Are variations in term premia related to the macroeconomy?

Gregory R. Duffee[*]

Haas School of Business
University of California – Berkeley

This Draft: June 26, 2007

ABSTRACT

To test whether expected excess bond returns are correlated with particular macroeconomic variables, the relevant null hypothesis is that expected excess returns are stochastic, persistent, and independent of the variables. However, current methods used to test this hypothesis—forecasting regressions and joint dynamic models of the term structure and macroeconomic variables—do not use this null. Their null is that excess returns are serially uncorrelated. This paper presents a dynamic model that satisfies the appropriate null. Simulation results show that finite-sample distributions of forecasting regressions under the appropriate null differ substantially from finite-sample distributions under the commonly-used null. Model estimation reveals that at most a small fraction of variation in expected excess returns is associated with inflation, output growth, and the short rate.

[*]Voice 510-642-1435, email duffee@haas.berkeley.edu. Address correspondence to 545 Student Services Building #1900, Berkeley, CA 94720. Thanks to Andrew Ang, Refet Gurkaynak, Monika Piazzesi, Oreste Tristani, and seminar participants at the Federal Reserve Board, Johns Hopkins, Washington University, Wharton, and an ECB-BIS workshop for helpful comments. This paper, along with the Matlab code used to estimate the model and construct the simulations, is at http://faculty.haas.berkeley.edu/duffee/.

1 Introduction

Tests of the expectations hypothesis document conclusively that premia on Treasury bonds vary with the shape of the term structure. Returns to long-term bonds less returns to short-term bonds can be predicted with spreads, including both the spread between forward rates and short-maturity yields as in Fama and Bliss (1987) and yield spreads as in Campbell and Shiller (1991).[1]

Campbell (1987) notes that spreads are powerful instruments for detecting variations in term premia because changes in expected excess returns to long-term bonds are automatically compounded in the prices of these bonds, and thus in the spread between long-term and short-term bond yields. Put differently, there is an accounting relation linking expected excess returns to forward rates. Yet the same accounting relation that makes spreads powerful instruments also makes them, in a sense, uninformative. Variations in expected excess returns can be detected with spreads regardless of the reasons for the variation, hence this evidence says nothing about the underlying determinants of term premia.

Beginning with Kessel (1965) and Van Horne (1965), economists have proposed various theories of time-varying term premia. Many theories imply that term premia are correlated with the state of the economy. For example, if term premia reflect risk compensation, premia will vary with the price of interest rate risk and the amount of interest rate risk. Plausible stories link both to the macroeconomy. Other theories, such as investor overreaction to information (see, e.g., Shiller et al. (1983)) are not as closely tied to economic conditions.

One way to help test these theories is to determine whether expected excess bond returns are correlated with measures of macroeconomic conditions such as economic activity, inflation, and indicators of monetary policy. Researchers use two methods to look for evidence of such correlations. The first follows Fama and Schwert (1977) by regressing excess bond returns on lagged macroeconomic variables. The second follows Ang and Piazzesi (2003) by estimating parsimonious models that specify the joint dynamics of the term structure and specified macroeconomic variables in a no-arbitrage setting. To oversimplify, the results of this research are mixed. Many tests find no relation between expected excess returns and a wide variety of macroeconomic variables. Others, especially recent work using either long-horizon return regressions or dynamic term structure models, find strong ties between term premia and the macroeconomy.

In this paper I argue that for the purpose of inferring a relation between term premia and the macroeconomy, all of these tests use an irrelevant null hypothesis. Either explicitly

[1] Term premia can vary either because of variations in expected excess returns or variations in conditional variances of yields, through Jensen's inequality. The focus in this paper, as in almost all of the literature on term premia, is on variations in term premia associated with the former channel.

or implicitly, existing research uses as its null the hypothesis that excess bond returns are uncorrelated across time. The alternative hypothesis is that expected excess returns vary with the macroeconomy. But that null hypothesis has already been strongly rejected. The current debate should not be about predictability of excess returns; it should be about the sources of predictability. A more relevant null hypothesis is that expected excess bond returns are stochastic, persistent, and independent of the macroeconomy. To simplify the exposition, I refer to the former null hypothesis as the restrictive null and the latter null hypothesis as the general null.

In principle, regression-based tests can incorporate the general null hypothesis by adjusting the covariance matrix of parameter estimates for persistence in the residuals. For these regressions, the important question is whether finite-sample properties of test statistics are similar to the asymptotic properties. The consequences for estimation of dynamic term structure models are more severe. Existing dynamic models typically rule out the general null hypothesis by construction. In other words, the models offer only a choice between term premia that are correlated with macroeconomic variables and term premia that are constant over time.

I develop a dynamic term structure model that satisfies the general null hypothesis. The model is nested in a broader model that satisfies the alternative hypothesis, in which term premia are imperfectly correlated with macro variables. I apply the model to the joint behavior of inflation, output growth, and Treasury yields. Using U.S. data from 1961 through 2005, I estimate the model imposing the general null hypothesis, as well as the model that does not impose this restriction. For comparison, I also estimate the more standard macro-finance dynamic term structure model, in which term premia vary only with macro variables.

I find that the standard model—which assumes that term premia depend only on inflation, output growth, and the short rate—is grossly inconsistent with the data. The general null—that term premia vary, but do not depend on these three variables—is statistically rejected in favor of the alternative hypothesis. However, the economic significance of the rejection is weak, in the sense that little of the variation in expected excess returns is associated with these three variables. This is consistent with Cochrane and Piazzesi (2005), who note that the factors that explain the vast majority of time-variation in yields are not important for explaining variations in expected excess returns. Moreover, forecasts of excess returns produced by the model satisfying the general null are more plausible than those produced by the alternative hypothesis. The alternative model appears to overfit substantially the behavior of expected returns during the Fed's monetarist experiment.

Armed with the estimates of these models, I reconsider the evidence of return-forecasting regressions. Monte Carlo simulations generate finite-sample distributions of the regressions' test statistics. These distributions are calculated both under the restrictive and general nulls. Finite-sample distributions based on the general null differ sharply from both their corresponding asymptotic distributions and the finite-sample distributions based on the restrictive null. Consider, for example, an out-of-sample test of a forecasting regression. Finite-sample rejection rates at five percent asymptotic critical values can exceed twenty percent when calculated using the general null, even though rejection rates are close to five percent when calculated using the restrictive null. The underlying source of the poor performance is that under the general null, variations in expected excess returns are highly persistent. Even test statistics produced with out-of-sample forecasting regressions have poor finite-sample properties, in part because the standard assumption that true residuals are orthogonal to each other does not hold.

The next section discusses forecasting regressions. It reviews both methodological approaches and existing evidence, then presents some new results. Section 3 discusses dynamic term structure models. It also reviews existing evidence, then presents a new dynamic model. The model is estimated in Section 4. Section 5 uses the model to construct finite-sample distributions of test statistics from forecasting regressions. The final section concludes.

2 Forecasting regressions

This section discusses the use of forecasting regressions to test whether expected excess bond returns covary with macroeconomic conditions. The first subsection describes the standard methodological approach and reviews earlier evidence. It concludes that the existing literature does not test the general null hypothesis against the alternative hypothesis that particular macro variables are correlated with expected excess bond returns. The second subsection helps to fill this gap in the literature. To preview the results, both in-sample and out-of-sample regressions indicate that annual excess returns are predictable with a combination of inflation, output growth, and the short rate. Regressions with quarterly excess returns do not support this conclusion.

2.1 The standard approach

The main goal of this strand of research is to identify variables that help predict future excess returns to bonds (and perhaps other assets). The earliest work includes Kessel (1965) and Van Horne (1965). Foreshadowing the debate to come, Kessel finds that term premia are positively associated with the level of interest rates and Van Horne finds the opposite;

both claim their results are consistent with economic theory.

The modern literature begins with Fama and Schwert (1977), who ask whether excess returns are forecastable with short-term nominal interest rates. They estimate a regression that can be written as

$$R_{i,t+1} - R_t^f = b_0 + b_1 R_t^f + e_{i,t+1} \tag{1}$$

where $R_{i,t+1}$ is the simple return to bond i from period t to period $t+1$ and R_t^f is the simple riskfree return from t to $t+1$, which is known at t.[2] The null hypothesis is that expected excess returns are constant, so b_1 is zero and the residuals are serially uncorrelated. They conclude that short-term interest rates cannot predict excess returns to Treasury bonds over return horizons ranging from one to six months.

During the 1970s and 1980s, researchers actively debated the existence of time-varying expected excess asset returns. The assumption of serially uncorrelated residuals is appropriate in that context, and is adopted in almost all of the articles summarized here. Nonetheless, Fama and Schwert calculate sample autocorrelations of fitted residuals and note that their persistence implies the presence of time-varying expected returns that are uncorrelated with the short-term interest rate.

Part of this early literature follows Fama (1976) by using measures of volatility to predict excess returns.[3] Such regressions appear in Shiller et al. (1983), Lauterbach (1989), and Klemkosky and Pilotte (1992). A broad summary of the results is that measures of volatility have only weak forecast power for excess returns. A more successful approach follows Fama (1984) in using information from forward rates to forecast returns. The classic references are Fama and Bliss (1987), Campbell (1987), and Stambaugh (1988). More recent evidence is in Cochrane and Piazzesi (2005). This research, conducted under the null that excess returns are unforecastable, conclusively rejects the null.

The same null hypothesis can be rejected using other forecasting variables. Keim and Stambaugh (1986) and Fama and French (1989) find that variables derived from stock prices predict both excess stock returns and excess bond returns.[4] An appealing interpretation of this result is that variations in term premia are driven by the business cycle, but by itself, the evidence is inconclusive. The accounting relation that limits our ability to interpret the forecast power of forward rates and spreads applies as well to the forecast power of variables

[2] They actually regress nominal returns on the riskfree return and ask whether the estimated coefficient differs from one.

[3] Unlike Fama and Schwert, Fama does not regress excess returns on truly predetermined variables. He regresses time-t excess returns on an estimate of time-t interest-rate volatility that uses information realized after time $t-1$.

[4] Most of this evidence is based exclusively on U.S. data. A notable exception is Ilmanen (1995), who uses term spreads and stock-price variables to predict excess returns to bonds issued by a variety of governments.

based on stock prices. The dividend/price decomposition described in Campbell and Shiller (1988) says that there is a mechanical relation between today's stock price and expectations of future returns. Thus variables constructed using today's stock prices will have forecast power for future excess bond returns as long as expected excess returns on stocks and bonds have common components, regardless of the reasons for the common components.

Ferson and Harvey (1991, 1993), Baker et al. (2003), and Ludvigson and Ng (2005) forecast excess bond returns using variables derived from the prices of risky securities and other variables that are related to macroeconomic conditions. For simplicity I refer to the former variables as price-based variables. In these regressions, variables related to the macroeconomy such as short-term interest rates, inflation, and measures of output growth often contribute to the forecasting power of the regression. The statistical evidence in Baker et al. (2003) and Ludvigson and Ng (2005) is particularly strong. Both papers look at excess returns over holding periods of at least a year. These two papers also break from the tradition of Fama and Schwert (1977) by explicitly accounting for general serial correlation in the regression residuals. However, their discussions of the finite-sample properties of their techniques rely on the restrictive null adopted by Fama and Schwert.

Even if we ignore statistical issues associated with these forecasting regressions, their results do not demonstrate that macroeconomic variables are correlated with expected excess bond returns. In the language of least squares, these regressions reveal partial correlations instead of unconditional correlations. The reason why these two measures differ is straightforward. Price-based variables are noisy measures of expected excess returns. For example, yield spreads depend on both expected excess bond returns and expected changes in future short rates. If the macroeconomic variables are correlated with the noise (e.g., today's short rate is correlated with expected changes in future short rates), they will help forecast excess returns in such regressions even if such variables are independent of expected excess returns. In order to be sure that the macro variables have independent forecasting power, they must appear in the regression without price-based variables.[5]

Aside from Fama and Schwert (1977), there is little direct evidence in the literature concerning the forecast power of exclusively non-price-based macroeconomic variables. The closest references are Friedman (1979) and Huizinga and Mishkin (1984). Friedman relates expected excess returns to macroeconomic activity, but he measures expected excess returns using forward rates less survey forecasts of short-term interest rates. Unlike Fama and Schwert, he finds term premia are related to short-term interest rates, but not to other macroeconomic measures. Huizinga and Mishkin use inflation to forecast real returns, but

[5]The methodology adopted in these papers is consistent with their primary objective, which is effectively to maximize the variation in excess bond returns that can be explained by lagged variables.

not excess returns, on a variety of assets. The next subsection helps fill this gap in the literature. It also gives us a benchmark with which to evaluate the role of the null hypothesis in forecasting regressions.

2.2 Some new evidence

This subsection uses inflation, output growth, and short-term interest rates to forecast excess bond returns. Tests of the hypothesis that these variables have no forecast power are conducted using both the restrictive null that excess returns are unforecastable and the general null that excess returns are stochastic and uncorrelated with the explanatory variables.

2.2.1 Data

The data are quarterly from 1961Q2 through 2005Q4. Inflation, denoted π_t, is the annualized log change in the GDP price deflator from quarter $t-1$ to quarter t. Output growth, denoted Δg_t, is the annualized log change in real GDP from quarter $t-1$ to quarter t. The nominal short rate, denoted r_t, is the annualized yield on the three-month Treasury bill as of the end of the quarter.

The literature on forecasting bond returns uses two types of excess returns. One approach follows Fama and Bliss (1987) by using annual log returns to zero-coupon Treasury bonds in excess of the yield on a one-year zero-coupon Treasury bond. The annual return horizon implies that the regressions must use either a fairly small number of observations or use overlapping observations. Another approach uses shorter-horizon returns, as in Keim and Stambaugh (1986). I consider both types of excess returns here.

Denote the annualized yield on an n-quarter zero-coupon bond at the end of quarter t by $y_t^{(n)}$. The log return to this bond over the next year (i.e., from the end of quarter t to the end of quarter $t+4$) less the yield on a one-year zero-coupon Treasury bond is

$$rx_{t,t+4}^{(n)} = \left(\frac{n}{4}\right) y_t^{(n)} - \quad \frac{n-4}{4} \quad y_{t+4}^{(n-4)} - y_t^{(4)}. \tag{2}$$

The lower case rx denotes a log return. The zero-coupon bond yields are from the Federal Reserve Board. The starting date of the sample is determined by the availability of these data.

I use monthly returns to maturity-sorted Treasury portfolios (from the Center for Research in Security Prices) to construct quarterly excess returns. Denoting the simple net return to portfolio p in month m of quarter $t+1$ as $R_{t+1(m)}^{p}$, the quarterly simple gross return from the end of quarter t to the end of quarter $t+1$ is

$$R^p_{t,t+1} \equiv \left(1 + R^p_{t+1(1)}\right)\left(1 + R^p_{t+1(2)}\right)\left(1 + R^p_{t+1(3)}\right). \tag{3}$$

This return corresponds to rolling over a position in portfolio p every month. Simple excess quarterly returns are produced by subtracting the simple gross return to a three-month Treasury bill that matures at the end of the quarter, or

$$RX^p_{t,t+1} = R^p_{t,t+1} - \exp(y^{(1)}_t/4). \tag{4}$$

The upper case RX denotes a simple return.

2.2.2 Regressions

Excess bond returns $rx^{(n)}_{t,t+4}$ and $RX^p_{t,t+1}$ are regressed on quarter-t values of inflation, output growth, and the short rate. The regressions for annual excess returns use overlapping observations. The regressions are estimated using the entire sample of 1961 through 2005 as well as the more recent sample of 1985 through 2005. The shorter sample is singled out because of the evidence of regime changes over the full sample, as discussed in more detail in Section 4.1. Regime changes do not invalidate these regressions because the orthogonality conditions are unaffected. However, they affect estimates of dynamic term structure models, and one of the goals of this exercise is to compare results from forecasting regressions with model-based estimates.

To limit the size of the tables I consider only two maturities. For annual returns, they are the bonds with original maturities of two and seven years. For quarterly returns, they are the portfolios with original maturities between two and three years and five and ten years.

For each regression, a Wald test is constructed of the hypothesis that the coefficients on the predetermined variables are all zero. This hypothesis is embedded in two different maintained hypotheses: The restrictive null that forecast errors are serially uncorrelated and the general null that forecast errors contain persistent components unrelated to the explanatory variables. For the restrictive null, the robust Hansen-Hodrick method is used to estimate the covariance matrix of parameter estimates (Hansen and Hodrick (1980), Ang and Bekaert (2006)). With this null the test is asymptotically distributed as a $\chi^2(3)$. For the general null, the method of Newey and West (1987) is used with four lags for quarterly returns and seven lags for annual returns. These choices of Newey-West lag lengths are arbitrary, but alternative choices do not lead to qualitatively different results. This test is also asymptotically distributed as a $\chi^2(3)$ if the lag length captures all of the serial correlation in the residual.

2.2.3 Forecasting out of sample

I use the out of sample forecast encompassing test of Ericsson (1992) as an additional test of these regressions. The following description of the procedure applies to forecasts of annual excess returns. The procedure for quarterly excess returns is slightly simpler because the observations do not overlap.

Estimate the return-forecasting regression using observations 1 through R of the macro variables and observations $rx_{1,5}^{(n)}$ through $rx_{R,R+4}^{(n)}$ of annual excess returns. Given the estimated parameters, forecast $rx_{R+4,R+8}^{(n)}$ using observation $R+4$ of the macro variables. Denote the realized forecast error by $u_{un,1}^{(n)}$, where the first subscript refers to a forecast error from an unrestricted regression. Then repeat this exercise using an additional observation, so that the new regression uses observations 1 through $R+1$, and so on. The result is a time series of one-step-ahead forecast errors $u_{un,t}^{(n)}$ with length $P = T - R - 7$, where T is the total number of quarters in the sample period. Construct a time series of restricted forecast errors $u_{r,t}^{(n)}$ using the same methodology, where the forecasting regression uses only a constant term. The first subscript refers to a forecast error from a restricted regression.

The test statistic is the t statistic of a regression of $u_{r,t}^{(n)}$ on $u_{r,t}^{(n)} - u_{un,t}^{(n)}$. No constant term is included in the regression. Under the restrictive null assumption that returns are serially uncorrelated (aside from induced serial correlation through the use of overlapping observations), the asymptotic distribution of the statistic is approximately standard normal.[6] The alternative hypothesis is that the statistic exceeds zero. Therefore the null of no forecastability is tested using the one-sided critical value for a normal distribution.

For the full sample 1961 through 2005, I use $R = 100$, thus $P = 72$ for annual excess returns and $P = 78$ for quarterly excess returns. I do not apply this procedure to the shorter sample because there are insufficient data to both reliably estimate the return-forecasting regression and construct a reasonably long time series of out of sample forecasts. I construct the t statistic with robust Hansen-Hodrick standard errors.

The distribution of this statistic under the general null is not known. More importantly, the statistic is not appropriate to tests of the general null because the forecasts are not truly out of sample. If both the forecasting variables and true expected excess returns are persistent, then in-sample predictability will correspond to out-of-sample predictability. Consider, for example, forecasting excess returns with inflation when the true data-generating

[6] The precise asymptotic distribution of the statistic depends on the asymptotic ratio of P/R, but the results of Clark and McCracken (2001) indicate that critical values from a standard normal distribution are reasonably accurate (although slightly conservative) for the case of three forecasting variables.

process implies that expected excess returns are determined by some independent, persistent variable ω_t. If ω_t and inflation are correlated in a sample, the in-sample regression will find that inflation forecasts expected excess returns. Then one-step-ahead expected excess returns and inflation are also likely to be correlated because both are persistent.

2.2.4 Results

Tables 1 and 2 present results for annual and quarterly return horizons, respectively. The results in Table 1 show over the full sample of 1961 through 2005, inflation, output growth, and the short rate collectively have substantial information about future excess returns. For both two-year and seven-year bonds, the R^2 in the full sample exceeds 10 percent, and statistical tests that the coefficients are all zero are rejected at the asymptotic one percent level. The choice of restrictive versus general null makes little difference to the statistical strength of this rejection. In addition, tests for out-of-sample forecastability reject the restrictive null. (The five percent critical value is 1.645.) Recall, though, that the properties of this test are unknown under the general null. The forecast power is less impressive in the most recent sample. Over 1985 through 2005, only excess returns to the two-year bond appear forecastable. Surprisingly, the statistical strength of the forecastability appears stronger under the general null than under the restrictive null.

The results of Table 2 temper the apparently strong evidence of Table 1. Using quarterly returns over 1961 through 2005, the joint statistical significance of the three forecasting variables is marginal at best. The test statistics for the in-sample regressions are all in the neighborhood of 10 percent asymptotic critical values. Evidence of forecastability from the out-of-sample tests is even weaker. For the 1985 through 2005 sample, any evidence of predictability from in-sample regressions disappears.

On balance, these results are muddled. We might be tempted to downweight the annual results relative to the quarterly results because of well-known statistical problems with regressions involving overlapping regressions. However, there may truly be more evidence of predictability using the annual excess returns because of the definition of "excess" for these returns (subtracting the one-year yield instead of the three-month yield). Both the annual and quarterly in-sample regressions are subject to the predictive regressions bias of Stambaugh (1999), but not the out-of-sample regressions.

Knowledge of the finite-sample distributions of the test statistics helps to better evaluate this regression evidence. Monte Carlo simulations are commonly used to evaluate the accuracy of asymptotic inference. To generate such simulations we need a joint model of the term structure, inflation, and output growth that satisfies the relevant null hypothesis. The development of such models is discussed in the next section. The finite-sample properties of the regressions estimated here are discussed in Section 5.

3 Dynamic term structure models

This section describes how dynamic term structure models are used in practice to draw inferences about the determinants of term premia; it also describes how they should be used to draw such inferences. The first subsection describes standard models and reviews earlier evidence. The second subsection explains in more detail how dynamic models can be used to test hypotheses about term premia. The third subsection develops a dynamic term structure model that satisfies both the general null hypothesis and (by relaxing some parameter restrictions) the alternative hypothesis that term premia are correlated with inflation, output growth, and the short rate.

3.1 The standard approach

Ang and Piazzesi (2003) construct a model that describes the joint dynamics of the term structure, inflation, and real activity, while simultaneously guaranteeing the absence of arbitrage opportunities in the bond market. This line of research has grown explosively in the past few years.[7] Before discussing the implications of these models for term premia, it is helpful to address a semantic issue. Ang and Piazzesi refer to inflation and real activity as "macro" variables, and distinguish them from three "latent" variables. Both macro and latent variables determine term structure dynamics. The language of this decomposition is unusual because the short-term interest rate is typically also viewed as a macro variable reflecting monetary policy. Evans and Marshall (2002) contrast the more typical decomposition with that of Ang and Piazzesi. This is purely semantic because the three latent variables in Ang and Piazzesi can be rotated into the short rate (perhaps observed with noise) and two other latent variables. The discussion in this section treats the short rate as a macro variable.

Following Ang and Piazzesi, a common dynamic modeling approach assumes that a low-dimensional state vector drives the joint dynamics of the term structure and a few macro variables such as inflation and/or output growth. Given parameter estimates of the model, we can calculate properties of the term structure, such as the fraction of variation in expected excess returns on an n-period bond that is attributable to shocks to each element of the state vector.

[7] Recent work includes Rudebusch and Wu (2004, 2005), Ang and Bekaert (2005), Dewachter and Lyrio (2004a), Dewachter et al. (2004b), and Hördahl et al. (2005a,b)

Depending on the chosen functional form, expected excess bond returns are said to be closely associated with short-term rates, inflation, output growth, or employment growth. For example, Ang, Dong, and Piazzesi (2005) argue that more than half of the variation in expected excess quarterly returns to five-year bonds is driven by the level of inflation. Ang, Piazzesi, and Wei (2005) find extremely strong statistical evidence linking both the level of the short rate and output growth to term premia. Law (2004) finds that all variation in term premia is driven by real economic activity, inflation, and monetary policy (the Fed funds rate).

These apparently strong, yet conflicting results about term premia raise a red flag. Even setting aside this concern, it is difficult to draw conclusions about term premia from this evidence because the models are not designed to address specific null hypotheses. The typical paper sets up its preferred model, specifies some parameter restrictions for tractability, then estimates the model. The implications of the parameter estimates are then summarized and interpreted. In particular, there is no discussion of what parameter restrictions are required for expected excess returns to be unforecastable with macro variables, nor are there tests of such restrictions. The next subsection describes some of the relevant issues.

3.2 Testing hypotheses about term premia

How can dynamic term structure models be used to test formally the hypothesis that a particular vector of observed macroeconomic variables f_t is unrelated to expected excess bond returns? To answer this question, it helps to recall how we use forecasting regressions to test this hypothesis. We regress a bond's excess return in $t+1$ on f_t. Under the null, future excess returns do not covary with f_t. The power of this test depends, in part, on the standard deviation of the innovation component of the bond's return. A larger standard deviation corresponds to reduced power because it corresponds to more noise in sample covariances.

Now consider testing this hypothesis using a complete dynamic model of both bond yields and the macro variables. In principle, this is a more powerful approach than a forecasting regression because the model exploits information from the bond's return innovation. This innovation can be decomposed into two types of news: news about future short rates and news about expected future excess returns. If expected excess returns are unrelated to f_t, then innovations in f_t are orthogonal to innovations in future expected excess returns. Thus under the null, contemporaneous covariances between innovations in f_t and news about expected future excess returns are zero. Hence the complete dynamic model can use information both from lead-lag covariances (as in forecasting regressions) and contemporaneous covariances.

Note, though, that neither return innovations nor the two types of news can be observed directly. Instead, they are all artifically constructed as parameterized functions of a state vector, where the functions are forced to be internally consistent (i.e., the generated return innovation must equal the sum of the two types of generated news). Thus any null and alternative hypotheses that the econometrician wants to test must be built into the functional form of news about expected excess returns. To test hypotheses concerning the relation between expected excess returns and the macro variables, the model needs to be sufficiently flexible to allow news about expected excess returns to have either a zero contemporaneous covariance with f_t (the null) or a nonzero covariance.

Unfortunately, almost all existing models used to describe the joint dynamics of bond yields and macro variables lack this flexibility. The models are so parsimonious that they rule out the possibility that term premia vary independently of the macro variables included in the model. For example, Ang, Dong, and Piazzesi (2005) use three state variables that are equivalent to output growth, inflation, and an unobserved short-term interest rate. Hence the only possible news about future expected excess returns must be macroeconomic news. Law (2004) allows for an additional state variable to capture similar dynamics, but imposes parameter restrictions that rule out the possibility of stochastic expected excess returns that are independent of the macroeconomy. An exception is Duffee (2006), who imposes no restrictions on the dimension of the state vector. He estimates part of the joint dynamics of inflation and the term structure and finds almost no relation between inflation and term premia.

The next subsection presents an example of a dynamic term structure model that is sufficiently flexible to allow expected excess returns to vary, either independently of inflation, output growth, and the short rate, or predictably with any of these variables. Although necessarily less parsimonious than the typical model in the literature, it is sufficiently tractable to estimate and to use in Monte Carlo simulations.

3.3 A new dynamic model

As in Ang and Piazzesi (2003), the period-t term structure and state of the economy is determined by a state vector with two types of factors. I refer to them as “macro” and “term premia” factors. The only role played by the term premia factors is to capture variations in expected excess returns that are unrelated to the macro factors.

3.3.1 Factors and factor dynamics

The macro factors are inflation, output growth, and the continuously-compounded short rate. They are stacked in a vector

$$f_t = \begin{pmatrix} \tilde{\pi}_t & \tilde{\Delta g_t} & \tilde{r}_t \end{pmatrix}'. \tag{5}$$

The tildes distinguish these factors from observed inflation, output growth, and the short rate. The relation between f_t and observed macro variables is established later. For now it is sufficient to note that a Kalman filter setting is used, so we can think of the difference between f_t and its observed counterpart as measurement error.

There are three term premia factors stacked in a vector ω_t. (The choice of three is dictated by the number of macro factors, as discussed in the context of equation (16) below.) The complete state vector is

$$x_t = \begin{pmatrix} f_t' & \omega_t' \end{pmatrix}'. \tag{6}$$

Because the short rate is included in x_t, the loading of the short rate on x_t has a simple form. Using standard affine term structure notation, it is

$$\tilde{r}_t = \delta_x' x_t, \qquad \delta_x' = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}. \tag{7}$$

The evolution of the state vector in this discrete-time model is described by a Gaussian vector autoregression. Formally, the dynamics are

$$x_t = \mu_x + K_x x_{t-1} + \Sigma_x \;{}_{x,t}, \epsilon \quad {}_{x,t} \sim N(0, I). \tag{8}$$

The specific matrices are given by

$$\mu_x = \begin{matrix} \mu_f \\ 0_{3\times1} \end{matrix}, \quad K_x = \begin{matrix} K_f & 0_{3\times3} \\ 0_{3\times3} & K_\omega \end{matrix}, \quad \Sigma_x = \begin{matrix} \Sigma_f & 0_{3\times3} \\ 0_{3\times3} & \Sigma_\omega \end{matrix}. \tag{9}$$

The matrices K_f K_ω, Σ_f, and Σ_ω are 3×3. Both Σ_f and Σ_ω are lower triangular. The process is assumed to generate stationary dynamics, so that the unconditional expectation of x_t is

$$E(x) = (I - K_x)^{-1}\mu_x. \tag{10}$$

With this specification, f_t and ω_t are independent. Therefore there is no information in the term premia factors about the evolution of the short rate. If investors were risk-neutral, bond prices would be determined exclusively by f_t. Thus the only role played by the term premia factors is to drive expected excess returns.

3.3.2 Bond pricing

The period-t price of a zero-coupon bond that pays a dollar at the end of period $t+n$ is given by the law of one price,

$$\tilde{P}_t^{(n)} = E_t\left(\tilde{P}_{t+1}^{(n-1)} M_{t+1}\right) \tag{11}$$

where M_t is the stochastic discount factor. Again, tildes represent true prices. Actual prices are observed with measurement error. The stochastic discount factor is

$$M_{t+1} = \exp \quad -h\tilde{r}_t - \lambda_t' \;{}_{x,t+1} - \frac{1}{2}\lambda_t'\lambda_t \tag{12}$$

where h is the length of a period (in years), and λ_t is the period-t compensation investors require to face factor risk. The functional form for λ_t is in the essentially affine class of Duffee (2002),

$$\Sigma_x \lambda_t = \lambda_0 + \lambda_1 x_t. \tag{13}$$

The parameterizations of the vector λ_0 and the matrix λ_1 are

$$\lambda_0 = \begin{matrix} \lambda_{0f} \\ 0_{3\times 1} \end{matrix} \tag{14}$$

and

$$\lambda_1 = \begin{matrix} \lambda_{1f} & I_{3\times 3} \\ 0_{3\times 3} & 0_{3\times 3} \end{matrix} \quad , \tag{15}$$

where λ_{0f} is a vector of length three and λ_{1f} is a 3×3 matrix. The identity matrix in the upper right quadrant of λ_1 is a normalization. (We can always transform a model with an arbitrary invertible matrix L in the upper right quadrant into a model with the identity matrix in the upper right quadrant by redefining the term premia factors as $\omega_t^* = L\omega_t$.)

An alternative, and perhaps more intuitive, representation of the compensation investors demand to face uncertainty in x_t is

$$\Sigma_x \lambda_t = \begin{matrix} \lambda_{0f} + \lambda_{1f} f_t + \omega_t \\ 0_{3\times 1} \end{matrix} \tag{16}$$

The top element on the right of (16) is the compensation investors demand to face macro risk. The compensation depends on the macro factors through λ_{1f} and on the term premia factors. Term premia factor i affects only the risk compensation for macro factor i. (This is why the number of term premia factors equals the number of macro factors.) Investors require no compensation to face uncertainty in the term premia factors.

Under the equivalent martingale measure, the dynamics of x_t are

$$x_t = \mu_x^q + K_x^q x_{t-1} + \Sigma_x \epsilon^q_{x,t}, \quad \epsilon^q_{x,t} \sim N(0, I), \tag{17}$$

where

$$\mu_x^q = \mu_x - \lambda_0, \quad K_x^q = K_x - \lambda_1. \tag{18}$$

Log bond prices are affine in the state vector. Using lower case to denote log prices, the notation is

$$\tilde{p}_t^{(n)} = A_n + B_n' x_t. \tag{19}$$

Solving recursively using the law of one price, the loadings of the log bond price on the factors are given by

$$B_n' = -h\delta_x' \left(I - K_x^q\right)^{-1} \left(I - \left(K_x^q\right)^n\right). \tag{20}$$

The constant term is

$$A_n = -h\delta_x' \left[nI - \left(I - K_x^q\right)^{-1} \left(I - \left(K_x^q\right)^n\right) \right] E^q(x) + \frac{1}{2} \sum_{i=1}^{n-1} B_i' \Sigma_x \Sigma_x' B_i, \quad n = 2, \ldots \tag{21}$$

with $A_1 = 0$. The notation $E^q(x)$ denotes the equivalent-martingale unconditional expectation of x and is the counterpart of (10).

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

$$\tilde{p}_{t+1}^{(n-1)} - \tilde{p}_t^{(n)} = h\tilde{r}_t + B_{n-1}'(\lambda_0 + \lambda_1 x_t) - \frac{1}{2} B_{n-1}' \Sigma_x \Sigma_x' B_{n-1} + B_{n-1}' \Sigma_x \epsilon_{x,t+1}. \tag{22}$$

The log of the gross expected return to an n-period bond from t to $t+1$ is

$$\log E_t \left(\tilde{P}_{t+1}^{(n-1)} / \tilde{P}_t^{(n)} \right) = h\tilde{r}_t + B_{n-1}'(\lambda_0 + \lambda_1 x_t). \tag{23}$$

The first term on the right of (23) is the riskfree return and the second is the time-varying compensation investors require to face uncertainty in x_t. The exposure to x_t is B_{n-1}' and the compensation per unit of x_t risk is, from (13), $\Sigma_x \lambda_t$.

3.3.3 From factors to observables

I use a state-space setting to relate the model's factors to observable macro variables and bond yields. At the end of each quarter, an econometrician observes inflation π_t, output growth Δg_t, the short rate r_t, and d yields on multiperiod bonds with maturities $n_1, \ldots, n_d$. Stack these observables in a vector

$$z_t = \begin{pmatrix} \pi_t & \Delta g_t & r_t & y_t^{(n_1)} & \ldots & y_t^{(n_d)} \end{pmatrix}'. \tag{24}$$

The relation between factors and observables is

$$z_t = \begin{pmatrix} 0_{3\times 1} \\ A_y \end{pmatrix} + \begin{pmatrix} I_{3\times 3} & 0_{3\times 3} \\ B_{fy} & B_{\omega y} \end{pmatrix} \begin{pmatrix} f_t \\ \omega_t \end{pmatrix} + \eta_t, \quad \eta_t \sim N(0, H). \tag{25}$$

The vector A_y and matrices B_{fy} and $B_{\omega y}$ are

$$A_y = -\frac{1}{h} \begin{pmatrix} \frac{1}{n_1} A_{n_1} \\ \ldots \\ \frac{1}{n_d} A_{n_d} \end{pmatrix}, \quad \begin{pmatrix} B_{fy} & B_{\omega y} \end{pmatrix} = -\frac{1}{h} \begin{pmatrix} \frac{1}{n_1} B'_{n_1} \\ \ldots \\ \frac{1}{n_d} B'_{n_d} \end{pmatrix}. \tag{26}$$

In the state-space setting the usual interpretation of η_t is measurement error. For inflation and output growth, a broader interpretation is more reasonable. Observed inflation consists of an underlying level of core inflation and transitory inflation shocks owing to short-lived factors such as temporary refinery capacity problems. Bond yields at the end of period t are unaffected by the transitory inflation shock in period t because investors know it will not persist. Similarly, observed output growth consists of a core component and a transitory component due to, say, weather-related shocks to consumer spending.

3.3.4 Discussion

This model will look familiar to those who follow the details of macro-finance dynamic models. If we remove the term premia factors, it is the Taylor rule model of Ang, Dong, and Piazzesi (2005).[8] The only difference between their model and the model here is the added generality to risk compensation. In their model, required compensation to face the risk of, say, inflation is determined by the levels of inflation, output growth, and the short rate. Here compensation is also allowed to depend on a latent factor that has dynamics independent of the macro factors. With the restriction $\lambda_{1f} = 0$ in (15) and (16), expected

[8] This is strictly true only after rotating their factors so that their latent factor is identical to the unobserved short rate, but this is without loss of generality.

excess bond returns are stochastic, persistent, imperfectly correlated across bond maturities, and independent of the macro factors. In this special case, which corresponds to the general null defined in the introduction of this paper, the magnitude of shocks to expected excess returns are determined by the volatility matrix Σ_ω and their persistence is determined by the feedback matrix K_ω. This restriction can be tested against the alternative hypothesis $\lambda_{1f} = 0$.

Naturally, the specification of risk premia in (16) is critical to distinguishing between macro and non-macro influences on term premia. But independence between the macro and term premia factors is also vital. To understand why, adopt the restriction of the general null $\lambda_{1f} = 0$ but replace the matrix of zeros in the upper right quadrant of K_x in (9) with free parameters. Then the evolution of the term premia factors depends only on term premia factors, but the evolution of the macro factors depends on both sets of factors.

With this alternative model, shocks to the macro factors are independent of expected excess returns at all leads and lags. Therefore a variance decomposition of expected excess returns assigns all of the variance to shocks to the term premia factors. But such a decomposition is not the right way to think about the model. A more appropriate perspective follows the projection decomposition of Bikbov and Chernov (2006). The intuition behind their projection is the same as the intuition of the forecasting regressions discussed in Section 2: is there information in macro factors about future excess returns? With this alternative model, expected excess returns from t to $t+1$ are no longer orthogonal to the history of macro factors $f_t, f_{t-1}, \ldots$, hence the model does not satisfy the general null notwithstanding the restriction $\lambda_{1f} = 0$.

A limitation of this model is that there is no additional information in the term structure that helps to forecast inflation, output growth, or the short rate. The history of these macro variables is sufficient to form minimum-variance forecasts. However, it is straightforward to relax this limitation without altering the interpretation of the term premia factors. Simply expand the vector of fundamentals f_t to include one or more latent factors, and expand the dimensions of the matrices in (9), (14), and (15) accordingly. These latent factors reflect information that investors have about the evolution of the macro factors that is not contained in the macro factors themselves. They are identified from the term structure. Under the general null, the additional factors do not affect risk premia. I do not attempt to estimate this expanded version here.

Another limitation of the model is its Gaussian structure. Although Gaussian models are standard in the macro-finance literature, innovations in observed bond yields are not homoskedastic. Gaussian models are used both because they are simple and because the research focuses on capturing the dynamics of expected excess returns. Gaussian models offer great flexibility in fitting these dynamics, while researchers have only recently worked out the mathematics of non-Gaussian term structure models with flexible specifications of risk premia. Cheridito et al. (2005) and Dai et al. (2006) develop flexible non-Gaussian specifications. Presumably the main methodological point of this paper—that the general null should be used instead of the restrictive null—can be implemented in such non-Gaussian models, but I do not investigate this issue.

Perhaps the most important objection to this model is that it offers no economic intuition for the presence of the term premia factors. The lack of intuition has led some readers to call this a nihilistic model of term premia. Buraschi and Jiltsov (2005) and Wachter (2006) are examples of the alternative approach, in which investors' preferences are expressly tied to the money supply or real consumption. But the model does not say that variations in term premia have no economic foundation. Instead, the model is a diagonostic tool to help us determine whether an econometrician has identified that foundation correctly. In this sense, the model is an intermediate step in the direction of a correctly specified economic model of premia, not an end in itself.

4 Model estimation

This section applies the framework of Section 3.3 to U.S. data. Three versions are estimated with maximum likelihood (ML). The first is the standard Taylor rule macro-finance model, where the term premia factors are excluded. In this version, expected excess returns vary only with the macro factors. The second version satisfies the general null hypothesis. Term premia factors are included, and are the only factors allowed to affect expected excess returns. The third version satisfies the alternative hypothesis, where all factors can affect expected excess returns. To simplify the exposition, I refer to these as the "standard," "general null," and "alternative" models respectively.

A roadmap to this long section is useful. The first subsection describes the data and the two sample periods over which the models are estimated. The second subsection summarizes the parameters to be estimated and the third explains the estimation procedure. The fourth subsection describes how information about the estimated models is presented in various tables. The fifth subsection discusses in detail the results for the full sample period 1961 through 2005. The sixth contains a briefer discussion of results for a shorter, more recent period.

A preview of the results may help to keep the big picture in sight. One clear conclusion is that the standard macro-finance model is markedly inferior—both statistically and economically—to the other models. Another conclusion is that the general null is statistically rejected in favor of the alternative model. But we should be very cautious about reading much into this statistical rejection. The economic significance of the rejection is small, in the sense that little of the variability in expected excess bond returns is associated with variability in inflation, output growth, or the short rate. More importantly, the rejection of the general null hypothesis appears to be due to overfitting. The greater flexibility of the alternative model leads to implausible estimates of expected excess returns.

4.1 Data and sample periods

The data on inflation, output growth, and the short rate are the same used in the forecasting regressions of Section 2.2.1. Yields on zero-coupon Treasury bonds with maturities ranging from one to ten years are from the Federal Reserve Board. The three models are estimated over both the full sample period 1961 through 2005 and the more recent period 1985 through 2005. When estimating over the full sample, I use yields on bonds with maturities of one, two, three, five, and seven years. When estimating over the more recent sample, I also use the yield on a ten-year bond. This yield is not available for every observation in the full sample.

The choice of sample period reflects a tradeoff between statistical power and economic plausibility. The implications of the model for term premia behavior rely on the assumption of parameter stability. The model requires that investors' expectations of future short-term interest rates are given by forecasts from a constant-parameter vector autoregression. Yet there is strong evidence that the past 50 years are not characterized by a single regime. The major break occurred at the beginning of Volcker's tenure as Chairman of the Federal Reserve Board.[9] The accompanying disinflation was largely completed by the end of 1984, although the dynamic regime-switching term structure models of Ang and Bekaert (2005) and Dai et al. (2005) find some evidence of instability after 1984.

Building a regime-switching model that satisfies the general null hypothesis is well beyond the scope of this paper. Therefore I simply estimate the model separately over the two periods and examine informally the economic plausibility of the results.

4.2 Summary of free parameters

The dynamics of the macro factors are determined by the parameters of a vector autoregression. These parameters are contained in the vector μ_f and the matrices K_f and Σ_f. To simplify estimation, the unconditional means of the macro factors are fixed to their sample

[9] See Gray (1996).

means. This restriction pins down μ_f given K_f. There are a total of 15 free parameters in K_f and Σ_f.

In the standard macro-finance version of the model, expected excess bond returns are determined both by the three elements of the vector λ_{0f} in (14) and by the nine elements of the matrix λ_{1f}. The version satisfying the general null hypothesis excludes λ_{1f} and adds the 15 parameters in the matrices K_ω and Σ_ω. The version satisfying the alternative hypothesis adds back the nine elements of λ_{1f}. Hence parameterizing expected excess returns for the standard model requires 12 free parameters, compared with 18 for the general null and 27 for the alternative hypothesis.

Each of these specifications can be reasonably described as too many parameters fitting too few observations. Individual parameters are estimated with little precision, making it impossible to make strong statements about the precise functional form of risk premia. For many purposes, more parsimonious specifications are preferred. But the goal here is to see which class of model is more consistent with the data, not to choose a preferred version of any one of these specifications. Thus I do not estimate any versions of these models that impose additional restrictions on expected excess returns.

I use a diagonal covariance matrix H of measurement error in (25). The standard deviations of measurement error of inflation and output growth are free parameters. Yields other than the riskfree rate share an identical standard deviation of measurement error, which is also a free parameter. The standard deviation of measurement error for the short-term interest rate is fixed at zero.

The main reason I rule out measurement error in the short-term rate is to ensure that the estimation results are informative about the behavior of real-life excess bond returns. The excess log return to an n-period zero-coupon bond is, in terms of its yield,

$$r_{e,t+1}^{(n)} = h[ny_t^{(n)} - (n-1)y_{t+1}^{(n-1)} - r_t]. \tag{27}$$

Decomposing this return into model-implied excess return and measurement error produces

$$r_{e,t+1}^{(n)} = \tilde{r}_{e,t+1}^{(n)} + h\left[n\eta_t^{(n)} - (n-1)\eta_{t+1}^{(n-1)} - \eta_{r,t}\right], \tag{28}$$

where $\tilde{r}_{e,t+1}^{(n)}$ is the model's counterpart to (27), $\eta_t^{(n)}$ is the measurement error for the n-period bond yield at time t, and $\eta_{r,t}$ is the contemporaneous measurement error of the short rate. The overall goal here is to better understand the properties of observed excess returns by studying the properties of model-implied excess returns. The smaller is the term in brackets in (28), the more relevant will be the results of this exercise.

In practice, setting the measurement error of the short rate to zero is sufficient to make the term in brackets very close to zero. Not only is $\eta_{r,t}$ identically zero, but the standard deviation of measurement error in the other bond yields is very small. (These results are discussed in more detail below.) If, however, the standard deviation of measurement error in the short rate is specified as a free parameter, it can exceed 40 basis points, depending on the sample period and the version of the model that is estimated. This choice allows the model to better fit observed inflation and output. Thus the choice made here is to give up some accuracy in the fitting these latter variables in order to better fit bond returns.

To summarize, the version of the model corresponding to the standard macro-finance setup has 30 free parameters, the version satisfying the general null has 36 free parameters, and the version satisfying the alternative hypothesis has 45 free parameters.

4.3 Estimation technique

Maximum likelihood estimation is implemented using the Kalman filter. Nonlinear optimization is required to find the solution. A common problem in estimation of dynamic term structure models is the bumpy surface of the likelihood function. The following procedure is used to find a global maximum. For each version of the model, one hundred different starting values are used. Given a starting value, five iterations of Simplex are used sequentially, followed by a derivative-based algorithm that uses analytic first derivatives. The parameter vector with the highest likelihood value across these 100 starting values is then used as a starting value for an additional five rounds of Simplex and derivative-based optimization.

In an effort to keep the optimization algorithm from exploring parts of the parameter space that are implausible, the likelihood value is set to a very large negative number if the parameters satisfy any of the following conditions. First, at least one eigenvalue of K_x^q is less than minus one. Second, for at least one of the bonds used in estimating the model, the unconditional mean yield exceeds 1.5 times the corresponding sample mean yield. Third, for at least one of the bonds used in estimating the model, the unconditional mean slope of the yield curve (mean bond yield less mean short rate) exceeds twice the corresponding sample mean. The Matlab code used to estimate the model is available on the author's web site.

4.4 A guide to the results

Six sets of parameter estimates are produced, corresponding to three models and two sample periods. To conserve space, I report parameter estimates only for the alternative model. Parameter estimates based on the full and shorter samples are in Tables 3 and 4, respectively.[10] Standard errors, computed using the outer product of first derivatives, are in parentheses.

[10]Estimates for the other models are available on the author's web site.

The only clear message to take from these tables is that, as noted in Section 4.2, there is not enough information in the sample to identify individual parameters associated with term premia. The standard errors on these estimates are extremely large. Tables 5 and 6 are more helpful in evaluating the models. The forms of these two tables are described here, while the results are discussed in the next two subsections.

Table 5 summarizes how the different versions of the model fit the data. We want to know which version has the best specification of compensation for bond risk. Thus two important metrics for evaluating the relative performance of the estimated models are the cross-sectional fit of bond yields and the ability to forecast future bond yields. In practice, ML estimation chooses parameters to optimally trade off accuracy in fitting bond yields with accuracy in fitting the macro variables. Hence misspecification of risk premia can (and in practice, does) show up in a poor fit of the macro variables. Therefore we should also evaluate the ability of the models to fit the macro data.

Table 5 reports log-likelihood values for each model. It also reports sample standard deviations of fitted measurement error for inflation, output growth, and the seven-year bond yield. (Recall that the short rate is observed without error.) The fitted measurement error for quarter t is defined as the actual quarter-t value less the filtered value, where the filtered value is based on information through quarter t. Finally, the table reports root mean squared errors for one-quarter-ahead forecasts of inflation, output growth, the short rate, and the yield on a seven-year bond. These errors are defined in the usual way for a Kalman filter.

Table 6 summarizes some of the sample properties of excess quarterly log bond returns implied by the estimated models.[11] The returns are constructed as follows. For each estimated model, there is a time series of the filtered state vector x_t. Filtered log prices, and therefore filtered yields, can be computed from this state vector using (19). Then a time series of filtered quarterly excess returns to an n-period bond can be computed using the filtered version of (27). The table reports sample means and standard deviations of excess log returns to both two-year and seven-year bonds.

An excess return is the sum of its conditional mean and its shock. Because the terms are uncorrelated, the excess return variance is the sum of the variances of the two components. Formally,

$$\mathrm{Var}\left(\tilde{r}^{(n)}_{e,t+1}\right) = B'_{n-1}\lambda_1 \mathrm{Var}(x_t)\lambda_1' B_{n-1} + B'_{n-1}\Sigma_x\Sigma_x' B_{n-1}. \tag{29}$$

The first term on the right of (29) is the variance of one-quarter-ahead expectations of

[11] We could also examine population properties of the models. But because most of the estimated models have dynamics that are close to unit roots, population properties, especially covariances, are not necessarily similar to in-sample properties. The population properties are highly sensitive to the estimated persistence of the state vector. Therefore these properties are estimated with little precision, making them uninformative measures for model comparison.

quarterly excess returns. Table 6 reports the ratio of this term (computed using the sample variance of the filtered state vector) to the sample variance of excess returns.

The table also reports the fraction of the sample variance of the conditional mean attributable to the sample variance of macro factors. For the standard model, this is identically one. For the general null, it is identically zero. For the alternative model, it is the part of the first term on the right of (29) that is attributable to the upper left (3 x 3) submatrix of the sample variance of x_t.[12]

Finally, Table 6 reports the theoretical first-order autocorrelation coefficient of one-quarter-ahead expectations of quarterly excess returns. This coefficient is

$$\text{AR(1) coefficient} = \frac{B'_{n-1}\lambda_1 K_x \text{Var}(x_t)\lambda'_1 B_{n-1}}{B'_{n-1}\lambda_1 \text{Var}(x_t)\lambda'_1 B_{n-1}}. \tag{30}$$

The results in these tables, along with some supporting evidence, are interpreted in the next two subsections.

4.5 Results for the full sample

Over the period 1961 through 2005, the standard macro-finance model does a very poor job fitting the data relative to the other two models. This is immediately obvious from a glance at the log-likelihoods in Table 5. The log-likelihood of the standard model is more than 650 below the log-likelihoods of the other models. Below I offer an economic interpretation of this failure.

Statistically, the general null is rejected in favor of the alternative model. The difference in log-likelihoods between the general null and alternative model is about 46. As reported in Table 3, this difference allows us to reject the general null in favor of the alternative model at any conventional significance level. Yet although the statistical rejection is strong, the economic significance of the rejection is not. According to Table 6, under the alternative hypothesis macro factors account for only a quarter of the expected excess return variances. The remainder is driven by term premia factors. More importantly, as we shall see, a detailed examination of forecasts from these models reveals that only the general null produces intuitively reasonable estimates of expected excess returns.

[12] In a finite sample, there is also a component attributable to the sample covariance between the macro factors and the term premia factors, but this component is small. In population, this component is zero.

4.5.1 Cross-sectional properties

The poor performance of the standard model over the full sample is most apparent in the model's ability to fit inflation. Table 5 reports that the standard deviation of inflation's measurement error is 1.89 percent with the standard model and only 0.50 percent with the general null. (It is slightly lower with the alternative model, at 0.43 percent.) Visual evidence of measurement errors is in Fig. 1. The figure plots actual inflation and output growth, along with filtered values from the models. Panel A compares actual inflation (black line) with filtered values from the standard model (blue line) and the general null (red line). Filtered values from the alternative model are similar to those from the general null, thus they are not displayed. Filtered values from the model satisfying the general null closely track actual inflation. Their sample correlation is 0.98. However, filtered values from the standard model often diverge substantially from actual inflation. Their correlation is only 0.63.

The intuition behind this evidence is straightforward. Expected excess returns fluctuate over time, but these fluctuations are not closely tracked by the levels of the macro variables. To break the link between macro variables and term premia, the general null and the alternative model include term premia factors. The standard model must resort to breaking the link between observed inflation and "inflation" as measured by the model.

Estimates from all three models describe output growth as the sum of a low-volatility, persistent process and a white-noise shock. Only the former component affects the term structure. Panel C of Fig. 1 displays actual output growth (black line) with filtered values from the standard model (blue line) and the general null (red line). Filtered values from the alternative model are similar to those from the general null. The correlation between actual and filtered values is about 0.2 for the standard model and 0.25 for the general null.

All of the models fit the cross-section of the term structure well. By construction, they all fit the short rate without error. Table 1 reports that measurement error at the long end is small. For the general null and alternative models, the standard deviation of measurement error is less than one-tenth of a basis point. Not surprisingly, the standard model is not as accurate, but its long-bond yield measurement error has a standard deviation of only eight basis points.

4.5.2 Forecasting

Table 5 documents that aside from the standard model's forecast of inflation (which is very poor owing to its high measurement error), the RMSEs of the competing models are all within a few basis points of each other. For example, the lowest RMSE for output growth is only four basis points less than the highest RMSE. Similar differences are seen in the RMSEs of the short rate and the seven-year bond yield.

Since the alternative model nests the general null, it is not surprising that its one-quarter-ahead forecasts of the seven-year bond yield are more accurate than those of the general null. However, forecasts from the standard model are *also* more accurate than those of the general null. At first glance, this is quite surprising, since the general null has three factors devoted to explaining variations in expected excess returns, while the standard model must explain these variations using macro factors.

Related surprising results are seen in Panel A of Fig. 2. The figure displays one-quarter-ahead expectations of the excess return to the seven-year bond. Forecasts from the standard model are in black, forecasts from the general null are in blue, and forecasts from the alternative model are in red. Forecasts from the standard and alternative models are much more variable than those from the general null. Their standard deviations are 1.5 percent, 2.0 percent, and 0.75 percent respectively. Again, given that the general null has so much flexibility in specifying term premia, why are its forecasts of excess returns so stable relative to the other models?

At the core of the answer is a point noted in Section 3.2. A dynamic term structure model requires that the innovation in a bond's return equals the sum of news about future short rates and news about future expected excess returns. Both types of news can be described as "physical measure" news. In particular, news about future short rates is news about the physical-measure expectation of mean short rates over the life of the bond. Because these expectations are determined entirely by macro variables, news about future short rates is driven only by shocks to inflation, output growth, and the short rate.

With both the standard model and the alternative model, shocks to macro variables are allowed to affect physical-measure expectations of future short rates differently than they affect equivalent-martingale expectations of future short rates. With the general null, these macro shocks are required to have identical effects on physical and equivalent-martingale expectations. This additional restriction gives the general null less flexibility in generating news about future short rates. In the data, imposing this restriction makes news about future short rates look a lot like the bond return's entire innovation. Thus there is not much left over to be explained by news about future expected excess returns. I illustrate this argument by taking a close look at the way these models interpret the events of 1980Q4.

4.5.3 What happened at the end of 1980?

Panel A of Table 7 reports that from 1980Q3 to 1980Q4 inflation jumped two percent, short rates jumped three percent, output growth jumped eight percent (recall that all these figures are annualized), and the seven-year yield rose about 50 basis points.

The panel also reports corresponding filtered values, along with the 1980Q3 forecasts of the 1980Q4 values. Recall that the models share the same functional form for macroeconomic dynamics. Nonetheless, the filtered values and forecasts differ because the parameters of these dynamics differ across the models. For example, the standard model predicts, as of 1980Q3, that the short rate will fall 50 basis points over the next quarter. The alternative model predicts a decline of only 20 basis points. These models also disagree about the expected future direction of the seven-year yield. The general null predicts it will fall about 20 basis points from 1980Q3 to 180Q4, the standard model predicts it will rise five basis points, and the alternative model predicts that it will increase about 20 basis points.

Panel B reports the implied realized shock to the seven-year bond yield. (We could also use the realized shock to the bond's return. Both can be decomposed into different kinds of news.) It ranges from 73 basis points, as calculated by the general null, to 33 basis points as calculated by the alternative model. Part of this shock is due to news about the mean short rate over the next seven years. Formally, this news is, for $n = 28$,

$$(E_t - E_{t-1}) \quad \frac{1}{n} \, {}_{i=0}^{n-1} \, r_{t+i} \quad = \frac{1}{n} \begin{pmatrix} 0 & 0 & 1 \end{pmatrix} (I - K_f)^{-1} \left(I - K_f^n \quad {}_{f,t}, \right. \tag{31}$$

where ${}_{f,t}$ is the realized shock to the macroeconomic variables. According to the general null, this news should raise the bond yield by 60 basis points. This leaves only 13 basis points of news of higher future expected excess returns.

By contrast, both the standard model and the alternative model indicate that news about future average short rates raises the bond yield by about 150 basis points. Therefore news about future expected returns *lowers* the yield by about one full percentage point. This is why (as seen in Fig. 2) both the standard model and the alternative model indicate a large drop in expected excess returns, to below -6 percent/quarter as of 1980Q4. According to the general null, the expected excess return to the bond in 1980Q4 is about 40 basis points/quarter. The forecasts from the standard and alternative models turn out to be more accurate than those of the general null: The seven-year bond yield jumps 300 basis points over the next three quarters.

The obvious interpretation of these contrasting results is that the standard and alternative models overfit the data. The parameter estimates of all of the models are blessed with hindsight, but the general null is less able to take advantage of this hindsight because of the restriction linking physical and equivalent-martingale expectations of the short rate.

Forecasts of long-run inflation offer additional evidence that the standard and alternative models are too accurate. Direct survey evidence concerning investors' beliefs is available from Blue Chip Economic Indicators. In October 1980, the ten-year inflation forecast (GDP deflator) from this survey was 8.25 percent/year.[13] The models can be used to calculate the expected mean ten-year inflation rate as of the end of September 1980. The general null produces a forecast of 9.5 percent/year, which is modestly higher than actual investors' forecasts. But the standard and alternative models produce dramatically lower forecasts of 5.4 percent and 5.0 percent respectively. Although inconsistent with what investors believed at the time, these low forecasts are more consistent with subsequent realizations (inflation averaged four percent from 1980Q4 through 1990Q3).

The overfitting apparent in this particular observation is an extreme example of the overfitting that occurs throughout the sample. Panel A in Fig. 2 provides convincing visual evidence that the fitted time series of expected excess returns for the standard and alternative models are implausible. For example, there is only one quarter for which the general null model produces an expected excess quarterly return to the seven-year bond of more than 2.5 percent (10 percent/year). There are 12 such quarters for the standard model and 23 for the alternative model. There are no quarterly observations for which the general null model produces an expected excess quarterly return less than -1.5 percent. There are 10 and 19 quarters for the standard and alternative models, respectively.

Thus, although the alternative model is statistically superior to the general null, its fit of the data is unreasonable. The general null is a more sensible interpretation of the data.

4.6 Results for the later sample

Interest rate behavior in the post-disinflationary period differs substantially from its full-sample behavior. However, from the perspective of model evaluation, there is little difference between the full sample and the shorter, more recent sample. Thus this discussion is relatively brief.

Four main points carry over from the full sample results. First, the standard macro-finance model is markedly inferior to both the general null and the alternative model. Over the period 1985 through 2005, the difference between the log-likelihood of the standard model and either of the other log-likelihoods exceeds 400. Second, the general null is statistically rejected in favor of the alternative model. The LR test in Table 4 rejects the null at all conventional significance levels. Third, the point estimates of the alternative model indicate that the macro factors account for relatively little of the variation in expected excess returns. The evidence in Table 6 indicates a range of 10 to 30 percent of the variance of one-quarter-ahead expected excess returns is attributable to variations in inflation, output growth, and the short rate.

[13] The data are available from the Philadelphia Federal Reserve at www.phil.frb.org/files/spf/cpie10.txt.

Fourth, the alternative model (but not the standard model) appears to overfit expected excess returns. Panel B of Fig. 2 plots one-quarter-ahead expectations of excess quarterly returns to a seven-year bond. It is apparent from the figure that the expected excess returns calculated using the alternative model are much more volatile than those calculated using the other two models. The standard deviation of these expectations is 0.89 percent for the alternative model, compared with 0.51 percent and 0.47 percent for the standard and general null, respectively. There are no quarters for which either the standard model or the general null has an expected excess quarterly return less than -1 percent (-4 percent/year). There are two such quarters for the alternative model. There are no quarters for which the standard model has an expected excess quarterly return greater than 2.5 percent (10 percent/year). The general null has three such quarters, while the alternative model has seven.

The main difference in the estimated models across the two sample periods appears in the cross-sectional fit of inflation. In contrast to the full sample, there were no significant, long-lived swings in inflation during the period 1985 through 2005. As seen in Panel B of Fig. 1, the models all characterize inflation during the later period as the sum of a highly persistent, low volatility latent process and substantial idiosyncratic noise. Table 5 reports that the standard deviation of measurement error in inflation is around 0.9 percent for each model. Correlations between observed inflation and the underlying latent process range from 0.26 for the general null to 0.39 for the alternative model.

One apparently puzzling result in Table 5 is that the RMSE for one-quarter-ahead forecasts of the seven-year bond yield is actually lower for the general null (0.69 percent) than for the alternative model (0.76 percent). How is this possible, given that the alternative model nests the general null? The reason is that these RMSEs are somewhat misleading. Forecast errors are calculated for observations 1985Q1 through 2005Q4, where the first forecast error is the difference between the filtered seven-year bond yield and the unconditional expected seven-year yield. All other forecast errors use conditional expectations. The unconditional yield curve implied by the estimates of the general null is closer to the sample mean yield curve than is the unconditional yield curve for the alternative model. However, the unconditional variances of yields are so large that the forecast error of the first observation has only a trivial effect on the log-likelihood. When the first observation is dropped in computing the RMSEs, the RMSEs for the general null and alternative model are 0.58 and 0.57 percent respectively.

For this shorter sample, as for the full sample, the general null appears to be a reasonable interpretation of the data. Although its conditional forecasts of future bond yields are slightly less accurate than the forecasts of the alternative model, the time series of expected excess returns produced by the alternative model seem to know too much about the future.

4.7 Further discussion

It is worth emphasizing that the results in this section apply only to inflation, output growth, and the short rate. They do not show that term premia are unrelated to other macroeconomic measures. Ludvigson and Ng (2005) argue that by focusing on a couple of variables thought to summarize the state of the economy, researchers can easily fail to uncover true predictability of excess returns. A narrower interpretation of the results here is more appropriate. The evidence shows that when we use a term structure model broader than the standard macro-finance framework, there is little evidence that the macro variables typically viewed as the main determinants of nominal bond yields—inflation, output growth, and the short rate—drive much variation in term premia.

One striking feature of the estimated models that satisfy the general null is extremely high persistence of expected excess returns. Table 6 reports the AR(1) coefficients of one-quarter-ahead expected excess returns are in the neighborhood of 0.9 based on full-sample estimates and in the neighborhood of 0.95 based on later-sample estimates. This high persistence has important implications for the finite-sample properties of forecasting regressions, as examined in the next section.

5 Finite-sample properties of forecasting regressions

This section studies finite-sample properties of forecasting regressions when the true data-generating process satisfies either the restrictive or general null hypothesis. Given a data-generating process, Monte Carlo simulations produce distributions of regression coefficients and associated test statistics. Two issues are addressed. First, for realistic sample sizes, how close are finite-sample distributions of test statistics to standard asymptotic distributions? Second, are finite-sample distributions associated with the restrictive null hypothesis, in which returns are completely unforecastable, similar to those associated with the general null hypothesis?

The main conclusion is that for the regressions estimated in Section 2.2, finite-sample distributions of test statistics under the general null have little in common with either asymptotic distributions of these tests or finite-sample distributions under the restrictive null. After adjusting the estimated test statistics for their finite-sample properties, all regression evidence of return predictability presented in Section 2.2 disappears.

5.1 Overview

Hodrick (1992) and Stambaugh (1999) discuss, in the context of forecasting stock returns, finite-sample properties of forecasting regressions under the restrictive null hypothesis that excess returns are serially uncorrelated. They identify two reasons why finite-sample properties diverge from asymptotic properties. First, typical choices of instruments used to forecast are both persistent and contemporaneously correlated with returns, leading to a predictive regressions bias. Second, estimates of the variability in parameter estimates are poor when overlapping observations are used in estimation. The additional problem introduced with the general null hypothesis is that the residual in a forecasting regression is serially correlated under the null.

The analysis here is restricted to studying the finite-sample properties of the in-sample and out-of-sample regressions estimated in Tables 1 and 2. Excess returns over annual and quarterly horizons to two-year and seven-year bonds are regressed on inflation, output growth, and the short rate. As in these tables, here annual returns are log returns and quarterly returns are simple returns. (Table 2 reports results for returns to maturity-sorted portfolios of bonds. The simulation evidence here uses returns to individual bonds.)

I use two "true" data-generating processes. The first satisfies the restrictive null and the second satisfies the general null. Parameters for the second process are the parameters for the general null estimated over 1985 through 2005. Parameters for the first process are identical, except that the term premia factors are excluded. Thus the two processes have identical finite-sample properties of the regressions' explanatory variables. The processes differ only in the serial correlation properties of the regression residuals.

Using estimates from the later sample instead of the full sample emphasizes the wedge between the restrictive and general nulls. The later-sample estimates imply higher persistence of expected excess returns than do the full-sample estimates. Greater persistence raises the variability of sample covariances between forecasting instruments and true residuals.

Regardless of the data-generating process, a simulation proceeds as follows. An initial draw of the state variables is taken from their unconditional multivariate normal distribution. Subsequent draws use their conditional multivariate normal distribution. The length of each simulation is 179 quarters, which is the length of the full sample studied in this paper. Regressions are estimated with OLS. For the restrictive null, the covariance matrix of the parameter estimates is computed using the robust Hansen-Hodrick approach. For the general null, the Newey-West procedure is used with three lags for quarterly return horizons and seven lags for annual return horizions. These choices mimic those used in the regressions reported in Tables 1 and 2. Out-of-sample regressions are implemented exactly as described in Section 2.2.3.

5.2 Results

The results of these simulations are easy to summarize. Regression tests that have good finite-sample properties under the restrictive null have poor finite-sample properties under the general null. Accordingly, it is important to rely neither on asymptotic properties, nor on finite-sample properties based on the restrictive null.

Detailed evidence, based on 5000 Monte Carlo simulations, is reported in Table 8. The first three columns specify the simulation: the type of regression (in-sample or out-of-sample), the return horizion, and the bond's maturity. The fourth and fifth columns concern the $\chi^2(3)$ tests of the restrictive null. The fourth column reports finite-sample rejection rates of tests of the null hypothesis when using the asymptotic five percent critical value. The fifth reports true finite-sample critical values at a five percent rejection rate. Columns six and seven contain the same information for the general null.

First consider the properties of these regressions under the restrictive null. The results are unsurprising. Differences between finite-sample and asymptotic rejection rates are larger for in-sample regressions than out-of-sample regressions. They are also larger for regressions that use overlapping observations than those that do not. Thus at one extreme are out-of-sample tests using non-overlapping observations. For these regressions (the final two columns in the table), finite-sample rejection rates at the asymptotic five percent critical value are about five percent. At the other extreme are in-sample regressions using quarterly observations of annual returns. For these regressions, finite-sample rejection rates at the asymptotic five percent critical value are around eighteen percent.

Under the general null hypothesis, the finite-sample properties of the regressions are substantially worse. For the in-sample tests, finite-sample rejection rates based on the general null are typically nearly twice the corresponding finite-sample rejection rates based on the restrictive null. Differences are more dramatic for out-of-sample tests. For these tests, finite-sample rejection rates based on the general null are typically between two and three times the corresponding finite-sample rejection rates based on the restrictive null. For example, out-of-sample tests using nonoverlapping observations have finite-sample rejection rates between ten and twenty percent at the asymptotic five percent critical value.

After correcting for the finite-sample properties of the test statistics, all statistical evidence of return forecastability in Tables 1 and 2 disappears. None of the joint tests has a test statistic anywhere near the true finite-sample critical values reported in Table 8. Hence properly interpreted, the regression evidence supports the same qualitative conclusion drawn from estimation of dynamic term structure models: the general null hypothesis is a reasonable description of the data.

6 Conclusion

Imagine that a researcher wishes to determine whether expected excess bond returns are correlated with a particular measure of macroeconomic activity; say, employment growth. One of the main messages of this paper is that it is hard to test this hypothesis accurately. The proper null hypothesis is that expected excess returns vary over time, but are uncorrelated with employment growth. Regression-based tests of this hypothesis, such as regressing excess bond returns on lagged employment growth, have finite-sample properties that are not close to standard asymptotic properties. Moreover, these finite-sample properties cannot be approximated accurately using a model that satisfies the restrictive null hypothesis of serially uncorrelated excess returns.

This paper develops a dynamic term structure framework that can be used to either test directly whether expected excess returns vary with specified measures of macroeconomic activity or to construct finite-sample distributions for regressions. From a modeling perspective, the key component of this framework is a set of latent factors that affect only expected excess returns and are independent of all other factors in the model. When applied to data over 1961 through 2005, the model indicates that expected excess returns are only weakly related to inflation, output growth, and the short rate.

References

Ang, Andrew, and Geert Bekaert, 2005, The term structure of real rates and expected inflation, Working paper, Columbia Business School.

Ang, Andrew, and Geert Bekaert, 2006, Stock return predictability: Is it there?, Working paper, Columbia Business School.

Ang, Andrew, Sen Dong, and Monika Piazzesi, 2005, No-arbitrage Taylor rules, Working paper, University of Chicago GSB.

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, 2005, What does the yield curve tell us about GDP growth?, *Journal of Econometrics*, forthcoming.

Baker, Malcolm, Robin Greenwood, and Jeffrey Wurgler, 2003, The maturity of debt issues and predictable variation in bond returns, *Journal of Financial Economics* 70, 261-291.

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

Bikbov, Ruslan, and Mikhail Chernov, 2006, No-arbitrage macroeconomic determinants of the yield curve, Working paper, Columbia Business School.

Buraschi, Andrea, and Alexei Jiltsov, 2005, Inflation risk premia and the expectations hypothesis, *Journal of Financial Economics* 75, 429-490.

Campbell, John Y., 1987, Stock returns and the term structure, *Journal of Financial Economics* 18, 373-399.

Campbell, John Y., and Robert J. Shiller, 1988, The dividend-price ratio and expectations of future dividends and discount factors, *Review of Financial Studies* 1, 195-228.

Campbell, John Y., and Robert J. Shiller, 1991, Yield spreads and interest rate movements: A bird's eye view, *Review of Economic Studies* 58, 495-514.

Cheridito, Patrick, Damir Filipović, and Robert L. Kimmel, 2005, Market price of risk specifications for affine models: theory and evidence, *Journal of Financial Economics*, forthcoming.

Clark, Todd E., and Michael W. McCracken, 2001, Tests of equal forecast accuracy and encompassing for nested models, *Journal of Econometrics* 105, 85-110.

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

Dai, Qiang, Ahn Le, and Kenneth J. Singleton, 2006, Discrete-time dynamic term structure models with generalized market prices of risk, Working paper, Stanford University GSB.

Dai, Qiang, Kenneth J. Singleton, and Wei Yang, 2005, Regime shifts in a dynamic term structure model of U.S. Treasury bond yields, Working paper, Stern School of Business.

Dewachter, Hans, and Marco Lyrio, 2004a, Macro factors and the term structure of interest rates, *Journal of Money, Credit, and Banking*, forthcoming.

Dewachter, Hans, Marco Lyrio, and Konstantijn Maes, 2004b, A joint model for the term structure of interest rates and the macroeconomy, *Journal of Applied Econometrics*, forthcoming.

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

Duffee, Gregory R., 2006, Term structure estimation without using latent factors, *Journal of Financial Economics* 79, 507-536.

Ericsson, Neil R., 1992, Parameter constancy, mean square forecast errors, and measuring forecast performance: An exposition, extensions, and illustration, *Journal of Policy Modeling* 14, 465-495.

Evans, Charles L., and David A. Marshall, 2002, Economic determinants of the nominal Treasury yield curve, Working paper, Federal Reserve Bank of Chicago.

Fama, Eugene F., 1976, Inflation uncertainty and expected returns on Treasury bills, *Journal of Political Economy* 84, 427-448.

Fama, Eugene F., 1984, The information in the term structure, *Journal of Financial Economics* 13, 509-528.

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

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

Fama, Eugene F., and G. William Schwert, 1977, Asset returns and inflation, *Journal of Financial Economics* 5, 115-146.

Ferson, Wayne E., and Campbell R. Harvey, 1991, The variation of economic risk premiums, *Journal of Political Economy* 99, 385-415.

Ferson, Wayne E., and Campbell R. Harvey, 1993, Explaining the predictability of asset returns, *Research in Finance* 11, 65-106.

Friedman, Benjamin M., 1979, Interest expectations versus forward rates: Evidence from an expectations survey, *Journal of Finance* 34, 965-973.

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

Hansen, Lars Peter, and Robert J. Hodrick, 1980, Forward exchange rates as optimal predictors of future spot rates: An econometric analysis, *Journal of Political Economy* 88, 829-853.

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

Hördahl, Peter, Oreste Tristani, and David Vestin, 2005a, A joint econometric model of macroeconomic and term structure dynamics, *Journal of Econometrics*, forthcoming.

Hördahl, Peter, Oreste Tristani, and David Vestin, 2005b, The term structure of inflation risk premia and macroeconomic dynamics, Working paper, European Central Bank.

Huizinga, John, and Frederic S. Mishkin, 1984, Inflation and real interest rates on assets with different risk characteristics, *Journal of Finance Papers and Proceedings* 39, 699-712.

Ilmanen, Antti, 1995, Time-varying expected returns in international bond markets, *Journal of Finance* 50, 481-506.

Keim, Donald B., and Robert F. Stambaugh, 1986, Predicting returns in the stock and bond markets, *Journal of Financial Economics* 17, 357-390.

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

Kim, Don H., and Athanasios Orphanides, 2005, Term structure estimation with survey data on interest rate forecasts, Working paper, Federal Reserve Board.

Klemkosky, Robert C., and Eugene A. Pilotte, 1992, Time-varying term premia on U.S. Treasury bills and bonds, *Journal of Monetary Economics* 30, 87-106.

Lauterbach, Beni, 1989, Consumption volatility, production volatility, spot-rate volatility, and the returns on Treasury bills and bonds, *Journal of Financial Economics* 24, 155-179.

Law, Peyron, 2004, Macro factors and the yield curve, Working paper, Stanford University.

Ludvigson, Sydney C., and Serene Ng, 2005, Macro factors in bond risk premia, Working paper, New York University.

Newey, Whitney K., and Kenneth D. West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703-708.

Rudebusch, Glenn D., and Tao Wu, 2004, A macro-finance model of the term structure, monetary policy, and the economy, Working paper, Federal Reserve Bank of San Francisco.

Rudebusch, Glenn D., and Tao Wu, 2005, Accounting for a shift in term structure behavior with no-arbitrage and macro-finance moels, *Journal of Money, Credit, and Banking*, forthcoming.

Shiller, Robert J., John Y. Campbell, and Kermit L. Schoenholtz, 1983, Forward rates and future policy: Interpreting the term structure of interest rates, *Brookings Papers on Economic Activity* 1983, 173-223.

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

Stambaugh, Robert F., 1999, Predictive regressions, *Journal of Financial Economics* 54, 375-421.

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

Wachter, Jessica A., 2006, A consumption-baesd model of the term structure of interest rates, *Journal of Financial Economics* 79, 365-399.

Table 1. Predicting annual excess bond returns

The return to a k-year zero-coupon Treasury bond from quarter t to quarter $t+4$ less the quarter-t yield on a one-year Treasury bond is regressed on the three-month T-bill yield as of the end of quarter t, the change in the log GDP price deflator from $t-1$ to t, and the change in log real GDP from $t-1$ to t. All variables are expressed in percent and the predetermined variables are expressed in annual terms. Two sets of standard errors are reported in parentheses. The first are generalized Hansen-Hodrick standard errors. The second are Newey-West standard errors, using seven lags. The column labeled "Joint test" reports Wald tests of the hypothesis that all coefficients equal zero. Asymptotic p-values, based on a $\chi^2(3)$ distribution, are in brackets. The final column reports a t-test, described in Section 2.2.3, that evaluates the contribution of the explanatory variables to out-of-sample forecasts of returns. Under the null, the asymptotic 5% critical value is 1.64. This statistic is constructed only for the full sample period.

Sample (# obs)	Maturity (years)	Inflation	Output growth	short rate	R^2	Joint test	P-val	Out of sample
1961:2–2005:4	2	−0.339	−0.107	0.229	0.14			2.09
(175)		(0.141)	(0.051)	(0.122)		13.09	[0.004]	
		(0.137)	(0.052)	(0.105)		12.63	[0.005]	
	7	−1.449	−0.279	0.641	0.12			1.67
		(0.528)	(0.197)	(0.572)		13.91	[0.003]	
		(0.498)	(0.208)	(0.472)		12.53	[0.006]	
1985:1–2005:4	2	−0.069	−0.213	0.220	0.16			
(80)		(0.202)	(0.088)	(0.125)		13.56	[0.004]	
		(0.196)	(0.075)	(0.106)		23.90	[0.000]	
	7	0.413	−0.071	0.392	0.02			
		(0.858)	(0.432)	(0.630)		0.96	[0.812]	
		(0.863)	(0.396)	(0.527)		1.64	[0.650]	

Table 2. Predicting quarterly excess bond returns

Excess quarterly simple returns to portfolios of Treasury bonds from the end of quarter t to the end of quarter $t+1$ are regressed on the three-month T-bill yield as of the end of quarter t, the change in the log GDP price deflator from $t-1$ to t, and the change in log real GDP from $t-1$ to t. All variables are expressed in percent and the predetermined variables are expressed in annual terms. Two sets of standard errors are reported in parentheses. The first are generalized Hansen-Hodrick standard errors. The second are Newey-West standard errors, using four lags. The column labeled "Joint test" reports Wald tests of the hypothesis that all coefficients equal zero. Asymptotic p-values, based on a $\chi^2(3)$ distribution, are in brackets. The final column reports a t-test, described in Section 2.2.3, that evaluates the contribution of the explanatory variables to out-of-sample forecasts of returns. Under the null, the asymptotic 5% critical value is 1.64. This statistic is constructed only for the full sample period.

Sample (# obs)	Maturities (years)	Inflation	Output growth	short rate	R^2	Joint test	P-val	Out of sample
1961:2–2005:4	$2 < \tau \leq 3$	−0.187	−0.041	0.136	0.04			0.95
(178)		(0.080)	(0.048)	(0.112)		5.92	[0.116]	
		(0.089)	(0.050)	(0.082)		6.17	[0.104]	
	$5 < \tau \leq 10$	−0.356	−0.053	0.213	0.04			0.86
		(0.146)	(0.083)	(0.188)		6.43	[0.093]	
		(0.150)	(0.083)	(0.143)		7.04	[0.071]	
1985:1–2005:4	$2 < \tau \leq 3$	−0.013	−0.049	0.114	0.03			
(83)		(0.192)	(0.066)	(0.084)		2.76	[0.431]	
		(0.155)	(0.059)	(0.082)		3.19	[0.363]	
	$5 < \tau \leq 10$	0.014	0.007	0.230	0.03			
		(0.381)	(0.137)	(0.171)		2.01	[0.570]	
		(0.288)	(0.128)	(0.167)		2.00	[0.573]	

Table 3. Estimates of a dynamic term structure model for 1961 through 2005

The joint dynamics of inflation π_t, output growth Δg_t, and the short rate r_t are described by the model of Section 3.3. The model is estimated with maximum likelihood over the period 1961Q2 through 2005Q4, using these data as well as yields on zero-coupon bonds with maturities of one, two, three, five, and seven years. Standard errors are in parentheses and the p-value of a likelihood ratio test is in brackets.

Panel A. Macro factor dynamics

	Uncon mean		K_f			$\Sigma_f \times 10^3$		Std dev of obs error $\times 10^2$
		π_{t-1}	Δg_{t-1}	r_{t-1}	π_t	Δg_t	r_t	
π_t	0.0375	0.769	−0.471	−0.014	6.457			0.672
	(–)	(0.340)	(0.843)	(0.079)	(0.836)			(0.076)
Δg_t	0.0337	0.032	1.061	0.018	−2.644	1.249		3.239
	(–)	(0.168)	(0.350)	(0.036)	(1.163)	(2.221)		(0.188)
r_t	0.0573	0.330	0.486	0.865	1.520	−8.696	5.434	0
	(–)	(0.414)	(1.038)	(0.103)	(2.086)	(3.759)	(6.011)	(–)

Panel B. Price of macro risk

	λ_{0f}		λ_{1f}		LR test of $\lambda_{1f} = 0$
		π_t	Δg_t	r_t	
π_t	−0.008	1.443	−0.007	0.093	91.20
	(0.248)	(2.032)	(4.935)	(0.477)	[.000]
Δg_t	0.004	−0.578	−0.011	−0.027	
	(0.097)	(0.799)	(1.935)	(0.191)	
r_t	0.018	−0.112	−0.361	−0.118	
	(0.042)	(0.356)	(0.863)	(0.074)	

Panel C. Term premia factor dynamics

		K_ω			$\Sigma_\omega \times 10^3$	
	ω_{1t-1}	ω_{2t-1}	ω_{3t-1}	ω_{1t}	ω_{2t}	ω_{3t}
ω_{1t}	1.480	0.282	0.362	1.502		
	(37.81)	(122.0)	(9.457)	(13.34)		
ω_{2t}	−0.200	0.310	−0.065	−0.222	0.150	
	(5.456)	(19.74)	(1.362)	(5.181)	(3.054)	
ω_{3t}	0.001	6.407	0.091	−3.983	−1.742	0.916
	(208.6)	(599.3)	(47.77)	(17.54)	(40.82)	(4.786)

Table 4. Estimates of a dynamic term structure model for 1985 through 2005

The joint dynamics of inflation π_t, output growth Δg_t, and the short rate r_t are described by the model of Section 3.3. The model is estimated with maximum likelihood over the period 1985Q1 through 2005Q4, using these data as well as yields on zero-coupon bonds with maturities of one, two, three, five, seven, and ten years. Standard errors are in parentheses and the p-value of a likelihood ratio test is in brackets.

Panel A. Macro factor dynamics

	Uncon mean	K_f π_{t-1}	Δg_{t-1}	r_{t-1}	$\Sigma_f \times 10^3$ π_t	Δg_t	r_t	Std dev of obs error $\times 10^2$
π_t	0.0241	0.906	−0.067	0.005	1.471			0.864
	(–)	(0.305)	(0.599)	(0.029)	(1.378)			(0.122)
Δg_t	0.0307	−0.233	0.479	−0.004	0.030	1.277		1.928
	(–)	(0.866)	(0.593)	(0.135)	(3.326)	(4.070)		(0.240)
r_t	0.0475	0.950	0.876	0.862	2.511	0.592	3.855	0
	(–)	(2.625)	(3.354)	(0.372)	(1.790)	(4.493)	(1.332)	(–)

Panel B. Price of macro risk

	λ_{0f}	λ_{1f} π_t	Δg_t	r_t	LR test of $\lambda_{1f} = 0$
π_t	−0.007	0.108	0.160	0.003	44.00
	(0.024)	(0.491)	(0.678)	(0.084)	[.000]
Δg_t	−0.001	0.244	−0.081	−0.025	
	(0.034)	(1.110)	(0.605)	(0.087)	
r_t	0.031	−0.753	−0.546	−0.016	
	(0.087)	(1.566)	(2.204)	(0.232)	

Panel C. Term premia factor dynamics

	K_ω ω_{1t-1}	ω_{2t-1}	ω_{3t-1}	$\Sigma_\omega \times 10^3$ ω_{1t}	ω_{2t}	ω_{3t}
ω_{1t}	0.571	0.111	−0.881	0.378		
	(20.53)	(23.91)	(46.39)	(0.726)		
ω_{2t}	0.102	0.529	−1.544	0.350	0.079	
	(33.88)	(39.50)	(77.02)	(1.652)	(0.234)	
ω_{3t}	−0.234	0.281	1.360	−0.013	−0.042	0.010
	(8.641)	(10.13)	(19.08)	(0.514)	(0.523)	(0.497)

Table 5. The performance of alternative term structure models

Three versions of a dynamic term structure model are estimated with maximum likelihood over two sample periods. Model 1 restricts term premia to be determined only by inflation, output growth, and the short rate. Model 2 restricts term premia to be orthogonal to these variables. Model 3 is unrestricted. Measurement error is defined as the actual value in quarter t less its filtered value. The table reports standard deviations of this measurement error for inflation, output growth, and the seven-year bond yield. The table also reports the root mean squared error of one-quarter-ahead forecasts of these variables and the short rate. The forecasts are produced using the Kalman filter.

	1961–2005 sample			1985–2005 sample		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Number of free parameters	30	36	45	30	36	45
Log likelihood	7318.3	7979.7	8025.3	4249.4	4664.3	4686.3
Standard deviation of measurement error						
Inflation (%)	1.89	0.50	0.43	0.85	0.93	0.87
Output growth (%)	3.29	3.24	3.23	1.80	1.79	1.94
7-year yield (b.p.)	4.83	0.08	0.08	5.31	0.02	0.02
Root mean squared error of one-quarter-ahead forecasts (%)						
Inflation	1.93	1.12	1.06	0.86	0.87	0.87
Output growth	3.25	3.22	3.21	1.91	1.96	1.96
Short rate	1.09	1.14	1.07	0.62	0.62	0.60
7-year yield	0.65	0.68	0.65	0.81	0.69	0.76

Table 6. Model-implied behavior of quarterly excess returns

Parameter estimates of a dynamic term structure model are used to calculate sample properties of quarterly log excess bond returns. The excess return in quarter t is the sum of its expectation as of $t-1$ and a shock. The table reports the fraction of the variance of excess returns that is attributable to the former channel. The variance of the conditional expectation is further decomposed into a component associated with variations in macro variables (inflation, output growth, and the short rate) and an orthogonal, non-macro component. The table reports the fraction attributable to the macro variables. It also reports the first-order serial correlation of the conditional expectation of log excess returns.

The model is estimated over two samples and in three versions. Model 1 restricts term premia to be determined only by inflation, output growth, and the short rate. Model 2 restricts term premia to be orthogonal to these variables. Model 3 is unrestricted.

	1961–2005 sample			1985–2005 sample		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Mean (%/q)						
2-year bond	0.24	0.24	0.25	0.45	0.45	0.45
7-year bond	0.41	0.40	0.39	1.21	1.20	1.20
Standard deviation (%/q)						
2-year bond	1.63	1.64	1.64	1.15	1.15	1.15
7-year bond	4.43	4.40	4.40	3.98	4.00	4.00
Fraction of return variance due to variation in conditional expectation						
2-year bond	0.09	0.03	0.23	0.02	0.03	0.05
7-year bond	0.11	0.03	0.20	0.02	0.01	0.05
Fraction of variance of conditional expectation due to variation in macro variables						
2-year bond	1	0	0.22	1	0	0.27
7-year bond	1	0	0.30	1	0	0.11
AR(1) of conditional expectation						
2-year bond	0.55	0.92	0.48	0.97	0.98	0.78
7-year bond	0.54	0.86	0.53	0.73	0.95	0.76

Table 7. A close look at 1980Q3 and 1980Q4

Parameter estimates of a dynamic term structure model are used to interpret the change in inflation, output growth, and interest rates from 1980Q3 to 1980Q4. Three versions of the model are estimated. Model 1 restricts term premia to be determined only by inflation, output growth, and the short rate. Model 2 restricts term premia to be orthogonal to these variables. Model 3 is unrestricted. Panel A reports observed values of inflation, output growth, the short rate (three-month T-bill), and the yield on a seven-year Treasury bond. The model assumes that all but the short rate are observed with noise, thus the true values are unobserved and must be filtered from the data. The panel reports Kalman filtered values for both 1980Q3 and 1980Q4, as well as Kalman filter forecasts, based on data through 1980Q3, of the 1980Q4 values. Panel B decomposes the associated shock to the seven-year bond yield into news about average short rates over the next seven years and news about risk premia.

Panel A. Applying the Kalman filter

	Inflation	Output growth	Short rate	7-year bond yield
Raw data				
1980Q3	8.97	−0.67	11.74	11.37
1980Q4	11.08	7.35	14.73	11.92
1980Q3 contemporaneous filtered values				
Model 1	6.92	1.90	11.74	11.35
Model 2	9.15	1.49	11.74	11.37
Model 3	8.83	1.15	11.74	11.37
1980Q3 forecast of 1980Q4 values				
Model 1	6.60	2.17	11.24	11.40
Model 2	9.19	1.38	11.03	11.19
Model 3	8.62	1.29	11.54	11.59
1980Q4 contemporaneous filtered values				
Model 1	8.43	−0.13	14.73	11.89
Model 2	10.54	1.52	14.73	11.92
Model 3	10.19	0.49	14.73	11.92

Panel B. Decomposing the 1980Q4 filtered shock to the seven-year bond yield

	Total shock	News about future short rates	Risk premia news related to macro factors	Risk premia news unrelated to macro factors
Model 1	0.49	1.39	−0.90	–
Model 2	0.73	0.60	–	0.13
Model 3	0.33	1.50	0.50	−1.68

Table 8. Finite-sample properties of regressions that forecast excess bond returns

This table summarizes results from 5000 Monte Carlo simulations. The "true" data-generating processes of yields, inflation, and output growth are based on an estimated dynamic term structure model. Under the restrictive null, expected excess returns are constant. Under the general null, they vary over time, but are independent of inflation, output growth, and the short rate.

For each simulation, 179 quarters of data are generated. Excess bond returns are calculated in two ways. The quarter $t+1$ simple excess return is the simple return to the n-year bond from t to $t+1$ less the contemporaneous simple return to the three-month bond. The annual log excess return from t to $t+4$ is the log return to the n-year bond less the quarter-t yield on a one-year bond.

Excess returns are regressed on the quarter-t values of inflation, output growth, and the short rate. The in-sample test statistic is a Wald test of the hypothesis that the coefficients are jointly zero. The table reports the empirical rejection rate using the five percent critical value for a $\chi^2(3)$ distribution, as well as the finite sample five percent critical value. Similar statistics are reported for the out-of-sample t-test of Ericsson (1992). This t-test has an asymptotic N(0,1) distribution and is discussed in more detail in the text.

			Restrictive null		General null	
Type of regression	Type of return	Maturity (years)	Rejection rate at 5% asy crit val	True 5% critical value	Rejection rate at 5% asy crit val	True 5% critical value
In-sample	Annual	2	0.180	14.19	0.312	23.79
In-sample	Annual	7	0.178	14.03	0.234	17.05
In-sample	Quarterly	2	0.103	9.84	0.265	19.31
In-sample	Quarterly	7	0.098	9.85	0.167	13.23
Out-of-sample	Annual	2	0.087	2.11	0.207	3.66
Out-of-sample	Annual	7	0.087	2.09	0.128	2.54
Out-of-sample	Quarterly	2	0.054	1.68	0.205	3.38
Out-of-sample	Quarterly	7	0.050	1.64	0.097	2.13

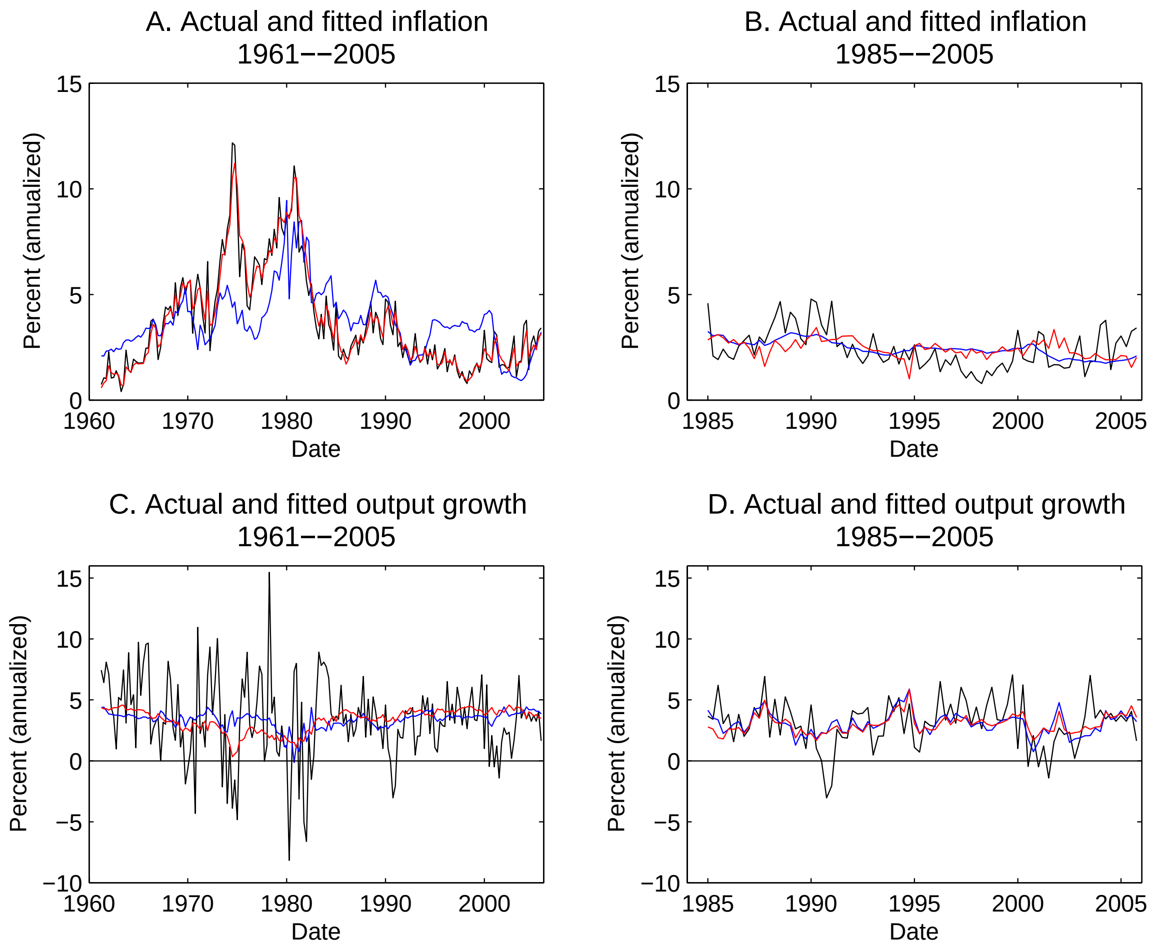

Fig. 1. A comparison of observed inflation and output growth with model-implied estimates. The black lines in each panel are actual inflation (Panels A and B) and output growth (Panels C and D) over the specified dates. The blue and red lines represent filtered values of latent, "core" inflation and output growth based on estimates of two dynamic term structure models of inflation, output growth, and the term structure. The model underlying the blue line is a standard macro-finance model, where term premia are entirely determined by inflation, output growth, and the short rate. The model underlying the red line is the "general null" model, where term premia are orthogonal to these variables. The models are estimated separately over the 1961–2005 and 1985–2005 samples.

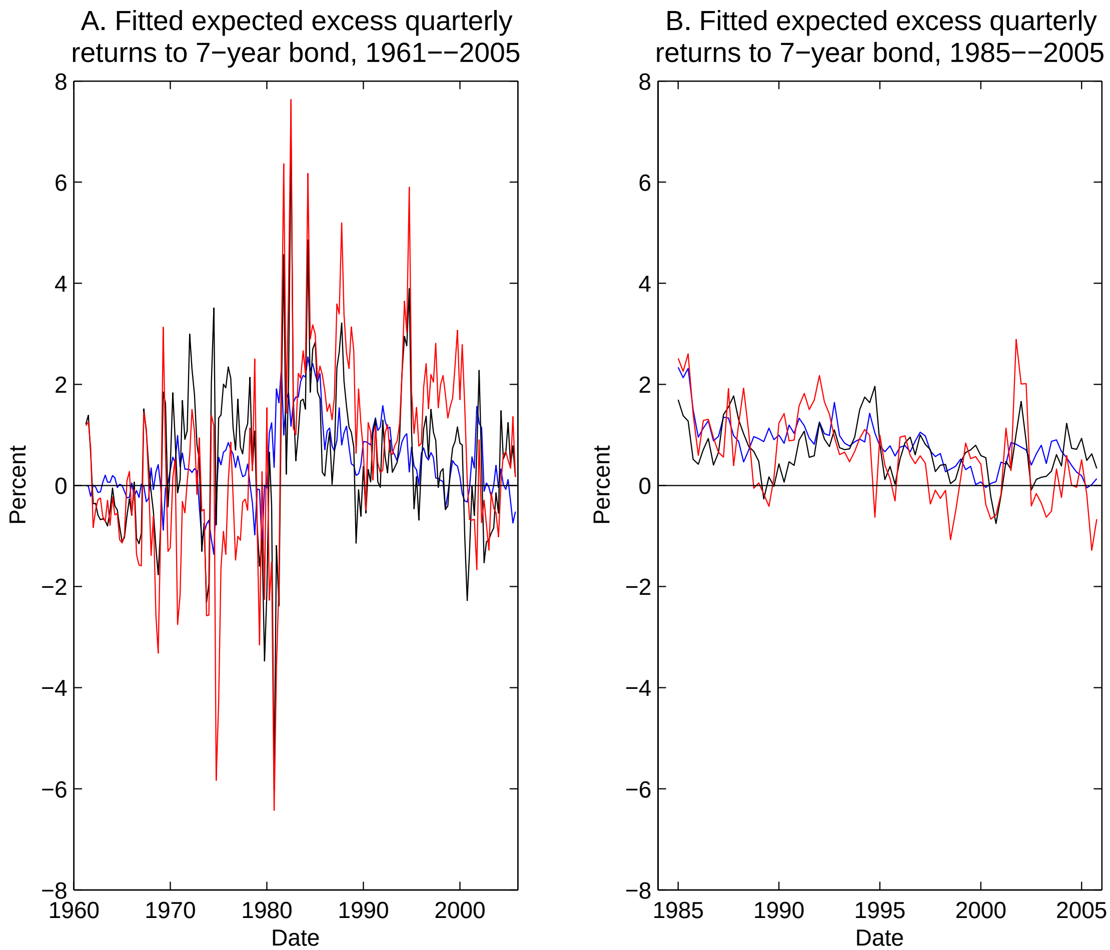

Fig. 2. Model-implied estimates of expected excess quarterly returns to a seven-year Treasury bond. The lines represent one-quarter-ahead expected excess returns (nominal return less return to a 90-day T-bill), where the expectations are based on estimates of three dynamic term structure models of inflation, output growth, and the term structure. One model (black line) is a standard macro-finance model, where term premia are entirely determined by the inflation, output growth, and the short rate. Another model (blue line) is the "general null" model, where term premia are orthogonal to these variables. The final model (red line) is unrestricted. The models are estimated separately over the 1961–2005 and 1985–2005 samples.