t^{(4)}, y_t^{(5)}\right]'$. Since this yield VAR implies return regressions, we use throughout a specification of ϕ_y, μ_y that is consistent with the single factor model for expected returns $E_t(hprx_{t+1}) = \tilde{a} + b(\gamma_0 + \gamma' f_t)$. The appendix shows how to calculate this restricted yield VAR and contrasts it with the unrestricted yield VAR, showing that the restrictions are small.

We search for linear combinations of the yield shocks $\eta_{t+1} = w' v_{t+1}^y$ to be our factors, covariances with which will exactly explain expected returns. We are looking for representations of the form

$$E_t(hprx_{t+1}) + \frac{1}{2}\sigma_t^2(hprx_{t+1}) = cov_t(hprx_{t+1}, \eta_{t+1}')\lambda_t \tag{23}$$

To simplify notation, fold the variance term in the constant,

$$\tilde{a} = a + \frac{1}{2}\sigma_t^2(hprx_{t+1}) \tag{24}$$

and write the covariance matrix of returns with yield shocks as

$$C = cov_t(hprx_{t+1}, v_{t+1}^{y\prime}).$$

Now we can write (23) as

$$\tilde{a} + b\left(\gamma_0 + \gamma' f_t\right) = Cw\lambda_t. \tag{25}$$

(a and b are 5×1, C is 4×5, w is 5×1, λ_t is a number.)

The single-index model

We start with the case $\tilde{a} = 0$, and identify a single shock that would exactly explain an exact single-factor model, including the constant. We will then identify a second shock to soak up whatever unconditional average return or variance term that is not explained by the first shock. As before, we pick the solution to (25)

$$C'(CC')^{-1}\left[b\left(\gamma_0 + \gamma' f_t\right)\right] = w\lambda_t$$

Clearly, λ_t will have to have a linear form,

$$\lambda_t = \delta\left(\gamma_0 + \gamma' f_t\right).$$

Then we need

$$C'(CC')^{-1}b = w\delta$$

δ is a number, and the left hand side is a 5×1 vector, so this equation ties down the weights w up to a normalization.

We normalize to unit variance shocks, so

$$w'E(vv')w = 1$$

Hence, we have the weights and the free parameter δ in the factor risk premium

$$\begin{aligned} \delta &= \sqrt{b'(CC')^{-1}CE(v^y v^{y\prime})C'(CC')^{-1}b} \\ w &= C'(CC')^{-1}b\,/\delta \end{aligned}$$

A two-index model

Next, we allow for $\tilde{a} \neq 0$. We will need a two index model. The procedure is exactly the same, though the algebra is less transparent because scalars are now vectors and matrices.

We are now looking for two shocks, or a 2×1vector $\eta_{t+1} = w' v^y_{t+1}$. We want a solution to (25),

$$\tilde{a} + b\left(\gamma_0 + \gamma' f_t\right) = C w \lambda_t. \tag{26}$$

in which w is a 5×2 matrix, and λ_t is a 2×2 matrix. Clearly, λ_t must be of the form

$$\lambda_t = c + d\left(\gamma_0 + \gamma' f_t\right)$$

where c and d are 2×1. Matching the constant and time-varying terms separately, we have

$$\tilde{a} = Cwc;\ b = Cwd$$

or

$$\left[\begin{array}{cc} \tilde{a} & b \end{array}\right] = Cw \left[\begin{array}{cc} c & d \end{array}\right].$$

We choose the usual solution,

$$C'(CC')^{-1}\left[\begin{array}{cc} \tilde{a} & b \end{array}\right] = w \left[\begin{array}{cc} c & d \end{array}\right]$$

$$C'(CC')^{-1}\left[\begin{array}{cc} \tilde{a} & b \end{array}\right]\left[\begin{array}{cc} c & d \end{array}\right]^{-1} = w \tag{27}$$

We choose the usual normalization to unit variance orthogonal shocks, $cov(\eta\eta') = I$ or

$$w' E(v^y v^{y\prime}) w = I_2.$$

The normalization implies

$$\left[\begin{array}{cc} c & d \end{array}\right]^{-1\prime}\left[\begin{array}{cc} \tilde{a} & b \end{array}\right]'(CC')^{-1} C E(v^y v^{y\prime}) C'(CC')^{-1}\left[\begin{array}{cc} \tilde{a} & b \end{array}\right]\left[\begin{array}{cc} c & d \end{array}\right]^{-1} = I_2$$

$$\left[\begin{array}{cc} \tilde{a} & b \end{array}\right]'(CC')^{-1} C E(v^y v^{y\prime}) C'(CC')^{-1}\left[\begin{array}{cc} \tilde{a} & b \end{array}\right] = \left[\begin{array}{cc} c & d \end{array}\right]'\left[\begin{array}{cc} c & d \end{array}\right] \tag{28}$$

Solving this equation for $\left[\begin{array}{cc} c & d \end{array}\right]$ requires and allows one more identification decision. We specify that the new shock–whose only purpose is to explain the constant – has a constant risk premium. Thus, we want

$$\begin{aligned}
\left[\begin{array}{cc} c & d \end{array}\right] &= \left[\begin{array}{cc} \lambda_0^{level} & \delta \\ \lambda_0^{slope} & 0 \end{array}\right] \\
\left[\begin{array}{cc} c & d \end{array}\right]'\left[\begin{array}{cc} c & d \end{array}\right] &= \left[\begin{array}{cc} \left(\lambda_0^{level}\right)^2 + \left(\lambda_0^{slope}\right)^2 & \lambda_0^{level}\delta \\ \lambda_0^{level}\delta & \delta^2 \end{array}\right]
\end{aligned}$$

(Of course, we don't know yet that the shocks will have level and slope interpretation, but it seems pointless to introduce another notation.) In this way, we can find c and d from a triangular factorization of the left hand side in (28)

In sum, we can now calculate the weights w_{level}, w_{slope} that define our shocks and the coefficients λ_0^{level}, λ_0^{slope}, δ that define our exact two-factor representation

$$\tilde{a} + b\left(\gamma_0 + \gamma' f_t\right) = C w_{level}\left[\lambda_0^{level} + \delta\left(\gamma_0 + \gamma' f_t\right)\right] + C w_{slope}\lambda_0^{slope} \tag{29}$$

The C matrix, the variance term and betas

We still have to find the covariance

$$C = cov_t(hprx_{t+1}, v_{t+1}^{y\prime})$$

of holding period return shocks with yield shocks. From the definition of excess holding period returns in terms of yields

$$\begin{aligned} hprx_{t+1}^{(n)} &= p_{t+1}^{(n-1)} - p_t^{(n)} - y_t^{(1)} && (30) \\ &= -(n-1)y_{t+1}^{(n-1)} + n y_t^{(n)} - y_t^{(1)}, && (31) \end{aligned}$$

we know that

$$cov_t(hprx_{t+1}^{(n)}, v_{t+1}^{y\prime}) = -(n-1)cov_t(y_{t+1}^{(n-1)}, v_{t+1}^{y\prime})$$

and thus

$$C = cov_t(hprx_{t+1}, v_{t+1}^{y\prime}) = -\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 4 & 0 \end{bmatrix}$$

We also have the σ^2 term in the constant,

$$\begin{aligned} cov(hprx_{t+1}, hprx_{t+1}') &= CE(v^y v^{y\prime})C' \\ \sigma_t^2(hprx_{t+1}) &= diag\left[CE(v^y v^{y\prime})C'\right] \end{aligned}$$

Rather than display the weights w that form the portfolios, it is more interesting to display how a shock affects yields and returns. We are looking for a representation

$$v_{t+1}^y = \Sigma\eta_{t+1} + \varepsilon_{t+1};\ E(\varepsilon_{t+1}\eta_{t+1}) = 0$$

The formula for OLS regression coefficients gives us Σ (with $\sigma^2(\eta) = 1$),

$$\Sigma = cov(v_{t+1}^y \eta_{t+1}) = cov(v_{t+1}^y v_{t+1}^{y\prime})w$$

The matrix Σ tells us how yields respond to the η shock. Figure 7 graphs the columns of Σ (one in this case, two in the next case). From (30), we can find how a shock η affects returns, i.e. the betas. Since C transforms from yields to returns, the answer is of course

$$hprx_{t+1} - \tilde{a} - b\left[\gamma_0 + \gamma' f_t\right] = C\Sigma$$

3.4 Expected return shocks

Saying that time-varying expected bond returns result from a time-varying premium on covariances with the level shock is not a complete answer. We want to know, what more fundamental risks does the level shock in bond yields represent? To answer this question, we examine a variety of other shocks in bond yields and macroeconomic variables.

Given that $\gamma' f$ is the state variable driving expected returns, it makes sense to ask whether innovations in $\gamma' f$ are important factors for bond returns. This follows Merton's (1973) logic that innovations to state variables for the investment opportunity set ought to show up as factors. Campbell (1996) Ferson and Harvey (1999) find that innovations to variables that forecast the market return can explain the cross-section of stock returns; Brennan Xia and Wang (2001) find that such innovations can explain the Fama-French size and book to market factors' ability to price the cross section of stocks. Perhaps a similar pattern will emerge for bonds.

To pursue this idea, we again start with the yield VAR, restricted so that the implied return regressions follow the single factor model. We orthogonalize the yield shock covariance matrix, to define shocks η_{t+1} such that $v_{t+1}^{y} = \Sigma \eta_{t+1}$, $E(\eta_{t+1}\eta_{t+1}') = I$. We construct the orthogonalization so that the first shock is the shock to the return forecasting factor $\gamma' f$. By orthogonalizing with the $\gamma' f$ shock first, we assign as much of the market price of risk to the $\gamma' f$ shock as possible. If the $\gamma' f$ shock is highly correlated with the "level" shock derived in the last section, we will assign the market prices of risk from the level shock to the $\gamma' f$ shock by this orthogonalization. We define the remaining shocks by an eigenvalue decomposition of the covariance matrix, which is the easiest way to construct a factor analysis.

Figure 9 presents the resulting decomposition of the yield variance-covariance matrix – the columns of the Σ in the representation

$$y_{t+1} = \mu_y + \phi_y y_t + \Sigma \eta_{t+1}.$$

The solid line presents the loadings of yields of each maturity on the expected return shock. The expected return shock turns out to have almost exactly the character of the "slope" shock identified above. It also contributes very little to the variance of yields. The second shock identified after the expected return shock has a "level" character, and captures the vast majority of yield shock variance.

Table 14 calculates the correlation between the expected return shock $\eta^{\gamma' f}$ and the single-index eigenvalue shocks derived in the last section. Together, the table and graph suggest that the expected return shock is not likely to take over the pricing role of the previous level shocks. The expected return shock has almost no correlation with the previous level shock. It is well correlated with the previous slope shock, and so may be able to take on that shock's minor role in pricing. By contrast, the new level shock remaining after the expected return shock is almost perfectly correlated with the previous level shock, and so is likely to take over that shock's strong role in pricing.

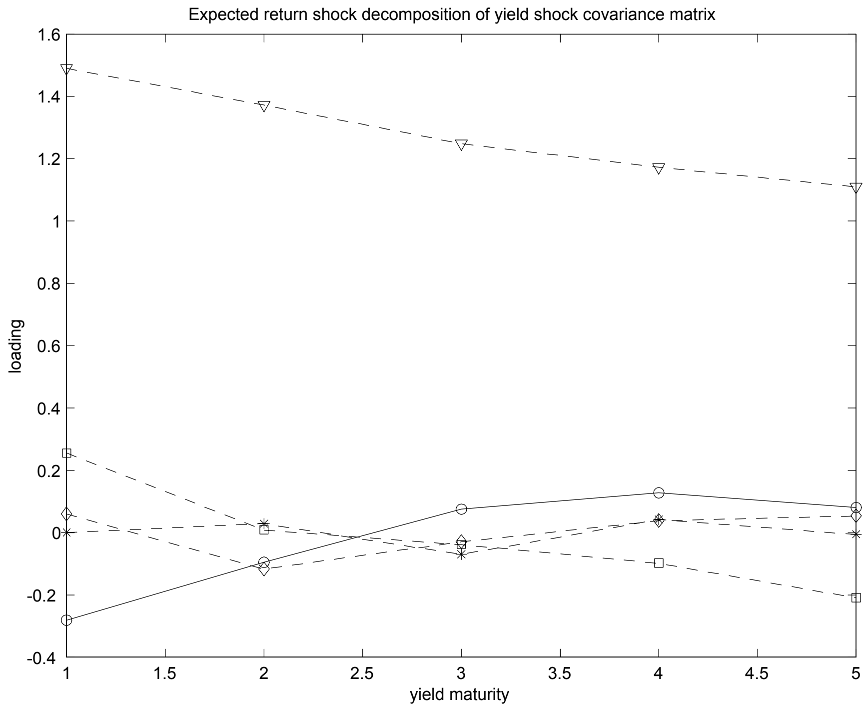

Figure 9: Yield shock covariance matrix starting with the expected return shock. The lines are the columns of Σ_y in the representation $y_{t+1} = \mu_y + \phi_y y_t + \Sigma_y \eta_{t+1}$, $E(\eta\eta') = I$. The first shock η is a shock to the expected return factor $\gamma' f$, normalized to unit variance. The remaining shocks are based on an eigenvalue decomposition of the variance covariance matrix, after removing the effect of the expected return shock. The solid line presents the first column of Σ_y, the response of yields at each maturity to an expected return shock.

		Eigenvalue shocks					Single index shocks	
		level	slope	y2	y3	y4	level	slope
Exp. return	$\gamma' f$	-0.03	0.69	-0.06	0.62	0.36	0.04	0.77
shocks	level	0.9996	0.02	-0.00	0.02	0.00	0.999	0.05

Table 14. Correlation of expected return shocks with eigenvalue decomposition shocks and shocks from the single-index model. In each case we start with a restricted yield VAR, $y_t = \mu_y + \phi_y y_{t-1} + v_t$, and express the shocks as $v_t = \Sigma\eta_t$; $E(\eta\eta') = I$. The eigenvalue shocks are identified by an eigenvalue decomposition of the regression error covariance matrix, $Q\Lambda Q' = E(vv')$, $\Sigma = Q\Lambda^{\frac{1}{2}}$. The expected return shocks are identified so the first shock is the standardized innovation to expected returns, $\eta_t^{\gamma' f} = k\left[\gamma' f_t - E_{t-1}\left(\gamma' f_t\right)\right]$, k chosen so $\sigma^2(\eta^{\gamma' f}) = 1$. The remaining shocks are identified by an eigenvalue decomposition (we only show the first such shock, in the "level" row.) The single index shocks are identified to exactly capture bond expected returns.

Table 15 presents the betas – how bond returns of each maturity respond to these shocks. As expected for a slope-shaped shock, the expected return shock generates a small positive return for short term bonds and a small negative return for long term bonds. However, the level shock has a much stronger effect on returns of all maturities.

What counts is betas times factor risk premia of course – the level shock might have no factor risk premium in this decomposition. The bottom half of Table 15 presents the factor risk premia. The risk premia associated with the expected return factor are small, and those associated with the level factor are large, and not much changed from Table 10.

Table 16 presents the bond expected return decomposition. The small spread in unconditional mean returns in the top panel is still almost entirely explained by covariance with the level shock. The expected return shock takes the place of the small curvature shock in Table 11. Most dramatically, when $\gamma' f$ is one standard deviation above its mean, the dramatic variation across bonds in conditional expected return is all driven by covariance with the level shock. The expected return shock contributes nothing.

In sum, this is a negative result. It is one worth exploring – innovations to expected returns are a natural candidate for priced risks. The level shock might have proxied for the expected return shock. Alas, the expected return shock is small, is nearly uncorrelated with the level shock, and accounts for none of the risk premium.

	$\gamma' f$	level			
	$100 \times C$: how returns load on shocks				
$hprx_{t+1}^{(2)}$	0.28	-1.49	-0.26	-0.06	-0.00
$hprx_{t+1}^{(3)}$	0.19	-2.75	-0.02	0.23	-0.06
$hprx_{t+1}^{(4)}$	-0.22	-3.75	0.11	0.09	0.21
$hprx_{t+1}^{(5)}$	-0.51	-4.69	0.39	-0.16	-0.17
	$\lambda_t = \lambda_0 + \delta\left(\gamma' f_t\right)$ factor risk premia				
λ_0	0.38	1.38	-0.02	-0.05	-0.65
δ	-4.3	-31.6	-3.25	4.10	18.6
$\lambda_0 + \delta E\left(\gamma' f_t\right)$	0.17	-0.20	-0.19	0.16	0.28
$\lambda_0 + \delta\left[E\left(\gamma' f_t\right) + \sigma\left(\gamma' f_t\right)\right]$	0.06	-1.01	-0.27	0.26	0.76

Table 15. Factor risk premia in bond returns. The first factor is an innovation to the expected return state variable $\gamma' f$. The remaining factors come from an eigenvalue decomposition of the yield innovation variance covariance matrix.

	$\gamma' f$	level				$E(hprx) + \frac{\sigma^2}{2}$
	Er from $\lambda_0 + \delta E(\gamma' f)$					
$hprx^{(2)}_{t+1}$	0.05	0.30	0.05	-0.01	-0.00	0.39
$hprx^{(3)}_{t+1}$	0.03	0.56	0.00	0.04	-0.02	0.61
$hprx^{(4)}_{t+1}$	-0.04	0.76	-0.02	0.01	0.06	0.78
$hprx^{(5)}_{t+1}$	-0.09	0.95	-0.07	-0.02	-0.05	0.72
	Effect on Er of 1σ increase in $\gamma' f$					
$hprx^{(2)}_{t+1}$	-0.02	1.21	0.02	-0.01	-0.00	1.19
$hprx^{(3)}_{t+1}$	-0.02	2.22	0.00	0.02	-0.03	2.20
$hprx^{(4)}_{t+1}$	0.02	3.03	-0.01	0.01	0.10	3.16
$hprx^{(5)}_{t+1}$	0.05	3.79	-0.03	-0.02	-0.08	3.72

Table 16. Decomposition of expected excess bond returns. The units are percent annual returns. The first five columns give the average returns due to each factor. In the top panel this is $\lambda_j \times C_{ij}$, where λ and C are given in Table 11. In the bottom panel this is the extra average return when $\gamma' f$ is one standard deviation above its mean. This is calculated as $\delta \times C_{ij} \times \sigma(\gamma' f)$. The last column gives the total amount of expected return there is to explain. The top panel gives the unconditional average return. The bottom panel gives the effect on expected return of a one standard deviation rise in $\gamma' f$. The $E(hprx) + \frac{\sigma^2}{2}$ column is equal to the row sum of the first five columns.

3.4.1 Details of the calculation

Denote the yield VAR

$$y_t = \mu_y + \phi_y y_{t-1} + v_t.$$

¿From $f_t^{(n-1\to n)} = -(n-1)y_t^{(n-1)} + ny_t^{(n)}$, we can recover forward rates from yields by $f_t = Dy_t$, with D a matrix of numbers given by equation (50) in the Appendix. Thus, the shock to $\gamma' f$ is $\gamma' D v_{t+1}$.We normalize to unit variance shocks, so the shock to $\gamma' f$ is

$$\eta_{t+1}^{\gamma' f} = \frac{\gamma' D}{\sqrt{\gamma' D E(vv') D' \gamma}} v_{t+1}$$

We find how each yield regression shock v_t is affected by the expected return shock $\eta_{t+1}^{\gamma' f}$ by finding the regression coefficients

$$\begin{aligned} v_{t+1} &= b\eta_{t+1}^{\gamma' f} + \varepsilon_{t+1} \\ b &= E(v_{t+1}\eta_{t+1}^{\gamma' f}) = \frac{E(vv')D'\gamma}{\sqrt{\gamma' D E(vv') D' \gamma}}. \end{aligned}$$

This calculation gives us the first column of Σ_y in our desired orthogonalization $v_{t+1} = \Sigma_y \eta_{t+1}$. We eigenvalue decompose what's left over. What's left over is

$$v_{t+1} - b\eta_{t+1}^{\gamma' f} = \left(I - \frac{E(vv')D'\gamma\gamma' D}{\gamma' DE(vv')D'\gamma} \right) v_{t+1},$$

and its covariance matrix is

$$\begin{aligned} E\left[\left(v_{t+1} - b\eta_{t+1}^{\gamma' f} \right) \left(v_{t+1} - b\eta_{t+1}^{\gamma' f} \right)' \right] &= \left(I - \frac{E(vv')D'\gamma\gamma' D}{\gamma' DE(vv')D'\gamma} \right) E(vv') \left(I - \frac{D'\gamma\gamma' DE(vv')}{\gamma' DE(vv')D'\gamma} \right) \\ &= E(vv') - \frac{E(vv')D'\gamma\gamma' DE(vv')}{\gamma' DE(vv')D'\gamma}. \end{aligned}$$

We decompose this matrix as $Q\Lambda Q'$. The second shock then captures the most variance subject to the constraint that it is orthogonal to $\eta^{\gamma' f}$; the third captures the most variance orthogonal to the first two and so forth. If the $\gamma' f$ shock happened to be the first eigenvalue of the covariance matrix, the remaining shocks identified in this way will be the remaining eigenvalues just as before.

3.5 Inflation and monetary policy shocks

The "level" shock in yields which seems to account for the bulk of the time-varying bond risk premium, is the reflection of some underlying macroeconomic shock. The question is, what is that shock? We have seen that it is not a shock to expected bond returns.

Since we are studying nominal bonds, inflation is a natural candidate. One of our biggest questions is, do you earn bond returns for holding *inflation* risk, or for the risk that *real* interest rates change? Has that premium changed as inflation volatility has declined so dramatically since the late 1970s? To that end, we add inflation to our VAR consisting of all Fama-Bliss yields and inflation risk.

Of course, inflation is fundamentally only a change of units. Thus inflation shocks will only be priced if the economy is non-neutral so that inflation shocks have real effects, or if inflation shocks are correlated with some underlying real shock. For example the government may choose inflation in response to real shocks, as a way of implementing a state-contingent default of nominal government debt.

To identify inflation shocks, we order them first, specifying that yields do respond contemporaneously to inflation shocks, but inflation does not respond contemporaneously to yields. We can give this identification a structural interpretation if we believe that goods prices move more slowly than bond prices. If this is true, the contemporaneous correlation of yield and inflation shocks is due to news moving from inflation to yields and not vice versa. We also choose this identification in order to assign as much of the risk premium to inflation as possible. One can always explain returns with returns, and shocks orthogonal to returns are not needed given the return shocks. Thus, if we orthogonalize inflation shocks last they get zero risk premium. By orthogonalizing inflation shocks first, we give them their best shot at explaining bond risk premia.

Monetary policy shocks are another natural candidate for macroeconomic shocks underlying the term structure. The federal funds rate is largely thought to be under the control of the Federal reserve. Furthermore, most analysts think that the Fed has the ability to control *real* short term rates, which are more likely to induce real holding period excess returns.

To identify monetary policy shocks, we add federal funds to the yield VAR, ordered first. Most monetary VARs define monetary policy shocks by forecasting federal funds rates with a smorgasbord of current and lagged macroeconomic variables and no interest rates. (For example, Christiano Eichenbaum and Evans 1999.) However, it seems important for our purposes to include yield information. If the bond markets know a fed funds change is coming, then it really isn't a shock, even if the change is unpredictable by macroeconomic variables. We could include a series of shocks identified by some other procedure rather than just federal funds. Alas, the policy shocks recovered from a detailed analysis in Piazzesi (2001) do not cover a long enough time period for us to use them.

The central issue in monetary VARs is orthogonalization; in our case whether a contemporaneous unpredictable movement in yields and the federal funds rate results from a policy shock that affects yields, or from a change in longer yields that causes the Fed to respond with a funds rate change. Fortunately for our purposes we do not have to take a stand on this issue. We orthogonalize the funds rate shock first, assigning all contemporaneous correlation to the funds rate shock. As with inflation, this gives monetary policy shocks their best chance to explain bond risk premia.

Let z_t denote either inflation, the log of one-year growth in the consumer price index, or the federal funds rate in month t. We run a VAR with z_t and yields, i.e. with

$$x_t = \left[\begin{array}{cccccc} z_t & y_t^{(1)} & y_t^{(2)} & y_t^{(3)} & y_t^{(4)} & y_t^{(5)} \end{array} \right],$$

we run

$$x_t = \phi x_{t-1} + v_t.$$

As usual, we use an annual horizon and overlapping monthly observations.

Inflation is not of much marginal use in forecasting bond yields or returns as reported in Table 10. The federal funds rate is even more useless, raising the R^2 in forecasting average returns from 0.397 to 0.398. For this reason, we keep the expected return model $E_t\left(hprx_{t+1}^{(n)}\right) = a + b\gamma' f_t$ intact rather than augment it with inflation or the funds rate. (Interestingly, the yields do help to forecast inflation and the funds rate.)

We define the inflation or monetary policy shock as the regression error of the z VAR equation, standardized to unit variance,

$$\eta_{t+1}^{z} = v_{t+1}^{z}/\sigma(v_{t+1}^{z}),\ z = ff \text{ or } \pi.$$

Table 17 shows how inflation and federal funds shocks affect yields. The pattern in each case is very much that of a "level" shock. This is a hopeful sign, as the level shock in yields accounted for most of the risk premium.

However, the correlation of the inflation shock with the "level" shock calculated from the eigenvalue decomposition of the yield-only VAR is only 0.46, foreshadowing that it will be an imperfect proxy for that shock. The Fed funds shock by contrast has an impressive 0.83 correlation with the level shock, as well as a decent -0.35 correlation with the slope shock, foreshadowing that it may well be able to stand in for those shocks in pricing.

	Response to π or ff shock						Correlation with y shocks	
	z_{t+1}	$y^{(1)}_{t+1}$	$y^{(2)}_{t+1}$	$y^{(3)}_{t+1}$	$y^{(4)}_{t+1}$	$y^{(5)}_{t+1}$	level	slope
$z = \pi$	1.65	0.77	0.64	0.54	0.50	0.47	0.46	-0.24
$z = ff$	2.06	1.35	1.15	0.99	0.91	0.84	0.83	-0.35

Table 17. Response of inflation, federal funds, and yields to a unit-variance shock to either inflation or federal funds, and correlation of inflation and fed funds shocks with level and slope shocks. We start with a VAR of inflation or fed funds and bond yields $x_t = \phi x_{t-1} + v_t$, $x_t = \left[\begin{array}{cc} z_t & y_t' \end{array}\right]'$, $z = \phi$ or ff. The z shock is the regression error, standardized to unit variance. The response of all shocks to the z shock is then given by $E(v_t v_t^z)/\sigma(v_t^z)$. The correlations give the correlation of z shocks with the level and slope shocks identified from an eigenvalue decomposition of the yield VAR shock covariance matrix.

The covariance of returns with the z shock is given by $cov(hprx^{(n)}_{t+1}, \eta^z_{t+1}) = (n-1)cov(v^{(n-1)}_{t+1}, \eta^z_{t+1})$. Now we have all the ingredients to calculate the inflation risk premia as in (15)-(17),

$$\begin{aligned}\lambda_0 &= cov(hprx_{t+1}, \eta^z_{t+1})' cov_t(hprx_{t+1}) \left(a + \frac{1}{2}\sigma^2_t(hprx_{t+1}) \right) \\ \delta &= cov(hprx_{t+1}, \eta^z_{t+1})' cov_t(hprx_{t+1}) b.\end{aligned}$$

Or estimates are, for the inflation shock, $\lambda_0 = 0.43$, $\delta = -12.3$ and for the federal funds shock, larger values $\lambda_0 = 0.92$, $\delta = -23.0$. As before, we bring these numbers to life by seeing how much of the conditional and unconditional expected return is accounted for by the inflation and federal funds shock factor risk premium.

Table 18 presents the results. The first row shows how returns load on the inflation and federal funds shocks. As with a level shock, longer and longer bonds have larger negative loadings on the inflation shock. The fed funds shock has larger responses, because it is better correlated with the level factor that accounts for the bulk of yield variation.

	$hprx_{t+1}^{(2)}$	$hprx_{t+1}^{(3)}$	$hprx_{t+1}^{(4)}$	$hprx_{t+1}^{(5)}$
C: how returns load on η^{π}	-0.77	-1.28	-1.63	-1.98
C: how returns load on η^{ff}	-1.35	-2.31	-2.99	-3.64
1. At $\gamma' f = E(\gamma' f)$:				
Target: $E(hprx) + \frac{1}{2}\sigma^2$	0.39	0.61	0.78	0.72
Er from π shock, $C\left[\lambda_0 + \delta E(\gamma' f)\right]$	0.08	0.14	0.18	0.22
Er from ff shock, $C\left[\lambda_0 + \delta E(\gamma' f)\right]$	0.31	0.54	0.70	0.85
2. Effect on Er of 1σ larger $\gamma' f$				
Target: $E_t hprx + \frac{1}{2}\sigma^2$	1.19	2.20	3.16	3.72
$E_t r$ from π shock, $C\delta\sigma(\gamma' f)$	0.21	0.35	0.45	0.54
$E_t r$ from ff shock, $C\delta\sigma(\gamma' f)$	1.11	1.90	2.46	2.99

Table 18. Contributions of inflation shock and federal funds shock to expected bond returns.

The unconditional (or conditional at $\gamma' f = E(\gamma' f)$) mean returns to be explained rise from 0.39 % to 0.72 % across maturity, in the "Target" row. The expected returns explained by the inflation premium are of the right sign, rising from 0.08 to 0.22%, but roughly a fourth too small across the board. Risk premia on bond shocks orthogonal to inflation (not shown) account for the rest, and again, the most important factor is a "level" factor. The federal funds shock does a much better job, however, neatly explaining the vast majority of the unconditional risk premium

When $\gamma' f$ rises by one standard deviation, conditional expected returns rise to much higher values, from 1.19% to 3.72%. The time-varying component of the inflation premium explains 0.21 to 0.54 percentage points of this rise. Again, the sign is right, but the magnitude is disappointingly low. Again, the "level" factor orthogonal to inflation (not shown) still explains most of the time-varying expected return. However, the federal funds shock is again very successful, explaining the majority of the time-varying expected return.

4 Why is this news?

We have found a single factor in bond yields, $\gamma' f$ that captures all of the substantial time-varying market price of risk. In the hundreds of studies of the term structure, how did this factor escape notice? We offer two explanations: 1) Most studies of the term structure focus on a few factors that explain the bulk of movements in bond yields or prices. The return forecasting factor does not appear in these exercises. You can ignore the return forecasting factor, imposing the expectations hypothesis or imposing much weaker expected return models, and you can describe bond *prices* or *yields* with great accuracy, though you will substantially miss bond *expected excess returns.* 2) Most authors study bond returns at a monthly horizon. However, the monthly data are not well described by an AR(1). Additional lags matter, in a way suggestive of i.i.d. measurement error. Monthly models raised to the 12th power completely miss the return forecastability. To see return forecastability at an annual horizon, you have to either look directly at the annual horizon, as we have, or consider non-Markovian representations of the bond data.

4.1 Yield factors do not capture return predictability

The simplest way to construct a factor decomposition is through an eigenvalue decomposition of the yield variance covariance matrix.

$$cov(yy') = Q\Lambda Q'$$

Then, we can define factors x_t by

$$\begin{aligned} y_t &= Qx_t \\ cov(x_t x_t') &= \Lambda^{\frac{1}{2}} \end{aligned}$$

This construction maximizes the variance of yields explained by each orthogonal factor in turn. Table 19 presents our factor decomposition for yields. (This is a factor decomposition of *yields*, not of *shocks*, based on the unconditional covariance matrix of yields, not of the shock covariance matrix as above.) Our calculation is based on the yield covariance matrix implied by the restricted yield VAR; there is an exact *one*-factor model $\gamma' f$ driving expected returns. (Since the restriction is not rejected, the results using an unrestricted yield VAR are of course very similar.)

Table 19 shows a familiar pattern. The "level" factor which moves all yields together has by far the largest variance. As one way to characterize the success of potential restricted-factor models, the "rmse" row of Table 19 calculates the root mean squared error that would result if you modeled yields as depending only the largest k factors. We can find market prices of risk λ_t so that each of these exercises also forms an affine model, based on the reduced-factor VAR representation, that prices the factors (linear combinations of yields) by construction.

The root mean squared error of all yields is 3.09%. If you use a model in which all yields are driven only by the level factor, you have only a 0.49% root mean squared error left. A second "slope" factor moves short rates up and long rates down, but has one seventh the standard deviation. The level and slope factors together are a very successful model, leaving only 19 basis points of root mean squared error. Then there are three small idiosyncratic factors. We have labeled them by the pattern of their weights. For example, the "2-4,5" factor raises the 2 year yield and lowers the 4,5 year yield. (Since they are orthogonal to the level factor, the remaining factors are essentially zero-cost portfolios). One could add the 2-4,5 factor and the 3-4,5 factor to obtain a "curvature" factor long 2, 3 and short 4, 5.

	level	slope	2-4,5	3-4,5	4-5
$\Lambda^{\frac{1}{2}}$	6.82	1.01	0.36	0.20	0.10
rmse (total=3.09)	0.49	0.19	0.10	0.04	0.00
$y^{(1)}$	0.48	0.82	-0.28	-0.09	-0.08
$y^{(2)}$	0.45	0.03	0.88	-0.15	0.03
$y^{(3)}$	0.44	-0.19	0.06	0.88	-0.04
$y^{(4)}$	0.44	-0.31	-0.28	-0.28	0.74
$y^{(5)}$	0.43	-0.43	-0.24	-0.36	-0.66

Table 19. Eigenvalue decomposition of yield variance-covariance matrix $Q\Lambda Q' = var(yy')$ from the restricted VAR. We start with the holding period return regression restricted to a single factor, $hprx_{t+1} = a + b\,(\gamma' f_t) + v^h_{t+1}$. We find the implied coefficients in a yield VAR, $y_{t+1} = \mu_y + \phi_y y_t + v^y_{t+1}$. Then, we compute $var(yy') = E(v^y v^{y\prime}) + \phi_y E(v^y v^{y\prime})\phi_y' + \phi_y^2 E(v^y v^{y\prime})\phi_y'^2 + ...$, and finally take the eigenvalue decomposition $var(yy') = Q\Lambda Q'$. The row market $\Lambda^{\frac{1}{2}}$ gives the square root of the eigenvalues. The row marked rmse gives the root mean squared error of a model that uses only the first k factors. It is calculated as the square root of the mean of the diagonal elements of $Q\Lambda^k Q'$, where Λ^k includes only the first k elements of the eigenvalue matrix Λ. The rows labeled $y^{(1)}...y^{(5)}$ give the eigenvectors Q which tells us how each yield loads on each shock.

Now, suppose you were faced with these yields. The most natural thing to do would be to summarize yields by a 2 factor model, with "level" and "slope" factors, and ignore the idiosyncratics. This would only result in a 19 basis point root mean squared error. In formal estimation, you might treat that as measurement error. At most, you'd consider a three factor model, leaving a 10 basis point root mean squared error.

If you were to focus on the big factors in this way, you would miss a substantial part of the predictability of returns. Table 20 demonstrates this point. The first row of Table 20 presents a regression of average holding period returns $\frac{1}{4}\sum_{n=2}^{5} hprx^{(n)}_{t+1}$ on the first, "level" factor. Even though this is by far the most important factor for yields, the return forecasting regression yields a miserable 0.07 R^2. The last column shows why. Here, we run a regression of the return-forecasting factor $\gamma' f_t$ on the level factor in yields. The coefficient in this regression is the same as for holding period returns, so we do not show it separately. The $R^2_{\gamma' f}$ captures how well we can approximate the $\gamma' f$ return forecasting factor with the level factor x_t^{level}. The 0.17 R^2 value says the answer is, "not well at all." This is not surprising. The return forecasting factor $\gamma' f$ is a tent-shaped function of forward rates. It is nearly orthogonal to a level factor in yields.

The second row of Table 20 regresses average excess returns and the return forecasting factor on the "level" and "slope" factors recovered from the yield variance-covariance matrix. We had found that $\gamma' f$ is correlated with slope measures of the term structure, and term structure slope variables are the classic return forecasters in the literature. We see some success in the regression coefficients. The -2.8 is much larger than 0.16, so the slope factor is far more important in forecasting returns. The return forecasting R^2 rises to 0.28, and the R^2 in explaining $\gamma' f$ rises to 0.71.

Still, the level and slope factors, which together leave only a 20 basis point pricing error, produce only a 0.28 R^2 in forecasting returns. As seen by the 0.71 $R^2_{\gamma' f}$ a substantial part of the return-forecasting factor is orthogonal to the level and slope factors in yields. Given the curved shape of γ, that is not too surprising.

Continuing this way, even the 2-4,5 and 3-4,5 idiosyncratics don't help all that much. To get the full 0.40 R^2 of the return forecasting factor, we have to look at *all* the yield factors. Even ignoring the last yield factor, leaving only a ridiculously small 4 basis point rmse pricing error, gives us only a 0.31 R^2!

level	slope	2-4,5	3-4,5	4-5	R^2_{hprx}	$R^2_{\gamma' f}$
0.20					0.07	0.17
0.16	-2.8				0.28	0.71
0.17	-2.7	0.61			0.28	0.72
0.16	-2.9	3.2	9.9		0.31	0.79
0.12	-2.8	3.9	6.7	-15.5	0.40	1.00

Table 20. Regressions of average excess returns $\frac{1}{4}\sum_{n=2}^{5} hprx_{t+1}^{(n)}$ and of the return forecasting factor $\gamma' f_t$ on factors x_t recovered from an eigenvalue decomposition of the variance-covariance matrix of yields in the restricted VAR. The coefficients are the same for both regressions. R^2_{hprx} gives the R^2 for the regression that forecasts returns with the factors x_t. $R^2_{\gamma' f}$ gives the R^2 in the regression of $\gamma' f_t$ on the factors. The factors x_t are defined and characterized in Table 19.

4.1.1 An expected return - expectations decomposition

The yield decomposition suggests that expected returns are not terribly important for understanding *yields*. The expectations hypothesis might give a good account of yields themselves, and getting market prices of risk wrong might have little impact on average pricing errors.

To quantify this impression, we construct a different factor model for yields, with $\gamma' f$ as the first factor. Then, we can see how well a restricted factor model behaves that ignores the $\gamma' f$ factor, and thus sets all expected returns to a constant.

Figure 10 presents the resulting decomposition. The lines marked "expectations hypothesis" are movements in yields driven by all the other factors, orthogonal to $\gamma' f$. Expected returns are, by construction, *constant* in this component of yields. The lines marked "expected returns" are movements in yields driven by the $\gamma' f$ factor. The two lines add up to the actual yields. The figure shows that the vast bulk of yield variation is quite well captured by the expectations hypothesis. Long yields are high when expected future short yields are high.

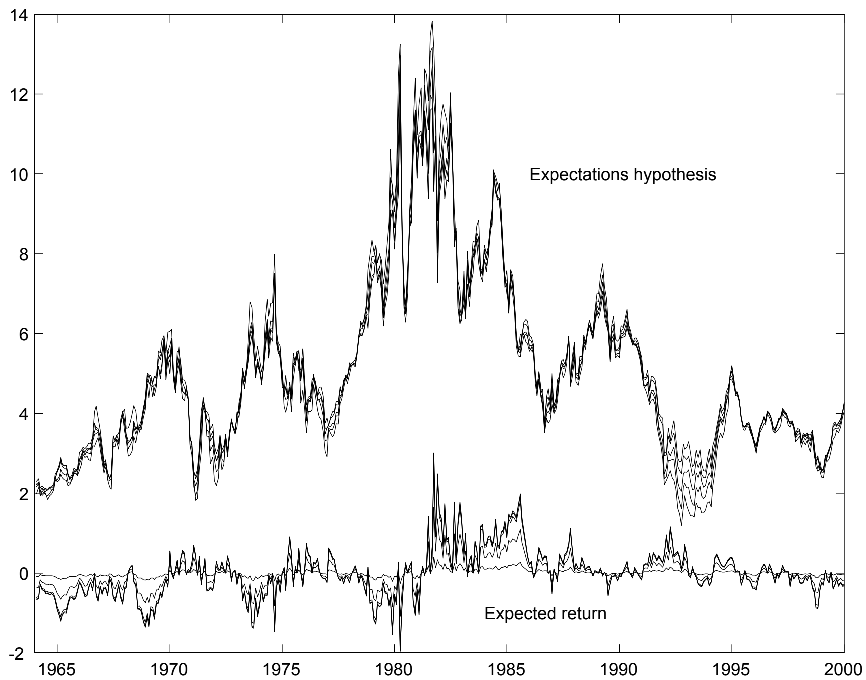

Figure 10: Decomposition of yields into a component with constant expected returns ('expectations hypothesis') and a component due to time-varying returns. We construct a factor model for yields $y_t = \Sigma x_t$ with $E(xx')$ diagnonal and the yield forecasting factor $\gamma' f$ as the first component, $\gamma' f_t = x_t(1)$. The "expected return" line is then $\Sigma(:,1)\,(\gamma' f_t)$ and the "expectations hypothesis" term is $\Sigma(:,2:5) * x_t(2:5)$.

The first row of Table 21 confirms this impression by computing root mean square pricing errors of various reduced factor representations. The total root mean square error is 3.09%. A one-factor model using the expected return factor is a miserable failure, leaving a 3.00% rmse pricing error. The expectations hypothesis factor model in the "all others" column, by contrast, leaves a 0.75% pricing error. As usual, the first two level and slope factors are the most important.

	total	just $\gamma' f$	All others	just level	level and slope
yields	3.09	3.00	0.75	0.82	0.77
yield spread	1.13	0.88	0.71	1.05	0.73

Table 21. Root mean square errors of yield factor models. "total" gives the root mean square error of yields, i.e. a factor model consisting of a constant. "just $\gamma' f$" uses only the expected return factor. "All others" uses all other factors, orthogonal to the expected return factor. "Just level" is a one-factor model using the first eigenvalue factor after the expected return factor; the loadings on this factor have a level pattern. "Level and slope" use the first two eigenvalue factors after the expected return factor.

We might suspect that the expectations hypothesis describes *yields* well, but misses *yield spreads.* After all, yield spreads were the central forecasting variable until $\gamma' f$ came along, and are correlated with the $\gamma' f$ factor. Figure 11 and the second row of Table 21 confirm this impression. The figure plots the spread between n year yields and one year yields, using the same data as in Figure 10. About half of the yield spread is still due to pure expectations hypothesis movements. Interestingly that component seems to account better for a slope in the yield curve. For example, in 1980 and 1987, the expectations hypothesis components of the spread vary across maturity, while the expected return components are similar for all maturities. In Table 21, the factor model using only $\gamma' f$ leaves an 0.88% rmse error on yield spreads, while all the other factors leave a 0.71 rmse pricing error.

4.1.2 Calculation

We proceed as we did in constructing the $\gamma' f$ factor in the innovation covariance matrix, but applied to the yield covariance matrix. Since $f = Dy$, our first factor is

$$x_t^{\gamma' f} = \gamma' \tilde{f}_t = \frac{\gamma' D}{\sqrt{\gamma' D cov(yy') D' \gamma}} \tilde{y}_t$$

where $\tilde{y} = y - E(y)$. We have normalized to a unit variance We find how each yield loads on this factor by finding the regression coefficients,

$$\begin{aligned} \tilde{y}_t &= b x_t^{\gamma' f} + \varepsilon_t \\ b &= \frac{cov(\tilde{y}_t, x_t^{\gamma' f})}{var(x_t^{\gamma' f 2})} = \frac{cov(yy') D' \gamma}{\sqrt{\gamma' D cov(yy') D' \gamma}} \end{aligned}$$

This calculation gives us the first column in our desired factor decomposition $y_t = \Sigma x_t$ We eigenvalue decompose what's left over. What's left over is

$$\tilde{y}_t - b x_t^{\gamma' f} = \left(I - \frac{cov(yy') D' \gamma \gamma' D}{\gamma' D cov(yy') D' \gamma} \right) \tilde{y}_t$$

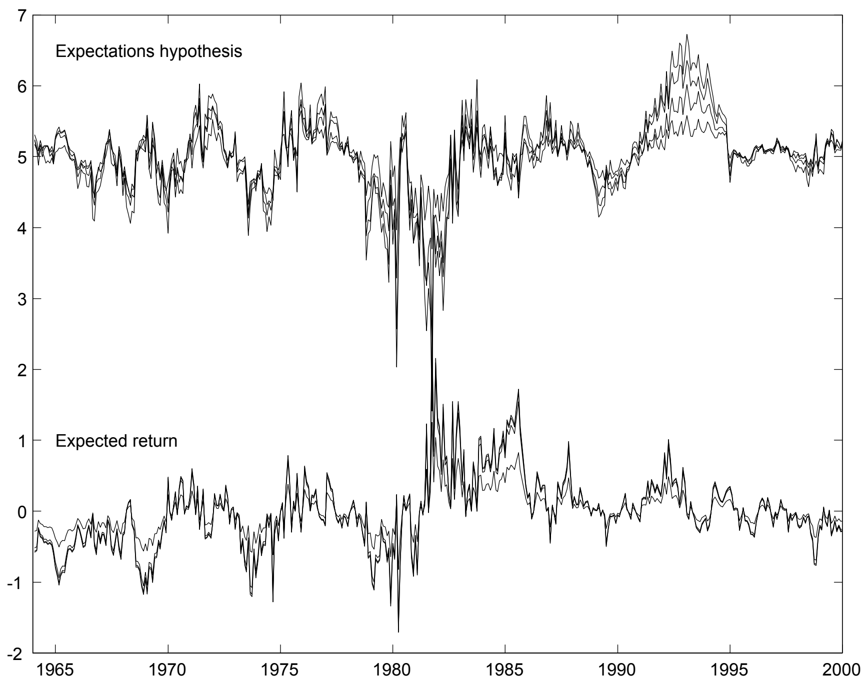

Figure 11: Decomposition of yield spreads into expected return and expectations hypothesis components. Each spread is taken over the one year rate, i.e. each line is $y_t^{(n)} - y_t^{(1)}$.

and its covariance matrix is

$$
\begin{aligned}
cov\left(\tilde{y}_t - bx_t^{\gamma' f}\right) &= \left(I - \frac{cov(yy')D'\gamma\gamma' D}{\gamma' Dcov(yy')D'\gamma}\right) cov(yy') \left(I - \frac{D'\gamma\gamma' Dcov(yy')}{\gamma' Dcov(yy')D'\gamma}\right) \\
&= cov(yy') - \frac{cov(yy')D'\gamma\gamma' Dcov(yy')}{\gamma' Dcov(yy')D'\gamma}
\end{aligned}
$$

We eigenvalue decompose this matrix to find the remaining factors.

4.1.3 Reconciling yield and expected-return forecasts

These results seem very strange. We made our computations using the restricted yield VAR. *By construction*, the yields satisfy an *exact* one-factor model for expected returns. If you start with the restricted yield VAR and find the implied regression of returns on yields (or forward rates) using $hprx_{t+1}^{(n)} = -(n-1)y_{t+1}^{(n-1)} + ny_t^{(n)} - y_t^{(1)}$, the implied regression of returns on yields satisfies $E_t hprx_{t+1} = a + b(\gamma' f_t) = a + b(\gamma' Dy_t)$ where D converts from yields to forward rates. $\gamma' Dy_t$ is a scalar, the single factor that drives expected returns. How can a *single* factor model for expected returns not show up in the dominant yield factors? Write the yield VAR

$$
y_{t+1} = \mu + \phi y_t + \Sigma \eta_{t+1}.
$$

The single factor for expected returns must imply strong restrictions on ϕ. How can this not show up in the yield factors?

There are three parts to the answer. First, factor structure in the shock covariance matrix $\Sigma\Sigma'$ is also vital to factor structure in yields. Second, if ϕ has a single factor structure, the loadings on a single factor rather than the factor itself will drive ϕ. Third, a single factor model for expected returns does not give a single factor model for ϕ in the first place.

1. The yield covariance matrix is driven by the shock covariance matrix $\Sigma\Sigma'$ as much as by ϕ. The yield variance covariance matrix is given by

$$cov(y_t, y_t') = \Sigma\Sigma' + \phi\Sigma\Sigma'\phi' + \phi^2\Sigma\Sigma'\phi'^2 + ... \tag{32}$$

Factor structure in the yield covariance matrix must come from factor structure in the innovations Σ *as well as* factor structure in the transition matrix ϕ. If ϕ has a factor structure but Σ does not, the first term will still give us a full rank forward covariance matrix. If Σ has a factor structure but ϕ does not, we have a stochastically singular system, but $var(ff')$ will typically be full rank.

We have already seen in Figure 15 that the yield VAR innovation covariance matrix has a very strong factor structure, with "level" "slope" and very small "idiosyncratic" shocks. Given this fact and (32), it's not surprising that the yield covariance matrix also has a very strong factor structure, with "level" "slope" and small idiosyncratic shocks, *no matter what* the factor structure of the transition matrix ϕ.

2. A single factor model for ϕ will not show up in factor structures for yields. As a very simple example to see this point, suppose $\phi = \beta\gamma'$, with β and γ both vectors. This is a single factor model for ϕ – the single portfolio $\gamma' y$ carries all forecasting information. Then

$$var(yy') = \Sigma\Sigma' + \beta\gamma'\Sigma\Sigma'\gamma\beta' + \beta\gamma'\beta\gamma'\Sigma\Sigma'\gamma\beta'\gamma\beta' + ...$$

To make the example even simpler, suppose that Σ is a vector – there is one "level" shock. Then $\gamma'\Sigma$ is a scalar and we can collapse the last term

$$var(yy') = \Sigma\Sigma' + \frac{(\gamma'\Sigma)^2}{1-(\gamma'\beta)^2}\beta\beta'.$$

In this case, the factor structure in yields is driven by the β *loadings* on the forecasting factor $\gamma' y$, and the covariance factor structure. It has *nothing* to do with the $\gamma' y$ factor that captures all forecasting information. In this case you would not recover $\gamma' y$ from the factors of the yield covariance matrix.

3. A single factor model for *returns* does not imply a single factor model for *yield* forecasts. Let's try. Starting with $E_t hprx_{t+1} = a + b\gamma' f_t$ we can write a single factor expected return model in terms of yields as $E_t hprx_{t+1} = a + b\alpha' y_t$ where $\alpha' = \gamma' D$ and D is the matrix that changes from forward rates to yields, (50) in the appendix. Now, using the definition of holding period return $hprx_{t+1}^{(n)} = -(n-1)y_{t+1}^{(n-1)} + ny_t^{(n)} - y_t^{(1)}$ we

can recover the yield forecast ϕ_y. The Appendix goes through the algebra. The answer ((52) in the Appendix) is that the first four rows of ϕ_y are given by

$$\phi_y(1:4,:) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & \frac{1}{4} \end{bmatrix} \left(\begin{bmatrix} -1 & 2 & 0 & 0 & 0 \\ -1 & 0 & 3 & 0 & 0 \\ -1 & 0 & 0 & 4 & 0 \\ -1 & 0 & 0 & 0 & 5 \end{bmatrix} - b\alpha' \right)$$

As you can see, even though the one factor structure $b\alpha'$ is hiding in there, it is subtracted from another matrix (that matrix adds $y^{(1)}$ back to the excess return) and the result no longer has a single factor structure.

4.1.4 Moral of the story

If yields or forward rates really do follow an *exact* factor structure then all state variables including $\gamma' f$ have to be functions of that exact factor structure. However, an important state variable like $\gamma' f$ can well be hidden in the small idiosyncratic factors that are often dismissed as minor specification or measurement errors. A factor like $\gamma' f$ can be the only important factor for explaining expected returns, and yet explain almost *none* of the variance of yields and forward rates. $\gamma' f$ has not been noticed before, because 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 captures the expected return facts in this data should include the level and slope factors *and* $\gamma' f_t$, even though inclusion of the latter will do almost nothing to reduce pricing errors.

4.2 A monthly model raised to the 12th power does not work

Most term structure estimates focus on a short horizon and then calculate implied values of longer-horizon statistics. For example, given that our data are monthly, the most natural approach is to estimate a monthly VAR and find implied values of annual statistics rather than follow our course, estimating annual VARs with overlapping monthly observations. A monthly VAR would have the additional advantage that we could include shorter maturity bond data.

It is also natural to consider today's yields or forward rates as sufficient state variables in the term structure, and thus that a VAR(1) representation for yields should be sufficient. If today's forward rate is the expected future spot rate, then a full set of forward rates is a sufficient set of state variables for the term structure; neither other variables nor additional lags can help. Measurement error, missing maturities or risk premia can overturn this hypothesis of course, so whether a VAR(1) will be sufficient is an empirical question. Still, the vast majority of term structure models are Markovian; current yields are the state variables, additional monthly lags of yields are not used.

In our data, by contrast, a monthly AR(1) raised to the 12th power is disastrously different from direct annual estimates. If you follow the usual approach, you will drastically underestimate the predictability of annual returns, and you will not see the beautiful single-factor structure.

The reason is that the monthly data do not follow a VAR(1). We have already seen that additional monthly lags help to forecast returns. The pattern we see in the data strongly suggests an ARMA(1,1) induced by i.i.d. measurement error on top of an underlying AR(1). We can reconcile monthly and annual estimates with this specification. This finding suggests that we will have to include additional lags of yields as state variables in order to extend the analysis to yields on bonds of less than one year maturity, or to successfully match our data with a term structure model specified at a horizon less than a year.

4.2.1 One-year regressions implied by monthly regressions

Consider the VAR in forward rates

$$f_{t+1} = \mu + \phi f_t + v_{t+1}$$

(The forward rate VAR is equivalent to the yield VAR. We study forward rates because the pattern of the coefficients in $\gamma' f$ is so pretty, and thus it's easy to see how well forward rate VAR coefficients approximate the pattern.) Since our data are monthly, we should be able to run a monthly VAR

$$f_{t+\frac{1}{12}} = \mu_m + \phi_m f_t + v^m_{t+\frac{1}{12}}$$

and then estimate the annual VAR by

$$\phi = \phi_m^{12}.$$

There may be some small sample differences in point estimates, but we should not see any disastrous difference in results. Since return forecasting regression is implied by the forward rate VAR, we should be able to calculate it from the ϕ_m^{12} coefficients as well as from the ϕ coefficients.

Figure 12 contrasts the direct estimates of the forward rate VAR, ϕ, with the values implied by the monthly regression, ϕ_m^{12}. In the top, you can see some of the pattern that occurs when we regress holding period returns on forward rates. If you look hard, you can see some of that pattern in the bottom panel, though quite distorted. Most strikingly, though, all the coefficients in the bottom panel are too small.

We're really after the return regressions, so Figure 13 presents the coefficients in the regression of one year holding period returns on forward rates. The familiar coefficients

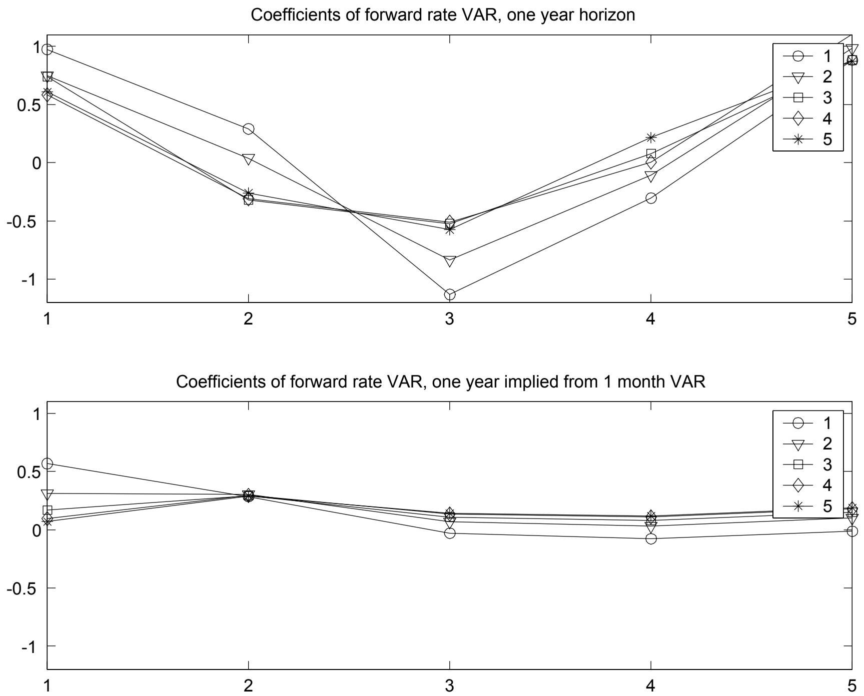

Figure 12: Coefficients in forward rate VAR at an annual horizon. Coefficients in the top panel are estimated directly, i.e.the lines are the rows of ϕ in $f_{t+1} = \mu + \phi f_t + v_{t+1}$. In the bottom panel, we estimate $f_{t+\frac{1}{12}} = \mu_m + \phi_m f_t + v^m_{t+\frac{1}{12}}$ and present the rows of ϕ_m^{12}.

in the top panel are calculated from the annual forward rate VAR ϕ, and are identical to direct estimates. The coefficients in the bottom panel are calculated from the implied forward rate VAR ϕ_m^{12}. The bottom panel is a real disaster. The coefficients implied from the monthly VAR are much too small. Most importantly, the pattern of the coefficients is distorted enough that you no longer see the beautiful one-factor structure that jumps out of the top panel.

Table 22 contrasts the R^2 in holding period return regression forecasts from the direct annual VAR, and from the monthly VAR raised to the 12th power. As you can see, the implied regression misses a great deal of the forecast power.

	$hprx^{(2)}$	$hprx^{(3)}$	$hprx^{(4)}$	$hprx^{(5)}$
R^2 from direct estimate, ϕ	0.38	0.39	0.41	0.38
R^2 from implied estimate, ϕ_m^{12}	0.28	0.28	0.29	0.25

Table 22. R^2 in return forecasting regressions $hprx^{(n)}_{t+1} = a^{(n)} + b^{(n)\prime} f_t + \varepsilon^{(n)}_{t+1}$. In the top row, the regression coefficients and R^2 are calculated from the annual VAR, ϕ. In the bottom row, the regression coefficients and R^2 are calculated from the monthly VAR raised to the 12th power, ϕ_m^{12}.

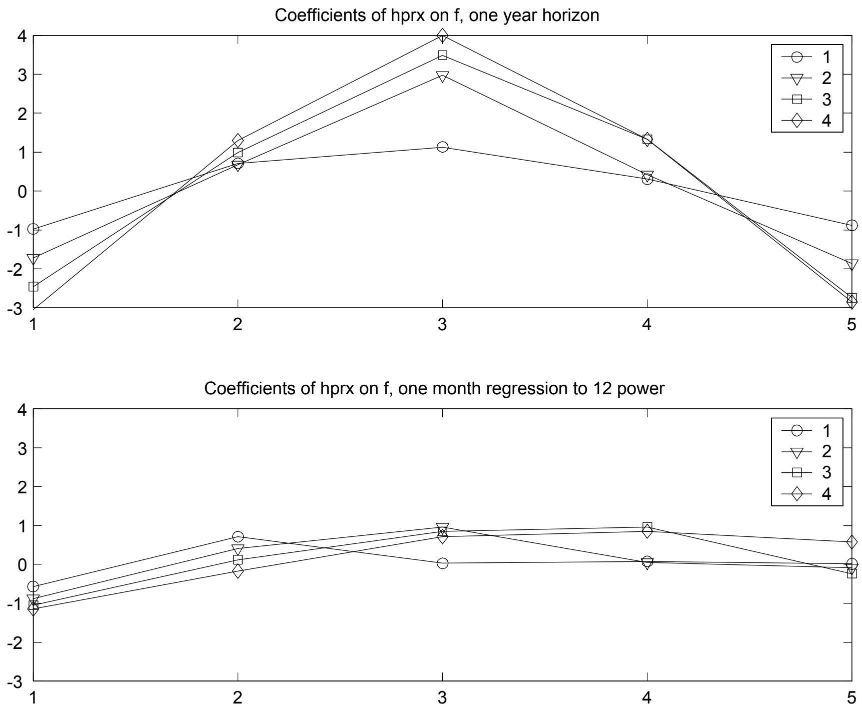

Figure 13: Coefficients in regressions of holding period excess returns $hprx_{t+1}^{(n)}$ on forward rates f_t. The coefficients in the top panel are constructed from the direct estimate of the annual VAR. They are identical to direct estimates of the regression coefficients. The coefficients in the bottom panel are constructed from the monthly forward rate VAR raised to the 12th power, ϕ_m^{12}.

4.2.2 A measurement error interpretation

This pattern in the data suggests an interpretation in terms of measurement error. Suppose a time series y_t really follows an AR(1), as most theoretical term structure models specify. However, suppose y_t is observed with i.i.d. measurement error ε_t. Then, the observed series x_t will follow an ARMA(1,1). Specifically, suppose

$$\begin{aligned} y_t &= \phi y_{t-1} + v_t \\ x_t &= y_t + \varepsilon_t. \end{aligned}$$

The autocovariances of the measured series x_t are

$$\begin{aligned} \sigma^2(x) &= \sigma^2(y) + \sigma^2(\varepsilon) \\ cov(x_t, x_{t-j}) &= \phi^j \sigma^2(y) \end{aligned}$$

The higher order autocorrelations fall off slower than the first autocorrelation $\rho_1 = \phi \frac{\sigma^2(y)}{\sigma^2(y)+\sigma^2(\varepsilon)}$ raised to the jth power. This is exactly what we see in comparing the one month and one year regressions. The sign of the coefficients is right, but the size is much too small.

Denote the ARMA(1,1) univariate representation of x_t by

$$x_t = \phi x_{t-1} + \eta_t - \theta \eta_{t-1}$$

The autoregressive representation of this ARMA(1,1) – what we see if we regress measured yields on lagged measured yields – is

$$x_t = (\phi - \theta)x_{t-1} + \theta(\phi - \theta)x_{t-2} + \theta^2(\phi - \theta)x_{t-3} + ... + \eta_t$$

This pattern is reminiscent of our explorations of additional lags in Table 8. The coefficients on additional lags had the same sign and pattern $(\phi - \theta)$, but declined with horizon.

The obvious next step is to estimate an ARMA(1,1) or more complex models with a richer specification of measurement error. However, our purpose in this paper is to characterize risk premia at an annual horizon, so we leave these calculations as an explanation why monthly estimates did not recover the return forecasting factor and a hint of the issues that must be confronted in order to extend the analysis to shorter maturities and hence, inevitably, a monthly estimation horizon.

5 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, $\gamma' f_t$. Then, expected excess returns on bonds of maturity n are $E_t(hprx_{t+1}^{(n)}) = a_n + b_n(\gamma_0 + \gamma' f_t)$, and even a_n is quite small.

Regressions of excess returns on this common factor show a much improved R^2. In contrast to Fama and Bliss' R^2 of about 17%, the R^2 on the common factor is about 40%, and 45% if we use a one-month moving average of the common factor $\gamma'(f_t + f_{t-1})$ to attenuate measurement error.

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

We relate time-varying expected returns to time-varying risk premia with constant covariances. Almost all of the spread in average bond returns is due to different covariances with the level shock, with a small but important contribution from the slope shock. However, all of the much larger spread in time-varying expected returns is due to a time-varying premium on the level shock. Innovations to the slope factor $\gamma' f$ and innovations to inflation do not explain the spread in expected returns across bonds. However, a time-varying risk premium for monetary policy shocks provides an excellent account of the time-varying returns we see in the data.

We show that the usual Hansen-Jagannathan (1991) approach to calculating discount factors that price asset returns by construction extends well to term structure models. In particular, our time-varying risk premia induce a conditionally homoskedastic affine term structure model that perfectly replicates bond yields and expected returns.

Why is this news? We think that two habits of the term structure literature have hid expected return variation. If you find an approximate k factor structure that does a good job of capturing yields or prices in monthly data, that model will almost certainly not show the large variation in expected returns at an annual horizon, or the factor $\gamma' f$ that drives that return. The expected return factor is small and poorly correlated with the level and slope factors that describe most prices and yields. A successful factor model that captures prices, yields and expected returns must contain level, slope and $\gamma' f$ factors, even though the latter will do almost nothing to improving pricing errors. Furthermore, the monthly data do not satisfy an AR(1), so implied annual regressions are far off their actual values.

References

[1] Ang, Andrew and Monika Piazzesi, 2001, "A No-Arbitrage Vector Autoregression of Term Structure Dynamics with Macroeconomic and Latent Variables," NBER working paper 8363.

[2] Brennan, Michael, Yihong Xia and Ashley Wang, 2001, "Intertemporal Asset Pricing and the Fama-French Three Factor Model," manuscript, UCLA and University of Pennsylvania.

[3] Cox, John C., Jonathan A. Ingersoll and Steven Ross, 1985, "A Theory of the Term Structure of Interest Rates," *Econometrica* 53, 385-408.

[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] Campbell, John Y. 1996, "Understanding Risk and Return," *Journal of Political Economy* 104, 298-345

[6] Christiano, Lawrence, Martin Eichenbaum and Charles Evans, 1999, "Monetary Policy Shocks: What Have We Learned and to What End?," in John Taylor and Michael Woodford, Eds., *Handbook of Macroeconomics* North-Holland.

[7] Dai, Qiang and Kenneth Singleton, 2001, "Expectation Puzzles, Time-varying Risk Premia, and Affine Models of the Term Structure," NBER working paper 8167, Forthcoming, *Journal of Financial Economics*

[8] Duarte, Jefferson, 1999, "The Relevance of the Parametrization of Price of Risk in Affine Term Structure Models," University of Chicago GSB Ph.D. dissertation.

[9] Duffee, Gregory R. 2002, "Term premia and interest rate forecasts in affine models," Forthcoming, *Journal of Finance* 57 (Formerly titled "Forecasting Future Interest Rates: Are Affine Models Failures?")

[10] Fama, Eugene F. 1990, "Term Structure Forecasts of Interest Rates, Inflation, and Real Returns," *Journal of Monetary Economics* 25, 59-76.

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

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

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

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

[15] 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

[16] Fisher, Mark ,1998, "A simple model of the failure of the expectations hypothesis." Working paper, Federal Reserve Bank of Atlanta.

[17] 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.*

[18] 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.

[19] Hansen, Lars Peter and Ravi Jagannathan, 1991, "Restrictions on Intertemporal Marginal Rates of Substitution Implied by Asset Returns", *Journal of Political Economy* 99, 225-262.

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

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

[22] Luttmer, Erzo, 1996, "Asset Pricing in Economies with Frictions," *Econometrica* 64, 1439-1467.

[23] Marshall, David, and Wendy Edelberg, 1996, "Monetary Policy Shocks and Long-term Interest Rates," *Federal Reserve Bank of Chicago Economic Perspectives*, 20,

[24] Merton, Robert C., 1973, "An Intertemporal Capital Asset Pricing Model," *Econometrica* 41,867-887.

[25] Piazzesi, Monika, 2001, "An Econometric Model of the Yield Curve with Macroeconomic Jump Effects," NBER Working Paper No. 8246

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

[27] 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.

6 Appendix

6.1 GMM approach to the factor model for expected bond returns

Define $\varepsilon_{t+1}^{(n)}$ as the forecast errors from the individual regressions; in the restricted model

$$\varepsilon_{t+1}^{(n)} = hprx_{t+1}^{(n)} - b_n \left(\gamma_0 + \gamma' f_t\right);$$

in the unrestricted model

$$\varepsilon_{t+1}^{(n)} = hprx_{t+1}^{(n)} - \left(\alpha_n + \beta_n' f_t\right);$$

Denote the forecast error from the average return regression

$$\bar{\varepsilon}_{t+1} = \frac{1}{4}\sum_n \varepsilon_{t+1}^{(n)} = \frac{1}{4}\sum_n hprx_{t+1}^{(n)} - \gamma_0 - \gamma' f_t,$$

where f_t denotes the right hand variables, the one year yield and forward rates,

$$f_t \equiv \left[\begin{array}{ccccc} y_t^{(1)} & f_t^{(1\to 2)} & f_t^{(2\to 3)} & f_t^{(3\to 4)} & f_t^{(4\to 5)} \end{array}\right]'.$$

We incorporate the two-step estimate by adding to the vector of moments. This approach allows us easily to impose the restriction $\sum b_n = 4$. Thus, the moments are

$$\begin{aligned} g_T' &= E\left[\begin{array}{cccccccccc} \varepsilon_{t+1}^{(2)} & \varepsilon_{t+1}^{(3)} & \varepsilon_{t+1}^{(4)} & \varepsilon_{t+1}^{(5)} & \varepsilon_{t+1}^{(2)} f_t' & \varepsilon_{t+1}^{(3)} f_t' & \varepsilon_{t+1}^{(4)} f_t' & \varepsilon_{t+1}^{(5)} f_t' & \frac{1}{4}\sum_n \varepsilon_{t+1}^{(n)} & \left(\frac{1}{4}\sum_n \varepsilon_{t+1}^{(n)}\right) f_t' \end{array}\right] \\ &= E\left[\begin{array}{cccc} \varepsilon_{t+1}' & \left(\varepsilon_{t+1} \otimes f_t\right)' & \bar{\varepsilon}_{t+1} & \bar{\varepsilon}_{t+1} f_t' \end{array}\right] \end{aligned}$$

We estimate the parameters of the restricted model with the following moment conditions:

$$\begin{aligned} E\left(\bar{\varepsilon}_{t+1}\right) &= 0 : \gamma_0 \\ E\left(\bar{\varepsilon}_{t+1} f_t\right) &= 0 : \gamma \\ E\left[\varepsilon_{t+1}^{(n)}\left(\gamma_0 + \gamma' f_t\right)\right] &= 0 : b_n \end{aligned}$$

The GMM formula for the standard error of the estimates is

$$\frac{1}{T}(ad)^{-1} a S a' (ad)^{-1\prime}$$

and the GMM formula for the covariance matrix of the sample moments is

$$cov(g_T) = \frac{1}{T}(I - d(ad)^{-1}a) S (I - d(ad)^{-1}a)'.$$

We form the χ^2 value for the overidentifying restrictions test by

$$\chi^2 = g_T' cov(g_T)^+ g_T \tag{33}$$

where + refers to a pseudoinverse, since the covariance matrix is singular. The degrees of freedom is the rank of $cov\,(g_T)$

The elements of these formulas are

$$a = \begin{array}{cc} & \begin{array}{cccccccccc} \varepsilon_{t+1}^{(2)} & \varepsilon_{t+1}^{(3)} & \varepsilon_{t+1}^{(4)} & \varepsilon_{t+1}^{(5)} & \varepsilon_{t+1}^{(2)} f_t' & \varepsilon_{t+1}^{(3)} f_t' & \varepsilon_{t+1}^{(4)} f_t' & \varepsilon_{t+1}^{(5)} f_t' & \bar{\varepsilon}_{t+1} & \bar{\varepsilon}_{t+1} f_t' \end{array} \\ \begin{array}{c} \gamma_0 \\ \gamma \\ b_2 \\ b_3 \\ b_4 \\ b_5 \end{array} & \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & I_5 \\ \gamma_0 & 0 & 0 & 0 & \gamma' & 0 & 0 & 0 & 0 & 0 \\ 0 & \gamma_0 & 0 & 0 & 0 & \gamma' & 0 & 0 & 0 & 0 \\ 0 & 0 & \gamma_0 & 0 & 0 & 0 & \gamma' & 0 & 0 & 0 \\ 0 & 0 & 0 & \gamma_0 & 0 & 0 & 0 & \gamma' & 0 & 0 \end{bmatrix} \end{array}$$

$$a = \begin{array}{cc} & \begin{array}{cccc} \varepsilon_{t+1}' & (\varepsilon_{t+1} \otimes f_t)' & \bar{\varepsilon}_{t+1} & \bar{\varepsilon}_{t+1} f_t' \end{array} \\ \begin{array}{c} \gamma_0 \\ \gamma \\ b \end{array} & \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & I_5 \\ I_4 \gamma_0 & I_4 \otimes \gamma' & 0 & 0 \end{bmatrix} \end{array}$$

(We indicate the moments and parameters outside the matrix for easier reference.)

$$d = -E \begin{array}{cc} & \begin{array}{cccccc} \gamma_0 & \gamma' & b_2 & b_3 & b_4 & b_5 \end{array} \\ \begin{array}{c} \varepsilon_{t+1}^{(2)} \\ \varepsilon_{t+1}^{(3)} \\ \varepsilon_{t+1}^{(4)} \\ \varepsilon_{t+1}^{(5)} \\ \varepsilon_{t+1}^{(2)} f_t \\ \varepsilon_{t+1}^{(3)} f_t \\ \varepsilon_{t+1}^{(4)} f_t \\ \varepsilon_{t+1}^{(5)} f_t \\ \bar{\varepsilon}_{t+1} \\ \bar{\varepsilon}_{t+1} f_{t+1} \end{array} & \begin{bmatrix} b_2 & b_2 f_t' & \gamma_0 + \gamma' f & 0 & 0 & 0 \\ b_3 & b_3 f_t' & 0 & \gamma_0 + \gamma' f & 0 & 0 \\ b_4 & b_4 f_t' & 0 & 0 & \gamma_0 + \gamma' f & 0 \\ b_5 & b_5 f_t' & 0 & 0 & 0 & \gamma_0 + \gamma' f \\ b_2 f_t & b_2 f_t f_t' & \gamma_0 f_t + f_t f_t' \gamma & 0 & 0 & 0 \\ b_3 f_t & b_3 f_t f_t' & 0 & \gamma_0 f_t + f_t f_t' \gamma & 0 & 0 \\ b_4 f_t & b_4 f_t f_t' & 0 & 0 & \gamma_0 f_t + f_t f_t' \gamma & 0 \\ b_5 f_t & b_5 f_t f_t' & 0 & 0 & 0 & \gamma_0 f_t + f_t f_t' \gamma \\ 1 & f_t' & 0 & 0 & 0 & 0 \\ f_t & f_t f_t' & 0 & 0 & 0 & 0 \end{bmatrix} \end{array}$$

$$d = - \begin{array}{cc} & \begin{array}{ccc} \gamma_0 & \gamma' & b' \end{array} \\ \begin{array}{c} \varepsilon_{t+1} \\ \varepsilon_{t+1} \otimes f_t \\ \bar{\varepsilon}_{t+1} \\ \bar{\varepsilon}_{t+1} f_{t+1} \end{array} & \begin{bmatrix} b & b \otimes E(f)' & I_4 \otimes (\gamma_0 + \gamma' E(f)) \\ b \otimes E(f) & b \otimes E(ff') & I_4 \otimes [\gamma_0 + E(f_t f_t') \gamma] \\ 1 & E(f)' & 0 \\ E(f) & E(ff') & 0 \end{bmatrix} \end{array}$$

If instead we allow a separate intercept a_n in each second step regression, we have

$$a = \begin{array}{cc} & \begin{array}{cccc} \varepsilon_{t+1}' & (\varepsilon_{t+1} \otimes f_t)' & \bar{\varepsilon}_{t+1} & \bar{\varepsilon}_{t+1} f_t' \end{array} \\ \begin{array}{c} \gamma_0 \\ \gamma \\ b \\ a \end{array} & \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & I_5 \\ 0 & I_4 \otimes \gamma' & 0 & 0 \\ I_4 & 0 & 0 & 0 \end{bmatrix} \end{array}$$

$$
d=-\begin{array}{cc} & \begin{array}{cccc} \gamma_0 & \gamma' & b' & a' \end{array} \\ \begin{array}{c} \varepsilon_{t+1} \\ \varepsilon_{t+1}\otimes f_t \\ \bar{\varepsilon}_{t+1} \\ \bar{\varepsilon}_{t+1} f_{t+1} \end{array} & \begin{bmatrix} 0 & b\otimes E(f)' & I_4\otimes\left[\gamma' E\left(f\right)\right] & I_5 \\ 0 & b\otimes E(ff') & I_4\otimes\left[E\left(f_t f_t'\right)\gamma\right] & I_4\otimes E(f_t) \\ 1 & E(f)' & 0 & 0 \\ E(f) & E(ff') & 0 & 0 \end{bmatrix} \end{array}
$$

We found the J_T tests (33) numerically unstable. Thus, we map the unrestricted model in to this setup so that we can form GMM based Wald tests of the restrictions. Denote the second step regressions

$$
hprx_{t+1}^{(n)}=\alpha_n+\beta_n' f_t+\varepsilon_{t+1}^{(n)}
$$

Thus, we can easily test the restrictions by testing whether $\alpha_n=b_n\gamma_0$ and/or $\beta_n=b_n\gamma'$ Let $\beta=\left[\begin{array}{cccc}\beta_2' & \beta_3' & \beta_4' & \beta_5'\end{array}\right]$Since there is no cross-effect to γ estimation, we only need the original set of moments and the a and d matrices are now

$$
a=\begin{array}{cc} & \begin{array}{cc} \varepsilon_{t+1}' & (\varepsilon_{t+1}\otimes f_t)' \end{array} \\ \begin{array}{c} \alpha \\ \beta \end{array} & \begin{bmatrix} I_4 & 0 \\ 0 & I_{20} \end{bmatrix} \end{array}
$$

$$
d=-\begin{array}{cc} & \begin{array}{cc} \alpha' & \beta' \end{array} \\ \begin{array}{c} \varepsilon_{t+1} \\ \varepsilon_{t+1}\otimes f_t \end{array} & \begin{bmatrix} I_4 & I_4\otimes E(f)' \\ I_4\otimes E(f) & I_4\otimes E(ff') \end{bmatrix} \end{array}
$$

6.2 Affine model

6.2.1 Proof of proposition 1

Proposition 1. Let X_t denote a vector of state variables that follows

$$
X_t=\mu+\phi X_{t-1}+\Sigma\varepsilon_t
$$

with i.i.d. normally distributed shocks ε_t and $E(\varepsilon_t\varepsilon_t')=I$. Let the short rate $y_t^{(1)}$ be included in the state vector X_t, $y_t^{(1)}=e_1' X_t$. Let m_{t+1} be given by

$$
m_{t+1}=e^{-y_t^{(1)}-\lambda_t'\lambda_t-\lambda_t'\varepsilon_t}
$$

where λ is a linear function of X_t, e.g.

$$
\lambda_t=\lambda_0+\lambda_1' X_t
$$

Then bond prices, generated by $e^{p_t^{(n)}}=E_t(m_{t+1}m_{t+2}...m_{t+n})$ are linear functions of the state variables, i.e. we can find A_n and B_n such that

$$
p_t^{(n)}=A_n+B_n' X_t. \tag{34}
$$

The affine model is equivalent to risk-neutral pricing with distorted probabilities.

$$\phi^* \equiv \phi - \Sigma\lambda_1 \tag{35}$$
$$\mu^* \equiv \mu - \Sigma\lambda_0. \tag{36}$$

We prove the proposition by simply grinding out the conditional expectation of the discount factor and finding the coefficients A_n,B_n. In the end, these parameters can be found recursively from

$$\begin{aligned} B_0 &= 0,\ A_0 = 0 \\ B_{n+1}^\top &= -e_1' + B_n^\top \phi^* && (37) \\ A_{n+1} &= A_n + B_n^\top \mu^* + \frac{1}{2} B_n^\top \Sigma\Sigma^\top B_n && (38) \end{aligned}$$

where μ^* and ϕ^* are defined from (35)-(36)

Algebra: We guess the form (34) and then show that the coefficients must obey (37)-(38). The price at time t of a $n+1$ period maturity bond is

$$P_t^{n+1} = E_t\left[m_{t+1} P_{t+1}^n\right]$$

Thus, we must have

$$\begin{aligned} \exp\left(A_{n+1} + B_{n+1}^\top\right) &= E_t\left[\exp\left(-r_t - \frac{1}{2}\lambda_t^\top\lambda_t - \lambda_t^\top\varepsilon_{t+1} + A_n + B_n^\top X_{t+1}\right)\right] && (39) \\ &= \exp\left(-r_t - \frac{1}{2}\lambda_t^\top\lambda_t + A_n\right) E_t\left[\exp\left(-\lambda_t^\top\varepsilon_{t+1} + B_n^\top X_{t+1}\right)\right] && (40) \end{aligned}$$

We can simplify the second term in (40):

$$\begin{aligned} & E_t\left[\exp\left(-\lambda_t^\top\varepsilon_{t+1} + B_n^\top X_{t+1}\right)\right] \\ =\ & E_t\left[\exp\left(-\lambda_t^\top\varepsilon_{t+1} + B_n^\top\mu + B_n^\top\phi X_t + B_n^\top\Sigma\varepsilon_{t+1}\right)\right] \\ =\ & E_t\left[\exp\left((-\lambda_t^\top + B_n^\top\Sigma)\varepsilon_{t+1} + B_n^\top\mu + B_n^\top\phi X_t\right)\right] \\ =\ & \exp\left(B_n^\top\mu + B_n^\top\phi X_t\right)\exp\frac{1}{2}(-\lambda_t^\top + B_n^\top\Sigma)(-\lambda_t^\top + B_n^\top\Sigma)^\top \\ =\ & \exp\left(B_n^\top\mu + B_n^\top\phi X_t\right)\exp\frac{1}{2}(\lambda_t^\top\lambda_t - 2B_n^\top\Sigma\lambda_t + B_n^\top\Sigma\Sigma^\top B_n) \end{aligned}$$

Now, continuing from (40):

$$\begin{aligned} A_{n+1} + B_{n+1}^\top X_t &= \left(-r_t - \frac{1}{2}\lambda_t^\top\lambda_t + A_n\right) + \left(B_n^\top\mu + B_n^\top\phi X_t\right) + \left(\frac{1}{2}\lambda_t^\top\lambda_t - B_n^\top\Sigma\lambda_t + \frac{1}{2}B_n^\top\Sigma\Sigma^\top B_n\right) \\ &= -r_t + A_n + B_n^\top\mu + B_n^\top\phi X_t - B_n^\top\Sigma\lambda_t + \frac{1}{2}B_n^\top\Sigma\Sigma^\top B_n \\ &= A_n + B_n^\top\mu - B_n^\top\Sigma\lambda_0 + \frac{1}{2}B_n^\top\Sigma\Sigma^\top B_n - e_1'X_t + B_n^\top\phi X_t - B_n^\top\Sigma\lambda_1 X_t \\ &= \left(A_n + B_n^\top\mu - B_n^\top\Sigma\lambda_0 + \frac{1}{2}B_n^\top\Sigma\Sigma^\top B_n\right) + \left(-e_1' + B_n^\top\phi - B_n^\top\Sigma\lambda_1\right) X_t. \end{aligned}$$

Matching terms, we obtain (37)-(38).

Iterating (37)-(38), we can also express the coefficients A_n, B_n in $p_t^{(n)} = A_n + B_n' X_t$ explicitly as

$$
\begin{aligned}
B_n^\top &= -e_1' \sum_{j=0}^{n-1} \phi_j^* = -e_1' \left(I - \phi^{*n}\right) \left(I - \phi^*\right)^{-1} && (41) \\
A_n &= \sum_{j=0}^{n-1} \left(B_j^\top \mu^* + \frac{1}{2} B_j^\top \Sigma \Sigma^\top B_j \right). && (42)
\end{aligned}
$$

Given prices, we can easily find formulas for yields, forward rates, etc. as linear functions of the state variable X_t. Yields are just

$$
y_t^{(n)} = -\frac{A_n}{n} - \frac{B_n^\top}{n} X_t.
$$

Forward rates are

$$
\begin{aligned}
f_t^{(n-1 \to n)} &= p_t^{(n-1)} - p_t^{(n)} \\
&= (A_{n-1} - A_n) - (B_{n-1} - B_n) X_t \\
&= A_n^f + B_n^f X_t
\end{aligned}
$$

We can find A^f and B^f from our previous formulas for A_n, B_n. From (41) and (42),

$$
B_n^f = e_1' \phi^{*n-1} \tag{43}
$$

$$
A_n^f = -B_{n-1}^\top \mu^* - \frac{1}{2} B_{n-1}^\top \Sigma \Sigma^\top B_{n-1} \tag{44}
$$

These formulas have quite a simple intuition. In a risk neutral economy, the price of a two period bond would be

$$
\begin{aligned}
e^{p_t^{(2)}} &= E_t \left(e^{-y_t^{(1)} - y_{t+1}^{(1)}} \right) \\
p_t^{(2)} &= E_t(-y_t^{(1)} - y_{t+1}^{(1)}) + \frac{1}{2} \sigma_t^2 \left(y_{t+1}^{(1)} \right)
\end{aligned}
$$

Now, $y_t^{(1)} = e_1' X_t$ and $y_t^{(1)} = e_1' \left(\phi X_t + \mu \right)$. Thus, in the risk-neutral economy, we expect

$$
p_t^{(2)} = \left[-e_1' \mu + \frac{1}{2} \sigma_t^2 (y_{t+1}^{(1)}) \right] - e_1' \left(I + \phi \right) X_t
$$

The actual formulas only differ by the difference between ϕ, μ and ϕ^*, μ^* as given by (35) and (36). Bonds in our economy are priced just as if we were in a risk-neutral ($\lambda = 0$) economy with modified probabilities μ, ϕ.

Note also in (35) and (36) that λ_0 contributes only to the difference between μ and μ^*, and thus contributes only to the *constant* term A_n in bond prices and yields. A homoskedastic discount factor can only give a constant risk premium. λ_1 contributes only to the difference between ϕ and ϕ^*, and only this parameter can affect the loading B_n of bond prices on state variables.

The forward rate formula (43) is even simpler. It says directly that the forward rate is equal to the expected value of the future spot rate and a Jensen inequality term, under the risk-neutral measure ϕ^*.

6.2.2 Proof of proposition 2

Proposition 2. Suppose that X_t contains a full set of prices, i.e. suppose that we can recover prices of 1 through N period bonds from X_t by $\begin{bmatrix} p_t^{(1)} & p_t^{(2)} & .. & p_t^{(N)} \end{bmatrix}' = PX_t$. Then, any λ_t that solves

$$E_t\left(hprx_{t+1}\right) = cov(hprx_{t+1}\varepsilon'_{t+1})\lambda_t + \frac{1}{2}\sigma_t^2\left(hprx_{t+1}\right) \tag{45}$$

form a self-consistent affine model; the predicted bond prices by $p_t^{(n)} = A_n + B'_n X_t$ are the same as those recovered directly by PX_t.

Of course if any λ that solves (45) works, the particular choice

$$\begin{aligned} \lambda_t &= C'(CC')^{-1}\left[E_t(hprx_{t+1}) + \frac{1}{2}\sigma_t^2(hprx_{t+1})\right]. \\ C &= cov(hprx_{t+1}\varepsilon'_{t+1}) \end{aligned}$$

also works. If the VAR consists of yields, then, for example, the third row of P which recovers $p^{(3)}$ would be $P'_3 = \begin{bmatrix} 0 & 0 & -3 & 0 & ... \end{bmatrix}$.

At an intuitive level, this proposition is obvious. We can always find a price by discounting its payoff at its own expected return. Thus, if our model correctly captures all the intermediate expected returns, it must also correctly capture any prices.

We prove the proposition by induction. We show that a discount factor with this λ_t prices an n period bond as a one period claim to an $n-1$ period bond. Suppose $A_{n-1} = 0$, $B_{n-1} = P_{n-1}$ i.e. the n period bond is correctly priced. Then from (38)-(37),

$$B'_n = -e'_1 + P'_{n-1}\phi^* = -e'_1 + P'_{n-1}\left(\phi - \Sigma\lambda_1\right) \tag{46}$$

$$A_n = P'_{n-1}(\mu - \Sigma\lambda_0) + \frac{1}{2}P'_{n-1}\Sigma\Sigma' P_{n-1} \tag{47}$$

We need to show that if $\lambda_t = \lambda_0 + \lambda_1 X_t$ solves (45), then $A_n = 0$ and $B_n = P_n$. From the basic definitions,

$$\begin{aligned}
hprx_{t+1} &= p_{t+1}^{(n-1)} - p_t^{(n)} - y_t^{(1)} \\
cov(hprx_{t+1}^{(n)}\varepsilon_{t+1}') &= cov(p_{t+1}^{(n-1)}\varepsilon_{t+1}') = P_{n-1}'\Sigma \\
\sigma_t^2(hprx_{t+1}^{(n)}) &= P_{n-1}'\Sigma\Sigma' P_{n-1} \\
E_t\left(hprx_{t+1}\right) &= E_t(p_{t+1}^{(n-1)}) - p_t^{(n)} - y_t^{(1)} = P_{n-1}'\left(\mu + \phi X_t\right) - P_n' X_t - e_1' X_t
\end{aligned}$$

Plugging it all in (45),

$$\begin{aligned}
P_{n-1}'\mu + P_{n-1}'\phi X_t - P_n' X_t - e_1' X_t &= P_{n-1}'\Sigma\lambda_t + \frac{1}{2}P_{n-1}'\Sigma\Sigma' P_{n-1} \\
P_n' X_t &= P_{n-1}'\mu - P_{n-1}'\Sigma\lambda_0 - \frac{1}{2}P_{n-1}'\Sigma\Sigma' P_{n-1} + \left(P_{n-1}'\phi - P_{n-1}'\Sigma\lambda_1 - e_1'\right) X_t
\end{aligned}$$

Collecting constant (A_n) terms and the terms (B_n) multiplying X_t, we have the required relations,

$$\begin{aligned}
0 &= P_{n-1}'\left(\mu - \Sigma\lambda_0\right) - \frac{1}{2}P_{n-1}'\Sigma\Sigma' P_{n-1} \\
P_n' &= P_{n-1}'\left(\phi - \Sigma\lambda_1\right) - e_1'.
\end{aligned}$$

6.2.3 Proof of proposition 3

Proposition 3. Among all market prices of risk λ_t *that price the available bond returns, or (equivalently) produce a self-consistent affine model, the market prices of risk defined by*

$$\lambda_t = C'(CC')^{-1}\left[E_t(hprx_{t+1}) + \frac{1}{2}\sigma_t^2(hprx_{t+1})\right] \tag{48}$$

produce the minimum value of $\lambda_t'\lambda_t$*, the discount factor with minimum volatility, and the minimum value of the maximum Sharpe ratio. They set to zero the prices of risk of any shock uncorrelated with bond returns.*

This is essentially an application of the Hansen Jagannathan (1991) bound. First we show that $\lambda_t'\lambda_t$ has all the indicated interpretations. Then, we show that the choice (48) is the minimum value subject to the constraint that λ_t satisfies

$$E_t\left(hprx_{t+1}\right) = cov(hprx_{t+1}\varepsilon_{t+1}')\lambda_t + \frac{1}{2}\sigma_t^2\left(hprx_{t+1}\right). \tag{49}$$

The discount factor is

$$m_{t+1} = e^{-y_t^{(1)} - \frac{1}{2}\lambda_t'\lambda_t - \lambda_t'\varepsilon_{t+1}}$$

Its variance is

$$\begin{aligned}\sigma_t^2(m_{t+1}) &= E_t(m_{t+1}^2) - [E_t(m_{t+1})]^2 \\ &= e^{-2y_t^{(1)}+\lambda_t'\lambda_t} - e^{-2y_t^{(1)}} \\ &= e^{-2y_t^{(1)}}\left[e^{\lambda_t'\lambda_t} - 1\right].\end{aligned}$$

Therefore, the conditional variance of the discount factor is monotonic in $\lambda_t'\lambda_t$. The maximum conditional Sharpe ratio of all assets priced by m_{t+1} is $\sigma_t(m_{t+1})/E_t(m_{t+1})$, and is minimized when $\lambda_t'\lambda_t$ is minimized.

Now, the minimization. The problem is simply $\min \lambda_t'\lambda_t$ s.t. (49). The first order conditions are

$$\lambda_t = cov(hprx_{t+1}\varepsilon_{t+1}')'\delta = C'\delta$$

where δ are Lagrange multipliers and $C = cov(hprx_{t+1}\varepsilon_{t+1}')$. Evaluating δ in the constraint (49),

$$E_t\left(hprx_{t+1}\right) = CC'\delta + \frac{1}{2}\sigma_t^2\left(hprx_{t+1}\right)$$

$$\delta = \left(CC'\right)^{-1}\left[E_t\left(hprx_{t+1}\right) - \frac{1}{2}\sigma_t^2\left(hprx_{t+1}\right)\right]$$

$$\lambda_t = C'\left(CC'\right)^{-1}\left[E_t\left(hprx_{t+1}\right) - \frac{1}{2}\sigma_t^2\left(hprx_{t+1}\right)\right]$$

If a shock is uncorrelated with returns, $cov(hprx_{t+1}\varepsilon_{t+1}^{(i)}) = 0$, then λ places no loading on that shock.

6.3 Forward VAR, yield VAR and return regressions

Denote

$$f_t = \begin{bmatrix} y_t^{(1)} \\ f_t^{(1\to 2)} \\ f_t^{(2\to 3)} \\ f_t^{(3\to 4)} \\ f_t^{(4\to 5)} \end{bmatrix},\ y_t = \begin{bmatrix} y_t^{(1)} \\ y_t^{(2)} \\ y_t^{(3)} \\ y_t^{(4)} \\ y_t^{(5)} \end{bmatrix};\ hprx_t = \begin{bmatrix} hprx_t^{(2)} \\ hprx_t^{(3)} \\ hprx_t^{(4)} \\ hprx_t^{(5)} \end{bmatrix}$$

Denote the return regression, forward rate VAR ,and yield VAR as

$$\begin{aligned}hprx_{r+1} &= a + Bf_t + \Sigma_h\eta_{t+1}^h \\ y_{t+1} &= \mu_y + \phi_y y_t + \Sigma_y\eta_{t+1}^y \\ f_{t+1} &= \mu_f + \phi_f f_t + \Sigma_f\eta_{t+1}^f\end{aligned}$$

These representations are related as follows. Let

$$D = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ -1 & 2 & 0 & 0 & 0 \\ 0 & -2 & 3 & 0 & 0 \\ 0 & 0 & -3 & 4 & 0 \\ 0 & 0 & 0 & -4 & 5 \end{bmatrix}; D^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 \\ \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \end{bmatrix} \tag{50}$$

Since $f_t^{(n-1\to n)} = p_t^{(n-1)} - p_t^{(n)} = -(n-1)y_t^{(n-1)} + ny_t^{(n)}$, we have

$$\begin{aligned} f_t &= Dy_t. \\ y_t &= D^{-1}f_t \end{aligned}$$

Applying D and D^{-1} to the yield and forward rate VARs, we relate the yield and forward rate VARs as follows

$$\begin{aligned} \mu_f &= D\mu_y;\ \mu_y = D^{-1}\mu_f \\ \phi_f &= D\phi_y D^{-1};\ \phi_y = D^{-1}\phi_f D \\ \Sigma_f \eta_{t+1}^f &= D\Sigma_y \eta_{t+1}^y;\ \Sigma_y \eta_{t+1}^y = D^{-1}\Sigma_f \eta_{t+1}^f \end{aligned}$$

Let

$$M = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 4 & 0 \end{bmatrix};\ N = \begin{bmatrix} -1 & 2 & 0 & 0 & 0 \\ -1 & 0 & 3 & 0 & 0 \\ -1 & 0 & 0 & 4 & 0 \\ -1 & 0 & 0 & 0 & 5 \end{bmatrix}$$

Since $hprx_{t+1}^{(n)} = p_{t+1}^{(n-1)} - p_t^{(n)} - y_t^{(1)} = -(n-1)y_{t+1}^{(n-1)} + ny_t^{(n)} - y_t^{(1)}$,

$$\begin{aligned} hprx_{t+1} &= -My_{t+1} + Ny_t \\ &= -M\left(\mu_y + \phi_y y_t + \Sigma_y \eta_{t+1}^y\right) + Ny_t \\ &= -M\mu_y - M\Sigma_y \eta_{t+1}^y + \left(N - M\phi_y\right) y_t \\ &= -M\mu_y - M\Sigma_y \eta_{t+1}^y + \left(N - M\phi_y\right) D^{-1} f_t \end{aligned}$$

Thus, given the yield VAR, the return regression is

$$\begin{aligned} a &= -M\mu_y \\ B &= \left(N - M\phi_y\right) D^{-1} \\ \Sigma_h \eta_{t+1}^h &= -M\Sigma_y \eta_{t+1}^y \end{aligned}$$

Inversion is a little trickier, because the holding period return regression does not have any information about $y_{t+1}^{(5)}, p_{t+1}^{(5)}, f_{t+1}^{(4\to5)}$. Hence, the return regression only identifies the first four rows of the VARs. Denote

$$M_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix};\ M_4^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & \frac{1}{4} \end{bmatrix};\ L = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

M_4 is the invertible part of M, and L lops off the last row of a vector or matrix. $L\mu_y$, $L\phi_y$, etc. are thus the first four rows of μ_y, ϕ_y, etc. We have $M = M_4 L$ and hence

$$\begin{aligned} a &= -M_4 L \mu_y \\ B &= \left(N - M_4 L \phi_y\right) D^{-1} \\ \Sigma_h \eta^h_{t+1} &= -M_4 L \Sigma_y \eta^y_{t+1} \end{aligned}$$

$$\begin{aligned} L\mu_y &= -M_4^{-1} a && (51) \\ L\phi_y &= M_4^{-1}(N - BD) && (52) \\ L\Sigma_y \eta^y_{t+1} &= -M_4^{-1} \Sigma_h \eta^h_{t+1} && (53) \end{aligned}$$

6.4 The restricted yield VAR

We construct our restricted yield VAR by estimating the restricted return regressions

$$hprx_{t+1} = a + b\gamma' f_t + v^h_{t+1}$$

and constructing the implied yield VAR from (51)-(52) using $B = b\gamma'$ and the a estimate from the restricted return regression. The last row of the yield VAR, forecasting $y^{(5)}_{t+1}$, is not restricted by the holding period return regression. The 5 year, holding period return $hprx^{(5)}_{t+1}$ depends on the price next year of 4 year bonds. Thus, we append the unconstrained $y^{(5)}_{t+1}$ equation to complete the yield VAR.

Figure 14 compares the restricted and unrestricted yield VAR coefficients ϕ_y.

You can see that the regression coefficients are quite close, but not exactly the same. The big visible difference is the coefficient of the one-year yield on the lagged two year yield, which is greater by 0.9 in the restricted estimate. The eigenvalue decomposition is hard to distinguish visually from Figure 15. The idiosyncratic shocks from the unrestricted VAR are slightly *less* important than those shown in Figure 15. The restricted regression cannot be statistically rejected, unsurprisingly, since this is a linear transformation of the holding period return regressions. Table A1 presents the R^2 values, and you can see that they are essentially the same.

	$y^{(1)}$	$y^{(2)}$	$y^{(3)}$	$y^{(4)}$	$y^{(5)}$
Restricted	0.624	0.678	0.718	0.741	0.757
Unrestricted	0.629	0.679	0.719	0.742	0.758

Table A1. R^2 from restricted and unrestricted yield VARs. The regression is

$$y_{t+1} = \mu_y + \phi_y y_t + v^y_{t+1}$$

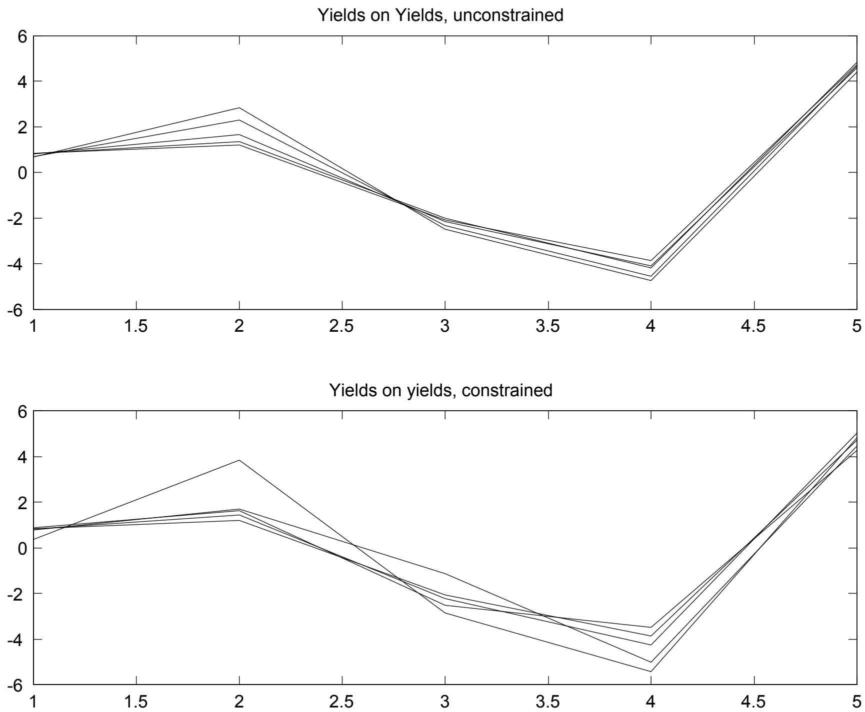

Figure 14: Regression coefficients ϕ_y in a yield VAR, $y_{t+1} = \mu_y + \phi_y y_t + \eta_{t+1}$The top panel shows unconstrained regression coefficients. The bottom panel shows coefficients implied by the constrained holding period return regression $hprx_{t+1} = a + b(\gamma' f_t) + v_{t+1}$ together with an unconstrained regression for the 5 year yield.

The top row is calculated using coefficients implied by the constrained holding period return regression

$$hprx_{t+1} = a + b(\gamma' f_t) + v_{t+1}^h$$

together with an unconstrained regression for the 5 year forward rate. The bottom row comes from an unrestricted OLS regression