Information in (and not in) the term structure

Gregory R. Duffee[*]

Johns Hopkins

First draft: March 2008
This Draft: October 28, 2010

ABSTRACT

Standard approaches to building and estimating dynamic term structure models rely on the assumption that yields can serve as the factors. However, the assumption is neither theoretically necessary nor empirically supported. This paper documents that almost half of the variation in bond risk premia cannot be detected using the cross section of yields. Fluctuations in this hidden component have strong forecast power for both future short-term interest rates and excess bond returns. They are also negatively correlated with aggregate economic activity, but macroeconomic variables explain only a small fraction of variation in the hidden factor.

[*]Voice 410-516-8828, email duffee@jhu.edu. Address correspondence to 440 Mergenthaler Hall, 3400 N. Charles St., Baltimore, MD 21218. The first version of this paper was written while the author was visiting Wharton. Thanks to Pierre Collin-Dufresne, Wayne Ferson, Scott Joslin, Sydney Ludvigson, Jacob Sagi, Dan Thornton, and seminar participants at the Federal Reserve Board, USC, Wharton, the 2008 NBER Summer Institute, and the 2009 WFA Annual Meetings for helpful conversations and comments.

1 Introduction

This paper advocates a significant change in the construction and estimation of multifactor term structure models. In a literature spanning more than two decades, researchers have almost universally assumed that the factors driving term structure dynamics can be represented as functions of yields. The assumption plays a critical role in all aspects of estimation. However, because it rules out a potentially important class of term structure dynamics, we need research methodologies that do not rely on the assumption.

The intuition behind the standard approach is so obvious that it is seldom mentioned. Investors' beliefs about future bond prices determine what investors are willing to pay for bonds. This suggests that today's term structure contains all information relevant to predicting both future returns to bonds and future bond yields. Put somewhat differently, the term structure follows a Markov process.

Empirical work exploits this Markov structure in many ways. It helps researchers choose the dimension of a model, because the same factors that determine the cross section of yields also determine yield dynamics. Therefore factor analysis of unconditional covariances among yields (the cross section) pins down the length of the state vector. It also simplifies considerably the search for time-varying expected bond returns, because it implies that time-t conditional expectations of returns can be expressed entirely in terms of forward rates observed at t. Other data are unnecessary to model yield dynamics. In addition, the one-to-one mapping from factors to yields implied by the Markov structure leads to tractable estimation of very complicated term structure models.

Yet recent empirical evidence calls this assumption into question. Ludvigson and Ng (2009) and Cooper and Priestly (2009) conclude that various measures of macroeconomic activity contain information about future excess bond returns that is not in forward rates. Cochrane and Piazzesi (2005) find that lagged forward rates contain information about future excess bond returns that is not in current forward rates. One possible explanation, as noted by Cochrane and Piazzesi, is that measurement error in yields obscures the Markov structure. In other words, these empirical results hinge on our inability to precisely observe yields. But plausible measurement error in Treasury yields is on the order of only a few basis points. Thus it is incumbent upon us to attempt to understand, from a formal perspective, why tiny measurement errors can cover up important information contained in the cross section of yields.

I show that it is easy to build a multifactor model in which one of the factors plays an important role in determining investors' expectations of future yields, yet has no effect on current yields. I refer to such a factor as a "hidden" factor, in the sense that a snapshot of the time-t yield curve conveys no information about it. A hidden factor has opposite effects on expected future interest rates and bond risk premia.

Consider, for example, news that raises risk premia and simultaneously leads investors to believe the Fed will soon cut short-term interest rates. The increase in risk premia induces an immediate increase in long-term bond yields, while the expected drop in short rates induces an immediate decrease in these yields. In a Gaussian term structure model, a single parameter restriction equates these effects, leaving the current term structure—but not expected future term structures—unaffected by the news. More generally, factors that drive risk premia and expected short rates in opposite directions can have arbitrarily small effects on the cross section of yields, yet large effects on yield dynamics.

This theoretical result, although not well-known, can be inferred from the existing term structure literature. Duffee (2002) contains an example in which the physical and equivalent-martingale dynamics are driven by state vectors with different dimensions. But its implications for empirical work have not been recognized until this paper, and contemporaneous and independent work by Joslin, Priebsch, and Singleton (2009). We take this idea in different directions. In a nutshell, I use filtering to ask whether there are hidden factors. Their work assumes the existence of two hidden factors that are linear combinations of observed inflation and industrial production, and estimate the resulting model using both yield and macroeconomic data.

I estimate a five-factor Gaussian term structure model using monthly Treasury yields from 1964 through 2007. The model's risk premia dynamics are parsimonious: a single "risk premium factor" determines the one-month-ahead risk premia on all bonds. The population properties imply that expected excess bond returns are highly volatile and that up to half of their variation is attributable to a hidden component of the risk premium factor. For example, a one standard deviation change in this hidden component lowers the expected one-year-ahead short rate by about 90 basis points. It raises the expected annual return to a five-year bond in excess of the yield on a one-year bond by more than two percent. The importance of the hidden component increases as the return horizon decreases.

Some of the point estimates have large confidence bounds. There is also evidence that the model's population properties overstate the amount of predictability in excess returns. Nonetheless, the case for the importance of hidden components of risk premia is strong both statistically and economically. Investors' expectations of future short rates, as measured by surveys, are low relative to current rates when the hidden component of risk premia is high, as the model predicts. The hidden component negatively covaries with various measures of aggregate economic activity, but it is a linguistic stretch to call it countercyclical. Measures of macroeconomic activity explain only a small fraction of variation in the hidden factor, a result that sharply distinguishes the model estimated here from the structure imposed *a priori* by Joslin et al. (2009). Similarly, the ability of the hidden factor to forecast excess returns is not captured, in a regression sense, by other macroeconomic or financial variables.

The next section presents and motivates a framework of term structure models with hidden and nearly-hidden factors. Section 3 estimates specific models within this framework and documents the magnitude of the hidden component of the risk premium factor. The hidden component is linked to investors' expectations and macroeconomic activity in Section 4. Concluding comments are in Section 5.

2 The modeling framework

This section explains why some important determinants of the yield dynamics may be undetectable in the cross section. To make this point in the starkest terms, I build a model in which n factors are necessary to model term structure dynamics, but only $n_0 < n$ factors appear in yields.

The model follows much of the modern term structure literature by abstracting from standard economic concepts such as utility functions and production technologies. Instead, both the short rate and the nominal pricing kernel are functions of a latent state vector.

2.1 A Gaussian model

I use a standard discrete time Gaussian term structure framework. The use of discrete time is innocuous. The role played by the Gaussian assumption is discussed in Section 2.6. The continuously-compounded one-period interest rate is r_t. This rate is expressed per period—if a period is a month, $r_t = 0.01$ corresponds to twelve percent per year. Interest rate dynamics are driven by a length-n state vector x_t. The relation between the short rate and the state vector is

$$r_t = \delta_0 + \delta_1' x_t. \tag{1}$$

The state vector has first-order Markov dynamics

$$x_{t+1} = \mu + K x_t + \Sigma \epsilon_{t+1}, \quad \epsilon_{t+1} | x_t \sim N(0, I). \tag{2}$$

The state vector is latent, hence identifying restrictions are typically imposed in estimation.

The period-t price of a zero-coupon bond that pays a dollar at $t+m$ is denoted $P_t^{(m)}$. The corresponding continuously-compounded yield is $y_t^{(m)}$. Bond prices satisfy the law of one price

$$P_t^{(m)} = E_t\left(M_{t+1}P_{t+1}^{(m-1)}\right) \tag{3}$$

where M_{t+1} is the pricing kernel. The pricing kernel has the log linear form

$$\log M_{t+1} = -r_t - \Lambda_t'\epsilon_{t+1} - \frac{1}{2}\Lambda_t'\Lambda_t. \tag{4}$$

The vector Λ_t is the compensation investors require to face shocks to state vector. The price of risk satisfies

$$\Sigma\Lambda_t = \lambda_0 + \lambda_1 x_t, \tag{5}$$

which is the essentially affine form introduced in Duffee (2002). Bonds are priced using the equivalent martingale dynamics

$$x_{t+1} = \mu^q + K^q x_t + \Sigma\epsilon_{t+1}^q, \tag{6}$$

where the equivalent martingale parameters are

$$\mu^q = \mu - \lambda_0, \quad K^q = K - \lambda_1. \tag{7}$$

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

$$y_t^{(m)} = A_m + B_m' x_t, \tag{8}$$

where the scalar A_m and the n-vector B_m are functions of the parameters in (1) and (6). The closed-form expression for loadings of yields on factors is

$$\begin{aligned} B_m' &= \frac{1}{m}\delta_1'\left(I + K^q + (K^q)^2 + \cdots + (K^q)^{m-1}\right) \\ &= \frac{1}{m}\delta_1'\left(I - K^q\right)^{-1}\left(I - (K^q)^m\right). \end{aligned} \tag{9}$$

Gaussian models are often used to forecast bond returns. Define the excess log return to an m-period bond over j periods as the bond's log return in excess of the return to a j-period bond. The excess return expressed in yields is

$$xr_{t,t+j}^{(m)} = -(m-j)y_{t+j}^{(m-j)} + my_t^{(m)} - jy_t^{(j)}. \tag{10}$$

This return can be written in terms of the yield function (8) and state dynamics as

$$\begin{aligned} xr_{t,t+j}^{(m)} = & \left[mA_m - (m-j)A_{m-j} - jA_j + \left(mB'_m - (m-j)B'_{m-j} - jB'_j \right) E(x) \right] \\ & + \left(mB'_m - (m-j)B'_{m-j}K^j - jB'_j \right) (x_t - E(x)) \\ & - (m-j)B'_{m-j}\Sigma \left[\sum_{i=0}^{j-1} K^i \epsilon_{t+j-i} \right]. \end{aligned} \tag{11}$$

The first term in square brackets on the right is the excess return's unconditional mean. The second term is the conditional deviation from this unconditional mean and the third term is the return shock.

2.2 Information in the cross section

Absent specific parameter restrictions, the period-t state vector can be inferred from a cross section of period-t bond yields. Stack the yields on n zero-coupon bonds in the vector y_t^a. We can write this vector as

$$y_t^a = \mathcal{A}^a + \mathcal{B}^a x_t \tag{12}$$

where $\mathcal{A}^a$ is a length-n vector containing A_m for each of the n bonds and $\mathcal{B}^a$ is a square matrix with rows B'_m for each bond. In general, $\mathcal{B}^a$ is invertible. Put differently, element i of the state vector affects the n bond yields in a way that cannot be duplicated by a combination of the other elements. With invertibility, the term structure contains the same information as x_t. We can write

$$x_t = (\mathcal{B}^a)^{-1} (y_t^a - \mathcal{A}^a). \tag{13}$$

Since x_t follows a first-order Markov process, the term structure of yields also follows a first-order Markov process.

Although this result is derived here in a Gaussian setting, it applies more generally to the class of affine term structure models. The entire empirical literature on dynamic term structure models (setting aside the current paper and Joslin et al. (2009)) takes it for granted. For example, the handbook treatment of Piazzesi (2009) does not mention that $\mathcal{B}^a$ may not invertible. The next subsection explains why, from an empirical perspective, invertibility is a very useful property.

2.3 The role of invertibility in empirical analysis

Invertibility allows us to infer the dimension of the state vector n from properties of the cross section of yields. One method, introduced by Stambaugh (1988), studies the predictability of excess bond returns. He infers n by using a condition equivalent to (13): conditional expectations of excess bond returns are functions of n forward rates. This methodology remains at the leading edge of the literature through Cochrane and Piazzesi (2005). Another method to infer n is factor analysis of the unconditional covariance matrix of yields or differenced yields. Litterman and Scheinkman (1991) conclude three factors explain, in a statistical sense, all but a negligible fraction of the variation in the term structure. Duffee (2002) and Brandt and Chapman (2003) use this result and (13) to justify the choice of $n = 3$.

Equation (13) implies that maximum likelihood estimation of affine term structure models requires only a panel of n yields and the density function of the state vector.[1] In fact, Piazzesi (2009) defines likelihood-based estimation of affine models in terms of (13). Pearson and Sun (1994) are the first to exploit this result. Chen and Scott (1993) expand the panel's cross section to d yields by assuming that n linear combinations of yields are observed without error and $d-n$ are observed with error. In the special case of Gaussian models, maximum likelihood estimation is also feasible when all d yields are observed with measurement error. Yet even with Gaussian models, estimation is simplified considerably when factors are treated as linear combinations of yields. Cochrane and Piazzesi (2008), Joslin, Singleton, and Zhu (2009), and Hamilton and Wu (2010) are recent applications that use (13) to estimate Gaussian models.

Invertibility implies that only yields are necessary to estimate affine models, but it does not rule out the use of other data. Ang and Piazzesi (2003) introduced macroeconomic variables into Gaussian term structure models, leading to an explosion of macro-finance research. This literature is not designed to produce more accurate term structure models, but rather to explicitly link the term structure to its fundamental determinants, such as inflation and monetary policy.

Although invertibility is widely assumed and useful, it need not hold. I now consider special cases of the Gaussian framework where $\mathcal{B}^a$ has rank less than n, so that the state vector cannot be extracted from the term structure. An example illustrates the mathematics and the economic intuition.

[1] The likelihood function for discretely observed observations may need to be evaluated numerically.

2.4 A two-factor example

Consider a two-factor Gaussian model. Because the latent factors in this model can be arbitrarily rotated, the state vector can be transformed into the short rate and some other factor, denoted h_t for "hidden." For this rotation, the dynamics of the state vector are (explicitly indicating the elements of the feedback matrix)

$$\begin{pmatrix} r_{t+1} \\ h_{t+1} \end{pmatrix} = \mu + \begin{pmatrix} k_{11} & k_{12} \\ k_{21} & k_{22} \end{pmatrix} \begin{pmatrix} r_t \\ h_t \end{pmatrix} + \Sigma\epsilon_{t+1}. \tag{14}$$

When k_{12} does not equal zero, time-t expectations of future short rates depend on both r_t and h_t. Thus we can think of h_t as all information about future short rates that is not captured by the current short rate.

If investors are risk-neutral, the level of h_t necessarily affects the term structure through expectations of future changes in the short rate. But if risk premia also vary with h_t, the net effect of h_t on yields is ambiguous. The restriction adopted in this example is that changes in risk premia exactly cancel expectations of future short rates, leaving yields unaffected by h_t. Formally, the requirement is $k_{12}^q = 0$, or $k_{12} = \lambda_{1(12)}$. Then the equivalent martingale dynamics of the state are

$$\begin{pmatrix} r_{t+1} \\ h_{t+1} \end{pmatrix} = \mu^q + \begin{pmatrix} k_{11}^q & 0 \\ k_{21}^q & k_{22}^q \end{pmatrix} \begin{pmatrix} r_t \\ h_t \end{pmatrix} + \Sigma\epsilon_{t+1}^q. \tag{15}$$

A glance at (15) reveals that under the equivalent martingale measure, the short rate follows a (scalar) first-order Markov process. The loading of the m-period bond yield on the state vector is, from (9),

$$B_m = \begin{pmatrix} \frac{1}{m}\left(1 - k_{11}^q\right)^{-1}\left(1 - \left(k_{11}^q\right)^m\right) \\ 0 \end{pmatrix}. \tag{16}$$

Therefore the matrix $\mathcal{B}^a$ in (12) cannot be inverted because it has a column of zeros. The factor h_t is hidden, in the sense that it has no effect on the period-t term structure. Even if an econometrician knows the parameters of the model, she cannot infer h_t from the cross section of yields at t. Nor can h_t be backed out of the price of some other fixed-income instrument, such as bond options.

Although the factor does not affect yields, investors observe it. They take it into account when setting bond prices and forming expectations of future yields (or equivalently, future returns to holding bonds). For concreteness, consider the case $k_{12} > 0$. Then for fixed r_t, an increase in h_t raises investors' expectations of future short rates. For example, consider macroeconomic news, such as unexpectedly high GDP growth, that raises the likelihood of future tightening by the Federal Reserve. If investors' willingness to bear interest risk does not change with h_t, this news raises current long-maturity bond yields. But with the restriction $k_{12} = \lambda_{1(12)}$, investors accept lower expected excess bond returns. The change in willingness to bear risk offsets exactly the news about expected future short rates, leaving yields unaffected.

The functional relation between expected excess returns and h_t can be seen in the formula for the expected excess log return, from t to $t+1$, on a bond with maturity m at period t. The period-t expectation is

$$\begin{aligned} E_t\left(xr_{t,t+1}^{(m)}\right) &\equiv my_t^{(m)} - (m-1)E_t\left(y_{t+1}^{(m-1)}\right) - r_t \\ &= mA_m - (m-1)A_{m-1} \\ &\quad +(1-k_{11}^q)^{-1}\left[(1-(k_{11}^q)^m) - \left(1-(k_{11}^q)^{(m-1)}\right)k_{11} - 1\right]r_t \\ &\quad -(1-k_{11}^q)^{-1}\left(1-(k_{11}^q)^{(m-1)}\right)k_{12}h_t. \end{aligned} \tag{17}$$

The final term in (17) captures the dependence of expected excess returns on h_t.

In this example, the short rate follows a two-factor Markov process under the physical measure and a one-factor Markov process under the equivalent martingale measure. A single parameter restriction is required to generate this structure. Armed with the intuition of this example, it is straightforward to proceed to the more general case in which the short rate follows an n-factor Markov process under the physical measure and an n_0-factor Markov process under the equivalent martingale measure, where $n_0 < n$. As in the two-factor case, a single parameter restriction is required for each hidden factor.

2.5 A canonical hidden-factor model

Latent state vectors in affine term structure models are inherently arbitrary. Dai and Singleton (2000) describe how they can be translated and rotated without observable consequences. The eigenvalues of K^q in (6) are invariant to these transformations. This analysis adopts the transformations that produce the ordered real Jordan form of K^q advocated by Joslin, Singleton, and Zhu (2009), hereafter cited as JSZ. For simplicity, I focus on the case in which all eigenvalues are real and distinct.

As shown by JSZ, an arbitrary state vector can be rotated and translated such that the parameters of the equivalent-martingale dynamics (6) are

$$\mu_q = 0, \quad K^q = D,$$

where D is a diagonal matrix with the real distinct eigenvalues along the diagonal and Σ is lower triangular. Note that each element of the state vector follows a univariate first-order Markov process under the Q measure. Innovations among the elements can be correlated. In JSZ the factors are scaled by setting δ_1 in (1) to a vector of ones. Here it is important to scale the factors by setting the diagonal elements of Σ to a vector of ones.

With this rotation, element i of the state vector is hidden, in the sense that it does not affect contemporaneous yields, if the element does not appear in the short-rate equation (1),

$$\delta_{1(i)} = 0. \tag{18}$$

If this restriction holds for one or more i, we can sort the elements of the state vector into the length-n_0 vector f_t and the length-n_1 vector h_t,

$$x_t = \begin{pmatrix} f_t' & h_t' \end{pmatrix}',$$

where (18) holds for h_t and does not hold for f_t. Then the vector h_t is hidden. The logic behind (18) is straightforward. It implies that r_t depends only on f_t. Similarly, the short rate at $t + \tau$ depends only on $f_{t+\tau}$. Since each factor follows a univariate Markov process under the equivalent martingale measure, the period-t equivalent martingale expectation of $f_{t+\tau}, \tau = 1, \ldots,$ depends only on f_t. Therefore period-t yields depend only on f_t.

As in the two-factor case, physical dynamics of the short rate depend on the entire state vector. The parameters of the physical dynamics (2) are

$$\mu = \lambda_0, \quad K = D + \lambda_1.$$

Fixed-income investors require compensation to face shocks to f_t. This required compensation will vary with h_t if the necessary elements of λ_1 are nonzero. This channel also allows h_t to contain information about the evolution of the short rate that is not in f_t.

This derivation requires minor adjustments if the eigenvalues of K^q are not all real and distinct. Using the ordered real Jordan form, the reader can easily derive the result that factors corresponding to pairs of complex eigenvalues must be hidden in pairs. No such restriction applies to factors corresponding to repeated real eigenvalues. One or more can be hidden, although the normalization of JSZ requires that hidden factors be ordered prior to non-hidden factors.

No fixed-income instrument has a price innovation driven by innovations in h_t. Thus, although investors may require compensation to face shocks to h_t, no information about this compensation is available in the fixed-income market. Hence a canonical model of fixed income can normalize risk premia on h_t to zero, which is why the dynamics of h_t under the equivalent martingale measure equal those under the physical measure.

The final normalizations necessary for a canonical model are scaling normalizations. The non-hidden vector f_t is scaled (and its sign is determined) by normalizing $\delta_f = \iota$, a vector of ones. The hidden vector h_t is scaled by normalizing the diagonal elements of Σ_h to a vector of ones.

2.6 The role of the Gaussian setting

Section 2.5 shows that with an appropriate restriction on a term structure model, only a subset of factors of an n-dimensional state vector affect bond yields. Models exhibiting unspanned stochastic volatility (USV), as described in Collin-Dufresne and Goldstein (2002), can be described similarly. Here I clarify the relation between the approach here and the USV approach.

In this model, short rate dynamics are described by an n-factor Markov process under the physical measure and an $(n_0 < n)$-factor Markov process under the equivalent martingale measure. All n_0 factors that appear in the equivalent martingale process affect bond yields. Hence we can say that under the equivalent martingale measure, the term structure follows an n_0 factor Markov process. By contrast, the USV framework is concerned only with the equivalent martingale measure. The physical measure is not specified. Under the equivalent martingale measure of a USV model, the short rate is determined by a n-dimensional state vector that follows a Markov process. Bond yields nonetheless do not depend on all n factors. (Prices of some other fixed-income instruments will depend on all n factors.) Thus under the equivalent martingale measure, the term structure does not follow a Markov process.

The economic interpretations of the two sets of parameter restrictions differ substantially. In this model, variations in expected future short rates are offset by variations in risk premia. With USV, variations in equivalent martingale expectations of future short rates are offset by variations in the Jensen's inequality component of bond yields. Stochastic volatility is thus critical to USV models (hence the name of the model class), but does not appear here.

Although USV models appear to have little in common with the model here, they can provide an alternative mechanism driving a wedge between the factors driving dynamics of yields and those driving the cross section of yields. Set risk premia to zero so that physical and equivalent martingale measures coincide. Then n factors are necessary to capture yield dynamics, while n_0 factors affect bond yields. I do not pursue this approach because the parameter restrictions necessary in a USV model are very tight.

One reason I use the Gaussian framework is to avoid complications associated with stochastic volatility. Reconsider the two-factor example of Section 2.4. If the conditional covariance matrix of factor innovations is allowed to be linear in h_t (a discrete-time approximation to a square-root diffusion model), then the level of h_t affects bond yields even when $k^q_{12} = 0$. Variations in risk premia can offset variations in expected future short rates, but do not offset variations in the Jensen's inequality component of yields. This problem does not arise in the two-factor example if conditional variances are allowed to depend on the short rate instead of h_t.

2.7 From theory to practice

Equation (18) is a knife-edge restriction when we take this term structure model literally. If all of the elements of δ_1 differ from zero by an arbitrarily small amount, then the exact mapping from factors to n yields in (12) implies that all factors can be inferred from the cross section using (13). A corollary of this observation is that a factor is either completely observable or completely hidden; there is no middle ground.

The knife-edge nature would not be a concern if the precise restriction (18) could be motivated economically. However, is far-fetched from an economic perspective. It would be a remarkable coincidence if there is a factor for which variations in expected future short rates are *exactly* offset by variations in required expected returns. What, then, is the practical relevance of hidden factors?

The answer lies in the real-world imperfections of the bond market that drive a wedge between theoretical bond yields and observed yields. Equation (12) implies that the unconditional covariance matrix of $d > n$ bond yields has a rank no greater than n. It equals n in the standard case and n_0 when (18) holds. Yet in the data, sample covariance matrices of zero-coupon bond Treasury yields are nonsingular for even large d; say, greater than ten. One interpretation of this result is that n is large, perhaps even infinite, as in Collin-Dufresne and Goldstein (2003). But from a variety of perspectives, it is more appealing to view bond yields as contaminated by small, transitory, idiosyncratic noise.

This noise is generated from three sources. First, there are market imperfections that distort bond prices, such as bid/ask spreads. Second, there are market imperfections that distort payoffs to bonds (and thus distort what investors will pay for bonds), such as special RP rates. Third, there are distortions created by the mechanical interpolation of zero-coupon bond prices from coupon bond prices.

I model the noise as classic measurement error. A vector of d period-t yields on bonds with maturities $m_1, \ldots, m_d$ is expressed as

$$y^o_t = \mathcal{A} + \mathcal{B}x_t + \eta_t, \quad \eta_t \sim N(0, I\sigma^2_\eta) \tag{19}$$

where the superscript on the left side denotes "observed" and η_t is a vector of measurement errors. For simplicity, in (19) the measurement error for each yield has the same variance. Element i of vector $\mathcal{A}$ is A_{m_i} and row i of matrix $\mathcal{B}$ is B'_{m_i}.

Equation (19) cannot be pushed to its logical limits. Since the measurement error is uncorrelated across maturities and time, (19) suggests that using either more points on the term structure or higher frequency data eliminates the effects of noise. Instead, the specification should be viewed as an approximation to a world in which noise dies out quickly and is roughly uncorrelated across the widely-spaced maturities used in empirical analysis.

Measurement error eliminates the knife-edge nature of (18). If $\delta_{1(i)}$ is in the neighborhood of zero, factor i's tiny contemporaneous effect on yields will be lost in the noise contaminating observed yields. This raises considerably the economic plausibility of hidden factors. It is easy to tell stories in which news has opposite effects on expected future short rates and investors' required expected excess returns. For example, the Taylor (1993) rule and its variants (see, e.g., Clarida, Galí, and Gertler (2000)) suggest that good news about future output is also news that future short rates are likely to rise. If willingness to bear interest rate risk covaries positively with the business cycle, the immediate effect of such news on bond yields is ambiguous and might be very close to zero.

More generally, measurement error creates partially hidden factors. Some, but not all, of the information in such factors can be inferred from the cross section of yields. Therefore when evaluating the economic importance of hidden factors, the economically interesting question is not whether (18) strictly holds. Instead, we should focus on a quantitative question: how big is the gap between the information contained in the state vector and the information contained in linear combinations of observed yields?

A reasonable method to measure the gap is to examine variances of conditional expectations. Consider, predictions of excess log returns to an m-period bond over j periods. Investors form their forecast using the n-vector x_t. Alternatively, we can forecast using linear combinations of d observed demeaned yields,

$$z_t = \mathcal{P}\left(y_t^o - E\left(y_t^o\right)\right). \tag{20}$$

In the empirical work that follows, $\mathcal{P}$ is chosen so that z_t is the first n principal components of bond yields. Forecasts formed with x_t are more accurate than those formed with z_t and thus more volatile. Their relative accuracy is measured with the variance ratio

$$\frac{\operatorname{Var}\left(E\left(xr_{t,t+j}^{(m)}|z_t\right)\right)}{\operatorname{Var}\left(E\left(xr_{t,t+j}^{(m)}|x_t\right)\right)}. \tag{21}$$

If the ratio is close to one, then z_t effectively spans x_t from the perspective of calculating conditional expectations of excess returns. In this case, hidden factors are economically small. But if the ratio is substantially less than one, partially hidden factors are economically important.

2.8 Implications for term structure estimation

Section 2.7 argues that we should consider seriously the possibility that one or more term structure factors are partially hidden from the cross section of yields. How, then, should we estimate dynamic term structure models? One requirement is to build necessary flexibility into the model through the dimension of the state vector. At least three state variables are needed to describe the cross section of Treasury yields. Hence if there is some variable hidden from the cross section, a model without a minimum of four state variables is misspecified.

Given a sufficiently flexible model, there are two broad paths to follow. They differ primarily in the data used in estimation. The direction taken in this paper is to infer the presence of hidden factors from the dynamics of yields, which I call a "yields-only" approach. Alternatively, yield data can be augmented by other data that contain independent information about factors that drive yield dynamics, which I call a "yields-plus" approach.

Yields-only estimation of models with hidden factors can be done with filtering. Pennacchi (1991) introduces filtering into affine term structure estimation. The usual motivation, as noted in Piazzesi (2009), is to extract information about the period-t state vector from the entire period-t cross section, thus avoiding the ad hoc assumption that exactly n yields are observed without error. But filtering also uses dynamics to infer this vector. Intuitively, filtering is equivalent to learning by the econometrician. The period-t forecast error is produced by both true period-t shocks and the error in the econometrician's $t-1$ prediction of the $t-1$ state vector. The cross sectional pattern of the period-t forecast errors helps the econometrician revise her prediction of the state vector at $t-1$ and form her prediction of the state vector at t.

In adition, it may be possible to infer the period-t state vector from period-t observations of non-yield data. Recall that hidden factors have equal and opposite effects on expected future short-term interest rates and risk premia. Data that depend separately on these two components (or weight them differently) will reveal such factors. Perhaps the most obvious choice is survey data on interest rate forecasts, such as that used by Kim and Orphanides (2005).[2] In line with the literature's recent focus on macro-finance models, Joslin et al. (2009)

[2] Kim and Orphanides use a three-factor model, which does not give them the flexibility to capture both the cross section and potential hidden factors.

use inflation and output growth. They assume two hidden factors are linear combinations of these variables.

The tradeoffs between a yields-only and a yields-plus approach are straightforward. Estimation using additional data is a more powerful approach, but also at greater risk of misspecification. Holding the sample length constant, and under the maintained hypotheses that the additional data are functions only of state vector that drives bond yields, and the data reveal otherwise hidden factors, the yields-plus approach will produce more precise estimates of the term structure model. It is more accurate to infer period-t factors from direct observation of variables at t than from teasing them out of dynamics.

However, samples of survey data are shorter than samples of bond yields. Long time series of macro data are available, but the requirement that the macro variables are spanned by the variables that drive yields is often problematic. In particular, the relation between the macroeconomy and time-varying risk premia is an active area of research, but one with with few uncontroversial conclusions. The yields-only approach does not take a stand on the relation between the term structure and the macroeconomy, and thus avoids the possibility of misspecifying the relation.

The empirical analysis in this paper is a bit of a hybrid. The next section uses a long sample of yields to estimate term structure models. Section 4 links a hidden factor uncovered through this estimation to both a shorter sample of surveys of interest rate forecasts and long samples of macroeconomic data.

3 Empirical analysis

This section constructs and estimates a reasonably parsimonious five-factor Gaussian term structure model. Only the most important of the five shocks has a time-varying price of risk. Constructing this model so that the term "most important" makes sense and no-arbitrage is satisfied requires additional structure that is developed in the first subsection.

The other subsections are devoted to an in-depth examination of the estimated five-factor model. Some features of three-factor and four-factor models are also discussed to help understand the marginal contribution of additional factors. Basic features of the data and the estimated factor loadings are presented in Section 3.2. The heart of the empirical discussion is Section 3.3, which discusses properties of excess bond returns. The most important result is that up to half of the population variance in risk premia is orthogonal to the cross section of observed yields. Section 3.4 discusses estimation of the time series of the hidden component of risk premia using filtering and smoothing. One of its conclusions is that the amount of predictability of excess returns over the data sample is less than that implied by the population properties of the model. The final subsection draws a related conclusion. It finds that although a five-factor model fits bond yield dynamics much better than a four-factor model, which in turn fits these dynamics much better than a three-factor model, the models are all equally accurate in forecasting future yields.

3.1 The estimated functional form

This subsection describes the restrictions on risk premia dynamics that are built into the model. One restriction that is *not* imposed is (18), which is necessary for an exactly hidden factor. In line with the discussion in Section 2.7, all factors are allowed to affect contemporaneous yields. The data tell us whether such effects are indistinguishable from measurement error. The economic importance of partially hidden factors is evaluated using the variance comparison in Section 2.7.

The model of risk premia dynamics imposed here is recommended by Duffee (2010). He finds that when fitted conditional Sharpe ratios are constrained to reasonable values, expected excess bond returns are largely explained by a specification in which investors require compensation to face shocks to the level and slope of the term structure. The compensation for slope risk is fixed, while the compensation for level risk varies with the shape of the term structure.

This risk premia specification is similar in spirit to Cochrane and Piazzesi (2008), who allow only level risk to be priced. There are two major differences with their approach. First, Cochrane and Piazzesi do not impose the no-arbitrage requirement that the yield-based factors driving the term are priced consistently with the model. Imposing that requirement is straightforward and done elsewhere, such as Ang, Piazzesi, and Wei (2006). Second, Cochrane and Piazzesi define level risk using sample principal components of observed yields. Their procedure does not ensure that level risk, as implied by model dynamics, is the priced risk. Here, principal components are defined using the model's dynamics.

In much of the term structure literature, the terms level and slope are based on principal components analysis of the covariance matrix of yields. For modeling convenience, here they are based on the covariance matrix of shocks to yields. The state vector is identified such that the shock to element i of the state vector x_t is the i^{th} principal component of the covariance matrix of shocks to yields. Factor dynamics under the equivalent-martingale measure are a special case of (6),

$$x_{t+1} = K^q x_t + \Omega^{1/2}\epsilon^q_{t+1}, \tag{22}$$

where diagonal matrix $\Omega^{1/2}$ contains the standard deviations of these orthogonal shocks. Appendix 1 describes the restrictions on K^q and Ω that produce this identification.

The factor dynamics under the physical measure are

$$x_{t+1} = \lambda_0 + (K^q + \lambda_1) x_t + \Omega^{1/2} \epsilon_{t+1}. \tag{23}$$

The vector λ_0 and the matrix λ_1 determine the dynamics of risk compensation. The specification here is

$$\lambda_0 = \begin{pmatrix} \lambda_{0(L)} & \lambda_{0(S)} & 0_{1\times(n-2)} \end{pmatrix}',$$

$$\lambda_1 = \begin{pmatrix} \lambda'_{1(L)} \\ 0_{(n-1)\times n} \end{pmatrix},$$

where $\lambda_{0(L)}$ and $\lambda_{0(S)}$ are scalars corresponding to level and slope risk, and $\lambda_{1(L)}$ is a length-n vector. Early work on risk dynamics, such as Duffee (2002), finds that variation in the price of level risk is necessary to capture the failure of the expectations hypothesis. Prior to Cochrane and Piazzesi (2005), this variation was typically captured by linking the price of level risk to the slope of the term structure. Cochrane and Piazzesi show that risk premia are better captured by an affine function of many points on the yield curve, hence the vector $\lambda_{1(L)}$ is unrestricted. Duffee (2010) notes that unconditional Sharpe ratios are higher for short-maturity bonds than long-maturity bonds, a pattern that is consistent with nonzero average prices for both level and slope risk. Hence $\lambda_{0(L)}$ and $\lambda_{0(S)}$ are free parameters.

Reasonable arguments support more flexible risk premia specifications. For example, standard ICAPM logic says that if $\lambda_{1(L)}$ is a free parameter vector, then investors should require compensation to face shocks to all factors. The reason is that each of these shocks change investment opportunity sets through their effects on the price of level risk. However, given our current understanding of term structure dynamics, the overfitting problem appears to swamp the advantages of more flexible functional forms.

This specification implies that a single linear combination of the state vector determines the compensation investors demand to face fixed-income risk from t to $t+1$. Call this linear combination the "risk premium factor," given by

$$RP_t \equiv \lambda'_{1(L)} x_t. \tag{24}$$

This specific linear combination contains all information relevant to predicting one-step-ahead excess returns. It is worth noting that this result does not generalize to predictions of j-step-ahead excess returns. Period-t expectations of multiperiod excess returns depend on period-t forecasts of the risk premium factor at $t+1$ and beyond. In general, the risk premium factor will not follow a univariate autoregressive process, thus the period-t value of the factor is not the best predictor of its future values.

3.2 Data, estimation technique, and factor loadings

The models are estimated using yields on zero-coupon Treasury bonds with maturities of three months, and one through five years. These data are from the Center for Research in Security Prices (CRSP). The yield on a three-month Treasury bill is from the Riskfree Rate file (bid/ask average). Artificially-constructed yields on zero-coupon bonds with maturities of one, two, three, four, and five years are from the Fama-Bliss file. Yields are observed at the end of each month from January 1964 through December 2007. An on-line appendix discusses how the results are affected when a ten-year bond yield is included in estimation. Briefly, because of a type of inherent misspecification in affine models, the resulting model is wildly unrealistic. The start of the sample coincides with that of Cochrane and Piazzesi (2005). The end predates the widespread financial crisis.

Estimation is with the Kalman filter, which produces correct conditional means and covariances in a Gaussian setting. The estimated parameter vector is described in Appendix 1. These parameters are a hybrid of parameters of the JSZ rotation and the principal components rotation. To simplify discussion of the estimated models, the principal components rotation (22) and (23) is used in the remainder of this section.

Table 1 reports the point estimates of the five-factor model. There are 44 nonzero parameters in the table, although there are only 29 free parameters. There are 15 restrictions built into these parameters from the requirement that the factors correspond to principal components. Most of the parameters are not intuitive, which is why the remainder of this section looks at the estimated model from a variety of more meaningful perspectives. Point estimates for the three-factor and four-factor models are available on the author's website.

Standard errors are in parentheses. They are constructed from Monte Carlo simulations. Each simulation begins by generating randomly 528 months of yields from the estimated model. The model is reestimated with maximum likelihood using these data and the parameter estimates are stored. This procedure is repeated 1000 times to construct the standard errors, as well as the confidence bounds displayed in Figure 1. The covariance matrix of the 44 parameter estimates has rank 29.

The standard errors and confidence bounds should be treated with caution. They are correct assuming the model is specified accurately. However, there is overwhelming evidence that bond returns, like many other returns observed in financial markets, are conditionally heteroskedastic. Hence extreme observations in the sample are likely to be less informative about the data-generating process than the Gaussian model implies.

The main objective of this empirical analysis is to understand how the factors affect both current yields and expected future yields. Therefore, to limit the size of this paper, an on-line appendix describes various unconditional properties of the model, such as means and standard deviations of yields. The main conclusion is that the model does a good job reproducing the relevant properties of yields used in estimation.

Figure 1 displays the loadings of observed bond yields on the factors. The solid lines are the analytic loadings from Equation (9), scaled by the standard deviations of factor shocks. These standard deviations are the diagonal elments of $\Omega^{1/2}$. Hence the vertical axis is the response of the yield curve to a one standard deviation factor shock, in annualized basis points. Corresponding sample values are also displayed. They are constructed in two steps. Fitted values of the state vector are first estimated with a Kalman smoothing algorithm. Then observed bond yields are regressed on the fitted state vector. The diamonds and circles are the regression coefficients, also scaled by the diagonal elements of $\Omega^{1/2}$. The dashed lines are 95 percent confidence bounds, produced by the Monte Carlo simulations.

There are two main conclusions to draw from this figure. First, the model reproduces the standard result that three factors drive almost all of the variation in yield innovations. A glance at the vertical scales is sufficient evidence. One standard deviation shocks to the first three factors correspond to yield innovations of 20 to 40 basis points for maturities up to five years. Corresponding values for the fourth and fifth factors are less than five basis points. Note that the loadings on the fifth factor increase substantially outside of the maturity range used to estimate the model. They are in the neighborhood of ten basis points for a ten-year yield. The on-line appendix discusses this property in greater detail.

Second, the model reproduces almost perfectly the sample loadings of observed bond yields on the factors, at least for those maturities used to estimate the model (between three months and five years). Moreover, the confidence bounds are practically indistinguishable from the point estimates over this maturity range.

One property of the figure deserve a detailed discussion: the shape of the loadings on the first principal component. This principal component is commonly called the "level" component. The loadings, displayed in Panel A, are not close to horizontal in the figure. The shape might trouble some readers who are more familiar with near-horizontal level factors. However, recall that the principal components decomposition used here is for shocks to yields. A model-implied principal components decomposition of yields (not displayed) has a near-horizontal first principal component.

For the purposes of building a model with risk premia defined in terms of principal components, using principal components of shocks is a better choice than using levels. One advantage is tractability, as discussed in Appendix 1. Another advantage is that the factor shocks are orthogonal by construction. Principal components of yields are unconditionally orthogonal, but their shocks are not. When factors are identified using principal components defined by the unconditional covariance matrix of yields, it does not make much sense to assume that the price of risk of shocks to one principal component varies over time while the price of risk of shocks to another principal component is time-invariant, perhaps fixed at zero. The risks are all correlated.[3]

3.3 A partially hidden risk premium factor

This subsection discusses population properties of excess bond returns implied by the five-factor model. The results are easy to summarize. At the point estimates, excess returns are highly predictable using full information; in other words, using the true state vector. A substantial fraction of this predictability (half, at the monthly horizon) is hidden from the cross-section of yields. Confidence bounds on the total amount of predictability are extremely large, but bounds on the fraction that is hidden are much tighter. The latter bounds allow us to conclude that hidden risk premia are an important feature of term structure dynamics.

The evidence is in Table 2. Recall the expression (11) for excess log bond returns in the Gaussian model. The table summarizes features of these returns to a five-year bond. The choice of five-year bond is arbitrary. Since a single factor drives risk premia, results for other maturities are similar. Point estimates and 95 percent confidence bounds are reported for the five-factor model. Point estimates for three-factor and four-factor versions are included to put the results for the five-factor model into context.

The first three columns contain population means of excess returns, population variances of true excess returns, and population variances of observed excess returns, which include measurement error. To show how to read the table, consider the column of means. Point estimates imply that the mean monthly excess return is 12 basis points per month, or about 1.25 percent per year. This exceeds the mean annual excess return of 0.84 percent because annual excess returns are calculated relative to the yield on a one-year bond, not a rolling position in one-month bonds. Ninety-five percent confidence bounds on these means are so large that they include negative mean returns. Put differently, there is so much sampling uncertainty in the model's dynamics that we cannot be sure the mean yield curve is positively sloped. Point estimates of the four-factor model produce similar means. The estimated three-factor model has much higher mean excess returns because estimates for that model imply a steeper unconditional yield curve.[4]

How much of the variation in true excess returns is forecastable? The answer depends on what information is used to predict returns. The model treats the state vector as observable by investors. Investors therefore know the second term on the right of (11), which is the predictable component of excess returns from t to $t+j$. According to the fourth column,

[3]Thanks to Stijn Van Nieuwerburgh for emphasizing this point to me.

[4]Detailed properties of the three-factor model are not reported in any table.

the variance of this term for monthly returns is 0.48 percent squared. Put differently, more than 13 percent of the total variance of excess returns from t to $t+1$ is predictable given the month-t state vector. The R^2 at the annual horizon is 45 percent. A glance at the table shows that the five-factor model implies much more predictability than do the three-factor and four-factor models.

Predictability of this magnitude seems implausible on economic grounds. One standard deviation increases in the conditional means are about 70 basis points and 4.3 percent at the monthly and annual horizons respectively. Since unconditional mean excess returns are quite low, these point estimates imply that conditional expected excess returns to a five-year bond are negative almost than 45 percent of the time. The predictability is also statistically unreliable. The low range of the 95 percent confidence bounds on predictable variances reduce the standard deviations of conditional expectations by about 40 percent. Thus there is strong statistical evidence of predictability, but we should not lean heavily on the the point estimates of extremely high predictability.

An econometrician does not observe the state vector, and thus must condition on weaker information. A common choice of conditioning information is the month-t cross section of observed yields, which are contaminated with measurement error. As mentioned in Section 2.7, the assumed independence of errors across maturities implies that the econometrician can completely avoid measurement error problems by observing an arbitrarily large number of yields. I avoid taking the model too literally and instead assume that the econometrician observes only those yields that are used to estimate the model. They are described in Section 3.2. The cross section of yields is then summarized with principal components.

From (19), the population covariance matrix of d observed yields is

$$\operatorname{Var}(y_t^o) = \mathcal{B}\operatorname{Var}(x_t)\mathcal{B}' + I\sigma_\eta^2. \tag{25}$$

Define the vector z_t as the first n principal components of y_t^o, based on (25). Using the notation of (20), the matrix $\mathcal{P}$ contains the loadings on y_t^o of these principal components. The variance of excess returns conditioned on z_t is, from (11),

$$\begin{aligned}&\operatorname{Var}\left(xr_{t,t+j}^{(m)}|z_t\right) = \\ &\left(mB_m' - (m-j)B_{m-j}'K^j - jB_j'\right)\operatorname{Var}(x_t|z_t)\left(mB_m' - (m-j)B_{m-j}'K^j - jB_j'\right)'.\end{aligned} \tag{26}$$

The variance on the right side of the equation is calculated in Appendix 2.

The penultimate column in Table 2 reports the population variance (26) for the observed yields used in estimating the model. The final column reports the variance ratio (21), which is the variability in excess returns conditioned on principal components relative to the variability conditioned on the state vector. According to the estimated five-factor model, the state vector contains much more information about future excess monthly returns than does the cross section of yields. The variance ratio (21) is only 0.53. Therefore close to half of the information in the state vector is hidden from the cross section. Even at the upper bound of the 95 percent confidence interval, the variance ratio is only 0.68. Hence there is strong statistical evidence that a large component of risk premia is hidden.

A closer look at this result is warranted. Using the model's notation described in Section 3.1, the one-period excess return can be expressed as

$$xr_{t,t+1}^{(m)} = \phi_{0,m} + \phi_{1,m} RP_t + \phi_{2,m}' \Omega^{1/2} \epsilon_{t+1}, \tag{27}$$

$$\begin{aligned}
\phi_{0,m} &= -(m-1)\left(B_{(L)m-1}\lambda_{0(L)} + B_{(S)m-1}\lambda_{0(S)}\right) - \frac{1}{2}(m-1)^2 B_{m-1}' \Omega B_m, \\
\phi_{1,m} &= -(m-1)B_{(L)m-1}, \\
\phi_{2,m} &= -(m-1)B_{m-1}.
\end{aligned} \tag{28}$$

In (28), the first two elements of the yield loading B_{m-1} are subscripted with (L) and (S), for level and slope. The scalar $\phi_{1,m}$ is the exposure of the bond's log price to shocks to the level of the term structure. Recall that the risk premium factor, defined in (24), captures time-varying compensation for these shocks. Therefore the second term on the right of (27) drives variations in the conditional expectation. The vector $\phi_{2,m}$ is the exposure of the bond's log price to shocks to the state vector, thus the third term is the return shock.

The risk premium factor is linked to the state vector through the vector of loadings $\lambda_{1(L)}$. Table 1 shows the risk premium factor loads heavily on the fourth and fifth elements of the state vector. As discussed in Section 3.2, one-standard-deviation shocks to these elements correspond to shocks to yields of less than five annualized basis points, which is less than the standard deviation of measurement error. Hence the information in these elements is obscured in the cross section. Shocks to these factors are less persistent than shocks to the first three factors. This explains why the variance ratio of Equation (21) is higher at the annual frequency, as reported in Panel B of Table 2, than it is at the monthly frequency.

Not surprisingly, a model with only three factors cannot reproduce the result that much of the variation in risk premia is hidden. The loading of yields on each of the factors are too large to be washed out by measurement error. Panel A of Table 2 shows that with a three-factor model, three principal components capture 97 percent of the predictability of the state vector. Somewhat surprisingly, the same 97 percent value shows up with the four-factor model, using four principal components. Section 3.5 explains the role of the fourth factor in the four-factor model.

3.4 Inferring bond risk premiums

Since the cross section of yields misses much of the information about conditional expected excess returns, how should an econometrician form forecasts? Because the Kalman filter uses information from dynamics, it produces more accurate estimates of the factors than does a purely cross-sectional approach. Two methods of extracting information from dynamics are embedded in the Kalman filter. One looks forward, using filtered estimates of the state at t to forecast returns from t to $t+j$. Another looks backwards from the end of the sample, interpreting history using smoothed estimates of the period-t state.

This subsection discusses how an econometrician should use these filtering and smoothing to infer risk premia over a particular sample. One of the conclusions is that, according to the estimated five-factor model, the sample period 1964 through 2007 was somewhat anomalous. In the sample, realized shocks to excess returns were negatively correlated with conditional expectations of these returns. This pattern drives a wedge between model-based forecasts and regression-based forecasts of excess returns.

When conditioning on factors estimated with the Kalman filter, we can form two types of expectations about future excess returns. If our goal is to better understand investors' expectations, we want to focus on an econometrician's expectations of investors' period-t expectations of excess returns. To simplify the discussion, focus on the single period return horizon. Using (27), investors' period-t expectations, which are conditioned on the true state variable, are

$$E\left(xr_{t,t+1}^{(m)}|x_t\right) = \phi_{0,m} + \phi_{1,m}RP_t. \tag{29}$$

Denote filtered and smoothed estimates of the state vector by $\hat{x}_t^f$ and $\hat{x}_t^s$ respectively. Since the risk premium factor is affine in the state vector, the econometrician's expectation of (29) conditioned on either filtered or smoothed factors is

$$\begin{aligned} &E\left[E\left(xr_{t,t+1}^{(m)}|x_t\right)|\hat{x}_t^l\right] = \phi_{0,m} + \phi_{1,m}\widehat{RP}_t^l, \\ &\widehat{RP}_t^l \equiv E\left(RP_t|\hat{x}_t^l\right) = E\left(\lambda_{1(L)}' x_t|\hat{x}_t^l\right) = \lambda_{1(L)}'\hat{x}_t^l, \qquad l \in \{f,s\}. \end{aligned} \tag{30}$$

Equation (30) uses the Kalman filter property that the expectation of the state conditioned on the filtered (smoothed) estimate equals the filtered (smoothed) estimate.

Now define variance ratios similar to forecast regression R^2s. The numerator is the variance of the conditional expectation of excess returns and the denominator is the total variance. Using (30), the ratios are

$$\frac{\phi_{1,m}^2 \mathrm{Var}\left(\widehat{RP}_t^l\right)}{\phi_{1,m}^2 \lambda_{1(L)}' \mathrm{Var}(x_t) \lambda_{1(L)} + \phi_{2,m}' \Omega \phi_{2,m} + (m^2 + (m-1)^2 + 1^2)\sigma_\eta^2}, \qquad l \in \{f, s\}. \tag{31}$$

The three terms in the denominator of (31) are the unconditional variance of the conditional mean, the variance of the true return shock, and the variance of measurement error's contribution to the observed return shock. This denominator differs from the denominator of (21), which excludes variance owing to measurement error. Here the denominator is written in terms of what the econometrician observes, not what investors observe.

Population variance ratios for monthly excess log returns to a five-year bond are displayed in Panel A of Table 3, in the columns labeled "Optimal filtering" and "Optimal smoothing." There is no analytic expression for the ratios. They are constructed with extremely long Monte Carlo simulations, where the econometrician is assumed to know the true parameters because she has an infinite time series of data available. For comparison, the full-information R^2 calculated from Table 2 is also reported. Recall that the full-information R^2 uses the variance of true excess returns in the denominator instead of the variance of observed excess returns.

For the five-factor model, the ratios using filtered and smoothed returns are 8.7 and 10.5 percent respectively. These ratios are substantially below the full-information value of 13.2 percent. Nonetheless, the Kalman filter produces more accurate forecasts of excess returns than does the cross section of yields. For the cross section, the corresponding value is 6.7 percent.

Smoothed estimates have more information than filtered estimates because smoothing infers the period-t state using, in part, return realizations in $t+1$ and beyond. If, say, excess returns in $t+1$ are unexpectedly high given filtered estimates of the period-t state, the risk premium factor at t was likely higher than the filtered estimate implied. Smoothing incorporates this information.

This property of smoothing leads to the second type of expected excess return conditioned an econometrician's information. Instead of forecasting investors' expectations, forecast excess returns directly with

$$E\left(xr_{t,t+1}^{(m)} | \hat{x}_t^l\right), \qquad l \in \{f, s\}. \tag{32}$$

With filtering, an application of iterated expectations to (30) produces (32); they are equivalent. With smoothing, iterated expectations does not apply because smoothed estimates depend on information revealed subsequent to period t. Put differently, the smoothed estimate of the risk premium factor is positively correlated with the return innovation in (27).

From an economic standpoint, the expectation (32) is not particularly interesting once we have (30), because it does not give us any additional insight into investors' beliefs. Nonetheless, it is useful to calculate in order to see whether it is a reasonable substitute for (30) in empirical work. It can be calculated with the predictive regression

$$xr_{t,t+1}^{(m)} = b_0 + b_1 \widehat{RP}_t^s + e_{t,t+1}^{(m)} \tag{33}$$

where orthogonality is imposed between the smoothed risk premium factor and the residual. The column in Table 3 labeled "OLS using smoothed" reports the R^2 for this regression. (For completeness, rsults are also reported for the regression using the filtered risk premium factor.) For the estimated five-factor model, the R^2 of 15.4 percent substantially exceeds the more relevant variance ratio reported in the "Optimal smoothing" column. It is not a good substitute. In fact, the R^2 even exceeds the full-information R^2. The reason, of course, is that "full information" refers to information available at t, while smoothed OLS uses information available after t.

The empirical properties discussed so far are population properties. It is also important to take a look at the 1964 through 2007 sample, especially because there are some substantial differences between the sample and population properties. Deriving the in-sample equivalents to the variance ratio (31) and the regression (33) is slightly complicated because the data sample does not contain monthly returns to a five-year bond. They are approximated by first generating filtered and smoothed estimates of the state vector for each month in the sample. Plugging these estimates into (23) produces filtered and smoothed versions of innovations to the state vector, which are labeled symmetrically with the estimated state vectors. The sample ratio that corresponds to (31) is

$$\frac{\phi_{1,m}^2 \hat{\text{Var}}\left(\widehat{RP}_t^l\right)}{\hat{\text{Var}}\left(\phi_{1,m}\widehat{RP}_t^l + \phi_{2,m}'\Sigma^{1/2}\hat{\epsilon}_{t+1}^l\right) + (m^2 + (m-1)^2 + 1)\sigma_\eta^2} \tag{34}$$

Hats on variances denote sample variances. The numerator is the sample counterpart to the numerator of (31). The first term in the denominator is the sample variance of the fitted true excess return, which is the sample counterpart of the first two terms in the denominator of (31). In (31), two variance terms are used because the state x_t and the innovation ϵ_{t+1} are treated as orthogonal in population. In (34) their sample correlation affects the sample variance of returns. The final terms in the denominators of (31) and (34) are identical. Because monthly excess returns are not actually observed in the sample, (34) assumes that the measurement errors in the sample have the properties of measurement error in the population.

Panel B of Table 3 reports the ratios (34), as well as R^2s for sample versions of (33). The R^2s reported in the "OLS" columns are modified versions of R^2s from regressions of fitted true excess returns in month $t+1$ on the estimated state vector in month t. The modification is an adjustment for measurement error following (34). For future reference, the tabel also reports log likelihoods for each of the estimated models.

The main result to take from Panel B is that for the five-factor model, all but the "Optimal filtering" sample ratio differ substantially from their corresponding population ratios in Panel A. The reason for the discrepancies is that in the sample, filtered month-t conditional expectations of excess returns are negatively correlated with month $t+1$ return shocks. Their sample covariance is about -0.3 of the sample variance of conditional expected returns. Hence in the sample, these conditional expectations are excessively volatile. In-sample predictive regressions using the filtered states are much less volatile because they impose in-sample orthogonality between predictions and residuals. Hence the predictive regression R^2 is only 0.043, compared with the optimal filtering variance ratio of 0.085. The same discrepancy affects in-sample results using smoothed states.

3.5 Three, four, and five factor models

How well do the three-factor, four-factor, and five-factor models forecast future bond yields? How well do they fit the cross section of yields? In a principal components rotation, the first three factors are typically level, slope, and curvature; what is the fourth factor? These questions are answered here. This subsection summarizes a detailed examination, most of which is not reported here owing to space constraints.

The most important and surprising result is that for any practical purpose, the forecasting abilities of the three models are equivalent. Differences among root mean squared forecasting errors are fractions of a basis point. The evidence is in the first two panels of Table 4. Forecast errors are defined as the yield at month $t+j$ less the Kalman filter forecast at t, which uses the filtered estimate of the state vector. They are expressed in annualized basis points. The table reports root mean squared (RMS) forecast errors at the one and twelve month horizons. At the one month horizon, RMS forecast errors are generally lower for the four-factor model than the three-factor model, although the magnitudes are tiny—around a tenth of a basis point. The five-factor model produces slightly smaller one-month-ahead RMS forecast errors at some maturities, and slightly higher RMS forecast errors at other maturities. At the twelve month horizon, it does a noticeably better job forecasting the three-month yield, and a noticeably worse job forecasting the five-year yield.

Although the forecasting performances are almost identical, the log-likelihoods increase substantially as the number of factors increases. Panel B of Table 3 shows that moving from a three-factor model to four and five-factor models raises the log-likelihood by 240 and an additional 50. The reason is that the higher dimensional models have, paradoxically, more parsimonious explanations of forecast errors. Recall that maximum likelihood rewards low-dimensional models of errors, through the determinant of their conditional covariance matrix. For example, a model that can explain forecast errors as the realization of a single shock will have a higher likelihood than a model that can explain the same forecast errors as the realization of two independent shocks.

Why are higher-dimensional models more parsimonious? First consider the four-factor model relative to the three-factor model. In four-factor model, the fourth factor is a curvature factor that is centered at a different maturity than the third factor. The third and fourth factors in the five-factor model behave similarly to those in the four-factor model, thus we can look at Figure 1 to see how they behave. The third factor is curvature centered at about the one-year maturity. In other words, yields on bonds with maturities around one year move away from yields at other maturities The fourth factor is curvature centered at two to three years. Yields on bonds with these maturities move away from yields on bonds with shorter and longer maturities.

The additional curvature factor produces smaller cross-sectional fitting errors. The third panel of Table 4 reports RMS cross-sectional errors in basis points of annualized yields. They are calculated using filtered estimates of the state vector. The fit of the four-factor model is about one to two basis points better than the fit of the three-factor model. These cross-sectional errors affect the likelihood through one-month-ahead forecast errors. Forecast errors that the three-factor model attributes to independent measurement error across six bonds are transformed by the four-factor model to a shock to the fourth factor—a lower-dimensional explanation.

The improvement associated with the five-factor model has little to do with cross-sectional fit. Without getting into gory details, the five-factor model attributes part of the month-$(t+1)$ forecast error as error in filtering the risk premium factor from t. This filtering error picks up part of what a four-factor model would attribute to a combination of independent shocks to level, slope, and two types of curvature.

The evidence in Table 4 raises an obvious question. Why does the five-factor model not forecast better than the other models? The population properties of these models, summarized in Tables 2 and 3, imply that the five-factor model is better than a lower-dimensional model at predicting excess bond returns. More accurate predictions of excess returns correspond to more accurate forecasts of future yields. The short answer is there is a mismatch between some of the population and sample properties, which suggests misspecification.

One mismatch, the sample correlation between conditional expectations of excess returns and return shocks, is mentioned in Section 3.4. Another mismatch is between the model-implied covariance matrix of factor innovations and the sample covariance matrix of filtered factor innovations. The latter is noticeably larger. Maximum likelihood chooses a relatively small covariance matrix to help fit the properties of conditional excess returns. Larger elements of the covariance matrix correspond to greater risk, for which investors must be compensated. Holding everything else constant, both average risk premia and the volatilities of risk premia will increase if elements of the covariance matrix increase.

This discussion leads to a cautionary conclusion. Researchers must be extremely careful drawing inferences from model-implied properties of term structure models. The models are unavoidably misspecified, and there is no straightforward methodology to detect where misspecification shows up in a model's parameters. If a researcher is interested in a particular feature of an estimated model, the researcher should obessively check a wide variety of that feature's implications against properties of the data.

4 A macroeconomic link?

The Holy Grail of the term structure literature is a testable, intuitive model linking yields to fundamental macroeconomic forces. This section argues that we are not close to finding it. There is no economically significant link between the five-factor model's hidden component of risk premia and a variety of macroeconomic time series. Similarly, little of the risk premia factor's forecasting power for excess returns is captured by macroeconomic instruments. The news is not entirely discouraging. There is strong statistical evidence that the hidden component of risk premia negatively covaries with aggregate economic activity. However, measures of macroreconomic activity do not capture the predictive power of the risk premium factor for excess returns.

4.1 Expectations of future short-term interest rates

I begin by describing the relation between the risk premium factor and the term structure. Assume that the risk premium factor is one standard deviation away from its mean.[5] What

[5]The sign of the risk premium factor is arbitrary. At the five-factor model's point estimates, an increase in RP_t corresponds to a decrease in expected excess returns. To simplify the discussion of the empirical results in this section, I reverse the sign, so that when I refer to an increase in the risk premium factor, I mean that expected excess returns are increasing.

does this tell us about period-t yields? What does it tell us about future short rates? These questions are answered by computing expectations conditioned on the risk premium factor. The relevant linear algebra is in Appendix 2. Results are displayed in Panels A and C of Figure 2. To help put a one standard deviation change in the risk premium factor into context, note that it corresponds to an increase of about 70 basis points in the expected excess monthly return to a five-year bond.

The figure shows that on average, an increase in expected excess returns is accompanied by an increase in the slope of the term structure. Short rates drop by 40 to 50 basis points, while long rates rise by about 35 basis points. The drop in short rates is expected to accelerate, with short rates dropping another 80 basis points over the next year. I emphasize "average" because the risk premium factor is determined by the entire state vector. The same value of the risk premium factor can be produced with a variety of realizations of the state vector. Each realization corresponds to a different term structure shape and a different path of expected future short rates.

The same values can be constructed for a one standard deviation change in the hidden component of the risk premium factor. Define the hidden component as the part unspanned by the vector z_t of n principal components of observed yields. Denote this hidden component with a tilde, as in

$$\widetilde{RP}_t \equiv RP_t - E(RP_t|z_t).$$

Again, the relevant linear algebra is in Appendix 2. Panel B of Figure 2 is the projection of contemporaneous yields given a one standard deviation change in the factor. This change is equivalent to an increase in expected excess monthly returns to a five-year bond of 47 basis points.

Consistent with the notion of "hidden," the yield curve has almost no reaction to the hidden component. By contrast, short rates are expected to drop by almost a full percentage point over the next year. This is (almost) tautological; if a bond's yield does not change, the only way excess returns can be high is if short rates are low.

The remainder of this section links the risk premium factor, and its hidden component, to data other than bond yields. Section 3.4 describes filtered and smoothed estimates of the factor. Define the estimated hidden component as the residual from regressing the smoothed factor on the first five principal components of observed bond yields,

$$\widehat{RP}_t^s = b_0 + b'\hat{z}_t + \widetilde{RP}_t^s.$$

A skeptical reader may be concerned that this hidden component is spurious; it is simply an ex-post overfitting of the sample. According to this view, when the hidden component implies high expected excess returns at t, investors at that time did not actually anticipate them. The Kalman filter smoothing overfits subsequent high realizations of excess returns, attributing part of them to a hidden factor.

To address this concern, I first confirm that investors' time-t expectations of future short rates are related to time-t estimate of the hidden component. At the end of the first month of every quarter since 1981Q3, participants in the Survey of Professional Forecasters are asked for their forecasts of the average level of the three-month Treasury bill during each of the next four quarters. Denote the quarter-t mean survey forecast of the three-month bill in quarter $t+j$ less the quarter-t bill yield as SPF_EXPECT(t,j). To align the bill yield with the survey timing, the quarter-t yield is defined as the three-month yield as of the end of the first month in the quarter. The continuously compounded yield from CRSP is converted to a discount basis to match the survey's yield convention.

Following the timing convention of yields, I define the quarterly value of the hidden component as the value of $\widetilde{RP}_t^s$ of the end of the first month in the quarter. To simplify interpretation of the estimated regression coefficients, this factor is normalized by its sample standard deviation. The sample period is 1981Q3 through 2007Q4.

The regression is

$$\widetilde{RP}_t^s = b_0 + b_1 \text{SPF\_EXPECT}(t,j) + e_{j,t} \tag{35}$$

for forecast horizons of one through four quarters ($j = 1, \ldots, 4$). Under the null hypothesis that the smoothed estimate of the hidden factor is spurious, the coefficient b_1 should be zero. Standard errors use the Newey-West adjustment for four lags of moving average residuals. Although the regression is more intuitive if the regressor and regressand are switched, there is a generated regressor problem when using the smoothed estimate of the hidden factor as the explanatory variable.

Table 5 reports point estimates, standard errors, R^2s, and the serial correlation of residuals at the first and fourth lags. The null hypothesis is overwhelmingly rejected. The point estimates are reliably negative, with asymptotic t statistics ranging from -2.5 to -5.0. The inverse of the point estimates imply that a one standard deviation change in the hidden component of the risk premium factor corresponds to expected declines in the three-month yield between 60 and 160 basis points.

These results confirm that investors' expectations of expected future yields are not spanned by the cross section of yields. Kalman filtering extracts part of investors' beliefs from yield dynamics. The next question is whether these beliefs are related to the macroeconomy.

4.2 Explaining the hidden factor with macro variables

It is easy to data-mine a spurious relation between bond risk premia and macroeconomic forces. There are hundreds of time series we could choose to explain variations in the hidden component of the risk premium factor. To reduce this danger, I follow Joslin et al. (2009) by choosing two prominent monthly time series: the growth of industrial production and CPI inflation.[6] I adopt their approach of smoothing each series by estimating univariate ARMA(1,1) processes. The smoothed month-t values used here are the ARMA month-t predictions of next month's values.

I also follow Ludvigson and Ng (2009, 2010), who construct principal components of more than 130 time series. They use the first eight principal components to predict excess bond returns. I use the eight series studied in Ludvigson and Ng (2010), which are available for the same 1964 through 2007 period used in this paper.[7] They label the most important component (in variance decomposition sense) a "real activity" factor because it is highly correlated with variables such as industrial production growth. In this sample, the correlation between the real activity factor and the growth of industrial production is 0.88.[8]

To explain the hidden component of bond risk premia with these variables, I estimate regressions of the form

$$\widetilde{RP}_t^s = b_0 + b' w_t + e_t,$$

where subsets of the ten time series are included in w_t. Because the residuals are highly serially correlated, standard errors and p-values use a Newey-West adjustment with 15 lags. Results are in Table 6.[9]

The first three rows of the table tell a consistent story. The hidden component of bond risk premia is countercyclical; it tends to be low when output and prices are growing faster than usual. The coefficients on industrial production growth, inflation, and real activity all have p-values against zero between two and five percent. However, the fraction of the hidden component that is explained is quite small; no more than ten percent, even when all eight of the Ludvigson-Ng factors are included.

The final two rows complicate matters. Ludvigson and Ng build a predictor of excess bond returns by searching over powers of their principal components. Based on this search, they use one nonlinear transformation—the third power of real activity. Adding this term raises the R^2 by seven to eight percentage points and results in overwhelmingly statistically

[6]They use core inflation, while I use CPI-U.

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

[8]I switch the sign of the real activity factor used by Ludvigson and Ng.

[9]The on-line appendix has a counterpart to Table 6 in which the dependent variable is replaced by the entire risk premium factor. A comparison of the two tables reveals that macroeconomic factors are more closely linked to the hidden component than to the entire risk premium factor.

significant coefficients on real activity and its third power.

If we take this result at face value, it points to an asymmetric relation between bond risk premia and macroeconomic activity. Yet over the 1964 through 2007 sample period, the nonlinear term acts more as a dummy variable that picks out a few months. The median absolute demeaned nonlinear term is about 1.3, but there are two months for which the absolute value exceeds 80. Two-thirds of the increase in R^2 associated with the nonlinear term is driven by the observation of May 1980. The hidden component for that month is at its sample maximum of four standard deviations above its mean, while the nonlinear term is at its sample minimum of 13 standard deviations below its mean. Owing to the sensitivity of the regression results to this single observation, it is safer to conclude that there is a statistically reliable but economically weak inverse relation between levels of economic activity and the estimated hidden component of bond risk premia. A nonlinear relation may exist but cannot be inferred reliably from this sample.

One interpretation of this empirical evidence is that there are both countercyclical and acyclical (non-macroeconomic) determinants of the hidden component of bond risk premia. Another is that the hidden component is entirely countercyclical; everything else is simply noise inherent in the construction of smoothed estimates of the hidden component. The latter perspective corresponds to the model of Joslin et al. (2009). The results in Table 6 cannot distinguish these possibilities. More relevant evidence is discussed next.

4.3 Comparisons to other forecasting variables

Is the predictive power of the fitted risk premium factor captured by other variables known to predict bond returns? I investigate this question by estimating return forecasting regressions with a variety of predictors. This analysis is performed largely in response to popular demand. To briefly summarize the results, the in-sample ability of the filtered risk premium factor to explain excess monthly bond returns is not materially affected by adding other well-known predictors to the forecasting regression. However, this factor does not capture all of the forecastibility of excess returns. In particular, the macro/finance principal components of Ludvigson and Ng contain substantial additional information about monthly excess returns over the 1964 through 2007 sample.

Before getting into the details, it is worth emphasizing that the modeling implications we can draw from these results are limited. We face three problems when interpreting the regressions. First, as discussed in Section 2.8, there is nothing in the model or estimation methodology that implies Kalman filtering of yields produces the most accurate forecasts. There may well be some non-yield variables that perform better. Conversely, if macroeconomic variables such as real activity span bond risk premia, but the econometrician observes the macro variables with error, filtered estimates of risk premia should contain additional information that predicts excess returns. Second, all of the predictive variables are chosen expressly for their documented ability to forecast over sample periods that overlap with the sample period here. Hence overfitting is an unavoidable problem. Third, forecasting regressions throw away the model-implied loadings of excess returns on the risk premium factor. Section 3.4 points out that in the 1964 through 2007 sample, there is a clear wedge between forecasts based on correct loadings and forecasts from OLS regressions.

With these caveats in mind, I turn to the regression specification. In all of the regressions, the dependent variable is the excess return to a portfolio of Treasury bonds from the end of month t to the end of month $t+1$. The bonds all have month-t maturities between five and ten years. Excess returns are constructed by subtracting the return to a portfolio of Treasury bonds with maturities under six months. Each regression includes as a predictive variable the month-t filtered estimate of the risk premium factor. One regression also includes the linear combination of forward rates introduced by Cochrane and Piazzesi (2005).[10] Another regression follows Joslin et al. (2009) by using smoothed inflation and industrial production growth as predictors. The variables are described in Section 4.2. Two other regressions use some or all of the eight principal components of many macroeconomic and financial time series, following Ludvigson and Ng (2009, 2010). These components are also described in Section 4.2.

Regression results are in Table 7. All of the predictive variables are normalized to have unit standard deviation. The dependent variable is measured in percent per month. Therefore the point estimate in the first listed regression implies that a one-standard-deviation change in the filtered risk premium factor at month t corresponds to an increase in the expected excess portfolio return of 35 basis points in month $t+1$. The t-statistic, in parenthesis, exceeds three, and the R^2 exceeds four percent. The t-statistic is adjusted for generalized heteroskedasticity. Of course, it cannot be adjusted for the fact that the filtered risk premium factor is expressly constructed to predict returns over the sample period.

The second regresson shows that the Cochrane-Piazzesi (CP) predictor provides no incremental explanatory power in the presence of the filtered risk premium factor. The estimated coefficient is economically and statistically insignificant. This result may be surprising, given the well-documented ability of CP to predict excess returns over much of this sample period. But the evidence in Cochrane and Piazzesi (2005) and subsequent work focuses on annual return horizons. In results not detailed here, I verify that the ability of the CP factor to explain monthly excess returns is stronger when more distant lags of CP are used. Put differ-

[10] I construct their predictor for the sample period here by following their recipe.

ently, in this data sample CP picks up a small, highly persistent variation in excess monthly bond returns. The risk premium factor picks up a larger, less persistent component.

The third regression shows that industrial production growth and inflation also provide little incremental explanatory power. Adding these two predictors raises the R^2 from 0.042 to only 0.049. The point estimates are not statistically different from zero at conventional levels, either individually or jointly. These results differ from those in Joslin et al. (2009), perhaps because of different sample periods (they use 1989 to 2008) and different yields (they use swap rates).

The fourth and fifth regressions regression use principal components from Ludvigson and Ng. The fourth includes a real activity factor (the first principal component) and a stock market factor (the eighth principal component). The latter is included because of the evidence dating back to Fama and French (1989) that variables such as dividend/price ratios can predict excess bond returns. The fifth regression includes all eight principal components. It does not include the third power of the real activity factor. For this sample, the nonlinear variable has no incremental explanatory power for excess monthly bond returns.

The table shows that both the real activity and stock market factors have economically and statistically significant forecasting power. Collectively, the other six principal components also contain substantial information about future excess returns. But little of this predictive power is related to the predictive power of the filtered risk premium factor. Adding real activity and the stock market factor raises the R^2 to 0.073. Adding the other six factors raises it to 0.124. In the fifth regression, the estimated coefficient on the filtered risk premium factor is 20 percent smaller than the univariate coefficient; 0.28 versus 0.35. It remains overwhelmingly statistically significant, again igoring the fact that the factor is constructed to predict returns.

As noted at the beginning of this discussion, these results cannot be pushed very far. But one reasonable conclusion is that the model of Joslin et al. (2009), in which industrial production growth and CPI inflation are perfectly observed and span hidden term structure factors, is an inaccurate description of the 1964 to 2007 sample period. A more plausible model in which macroeconomic activity and inflation span hidden factors but are observed with noise is not ruled out, although not given much support here.

Nor do these results rule out the distinct possibility that the model and forecasting regressions are misspecified because of structural shifts during the Federal Reserve's monetarist experiment period. The risk of such misspecification drives the choice in Joslin et al. (2009) to restrict attention to the 1989 through 2008 period. Using the terminology of Section 2.8, I leave as a open question whether a yields-only model or a yields-plus model better fits that particular data sample.

5 Conclusion

In the context of a Gaussian dynamic term structure model, this paper documents that an economically important component of bond risk premia is hidden from the cross section of yields. This component contains substantial information about expected future yields but has a negligible immediate effect on the term structure. The factor is related to aggregate economic activity, but the strength of the relation is fairy modest.

An important lesson to draw from this evidence is that an econometrician should not rely on estimation techniques that extract information exclusively from the cross section. Such techniques are standard in the literature on dynamic term structure models. Instead, she needs to build models that accommodate hidden factors, and use estimation techniques that are robust to the presence of these factors. The method adopted here is filtering, which uses information from yield dynamics to infer factor properties. Another potentially valuable approach is to use information from sources other than bond yields. Unfortunately, the macroeconomic variables considered here appear to contain relatively little information about hidden components of risk premia for the sample 1964 through 2007.

Appendix 1. Model parameterization details

This appendix derives the restrictions that define the factors as principal components. Begin with the preferred rotation of JSZ, denoting the state vector with an asterisk to contrast it with a rotation based on principal components. Under the equivalent-martingale measure, the state's dynamics are

$$x^*_{t+1} = Dx^*_{t+1} + \Sigma\epsilon^q_{t+1}, \quad \epsilon^q \sim MVN(0, I),$$

where Σ is lower triangular and D is a diagonal matrix with distinct real eigenvalues on the diagonal. I skip discussion of the more general cases of repeated or complex eigenvalues because the estimated eigenvalues in the empirical work here are all real and distinct. Following JSZ, the normalized short-rate equation is

$$r_t = \delta_0 + \iota' x^*_t$$

where ι is a vector of ones. For ease of reference, denote the parameters of the equivalent-martingale dynamics as

$$\rho^q = \begin{pmatrix} \text{diag}(D)' & s' & \delta_0 \end{pmatrix}',$$

where s is the vector of parameters of the lower triangular matrix Σ. There are $n + n(n+1)/2 + 1$ free parameters in ρ^q.

The next step is to define a new state vector rotation using these parameters. The first element of the new state vector has an innovation equal to the first principal component of innovations to bond yields, the second element has an innovation equal to the second principal component, and so on. The yields used to compute these components are those on v bonds with maturities $M = \{m_1, \ldots, m_v\}$. These are not necessarily the bonds used to estimate the model. The rotation used in this paper is based on 60 bonds with maturities from one to sixty months.

Stack the true yields on these bonds (i.e., yields uncontaminated by measurement error) in the vector y^*_t. The relation between the yield vector and the original state vector is

$$y^*_t = A^* + B^* x^*_t,$$

where element i of the v-vector A^* is A_{m_i} and row i of the $v \times n$ matrix B^* is B'_{m_i}. Both A^* and B^* are determined only by ρ^q and the choice of maturities M.

The new state vector is defined by principal components of innovations to y^*_t. The covariance matrix of these innovations is

$$\text{Cov}_t(y^*_{t+1}) = B^*\Sigma\Sigma B^{*\prime}.$$

Assume $v \geq n$. Then the rank of this covariance matrix is n unless (18) is satisfied. Since this restriction is not imposed, assume the rank is n. Then the matrix can be decomposed into principal components as

$$\text{Cov}_t(y^*_{t+1}) = P\Omega P',$$

where the $v \times n$ matrix P contains the loadings of the yields on the n principal components and the diagonal $n \times n$ matrix Ω contains the eigenvalues ordered by magnitude. These matrices depend only on ρ^q and M. Define the $n \times n$ matrix Γ as

$$\Gamma = P' B^*.$$

Define the new factor vector as

$$x_t \equiv P'(y_t^* - A^*) = \Gamma x_t^*.$$

The mapping from this state vector to the short rate is

$$r_t = \delta_0 + \delta_1' x_t, \qquad \delta_1' = \iota' \Gamma^{-1}.$$

Equation (22) contains the equivalent martingale dynamics, where

$$K^q = \Gamma D \Gamma^{-1}.$$

This principal components factor rotation is defined entirely in terms of the equivalent-martingale parameters ρ^q and the maturity vector M. Physical measure dynamics are irrelevant. This property is why the principal components are calculated using covariances of yield innovations. Principal components calculated using covariances of, say, yields depend on both equivalent-martingale and physical measure dynamics. The approach taken here simplifies considerably estimation of the model.

The free parameters linking the equivalent martingale and physical measures are the scalars $\lambda_{0(L)}$ and $\lambda_{0(S)}$ and the n-vector $\lambda_{1(L)}$. The final model parameter is the standard deviation of measurement error in bond yields in (19). Therefore the set of parameters that describe the entire term structure is

$$\rho = (\rho^{q\prime} \quad \lambda_{0(L)} \quad \lambda_{0(S)} \quad \lambda_{1(l)}' \quad \sigma_\eta)'.$$

There are a total of $2n + n(n+1)/2 + 4$ free parameters.

Appendix 2. Linear algebra derivations

Section 3.3 uses the expectation of the state vector conditioned on principal components of observed yields. Denote by z_t the vector of the first n principal components, written in the form (20). The variance of z_t is, using (19),

$$\mathrm{Var}(z_t) = \mathrm{Var}(\mathcal{P}\left(\mathcal{B}x_t + \eta_t\right)) = \mathcal{P}\mathcal{B}\mathrm{Var}(x_t)\mathcal{B}'\mathcal{P}' + \mathcal{P}\mathcal{P}'\sigma_\eta^2.$$

The projection of the state vector on these principal components is

$$\begin{aligned}
E(x_t|z_t) &= E(x_t) + \mathrm{Cov}(x_t, z_t)\mathrm{Var}(z_t)^{-1} z_t \\
&= E(x_t) + \mathrm{Cov}\left(x_t, \mathcal{P}\mathcal{B}x_t + \mathcal{P}\eta_t\right)\mathrm{Var}(z_t)^{-1} z_t \\
&= E(x_t) + \mathrm{Var}(x_t)\mathcal{B}'\mathcal{P}'\mathrm{Var}(z_t)^{-1} z_t.
\end{aligned}$$

The variance of this conditional expectation is

$$\mathrm{Var}\left(E(x_t|z_t)\right) = \mathrm{Var}(x_t)\mathcal{B}'\mathcal{P}'\mathrm{Var}(z_t)^{-1}\mathcal{P}\mathcal{B}\mathrm{Var}(x_t).$$

Section 4 uses projections of contemporaneous yields and future short rates on the risk premium factor. They are expressed as deviations from unconditional means. Yield projections are

$$\begin{aligned} E\left(y_t^{(m)} - E\left(y_t^{(m)}\right) \middle| RP_t - E(RP_t)\right) &= B_m' E\left(x_t - E(x_t) \middle| RP_t - E(RP_t)\right) \\ &= B_m' E\left(x_t - E(x_t) \middle| \lambda_{1(L)}'(x_t - E(x_t))\right) \\ &= B_m' \mathrm{Cov}(x_t, x_t' \lambda_{1(L)}) \mathrm{Var}(\lambda_{1(L)}' x_t)^{-1} (RP_t - E(RP_t)) \\ &= B_m' \mathrm{Var}(x_t) \lambda_{1(L)} \left(\lambda_{1(L)}' \mathrm{Var}(x_t) \lambda_{1(L)}\right)^{-1} (RP_t - E(RP_t)). \end{aligned}$$

For projections of the j-ahead future short rate on the risk premium, use E_t to denote the expectation conditioned on all time-t information. The projections are

$$\begin{aligned} E\left(r_{t+j} - E(r_t) \middle| RP_t - E(RP_t)\right) &= \delta_1' E\left(x_{t+j} - E(x_t) \middle| \lambda_{1(L)}'(x_t - E(x_t))\right) \\ &= \delta_1' E\left(E_t(x_{t+j} - E(x_t)) \middle| \lambda_{1(L)}'(x_t - E(x_t))\right) \\ &= \delta_1' E\left(K^j(x_t - E(x_t)) \middle| \lambda_{1(L)}'(x_t - E(x_t))\right) \\ &= \delta_1' K^j \mathrm{Var}(x_t) \lambda_{1(L)} \left(\lambda_{1(L)}' \mathrm{Var}(x_t) \lambda_{1(L)}\right)^{-1} (RP_t - E(RP_t)) \end{aligned}$$

Section 4 also uses projections on the hidden component of the risk premium factor. Construct the hidden component using

$$\widetilde{RP}_t \equiv RP_t - E(RP_t|z_t) = \lambda_{1(L)}' \left(x_t - E(x_t|z_t)\right).$$

The variance of the hidden component is

$$\begin{aligned} \mathrm{Var}\left(\widetilde{RP}_t\right) = \lambda_{1(L)}' \Big[&\left(I - \mathrm{Var}(x_t)\mathcal{B}'\mathcal{P}'\mathrm{Var}(z_t)^{-1}\mathcal{P}\mathcal{B}\right) \mathrm{Var}(x_t) \left(I - \mathcal{B}'\mathcal{P}'\mathrm{Var}(z_t)^{-1}\mathcal{P}\mathcal{B}\mathrm{Var}(x_t)\right) \\ &+ \mathrm{Var}(x_t)\mathcal{B}'\mathcal{P}'\mathrm{Var}(z_t)^{-1}\mathcal{P}\mathcal{P}'\mathrm{Var}(z_t)^{-1}\mathcal{P}\mathcal{B}\mathrm{Var}(x_t)\sigma_\eta^2 \Big] \lambda_{1(L)}. \end{aligned}$$

The covariance of the state vector with the hidden component is

$$\mathrm{Cov}(x_t, \widetilde{RP}_t) = \mathrm{Var}(x_t) \left(I - \mathcal{B}'\mathcal{P}'\mathrm{Var}(z_t)^{-1}\mathcal{P}\mathcal{B}\mathrm{Var}(x_t)\right) \lambda_{1(L)}.$$

The projection of contemporaneous yields on the hidden component of the risk premium factor is

$$E\left(y_t^{(m)} - E\left(y_t^{(m)}\right) \middle| \widetilde{RP}_t\right) = B_m' \mathrm{Cov}(x_t, \widetilde{RP}_t) \mathrm{Var}(\widetilde{RP}_t)^{-1} \widetilde{RP}_t.$$

The projection of future short rates on the hidden component is

$$E\left(r_{t+j} - E(r_t) \middle| \widetilde{RP}_t\right) = \delta_1' K^j \mathrm{Cov}(x_t, \widetilde{RP}_t) \mathrm{Var}(\widetilde{RP}_t)^{-1} \widetilde{RP}_t.$$

References

Ang, Andrew, and Monika Piazzesi, 2003, A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables, *Journal of Monetary Economics* 50, 745-787.

Ang, Andrew, Monika Piazzesi, and Min Wei, 2006, What does the yield curve tell us about GDP growth?" *Journal of Econometrics* 131, 359-403.

Brandt, Michael W., and David A. Chapman, 2003, Comparing multifactor models of the term structure, Working paper, Duke University.

Chen, Ren-Raw, and Louis Scott, 1993, Maximum likelihood estimation for a multifactor equilibrium model of the term structure of interest rates, *Journal of Fixed Income* 3, 14-31.

Clarida, Richard, Jordi Galí, and Mark Gertler, 2000, Monetary policy rules and macroeconomic stability: evidence and some theory, *Quarterly Journal of Economics* 115, 147-180.

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.

Collin-Dufresne, Pierre, and Robert S. Goldstein, 2003, Generalizing the affine framework to HJM and random field models, Working paper, Haas School of Business, University of California at Berkeley.

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.

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

Duffee, Gregory R., 2009, Forecasting with the term structure: The role of no-arbitrage restrictions, Working paper, Johns Hopkins.

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

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

Fama, Eugene F., and Kenneth R. French, 1989, Business conditions and expected returns on stocks and bonds, *Journal of Financial Economics* 25, 23-49.

Gurkaynak, Refet S., Brian Sack, and Jonathan H. Wright, 2006, The U.S. Treasury yield curve: 1961 to the present, working paper, Federal Reserve Board.

Hamilton, James D., and Jing Wu, 2010, Identification and estimation of affine-term-structure models, Working paper, UCSD.

Joslin, Scott, Marcel Priebsch, and Kenneth J. Singleton, 2009, Risk premium accounting in macro-dynamic term structure models, Working paper, Stanford GSB.

Joslin, Scott, Kenneth J. Singleton, and Haoxiang Zhu, 2009, Characterizing Gaussian DTSMs: Why seemingly different models are (nearly) identical, Working paper, Stanford GSB.

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.

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.

Pearson, Neil D., and Tong-sheng Sun, 1994, Exploiting the conditional density in estimating the term structure: An application to the Cox, Ingersoll, and Ross model, *Journal of Finance* 49, 1279-1304.

Pennacchi, George G., 1991, Identifying the dynamics of real interest rates and inflation: evidence using survey data, *Review of Financial Studies* 4, 53-86.

Piazzesi, Monika, 2009, Affine term structure models, *Handbook of Financial Econometrics* 1, Yacine Ait-Sahalia and Lars Hansen, eds., North-Holland, forthcoming.

Stambaugh, Robert F., 1988, The information in forward rates, *Journal of Financial Economics* 21, 41-70.

Taylor, John B., 1993, Discretion versus policy rules in practice, *Carnegie-Rochester Conference Series on Public Policy* 39, 195-214.

Table 1. An estimated dynamic term structure model

The short rate is an affine function of a length-five state vector x_t. Elements of the state vector are principal components of shocks to the term structure, ordered in decreasing importance. The equivalent-martingale and physical dynamics are

$$x_{t+1} = K^q x_t + \Omega^{1/2}\epsilon^q_{t+1}, \quad x_{t+1} = \lambda_0 + (K^q + \lambda_1)x_t + \Omega^{1/2}\epsilon_{t+1},$$

where Ω is diagonal. The first two elements of λ_0 and the first row of λ_1, denoted $\lambda_{1(L)}$, are free; all other elements are zero.

Yields on bonds with maturities of three months and one through five years are observed with iid measurement error with standard deviation σ_η. Kalman filter estimation uses month-end yields from 1964 through 2007. The table reports parameter estimates and standard errors. The standard errors are computed from Monte Carlo simulations under the null hypothesis that the estimated model is true.

	Factor				
	1	2	3	4	5
Loading of short rate on factors	0.128	0.398	0.509	0.267	0.210
	(0.006)	(0.014)	(0.020)	(0.026)	(0.0240
K^q	1.002	−0.218	−0.342	−0.230	−0.142
	(0.003)	(0.010)	(0.018)	(0.019)	(0.020)
	0.020	0.869	−0.485	−0.330	−0.349
	(0.004)	(0.015)	(0.026)	(0.039)	(0.030)
	0.009	−0.016	0.794	−0.334	−0.251
	(0.002)	(0.006)	(0.027)	(0.033)	(0.033)
	0.000	0.003	0.039	0.980	−0.207
	(0.001)	(0.002)	(0.010)	(0.020)	(0.011)
	0.000	−0.001	0.001	0.021	1.023
	(0.000)	(0.001)	(0.007)	(0.009)	(0.016)
$\text{diag}(\Omega^{1/2}) \times 10^4$	26.772	7.615	3.155	1.072	0.419
	(0.915)	(0.249)	(0.140)	(0.101)	(0.069)
$\lambda_0 \times 10^4$	3.427	−1.855	0.0	0.0	0.0
	(2.903)	(0.341)	(-)	(-)	(-)
$\lambda_{1(L)}$	−0.074	0.336	0.712	2.244	6.506
	(0.027)	(0.087)	(0.288)	(0.734)	(1.486)
Constant term in short rate ($\times 10^2$)		0.484			
		(0.081)			
Std dev of measurement error ($\times 10^4$)		0.492			
		(0.010)			

Table 2. Model-implied population properties of excess bond returns

Gaussian term structure models with n factors are estimated with the Kalman filter. True yields are affine functions of the unobserved factors. Observed yields are contaminated with iid measurement error. This table reports population properties of excess monthly and annual log returns to a five-year bond. Excess returns are calculated by subtracting continuously-compounded yields on one-month and one-year bonds, respectively. Observed returns differ from true returns because of the measurement error. Conditional expectations of future excess true returns are calculated using either the state vector (full info) or n principal components of observed yields. The "Ratio" column reports the ratio of the variance of the former conditional expectation to the variance of the latter conditional expectation. Two-sided 95 percent confidence bounds are reported for the five-factor model. They are computed from Monte Carlo simulations. Means are in percent and variances are in percent squared.

Factors (n)	Mean	Total variance: True returns	Total variance: Observed returns	Variance of conditional expectation: Using full info	Variance of conditional expectation: Using n PCs	Ratio
A. Monthly returns						
5	0.12	3.62	3.79	0.48	0.25	0.53
	[−0.03 0.26]	[3.16 4.03]	[3.33 4.20]	[0.17 0.65]	[0.09 0.35]	[0.36 0.68]
4	0.15	3.62	3.83	0.19	0.18	0.97
3	0.26	3.60	3.89	0.15	0.15	0.97
B. Annual returns						
5	0.84	40.62	40.76	18.28	12.88	0.70
	[−0.55 2.15]	[26.50 52.87]	[26.65 53.00]	[6.80 26.20]	[3.34 18.76]	[0.55 0.86]
4	0.90	31.59	31.77	7.90	7.81	0.99
3	2.29	31.18	31.43	6.56	6.52	0.99

Table 3. Model-implied R^2s of monthly excess bond returns

Gaussian term structure models with n factors are estimated with the Kalman filter. True yields are affine functions of the unobserved factors. Observed yields are contaminated with iid measurement error. This table reports model-implied ratios of predicted/total variances for log excess monthly returns to five-year zero-coupon bond. Ratios are calculated both in population and in the data sample of 1964 through 2007. In Panel A, the column labeled "full info" defines the ratio using true returns predicted with the true factors. All other columns use observed returns and factors extracted with the Kalman filter. "Optimal filtering" and "optimal smoothing" use the model to form forecasts of returns with either filtered or smoothed estimates of the factors. The columns labeled "OLS" use predictive regressions to form forecasts of excess returns with either filtered or smoothed factors as regressors. Panel B also reports the log likelihoods of each estimated model.

A. Population ratios

Factors (n)	Full info	Optimal filtering	Optimal smoothing	OLS using filtered	OLS using smoothed
5	0.132	0.087	0.105	0.087	0.154
4	0.051	0.047	0.048	0.048	0.051
3	0.042	0.039	0.039	0.038	0.039

B. Sample ratios

Factors (n)	Log likelihood	Optimal filtering	Optimal smoothing	OLS using filtered	OLS using smoothed
5	26,711.85	0.085	0.084	0.043	0.131
4	26,660.15	0.034	0.034	0.036	0.039
3	26,419.38	0.033	0.035	0.035	0.037

Table 4. Time-series and cross-sectional accuracy of term structure models

Gaussian term structure models with n factors are estimated with the Kalman filter. The sample period is January 1964 through December 2007. The table reports root mean squared conditional forecast errors for various bonds at the one-month and twelve-month horizons. In both cases, errors are defined as observed yields less yields implied by the month-t filtered estimate of the factors. Conditional forecast errors use observed yields at $t + j, j = 1, 12$, and cross-sectional errors use observed yield at t. Errors are in basis points of annualized yields.

Factors	3 mon	1 year	2 years	3 years	4 years	5 years
One month ahead						
3	54.05	51.78	46.14	42.70	40.57	38.40
4	54.03	51.80	46.08	42.61	40.50	38.27
5	53.73	52.30	45.93	42.65	40.67	38.10
Twelve months ahead						
3	176.09	166.53	149.92	135.42	127.71	121.68
4	176.72	166.61	149.21	134.35	126.16	119.75
5	172.11	167.89	149.27	134.82	126.35	122.05
Cross-sectional						
3	2.63	6.95	6.54	5.96	5.79	5.70
4	1.66	4.09	5.61	4.95	5.57	4.41
5	1.35	3.23	4.40	4.56	4.81	3.57

Table 5. The relation between survey forecasts and model-implied hidden risk premia

A five-factor term structure model is estimated with the Kalman filter. Bond risk premia are constrained to vary with a single risk premium factor. Smoothed estimates of the risk premium factor are regressed on five principal components of observed yields. The residual from this regression is the hidden risk premium component. It is normalized to have a unit standard deviation.

Quarterly observations of expectations of future Treasury bill yields are from the Survey of Professional Forecasters. The data used are quarter-t mean survey forecasts of the three-month T-bill yield during quarters $t + j, j = 1, \ldots 4$, expressed in annual percentage points. The contemporaneous three-month yield is subtracted from the forecasts to produce forecasted changes in the yield. The hidden risk premium component is regressed on contemporaneous expectations of j-quarter ahead changes in Treasury yields. All regressions are estimated from 1981Q3 through 2007Q4 (106 quarterly observations). Newey-West standard errors are in parentheses, adjusted for four lags of moving average residuals.

	Quarters ahead (j)			
	1	2	3	4
Coef	−1.733	−1.140	−0.844	−0.637
	(0.347)	(0.378)	(0.333)	(0.251)
Serial corr of residual, 1st lag	0.40	0.45	0.46	0.46
Serial corr of residual, 4th lag	−0.12	−0.17	−0.16	−0.12
R^2	0.25	0.17	0.14	0.12

Table 6. Projections of the hidden risk premium factor on macroeconomic variables, 1964 through 2007

A five-factor term structure model is estimated with the Kalman filter. Bond risk premia are constrained to vary with a single risk premium factor. The "hidden" part of the factor is orthogonal to principal components of observed yields. Monthly smoothed estimates of the hidden component are regressed on contemporaneous realizations of other variables. Industrial production growth and CPI inflation are both month-t predictions of month-$(t+1)$ values, from individual ARMA(1,1) models. Ludvigson-Ng construct eight principal components of many macro and financial time series. The first is a "real activity factor," which here is normalized to positively covary with industrial production growth. Each variable used in the table is normalized to have a unit standard deviation. The table reports point estimates and t-statistics. The latter are adjusted for 15 lags of moving average residuals. P-values of joint tests that coefficients on Ludvigson-Ng factors two through eight equal zero are in square brackets. The column labeled ρ_{15} is the serial correlation of residuals at the 15th lag. The sample is January 1964 through December 2007.

Include LN factors 2-8?	Ind. prod. growth	Inflation	LN real activity	LN (real activity)^3	ρ_{15}	R^2
No	-0.20	-0.27	-	-	-0.02	0.08
	(-2.19)	(-2.05)				
No	-	-	-0.22	-	0.06	0.05
			(-1.98)			
Yes	-	-	-0.22	-	0.07	0.10
[0.019]			(-2.08)			
No	-	-	-0.46	0.36	0.10	0.13
			(-5.06)	(4.03)		
Yes	-	-	-0.47	0.37	0.10	0.17
[0.001]			(-5.35)	(3.66)		

Table 7. One-month-ahead forecasting regressions, 1964 through 2007

The excess return to a portfolio of Treasury bonds from the end of month t to the end of month $t+1$ is forecast with a variety of month-t variables. Two are designed to predict excess returns; an estimated risk premium factor filtered from estimation of a five-factor term structure model, and the Cochrane-Piazzesi predictor derived from a cross-section of Treasury yields. Smoothed industrial production growth, smoothed CPI inflation, and Ludvigson-Ng factors are described in the notes to Table 5. The eighth Ludvigson-Ng factor is a "stock market" factor. Each predictor is normalized to have a unit standard deviation. The table reports point estimates and t-statistics. The latter are adjusted for generalized heteroskedasticity. The p-value of the joint test that coefficients on Ludvigson-Ng factors two through seven equal zero is in square brackets. The bond portfolio has maturities ranging from five to ten years. The sample is February 1964 through December 2007.

Include LN factors 2-7?	Filtered factor	Cochrane/ Piazzesi	Ind. prod. growth	Inflation	LN real activity	LN stock market	R^2
No	0.35	-	-	-	-	-	0.042
	(3.24)						
No	0.32	0.06	-	-	-	-	0.043
	(2.98)	(0.47)					
No	0.32	-	−0.10	−0.13	-	-	0.049
	(3.11)		(−1.40)	(−1.09)			
No	0.33	-	-	-	−0.18	0.24	0.073
	(3.06)				(−2.17)	(3.45)	
Yes	0.28	-	-	-	−0.19	0.24	0.124
[0.001]	(2.82)				(−2.22)	(3.45)	

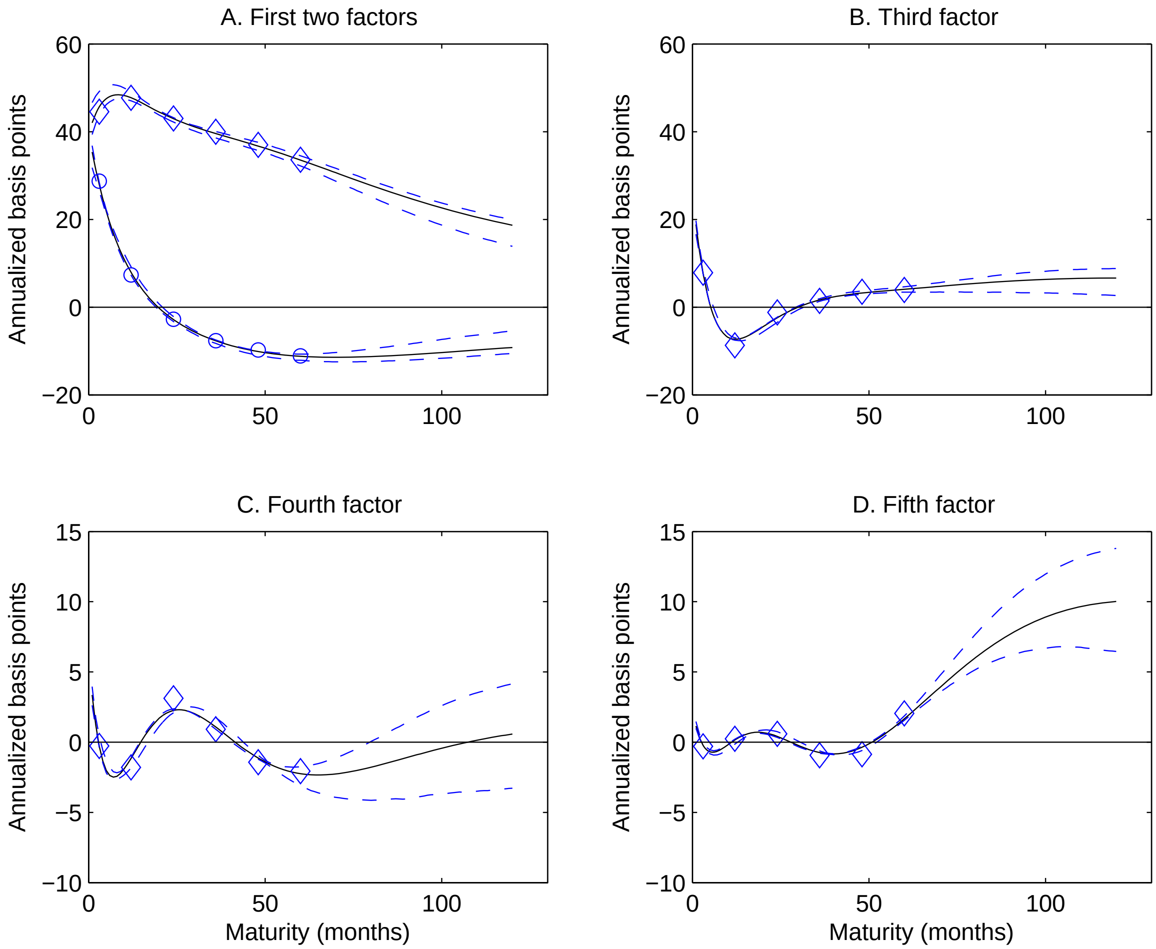

Fig. 1. Estimated yield loadings for a five-factor Gaussian term structure model estimated with monthly data from 1964 to 2007. The factors are principal components of shocks to the term structure. They are scaled by estimated standard deviations of the shocks. The diamonds are coefficients from regressions of observed yields on smoothed estimates of the factors. The dashed lines are two-sided 95 percent confidence intervals calculated from Monte Carlo simulations. Note the vertical scales of Panels A and B differ from those of Panels C and D.

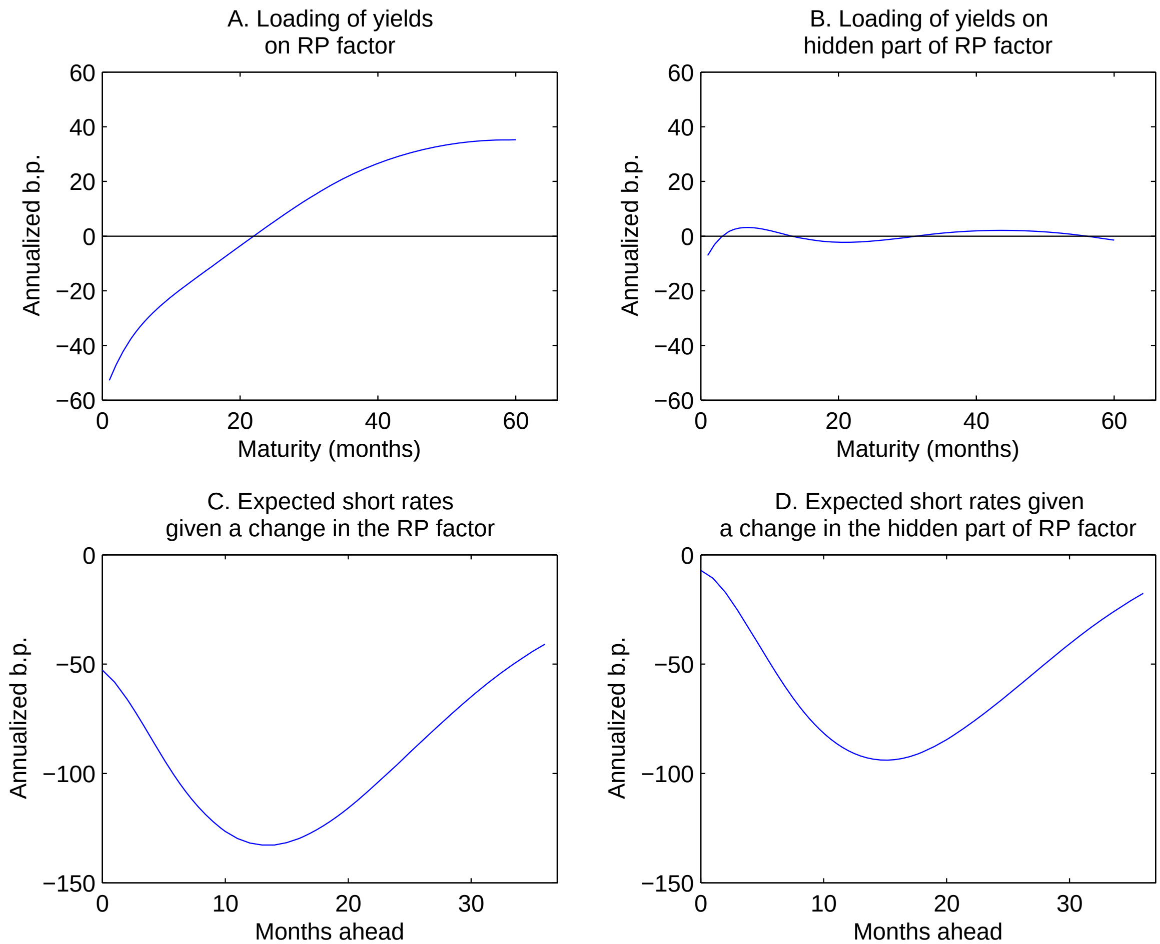

Fig. 2. Model-implied effects of a risk premium factor. A Gaussian five-factor term structure model is estimated with monthly data from 1964 to 2007. A single "RP" factor determines risk premia on all bonds. The hidden part of the factor is the residual from projecting the factor on five principal components of observed yields. Panels A and B display loadings of yields on one standard deviation changes in the factor and its hidden part, respectively. Panels C and D are expected changes in the one-month spot rate conditioned on the same changes.