Forecasting interest rates

Gregory R. Duffee
Johns Hopkins University

This version July 2012

Abstract

This chapter discusses what the asset-pricing literature concludes about the forecastability of interest rates. It outlines forecasting methodologies implied by this literature, including dynamic, no-arbitrage term structure models and their macro-finance extensions. It also reviews the empirical evidence concerning the predictability of future yields on Treasury bonds and future excess returns to holding these bonds. In particular, it critically evaluates theory and evidence that variables other than current bond yields are useful in forecasting.

JEL Code: G12
Key words: Term structure, affine models, predicting bond returns, predicting bond yields

Prepared for the *Handbook of Economic Forecasting*, Vol. 2, edited by Allan Timmermann and Graham Elliott, to be published by Elsevier. Some of the material in this chapter also appears in "Bond pricing and the macroeconomy," edited by George Constantinides, Milt Harris, and Rene Stulz, a chapter prepared for the *Handbook of the Economics of Finance*, also to be published by Elsevier. Thanks to two anonymous referees and many seminar participants for helpful comments and conversations. Voice 410-516-8828, email duffee@jhu.edu. Address correspondence to 440 Mergenthaler Hall, 3400 N. Charles St., Baltimore, MD 21218.

1 Introduction

How are interest rates on Treasury securities likely to change during the next month, quarter, and year? This question preoccupies financial market participants, who attempt to profit from their views. Policymakers also care. They attempt to predict future rates (and attempt to infer market participants' predictions) to help choose appropriate monetary and fiscal policies. More relevant for this chapter, academics use interest rate forecasts to help predict related variables, such as real rates, inflation, and macroeconomic activity. They also build term structure models that link interest rate forecasts to the dynamics of risk premia.

This chapter describes and evaluates an approach to forecasting that is grounded in finance. Interest rates are functions of asset prices, thus their dynamics can be studied using tools of asset pricing theory. The theory is particularly powerful when applied to Treasury yields, since the underlying assets have fixed payoffs (unlike, say, stocks). Nonetheless, there are known limitations to this approach, and important questions that remain open.

The most immediate implication of finance theory is that investors' beliefs about future prices are impounded into current prices. Therefore forecasts made at time t should be conditioned on the time-t term structure. An incontrovertible conclusion of the literature is that the time-t term structure contains substantial information about future changes in the slope and curvature of the term structure. It is less clear whether the term structure has information about future changes in the overall level of the term structure. This chapter argues that according to the bulk of the evidence, the level is close to a martingale.

Overfitting is always a concern in forecasting. Again, this chapter takes a finance-based approach to addressing this issue. Forecasts of future interest rates are also forecasts of future returns to holding bonds. Forecasts that imply substantial predictable variation in expected excess returns to bonds (i.e., returns less the risk-free return) may point to overfitting. For example, big swings in expected excess returns from one month to the next are hard to reconcile with risk-based explanations of expected excess returns.

Gaussian dynamic term structure models are the tool of choice to describe joint forecasts of future yields, future returns, and risk premia. These models impose no-arbitrage restrictions. Sharpe ratios implied by estimated models are helpful in detecting overfitting, and restrictions on the dynamics of risk premia are a natural way to address overfitting. One of the important open questions in the literature is how these restrictions should be imposed. The literature takes a variety of approaches that are largely data-driven rather than driven by economic models of attitudes towards risk.

Macroeconomic variables can be added to a dynamic term structure model to produce a macro-finance model. From the perspective of forecasting interest rates, this type of extension offers many opportunities. It allows time-t forecasts to be conditioned on information other than the time-t term structure. In addition, the dynamics of interest rates are tied to the dynamics of the macro variables, allowing survey data on their expected values to be used in estimation. Finally, macro-finance models allow restrictions on risk premia to be expressed in terms of fundamental variables such as economic activity and consumption growth.

Unfortunately, standard economic explanations of risk premia fail to explain the behavior of expected excess returns to bonds. In the data, mean excess returns to long-term Treasury bonds are positive. Yet traditional measures of risk exposure imply that Treasury bonds are not assets that demand a risk premium. Point estimates of their consumption betas are negative and point estimates of their CAPM betas are approximately zero. Moreover, although expected excess returns to bonds vary over time, these variations are unrelated to interest rate volatility or straightforward measures of economic growth. These facts are a major reason why applied models of bond risk premia shy away from approaches grounded in theory.

Recent empirical work concludes that some macro variables appear to contain substantial information about future excess returns that is not captured by the current term structure. Some, but not all, of this evidence is consistent with a special case of macro-finance models called hidden-factor models. The only original contribution of this chapter is to take a skeptical look at this evidence. Based on the analysis here, it is too soon to conclude that information other than the current term structure is helpful in forecasting future interest rates, excess returns, and risk premia.

2 Forecasting methods from a finance perspective

Figure 1 displays a panel of yields derived from prices of nominal Treasury bonds. The displayed yields are, for the most part, not yields on actual Treasury securities. The figure diplays zero-coupon bond yields. These yields are the objects of interest in most academic work. The Treasury Department issues both zero-coupon and coupon bonds. The former are Treasury bills, which have original maturities no greater than a year. The latter are Treasury notes and bonds. Academics typically use zero-coupon yields interpolated from yields on Treasury securities. (This chapter uses the terms “yield” and “interest rate” interchangeably.) The interpolation is inherently noisy. Bekaert, Hodrick, and Marshall (1997) estimate that the standard deviation of measurement error is in the range of seven to nine basis points of annualized yield for maturities of at least a year.

The yields in Figure 1 are yields on actual three-month Treasury bills, zero-coupon yields on hypothetical bonds with maturities from one to five years constructed by the Center for Research in Security Prices (CRSP), and the yield on a zero-coupon hypothetical ten-year bond constructed by staff at the Federal Reserve Board following the procedure of Gurkaynak, Sack, and Wright (2007). Yields are all continuously compounded. The CRSP data are month-end from June 1952 through December 2010. Until the 1970s, the maturity structure of securities issued by the Treasury did not allow for reliable inference of the ten-year zero-coupon yield. The first observation used here is January 1972.

A glance at the figure suggests that yields are cointegrated. More precisely, spreads between yields on bonds of different maturities are mean-reverting, but the overall level of yields is highly persistent. A robust conclusion of the literature is that Standard tests cannot reject the hypothesis of a unit root in any of these yields. From an economic perspective it is easier to assume that yields are stationary and highly persistent rather than truly nonstationary. Econometrically these alternatives are indistinguishable over available sample sizes.

By contrast, another robust conclusion of the literature is that spreads are stationary. For example, the handbook chapter of Martin, Hall, and Pagan (1996) shows there is a single cointegrating vector in Treasury yields. Not surprisingly, early academic attempts to model the dynamic behavior of bond yields used cointegration techniques. It is helpful to set up some accounting identities before discussing the logic and limitations of a cointegration approach to forecasting.

2.1 Notation and accounting identities

Consider a zero-coupon bond that matures at $t+n$ with a payoff of a dollar. Denote its time-t price and yield by

$$\begin{aligned} P_t^{(n)} &: \text{ Price} \\ p_t^{(n)} &: \text{ Log price} \\ y_t^{(n)} &: \text{ Continuously compounded yield, } y_t^{(n)} \equiv -\frac{1}{n} p_t^{(n)}. \end{aligned}$$

The superscript refers to the bond's remaining maturity. Denote the return to the bond from t to $t+1$, when its remaining maturity is $n-1$, by

$$\begin{aligned} R_{t,t+1}^{(n)} &: \text{ gross return to the bond from } t \text{ to } t+1,\ R_{t,t+1}^{(n)} \equiv P_{t+1}^{(n-1)}/P_t^{(n)} \\ r_{t,t+1}^{(n)} &: \text{ log return, } r_{t,t+1}^{(n)} \equiv \log R_{t,t+1}^{(n)}. \end{aligned}$$

The log return to the bond in excess of the log return to a one-period bond is denoted

$$xr_{t,t+1}^{(n)} : \text{log excess return}, \;\; xr_{t,t+1}^{(n)} \equiv r_{t,t+1}^{(n)} - y_t^{(1)}.$$

The yield on a bond can be related to future bond returns in two useful ways. The first links the bond's current yield to the bond's yield next period and the excess return to the bond. The relation is

$$y_t^{(n)} = y_t^{(1)} + \frac{n-1}{n}\left(y_{t+1}^{(n-1)} - y_t^{(1)}\right) + \frac{1}{n} xr_{t,t+1}^{(n)} \tag{1}$$

This is an accounting identity; the $t+1$ realizations on the right side must equal the time-t value on the left. For example, a higher yield at $t+1$ implies a lower price at $t+1$, and thus a lower realized excess return. The second accounting identity links the current yield to the sum, during the life of the bond, of one-period yields and excess returns:

$$y_t^{(n)} = \frac{1}{n}\sum_{j=0}^{n-1} y_{t+j}^{(1)} + \frac{1}{n}\sum_{j=0}^{(n-1)} xr_{t+j,t+j+1}^{(n-j)}. \tag{2}$$

In words, holding a bond's yield constant, a higher average short rate over the life of the bond corresponds to lower realized excess returns.

Conditional expectation versions of these identities are

$$y_t^{(n)} = y_t^{(1)} + \frac{n-1}{n}\left[E_t\left(y_{t+1}^{(n-1)}\right) - y_t^{(1)}\right] + \frac{1}{n}E_t\left(xr_{t,t+1}^{(n)}\right) \tag{3}$$

and

$$y_t^{(n)} = \frac{1}{n}E_t\left(\sum_{j=0}^{n-1} y_{t+j}^{(1)}\right) + \frac{1}{n}E_t\left(\sum_{j=0}^{(n-1)} xr_{t+j,t+j+1}^{(n-j)}\right). \tag{4}$$

These conditional expectations are also identities. They hold regardless of the information set used for conditioning, as long as the set contains the yield $y_t^{(n)}$. In particular, these equations hold for investors' information sets and econometricians' information sets, which may differ.

2.2 Cointegration

Campbell and Shiller (1987) motivate a cointegration approach to modeling the term structure. They make the simplifying assumption that the weak form of the expectations hypothesis holds. Then the conditional expectation (4) can be written as

$$y_t^{(n)} = \frac{1}{n} E_t \left[\sum_{j=0}^{n-1} y_{t+j}^{(1)} \right] + c^{(n)} \tag{5}$$

where $c^{(n)}$ is a maturity-dependent constant. The spread between the yield on an n-period bond and the one-period yield is then, after some manipulation,

$$S_{n,1} \equiv y_t^{(n)} - y_t^{(1)} = \sum_{j=1}^{n-1} (n-j) E_t \left[y_{t+j}^{(1)} - y_{t+j-1}^{(1)} \right] + c^{*(n)}. \tag{6}$$

Spreads are sums of expected first differences of one-period yields. Therefore spreads are $I(0)$ if one-period yields are $I(1)$.

Campbell and Shiller examine monthly observations of one-month and 20-year bond yields over the period 1959 to 1983. They cannot reject the hypotheses that yields are $I(1)$ and the spread is $I(0)$. Hence they advocate an error-correction model (ECM) to fit yield dynamics. For a vector of bond yields y_t and a linearly independent vector of spreads S_t, an ECM(p) representation is

$$\Delta y_t = \sum_{i=1}^{p-1} \Theta_i \Delta y_{t-i} + B S_t + \varepsilon_{t+1} \tag{7}$$

where Θ_i and B are matrices. The intuition is straightfoward and does not depend on the weak form of the expectations hypothesis. Investors impound their information about future short-term rates in the prices (and hence yields) of long-term bonds. If investors have information about future rates that is not captured in the history of short-term rates, then yield spreads will help forecast changes in short-term rates.

The first application of this type of model to forecasting interest rates is Hall, Andersen, and Granger (1992). Researchers continue to pursue advances in this general methodology. When many yields are included in the cross-section, the dimension of the estimated model is high. This can lead to overfitting and poor out-of-sample forecasts. Bowsher and Meeks (2008) use cubic splines to fit the cross-section. The knots of the spine are modeled with an ECM. Almeida, Simonsen, and Vicente (2012) take a similar approach, interpreting the splines as a way to model partially segmented markets across Treasury bonds.

The ECM approach is based on asset-pricing theory. However, the theory is not pushed to its sensible conclusion. Spreads help forecast future yields because investors put their information into prices. The ECM approach does not recognize that period-t bond prices (yields) are determined based on *all* information that investors at t have about future interest rates.

2.3 The term structure as a first-order Markov process

Asset prices incorporate all information available to investors. This fact leads to a more parsimonious approach to modeling interest rates than an ECM. We first look at the general statement of the result, then consider some special cases and caveats.

Assume that all of the information that determines investors' forecasts at t can be summarized by a p-dimensional state vector x_t. More precisely, x_t contains all information investors at t use to predict one-period bond yields and excess returns to multi-period bonds for all future periods $t+1, t+2$ and so on. Substitute this assumption into the identity (4) to produce

$$y_t^{(n)} = \frac{1}{n} \sum_{j=1}^{n-1} E\left(y_{t+j}^{(1)} | x_t\right) + \frac{1}{n} \left(\sum_{j=1}^{(n-1)} E\left(xr_{t+j-1,t+j}^{(n-j)} | x_t\right) \right). \quad (8)$$

It is trivial to see that the yield on the left side cannot be a function of anything other than the state vector, since only x_t shows up on the right side. Hence we can write

$$y_t^{(n)} = y(x_t; n).$$

This is simply a statement that investors determine the price of the n-period bond based on the predictions of short rates during the life of the bond and expected returns to the bond in excess of these short rates.

Stack time-t yields on bonds with different maturities in a vector y_t. These yields are a function of the state vector,

$$y_t = f\left(x_t; \mathcal{N}\right)$$

where the maturities of the bonds are in the vector $\mathcal{N}$. The key step in the derivation is to assume there exists an inverse function such that

$$x_t = f^{-1}\left(y_t; \mathcal{N}\right). \tag{9}$$

The inverse function exists if yields contain the same information as x_t. Therefore the rank of $\partial f / \partial x_t'$ must be p. A necessary condition is that there are at least p yields in the vector y_t. There are standard technical conditions associated with this result, but the intuition is straightforward. If each element of x_t has its own unique effect on the time-t yield curve, the yield curve can be inverted to infer x_t. Put differently, the time-t yield curve contains all information necessary to predict future values of x_t, and thus future yield curves. This allows us to write

$$E_t(y_{t+k}) = g\left(y_t; \mathcal{N}\right) \tag{10}$$

for some function $g(\cdot)$ that is determined by the mappings from factors to expected future one-period yields and excess bond returns.

A slightly simplistic interpretation of (10) is that both x_t and y_t must follow first-order Markov processes. Intuitively, x_t is Markov because it is defined as the set of information relevant to forming conditional expectations for all future horizons. If information at time t other than x_t were helpful in predicting future values of x_t, then investors could use that additional information to refine their conditional expectations of future yields. Similarly, everything investors know about expected future yields shows up in the period-t term structure. This intuition does not quite correspond to the statement that x_t and y_t are Markov because (10) says nothing about conditional higher moments. For example, the vector x_t need not contain all information relevant to forming conditional covariances among yields. Collin-Dufresne and Goldstein (2002) construct a general framework in which state variables that drive variations in conditional volatilities of yields do not affect the cross section of yields.

Some readers may find it useful to compare this statement with the expectations hypothesis. The weak form of the expectation hypothesis states that time-t forward rates differ from expected future yields by maturity-dependent constants. Forecasting with the weak form requires a parameterized model, but it is a trivial model: the only unknown parameters are the constants. When we step outside the weak form of the expectations hypothesis, we recognize that forward rates incorporate both expected future yields and time-varying expected excess returns. Equation (8) puts structure on expected excess returns by requiring that they depend on at most p state variables. Equation (9) implicitly says that by looking at the entire yield curve, we can disentangle shocks to expected excess returns from shocks to expected future yields. Disentangling these shocks requires estimation of the function $g(\cdot)$ in (10).

There is a potentially important hole in this derivation. It is possible that some cancellation may occur on the right side of (8). In other words, there may be a state variable that has equal and opposite effects on expected future short rates and expected future excess bond returns. This will reduce the dimension of the state vector that determines yields. "Hidden factor models," introduced by Duffee (2011a) and Joslin, Priebsch, and Singleton (2010), capture this idea. State variables that drive this kind of variation drop out of the left side, hence the period-t cross-section of yields does not contain all information relevant to forecasting future yields. We defer considering hidden factor models until Section 5.3.

2.4 Methodological implications

The previous argument has an important implication for the choice of forecasting methodology. When forming forecasts as of time-t, the use of information other than time-t yields requires a compelling justification. Such information includes yields dated prior to t, measures of inflation, central bank policy announcements, and economic activity. All of these variables are related to future yields. They may also be key components of the fundamental economic determinants of yields. But all of the information in these variables is already embedded in the time-t yield curve.

One way to justify using non-yield information is ease of estimation. To take an extreme example, we can form a forecast of next year's average three-month Treasury bond yield either by estimating a model or by reading a forecast off of survey responses of financial market participants. The latter approach is easier and probably more accurate. Unfortunately, samples of survey data for interest rate forecasts are not as long as those for forecasts of important macro variables such as inflation. Moreover, surveys typically ask questions about only one or two maturities on the yield curve. Section 5.1 discusses alternate uses of survey data.

A less extreme example is parameter stability over time. If there is a good reason to believe that the mapping from current yields to expected future yields is unstable over time, while the mapping from, say, current inflation to expected future yields is stable, then it makes sense to use inflation data to forecast yields. The literature contains considerable evidence for parameter instability in term structure models, but it is silent on the relative stability of forecasts using yields versus forecasts using other variables.

Another justification for the use of non-yield information is that we may believe investors at time t were using the wrong model to forecast yields. For example, they may have ignored a variable that we, as econometricians looking back in time, know is important. Yet another justification is the belief that the length of the state vector that drives expectations exceeds the number of available yields in the cross-section. Finally, measurement error can obscure the information in observed time-t yields. In practice, this is implausible because the measurement error in yields is very small relative to the measurement error in other relevant variables, such as observed inflation and output growth.

It is also important to recognize that a Markov process for yields does not imply that yields follow a first-order VAR. Markov processes may be nonlinear. For example, there may be occasional regime shifts in yield dynamics. When yields follow a Markov process, the time-t regime can be backed out of time-t yields. But extracting this information requires a regime-shifting model rather than a VAR(1).

Even if a VAR(1) is a reasonable model of yields, we must decide how to compress the information in the cross-section of yields. Yields on bonds of similar maturities are very highly correlated. Therefore a standard approach is to extract common factors from yields and apply a VAR to the factors. This procedure is the subject of the next section.

3 Regression approaches to forecasting Treasury yields

For forecasting purposes, an important difference between Treasury yields and macro variables such as GDP is the role of the cross-section. Since yields on bonds of different maturities are highly correlated, it does not make much sense to estimate unrelated forecasting regressions for each yield.[1] An almost universal approach to regression-based forecasting of Treasury yields is to first compress the cross-section into a low-dimensional vector, then use regressions to forecast elements of the vector. Forecasts of individual yields are determined by the cross-sectional mapping from the vector to the individual yield. We begin with the compression technique of principal components.

[1] However, the important early contribution to forecasting of Fama and Bliss (1987) adopts this method.

3.1 Principal components of yields

Following the spirit of Litterman and Scheinkman (1991), researchers often summarize term structures by a small set of linear combinations of yields.[2] An uncontroversial conclusion of the term structure literature is that the first few principal components of the covariance matrix of yields capture almost all of the variation in the term structure. Table 1 illustrates this for the sample 1972 through 2010. The table reports the unconditional standard deviations of the seven bond yields over this sample. It also reports the standard deviations of the residuals after taking out the first n principal components. Standard deviations of residuals from using three principal components range from five to eleven basis points of annualized yields, which is roughly the same range as the measurement error described by Bekaert, Hodrick, and Marshall (1997).

These first three principal components are commonly called "level," "slope," and "curvature" respectively. The motivation behind the labels is illustrated in Figure 2, which is constructed from the sample 1972 through 2010. A change in the first principal component corresponds to a roughly level change in yields of different maturities. A change in the second component moves short-maturity yields in the opposite direction of long-maturity yields. A change in the third component moves the short end and the long end together, away from intermediate maturities.

The figure includes loadings on the fourth and fifth components. From the scale of the figure, it is clear that level, slope, and curvature account for almost all of the variation in term structure shapes. Therefore forecasting future Treasury term structures is, for most purposes, equivalent to forecasting the first three principal components of yields.

Equation (10) requires that expectations of future yields are determined by current yields. We implement this in a regression setting by forecasting future principal components with current principal components. One method uses separate regressions for each future horizon k. Carriero, Kapetanios, and Marcellino (2012) follow this path. In this chapter a VAR(1)

[2] Litterman and Scheinkman use factor analysis rather than principal components.

is the primary tool for forecasting. Section 2.4 makes it clear that (10) does not imply that the yield curve follows a VAR(1) process. Nonetheless, it is the most obvious model with which to generate forecasts across various horizons that are consistent with each other. A VAR(1) applied to principal components allows us to use time-t principal components to predict time-$(t+k)$ principal components. The mapping from principal components to yields translates these forecasts to expected future yields.

We want to limit the number of principal components included in the VAR to avoid overfitting. The time-t term structure is almost entirely described by level, slope, and curvature, but it does not necessarily follow that a VAR(1) that includes only these first three principal components is an appropriate model.[3] Other principal components may have information about expected future values of level, slope, and curvature. Cochrane and Piazzesi (2005) point out that a factor does not need to have a large effect on the time-t term structure in order to play an important role in time-t expectations of future yields and expected returns. Roughly, the factor could have partially offsetting effects on the right side of (8). Therefore the appropriate number of principal components to include in the VAR is an empirical question.

3.2 Forecasting principal components

How much information is in the time-t term structure about future values of level, slope, and curvature? Stack the first five principal components of the term structure in the vector $\tilde{\mathcal{P}}_t$ and denote individual principal components by $\tilde{\mathcal{P}}_{i,t}$. A tilde represents observed data, which may be contaminated by measurement error or market imperfections. The forecasting regressions are

$$\tilde{\mathcal{P}}_{i,t+k} - \tilde{\mathcal{P}}_{i,t} = b_0 + b_1' \tilde{\mathcal{P}}_t + e_{i,t,t+k}$$

[3]Principal components are unconditionally orthogonal, but not necessarily conditionally orthogonal. Therefore a VAR(1) process for principal components will generally not equal a set of univariate AR(1) processes for the individual components.

Table 2 reports results for forecasts of the first three principal components. The regressions are estimated for two sample periods and one-month, three-month, and twelve-month horizons. Of course, the longer-horizon regressions are superfluous if we are confident that the VAR(1) is an accurate description of monthly yield dynamics. Only non-overlapping observations are used for the latter two horizons. Overlapping observations contain somewhat more information. They also can have large discrepancies between finite-sample and asymptotic statistical properties, especially in a near unit-root setting. The choice here is to give up some of the information in the sample in exchange for more reliable test statistics.

There are three main conclusions to draw from Table 2. First, there is no statistical evidence that changes in the first principal component (level) are forecastable. For both the 1952–2010 and 1970–2010 samples, tests of the hypothesis that all five coefficients of b_1 are zero cannot be rejected at the five percent level. (The covariance matrix of the parameter estimates is adjusted for generalized heteroskedasticity.) Only one p-value is less than ten percent, and that is for the annual-horizon regression in the shorter sample. This p-value should not be given much weight since there are only 38 observations for this regression.

Second, slope and curvature are unquestionably forecastable. The p-values are approximately zero for both sample periods and all forecasting horizons. At the quarterly horizon, about 20 (30) percent of the variation in slope (curvature) is predictable.

Third, factors other than the first three do not contribute much to forecasts of slope and curvature. For slope forecasts, tests of the hypothesis that the coefficients on the fourth and fifth principal components equal zero do not come close to rejection. The statistical evidence for curvature forecasts is mixed. In particular, these components appear to help predict quarter-ahead curvature. But recall from Figure 2 that curvature contributes relatively little to variations in the term structure over time. Moreover, the incremental R^2 of the fourth and fifth principal components to curvature forecasts is economically small; between three and five percentage points at the quarterly horizon. (These details are not reported in any table.)

These results, combined with the idea that parsimonious dynamic models are better than complicated models, suggest that a reasonable forecasting model is a VAR(1) for level, slope, and curvature, combined with restriction that that level is a random walk, unforecastable by any of the principal components. The mapping from these factors to observed yields is

$$\tilde{y}_t^{(n)} = \sum_{i=1}^{3} \beta_{i,n} \tilde{\mathcal{P}}_{i,t} + \epsilon_{n,t}. \tag{11}$$

The residual picks up cross-sectional variation in yields that is unexplained by the first three principal components. Duffee (2011b) takes this approach in forecasting future yields and concludes that the model works well in pseudo out-of-sample forecasting, relative to various other forecasting methods.

The conclusion that the level is unforecastable may appear to rely heavily on a misspecified model. There have been important shifts in the data-generating process of short rates during the past fifty years. The most prominent is the Federal Reserve's monetarist experiment from 1979 to 1983. Stationary regime-shifting processes can appear nonstationary from the perspective of a model that does not include regime shifts.

A comprehensive analysis of regime shifts is beyond the scope of this chapter. However, it is worth noting that the unpredictability of the level factor is robust to allowing for a simple kind of structural change. In results not detailed in any table, I estimate a VAR(1) on the first three principal components of yields for the sample 1952 through 2010. Chow-type dummies are included for the period 1979:1 through 2003:12. The results overwhelmingly reject the null hypothesis of constant coefficients over the full sample. In particular, during the monetarist period the level factor exhibits strong mean reversion, both economically and statistically. But outside of this monetarist period, the hypothesis that the level is unpredictable is not rejected at the ten percent level.

3.3 Related models

The previous subsection describes a simple model of the term structure. Three principal components follow a VAR(1), and yields are related to the principal components through the cross-sectional mapping (11). This model does not attempt to impose any no-arbitrage conditions, other than the Markov intuition of Section 2.3. Researchers have developed other models with VAR(1) factor dynamics and mappings from factors to yields that also ignore no-arbitrage considerations.

Perhaps the best-known of these models is the contribution of Diebold and Li (2006). They build a dynamic version of the term structure introduced by Nelson and Siegel (1987). The cross-section of the term structure is summarized by the level, slope, and curvature factors of the Nelson-Siegel model. These three factors are assumed to follow a VAR(1). Diebold and Li find that the dynamics with the best pseudo out-of-sample properties are those without any feedback among the factors. In other words, level, slope, and curvature follow univariate AR(1) processes. The forecast accuracy of their model is discussed in Section 3.5.

Other methods for compressing the information in yields can also be combined with a VAR(1). Perhaps the simplest approach is to choose a few bonds with maturities across the term structure—say, three months, two years, and ten years—and use the yields on these bonds as factors. This method ignores information in other bond yields. Other methods use information from the entire term structure. For example, the knot points of the cubic spline used by Bowsher and Meeks (2008) could be fit to a VAR(1) instead of the ECM model described in Section 2.2. Almeida and Vicente (2008) use Legendre polynomials to describe the cross-section of yields and fit them to a VAR(1).

The Markov logic of Section 2.4 implies that there is a high bar to including non-yield variables in forecasting regressions. Simply noting that expected future yields should depend on, say, inflation or macroeconomic activity is not sufficient. Since this information is already in contemporaneous yields, including both yields and macro variables in a regression is a recipe for overfitting, in both in-sample and pseudo out-of-sample forecasts.

Perhaps not surprisingly, this logic is routinely ignored. A popular forecasting approach uses a VAR that includes both compressed information from the term structure and compressed information from a large panel of macro variables. Recent examples are Koopman and van der Wel (2011) and Favero, Niu, and Sala (2012). Empirical evidence linking macro variables with term structure forecasts is discussed in Section 5. Much of the empirical work involving macro variables and the term structure focuses on forecasting future returns to bonds rather than future yields. The equivalence of these perspectives is discussed next.

3.4 Predictable excess returns

Forecasts of future yields using current yields (and perhaps other variables) are necessarily also forecasts of expected log returns to bonds. A derivation based on Campbell and Shiller (1991) makes this point starkly. The relevant accounting identity is

$$y_{t+1}^{(n-1)} - y_t^{(n)} = \frac{1}{n-1} \quad y_t^{(n)} - y_t^{(1)} \quad - \frac{1}{n-1} xr_{t,t+1}. \tag{12}$$

Campbell and Shiller (1991) use this equation as a test of the weak form of the expectations hypothesis by taking time-t expectations, then fixing the expected excess return to a constant. They test whether the OLS coefficient on the first term on the right side equals one, and conclusively reject the hypothesis.

The left side is the change in the yield on a bond from t to $t + 1$. For reasonably long maturities (say, a five-year bond), variations over time in the left side are very close to variations in the first principal component—the level—of the term structure. Therefore according to the results of Table 2, the left side is close to unforecastable with time-t yields. Since the left side equals the right side, the right side must also be unforecastable with time-t yields. The first term on the right side is a measure of the slope of the term structure, which varies widely over time; recall Figure 1. Since the sum on the right is unforecastable, the second term, excess returns to the bond, must also be strongly forecastable and positively correlated with the slope of the term structure.

This implication is confirmed with excess return regressions. Monthly simple returns to maturity-sorted portfolios of Treasury bonds are available from CRSP. Excess returns are constructed by subtracting the return to the shortest-maturity portfolio, which contains bonds with maturities less than six months. Excess returns for horizons longer than a month are constructed by compounding simple returns to the portfolios, then subtracting the compounded return to the shortest-maturity portfolio.

Excess returns from month t to $t+k$ are regressed on the first five principal components of the term structure at month t. Formally, the regressions have the form

$$\tilde{R}^e_{i,t,t+k} = b_0 + b_1' \tilde{\mathcal{P}}_t + e_{p,t,t+k}, \tag{13}$$

where the notation on the left side indicates the observed simple excess return to portfolio i from t to $t+k$. As in Table 2, the regressions are estimated for monthly, quarterly, and annual horizons, and only non-overlapping observations are used for the latter two horizons.

Point estimates for two portfolios are displayed in Table 3. There are two strong conclusions to draw from the results. First, the level of the term structure is unrelated to future excess returns. None of the t-statistics comes close to rejecting the hypothesis that the point estimate is zero. Second, a less-steep slope (larger value of the second principal component) corresponds to lower future excess returns. A one standard deviation increase in the second principal component corresponds to a decrease of about 0.25 percent in next month's excess return to bonds with maturities between five and ten years. The coefficients scale almost linearly in the return horizon, at least up to the annual horizon considered here.

Less clear are the links between other principal components and future excess returns. There is reasonably strong statistical evidence that greater curvature (an increase in short-maturity and long-maturity yields, a decrease in intermediate-maturity yields) predicts lower excess returns. The magnitude of the predictability is about half of that associated with the slope of the term structure. Three of the four t-statistics for the 1952–2010 sample allow rejection at the five percent level of the null hypothesis that the coefficient is zero.

Table 4 reports the R^2s of these and related forecasting regressions. At the monthly horizon, roughly five percent of the variation in excess returns is predictable. The figure rises to roughly ten percent at the quarterly horizon. The R^2s for annual horizons are intriguing. For the two sample periods emphasized in this paper, the R^2s are between 25 and 30 percent. However, Table 4 also includes results for the sample 1964 through 2003, which matches the sample period studied by Cochrane and Piazzesi (2005). They report R^2s up to 44 percent when annual excess returns are predicted by five forward rates. In Table 4 the R^2s for the 1964 through 2003 sample exceed 50 percent, both because the predictors are slightly different than those used by Cochrane and Piazzesi (2005) and because the regressions use non-overlapping observations. There is unquestionably overfitting in these regressions, since only 39 observations are explained by five variables. But the annual regression for 1972 through 2010 has fewer observations and yet a much smaller R^2. Section 7 discusses this result in detail. In this section it suffices to note that the high R^2s realized over the sample period of Cochrane and Piazzesi are not matched in the full sample.

3.5 Forecast accuracy evaluation

At this point it is appropriate to discuss how we should evaluate forecasting models. In this Handbook, Clark and McCracken (2012) emphasize the important role of pseudo out-of-sample forecasts in econometric evaluation of forecasting models. They discuss two reasons for this focus. First, forecasting models are often used to forecast (truly) out of sample, thus evaluation methodologies that mimic this procedure are sensible. Second, in-sample overfitting via data-mining is often detected with pseudo out-of-sample tests.

Both explanations apply to forecasts of bond yields, and pseudo out-of-sample forecast accuracy plays a major role in much of the relevant research. Two particularly influential contributions to the term structure literature recommend specific models based on their pseudo out-of-sample performance. Duffee (2002) and Diebold and Li (2006) conclude that their respective term structure models have pseudo out-of-sample forecasts of future yields that are more accurate than random walk forecasts. Their results are a major reason why much of the subsequent forecasting literature has explored variants of their models.

Unfortunately, neither conclusion is robust. Carriero (2011) finds the Diebold-Li model performs poorly out-of-sample for the period 1983 through 2009, a result driven by the forecast errors in the latter part of the sample. Duffee (2011b) finds that forecasts from both models are inferior to random walk forecasts when the data samples in the original papers are expanded to include more recent observations. In other words, the pseudo out-of-sample performance of the models differs substantially from the true out-of-sample performance. Only the passage of additional time will reveal whether the conclusions of Duffee (2002) and Diebold and Li (2006) are examples of pseudo out-of-sample data-mining or whether the models' recent performance is anomalous.

This chapter advocates a different tool for evaluation: the conditional Sharpe ratio. The main advantage of the conditional Sharpe ratio is that it is a tool for ex ante evaluation of a forecast. Unlike out-of-sample forecasting errors, Sharpe ratios do not depend on sampling error inherent in out-of-sample realizations.

The conditional Sharpe ratio for asset or portfolio i over the horizon from t to $t+k$ is

$$s_{i,t,t+k} = \frac{E_t \left(R^e_{i,t,t+k} \right)}{\sqrt{\text{Var}_t \left(R^e_{i,t,t+k} \right)}}. \tag{14}$$

Fitted conditional Sharpe ratios that are implausibly large are evidence of overfitting. For the forecasting regressions (13), the numerator of (14) is the fitted value of the regression. If the return innovations are homoskedastic, the denominator is the standard deviation of the regression's residuals.

Figure 3 displays fitted values of (14) for the portfolio of Treasury bonds with maturities between five and ten years. The fitted values are from monthly regressions for the sample 1952 through 2010. The sample mean conditional Sharpe ratio is only 0.06, or about 0.24 when multiplied by the square root of twelve to put it in annual terms. The fitted Sharpe ratios fluctuate substantially over time, reaching a maximum of 1.3 (4.7 in annual terms) and a minimum of -1.0 (-3.3 in annual terms), both during the Fed's monetarist experiment period. Since conditional volatilities were relatively high during this same period, the assumption of homoskedasticity exaggerates the fluctuations in conditional Sharpe ratios. But even outside of this period, the patterns in Figure 3 are hard to reconcile with our intuition about risk premia. For example, in the late 1960s and early 1970s, fitted conditional Sharpe ratios frequently flip back and forth from about one to minus one (in annual terms).

Although fitted Sharpe ratios can reveal overfitting, it is not easy to address the problem using finance theory at the level of accounting identities discussed in Section 2. If, say, a shrinkage principle is used to reduce the predictability of excess returns, the accounting identities (3) and (4) imply that forecasts of future yields will be affected in some way. But the effects could work through conditional variances of expected future yields, covariances among expected future yields, or covariances between expected future yields and current expected excess returns. There is insufficient structure to distinguish among these possibilities. For this additional structure, we turn to term structure models that impose no-arbitrage restrictions on yield dynamics.

4 A dynamic term structure framework

This section presents the workhorse no-arbitrage framework used for forecasting. It first appears in Ang and Piazzesi (2003). In contrast to the early term structure models that are expressed in a continuous-time setting, the setting here builds on the discrete-time model of Backus and Zin (1994). The key assumptions are that interest rate dynamics are linear and homoskedastic with Gaussian shocks. Before presenting the model, it is worth discussing why a model with such obviously counterfactual assumptions lays a central role in the literature.

It is easy to find evidence of nonlinear, non-Gaussian dynamics. For example, Figure 1 shows that conditional variances of yields vary substantially through time, reaching their peak during the monetarist experiment of 1979 through 1982. Gray (1996) concludes that a model of time-varying mean reversion and time-varying GARCH effects fits the dynamics of the short-term interest rate. The figure also shows that the zero bound imposes nonlinear dynamics on yields that bind in the last two years of the sample. Unfortunately, tractability must sometimes trump truth in dynamic term structure models. Dynamic term structure models describe the evolution of bond prices over time. This description requires (a) a method for computing bond prices that satisfy no-arbitrage; and (b) transition densities of prices from t to $t + \tau$. Researchers naturally restrict their attention to models for which (a) and (b) are computationally feasible.

The best-known class of models with tractable bond pricing is the affine class of Duffie and Kan (1996). This class includes both homoskedastic (Gaussian) and heteroskedastic models. Dai and Singleton (2000) and Duffee (2002) combine this affine class with linear dynamics of the underlying state vector to produce the "completely affine" and "essentially affine" classes respectively. One of the conclusions in Duffee (2002) is that only the Gaussian models in this class are sufficiently flexible to generate plausible forecasts of future yields. Gaussian no-arbitrage models are easy to understand and use, and can generate complicated yield dynamics. Although the models have fixed conditional variances, this is typically not a concern of the forecasting literature. This literature focuses on predicting yields and excess returns rather than constructing conditional second moments.

Recent research has attempted to find alternatives to the Gaussian class. Cheridito, Filipović, and Kimmel (2007) extend the essentially affine class to give non-Gaussian versions greater flexibility. A nonlinear tweak to the completely affine class is introduced by Duarte (2004). The quadratic class, which has nonlinear dynamics, is developed by Leippold and Wu (2002) and Ahn, Dittmar, and Gallant (2002). Dai, Singleton, and Yang (2007) and Ang, Bekaert, and Wei (2008) construct models with regime switches along certain dimensions. A fairly general nonlinear framework with affine pricing is developed by Le, Singleton, and Dai (2010).

These and other approaches show promise, but none has gained much traction in the applied literature. The limited existing evidence suggests that further exploration is warranted. For example, Almeida, Graveline, and Joslin (2011) find that a stochastic volatility model in the extended affine class generates more accurate in-sample forecasts of yields than a comparable Gaussian model. Almeida and Vicente (2008) draw favorable conclusions about the forecast accuracy of no-arbitrage models that describe the term structure with Legendre polynomials. But as of this writing, the workhorse framework is the Gaussian essentially affine setting.

4.1 The framework

The continuously-compounded nominal one-period interest rate is r_t. This can also be written as $y_t^{(1)}$, the yield on a one-period bond. The short rate is a function of a length-p state vector, denoted x_t. The function is

$$r_t = \delta_0 + \delta_1' x_t. \tag{15}$$

The state vector has first-order Markov dynamics

$$x_{t+1} = \mu + K x_t + \Sigma \; _{t+1}, \epsilon \quad _{t+1} \sim N(0, I). \tag{16}$$

Without loss of generality, Σ is lower triangular. The parameters of (16) are common knowledge. An implication of this assumption is that an econometrician cannot hope to uncover a trading strategy with expected returns that exceed the compensation required for bearing the strategy's risk. All expected excess returns must be explained by risk compensation

Using the notation of Section 2.1, no-arbitrage implies

$$P_t^{(n)} = E_t \quad M_{t+1} P_{t+1}^{(n-1)} \tag{17}$$

where M_{t+1} is the strictly positive pricing kernel, or stochastic discount factor (SDF). The SDF is assumed to have the log-linear form

$$\log M_{t+1} = -r_t - \Lambda_t' \ {}_{t+1} - \frac{1}{2}\Lambda_t'\Lambda_t. \tag{18}$$

The vector Λ_t is the compensation investors require to face the unit normal shocks ${}_{t+1}$. The price of risk is a function of the state vector, thus it varies over time according to

$$\Sigma\Lambda_t = \lambda_0 + \lambda_1 x_t. \tag{19}$$

The left side of (19) is the compensation investors require to face shocks to the state vector. Bonds are priced using the equivalent martingale dynamics

$$x_{t+1} = \mu^q + K^q x_t + \Sigma \ {}_{t+1}^{q}, \tag{20}$$

where the equivalent martingale parameters are

$$\mu^q = \mu - \lambda_0, \quad K^q = K - \lambda_1. \tag{21}$$

The discrete-time analogues of the restrictions in Duffie and Kan (1996) imply that zero-coupon bond prices can be written as

$$p_t^{(n)} = \mathbb{A}_n + \mathbb{B}_n' x_t. \tag{22}$$

The loading of the log price on the state vector is

$$\mathbb{B}_n' = -\delta_1' \left(I - K^q\right)^{-1} \left(I - \left(K^q\right)^n\right) \tag{23}$$

and the constant term satisfies the difference equation

$$\mathbb{A}_1 = -\delta_0, \quad \mathbb{A}_{n+1} = -\delta_0 + \mathbb{A}_n + \mathbb{B}_n'\mu^q + \frac{1}{2}\mathbb{B}_n'\Sigma\Sigma'\mathbb{B}_n. \tag{24}$$

Zero-coupon yields are written as

$$y_t^{(n)} = A_n + B_n' x_t, \tag{25}$$

$$A_n = -\frac{1}{n}\mathbb{A}_n, \qquad B_n = -\frac{1}{n}\mathbb{B}_n. \tag{26}$$

The log excess return to an n-period bond from t to $t+1$ is

$$xr_{t,t+1}^{(n)} = \mathbb{B}_{n-1}' \left(\lambda_0 + \lambda_1 x_t\right) - \frac{1}{2}\mathbb{B}_{n-1}'\Sigma\Sigma'\mathbb{B}_{n-1} + \mathbb{B}_{n-1}'\Sigma\ _{t+1}. \tag{27}$$

To understand this equation, begin with the final term on the right side, which is the return shock. The state-vector shock is $\Sigma\ _{t+1}$. An n-period bond at t is an $(n-1)$-period bond at $t+1$, and the sensitivity of its log price to the state vector is $\mathbb{B}_{n-1}$. Investors require compensation to face innovations in the state vector. The risk compensation for holding the bond is the first term on the right, which is the exposure to the state-vector shock times the conditional compensation for the shock. The middle term on the right is a Jensen's inequality adjustment for log returns.

4.2 Empirical estimation

Assume that we observe a panel of zero-coupon Treasury bonds. There are T time series observations of $d \geq p+1$ yields, stacked in vectors $\tilde{y}_t, t = 1, T$. Recall the tilde denotes observed yields, which here must be distinguished from model-implied yields. The maturities of the bonds are $n_1, \ldots, n_d$, which are fixed across time. There must be at least $p+1$ yields to identify the model's parameters. Using the language of the Kalman filter, the measurement equation is

$$\tilde{y}_t = \mathbf{A} + \mathbf{B}x_t + \eta_t, \qquad \eta_t \sim N(0, \Omega). \tag{28}$$

The mapping from the state vector to yields satisfies no-arbitrage, or

$$\mathbf{A} = \begin{pmatrix} A_{n_1} & \ldots & A_{n_d} \end{pmatrix}',$$

$$\mathbf{B} = \begin{pmatrix} B_{n_1} & \ldots & B_{n_d} \end{pmatrix}',$$

where A_n and B_n are defined by (26). The error term in (28) accounts for deviations from the exact no-arbitrage prices. The interpretation of this term that is consistent with the no-arbitrage formulas is measurement error. In reality, it also picks up model misspecification, such as too few common factors, and components of yields related to market imperfections. Another way to write the measurement equation is

$$\tilde{y}_t = y_t + \eta_t, \tag{29}$$

$$y_t = \mathbf{A} + \mathbf{B}x_t. \tag{30}$$

The notation y_t, without the tilde, indicates yields uncontaminated by measurement error (or model misspecification and market imperfections).

The model can be estimated with maximum likelihood (ML) using the Kalman filter. Because the state vector is latent (i.e., not tied to anything observable), the model is unidentified unless identifying restrictions are imposed. Examples of identifying restrictions are discussed next.

4.3 What are the factors?

What determines the dynamic behavior of the term structure? Although this is an important question, it is vacuous in the context of the simple factor model here. The question is effectively asking how we should interpret elements of the state vector. But the state vector is arbitrary. An observationally equivalent model is produced by scaling, rotating, and translating the state vector. Associated with each rotation is a different set of parameters of the transition equation (16) and the measurement equation (28).

Define such a transformation as

$$x_t^* = \underbrace{\Gamma_0}_{p\times 1} + \underbrace{\Gamma_1}_{p\times p} x_t \tag{31}$$

where Γ_1 is nonsingular. Following Dai and Singleton (2000), an observationally equivalent model replaces x_t with x_t^*, and replaces the parameters of (16) with

$$K^* = \Gamma_1 K \Gamma_1^{-1}, \quad \mu^* = \Gamma_1 \mu + (I - K^*)\Gamma_0, \quad \Sigma^* = \text{chol}\left(\Gamma_1 \Sigma \Sigma' \Gamma_1'\right),$$

where chol($\cdot$) indicates a Cholesky decomposition. Similarly, the parameters of the equivalent-martingale dynamics (20) are replaced with

$$K^{q*} = \Gamma_1 K^q \Gamma_1^{-1}, \quad \mu^{q*} = \Gamma_1 \mu^q + (I + K^{q*})\,\Gamma_0.$$

There are many ways to identify the state vector and thus identify the model's parameters. One way is to restrict the K matrix to a diagonal matrix, set μ to zero, and set the diagonal elements of Σ equal to one. These define a rotation, translation, and scaling respectively.[4] Another approach equates the state vector with linear combinations of yields.

Consider any $p \times d$ matrix $\mathcal{L}$ with rank p. (Recall $p < d$.) Premultiply (30) by $\mathcal{L}$ to

[4]This is slightly stronger than identification, since it imposes the restriction that the eigenvalues of K are real.

produce

$$\mathcal{L}y_t = \mathcal{L}\mathbf{A} + \mathcal{L}\mathbf{B}x_t. \tag{32}$$

As long as $\mathcal{L}\mathbf{B}$ is invertible, the left side of equation (32) defines a new state vector. A simple example is a diagonal $\mathcal{L}$ with p diagonal elements equal to one and the remainder equal to zero. This choice produces a state vector equal to p "true" yields, or yields uncontaminated by measurement error. Normalization restrictions for arbitrary $\mathcal{L}$ are described in Joslin, Singleton, and Zhu (2011).

Another useful choice of $\mathcal{L}$ is based on principal components of yields. For a given data sample, the first p principal components can be written as

$$\tilde{\mathcal{P}}_t = \mathcal{L}\tilde{y}_t,$$

where $\mathcal{L}$ is a $p \times d$ matrix of loadings of the sample principal components on the observed yields. Figure 2 is a graphical illustration of this matrix for one sample. Then the factors correspond to "level," "slope," and so on. Regardless of the choice of $\mathcal{L}$, this kind of state vector rotation emphasizes that the term structure model is really a model of one set of derivative instruments (yields) explaining another set of derivative instruments (more yields).

4.4 No-arbitrage restrictions and Sharpe ratios

No-arbitrage restrictions ensure that bonds are priced consistently with each other: equal risk receives equal compensation. The restrictions say nothing about the appropriate magnitude of the compensation, or how it should vary over time. Additional restrictions can be imposed on a no-arbitrage model to put more structure on risk premia. With this Gaussian model, the polar cases are unrestricted risk premia—the vector λ_0 and the matrix λ_1 in (19) are free parameters—and risk-neutrality, which sets both to zero.

It might seem natural to estimate the model with unrestricted risk premia, but Joslin, Singleton, and Zhu (2011) and Duffee (2011b) show that this version of the no-arbitrage model is of no value in forecasting. There are two pieces to the argument, and both are worth describing in some detail.

Joslin, Singleton, and Zhu (2011) conclude that no-arbitrage restrictions, in the absence of restrictions on risk premia, have no bite when estimating conditional expectations of future values of the state vector. These conditional expectations are determined by the parameters μ and K of the true, or physical measure, dynamics (16). No-arbitrage restrictions boil down to the existence of equivalent martingale dynamics, given by (20). When risk premia are unrestricted, the parameters μ^q and K^q of the equivalent martingale dynamics are unrelated to their physical-measure counterparts. Although the two measures share volatility parameters, in a Gaussian setting these parameters do not affect ML estimates of μ and K. If the state vector is rotated to, say, the first p principal components of yields, then the ML estimates of μ and K are essentially indistinguishable from VAR estimates.

No-arbitrage restrictions *do* affect the mapping from the state vector to yields; in other words, they bite. Absent the no-arbitrage restrictions, the parameters $\mathbf{A}$ and $\mathbf{B}$ of the measurement equation (28) are restricted only by normalization requirements. But Duffee (2011b) points out that these parameters can be estimated with extremely high precision even if no-arbitrage restrictions are not imposed, since the measurement equation amounts to a regression of yields on other yields. The R^2s of such regressions are very close to one. In practice, imposing the restrictions even when they are true does not improve estimates of the measurement equation.

In a nutshell, the Gaussian no-arbitrage model, absent additional restrictions on risk premia, offers no advantages over a simple regression-based approach. First estimate a VAR on the first p principal components, then estimate cross-sectional regressions of yields on the components. Nonetheless, the no-arbitrage model is valuable because the estimated properties of risk premia, and hence fitted conditional Sharpe ratios, can be used to evaluate the reasonableness of the regression-based approach.

This analysis of Sharpe ratios goes beyond the fitted Sharpe ratios (14) calculated with forecasting regressions. The forecasting regressions produce Sharpe ratios for portfolios of Treasury bonds. This contrasts sharply with a no-arbitrage model, which can be used to determine Sharpe ratios for any dynamic trading strategy in fixed income, including all strategies that attain the maximum Sharpe ratio.

With the log-normal SDF, the maximum conditional Sharpe ratio for simple returns is

$$s_t^{max} = \sqrt{e^{\Lambda_t' \Lambda_t} - 1}. \tag{33}$$

Duffee (2010) explains why bond portfolios do not attain this maximum Sharpe ratio. The intuition is that an asset with the maximum Sharpe ratio has a return that is perfectly negatively correlated with the SDF. Log returns to bonds can be perfectly negatively correlated with the log of the SDF. If the SDF and bond returns both have low volatility, this is almost the same as a perfect negative correlation in levels. But with high volatilities, there can be a substantial wedge between Sharpe ratios of bonds and the maximum Sharpe ratio implied by a Gaussian term structure model. The higher the Sharpe ratios for bond portfolios, the more volatile is the SDF, and therefore the larger the wedge there is between Sharpe ratios for bond portfolios and the maximum Sharpe ratio.

Fitted maximum Sharpe ratios for high-dimensional Gaussian models can be unrealistically high. Using a data sample of Treasury yields from 1971 to 2008, Duffee (2010) estimates Gaussian no-arbitrage models with two, three, four, and five factors. No restrictions are placed on risk premia. He calculates a monthly mean of (33) for each model. Annual-horizon maximum Sharpe ratios are about 2.7 for the four-factor model, a number that seems quite large until it is compared to the five-factor model figure of about 10^{30}. Sharpe ratios for two-factor and three-factor models are reasonable.

This Sharpe ratio evidence points to overfitting with a four-factor model and severe overfitting with a five-factor model. One simple way to solve the overfitting problem is to simply build forecasting models with only three factors, sacrificing the ability to capture more obscure information in the term structure. Another approach is to impose restrictions on risk premia dynamics.

4.5 Restrictions on risk premia dynamics

Restrictions on risk premia increase the precision of estimates of physical dynamics (16). The reason is that equivalent-martingale dynamics (20) are estimated with high precision, and risk premia restrictions tighten the relation between physical and equivalent-martingale dynamics. Thus imposing correct restrictions on risk premia should improve forecast accuracy.

Risk premia are determined by the p-vector λ_0 and the $p \times p$ matrix λ_1. The compensation investors receive for facing a unit shock to element i of the state vector is an affine function of the state vector. The constant term of this risk compensation is element i of the vector λ_0, while the sensitivity of risk compensation to the state vector is row i of λ_1.

There are many ways to impose restrictions on the elements of λ_0 and λ_1. One is to set to zero any parameters that are statistically indistinguishable from zero. A more formal frequentist approach is adopted by Christensen, Lopez, and Rudebusch (2010) and Joslin, Priebsch, and Singleton (2010), who use various information criteria. Bauer (2011) adopts a Bayesian approach to estimating these parameters. Another is to apply a shrinkage principle, setting the parameters to a weighted average of their unrestricted estimates and their risk-neutral value of zero. A simple example of this is in Cochrane and Piazzesi (2008). Duffee (2010) takes yet another approach, estimating parameters subject to a restriction on maximum Sharpe ratios.

Another common approach imposes *a priori* restrictions on the elements of λ_0 and λ_1 based on a combination of tractability and beliefs about risk premia. For example, both Cochrane and Piazzesi (2008) and Duffee (2011a) assume that only shocks to the level of the term structure have time-varying risk compensation. For Duffee's chosen factor rotation, this restriction implies that only one row of λ_1 is nonzero. Cochrane and Piazzesi go further and choose a rotation for which a single factor drives time-variation in a single price of risk. A qualitatively similar approach is taken by Ang, Bekaert, and Wei (2008), who choose a factor rotation such that risk premia depend only on one of the factors. A weaker assumption that does not rely on a specific factor rotation is that the rank of λ_1 is less than p. Joslin, Singleton, and Zhu (2011) describe how to impose this restriction in ML estimation.

An indirect method of imposing restrictions on risk premia dynamics is to impose restrictions on the physical dynamics (16). One example is Christensen, Diebold, and Rudebusch (2010), which is a dynamic model of the three factors of Nelson and Siegel (1987).[5] The version of their model with the most accurate forecasts imposes the restrictions that the factors all follow univariate autoregressive processes. These restrictions implicitly restrict risk premia dynamics.

Noticeably absent from this discussion of risk premia is any mention of fundamental determinants of risk compensation. The framework of Section 4 abstracts from consumption, inflation, output growth, and all other macroeconomic variables. With a yields-only model, it is difficult to impose restrictions that motivated by workhorse asset-pricing models of risk premia. The model must be extended to include macroeconomic variables.

5 Macro-finance models

The primary motivation behind macro-finance models is to ground term structure dynamics in the dynamics of variables, such as inflation, that are viewed as fundamental determinants of yields. A secondary motivation is to improve forecasts of yields out-of-sample by reducing the overfitting problem.

Macro-finance models typically follow Ang and Piazzesi (2003) by expanding the measurement equation of the yields-only framework. Assume we observe some variables at time t, stacked in a vector $\tilde{f}_t$. In principle, this vector may contain any variable other than the price

[5] Christensen et al. (2010) slightly modify the Nelson-Siegel factors to satisfy no-arbitrage. Their model does not fit into the framework of Section 4 because it is written in continuous time. The version in Diebold and Li (2006) is in discrete time, but does not satisfy no-arbitrage.

of a fixed-income derivative. Usually it contains macro variables, but it may also contain survey data.

The macro-finance measurement equation is

$$\begin{pmatrix} \tilde{y}_t \\ \tilde{f}_t \end{pmatrix} = \begin{pmatrix} \mathbf{A}_y \\ \mathbf{A}_f \end{pmatrix} + \begin{pmatrix} \mathbf{B}_y \\ \mathbf{B}_f \end{pmatrix} x_t + \eta_t. \tag{34}$$

No-arbitrage restrictions apply to the vector $\mathbf{A}_y$ and the matrix $\mathbf{B}_y$. No such restrictions apply to the vector $\mathbf{A}_f$ and the matrix $\mathbf{B}_f$, although other restrictions may be imposed.

The key assumption built into (34) is that the same state vector that determine the cross-section of yields also determine the additional variables. This is a practical rather than a theoretical restriction. In theory, even if the term structure and the additional variables had nothing to do with each other, the state vector could be expanded to include the necessary state variables. But in practice, researchers use state vectors with three, four, or five factors.

5.1 Survey data

One of the most promising applications of the macro-finance framework has nothing to do with uncovering the fundamental determinants of yields or risk premia. Instead, it uses information from surveys to improve estimates of the model's dynamics. Investors' expectations of future interest rates are measured directly with surveys. These survey-based expectations can be included in $\tilde{f}_t$. The model's parameters can then be estimated while imposing the restriction that the survey expectations equal the model's expectations plus noise, as in Kim and Orphanides (2005).

An extension of this approach is Chernov and Mueller (2011). They include both observed inflation and survey expectations of future inflation to a yields-only model. Including inflation forces yields to share dynamics with inflation, and including inflation expectations helps to pin down those shared dynamics. Their use of survey data is consistent with the results of Ang, Bekaert, and Wei (2007), who find that surveys produce more accurate forecasts of inflation than models.

5.2 The factors of macro-finance models

An important conceptual difficulty with macro-finance models is that the macro variables of interest, such as inflation, aggregate consumption, and output growth, are not spanned by the term structure. Recall from (32) that the length-p state vector can be rotated into p linearly independent combinations of "true" yields. The only requirement is that the mapping from the original state vector to the linear combinations is invertible. This result implies that, aside from cross-sectional measurement error, regressions of the macro factors on p linear combinations of contemporaneous yields should produce R^2s of one.

We do not observe variables in the absence of the cross-sectional error. The equivalent regressions using observed variables tell us about the magnitude of this error. For typical variables included in macro-finance models, the R^2s are on the wrong side of 1/2. To illustrate the problem, consider quarterly observations of Treasury yields, inflation, per capita consumption growth, and industrial production growth. Quarter-t inflation is measured by the log change in the CPI from the final month of quarter $t-1$ to the final month of quarter t. Per capita consumption growth is measured by the log change in real consumption per person on nondurables and services from quarter $t-1$ to quarter t. Consumption data are constructed following Piazzesi and Schneider (2007). Bond yields are for maturities of three months and one through five years. The sample period is 1952Q2 through 2010Q4.

Table 5 reports R^2s of cross-sectional regressions of the macro variables on the first five principal components of the Treasury yields. The implicit macro-finance model is one with five factors. The R^2s range from 3 percent (industrial production) to 45 percent (inflation). As mentioned above, one interpretation of these low R^2s is measurement error. We know from Section 3.1 that the measurement error of yields is tiny relative to the variability of yields over this sample. Thus if the five-factor model is correctly specified, the observed macro variables are primarily noise.

This seems implausible. But there is more direct evidence that the problem is misspecification. If yields are measured perfectly and the macro variables are measured with noise, then p linear combinations of yields contain all information relevant to forecasting future values of the macro variables. Lags of the macro variables should have no additional explanatory power. To test this implication, one-quarter-ahead values of inflation, consumption growth, and industrial production growth are first predicted with five principal components of yields, then with just the lagged value of the macro variable. Finally, all six explanatory variables are included.

Table 5 reports R^2s for the first two regressions and reports t-statistics on the lagged macro variable for the third regression. The hypothesis that the term structure contains all information relevant for forecasting future values is overwhelmingly rejected for all three macro variables. In fact, the term structure has less information about future inflation and industrial production growth than is contained in the single lag of the forecasted variable.

At first glance, a sensible interpretation of this evidence is that the macro variables all contain some component that is orthogonal to term structure dynamics. But there is another interpretation, suggested by regressions that forecast excess bond returns. Following the logic that the state vector can be rotated into linear combinations of yields, all information about future excess bond returns should be contained in current yields.

Recent evidence undercuts this implication. Cooper and Priestly (2009) finds that the output gap forecasts excess bond returns even when information from the term structure is included in the forecasting regression. Ludvigson and Ng (2009, 2010), construct principal components of more than 130 macroeconomic and financial time series. They use the first eight principal components to predict excess bond returns. Cieslak and Povala (2011) construct a weighted average of past inflation, where the weights decline very slowly. Empirically, both the principal components and the measure of past inflation contain substantial information about future excess returns that is not contained in the term structure.

We take a close look at some of this evidence in Section 7. In the next subsection we take the evidence at face value and consider its implications.

5.3 Hidden factors

Duffee (2011a) and Joslin, Priebsch, and Singleton (2010) develop a restricted form of Gaussian no-arbitrage models called hidden factor models. A key feature of a hidden factor model is that no linear combinations of yields can serve as the model's state vector. Mathematically, the matrix $\mathcal{L}\mathbf{B}$ in (32) is not invertible for any choice of $\mathcal{L}$. Somewhat less formally, the term structure of bond yields is not a first-order Markov process. Investors have information about future yields and future excess returns that is not impounded into current yields.

The intuition behind a hidden factor model is easily expressed using equation (8). If there is some state variable (or linear combination of state variables) that has equal and opposite effects on time-t expectations of future short-term rates and time-t expectations of future excess returns, that state variable (or linear combination) will not affect the time-t term structure. Put differently, the state vector that determines yields is smaller than the state vector that determines expected future yields. Duffee (2011a) shows that a single parameter restriction on the Gaussian no-arbitrage model produces a hidden factor.

Hidden factors can, in principle, explain all of the results in Table 5, as long as the factor(s) that are hidden from the term structure are revealed in macroeconomic data. For example, imagine that economic growth suddenly stalls. Investors anticipate that short-term rates will decline in future months. But investors' risk premia also rise because of the downturn. These effects move long-term bond yields in opposite directions. If they happen to equal each other in magnitude, the current term structure is unaffected. Nothing in the term structure predicts what happens next to either economic growth or bond yields. However, the lower economic growth will predict higher future excess bond returns.

This kind of story motivates the empirical approach of Joslin, Priebsch, and Singleton (2010), who build a model with two hidden factors that are revealed in economic activity and inflation. Duffee (2011a) takes a different approach, using only yields and filtering to infer the presence of a hidden factor. Both conclude hidden factors are important, as do Chernov and Mueller (2011), who use survey expectations of inflation to reveal such a factor. This preliminary evidence indicates that hidden factor models should be taken seriously.

However, this evidence can also be used as a poor justification for data-mining. The spanning requirement of models without hidden factors reduces significantly the ability of researchers to fish for variables that forecast excess returns. Hidden factors remove this constraint. If a researcher finds a variable that has statistially significant forecast power when information from the term structure is included in the forecasting regression, they can simply say that yet another hidden factor is uncovered.

Even with hidden factor models, data-mining would not be an important problem if we understood the economic fundamentals behind risk premia in the bond market. Armed with this knowledge, we could require that the only variables used to forecast excess returns be those linked to the fundamental determinants of risk premia. Unfortunately, the state of the art in term structure research tells us little about these fundamental determinants.

6 Economic fundamentals and risk premia

The workhorse models used in finance to explain risk premia are extensions of the classic case of a representative agent with power utility and an aggregate consumption endowment. An asset's conditional expected excess return depends on the return's conditional covariance with consumption growth (power utility and habit formation as in Campbell and Cochrane (1999)) and with the return to total wealth (recursive utility as in Epstein and Zin (1989)). To help judge whether these models can explain the dynamics of bond risk premia, this section presents two types of empirical evidence. The first subsection looks at consumption and CAPM betas. The second considers whether time-varying expected excess returns are related to macroeconomic conditions.

6.1 Bond return betas

This empirical analysis uses quarterly returns to match up with quarterly observations of consumption growth. Excess bond returns are measured by the return to a portfolio of Treasury bonds with maturities between five and ten years. Quarter-end to quarter-end returns are simple returns cumulated from monthly returns from CRSP. The simple return to a three-month T-bill is subtracted to form excess returns. Excess returns to the aggregate stock market are constructed in the same way, using the CRSP value-weighted index.

Table 6 reports correlations among these excess returns and log per capita consumption growth. The table also includes the same series of quarterly inflation and log growth in industrial production that are used in Section 5.2. The sample is 1952Q2 through 2010Q4. The most important information in the table is that at the quarterly frequency, excess Treasury bond returns are countercyclical. The correlations with consumption growth and industrial production growth are both negative.

This pattern may surprise some readers. The influential handbook chapter of Campbell (2003) reports that the correlation between excess bond returns and consumption growth is slightly positive for the U.S., although the evidence for other countries is mixed. Campbell uses the "beginning of period" assumption for consumption growth. In Table 6 that is the correlation between excess bond returns and the lead of consumption growth, which is slightly positive. The motivation behind the beginning of period assumption is that aggregate stock returns are more closely correlated with future consumption growth than contemporaneous consumption growth. The usual interpretation is that the shock underlying the aggregate stock return is not fully incorporated into measured consumption until the next quarter. But that is not the case for the shock underlying bond returns; the immediate reaction of both consumption and industrial production is much stronger than the one-quarter-ahead reaction.

The table also reports that the correlation between aggregate stock returns and aggregate bond returns is close to zero. This fact, combined with the negative correlation with consumption growth, spells trouble for simple consumption-based models of the term structure. On average, the nominal yield curve in the U.S. slopes up. This shape implies that expected excess returns to Treasury bonds are positive. But these correlations suggest that the risk premium should be negative.

More formal evidence is in Table 7, which reports results of regressing excess bond returns on either contemporaneous consumption growth or contemporaneous aggregate stock returns. The table also reports the sample mean of excess returns. Over the full sample 1952 through 2010, the consumption beta of excess bond returns is around minus one, while the stock market beta is almost exactly zero. Neither coefficient is statistically different from zero. Results are also displayed for the first and second halves of the sample. In one sense, these samples differ substantially from each other: the mean excess return in the first half is weakly negative, while the mean excess return in the second half is strongly positive. Yet in both cases, the coefficients are statistically indistinguishable from zero, the consumption beta is negative, and the stock market beta is almost zero. The table also contains results for a fairly homogeneous sample period, the Greenspan era of 1987 through 2006. The consumption beta over this period is about -1.8, while mean excess returns remain strongly positive.

There are, of course, a variety of ways to reconcile these betas with a positively-sloped term structure. For example, it is possible that conditional betas covary with conditional risk premia in a manner that generates both unconditionally positive mean excess returns and unconditionally negative consumption betas. In addition, the aggregate stock return is not the return to total wealth. It is possible that the beta with respect to the return to total wealth is positive even though the beta with respect to the stock market is zero. Moreover, there are alternative utility-based models that introduce heterogeneous agents, heterogeneous goods, and different utility specifications. An evaluation of these approaches is outside the scope of this chapter. The handbook chapter Duffee (2012) discusses in detail some of the theory and evidence relating the term structure and the macroeconomy.

6.2 Predictable variation of excess bond returns

Beginning with Kessel (1965) and Van Horne (1965), economists have proposed various theories of time-varying expected excess bond returns. Naturally, many theories imply that this variation is correlated with the state of the economy. Here we look at some empirical evidence to help us evaluate the theories.

For this exercise, monthly excess returns to nominal Treasury bonds are defined as the simple return to a portfolio of bonds with maturities between five and ten years less the return to a portfolio with maturities under six months. The data are from CRSP. The forecasting regressions take the form

$$\tilde{R}^{e}_{bond,t,t+1} = b_0 + b_1' \tilde{X}_t + e_{bond,t,t+1}, \tag{35}$$

where $\tilde{X}_t$ is a vector of variables known at t. The focus on monthly returns is somewhat at odds with much of the recent literature that follows Cochrane and Piazzesi (2005) by studying annual returns. But the goal here is not to generate large R^2s, but rather to judge the statistical significance of predictability. When necessary, here we follow the intuition of Hodrick (1992) by predicting monthly returns with annual averages of predictors rather than predicting annual returns with month-t values of the predictors.

Panel A of Table 8 reports the results of a variety of forecasting regressions. Consistent with the evidence in Tables 3 and 4, the statistical evidence for predictability using information in the term structure is overwhelming. The slope of the term structure, measured by the five-year yield less the three-month yield, is positively correlated with future excess returns.

Although excess returns are clearly predictable, a couple of obvious choices of macroeconomic determinants of risk premia have no predictive power. A measure of the conditional standard deviation of long-term bond yields is constructed using squared daily changes in yields. This measure has no ability to forecast excess bond returns; the p-value is about 0.5.

(Owing to the availability of daily observations of long-term yields, the sample period for this regression begins with 1962.) Many theories imply that risk premia are countercyclical. Yet over the 1952–2010 period, lagged changes in log industrial production have no ability to forecast excess returns, whether the changes are from month $t-1$ to month t or from month $t-12$ to month t.

The next regression uses the eight principal components of macroeconomic and financial times series constructed by Ludvigson and Ng (2010). Their data are available for 1964 through 2007. Table 8 reports that when these eight principal components are used to forecast monthly excess returns, the adjusted R^2 is more than eight percent. The hypothesis that the coefficients are jointly zero is rejected at any conceivable confidence level. Note that this regression does not include any information from the term structure. This evidence concerns the overall forecast power of the eight principal components, not their incremental explanatory power.

Although this evidence is certainly encouraging to researchers searching for links between risk premia and the macroeconomy, there are three important caveats. First, it is not clear how to map much of this predictability to specific macroeconomic concepts. Ludvigson and Ng call the first principal component "real activity" because it is highly correlated with standard measures of real activity. For example, its correlation with log-differenced industrial production exceeds 0.8. But relatively little of the forecast power of the principal components comes from the measure of real activity. Table 8 reports that the adjusted R^2 using only the first principal component is about 1.6 percent. Interpreting the other principal components is more problematic.

Second, the sample period 1964 through 2007, although very long, may be anomalous. Table 8 uses this shorter sample to redo the forecasting regression with industrial production growth. With the shorter sample, the p-value for growth in industrial production drops from 0.5 to 0.02. Since the first principal component is highly correlated with the growth in industrial production, this result suggests that the predictability associated with Ludvigson and Ng's real activity factor may be sample-specific.

Third, the predictability of excess bond returns is not accompanied by similar predictability of excess stock returns. In models that link time-varying bond risk premia to the macroeconomy, this variation is typically accompanied by varying equity risk premia. But Panel B of Table 8 reports that the shape of the term structure contains little information about future aggregate equity excess returns. These regressions take the same form as (35), with excess stock returns replacing excess bond returns on the left side. The eight principal components of macroeconomic and financial variables collectively have substantial forecasting power, but the coefficients on the principal components do not line up with those that forecast bond returns. The correlation between the fitted values of the regressions is only 0.39. For the aggregate stock excess return, the coefficient on real activity is statistically indistinguishable from zero.

7 A robustness check

Is there really information about future excess returns that is not captured by the current term structure? Section 5.2 briefly discusses empirical evidence that macroeconomic variables have incremental explanatory power. We take a closer look at that evidence here. We also consider one of the results in Cochrane and Piazzesi (2005), hereafter CP. The material in this section is new to the literature.

In an influential paper, CP find that five month-t forward rates constructed with yields on bonds with maturities of one through five years contain substantial information about excess bond returns over the next year (the end of month t to the end of month $t+12$). Their evidence is surprising along many dimensions. In particular, they conclude that lags of forward rates (dated $t-1$ and earlier) contain information about the excess returns that is not in month t forward rates.

As CP note, this result is inconsistent with the logic that the time-t term structure contains all information relevant to forecasting future yields and excess excess returns. In this sense, the evidence is similar to the evidence that macro variables help predict excess returns. Hidden factors help justify the macro evidence. But even if there are hidden factors (a theoretical development not available to CP), it is hard to understand why the information is hidden from the time-t term structure but not hidden in earlier term structures. CP suggest measurement error in yields that is averaged away by including additional lags of yields.

Rather than examine the annual horizon returns of CP, here we look at monthly returns. Consider forecasting monthly excess returns using both the time t term structure and lagged information, as in

$$\tilde{R}^{e}_{bond,t,t+1} = b_0 + b_1' \tilde{\mathcal{P}}_t + b_2' \sum_{i=0}^{11} \tilde{F}_{t-i} + e_{bond,t,t+1}. \tag{36}$$

In (36), the notation $\tilde{F}_t$ refers to a vector of forward rates observed at month t,

$$\tilde{F}_t = \begin{pmatrix} \tilde{F}_{0,1,t} & \tilde{F}_{1,2,t} & \tilde{F}_{2,3,t} & \tilde{F}_{3,4,t} & \tilde{F}_{4,5,t} \end{pmatrix}',$$

where $\tilde{F}_{j-1,j,t}$ is the forward rate from year $j-1$ to year j as of month t. The notation $\tilde{\mathcal{P}}_t$ refers to the vector of five principal components of the term structure. If the principal components are excluded from this regression, it is the Hodrick (1992) version of the regressions in CP. Instead of using month-t forward rates to forecast annual excess returns, the sum of the previous twelve months of forward rates are used to forecast monthly excess returns. The excess bond return is for a portfolio of Treasury bonds with maturities between five and ten years.

The regression is estimated for the 1964 through 2003 sample studied by CP and for the full 1952 through 2010 sample. Two versions are estimated. One excludes the forward rates and the other includes both sets of explanatory variables. Results are in Table 9, as are other results that are discussed below.

When the forward rates are excluded, the R^2 for the 1964 through 2003 sample is 5.8 percent. (This duplicates one of the regressions reported in the last line of Table 4.) Including lagged forward rates raises the R^2 to 11.0 percent. The null hypothesis that the coefficients on the forward rates are all zero is overwhelmingly rejected. However, over the full sample, the incremental explanatory power of the forward rates is modest; the comparable R^2s are 5.0 percent and 6.6 percent respectively. For the full sample, the hypothesis that the coefficients on the forward rates are all zero cannot be rejected at the five percent level.

These results are qualitatively similar to the forecasting regression R^2s in Table 4 and discussed in Section 3.4. In Table 4, five principal components of yields are used to predict excess returns at horizons of one, three, and twelve months. The R^2s for one month are approximately the same for the three sample periods 1952–2010, 1972–2010, and 1964–2003. The same result holds for the three-month horizon. But at the annual horizon, the 1964 through 2003 period stands out, with R^2s that are about 20 percentage points greater than the R^2s for the other two periods. The results of Table 9 explain the discrepancy, at least in a mechanical sense. Over the 1964 to 2003 sample, lagged term structures have incremental predictive power for monthly excess returns. Thus the forecast power of lagged term structures is unusual in a statistical sense: it is overwhelmingly significant over most of the 1952 through 2010 sample, but insignificant over the entire period.

The forecast power of lagged term structures is unusual in another way. Recall the discussion in Section 3.2 that future changes in the level of the term structure are unforecastable with the current term structure. Only changes in the slope and curvature are forecastable. Yet over the 1964 through 2003 sample, lagged forward rates predict future changes in the level.

The evidence is again in Table 9. The relevant regressions replace the excess return on the left side of (36) with changes in the first three principal components of yields. Over the sample period of CP, the null hypothesis that lagged forward rates have no incremental explanatory power for changes in the level is rejected at the three percent level. Similar hypotheses for slope and curvature are not rejected. Over the full sample period, lagged forward rates have no ability to forecast the level of yields.

The level of the term structure is a random walk: it exhibits no mean reversion. But during the period 1964 through 2003, lagged term structures contained information about future increments to the level. This is why the lags have incremental explanatory power for excess returns. The next major result of this section is that the same statement applies to macro variables known to forecast excess returns.

Denote the eight principal components of macro and financial variables constructed by Ludvigson and Ng (2010) by $\tilde{\mathcal{M}}_t$.[6] The excess return forecasting regression is

$$\tilde{R}^e_{bond,t,t+1} = b_0 + b_1' \tilde{\mathcal{P}}_t + b_2' \tilde{\mathcal{M}}_t + e_{bond,t,t+1}. \tag{37}$$

The sample period is 1964 through 2007.

The first panel of Table 10 reports the results. When the Ludvigson-Ng data are excluded, the R^2 is 5.5 percent. When they are included, the R^2 jumps to 13.8 percent, and the hypothesis that the coefficients on the principal components are all zero is overwhelmingly rejected. The next set of regressions replace the excess return on the left side of (37) with changes in level, slope, and curvature. The macro/financial factors have substantial incremental forecast power for the level of the term structure and none for either slope or curvature.

Next, replace the macro/financial principal components with the measure of lagged inflation constructed by Cieslak and Povala (2011).[7] The sample period is 1972 through 2009, determined by the availability of the inflation measure. The results are in the second panel of Table 10. Including this single predictive variable raises the R^2 of excess returns from 6.5 percent to 10.6 percent. The hypothesis that the coefficient is zero is overwhelmingly rejected. The same result holds for forecasts of changes in the level of the term structure. Again, the variable has no incremental explanatory power for slope or curvature.

These results raise two important questions about the robustness of the results of Ludvig-

[6]Thanks to the authors for making their data available.

[7]Thanks the the authors for making their data available.

son and Ng (2009, 2010) and Cieslak and Povala (2011). First, how reliable is the statistical inference? Drawing correct inferences inferences about the finite-sample behavior of a random walk process is often difficult. Consider, for example, Figure 4, which displays the three-month yield, the five-year yield, and the lagged inflation measure of Cieslak and Povala. Over the sample period 1970 through 2010, the level of yields rises steadily for 12 years, then drops for the next 30 years. The behavior of the lagged inflation measure is similar. This kind of pattern can easily produce spurious predictability. Both Ludvigson and Ng (2009, 2010) and Cieslak and Povala (2011) are concerned about statistical reliability, and employ pseudo out-of-sample tests. But true out-of-sample tests require patience.

Second, what kind of hidden-factor story is consistent with this evidence? Macro stories are typically cyclical. But to explain the patterns here, a hidden-factor story has to explain why macro news today corresponds to an expected future permanent shock to short rates. Moreover, the shock has to temporarily change risk premia in the opposite direction.

To answer the question posed at the beginning of the section: perhaps not. There are good theoretical and empirical reasons to be skeptical of the evidence that there is information about expected excess returns beyond that contained in bond yields.

8 Concluding comments

Finance theory provides some specific guidance when forming forecasts of future interest rates. Nonetheless, important questions remain open. The Holy Grail of this literature is a dynamic model that is parsimonious owing to economically-motivated restrictions. The requirement of no-arbitrage is motivated by economics, but by itself it is too weak to matter. Economic restrictions with bite require, either directly or indirectly, that risk premia dynamics be tied down by economic principles. There no consensus in the literature about how this should be done. To date, no restrictions that come out of our workhorse models of asset pricing appear to be consistent with the observed behavior of Treasury bond yields.

Another open question is whether any variables contain information about future interest rates that is not already in the current term structure. Recent empirical work suggests that both lagged bond yields and certain macroeconomic variables have incremental information, but this chapter argues that the robustness of these results is not yet known. A closely related question is whether changes in the overall level of the term structure are forecastable. Again, recent empirical work suggests (at least implicitly) they are forecastable, but the nature of this forecastability raises some statistical and economic problems that are not yet resolved.

References

Ahn, Dong-Hyun, Robert F. Dittmar, and A. Ronald Gallant, 2002, Quadratic term structure models: Theory and evidence, *Review of Financial Studies* 15, 243-288.

Almeida, Caio, Jeremy J. Graveline, and Scott Joslin, 2011, Do interest rate options contain information about excess returns?, *Journal of Econometrics* 164, 35-44.

Almeida, Caio, Axel Simonsen, and José Vicente, 2012, Forecasting bond yields with segmented term structure models, Working paper, Central Bank of Brazil.

Almeida, Caio, and José Vicente, 200, The role of no-arbitrage on forecasting: Lessons from a parametric term structure model, Working paper, Central Bank of Brazil.

Ang, Andrew, Geert Bekaert, and Min Wei, 2007, Do macro variables, asset markets or surveys forecast inflation better?, *Journal of Monetary Economics* 54, 1163-1212.

Ang, Andrew, Geert Bekaert, and Min Wei, 2008, The term structure of real rates and expected inflation, *Journal of Finance* 63, 797-849.

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.

Backus, David K., and Stanley E. Zin, 1994, Reverse engineering the yield curve, NBER Working paper 4676.

Bauer, Michael D., 2011, Term premia and the news, Federal Reserve Bank of San Francisco Working Paper 2011-03.

Bekaert, Geert, Robert J. Hodrick, and David A. Marshall, 1997, On biases in tests of the expectations hypothesis of the term structure of interest rates, *Journal of Financial Economics* 44, 309-348.

Bowsher, Clive G., and Roland Meeks, The dynamics of economic functions: Modelling and forecasting the yield curve, 2008, Working Paper, Federal Reserve Bank of Dallas.

Campbell, John Y., Consumption-based asset pricing, in George Constantinides, Milton Harris, and Rene Stulz., Eds., *Handbook of the Economics of Finance Vol. 1B*, North-Holland, Amsterdam, 803-887.

Campbell, John Y., and John H. Cochrane, 1999, By force of habit: A consumption-based explanation of aggregate stock market behavior, *Journal of Political Economy* 107, 205-251.

Campbell, John Y., and Robert J. Shiller, 1987, Cointegration and tests of present value models, *Journal of Political Ecoomy* 95, 1062-1088.

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.

Carriero, Andrea, 2011, Forecasting the yield curve using priors from no-arbitrage affine term structure models, *International Economic Review* 52, 425-460.

Carriero, Andrea, George Kapetanios, and Massimiliano Marcellino, 2012, Forecasting government bond yields with large Bayesian vector autoregressions, *Journal of Banking and Finance* 36, 2026-2047.

Cheridito, Patrick, Damir Filipović, and Robert L. Kimmel, 2007, Market price of risk specifications for affine models: Theory and evidence, *Journal of Financial Economics* 83, 123-170.

Chernov, Mikhail, and Philippe Mueller, 2011, The term structure of inflation expectations, *Journal of Financial Economics*, forthcoming.

Christensen, Jens H.E., Francis X. Diebold, and Glenn D. Rudebusch, 2010, The affine arbitrage-free class of Nelson-Siegel term structure models, *Journal of Econometrics*, forthcoming.

Christensen, Jens H.E., Jose Lopez, and Glenn D. Rudebusch, 2010, Inflation expectations and risk premiums in an arbitrage-free model of nominal and real bond yields, *Journal of Money, Credit, and Banking* 42, 143-178.

Cieslak, Anna, and Pavol Povala, 2011, Understanding bond risk premia, Working paper, Kellogg School of Management.

Clark, Todd E., and Michael W. McCracken, 2012, Advances in Forecast Evaluation, this Handbook.

Cochrane, John H., and Monika Piazzesi, 2005, Bond risk premia, *American Economic Review* 95, 138-160.

Cochrane, John H., and Monika Piazzesi, 2008, Decomposing the yield curve, Working paper, Chicago Booth.

Collin-Dufresne, Pierre, and Robert S. Goldstein, 2002, Do bonds span the fixed income markets? Theory and evidence for unspanned stochastic volatility, *Journal of Finance* 57, 1685-1730.

Cooper, Ilan, and Richard Priestly, 2009, Time-varying risk premiums and the output gap, *Review of Financial Studies* 22, 2801-2833.

Dai, Qiang, and Kenneth J. Singleton, 2000, Specification analysis of affine term structure models, *Journal of Finance* 55, 1943-1978.

Dai, Qiang, Kenneth J. Singleton, and Wei Yang, Regime shifts in a dynamic term structure model of U.S. Treasury bond yields, *Review of Financial Studies* 20, 1669-1706.

Diebold, Francis X., and Canlin Li, 2006, Forecasting the term structure of government bond yields, *Journal of Econometrics* 130, 337-364.

Duarte, Jefferson, 2004, Evaluating an alternative risk preference in affine term structure models, *Review of Financial Studies* 17, 379-404.

Duffee, Gregory R., 2002, Term premia and interest rate forecasts in affine models, *Journal of Finance* 57, 405-443.

Duffee, Gregory R., 2010, Sharpe ratios in term structure models, Working paper, Johns Hopkins.

Duffee, Gregory R., 2011a, Information in (and not in) the term structure, *Review of Financial Studies* 24, 2895-2934.

Duffee, Gregory R., 2011b, Forecasting with the term structure: The role of no-arbitrage restrictions, Working paper, Johns Hopkins.

Duffee, Gregory R., 2012, Bond pricing and the macroeconomy, in G.M. Constantinides, M. Harris and R. Stulz, Eds., *Handbook of the Economics of Finance*, forthcoming.

Duffie, Darrell, and Rui Kan, 1996, A yield-factor model of interest rates, *Mathematical Finance* 6, 379-406.

Epstein, Larry G., and Stanley E. Zin, 1989, Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework, *Econometrica* 57, 937-969.

Fama, Eugene F., and Robert R. Bliss, 1987, The information in long-maturity forward rates, *American Economic Review* 77, 680-692.

Faverno, Carlo A., Linlin Niu, and Luca Sala, 2012, Term structure forecasting: no-arbitrage restrictions versus large information set, *Journal of Forecasting* 31, 124-156.

Gray, Stephen F., 1996, Modeling the conditional distribution of interest rates as a regime-switching process, *Journal of Financial Economics* 42, 27-62.

Gurkaynak, Refet S., Brian Sack, and Jonathan H. Wright, 2007, The U.S. Treasury yield curve: 1961 to the present, *Journal of Monetary Economics* 54, 2291-2304.

Hall, Anthony D., Heather M. Anderson and Clive W.J. Granger, 1992, *Review of Economics and Statistics* 74, 116-126.

Hodrick, Robert J., 1992, Dividend yields and expected stock returns: alternative procedures for inference and measurement, *Review of Financial Studies* 5, 357-386.

Joslin, Scott, Marcel Priebsch, and Kenneth J. Singleton, 2010, Risk premiums in dynamic term structure models with unspanned macro risks, Working paper, Stanford GSB.

Joslin, Scott, Kenneth J. Singleton, and Haoxiang Zhu, 2011, A new perspective on Gaussian dynamic term structure models, *Review of Financial Studies* 24, 926-970.

Kessel, Reuben A., 1965, The cyclical behavior of the term structure of interest rates, NBER occasional paper 91.

Kim, Don H., and Athanasios Orphpanides, 2005, Term structure estimation with survey data on interest rate forecasts, FEDS working paper 2005-48, Federal Reserve Board.

Koopman, Siem Jan, and Michel van der Wel, 2011, Forecasting the U.S. term structure of interest rates using a macroeconomic smooth factor dynamic model, Working paper, Tinbergen Institute.

Le, Ahn, Kenneth J. Singleton, and Qiang Dai, 2010, Discrete-time affine^Q term structure models with generalized market prices of risk, *Review of Financial Studies* 23, 2184-2227.

Leippold, Markus, and Liuren Wu, 2002, Asset pricing under the quadratic class, *Journal of Financial and Quantitative Analysis* 37, 271-295.

Litterman, Robert, and Jose Scheinkman, 1991, Common factors affecting bond returns, *Journal of Fixed Income* 1, 54-61.

Ludvigson, Sydney C., and Serene Ng, 2009, Macro factors in bond risk premia, *Review of Financial Studies* 22, 5027-5067.

Ludvigson, Sydney C., and Serene Ng, 2010, A factor analysis of bond risk premia, *Handbook of Applied Econometrics*, forthcoming.

Martin, Vance, Anthony D. Hall, and Adrian R. Pagan, 1996, Modelling the term structure, *Handbook of Statistics* 14, G.S. Maddala and C.R. Rao, Eds., 91-118.

Nelson, Charles R., and Andrew F. Siegel, 1987, Parsimonious modeling of yield curves, *Journal of Business* 60, 473-489.

Piazzesi, Monika, and Martin Schneider, 2007, Equilibrium yield curves, *NBER Macroeconomics Annual 2006*, Cambridge, MA: MIT Press, 389-442.

Van Horne, James, 1965, Interest-rate risk and the term structure of interest rates, *Journal of Political Economy* 73, 344-351.

Table 1. Cross-sectional fitting errors using principal components

Principal components of zero-coupon nominal bond yields are calculated for the sample January 1972 through December 2010. The maturities range from three months to ten years. The table reports cross-sectional residuals from using the first n components to fit the cross-section of yields, from $n = 0$ to $n = 5$. The values for $n = 0$ are the sample standard deviations of the yields. Standard deviations measured in basis points of annualized yields.

Number of components	3 mon	1 yr	2 yr	3 yr	4 yr	5 yr	10 yr
0	322.2	319.7	310.3	298.9	290.0	281.3	250.8
1	75.3	44.0	16.3	18.8	30.2	38.9	65.0
2	21.4	14.5	14.6	12.4	9.4	6.4	22.1
3	5.0	10.9	6.1	6.2	8.1	6.4	7.4
4	0.5	3.4	6.0	5.0	5.2	5.6	1.9
5	0.2	0.3	2.6	4.9	4.4	5.1	1.2

Table 2. Predicting level, slope, and curvature with the term structure

Principal components of month-end observations of nominal bond yields are constructed for the specified samples. Changes in these principal components from month t to months $t+1$, $t+3$, and $t+12$ are forecast in-sample with OLS using the month-t values of the first five principal components. The table reports R^2s of these regressions and p-values of tests of the hypothesis that all five coefficients are zero. For forecasts of the second and third principal components, the table also reports p-values of tests of the hypothesis that the coefficients on the fourth and fifth principal components are zero. The tests use the White heteroskedasticity-consistent covariance matrix.

The 1952 through 2010 sample uses yields with maturities of three months and one through five years. For this sample the forecasting regressions use 702 monthly observations from 1952:7 through 2010, 234 quarterly observations from 1952Q3 through 2010Q4, and 58 annual observations from 1953 through 2010. The 1972 through 2010 sample adds a ten-year maturity. For this sample the regressions use 467 monthly, 155 quarterly, and 38 annual observations.

	First PC		Second PC			Third PC		
Freq	R^2	1-5 Test	R^2	1-5 Test	4-5 Test	R^2	1-5 Test	4-5 Test
1952 through 2010 sample								
M	0.016	0.219	0.077	0.000	0.946	0.125	0.000	0.048
Q	0.038	0.192	0.198	0.000	0.842	0.300	0.000	0.011
A	0.143	0.247	0.549	0.000	0.229	0.496	0.000	0.993
1972 through 2010 sample								
M	0.029	0.125	0.082	0.002	0.784	0.089	0.000	0.246
Q	0.055	0.160	0.220	0.000	0.580	0.255	0.000	0.005
A	0.161	0.062	0.582	0.000	0.439	0.469	0.000	0.406

Table 3. Forecasting excess returns to Treasury bonds with the yield curve

Excess returns to maturity-sorted portfolios of Treasury bonds from month t to month $t+k$ are regressed on the month-t values of the first five principal components of the Treasury term structure. The horizons are one month, one quarter, and one year. Only non-overlapping observations are used. Returns are expressed in percent. The principal components are all normalized to have unit standard deviation. The table reports parameter estimates and asymptotic t-statistics. The covariance matrices of parameter estimates are adjusted for generalized heteroskedasticity.

Freq	Maturity bucket	PC 1	PC 2	PC 3	PC 4	PC 5
1952 through 2010 sample						
M	2-3 years	0.028	−0.110	−0.068	0.050	−0.027
		(0.54)	(−2.06)	(−1.97)	(1.43)	(−0.47)
M	5-10 years	0.041	−0.257	−0.139	0.199	−0.011
		(0.44)	(−3.03)	(−2.08)	(2.60)	(−0.12)
Q	5-10 years	0.206	−0.730	−0.573	0.183	−0.056
		(0.67)	(−2.26)	(−3.26)	(0.87)	(−0.17)
A	5-10 years	0.861	−2.799	−0.833	−0.187	−1.254
		(1.13)	(−3.86)	(−1.48)	(−0.22)	(−1.49)
1972 through 2010 sample						
M	2-3 years	0.012	−0.117	−0.082	0.082	−0.081
		(0.18)	(−1.84)	(−1.59)	(1.20)	(−0.96)
M	5-10 years	0.005	−0.285	−0.151	0.237	−0.217
		(0.04)	(−2.73)	(−1.63)	(1.88)	(−1.70)
Q	5-10 years	0.137	−0.847	−0.572	0.630	−0.057
		(0.37)	(−2.58)	(−2.19)	(1.49)	(−0.10)
A	5-10 years	0.396	−2.639	−1.103	−0.027	−2.096
		(0.40)	(−2.77)	(−1.58)	(−0.02)	(−1.96)

Table 4. R^2s of excess-return forecasting regressions

Excess returns to maturity-sorted portfolios of Treasury bonds from month t to month $t+k$ are regressed on the month-t values of the first five principal components of the Treasury term structure. The horizons are one month, one quarter, and one year. Only non-overlapping observations are used. The table reports R^2s for various combinations of sample periods and maturity-sorted portfolios.

Sample period	Maturity bucket	Month	Quarter	Annual
1952–2010	2-3 years	0.030	0.075	0.252
1972–2010	2-3 years	0.040	0.098	0.243
1964–2003	2-3 years	0.036	0.097	0.526
1952–2010	5-10 years	0.050	0.107	0.309
1972–2010	5-10 years	0.064	0.132	0.343
1964–2003	5-10 years	0.058	0.129	0.525

Table 5. Explaining macroeconomic variables with the term structure

The term structure at the end of quarter t is summarized by its first five principal components. CPI inflation is measured from the last month of quarter $t-1$ to the last month of quarter t. Industrial production growth is measured similarly. Log aggregate per capita consumption growth is measured from quarter $t-1$ to quarter t. The table reports R^2s of regressions of the macro variables on the five principal components of the term structure. Regressions are both cross-sectional and one-quarter-ahead forecasts. Univariate R^2s are for AR(1) regressions of the macro variables. The column labeled "T-statistic" is the t-statistic on the lagged macro variable when it is included in a forecasting regression with the five principal components. It is adjusted for generalized heteroskedasticity. The data sample is 1952Q2 through 2010Q4.

	Cross section	Forecasting, principal components	Forecasting, univariate	T-statistic
Inflation	0.451	0.429	0.376	3.54
Consumption growth	0.090	0.110	0.143	4.69
IP growth	0.032	0.081	0.146	3.87

Table 6. Sample correlations, 1952 through 2010

Excess returns to a portfolio of long-term Treasury bonds and the aggregate stock market are measured from the end of quarter $t-1$ to the end of quarter t. Contemporaneous per capita consumption growth is the log change in aggregate per capita consumption from quarter $t-1$ to quarter t, Δc_t. Industrial production growth and CPI inflation are both log changes from the final month of quarter $t-1$ to the final month of quarter t, ΔIP_t and π_t respectively. The sample period is 1952Q2 through 2010Q4.

	Excess bond ret	Excess stock ret	Δc_t	Δc_{t+1}	ΔIP_t	ΔIP_{t+1}	π_t
Excess bond return	1.00						
Excess stock return	0.06	1.00					
Δc_t	−0.13	0.20	1.00				
Δc_{t+1}	0.07	0.26	0.38	1.00			
ΔIP_t	−0.28	0.04	0.50	0.26	1.00		
ΔIP_{t+1}	−0.13	0.36	0.42	0.52	0.38	1.00	
π_t	−0.29	−0.15	−0.25	−0.30	−0.13	−0.19	1.00

Table 7. Consumption and CAPM betas for excess bond returns

Excess returns to a portfolio of long-term Treasury bonds and the aggregate stock market are measured from the end of quarter $t-1$ to the end of quarter t. Contemporaneous consumption growth is the log change in aggregate per capita consumption from quarter $t-1$ to quarter t, Δc_t. The table reports coefficients of regressions of excess bond returns on either consumption growth or stock returns. Asymptotic t-statistics, in parentheses, use a Newey-West adjustment with two lags. The units are percent per quarter.

Sample	Constant	Consumption growth	aggregate stock return
1952:2–2010:4	0.40		
	(1.91)		
	0.77	−0.86	
	(2.41)	(−1.81)	
	0.36		0.02
	(1.64)		(0.63)
1952:2–1980:4	−0.15		
	(−0.67)		
	0.29	−0.88	
	(0.66)	(−1.40)	
	−0.27		0.07
	(−1.14)		(1.93)
1981:1–2010:4	0.92		
	(2.90)		
	1.11	−0.51	
	(2.52)	(−0.65)	
	0.94		−0.01
	(2.88)		(−0.27)
1986:1–2007:4	0.74		
	(2.64)		
	1.47	−1.76	
	(3.06)	(−1.97)	
	0.85		−0.06
	(2.99)		(−1.38)

Table 8. Predictability of excess returns to bonds and stocks

Monthly excess returns to long-term nominal Treasury bonds and the aggregate stock market are regressed on a variety of predetermined variables. The column labeled "Test Stat" is the p-value of the test that the coefficient(s) are all zero. The covariance matrix of the estimated coefficients uses the Newey-West adjustment for five lags of moving average residuals. For regressions with a single predictor, the sign of the estimated coefficient is reported.

Predictors	Sample	Adj. R^2	Test stat	Sign
A. Excess bond returns				
Slope	1952:7 – 2010:12	0.0256	0.002	Pos
Conditional SD of yields	1962:3 – 2010:12	0.0007	0.486	Pos
Log change of industrial prod.	1952:7 – 2010:12	0.0016	0.111	Neg
12-month average of log change of industrial prod.	1952:7 – 2010:12	0.0007	0.225	Neg
8 principal components of of macro/financial variables	1964:2 – 2008:1	0.0864	0.000	—
"Real activity" principal component of macro/financial variables	1964:2 – 2008:1	0.0161	0.002	Neg
Log change of industrial prod.	1964:2 – 2008:1	0.0066	0.017	Neg
B. Excess stock returns				
Slope	1952:7 – 2010:12	0.0047	0.066	Pos
Log change of industrial prod.	1952:7 – 2010:12	−0.0012	0.774	Pos
8 principal components of of macro/financial variables	1964:2 – 2008:1	0.0408	0.000	—
"Real activity" principal component of macro/financial variables	1964:2 – 2008:1	0.0018	0.245	Neg

Table 9. The incremental explanatory power of lagged term structures

The month-t term structure is summarized in two ways: the first five principal components of yields and five forward rates from year-ahead $j-1$ to year-ahead j, $j = 1, \ldots, 5$. The forward rates are then averaged over months $t-11, \ldots, t$. This term structure information is used to predict excess returns to a portfolio of Treasury bonds from month t to $t+1$. It is also used to predict changes in the level, slope, and curvature of the term structure (first three principal components) from t to $t+1$. The table reports R^2s for the forecasting regressions. It also reports the p-value of a test that the averaged forward rates have no incremental explanatory power when the five principal components are included in the regression. The covariance matrices of the parameter estimates are adjusted for generalized heteroskedasticity.

Forecasted Variable	Five principal components	Principal components and average forward rates	Test of forwards
1964 through 2003 sample			
Excess returns	0.058	0.110	0.001
First PC (level)	0.020	0.066	0.023
Second PC (slope)	0.079	0.098	0.336
Third PC (curve)	0.129	0.147	0.098
1952 through 2010 sample			
Excess returns	0.050	0.066	0.094
First PC (level)	0.016	0.028	0.322
Second PC (slope)	0.077	0.097	0.101
Third PC (curve)	0.125	0.145	0.015

Table 10. The incremental explanatory power of macroeconomic variables

The month-t term structure is summarized by the first five principal components of yields. These principal components are used to predict excess returns to a portfolio of Treasury bonds from month t to $t+1$. Other predictors are the eight principal components of many macroeconomic time series, constructed by Ludvigson and Ng (2010), and a weighted average of past inflation constructed by Cieslak and Povala (2011). These variables are also used to predict changes in the level, slope, and curvature of the term structure (first three principal components) from t to $t+1$. The table reports R^2s for the forecasting regressions. It also reports the p-value of a test that other predictors have no incremental explanatory power when the five principal components are included in the regressions. The covariance matrices of the parameter estimates are adjusted for generalized heteroskedasticity.

Forecasted Variable	Five yield PCs	Yield PCs and macro PCs	Test of macro PCs
1964:1 through 2007:12 sample			
Excess returns	0.058	0.138	0.000
First PC (level)	0.021	0.140	0.000
Second PC (slope)	0.080	0.117	0.056
Third PC (curve)	0.130	0.159	0.232

Forecasted Variable	Five yield PCs	Yield PCs and past inflation	Test of past inflation
1971:11 through 2009:12 sample			
Excess returns	0.065	0.106	0.000
First PC (level)	0.027	0.056	0.000
Second PC (slope)	0.078	0.079	0.406
Third PC (curve)	0.130	0.131	0.757

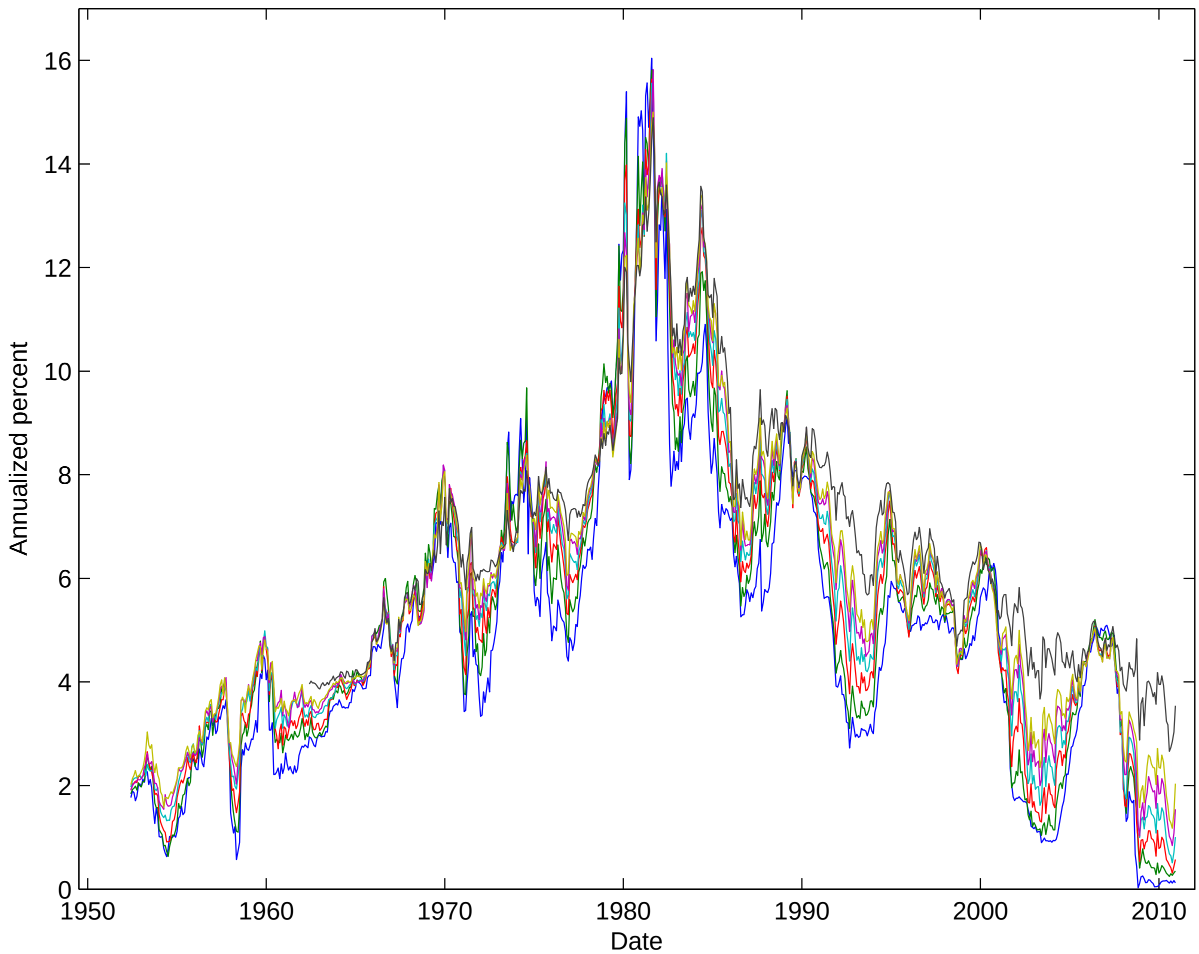

Fig. 1. Yields on nominal Treasury zero-coupon bonds. The maturities range from three months to ten years.

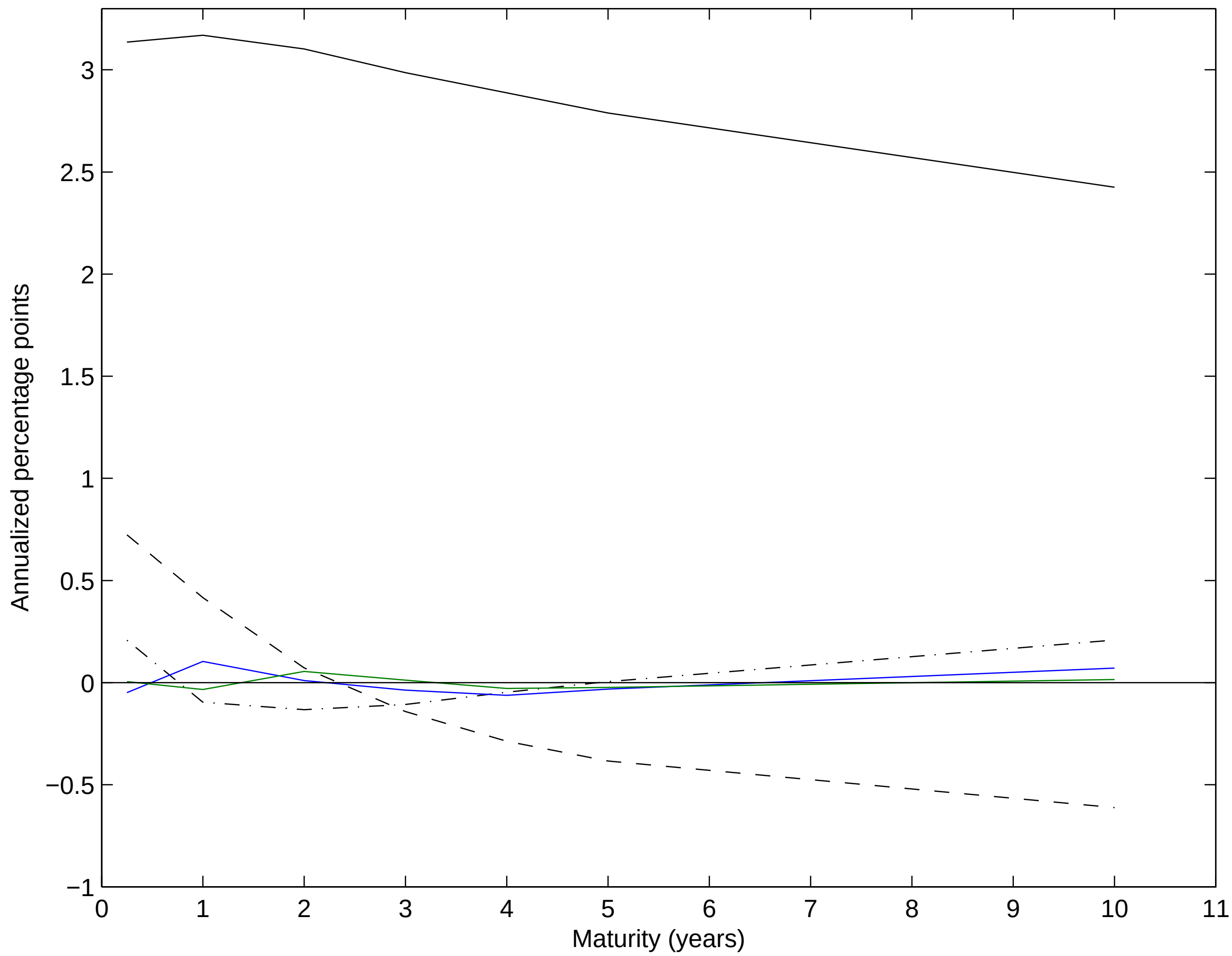

Fig. 2. Principal components of nominal bond yields. The figure displays loadings of yields on their first five principal components, multiplied by the sample standard deviations of the components. The first, second, and third principal components, termed "level," "slope," and "curvature," are displayed with solid, dashed, and dotted-dashed lines. The data sample is January 1972 through December 2010.

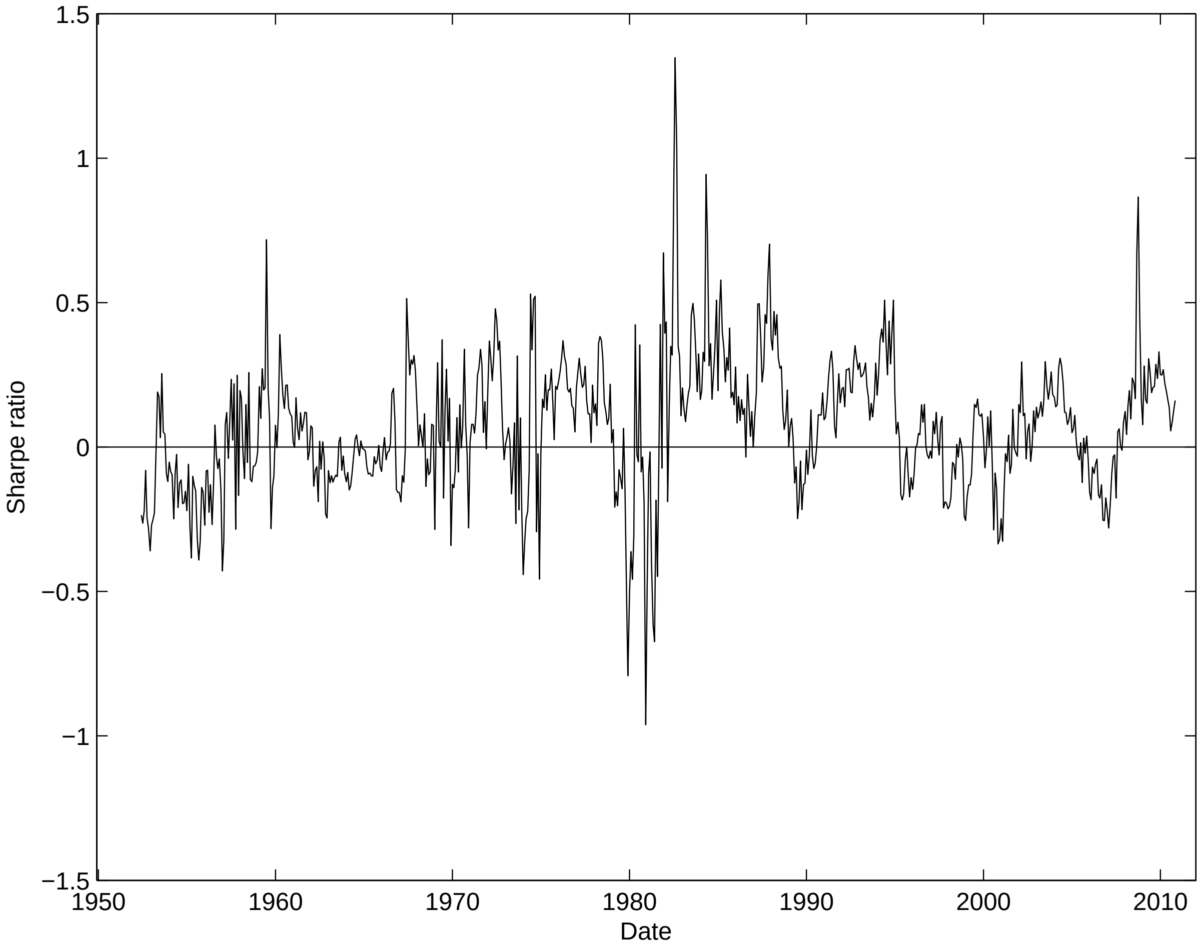

Fig. 3. Conditional monthly Sharpe ratios. The figure displays fitted conditional Sharpe ratios from month t to month $t+1$ for a portfolio of Treasury bonds with maturities between five and ten years. Monthly excess returns are predicted using the first five principal components of Treasury yields. The Sharpe ratios are calculated assuming that conditional standard deviations of excess returns are constant over time.

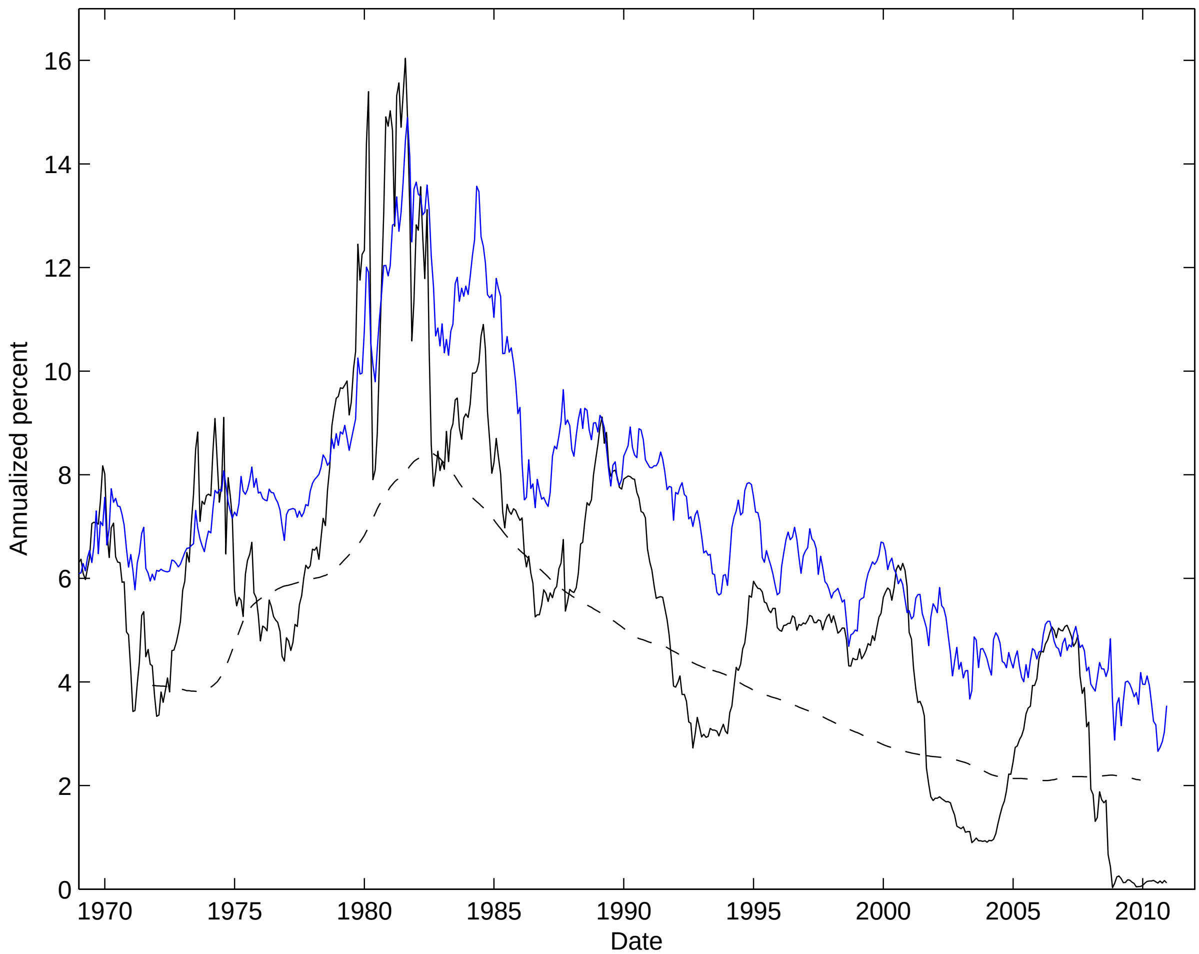

Fig. 4. Treasury bond yields and a backward-looking measure of long-run inflation. The figure plots yields on three-month and ten-year zero-coupon Treasury bonds. It also plots a weighted average of past inflation, denoted τ^{CPI}, constructed by Cieslak and Povala (2011).