Bond Risk Premia

John H. Cochrane and Monika Piazzesi*

Abstract

We study time variation in expected excess bond returns. We run regressions of one-year excess returns on initial forward rates. We find that a single factor, a single tent-shaped linear combination of forward rates, predicts excess returns on 1-5 year maturity bonds with R^2 up to 0.44. The return-forecasting factor is countercyclical and forecasts stock returns. An important component of the return-forecasting factor is unrelated to the level, slope, and curvature movements described by most term structure models. We document that measurement errors do not to affect our central results. (JEL G0, G1, E0, E4)

We study time-varying risk premia in US government bonds. We run regressions of one-year excess returns – borrow at the one-year rate, buy a long term bond, and sell it in one year – on five forward rates available at the beginning of the period. By focusing on excess returns, we net out inflation and the level of interest rates, so we focus directly on real risk premia in the nominal term structure. We find R^2 values as high as 44 percent. The forecasts are statistically significant, even taking into account the small-sample properties of test statistics, and they survive a long list of robustness checks. Most importantly, the pattern of regression coefficients is the same for all maturities. A *single* "return-forecasting factor," a single linear combination of forward rates or yields, describes time-variation in the expected return of *all* bonds.

This work extends Eugene Fama and Robert Bliss's (1987) and John Campbell and Robert Shiller's (1991) classic regressions. Fama and Bliss found that the spread between the n-year forward rate and the 1-year yield predicts the 1-year excess return of the n-year bond, with R^2 about 18 percent. Campbell and Shiller found similar results forecasting yield changes with yield spreads. We substantially strengthen this evidence against the

1

expectations hypothesis. (The expectations hypothesis that long yields are the average of future expected short yields is equivalent to the statement that excess returns should not be predictable.) Our p-values are much smaller, we more than double the forecast R^2, and the return-forecasting factor drives out individual forward or yield spreads in multiple regressions. Most importantly, we find that the *same* linear combination of forward rates predicts bond returns at all maturities, where Fama and Bliss and Campbell and Shiller relate each bond's expected excess return to a different forward spread or yield spread.

Measurement error

One always worries that return forecasts using prices are contaminated by measurement error. A spuriously high price at t will seem to forecast a low return from time t to time $t+1$; the price at t is common to left and right hand sides of the regression. We address this concern in a number of ways. First, we find that the forecast power, the tent shape, and the single-factor structure are all preserved when we lag the right hand variables, running returns from t to $t+1$ on variables at time $t-i/12$. In these regressions, the forecasting variables (time $t-i/12$ yields or forward rates) do not share a common price with the excess return from t to $t+1$. Second, we compute the patterns that measurement error can produce, and show they are not the patterns we observe. Measurement error produces returns on n-period bonds that are forecast by the n-period yield. It does not produce the single-factor structure; it does not generate forecasts in which (say) the five-year yield helps to forecast the two-year bond return. Third, the return-forecasting factor predicts excess stock returns with a sensible magnitude. Measurement error in bond prices cannot generate this result.

Our analysis does reveal some measurement error, however. Lagged forward rates also help to forecast returns in the presence of time-t forward rates. A regression on a moving average of forward rates shows the same tent-shaped single factor, but improves R^2 up to 44 percent. These results strongly suggest measurement error. Since bond prices are time-t expectations of future nominal discount factors, it is very difficult for any economic model of correctly measured bond prices to produce dynamics in which lagged yields help to forecast

2

anything. However, if the risk premium moves slowly over time but there is measurement error, moving averages will improve the signal to noise ratio on the right hand side.

These considerations together argue that the core results – a single roughly tent-shaped factor that forecasts excess returns of all bonds, and with a large R^2 – are not driven by measurement error. Quite the contrary: to see the core results you have to take steps to mitigate measurement error. A standard monthly AR(1) yield VAR raised to the 12th power misses most of the one-year bond return predictability and completely misses the single-factor representation. To see the core results you must look directly at the one-year horizon, which cumulates the persistent expected return relative to serially uncorrelated measurement error (or use more complex time series models) and you see it better with a moving average right hand variable.

The single-factor structure is statistically rejected when we regress returns on time-t forward rates. However, the single factor explains over 99.5 percent of the variance of expected excess returns, so the rejection is tiny on an economic basis. Also, the statistical rejection shows the characteristic pattern of small measurement errors: tiny movements in n-period bond yields forecast tiny additional excess return on n-period bonds, and this evidence against the single-factor model is much weaker with lagged right hand variables. We conclude that the single-factor model is an excellent approximation, and may well be the literal truth once measurement errors are accounted for.

Term structure models

We relate the return-forecasting factor to term structure models in finance. The return-forecasting factor is a symmetric, tent-shaped linear combination of forward rates. Therefore, it is unrelated to pure slope movements: a linearly rising or declining yield or forward curve gives exactly the same return forecast. Important variation in the return-forecasting factor, and an important part of its forecast power, is unrelated to the standard "level," "slope" and "curvature" factors that describe the vast bulk of movements in bond yields and thus form the basis of most term structure models. The 4-5 year yield spread, though a tiny

factor for yields, provides important information about the expected returns of all bonds. The increased power of the return-forecasting factor over three-factor forecasts is statistically and economically significant.

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

One may ask, "How can it be that the *five*-year forward rate is necessary to predict the returns on *two*-year bonds?" This natural question reflects a subtle misconception. Under the expectations hypothesis, yes, the n-year forward rate is an optimal forecast of the one-year spot rate $n-1$ years from now, so no other variable should enter that forecast. But the expectations hypothesis is false, and we're forecasting one-year excess returns not n-year spot rates anyway. Once we abandon the expectations hypothesis (so that returns are forecastable at all), it is easy to generate economic models in which many forward rates are needed to forecast one-year excess returns on bonds of any maturity. We provide an explicit example. The form of the example is straightforward: Aggregate consumption and inflation follow time-series processes, and bond prices are generated by expected marginal utility growth divided by inflation. The discount factor is conditionally heteroskedastic, generating a time-varying risk premium. In the example, bond prices are linear functions of state variables, so this example also shows that it is straightforward to construct affine models, models in the style of Darrell Duffie and Rui Kan (1996) which dominate the term structure literature,

that reflect this or related patterns of bond return predictability. A crucial feature of the example, but an unfortunate one for simple story-telling, is that the discount factor must reflect five state variables, so that five bonds can move independently. Otherwise, one can recover (say) the five-year bond price exactly from knowledge of the other four bond prices, and multiple regressions are impossible.

Related literature

Our single-factor model is similar to the "single index" or "latent variable" models used by Lars Hansen and Robert Hodrick (1983) and Wayne Ferson and Michael Gibbons (1985) to capture time-varying expected returns. Robert Stambaugh (1988) ran regressions similar to ours of 2-6 month bond excess returns on 1-6 month forward rates. After correcting for measurement error by using adjacent rather than identical bonds on the left and right hand side, Stambaugh found a tent-shaped pattern of coefficients similar to ours (Figure 2, page 53). Stambaugh's result confirms that the basic pattern is not driven by measurement error. Antti Ilmanen (1995) ran regressions of monthly excess returns on bonds in different countries on a term spread, the real short rate, stock returns, and bond return betas.

I. Bond return regressions

A. Notation

We use the following notation for log bond prices:

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

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

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

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

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

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

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

We denote excess log returns by

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

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

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

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

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

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

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

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

B. Excess return forecasts

We run regressions of bond excess returns at time $t+1$ on forward rates at time t. Prices, yields and forward rates are linear functions of each other, so the forecasts are the same for any of these choices of right hand variables. We focus on a one-year return horizon. We use the Fama-Bliss data (available from CRSP) of 1 through 5 year zero coupon bond prices, so we can compute annual returns directly.

We run regressions of excess returns on all forward rates,

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

The top panel of Figure 1 graphs the slope coefficients $\begin{bmatrix} \beta_1^{(n)} & \cdots & \beta_5^{(n)} \end{bmatrix}$ as a function of maturity n. (The Appendix includes a table of the regressions.) The plot makes the pattern clear – *the same function of forward rates forecasts holding period returns at all maturities. Longer maturities just have greater loadings on this same function.*

This beautiful pattern of coefficients cries for us to describe expected excess returns of all maturities in terms of a single factor, as follows:

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

b_n and γ_n are not separately identified by this specification, since you can double all the b and halve all the γ. We normalize the coefficients by imposing that the average value of b_n is one, $\frac{1}{4} \sum_{n=2}^{5} b_n = 1$.

We estimate (2) in two steps. First, we estimate the γ by running a regression of the average (across maturity) excess return on all forward rates,

$$(3) \qquad \frac{1}{4} \sum_{n=2}^{5} rx_{t+1}^{(n)} = \gamma_0 + \gamma_1 y_t^{(1)} + \gamma_2 f_t^{(2)} + \dots + \gamma_5 f_t^{(5)} + \bar{\varepsilon}_{t+1},$$

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

The second equality reminds us of the vector and average (overbar) notation. Then, we estimate b_n by running the four regressions

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

The single-factor model (2) is a restricted model. If we write the unrestricted regression coefficients from equation (1) as 4×6 matrix β, the single-factor model amounts to the restriction $\beta = b\gamma^\top$. A *single* linear combination of forward rates $\gamma^\top f_t$ is the state variable for time-varying expected returns of *all* maturities.

Table 1 presents the estimated values of γ and b, standard errors and test statistics. The γ estimates in Panel A are just about what one would expect from inspection of Figure 1. The loadings b_n of expected returns on the return-forecasting factor $\gamma^\top f$ in Panel B increase smoothly with maturity. The bottom panel of Figure 1 plots the coefficients of individual-bond expected returns on forward rates, as implied by the restricted model; i.e. for each n, it presents $\begin{bmatrix} b_n\gamma_1 & \cdots & b_n\gamma_5 \end{bmatrix}$. Comparing this plot with the unrestricted estimates of the top panel, you can see that the single-factor model almost exactly captures the unrestricted parameter estimates. The specification (2) constrains the constants $(b_n\gamma_0)$ as well as the regression coefficients plotted in Figure 1, and this restriction also holds closely. The unrestricted constants are $(-1.62, -2.67, -3.80, -4.89)$. The values implied from $b_n\gamma_0$ in Table 1 are similar, $(0.47, 0.87, 1.24, 1.43) \times (-3.24) = (-1.52, -2.82, -4.02, -4.63)$. The restricted and unrestricted estimates are close statistically as well as economically. The largest t-statistic for the hypothesis that each unconstrained parameter is equal to its restricted value is 0.9 and most of them are around 0.2. Section V. considers whether the restricted and unrestricted coefficients are *jointly* equal, with some surprises.

The right half of Table 1B collects statistics from unrestricted regressions (1). The unrestricted R^2 in the right half of Table 1B are essentially the same as the R^2 from the restricted model in the left half of Table 1B, indicating that the single-factor model's restrictions –

that bonds of each maturity are forecast by the *same* portfolio of forward rates – do little damage to the forecast ability.

C. Statistics and other worries

Tests for joint significance of the right hand variables are tricky with overlapping data and highly cross-correlated and autocorrelated right hand variables, so we investigate a number of variations in order to have confidence in the results. The bottom line is that the five forward rates are jointly highly significant, and we can reject the expectations hypothesis (no predictability) with a great deal of confidence.

We start with the Hansen-Hodrick correction, which is the standard way to handle forecasting regressions with overlapping data. (See the Appendix for formulas.) The resulting standard errors in Table 1 in Panel A ("HH, 12 lags") are reasonable, but this method produces a $\chi^2(5)$ statistic for *joint* parameter significance of 811, far greater than even the 1 percent critical value of 15. This value is suspiciously large. The Hansen-Hodrick formula does not necessarily produce a positive definite matrix in sample; while this one is positive definite, the 811 χ^2 statistic suggests a near-singularity. A χ^2 statistic calculated using only the diagonal elements of the parameter covariance matrix (the sum of squared individual t-statistics) is only 113. The 811 χ^2 statistic thus reflects *linear combinations* of the parameters that are apparently–but suspiciously–well measured.

The "NW, 18 lags" row of Table 1A uses the Newey-West correction with 18 lags instead of the Hansen-Hodrick correction. This covariance matrix is positive definite in any sample. It underweights higher covariances, so we use 18 lags to give it a greater chance to correct for the MA(12) structure induced by overlap. The individual standard errors in Table 1A are barely affected by this change, but the χ^2 statistic drops from 811 to 105, reflecting a more sensible behavior of the off-diagonal elements. 105 is still a long way above the 1 percent critical value of 15, so we still decisively reject the expectations hypothesis. The individual (unrestricted) bond regressions of Table 1B also use the NW, 18 correction, and reject zero

coefficients with χ^2 values near 100.

With this experience in mind, the following tables all report HH 12 lag standard errors, but use the NW, 18 lag calculation for joint test statistics.

Both Hansen-Hodrick and Newey-West formulas correct "nonparametrically" for arbitrary error correlation and conditional heteroskedasticity. If one knows the pattern of correlation and heteroskedasticity, formulas that impose this knowledge can work better in small samples. In the row labeled "Simplified HH," we ignore conditional heteroskedasticity, and we impose the idea that error correlation is due only to overlapping observations of homoskedastic forecast errors. This change raises the standard errors by about a third, and lowers the χ^2 statistic to 42, which is nonetheless still far above the 1 percent critical value.

As a final way to compute asymptotic distributions, we compute the parameter covariance matrix using regressions with nonoverlapping data. There are 12 ways to do this, January to January, February to February, and so forth, so we average the parameter covariance matrix by over these 12 possibilities. We still correct for heteroskedasticity. This covariance matrix is larger than the true covariance matrix, since by ignoring the intermediate though overlapping data we are throwing out information. Thus, we see larger standard errors as expected. The χ^2 statistic is 23, still far above the 1 percent level. Since we soundly reject using a too-large covariance matrix, we certainly reject using the correct one.

The small-sample performance of test statistics is always a worry in forecasting regressions with overlapping data and highly serially correlated right hand variables (e.g., Geert Bekaert et al. 1997), so we compute three small sample distributions for our test statistics. First, we run an unrestricted 12 monthly lag vector autoregression of all 5 yields, and bootstrap the residuals. This gives the "12 Lag VAR" results in Table 1, and the "Small T" results in the other tables. Second, to address unit and near-unit root problems we run a 12 monthly lag VAR that imposes a single unit root (one common trend) and thus four cointegrating vectors. Third, to test the expectations hypothesis ("EH" and "Exp. Hypo." in the tables), we run an unrestricted 12 monthly lag autoregression of the one year yield, bootstrap the

residuals, and calculate other yields according to the expectations hypothesis as expected values of future one year yields. See the Appendix for details.

The small-sample statistics based on the 12 lag yield VAR and the cointegrated VAR are almost identical. Both statistics give small-sample standard errors about a third larger than the asymptotic standard errors. We compute "small sample" joint Wald tests by using the covariance matrix of parameter estimates across the 50,000 simulations to evaluate the size of the sample estimates. Both calculations give χ^2 statistics of roughly 40, still convincingly rejecting the expectations hypothesis. The simulation under the null of the expectations hypothesis generates a conventional small-sample distribution for the χ^2 statistics. Under this distribution, the 105 value of the NW, 18 lags χ^2 statistic occurs so infrequently that we still reject at the 0 percent level. Statistics for unrestricted individual-bond regressions (1) are quite similar.

One might worry that the large R^2 come from the large number of right hand variables. The conventional adjusted $\bar{R}^2$ is nearly identical, but that adjustment presumes i.i.d. data which is not valid in this case. The point of adjusted $\bar{R}^2$ is to see whether the forecastability is spurious, and the χ^2 is the correct test that the coefficients are jointly zero. To see if the increase in R^2 from simpler regressions to all five forward rates is significant, we perform χ^2 tests of parameter restrictions in table 4 below.

To assess sampling error and overfitting bias in R^2 directly (sample R^2 is of course a biased estimate of population R^2), Table 1 presents small-sample 95 percent confidence intervals for the unadjusted R^2. Our 0.32-0.37 unrestricted R^2 in Table 1B lie well above the 0.17 upper end of the 95 percent R^2 confidence interval calculated under the expectations hypothesis.

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

D.   Fama-Bliss regressions

Fama and Bliss (1987) regressed each excess return against the same maturity forward spread, and provided classic evidence against the expectations hypothesis in long-term bonds. Forecasts based on yield spreads such as Campbell and Shiller (1991) behave similarly. Table 2 updates Fama and Bliss's regressions to include more recent data. The slope coefficients are all within one standard error of 1.0. Expected excess returns move essentially one-for-one forward spreads. Fama and Bliss's regressions have held up well since publication, unlike many other anomalies.

The multiple regressions and the single-factor model in Table 1 provide stronger evidence against the expectations hypothesis than do the updated Fama-Bliss regressions in Table 2 in many respects. Table 1 shows stronger χ^2 rejections for all maturities, and more than double Fama and Bliss's R^2. The Appendix shows that the return-forecasting factor drives out Fama-Bliss spreads in multiple regressions. Of course, the multiple regressions also suggest the attractive idea that a *single* linear combination of forward rates forecasts returns of *all* maturities, where Fama and Bliss, and Campbell and Shiller, relate each bond's expected return to a different spread.

E.   Forecasting stock returns

We can view a stock as a long-term bond plus cashflow risk, so any variable that forecasts bond returns should also forecast stock returns, unless a time-varying cashflow risk premium happens exactly to oppose the time-varying interest rate risk premium. The slope of the term structure also forecasts stock returns, as emphasized by Fama and French (1989), and this is important confirmation that the bond return forecast corresponds to a risk premium and not to a bond-market fad or measurement error in bond prices.

The first row of Table 3 forecasts stock returns with the bond return forecasting factor $\gamma^\top f$. The coefficient is 1.73, and statistically significant. The 5-year bond in Table 1 had a coefficient of 1.43 on the return-forecasting factor, so the stock return corresponds to a

somewhat longer duration bond, as one would expect. The 0.07 R^2 is less than for bond returns, but we expect a lower R^2 since stock returns are subject to cash flow shocks as well as discount rate shocks.

Regressions 2-5 remind us how the dividend yield and term spread forecast stock returns in this sample. The dividend yield forecasts with a 5 percent R^2. The coefficient is economically large: a one-percentage point higher dividend yield results in a 3.3 percentage point higher return. The R^2 for the term spread in the third regression is only 2 percent. The fourth regression suggests that the term spread and dividend yield predict different components of returns, since the coefficients are unchanged in multiple regressions and the R^2 increases. Neither d/p nor the term spread are statistically significant in this sample. Studies that use longer samples find significant coefficients.

The fifth and sixth regressions compare $\gamma^\top f$ with the term spread and d/p. The coefficient on $\gamma^\top f$ and its significance are hardly affected in these multiple regressions. The return-forecasting factor drives the term premium out completely.

In the seventh row, we consider an unrestricted regression of stock excess returns on all forward rates. Of course, this estimate will be noisy, since stock returns are more volatile than bond returns. All forward rates together produce an R^2 of 10 percent, only slightly more than the $\gamma^\top f$ R^2 of 7 percent. The stock return forecasting coefficients recover a similar tent shape pattern (not shown.) We discuss the eighth and ninth rows below.

II. Factor models

A. Yield curve factors

Term structure models in finance specify a small number of factors that drive movements in all yields. Most such decompositions find "level," "slope," and "curvature" factors that move the yield curve in corresponding shapes. Naturally, we want to connect the return-forecasting factor to this pervasive representation of the yield curve.

13

Since γ is a symmetric function of maturity, it has nothing to do with pure slope movements; linearly rising and declining forward curves and yield curves give rise to the same expected returns. (A linear yield curve implies a linear forward curve.) Since γ is tent-shaped, it is tempting to conclude it represents a curvature factor, and thus that the curvature factor forecasts returns. This temptation is misleading, because γ is a function of *forward rates*, not of yields. As we will see, $\gamma^\top f$ is not fully captured by *any* of the conventional yield-curve factors. It reflects a 4-5 year yield spread that is ignored by factor models.

Factor loadings and variance

To connect the return-forecasting factor to yield curve models, the top left panel of Figure 2 expresses the return-forecasting factor as a function of yields. Forward rates and yields span the same space, so we can just as easily express the forecasting factor as a function of yields,[2] $\gamma^{*\top} y_t = \gamma^\top f_t$. This graph already makes the case that the return-forecasting factor has little to do with typical yield curve factors or spreads. The return-forecasting factor has no obvious slope, and it is is curved at the *long* end of the yield curve, not the short-maturity spreads that constitute the usual curvature factor.

To make an explicit comparison with yield factors, the top right panel of Figure 2 plots the loadings of the first 3 principal components (or factors) of yields. Each line in this graph represents how yields of various maturities change when a factor moves, and also how to construct a factor from yields. For example, when the "level" factor moves, all yields go up about 0.5 percentage points, and in turn the level factor can be recovered from a combination that is almost a sum of the yields. (We construct factors from an eigenvalue decomposition of the yield covariance matrix. See the Appendix for details.) The slope factor rises monotonically through all maturities, and the curvature factor is curved at the *short* end of the yield curve. The return-forecasting factor in the top left panel is clearly not related to any of the first three principal components.

The level, slope, curvature, and two remaining factors explain in turn 98.6, 1.4, 0.03, 0.02 and 0.01 percent of the variance of yields. As usual, the first few factors describe

the overwhelming majority of yield variation. However, these factors explain in turn quite different fractions, 9.1, 58.7, 7.6, 24.3, and 0.3 percent, of the variance of $\gamma^\top f$. 58.7 means that the slope factor explains a large fraction of $\gamma^\top f$ variance. The return-forecasting factor $\gamma^\top f$ is correlated with the slope factor, which is why the slope factor forecasts bond returns in single regressions. However, 24.3 means that the fourth factor, which loads heavily on the 4-5 year yield spread and is essentially unimportant for explaining the variation of *yields*, turns out to be very important for explaining *expected returns*. We expect this feature from the top left panel of Figure 2 shows, $\gamma^\top f$ is related to movements in 4-5 year yields.

Forecasting with factors and related tests

Table 4 asks the central question: how well can we *forecast* bond excess returns using yield curve factors in place of $\gamma^\top f$? The level and slope factor together achieve a 22 percent R^2. Including curvature brings the R^2 up to 26 percent. This is still substantially below the 35 percent R^2 achieved by $\gamma^\top f$, i.e. achieved by including the last two other principal components.

Is the increase in R^2 statistically significant? We test this and related hypotheses in Table 4. We start with the slope factor alone. We run the restricted regression

$$\overline{rx}_{t+1} = a + b \times \text{slope}_t + \varepsilon_{t+1} = a + b \times \left(q_2^\top y_t \right) + v_{t+1}$$

where q_2 generates the slope factor from yields. We want to test whether the restricted coefficients a, $(b \times q_2)$ are jointly equal to the unrestricted coefficients γ^*. To do this is, we add 3 yields to the right hand side, so that the regression is again unconstrained, and exactly equal to $\gamma^\top f_t$,

(4) $$\overline{rx}_{t+1} = a + b \times \text{slope}_t + c_2 y_t^{(2)} + c_3 y_t^{(3)} + c_4 y_t^{(4)} + c_5 y_t^{(5)} + \bar{\varepsilon}_{t+1}.$$

Then, we test whether c_2 through c_5 are jointly equal to zero.[3] (So long as the right hand variables span all yields, the results are the same no matter which extra yields one includes.)

The hypothesis that slope, or any combination of level, slope and curvature, are enough to forecast excess returns is decisively rejected. For all three computations of the parameter covariance matrix, the χ^2 values are well above the five percent critical values and the p-values are well below one percent. The difference between 22 percent and 35 percent R^2 is statistically significant.

To help understand the rejection, the bottom left panel in Figure 2 plots the restricted and unrestricted coefficients. For example, the coefficient line labeled "level & slope" represents coefficients on yields implied by the restriction that only the level and slope factors forecast returns. The figure shows that the restricted coefficients are well outside *individual* confidence intervals, especially for 4 and 5 year maturity. The rejection is therefore straightforward, and does not rely on mysterious off-diagonal elements of the covariance matrix or linear combinations of parameters.

In sum, although level, slope, and curvature together explain 99.97 percent of the variance of yields, we still decisively reject the hypothesis that these factors alone are sufficient to forecast excess returns. The slope and curvature factors, curved at the *short* end, do a poor job of matching the unrestricted regression which is curved at the *long* end. The tiny 4-5 year yield spread is important for forecasting *all* maturity bond returns.

Simple spreads

Many forecasting exercises use simple spreads rather than the factors common in the affine model literature. To see if the tent-shaped factor really has more information than simple yield spreads, we investigated a number of restrictions on yields and yield spreads.

Many people summarize the information in Fama and Bliss (1987) and Campbell and Shiller (1991) by a simple statement that yield spreads predict bond returns. The "$y^{(5)} - y^{(1)}$" row of Table 4 shows that this specification gives the 0.15 R^2 value typical of Fama-Bliss or Campbell-Shiller regressions. However, the restriction that this model carries all the information of the return-forecasting factor is decisively rejected.

By letting the one and five year yield enter separately in the next row of Table 4, we allow

a "level" effect as well as the 5-1 spread ($y^{(1)}$ and $y^{(5)}$ is the same as $y^{(1)}$ and $\left[y^{(5)} - y^{(1)}\right]$). This specification does a little better, raising the R^2 value to 0.22, and cutting the χ^2 statistics down, but it is still soundly rejected. The 1 and 5 year yield carry about the same information as the level and slope factors above.

To be more successful, we need to add yields. The most successful three-yield combination is the 1, 4, and 5 year yields as shown in the last row of Table 4. This combination gives an R^2 of 33 percent, and it is not rejected with two of the three parameter covariance matrix calculations. It produces the right pattern of 1,4, and 5 year yields in graphs like the bottom left panel of Figure 2.

Fewer maturities

Is the tent-shape pattern robust to the number of included yields or forward rates? After all, the right hand variables in the forecasting regressions are highly correlated, so the pattern we find in multiple regression coefficients may be sensitive to the precise set of variables we include. The bottom right hand panel of Figure 2 is comforting in this respect: as one adds successive forward rates to the right hand side, one slowly traces out the tent-shaped pattern.

Implications

If yields or forward rates followed an *exact* factor structure then all state variables including $\gamma^\top f$ would be functions of the factors. However, when as in fact yields do not follow an exact factor structure, an important state variable like $\gamma^\top f$ can be hidden in the small factors that are often dismissed as minor specification errors. This observation suggests a reason why the return-forecast factor $\gamma^\top f$ has not been noticed before. Most studies *first* reduce yield data to a small number of factors and *then* look at expected returns. To see expected returns, it's important *first* to look at expected returns and *then* investigate reduced factor structures. A reduced factor representation for yields that is to capture the expected return facts in this data must include the return-forecasting factor $\gamma^\top f$ *as well as* yield curve factors such as level, slope and curvature, even though inclusion of the former will do almost nothing to fit yields, i.e. to reduce pricing errors.

B. Affine models

It has seemed trouble enough to modify term structure models to incorporate Fama-Bliss or Campbell-Shiller predictability (Gregory Duffee 2002, Qiang Dai and Kenneth Singleton 2002). Is it that much harder to incorporate the much greater predictability we find into such models? The answer is no. In fact, it is trivial: one can *construct* market prices of risk that generate *exactly* our return regressions in an affine model. This discussion also answers the question "is there any economic model that generates the observed pattern of bond return forecastability?"

Our task is to construct a time series process for a nominal discount factor (pricing kernel, transformation to risk-neutral measure, marginal utility growth, etc.) M_{t+1} that generates bond prices with the required characteristics via $P_t^{(n)} = E_t(M_{t+1}M_{t+2}\cdots M_{t+n})$. With $M_{t+1} = \beta u'(c_{t+1})/u'(c_t) \times 1/\pi_{t+1}$, we can as easily express the primitive assumptions with preferences and a time-series process for consumption growth and inflation. However, since we do not here compare bond prices to consumption and inflation data, we can follow the affine model tradition and specify the time-series process for M_{t+1} directly.

We want to end up with bond prices that satisfy the return-forecasting regression

$$(5) \qquad\qquad rx_{t+1} = \beta f_t + \varepsilon_{t+1}; \ cov(\varepsilon_{t+1}\varepsilon_{t+1}^\top) = \Sigma.$$

We work backwards from this end. Consider a discount factor of the form

$$(6) \qquad\qquad M_{t+1} = \exp\left(-y_t^{(1)} - \frac{1}{2}\lambda_t^\top \Sigma \lambda_t - \lambda_t^\top \varepsilon_{t+1}\right),$$

with normally distributed shocks ε_{t+1}. (We're constructing a model, so we can specify the distribution.) From $1 = E_t(M_{t+1}R_{t+1})$, one-period log excess returns must then obey

$$(7) \qquad\qquad E_t(rx_{t+1}^{(n)}) + \frac{1}{2}\sigma_t^2(rx_{t+1}^{(n)}) = cov_t(rx_{t+1}^{(n)}, \varepsilon_{t+1}^\top)\lambda_t.$$

The time-varying discount-factor coefficients λ_t are also the "market prices of risk" that determine how much a unit of covariance translates into an expected return premium. Now, in the notation of regression (5), condition (7) is

$$\beta f_t + \frac{1}{2} diag(\Sigma) = \Sigma \lambda_t.$$

Thus, we can ensure that the model represents the one period return regression correctly by the form (6) with the choice

(8)
$$\lambda_t = \Sigma^{-1} \left[\beta f_t + \frac{1}{2} diag(\Sigma) \right].$$

(This is the log version of Lars Hansen and Ravi Jagannathan's 1991 discount factor construction.)

The discount factor (6) is the basis of an affine term structure model, and that model generates exactly the return regression (5). (The model is a special case of the Andrew Ang and Monika Piazzesi 2003 discrete time exponential-Gaussian model.) Here is what that statement means. Write the VAR for prices corresponding to the return regression (5) as

(9)
$$p_{t+1} = \mu + \phi p_t + v_{t+1}; \ cov(v_{t+1}, v_{t+1}^\top) = V.$$

Since returns, yields, and forwards are all linear functions of each other, this log-price VAR carries all the information of the return regressions (5). Conversely, one can recover the first four rows of ϕ from the return regressions, since $rx_{t+1}^{(n)} = p_{t+1}^{(n-1)} - p_t^{(n)} + p_t^{(1)}$. The return shocks ε_{t+1} are exactly the first four price shocks v_{t+1}, and the return covariance matrix Σ is the first four rows and columns of the price shock covariance matrix V.

Now, forget that p_t in (9) represents prices. Treat (9) as a time-series process for a general vector of unobserved state variables, with normally distributed shocks. Suppose that the discount factor is related to the state variables by (6) and (8). (To write this out,

include the linear transformation from prices p_t to f_t and $y_t^{(1)}$ on the right hand sides of (6) and (8), so that (6) and (8) are specified in terms of the state variables p_t.)

Now we have a time-series process for the discount factor. We want to know, what prices are *generated* by this model? Is $\log E_t(M_{t+1}M_{t+2}\cdots M_{t+n})$ equal to the state variable $p_t^{(n)}$? The answer is yes. The prices that come *out* of the model are precisely the same as the the state variables that go *in* the model. In this way, we have, in fact, constructed an affine (prices are linear functions of state variables) model that generates the return regression.

The logic of the proof of this statement is transparent. We have constructed the discount factor (6) to exactly capture the one period yield $y_t^{(1)}$ and one period expected excess returns $E_t(rx_{t+1}^{(n)})$. But any price can always be trivially expressed as its payoff discounted by its expected return, so if a model correctly captures the latter it must correctly capture prices. Alas, the algebra required to show this simple point takes some space, so we relegate it to the Appendix. The Appendix also discusses the affine model in greater detail.

Affine models seem puzzling to newcomers. Why start with a price VAR and go through all this bother to end up in the same place? The point of an affine model is to exhibit a stochastic discount factor (or pricing kernel) consistent with the bond price dynamics in (9). We can use this model to predict prices for bonds of other maturities or to predict the prices of term structure derivatives in a way that leaves no arbitrage relative to the included bonds. The affine model by itself does not restrict the time-series process for yields. If one desires further restrictions on the time-series process for the data, such as a k-factor structure, one can simply add structure to the dynamics (9).

The discount factor also exhibits a possible time series process for marginal utility growth. It shows that there is *an* economic model that generates our bond return forecastability. However, while it is tempting to look at the time series properties of the discount factor M_{t+1} and try to relate them to aggregate consumption, inflation, and macroeconomic events, this is not a simple inquiry, as examination of Hansen–Jagannathan (1991) discount factors for stocks does not quickly lead one to the correct economic model. The result is alas *a*

candidate marginal utility process not *the* marginal utility growth process.

This example does no more than advertised: it is a discrete-time affine term structure model that reproduces the pattern of bond return predictability we find in the data at an annual horizon. It is not necessarily a good general-purpose term structure model. We have not specified how to fill in higher frequencies, what η, ψ in $p_{t+1/12} = \eta + \psi p_t + v_{t+1/12}$ imply (9) (i.e. $\psi^{12} = \phi$), or, better, what continuous-time process $dp_t = \mu(\cdot)dt + \sigma(\cdot)dz$ does so, *and* correctly fits the data, including conditional heteroskedasticity and the non-Markovian structure we find below. We have not specified what the monthly or instantaneous market prices of risk and discount factor look like that generate (6) at an annual horizon. (Pierre Collin-Dufresne and Robert Goldstein 2003 write a term structure model that incorporates our forecasts and addresses some of these issues.)

III.   Lags and measurement error

A.   Single-lag regressions

Are lagged forward rates are useful for forecasting bond excess returns? Measurement error is the first reason to investigate this question. A spuriously high price at t will erroneously indicate that high prices at t forecast poor returns from t to $t + 1$.   If the measurement error is serially uncorrelated, however, the forecast using a one month lag of the forward rates is unaffected by measurement error, since the price at time t is no longer common to left and right hand sides of the regression.[4] Therefore, we run regressions of average (across maturity) returns $\overline{rx}_{t+1}$ on forward rates $f_{t-i/12}$ that are lagged by i months.

Figure 3 plots the coefficients. The basic tent shape is unaltered at the first and second lags. As we move the right-hand variables backward through time, the shape moves slightly to the right, and this pattern continues through the first year of additional lags (not shown). The reason for this shift is that the return-forecasting factor $\gamma^\top f$ is not Markovian; if we predict $\gamma^\top f_t$ from $f_{t-i/12}$, the coefficients do not follow the tent-shaped pattern. Thus, as

we forecast returns with longer lags of forward rates, the coefficients mix together the tent-shaped pattern by which f_t forecasts returns rx_{t+1} with the pattern by which $f_{t-i/12}$ predicts $\gamma^\top f_t$. The R^2 decline slowly with horizon, from 35 percent at lag 0 to 35, 32, and 31 percent at lags 1, 2, and 3, and the coefficients are highly jointly significant. The same pattern holds in forecasts of individual bond returns $rx_{t+1}^{(n)}$ on lagged forward rates (not shown): the single-factor model is as evident using lagged forward rates as it is using the time-t forward rates.

We conclude that serially uncorrelated measurement errors, or even errors that are somewhat correlated over time, are not the reason for the forecastability we see, including the large R^2, the single-factor structure and the tent-shaped pattern of coefficients. The most natural interpretation of these regressions is instead that $\gamma^\top f$ reflects a slow-moving underlying risk premium, one almost as well revealed by $f_{t-i/12}$ as it is by f_t.

Measurement error also does not easily produce the tent-shaped return-forecasting factor or a single-factor model; it does not produce regressions in which (say) the *five*-year yield helps to forecast the *two*-year bond return. Figure 4 plots the entirely spurious coefficients that result from $rx_{t+1}^{(n)} = \beta^{(n)} f_t + \varepsilon_{t+1}^{(n)}$ if excess returns are truly unpredictable, but there is measurement error in prices or forward rates. (See the Appendix for details.) The coefficients are step functions of forward rates, not tents. The coefficients for the n-period bond excess return simply load on the n-year yield and the one-year yield. The pattern of coefficients differs completely across maturity rather than display a single-factor structure. The only way to generate a single-factor prediction from measurement error is if measurement error at time t is correlated with measurement error one year later at time $t+1$, and in just the right way across bonds, which seems rather far-fetched.

B.   Multiple-lag regressions

Do multiple lags of forward rates help to forecast bond returns? We find that they do, and again with an attractive factor structure. We started with unrestricted regressions.

We found that multiple regression coefficients displayed similar tent shapes across maturity, much like the single-lag regression coefficients of Figure 3, and once again bonds of different maturity had the same pattern of regression coefficients blown up by different amounts. These observations suggest a single-factor structure across time as well as maturity,

$$(10) \qquad rx_{t+1}^{(n)} = b_n \gamma^\top \left[\alpha_0 f_t + \alpha_1 f_{t-\frac{1}{12}} + \alpha_2 f_{t-\frac{2}{12}} + ... + \alpha_k f_{t-\frac{k}{12}} \right] + \varepsilon_{t+1}^{(n)}.$$

We normalize to $\sum_{j=0}^{k} \alpha_j = 1$ so that the units of γ remain unaffected. Since we only add one parameter (α_k) per new lag introduced, this specification gives a more believable forecast than an unrestricted regression with many lags. Since the single factor restriction works well across maturities n, we present only the results for forecasting average returns across maturity, corresponding to Table 1A,

$$(11) \qquad \overline{rx}_{t+1} = \gamma^\top \left[\alpha_0 f_t + \alpha_1 f_{t-\frac{1}{12}} + \alpha_2 f_{t-\frac{2}{12}} + ... + \alpha_k f_{t-\frac{k}{12}} \right] + \bar{\varepsilon}_{t+1}.$$

We can also write the regression as

$$(12) \qquad \overline{rx}_{t+1} = \alpha_0 \left(\gamma^\top f_t \right) + \alpha_1 \left(\gamma^\top f_{t-\frac{1}{12}} \right) + \alpha_2 \left(\gamma^\top f_{t-\frac{2}{12}} \right) + ... + \alpha_k \left(\gamma^\top f_{t-\frac{k}{12}} \right) + \bar{\varepsilon}_{t+1};$$

we can think of the restricted model as simply adding lags of the return-forecasting factor $\gamma^\top f_t$.

Table 5 presents estimates[5] of this model. The R^2 rise from 35 to 44 percent once we have added 3 additional lags. The γ estimates change little as we add lags. They do shift to the right a bit, as we found for the single lag regressions in Figure 3. The restriction that γ is the same for each lag is obviously not one to be used for extreme numbers of lags, but there is little credible gain to generalizing it for the current purpose.

With two lags, the α estimate wants to use a moving average of the first and second lag. As we add lags, a natural pattern emerges of essentially geometric decay. The α and the

marginal forecast power of additional lags that they represent are statistically significant out to the fourth additional lag.

The b loadings across bonds (not shown) are essentially unchanged as we add lags, and the R^2 of the model that imposes the single-factor structure across maturity are nearly identical to the R^2 of unrestricted models, as in Table 1B. Returns on the individual bonds display the single-factor structure with the addition of lagged right hand variables.

Multiple lags also help when we forecast stock excess returns. As indicated by row 8 in Table 3, the R^2 for stocks rises from 7 percent to 12.5 percent when we include 3 additional lags. The slope coefficient rise to 2.11, which corresponds well to a bond substantially longer than 5 years. As shown in the last row, the multiple regression with term spreads and d/p is also stronger than its counterpart with no lags.

C.   The failure of Markovian models

The importance of extra monthly lags means that a VAR(1) *monthly* representation, of the type specified by nearly every explicit term structure model, does not capture the patterns we see in *annual* return forecasts. To see the annual return forecastability, one must either look directly at annual horizons as we do, or one must adopt more complex time-series models with lags.

To quantify this point, we fit an unconstrained monthly yield VAR, $y_{t+1/12} = Ay_t + \varepsilon_{t+1/12}$, and we find the implied annual VAR representation $y_{t+1} = A^{12}y_t + u_{t+1}$. Yields, forward rates, and returns are all linear functions of each other, so we can calculate return forecasts and R^2 directly from this annual VAR representation. Figure 5 presents the *annual* excess return forecasting coefficients for individual bonds implied by this *monthly* VAR(1). The coefficients here are nothing at all like the tent shape we find in the direct annual horizon regressions or VAR. They are most reminiscent of the coefficients resulting from pure measurement error in Figure 4: the n-year yield is the most important forecasting variable for the one year return on the n-year bond, and the forward rate coefficients, cumulating these, look like step

functions. Most importantly, these coefficients hide the single-factor structure: no single combination of yields or forward rates summarizes return forecastability of bonds across different maturity. The forecast R^2 are cut to 22-26 percent instead of 32-37 percent – and given the pattern, one would be well justified in dismissing the whole thing as measurement error. (The implied forecasts of average (across maturity) returns $\overline{rx}_{t+1}$ do reveal the usual shapes, as one might guess by visually averaging the lines in Figure 5, but in much attenuated form, and also with only 24 percent R^2.) By focusing directly on an annual horizon, our original regressions raise the signal-to-noise ratio and uncover the underlying return forecast. (One can alternatively use more complex models to get the same result. For example, the 12 lag monthly VAR we use for bootstraps does imply the tent-shaped single-factor model and the large forecast R^2.)

This observation suggests a second reason why such an important feature of the yield curve could go unrecognized for so long. Yield curve modelers almost always specify and estimate a high frequency model, using monthly or finer data, and never include lags. If they look at all at annual horizon return forecasts, they find the implied forecasts from these high-frequency Markovian specifications. But since the lags *do* matter, these implied annual forecasts miss much predictability and completely miss the one-factor structure.

D.   Summary and implications

Additional lags matter. But additional lags are awkward bond forecasting variables. Bond prices are time-t expected values of future discount factors, so a full set of time-t bond yields should drive out lagged yields in forecasting regressions:[6] $E_t(x)$ drives out $E_{t-1/12}(x)$ in forecasting any x. Bond prices *reveal* all other important state variables. For this reason, term structure models do not include lags.

The most natural interpretation of all these results is instead that time-t yields (or prices, or forwards) truly are sufficient state variables, but there are small measurement errors that are poorly correlated over time. For example, if a true variable $x_t = \gamma^\top f_t$ is an AR(1)

$x_t = \rho x_{t-1} + u_t$, but is measured with error $\tilde{x}_t = x_t + v_t$, v_t i.i.d., then the standard Kalman filter says that the best guess of the true x_t is a geometrically weighted moving average $E(x_t | \tilde{x}_t, \ \tilde{x}_{t-1}, \ \tilde{x}_{t-2}, \ ...) = b \sum_{j=0}^{\infty} \phi^j \tilde{x}_{t-j}$. This is the pattern that we find in Table 5.

However, measurement error clearly is not the driving force behind return forecastability or the nature of the return-forecasting factor. The tent-shaped single factor is revealed and the R^2 is still large using single lags of forward rates, which avoid the effects of measurement error. Measurement error produces multifactor models (each return forecast by its own yield), not a single-factor model. We reach the opposite conclusion: Measurement error *conceals* the underlying single-factor model unless one takes steps to mitigate it – for example looking directly at annual horizons as we do – and even then the measurement error cuts down the true R^2 from over 44 percent to "only" 35 percent.

IV.   History

One wants to see that the regression is robust, that it is driven by clear and repeated stylized facts in the data, that it is not an artifact of one or two data points. To this end, we examine the historical experience underlying the regressions.

Figure 6 plots the forecast and actual average (across maturity) excess returns, along with the history of yields. The top line averages the Fama-Bliss forecasts from Table 2 across maturity. The second line gives the return-forecasting factor. We present the forecast using three lags, $E_t(\overline{rx}_{t+1}) = \gamma^\top \sum_{j=0}^{3} \alpha_j f_{t-\frac{j}{12}}$; the forecast using no lags is nearly indistinguishable. The ex-post return line is shifted to the left one year so that the forecast and its result are plotted at the same time period.

The graph verifies that the forecasting power of $\gamma^\top f$ and the improved R^2 over the Fama-Bliss and other slope-based regressions is not driven by outliers or suspicious forecasts of one or two unusual data points. The return forecast has a clear business cycle pattern, high in troughs and low at peaks. The results do not depend on full-sample regressions, as the

forecast using real-time regressions (from 1964 to t, not shown) is quite similar.

Comparing the return-forecasting factor and the forward spreads embodied in the Fama-Bliss regressions, we see that they agree in many instances. (Forecasts based on the slope of the term structure or simple yield spreads are quite similar to the Fama-Bliss line.) One pattern is visible near 1976, 1982, 1985, 1994, and 2002: In each case there is an upward sloping yield curve that is *not* soon followed by rises in yields, giving good returns to long-term bond holders.

However, the figure also shows the much better fit of the return-forecasting factor in many episodes, including the late 1960s, the turbulent early 1980s, and the mid 1990's. The return-forecasting factor seems to know better when to get out; when the long-anticipated (by an upward sloping yield curve) rise in rates will actually come, leading to losses on long term bonds. This difference is dramatic in 1983, 1994, and 2003.

What information is $\gamma^\top f$ using to make these forecasts? To answer this question, Figure 7 plots the forward curves at selected peaks (solid) and troughs (dashed) of the return-forecasting factor. These dates are marked by vertical lines in Figure 6. In December 1981, the return-forecasting factor signaled a good excess return on long term bonds, though the yield curve was nearly flat and the Fama-Bliss regression forecast nothing. Figure 7 shows why: though flat, the forward curve has a nice tent shape, which multiplied by γ gives a strong forecast. As shown in Figure 6, that forecast was correct. By April of 1983, however, $\gamma^\top f$ forecast essentially no excess return on long term bonds (Figure 6), despite the strong upward slope, because the yield curve curves up (Figure 7). The Fama-Bliss and other slope forecasts are optimistic, but the return-forecasting factor proves right, and long term bonds have worse than -5 percent excess return as interest rates rise over the next year.

In August of 1985, the level and slope of the term structure are about the same as they were in April of 1983, and the slope-based forecast is the same, mildly optimistic. But the return-forecasting factor predicts a strong, nearly 10 percent expected excess return, and this return is borne out ex-post. What's the difference? April 1983 and August 1985 have

the same level and slope, but April 1983 is overall flat, with a slight upward curve at the long end, while August 1985 curves down. Interacted with the tent-shaped γ operator, we get the correct strong, positive, and correct forecast.

The peak and trough of April 1992 and August 1993 provide a similar comparison. In April 1992, the yield curve is upward sloping and has a tent-shape, so both the return-forecasting factor and the slope-based forecasts are optimistic. During the subsequent year, yields actually went *down*, giving good returns to long-term bond holders. But by August 1993, the slope is still there but the forward curve has lost its tent-shape. The Fama-Bliss and other slope forecasts are still positive, but the return-forecasting factor gives the order to bail out. That order proves correct, as yields do rise sharply in the following year giving terrible returns to long-term bond holders. The same pattern has set up through the recession of 2000, and the slope forecast and $\gamma^\top f$ disagree once again at the end of the sample.

In sum, the pattern is robust. A forecast that looks at the tent-shape, ignoring the slope, has made better forecasts in each of the important historical episodes.

V. Tests

A. Testing the single-factor model

The parameters of the unrestricted $(rx_{t+1} = \beta f_t + \varepsilon_{t+1})$ return forecasting regressions and those of the restricted single-factor model $(rx_{t+1} = b\left(\gamma^\top f_t\right) + \varepsilon_{t+1})$ are very close *individually* (Figure 1), and well inside one-standard error bounds, but are they *jointly* equal? Does an overall test of the single-factor model reject?

The moments underlying the unrestricted regressions (1) are the regression forecast errors multiplied by forward rates (right hand variables),

$$(13) \qquad\qquad g_T(\beta) = E\left(\varepsilon_{t+1} \otimes f_t\right) = 0.$$

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

$$E\left[\left(1_4^\top \varepsilon_{t+1}\right) \otimes f_t\right] = 0, \tag{14}$$

$$E\left[\varepsilon_{t+1} \otimes \left(\gamma^\top f_t\right)\right] = 0. \tag{15}$$

We use moments (14)-(15) to compute GMM standard errors in Table 1. The restricted model $\beta = b\gamma^\top$ does not set all the moments (13) to zero, $g_T(b\gamma^\top) \neq 0$. We can compute the "J_T" χ^2 test that the remaining moments are not too large. To do this, we express the moments (14)-(15) of the single-factor model as linear combinations of the unrestricted regression moments (13), $a_T g_T = 0$. Then we apply Lars Hansen's (1982) Theorem 3.1 (Details in the Appendix). We also compute a Wald test of the joint parameter restrictions $\beta = b\gamma^\top$. We find the GMM distribution $cov\left[vec(\beta)\right]$, and then compute the χ^2 statistic $\left[vec(b\gamma^\top) - vec(\beta)\right]^\top cov\left[vec(\beta)\right]^{-1} \left[vec(b\gamma^\top) - vec(\beta)\right]$. ($vec$ since β is a matrix.)

Table 6 presents the tests. Surprisingly, given how close the parameters are in Figure 1, the tests in the first two rows all decisively reject the single-factor model. The NW, 18 S matrix again produces suspiciously large χ^2 values, but tests with the simplified S matrix, the S matrix from nonoverlapping data, and the small sample also give strong rejections.

When we consider a forecast lagged one month however, the evidence against the single-factor model is much weaker. The asymptotic χ^2 values are cut by factors of 5-10. The simple S and nonoverlapping tests no longer reject. The small-sample χ^2 values also drop considerably, in one case no longer rejecting.

B. Diagnosing the rejection

To understand the single-factor model rejection with no lags, we can forecast single-factor model failures. We can estimate $\tilde{\Gamma}$ in ,

$$rx_{t+1} - b \times \overline{rx}_{t+1} = \tilde{\Gamma}^\top f_t + w_{t+1}. \tag{16}$$

The left hand side is a portfolio long the excess return of the nth bond, and short b_n times the average of all excess returns. The restriction of the single-factor model is precisely that such portfolios should not be forecastable. (Since $E_t\left(\overline{rx}_{t+1}\right) = \gamma^\top f_t$, we can equivalently put $rx_{t+1} - b \times \gamma^\top f_t$ on the left hand side. Here, we check whether individual forward rates can forecast a bond's return, above and beyond the constrained pattern $b\gamma^\top f_t$.)

These regressions amount to a characterization of the unconstrained regression coefficients in $rx_{t+1} = \beta f_t + \varepsilon_{t+1}$. The coefficients $\tilde{\Gamma}$ are simply the difference between the unconstrained and constrained regression coefficients, $\tilde{\Gamma}^\top = \beta - b\gamma^\top$. The Wald test in Table 6 is precisely a test of joint significance of these coefficients. Here, we examine the coefficients for interpretable structure.

Table 6 presents regressions of the form (16). We use yields rather than forward rates on the right hand side, as yields paint a clearer picture. A pattern emerges: The diagonals (emphasized by boldface type) are large. If the two-year yield is a little high (first row, 0.80) relative to the other yields, meaning that the two-year bond price is low, then two-year bond returns will be better than the one factor model suggests, $rx_{t+1}^{(2)} - b_2\overline{rx}_{t+1}$ will be large. Similarly, if the 3, 4, or 5-year yields are higher than the others (2.36, 1.84, 0.72), then the 3, 4 and 5 year bonds will do better than the single-factor model suggests. The forecasts are *idiosyncratic*. There is no single factor here: each bond is forecast by a different function of yields, and the pattern is quite similar to the pattern induced by measurement error in the bottom panel of Figure 4.

These forecasts are *statistically* significant, with some impressive t-statistics, and hence they cause the statistical rejection of the single-factor model with no lags. The R^2 from 0.12 to 0.37 are at least as good as the Fama-Bliss forecasts and sometimes as good as the single-factor forecasts of overall excess returns.

So why did the single-factor model look like such a good approximation? Because the deviations from the single-factor model are *tiny*. The standard deviation of expected portfolio returns $\sigma(\tilde{\gamma}^\top y)$ ranges from 17 to 21 basis points, and that of the ex-post portfolio returns

σ(lhs) ranges from 30 to 61 basis points. By contrast, the standard deviation of expected excess returns $\sigma(b^{(n)}\gamma^\top y)$ ranges from 1.12 to 3.45 percentage points, and that of ex-post excess returns ranges from 1.93 to 6.00 percentage points. Tiny returns are forecast by tiny yield spreads, but with good R^2 and statistical significance.

As another way to capture the structure of expected returns, we perform a principal components analysis of *expected* returns by an eigenvalue decomposition of $cov(\beta f_t f_t^\top \beta^\top)$. (Details in the Appendix.) The first principal component is almost exactly the return-forecast factor $\gamma^\top$, and bond returns load on it with almost exactly the b loadings we found above. This first principal component is by far the largest: The standard deviations of the principal components are 5.16, 0.26, 0.20 and 0.16 percentage points. As fractions of variance, they account for 99.51, 0.25 0.15, and 0.09 percent of the total variance of expected returns. Thus, the first factor, $\gamma^\top f$, totally dominates the variance of expected returns.

This is the heart of our result: If one forms an eigenvalue decomposition of the covariance matrix of yields, yield changes, prices, forwards, ex-post returns, or just about any other characteristic of the term structure, one obtains "level," "slope" and "curvature" components, which account for almost all variation. If one forms an eigenvalue decomposition of the covariance matrix of *expected excess returns*, however, the tent-shaped $\gamma^\top f$, poorly related to level, slope and curvature, is by far the dominant component.

But the remaining components *are* statistically significant, and that is why the single-factor model with no lags is rejected. To see them, we have to finely tune our microscope. If we forecast the returns of portfolios $\delta^\top rx_{t+1}$, using portfolio weights δ with $\delta^\top b = 0$, then the single-factor model predicts $E_t(\delta^\top rx_{t+1}) = \delta^\top b\left(\gamma^\top f_t\right) = 0$. The left hand side of (16) and Table 6 give one simple set of such portfolios. This prediction turns out to be false, these portfolios can be predicted, with patterns nothing like γ. But with any other set of portfolios, with portfolios $\delta^\top rx_{t+1}$ with $\delta^\top b$ even slightly different from 0, the first factor will overwhelm the smaller additional factors, so the portfolio will be forecast with a pattern very close to $\gamma^\top f_t$.

Repeating regressions of the form of Table 7 with lagged forward rates, we obtain much smaller forecasts; some t statistics are a bit above 2, as some of the tests in Table 6 still suggested statistical rejections, but as in Table 6, this depends on how one calculates the test statistics. Most importantly, there is no interpretable pattern to the coefficients, which leads us further to discount evidence against the single-factor model when we lag the right hand variables.

C.   Interpretation

What should we make of these additional return-forecasting factors? They are very small. The factors represent small movements of the bond yields, and they forecast small returns to the corresponding portfolios. They are also idiosyncratic; there is no common structure. When the nth bond price is a bit low (yield is a bit high), that bond has a high subsequent return. Furthermore, the phenomenon only lasts one month, as the evidence against the single-factor model with a lagged forecasting variable is much weaker. Thus, the failures seem to represent one-month serially uncorrelated pricing discrepancies.

Even though they are small, they are not necessarily "economically insignificant." The portfolios are zero-cost, so a huge short position matched by a huge long position can make a lot of money from a small pricing discrepancy with a 35 percent R^2. A bond trader's business is precisely to pick individual bonds that are a little over- or underpriced relative to a yield curve, hedge and leverage as much as possible, and wait for the small price fluctuation to melt away. One needs to ask whether 20 basis point pricing discrepancies survive real-world transactions costs of course, but our interpolated zero-coupon bond data are too far removed from transactions data to really answer this question.

On the other hand, they could simply be measurement errors, and we favor this interpretation. The facts are entirely consistent with a measurement error interpretation. Serially uncorrelated measurement error will cause multiple lags to help to forecast bond returns, it will cause a rejection of the single-factor model with zero lags, a failure to reject the single-

factor model with lagged right hand variables, and it will cause the idiosyncratic nature of the forecasts of failures seen in Table 6. If the failure represents real pricing anomalies, there is no reason those anomalies should be idiosyncratic, rather than follow structures with additional factors that move bonds of all maturities. And, most tellingly, additional factors in expected returns cannot induce the non-Markovian structure that additional lags help to forecast returns.

In any case, the *economically* interesting variation in expected bond returns (to all but highly leveraged, low transactions cost traders) is clearly represented by the single-factor model, which explains 99.5 percent of the variance of bond expected returns as forecast by forward rates. 99.5 percent of the variance of bond expected returns as forecast by forward rates.

D.   Two step vs. efficient estimates

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

The crucial question is, what moments will GMM use to choose γ, the linear combination of forward rates that forms the single factor? Once the single-factor parameter γ is estimated, even efficient GMM estimates the remaining b coefficients by regressions of each return on $\gamma^\top f_t$. In turn, taking linear combinations of moments is the same thing as forming a portfolio, so the crucial question becomes, "which single portfolio of excess returns $\delta^\top rx_{t+1}$ will efficient GMM regress on all forward rates to estimate γ?" The answer is that efficient GMM pays attention to *well-measured* linear combinations of moments, guided by S, not "large" or

"economically interesting" moments.

For example, suppose that the third row of Table 7A forecasts its corresponding linear combination of returns $rx_{t+1}^{(4)} - b_4 \overline{rx}_{t+1}$ with 100 percent R^2. This moment is exactly measured, so efficient GMM will estimate parameters γ to fit this regression exactly – it will report the third row of Table 7A as the "single-factor model." Efficient GMM will *completely miss* the economically interesting first factor that describes 99.5 percent of the variance of expected returns.

In our data, the R^2 for all expected return "factors" are roughly comparable (Table 7B), and thus all are about as well measured. That fact means that efficient GMM pays about equal attention to all the regressions in Table 7B along with the regression of $\overline{rx}_{t+1} = \gamma^\top f_t + \bar{\varepsilon}_{t+1}$, producing a single factor that is roughly an average of γ and the rows of Table 7B. The resulting single factor explains very little of the variance of expected excess returns.

We want a GMM estimate of the approximate single factor that explains most of the variance of expected returns, not the one that minimizes the best measured, even if tiny, moments. For that purpose, we GMM to pay attention to a portfolio such as $1^\top rx_{t+1}$, as in the two step procedure. The return-forecasting factor is so dominant that it doesn't really matter which linear combination we choose, so long as we keep GMM from paying attention to the special linear combinations $\delta^\top rx_{t+1}$ with $\delta^\top b = 0$ that produce our very small, but statistically significant, additional factors.

VI.   Concluding remarks

The Appendix presents many additional results and robustness checks. We show that the return forecasts are stable across subsamples, they survive in real-time estimates, they generate trading rule profits, and they are verified in a different dataset. The Appendix also shows how $\gamma^\top f$ forecasts the short rate of interest; it presents individual-bond forecasting regressions; and it details the principal components analysis of expected excess returns. The

Appendix also documents the computations and gives an detailed presentation of the affine model.

This analysis is still admittedly incomplete in many respects. We only examine 1-5 year maturity bonds at a one year horizon. Understanding how expected returns and interest rate forecasts vary across investment horizon, and extending the analysis to all maturities are of course both important issues. However, these extensions require some subtle time-series analysis, probably including an explicit treatment of measurement errors. They also require extending the data beyond zero coupon bonds at one year spaced maturities, and integration with an explicit term structure model when returns cannot be measured directly. We do not mention variances or covariances, and hence we do not mention Sharpe ratios or optimal portfolios. Modeling time-varying second moments is likely to require as much effort as modeling the first moments in this paper. Finally, we do not tie the time-varying premia to macroeconomic or monetary fundamentals. In particular, we do not offer a deep interpretation of what the tent-shaped function of forward rates or the 4-5 year yield spread means, other to note that the result is suggestively correlated with business cycles. All these and more are important extensions, but we have strained space and the reader's patience enough for now.

References

[1] Ang, Andrew and Monika Piazzesi. "A No-Arbitrage Vector Autoregression of Term Structure Dynamics with Macroeconomic and Latent Variables." *Journal of Monetary Economics*, 2003, (50), pp. 745-787.

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

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

[4] Collin-Dufresne, Pierre and Robert Goldstein. "Generalizing the Affine Framework to HJM and Random Field Models." Working paper, Berkeley, 2003.

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

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

[7] Duffie, Darrell and Rui Kan. "A Yield-Factor Model of Interest Rates." *Mathematical Finance*, 1996, (6), pp. 379-406.

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

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

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

[11] Hansen, Lars Peter. "Large Sample Properties of Generalized Method of Moments Estimators." *Econometrica*, 1982, (50), pp. 1029-1054.

[12] Hansen, Lars Peter and Robert J. Hodrick. "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: University of Chicago Press, 1983.

[13] Hansen, Lars Peter and Ravi Jagannathan. "Implications of Security Market Data for Models of Dynamic Economies." *Journal of Political Economy*, 1991, (99), pp. 225-62.

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

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

Notes

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

[2]The yield coefficients γ^* are given from the forward rate coefficients γ by

$$\gamma^{*\top}y = (\gamma_1 - \gamma_2)y^{(1)} + 2(\gamma_2 - \gamma_3)y^{(2)} + 3(\gamma_3 - \gamma_4)y^{(3)} + 4(\gamma_4 - \gamma_5)y^{(4)} + 5\gamma_5 y^{(5)}.$$

This formula explains the big swing on the right side of the graph: The tent-shaped γ are multiplied by maturity, and the γ^* are based on differences of the γ.

[3]In GMM language, the unrestricted moment conditions are $E\left[y_t\bar{\varepsilon}_{t+1}\right]$. The restrictions set linear combinations of these moments to zero, $E\left[\bar{\varepsilon}_{t+1}\right]$ and $q_2^\top E\left[y_t\bar{\varepsilon}_{t+1}\right]$ in this case. The Wald test on c_2 through c_5 in (4) is identical to the overidentifying restrictions test that the remaining moments are zero.

[4]Stambaugh (1988) addressed the same problem by using different bonds on the right and left hand side. Since we use interpolated zero-coupon yields, we cannot use his strategy.

[5]For a given set of α, we estimate γ by running regression (11). Then, fixing γ, we estimate the α by running regression (12). When this procedure has iterated to convergence, we estimate bs from (10). The moment conditions are the orthogonality conditions corresponding to regressions (10)-(12). The parameters enter nonlinearly, so a search or this equivalent iterative procedure are necessary to find the parameters.

[6]Of course, the true state vector might use more than 5 yields, so lags enter because they are correlated with omitted yields. This seems to us like a remote possibility for explaining the results.

Table 1. Estimates of the single-factor model

A. Estimates of the return-forecasting factor, $\overline{rx}_{t+1} = \gamma^\top f_t + \bar{\varepsilon}_{t+1}$

	γ_0	γ_1	γ_2	γ_3	γ_4	γ_5	R^2	$\chi^2(5)$
OLS estimates	-3.24	-2.14	0.81	3.00	0.80	-2.08	0.35	
Asymptotic (Large T) distributions								
HH, 12 lags	(1.45)	(0.36)	(0.74)	(0.50)	(0.45)	(0.34)		811.3
NW, 18 lags	(1.31)	(0.34)	(0.69)	(0.55)	(0.46)	(0.41)		105.5
Simplified HH	(1.80)	(0.59)	(1.04)	(0.78)	(0.62)	(0.55)		42.4
No overlap	(1.83)	(0.84)	(1.69)	(1.69)	(1.21)	(1.06)		22.6
Small-sample (Small T) distributions								
12 lag VAR	(1.72)	(0.60)	(1.00)	(0.80)	(0.60)	(0.58)	$[0.22, 0.56]$	40.2
Cointegrated VAR	(1.88)	(0.63)	(1.05)	(0.80)	(0.60)	(0.58)	$[0.18, 0.51]$	38.1
Exp. Hypo.							$[0.00, 0.17]$	

B. Individual-bond regressions

		Restricted, $rx_{t+1}^{(n)} = b_n\left(\gamma^\top f_t\right) + \varepsilon_{t+1}^{(n)}$				Unrestricted, $rx_{t+1}^{(n)} = \beta_n f_t + \varepsilon_{t+1}^{(n)}$			
n	b_n	Large T	Small T	R^2	Small T	R^2	EH	Level R^2	$\chi^2(5)$
2	0.47	(0.03)	(0.02)	0.31	$[0.18, 0.52]$	0.32	$[0, 0.17]$	0.36	121.8
3	0.87	(0.02)	(0.02)	0.34	$[0.21, 0.54]$	0.34	$[0, 0.17]$	0.36	113.8
4	1.24	(0.01)	(0.02)	0.37	$[0.24, 0.57]$	0.37	$[0, 0.17]$	0.39	115.7
5	1.43	(0.04)	(0.03)	0.34	$[0.21, 0.55]$	0.35	$[0, 0.17]$	0.36	88.2

Notes: The 10 percent , 5 percent and 1 percent critical values for a $\chi^2(5)$ are 9.2, 11.1 and 15.1 respectively. All p-values are less than 0.005. Standard errors in parentheses (), 95% confidence intervals for R^2 in square brackets []. Monthly observations of annual returns, 1964-2003.

Table 2. Fama-Bliss excess return regressions

Maturity n	β	Small T	R^2	$\chi^2(1)$	p-val	EH p-val
2	0.99	(0.33)	0.16	18.4	$\langle 0.00 \rangle$	$\langle 0.01 \rangle$
3	1.35	(0.41)	0.17	19.2	$\langle 0.00 \rangle$	$\langle 0.01 \rangle$
4	1.61	(0.48)	0.18	16.4	$\langle 0.00 \rangle$	$\langle 0.01 \rangle$
5	1.27	(0.64)	0.09	5.7	$\langle 0.02 \rangle$	$\langle 0.13 \rangle$

Notes: The regressions are $rx_{t+1}^{(n)} = \alpha + \beta \left(f_t^{(n)} - y_t^{(1)} \right) + \varepsilon_{t+1}^{(n)}$. Standard errors are in parentheses "()", probability values in angled brackets "<>". The 5 percent and 1 percent critical values for a $\chi^2(1)$ are 3.8 and 6.6.

Table 3. Forecasts of excess stock returns

	Right hand variables	$\gamma^\top f$	(t-stat)	d/p	(t-stat)	$y^{(5)} - y^{(1)}$	(t-stat)	R^2
1	$\gamma^\top f$	1.73	(2.20)					0.07
2	D/p			3.30	(1.68)			0.05
3	Term spread					2.84	(1.14)	0.02
4	D/p and term			3.56	(1.80)	3.29	(1.48)	0.08
5	$\gamma^\top f$ and term	1.87	(2.38)			-0.58	(-0.20)	0.07
6	$\gamma^\top f$ and d/p	1.49	(2.17)	2.64	(1.39)			0.10
7	All f							0.10
8	Moving average $\gamma^\top f$	2.11	(3.39)					0.12
9	MA $\gamma^\top f$, term, d/p	2.23	(3.86)	1.95	(1.02)	-1.41	(-0.63)	0.15

Notes: The left hand variable is the one-year return on the value-weighted NYSE stock return, less the 1-year bond yield. Standard errors use the Hansen-Hodrick correction.

Table 4. Excess return forecasts using yield factors and individual yields

Right hand variables	R^2	NW, 18 χ^2	NW, 18 p-value	Simple S χ^2	Simple S p-value	Small T χ^2	Small T p-value	5 percent crit. value
Slope	0.22	60.6	$\langle 0.00 \rangle$	22.6	$\langle 0.00 \rangle$	24.9	$\langle 0.00 \rangle$	9.5
Level, slope	0.24	37.0	$\langle 0.00 \rangle$	20.5	$\langle 0.00 \rangle$	18.6	$\langle 0.00 \rangle$	7.8
Level, slope, curve	0.26	31.9	$\langle 0.00 \rangle$	17.3	$\langle 0.00 \rangle$	16.7	$\langle 0.00 \rangle$	6.0
$y^{(5)} - y^{(1)}$	0.15	85.5	$\langle 0.00 \rangle$	30.2	$\langle 0.00 \rangle$	33.2	$\langle 0.00 \rangle$	9.5
$y^{(1)}, y^{(5)}$	0.22	45.7	$\langle 0.00 \rangle$	24.6	$\langle 0.00 \rangle$	22.2	$\langle 0.00 \rangle$	7.8
$y^{(1)}, y^{(4)}, y^{(5)}$	0.33	9.1	$\langle 0.01 \rangle$	4.6	$\langle 0.10 \rangle$	4.9	$\langle 0.09 \rangle$	6.0

Notes: The test is $c = 0$ in regressions $\overline{rx}_{t+1} = a + bx_t + cz_t + \overline{\varepsilon}_{t+1}$ where x_t are the indicated right hand variables and z_t are yields such that $\{x_t, z_t\}$ span all five yields.

Table 5. Additional Lags.

A. γ estimates

Max lag	const	$y^{(1)}$	$f^{(2)}$	$f^{(3)}$	$f^{(4)}$	$f^{(5)}$	R^2
0	-3.24	-2.14	0.81	3.00	0.80	-2.08	0.35
1	-3.22	-2.44	1.07	3.68	1.18	-3.11	0.41
2	-3.18	-2.56	1.33	3.47	1.76	-3.62	0.43
3	-3.20	-2.61	1.43	3.36	2.17	-3.98	0.44

B. α estimates and t statistics

Max lag	α_0	α_1	α_2	α_3	α_0	α_1	α_2	α_3
0	1.00				(8.44)			
1	0.50	0.50			(8.95)	(6.60)		
2	0.38	0.35	0.28		(6.81)	(6.45)	(4.46)	
3	0.31	0.29	0.20	0.21	(4.82)	(7.17)	(4.53)	(3.07)

Notes: The model is $rx_{t+1}^{(n)} = \gamma^\top \left[\alpha_0 f_t + \alpha_1 f_{t-\frac{1}{12}} + \alpha_2 f_{t-\frac{2}{12}} + \ldots + \alpha_k f_{t-\frac{k}{12}} \right] + \varepsilon_{t+1}^{(n)}$.

Table 6. GMM tests of the single-factor model

Lag i	Test	NW, 18		Simple S		No overlap		Small sample	
		χ^2	p-value	χ^2	p-value	χ^2	p-value	χ^2	p-value
0	J_T	1269	$\langle 0.00 \rangle$	110	$\langle 0.00 \rangle$	87	$\langle 0.00 \rangle$	174	$\langle 0.00 \rangle$
0	Wald	3460	$\langle 0.00 \rangle$	133	$\langle 0.00 \rangle$	117	$\langle 0.00 \rangle$	838	$\langle 0.00 \rangle$
1	J_T	157	$\langle 0.00 \rangle$	19.8	$\langle 0.18 \rangle$	22.1	$\langle 0.11 \rangle$	86.4	$\langle 0.00 \rangle$
1	Wald	327	$\langle 0.00 \rangle$	20.0	$\langle 0.18 \rangle$	24.5	$\langle 0.06 \rangle$	74.4	$\langle 0.00 \rangle$
2	J_T	134	$\langle 0.00 \rangle$	20.4	$\langle 0.16 \rangle$	22.7	$\langle 0.09 \rangle$	80.4	$\langle 0.00 \rangle$
2	Wald	240	$\langle 0.00 \rangle$	20.5	$\langle 0.15 \rangle$	23.9	$\langle 0.07 \rangle$	22.8	$\langle 0.05 \rangle$

Notes: Tests the single-factor model $rx_{t+1} = b\gamma^\top f_{t-i/12} + \varepsilon_{t+1}$ against the unrestricted model $rx_{t+1} = \beta f_{t-i/12} + \varepsilon_{t+1}$. The 5 percent and 1 percent critical values for all tests are 25.0 and 30.6 respectively. J_T gives the χ^2 test that all moments of the unrestricted regression $E(f_{t-\frac{i}{12}} \otimes \varepsilon_{t+1})$ are equal to zero. Wald gives the Wald test of the joint parameter restrictions $b\gamma^\top = \beta$. Column headings give the construction of the S matrix. "Small sample" uses the covariance matrix of the unrestricted moments $E(f_{t-\frac{i}{12}} \otimes \varepsilon_{t+1})$ across simulations of the 12 lag yield VAR to calculate $g_T^\top cov(g_T)^+ g_T$ and it uses parameter covariance matrix $cov(\beta)$ of unconstrained estimates across simulations in Wald tests.

Table 7. Forecasting the failures of the single-factor model

A. Coefficients and t-statistics

Right hand variable

Left hand var.	const.	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(3)}$	$y_t^{(4)}$	$y_t^{(5)}$
$rx_{t+1}^{(2)} - b_2\overline{rx}_{t+1}$	-0.11	-0.20	**0.80**	-0.30	-0.66	0.40
(t-stat)	(-0.75)	(-1.43)	**(2.19)**	(-0.90)	(-1.94)	(1.68)
$rx_{t+1}^{(3)} - b_3\overline{rx}_{t+1}$	0.14	0.23	-1.28	**2.36**	-1.01	-0.30
(t-stat)	(1.62)	(2.22)	(-5.29)	**(11.24)**	(-4.97)	(-2.26)
$rx_{t+1}^{(4)} - b_4\overline{rx}_{t+1}$	0.21	0.20	-0.06	-1.18	**1.84**	-0.82
(t-stat)	(2.33)	(2.39)	(-0.33)	(-8.45)	**(9.13)**	(-5.48)
$rx_{t+1}^{(5)} - b_5\overline{rx}_{t+1}$	-0.24	-0.23	0.55	-0.88	-0.17	**0.72**
(t-stat)	(-1.14)	(-1.06)	(1.14)	(-2.01)	(-0.42)	**(2.61)**

B. Regression statistics

Left hand var.	R^2	$\chi^2(5)$	$\sigma(\tilde{\gamma}^\top y)$	$\sigma(\text{lhs})$	$\sigma(b^{(n)}\gamma^\top y)$	$\sigma(rx_{t+1}^{(n)})$
$rx_{t+1}^{(2)} - b_2\overline{rx}_{t+1}$	0.15	41	0.17	0.43	1.12	1.93
$rx_{t+1}^{(3)} - b_3\overline{rx}_{t+1}$	0.37	151	0.21	0.34	2.09	3.53
$rx_{t+1}^{(4)} - b_4\overline{rx}_{t+1}$	0.33	193	0.18	0.30	2.98	4.90
$rx_{t+1}^{(5)} - b_5\overline{rx}_{t+1}$	0.12	32	0.21	0.61	3.45	6.00

Notes: In Panel A we use the estimates b from Table 1 to construct left hand variables; if the single-factor model $rx_{t+1} = b\gamma^\top f_t + \varepsilon_{t+1}$ holds, these portfolio returns should not be predictable. In Panel B, $\chi^2(5)$ gives the χ^2 statistic for joint significance of all right hand variables, excluding the constant. The 5 percent critical value is 11.07; p values are all below 1 percent. $\sigma(\tilde{\gamma}^\top y)$ gives the standard deviation of the right hand side variables, and $\sigma(\text{lhs})$ gives the standard deviation of the left hand variables. $\sigma(b^{(n)}\gamma^\top y)$ gives the standard deviation of the single-factor model forecast of $rx_{t+1}^{(n)}$ for comparison, and $\sigma(rx_{t+1}^{(n)})$ gives the standard deviation of that left-hand variable for comparison.
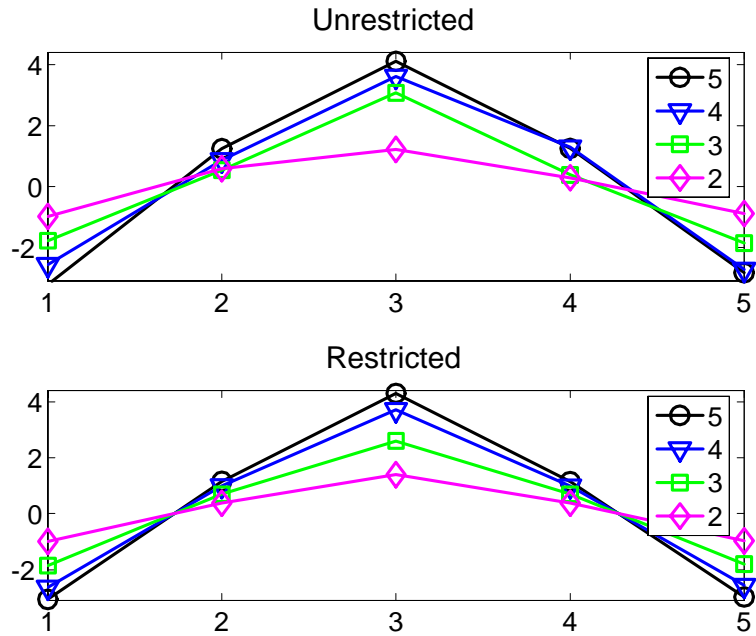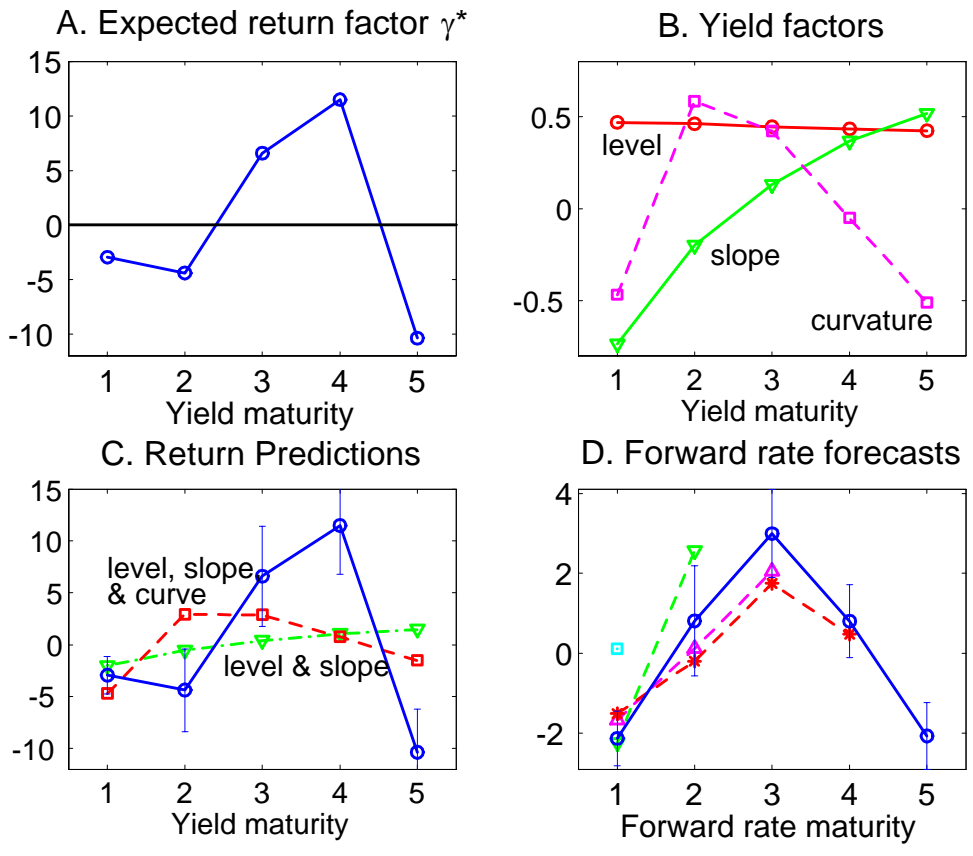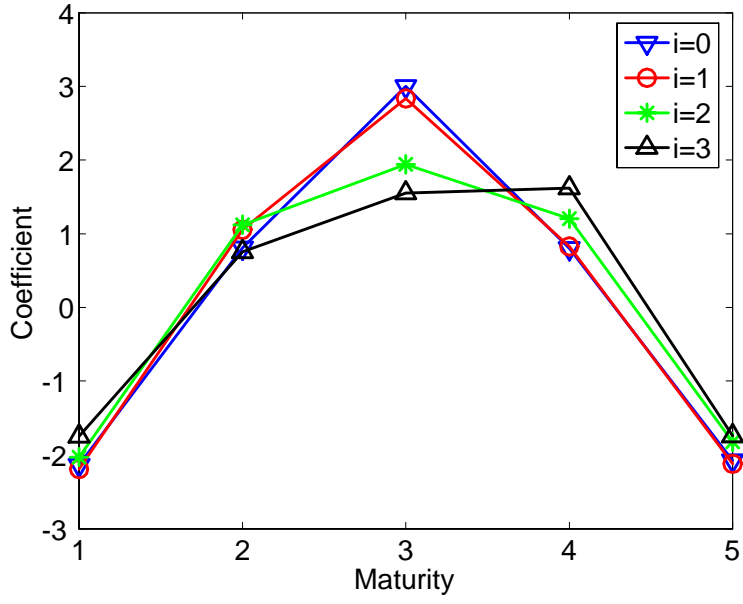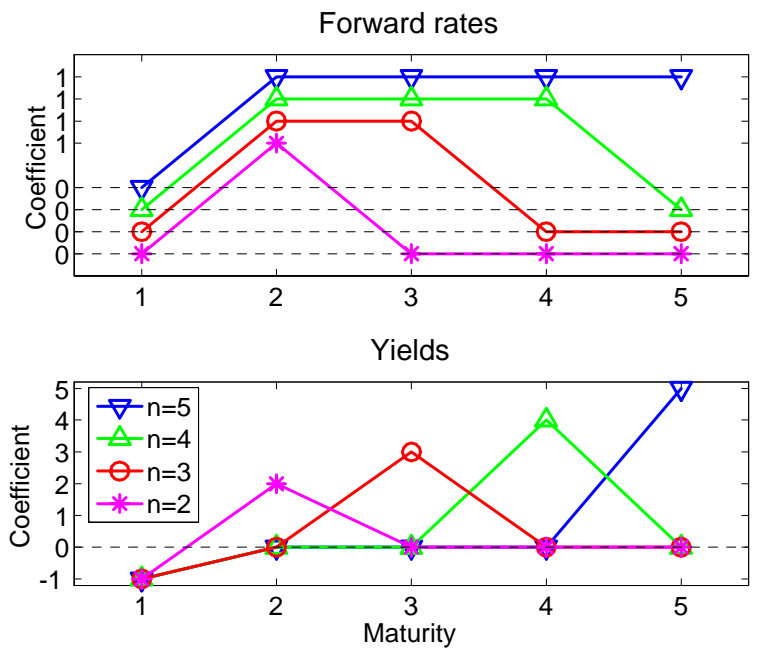
Figure 1:
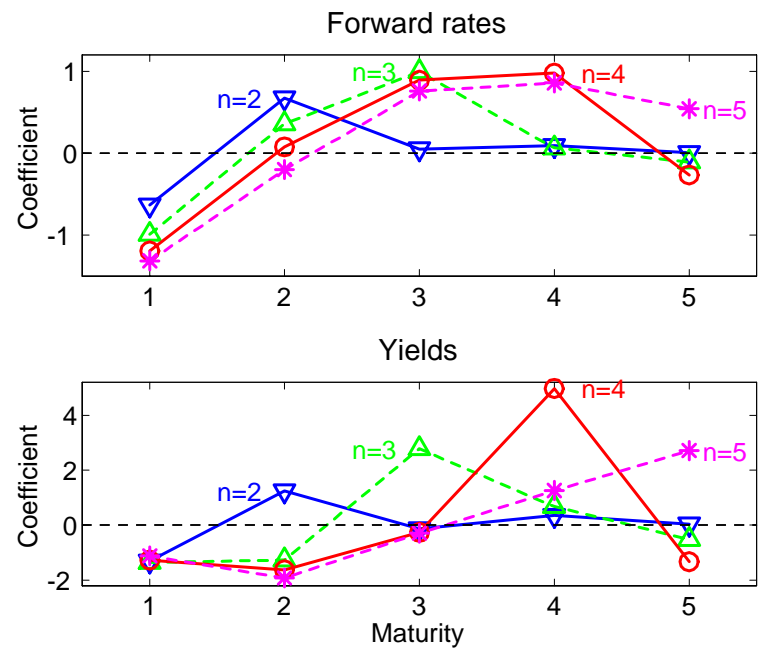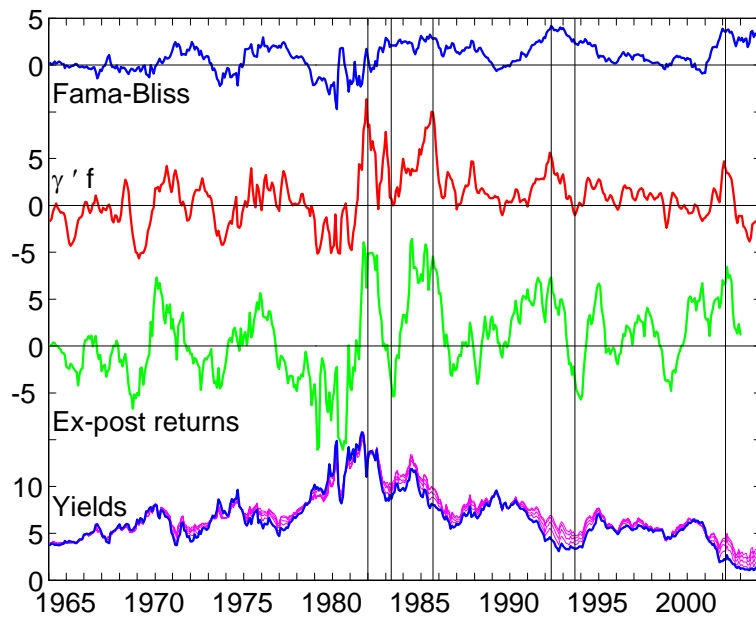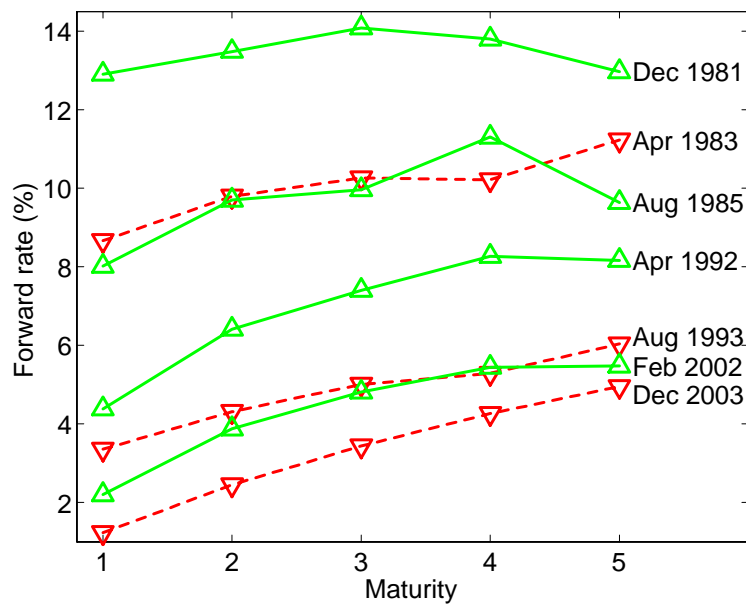
Figure 2:

Figure 3:

Figure 4:

Figure 5:

Figure 6:

Figure 7:

Figure Captions

Figure 1. Regression coefficients of one-year excess returns on forward rates. The top panel presents estimates β from the unrestricted regressions (1). The bottom panel presents estimates $b\gamma^\top$ from the single-factor model (2). The legend (5, 4, 3, 2) gives 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.

Figure 2. Factor models. Panel A shows coefficients γ^* in a regression of average (across maturities) holding period returns on all yields, $\overline{rx}_{t+1} = \gamma^{*\top} y_t + \varepsilon_{t+1}$. Panel B shows the loadings of the first three principal components of yields. Panel C shows the coefficients on yields implied by forecasts using factors to forecast excess returns. Panel D. shows coefficient estimates from forecasts using one, two, three, four and all five forward rates.

Figure 3. Coefficients in regressions of average (across maturity) excess returns on lagged forward rates, $\overline{rx}_{t+1} = \gamma^\top f_{t-\frac{i}{12}} + \varepsilon_{t+1}$.

Figure 4. Coefficients in regressions of bond excess returns on forward rates (top) and yields (bottom) generated by pure measurement error that is uncorrelated over time.

Figure 5. Regression coefficients of one year excess returns on forward rates (top) and yields (bottom) implied by a monthly VAR.

Figure 6. Forecast and actual average (across maturity) excess returns. The ex-post return line is shifted to the left one year so that the realization lines up with the forecast. The one-year yield in the bottom set of lines is emphasized. The vertical lines mark dates at which Figure 7 plots forward curves.

Figure 7. Forward rate curve on specific dates. Upward pointing triangles with solid lines are high return forecasts; downward pointing triangles with dashed lines are low return forecasts.