What does the Yield Curve imply about Investor Expectations?

Eric Gaus[1] **and Arunima Sinha**[2]

Abstract

We use daily data to model investors' expectations of U.S. yields, at different maturities and forecast horizons. We consider two adaptive learning algorithms to characterize the conditional yield forecasts. Our framework yields the first empirical estimates of the pace of learning by investors. The superior performance of the endogenous learning mechanism suggests that investors account for structural change and respond to significant, persistent deviations by modifying the amount of information they use. Our results provide strong empirical motivation to use the class of adaptive learning models considered here for modeling and analyzing expectation formation by investors.

JEL classifications: E43, E47, D83, C5

Keywords: Adaptive learning, Investor beliefs, Inflation Expectations, Excess returns

.

.

.

.

.

.

.

.

.

.

[1]Department of Economics, Haverford College, 370 Lancaster Ave, Haverford, PA 19041. E-mail: egaus@haverford.edu.

[2]Department of Economics, Fordham University. 113 West 60th street, NY, NY 10023. Email: asinha3@fordham.edu. We would like to thank the participants at the Society for Computational Economics (Oslo), Expectations in Dynamic Macroeconomic Models (Bank of Finland) and the American Economic Association 2016 meeting (San Francisco) for comments and suggestions. Discussions with Bruce Preston, Ricardo Reis and Michael Woodford were extremely helpful. All errors are our own.

1 Introduction

> [T]he Federal Reserve's ability to influence economic conditions today depends critically on its ability to shape expectations of the future, specifically by helping the public understand how it intends to conduct policy over time, and what the likely implications of those actions will be for economic conditions. (Vice-Chair Janet Yellen, At the Society of American Business Editors and Writers 50th Anniversary Conference, Washington, D.C., April 4, 2013)

Investor expectations about the term structure of yields are central to the conduct of monetary policy. Influencing these expectations through the different instruments available to the Federal Reserve, has been important during the Great Moderation. During the Great Recession and its aftermath, this strategy has been at the forefront of the central bank's policy. As the accommodative monetary policy stance of the Federal Reserve kept the federal funds rate at the zero-lower bound from December 2008 to November 2015, one of the main channels through which monetary policy affected longer yields (and the subsequent consumption and savings decisions of economic agents), was by affecting the formation of conditional expectations by market investors.

The contribution of the present analysis is to characterize the expectations formation process of market investors about the term structure of yields at different forecasting horizons. We further explore whether differences exist in the expectations formation process between the Great Moderation and the Great Recession periods, i.e., during periods of low and high macroeconomic volatility. To do this, we develop a novel methodology to model the evolution of investor beliefs using daily data on the U.S. nominal yield curve. Using the Great Moderation as the baseline period, we extend the results to include the Great Recession. Our analysis allows for the comparison of investor beliefs about the entire yield curve, across a cross-section of forecast horizons.

Our strategy is briefly described as follows: we use the daily yield curve factors estimated by Gürkaynak, Sack and Wright (2007; henceforth, GSW) to construct yield forecasts. Following recent studies[3], we first construct conditional expectations of yields (and associated latent factors) using a vector auto-regressive (VAR) model with constant coefficients. We evaluate the forecasting performance of the model, and a series of rationality tests of the

[3] Examples include Diebold and Li (2006) and Aruoba, Diebold and Rudebusch (2006).

implied forecasting errors confirm that these errors are biased, systematic, and correlated with revisions in yield forecasts. In addition to these findings on the forecast errors, the framework also imposes the restriction that investors must be placing identical weights on past information while forecasting the short and long yields over different forecasting horizons. Thus, it does not allow investors to endogenously adapt to any structural breaks that they might perceive in the evolution of the average yields, or the yield curve slope.

The above results motivate our hypothesis that market investors are using other models of expectations formation. Theoretical analyses, such as Piazzesi, Salomao and Schneider (2015) and Sinha (2016), incorporate adaptive learning into the expectations formation of optimizing agents in models of the yield curve. The implied term structures are more successful at matching the properties of the empirical yield curve, relative to models with time-invariant beliefs. Therefore, we explore a class of adaptive learning models for the formation of conditional forecasts of the nominal and real yield curve factors, and subsequent yields: constant gain learning (CGL) and an endogenous learning (EGL) algorithm that we develop here. The main innovation is that investors are now allowed to vary the weights they place on past information about yields; they are also able to adapt to the size of large and persistent deviations observed in the yield curve factors data.

We find that there are significant improvements in forecasting performance of the model with the learning processes. Our results are based on the implied forecasts for the two sample periods, and the methodology characterizes the speed of learning by market investors using high frequency data, This, to our knowledge, is a first for the adaptive learning literature. The parameters of the learning models - the updating coefficient or the "gain", the adjustment factor in case of large deviations, and the time period used to compute deviations with respect to historical data, are all estimated from the daily yield curve data, for different forecasting horizons. The main result is that at different pairs of yield maturity and forecast horizon, the endogenous learning forecast improves upon the constant gain algorithm. For example, at the 1-month forecasting horizon, for the nominal 1-year yield, endogenous learning improves upon the constant gain mean square forecast error by 36%; at the 6-month horizon, the improvement is close to 18%. This improvement in the performance persists across yield maturities as well.

We investigate the forecasts obtained from the learning models on two separate dimensions. First, we subject the forecast errors to the same set of rationality tests, as done for the constant-coefficients model. The results show that forecast errors from both the learning models improve upon the findings for the time-invariant case: the errors are unbiased (or significantly less so), the predictive content of current forecasts for the errors is either insignificant or much reduced, and systematic nature of the forecast errors is also greatly diminished for the 1- and 10-year yields. Thus, we conclude that the forecast errors of the learning model improve upon those from the constant-coefficients model.

We then test whether the out-of-sample learning forecasts improve upon random walk forecasts[4]. Since yields are very persistent, the random walk model is difficult to out-perform in these out-of-sample forecasts. At the shortest forecasting horizon (1-month), we find that this is still the case. However, at the longer forecasting horizons, the learning models improve upon the random walk forecasts. For example, the EGL 6-month forecast for the 10-year yield improves upon the random walk model by approximately 13%.

In addition to the superior performance of the EGL mechanism on the above dimensions, the estimation of the endogenous gain parameters yields several insights into the expectations formation process of agents: (a) the implied conditional expectations of investors display substantial time-variation and adapt to large deviations in the data during periods of low and high macroeconomic volatility (we present estimates of the gains from the Great Moderation as well as the Great Recession to demonstrate this); (b) while constructing forecasts of the 10-year yield, the investors' expectations are largely invariant to large deviations in the observed data, and they account for structural change to a significantly smaller degree, relative to the 1- and 5-year yields, during the Great Moderation. This contrasts with the expectations formation during the Great Recession, where the 1- and 3-month forecasts of the 10-year nominal yield give more weight to the more recent level and slope factor data. This suggests that during periods of low macroeconomic volatility, investors do not expect structural change in the data for the long-term yield, but become much more attentive during highly volatile periods. Thus, monetary policy actions that target the long-end of the yield curve during a recession may be more successful at influencing the savings and investment decisions of agents. This result supports findings elsewhere in the literature: for example, Coibion and Gorodnichenko (2015) find that survey forecasters exhibit greater information rigidities during the Great Moderation, compared to the earlier recessions. Using

[4]In the accompanying appendix, we also compare the learning forecasts with those of the Diebold-Li (2006) model.

observed data on the yield curve, we show that investors are, in fact, forming conditional expectations differently over the business cycle. Thus, the mechanism offers a tractable way of incorporating time-variation in expectations formation, and can aid in policy analysis: less attention to more recent data during recessions (for different yield maturities and forecast horizons) indicates that investors will less more rapidly to policy changes during recessions.

Our methodology allows for investors to allow the gains to vary across different yield curve factors and across forecast horizons; this provides a more intuitive way to allow for the investors to update their information. For example, while forming forecasts, the investors may place more or less weight on the history of the level of yields, than on the slope of the yield curve. If they believe that there were several structural breaks in the average level of the yield curve, they may prefer to place more weight on the recent past observations, instead of the longer history. If such breaks are not perceived to exists in the yield curve slope, the investors may place almost equal weight on past observations. These gain parameters are therefore central to the bounded rationality approach, since they determine the persistence in expectations formation, and how investors will react to permanent versus transitory shocks. In this analysis, we use fixed baseline time periods (for the Great Moderation and the Great Recession period) to find the optimal gains.

Given the success of the endogenous learning mechanism in modeling conditional forecasts of investors, we then consider the implications of our framework for two important aspects of the term structure of interest rates: first, what are the inflation expectations implied by the endogenous learning model? Second, how far does the EGL mechanism go in matching the patterns observed in survey data and other data on expected excess returns?

Inflation expectations for 5- and 10-years at different forecast horizons, are derived using the difference between the conditional expectations of the nominal and corresponding TIPS yields. We find that up until the middle of the 2006, the 1-, 3- and 6-month inflation expectations kept pace with each other. However, by the beginning of 2007, there was a significant divergence in the inflation expectations for the 1- month relative to the others. The period at the start of the financial crisis was also marked by an enormous increase in the uncertainty of inflation expectations, both for the 5- and 10-year yields. We further consider the correlations of the inflation expectations obtained from the endogenous learning model with two measures of inflation expectations - the model-based inflation expectations derived by the Federal Reserve Bank of Cleveland, and the 10-year inflation expectations from the

Survey of Professional Forecasters (SPF).

To examine the implications for excess returns, we first use survey data from the SPF to derive the excess returns for ten-year nominal yields at different forecasting horizons. The excess returns are then constructed in a similar manner from the learning models. The endogenous learning mechanism shows a higher correlation with survey expected excess returns, relative to the constant gain mechanism. We also construct expected excess returns from the Cochrane and Piazzesi (2005) model, and compute the correlations of the learning model-implied excess returns with this model's excess returns.

This paper is organized as follows: section two gives a brief overview of the literature. The factor model for the nominal yield curve, and tests for systematic relationships between the forecast errors and revisions are described in section three. Section four discusses the different learning mechanisms and section five presents the numerical results, along with a discussion of the optimization routines. We also discuss the findings in the context of other endogenous learning mechanisms here. Further tests of the learning model are discussed in section six, including comparisons with the random walk model. The findings for inflation expectations and expected excess returns are described in section seven and section eight concludes.

2 Related Literature

Several analyses have used the Nelson-Siegel-Svensson parameterization for fitting the yield curve. The U.S. nominal yield curve data used here is drawn from the yield curves estimated by GSW (2007) based on this spline approach. There are other widely-used frameworks for modeling the term structure as well.[5] However, the focus of the present paper is to extract the process which best approximates the evolution of the yield curve factors, instead of analyzing different models of yield curve estimation. Thus, we choose a flexible, parsimonious

[5] For example, Aruoba, Diebold and Rudebusch (2006) estimate the yield curve using the Nelson-Siegel approach, and estimate the evolution of the yield and factor jointly. Diebold and Li (2006) propose a dynamic version of the approach. These analyses use the original three-factor model of Nelson and Siegel (1987). The Svensson (1994) model extends this framework and incorporates additional flexibility in the shape of the yield curve. A survey of the different models of the term structure and their relative forecasting performances is conducted by Pooter (2007). A more recent approach has introduced the restrictions used in affine arbitrage-free models of the term structure, which suffer from poor forecasting performance, into the spline based methods (Christensen, Diebold and Rudebusch, 2011).

framework that is widely used for modeling the term structure, and analyze the conditional forecasts implied by this approach.

Our study is related to the recent work that has introduced time variation in the estimation of yield curve forecasts. Bianchi, Mumtaz and Surico (2009) model the U.K. nominal yield curve using the Nelson-Siegel-Svensson approach; the authors also use a time-varying process for the evolution of the factors. In their model, a regime-switching model for the evolution of the factors is specified. Duffee (2011) develops a three-factor term structure model, in which the factors are the first three components of yields. A random walk on the first principal component (corresponding to the level) is imposed; the other two factors are assumed to be stationary. Van Dijk, Koopman, Wel and Wright (2014) also impose non-stationarity on the level component of the Nelson-Siegel model, wherein the authors consider autoregressive specifications with a time-varying unconditional mean or a "shifting endpoint". They present three approaches to model the shifting endpoint: exponential smoothing; survey forecasts of interest rates, output and inflation and exponentially smoothed realizations of the macroeconomic variables. Both the latter papers establish the superior performance of the respective time-varying models in out-of-sample yield forecasts. While the motivation of the present analysis is similar, we allow for time-varying coefficients in all the factors in the term structure model. This allows us to investigate whether the importance of the yield curve level vis a vis the slope remains the same across different periods and forecast horizons. Also, the investors are assumed to entirely rely on the yield curve time series, without assuming a specific form of dependence on different macroeconomic variables. The focus of the present exercise is to characterize investor expectations about the different aspects of the yield curve. While we present the in-sample forecasting errors below, out-of-sample forecasting is not the main objective of the analysis.

Recent analyses have also investigated the forecasting performance of the Nelson-Siegel model with time-varying models. Starting with the dynamic Nelson-Siegel model, Chen and Niu (2014) construct forecasts of yields by optimally selecting the sample period over which parameters are approximately constant for the factor dynamics. The constructed adaptive dynamic Nelson-Siegel (ADNS) model constructed by the authors results in better forecasts of the yield curve compared with the dynamic Nelson-Siegel model. The algorithm is able to detect structural breaks in the factor series, and this leads to the improvements in forecast performance. Xiang and Zhu (2013) develop a regime-switching version of the dynamic

Nelson-Siegel model. In this framework, latent yield factors are assumed to follow a Markov-switching VAR process, and the optimal number of regimes (found to be two regimes based on daily U.S. Treasury yield data) are determined in the estimation process. The two-regime Nelson-Siegel model is found to generate superior yield forecasts, especially at the shorter end of the yield curve. Although forecasting performance is not the focus of the present analysis, our work is related to both these studies as we also posit that the investors are forming expectations using a time-varying coefficients model.

In order to discipline the time-varying parameters in our analysis, we use variants of the adaptive learning algorithm. Other analyses have incorporated the adaptive learning framework in the optimizing agents' expectations formation to derive the yield curve in partial and general equilibrium models, to improve the fit of the model relative to empirical observations of the term structure. Laubach, Tetlow and Williams (2007) allow investors to re-estimate the parameters of their term structure model – both those determining the point forecasts of yields, and the parameters describing economic volatility – based on incoming data. Kozicki and Tinsley (2001) and Dewachter and Lyrio (2006) use changing long-run inflation expectations as an important factor characterizing the yield curve. Fuhrer (1996) finds that estimating changing monetary policy regimes is important for the success of the Expectations Hypothesis of the term structure. Piazzesi, Salomao and Schneider (2015) decompose expected excess returns into the returns implied by the statistical VAR model and survey expectations, used as an approximation for subjective investor expectations. Survey expectations are found to be significantly more volatile compared to model implied returns. Giacoletti, Laursen and Singleton (2014) estimate a dynamic term structure model in which the investor learns about the joint distribution of the yield curve and the macroeconomy. The common theme of these analyses is the incorporation of subjective beliefs in explaining characteristics of the empirical term structure. The distinguishing feature of our analysis is we use the term structure data to estimate the process that produces the best in-sample forecasts at different forecast horizons and maturities, to approximate the expectations process of agents. We also allow the investors to endogenously learn from their past errors.[6]

While the above mentioned analyses have primarily used constant-gain adaptive learn-

[6] In contrast, the analysis of Piazzesi, Salomao and Schneider (2015) directly imposes the constant-gain learning model on the expectations formations process of optimizing agents, and analyzes the subsequent forecasts. In this case, investors form beliefs over different forecast horizons and yield maturities using the same updating parameter.

ing, endogenous learning algorithms have also been previously incorporated by Marcet and Nicolini (2003) and Milani (2007a). In the former analysis, the authors incorporate bounded rationality in a monetary model; the agents switch between using a constant gain and a decreasing gain algorithm. They are successfully able to explain the recurrent hyperinflation across different countries during the 1980s. One of the main contributions of this analysis is to present a tractable endogenous gain algorithm, in which the optimizing agents are able to adjust their gain parameters in response to significant deviations from the historical mean. Here, the size of the gain responds to the deviations; in Milani (2007a), the agents switch between constant gains based on the historical average of the forecasting errors. Our work is closely related to Gaus (2014), who proposes a variant of the endogenous gain learning mechanism, in which the agents adjust the gain coefficient in response to the deviations in observed coefficients. Kostyshyna (2013) develops an adaptive step-size algorithm to model time-varying learning in the context of hyperinflations.

3 Factor Model and the Performance of Implied Yield Forecasts

GSW (2007) model the zero-coupon yield curve for 1972 − 2011 using the Nelson-Siegel-Svensson approach:

$$\begin{aligned} y_t^n &= \beta_0 + \beta_1 \frac{1-\exp\left(\frac{-n}{\tau_1}\right)}{\frac{n}{\tau_1}} + \beta_2 \left[\frac{1-\exp\left(\frac{-n}{\tau_1}\right)}{\frac{n}{\tau_1}} - \exp\left(\frac{-n}{\tau_1}\right)\right] \\ &\quad + \beta_3 \left[\frac{1-\exp\left(\frac{-n}{\tau_2}\right)}{\frac{n}{\tau_2}} - \exp\left(\frac{-n}{\tau_2}\right)\right]. \end{aligned} \tag{1}$$

Here y_t^n is the zero-coupon yield of maturity n months at time t, β_0 approximates the level of the yield curve, β_1 approximates its slope, β_2 the curvature and β_3 the convexity of the curve. The latter captures the hump in the yield curve at longer maturities (20 years or more). When $\beta_3 = 0$, the specification in (1) reduces to the Nelson-Siegel (1987) form. This functional form is a parsimonious representation of the yield curve.[7] The estimates for this

[7]See Pooter (2007) for an overview of the methods and forecast comparison.

nominal curve are updated daily, and are available from January 1972 on the Federal Reserve Board website. The parameters in (1), which are $\beta_0, \beta_1, \beta_2, \beta_3, \tau_1$ and τ_2, are estimated using maximum likelihood by minimizing the sum of squared deviations between the actual Treasury security prices and the predicted prices.[8] In our analysis below, we will be using these daily factors estimated by GSW.

To construct yield forecasts using the representation in (1), it must be amended with a process for the evolution of the factors[9]:

$$
\begin{aligned}
\mathbf{y}_t &= \mathbf{X}_t \boldsymbol{\beta}_t + \boldsymbol{\varepsilon}_t & \text{(2a)} \\
\boldsymbol{\beta}_t &= \boldsymbol{\mu} + \Phi \boldsymbol{\beta}_{t-1} + \boldsymbol{\eta}_t. & \text{(2b)}
\end{aligned}
$$

Here $\mathbf{y}_t$ is the $n \times 1$ vector of yields, $\mathbf{X}_t$ is a $n \times 4$ vector of the regressors in (1), $\boldsymbol{\beta}_t$ is a 4×1 vector of the factors, μ is the intercept and Φ denotes the dependence of the factors on past values. Since the parameters τ_1 and τ_2 are jointly estimated using the maximum likelihood approach, the $\mathbf{X}_t$ vector is time-varying. Also, $var(\varepsilon_t) = H$ is a diagonal $n \times n$ matrix, and $var(\eta_t) = Q$ is a 4×4 diagonal matrix. The factor errors are assumed to be distributed as a normal, with mean zero.[10] We will consider this as the benchmark model for factor evolution.

The forecasts of the yields are constructed as follows:

$$
\begin{aligned}
E_t \mathbf{y}_{t+h} &= E_t \mathbf{X}_t \hat{\boldsymbol{\beta}}_{t+h} = \mathbf{X}_t E_t \hat{\boldsymbol{\beta}}_{t+h} & \text{(3a)} \\
E_t \hat{\boldsymbol{\beta}}_{t+h} &= \left[I_4 - \hat{\Phi}^h \right] \left[I_4 - \hat{\Phi} \right]^{-1} \boldsymbol{\mu} + \hat{\Phi}^h \boldsymbol{\beta}_t, & \text{(3b)}
\end{aligned}
$$

where h is the forecast horizon. Here, the second equality in (3a) holds since we use estimated values of the parameters τ_1 and τ_2 at time t, while forming the conditional forecasts.

[8] The prices are weighted by the inverse of the duration of the securities. Underlying Treasury security prices in the Gürkaynak, Sack and Wright estimation are obtained from CRSP (for prices from 1961 - 1987), and from the Federal Reserve Bank of New York after 1987.

[9] This is the two-step estimation of yields and factors (Diebold and Li (2006) and Aruoba, Diebold and Rudebusch (2006)).

[10] In the estimation, the cross covariances in η_t are set to zero.

3.1 Tests of the Forecast Errors

Since the model for factor evolution in (2b), and implied conditional yield forecasts in (3a) have been widely used in the literature, we first test the forecast errors implied by this framework. The underlying hypothesis in these analyses is that the framework in (2b) is the "true" model for factor evolution. In this case, the forecasts of yields would be rational; that is, they satisfy the null hypotheses of unbiasedness and efficiency. Thomas (1999) presents a survey of the literature that examines the rationality of inflation forecasts reported by different surveys, and these tests are used to analyze the rationality of the forecasts from the benchmark model. For the following tests, the sample period from $1985-2000$ is considered. The forecasts are constructed for the next four years, using a rolling data window. At each step, the 1-, 3- and 6-month ahead forecasting errors are constructed. This exercise uses data at the daily frequency, and the forecast errors at maturity n and horizon h are defined as the difference between the realized yields, and the conditional expected yields from (3a).

3.1.1 Are the Forecast Errors Unbiased?

In order to test whether the model specification in (2b) leads to unbiased forecasts, the following regression is considered:

$$y^n_{t+h} - E_t y^n_{t+h} = \alpha^n + e^n_{t,t+h}, \tag{4}$$

for forecast horizons $h = 1, 3$ and 6 months.[11] Here $E_t y^n_{t+h}$ is the expectation at time t of the yield of maturity n, h periods into the future. The errors corresponding to the regressions for different yield maturities are denoted by e^n_{1t}. The coefficients for the different yield maturities and forecast horizons are shown in the first panel of table 1. The null hypothesis of unbiasedness requires $\alpha^n_1 = 0, \forall n$. The coefficients in this panel show that for the 1-year yield maturity, as the forecast horizon increases, the implied conditional forecasts of yields overshoot the realized yields. For the 5- and 10-year yields, the model undershoots the implied yields, but as the forecast horizon increases, the conditional forecasts are larger than the actual yields.

[11] This is equivalent to the specification considered by Thomas (1999), and is used by Mankiw, Reis and Wolfers (2004).

3.1.2 Are the Forecast Errors Efficient?

We test whether there is information in the forecast of the yields which can help to predict the forecast error:

$$y^n_{t+h} - E_t y^n_{t+h} = \alpha^n + \beta^n E_t y^n_{t+h} + e^n_{t,t+h}. \tag{5}$$

Under the null hypothesis, $\alpha^n = 0$ and $\beta^n = 0$. This implies that the forecasts themselves have no predictive content for forecast errors. The coefficients in the second panel of table 1 show that this hypothesis is rejected for the yield maturities considered, across the different forecast horizons.

3.1.3 Are the Forecast Errors Systematic?

If (2b) is the true model for the evolution of the factors, then the implied yield forecasts must correspond to the "true" forecast. In this case, the forecast errors must be uncorrelated with the revision in forecast yields. That is, in the following regression:

$$y^n_{t+h} - E_t y^n_{t+h} = \alpha^n + \beta^n \left(E_t y^n_{t+h} - E_{t-1} y^n_{t+h} \right) + e^n_{t,t+h} \tag{6}$$

the intercept and slope coefficients must be statistically not different from zero.[12] The coefficients from the regression in (6) are reported in the third panel of table 1. Several patterns of interest emerge from the coefficient estimates. The slope coefficients are statistically different from zero, implying that the ex-post forecast errors are systematically predictable from the ex-ante forecast revisions. There is also a qualitative difference in how the forecast errors respond to forecast revisions at various horizons. At the longest forecast horizon considered, the slope coefficient is positive, implying that the yield forecasts implied by the model were lower than observed yields.

[12] This is similar to the test used by Coibion and Gorodnichenko (2012) as a test for full-information rational expectations. The authors map the estimates of the slope coefficients which they obtain from a regression of inflation forecast errors on the inflation forecast revisions in survey data to theoretical models of asymmetric information.

Yield Maturity	$h = 1$ month		$h = 3$ months		$h = 6$ months	
	α	β	α	β	α	β
	Test 1: $y_{t+h} - E_t y_{t+h} = \alpha + error_t$					
1 year	-2.1764 (0.04)	-	-3.5495 (0.05)	-	-5.2820 (0.08)	-
5 years	0.6366 (0.02)	-	-0.5364 (0.03)	-	-1.9979 (0.05)	-
10 years	1.9427 (0.02)	-	0.7984 (0.03)	-	-0.6240 (0.04)	-
	Test 2: $y_{t+h} - E_t y_{t+h} = \alpha + \beta E_t y_{t+h} + error_t$					
1 year	1.8125 (0.12)	-0.9225 (0.02)	2.4473 (0.12)	-1.0533 (0.02)	3.1419 (0.10)	-1.1353 (0.01)
5 years	2.3957 (0.04)	-0.5669 (0.01)	2.5063 (0.06)	3.7128 (0.03)	3.0019 (0.07)	-0.8723 (0.01)
10 years	3.8036 (0.02)	-0.6499 (0.00)	-0.5364 (0.03)	-0.7276 (0.00)	3.8714 (0.04)	-0.8286 (0.00)
	Test 3: $y_{t+h} - E_t y_{t+h} = \alpha + \beta \left(E_t y_{t+h} - E_{t-1} y_{t+h} \right) + error_t$					
1 year	-0.0000 (0.00)	3.0649 (0.02)	0.0357 (0.00)	0.5958 (0.00)	0.0496 (0.00)	0.7051 (0.00)
5 years	0.0489 (0.00)	-0.5694 (0.02)	0.0505 (0.00)	0.1743 (0.00)	0.0539 (0.00)	0.3776 (0.00)
10 years	0.0801 (0.00)	-2.4728 (0.02)	0.0818 (0.00)	-0.0691 (0.00)	0.0722 (0.00)	0.1809 (0.00)

Table 1: Testing Forecast Errors for Nominal Yield Curve Factors

Note: The above coefficient estimates are reported using daily data on the latent factors, for the period 1985-2000. The standard errors are shown for the corresponding coefficients in brackets. These coefficients are statistically significant at the 5% level.

3.1.4 Forecast Errors from the Survey Data

For comparison, it is useful to analyze the performance of expectations of yields reported by the Survey of Professional Forecasters (SPF) using the above tests. SPF data on median forecasts of the 10-year Treasury yield and 3-month Treasury bills are available. We construct the regressions in (4), (5) and (6) using the forecasts at the 6- and 12-month forecast horizons.[13] The results are shown in three panels in table 2. The null of unbiasedness is strongly rejected for the 3-month Treasury bills. The median forecasts of the Treasury bills and the 10-year bonds are found to have strong predictive power for the forecast errors, and the forecast revisions are related to the forecast errors in a statistically significant manner.[14]

[13]This regression is constructed using the monthly forecasts reported by the SPF.

[14]SPF forecasts are only available monthly, and the expectations are reported at the quarterly horizons.

Yield Maturity	$h=3$ months α	β	$h=1$ year α	β
	Test 1: $y_{t+h} - E_t y_{t+h} = \alpha + error_t$			
T-bill	-0.1288*** (0.05)	-	-0.2305 (0.18)	-
10 year	-0.1220 (0.09)	-	-0.2305 (0.18)	-
	Test 2: $y_{t+h} - E_t y_{t+h} = \alpha + \beta E_t y_{t+h} + error_t$			
T-bill	0.3201* (0.19)	-0.0794** (0.03)	6.8827*** (1.17)	-1.1136*** (0.18)
10 year	2.0809*** (0.77)	-0.3472*** (0.12)		
	Test 3: $y_{t+h} - E_t y_{t+h} = \alpha + \beta \left(E_t y_{t+h} - E_{t-1} y_{t+h}\right) + error_t$			
T-bill	-0.1040** (0.04)	0.3636*** (0.09)	-0.2135 (0.18)	-0.3735 (0.48)
10 year	-0.1209 (0.10)	0.2493 (0.21)		

Table 2: Testing Forecast Errors for SPF Data

Note: The SPF median forecasts are reported monthly, and data from 1992Q2-2002-Q4 is used here. *** denotes significance at the 1% level, ** at the 5% level and * at the 10% level

4 Construction of Yield Forecasts under Alternative Learning Models

In this section, investors are allowed to update their estimates of the parameters (μ, Φ), as new information becomes available. That is, in contrast to (2b), this process is represented using a time-varying VAR model (with the coefficients being updated using different learning schemes):

$$\boldsymbol{\beta}_t = \boldsymbol{\mu}_{t-1} + \Phi_{t-1}\boldsymbol{\beta}_{t-1} + \boldsymbol{\eta}_t. \tag{7}$$

The timing is as follows: at time t, the GSW estimates of $(\beta_0, \beta_1, \beta_2, \beta_3, \tau_1, \tau_2)$ are used, and to construct forecasts of the yields at 1-, 3- and 6-month horizons, the investors use the learning processes described below to determine (μ_t, Φ_t). Once the parameters (μ_t, Φ_t) are estimated, they are used for constructing the conditional yield forecasts. At time $t+1$ the process is repeated, and updated estimates of (μ_{t+1}, Φ_{t+1}) are used to construct the forecasts of yields and corresponding forecast errors. For each factor $\beta_i, i \in \{0, 1, 2, 3\}$, the coefficients $\Omega_{i,t} = \left(\mu_{i,t}, \Phi_{i,t}\right)$ are updated as:

$$\begin{aligned} \begin{pmatrix} \mu_{i,t} \\ \phi_{i,t} \end{pmatrix} &= \begin{pmatrix} \mu_{i,t-1} \\ \phi_{i,t-1} \end{pmatrix} + g_i R_{i,t-1}^{-1} q_{i,t-1} \left[\beta_{i,t} - \begin{pmatrix} \mu_{i,t-1} \\ \phi_{i,t-1} \end{pmatrix}' q_{i,t-1} \right] \\ R_{i,t} &= R_{i,t-1} + g_i \left[q_{i,t-1} q_{i,t-1}' - R_{i,t-1} \right] \end{aligned} \tag{8}$$

where $q_{i,t-1} = \left(1, \beta_{i,t}\right)_{t=0}^{\prime t-1}$, g_i is the weight the investors assign to the forecast errors made and $\beta_{i,t}$ is the latent factor derived at time t using the maximum likelihood procedure. Finally, the forecasts of the yields are given by:

$$\begin{aligned} E_t \mathbf{y}_{t+h} &= \mathbf{X}_t E_t \hat{\boldsymbol{\beta}}_{t+h} \\ E_t \hat{\boldsymbol{\beta}}_{t+h} &= \left[I_4 - \hat{\Phi}_{t-1}^h \right] \left[I_4 - \hat{\Phi}_{t-1} \right]^{-1} \boldsymbol{\mu}_{t-1} + \hat{\Phi}_{t-1}^h \boldsymbol{\beta}_t. \end{aligned} \tag{9}$$

The only distinction from (3a) is that the coefficients (μ_t, Φ_t) are updated over time. We make the assumption that while making conditional forecasts at time t, the investors do not allow for the possibility that they will revise their estimates of (μ, Φ).[15] The two updating schemes that we consider are described below.

4.1 Constant gain learning

With constant gain learning (CGL), the gain parameter g is fixed. CGL has been a widely used method for characterizing the expectations formation for optimizing agents. In contrast

[15] This is the anticipated utility assumption (Kreps, 1988).

to the constant-coefficients model, investors can now allow for structural changes in the data they are forecasting, by placing an exponentially decaying weight on the history of observations. However, this process does not allow them to modify the weights they place on past data, in case they observe actual data realizations that are significantly different. That is, at any point in time, the agents will continue to place the same weight on an observation n quarters ago, that they did before. Due to this characteristic of CGL, the technique is limited in explaining the behavior of macroeconomic variables, such as the high inflation in 1970s, and the subsequent behavior of the series during the Great Moderation. These observations motivate us to propose the following learning techniques.

4.2 Endogenous gain learning

Under endogenous learning, the investors continue to use the law of motion for the factors in (7), along with the updating equation in (8). However, the gain is no longer held fixed for the entire sample. Under endogenous learning (EGL) the gain switches according to the specification below:

$$g_t = \bar{g}_{lb} + \bar{g}_{sf} \frac{\left| \frac{\Omega_t - \bar{\Omega}_k}{\sigma_\Omega} \right|}{1 + \left| \frac{\Omega_t - \bar{\Omega}_k}{\sigma_\Omega} \right|}. \tag{10}$$

Here $\bar{\Omega}$ is the average of the k most recent coefficients and σ_Ω is the standard deviation of these k coefficients. The lower bound of the endogenous gain is $\bar{g}_{lb}$, and $\bar{g}_{sf}$ is the scaling factor. In this variant of endogenous learning, if the recent coefficient estimate (Ω_t) is close to the mean ($\bar{\Omega}_k$), then $g_t = \bar{g}_{lb}$. However, as the realization of Ω_t diverges from $\bar{\Omega}_k$, the gain approaches $\bar{g}_{lb} + \bar{g}_{sf}$. Therefore, as long as $\bar{g}_{sf} < 1$ and $\bar{g}_{lb} + \bar{g}_{sf} < 1$, g_t will be bounded between zero and one. The novel feature of this learning mechanism is that it allows the investors to endogenously switch or adjust their beliefs and permits them to change the weights they place on past data, in response to new information. Investors are allowed to increase or decrease the value of the gain in times when their coefficient estimates are different from the recent past; the size and sign of this adjustment will be determined in the estimation below. This algorithm was originally developed in Gaus (2014). The comparative numerical results below are presented for the CGL and gain specification following (10). The estimation of the gain parameters for (8) and (10) are discussed below in section 5.1 below.

It is useful to note here that this algorithm allows investors to place greater (or smaller) weight on new information in periods of large deviations, and therefore vary the degree to which they are becoming more (or less) attentive to the recent data is estimated from the yield curve data. In our estimation strategy (described in section 5.1 below), parameters $\bar{g}_{lb}$, $\bar{g}_{sf}$ and k are estimated from the baseline period. Thus, if the data implies that investors pay the same attention to the past data in periods of large deviations as "normal" times, then the endogenous learning algorithm will be flexible enough to accommodate this.

The EGL can be further understood in the context of the learning rule adopted by Marcet and Nicolini (2003). In that exercise, the learning mechanism is one in which decreasing gain (or standard least squares learning) is used in stable periods, and the agents switch to using constant gain in periods of "instability". Thus, the expectations formation process is endogenous to the model, which successfully accounts for recurrent hyperinflations in the 1980s. Recent work by Carvalho, Eusepi, Moench and Preston (2015) uses a learning mechanism similar to the Marcet and Nicolini (2003) setup to explain the behavior of long-run inflation expectations in the United States; the authors are able to successfully explain why inflation expectations became unanchored in the 1970s. In the EGL framework of the present paper, a similar strategy is followed, but now the agents are also able to adjust the size of the gain parameter to the magnitude of the instability.

5 Evaluation of the Models and Implications for Investor Expectations

There are three aspects of investor expectations that we will analyze. First, for a fixed yield maturity, how do investors form conditional forecasts over different forecasting horizons? That is, do they hold their beliefs constant while making forecasts over the short- and medium-term, or do the beliefs depend on the forecasting horizon? Second, when the forecasting horizon is held constant, do investors keep their beliefs constant while making forecasts about the one- and ten-year yields, or are these beliefs varying? Finally, we explore the expectations formation process for real yields using data on Treasury Inflation Protected Securities (TIPS), and investigate if these are substantially different from the analogous processes for nominal yields. The results presented below will provide a framework for analyzing the beliefs of investors on these dimensions.

We first consider the performance of the different models of expectations formation for the Great Moderation period, and the analysis is later expanded to compare forecasts for the Great Recession. The models' forecasting performance is evaluated by comparing their mean square forecast errors (MSFEs), and then the implications of these results for modeling investor expectations are discussed. The sample period for nominal yields is January 1980 to December 1992. The in-sample forecasts are constructed for the one-, five- and ten-year yields, at the one-, three- and six-month horizons. These horizons are set to match (on average) the number of trading days. For example, for constructing the one-month ahead forecast, the number of days is set at 21. Before discussing the model evaluation across different time periods in section 5.2 below, we describe the mechanism used to compute the optimal gains used in the different learning mechanisms.[16]

Finally, we investigate the performance of the learning model on two other dimensions in section 6 below: we compare the out-of-sample learning forecasts relative to the random walk model[17], and analyze the learning model-implied forecast errors using the tests from section 3.1.

5.1 Determination of the Gain Parameters

In order to allow investors to update their coefficients of Ω_t, using the constant-gain algorithm described above, the initial values of the gain parameters must be set. We allow the investors to use different gains for the four latent factors[18]. Thus, the investors are no longer constrained to using the same gains for the level, slope and curvature of the yield curve. For the Great Moderation period, the sample period from January 1980 to December 1992 is used to find the optimal constant gain, as well as the parameters of the endogenous learning process, for the latent factors. These are shown in table 3[19] for the three different forecasting horizons and the 1-year yield. The gains for the 5- and 10-year yields are shown in tables 4 and 5 respectively. To analyze the implications for the Great Recession period, the baseline

[16] In section 6.2 below, we also present the out-of-sample forecasts of the learning models, relative to the random walk model.

[17] Performance relative to the Diebold-Li (2006) model is also explored in the appendix.

[18] The corresponding initial values are available upon request

[19] These values are at the lower end of the gain values used in the literature. For example, Eusepi and Preston (2013) use a gain of 0.002 in a RBC model, while Milani (2007b) estimates a gain of 0.02 using a DSGE model for the U.S. economy. However, these analyses use quarterly data, in contrast to the daily time series used here.

period used to estimate the values of the learning parameters is July 2006 to June 2009

The optimization routine minimizes the root mean squared forecasting error over the parameters of the learning processes in (8) and (10). For the constant gain algorithm, this is g_i, for $i = \{0, 1, 2, 3\}$, and for the endogenous learning algorithm, k, $\bar{g}_{lb}$ and $\bar{g}_{sf}$ for the different factors. Optimal values of the parameters are estimated for each of the three forecasting horizons (1, 3 and 6 months). To our knowledge, our paper is the first to provide estimates of the gain parameter, using macroeconomic data observed at a daily frequency and varying forecast horizons. Given the superior performance of the EGL mechanism shown in section 5.1.1 below, we concentrate on discussing the pattern in the gain parameters for this process here.

We note that at the shortest forecasting horizon (1 month), the investors adjust their gain on the level factor to pay more attention on the recent observations for the 1- and 5-year yields. Data on the slope is less heavily weighted for the 1-year and weighted more for the 5-year yield. For longer forecast horizons (3- and 6-months), the pattern is reversed for the 1-year yield. For the 5-year, the adjustment factor remains positive for these horizons. That is, investors become more attentive to the recent observations of the yield curve slope factor. The other main finding is that the gains are negligible for the level and slope factors for the 10-year yield at the 1- and 3-month forecast horizons. This implies that investors are not changing their beliefs or not accounting for structural changes while forecasting at the long end of the yield curve. For the remaining two factors, the predominant trend is that while the lower bound gain in the EGL mechanism is positive, the adjustment factor is substantially negative: that is, in periods of large deviations, the investors appear to be paying very little attention to the more recent data. This exercise suggests that monetary policy actions, which target the short or long end of the yield curve, will have asymmetric effects on the conditional yield forecasts made by market investors. If investors are not weighting the recent observations of the yield curve level and slope for constructing their forecasts of the 10-year yield, then the monetary authority will need to take this into account to determine the effects of the policy action on their long-term savings and investment decisions. As shown in the first column of 5, the constant gain algorithm will be unable to capture this dimension of investor expectations.

During the Great Recession, we find that the $\bar{g}_{lb}$ parameter of the EGL scheme is lower for the different factors at the various forecast horizons for the 1- and 5-year yields. The scaling factor, $\bar{g}_{sf}$, is also close to zero or negative to the different yield maturity and forecasting horizon pairs. These estimates suggest that during periods of high volatility, market investors pay much less attention to the recent observations, and this pattern is magnified during periods of large deviations ($\bar{g}_{sf}$ is negative). We also note that for the 1-month horizon, the endogenous gains are substantially lower than the constant gain counterparts. The gain parameters corresponding to the slope and curvature factor for the 10-year yield, however, are found to be larger than their counterparts for the Great Moderation. These findings suggest that during the Great Recession, investors were more attentive to the recent observations for the yield curve slope and curvature factors while forecasting the long-term yield. Thus, policy actions at the long end of the term structure may have been more effective in affecting investor expectations. This finding is similar to the Swanson and Williams (2014) hypothesis that medium- and long-term yields continued to respond to macroeconomic news even after the zero-lower bound was put in place.

Optimal Values of Gain Parameters								
	Great Moderation				Great Recession			
Factors	CGL	EGL			CGL	EGL		
		$\bar{g}$	$\bar{g}^{sf}$	k		$\bar{g}$	$\bar{g}^{sf}$	k
Forecasting horizon $h = 1$ month								
β_0	0.052	0.114	0.008	19	0.122	0.042	0.003	19
β_1	0.040	0.124	-0.009	19	0.112	0.020	-0.006	19
β_2	0.108	0.017	0.061	19	0.122	0.051	-0.019	19
β_3	0.108	0.242	-0.242	19	0.126	0.053	0.004	19
Forecasting horizon $h = 3$ months								
β_0	0.118	0.128	-0.010	54	0.079	0.087	-0.000	57
β_1	0.115	0.118	0.023	54	0.081	0.096	-0.006	57
β_2	0.121	0.108	0.027	54	0.078	0.076	0.010	57
β_3	0.120	0.131	-0.020	54	0.055	0.088	-0.015	57
Forecasting horizon $h = 6$ months								
β_0	0.110	0.117	-0.033	115	0.014	0.012	0.003	114
β_1	0.110	0.099	0.045	115	0.017	0.013	0.005	114
β_2	0.115	0.231	-0.231	115	0.065	0.066	-0.048	114
β_3	0.117	0.234	-0.234	115	0.012	0.018	-0.011	114

Table 3: Optimal Values of the Gain Parameter

Note: These are the optimal gain values for constant gain (CGL) and endogenous gain (EGL), for the one-year yield, for the two sample periods.

Optimal Values of Gain Parameters								
	Great Moderation				Great Recession			
Factors	CGL	EGL			CGL	EGL		
		$\bar{g}$	$\bar{g}^{sf}$	k		$\bar{g}$	$\bar{g}^{sf}$	k
Forecasting horizon $h = 1$ month								
β_0	0.039	0.110	0.010	17	0.029	0.000	0.006	20
β_1	0.075	0.111	0.020	17	0.040	0.009	-0.007	20
β_2	0.108	0.192	-0.150	17	0.009	0.000	0.020	20
β_3	0.095	0.239	-0.239	17	0.050	0.023	0.006	20
Forecasting horizon $h = 3$ months								
β_0	0.026	0.141	-0.023	59	0.000	0.000	0.000	52
β_1	0.058	0.119	0.027	59	0.002	0.036	0.004	52
β_2	0.119	0.120	0.005	59	0.002	0.083	0.016	52
β_3	0.118	0.125	-0.005	59	0.003	0.008	0.006	52
Forecasting horizon $h = 6$ months								
β_0	0.083	0.094	-0.022	119	0.000	0.000	0.005	124
β_1	0.122	0.117	0.005	119	0.003	0.017	-0.006	124
β_2	0.110	0.131	-0.115	119	0.002	0.000	0.004	124
β_3	0.122	0.237	-0.235	119	0.010	0.009	0.009	124

Table 4: Optimal Values of the Gain Parameter

Note: These are the optimal gain values for constant gain (CGL) and endogenous gain (EGL), for the five-year yield, for the two sample periods.

Optimal Values of Gain Parameters								
	Great Moderation				Great Recession			
Factors	CGL	EGL			CGL	EGL		
		$\bar{g}$	$\bar{g}^{sf}$	k		$\bar{g}$	$\bar{g}^{sf}$	k
Forecasting horizon $h = 1$ month								
β_0	0.000	0.000	0.000	7	0.000	0.000	0.008	10
β_1	0.004	0.000	0.018	7	0.189	0.219	-0.024	10
β_2	0.108	0.000	0.046	7	0.004	0.185	-0.015	10
β_3	0.095	0.213	-0.200	7	0.029	0.166	0.022	10
Forecasting horizon $h = 3$ months								
β_0	0.000	0.000	0.000	62	0.000	0.000	0.005	59
β_1	0.003	0.000	0.008	62	0.004	0.087	-0.001	59
β_2	0.119	0.055	0.139	62	0.003	0.086	0.017	59
β_3	0.118	0.176	-0.106	62	0.060	0.005	0.006	59
Forecasting horizon $h = 6$ months								
β_0	0.000	0.000	0.000	111	0.000	0.000	0.006	112
β_1	0.122	0.117	0.005	111	0.004	0.019	-0.006	112
β_2	0.108	0.000	0.029	111	0.003	0.089	0.043	112
β_3	0.122	0.238	-0.238	111	0.024	0.036	0.018	112

Table 5: Optimal Values of the Gain Parameter

Note: These are the optimal gain values for constant gain (CGL) and endogenous gain (EGL), for the ten-year yield, for the two sample periods.

5.2 Model Evaluation

We examine the evolution of expectations during the Great Moderation and the Great Recession, as well as analyze the implications of using yields from Treasury Inflation Protected Securities.

5.2.1 Investor Expectations during the Great Moderation

A comparison of the two learning models, on the basis of the MSFEs derived from the in-sample conditional yield forecasts is presented in table 6. The statistical significance of the difference in forecasts is constructed using the Diebold-Mariano (1995) statistic. The null hypothesis of the test statistic is that there is no difference in the forecast accuracy of the two competing models; a rejection of the null implies that the two forecasting models have statistically different forecasting performances. The dominant trend evident from table 6 is that the MSFEs from the endogenous learning model are lower than those derived from constant gain at all forecasting horizons and yield maturities. This indicates that the market investors are, in fact, responding to deviations in the data, and adjusting the weights placed on the more recent observations. We also find that the level of the MSFEs is the lowest for the long-term yield (10 years), across the forecasting horizons. The largest gains in forecasting performance occurs for the 1-year yield. We also note that both the learning models improve upon the constant-coefficients model, and the relative MSFEs are shown in the appendix.

In our view, the above results suggest the following implications. First, incorporating time-variation in the formation of investors' conditional forecasts leads to significant forecasting improvements. These results are robust across forecasting horizons, as well as yield maturities. Second, a large literature has used constant gain learning to model investor beliefs in theoretical frameworks. This framework may not be able to capture the belief formation process adequately, even during the Great Moderation. Adopting the endogenous learning algorithms proposed above provides an intuitive manner to model investor beliefs during periods of low volatility, as well as of high macroeconomic volatility (as discussed for the Great Recession below).

5.2.2 Investor Expectations during the Great Recession

As before, the MSFE is used to compare the forecasting performance across different models. The results for the different models are presented in the third and fourth columns of table 6. Unlike the Great Moderation period, we find that the substantive improvements even at the 10-year yield across the forecasting horizons. The other main observation is that the MSFEs are smaller for the Great Recession period, which we attribute to the smaller data sample period. Even for the periods of higher macroeconomic volatility, the constant gain learning approach is unable to capture the shifts in beliefs as done by the endogenous learning mechanism. The analysis of monetary policy actions through the lens of these CGL frameworks, may therefore, be an incomplete representation of investors' conditional forecasts.

Yield	Great Moderation		Great Recession	
Maturity	MSFE-CGL	MSFE-EGL	MSFE-CGL	MSFE-EGL
		Forecasting horizon $h = 1$ month		
1 year	5.762	4.235*	0.438	0.401
5 years	3.913	3.031*	1.143	0.755*
10 years	2.987	2.634	2.329	1.836*
		Forecasting horizon $h = 3$ months		
1 year	6.499	5.982	0.445	0.392*
5 years	4.130	3.731*	0.941	0.688*
10 years	2.615	2.405	2.205	1.615*
		Forecasting horizon $h = 6$ months		
1 year	6.427	5.456*	0.284	0.253
5 years	3.851	3.628	0.884	0.796*
10 years	2.407	2.280*	2.200	1.836*

Table 6: Evaluating the Conditional Forecasts

Note: These are the mean square forecast error (MSFE) values for constant gain (CGL) and endogenous gain learning (EGL) models, at the three forecasting horizons. The starred values here show the EGL forecasts that are statistically superior to CGL, as implied by the Diebold-Mariano test. The null hypothesis of the test is that the forecasts are statistically indistinguishable.

5.2.3 Investor Expectations from TIPS Yields

We also use data from Treasury Inflation Protected Securities (TIPS) to estimate the learning parameters for real yields. The strategy for estimating the parameters mechanisms for these TIPS yields is the same as followed above. We use the factors estimated by Gürkaynak, Sack and Wright (2010) However, only the 5- and 10-year yield data is used in the estimation exercise. Given the shorter data sample for TIPS, the estimates are presented for the sample period August 24, 2004 to June 30, 2008. The comparative forecast results are presented in table 9 (the Diebold-Mariano test statistics are used to indicate statistically superior forecasts), and the corresponding optimal gains are shown in tables 7 and 8. As for the nominal yields, the MSFEs suggest that the endogenous learning mechanism generates substantial improvements in the conditional forecasts, relative to the constant gain process. The optimal gains for the endogenous learning scheme suggest that investors are revising their expectations about the level and slope factors of the 5-year TIPS yields much less than the nominal 5-year counterparts for the Great Moderation period. On the other hand, for the 10-year conditional forecasts, both the TIPS and nominal yields suggest that investors are taking new information into account very slowly.

Factors	Optimal Values of Gain Parameters			
	CGL	EGL		
		$\bar{g}$	$\bar{g}^{sf}$	k
	Forecasting horizon $h = 1$ month			
β_0	0.071	0.005	0.004	17
β_1	0.134	0.138	0.004	17
β_2	0.103	0.023	-0.020	17
β_3	0.094	0.010	0.006	17
	Forecasting horizon $h = 3$ months			
β_0	0.017	0.069	0.011	51
β_1	0.039	0.097	0.001	51
β_2	0.014	0.089	0.010	51
β_3	0.022	0.094	-0.022	51
	Forecasting horizon $h = 6$ months			
β_0	0.062	0.000	0.006	123
β_1	0.103	0.094	0.036	123
β_2	0.049	0.008	-0.008	123
β_3	0.086	0.010	0.011	123

Table 7: Optimal Values of the Gain Parameter for TIPS Yields

Note: These are the optimal gain values for constant gain (CGL) and endogenous gain (EGL), for the TIPS five-year yields, for the 2004-2008 sample period.

Factors	Optimal Values of Gain Parameters			
	CGL	EGL		
		$\bar{g}$	$\bar{g}^{sf}$	k
	Forecasting horizon $h = 1$ month			
β_0	0.130	0.000	0.006	13
β_1	0.131	0.106	0.011	13
β_2	0.119	0.027	-0.021	13
β_3	0.114	0.001	0.014	13
	Forecasting horizon $h = 3$ months			
β_0	0.000	0.000	0.004	52
β_1	0.002	0.092	0.013	52
β_2	0.001	0.110	0.018	52
β_3	0.039	0.002	0.012	52
	Forecasting horizon $h = 6$ months			
β_0	0.000	0.000	0.006	125
β_1	0.004	0.161	-0.011	125
β_2	0.002	0.139	0.028	125
β_3	0.033	0.098	0.036	125

Table 8: Optimal Values of the Gain Parameter

Note: These are the optimal gain values for constant gain (CGL) and endogenous gain with (EGL), for the TIPS ten-year yields, for the 2004-2008 sample period.

Yield	MSFE	
Maturity	MSFE-CGL	MSFE-EGL
	Forecasting horizon $h = 1$ month	
5 years	0.958	0.862
10 years	2.525	2.092*
	Forecasting horizon $h = 3$ months	
5 years	0.967	0.960
10 years	2.234	2.128
	Forecasting horizon $h = 6$ months	
5 years	0.851	0.808*
10 years	1.734	1.206*

Table 9: Evaluating the Conditional Forecasts for TIPS Yields

Note: These are the mean square forecast error (MSFE) values for constant gain (CGL) and endogenous gain learning (EGL) models, at the three forecasting horizons, for TIPS yields. The starred values here show the EGL forecasts that are statistically superior to CGL, as implied by the Diebold-Mariano test. The null hypothesis of the test is that the forecasts are statistically indistinguishable.

6 Further Tests of the Learning Models

In the following section, we use two other metrics to evaluate the learning models. We first investigate the forecast errors implied by the CGL and EGL models with respect to the rationality tests constructed in section 3.1 above. Next, we use the random walk model to construct out-of-sample forecasts, and compare these with the equivalent forecasts from the learning framework. In the accompanying appendix, we also compare the performance of the learning models with the Diebold and Li (2006) model[20].

[20] For this exercise, we use the Fama-Bliss monthly yields for estimating the three-factor version of the Nelson-Siegel (1987) yield curve model. The out-of-sample learning forecasts are then constructed, and

6.1 Forecast Errors from the Learning Models

In section 3.1 above, we examined the performance of the forecast errors obtained from the constant-coefficients model, with respect to a series of rationality tests. In this section, we consider how the forecast errors obtained from the learning models perform in these tests[21]. The results from the constant gain and endogenous gain learning models are shown in tables 10 and 11, and we discuss the test-wise results below.

For the first test, whether the forecast errors are unbiased, both the learning algorithms outperform the constant-coefficients model. The coefficients for the intercept parameter are either not significant or the size of the coefficients is smaller. The exception here is the 1- and 3-month ahead forecasts for the 5-year yield.

The second test measures whether the forecast errors are efficient. In terms of the slope coefficient, that is, whether the forecast errors are predicted by the forecasts themselves, the learning models also do better. For the 1- and 5-year yields, this is evident across the three forecasting horizons. For the 10-year yield, the improvement is evident at the 1-month horizon; the size of the slope coefficients is similar to the constant coefficients model at the 3- and 6-month horizon (for both constant and endogenous learning models).

Finally, the third test measures whether the forecast errors are systematic. For the 1-year yield, the slope coefficients of both the learning models are statistically insignificant, or smaller than the constant coefficients model. For the 5-year yield, learning models outperform the model at the 1-month horizon. For the 10-year yield, a similar pattern in observed in the intercept coefficients (for all forecasting horizons), and the slope coefficients for the 1-month forecasting horizon. These results suggest that the forecast errors from the learning models significantly improve upon the constant-coefficients model.

---

compared with those implied by the Diebold-Li model.

[21] These are the in-sample forecast errors for the learning model, based on the gains estimated for the Great Moderation period.

Yield Maturity	$h = 1$ month α	β	$h = 3$ months α	β	$h = 6$ months α	β
	Test 1: $y_{t+h} - E_t y_{t+h} = \alpha + error_t$					
1 year	-1.5849* (0.12)	-	-1.6938* (0.12)	-	-1.6316* (0.17)	-
5 years	-0.7037* (0.21)	-	-0.6571* (0.19)	-	-0.7573* (0.22)	-
10 years	0.0437 (0.28)	-	-0.1691 (0.23)	-	0.0733 (0.04)	-
	Test 2: $y_{t+h} - E_t y_{t+h} = \alpha + \beta E_t y_{t+h} + error_t$					
1 year	-0.5540 (0.92)	-0.1230 (0.10)	-0.8119 (1.01)	-0.1018 (0.11)	2.3918 (0.98)	-0.4716* (0.11)
5 years	3.7053* (0.43)	-0.7335* (0.04)	3.3985* (0.45)	-0.6985* (0.05)	3.6435* (0.46)	-0.7257* (0.05)
10 years	4.3957* (0.30)	-0.8675* (0.03)	4.6441* (0.29)	-0.7803* (0.03)	3.4918* (0.29)	-0.8784* (0.03)
	Test 3: $y_{t+h} - E_t y_{t+h} = \alpha + \beta \left(E_t y_{t+h} - E_{t-1} y_{t+h}\right) + error_t$					
1 year	-1.5977* (0.12)	-0.3463 (0.16)	-1.6999* (0.12)	-0.3902 (0.18)	-1.6964* (0.13)	-0.4478 (0.11)
5 years	-0.5283* (0.08)	-0.3768* (0.12)	-0.5546* (0.17)	-0.3740* (0.11)	-0.5764* (0.17)	-0.4730* (0.11)
10 years	0.0893 (0.26)	-0.4454* (0.12)	-0.1686 (0.22)	-0.2528 (0.03)	0.1414 (0.26)	-0.5189* (0.11)

Table 10: Testing Forecast Errors for Nominal Yield Curve Factors

Note: The above coefficient estimates are reported for the CGL yields, at the three forecasting horizons. The standard errors are shown for the corresponding coefficients in brackets.

Yield Maturity	$h=1$ month		$h=3$ months		$h=6$ months	
	α	β	α	β	α	β
	Test 1: $y_{t+h} - E_t y_{t+h} = \alpha + error_t$					
1 year	-1.6505*	-	-1.6501*	-	-1.7223*	-
	(0.11)		(0.12)		(0.15)	
5 years	-0.6424*	-	-0.6318*	-	-0.6802*	-
	(0.20)		(0.18)		(0.21)	
10 years	0.1095	-	-0.1695	-	0.1171	-
	(0.27)		(0.22)		(0.28)	
	Test 2: $y_{t+h} - E_t y_{t+h} = \alpha + \beta E_t y_{t+h} + error_t$					
1 year	-1.5699	-0.0093	-1.2729	-0.0395	2.0900	-0.4387*
	(0.98)	(0.11)	(1.01)	(0.11)	(0.86)	(0.09)
5 years	3.5313*	-0.7119*	3.4092*	-0.6996*	3.4142*	-0.6962*
	(0.42)	(0.04)	(0.44)	(0.04)	(0.43)	(0.04)
10 years	4.3677*	-0.8631*	4.7532*	-0.7855*	4.4611*	-0.8741*
	(0.30)	(0.03)	(0.30)	(0.03)	(0.29)	(0.03)
	Test 3: $y_{t+h} - E_t y_{t+h} = \alpha + \beta\left(E_t y_{t+h} - E_{t-1} y_{t+h}\right) + error_t$					
1 year	-1.6571*	-0.4194	-1.6222	-0.3989*	-1.7339*	-0.4966*
	(0.11)	(0.17)	(0.12)	(0.14)	(0.13)	(0.09)
5 years	-0.5357*	-0.4330*	-0.5935*	-0.4232*	-0.4948*	-0.4695*
	(0.17)	(0.10)	(0.17)	(0.11)	(0.16)	(0.09)
10 years	0.1558	-0.4542*	0.2231	-0.3761*	0.1765	-0.4922*
	(0.26)	(0.12)	(0.21)	(0.13)	(0.26)	(0.10)

Table 11: Testing Forecast Errors for Nominal Yield Curve Factors

Note: The above coefficient estimates are reported for the EGL yields, at the three forecasting horizons. The standard errors are shown for the corresponding coefficients in brackets.

6.2 Forecast Performance relative to the Random Walk Model

We compare the out-of-sample forecasting performance of the learning models with the random walk model. As daily yields are very persistent, the random walk has been shown to be difficult to beat in out-of-sample forecasts. We investigate the out-of-sample forecasts for the Great Moderation period below.

The random walk model implies a "no-change" forecast for the individual yields. In this case, the $h-$day ahead forecast of the $n-$period yield in time t is simply the time t observation. For the learning models during the Great Moderation period, the sample period from January 1980 to December 1992 is used to find the optimal constant gain, as well as the parameters of the endogenous learning process as noted in the estimation section 5.1 above. The out-of-sample forecasts are constructed for the yield maturities at the different forecasting horizons for the period January to December 1993. Table 12 presents the MSFEs of the learning models, as a fraction of the MSFE of the random walk model. In this table, the Diebold-Mariano test is used to compare the forecasting performance of the learning models relative to the random walk. When the ratio of the MSFEs is smaller than one, then the respective learning model predicts better out-of-sample forecasts than the random walk, and the starred values indicate that this improvement is statistically significant according to the Diebold-Mariano test.

The results indicate that at the 1-month forecasting horizon, there are no significant improvements in forecasting performance of the learning models, relative to the random walk model. The dominance of the random walk model at the short forecasting horizon has been repeatedly found in the literature (some examples include Moench (2008), Chen and Niu (2014) and Xiang and Zhu (2013)). At the longer forecasting horizons, however, the learning models outperform the random walk model in a statistically significant manner.

Yield	Great Moderation	
Maturity	MSFE CGL/RW	MSFE EGL/RW
	Forecasting horizon $h = 1$ month	
1 year	1.3525	1.3196
5 years	1.0091	1.0079
10 years	1.0147	1.0799
	Forecasting horizon $h = 3$ months	
1 year	1.0256	0.9443*
5 years	0.9377*	0.9285*
10 years	0.9491	1.0435
	Forecasting horizon $h = 6$ months	
1 year	1.2500	0.9401*
5 years	0.8911*	0.9374*
10 years	0.9175*	0.8764*

Table 12: Evaluating the Conditional Forecasts

Note: These are the ratios of the mean square forecast error (MSFE) values for constant gain and endogenous learning models relative to the random walk model, at the three forecasting horizons. A value less than one indicates that the learning model performs better. Statistical significance of an improved forecast is measured using the Diebold-Mariano test. The starred values indicate significance at the 10% level.

7 Applying the Learning Mechanisms

Given the above findings on the superior performance of the endogenous learning mechanism, we investigate the implication of this process for forming conditional forecasts using inflation expectations and survey data on expected excess returns.

7.1 Inflation Expectations

Given our estimation of the conditional forecasts for the nominal and TIPS yields, we further examine the inflation expectations implied by the difference of the conditional forecasts between the nominal and TIPS yields. The evolution of inflation expectations for the 5-year and 10-year horizon are presented in figures 1 and 2.

The 5- and 10-year inflation forecasts implied by the endogenous learning mechanism show several interesting features; these are reported for the three different forecast horizons considered. From the 5-year figures, we note that until the end of 2006, the inflation forecasts for the different forecasting horizons move in tandem. However, starting in early 2007, the 3- and 6-month forecasts begin to diverge from the 1-month forecasts. By the end of 2007, these begin to move together again, but are marked by a sharp increase in the volatility. This increase in volatility is also mirrored in the 10-year inflation forecasts. These results suggest that before the start of the financial crisis, the 5- and 10-year inflation forecasts of market investors moved together across different forecast horizons, and showed very little variance. However, as the financial upheaval began to take root in early 2007, investors became much more uncertain about inflation in the longer horizons (3- and 6-month). During the first period of the crisis (December 2007 to June 2008), the increased market uncertainty is reflected in the implied inflation forecasts.

Given these inflation expectations from the endogenous learning model, we further investigate the correlation of the forecasts with survey expectations. We consider the inflation expectations derived from the Federal Reserve Bank of Cleveland model of inflation expectations and the Survey of Professional Forecasters. The Cleveland Fed model is an affine model of the nominal and real term structures, which uses data on inflation swap rates, nominal Treasury yields, and survey forecasts of inflation for estimating the model parameters (Haubrich, Pennacchi and Ritchken, 2012). The implied 5- and 10-year inflation forecasts are based on monthly data. For the August 2004 - June 2008 period, the correlation between the 5-year Cleveland Fed inflation forecasts and those generated by the endogenous learning model is 0.37 and the correlation with the 10-year inflation forecasts is 0.28.[22]

The Survey of Professional Forecasters reports the forecasts of the annual average rate of headline CPI inflation for 10 years (the year in which the survey is conducted is also

[22] For the purpose of the computing the correlations, the end of month endogenous learning inflation forecasts are considered.

included). This estimate of long-term inflation expectations has been reported since the last quarter of 1991. The Survey also began to report 5-year inflation expectations starting in the third quarter of 2005. As the SPF inflation forecasts are reported every quarter, we only use the correlation with the 10-year forecasts for the August 2004-June 2008 period. The correlation for this relatively short time sample is 0.17.

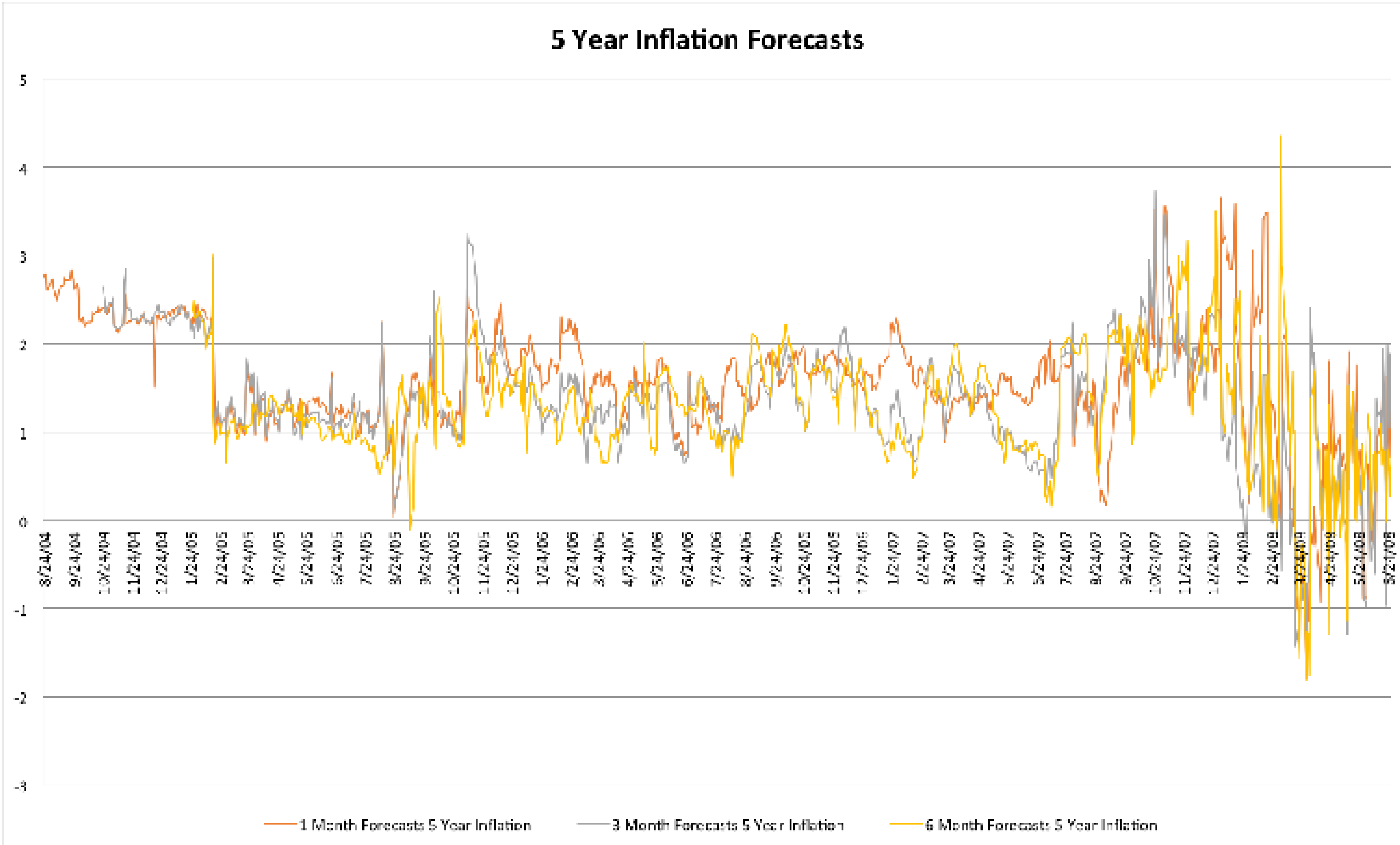

Figure 1: Evolution of 5-year Inflation Expectations

Note: This figure shows the 5-year inflation expectations derived from the endogenous learning (EGL) process at various forecast horizons.

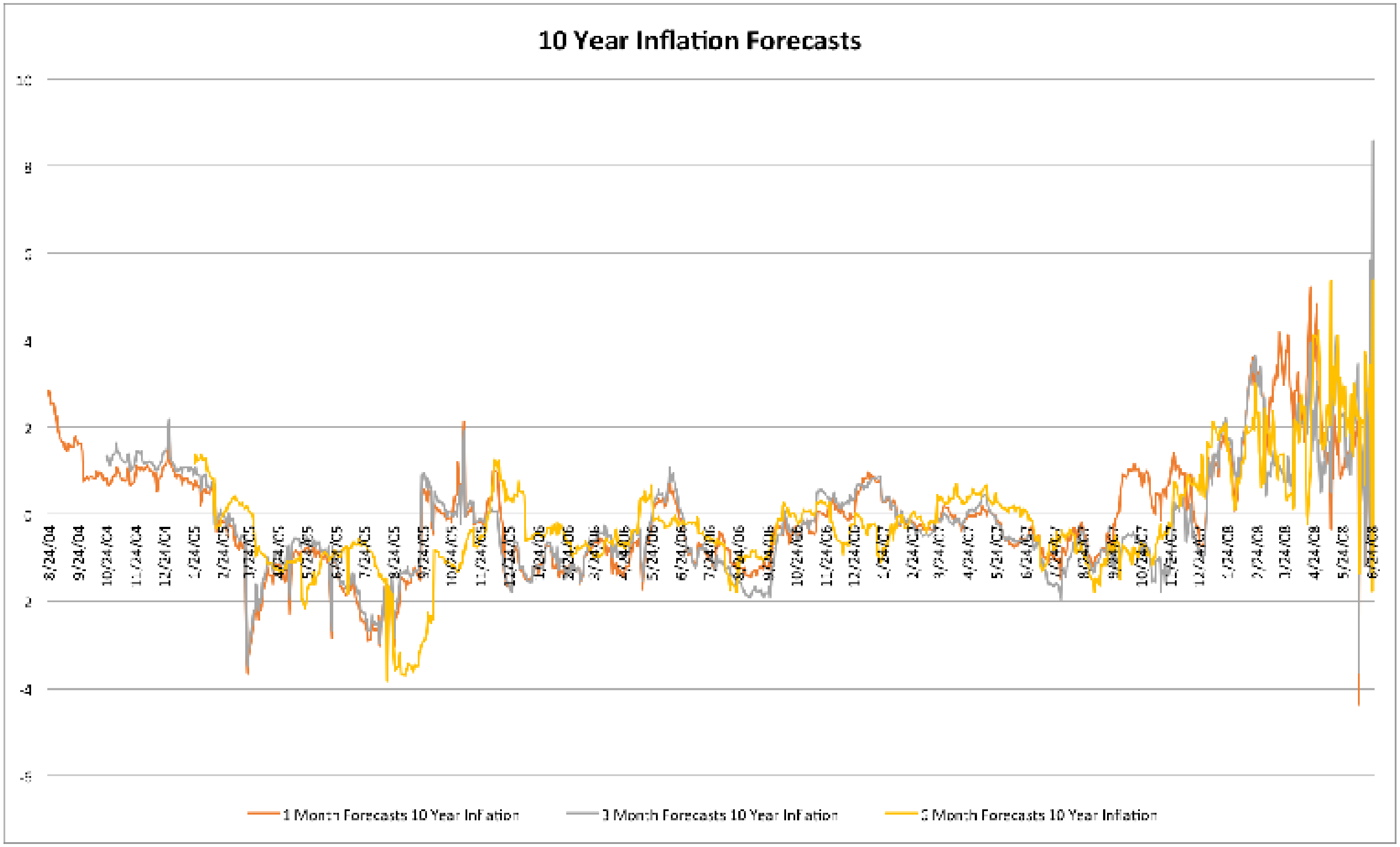

Figure 2: Evolution of 10-year Inflation Expectations

Note: This figure shows the 10-year inflation expectations derived from the endogenous learning (EGL) process at various forecast horizons.

7.2 Explaining Expected Excess Returns

Predictable patterns in excess returns for nominal yields in the U.S. have been well documented. Piazzesi, Salomao and Schneider (2015) and Dick, Schmeling and Schrimpf (2013) document the patterns in excess returns using survey data; the patterns are compared with those generated by an affine factor model in the former approach. In this section, we first use data from the Survey of Professional Forecasters to estimate the expected excess returns for the ten-year yield. We compare the implications of analogous expected excess returns the learning models with the survey returns. We also show the correlations of the learning model excess returns with the excess returns computed by Cochrane and Piazzesi (2005) to provide another point of comparison.

The excess return for yield maturity n, at time t, for horizon h is given by:

$$E_t\left[rx_{t,t+h}^{(n+h)}\right] = E_t\left[p_{t+h}^{(n)}\right] - p_t^{(n+h)} - y_t^h, \tag{11}$$

here p_t^n is the price of the zero-coupon security at time t of maturity n quarters. Then, in terms of the yields:

$$E_t\left[rx_{t,t+h}^{(n+h)}\right] = -nE_t\left[y_{t+h}^{(n)}\right] + (n+h)\, y_t^{(n+h)} - y_t^{(h)}. \tag{12}$$

With SPF data, we can compute the following, for $n = 40$ quarters ($10 - year$ yield) and $h = 4$ (1-year ahead forecasts)

$$\begin{aligned} E_t\left[rx_{t,t+h}^{(n+h)}\right] &= -nE_t\left[y_{t+h}^{(n)}\right] + (n+h)\, y_t^{(n+h)} - y_t^{(h)} \\ &= -40E_t\left[y_{t+4}^{(40)}\right] + (40+4)\, y_t^{(40+4)} - y_t^{(4)}. \end{aligned} \tag{13}$$

In order to compute excess returns for the ten-year yield for forecasting horizons h, for the yields $y_t^{(40+4)}$ and $y_t^{(4)}$, we use the eleven-year and one-year yield from the Gürkaynak, Sack and Wright (2007) data. $E_t\left[y_{t+4}^{(10)}\right]$ is the SPF expected value of the 10-year yield at the 1-year horizon. The same methodology is used to construct expected excess returns from the learning models; in this case, the $E_t\left[y_{t+4}^{(10)}\right]$ is computed using the conditional forecasts described above. In this section, we generate expected excess returns from survey data and the theoretical models for the period 1993 to 2008. For the learning models, the gain parameters are set using the optimal gains derived for the Great Moderation period (we use the gains shown in table 3).

The evolution of 10-year expected excess returns for the constant and endogenous gain algorithms are shown in figure 3. For the common sample period, we find that the correlation of the CGL excess returns, with the SPF excess returns are 0.08. This correlation is 0.18 for the EGL excess return. Thus, the endogenous learning algorithm approximates investor expected excess returns more closely than the constant-gain learning algorithm.

To provide another benchmark comparison, we also compare the expected excess returns implied by the learning models with the expected excess returns of Cochrane and Piazzesi (2005). In this study, the authors find that expected excess returns are time-varying and the annual returns on 1- to 5-year bonds are well explained by a tent-shaped factor, which is a linear combination of forward rates. This tent factor is also found to be counter-cyclical. We construct the 5-year expected excess returns from the learning models, at the comparable horizon[23], and examine the correlation with the 5-year expected excess returns of Cochrane and Piazzesi (2005). The correlation of the Cochrane and Piazzesi (2005) expected excess return with the CGL and EGL return is 0.23 and 0.33 respectively. The learning algorithms also show similar counter-cyclicality as the Cochrane and Piazzesi (2005) returns: for 1993-2003, the correlation of the Cochrane-Piazzesi 5-year expected excess returns with the cyclical component of real GDP[24] is -0.44; for the CGL and EGL returns, this is -0.43 and -0.41 respectively. Piazzesi, Schneider and Salomao (2015) also document that the expected excess returns computed from a constant-coefficients VAR model, and subjective expected excess returns (derived from the Blue Chip and Goldsmith-Nagan survey forecasts) show similar countercyclical behavior.

[23] Cochrane and Piazzesi (2005) construct expected excess returns using data from 1965 to 2003 (although their dataset starts in 1952, they consider the later start date due to unreliability of initial data). To compare with the learning excess returns, we present correlations between 1993 and 2003.

[24] The real GDP cyclical component is derived using the Hodrick-Prescott filter.

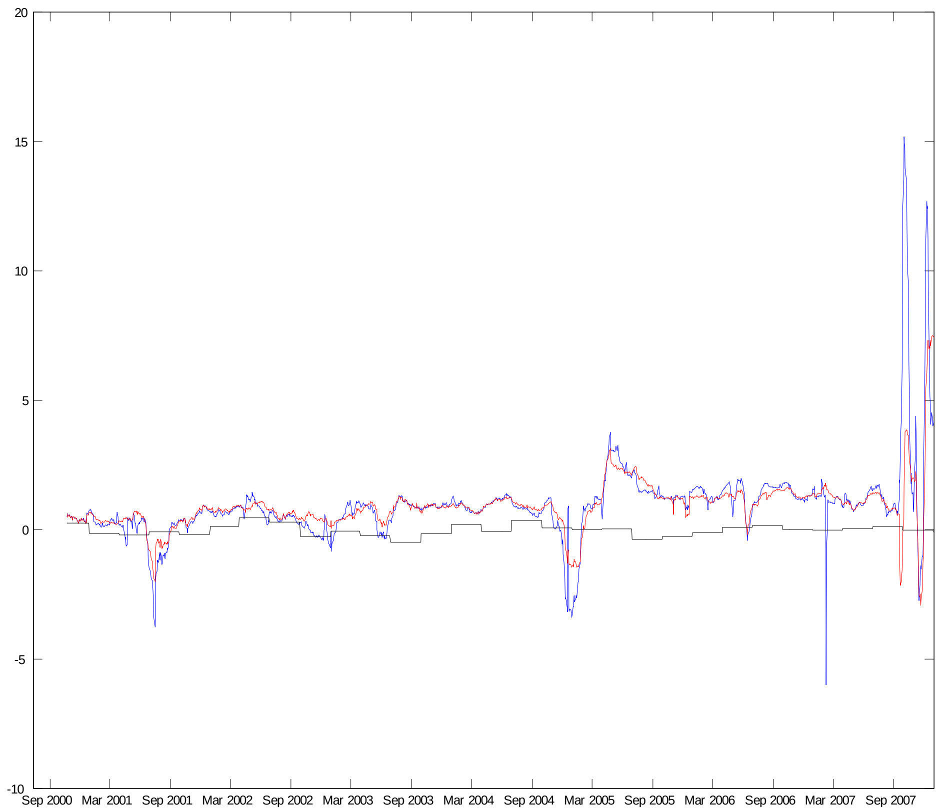

Figure 3: Evolution of Expected Excess Returns

Note: This figure shows the evolution of expected excess returns for the ten-year yield, derived from the survey data (from the Survey of Professional Forecasters) and the learning models. The SPF returns are shown by the black line, the CGL returns by the blue line, and the EGL returns by the red line.

8 Conclusion

The empirical analysis conducted above provides macroeconomists and financial economists a glimpse at how subjective expectations might evolve over time. The specific application to yield curves, contributes to discussions about central bank policy. Central bankers try to influence the economy using the short-term yields. Whether the transmission mechanism (to the long end of the curve) occurs as posited by bankers is still a matter of debate. While constructing forecasts, if expectations of investors about future short yields are not rational, and are more persistent than policy makers expect them to be, then long yields may not move as much as anticipated. The above analysis attempts to show that forecasting using the Nelson-Siegel-Svensson model of the yield curve can be improved upon by allowing for a process for factor evolution that incorporates time-varying parameters, instead of a constant-coefficients VAR model. Two alternative models of expectations formation are suggested here, the constant-gain learning process and endogenous gain, and these are found to improve upon the forecasting performance relative to the constant-coefficients model. The improvements in forecasting occurs during periods of low volatility, as well as during the financial crisis period.

References

[1] Aruoba, B., F.X. Diebold and G.D. Rudebusch. 2006. "The Macroeconomy and the Yield Curve: A Dynamic Latent Factor Approach." *Journal of Econometrics,* 131: 309-338.

[2] Bianchi, F., H. Mumtaz and P. Surico. 2009. "The Great Moderation of the Term Structure of U.K. Interest Rates." *Journal of Monetary Economics,* 56: 856-71.

[3] Carvalho, C., S. Eusepi, B. Preston and E. Moench. 2015. "In Search of a Nominal Anchor: What Drives Long-Term Inflation Expectations?" Working paper.

[4] Chen, Y. and L. Niu. 2014. "Adaptive dynamic Nelson–Siegel term structure model with applications." *Journal of Econometrics,* 180(1): 98-115.

[5] Christensen, J.H.E., F.X. Diebold and G.D. Rudebusch. 2011. "The Affine Arbitrage-Free Class of Nelson-Siegel Term Structure Models." *Journal of Econometrics,* 164: 4-20.

[6] Cochrane, J. H., and M. Piazzesi. 2005. "Bond Risk Premia." *American Economic Review*, 95(1): 138-160.

[7] Coibion, O., and Y. Gorodnichenko. 2015. "Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts." *American Economic Review,* 105: 2644–2678.

[8] Dewachter, H., and M. Lyrio. 2006. "Macro Factors and the Term Structure of Interest Rates." *Journal of Money, Credit and Banking,* 38: 119-140.

[9] Diebold, F.X. and C. Li. 2006. "Forecasting the Term Structure of Government Bond Yields." *Journal of Econometrics,* 130: 337-364.

[10] Diebold, F.X. and R. S. Mariano. 1995. "Comparing Predictive Accuracy." *Journal of Business and Economic Statistics*, 13: 253-265.

[11] Dick, C. D., M. Schmeling and A. Schrimpf. 2013. "Macro-expectations, aggregate uncertainty, and expected term premia." *European Economic Review*, 58(C): 58-80.

[12] Duffee, G.R. 2011. "Forecasting with the term structure: The role of no-arbitrage restrictions." Working paper, Johns Hopkins University.

[13] Eusepi, S. and B. Preston. 2011. "Expectations, Learning and Business Cycle Fluctuations." *American Economic Review,* 101(6): 2844-72.

[14] Fuhrer, J. 1996. "Monetary Policy Shifts and Long-Term Interest Rates." *The Quarterly Journal of Economics,* 111(4): 1183-1209.

[15] Gaus, E. 2014. "Time-Varying Parameters and Endogenous Learning Algorithms." Working paper, Ursinus College.

[16] Giacoletti, M., K. Laursen and K.J. Singleton. 2014. "Learning the Macro-Dynamics of U.S. Treasury Yields with Aribitrage-free Term Structure Models." Working paper, Stanford University Graduate School of Business.

[17] Gürkaynak, R.S., B. Sack and J.H. Wright. 2007. "The U.S. Treasury Yield Curve: 1961 to the Present." *Journal of Monetary Economics,* 54: 2291-2304.

[18] Gürkaynak, R.S., B. Sack and J.H. Wright. 2010. "The TIPS Yield Curve and Inflation Compensation." *American Economic Journal: Macroeconomics*, 2(1): 70-92

[19] Haubrich, J.G., G. Pennacchi and P. Ritchken. 2012. "Inflation Expectations, Real Rates and Risk Premia: Evidence from Inflation Swaps." *Review of Financial Studies,* 25(5): 1588-1629.

[20] Kostyshyna, O. 2013. "Application of an Adaptive Step-Size Algorithm in Models of Hyperinflation." *Macroeconomic Dynamics,* 16: 355–375.

[21] Kozicki, S. and P. A. Tinsley. 2001. "Shifting Endpoints in the Term Structure of Interest Rates." *Journal of Monetary Economics,* 47(3): 613-652.

[22] Kreps, D. 1998. "Anticipated Utility and Dynamic Choice." *1997 Schwartz Lecture. In: D.P. Jacobs, E. Kalai and M. Kamien, Editors, Frontiers of Research in Economic Theory,* Cambridge University Press, Cambridge, England.

[23] Laubach, T., R. J. Tetlow and J. C. Williams. 2007. "Learning and the Role of Macroeconomic Factors in the Term Structure of Interest Rates". Working paper.

[24] Mankiw, N.G., J. Wolfers and R. Reis. 2004. "Disagreement about Inflation Expectations," *NBER Macroeconomics Annual 2003*: 209-248

[25] Marcet, A. and J.P. Nicolini. 2003. "Recurrent Hyperinflations and Learning." *American Economics Review,* 93(5): 1476-1498.

[26] Milani, F. 2007a. "Learning and Time-Varying Macroeconomic Volatility", Working paper.

[27] Milani, F. 2007b. "Expectations, Learning and Macroeconomic Persistence", *Journal of Monetary Economics,* 54: 2065-2082.

[28] Moench, E. 2008. "Forecasting the yield curve in a data-rich environment: A no-arbitrage factor-augmented VAR approach." *Journal of Econometrics*, 146(1): 26-43.

[29] Nelson, C.R. and A.F. Siegel. 1987. "Parsimonious Modeling of Yield Curve." *Journal of Business,* 60: 473-489.

[30] Piazzesi, M., J. Salomao and M. Schneider. 2015. "Trend and Cycle in Bond Premia", Working paper.

[31] Pooter, M. de. 2007. "Examining the Nelson-Siegel Class of Term Structure Models," *Discussion Papers 07-043/4.* Tinbergen Institute.

[32] Rudebusch, G.D. and T. Wu. 2007. "Accounting for a Shift in Term Structure Behavior with No-Arbitrage and Macro-Finance Models", *Journal of Money, Credit and Banking,* 39: 395-422.

[33] Sinha, A. 2016. "Learning and the Yield Curve." *Journal of Money, Credit and Banking,* 48(2-3): 513-547.

[34] Svensson, L. 1994. “Estimating and Interpreting Forward Rates: Sweden 1992-4.” *National Bureau of Economic Research Working Paper 4871.*

[35] Thomas, L. 1999. "Survey Measures of Expected U.S. Inflation." *Journal of Economic Perspectives,* 13(4): 125-144.

[36] van Dijk D., S. J. Koopman, M. van der Wel and J.H. Wright. 2014. "Forecasting interest rates with shifting endpoints." *Journal of Applied Econometrics*, 29: 693–712.

[37] Xiang, J. and X. Zhu. 2013. "A Regime-Switching Nelson–Siegel Term Structure Model and Interest Rate Forecasts." *Journal of Financial Econometrics,* 11(3): 522-555.