Are variations in term premia related to the macroeconomy?

Gregory R. Duffee*

Haas School of Business
University of California – Berkeley

This Draft: March 27, 2006

ABSTRACT

To test the hypothesis that expected excess bond returns are correlated with macroeconomic variables such as inflation, the relevant null hypothesis is that expected excess returns are stochastic, persistent, and independent of such 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. Instead, they use the null hypothesis that excess returns are serially uncorrelated. This paper presents a dynamic model that satisfies the appropriate null hypothesis. Simulations lead to the conclusion that finite-sample distributions of forecasting regressions under the appropriate null differ substantially from finite-sample distributions under the commonly-used null. Direct estimation of the model reveals that only a small fraction of the variation in expected excess returns is associated with inflation, output growth, or 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, Orestes Tristani, and seminar participants at the Federal Reserve Board and the ECB-BIS Workshop on Macrofinance Modeling of the Term Structure of Interest Rates for helpful comments. This paper will eventually appear at http://faculty.haas.berkeley.edu/duffee/.
1 Introduction

Tests of the expectations hypothesis document conclusively that term 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 of them 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 variables that proxy for the economic fundamentals of choice. 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 long-horizon return regressions and 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

¹Term 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; it should be about the source of that 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 construct 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 vary with macro variables. I apply the model in two ways. First, I estimate the broader version using data from 1955 through 2004. The results show that expected excess returns vary substantially through time, but little of this variation is associated with inflation, output growth, or the short rate. Moreover, the relation between expected excess returns and these macro variables disappears after 1984. Qualitatively, this conclusion is similar to one drawn in Cochrane and Piazzesi (2005). They note that the factors that explain the vast majority of time-variation in yields are not important for explaining variations in expected excess returns.

Second, I estimate the version of the model that satisfies the general null hypothesis. Monte Carlo simulation of the estimated model generates finite-sample distributions of test statistics associated with excess return forecasting regressions. These distributions are calculated both under the restrictive null and the general null.

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. For univariate forecasting regressions, empirical rejection rates at 5 percent asymptotic critical values are about twice as large with the general null as they are with the restrictive null. 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.

In practice, correcting test statistics for the properties of finite samples has important effects on inference. When annual excess returns to long-term bonds are regressed on inflation, output growth, and the short rate, asymptotic critical values under the restrictive null strongly reject the null that the macro variables have no forecast power. But using finite-sample critical values overturns this conclusion.

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 for a relation between expected excess bond returns and macroeconomic conditions. The first subsection describes the standard methodological approach and reviews earlier evidence. It concludes that nothing in the existing literature tests 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 suggest that excess returns are predictable with a combination of inflation, output growth, and the short rate.

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

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

(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 $t$. 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 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) in using measures of volatility to predict excess returns. 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 based on stock prices predict both excess stock returns and excess bond returns. 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 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 a combination of variables based on prices of risky securities.

---

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.
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 ex-
cess 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. The reason is that the 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
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 con-
ducted 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.

5The 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.
2.2.1 Data

The data are quarterly from 1955Q1 through 2004Q4. Inflation, denoted $\pi_t$, is the annualized log change in the GDP price deflator from quarter $t - 1$ to quarter $t$. Output growth, denoted $\Delta g_t$, 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

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

The lower case $rx$ denotes a log return. These data are from the Center for Research in Security Prices (CRSP).

I also use monthly returns to maturity-sorted Treasury portfolios (also from CRSP) 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

$$R_{t,t+1}^p = \left(1 + R_{t+1(1)}^p\right)\left(1 + R_{t+1(2)}^p\right)\left(1 + R_{t+1(3)}^p\right).$$ (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

$$RX_{t,t+1}^p = R_{t,t+1}^p - \exp(r_t/4).$$ (4)

The upper case $RX$ denotes a simple return.
2.2.2 Regressions

Excess bond returns $r_{x_{t,t+4}}^{(n)}$ and $RX_{t,t+1}^{P}$ are regressed on quarter-$t$ values of inflation, output growth, and the short rate. The regressions are estimated using the entire sample of 1955 through 2005 as well as the more recent sample of 1985 through 2004. The shorter sample is singled out because of the strong 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.

The regressions for annual excess returns use overlapping observations. 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 five 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 $r_{x_{1,5}}^{(n)}$ through $r_{x_{R,R+4}}^{(n)}$ of annual excess returns. Given the estimated parameters, forecast $r_{x_{R+4,R+8}}^{(n)}$ 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 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 1955 through 2004, I use \( R = 120 \), thus \( P = 73 \) for annual excess returns and \( P = 79 \) for quarterly excess returns. I do not apply this procedure to the shorter sample 1985 through 2004 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 process implies that expected excess returns are determined by some independent, persistent variable \( \omega_t \). If \( \omega_t \) 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 1955 through 2004, inflation, output growth, and the short rate collectively have substantial information about future excess returns. For both two-year and five-year bonds, the \( R^2 \) in the full sample is 14 percent, and \( \chi^2(3) \) tests that the coefficients are all zero are rejected at the two percent level. The choice of restrictive versus general null makes little difference to the statistical strength of this

---

\(^6\)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.
rejection. If anything, the rejection is slightly stronger under the general null. In addition, the test for out-of-sample forecastability strongly rejects the restrictive null. (The 5 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 2004, only excess returns to the two-year bond appear forecastable. Again, the statistical strength of the forecastability appears stronger under the general null than under the restrictive null.

The results for quarterly returns in Table 2 are less clearcut. For the in-sample regressions over 1955 through 2004, the joint statistical significance of the three forecasting variables is marginal. Under the restrictive null, we cannot reject at the 5 percent level the hypothesis of no forecastability. Under the general null, the hypothesis is rejected for the longer-maturity portfolio but not for the shorter-maturity portfolio. Over the more recent sample, there is no evidence of in-sample forecastability using either null. By contrast, the test of out-of-sample forecastability strongly rejects the restrictive null.

On balance, these results are a bit of a muddle. 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 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

Dynamic term structure models exploit the special features of bonds: fixed cash flows and a term structure. This section describes how dynamic term structure models are used to draw inferences about the determinants of term premia. The first subsection describes the standard methodological approach, and reviews earlier evidence. The second 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. 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 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 answer questions such as the fraction of variation in expected excess returns on a $n$-period bond that is attributable to shocks to each element of the state vector.

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).

The fact that the literature contains such apparently strong, yet conflicting results is sufficient to create concern about the robustness of the results. A more concrete concern is the model-imposed requirement that the only possible determinants of term premia are the elements of the state vector. Parsimonious specifications of the state vector make it almost

---

7Recent 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)
8The evidence is in the estimates of $\lambda_1$ reported in their Table 4.
impossible for term premia to vary independently of macroeconomic variables. 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. Thus term premia must either be correlated with macroeconomic conditions or constant over time. 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.

More broadly, drawing conclusions from these papers about term premia dynamics is difficult because specific null hypotheses involving term premia are seldom stated or tested. The typical paper constructs a model, specifies some parameter restrictions to allow for tractable estimation, then estimates the model. The implications of the resulting model are then summarized and interpreted. There is no discussion of what parameter restrictions are required for expected excess returns to be unforecastable with macro variables, let alone tests of such restrictions.

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.

### 3.2 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 capture variations in expected excess returns that are unrelated to the macro factors.

#### 3.2.1 Factors and factor dynamics

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

\[
\mathbf{f}_t = \begin{pmatrix}
\tilde{\pi}_t \\
\tilde{\Delta} g_t \\
\tilde{r}_t
\end{pmatrix}'.
\]
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 \( \omega_t \). (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

\[
x_t = \left( \begin{array}{c} f'_t \\ \omega'_t \end{array} \right)'.
\]  

(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

\[
\check{r}_t = \delta'_x x_t, \quad \delta_x = \left( \begin{array}{cccc} 0 & 0 & 1 & 0 & 0 \end{array} \right)'.
\]

(7)

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

\[
x_t = \mu_x + K_x x_{t-1} + \Sigma_x \epsilon_{x,t}, \quad \epsilon_{x,t} \sim N(0, I).
\]

(8)

The specific matrices are given by

\[
\mu_x = \left( \begin{array}{c} \mu_f \\ 0_{3\times 1} \end{array} \right), \quad K_x = \left( \begin{array}{cc} K_f & 0_{3\times 3} \\ 0_{3\times 3} & K_\omega \end{array} \right), \quad \Sigma_x = \left( \begin{array}{cc} \Sigma_f & 0_{3\times 3} \\ 0_{3\times 3} & \Sigma_\omega \end{array} \right).
\]

(9)

The matrices \( K_f, K_\omega, \Sigma_f, \text{ and } \Sigma_\omega \) are 3 \times 3. Both \( \Sigma_f \) and \( \Sigma_\omega \) are lower triangular. The process is assumed to generate stationary dynamics, so that the unconditional expectation of \( x_t \) is

\[
E(x) = (I - K_x)^{-1} \mu_x.
\]

(10)

With this specification, \( f_t \) and \( \omega_t \) 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 only by \( f_t \). Thus the only role played by the term premia factors is to drive bond risk premia.
3.2.2 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,

\[
\tilde{P}_t^{(n)} = E_t \left( \tilde{P}_{t+1}^{(n-1)} M_{t+1} \right)
\]  \hspace{1cm} (11)

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

\[
M_{t+1} = \exp \left( - h \tilde{r}_t - \lambda_t^t \epsilon_{x,t+1} - \frac{1}{2} \lambda^t_t \lambda_t^t \right)
\]  \hspace{1cm} (12)

where \(h\) is the length of a period (in years), and \(\lambda_t\) is the period-\(t\) compensation investors require to face \(\epsilon_{x,t+1}\) risk. The functional form for \(\lambda_t\) is in the essentially affine class of Duffee (2002),

\[
\Sigma_x \lambda_t = \lambda_0 + \lambda_1 x_t.
\]  \hspace{1cm} (13)

In (13), \(\lambda_0\) is a vector and \(\lambda_1\) is a matrix. The parameterizations of \(\lambda_0\) and \(\lambda_1\) are

\[
\lambda_0 = \begin{pmatrix} \lambda_{0f} \\ 0_{3 \times 1} \end{pmatrix}
\]  \hspace{1cm} (14)

and

\[
\lambda_1 = \begin{pmatrix} \lambda_{1f} & I_{3 \times 3} \\ 0_{3 \times 3} & 0_{3 \times 3} \end{pmatrix}
\]  \hspace{1cm} (15)

where \(\lambda_{0f}\) is a vector of length three and \(\lambda_{1f}\) is a \(3 \times 3\) matrix.

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

\[
\Sigma_x \lambda_t = \begin{pmatrix} \lambda_{0f} + \lambda_{1f} \tilde{f}_t + \omega_t \\ 0_{3 \times 1} \end{pmatrix}
\]  \hspace{1cm} (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 \(\lambda_{1f}\) 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

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

where

$$\mu_x^q = \mu_x - \lambda_0, \quad K_x^q = K_x - \lambda_1.$$  \hspace{1cm} (18)

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

$$\tilde{p}_t^{(n)} = A_n + B'_n x_t.$$  \hspace{1cm} (19)

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

$$B'_n = -h \delta'_x (I - K_x^q)^{-1} (I - (K_x^q)^n).$$  \hspace{1cm} (20)

The constant term is

$$A_n = -h \delta'_x \left[ nI - (I - K_x^q)^{-1} (I - (K_x^q)^n) \right] E^q(x) + \frac{1}{2} \sum_{i=1}^{n-1} B'_i \Sigma_x \Sigma'_x B_i, \quad n = 2, \ldots$$  \hspace{1cm} (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

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

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

$$\log E_t \left( \tilde{p}^{(n-1)}_{t+1} / \tilde{p}^{(n)}_t \right) = h \tilde{r}_t + B'_{n-1} (\lambda_0 + \lambda_1 x_t).$$  \hspace{1cm} (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 $B'_{\tau-1}$ and the compensation per unit of $x_t$ risk is, from (13), $\Sigma_x \lambda_t$.

### 3.2.3 From factors to observables

I use a state-space setting to relate model’s factors to observable macro factors and bond yields. At the end of each quarter, an econometrician observes inflation $\pi_t$, output growth $\Delta g_t$, the short rate $r_t$, and $d$ yields on multiperiod bonds with maturities $n_1, \ldots, n_d$. Stack
these observables in a vector
\[ z_t = \left( \pi_t \ \Delta g_t \ r_t \ y_t^{(n_1)} \ \ldots \ y_t^{(n_d)} \right)' \tag{24} \]

The relation between factors and observables is
\[ z_t = \begin{pmatrix} 0_{3 \times 1} \\ A_y \end{pmatrix} + \begin{pmatrix} I_{3 \times 3} & 0_{3 \times 3} \\ B_{fy} & B_{\omega y} \end{pmatrix} \begin{pmatrix} f_t \\ \omega_t \end{pmatrix} + \eta_t, \ \eta_t \sim N(0, H). \tag{25} \]

The vector \( A_y \) and matrices \( B_{fy} \) and \( B_{\omega y} \) are
\[ A_y = \begin{pmatrix} -\frac{1}{h_{n_1}} A_{n_1} \\ \ldots \\ -\frac{1}{h_{n_d}} A_{n_d} \end{pmatrix}, \quad \begin{pmatrix} B_{fy} & B_{\omega y} \end{pmatrix} = \begin{pmatrix} -\frac{1}{h_{n_1}} B'_{n_1} \\ \ldots \\ -\frac{1}{h_{n_d}} B'_{n_d} \end{pmatrix}. \tag{26} \]

In the state-space setting the usual interpretation of \( \eta_t \) 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.

### 3.2.4 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).\(^9\) 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 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 \( \Sigma_\omega \) and their persistence is determined by the feedback matrix \( K_\omega \). This restriction can be tested against the alternative hypothesis

---
\(^9\)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.
\( \lambda_{1f} \neq 0. \)

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 themselves, factors, but the evolution of the macro factors depends on both sets of factors. With this alternative model, 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 \( \lambda_{1f} = 0. \)

The model 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. But the model does not say that variations in term premia have no economic foundation. Instead, the model is a diagnostic 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 discusses the results of estimating the dynamic term structure model. The emphasis is on the estimate-implied behavior of expected excess returns. In particular, is there statistically reliable evidence that the macro variables inflation, output growth, and the short rate are related to expected excess returns? What is the economic significance of this predictability? Do the answers to these questions depend on the sample period? A brief summary of the results is that over the 1955 through 2004 period, there is statistically strong evidence that these variables (particularly inflation) predict excess bond returns. However, the economic significance of the predictability is quite weak. Moreover, all statistical and economic significance is absent in the 1985 through 2004 period.
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 bond yields with maturities of one through five years are from CRSP.

Recall that the forecasting regressions are estimated over the full period 1955 through 2004 and the more recent period 1985 through 2004. The dynamic term structure model is estimated over the same two periods, but the results for the full sample should not be given great weight. The implications of the model for term premia behavior critically rely on the assumption of parameter stability. The model requires that investor expectations of future short-term interest rates are given by forecasts from a constant-parameter vector autoregression. 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.\textsuperscript{10} The accompanying disinflation was largely completed by the end of 1984. The assumption here is that 1985 through 2004 constitutes a single regime, although the dynamic regime-switching term structure models of Ang and Bekaert (2005) and Dai et al. (2005) find some evidence of instability over this period.

4.2 Summary of free parameters

The dynamics of the macro factors are determined by the parameters of a vector autoregression. There are a total of 18 parameters in the vector $\mu_f$ and the matrices $K_f$ and $\Sigma_f$. Mean expected excess returns are determined by the three elements of the vector $\lambda_{0f}$ in (14). The part of expected excess returns that are independent of the macro factors are determined by the 15 parameters in $\Sigma_{\omega}$. If expected excess returns are also allowed to vary with macro factors, then $\lambda_{1f}$ is a free matrix with nine parameters.

The covariance matrix $H$ of measurement error in (25) must also be parameterized. I choose a simple diagonal specification for $H$. This adds up to eight parameters (standard deviations of measurement error); one for each of the macro factors and the five zero-coupon bond yields. However, I fix the standard deviation of measurement error for the short-term interest rate to zero, leaving only seven free parameters.

This choice reflects two practical objectives. First, I want to push the results in the direction of explaining the term structure with inflation and output growth. As noted by Ang, Dong, and Piazzesi (2005), estimates of dynamic term structure models often imply that the term structure is almost entirely driven by the short rate (or some latent factor that is determined from bond yields), giving non-yield factors almost no role. By assuming

\textsuperscript{10}See Gray (1996).
that core inflation and output growth are measured with error, the model is not forced to fit the term structure to every blip in these variables. In contrast, any idiosyncratic shocks to the short rate are forced to affect longer-term yields. Second, the Monte Carlo simulations that use the estimated parameters are easier to execute and interpret if the short rate is not measured with error.

4.3 Estimation technique

I estimate both restricted and unrestricted versions of the model with maximum likelihood (ML). The state-space representation is implemented easily with the Kalman filter. The restricted version imposes $\lambda_{1f} = 0$, so that macro variables are assumed independent of expected excess returns. A likelihood ratio (LR) test of the hypothesis $\lambda_{1f}$ is used to evaluate the statistical significance of the restriction. Nonlinear optimization is required to find the ML solution. To find the global maximum, 20 starting values are randomly generated. For each starting value, simplex is used to determine a well-behaved neighborhood of the parameter estimates. A derivative-based algorithm is used to improve the accuracy of the estimates.

Both the $p$-value of the LR test and standard errors of the parameter estimates are calculated with 500 Monte Carlo simulations. More precisely, the data generating process of an assumed true model is used to construct a random sample of macro variables and bond yields. (The length of the sample is 200 quarters for full-sample simulations and 80 quarters for later-sample simulations.) For all Monte Carlo simulations, the ML estimates of the restricted model are used as the assumed true model. This is an important point that deserves emphasis. Even when constructing standard errors for the unrestricted parameter estimates, I assume that truth satisfies the null hypothesis $\lambda_{1f} = 0$. For example, when constructing standard errors and $p$-values of the LR test for full-sample estimates, the ML estimates of the restricted full-sample model is used generate a random samples of 200 quarters of data. Using these data, a draw from the distribution of restricted (unrestricted) parameter estimates is determined by ML estimation of the restricted (unrestricted) model. A draw from the distribution of likelihood ratio tests is twice the difference in the log-likelihoods of these two ML estimates.
4.4 An overview of the results

Tables 3 and 4 report unrestricted parameter estimates for the 1955–2004 and 1985–2004 periods, respectively. The tables also report the LR test of the null hypothesis. To conserve space, parameter estimates of the restricted models are not reported. Rather than discuss the individual parameter estimates (which are not particularly intuitive), I summarize the implied joint behavior of macro variables and bond yields implied by the estimates.

Fig. 1 displays the time series of observed inflation, output growth, and the five-year bond yield. Inflation and output growth are plotted along with filtered estimates of latent, “core” inflation and output growth. The filtered values are calculated using unrestricted parameter estimates. For both the full sample (panel A) and the more recent sample (panel B), observed and latent inflation closely track each other. By contrast, latent output growth is smoother than observed output. In the full sample, the standard deviations of observed and latent output growth are 3.7 percent and 2.0 percent, respectively. In the later sample, the corresponding standard deviations are 2.0 percent and 1.0 percent.

Panels E and F compare the observed five-year bond yield with the model-implied “risk-neutral” yield. The risk-neutral yield is calculated using the filtered estimates of the factors and the parameters of the unrestricted model, but setting $\lambda_0$ in (14) and $\lambda_1$ in (15) to zero. Thus this risk-neutral yield is the yield on the bond given by combining the joint physical dynamics of the macro variables with valuation by risk-neutral investors.

The panels show the spread between observed and risk-neutral yields varies substantially over time. In Panel E, the spread is close to zero from 1955 through 1979. The spread then jumps to close to 450 basis points during 1980 through 1984. Both panels show that the spread falls to zero during 1985 through 2004. The average spread based on the full-sample results is 197 basis points over this period; it is 139 basis points based on the later-sample results.

This evidence conveys a visual sense of a point made more formally in the next subsection: variations in risk premia are not closely tied to inflation, output growth, or the short rate. Inflation and the short rate generally increase during 1955 through 1979, but the spread between observed and risk-neutral yields varies little over this period. The spread is very high when inflation and the short rate are high during 1980 and 1981, but the spread falls much more slowly than these variables during the disinflationary period.

---

11 Standard deviations of measurement error for the bond yields are not reported. The largest is only 12 basis points. All others are less than 10 basis points.
4.5 The predictability of expected excess returns

Table 5 summarizes the behavior of excess bond returns implied by model estimates. To focus the discussion, I restrict my attention to quarterly excess returns to two-year and five-year bonds. Recall that four sets of parameters are estimated. In the table, the results are broken down by sample period in the first column and by type of estimated model in Panels A and B.

To illustrate the form of the table, consider the first row in Panel A, which is based on the parameter estimates of the restricted model estimated over the full sample. The “Mean” column reports that the point estimate of the unconditional mean of log excess quarterly returns to a two-year bond is 0.03 percent, or three basis points. This is not a sample mean. Instead, it is the population mean derived from (23) and (10). Formally, the mean excess return to holding an \( n \)-quarter bond for one quarter is

\[
\text{mean excess quarterly return} = B' n^{-1} (\lambda_0 + \lambda_1 (I - K_x)^{-1} \mu_x).
\]  

(27)

The number in parentheses below 0.03 is a simulation-based estimate of the bias in this point estimate. It is the mean, across the Monte Carlo simulations described in Section 4.3, of corresponding point estimates of unconditional mean excess returns. The value 0.06 means that when the true model has an unconditional mean excess quarterly return to the two-year bond of three basis points, the expected model-based estimate of the unconditional mean excess return is six basis points. The same simulations are used to construct a [5%, 95%] confidence range for the point estimate. By this metric, the uncertainty in the point estimate is high, ranging from −41 basis points to 43 basis points.

The “Std dev” column reports the model-implied unconditional standard deviation of the quarterly log excess return. This is the unconditional standard deviation of (22) excluding the risk-free component, or the square root of

\[
\text{variance of excess quarterly return} = B' n^{-1} \lambda_1 \text{Var}(x_t) \lambda_1' B n^{-1} + B' n^{-1} \Sigma_x \Sigma_x' B n^{-1}.
\]  

(28)

In (28), the unconditional variance of \( x_t \) is determined by the VAR dynamics of \( x_t \). For the full-sample restricted model, the unconditional standard deviation of the excess quarterly return to the two-year bond is 1.77 percent.

The first term on the right of (28) is the unconditional variance of one-quarter-ahead expectations of quarterly excess returns. The second term is the unconditional variance of return shocks. The next column in the table reports that 16 percent of the variance of quarterly excess returns to the two-year bond is attributable to variations in the conditional
mean. Put differently, the square root of the first term on the right of (28) is 71 basis points.

The unconditional variance of the conditional mean can be decomposed into a component associated with the macro variables and a component associated with the latent term premia variables.\textsuperscript{12} The penultimate column reports the fraction of this variance associated with the macro variables. The estimated models in Panel A impose the restriction $\lambda_{1f} = 0$, so this column is identically zero.

The final column reports the population first-order autocorrelation coefficient of one-quarter-ahead expectations of quarterly excess returns. This coefficient is

$$\text{AR}(1) \text{ coefficient} = \frac{B'_{n-1}\lambda_1 K_x \text{Var}(x_t)\lambda'_1 B_{n-1}}{B'_{n-1}\lambda_1 \text{Var}(x_t)\lambda'_1 B_{n-1}}. \quad (29)$$

There are two broad points to take from Panel A of Table 5. First, under the null hypothesis that macro variables cannot forecast excess bond returns, a substantial component of excess bond returns is nonetheless predictable. The theoretical $R^2$ of a regression forecasting quarterly excess returns is around 16 percent over the full sample and around 12 percent in the later sample. (The explanatory variables in this regression are the latent term premia factors.) Second, the persistence of these predictable components is very high in the later sample. The first-order autocorrelation for predictable quarterly returns is around 0.9 over this period. By contrast, the full-sample results imply that the predictable components are much less persistent; autocorrelation coefficients are around 0.5.

Panel B uses estimates of both the unrestricted and restricted models to address the relation between expected excess returns and the macro variables. The first row in this panel indicates that according to the unrestricted, full-sample parameter estimates, 21 percent of the variance of conditional mean excess two-year bond return is attributable to variations in the macro variables. An estimate of the unrestricted model will attribute some of the variation in this conditional mean to macro variables even if the true model exhibits no such relation. Using Monte Carlo simulations of the restrictive model, the distribution of this estimated fraction has a mean of 7 percent and a confidence range of 0 through 18 percent. Based on this distribution, we can reject the null hypothesis. Of course, this rejection is based on only one aspect of the model’s behavior. A more appropriate test is the LR test in Table 3, which overwhelmingly rejects the null.

Although the statistical rejection is strong, the economic significance of the relation between macro variables and expected excess returns is weak. Even if the full-sample, un-

\textsuperscript{12}The decomposition is constructed by calculating the decomposition of $\text{Var}(x_t)$, then pre- and post-multiplying the result by $B'_{n-1}\lambda_1$. Because the macro variables are independent of the term premia variables, the ordering of the variables in this decomposition is irrelevant.
restricted model point estimates are taken at face value, the theoretical $R^2$ of a regression in which macro variables forecast quarterly excess returns to the two-year bond is only four percent. The corresponding $R^2$ for the five year bond is only two percent. Because of the upward bias in these point estimates, the true $R^2$s are likely to be substantially smaller.

Fig. 2 graphically displays the relation between the macro variables and expected excess returns. The figure shows the effect on expected quarterly excess returns if the specified macro variable is one percentage point higher, holding the other macro variables constant. The solid lines are the point estimates, calculated from estimated parameters of the unrestricted model. The outer dashed lines are the [5%, 95%] confidence bounds on these point estimates. The middle dashed line is the bias in the point estimates. These are calculated using Monte Carlo simulations of the restricted model.

Panel A shows that based on full-sample results, the expected quarterly excess return to a five-year bond declines 36 basis points for every one percentage point increase in inflation. (This is the value of the solid line at the five-year maturity point on the $x$ axis.) Under the null hypothesis, the true value is zero. The middle dashed line shows that there is minimal bias in these point estimates if the null hypothesis is true. The point estimates are well outside of the confidence bounds, consistent with the strong statistical rejection of the null hypothesis over the full sample. Panels C and E show that the relation between the other macro variables and expected excess returns is weaker, both statistically and economically. Qualitatively, all of these results are similar to the regression results in Table 2.

There is another message in Panel E, which considers the effect of a one percentage point increase in the short rate. Under the null of no effect, the coefficient is biased up. The reason is the predictive regressions bias of Stambaugh (1999). The short rate is highly persistent, and innovations to the short rate and realized bond returns are negatively contemporaneously correlated. Thus in a finite sample, the short rate will appear to be positively correlated with future excess returns.

In the post-1984 sample, all evidence that the macro variables forecast excess bond returns disappears. The point estimates in Table 5 indicate that around 10 percent of the variance in the conditional means of the excess bond returns are attributable to variations in the macro variables. Ignoring the bias in these point estimates, the theoretical $R^2$ of a regression in which macro variables forecast quarterly excess returns to the two-year bond is close to one percent. But the point estimate of 10 percent is actually smaller than mean fraction from Monte Carlo simulations of the null. The LR test in Table 4 does not come close to rejecting the null hypothesis. Finally, Fig. 2 shows that the effects of the individual macro variables on expected excess returns are all small and well inside the null hypothesis confidence bounds.
5 Finite-sample properties of forecasting regressions

This section studies finite-sample properties of forecasting regressions when the true data-generating process satisfies the general null hypothesis. Given a data generating process, Monte Carlo simulations generate 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?

Section 5.1 uses univariate regressions to study the role of the null hypothesis. The main conclusion is that finite-sample distributions of \( t \)-tests under the general null have little in common with either asymptotic distributions of these tests or finite-sample distributions under the restrictive null. Section 5.2 uses the general null hypothesis to calculate finite sample test statistics for the regressions presented in Section 2.2.4. After adjusting for finite-sample properties, the seemingly strong evidence of return predictability in Section 2.2.4 disappears.

5.1 Univariate forecasting regressions

Denote the log excess return to some asset from \( t \) to \( t + 1 \) as \( r_{x,t,t+1} \). Consider a univariate relation between excess returns over a holding period of length \( k \) and a lagged variable \( w_t \), as in

\[
\sum_{i=0}^{k-1} r_{x,t+k,t+k+1} = \beta_0 + \beta_1 w_t + \theta_{t,t+k}.
\]  

(30)

Estimation of (30) with OLS produces parameter estimates \( b_0 \) and \( b_1 \) and a \( t \)-test of the hypothesis that \( \beta_1 = 0 \). In the context of stock returns, Hodrick (1992) and Stambaugh (1999) discuss finite-sample properties of such regressions under the restrictive null hypothesis that single-period log excess returns are serially uncorrelated. They identify two reasons why finite-sample properties diverge from asymptotic properties. First, typical choices of \( w_t \) are both persistent and contemporaneously correlated with returns, leading to a predictive regressions bias. Second, estimates of the uncertainty in the estimate of \( \beta_1 \) are poor when \( k > 1 \) and overlapping observations are used in estimation. The additional problem introduced with the general null hypothesis is that the residual \( \zeta_{t,t+1} \) is serially correlated.

How important is serial correlation of the residual? Naturally, the answer depends on both the properties of the excess return and the forecasting variables. Here, I restrict my attention to quarterly excess returns to a two-year bond and forecasts made with either inflation or the short rate. I construct finite sample distributions of \( b_1 \) and its associated
5.1.1 The data-generating processes

The first data-generating process satisfies the general null hypothesis. It is the restricted version ($\lambda_{1f} = 0$) of the dynamic model estimated in the previous section. Simulations are produced using two sets of true parameter values, corresponding to estimates from the 1955–2004 period and the 1985–2004 period. Simulations generate observations of inflation, the short rate, and yields on bonds with maturities of 2 years and 1 3/4 years (to calculate quarterly returns).

Recall that because of measurement error, the model allows for a distinction between observed (measured with error) and latent (no measurement error) inflation, output growth, and bond yields. The simulations include measurement error in inflation and output growth but not in yields. There are two reasons to exclude measurement error from yields. First, estimates of standard deviations of measurement error in bond yields are small, as noted in footnote 11. Second, the real role played by measurement error in yields is that of a catch-all, picking up model misspecification. Here we treat the model as correct, thus including measurement error in bond yields simply introduces noise into bond returns.

The second data-generating process satisfies the restrictive null hypothesis that single-period excess returns are serially uncorrelated. There is no obvious choice of process here. One possibility is to use the same model used with the general null hypothesis, but exclude the latent term premia factors. Another possibility is to use a much less elaborate term structure model, along the lines of the processes used by Bekaert et al. (1997) in their analysis of finite-sample properties of tests of the expectations hypothesis. I use a simpler approach here that does not require a term structure model. Following Stambaugh (1999), the model is (in the case of inflation)

$$rw_{t,t+1} = \mu_r + \theta_{t+1},$$

$$\pi_{t+1} = \mu_w + \rho \pi_t + \nu_{t+1},$$

$$\begin{pmatrix} \theta_t \\ \nu_t \end{pmatrix} = \text{MVN} \left( 0, \begin{pmatrix} \sigma^2_{\theta} & \sigma_{\theta\pi} \\ \sigma_{\pi\nu} & \sigma^2_{\nu} \end{pmatrix} \right).$$

Finite-sample properties of regressions using the short rate are examined using the same process, replacing inflation in (32) with the short rate. The “true” parameters for this joint process are taken from sample values over 1955 through 2004 and 1985 through 2004. Because the data used in this paper do not include quarterly excess returns to a two-year
zero-coupon bond, I use as a proxy quarterly excess returns to a portfolio of coupon bonds with maturities between two and three years. These data are discussed in more detail in Section 2.2.1.\footnote{An alternative proxy is the time series of filtered values of the quarterly excess return to a two-year zero-coupon bond using estimates of the restricted version of the dynamic term structure model from Section 4.4. With this proxy, the empirical rejection rates of the hypothesis that $\beta_1 = 0$ are almost identical to the results presented in the paper.}

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 200 quarters, hence the number of observations available for the regression is $200 - k$. I use quarterly ($k = 1$) and annual ($k = 4$) return horizons.

Regressions are estimated with OLS. For the restrictive null, robust Hansen-Hodrick $t$-statistics are used. For the general null, the Newey-West $t$-statistics are used with $k + 3$ lags. In principle, robust $t$ statistics are unnecessary because the true data-generating processes are homoskedastic. But the point of this exercise is to shed light on the properties of tests that are used in practice. Robust test statistics are necessary in practice because real-world dynamics exhibit conditional heteroskedasticity. Unfortunately, there is a tradeoff in dynamic term structure models between flexibility in specifying conditional covariances and flexibility in specifying drifts. The analysis here thus uses the special case of homoskedasticity to study the behavior of tests that are robust to heteroskedasticity.

5.1.2 Results

Table 6 presents results based on 5000 simulations. The first four columns specify the type of simulation: the return horizon, the explanatory variable in the regression, the sample period used to determine the true parameters, and the type of data-generating process. For regressions that use inflation, the table reports results using parameter estimates from both sample periods. Inspection of the table reveals that the choice of sample period has little effect on the simulation results. Accordingly, for the regressions that use the short rate, only results based on parameters from 1985 through 2004 are shown.

The mean coefficients are all positive because of the predictive regressions bias. Bond returns are negatively contemporaneously correlated with shocks to both inflation and the short rate. In addition, both inflation and the short rate are persistent. Over the full sample, their first-order autocorrelation coefficients are 0.89 and 0.93 respectively. In combination, these features lead to an upward-biased point estimate.

Notwithstanding this bias, asymptotic and finite-sample distributions under the restrictive null hypothesis are similar when using quarterly returns. The first two rows in the table
show that with this null, the hypothesis that inflation forecasts excess returns is rejected at the 5 percent asymptotic critical value in less than 6 percent of the simulations. The fourth row from the bottom shows that when forecasting with the short rate, the corresponding empirical rejection rate is less than 8 percent. Naturally, the divergence between these two distributions is larger for annual returns. Empirical rejection rates are between 8 and 9 percent when using inflation and around 11 percent when using the short rate.

With the general null hypothesis, the finite-sample properties of the regressions are substantially worse. Across both return horizons and both forecasting variables, the empirical rejection rates at the 5 percent asymptotic level are approximately twice as high as the empirical rejection rates under the restrictive null. For example, regressions of quarterly returns on inflation reject the null hypothesis at the 5 percent asymptotic level in 11 percent of the simulations. When using the short rate to forecast, the empirical rejection rate rises to 19 percent.

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.

5.2 Multivariate forecasting regressions

This subsection presents additional evidence on the finite-sample properties of forecasting regressions. The regressions of interest are the in-sample and out-of-sample regressions used in Section 2.2, where excess returns are regressed on inflation, output growth, and the short rate. Monte Carlo simulations are used to calculate empirical rejection rates and finite-sample critical values for these regressions under the general null hypothesis.

The data-generating process for all of these simulations is the restricted version of the dynamic term structure model estimated over the sample 1985 through 2004. Mimicking the results in Tables 1 and 2, samples of both 200 and 80 quarters are simulated. The annual log excess return regressions of Table 1 are duplicated with these simulations. The quarterly simple excess return regressions of Table 2 cannot be simulated exactly because the data used in Table 2 are returns to portfolios of coupon bonds. In the simulations, I use two-year and five-year zero-coupon bonds instead of coupon bond portfolios with maturities from two to three years and five to ten years, respectively.

---

14 To be precise, the asymptotic distribution is the distribution appropriate for the restrictive null hypothesis. Because the Newey-West correction does not capture all of the serial correlation in the residual, the asymptotic distribution under the general null hypothesis is not known.

15 Results using the model estimated over the sample 1955 through 2004 differ somewhat from those presented here, but not in a systematic way.
Simulation results are displayed in Table 7. The first four columns specify the type of regression (in-sample or out-of-sample), the return horizon, the length of the simulation, and the bond’s maturity. The final two columns contain properties of the Wald test that the coefficients on inflation, output growth, and the short rate are jointly zero. The test has an asymptotic $\chi^2(3)$ distribution under the restrictive null hypothesis. The 5 percent critical value for this test is 7.81. The fifth column reports the empirical rejection rate at this critical value, and the final column reports the true critical value required for a 5 percent rejection rate.

Given the evidence of the previous subsection, it is not surprising that the finite-sample properties of the in-sample regressions are not close to their asymptotic properties. For example, the first two rows of the table report that full-sample regressions (200 quarterly observations) on annual returns have empirical rejection rates in excess of 20% at the 5% asymptotic critical value. A test statistic in the neighborhood of 15 is necessary to reject the general null hypothesis at the correct 5% level. With 80 quarterly observations, the corresponding empirical rejection rates exceed 40%. Using quarterly return horizons, differences between asymptotic and finite-sample distributions are not quite as dramatic. Nonetheless, empirical rejection rates at the 5% level are about 15% with 200 quarterly observations and about 25% with 80 quarterly observations. Distributions of test statistics for out-of-sample regressions are better behaved. (Recall these statistics are computed only for samples of 200 quarterly observations.) Empirical rejection rates at the asymptotic 5% level are between 7 and 12 percent.

The main message in these results is that there is no reliable statistical evidence of return forecastability in the regressions of Tables 1 and 2. Recall that in Table 1, tests of annual return forecastability strongly reject the null of no forecastability using asymptotic critical values. If, however, we use the true 5% critical values computed in Table 7, none of the test statistics reported in Table 1 are significantly different from zero. This conclusion holds for both in-sample and out-of-sample tests. Of the statistics reported for quarterly returns in Table 2, only one differs from zero at the true 5% level; the out-of-sample forecasts of the five-year bond return. It is hard to put much weight on this single test statistic, given that none of the other tests support the same conclusion.

Recall that in Section 4, a likelihood ratio test strongly rejected (using the finite-sample distribution of the test) the general null hypothesis over the 1955–2004 sample. What explains the failure to reject this null with forecasting regressions? If the model is correctly specified, ML estimation of the model is a more powerful method of testing the null hypothesis. But the evidence of regime changes mentioned in Section 4.1 is inconsistent with the model. The regressions, although less powerful, are more robust. A fair conclusion is that
there is no statistically reliable evidence that variations in expected excess bond returns are related to inflation, output growth, or the short rate. More importantly, even if there truly is a link between these macroeconomic variables and term premia, the economic significance is small.

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 1955 through 2004, the model indicates that expected excess returns are only weakly related to inflation, output growth, and the short rate.
References


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


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.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Maturity (years)</th>
<th>Output growth</th>
<th>Short rate</th>
<th>$R^2$</th>
<th>Joint test</th>
<th>P-val</th>
<th>Out of sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>1955:1–2004:4</td>
<td>2</td>
<td>-0.327</td>
<td>0.267</td>
<td>0.14</td>
<td>11.39</td>
<td>[0.010]</td>
<td>1.96</td>
</tr>
<tr>
<td>(196)</td>
<td></td>
<td>(0.144)</td>
<td>(0.117)</td>
<td></td>
<td>(0.101)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>-1.091</td>
<td>0.647</td>
<td>0.14</td>
<td>11.02</td>
<td>[0.012]</td>
<td>1.87</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.413)</td>
<td>(0.395)</td>
<td></td>
<td>(0.329)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.396)</td>
<td>(0.122)</td>
<td></td>
<td>(0.120)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1985:1–2004:4</td>
<td>2</td>
<td>0.037</td>
<td>0.138</td>
<td>0.10</td>
<td>8.68</td>
<td>[0.034]</td>
<td></td>
</tr>
<tr>
<td>(76)</td>
<td></td>
<td>(0.241)</td>
<td>(0.148)</td>
<td></td>
<td>(0.120)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.234)</td>
<td>(0.120)</td>
<td></td>
<td>(0.109)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.282</td>
<td>0.254</td>
<td>0.03</td>
<td>1.82</td>
<td>[0.610]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.744)</td>
<td>(0.558)</td>
<td></td>
<td>(0.462)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.727)</td>
<td>(0.264)</td>
<td></td>
<td>(0.260)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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.

<table>
<thead>
<tr>
<th>Sample (2 obs)</th>
<th>Maturities (years)</th>
<th>Inflation</th>
<th>Output growth</th>
<th>Short rate</th>
<th>$R^2$</th>
<th>Joint test</th>
<th>P-val sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>1955:1–2004:4</td>
<td>$2 &lt; \tau \leq 3$</td>
<td>-0.193</td>
<td>-0.047</td>
<td>0.143</td>
<td>0.04</td>
<td>7.04</td>
<td>0.070</td>
</tr>
<tr>
<td>(199)</td>
<td></td>
<td>(0.078)</td>
<td>(0.036)</td>
<td>(0.103)</td>
<td></td>
<td></td>
<td>(0.078)</td>
</tr>
<tr>
<td></td>
<td>$5 &lt; \tau \leq 10$</td>
<td>-0.374</td>
<td>-0.065</td>
<td>0.243</td>
<td>0.04</td>
<td>7.70</td>
<td>0.053</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.142)</td>
<td>(0.062)</td>
<td>(0.174)</td>
<td></td>
<td></td>
<td>(0.142)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.147)</td>
<td>(0.061)</td>
<td>(0.132)</td>
<td></td>
<td></td>
<td>(0.147)</td>
</tr>
<tr>
<td>1985:1–2004:4</td>
<td>$2 &lt; \tau \leq 3$</td>
<td>0.037</td>
<td>-0.032</td>
<td>0.084</td>
<td>0.02</td>
<td>1.68</td>
<td>0.641</td>
</tr>
<tr>
<td>(79)</td>
<td></td>
<td>(0.217)</td>
<td>(0.070)</td>
<td>(0.091)</td>
<td></td>
<td></td>
<td>(0.217)</td>
</tr>
<tr>
<td></td>
<td>$5 &lt; \tau \leq 10$</td>
<td>0.029</td>
<td>0.022</td>
<td>0.211</td>
<td>0.02</td>
<td>1.66</td>
<td>0.646</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.430)</td>
<td>(0.142)</td>
<td>(0.187)</td>
<td></td>
<td></td>
<td>(0.430)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.340)</td>
<td>(0.134)</td>
<td>(0.181)</td>
<td></td>
<td></td>
<td>(0.340)</td>
</tr>
</tbody>
</table>
Table 3. Estimates of a dynamic term structure model for 1955 through 2004

The joint dynamics of inflation $\pi_t$, output growth $\Delta g_t$, and the short rate $r_t$ are described by the model of Section 3.2. The model is estimated with maximum likelihood over the period 1955Q1 through 2004Q4, using these data as well as yields on zero-coupon bonds with maturities from one to five years. Standard errors are in parentheses and the $p$-value of a likelihood ratio test is in brackets. They are calculated with Monte Carlo simulations.

Panel A. Macro factor dynamics

<table>
<thead>
<tr>
<th></th>
<th>$\mu_f \times 10^2$</th>
<th>$\pi_{t-1}$</th>
<th>$K_f$</th>
<th>$\Delta g_{t-1}$</th>
<th>$r_{t-1}$</th>
<th>$\Sigma_f \times 10^3$</th>
<th>Std dev of obs error $\times 10^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\pi_t$</td>
<td>-0.508</td>
<td>1.019</td>
<td>0.133</td>
<td>-0.001</td>
<td>5.353</td>
<td></td>
<td>0.698</td>
</tr>
<tr>
<td></td>
<td>(0.269)</td>
<td>(0.052)</td>
<td>(0.046)</td>
<td>(0.030)</td>
<td>(1.108)</td>
<td></td>
<td>(0.072)</td>
</tr>
<tr>
<td>$\Delta g_t$</td>
<td>2.545</td>
<td>0.020</td>
<td>0.546</td>
<td>-0.212</td>
<td>-9.868</td>
<td>16.588</td>
<td>2.720</td>
</tr>
<tr>
<td></td>
<td>(0.925)</td>
<td>(0.192)</td>
<td>(0.131)</td>
<td>(0.105)</td>
<td>(4.529)</td>
<td>(5.371)</td>
<td>(0.568)</td>
</tr>
<tr>
<td>$r_t$</td>
<td>0.014</td>
<td>0.168</td>
<td>0.053</td>
<td>0.853</td>
<td>0.055</td>
<td>5.465</td>
<td>8.494</td>
</tr>
<tr>
<td></td>
<td>(0.285)</td>
<td>(0.050)</td>
<td>(0.034)</td>
<td>(0.033)</td>
<td>(1.426)</td>
<td>(1.685)</td>
<td>(2.552)</td>
</tr>
</tbody>
</table>

Panel B. Price of macro risk

<table>
<thead>
<tr>
<th></th>
<th>$\lambda_{0f}$</th>
<th>$\lambda_{1f}$</th>
<th>$\pi_t$</th>
<th>$\Delta g_t$</th>
<th>$r_t$</th>
<th>LR test of $\lambda_{1f} = 0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\pi_t$</td>
<td>-0.647</td>
<td>-0.365</td>
<td>0.164</td>
<td>0.122</td>
<td></td>
<td>38.64</td>
</tr>
<tr>
<td></td>
<td>(0.099)</td>
<td>(0.029)</td>
<td>(0.055)</td>
<td>(0.096)</td>
<td></td>
<td>[0.00]</td>
</tr>
<tr>
<td>$\Delta g_t$</td>
<td>1.554</td>
<td>0.420</td>
<td>-0.950</td>
<td>-0.580</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.161)</td>
<td>(0.081)</td>
<td>(0.103)</td>
<td>(0.204)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$r_t$</td>
<td>-0.010</td>
<td>-0.062</td>
<td>0.169</td>
<td>0.050</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.033)</td>
<td>(0.052)</td>
<td>(0.035)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Panel C. Term premia factor dynamics

<table>
<thead>
<tr>
<th></th>
<th>$K_{\omega}$</th>
<th>$\Sigma_{\omega} \times 10^3$</th>
<th>$\omega_{1t}$</th>
<th>$\omega_{2t}$</th>
<th>$\omega_{3t}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega_{1t}$</td>
<td>1.210</td>
<td>3.230</td>
<td>0.166</td>
<td>0.054</td>
<td>0.737</td>
</tr>
<tr>
<td></td>
<td>(0.138)</td>
<td></td>
<td>(0.054)</td>
<td>(0.075)</td>
<td>(23.07)</td>
</tr>
<tr>
<td>$\omega_{2t}$</td>
<td>-1.054</td>
<td>-13.728</td>
<td>1.077</td>
<td>1.536</td>
<td>16.693</td>
</tr>
<tr>
<td></td>
<td>(0.222)</td>
<td></td>
<td>(0.075)</td>
<td>(1.536)</td>
<td>(41.17)</td>
</tr>
<tr>
<td>$\omega_{3t}$</td>
<td>-0.135</td>
<td>-1.703</td>
<td>0.085</td>
<td>0.186</td>
<td>3.495</td>
</tr>
<tr>
<td></td>
<td>(0.089)</td>
<td></td>
<td>(0.025)</td>
<td>(0.117)</td>
<td>(0.577)</td>
</tr>
</tbody>
</table>
Table 4. Estimates of a dynamic term structure model for 1985 through 2004

The joint dynamics of inflation $\pi_t$, output growth $\Delta g_t$, and the short rate $r_t$ are described by the model of Section 3.2. The model is estimated with maximum likelihood over the period 1985Q1 through 2004Q4, using these data as well as yields on zero-coupon bonds with maturities from one to five years. Standard errors are in parentheses and the $p$-value of a likelihood ratio test is in brackets. They are calculated with Monte Carlo simulations.

Panel A. Macro factor dynamics

<table>
<thead>
<tr>
<th></th>
<th>$\mu_f \times 10^2$</th>
<th>$\pi_t$</th>
<th>$K_f$</th>
<th>$\Delta g_{t-1}$</th>
<th>$r_{t-1}$</th>
<th>$\Sigma_f \times 10^3$</th>
<th>Std dev of obs error $\times 10^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\pi_t$</td>
<td>0.768</td>
<td>0.546</td>
<td>-0.050</td>
<td>0.093</td>
<td>5.131</td>
<td>0.477</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.415)</td>
<td>(0.100)</td>
<td>(0.088)</td>
<td>(0.054)</td>
<td>(1.147)</td>
<td>(0.179)</td>
<td></td>
</tr>
<tr>
<td>$\Delta g_t$</td>
<td>0.432</td>
<td>0.257</td>
<td>0.868</td>
<td>-0.120</td>
<td>-3.608</td>
<td>5.666</td>
<td>1.692</td>
</tr>
<tr>
<td></td>
<td>(0.524)</td>
<td>(0.098)</td>
<td>(0.092)</td>
<td>(0.063)</td>
<td>(1.721)</td>
<td>(1.377)</td>
<td>(0.145)</td>
</tr>
<tr>
<td>$r_t$</td>
<td>-1.503</td>
<td>0.333</td>
<td>0.372</td>
<td>0.901</td>
<td>2.162</td>
<td>1.642</td>
<td>3.534</td>
</tr>
<tr>
<td></td>
<td>(0.474)</td>
<td>(0.101)</td>
<td>(0.093)</td>
<td>(0.059)</td>
<td>(0.748)</td>
<td>(0.704)</td>
<td>(0.801)</td>
</tr>
</tbody>
</table>

Panel B. Price of macro risk

<table>
<thead>
<tr>
<th></th>
<th>$\lambda_0f \times 10^2$</th>
<th>$\lambda_{1f}$</th>
<th>$\lambda_{1f} \times 10^2$</th>
<th>$\lambda_{1f} \times 10^2$</th>
<th>$\lambda_{1f} \times 10^2$</th>
<th>$\lambda_{1f} \times 10^2$</th>
<th>$\lambda_{1f} \times 10^2$</th>
<th>$\lambda_{1f} \times 10^2$</th>
<th>$\lambda_{1f} \times 10^2$</th>
<th>$\lambda_{1f} \times 10^2$</th>
<th>$\lambda_{1f} \times 10^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\pi_t$</td>
<td>-0.002</td>
<td>0.042</td>
<td>-0.117</td>
<td>-0.162</td>
<td>5.66</td>
<td>[0.76]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.007)</td>
<td>(0.070)</td>
<td>(0.080)</td>
<td>(0.053)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Delta g_t$</td>
<td>0.129</td>
<td>-0.136</td>
<td>0.191</td>
<td>0.198</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.292)</td>
<td>(0.057)</td>
<td>(0.074)</td>
<td>(0.054)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$r_t$</td>
<td>-0.300</td>
<td>0.049</td>
<td>0.026</td>
<td>-0.037</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.344)</td>
<td>(0.078)</td>
<td>(0.057)</td>
<td>(0.044)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Panel C. Term premia factor dynamics

<table>
<thead>
<tr>
<th></th>
<th>$K_\omega$</th>
<th>$\Sigma_\omega \times 10^3$</th>
<th>$\omega_{1t}$</th>
<th>$\omega_{2t}$</th>
<th>$\omega_{3t}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega_{1t}$</td>
<td>0.703</td>
<td>4.292</td>
<td>-0.110</td>
<td>-0.067</td>
<td>18.22</td>
</tr>
<tr>
<td></td>
<td>(0.111)</td>
<td>(18.22)</td>
<td>(0.103)</td>
<td>(0.361)</td>
<td>(0.094)</td>
</tr>
<tr>
<td>$\omega_{2t}$</td>
<td>-0.114</td>
<td>-3.693</td>
<td>0.905</td>
<td>0.276</td>
<td>11.60</td>
</tr>
<tr>
<td></td>
<td>(0.094)</td>
<td>(11.60)</td>
<td>(0.079)</td>
<td>(0.252)</td>
<td>(0.054)</td>
</tr>
<tr>
<td>$\omega_{3t}$</td>
<td>0.069</td>
<td>-0.512</td>
<td>-0.024</td>
<td>0.871</td>
<td>0.645</td>
</tr>
<tr>
<td></td>
<td>(0.054)</td>
<td>(0.645)</td>
<td>(0.046)</td>
<td>(0.121)</td>
<td>(0.046)</td>
</tr>
</tbody>
</table>
Table 5. Model-implied behavior of quarterly excess bond returns

Parameter estimates of a dynamic term structure model are used to calculate properties of quarterly log excess bond returns. The excess return in quarter $t$ is the sum of its expectation as of $t-1$ and an orthogonal 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. The final column reports the first-order serial correlation of the conditional expectation of log excess returns.

The model is estimated over two samples and in both restricted and unrestricted versions. The restriction is that macro variables are unrelated to variations in conditional expected excess returns. Means (in parentheses) and [5% 95%] confidence intervals of the reported properties are calculated with simulations, where the assumed true model is the restricted version for the relevant sample.

<table>
<thead>
<tr>
<th>Sample period</th>
<th>Maturity (quarters)</th>
<th>Mean (%/q)</th>
<th>Std dev (%/q)</th>
<th>Fraction of return var due to cond expect</th>
<th>Fraction of cond expect due to macro</th>
<th>AR(1) of cond expect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel A: Estimates from the restricted model</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1955–2004</td>
<td>8</td>
<td>0.03</td>
<td>1.77</td>
<td>0.16</td>
<td>0</td>
<td>0.55</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.06)</td>
<td>(1.75)</td>
<td>(0.15)</td>
<td>(0.49)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>[−0.41 0.43]</td>
<td>[1.62 1.89]</td>
<td>[0.12 0.19]</td>
<td>[0.35 0.62]</td>
<td></td>
</tr>
<tr>
<td>1955–2004</td>
<td>20</td>
<td>−0.04</td>
<td>3.68</td>
<td>0.17</td>
<td>0</td>
<td>0.55</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.02)</td>
<td>(3.64)</td>
<td>(0.16)</td>
<td>(0.52)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>[−0.86 0.68]</td>
<td>[3.37 3.95]</td>
<td>[0.12 0.20]</td>
<td>[0.39 0.63]</td>
<td></td>
</tr>
<tr>
<td>1985–2004</td>
<td>8</td>
<td>0.07</td>
<td>1.15</td>
<td>0.13</td>
<td>0</td>
<td>0.94</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.08)</td>
<td>(1.12)</td>
<td>(0.11)</td>
<td>(0.92)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>[−0.16 0.34]</td>
<td>[1.00 1.25]</td>
<td>[0.06 0.17]</td>
<td>[0.82 0.97]</td>
<td></td>
</tr>
<tr>
<td>1985–2004</td>
<td>20</td>
<td>0.15</td>
<td>2.94</td>
<td>0.10</td>
<td>0</td>
<td>0.87</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.16)</td>
<td>(2.87)</td>
<td>(0.09)</td>
<td>(0.82)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>[−0.34 0.66]</td>
<td>[2.52 3.24]</td>
<td>[0.05 0.13]</td>
<td>[0.64 0.94]</td>
<td></td>
</tr>
<tr>
<td>Panel B: Estimates from the unrestricted model</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1955–2004</td>
<td>8</td>
<td>0.27</td>
<td>1.75</td>
<td>0.21</td>
<td>0.21</td>
<td>0.63</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.06)</td>
<td>(1.75)</td>
<td>(0.16)</td>
<td>(0.07)</td>
<td>(0.50)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[−0.41 0.44]</td>
<td>[1.61 1.89]</td>
<td>[0.13 0.20]</td>
<td>[0.00 0.18]</td>
<td>[0.33 0.63]</td>
</tr>
<tr>
<td>1955–2004</td>
<td>20</td>
<td>0.35</td>
<td>3.63</td>
<td>0.17</td>
<td>0.14</td>
<td>0.57</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.01)</td>
<td>(3.64)</td>
<td>(0.17)</td>
<td>(0.05)</td>
<td>(0.53)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[−0.86 0.68]</td>
<td>[3.33 3.96]</td>
<td>[0.13 0.22]</td>
<td>[0.00 0.15]</td>
<td>[0.38 0.66]</td>
</tr>
<tr>
<td>1985–2004</td>
<td>8</td>
<td>0.12</td>
<td>1.18</td>
<td>0.17</td>
<td>0.09</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.07)</td>
<td>(1.12)</td>
<td>(0.14)</td>
<td>(0.23)</td>
<td>(0.87)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[−0.16 0.33]</td>
<td>[1.00 1.26]</td>
<td>[0.07 0.23]</td>
<td>[0.03 0.51]</td>
<td>[0.72 0.96]</td>
</tr>
<tr>
<td>1985–2004</td>
<td>20</td>
<td>0.26</td>
<td>3.02</td>
<td>0.13</td>
<td>0.11</td>
<td>0.88</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.15)</td>
<td>(2.88)</td>
<td>(0.12)</td>
<td>(0.24)</td>
<td>(0.80)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[−0.37 0.68]</td>
<td>[2.53 3.22]</td>
<td>[0.07 0.18]</td>
<td>[0.03 0.54]</td>
<td>[0.62 0.92]</td>
</tr>
</tbody>
</table>
Table 6. Finite-sample properties of univariate forecasting regressions

This table summarizes results from 5000 Monte Carlo simulations. Excess returns to two-year bonds from quarter-end $t$ to quarter-end $t+k$ are regressed on the quarter-end observation of either inflation or the short rate. The “true” data-generating process depends on the choice of null hypothesis. Under the restrictive null of no return predictability, the forecasting variable follows an AR(1) process and quarterly returns are iid. Under the general null that returns are truly forecastable, but are independent of lagged inflation and the short rate, the data-generating process is a dynamic term structure model. The parameters used in these processes are based on the sample period listed in “Source of true params.”

For each simulation, 200 quarters of data on are generated. Quarterly and annual ($k = 1$ and $k = 4$) return horizons are used. Regressions with annual returns use overlapping observations. The hypothesis that the coefficient on the forecasting variable is zero is tested using a $t$-statistic. Under the restrictive null, the statistic uses generalized Hansen-Hodrick standard errors. Under the general null, the statistic uses Newey-West standard errors with $k + 3$ lags. The table reports the mean and standard deviation of the estimated parameters, the mean standard error, and the empirical rejection rate of the hypothesis that the coefficient is zero, using the 5% asymptotic critical value.

<table>
<thead>
<tr>
<th>Type of return</th>
<th>Forecast variable</th>
<th>Source of true params</th>
<th>Null hypothesis</th>
<th>Mean coef</th>
<th>Std dev of coef</th>
<th>Mean std err</th>
<th>Rejection rate at 5% asy crit val</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quarterly</td>
<td>Inflation</td>
<td>1955-2004</td>
<td>restrictive</td>
<td>0.006</td>
<td>0.061</td>
<td>0.060</td>
<td>0.055</td>
</tr>
<tr>
<td>Quarterly</td>
<td>Inflation</td>
<td>1985-2004</td>
<td>restrictive</td>
<td>0.007</td>
<td>0.113</td>
<td>0.112</td>
<td>0.059</td>
</tr>
<tr>
<td>Annual</td>
<td>Inflation</td>
<td>1955-2004</td>
<td>restrictive</td>
<td>0.020</td>
<td>0.226</td>
<td>0.207</td>
<td>0.087</td>
</tr>
<tr>
<td>Annual</td>
<td>Inflation</td>
<td>1985-2004</td>
<td>restrictive</td>
<td>0.022</td>
<td>0.337</td>
<td>0.316</td>
<td>0.082</td>
</tr>
<tr>
<td>Quarterly</td>
<td>Inflation</td>
<td>1955-2004</td>
<td>general</td>
<td>0.009</td>
<td>0.075</td>
<td>0.062</td>
<td>0.109</td>
</tr>
<tr>
<td>Quarterly</td>
<td>Inflation</td>
<td>1985-2004</td>
<td>general</td>
<td>0.027</td>
<td>0.103</td>
<td>0.087</td>
<td>0.113</td>
</tr>
<tr>
<td>Annual</td>
<td>Inflation</td>
<td>1955-2004</td>
<td>general</td>
<td>0.032</td>
<td>0.275</td>
<td>0.195</td>
<td>0.176</td>
</tr>
<tr>
<td>Annual</td>
<td>Inflation</td>
<td>1985-2004</td>
<td>general</td>
<td>0.098</td>
<td>0.357</td>
<td>0.282</td>
<td>0.153</td>
</tr>
<tr>
<td>Quarterly</td>
<td>Short rate</td>
<td>1985-2004</td>
<td>restrictive</td>
<td>0.041</td>
<td>0.073</td>
<td>0.066</td>
<td>0.077</td>
</tr>
<tr>
<td>Annual</td>
<td>Short rate</td>
<td>1985-2004</td>
<td>restrictive</td>
<td>0.155</td>
<td>0.271</td>
<td>0.235</td>
<td>0.112</td>
</tr>
<tr>
<td>Quarterly</td>
<td>Short rate</td>
<td>1985-2004</td>
<td>general</td>
<td>0.026</td>
<td>0.058</td>
<td>0.041</td>
<td>0.189</td>
</tr>
<tr>
<td>Annual</td>
<td>Short rate</td>
<td>1985-2004</td>
<td>general</td>
<td>0.100</td>
<td>0.227</td>
<td>0.154</td>
<td>0.211</td>
</tr>
</tbody>
</table>
This table summarizes results from 5000 Monte Carlo simulations. The true data-generating process of yields, inflation, and output growth is based on the maximum likelihood estimate of a dynamic term structure model. The model implies that expected excess returns are stochastic and persistent, but are independent of lagged inflation, output growth and the short rate.

For each simulation, either 80 or 200 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 statistic uses the Newey-West adjustment with three lags for quarterly returns and seven lags for annual returns. The table reports the empirical rejection rate using the 5% critical value for a $\chi^2(3)$ distribution (7.81), as well as the finite sample 5% critical value. Similar statistics are reported for the out-of-sample $t$-test of Ericsson (1992). This $t$-test, which is constructed only for the longer simulations, has an asymptotic $N(0,1)$ distribution and is discussed in more detail in the text.

<table>
<thead>
<tr>
<th>Type of regression</th>
<th>Type of return</th>
<th>Length of simulation (quarters)</th>
<th>Maturity (years)</th>
<th>Rejection rate at 5% asy crit val</th>
<th>True 5% critical value</th>
</tr>
</thead>
<tbody>
<tr>
<td>In-sample</td>
<td>Annual</td>
<td>200</td>
<td>2</td>
<td>0.245</td>
<td>16.45</td>
</tr>
<tr>
<td>In-sample</td>
<td>Annual</td>
<td>200</td>
<td>5</td>
<td>0.208</td>
<td>14.61</td>
</tr>
<tr>
<td>In-sample</td>
<td>Annual</td>
<td>80</td>
<td>2</td>
<td>0.441</td>
<td>34.42</td>
</tr>
<tr>
<td>In-sample</td>
<td>Annual</td>
<td>80</td>
<td>5</td>
<td>0.409</td>
<td>30.84</td>
</tr>
<tr>
<td>In-sample</td>
<td>Quarterly</td>
<td>200</td>
<td>2</td>
<td>0.182</td>
<td>13.20</td>
</tr>
<tr>
<td>In-sample</td>
<td>Quarterly</td>
<td>200</td>
<td>5</td>
<td>0.137</td>
<td>11.14</td>
</tr>
<tr>
<td>In-sample</td>
<td>Quarterly</td>
<td>80</td>
<td>2</td>
<td>0.293</td>
<td>19.24</td>
</tr>
<tr>
<td>In-sample</td>
<td>Quarterly</td>
<td>80</td>
<td>5</td>
<td>0.247</td>
<td>16.82</td>
</tr>
<tr>
<td>Out-of-sample</td>
<td>Annual</td>
<td>200</td>
<td>2</td>
<td>0.116</td>
<td>2.36</td>
</tr>
<tr>
<td>Out-of-sample</td>
<td>Annual</td>
<td>200</td>
<td>5</td>
<td>0.096</td>
<td>2.22</td>
</tr>
<tr>
<td>Out-of-sample</td>
<td>Quarterly</td>
<td>200</td>
<td>2</td>
<td>0.112</td>
<td>2.25</td>
</tr>
<tr>
<td>Out-of-sample</td>
<td>Quarterly</td>
<td>200</td>
<td>5</td>
<td>0.066</td>
<td>1.80</td>
</tr>
</tbody>
</table>
Fig. 1. A comparison of observed inflation, output growth, and bond yields with model-implied estimates. The solid lines in each panel are actual inflation (Panels A and B), output growth (Panels C and D), and the five-year zero-coupon bond yield (Panels E and F) over the specified dates. The dashed lines in Panels A through D represent filtered values of latent, “core” inflation and output growth based on estimates of a dynamic model of inflation, output growth, and the term structure. The model is estimated separately over the 1955–2004 and 1985–2004 samples. The dashed lines in Panels E and F represent yields on the five-year bond if investors were risk-neutral and their expectations of future short-term interest rates were those implied by the dynamic model.
Fig. 2. Model-implied estimates of the effects of inflation, output growth, and short-term interest rates on expected excess bond returns. A dynamic model of inflation, output growth, and the term structure is estimated separately over the 1955–2004 and 1985-2004 samples. The horizontal axes in the panels represent a bond’s maturity and the vertical axes represent expected excess quarterly returns. The solid lines in each panel are point estimates of the change in expected excess returns given a one percentage point increase in the specified variable—infation, output growth, or the short rate—holding the other variables constant. The dashed lines are the bias and [5%, 95%] confidence bounds on these point estimates under the null hypothesis that none of these variables can truly predict excess bond returns.