Decomposing the Yield Curve

John H. Cochrane*and Monika Piazzesi

March 13, 2008

Abstract

We construct an affine model that incorporates bond risk premia. By understanding risk premia, we are able to use a lot of information from well-measured risk-neutral dyanmics to characterize real expectations. We use the model to decompose the yield curve into expected interest rate and risk premium components. We characterize the interesting term structure of risk premia – a forward rate reflects expected excess returns many years into the future, and current slope and curvature factors forecast future expected returns even though they do not forecast current returns.

*Graduate School of Business, University of Chicago and NBER. Address: 5807 S. Woodlawn, Chicago, IL 60637. Email: john.cochrane@chicagogsb.edu. We thank David Backus and Jonathan Wright for helpful comments. We gratefully acknowledge research support from the CRSP and from an NSF grant administered by the NBER. A version of this paper with color graphics are available on the authors' websites.

1 Introduction

We construct and estimate a multifactor affine model of the term structure of interest rates that incorporates the lessons we learned about bond risk premia in Cochrane and Piazzesi (2005). Most of all, in that paper, we found that a single "factor" – a single linear combination of yields or forward rates – captures all of the economically-interesting variation in one-year expected excess returns for bonds of all maturities. We integrate this "return-forecasting factor" into an affine model. The affine model extends our understanding of risk premia over time – what do forecasts of two-year returns look like? It ties current yields or forward rates to long-term expectations of future interest rates and risk premia. We use this model to decompose the yield curve, to answer the question, "how much of a given yield curve corresponds to expectations of future interest rates, and how much corresponds to risk premia?" Of course, the affine model with risk premia should be useful in a wider variety of applications.

One may ask, why bother with the structure of an affine model? Just forecast interest rates, and define the risk premium as the residual of observed forward rates from this forecast. This approach leads to large statistical and specification uncertainty. Different but equally plausible ways of forecasting interest rates give wildly different answers. To give a sense of this uncertainty, Figure 1 forecasts one-year rates 1,2,3,... years into the future. Each forecast is made using a VAR of the 5 Fama-Bliss (1987) forward rates. The top panel is a simple VAR in levels. The bottom panel forecasts *changes* in forward rates from forward-spot *spreads*, treating forward rates as a set of 5 cointegrated variables.

The levels VAR of the top panel indicates relatively quick mean-reversion. Most variation in forward rates from the sample mean interest rate of about 6% will be labeled risk premium. Obviously, the location of the "mean" and suspicions that it might shift over time give much pause to that conclusion. The error-correction representation of the bottom panel essentially allows such a shifting mean. It gives a bit of interest-rate forecastability, but not much. Clearly, forward rates will show much less variation in risk premia in that representation. But do we really believe there is so little forecastability in one-year rates? Most of all, must such a huge difference in results come down to arbitrary specification choices? (Or, worse, a battery of inconclusive unit root tests?) The large sampling error of 10 year interest rate forecasts from any regression only muddies the picture even more.

The structure of the affine model allows us to infer a lot about the *dynamics* of yields from the *cross-section*. The risk-neutral dynamics that fit the cross-section of yields or forward rates are measured with very great precision, since the only errors are the 10 bp or so "measurement errors" that remain after one fits 3 or 4 factors to a cross section of yields. We start with estimates of the risk-neutral dynamics, and add market prices of risk to understand the true-measure dynamics. A detailed investigation of market prices of risk reduces this choice to one parameter, that only affects two elements of the 4×4 factor transition matrix. Thus, we are able to learn a lot about true-measure dynamics from the cross section.

For example, movements in yields are dominated by a "level factor" in which all yields

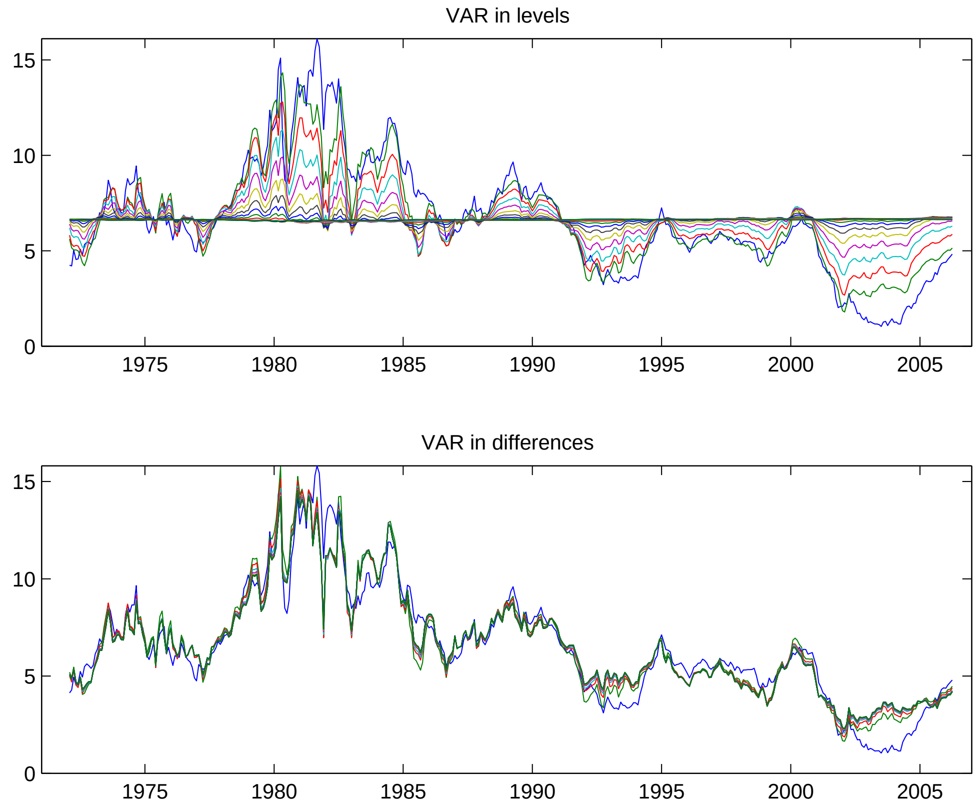

Figure 1: Forecasts of the one-year rate. The top panel uses VAR at one-year horizon of the 5 Fama-Bliss forward rates. The bottom panel forecasts changes in forward rates with forward-spot spreads.

move together by approximately the same amount. To make sense of this, any model must assign risk-neutral probabilities in which interest rates have a large permanent component. The "level factor" must have an autoregressive coefficient near one. We show how market prices of risk cannot alter this conclusion, so we estimate a very persistent real-measure transition matrix as well, with results that look more like the cointegrated specification of Figure 1, even though the model is written in levels.

Sampling uncertainty remains, of course, primarily coming from uncertainty about the true means of the factors, and to a lesser extent from the market price of risk. But it only affects the model's conclusions in very restricted ways.

Sampling uncertainty aside, the affine model is useful in order to characterize and understand point estimates of risk premia. Risk premia are large, measured either way. Expected future interest rates are quite different from forward rates. A pure forecasting approach leaves them as simple undigested residuals. The affine model connects many elements of a time-series representation of bond yields. First, we find that variation over time in expected returns of all maturities can be captured in a single state variable, which we dub the "return-forecasting factor", and which is not spanned by standard level, slope, and curvature factors.

Second, we find that slope and curvature movements forecast future movements in the return-forecasting factor. Expected returns are not just an AR(1) process that one can tack on orthogonally to a standard term structure model. If the slope or curvature of the term structure is large, *future* risk premia are large, even if *current* risk premia are zero. Since today's 10 year forward rate reflects 10 years of return premia, these factors are important for understanding long term forward rates. Understanding this *term structure of risk premia* is a central theme.

Third, we find that time-varying expected bond returns correspond entirely to compensation for exposure to risk of a "level" shock. The market prices of risk of expected-return, slope, and curvature shocks are almost exactly zero. Thus, our representation of time-varying risk premia, which potentially requires 16 numbers (a 4×4 matrix of loadings of each factor on each shock) turns out to require only *one* parameter. Market prices of risk *depend* on only one variable (the return-forecasting factor) and are earned in *compensation* for exposure only to one shock (the level shock), and only a single parameter (λ_{0l}) describes the entire transformation from risk-neutral to actual dynamics.

This last finding also paves the way to an economic understanding of risk premia in the term structure. If one wants to understand expected returns as compensation for exposure to macroeconomic shocks, those shocks must be shocks that have "level" effects on the term structure. As a counterexample, monetary policy shocks are typically estimated to have a "slope" effect on the term structure. We at least learn one variable that is *not* responsible for risk premia in the term structure.

1.1 Literature

The current literature documenting the failure of the expectations hypothesis goes back to the 1980s. Fama and Bliss (1987) and Campbell and Shiller (1991) showed that the particular forward rates that should forecast spot rates do not do so at a one-year horizon, and instead they forecast excess returns. However, Fama and Bliss also showed that forward-spot spreads do seem to forecast changes in interest rates at longer horizons, so the total impact on *yield* premia remains to be seen. Stambaugh (1988) and Cochrane and Piazzesi (2005) extend the finding of time-varying expected bond returns by forecasting returns with all available yields, not just single yields with specific maturities, finding substantially more return predictability. Piazzesi and Swanson (2004) document large risk premia in the short-term Federal Funds futures market.

The paper closest to ours is Kim and Wright (2005). Kim and Wright fit a three-factor constant-volatility affine model to weekly bond data, with the same purpose of decomposing observed yield and forward curves into expectations of future interest rates and risk premia. In particular, they find that the 50 basis point decline in the 10 year zero coupon yield between June 29 2004 and July 29 2005 includes an 80 basis point decline in term premium, implying a 30 basis point *rise* in the 10 year average of expected future interest rates. They estimate a 150 basis point decline in the 10 year instantaneous forward rate, of which 120 basis points are a decline in the corresponding risk premium. (p. 11.)

The main difference is implementation. We estimate our model with a stronger focus on matching direct estimates of one-year expected returns. In fact, we judiciously select the four factors in our yield curve model so that it can exactly match the OLS regression results for excess bond returns. These factors turn out to be traditional level, slope and curvature factors combined with the return-forecasting factor from Cochrane and Piazzesi (2005). Kim and Wright specify three latent factors and focus on matching interest rates forecasts at various horizons. We explore risk premia in great detail. Kim and Wright estimate weekly dynamics, raising them to the 520th power to find implications for 10 year forward rates. We estimate annual dynamics, check them against direct long-term forecasts, and investigate a cointegration specification for tying down the long run. Kim and Wright's estimation follows Kim and Orphanides (2005) in treating Blue Chip Financial Forecasts of interest rates as observable conditional expected values of future interest rates, at least after correcting for possibly autocorrelated measurement errors. Kim and Wright also have a nice summary of stories told about the recent yield curve movements.

Rudebusch, Swanson, and Wu (2006) fit "Macro-Finance" models to the yield curve. These models are based on observable macroeconomic factors rather than just bond yields. As a result they do not always fit the yield curve – expected interest rates plus risk premium can fail to add up to the observed yield. The models amount to very sophisticated regressions of yields on macroeconomic variables. They find two such models fit well before 2004, but produce large negative residuals for yields in the recent period, which they argue is a measure that something unexplained or unusual is in fact going on, that there is in this sense a "conundrum." They investigate a number of popular stories and quantitatively evaluate a number of them. In particular, they notice that declines in various measures of volatility correlate well with the "conundrum." Since their model does not have time-varying volatility, this finding suggests that term structure models that include time-varying volatilities may be important for understanding this period.

Kozicki and Tinsley (2001) emphasize as we do how much long-horizon interest rate forecasts depend on the specification of the long-run time series properties one imposes, such as level-stationarity vs. unit roots as in our Figure 1, and how little the time-series data help to distinguish these since both models fit short run dynamics equally well. Our hope is that fitting the cross-section via an affine model reduces some of this uncertainty.

2 Yields, risk premia and affine model

2.1 Notation, expectations hypothesis, and risk premia

2.1.1 Notation

Denote prices and log prices by

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

The log yield is

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

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

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

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

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

We denote excess log returns over the one-period rate by

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

2.1.2 Expectations hypothesis and risk premia

There are three conventional ways of capturing yield curve relationships. 1) *The long-term yield is average of expected future short term rates plus a risk premium* 2) *The forward rate is the expected future short rate plus a risk premium, and* 3) *The expected one period return on long term bonds equals the expected return on short term bonds plus a risk premium.* In equations,

$$\begin{aligned} y_t^{(n)} &= \frac{1}{n} E_t \left(y_t^{(1)} + y_{t+1}^{(1)} + \ldots + y_{t+n-1}^{(1)} \right) + rpy_t^{(n)} && (1) \\ f_t^{(n)} &= E_t(y_{t+n-1}^{(1)}) + rpf_t^{(n)} && (2) \\ E_t(r_{t+1}^{(n)}) &= y_t^{(1)} + rpr_t^{(n)} && (3) \end{aligned}$$

One can think of these equations as definitions of the respective risk premia. These three statements are equivalent, in the sense that if one equation holds with zero risk premium or a risk premium that is constant over time, all the other equations hold with zero risk premium or a risk premium that is constant over time as well.

Each statement has a portfolio interpretation. The yield-curve risk premium is the average expected return from holding a n-year bond to maturity, financed by a sequence of one-year bonds. The forward risk premium is the expected return from holding an n-year bond to maturity, financed first by holding an $n-1$ year bond to maturity and then with a one-year bond from time $t+n-1$ to time $t+n$; equivalently it is the expected return from planning to borrow for a year in the future spot market and contracting today to lend in the forward market. The return premium is the premium from holding an n-year bond for one year, financed with a one-year bond.

Since the forward rate translates so transparently into expected future interest rates, we focus our analysis on the forward rate curve, rather than the more conventionally-plotted yield curve. Obviously, the two curves carry the same information.

2.1.3 Relating return and yield curve risk premia

The yield, forward, and return risk premia are not the same objects, but each can be derived from the other. In particular, as we digest the information in the forward curves at each date, we want to understand the connection between risk premia in the forward curve – the difference between forward rate and expected future spot rate – to the one-period return risk premium about which we have more information and intuition.

By simply rearranging the definitions, the n-year forward rate equals the future one-year yield plus the difference between n-1 year returns on an n-year bond and an n-1 year bond,

$$f_t^{(n)} = y_{t+n-1}^{(1)} + r_{t,t+n-1}^{(n)} - r_{t,t+n-1}^{(n-1)}.$$

Since this identity holds ex-post, we can take expectations and express it ex-ante as well, splitting the n-year forward rate to expected future one-year rate plus the risk premium, which is the difference in expected return between an n-year bond held for n-1 years and an n-1 year bond held to maturity,

$$f_t^{(n)} = E_t\left(y_{t+n-1}^{(1)}\right) + E_t\left(r_{t,t+n-1}^{(n)} - r_{t,t+n-1}^{(n-1)}\right). \quad (4)$$

Figure 2 illustrates, and emphasizes how this relation is simply an identity.

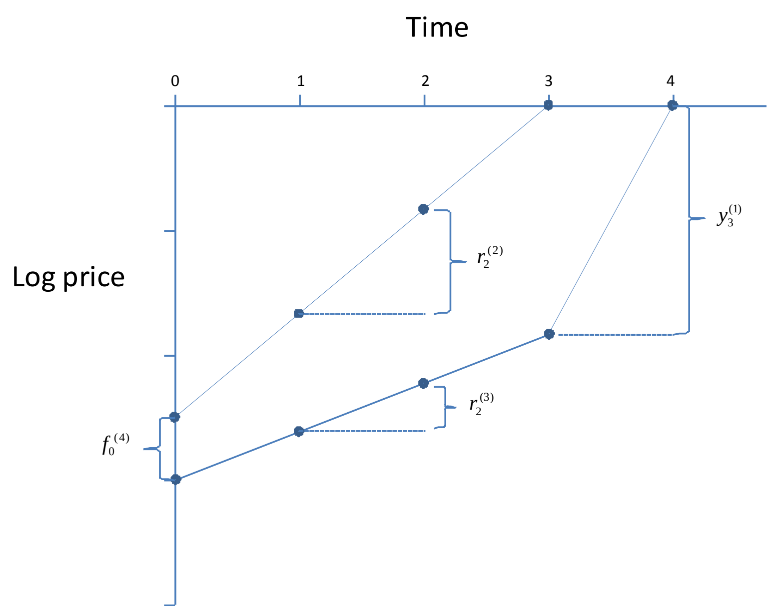

Figure 2: Example, following the price of 3 and 4 year bonds over time. The 4 year forward rate is split into the one year rate in year 3 and risk premiums.

We want to relate bond prices ultimately to *one*-year risk premia, defined as one-year expected returns in excess of the one-year rate. Therefore, we break the $n-1$-year returns to a sequence of one-year bond returns, also illustrated in Figure 2. The main representation which we explore is then given by

$$f_t^{(n)} - E_t\left(y_{t+n-1}^{(1)}\right) = E_t\left(rx_{t+1}^{(n)} - rx_{t+1}^{(n-1)}\right) + E_t\left(rx_{t+2}^{(n-1)} - rx_{t+2}^{(n-2)}\right) + \ldots + E_t\left(rx_{t+n-2}^{(2)}\right). \quad (5)$$

In portfolio terms, the forward-spot spread breaks down to a sequence of one-year investments: This year, buy an n period bond, short an n-1 period bond. Next year, buy an n-1 year bond, short an n-2 year bond, and so on and so forth. Relation (5) expresses the overall forward premium as the sum of these investments.

Similarly, the yield curve risk premium – the difference between current yield and the average of expected future short rates – is the average of expected future return premia of declining maturity,

$$y_t^{(n)} - \frac{1}{n} E_t \left(y_t^{(1)} + y_{t+1}^{(1)} + .. + y_{t+n-1}^{(1)} \right) = \frac{1}{n} \left[E_t \left(rx_{t+1}^{(n)} \right) + E_t \left(rx_{t+2}^{(n-1)} \right) + \ldots + E_t \left(rx_{t+n-1}^{(2)} \right) \right]. \tag{6}$$

This somewhat simpler relationship between risk premia comes at the expense of a slightly more complex relationship between yields and expected future interest rates in the expectations portion of the decomposition.

Formulas (4), (5) and (6) emphasize two important points. First, the forward premium and the yield premium depend on expected *future* return premia as well as current return premia. We will find interesting dynamics in return premia: they do not just follow exponential decay. To understand forward or yield curve risk premia, we have to understand the interesting *term structure of risk premia*.

Second, one gets the sense from empirical work (Fama and Bliss (1987), Chinn and Meredith (2005), Boudoukh, Richardson and Whitelaw 2006)) that "the expectations hypothesis works in the long run," and that this might be a sensible restriction to impose. The formulas reveal that this is not likely to be the case. The forward-spot spread cumulates expected returns. Thus, if there is a positive risk premium, say $E_t \left(rx_{t+1}^{(n)} \right) - E_t \left(rx_{t+1}^{(n-1)} \right) > 0$ early in the horizon, we would need to see an exactly offsetting *negative* risk premium later in the investment horizon, in order to recover $f_t^{(n)} - E_t \left(y_{t+n-1}^{(1)} \right)$ for large n. It seems strained to think that economic risk premia always change signs and integrate to zero. Instead, one might naturally think (and we find) that risk premia die out exponentially as the horizon increases. This natural pattern means that forward rates remain different from expected one-year rates at any maturity. We can expect them to parallel each other at long maturities, not to converge.

3 Affine Model Outline

Here we specify of the affine model, and outline our empirical procedure. The next few sections discuss each element of the specification and empirical procedure in detail.

3.1 Model

We specify four observable factors, constructed from forward rates. The first factor x_t is a slight refinement of the the "bond-return forecasting factor" described in Cochrane and Piazzesi (2005). It is a linear combination of yields or forward rates, formed as the largest eigenvector of the expected-return covariance matrix. The remaining yield-curve factors are level, slope and curvature factors, estimated from an eigenvalue decomposition of the forward-rate covariance matrix, orthogonalized with respect to the return-forecasting factor x_t. (We run regressions $f_t = a + bx_t + e_t$, and form factors from e_t.) Thus, our factors are

$$X_t = \left[\begin{array}{cccc} x_t & level_t & slope_t & curve_t \end{array} \right]'.$$

The main observation of Cochrane and Piazzesi (2005) is that expected returns of all maturities move together, so a single variable x_t can describe the time-variation of expected returns. Even if that variable were spanned by conventional level, slope and curvature factors, it would be useful for our purpose to isolate it and then reorthogonalize the remaining factors. In fact, as we document below, x_t is not spanned by level, slope, and curvature, so it is especially important to include it separately. Equivalently, curvature is not spanned by x_t, level and slope, so we need four factors to describe the cross section of bond prices as well as conventional three-factor models.

To construct the model, we use the discrete-time homoskedastic exponential-affine structure from Ang and Piazzesi (2003). We specify dynamics of the factors,

$$X_{t+1} = \mu + \phi X_t + v_{t+1};\ E\left(v_{t+1}v_{t+1}'\right) = V, \tag{7}$$

with normally distributed shocks. We specify that the log nominal discount factor is a linear function of the factors,

$$\begin{aligned} M_{t+1} &= \exp\left(-\delta_0 - \delta_1' X_t - \frac{1}{2}\lambda_t' V \lambda_t - \lambda_t' v_{t+1}\right) \\ \lambda_t &= \lambda_0 + \lambda_1 X_t. \end{aligned} \tag{8}$$

$\mu, \phi, V, \delta_0, \delta_1, \lambda_0$, and λ_1 are parameters which we pick below. The time-varying market price of risk λ_t generates a conditionally heteroskedastic discount factor, which we need to capture time-varying expected excess returns.

We calculate the model's predictions for bond prices recursively,

$$\begin{aligned} p_t^{(1)} &= \log E_t\left(M_{t+1}\right) = -\delta_0 - \delta_1' X_t \\ p_t^{(n)} &= \log E_t\left[M_{t+1} \exp\left(p_{t+1}^{(n-1)}\right)\right]. \end{aligned}$$

Calculating the expectations takes some algebra. The Appendix to Cochrane and Piazzesi (2005), available on our websites, goes through the algebra in detail. We summarize the results as follows: Log prices are linear ("affine") functions of the state variables,

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

The coefficients A_n and B_n can be computed recursively:

$$\begin{aligned} A_0 &= 0;\ B_0 = 0 \\ B_{n+1}' &= -\delta_1' + B_n' \phi^* \\ A_{n+1} &= -\delta_0 + A_n + B_n' \mu^* + \frac{1}{2} B_n' V B_n \end{aligned} \tag{9}$$

where μ^* and ϕ^* are defined as

$$\phi^* \equiv \phi - V\lambda_1 \tag{10}$$
$$\mu^* \equiv \mu - V\lambda_0. \tag{11}$$

This calculation gives us *loadings* – how much a price moves when a factor moves.

Having found prices, we can find the forward rate loadings, on which we focus,

$$f_t^{(n)} = A_n^f + B_n^{f\prime} X_t \tag{12}$$

where

$$B_n^{f\prime} = \delta_1' \phi^{*n-1} \tag{13}$$
$$A_n^f = \delta_0 - B_{n-1}' \mu^* - \frac{1}{2} B_{n-1}' V B_{n-1}. \tag{14}$$

We find this A^f and B^f from our previous formulas for A and B and $f_t^{(n)} = p_t^{(n-1)} - p_t^{(n)} = (A_{n-1} - A_n) + (B_{n-1}' - B_n')X_t$. We recognize in (13) the risk-neutral expectation of future one-year rates. Thus, μ^* and ϕ^* are "risk-neutral dynamics."

With prices and forward rates, returns, expected returns, etc. all follow as functions of the state variables. Expected returns are particularly useful for us

$$\begin{aligned} E_t\left(rx_{t+1}^{(n)}\right) &= B_{n-1}' V \lambda_0 - \frac{1}{2} B_{n-1}' V B_{n-1} + B_{n-1}' V \lambda_1 X_t \\ &= B_{n-1}' \left(\mu - \mu^*\right) - \frac{1}{2}\sigma^2(rx_{t+1}^{(n)}) + B_{n-1}' \left(\phi - \phi^*\right) X_t. \end{aligned} \tag{15}$$

We can also write the forward premium in terms of risk-neutral vs. actual probabilities as

$$\begin{aligned} f_t^{(n)} &= \delta_0 + \delta_1' \left(I + \phi^* + \phi^{*2} + ... + \phi^{*n-1}\right) \mu^* + \phi^{*n-1} X_t - \frac{1}{2} B_{n-1}' V B_{n-1} \\ E_t y_{t+n-1}^{(1)} &= \delta_0 + \delta_1' \left(I + \phi + \phi^2 + ... + \phi^{n-1}\right) \mu + \phi^{n-1} X_t \end{aligned}$$

3.2 Parameters

As usual in term structure models, only the *risk-neutral* dynamics matter for determining loadings. The relation between either prices (9) or forward rates (13)-(14) and state variables X_t, i.e. the coefficients A_n, B_n or A_n^f, B_n^f, depend only on the risk-neutral versions ϕ^* and μ^* of the transition dynamics, defined in (10)-(11), and the covariance matrix V.

We take V from the innovation covariance matrix of an OLS estimate of the factor dynamics (7). (We do not directly use the OLS estimated μ and ϕ, but we will contrast that estimate with the μ and ϕ estimated by our procedure. Updating V and reestimating makes very little difference.) Then, we choose risk-neutral dynamics ϕ^* and μ^* to match the *cross-section* of forward rates. Since (13)-(14) nonlinear, this requires a search: we choose

μ^* and ϕ^* to minimize the sum of squared differences between model predictions and actual forward rates:

$$\min_{\mu^*,\phi^*} \sum_{n=1}^{N} \sum_{t=1}^{T} \left(A_n^f + B_n^{f\prime} X_t - f_t^{(n)} \right)^2 \tag{16}$$

The 4-factor affine model produces a very good cross-sectional fit.

Market prices of risk λ_0 and λ_1 matter for our central question, evaluating risk premia and finding true-measure expected interest rates and returns. Our inquiry leads to a restricted specification:

$$\begin{aligned} \lambda_t &= \lambda_0 + \lambda_1 X_t \\ \lambda_t &= \begin{bmatrix} 0 \\ \lambda_{0l} \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 \\ \lambda_{1l} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_t \\ level_t \\ slope_t \\ curve_t \end{bmatrix} \end{aligned} \tag{17}$$

The fact that all columns of λ_1 are zero other than the first means that expected returns only vary *over time* with the return-forecasting factor x_t. That is, of course, the whole point of labeling x_t the return-forecasting factor. The restriction that only the second *row* of λ_t is nonzero means that risk premia are earned only as compensation for exposure to level shocks. As we will show, the data overwhelmingly support this restriction, because risk premia rise essentially linearly with maturity, as does the return covariance with the level shock.

With these restrictions, we can easily estimate λ_{0l} and λ_{1l}. Via (15), any return-forecasting regression has a constant that depends only on λ_{0l} and a slope coefficient that only depends on λ_{1l}. We choose a particularly convenient portfolio to run this regression, and pick λ_{1l} to match the regression coefficient. We choose λ_{0l} to match the means of the factors. Given λ_0, λ_1, we can now construct true-measure dynamics $\mu = \mu^* + V\lambda_0$ and $\phi = \phi^* + V\lambda_1$, and we can use the model to decompose the yield curve.

This is a method-of-moments estimation, so we use a GMM framework to evaluate asymptotic sampling error.

4 Affine model specification and estimation

This section outlines the reason for and details of each step of our affine model specification and estimation.

4.1 Data

It is convenient to summarize the yield curve by the yields of zero-coupon bonds at annually spaced maturities. However, the underlying treasury bonds are coupon bonds with varying maturities, so some fitting has to be done. We use two data sets. First, we use the Fama and Bliss (1985) data on 1-5 year maturity zero coupon bond prices. This data is well known, allowing a good comparison to previous work. Importantly, as we will see, the Fama and Bliss data is not smoothed across maturities. It also has a long time span which is particularly important for fitting risk premia. It is tempting to work with cleaner swap data, but if there is only one recession in a data set, it's hard to say much at all about recession-related risk premia.

To incorporate longer maturities into the analysis, we also use the new Gürkaynak, Sack, and Wright (2006) zero-coupon treasury yields. We use the data including maturities up to 15 years starting in 1971. By looking at longer maturities, we can check the behavior of our forecasts and factor loadings at longer horizons. It is also important to check that the broad brush of results hold up in a different dataset. The disadvantage, for some purposes, of this dataset is that it consists of a fitted function which smooths across maturities. Gürkaynak, Sack, and Wright estimate the Svennson (1994) six-parameter function for instantaneous forward rates

$$f^{(n)} = \beta_0 + \beta_1 e^{-n/\tau_1} + \beta_2 (n/\tau_1) e^{-n/\tau_1} + \beta_3 (n/\tau_2) e^{-n/\tau_2}.$$

The parameters $\beta_0, \beta_1, \beta_2, \beta_3, \tau_1, \tau_2$ are different on each date. β_3 and τ_2 only appear starting in 1980, and jump in at unfortunately large values. This fact means that there can only be one hump in the yield curve before that date. All of their yield, price, etc. "data" derive from this fitted function. Granted the function is quite flexible, but by fitting a function there are 6 degrees of freedom in 15 maturities; this is really a "6 factor model" of the term structure. Yield curve models will be evaluated by how closely they match the functional form, not necessarily by how well they match the underlying data. Finally, this functional form cannot be generated by standard yield curve models. Since the asymptotic ($n \to \infty$) forward rate and yield vary over time, there is an asymptotic arbitrage opportunity.

Now, the differences between the GSW and FB forward rate data are quite small on most dates given date, so for many purposes the slight smoothing in GSW data may make no difference. But excess return forecasts imply multiple differences of underlying price data – the excess return is $rx_{t+1}^{(n)} = p_{t+1}^{(n-1)} - p_t^{(n)} + p_t^{(1)}$, a forward-spot spread is $f_t^{(n)} - y_t^{(1)} = p_t^{(n-1)} - p_t^{(n)} + p_t^{(1)}$, and using all five prices on the right hand side of a return forecast allows for further differencing – so small amounts of smoothing have the potential to lose a lot of information in forecasting exercises.

4.2 One-year return forecasts mixing FB and GSW data

In Cochrane and Piazzesi (2005), we characterized one-period return risk premia by running regressions of excess returns on all of the Fama-Bliss forward rates,

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

We found that a single factor captured all the economically interesting variation in expected returns. To an excellent approximation we can write

$$\begin{aligned} r_{t+1}^{(n)} &= b_n \left[\gamma_0 + \gamma_1 y_t^{(1)} + \gamma_2 f_t^{(2)} + ... + \gamma_5 f_t^{(5)} \right] + \varepsilon_{t+1}^{(n)} && (18) \\ &= b_n \left[\gamma' f_t \right] + \varepsilon_{t+1}^{(n)}. && (19) \end{aligned}$$

The *single* "factor" or linear combination of forward rates (or yields) $\gamma' f_t$ captures expected returns across *all* maturities.

Our first task is to extend the specification of return forecasts to incorporate the longer maturities in the Gürkaynak, Sack, and Wright (GSW) data set. The conclusion of this section is that the best forecast of the GSW excess returns remains to regress them on a three month average of the 5 FB forward rates, and we use that specification of one-year risk premia in the rest of the paper.

We clearly need *some* reduction of right hand variables. Regressions using the 5 Fama-Bliss prices, yields, or forward rates on the right hand side already raise the specter of multicollinearity. Regressions using 15 maturities on the right hand side, generated from a six-factor model, are obviously hopeless.

Table 1 presents the R^2 values of forecasting one-year returns in the GSW data from forward rates. We evaluate overall forecastability by the regression of average (across maturity) returns

$$\overline{rx}_{t+1} = \frac{1}{14} \sum_{n=2}^{15} rx_{t+1}^{(n)}$$

on forward rates. The same patterns hold for individual-maturity bond returns.

Right hand variables	rx_{t+1}^{GSW} on f_t^{GSW}	f_t^{FB}	rx_{t+1}^{FB} on f_t^{FB}	rx_{t+1}^{GSW} on $f_t^{FB} - y_t^{(1)}$
f1–f5	0.29	0.35	0.33	0.32
f1, f3, f5	0.26	0.33	0.31	0.29
f1-f15	0.38			
3 mo. MA, f1–f5	0.31	0.46	0.42	0.44
3 mo. MA, f1–f15	0.48			

Table 1. R^2 values for forecasting average (across maturity) excess returns $\overline{rx}_{t+1} = \frac{1}{14}\sum_{n=2}^{15} rx_{t+1}^{(n)}$. GSW denotes Gürkaynak, Sack, and Wright (2006) forward rate data. FB denotes Fama Bliss (1986) data, updated by CRSP. f1-f5 uses one to five year forward rates; f1, f3, f5 use only the one, three, and five year forward rates; f1-f15 uses one to 15 year forward rates. 3 mo. MA uses three month moving averages of the right hand variables. Overlapping monthly observations of one-year return forecasts 1971-2006.

In the first row, we forecast GSW returns with the first 5 forward rates, as in our 2005 paper. The R^2 is a respectable 0.29. However, the regression coefficients (not shown) have the strong W shape suggestive of multicollinearity rather than the tent-shape we found in FB data. When we regress the *GSW* returns on the *FB* forward rates, we obtain much better R^2 of 0.35, and exactly the same tent-shaped pattern of coefficients that we find regressing FB returns on FB forward rates. This is even better than the 0.33 R^2 we find in regressing FB forward rates on their own returns.

This finding suggests ill effects of smoothing across maturities in the GSW data. (The top left panel of Figure 17, discussed below, plots a revealing data point.) The GSW smoothing removes some measurement error along with the forecasting signal, which is why when we use it to measure *returns* on the left-hand side, it measures those returns a little more cleanly and delivers a higher R^2 than the FB returns.

In the second row of Table 1, we use only the 1, 3, and 5 year maturities to forecast. This forecast loses very little R^2, and the coefficients show the familiar tent shape (not shown). Though the extreme multicollinearity has been eliminated, the FB forwards still forecast the GSW returns better than do the GSW forwards.

In the third row, we report what might seem the natural extension of our 2005 approach, forecasting the GSW return with all 15 GSW forward rates on the right hand side. The R^2 rises to an apparently attractive 0.38. However, the coefficients (not shown) are an uninterpretable combination of strong positive and negative numbers, the clear sign of extreme multicollinearity. When one smooths out independent variation in forward rates, as the GSW 6-factor fit does, then one loses the ability to run multiple regressions with 15 or even 5 right hand variables, and measure the separate contributions of each right hand variable with any precision. In addition the 0.38 R^2 (unadjusted) is barely larger than the 0.35 R^2 obtained from the 5 Fama-Bliss forward rates.

In our 2005 paper, we found that three-month moving averages of forward rates had better forecast power than the last month alone, which we attributed to small i.i.d. measurement errors. The last two rows of Table 1 show forecasts with these moving averages. All of the R^2 improve. In particular, the 5 FB forward rates give an impressive 0.46 R^2. Interestingly, the R^2 using the GSW data also improve, suggesting that their cross-sectional smoothing does not render their forward rates Markovian either.

Looking at individual maturities (not shown) all of these regressions show the one-factor structure, which (as opposed to the shape of the coefficients) is the major message of our 2005 paper. Whatever the linear combination of forward rates is that forecasts excess returns, the *same* linear combination on the right hand side forecasts returns of *any* maturity on the left hand side.

We conclude that the 5 Fama-Bliss forward rates capture all the information in the 15 GSW forward rates about future returns, and more. We therefore construct our risk premium estimates for the GSW data by running regressions on the FB data.

Forecasting the *FB* returns on the left hand side produces nearly identical coefficients and excess return forecasts (not shown) to go with the quite similar R^2 values shown in the rx_{t+1}^{FB} column. There seems to be little problem with using GSW vs FB *returns*, i.e. left-hand side variables.

One may worry about level effects in expected returns as much as one does in yield forecasting. Perhaps the success of all these forecasts just comes down to saying expected returns are high when interest rates are high, i.e. exploiting ex-post knowledge that the 1980s were a great decade for long-term bondholders. The final column of Table 1 forecasts excess returns using only forward rate *spreads*, thus ruling out the level effect. We see that the R^2 is hardly lower, and plots (quite similar to Figure 4 below) of the fitted values are nearly identical. This short-term return forecast is not driven by a level effect.

4.3 The return-forecasting factor x

Having isolated the best set of right-hand variables, we study expected excess returns of individual bonds in detail. Given the results of Table 1, we use the *Fama-Bliss* data to construct the return-forecasting factor. We regress the 14 GSW excess returns on a three month moving average of the five FB forward rates,

$$\begin{aligned} rx_{t+1}^{GSW} &= \alpha + \beta f_t^{FB} + \varepsilon_{t+1} \\ (14 \times 1) &= (14 \times 1) + (14 \times 5)(5 \times 1) + (14 \times 1) \end{aligned} \tag{20}$$

Using the Fama-Bliss data on the right-hand side solves the issue of multicollinearity in the GSW data. We do not get to learn how longer maturity forward rates might enter into the return-forecasting function, i.e. how the tent-shape pattern across the first five forward rates is modified by or extended to longer maturities, but it is clear from Table 1 that there are not enough degrees of freedom left in the GSW data for us to do that. We do have a right-hand variable that significantly forecasts the GSW returns, and that is all that the factor model requires. (The R^2 results of Table 1 and discussion of coefficient patterns extend to the individual maturities.)

We pursue one variation. One should be concerned with any regression that has the *level* of yields or forwards on the right hand side, as those variables have a large increase to 1980 and decrease thereafter. A regression might just be picking up these two data points. Technically, the level of yields or forwards contains a component that if it does not have a pure unit root has a root very close to unity. With these problems in mind, we also construct a return-forecasting factor that only uses *spread* information. We denote spreads with a tilde,

$$\begin{aligned} \tilde{f}_t^{(n)} &= f_t^{(n)} - y_t^{(1)} \\ \tilde{f}_t &= \left[\begin{array}{cccc} \tilde{f}_t^{(2)} & \tilde{f}_t^{(3)} & \tilde{f}_t^{(4)} & \cdots \end{array} \right] \end{aligned}$$

Then the spread specification of the return forecast is

$$\begin{aligned} rx_{t+1}^{GSW} &= \alpha + \beta \tilde{f}_t^{FB} + \varepsilon_{t+1}. \\ (14 \times 1) &= (14 \times 1) + (14 \times 4)(4 \times 1) + (14 \times 1) \end{aligned} \tag{21}$$

Given our concern about overfitting regressions using "level" right hand variables, we adopt this as the default specification, though the results are quite similar done either way.

Next, using either specification, we form a factor decomposition of *expected* returns by eigenvalue-decomposing the covariance matrix of expected returns

$$Q_r \Lambda_r Q_r' = cov\left[E_t(rx_{t+1}^{GSW})\right] = cov(\beta \tilde{f}_t^{FB}).$$

(The Appendix briefly reviews factor models formed by eigenvalue decompositions.)

Figure 3 plots the loadings – the columns of Q_r – in this exercise. These loadings tell us how much expected excess returns of each maturity move when the corresponding factor moves; they are regression coefficients of expected excess returns on the factors. The results (not shown) are almost exactly the same for the level specification, based on $cov(\beta f_t^{FB})$.

The caption gives the standard deviations of the factors $\sigma_i = \sqrt{\Lambda_{ri}}$ and the fractions of variance $\sigma_i^2 = \Lambda_{ri} / \sum_j \Lambda_{rj}$ explained by the factors. We find that the first factor utterly dominates this covariance matrix of expected returns, accounting for 99.9% of the variance of expected returns. We change almost nothing by imposing a single-factor model that this first factor describes movements in expected excess returns of *all* 14 maturities.

As shown in Figure 3, the dominant factor affects all maturities in the same direction, and its effect on expected returns rises almost linearly with maturity. Expected excess returns apparently all move in lockstep. Longer duration bonds' expected returns are more sensitive to changes in risk premia. This finding is reminiscent of the response of *ex-post* returns to changes in a "level" factor: longer duration bonds have larger responses to level movements. However, ex-post returns also move in response to changes in slope and curvature factors, and that sort of response is missing in expected returns.

The other factors seem to have pretty and perhaps economically interpretable loadings. However, the GSW data construction by smooth functions guarantees that even noise will be pretty in this data set. Doing the same exercise with FB return data, the remaining factors are clearly idiosyncratic movements in each maturity, as would be caused by pure measurement or iid pricing errors.

We conclude that a *single* factor accounts for all of the economically-interesting[1] variation in expected excess returns, and we follow that lead in constructing our factor representation. Thus, we define the return-forecasting factor by the eigenvector corresponding to this eigenvalue,

$$x_t = q_r' E_t\left(rx_{t+1}\right) = q_r'\left(\alpha + \beta \tilde{f}_t^{FB}\right) = q_r'\alpha + \gamma' \tilde{f}_t^{FB}$$

where q_r denotes the column of Q_r corresponding to the largest eigenvalue, and the last equality introduces the notation $\gamma' = q_r'\beta$.

Since the β have a tent-shaped form and q_r weights them all positively, $\gamma' = q_r'\beta$ is a familiar tent-shaped function of forward rates. Since Q_r is orthogonal, $q_r'q_r = 1$, and the point of the one-factor model is that the regression coefficients of each maturity return on forward rates are all proportional, $\beta \approx q_r\gamma'$ so then $q_r'\beta \approx q_r q_r'\gamma' = \gamma'$.

If we start with the regression forecast of each excess return,

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

and we premultiply by q_r', we obtain

$$E_t\left(q_r' rx_{t+1}\right) = q_r'\alpha + q_r'\beta f_t = x_t \tag{22}$$

Thus, the return-forecast factor x_t is the linear combination of forward rates that forecasts the portfolio $q_r' rx_{t+1}$. We will use this fact to choose affine model parameters so that $E_t\left(q_r' rx_{t+1}\right)$ has a constant of zero and a slope on x_t of one.

[1] The adjective is necessary because we cannot *statistically* reject the presence of additional factors. Minute movements may nonetheless be well-measured. Cochrane and Piazzesi (2005) has an extended discussion of this point.

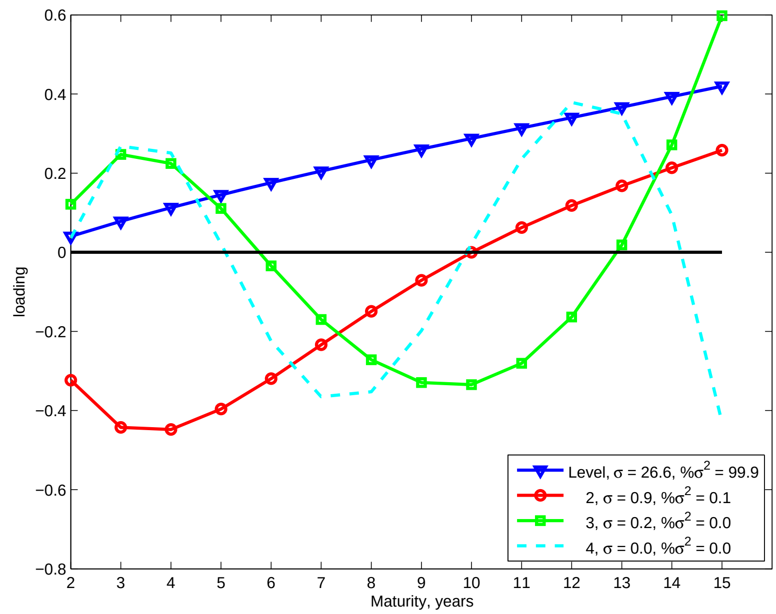

Figure 3: Factors in the covariance matrix of expected returns, regressing the 15 GSW returns on the 5 FB forward spreads.

The single-factor restriction then says that other expected returns are, to an excellent approximation, linear functions of this expected return, with q_r as the constants of proportionality:

$$\begin{aligned}
E_t(rx_{t+1}) &= \alpha + \beta f_t \\
&= \alpha + q_r \gamma' f_t \\
&= \alpha + q_r (x_t - q_r' \alpha) \\
E_t(rx_{t+1}) &= (I - q_r q_r') \alpha + q_r x_t
\end{aligned} \tag{23}$$

Figure 4 plots the return-forecasting factor x_t, calculated based on the levels and spreads of forward rates. The plot shows the business-cycle nature of the premium in the 1980s, and the early 1990s and 2000s. It also shows interesting business-cycle and inflation-related premiums in the 1970s. The plot verifies that the spread and level specifications give nearly identical results.

4.4 The return-forecast factor is not subsumed in level, slope and curvature

One would hope that this return-forecast factor x_t can be expressed as a linear combination of conventional level, slope, and curvature factors. Then, we could describe yield curve dynamics only in terms of these traditional factors. We would use our understanding of expected excess returns only to impose structure on the three-factor dynamics. This hope

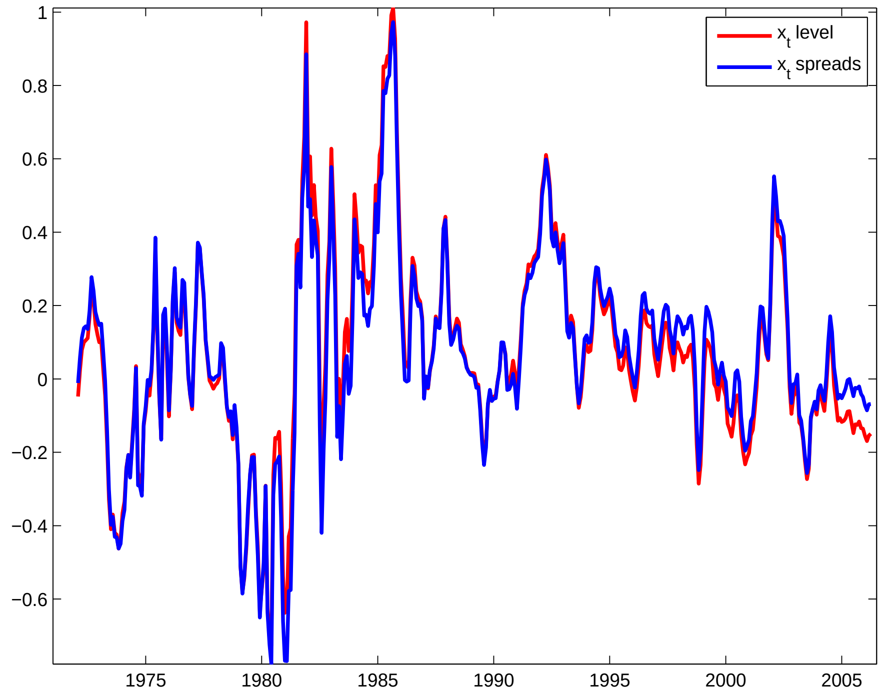

Figure 4: Return-forecasting factor $x_t = \gamma' f_t$ constructed from all forward rates (level) and using only forward spreads (spreads).

turns out not to be fulfilled: the return-forecast factor is poorly spanned by level, slope and curvature factors, so if we want to capture risk premia, we must include it as an extra factor.

The central question is, how well can we forecast excess returns using conventional yield curve factors? In Table 2, we compare the forecasting performance of the return-forecast factor with that of these conventional yield curve factors. We form the standard level, slope, and curvature factors by an eigenvalue decomposition of the covariance matrix of forward rates, $Q\Lambda Q' = cov(f_t)$. The factors are formed by $z_{it} = Q(:,i)' f_t$ where $Q(:,i)$ denotes the columns of Q corresponding to the three largest eigenvalues. The labels "level", "slope", and "curvature" come from the shape of the loadings (and weights) $Q(:,i)$. The first three eigenvalues explain in turn 97.73, 2.16 and 0.10 percent of the variance of forward rates (this is $\Lambda_{ii}/\sum_j \Lambda_{jj}$), adding up to 99.99% of the variance of forward rates explained. By any factor-model fitting standards, these three factors appear sufficient on their own to capture movements in the forward curve. In fact, the curvature factor is close to the threshold that one would drop it, since it explains only 0.10 % of the variance of forward rates. We keep it, to verify that the return-forecast factor is not just a proxy for this conventional curvature factor.

	x_t	$level$	$slope$	$curve$	$factor4$	$factor5$	R^2
% of var(f)		97.7	2.2	0.10	0.01	0.00	
	0.24						0.46
	(9.00)						
		0.19					0.03
		(0.80)					
		0.21	-1.72				0.13
		(0.95)	(-2.51)				
		0.22	-1.74	4.79			0.25
		(1.11)	(-2.72)	(2.82)			
		0.20	-1.70	4.73	-0.28	-15.49	0.34
		(1.13)	(-2.60)	(2.80)	(-0.09)	(-2.87)	

Table 2. Regression coefficients, t-statistics and R^2 for forecasting the average (across maturity) excess return $\overline{rx}_{t+1}$ in GSW data, based on the return-forecast factor x_t and eigenvalue-decomposition factors of forward rates. The top row gives the fraction of variance explained, $100 \times \Lambda_i / \sum_{j=1}^{15} \Lambda_j$. Monthly observations of annual returns 1971-2006. Standard errors include a Hansen-Hodrick correction for serial correlation due to overlap.

We focus on the forecasts of average (across maturity) returns $\overline{rx}_{t+1}$. The results are quite similar with other portfolios or individual bonds. Table 2 starts with the return-forecasting factor, which is highly significant (in all cases, ignoring estimation error in factor formation) and generates an 0.46 R^2. By contrast, the dominant (97.7% of variance) level factor does nothing to forecast returns. Slope does forecast returns, with a 0.13 R^2, and curvature also helps, raising the R^2 to 0.25. The return forecast factor *is* correlated with these conventional yield curve factors, and conventional factors do capture *some* of the return forecastability in the data. But considerable orthogonal movement in expected return remains. To span this movement without using the return-forecasting factor directly, one must include additional, usually ignored yield curve factors. For example, Table 2 shows that the tiny fifth factor significantly helps to forecast returns (t $= -2.87$), and raises R^2 from 0.25 and 0.34. But this still is not all the predictability in the data. The Fama-Bliss forward rates underlying the return-forecasting factor evidently include additional small "factors" that are set to zero in GSW's smoothing algorithm, raising the R^2 another 12 percentage points. Forming level, slope and curvature factors from yields rather than forward rates gives quite similar results, as does imposing a unit root-cointegration structure in which the level factor is a parallel shift and the remaining factors derive from an eigenvalue decomposition of the covariance of yield or forward spreads.

In sum, Table 2 verifies that, in order to construct a factor model that reflects what we know about return forecasting, we have to explicitly include the return-forecast factor, along with the remaining yield curve factors. Clearly, it would be a mistake to ignore the leftover portion of the return-forecast factor as we conventionally ignore the fourth and fifth factors in representing the level of yields of forward rates, and as we have ignored factors past the first one in representing the covariance of expected excess returns.

4.5 Constructing factors

To capture the yield curve in a way that captures return premia, then, we need to augment conventional yield curve factors by explicitly including the return-forecasting factor. Our first step is to construct data on factors whose movements will describe the yield curve.

Since eigenvalue decompositions are so easy and insightful, we proceed by minimally modifying that procedure. We start with the return forecast factor

$$x_t = q_r' \left(\alpha + \beta \tilde{f}_t^{FB} \right)$$

as defined above. We include the constant so that this factor has directly the interpretation as the expected return of the portfolio $q_r' rx_{t+1}$, i.e. so $q_r' rx_{t+1} = 0 + 1x_t + \varepsilon_{t+1}$, but the location of constants is irrelevant as any constants added here will be subtracted later.

We define the remaining yield curve factors by an eigenvalue decomposition of the forward covariance matrix, after orthogonalizing with respect to the return-forecasting factor. As Table 1 shows, the usual level, slope and curvature factors are correlated with x_t, and also forecast returns. By orthogonalizing these factors, we create factors that are uncorrelated with each other (always convenient), plus we have the attractive restriction that the factors other than x_t *do not* forecast one-year returns *at all.* We run regressions

$$f_t^{GSW} = c + dx_t + e_t.$$

(We write c and d to save a and b for loadings, below.) Then we take the covariance decomposition of the residual,

$$Q\Lambda Q' = cov(e_t),$$

and we define the remaining factors as

$$\begin{aligned} level_t &= Q(:,1)'\,(c + e_t) \\ slope_t &= Q(:,2)'\,(c + e_t) \\ curve_t &= Q(:,3)'\,(c + e_t) \end{aligned}$$

The labels "level" "slope" and "curvature" come ex-post from the shapes of the resulting loadings, i.e. how much each forward rate moves when a factor moves, or from the similar shapes of the weights $Q(:,i)$ by which we form factors from forward residuals.

4.6 Fitting forwards – risk-neutral dynamics

At this point, we have constructed observable factors

$$X_t = \left[\begin{array}{cccc} x_t & level_t & slope_t & curve_t \end{array} \right]'.$$

The next step is to pick parameters $\delta_0, \delta_1, V, \mu^*, \phi^*$ of the affine model in its risk-neutral form. Given these parameters, we can calculate the model's predicted loadings A_n^f, B_n^f (13)-(14), i.e. how each forward rate should load on factors by

$$f_t^{(n)} = A_n^f + B_n^{f\prime} X_t.$$

We pick V from an OLS estimate of transition dynamics

$$X_{t+1} = \mu + \phi X_t + v_{t+1};\ E\left(v_{t+1}v_{t+1}'\right) = V. \tag{24}$$

We specify the model to have mean-zero factors, and therefore $\mu = 0$. Therefore, we de-mean the factors before proceeding. This is important to obtaining reasonable results. We will have ϕ and ϕ^* with eigenvalues near one, and in fact in some estimates we do not impose eigenvalues less than one so point estimates can have larger eigenvalues. The model's factor mean $E(X) = (I - \phi)^{-1}\mu$ is thus a very sensitive function of μ. And, as we have known since the first Figure, long-term means are important to long-term forward rate decompositions. By de-meaning the factors before starting, we make sure the model produces the sample means of the factors, or at least that $(I - \phi)E(X) = \mu$. In essence, we are parameterizing the model by the factor means $E(X)$ rather than μ. We therefore will parameterize sample uncertainty in terms of uncertainty about $E(X)$.

With $E(X) = 0$, $\mu = 0$, and the structure on market prices of risk described in (32) below, we have

$$\begin{aligned} \mu^* &= cov(v, v_l)\lambda_{0l} \\ (4 \times 1) &= (4 \times 1)(1 \times 1) \end{aligned}$$

Thus, we search over parameters $\{\delta_0, \delta_1, \lambda_{0l}, \phi^*\}$.

We search over these paremeters to minimize the sum of squared differences between model predictions and actual forward rates.

$$\min \sum_{n=1}^{N} \sum_{t=1}^{T} \left(A_n^f + B_n^{f\prime} X_t - f_t^{(n)}\right)^2 \tag{25}$$

As a benchmark, we compute OLS regressions of forward rates on factors,

$$f_t^{(n)} = a_n + b_n' X_t + \varepsilon_t. \tag{26}$$

The coefficients in this regression are the best any linear factor model can do to attain the objective (25), so the comparison gives us some sense of how much the affine model structure is affecting the results. The a_n, b_n do not embody the restrictions of an affine model. Thus, we can also think of the minimization as finding the affine factor model *loadings* that are closest to the regression-based loadings, weighted by the covariance matrix of the factors.

The errors in (25) are strongly correlated across maturities. If there were any difficulty in this fit, or to do maximum likelihood, we should minimize a weighted sum. However, our fitted forward rates are so close to the best achievable (26) that this refinement makes little difference to the results.

Figure 5 presents the estimated loadings B_n^f and regression-based loadings b_n. These loadings answer the question "if a factor moves, how much does each forward rate $f^{(n)}$ move?" A movement in the return-forecast factor x_t sends short rates down and all long rates up. The "level", "slope" and "curvature" factors are so named because of the familiar and sensible shapes of the loadings. The level factor moves all forwards in the same direction. The slope factor moves short rates up and long rates down, but with a much smoother pattern than does the return-forecast factor x. The curve factor induces a curved shape to forward rates. Since the return-forecast factor explains so little of the variance of forward rates, the other factors are not much altered by orthogonalizing with respect to the return-forecast factor.

These loadings – how much each forward rate moves when a factor moves – are not the same as the weights – how we construct each factor from forward rates. Expected returns and x are still constructed by the familiar tent-shaped function of forward rates. The largest eigenvector of the covariance matrix of expected returns q_r *is* both the weight and the loading of expected *excess returns* on the factors, but neither the weight nor loading relating forward rates to factors. Similarly, the remaining factors are constructed from residuals in a regression of forward rates on expected return factors, so the weights and loadings are equal for those residuals, but not for the underlying forward rates.

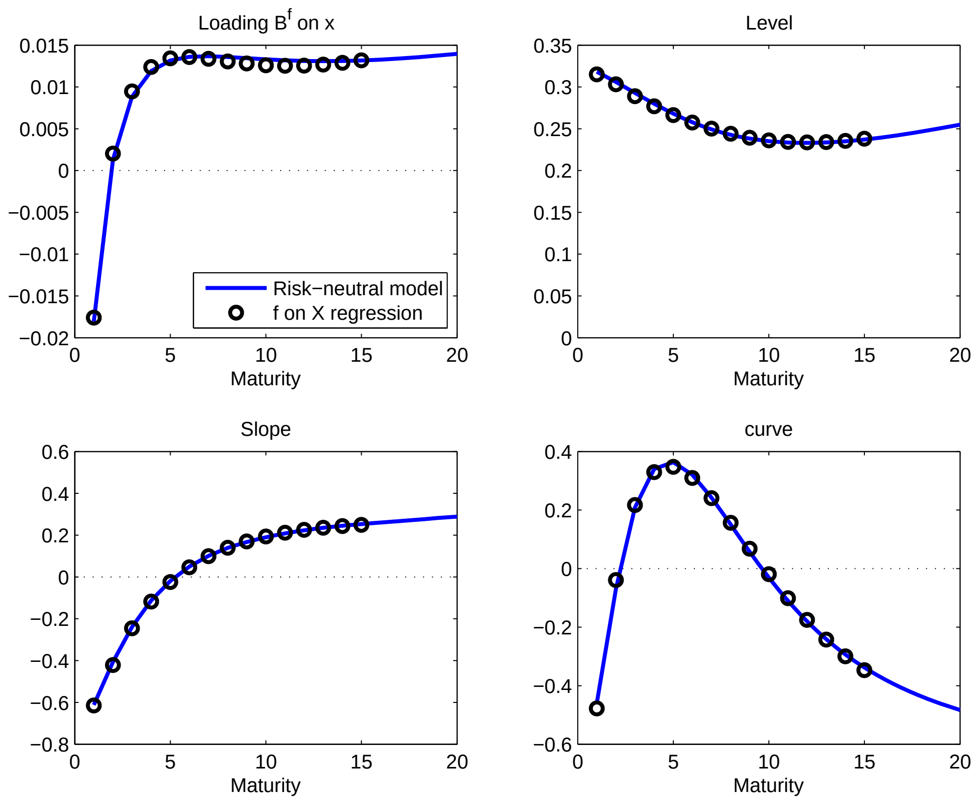

Figure 5: Affine model loadings, B^f in $f^{(n)} = A^f + B^{f\prime}X_t$. The line gives the loadings of the affine model, found by searching over parameters $\delta_0, \delta_1, \mu^*, \phi^*$. The circles give regression coefficients of forward rates on the factors.

Figure 5 shows that the "affine model" loadings give a very good approximation to the regression loadings. The structure of the affine model is clearly flexible enough to give an excellent fit to the data. We present the estimates of the risk-neutral dynamics μ^*,ϕ^* below and contrast them with estimates of the true-measure dynamics μ, ϕ.

Table 3 quantifies this observation with measures of model fit – how close forward rates predicted by the model $f_t^{(n)} = A_n^f + B_n^{f\prime} X_t$ are to actual forward rates. The "Model" columns give the result of the minimization (25), while the "Regression" columns give the result of unconstrained regression (26). "Model-Reg." characterizes the difference between the affine model and regressions. The mean squared errors are less than 20 bp, with somewhat lower mean absolute deviations.

	M. S. E.			M. A. D.		
	Model	Regression	Model-Reg.	Model	Regression	Model-Reg.
Baseline model	16.3	9.8	13.0	12.3	6.2	10.4
Force eig<1	17.0	9.8	13.9	13.0	6.2	11.1
OLS-affine	171	9.8	136	136	6.2	135

Table 3. Affine model fits in basis points (1 = 0.01%) . M.S.E. gives the mean squared error, $\left[\frac{1}{15}\sum_{n=1}^{15}\frac{1}{T}\sum_{t=1}^{T}\left(f_t^{(n),model} - f_t^{(n)}\right)^2\right]^{\frac{1}{2}}$. M.A.D. gives mean absolute deviations, $\frac{1}{15}\sum_{n=1}^{15}\frac{1}{T}\sum_{t=1}^{T}\left\|\left(f_t^{(n),model} - f_t^{(n)}\right)\right\|$. The "model" is the affine model prediction $f_t^{(n),model} = A_n^f + B_n^{f\prime} X_t$. "Regression" is the fitted value of a regression of data on the four factors, $f_t^{(n)} = a_n + b_n' X_t + \varepsilon_t$. "Model-Reg." gives the difference between the two fitted values, rather than model - data. Force Eig<1 describes the model in which we force the maximum eigenvalue of ϕ^* to be less than one.

4.7 Market prices of risk

We choose market prices of risk λ_0, λ_1 to match the cross-section of bond expected returns. From $1 = E_t\left(M_{t+1}R_{t+1}\right)$ and (8) expected returns follow

$$E_t\left[rx_{t+1}\right] + \frac{1}{2}\sigma^2(rx_{t+1}) = cov(rx_{t+1}, v_{t+1}')\left(\lambda_0 + \lambda_1 X_t\right). \tag{27}$$

(We have specified a conditionally homoskedastic model, so σ_t and cov_t do not need t subscripts.) Thus λ_t gives the "market price of risk" of the v_{t+1} shocks, i.e. it says how much an expected return must rise to compensate for covariance of that return with a given shock.

We specified the regression-based single-factor model for expected returns,

$$E_t\left(rx_{t+1}\right) = (I - q_r q_r')\alpha + q_r x_t. \tag{28}$$

This model means that all but the first column of λ_1 must equal zero, so that expected returns in (27) only depend on the return-forecast factor x_t. Putting together (27) and (28), and denoting by λ_{1x} the first column of λ_1, we have

$$\begin{aligned}(I - q_r q_r')\alpha + \frac{1}{2}\sigma_t^2(rx_{t+1}) &= cov(rx_{t+1}, v_{t+1}')\lambda_0 && (29)\\ q_r &= cov(rx_{t+1}, v_{t+1}')\lambda_{1x} && (30)\\ (14 \times 1) &= (14 \times 4)(4 \times 1). &&\end{aligned}$$

The first equation describes the constant term, or, roughly, how unconditional mean returns are generated by covariances with shocks times market prices of risk. The second equation describes how the time-varying component of mean returns is generated by covariances with shocks times market prices of risk. Each is a four-factor model; covariance of returns with innovations to each of the four factors can generate mean returns.

To understand these market prices of risk, we can use the time-series of excess returns to estimate the variance term, and that along with the time-series of factor shocks v_{t+1} to estimate the covariance $cov_t(rx_{t+1}, v'_{t+1})$. Figure 6 presents the the left-hand, "expected return" side of (30), represented by the solid black line labeled q_r, and the covariances of returns with shocks to each of the factors –the covariance terms in the right hand side of (30), represented by the dashed colored lines with triangles. The job of the cross-sectional regression is to find the linear combination of the four covariance lines that most closely matches the expected return line. The graph shows clearly that the expected return line is already almost a perfect match to the covariances with the level shock. Both rise nearly linearly with maturity. Longer maturity bonds have *expected* returns that vary more strongly with the state variable x_t – hence the rise of the expected return or q_r line – and longer maturity bonds are more affected *ex-post* when the level shock hits – hence the rise in the $cov(r, level)$ line. By contrast, the covariance of returns with the slope and return-forecast factors have completely the wrong shape. Thus, it's clear from the graph that we will obtain an excellent approximation by setting the market prices of risk of expected-return, slope, and curvature shocks to zero, and understanding all time-varying risk premia as compensation for covariance with the level shock. Figure 6 presents the fitted value with this restriction, the multiple of $cov(r, level)$ closest to the q_r line, with slope λ estimated by cross-sectional regression, and it is so close to q_r as to be nearly visually indistinguishable. Unsurprisingly, the simlar graph of the left hand side of (29) and covariances shows the same pattern, so that the market prices of risk in λ_0 other than level shock can be set to zero.

In sum, it seems an excellent approximation to specify that *the time-varying market price of risk comes only from covariance of returns with the level shock.* This finding is the first step in the important task of coming to an *economic* understanding of bond risk premia. Eventually, we want to understand what *macroeconomic* shocks underlie bond risk premia. Now we know at least that those shocks must be shocks that have *level* effects on the forward curve. For a counterexample, monetary policy shocks are usually estimated to have a *slope* effect on the yield curve, affecting short term rates, but either leaving long term rates alone or even sending them in the other direction. With such an estimate, the bond risk premium will *not* be earned as compensation for the risk of monetary policy shocks.

We have now simplified market prices of risk greatly in *two* dimensions: the market prices of the return-forecast, slope and curvature shocks are zero, and the market prices only vary as a function of the return-forecast factor. An equation is worth a thousand words here: the

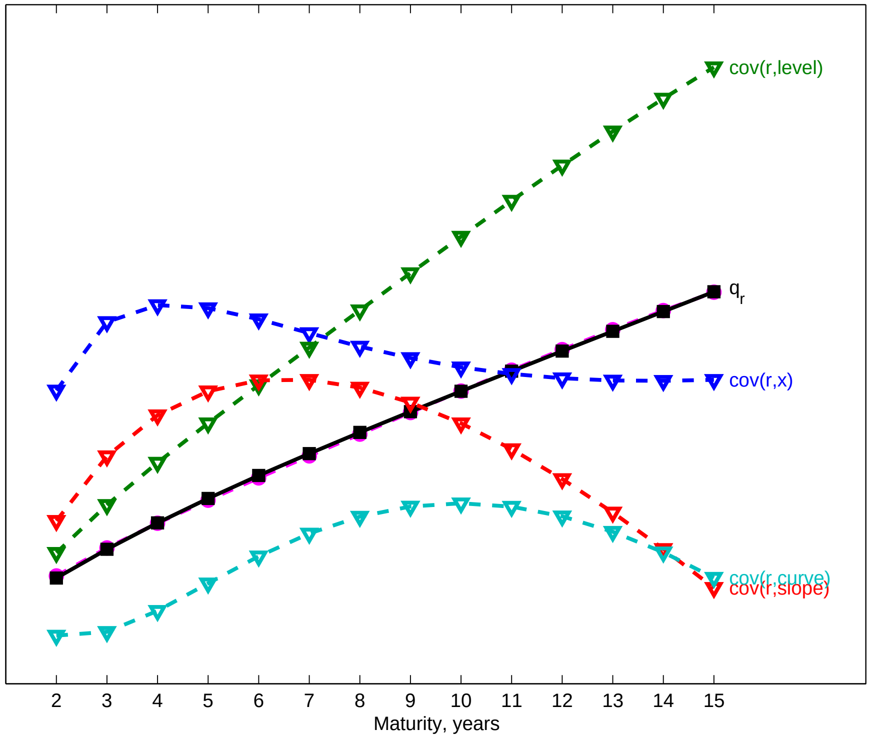

Figure 6: Loading q_r of expected excess returns on the return-forecasting factor x_t, covariance of returns with factor shocks, and fitted values. The fitted value is the OLS cross-sectional regression of q_r on $cov(r, v_{level})$, the dashed line nearly colinear with q_r. The covariance lines are rescaled to fit on the graph.

market prices of risk in (8) and (10)-(11) take the form

$$\lambda_t = \lambda_0 + \lambda_1 X_t \tag{31}$$

$$\lambda_t = \begin{bmatrix} 0 \\ \lambda_{0l} \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 \\ \lambda_{1l} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_t \\ level_t \\ slope_t \\ curve_t \end{bmatrix} \tag{32}$$

The single-factor expected return model only requires that the last three columns of λ_1 are zero. The restriction that the first and last two *rows* of λ_0, λ_1 are zero is a separate restriction, that can be imposed or not, and which comes from examination of expected returns vs. covariances as in Figure 6 and the corresponding regressions.

4.8 Market prices of risk: estimates

The natural procedure, then, is to estimate market prices of risk (now down to two numbers), so that the *model* reproduces the forecasting *regressions* that describe bond expected returns.

The affine model predicts expected returns

$$E_t\left(rx_{t+1}^{(n)}\right)=\left[B_{n-1}'V\lambda_0-\frac{1}{2}B_{n-1}'VB_{n-1}\right]+\left[B_{n-1}'V\lambda_1\right]X_t$$

(again, we relegate the algebra to the appendix). Given the simplification of (32), we can write this equation as

$$E_t\left(rx_{t+1}^{(n)}\right)=\left[B_{n-1}'cov(v_t,v_t^{level})\lambda_{0l}-\frac{1}{2}B_{n-1}'VB_{n-1}\right]+\left[B_{n-1}'cov(v_t,v_t^{level})\right]\lambda_{1l}x_t \qquad (33)$$

We have 14 expected returns, each a function of a constant and x_t, which we wish to match with two numbers λ_{0l} and λ_{1l}. Anything we do will amount to choosing one portfolio to match. The most natural portfolio is the one weighted by q_r, since it recovers the return-forecasting factor x_t, i.e.

$$E_t\left(q_r'rx_{t+1}^{(n)}\right)=x_t.$$

(See equation (22).) Substituting in (33), we therefore can estimate the market price of risk λ_{1l} easily by setting the regression coefficient of $q_r'rx_{t+1}^{(n)}$ on x_t to one,

$$1=q_r'Bcov(v_t,v_t^{level})\lambda_{1l}$$

i.e.,

$$\lambda_{1l}=\frac{1}{q_r'Bcov(v_t,v_t^{level})} \qquad (34)$$

where B is the 14×4 matrix whose nth row is B_{n-1}

We could similarly estimate the constant portion of the market price of risk as the value that sets the intercept in the forecasting regression of $q_r'rx_{t+1}^{(n)}$ on x_t to zero,

$$q_r'Bcov(v_t,v_t^{level})\lambda_{0l}=\frac{1}{2}q_r'\left[B_{n-1}'VB_{n-1}\right]$$

and hence

$$\lambda_{0l}=\frac{1}{2}q_r'\left[B_{n-1}'VB_{n-1}\right]\lambda_{1l}$$

where the notation $[\cdot]$ means to make a vector out of the enclosed quantity over the index n.

However, we estimated λ_{0l} above to choose the risk-neutral mean μ^* in such a way that $\mu=0$ exactly, i.e. that the model produces the demeaned factors. We check that the two estimates are trivially different, and that the intercept of the $q_r'rx_{t+1}^{(n)}$ is not importantly different from zero.

With λ_{0l} and λ_{1l} estimated, we now can recover true-measure dynamics $\phi=\phi^*+V\lambda_1$. We have $\mu=0$ already. All the parameters are estimated, and we can now use the model.

5 Estimates and decompositions

5.1 Transition matrices μ, ϕ, and impulse-responses

Table 4 presents estimates of the transition matrices, μ and ϕ. Figures 7 and 8 present corresponding impulse-response functions. These functions are the responses $X_{t+1} = \phi X_t$ that follow from each element of X being set to one with the others zero in turn. We have not orthogonalized the shocks (though only level and return forecast shocks have much correlation; as usual, level slope and curvature factors defined from forward rates are quite similar to similar factors defined from innovations, so orthogonal factors have nearly orthogonal shocks), and these response functions do not have a structural interpretation. The point of the response functions is merely a simple way to digest the dynamics, as an alternative to staring at a 4×4 matrix of numbers.

	$100\times\mu$	x	level	slope	curve	mse	R^2
Risk-neutral:	μ^*		ϕ^*				
x	-1.39	0.35	-0.02	-1.05	8.19	25.4	
level	0.68	0.03	0.98	-0.21	-0.22	4.77	
slope	0.00	0.00	-0.02	0.76	0.77	1.73	
curve	0.00	0.00	-0.01	0.02	0.70	0.65	
Actual:	μ		ϕ				
x	0	*0.61*	-0.02	-1.05	8.19	27.0	
level	0	*-0.09*	0.98	-0.21	-0.22	3.65	
slope	0	*-0.00*	-0.02	0.76	0.77	1.74	
curve	0	*0.00*	-0.01	0.02	0.70	0.65	

Table 4. Estimates of model dynamics, μ and ϕ in $X_{t+1} = \mu + \phi X_t + v_{t+1}$. |eig| gives the eigenvalues of ϕ, in order. mse is mean-squared error.

	x	level	slope	curve
x	1	-0.51	0.16	-0.07
level		1	-0.07	0.06
slope			1	-0.01
curve				1

Table 5. Correlation matrix of shocks V, in $X_{t+1} = \mu + \phi X_t + v_{t+1}, V = cov(v_{t+1}v'_{t+1})$. OLS estimates

The first panel of Table 4 and Figure 7 present the risk-neutral dynamics ϕ^*. A shock to the return-forecast factor x dissipates quickly, with a 0.35 correlation coefficient. (first column of Table 4, top left graph of Figure 7.) It has very little effect on the other factors.

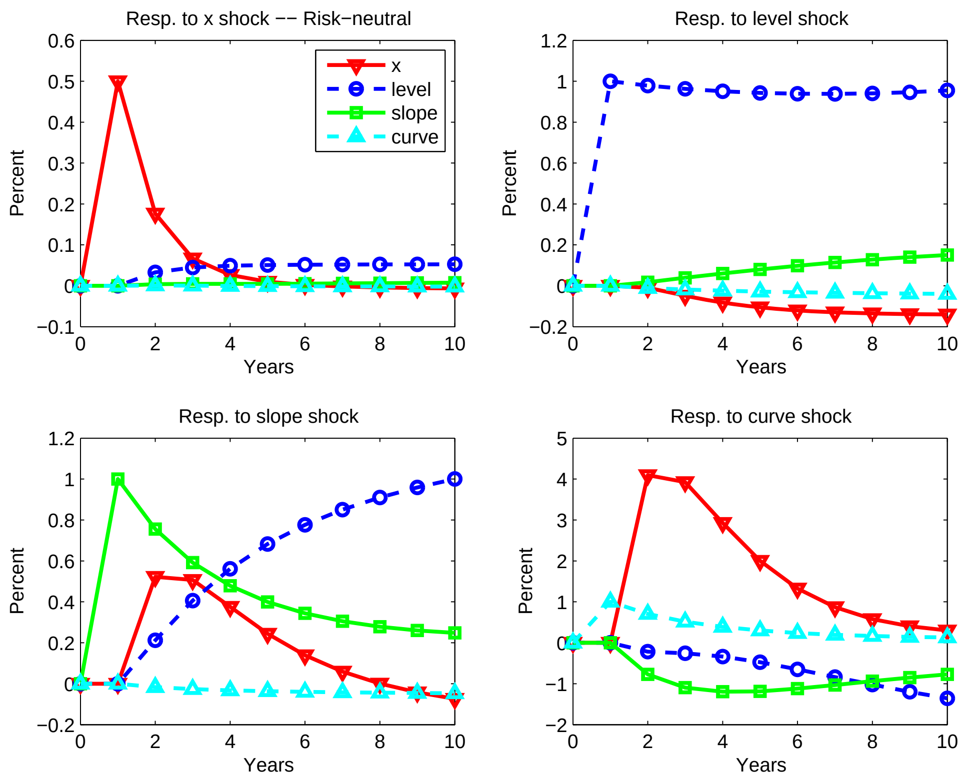

Figure 7: Impulse-response function of risk-neutral transition matrix ϕ^*. The x response is divided by two to fit on the same scale.

A movement in the level factor is essentially a "parallel shift" sending all forward rates up and lasting essentially forever with a 0.98 coefficient (second column of Table 4, top right panel of Figure 7). A slope shock decays at medium pace with a 0.76 coefficient. However, it also sets off movements in the return-forecast factor and level. A curvature shock also decays moderately with a 0.70 coefficient, and sets off movements in all the other factors.

The extreme persistence of the level shock (top right panel of Figure 7) is important. It is a natural finding: nearly "parallel shifts" or a "level factor" dominate yield curve movements in just about any characterization. Risk neutral dynamics must make sense of this movement via the expectations hypothesis (plus minor variance terms). The only way to make sense of a level movement in yields or forward rates is to suppose that expected future interest rates also rise in parallel, i.e. that dynamics have a root very close to unity. A motivating issue of our investigation, starting with Figure 1, is whether to model interest rate forecasts as mean-reverting or containing an important random walk component. The risk-neutral dynamics suggest something close to a random walk, for clear reasons: this is the only way to generate a level factor. The remaining question is, what do our estimates of true-measure ϕ dynamics say?

The middle panel of Table 4 and Figure 8 presents these central estimates of ϕ. Again, we form ϕ^* first to match the cross-section of forward rates, and then we estimate $\phi = \phi^* + V\lambda_1$,

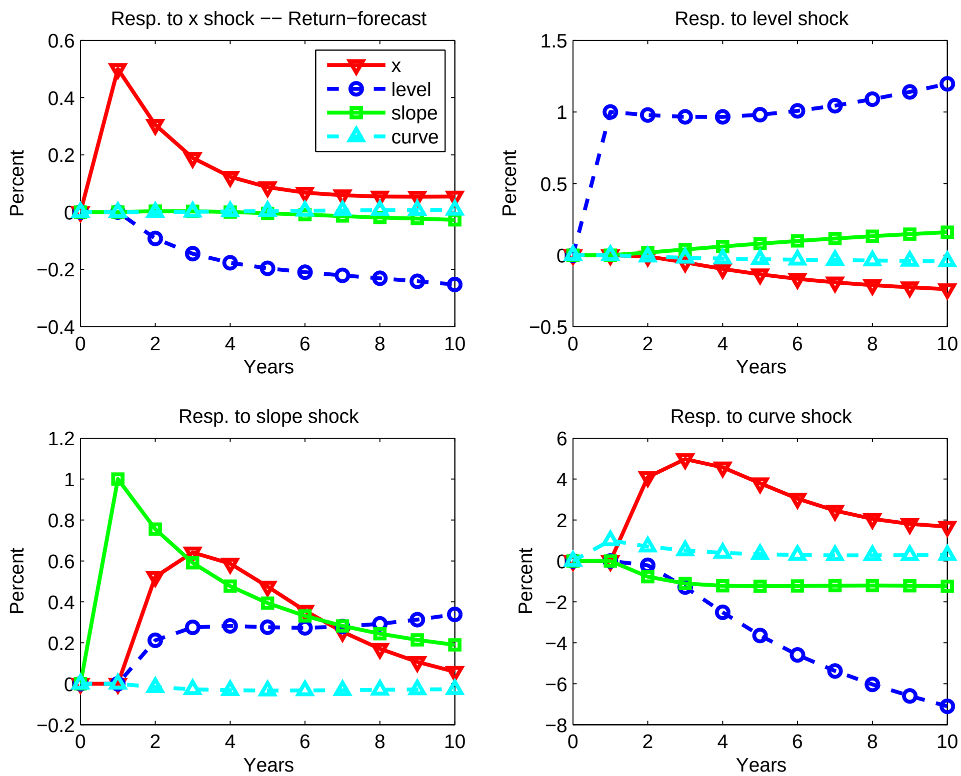

Figure 8: Impulse-response function of the real-measure transition matrix ϕ. The x response is divided by two to fit on the same scale.

and (32) means this has the restricted form:

$$\phi = \phi^* + \begin{bmatrix} cov(v_x, v_l) & 0 & 0 & 0 \\ cov(v_l, v_l) & 0 & 0 & 0 \\ cov(v_s, v_l) & 0 & 0 & 0 \\ cov(v_c, v_l) & 0 & 0 & 0 \end{bmatrix} \lambda_{0l}.$$

Thus ϕ *and* ϕ^* *only differ in their first column*, and that in a constrained way, according to the covariance of shocks with the level shock. Finally, the bottom two elements are nearly zero – shocks other than x and level are nearly uncorrelated (Table 5) – so ϕ and ϕ^* really only differ in the first two elements. We italicize the first column of ϕ in Table 4 to emphasize that the remaining columns are identical to the risk-neutral columns.

This relationship is important. Estimates of transition dynamics built on time-series evidence are subject to substantial statistical uncertainty. Estimates of the risk-neutral dynamics are, by contrast, measured with high precision. The only error is the 16 bp or so fitting or "measurement" error in the regression of contemporaneous forward rates on factors – this is not a forecasting regression – and that error can be made arbitrarily small by increasing the number of factors. If the data followed an exact factor model, we would know the risk-neutral dynamics with *perfect* precision, even in a short time-series with substantial uncertainty in transition-matrix estimates.

If we allow arbitrary market prices of risk, all this precision vanishes in our estimate of ϕ. However, the compelling structure we are put on market prices of risk means that all but the first column of ϕ – and, really, all but the top two elements – are measured from the cross-sectional, risk-neutral dynamics ϕ^* ignoring *any* time-series evidence.

In particular, the $\phi_{l,l}$ element must remain unchanged at 0.98 (middle panel of Table 4, top right plot of Figure 8), so level shocks still have very long-lasting effects. To avoid this conclusion; to get mean-reversion like we find in the OLS dynamics, we would need expected returns to vary with the level factor. The finding that they do not – that 99.9% of the variance of expected returns comes from the x factor – drives our estimate of a much more persistent level factor than indicated by OLS dynamics.

The remaining elements of the actual dynamics ϕ (Middle panel, Table 4, Figure 8) are only slightly changed from the risk-neutral counterparts. Of course these changes are important, as if there are no changes then returns are not forecastable and there are no risk premia. The return-forecast factor is a bit more persistent, with a 0.61 own coefficient rather than 0.35, and a more drawn-out response. As the $\phi_{l,x}$ coefficient is increased, a return-forecast x shock is now followed by a slow decline in the level factor. The pattern of the other response functions is broadly unchanged. There is an interesting block structure: Level shocks have no effect on anything else (top right). Return forecast shocks persist and set of a small level effect, with no effect on the other factors. Slope shocks affect slope, level, and return-forecast, but not curve. Curve shocks affect all of the factors.

Life would be simple if expected returns followed a simple AR(1) process, disconnected from other factor movements. Then we could largely avoid questions of the "term structure of risk premia." When current expected returns are high, so would be future expected returns, and forward rates of all maturities would be higher than spot rates. We could legitimately talk about "the" forward premium. All the terms of the forward premium decomposition (5) or yield premium decomposition (6) would move together. The fact that slope and curve shocks feed back to the expected return factor shows that we cannot simplify the world in this way. If slope is high, even if *current* expected returns are zero (no x movement), we forecast that *future* expected returns will decline, and vice-versa for curvature (see the x responses to slope and curve shocks in Figure 8, or the loadings in the x rows of Table 3). The forward and yield premium decompositions can feature interesting term-structures of risk premia. Since this feature comes from the well-measured second through fourth columns of ϕ^*, unaffected by market prices of risk, it is not easily avoidable by appeal to sampling error.

The responses of the risk-neutral ϕ^* and estimated ϕ dynamics do not seem to be settling down after the 10 years plotted in Figures 7 and 8. In fact, the maximum eigenvalue of ϕ^* is 1.02 and the maximum eigenvalue of ϕ is 1.05, both slightly larger than one. If bonds of all maturity are traded, an eigenvalue ϕ^* above one leads to an arbitrage opportunity, so can be ruled out. We can easily impose that the eigenvalues of ϕ^* are less than one in our estimation procedure. Doing so produces a very slight deterioration in cross sectional fit, seen in Table 3. Eigenvalues of ϕ^* slightly greater than one let the model fit the slight upward curve in the x and level loadings, the top two panels of Figure 5. Imposing an eigenvalue less than one makes the right side of these loadings a straight line. Since the long-term data are fitted from a statistical model that does not impose eigenvalues less than one, these may well be spurious phenomena. However, almost nothing of substance for predictions out to 20 year time or maturity is changed by imposing eigenvalues of ϕ^* less than one vs. leaving the estimate as is, so we leave the simple result. Obviously, taking extreme long-maturity limits of the estimates is dangerous in any case. Given either estimate of ϕ^*, the largest eigenvalue of ϕ is a gently rising function of the market price of risk. Using a lower market price of risk makes almost no difference. We can obtain an eigenvalue of ϕ less than one only by forcing the eigenvalue of ϕ^* to be correspondingly lower than one. Again, this improves the esthetics of the right hand side of the graphs, but makes essentially no difference to our results.

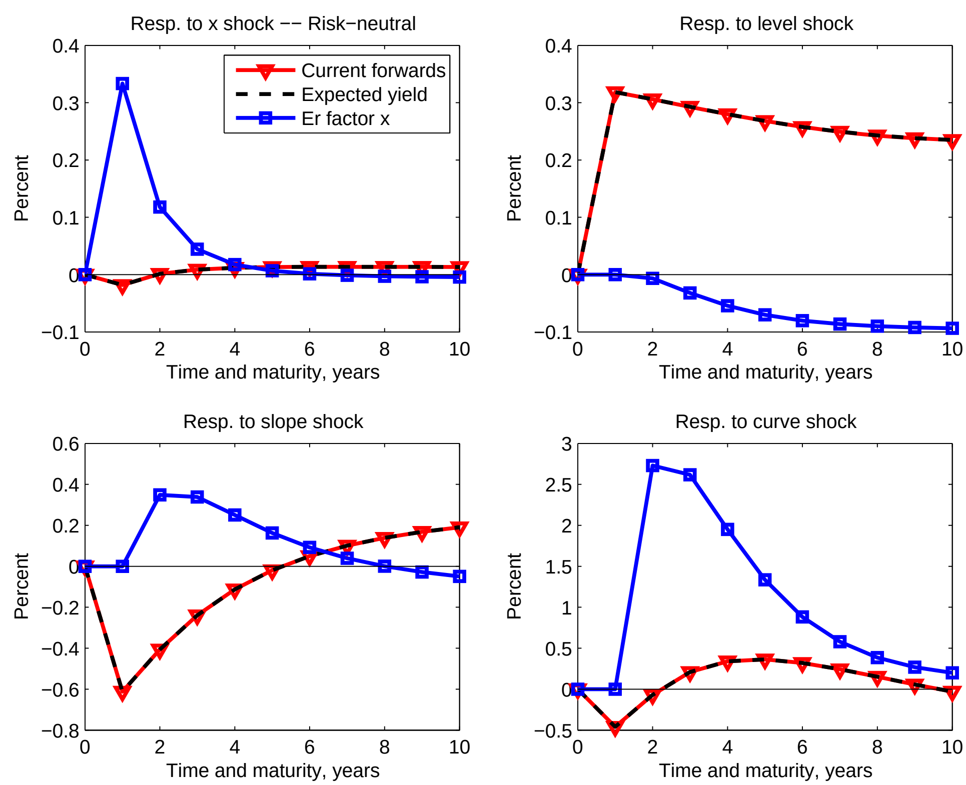

Figure 9: Response of expected future one-year yield, current forward rates, and return-forecast factor x to unit shocks, under the risk-neutral transition matrix ϕ^*.

As another perhaps more intuitive way to understand the dynamics, Figures 9 and 10 plot the responses of expected one year yields $E_t y^{(1)}_{t+n-1}$, current forward rates $f^{(n)}_t$ and the expected future return-forecast factor x_t to each of the factor shocks. Under risk-neutral dynamics in Figure 9, the response of expected future yields is, of course, exactly the same as the response of current forward rates. A change in the return forecast factor quietly decays (top left panel). A change in the level factor sends current forward rates and expected future interest rates up for a long maturity and time, of course. Slope and curvature shocks lead to sensibly shaped movements in expected future rates and current yields; each leads to substantial movements in the return-forecast factor as seen above.

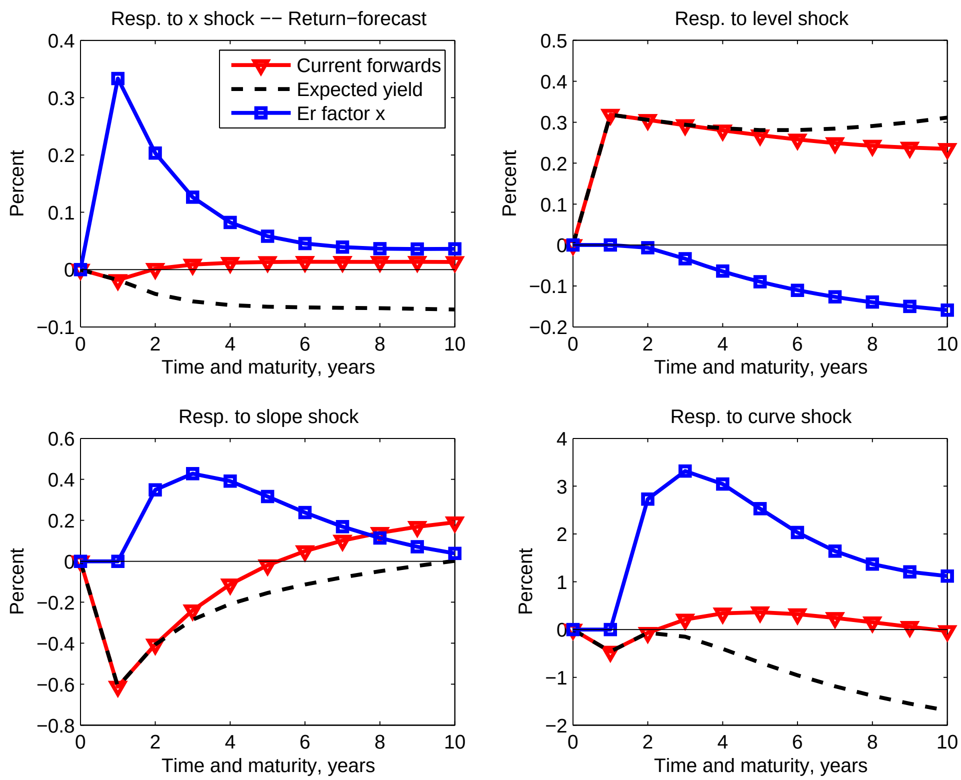

Figure 10: Response of expected future one-year yield, current forward rates, and return-forecast factor x to unit shocks, under the estimated transition matrix ϕ.

In our estimate of the real transition matrix ϕ, Figure 10, we can see the crucial divergence between current forward rates and expected future interest rates. This is the central question for a yield curve decomposition, and in this figure we can see how the forward curve decomposition at any date depends on factor configurations at that date. If there is a large return-forecast factor x_t, and the other factors are zero (top left panel), we see that current forwards are nearly constant. Expected future interest rates decline, however, so forwards exceed expected future interest rates. As the expected return x reverts, the forward and expected yield curves become parallel, all as one expects from the forward decomposition (5).

However, movement in the other factors can also lead to a divergence between forward rates and expected future interest rates, despite no *current* risk premium. A level shock leads to a slow decline in future risk premia, so forwards do not rise as much as expected future yields. Here is the source of the slightly greater eigenvalue of actual vs. risk neutral dynamics – since expected returns decline in response to the level shock, expected future yields rise *even more* than forward rates, which follow the "level" pattern.

More importantly, a large slope or curvature set off a rise in *future* expected returns so forwards move more than yields. These are crucially important responses in our decompositions. To get a large spread between forward rates and expected future spot rates, we need not so much a large *current* expected return, seen in the x factor, but large slope or curvature, which can forecast large and, more importantly, long-lived, movements in expected *future* one-year return premia.

5.2 Decompositions

We can now use the estimated model to decompose the yield curve. At any moment in time, how much of the forward rate corresponds to expected future interest rates, and how much to risk premia? What is the term structure of such risk premia – are forward rates responding to current expected excess returns, or to expectations of future expected excess returns?

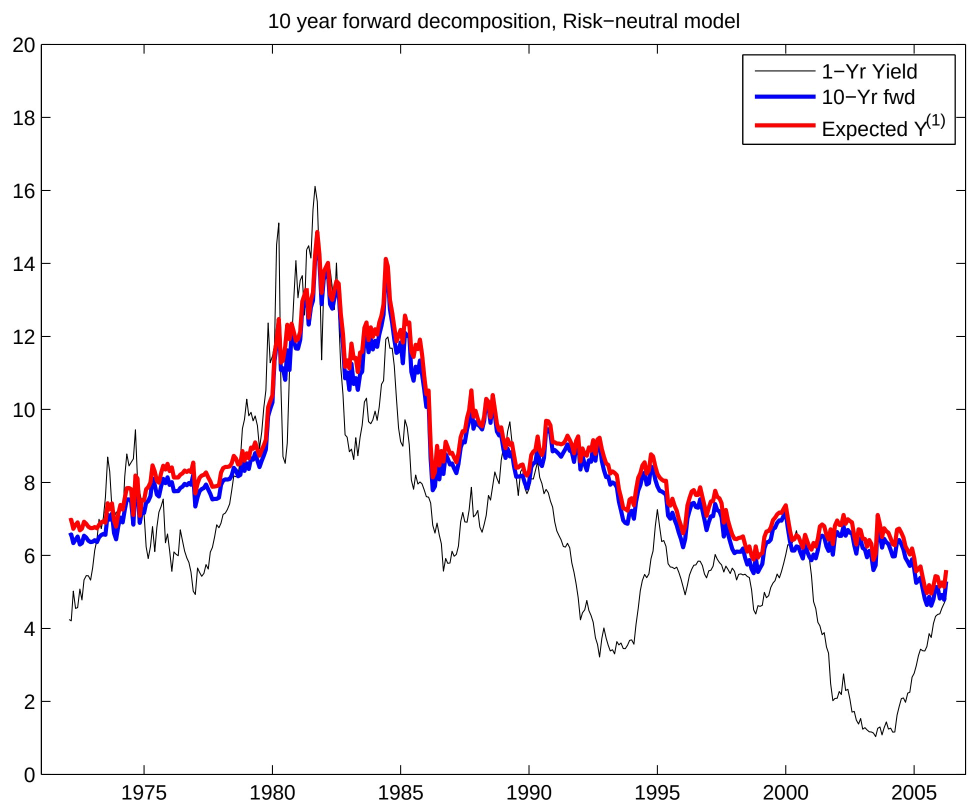

Figure 11: Current one-year rate, 10 year forward rate $f_t^{(n)}$, and expected one-year rate $E_t y_{t+9}^{(1)}$, generated from the risk-neutral model.

Figure 11 starts by plotting the current one-year rate, $y_t^{(1)}$, the 10 year forward rate $f_t^{(10)}$ and the expected one-year rate $E_t y_{t+9}^{(1)}$ as generated by the risk-neutral model. Risk neutral isn't quite expectations hypothesis because of $\frac{1}{2}\sigma^2$ terms, but the plot shows that these terms are small, and do not matter much for the rest of our discussion

Figures 12 and 13 present the corresponding decompositions computed from our estimated affine model. You can see sharp divergences between forward rates and expected future interest rates. Far from reverting quickly to the mean as in the top panel of Figure

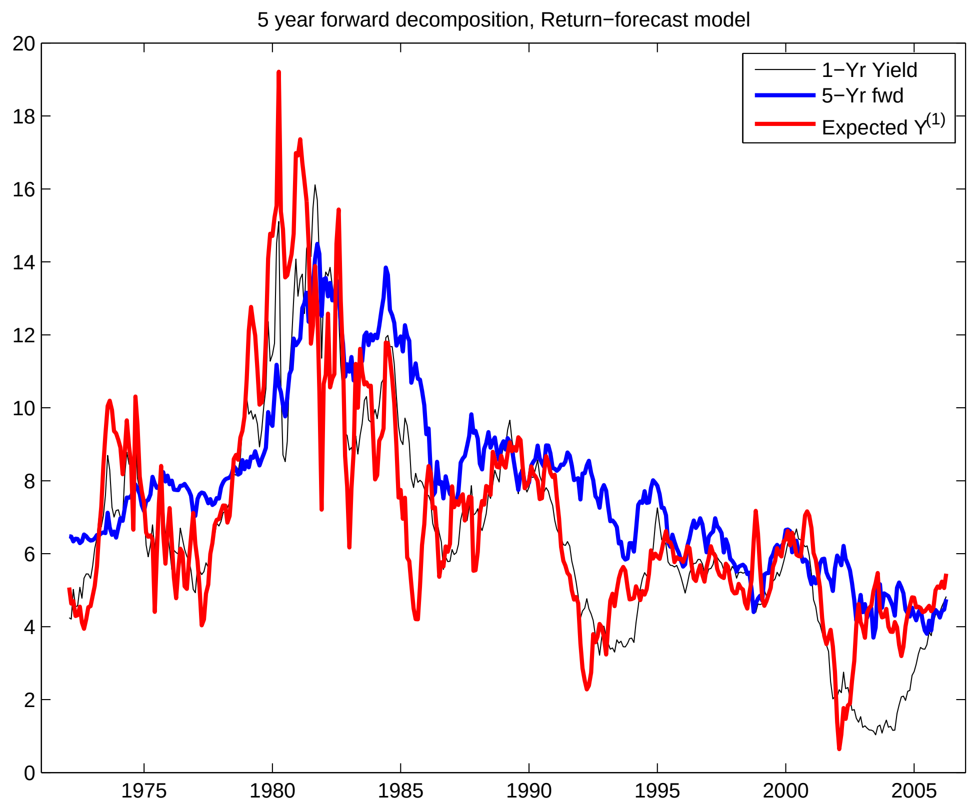

Figure 12: Current one-year rate $y_t^{(1)}$, 5 year forward rate $f_t^{(5)}$, and expected one-year rate $E_t y_{t+4}^{(1)}$ computed from estimated affine model.

1, expected future interest rates track current rates through the three rate declines of the recent sample. In fact, expected future rates decline *faster* than current rates – the expected future rates move even more than the cointegrated representation of the bottom panel of Figure 1. This is a reasonable finding: market participants quite plausibly knew rates would decline further in these episodes. These are periods of large current expected returns, as we found in our 2005 paper. With forward rates that do not move much, high expected returns means expectations of further yield declines. However, past the bottom, expected interest rates rise quickly, eliminating the forward premium. Again, this is the pattern we have seen in the return-forecast factor. It says to "get out" after the bottom, even when forward-spot spreads are still forecasting return premia. In the late 1970s, by contrast, the model says that investors expected interest rates to climb above current rates, so the comparatively low forward rates implied a negative risk premium on long term bonds.

Are the large risk premia seen in these figures due to this year's expected returns, or to expectations of future premia? Figure 14 presents the terms of the forward-rate decomposition (5) along with the spread between the 5 year forward rate $f_t^{(5)}$ and the expected one-year rate $E_t y_{t+4}^{(1)}$. Each line represents the cumulated terms of (5). The line closest to zero is the contribution of one-year expected returns $E_t \left(rx_{t+1}^{(5)} - rx_{t+1}^{(4)} \right)$; the next line adds

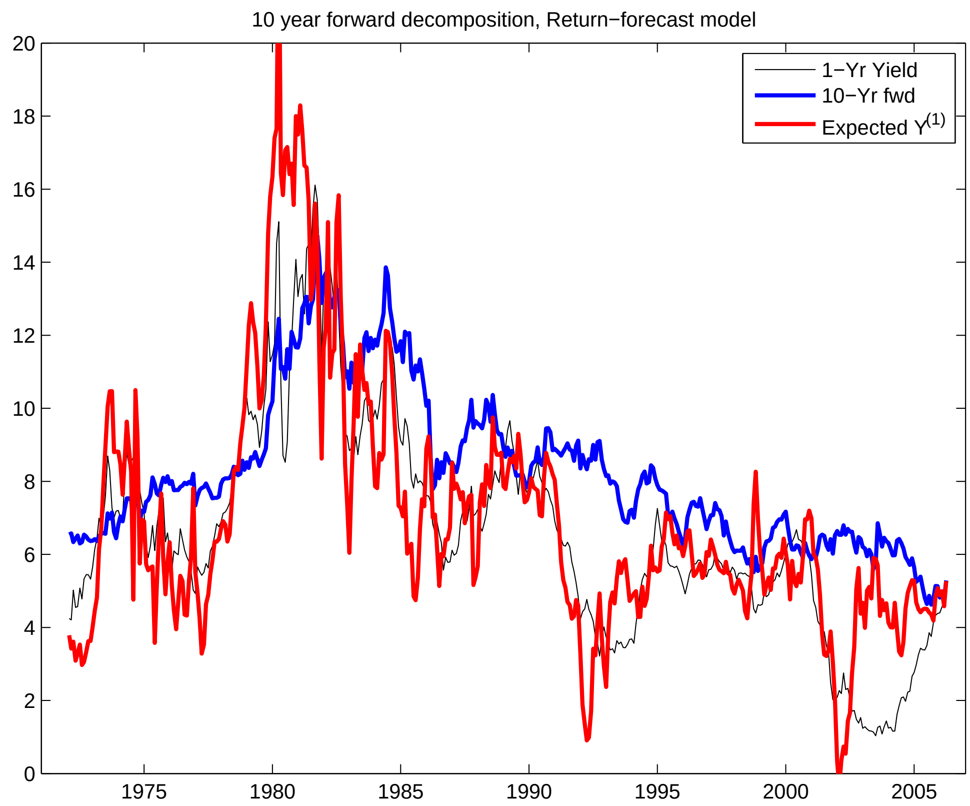

Figure 13: Current one-year rate $y_t^{(1)}$, 10 year forward rate $f_t^{(10)}$, and expected one-year rate $E_t y_{t+9}^{(1)}$ computed from estimated affine model.

to that the second year of expected returns $E_t\left(rx_{t+2}^{(4)} - rx_{t+2}^{(3)}\right)$ and so forth. The forward rate and expected one-year rate are the same as plotted Figure 12. As the figure shows, the 5 year risk premium is typically due to expectations of one-year return premia many years into the future. In typical times of large premium, the first year risk premium only accounts for 20% or so of the total premium. Since the return-forecast factor mean-reverts fairly quickly, these times are times in which the constellation of other factors implies large future risk premia. In particular, these are typically times of a large positive slope, which leads to large expected future risk premia as seen in the impulse-responses. A big forward spread needs a strong slope *and* a large current return forecast.

Finally, Figure 15 presents a spectrum of interest rate forecasts, as a counterpart to Figure 1. The large root imposed by the risk-neutral dynamics means that, although the model is specified as a VAR(1) in levels, interest rate forecasts end up looking a lot more like the unit root/cointegrated representation of the bottom part of Figure 1. In fact, more so: most notably in the interest rate declines going into the last three recessions, expected future one-year rates decline even beyond current rates, where Figure 1 still shows modest rises.

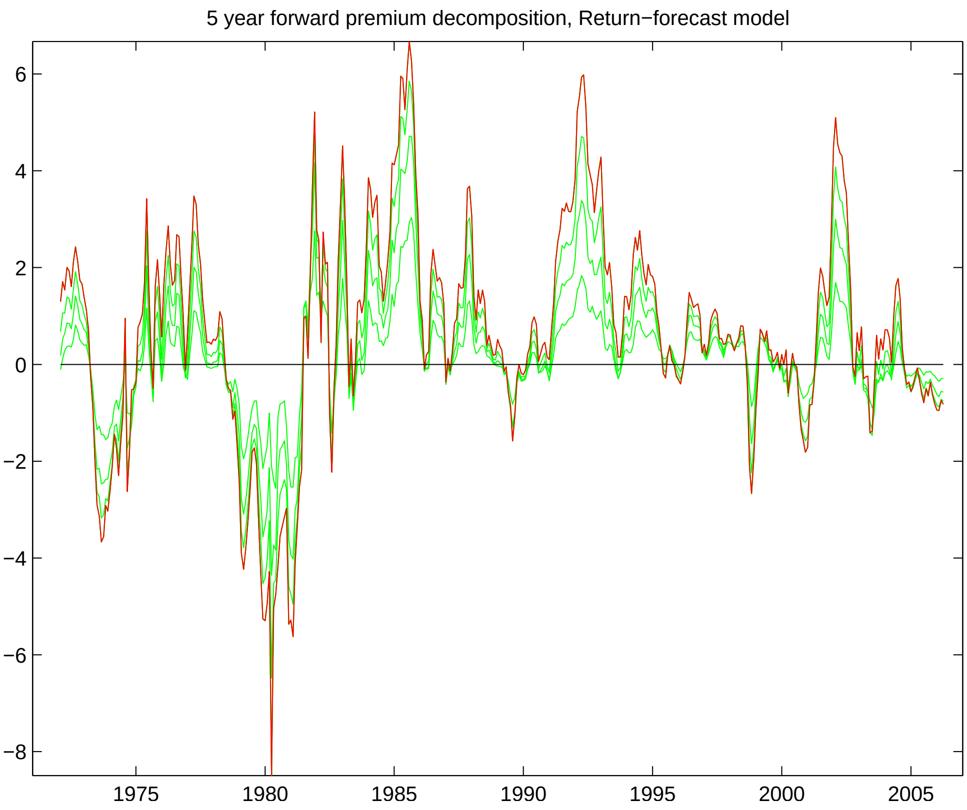

Figure 14: Spread between five year forward rate and expected one-year rate $f_t^{(5)} - E_t y_{t+4}^{(1)}$, and the terms of its decomposition into one-year risk premia. The line closest to the origin gives the first year premium, $E_t\left(rx_{t+1}^{(5)} - rx_{t+1}^{(4)}\right)$; the next line adds the second year premium $E_t\left(rx_{t+2}^{(4)} - rx_{t+2}^{(3)}\right)$ and so forth.

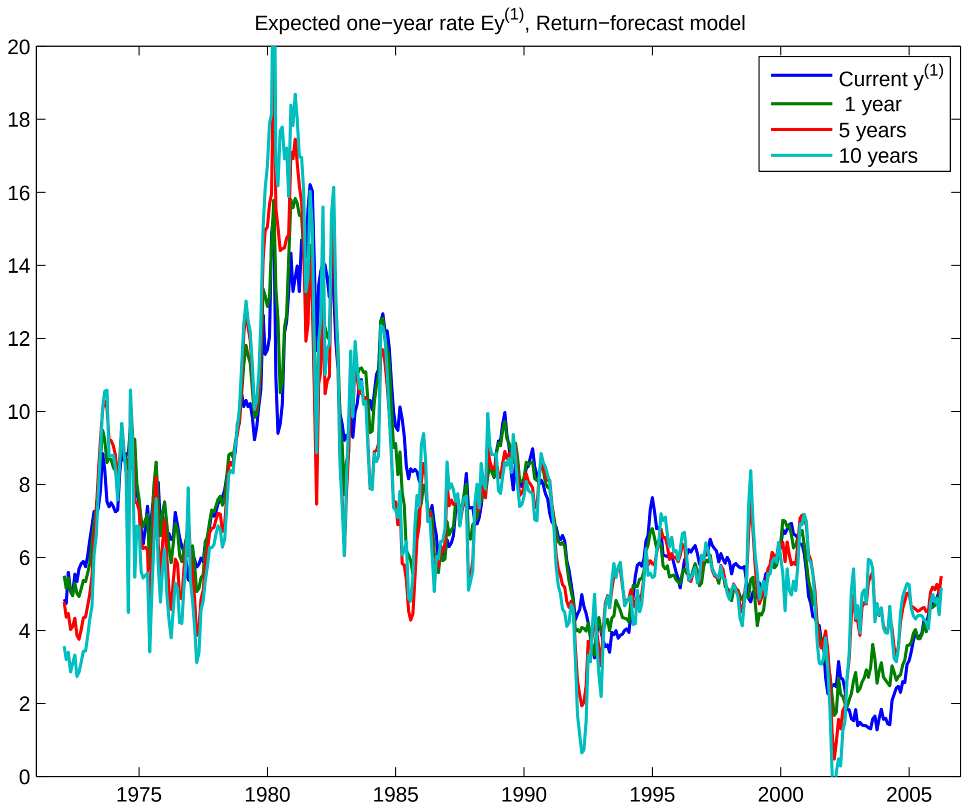

Figure 15: Forecasts of one-year rates $Ey^{(1)}_{t+n}$ from the affine model

5.3 The conundrum

The decomposition of the yield curve in the last few years has attracted a lot of attention. Most famously, in February 2005 testimony[2], Federal Reserve Chairman Alan Greenspan called the lowering of long-term yields while short term rates rose a "conundrum". He elaborated in a speech[3] on June 6 2005, opining that "The pronounced decline in U.S. Treasury long-term interest rates over the past year despite a 200-basis-point increase in our federal funds rate is clearly without recent precedent." The importance of separating expectations from risk premiums in this context was made clear in a speech[4] by Governor Donald Kohn in July 2005:

> Nothing better illustrates the need to properly account for risk premiums than the current interest rate environment: To what extent are long-term interest rates low because investors expect short-term rates to be low in the future... and to what extent do low long rates reflect narrow term premiums, perhaps induced by well-anchored inflation expectations or low macroeconomic volatility?

Chairman Bernanke continued the inquiry. In a widely-reported speech[5] on March 20 2006, he made clear how we separate yield curves into "expectation" and "risk premium" components, and how stories about "demands" by various agents are the same thing as a risk premium. He continued

> According to several of the most popular models, a substantial portion of the decline in distant-horizon forward rates over recent quarters can be attributed to a drop in term premiums ...the decline in the premium since last June 2004 appears to have been associated mainly with a drop in the compensation for bearing real interest rate risk.

He cited Kim and Wright (2005) here, undoubtedly the first-ever citation by a sitting Federal Reserve Chairman to an "Arbitrage-Free Three-Factor Term Structure Model" in a public speech. Chairman Bernanke also explained why he thinks the issue is important:

> What does the historically unusual behavior of long-term yields imply for the conduct of monetary policy? The answer, it turns out, depends critically on the source of that behavior. To the extent that the decline in forward rates can be

[2]Testimony of Chairman Alan Greenspan Federal Reserve Board's semiannual Monetary Policy Report to the Congress Before the Committee on Banking, Housing, and Urban Affairs, U.S. Senate February 16, 2005. All Federal Reserve official's testimony and speeches are available at http://www.federalreserve.gov/newsevents.htm

[3]Remarks by Chairman Alan Greenspan Central Bank panel discussion to the International Monetary Conference, Beijing, People's Republic of China (via satellite) June 6, 2005.

[4]Remarks by Governor Donald L. Kohn at the Financial Market Risk Premiums Conference, Federal Reserve Board, Washington, D.C. July 21, 2005.

[5]Remarks by Chairman Ben S. Bernanke before the Economic Club of New York, New York, New York March 20, 2006.

traced to a decline in the term premium, perhaps for one or more of the reasons I have just suggested, the effect is financially stimulative and argues for greater monetary policy restraint, all else being equal.....

However, if the behavior of long-term yields reflects current or prospective economic conditions, the implications for policy may be quite different–indeed, quite the opposite. The simplest case in point is when low or falling long-term yields reflect investor expectations of future economic weakness. Suppose, for example, that investors expect economic activity to slow at some point in the future. If investors expect that weakness to require policy easing in the medium term, they will mark down their projected path of future spot interest rates, lowering far-forward rates and causing the yield curve to flatten or even to invert....

One doesn't have to agree with the conclusions about optimal monetary policy to accept Chairman Bernanke's endorsement that the issue is important.

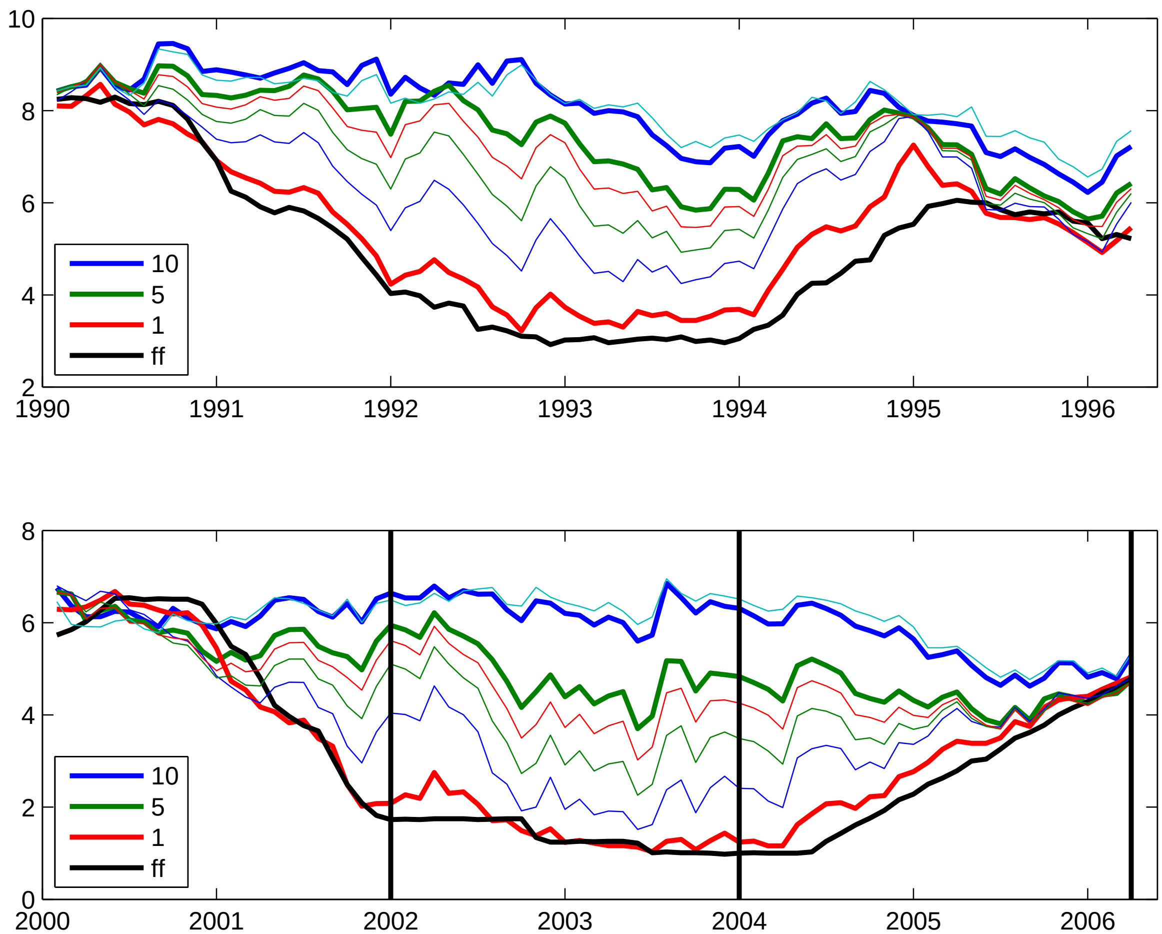

Figure 16: Forward rates in two recessions. The federal funds rate, 1-5, 10 and 15 year forward rates are plotted. Federal funds, 1, 5 and 10 year forwards are emphasized. The vertical lines in the lower panel highlight specific dates that we analyze more closely below.

It is unfortunate for the excitement of our application, but fortunate for our statistical approach that imposes some kind of stationarity so we can analyze the present by exploiting historical correlations, that recent history is in fact not at all that unusual. Figure 16 contrasts the behavior of forward rates through the last two recessions, conveniently occurring almost exactly a decade apart. The pattern is strikingly similar. Short-term yields and forwards decline, spreads widen, and then yields and forwards recover as spreads tighten again. Long term forwards decline overall. The only difference is the exact timing of the forward rate movements.

"What conundrum?," we are tempted to ask. First, long forwards *should* fall when the Fed tightens. Tighter policy now means lower inflation later, and thus *lower* nominal rates in 10 years. In 1994, the opposite nearly one-for-one *rise* of long forwards with rises in the Federal Funds rate was viewed as a conundrum for just this reason. Second, the term premium *should* be negative. In a world with stable inflation, so that interest rate variation comes from variation in *real* rates, long-term bonds are safer investments for long-term investors. Were we Fed Chairs testifying to Congress with the plots of Figure 16 in hand, we would be tempted to say simply that markets believe inflation is conquered, and you're welcome.

This isn't a policy paper however. The "conundrum" provides an interesting historical episode on which to evaluate our method for decomposing the yield curve, and Figure 17 gives our model's analysis of this period.

The top panels and the bottom left panel present the cross-section of forward rates on the three indicated dates, which are also marked by vertical lines in the bottom right panel. "Fwds" gives the forward curve on each date and "$Ey^{(1)}$ gives the expected future one-year rate, $Ey^{(1)}_{t+j}$ as a function of j. If the expectations hypothesis held, these two lines would coincide. The forward premium for each maturity is the vertical distance between "Fwds" and "$Ey^{(1)}$" lines.

$E(rx^{(10)}/2)$ gives the expected return of a 10 year bond through time, $E_t(rx^{(10)}_{t+j}/2)$ as a function of j. We divide by two so that this line fits on the same scale as the others. The point of this line is to understand the term structure of risk premia, via the decomposition (5),

$$f^{(n)}_t - E_t\left(y^{(1)}_{t+n-1}\right) = E_t\left(rx^{(n)}_{t+1} - rx^{(n-1)}_{t+1}\right) + E_t\left(rx^{(n-1)}_{t+2} - rx^{(n-2)}_{t+2}\right) + \ldots + E_t\left(rx^{(2)}_{t+n-2}\right). \tag{35}$$

In general, understanding this decomposition for each maturity at a give date requires us to plot a lot of lines. Fortunately, the structure of the affine model allows a much simpler understanding. First, expected excess returns all depend on the single state variable x_t, so the risk premium terms on the right hand side of (35) are linear functions of x_t, $E_t\left(x_{t+1}\right)$, $E_t\left(x_{t+2}\right)$, and so forth. Second, the loading of $E_t\left(rx^{(n)}_{t+1}\right)$ on x_t is nearly a linear function of n, so to a good approximation $E_t\left(rx^{(n)}_t - rx^{(n-1)}_t\right) \approx a + bx_t$ is the same for all n. Thus, ignoring the constant to focus on variation over time, we have to a very good approximation

$$f^{(n)}_t = E_t\left(y^{(1)}_{t+n-1}\right) + ax_t + aE_t\left(x_{t+1}\right) + aE_t\left(x_{t+2}\right) + .. + aE_t\left(x_{t+n-2}\right).$$

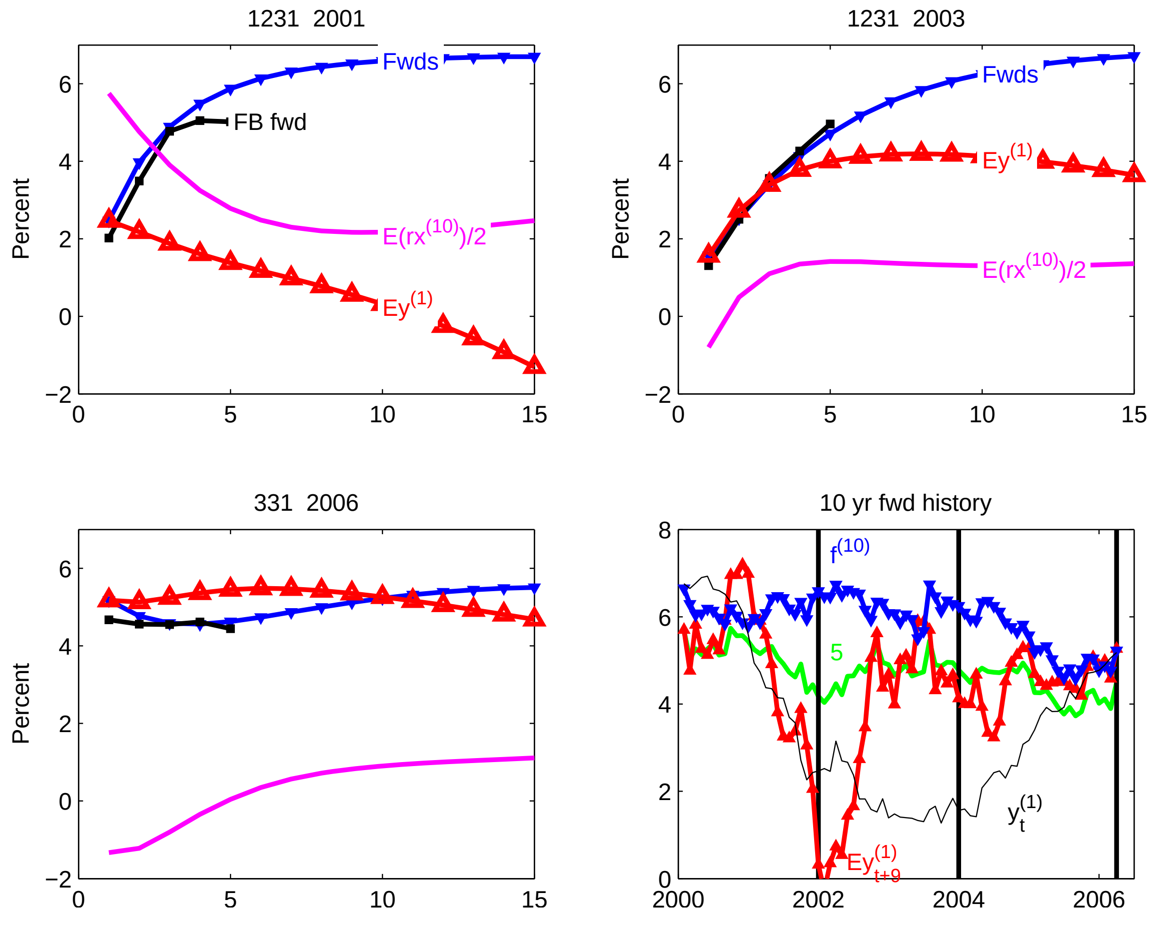

Figure 17: Decomposition of the 10 year forward rate. The top panels and the bottom left panel present the cross-section of forward rates on the three indicated dates, which are also marked by vertical lines in the bottom right panel. Fwds gives the forward curve on each date, FB fwd gives the Fama-Bliss forward curve on the same dates. $Ey^{(1)}$ gives the expected future one-year rate, $Ey^{(1)}_{t+j}$ as a function of j. $E(rx^{(10)}/2)$ gives the expected return of a 10 year bond through time, $E_t(rx^{(10)}_{t+j}/2)$ as a function of j. In the bottom right panel, $f^{(10)}$ is the 10 year forward rate, $y^{(1)}_t$ is the current one-year rate, $Ey^{(1)}_{t+9}$ is the expected value at each date of the future one-year rate, i.e. $E_t(y^{(1)}_{t+9})$ as a function of t, and 5 gives $E_t(y^{(1)}_{t+9})$ plus the first five years of risk premia in the forward decomposition, i.e. $E_t(y^{(1)}_{t+9})$ $+E_t(rx^{(10)}_{t+1} - rx^{(9)}_{t+1}) + E_t(rx^{(9)}_{t+2} - rx^{(8)}_{t+2}) + ... + E_t(rx^{(6)}_{t+5} - rx^{(5)}_{t+5})$.

Therefore, we can understand the term structure of the forward premium - the extent to which the difference between the 10 year forward and the expected one year rate 9 years in the future depends on *this* year's expected excess returns rather than expectations of *next* year's excess returns, etc.– by just tracking forecasts of the return-forecasting variable x_t, or by tracking forecasts of any single expected excess return that is a linear function of x_t. Figure 17 does this by presenting expectations of the 10 year bond's expected excess return.

The bottom right panel of Figure 17 tracks a smaller number of variables through time. $f^{(10)}$ is the 10 year forward rate, $y_t^{(1)}$ is the current one-year rate, and $Ey_{t+9}^{(1)}$ is the expected value at each date of the future one-year rate, i.e. $E_t(y_{t+9}^{(1)})$ as a function of t. Thus, in this panel the 10 year forward premium is given by the vertical distance between $f^{(10)}$and $Ey_{t+9}^{(1)}$ lines. Adding all the terms of the forward decomposition (5) clutters up the graph too much, so the line marked "5" gives the first 5 terms of that decomposition, i.e. $E_t(y_{t+9}^{(1)})$ $+E_t(rx_{t+1}^{(10)} - rx_{t+1}^{(9)}) + E_t(rx_{t+2}^{(9)} - rx_{t+2}^{(8)}) + ... + E_t(rx_{t+5}^{(6)} - rx_{t+5}^{(5)})$. The vertical distance between the "5" line and $Ey_{t+9}^{(1)}$ gives the contribution to the 10 year forward premium from the *first* five years of expected returns, while the vertical distance between the "$f^{(10)}$" line and the "5" line gives the contribution to the 10 year forward premium from the remaining (years 6-9) years of expected returns.

Having met the players, we can now tell the story. Start with December 2001, the first slice of the bottom right panel and the date plotted in the top left panel of Figure 17. The one-year risk premium was large, represented by the large value $E_t(rx_{t+1}^{(10)})$ in the left-hand end of the top left panel (in fact, x reaches its peak in this sample on January 2002). This fact translates to the large difference between the first forecasted one year rate $E_t\left(y_{t+1}^{(1)}\right)$ and its corresponding forward rate, shown in the second data points of the top left panel. The model was right on this occasion, as December 2001 was a great time to buy long-term bonds. Interest rates continued to decline for two years, despite high long-term yields and forward rates, generating great returns for long-term bond holders. Our measure of the one-year risk premium is so high because the Fama-Bliss forward rate data show a strong "tent shape", which is the signal of high expected returns. This date is particularly interesting because it is one of the few dates in which the Fama-Bliss forward curve and the GSW forward curve are substantially different. The GSW forward curve smooths over the decline in 4 and 5 year forwards at this date, but in doing so it throws out the crucial information that the FB forward curves use to forecast expected returns. Data points like these are why we found in Table 1 that Fama-Bliss data do a better job of forecasting even GSW returns. Notice that FB and GSW data are nearly identical on the other dates in this analysis.

However, the one-year return risk premium is not expected to stay constant. For the first few years the risk premium (measured by the $E(rx^{(10)})/2$ line) declines quickly, following the response to an x shock picture we might expect from the top left panel of the impulse-response Figure 8. However, expected returns then stop declining at about 2%, and stay at about 2% out to a 15 year maturity. Both slope and curvature are large in December 2001, as you can see from eyeballing the forward curve, and large slope and curvatures set off large *future* risk premia, even when memory of the current x_t has melted away, as we saw in the bottom panels of the impulse-response functions Figure 8. Each year of expected large returns must correspond to further *divergence* between forward curve and expected one-year rates, and the expected one-year rate curve keeps going down. The conjunction of large return-forecast (tent), with large slope and curvature factors makes this date a perfect storm for forward premia, and in fact this date is one of the largest forward premiums seen anywhere in the sample.

The history of the bottom panel of Figure 17 shows the events that brew this storm. Interest rates $y_t^{(1)}$ have been falling while the 10 year forward stayed relatively constant. As current (x) and future (slope, curve) risk premia rose, the expected future interest rate has fallen even faster than the one-year rate. The "5" line shows that the rise in one-year risk premia accounts for the bulk, but not all of the large risk premium on this date.

The sharp decline in the $Ey_{t+9}^{(1)}$ curve through 2001 is notable. Expected future interest rates fell more than current interest rates. This seems to us a a sensible reading of history (and certainly more reasonable than a constant 6% expectation as suggested by the top panel of Figure 1). In late 2002, the Fed had been following a sharp and unprecedented decline in Federal funds rates (see Figure 16). The economy was in a recession following a sharp stock market decline. It was not clear how low the Fed would go or for how long. There was much talk about deflation and the dangers of repeating Japan's experience. Economists were writing papers about helicopter drops, fiscal expansion, and impact of a zero lower bound on nominal interest rates for the conduct of monetary policy.

By December 2003, things have settled down considerably. We have now had two years of stable low interest rates and steady macroeconomic growth. In the top right panel of Figure 17 we see the forward curve is still upward sloping and curved, but it has completely lost the tent shape that generates one-period expected returns. The Fama Bliss data now, as typically, agree with the GSW data and the return-forecast factor x (not shown) is in fact very slightly negative. Once again, this proved a correct forecast, and return forecasts based on the slope of the yield curve proved wrong: interest rates were poised to rise, and long-term bond investors did not make good returns in 2004. Following this low one-year risk premium, expected future interest rates track the forward curve almost exactly for the first few years. There is no risk premium in 3 year forward rates and nearly none in 4 year forward rates. However, the large slope (about its maximum value in the sample on this date) forecasts a rise in *future* expected returns. Thus, starting in year 4, the forward curve and expected interest rate curve diverge, leaving a substantial 10 year forward premium even though there was no 4 year forward premium. This is a good date on which to remember there is no single "forward premium," there is a *term structure* of forward premia.

The bottom right panel shows how we got to this date. Interest rates $y_t^{(1)}$ declined slowly in 2002 and 2003, and the 10 year forward rate again didn't change much. The tent factor melted away, and so our measure of expected interest rates $Ey_{t+9}^{(1)}$ rose quickly. Almost the entire 10 year forward premium comes from years 6-10 in the bottom right panel. Again, the quick rise in expected interest rates $Ey_{t+9}^{(1)}$ seems plausible to us. It was clear the economy would recover, interest rate rises were on their way, repeating the 1994 experience. What is unusual is that our factor model is able to read these expectations from bond data.

At the end of our sample, in March 2006, the forward curve is essentially flat, and so are expected future interest rates (bottom left panel of Figure 17). There is no tent, no slope and no curvature. Slightly negative expected current returns send expected interest rates a bit above the forward curve for short maturities – there is a slight negative forward premium up to 10 years– but slightly positive expected future returns have the opposite effect for larger maturities.

So, what "caused" the decline in the 10 year forward rate from 2004 to 2006, or at least how do we decompose it to interest rate expectations and risk premiums? Following the history in the bottom right panel of Figure 17, we see the rise in interest rates, together with the decline in the 10 year forward rate that sparked the "conundrum." Broadly speaking, the decline in forward rates can, in fact, be chalked up to a decline in risk premium. From Dec 2003 to March 2006 the expected one-year rate $Ey_{t+9}^{(1)}$ only rises about a percentage point, while the forward rate declines. We see more: almost all of the decline in risk premium happened in the first year 2004. By Jan 2005, the 10 year forward and expected one-year rate are nearly equal, and track each other well. Interestingly, the zero overall risk premium corresponds through the year, as on the last date plotted in the bottom right panel, to a slightly negative near-term premium and a slightly positive longer-term premium, as the green "5" line is below both forward and expected one-year curves in the last year.

However, one can just as well summarize the plot with the opposite conclusion: the 10 year forward premium fell two percentage points from Jan 2005 to March 2006, *as it always does* in a period of rising rates and macroeconomic growth, and exactly as it did in previous episodes. Thus the forward rate would not have fallen if expected one-year rates had risen *as they usually do* in such periods. One can then debate whether one interprets the failure of expected interest rates to rise as a successful anchoring of inflation expectations or a mysterious delinking of long from short rates, as Greenspan apparently was doing.

However, it's equally clear that this is far from a puzzling event. A one or two percent forward premium in either direction is small compared to the 7 percent forward premium in December 2001. Small and even negative forward premiums are common at this stage of the business / interest rate cycle, as we saw in the full-sample 10 year forward rate decomposition of Figure 13.

6 Variations

6.1 Less risk premium

The estimate of the market price of risk λ_{1l} is the central source of sampling uncertainty in our procedure. It comes, fundamentally from the size of regression return forecasts

$$q_r' rx_{t+1}^{(n)} = \alpha + \beta x_t + \varepsilon_{t+1}$$

In addition to pure sampling error, this regression (and the formation of the x factor) is likely to be overfit and hence biased upwards. Fortunately, it is a single parameter λ_{0l}. To investigate its effect, we present here some key graphs that use half the estimated market price of risk. As one might expect, they come out essentially halfway between the risk-neutral results and the results we have presented so far. If you thought those results looked "overfit", there is a simple way to cure that feeling. This calculation gives a sense of the sampling uncertainty in the final results, or what the result of a more principled lowering of the estimated risk premium based on small-sample biases or Bayesian shrinkage towards risk neutrality might produce.

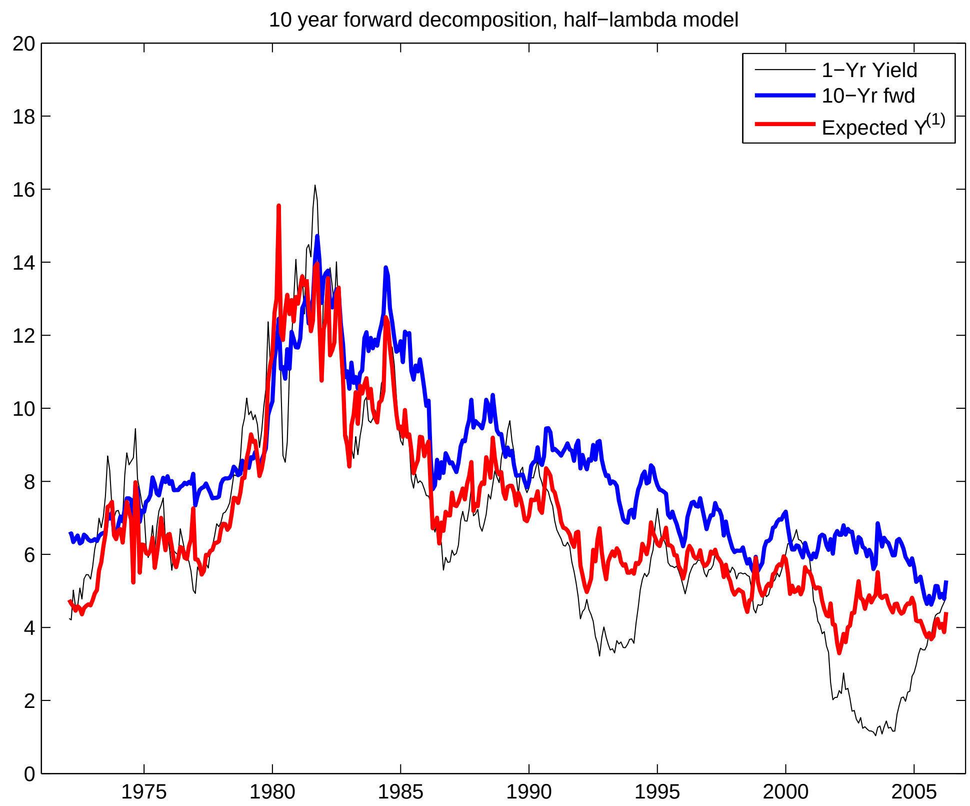

Figure 18: Current one-year rate $y_t^{(1)}$, 10 year forward rate $f^{t(10)}$, and expected one-year rate $E_t y_{t+9}^{(1)}$ computed from estimated affine model using half the estimated market price of risk λ_{0l}.

Figure 18 presents the 10 year forward rate and expected one year rate, mirroring Figure 13, and Figure 19 presents a detailed look at the 2000-2006 period, mirroring Figure 17.

In Figure 18 expected future interest rates move now essentially one for one with current interest rates from 1972 to 1975, rather than getting ahead of the rises and declines as they did with the full estimated risk premium. There is still nothing like quick reversion to an unconditional mean as seen in the OLS estimates. Through the three dips of 1987, 1993, and 2003, we now see expected one year rates stop their decline, rather than fall all the way down and past the one-year rate as they did with the full risk premium.

In Figure 19, the lower risk premium means relatively stable expected interest rates in December 2001 (top left) rather than the continued decline that this data point (with one of the largest forward premia in the whole sample) presented with the full risk premium. The other dates are similar to the previous findings, but with expected interest rates that are halfway to their corresponding forward curves, with about half the premium.

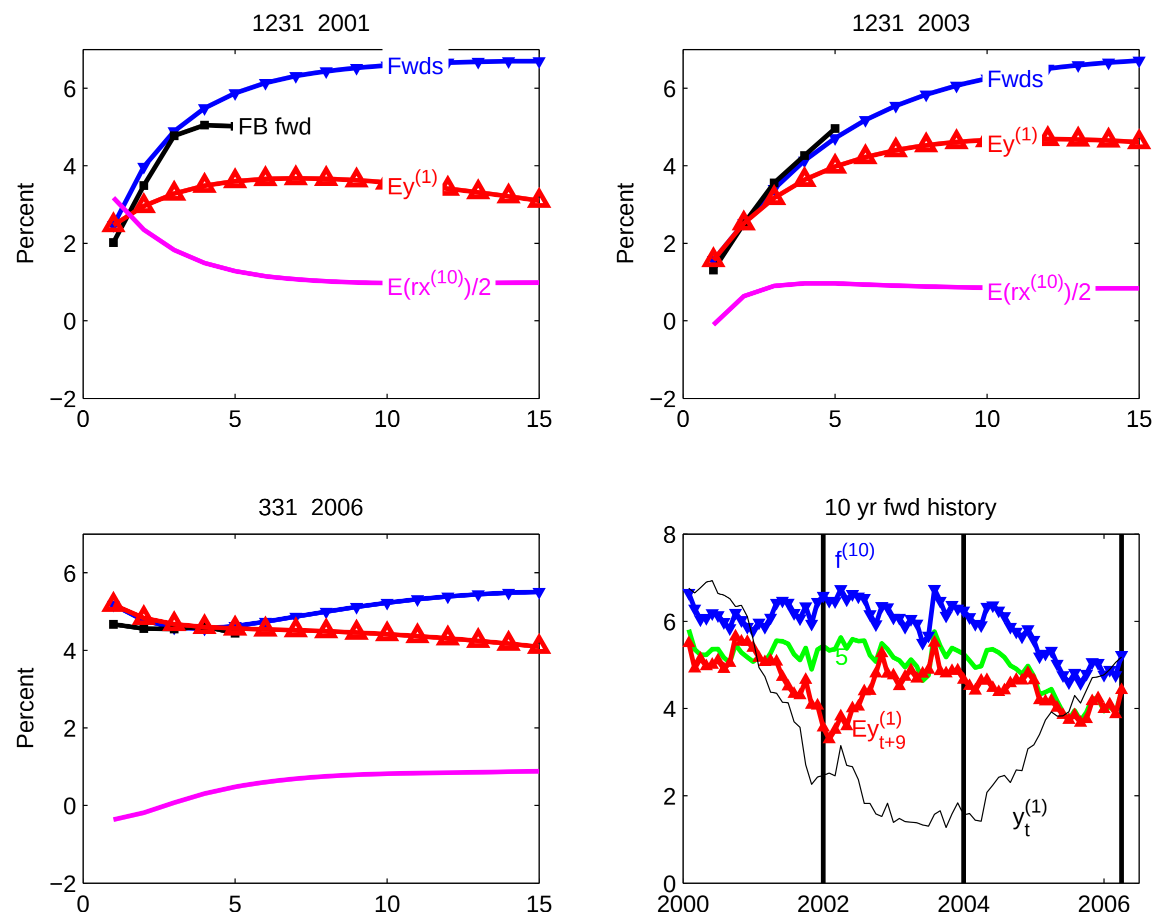

Figure 19: Recent history forward rate decomposition, using the affine model with half the estimated risk premium λ_{0l}.

6.2 Comparison with OLS dynamics

We have found the risk-neutral dynamics ϕ^* to fit forward rates, and then carefully worked back to the true dynamics ϕ by careful examination of market prices of risk. One could follow the opposite tack – estimate the true dynamics by simple OLS regression $X_{t+1} = \mu + \phi X_t + v_{t+1}$. (Figure 1 used raw data. Here, we investigate OLS estimates of factor dynamics after forming our factors, including the return-forecasting factor x.)

Table 6 presents the OLS - estimated dynamics, Table 7 presents standard errors, and Figure 20 presents the impulse-response functions, which we can compare with Figures 7 and 8. The broad pattern of the impulse-responses is gratifyingly similar. The major difference is the persistence of the level shock. Table 6 finds $\phi_{22} = 0.89$ and Figure 20 shows strong mean-reversion in the level shock, where we estimated $\phi_{22} = \phi_{22}^* = 0.98$ and an essentially permanent response above. OLS estimates of near-unit roots are notoriously biased downward, so the difference makes sense.

The OLS t statistics also document the substantial sampling uncertainty in a time-

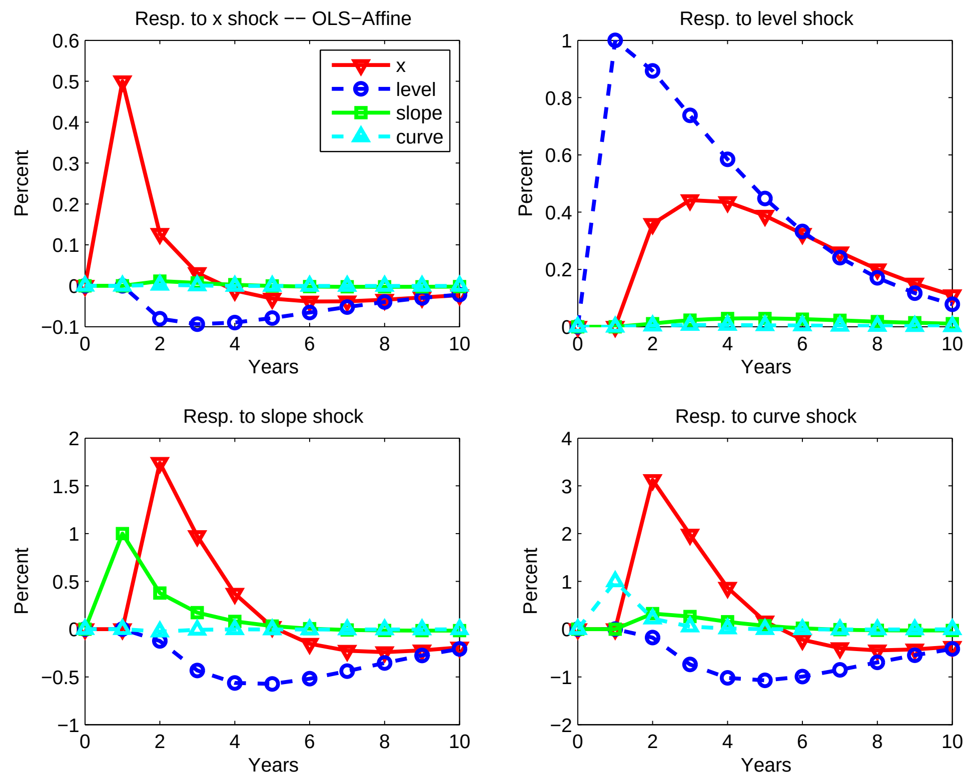

Figure 20: Impulse-response function for the OLS estimate of the transition matrix ϕ. The x response is divided by two to fit on the same scale.

series procedure. The coefficients of x_t on slope$_{t-1}$ and curvature$_{t-1}$, which are central to understanding long-term forward premia, have t stats near 2.

	100×μ	x	level	slope	curve	mse	R^2
	μ		ϕ				
x	-0.84	0.25	0.72	-3.48	6.25	24.1	0.20
level	-0.10	-0.08	0.89	0.12	-0.18	3.52	0.84
slope	-0.02	-0.01	-0.01	0.38	-0.33	1.47	0.21
curve	-0.02	0.00	0.00	0.22	0.22	0.57	0.10

Table 6. Estimates of model dynamics, μ and ϕ in $X_{t+1} = \mu + \phi X_t + v_{t+1}$. |eig| gives the eigenvalues of ϕ, in order. mse is mean-squared error.

	100×μ	x	level	slope	curve
	μ		ϕ		
x	-0.22	1.69	1.81	-2.03	1.73
level	-0.17	-4.41	14.7	0.47	-0.28
slope	-0.06	-1.49	-0.38	2.53	-1.63
curve	-0.25	0.86	0.41	0.70	2.20

Table 7. T- statistics. OLS t-statistics are computed with a Hansen-Hodrick correction for serial correlation due to overlap

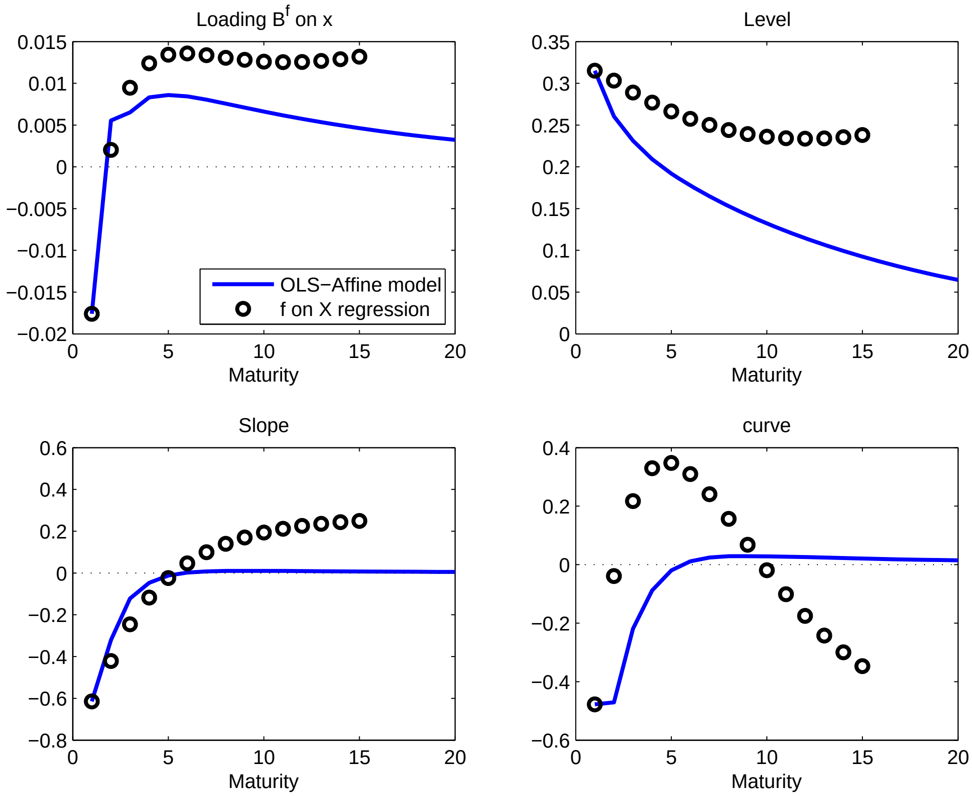

Figure 21: Loadings B^f of an affine model constructed from an OLS estimate of the transition matrix ϕ, using market price of risk estimates to construct $\phi^* = \phi - V\lambda_1$.

We can complete the affine model, taking the OLS estimate of the *true* dynamics, and using our estimates of market prices of risk to work back to the risk-neutral dynamics ϕ^*, rather than follow the opposite procedure as we have done above. Figure 21 presents the resulting affine-model loadings. Initially, this looks like a disaster compared to the beautiful fit of Figure 5. Table 3 shows that the fit is indeed much worse, with 1-2 percentage point errors rather than 10-20 basis point errors. However, this is the result of an estimate that places *no* effort whatsoever in fitting the cross-section of forward rates. Given that fact, it is in a sense remarkable that anything even vaguely reasonable comes out. Nonetheless, any model-fitting procedure that places weight on fitting the cross-section as well as the time series will obviously give up quickly on OLS dynamics to better fit the cross section, especially with the possibility of 15 bp measurement errors in sight.

The quicker mean-reversion means that the OLS dynamics give a fundamentally different picture of risk premia. Figure 22 presents the 10 year forward rates along with the OLS expectation of the corresponding future one-year yield. These VARs do incorporate the

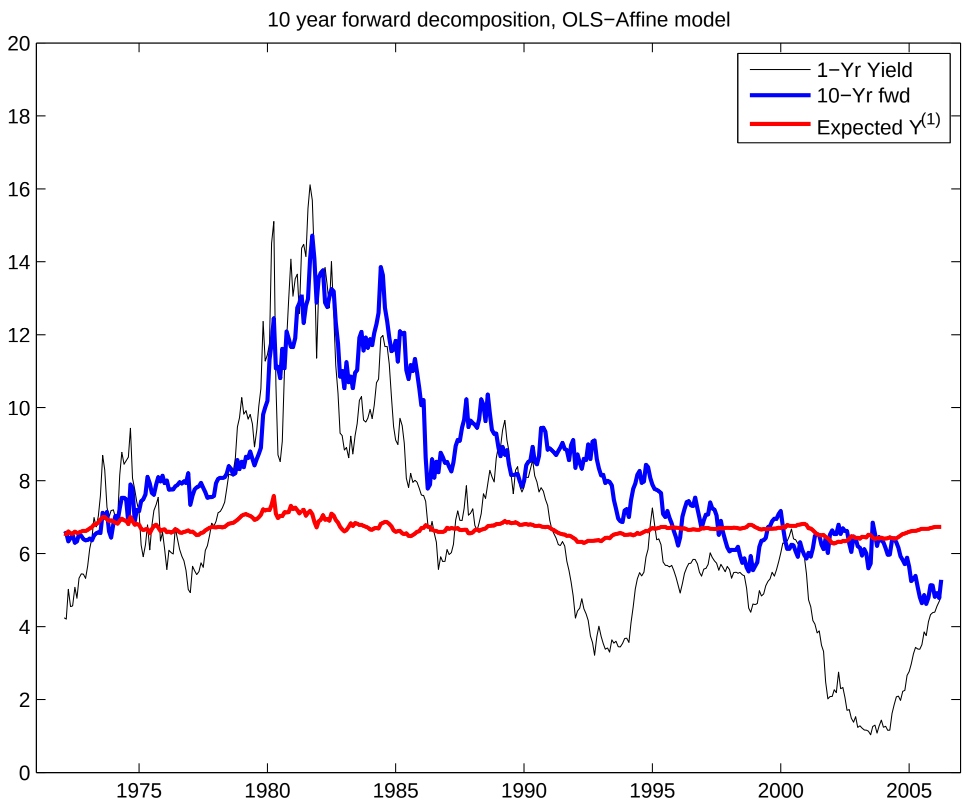

Figure 22: One year rate, 10 year forward rate $f_t^{(10)}$, and expected one-year rate $E_t y_{t+9}^{(1)}$, generated from an OLS estimate of factor dynamics $X_t = \mu + \phi X_{t-1} + v_t$.

return-forecasting factor, so there is a bit more variation in long-term expected interest rates. However, the pattern of quick reversion to the unconditional (sample!) mean is clearly visible.

We can see the decomposition for the recent period by looking at the right hand portion of Figure 22. Expected one-year rates quickly converge to the roughly 6% unconditional mean. Forward rates happen to be near 6% in the recent period, so we see little premium, but all the recent small decline in forward rates is attributed to a decline – to negative values – in that premium.

Aside from the appeal to sampling error downward bias of large roots, we can consider the plausibility of results. Contrast the 10 year forward decompositions in 22 and 13. The OLS estimates ask us to believe that forward rates throughout the high interest rate period of the 1980s corresponded to tremendous negative risk premia, since investors all knew interest rates would soon return to the average of about 6 percent. They ask us to believe that long-term expected interest rates are essentially unaffected by current experience, including both the effects of varying inflation in the 1970s and the business cycle dips of the last 20 years.

7 Conclusions and questions

We construct an affine model of the term structure of interest rates with risk premia, that allows us to decompose the yield curve at any moment in time into expected future interest rates and risk premia. We estimate the risk-neutral dynamics from the cross-section of forward rates. A detailed examination of market prices of risk leads us to an important restriction: expected returns vary over time only in response to a single "return-forecast" factor, and are only earned in compensation for covariance with level shocks. As a result, a single parameter controls the transformation from risk neutral to actual dynamics, and that parameter is measured by running a simple return-forecasting regression. Therefore, we are able to estimate actual dynamics by using all the information in the cross-section, and this single parameter.

We find plausible but unusual results. Long-term expectations of one-year rates track current one-year rates, and often get ahead of them, declining more than one-year rates when one year rates decline for example. They do not revert quickly to a mean, as a simple OLS estimate of interest rate dynamics would suggest, nor do they follow current forward rates as a risk-neutral or expectations hypothesis would suggest. They behave more closely to a cointegrated system. This is sensible, but it results in very large estimates of forward risk premia, given that forward rates do not move much over time. The key to the result is that risk-neutral dynamics require eigenvalues very near to one, to account for a level factor in yields or forward rates, and our transformation to real dynamics cannot do much to change the very persistent nature of shocks.

We connect forward premia – the difference between a 10 year forward rate and the expected one-year rate in nine years – to one-period expected return premia. A forward premium can result from a high *current* expected return, or an expectation that one-period returns will carry a premium in the future. There is an interesting and time-varying *term structure of risk premia*, which we characterize. We find that the one-period return forecasting factor x declines swiftly. However, orthogonal slope and curvature movements of the term structure, though unrelated to *current* expected returns, nonetheless forecast rises in long-term *future* expected returns, and thus contribute to forward rate risk premia.

We have focused on point estimates, with some loose discussion of sampling error. Obviously, the model needs a careful analysis of sampling error, induced from uncertainty about factor means and market prices of risk especially, and documentation of the idea that going from risk-neutral to real measure with constrained market prices of risk really does reduce sampling uncertainty.

The natural next step is to incorporate other information about long-term interest rate expectations. We can take that information from real bonds, such as the TIPS yield curve, inflation itself, or price or exchange rate data.

It is certainly plausible that the nature of the term structure of interest rates changes in different inflation environments. Long term nominal bonds have less volatile real returns than rolling over short rates when inflation is steady and real rates vary, and conversely when inflation is volatile and real rates vary less. Therefore, we expect a negative forward premium when inflation is steady, and a positive forward premium when inflation is more volatile. When inflation is incorporated, one should allow for changes over time in the nature of term premia.

One wants eventually a macroeconomic understanding of risk premia in the term structure. We have advanced (at least our own) understanding of risk premia by connecting them across maturity, and understanding how level, slope, and curvature movements forecast future risk premia. We understand that risk premia move on a single factor, and are earned as compensation for exposure to another single factor, level shifts. This is a start, and indicates that whatever macroeconomic events drive bond risk premia they must also correspond to level shocks, and will be forecastable by the same term-structure factors. But their identity remains an important question for economic investigation.

References

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

Boudoukh, Jacob, Matthew Richardson and Robert F. Whitelaw, 2006, "The Information in Long-Maturity Forward Rates: Implications for Exchange Rates and the Forward Premium Anomaly," Manuscript, Stern School of Business.

Gallmeyer, Michael F., Burton Hollifield, Francisco Palomino, and Stanely Zin, 2006, "Arbitrage-Free Bond Pricing with Dynamic Macroeconomic Models" Manuscript, Carnegie Mellon University.

Chinn, Menzie David, and Guy Meredith, 2005, "Testing Uncovered Interest Parity at Short and Long Horizons during the Post-Bretton Woods Era," NBER Working Paper No. 11077.

Cochrane, John H. and Monika Piazzesi, 2005, "Bond Risk Premia," *American Economic Review* 95, 138-160.

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.

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

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

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

Fama, Eugene F., 2006, "The Behavior of Interest Rates" *Review of Financial Studies*, 19, 359-379.

Gürkaynak, Refet S., Brian Sack and Jonathan H. Wright, 2006, "The U.S. Treasury Yield Curve: 1961 to the Present" Manuscript, Federal Reserve Board of Governors

Kim, Don H. and Athanasios Orphanides, 2005, "Term Structure Estimation with Survey Data on Interest Rate Forecasts." *Finance and Economics Discussion Series* 2005-48. Washington: Board of Governors of the Federal Reserve System.

Kim, Don H. and Jonathan H. Wright, 2005, "An Arbitrage-Free Three-Factor Term Structure Model and the Recent Behavior of Long-Term Yields and Distant-Horizon Forward Rates," *Finance and Economics Discussion Series* 2005-33. Washington: Board of Governors of the Federal Reserve System.

Kozicki, Sharon, and P.A. Tinsley (2001), "Shifting Endpoints in the Term Structure of Interest Rates," *Journal of Monetary Economics* 47, 613-652.

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

Piazzesi, Monika and Eric T. Swanson 2004. "Futures Prices as Risk-Adjusted Forecasts of Monetary Policy." NBER Working Paper 10547, University of Chicago.

Rudebusch, Glenn D., Eric T. Swanson, and Tao Wu, 2006, "The Bond Yield 'Conundrum' from a Macro-Finance Perspective" Federal Reserve Bank of San Francisco Working Paper 2006-16.

8 Appendix

8.1 Eigenvalue-based factor models

The eigenvalue decomposition gives a very quick way to form factor models. Given an $N \times 1$ vector of mean-zero random variables x, with $cov(x, x') = \Sigma$, we form $Q\Lambda Q' = cov(x)$ by the eigenvalue decomposition. If X is a $T \times N$ matrix of data on x, then [Q,L] = eig(cov(x)) in matlab. Λ is diagonal and Q is orthonormal, $QQ' = Q'Q = I$.

If we form "factors" by $z = Q'x$, we have $cov(z, z') = Q'Q\Lambda Q'Q = \Lambda$, i.e. the z are uncorrelated with each other. The columns of Q thus express how to construct factors from the data on x. We can then write $x = Qz$, and the columns of x also give "loadings" that describe how each x moves if one of the z moves.

If some of the diagonals Λ are zero, then we express all movements in x by reference to only a few underlying factors. For example, if only the first Λ is nonzero, then we can express $x = Q(:,1)z_1$ where $Q(:,1)$ denotes the first column of Q. In practice, we often find that many of the diagonals of Λ are very small, so setting them to zero and fitting x with only a few factors leads to an excellent approximation.

Since the factors are orthogonal we have $x = Q(:,1)z_1 + Q(:,2)z_t + ...$ Thus, the factor model amounts to a regression fit of x on the factors z. If some diagonals of Λ are small, the variance of the corresponding z are small, and this regression fits with an excellent R^2.

Factors constructed in this way solve in turn the question "what linear combinations of x has maximum variance, subject to the constraint that the sum of squared weights is one and each linear combination is orthogonal to the previous ones?" In equations, each column of Q satisfies

$$\max \left[var \left(Q_i' x \right) \right] \text{ s.t. } Q_i' Q_i = 1,\ Q_i' Q_j = 0,\ j < 1$$

We often compute the "fraction of variance of x explained by the first k factors," as the ratio of included to excluded eigenvalues. $\sum_{i=1}^{k} \Lambda_{ii} / \sum_{i=1}^{N} \Lambda_{ii}$. Since x is a vector, this means the fraction of the sum of the variances of each element. $\sum_i var(x_i) = Trace(Q'\Lambda Q') = \sum_{ii} \Lambda_{ii}$, so if we form a factor model by setting some Λ_{ii} to zero, the sum of the variances of the fitted forward rates is the sum of the retained eigenvalues.