

Learning and the Role of Macroeconomic Factors in the Term Structure of Interest Rates

Thomas Laubach, Robert J. Tetlow, and John C. Williams *

October 10, 2007

*Board of Governors of the Federal Reserve System, tlaubach@frb.gov (corresponding author); Board of Governors of the Federal Reserve System, rtetlow@frb.gov; and Federal Reserve Bank of San Francisco, john.c.williams@sf.frb.org. The views expressed herein are those of the authors and do not necessarily reflect those of the Board of Governors of the Federal Reserve System, its staff, or the management of the Federal Reserve Bank of San Francisco.

Abstract

Models of the term structure based on only observable variables have had limited success in explaining movements in longer-term interest rates. A key assumption in much of this literature is that agents know all the parameters describing the model of the economy and that these parameters are fixed for all time. In this paper, we relax both of these assumptions and assume that agents regularly re-estimate the parameters of their models—both those determining the point forecasts and those describing economic volatility—based on incoming data. In this way, we allow for the real-time problem of pricing assets based on the information set available at the time. In addition, we allow for discounting of past data reflecting a concern on the part of agents for structural change in the economy. We find that the learning model with discounting does a much better job at explaining longer-term yields than an equivalent model with constant coefficients estimated over the full sample; in particular, the deviations from the expectations hypothesis are much smaller on average with our learning model. We then estimate the model term premia imposing an affine arbitrage-free structure. We show that incorporating learning improves the in-sample fit and forecasting performance of the model. Learning also implies time variation in the real-time estimates of macroeconomic volatility that is absent in standard macroeconomic models with fixed coefficients. We find that these shifts in estimates of macroeconomic volatility help explain movements in term premia over the past half century. More generally, our analysis highlights the importance of taking into account the information sets of investors in understanding the determinants of bond prices.

JEL Codes: D83, D84, G12

Keywords: affine term structure models, learning, structural change.

1 Introduction

Term structure models in which yields and market prices of risk are linear (“affine”) functions of underlying state variables have been shown to be very successful in explaining several facts about the term structure that have long puzzled researchers in finance. Typically, these affine term structure models have relied on unobserved factors to explain bond yields. To finance professionals this reliance on unobserved factors is of little consequence; they are primarily concerned with fit. To macroeconomists and monetary economists, however, the difficulty in interpreting the origins and dynamic behavior of these unobserved factors is problematic and obscured by their atheoretical nature. The recent macro finance literature has sought to reintroduce a central role for macroeconomic influences on the term structure of interest rates. In particular, Ang and Piazzesi (2003) show that macroeconomic variables have explanatory power for bond yields, but that latent variables still explain a great deal of the variation in yields.¹

Because interest rates, especially at long maturities, depend crucially on expectations of future short-term interest rates, how agents form expectations plays an important role in term structure models with macro factors. Most research in this area is conducted under the assumption that the relationships between macroeconomic variables and interest rates are constant over time and that these are known with certainty by all agents. For example, the typical assumption in the literature is that expectations of future interest rates at all times are made using forecasts from a model estimated on the full sample of data, including data not yet observed by the agent (see, for example, Diebold, Rudebusch, and Aruoba (2006)). However, neither of these assumptions is tenable. The presence of structural breaks among key U.S. macroeconomic variables over the post-war period is well documented; see, e.g., Stock and Watson (1996, 2007). In such an environment, agents’ perceptions of economic relationships will evolve over time as well, and will depend on the information contained in the data. Accordingly, in this paper, we reexamine the role of macroeconomic determinants of interest rates, but do so in a model where agents are learning over time and face the real-time problem of pricing bonds with incomplete information about the macroeconomic environment.

In a realistic setting where the economic environment changes in unpredictable ways

¹See also Rudebusch and Wu (2003) and Dewachter and Lyrio (2006) and references therein for other macro finance models with latent factors.

and agents learn about the economy from observed data, the formation of expectations of future interest rates and the perceptions of risk in the economy will change over time.² As a result, models that falsely assume a time-invariant economy with perfect knowledge will generate expectations of future interest rates and perceptions of risk that are mismeasured, contaminating the model’s predictions of bond yields. The effects of mismeasurement manifest themselves in the comparison of the performance of standard models that assume a fixed known structure, to models that incorporate expectations data as an input to the model. For example, Bernanke, Reinhart, and Sack (2004) develop a model without latent factors that fits the yields data well.³ Their model includes data on survey expectations of inflation and expectations of future interest rates from futures markets in their model, which provide useful explanatory power for yields. However, their approach does not provide an explanation for the observed evolution of expectations.

In this paper, we relax the assumptions of a fixed economic environment and perfect knowledge and develop a model with learning and apply it to the real-time pricing of bonds. As is typical in the literature, we assume agents form forecasts of short-term interest rates using a small scale vector autoregressive (VAR) model. We assume agents recursively reestimate this model each period using the available data and use this model to form expectations of future interest rates, as well as to estimate the degree of macroeconomic volatility. To take account of the concern for structural change, we allow for discounting of past data in the estimation of the forecasting model. We evaluate the performance of this model against a benchmark model where the VAR parameters are estimated once and for all using the full sample of data. In this sense, our approach is related to that of Kozicki and Tinsley (2001) and Dewachter and Lyrio (forthcoming), who incorporated time variation in model intercepts as a source for explaining term structure dynamics. However, our approach differs from theirs in that ours allows for updating beliefs of all model parameters based on simple regressions, rather than imposing a specific structure on the estimation of a subset of parameters.

We evaluate our model on four criteria, three of which are formal and the fourth less so. Among the three formal criteria, we first analyze the magnitude and patterns of deviations from the expectations hypothesis assuming constant term premia. Existing term struc-

²Expectations of future interest rates can differ substantially from those implied by rational expectations, as shown in Orphanides and Williams (2005, 2006) and Beechey (2007).

³See also Kim and Orphanides (2005).

ture models generally imply sizable movements in term premia over time that are difficult to justify on theoretical grounds based on standard models of preferences and technology. Therefore, all else equal, we view a reduction in the magnitude of fluctuations in term premia as a desirable characteristic of a macro model of the term structure. In this regard our approach differs from strictly finance oriented criteria which are more interested in in-sample fit and the identification of associated arbitrage opportunities. At the same time, we recognize that nominal Treasury securities carry some risk and therefore term premia are unlikely to be strictly constant at all time. We therefore adopt a second criterion, the in-sample fit of the estimated model, allowing for time-varying term premia that depend on macroeconomic conditions under the restriction of no arbitrage. Thirdly, we compare the forecasting performance of the learning model with other existing models of the term structure. As shown by Mönch (2005), many term structure models that exhibit excellent fit in terms of current-period forecasts perform unsatisfactorily in forecasting yields several months to a year ahead. From the perspective of macro and monetary economists, forecasting ability is an important test for models and we therefore include it as our third criterion.

As compelling as these formal criteria might be, to us a less formal criterion is equally interesting, and that is consistency with monetary history. It is well known that there have been changes in the conduct of monetary policy. And monetary authorities are, after all, the drivers of the short-term interest rates that underly the expectations hypothesis. It stands to reason, therefore, that a complete description of fluctuations in bond yields needs to embrace the monetary facts on the ground, including the history of monetary policy. Accordingly, we will devote a bit of space later on to how we think our results accord with received wisdom in this area.

Turning to our results, we find that the learning model with discounting does a much better job at explaining longer-term yields than an equivalent model with constant coefficients estimated over the full sample; in particular, the deviations from the expectations hypothesis are much smaller on average with our learning model. The learning model is able to track the large upward movement in yields in the 1970s and the early 1980s and the downward movement in yields during the subsequent twenty years. Interestingly, the model estimated with fixed coefficients does especially poorly in the middle and later periods. In addition, our learning model's implied path of long-run inflation expectations mimics in

important regards the gradual decline in these expectations observed in the past quarter century; the constant-coefficient model in contrast implies that these expectations should have been roughly constant.

The learning model has important implications for the evolution of term premia. Unlike many other term structure models that imply a sizable downward trend in term premia over the past 25 years, our model's predictions for risk-neutral rates accord reasonably well with the broad contours of movements in actual yields, implying a much reduced underlying trend in term premia. That said, we find a non-trivial role for time-varying risk premia in determining yields. Following Ang and Piazzesi (2003), we model the market prices of risk as linear time-invariant functions of the state of the economy as captured by the three macro variables in the forecasting VAR at each point in time. We then estimate the model term premia imposing an affine arbitrage-free structure. The learning model naturally generates time variation in the real-time estimates of macroeconomic volatility that is absent in standard macroeconomic models with fixed coefficients. We find that these shifts in estimates of macroeconomic volatility help explain movements in term premia over the past half century.

The remainder of the paper is structured as follows. Section 2 discusses the VAR model used for forecasting and the term-structure model. Section 3 presents the real-time estimates of the VAR and resulting risk-neutral yields. Section 4 reports the estimation results for the arbitrage-free model of the term premium and analyzes the implications of learning for the evolution of the term premia. Section 5 examines the forecast performance of the learning model and compares it to that of other term structure models. Section 6 examines the robustness of our main findings to different calibrations of the learning process. Section 7 concludes.

2 Methodology

In this section we develop the model that we will use to explain and interpret the evolution of the U.S. term structure of interest rates over the past five decades. The model consists of two parts: a VAR model of the macroeconomy that agents reestimate each period and then use to form expectations of future short-term interest rates, and an affine term structure model that uses at each point in time the agents' current estimate of this VAR to price

yields at any maturity in an arbitrage-free manner.

Throughout what follows, we assume that agents are working with an anticipated utility model (Kreps 1997), which means that agents recurrently maximize the expected net present value of a future stream of returns with respect to a model that is recursively re-estimated. At each period, they allow for further time variation in their perceived law of motion for the economy, but they behave as though any such variation is a matter of sampling error. Anticipated utility is a commonly invoked simplification to a more elaborate (and infeasible) model wherein agents optimize their choices conditional on a full slate of possible future models. The assumption of anticipated utility can either be justified as a reasonable characterization of the kind of bounded rationality that the bond pricing agents in our model face, or as an approximation of the more complicated problem. Sargent and Williams (2005) show that, at least in some contexts, anticipated utility is indeed an excellent approximation.

The remainder of this section discusses the two parts of our model and the criteria on which we focus for evaluating competing models.

2.1 A VAR Model of Learning about the Macroeconomy

We approach the problem of model specification and prediction from the perspective of someone who prices financial assets in a quasi-real-time environment.⁴ One implication of our emphasis on the real-time aspect is that we use ideas and techniques available to market participants at the time. Hence we focus on a simple, parsimonious VAR model of inflation, real activity, and the short-term interest rate as the basic model for approximating agents' expectations formation for future short-term interest rates. Although the techniques by which the Federal Reserve implemented monetary policy underwent substantial changes over the past five decades, the basic fact that short-term interest rates changed systematically with inflation and real activity was widely understood by the 1960s.

A second implication of our emphasis on modeling real-time expectations formation is that agents are assumed to reestimate the VAR coefficients every period, following in the literature on learning, including Marcet and Sargent (1989) and Evans and Honkapohja (2001). Consistent with the extant literature, we assume that agents protect themselves

⁴We say "quasi-real-time environment" because we do not have a complete set of vintage data that would allow us to take proper account of data revisions. That said, as we shall see, only one of the series we use is subject to revision for anything other than seasonal adjustment.

from the possibility of unobserved shifts or drifting in the data generation process by discounting past data, as described below.

Let the vector of endogenous variables be $x_t \equiv [\pi_t \ q_t \ r_t]'$ where π is a measure of inflation, q a measure of real activity, and r is the one-period risk-free interest rate; the data used in this study will be described in the following section. Then at each date $t = 1, \dots, T$ the vintage- t VAR with p lags can be written in companion form as

$$X_\tau = \mu_t + \Phi_t X_{\tau-1} + \bar{u}_\tau, \quad \tau = 1, \dots, t \quad (1)$$

where $X_t = [x'_t \dots x'_{t+1-p}]'$, the time subscripts on μ and Φ signify that we will be allowing time variation in the VAR coefficients, and where

$$\bar{u}_\tau = \begin{bmatrix} u_\tau \\ 0_{3(p-1) \times 1} \end{bmatrix}. \quad (2)$$

Let

$$u^t = \begin{bmatrix} u'_1 \\ \vdots \\ u'_t \end{bmatrix} \quad (3)$$

denote the matrix of residuals up to date t that are associated with the parameter estimates μ_t , Φ_t , and let $\Sigma_t \equiv u^t u^t / t$.

We estimate the VAR recursively, allowing for limited memory; this is sometimes known as discounted recursive least squares (DRLS). Letting $\Gamma_t \equiv [\mu_t \ \Phi_t]$ and $\bar{X}_t = [1 \ X'_t]'$, the recursive updating formula for the VAR parameters is

$$\Gamma_t = \Gamma_{t-1} + P_{t-1} \bar{X}_t' (\bar{X}_t P_{t-1} \bar{X}_t' + \rho)^{-1} (\bar{X}_t - \Gamma_{t-1} \bar{X}_{t-1}) \quad (4)$$

and the precision matrix P_t is updated recursively according to the formula

$$P_t = P_{t-1} [I - \bar{X}_t (\bar{X}_t P_{t-1} \bar{X}_t' + \rho)^{-1} \bar{X}_t' P_{t-1}] \rho^{-1} \quad (5)$$

Equation (4) differs from standard recursive least squares, as in Harvey (1989) for example, solely by the addition of the term ρ which allows for an exponential weighting scheme such that the influence of old data is discarded at rate $\rho \leq 1$. For this reason, ρ is sometimes called the “forgetting factor”. A forecaster would downweight older data to the extent that he or she suspected that undetectable breaks in the data generating process occur with some

regularity. Besides providing what, in our view, is an intuitive mechanism for the evolution of agents' forecast models, DRLS also has the property that it converges to constant-gain learning, a form of learning that has been studied extensively in the literature (see, e.g., Evans and Honkapohja (2001)). Specifically, the gain term $\frac{1-\rho}{1-\rho^t}$ converges to $1 - \rho$ as the sample size t increases. The idea behind this modeling decision is that while the academic may concern himself or herself with testing the hypothesis that the structural coefficients of a VAR are constant, the bond trader is more likely to constantly update his or her estimates and allow whatever time variation the job of prediction requires.

2.2 The affine term structure model

Our model of the term structure of nominal Treasury zero-coupon yields is of the “essentially affine” form (Duffee, 2002), in which both yields and market prices of risk are linear functions of the state, X_t . The novel element in our analysis is that although the parameters governing the market price of risk, which we take to be a reduced-form for underlying preferences, are assumed to be time-invariant, the relationship between yields and the state vector varies over time in consequence of the time-varying forecasting model used by agents.⁵

Let P_{nt} denote the price of a zero-coupon bond with n periods to maturity, and let $y_t^n = -\log P_{nt}$ denote the continuously compounded n -period zero-coupon yield. The nominal holding period return between periods t and $t + 1$ on this bond is given by $1 + R_{n,t+1} = P_{n-1,t+1}/P_{nt}$. Absence of arbitrage implies the existence of a stochastic discount factor M_{t+1} with the property that $1 = E_t[(1 + R_{i,t+1})M_{t+1}]$ for any asset i with (nominal) gross return $1 + R_{i,t+1}$. Because we are interested in pricing nominal zero coupon bonds, we focus throughout on nominal stochastic discount factors. In the case of zero-coupon bonds, this implies

$$P_{nt} = E_t[P_{n-1,t+1}M_{t+1}] \quad (6)$$

Equation (6), together with a process for the stochastic discount factor M_{t+1} and the identity $P_{0t} = 1$, allows us to price bonds for any maturity n .

If investors were risk-neutral, the stochastic discount factor would reduce to $M_{t+1} = \exp(-r_t)$, and the expectations hypothesis of the term structure would hold. By contrast, with risk sensitive investors, the stochastic discount factor is given by $M_{t+1} = \exp(-r_t)\xi_{t+1}$,

⁵The derivation of the term structure model follows the exposition in Campbell *et al.* (1997, Ch. 11) and Ang and Piazzesi (2003).

where ξ_t denotes the equivalent martingale measure (or Radon-Nikodym derivative) converting the risk-neutral to the physical (or data-generating) measure.⁶ We follow Ang and Piazzesi (2003) and assume that ξ_{t+1} follows the process

$$\xi_{t+1} = \xi_t \exp \left(-\frac{1}{2} \lambda_t' \lambda_t - \lambda_t' \bar{\epsilon}_{t+1} \right) \quad (7)$$

where λ_t is the vector of time-varying market prices of risk associated with the sources of uncertainty represented by the structural VAR innovations $\bar{\epsilon}_{t+1}$. These innovations are derived from the reduced-form VAR residuals \bar{u}_{t+1} defined in (2) through a standard recursive identification scheme

$$\bar{u}_\tau = \Sigma_t^{.5} \bar{\epsilon}_\tau, \quad \tau = 1, \dots, t \quad (8)$$

where $\Sigma_t^{.5}$ is a $3p \times 3p$ matrix with the Cholesky factor of Σ_t in its upper left corner and 0 elsewhere.

Following Duffee (2002), the prices of risk are assumed to follow the affine process

$$\lambda_t = \lambda_0 + \lambda_1 X_t \quad (9)$$

where the vector λ_0 and the matrix λ_1 are commensurate with X_t . Under the assumption that the VAR innovations $\bar{\epsilon}_{t+1}$ are normally distributed, both ξ_{t+1} and the pricing kernel

$$\begin{aligned} M_{t+1} &= \exp(-r_t) \xi_{t+1} / \xi_t \\ &= \exp \left(-r_t - \frac{1}{2} \lambda_t' \lambda_t - \lambda_t' \bar{\epsilon}_{t+1} \right), \end{aligned} \quad (10)$$

are log-normal too.⁷ Letting $f(\bar{\epsilon}_{t+1})$ denote the density of the structural VAR innovations, the Radon-Nikodym derivative $\xi_{t+1}(\lambda_t)$ has the effect of shifting the mean of

⁶The existence of the equivalent martingale measure in the absence of arbitrage, and the change of measures under the Girsanov theorem, are discussed in Duffee (2001, Ch. 6).

⁷Duffee (2002) shows that the “essentially affine” form (9) for the market price of risk has several advantages over “completely affine” models in which market prices of risk are proportional to the variance of the underlying sources of risk. In particular, he shows that essentially affine models perform much better in out-of-sample forecasts than completely-affine models. A drawback of the essentially affine model is that it has proved challenging to derive a stochastic discount factor of the form (10) from first principles. For example, under the assumption that consumption growth Δc_t is distributed iid $\sim N(0, \sigma^2)$, CRRA utility $u(C_t) = C_t^{1-\gamma}/(1-\gamma)$ implies a log real discount factor $m_{t+1} = -r_t - \frac{(\gamma\sigma)^2}{2} - \gamma\Delta c_{t+1}$, and hence the market price of risk γ is constant. The same is true in a model with a multiplicative exogenous consumption habit (Bekaert et al., 2006). Ravenna and Seppälä show that time-varying risk premia can be derived in a new-Keynesian model by using third-order approximations to the model’s structural equations instead of log-linear approximations, but this discount factor is not of the form (10).

the distribution of $\bar{\epsilon}_{t+1}$, so that $\xi_{t+1}(\lambda_t)f(\bar{\epsilon}_{t+1}) = f(\tilde{\epsilon}_{t+1})$ where $\tilde{\epsilon}_{t+1} = \bar{\epsilon}_{t+1} + \lambda_t$. The law of motion (1) of the VAR can then be rewritten as

$$\begin{aligned} X_\tau &= (\mu_t - \Sigma_t^5 \lambda_0) + (\Phi_t - \Sigma_t^5 \lambda_1) X_{\tau-1} + \Sigma_t^5 \tilde{\epsilon}_\tau \\ &= \tilde{\mu}_t + \tilde{\Phi}_t X_{\tau-1} + \tilde{\epsilon}_\tau \end{aligned} \quad (11)$$

with the risk-adjusted parameters given by $\tilde{\mu}_t = \mu_t - \Sigma_t^5 \lambda_0$ and $\tilde{\Phi}_t = \Phi_t - \Sigma_t^5 \lambda_1$.

Because the short rate, $r_t = y_t^1$, is part of the VAR, the short-rate process

$$r_t = \delta_0 + \delta_1 X_t \quad (12)$$

is given by $\delta_0 = 0$ and δ_1 selecting the element of X_t corresponding to r_t . Substituting from (10) into (6) and performing some calculations described in Appendix A, we obtain

$$P_{nt} = \exp(\bar{A}_{n,t} + \bar{B}'_{n,t} X_t) \quad (13)$$

where the sequences of scalars $\{\bar{A}_{n,t}\}_{t=1,\dots,T}$ and vectors $\{\bar{B}_{n,t}\}_{t=1,\dots,T}$ follow the recursions

$$\begin{aligned} \bar{A}_{n+1,t} &= \bar{A}_{n,t} + \bar{B}'_{n,t}(\mu_t - \Sigma_t^5 \lambda_0) + \frac{1}{2} \bar{B}'_{n,t} \Sigma_t \bar{B}_{n,t} - \delta_0 \\ &= \bar{A}_{n,t} + \bar{B}'_{n,t} \tilde{\mu}_t + \frac{1}{2} \bar{B}'_{n,t} \Sigma_t \bar{B}_{n,t} - \delta_0 \end{aligned} \quad (14)$$

$$\begin{aligned} \bar{B}'_{n+1,t} &= \bar{B}'_{n,t}(\Phi_t - \Sigma_t^5 \lambda_1) - \delta'_1 \\ &= \bar{B}'_{n,t} \tilde{\Phi}_t - \delta'_1 \end{aligned} \quad (15)$$

with $\bar{A}_{1,t} = -\delta_0$, $\bar{B}_{1,t} = -\delta_1$. The continuously compounded zero-coupon yield $y_t^n = -\log P_{nt}$ is then given by

$$y_t^n = A_{n,t} + B'_{n,t} X_t \quad (16)$$

where $A_{n,t} = -\bar{A}_{n,t}/n$, $B_{n,t} = -\bar{B}_{n,t}/n$. The second lines in each of the recursions (14) and (15) imply that current bond yields are the average of expected future short rates projected under the risk-adjusted law of motion of X_t (11) plus a Jensen's inequality term arising from the fact that we are describing expectations of log returns.

In comparison to the existing literature on term structure models with macroeconomic variables among the factors, the key innovation is the effect of variation in the vector μ_t and the matrices Φ_t and Σ_t on the relationship between states and yields. One point to note is the effect of time variation in Σ_t : this not only interacts with the coefficients of the market prices of risk through the terms $\Sigma_t^5 \lambda_0$ in (14) and $\Sigma_t^5 \lambda_1$ in (15), but also affects

yields in the case of risk-neutrality (i.e. when $\lambda_t = 0$) through the Jensen's inequality term $\frac{1}{2}\bar{B}'_{n,t}\Sigma_t\bar{B}_{n,t}$ in (14). In section 4 we will show that time variation in Σ_t plays an important role for our results.

2.3 Maximum likelihood estimation and model evaluation

Let Y_t denote the m -dimensional vector of yields to which the parameters of the term structure model are fitted (as explained below, in our case $Y_t = [y_t^3, y_t^{12}, y_t^{36}, y_t^{60}]'$). To avoid stochastic singularity, we assume that the elements of Y_t are observed with i.i.d. Gaussian measurement error u_t , i.e.

$$Y_t = A_t + B_t X_t + u_t \quad (17)$$

where the vectors A_t and matrices B_t select the $A_{n,t}$ and $B_{n,t}$ of corresponding maturity. Let Δ denote the diagonal covariance matrix of the measurement errors. The log likelihood of the vector of observed yields is then given by

$$\mathcal{L} = -\frac{Tm}{2}\log(2\pi) - \frac{T}{2}\log|\Delta| - \frac{1}{2}\sum_{t=1}^T u_t' \Delta^{-1} u_t \quad (18)$$

The log likelihood is being maximized with respect to the 3 non-zero elements of λ_0 , the upper 3×3 block of λ_1 , and the four diagonal elements of Δ .

While, from an econometric point of view, the log likelihood of the estimated term structure model is the natural criterion for evaluating competing models, it suffers from the drawback that much of the fit is being achieved through time-variation in term premia which are by definition unobservable. Such term premia are constrained by the assumption of arbitrage-free pricing, but they nonetheless constitute important degrees of freedom, relative to the standard expectations hypothesis (EH) of the term structure which holds that such term premia should be constant. As we discussed in the introduction, we therefore complement the evaluation of our model and various competitor models, by two additional criteria. First, we ask that departures from the expectations hypotheses of the term structure be as small as possible. We implement this criterion by measuring the standard deviation of the difference between the “synthetic” n -period yield generated from a given VAR – computed as the average of projected one-period yields over the next n periods – and the actual n -period yield. In other words, we would judge one model to be superior to another (on average, over certain maturities) if it minimizes the standard deviation of

deviations from the EH, all else equal. This criterion is closely related to requiring that variation in the term premia as measured by the estimated *term structure* model should be as small as possible. While we not go so far as to argue that term premia must be constant, our criterion reflects our prior view that a model which has to rely predominantly on time-variation in term premia to explain the historical evolution of yields is of lesser value than one that does not.

Our third, formal criterion is based on out-of-sample forecast performance. As we just argued, time-varying term premia provide additional degrees of freedom in fitting the data, and could thus potentially lead to overfitting. As usual, overfitting should manifest itself in a deterioration in forecast performance. In section 5 we therefore examine the forecasting performance of our model relative to several competitor models.⁸

3 Estimation and Results for the VAR

In this section we first describe the data used in this study. We then present estimation results for our VAR model of the macroeconomy.

3.1 The data

We use the Fama-Bliss dataset of 1- and 3-month “risk-free” Treasury bill rates and nominal zero-coupon Treasury yields at maturities of 12, 36, and 60 months to estimate the parameters of our term structure model. The dataset consists of observations based on prices recorded on the last trading day of the month, beginning in June 1952 and ending in December 2006. Although the macroeconomic literature tends to use the federal funds rate as the instrument of monetary policy, for our purposes, a T-bill rate is preferable because, like yields at longer maturities but unlike the federal funds rate, the T-bill rate is not subject to default risk.

⁸There are additional criteria that have been used in the literature for evaluating the plausibility of alternative models. For example, Dai and Singleton (2002) and Hördahl et al. (2006) ask whether interest rate data simulated from the term structure model produce the same failures of the expectations hypothesis as documented in Campbell and Shiller (1991) and Campbell et al. (1997). Another possibility would be to consider the predictive power of the term structure for GDP growth examined, e.g., in Ang et al. (2006). In our model, it would be particularly interesting to ask how this predictive power has changed over time, in light of recent evidence that it has declined since the early 1980s. We leave these alternative tests for future work.

Commensurate with the yield data, we use a 3-variable, monthly VAR model to represent the macroeconomy. As discussed before, our objective is to approximate with an "on-line algorithm" what private-sector economists and bond traders might have done, in real time, to price bonds. In addition to the 1-month T-bill yield, we need measures of inflation and real activity. With respect to the former, a key issue is which measure market participants might have thought the Federal Reserve would have been most likely to focus on. This consideration suggests a focus on "core" inflation; that is, all-items inflation excluding the volatile food and energy components. We therefore elect to work with core CPI inflation⁹ Core CPI data are available back to January 1957; for the earlier years of our sample we splice all-items CPI to the core data. For our measure of real activity, we use payroll employment growth.¹⁰

Operationally, our price and employment data need to be made stationary, and so conversion into rates of change and time averaging needs to be considered. Monthly growth rates have the advantage of focusing on the "news" in the data, instead of contaminating it with old information, such as in the 12-month growth rates used, e.g., by Ang and Piazzesi (2003), Hördahl, Tristani, and Vestin (2006) and Rudebusch and Wu (2003). On the other hand, one-month growth rates are very volatile and therefore create high-frequency noise that agents may have smoothed. A particular problem arises due to our use of core CPI as inflation measure. Because the BLS reports the levels of CPI series only to precision of one decimal, in the early years of our sample (when the level of the index was well below the base level of 100 in the years 1982-84) the one-month growth rates suffer from "granularity." We therefore choose 3-month growth rates for the core CPI and 1-month growth rates for

⁹Although inflation based on the deflator of personal consumption expenditures excluding food and energy ("core PCE") has received more attention since the mid-1990s, this measure has the drawback that it is subject to greater changes in definition and other revisions over time than CPI-based measures. That said, the Bureau of Labor Statistics (BLS) began publishing core CPI data only in 1978, we nonetheless use it throughout our sample based on the view that market participants would have been able to make their own adjustments for the energy price movements in the early 1970s.

¹⁰Possible alternatives would have been to use some measure of the output gap or the capacity utilization rate. The output gap, however, depends on arbitrary detrending procedures, many of which are "two-sided" in nature and not applicable for real-time work. The capacity utilization rate is subject to substantial revisions and covers a small and decreasing share of the real economy. A more obvious alternative is the unemployment rate, which has the advantage that it is not subject to revision. However, empirical work by Gürkaynak et al. (2005) shows that financial markets respond much more strongly to surprises in payroll employment than they do to surprises in the unemployment rate.

payroll employment.

3.2 Time-varying VAR estimates

The fact that our VAR is re-estimated every period and that it is used for forecasting short-term yields far into the future obliges us to keep it as parsimonious as possible so as to reduce the (in any case substantial) number of parameters to be estimated. For this reason, we restrict ourselves to three variables, one representing real variables, one nominal, and one for short-term interest rates. Similarly, our prior is for shorter lags over longer; with these considerations in mind, based on likelihood ratio tests, we selected four lags. In estimating the VAR, we use the seven-year sample from January 1953 to December 1958 to obtain initial estimates of the parameters μ and Φ of the VAR. For a given choice of ρ we obtain these estimates by weighted least squares. We then use equations (4) and (5) to update the VAR parameters based on each month of additional data, period by period, beginning in January 1959 and ending in December 2006.

One feature of this simple algorithm is that in several instances, especially towards the end of the 1960s and the end of the 1970s, the largest eigenvalue of the matrix, Φ , exceeds 1. Because we use our VAR to form short-rate projections up to 60 months into the future, explosive eigenvalues lead to unrealistically extreme short-rate projections. Thus, consistent with the extant literature on learning, we employ what is known as a projection facility: that is, we assume that if, at any date, the largest eigenvalue of the estimated Φ exceeds our limit of 0.999, the previous period's VAR parameter estimates are retained; then, in the subsequent periods, the parameters are re-estimated using weighted least-squares from the outset at January 1953, and again the same eigenvalue check is performed.¹¹

We choose ρ to minimize the one-step-ahead forecast errors of the VAR. This is perhaps the most natural criterion for model selection for VARs used for prediction. Results based on this criterion are shown in the upper panel of Table 1, under the label “1-step-ahead forecast RMSE.” The first row of the upper panel in Table 1 shows as benchmark the RMSE obtained from estimating a VAR over the full sample. As shown in the second row, these RMSEs are only marginally lower when we recursively estimate our VAR, i.e. when we

¹¹The re-estimation with weighted least squares is equivalent to starting over with a potentially very long training sample. It ensures that the triggering of the projection facility does not contaminate, in some sense, subsequent periods' parameter estimates.

Table 1: Calibration of ρ

Estimation Method	1-step-ahead Forecast RMSE			
	π	Δn	y^1	Wt. ave.
Full-sample (constant VAR)	0.75	2.10	0.61	1.03
Learning ($\rho = 1$)	0.73	2.07	0.58	1.00
Learning ($\rho = 0.978$)	0.62	1.73	0.54	0.88
	SD of Deviation from EH			
	y^3	y^{12}	y^{36}	y^{60}
Full-sample (constant VAR)	0.40	0.70	1.16	1.43
Learning ($\rho = 1$)	0.40	0.67	1.03	1.21
Learning ($\rho = 0.978$)	0.46	0.71	0.91	0.91

allow for parameter variation over time, but do not discount past data ($\rho = 1$). In this case, our estimation is equivalent to repeated OLS estimation with an expanding sample, except for the rejection of estimates that imply explosive VAR dynamics. However, this occurs on only two occasions (0.3 percent of the sample of 576 estimates). The third row shows the RMSEs for the value of ρ that minimizes the weighted average of the variances of the three forecast errors, with the weights equal to the inverse of the variance implied by the learning model with $\rho = 1$.¹² These RMSEs are up to 20 percent below the ones shown in row 1 for the constant-coefficient VAR. The value of $\rho = .978$ is fairly low, although not that much lower than the value of 0.9875 used by Sargent (1999). Our value implies that observations that are three years in the past receive about half as much weight as the most recent one, and observations that are 10 years old receive only a weight of 7 percent. Not surprisingly, the lower the value of ρ , the higher the incidence of triggers of the projection facility. However, over the range from 1 to 0.978 studied here, the incidence never rises above five percent.

As a check on our results, we also calibrated ρ to minimize the standard deviation of the differences between synthetic yields for 3, 12, 36, and 60 month bonds and the corresponding observed yields, as discussed in the previous section. The results for ρ based on this second criterion were very close to the base-case results.¹³ In the bottom panel of the table, we

¹²We weigh the variances by the inverse of the respective full-sample variances, as otherwise the very volatile 1-month employment growth rates would receive most of the weight in the minimization.

¹³The optimal ρ chosen using the alternative criterion was 0.975.

retain the value of $\rho = 0.978$ while computing results for our alternative criterion, the “standard deviation of deviations from the expectations hypothesis.” Like in the top panel, allowing for learning over time in the simplest possible way – expanding sample or recursive least squares – improves the fit of the model, particularly at longer maturities. Also like in the top panel, the improvement is more pronounced when discounting is permitted.¹⁴

The 1970s and early 1980s was a tumultuous period for the U.S. economy and interest rates in particular, and the macroeconomic seas have calmed considerably since then. We therefore evaluate the performance of the learning model in two subsamples, one running from 1959 through 1983, and the second starting in 1984, a commonly used date for the beginning of the “Great Moderation.” The objective is to see whether the results based on the full sample are specific to an earlier period of history and to see how the two models perform in perhaps a more representative period of the past 23 years. Table 2 reports summary statistics for the two sub-samples (and the full 1959-2006 sample) for the baseline learning model and the model using the constant-coefficient VAR estimated on the full sample. The upper panel of the table reports the root-mean-squared deviations of the model-generated risk neutral rates from the actual yields. To make these calculations consistent with those on the full sample, we assume constant term premia over 1959-2006.

Under the assumption of constant term premia, the learning model clearly dominates the constant-coefficients VAR model in both subsamples, especially for longer-term yields. Not surprisingly, the fit to the data of both models is appreciably better in the subsample starting in 1984. However, the better performance of the learning model in both subsamples explains its clear dominance on the combined full sample.

The learning model generates significant movements in long-run inflation expectations broadly consistent with the available data, but totally absent from the constant-coefficient VAR estimated on the full sample. Figure 1 shows the simulated projections, at each date, from the time-varying VAR model of average inflation five to ten years ahead, the solid line. It compares these predictions with those from the same VAR estimated assuming time-invariant coefficients, the dash-dotted line. To assess the adequacy of these forecasts, the dashed line shows the forecast of inflation 5-to-10 years ahead from a survey of financial

¹⁴Of course, there is no guarantee that either of these two criteria is the best recursive characterization of the updating model used by agents in real time. We shall show below that within the class of simple, recursive algorithms, it performs well. That said, we leave open to future research the possibility of forgetting factors that are changing with real-time forecasting performance.

Table 2: Results for Sub-samples

Sample	RMS Deviations from EH			
	y^3	y^{12}	y^{36}	y^{60}
Learning model				
1959-1983	0.5563	0.8037	0.9926	1.0073
1984-2006	0.3233	0.5837	0.8026	0.7928
1959-2006	0.4596	0.7069	0.9066	0.9108
Full-sample VAR				
1959-1983	0.4546	0.7730	1.2766	1.5550
1984-2006	0.3244	0.6200	1.0258	1.2756
1959-2006	0.3975	0.7039	1.1632	1.4280

Note: RMS of deviations from expectations hypothesis computed assuming constant term premia over 1959-2006.

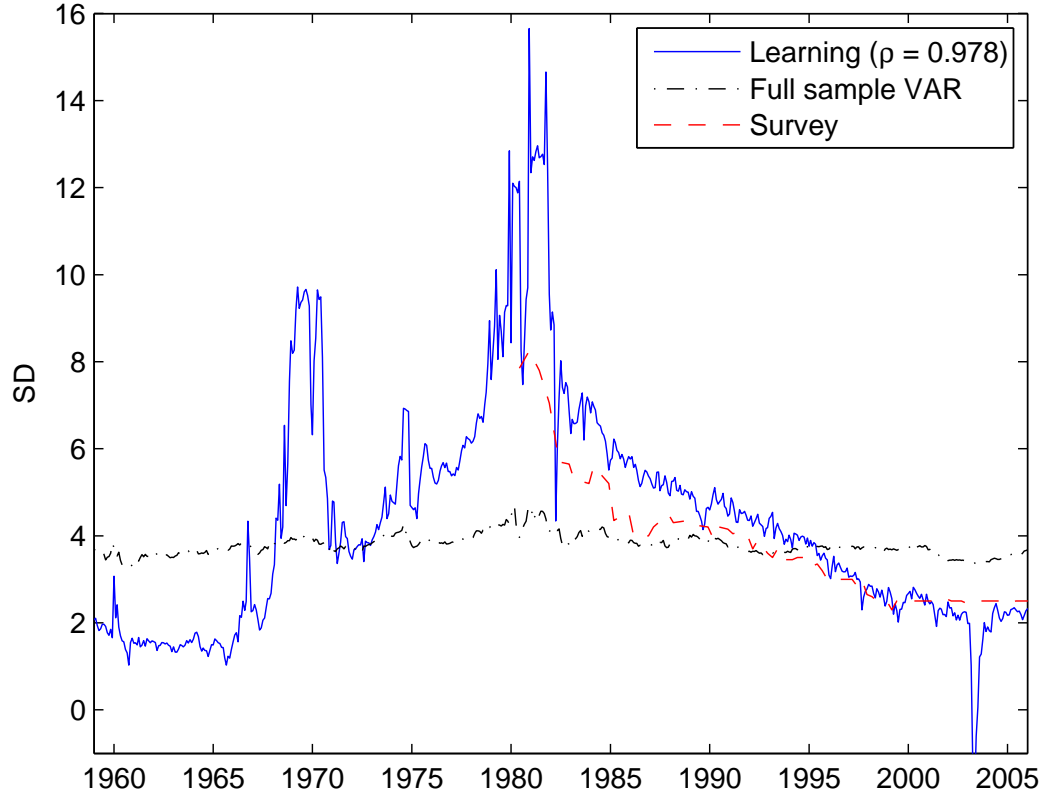
market participants and professional forecasters.¹⁵ Three notable facts can be gleaned from this figure. First, our learning model does a good job of mimicking the movements of the survey. Considering that fitting the survey was not a criterion for model selection, this is a noteworthy accomplishment for the model. Second, the cost of assuming that there has been no time variation in model coefficients is large: the constant coefficient model performs very poorly.¹⁶

Third, during certain periods the solid line shows a high degree of sensitivity of predicted inflation to innovations in the data, particularly in the late 1960s, the period from about 1979-1983, and a brief period early in the new milenium. The second of these periods is

¹⁵To be precise, the underlying data is quarterly and is interpolated to monthly frequency. The data from 1981Q2 to 1991Q1 are from the survey of financial market participants conducted by Richard Hoey of Drexel Burnham Lambert. The survey participants were asked for their expectation of CPI inflation for the period 5 to 10 years ahead. After 1991Q1 the data is from the Survey of of Professional Forecasters, conducted by the Federal Reserve Bank of Philadelphia, asking participants for their expectation of CPI inflation on average over the next 10 years.

¹⁶One reaction to this finding is that it merely reflects the benefits of a free parameter. While there is something to this view, it is worth noting that the learning model at each point of time has less data to work with than does the constant coefficient VAR which uses the full sample data. If the constant coefficient model is true, the benefit of the added data for sharpening estimates should be large.

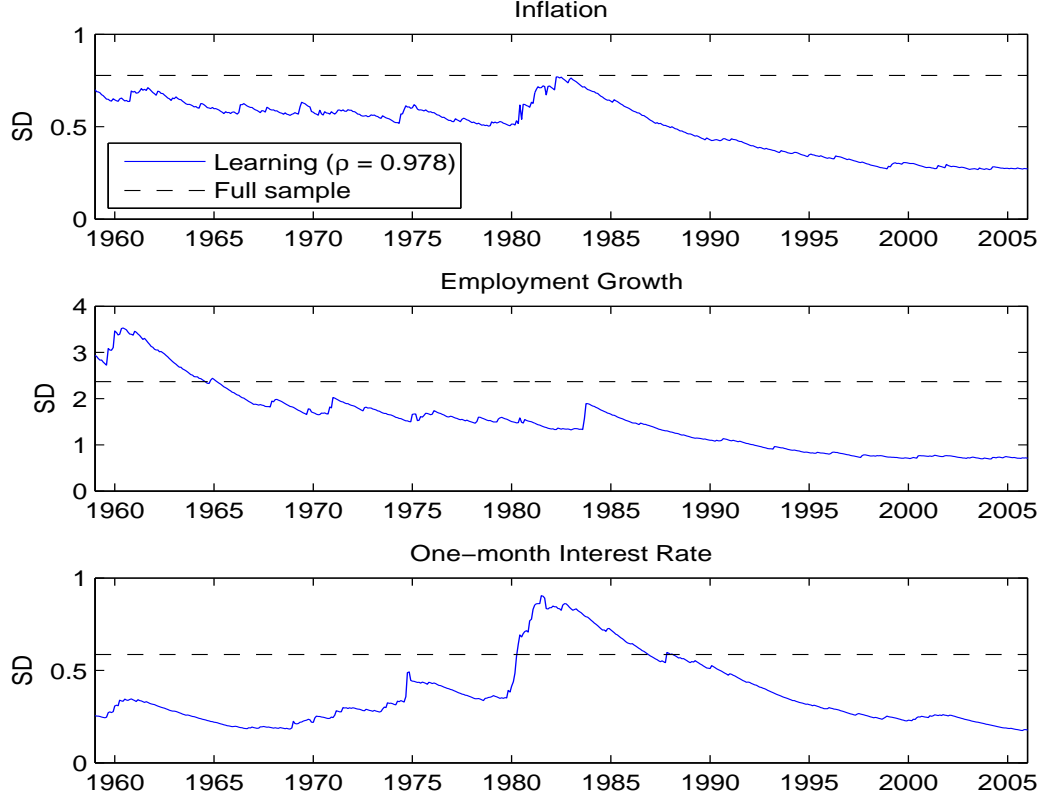
Figure 1: Inflation Expectations 5 to 10 Years Ahead



described by Goodfriend (2005) as one of “inflation scares,” periods where doubts about the Fed’s determination to control inflation resulted in outsized jumps in expected inflation (Goodfriend does not discuss the earlier period). The 2003 period is well known as one where inflation was surprisingly low and some market participants perceived the Fed as actively considering “unconventional policy tools” in case the zero lower bound on the Federal Funds rate should become binding. The concern about deflation, like the downward spike of projected inflation in 2003, would turn out to be short lived. All told, the model’s predictions of volatile long-term inflation expectations occur during periods that line up with well-known events.

The learning model implies significant time variation in the perceived degree of macroe-

Figure 2: Standard Deviations of VAR Residuals



economic volatility. The solid lines in Figure 2 show the standard deviations of the learning model's residuals as extracted from the diagonal of the time-varying covariance matrix, Σ_t . These are compared with their counterparts from the constant coefficient VAR, the dashed lines. The solid lines show sizable time variation, consistent with the findings of Stock and Watson (2007) among others. Moreover, the timing of declines in the standard deviations of the shocks lines up with the timing of the onset of the so-called Great Moderation, as in Kim, Nelson and Piger (2004), in that the standard deviation of the innovation to our real variable, employment growth, begins to decline markedly in about 1984, and the decline in the standard deviation of inflation innovations sets in in about 1983. Perhaps the most noteworthy finding, however, is the behavior of the T-bill rate residual. It shows the most

Table 3: Estimates of Term Structure Model Parameters

	Risk price parameters			
	λ_0	λ_1		
		π	Δn	r
λ_π	-0.6540 (0.0758)	0.1197 (0.0188)	0.2739 (0.0229)	-0.0501 (0.0182)
$\lambda_{\Delta n}$	0.4188 (0.0644)	-0.1658 (0.0181)	0.1152 (0.0176)	0.1129 (0.0193)
λ_r	-1.2110 (0.0442)	0.0653 (0.0068)	-0.0164 (0.0090)	0.0343 (0.0078)
Yield Measurement Error RMSE				
	y^3	y^{12}	y^{36}	y^{60}
	0.7036 (0.0178)	0.8357 (0.0541)	0.8547 (.1014)	0.8437 (0.0750)

Note: Standard errors in parentheses.

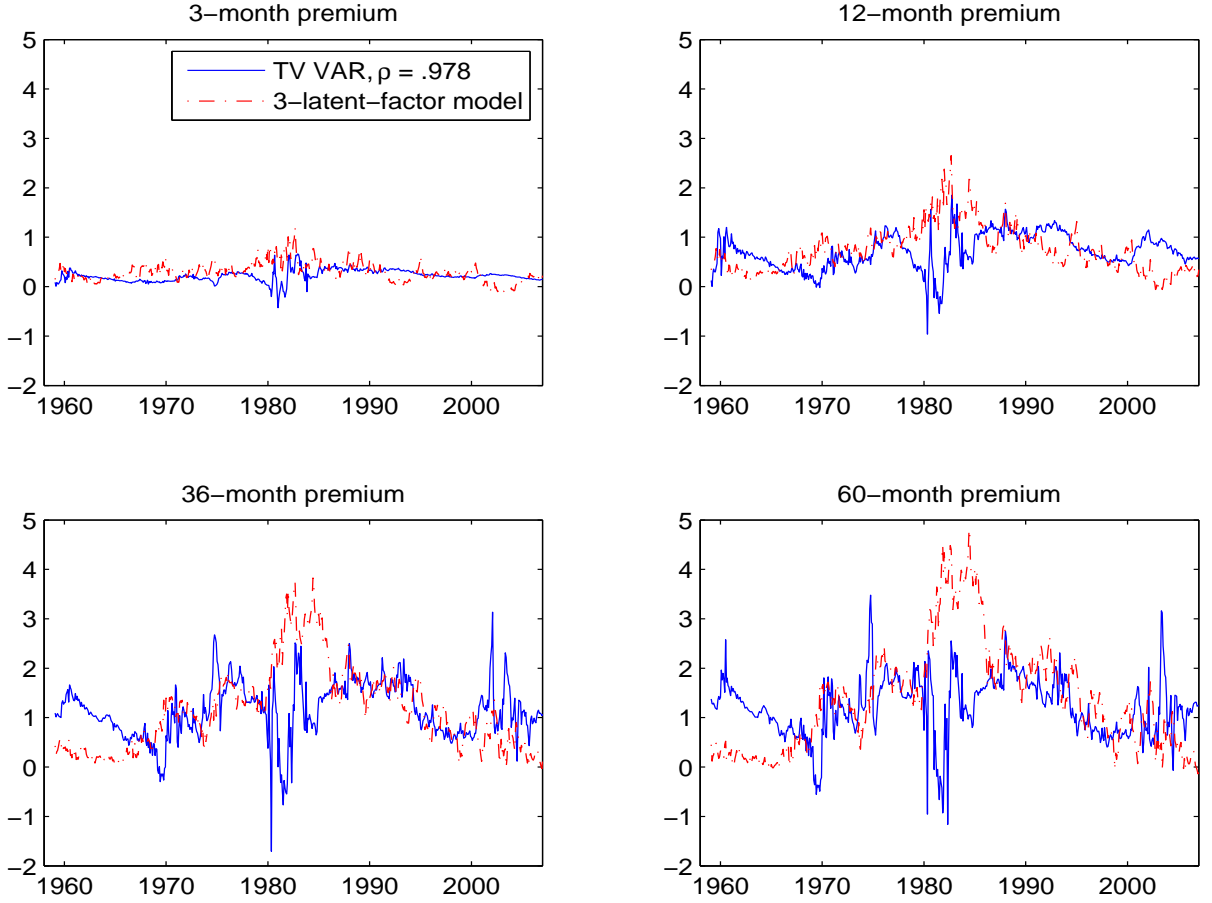
time variation of the three variables. It also shows its largest rise slightly before that of the inflation innovation. Overall the story told by these residuals is consistent with the view that highly variable perceived shocks to inflation and real variables had effects on the economy to which monetary policy did not respond sufficiently strongly. Then, beginning in 1979, the regime began to change – in fits and starts. The October 1979 change in operating rules, combined with an absence of relevant historical data from which to judge the new regime, resulted in confusion, as measured by the large perceived shocks to the T-bill rate. Over time, both policy and perceptions settled down and the measured standard deviations of all shocks diminished.

4 Estimation and Results for the Term Structure Model

In this section we present results from estimating the term structure model, taking the estimated VAR discussed in the previous section as input. We also discuss the stability of our main results across subsamples.

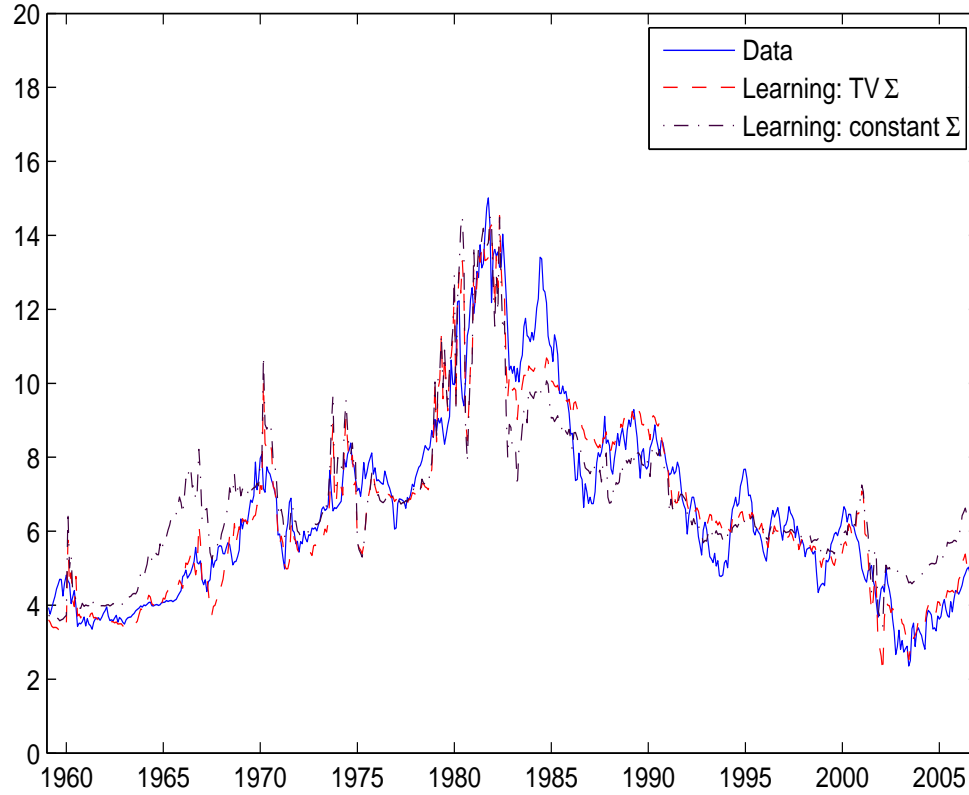
We now embed the estimated VAR model into the term structure model as described in section 2.2 and estimate the remaining parameters. Table 3 shows the maximum likelihood

Figure 3: Term Premia



estimates of the elements of λ_0 and λ_1 governing the evolution of the market price of risk as well as the square roots of the diagonal elements of the covariance matrix Δ defined in (18). Two points related to the RMSEs of the yields are worth noting. First, for the shorter-horizon yields (3 and 12 months) these RMSEs are larger than the ones reported in Table 1, based on the VAR model alone. Thus, adding time-varying term premia to the expectations component does not improve the fit of the model for these yields, reflecting the restrictions that the model imposes on the evolution of the term premia. The opposite is true for the long-horizon yields: here adding the term premia implied by the term structure model helps the fit. The second point is that these RMSEs are large relative to RMSEs based on latent factor models in the finance literature. For example, a model with three

Figure 4: 60-month Yield and Fitted Values with Time-Varying and Constant Σ



latent and no observable factors (such as the “yields only” model of Ang and Piazzesi (2003) estimated using our yields data set results in measurement error RMSEs of between 10 and 25 basis points, as compared to the 70-85 basis points range in our model. However, as pointed out in the introduction, our model is not designed to maximize in-sample fit, and compares favorably with latent-factor models in out-of-sample forecasts at longer horizons.

Figure 3 shows the term premia implied by our model at the four maturities of the yields included in estimation (the solid lines), and compares them to the term premia implied by the term structure model using three latent factors (the dashed lines). It is clear from the figure that during the episodes in which our VAR exhibits near-explosive behavior (1969-70 and 1980-81), the term premia are implausibly low and in fact turn negative. Apart

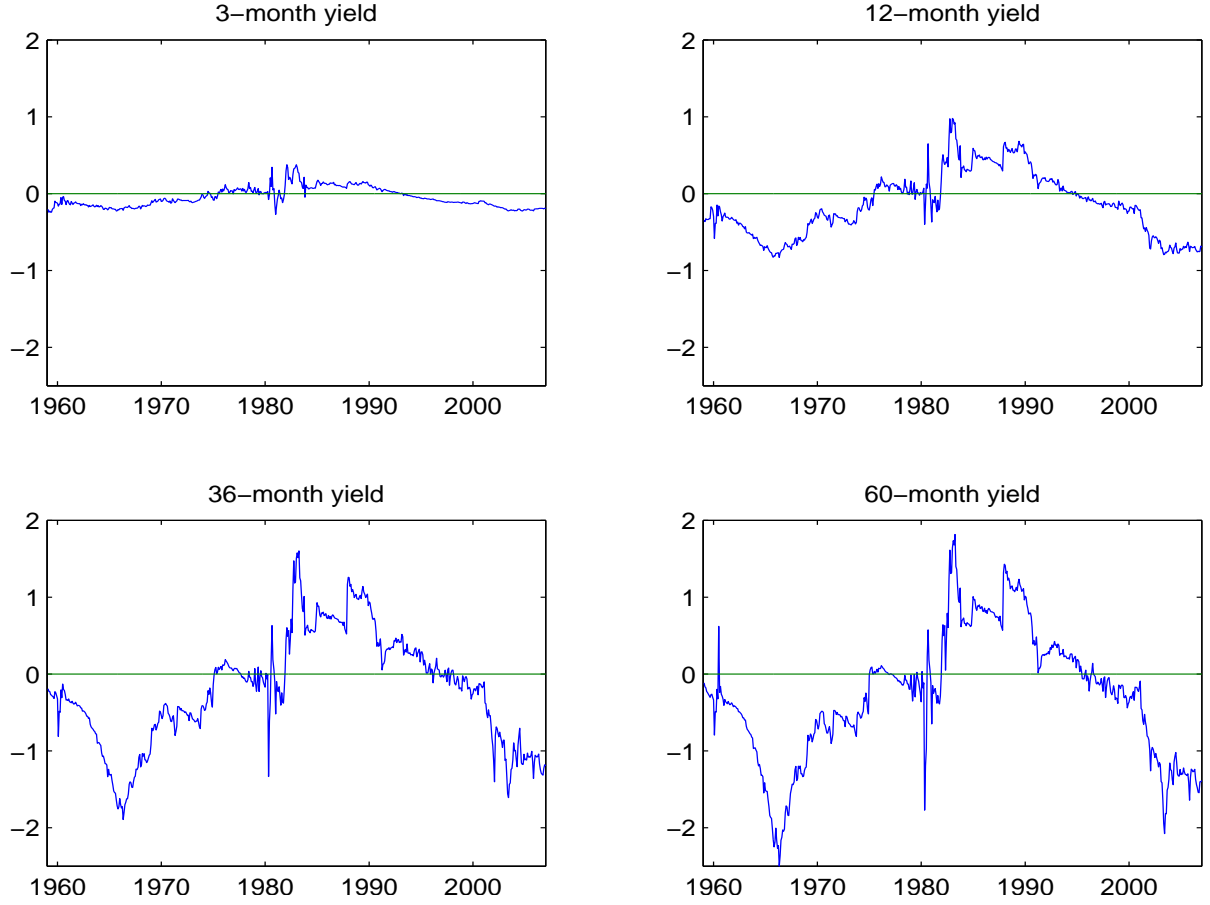
from these two windows, in some periods they tell an interestingly different story from the term premia implied by the latent factor model. In particular, during the period 1982 to 1986 the latent-factor model explains the movements of longer-horizon yields through term premia in the range of 4 percent or more. By contrast, in our model term premia, although higher than in the late 1960s or in the past 15 years, are not nearly as elevated as those in the latent-factor model. Note that this is a period during which the long-horizon inflation expectations of our VAR shown in Figure 1 track closely the survey expectations, providing independent evidence that our VAR provides a good approximation to agents' expectations formation. Evidently, the expectations component is able to explain a larger fraction of the variation in yields in our model than it is in the latent-factor model, suggesting that time variation in the law of motion of the factors may be important to the pricing of bonds. We view the fact that our model requires less time variation in term premia as an appealing feature.

Figure 4 illustrates the contribution made by the variation over time in the estimate of the covariance matrix Σ_t of the VAR residuals. The solid line shows the 60-month yield, and the dashed line the fitted values from our model. The dashed-dotted line is computed by keeping Σ_t constant in the recursions (14) and (15) at its average taken over the 576 estimates of Σ_t . As discussed in section 2, although the market prices of risk λ_t themselves are not functions of the second moments of the data, the pricing of bonds depends on those moments both through interaction terms with λ_0 and λ_1 and through Jensen's inequality terms that operate independent of agents' risk attitudes. As Figure 4 makes clear, the fact that our model allows for time variation in the perceived volatility of the economy helps explain yields, especially during the 1960s, the mid-1980s, and the period since 2002.

The contribution made by the time-varying matrices Σ_t (i.e. the difference between the dashed and dashed-dotted line in Figure 4) is shown, for all four yields, in Figure 5. This figure tells a story broadly similar to the lower panel of figure 2, which shows the standard deviation of the VAR residual to the interest rate equation. The contribution (relative to keeping Σ fixed at its sample average) falls through about 1967, then climbs, rises sharply in late 1981, remains elevated until about 1991, and then declines substantially. Roughly 300 basis points of the decline in 5-year yields since the 1980s is attributed to the decline in the perceived volatility of the economy.

Our model of perceived time variation in the structure of the economy is of particular

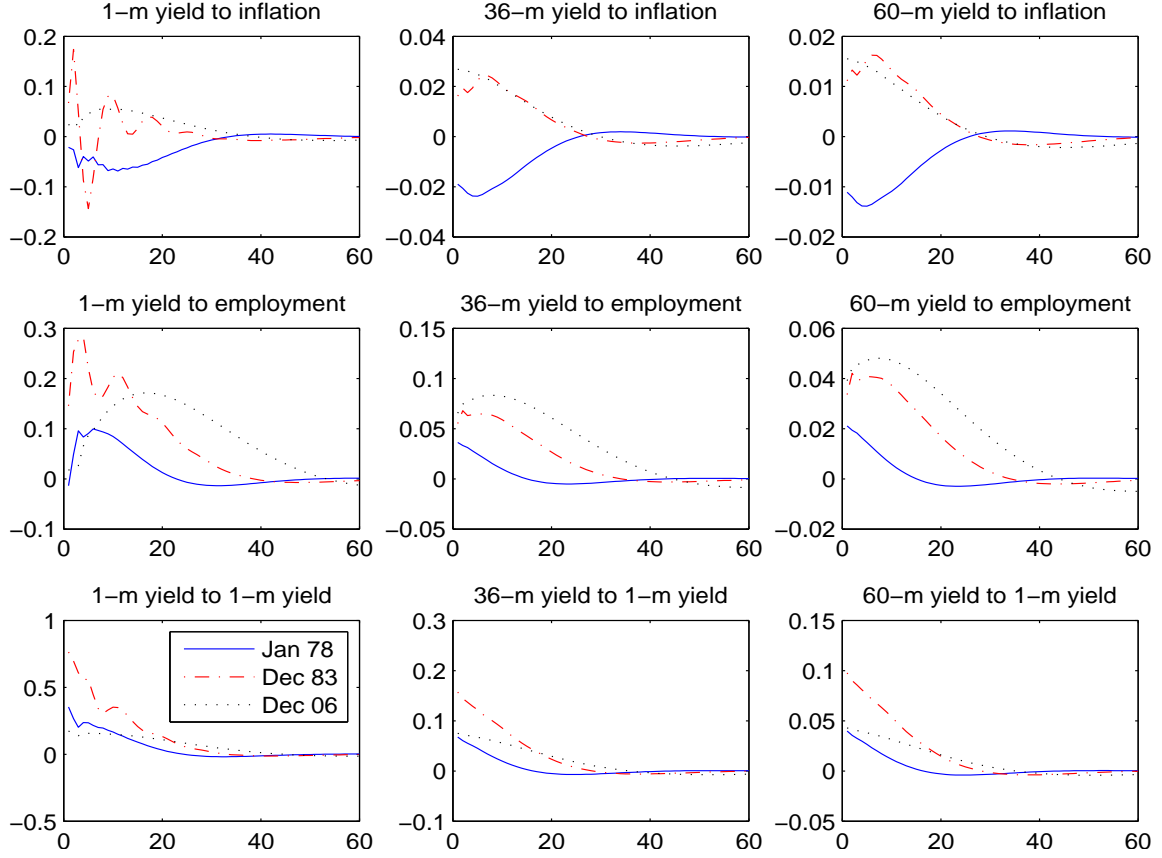
Figure 5: Contribution to Fitted Yields from Time Variation in Σ



interest in view of the large literature documenting time variation in the conduct of monetary policy. Figure 6 provides a perspective by showing impulse responses of the 1-month yield (the left column) and the transmission of these responses to longer-term yields (the right two columns) to the orthogonalized shocks ϵ_t implied by our identification scheme.¹⁷ We present IRFs from early 1978, an episode often associated with failure of monetary policy to respond sufficiently forcefully to inflation; from late 1983, a time by which the Volcker disinflation was close to complete; and from December 2006, our sample end. In fact, the upper left panel appears consistent with the view that monetary policy failed to respond to inflation shocks in 1978, while for the later dates the response is positive. Similarly, the

¹⁷The IRFs of the longer-term yields are computed as described in Ang and Piazzesi (2003).

Figure 6: Impulse Responses of Yields

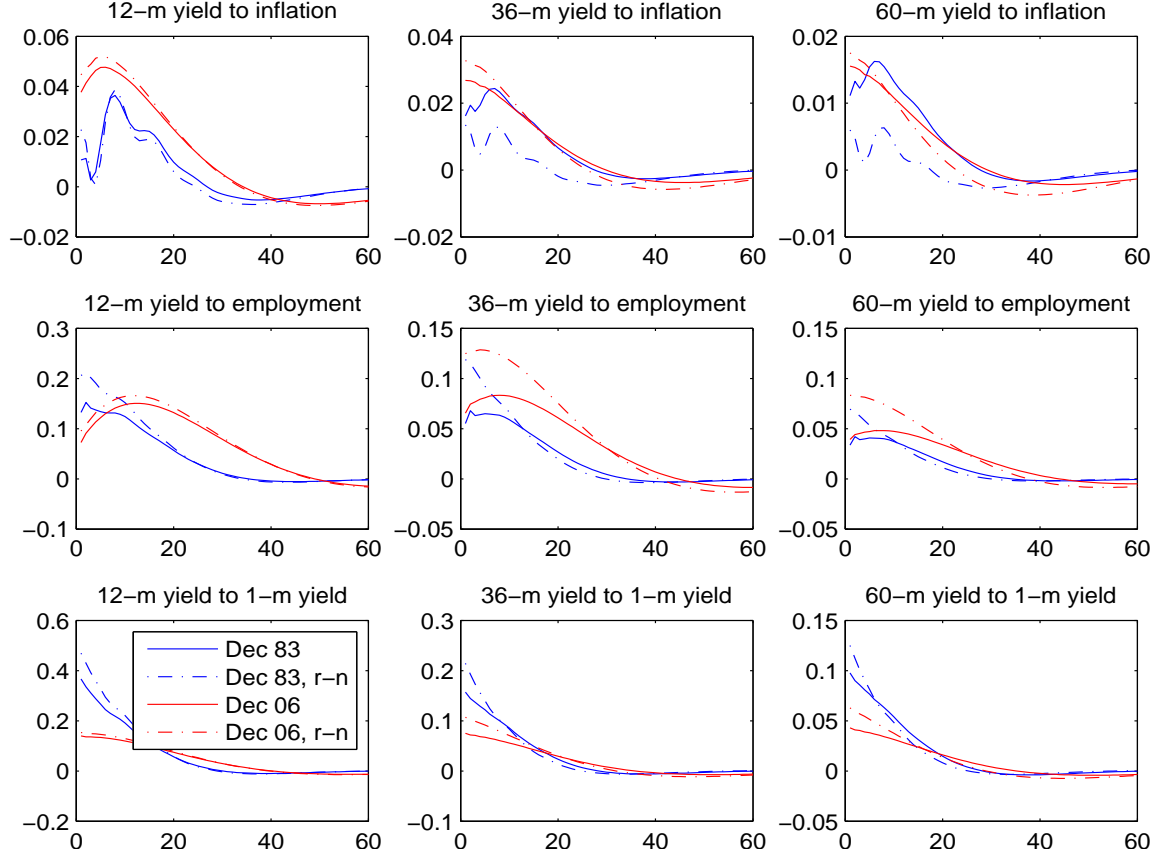


response to employment growth is more forceful for the later two dates than for 1978.

The responses of the longer-term yields shown in the right two columns of Figure 6 can be due either to shifts in expectations of future short rates or to changes in term premia. In Figure 7 we disentangle these two effects by computing IRFs for longer yields that would obtain under risk neutrality. In 1983 (the blue lines), the risk-neutral responses to inflation shocks (the dashed-dotted lines) are substantially below the total IRFs, indicating that at that time an inflation shock was transmitted to longer maturity yields primarily through an increase in compensation for inflation risk; in 2006 (the red lines), the risk-neutral responses to inflation shocks are nearly identical to the overall responses.

Finally, we again consider the sub-sample performance of the model, this time including the estimated term premia described above. Table 4 reports summary statistics for the two

Figure 7: Impulse Responses: Contributions from Term Premia



sub-samples (and the full 1959-2006 sample) for the baseline learning model and the model using the constant-coefficient VAR estimated on the full sample. Note that the parameters describing the price of risk are estimated on the full sample.

The learning model including time-varying term premia dominates the constant-coefficients VAR model in the latter subsample, with the fit of shorter yields comparable but the fit of longer-term yields much better in the learning model. The relative performance of the learning model in the earlier subsample is mixed: it does better at longer yields but worse on shorter-term yields. This pattern in the earlier subsample drives the full-sample results. As noted above, the learning model breaks down at times in the 1970s and early 1980s due to the extreme swings in inflation and interest rates.

In summary, these results show that the learning model's performance, relative to the

Table 4: Results for Sub-samples

	Measurement Error RMSE			
	y^3	y^{12}	y^{36}	y^{60}
Learning model				
1959-1983	0.8837	0.9681	0.8818	0.8664
1984-2006	0.4291	0.6624	0.8280	0.8183
1959-2006	0.7036	0.8357	0.8547	0.8437
Full-sample VAR				
1959-1983	0.7473	0.8120	0.9268	1.0064
1984-2006	0.4203	0.6331	0.9049	1.0760
1959-2006	0.6128	0.7318	0.9164	1.0403

Note: RMS of deviations from expectations hypothesis
computed assuming constant term premia over 1959-2006.

constant-coefficient VAR, is much better in the post-1983 sample than implied by the full-sample results. To the extent that this more recent subsample is more representative of the macroeconomic and monetary policy environment expected to exist in the future, these results provide strong support for the usefulness of the learning model in modeling yields.

5 Forecast Performance

Up to now we have focussed on measures of in-sample fit as criteria for evaluating the correspondence between the predictions of the model and the data. We now turn to the out-of-sample forecasting ability of the learning model and compare it to that of other terms structure models in the literature. As is typical in the literature, we examine forecast horizons of between one, six, and twelve months. Arguably, the multi-period forecasts provide a better reading of the model's ability to describe the underlying dynamic behavior of yields compare than the current period fit.

Table 5 reports root-mean-squared errors obtained from out-of-sample multi-period forecasts of yields from several models. Let $\hat{y}_{t,k}^n$ denote the k -period-ahead forecast for the n -period yield made in period t . The k -period-ahead forecast error for the n -period yield is then given by $y_{t+k}^n - \hat{y}_{t,k}^n$. The models are a random walk ($\hat{y}_{t,k}^n = y_t^n$), the Nelson-Siegel

Table 5: Out-of-Sample RMS Forecast Errors (2000 - 2006)

Yield Maturity	Random Walk	Diebold-Li NS AR	Duffee $A_0(3)$	Learning Macro	Full-sample Macro
1-month ahead forecasts					
12	.24	.29	.28	.80	.77
36	.30	.30	.35	.98	.87
60	.30	.33	.31	.84	.96
6-month ahead forecasts					
12	.92	.96	.93	1.14	1.30
36	.80	.78	.79	1.13	1.11
60	.71	.73	.72	0.93	1.05
12-month ahead forecasts					
12	1.70	1.72	1.69	1.71	1.98
36	1.29	1.27	1.25	1.35	1.54
60	1.03	1.04	1.04	1.05	1.32

AR model of Diebold and Li (2006), Duffee’s (2002) essentially-affine model with three latent factors, and the two macro models with no latent factors described in this paper (labeled “Macro-Learning” and “Full-sample Macro”). The forecasts start in 2000:1 for yields in 2000:2, 2000:7, and 2000:12. All forecasts are evaluated against outcomes through 2006:12.¹⁸ The parameters of the Nelson-Siegel model are reestimated each period; those of the $A_0(3)$ model and the time-invariant parameters of the macro models have been estimated through 1999:12.

As seen clearly in the upper two panels of the table, the two macro models do a much worse job at forecasting yields at one-month horizons than term structure models that incorporate latent factors. This result is consistent with similar comparisons in Mönch (2005) and reflects the superior current-period fit of latent factor models in comparison to models that rely only on observed macro factors. The latent factor models also do somewhat better than the macro models at six-month-ahead forecasts. Interestingly, none of these models clearly outperforms a random walk at the one- and six-month forecast horizons.

¹⁸The results are qualitatively similar for alternative forecast samples from the past 12 years.

The learning model performs about as well as the pure finance models and a random walk at 12-month forecasts horizons and is superior to the full-sample constant-coefficient model at this horizon. Evidently, the learning model captures the underlying movements in yields about as well as atheoretic models. The learning model’s forecasting performance likely can be improved by using factor-augmented VARs (FAVAR) for forecasting as in recent papers by Mönch (2005) and de Pooter, et al. (2007). We leave the incorporation of learning in the context of FAVARs to future work.

6 Robustness Analysis

6.1 Sensitivity to the Discount Factor

In this section, we consider alternative calibrations of the discount factor, ρ used in updating the VAR estimates. The first four columns of Table 6 report the standard deviations of the differences between the model-generated risk-neutral yields and the actual yields for various values of ρ . The next four columns of the table report the RMSE of the model residuals, where the model prediction includes the estimated term premium. The final column reports the log likelihood, \mathcal{L} , associated with the estimated model with the specified value of ρ . For comparison, results for the learning model under the baseline value of $\rho = 0.978$ and for the model with a constant-coefficient VAR estimated on the full sample of data are reported in the final two rows of the table.

The discount factor ρ that minimizes the sum of variances of the deviations from the expectations hypothesis (assuming constant term premia) is 0.975, slightly lower than the baseline value of 0.978. The standard deviations of the deviations from the expectations hypothesis are nearly identical for these two values of ρ . The magnitude of the deviations from the expectations hypothesis is not sensitive to values of ρ between .970 and .985, but these deviations become much larger as the discount factor approaches unity. Nonetheless, even with no discounting, the learning model does better than the full-sample, constant-coefficient VAR based on these statistics.

Similarly, the “fit” of the model is maximized when ρ is slightly above our baseline calibrated value. The fit declines sharply in the absence of discounting. By this measure, the model fit is superior to the constant-coefficient VAR estimated on the full sample of the data as long as ρ is equal or less than the baseline calibrated value. The likelihood

Table 6: Sensitivity of results to ρ

ρ	SD of Deviation from EH				Residual RMSE				\mathcal{L}
	y^3	y^{12}	y^{36}	y^{60}	y^3	y^{12}	y^{36}	y^{60}	
0.970	0.44	0.71	0.91	0.94	0.71	0.81	0.82	0.85	13,593.4
0.975	0.44	0.71	0.91	0.93	0.70	0.82	0.83	0.83	13,599.2
0.980	0.45	0.70	0.95	0.98	0.70	0.84	0.85	0.86	13,539.9
0.985	0.44	0.69	0.96	0.98	0.70	0.86	0.87	0.85	13,541.4
0.990	0.42	0.71	1.09	1.23	0.70	0.88	0.86	0.83	13,539.7
0.995	0.41	0.68	1.05	1.19	0.71	0.90	0.86	0.83	13,521.2
1.000	0.39	0.67	1.02	1.19	0.69	0.91	1.08	1.26	13,148.1
0.978	0.46	0.71	0.91	0.91	0.70	0.84	0.84	0.84	13,559.4
Full sample	0.40	0.70	1.16	1.43	0.61	0.73	0.92	1.04	13,555.8

is maximized at $\rho = 0.975$, the same value found to yield the closest fit in terms of the expectations hypothesis.

7 Conclusions and Further Research

The main finding of our analysis is that incorporating real-time learning in a macro finance model has important implications for the path of risk-neutral yields and the price of risk compared to a standard model with fixed coefficients. On purely theoretical grounds, we view the learning model as providing a more realistic description of the real-time problem of pricing assets in an environment of structural change and imperfect knowledge. Our analysis also indicates that the learning model provides a better fit to the data than that from the standard approach of assuming constant coefficients estimated over the full sample. Moreover, the results from the learning model also appear to correspond more closely to the history of the conduct of monetary policy and economic events of the past several decades.

The analysis in this paper has focused on measures of in-sample fit and forecasting performance for the evaluation of alternative models of the term structure. This represents the tip of the iceberg in terms of term structure model evaluation. As mentioned earlier, recent research has spawned a rich set of theoretical restrictions and empirical regularities that can be applied to evaluate models of the term structure and in future work, we intend to analyze the performance of term structure models with learning with these criteria.

For the sake of parsimony and clarity, we have limited ourselves to a very simple recursive updating formula to represent the evolution of beliefs. In doing so, we have abstracted from number of important issues. Most importantly, we have assumed a constant value for the discount factor that was chosen based on a measure of fit of the VAR over the dull sample. In practice, the discount factor could vary over time depending on the perceived behavior of the economy. In addition, during periods of announced monetary policy regime changes, such as occurred in 1979, agents presumably would not treat past data the same as at other periods and such modifications to the learning process can be incorporated. Similarly, we have assumed that the risk price parameters are fixed over time; one could allow for time variation in these parameters as well.

References

- [1] Ang, A., M. Piazzesi and M. Wei (2006). “What does the yield curve tell us about GDP growth?” *Journal of Econometrics* 131 (1-2), March - April, 359-403.
- [2] Ang, A. and M. Piazzesi (2003). “A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables.” *Journal of Monetary Economics* 50 (4), April, 745-787.
- [3] Beechey, M. (2007). “Learning and the term structure of interest rates.” Manuscript, Federal Reserve Board, June.
- [4] Bekaert, G., S. Cho and A. Moreno (2006). “New-Keynesian macroeconomics and the term structure.” Manuscript, Columbia University, June.
- [5] Bernanke, B., V. Reinhart and B. Sack (2004). “Monetary policy alternatives at the zero bound: An empirical Assessment.” *Brookings Papers on Economic Activity* 2, 1-100.
- [6] Branch, W. and G. Evans (2006). “A simple recursive forecasting model.” *Journal of Monetary Economics* 91 (2), May, 158-166.
- [7] Cochrane, John H. (2006). “Financial markets and the real economy.” Manuscript, University of Chicago, December.
- [8] Campbell, J., A. Lo, and C. MacKinlay (1997). *The econometrics of financial markets*. Princeton, NJ: Princeton University Press.
- [9] Campbell, J. and R. Shiller (1991). “Yield spreads and interest rate movements: A bird’s eye view.” *Review of Economic Studies* 58 (3), May, 495-514.
- [10] Cogley, T. (2005). “Changing beliefs and the term structure of interest rates: cross-equation restrictions with drifting parameters.” *Review of Economic Dynamics* 8 (2), April, 420-451.
- [11] Cogley, T. and T. Sargent (2001). “Evolving post World War II US inflation dynamics.” *NBER Macroeconomics Annual* 16, 331-373.

- [12] Cogley, T. and T. Sargent (2005). "Drifts and volatilities: monetary policies and outcomes in the post WWII US." *Review of Economic Dynamics* 8 (2), April, 262-302.
- [13] Dai, Q. and K. Singleton (2002). "Expectations puzzles, time-varying risk premia, and affine models of the term structure." *Journal of Financial Economics* 63 (3), March, 415-441.
- [14] Dewachter, H. and M. Lyrio (2006). "Macro factors and the term structure of interest rates." *Journal of Money, Credit, and Banking* 38 (1), February, 119-140.
- [15] Dewachter, H. and M. Lyrio (forthcoming). "Learning, macroeconomic dynamics and the term structure of interest rates." In J. Campbell (ed.), *Asset prices and monetary policy*, Chicago, IL: The University of Chicago Press.
- [16] Diebold, F., G. Rudebusch and S. Aruoba (2006). "The macroeconomy and the yield curve: a dynamic latent factor approach." *Journal of Econometrics* 131 (1-2), March - April, 309-338.
- [17] Duffee, G. (2002). "Term premia and interest rate forecasts in affine models." *Journal of Finance* 57 (1), February, 405-443.
- [18] Duffie, D. (2001). *Dynamic asset pricing theory*. Princeton, NJ: Princeton University Press.
- [19] Evans, G. and S. Honkapohja (2001). *Learning and expectations in macroeconomics*. Princeton, NJ: Princeton University Press.
- [20] Evans, C. and D. Marshall (2007). "Economic determinants of the nominal treasury yield curve." *Journal of Monetary Economics* 54 (7), October, 1986-2003.
- [21] Fuhrer, J. (1996). "Monetary policy shifts and long-term interest rates." *Quarterly Journal of Economics* 111 (4), November, 1183-1209.
- [22] Goodfriend, M. (2005). "Inflation targeting in the United States." in B. Bernanke and M. Woodford (eds.) *The Inflation-Targeting Debate*. Chicago: University of Chicago Press, 311-337.

- [23] Gürkaynak, R., B. Sack and E. Swanson (2005). "The sensitivity of long-term interest rates to economic news: evidence and implications for macroeconomic models." *American Economic Review* 95 (1), 425-436.
- [24] Harvey, Andrew (1989). *Forecasting, structural time series models and the Kalman filter*. Cambridge: Cambridge University Press.
- [25] Hördahl, P., O. Tristani and D. Vestin (2005). "The yield curve and macroeconomic dynamics." Manuscript, December.
- [26] Hördahl, P., O. Tristani and D. Vestin (2006). "A joint econometric model of macroeconomic and term-structure dynamics." *Journal of Econometrics* 131 (1-2), March - April, 405-444.
- [27] Kim, D. and A. Orphanides (2005). "Term Structure Estimation with Survey Data on Interest Rate Forecasts." Finance and Economics Discussion Series (FEDS) working paper 2005-48, Board of Governors of the Federal Reserve, October.
- [28] Kozicki, S. and P. Tinsley (2001). "Shifting endpoints in the term structure of interest rates." *Journal of Monetary Economics* 47 (3), June, 613-652.
- [29] Kreps, D. (1997). *Anticipated utility and dynamic choice*. Cambridge: Cambridge University Press.
- [30] Marcet, A. and T. Sargent (1989). "Convergence of least-squares learning mechanisms in self-referential linear stochastic models." *Journal of Economic Theory* 48 (2), August, 337-368.
- [31] Mönch, E. (2005). "Forecasting the yield curve in a data-rich environment: A no-arbitrage factor-augmented VAR approach." ECB Working Paper No. 544, November.
- [32] Kim, C., Nelson, C. and J. Piger (2004). "The less-volatile US economy: a Bayesian investigation of timing, breadth, and potential explanations." *Journal of Business and Economic Statistics* 22 (1), 80-93.
- [33] Orphanides, A. and J. Williams (2005), "Inflation scares and forecast-based monetary policy." *Review of Economic Dynamics* 8 (2), April 2005, 498-527.

- [34] Orphanides, A. and J. Williams (2006). “Inflation targeting under imperfect knowledge.” In F. Mishkin and K. Schmidt-Hebbel (ed.), *Monetary policy under inflation targeting*, Santiago, Chile: Central Bank of Chile.
- [35] Piazzesi, M. and M. Schneider (2006). “Equilibrium yield curves.” *NBER Macroeconomics Annual* 21, 389-442.
- [36] Ravenna, F. and J. Seppälä (2007). “Monetary policy and rejections of the expectations hypothesis.” Manuscript, University of California at Santa Cruz, May.
- [37] Rudebusch, G. and T. Wu (2003). “A macro-finance model of the term structure, monetary policy, and the economy.” Federal Reserve Bank of San Francisco working paper no. 2003-17, December.
- [38] Sargent, T. (1999). *The conquest of American inflation*. Princeton: Princeton University Press.
- [39] Sargent, T. and N. Williams (2005). “Impacts of priors on convergence and escapes from Nash inflation.” *Review of Economic Dynamics* 8 (2), April, 360-391.
- [40] Stock, J. and M. Watson (1996). “Evidence on structural instability in macroeconomic time-series relations.” *Journal of Business and Economic Statistics* 14 (1), January, 11-30.
- [41] Stock, J. and M. Watson (2007). “Why has US inflation become harder to forecast?” *Journal of Money, Credit, and Banking* 39, February, 3-33.
- [42] Tetlow, R. and P. von zur Muehlen (2001). “Simplicity versus optimality: The choice of monetary policy rules when agents must learn.” *Journal of Economic Dynamics and Control* 25 (1-2), January, 245-279.