What does the Yield Curve imply about Investor Expectations?

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

March 2014

Abstract

We find that investors' expectations of U.S. nominal yields, at different maturities and forecast horizons, exhibit significant time-variation during the Great Moderation. Nominal zero-coupon bond yields for the U.S. are used to fit the yield curve using a latent factor model. In the benchmark model, the VAR process used to characterize the conditional forecasts of yields has constant coefficients. The alternative class of models assume that investors use adaptive learning, in the form of a constant gain algorithm and different endogenous gain algorithms, which we propose here. Our results indicate that incorporating time-varying coefficients in the conditional forecasts of yields lead to large improvements in forecasting performance, at different maturities and horizons. These improvements are even more substantial during the Great Recession. We conclude that our results provide strong empirical motivation to use the class of adaptive learning models considered here, for modeling potential investor expectation formation in periods of low and high volatility, and the endogenous learning model leads to significant improvements over the benchmark in periods of high volatility. A policy experiment, which simulates a surprise shock to the level of the yield curve, illustrates that the conditional forecasts of yields implied by the learning models do significantly better at capturing the response observed in the realized yield curve, relative to the constant-coefficients model. Furthermore, the endogenous learning algorithm does well at matching the time-series patterns observed in expected excess returns implied by the Survey of Professional Forecasters.

JEL classifications: E43, E47, D83, C5

Keywords: Adaptive learning, Investor beliefs, Monetary policy, Excess returns

---

[1] Department of Business and Economics, Ursinus College, 601 East Main Street, Collegeville, PA 19426. E-mail: egaus@ursinus.edu.

[2] Department of Economics, Leavey School of Business, Santa Clara University, 500 El Camino Real, Santa Clara, CA 95053. Email: asinha1@scu.edu. Support from the Leavey Research Grant for 2011-12 is gratefully acknowledged.

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)

Expectations of investors in the economy about the term structure of yields are central to the conduct of monetary policy. Influencing these expectations, using the different instruments available to the Federal Reserve, has been important during the Great Moderation. During the Great Recession, this strategy has been at the forefront of the central bank's policy: as the accommodative monetary policy stance of the Federal Reserve has kept the federal funds rate at the zero-lower bound since 2008, one of the main channels through which monetary policy can affect longer yields (and the subsequent consumption and savings decisions of economic agents), is by affecting the formation of conditional expectations by market investors.

Recognizing the importance of the formation of investor expectations, this paper asks three questions: how are the conditional expectations of investors about yields at different forecasting horizons formed, how do they change over time, and are there are significant differences in the formation of these beliefs during periods of high and low volatility? To answer these, we use U.S. nominal yield curve data to develop a novel methodology to model the evolution of investor beliefs. The Great Moderation is used as the baseline period, and the results are extended 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: following recent studies[3], a latent factor model is used to fit the U.S. nominal yield curve; in the model, implied conditional expectations of yields (and associated latent factors) are formed using a vector auto-regressive (VAR) model with constant coefficients. The forecasting performance of the model is evaluated, and a series of tests of rationality of the forecasting errors implied by this model confirm that these

---

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

errors are biased, systematic, and correlated with revisions in yield forecasts. Additionally, this framework imposes the restriction that investors must be placing identical weights on past information while forecasting the short and long yields. The model also implies that the investors must be using constant coefficients to form expectations 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 (2013) and Sinha (2013), 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 alternative specifications for the formation of conditional forecasts of the yield curve factors, and subsequent yields. A class of adaptive learning models are considered for expectations formation: constant gain learning (CGL) and variants of an endogenous learning algorithm that we develop. The main innovation is that investors are now allowed to vary the weights they place on past information about yields, and they are also able to change these weights in response to large and persistent deviations observed in the yield curve factors. There are significant improvements in forecasting performance of the model, at different forecasting horizons and yield maturities. For example, at the one-month forecasting horizon, the CGL forecast of the ten-year yield improves on the benchmark by 6%; at the six-month horizon, the improvement is 25%. Similar improvements are found in the one- and five-year yields. The main conclusions we draw are the following: (a) the implied conditional expectations of investors display significant time-variation during the Great Moderation; (b) investors place asymmetric weights on past information while forming expectations about future long and short yields, for a fixed forecasting horizon; and (c) for a fixed yield maturity, the conditional expectations for over different forecasting horizons will also place varying weights on past information. In addition, the results from the endogenous learning schemes suggest that when investors are making conditional forecasts at the shorter forecasting horizon, the process that best describes their expectations formation allows their beliefs to switch between different "regimes", even though the underlying state variable is assumed not to follow a regime-switching process. On observing large deviations in their coefficient estimates, we find that the investors begin to place greater weight on the past history of observations than before. However, for the longer forecasting horizons, on observing large deviations in the time series of the yield curve factors from the past, the investors may become more inattentive to the past data. Thus, there is an asymmetry in the formation of conditional forecasts at the short versus long forecasting horizon, which has important implications for analyzing the effects of monetary policy actions on the yield curve. For example, if conditional forecasts about the long yield at a fixed forecasting horizon respond much more sluggishly to new information than forecasts for the short yield, the policy action at the short end of the term structure will transmit to the long end at a slower pace than predicted by the constant coefficients models.

Our results for the Great Moderation and the financial crisis period suggest that the beliefs of investors can be best modelled using endogenous gain learning. The forecasting performances of the different learning algorithms during the low volatility period are similar; however, during periods of high volatility, there are large improvements in the conditional forecasts of the endogenous learning mechanism, relative to the constant gain process. The endogenous learning process does better in accounting for the movements in the observed time series of the latent factors (these approximate the level of the yield curve, its slope, curvature and convexity), which are used to form the forecasts of the yields. It is able to do this since it allows the beliefs to endogenously adapt to large deviations observed, and place more weight on past information than in periods of low volatility.

Our empirical strategy also contributes to the literature on adaptive learning. In the learning processes, the key parameter of interest is the updating or gain coefficient. The gains are allowed to vary across the different factors; this provides a general 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. To our knowledge, this is the first paper to extract optimal gains from (relatively) high frequency data. While the magnitude of the optimal gains derived from daily data is small, there are large improvements in forecasting performance.

Given the success of the endogenous learning mechanism in modeling conditional forecasts of investors, we consider the implications of our framework for two important aspects of the term structure of interest rates: first, how does the effect of a sudden policy announcement, which lowers the average level of yields, affect the term structure and the conditional forecasts, when agents use endogenous learning to form these forecasts; and second, do the adaptive learning mechanisms considered here for explain the patterns observed in survey data for excess returns.

For the first implication, we consider the following policy experiment. Since August 2011, the statements of the Federal Open Markets Committee meetings have included calendar-based forward guidance about the length of the accommodative monetary policy stance. We ask the following question: suppose that in a FOMC announcement, there is a surprise lengthening of the accommodative monetary policy stance, and a sudden drop in the level of the yield curve. In this case, what are the one-month ahead forecasts of the yields (at different maturities) implied by the constant-coefficients model, and the learning algorithms? We find that forecasts implied by the constant-coefficient model are unable to capture the observed changes in the yield curve, and the implied yield curve slope remains flat. This is in contrast to the observed yield curve. On the other hand, the constant gain and endogenous learning schemes perform significantly better. Intuitively, the time-invariant coefficients model does not fully account for the reaction of investors to the surprise information; in contrast, the learning mechanisms allow them to weight the new information differently following the shock.

To examine the implications for excess returns, we first use survey data from the Survey of Professional Forecasters (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 constant coefficients and learning models. The endogenous learning mechanism does significantly better at matching the observed patterns in survey expected excess returns, relative to the constant gain mechanism.

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. The policy experiment and findings for expected excess returns are described in sections six and seven respectively, and section eight concludes.

2 Related Literature

There are three strands of the literature that are relevant for the purposes for this paper. The first is the extant analyses that have used the Nelson-Siegel-Svensson parameterization for fitting the yield curve. The database used here is drawn from the nominal yield curves estimated by Gürkaynak, Sack and Wright (2007) based on this spline approach. 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[4]. 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). In contrast to these, the focus of this 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. If a time-varying process for the factors is found to perform better than the VAR for forecasting purposes, then it will have implications for the evolution of the discount factor over time. The work of Bianchi, Mumtaz and Surico (2009) models the U.K. nominal yield curve using the Nelson-Siegel-Svensson approach, and specify a time-varying process for the evolution of the factors. They find that the factors of the yield curve showed greater volatility before inflation targeting was adopted in the U.K. in 1992.

Time-varying beliefs have been widely incorporated into partial and general equilibrium models of the yield curve to match characteristics of the data: Laubach, Tetlow and Williams (2007) allow investors to re-estimate the parameters of their term structure model – both

---

[4] 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.

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 (2013) 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. The authors use constant-gain learning to describe these expectations, and the excess returns implied by the learning model capture movements in the empirical data better. The common theme of these analyses is the incorporation of subjective beliefs in explaining characteristics of the empirical term structure.

Finally, endogenous learning algorithms have been previously introduced in the literature 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. In Milani (2007a), the agents switch between gains based on the historical average of t he forecasting errors, instead of a fixed value. Gaus (2013) 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 (2012) 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

The zero-coupon yield curve for $1972 - 2011$ is modeled using the Nelson-Siegel-Svensson approach:

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

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 has been used by Gürkaynak, Sack and Wright (2007) to construct the zero-coupon yield curve, and is a parsimonious representation of the yield curve.[5] The estimates for this 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.[6]

To construct yield forecasts using the representation in (1), it must be amended with a process for the evolution of the factors. Diebold and Li (2006) and Aruoba, Diebold and Rudebusch (2006) specify the two-step estimation of yields and factors:

---

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

[6] 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.

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

Here $\mathbf{y}_t$ is the 3×1 vector of yields, $\mathbf{X}_t$ is a 4×1 vector of the regressors in (1)[7], $\boldsymbol{\beta}_t$ is a 4×1 vector of the factors, μ is the intercept and Φ denotes the dependence of the factors on past values. We will consider this as the benchmark model for factor evolution. The variance-covariance matrices given by:

$$var(\varepsilon_t) = H = \begin{pmatrix} \sigma_1^2 & 0 & 0 \\ ... & ... & ... \\ 0 & 0 & \sigma_n^2 \end{pmatrix}; \; var(\eta_t) = Q = \begin{pmatrix} \omega_{11}^2 & \omega_{12}^2 & \omega_{13}^2 \\ ... & ... & ... \\ \omega_{n1}^2 & \omega_{n2}^2 & \omega_{n3}^2 \end{pmatrix}. \quad (3)$$

The factor errors are assumed to be distributed as a normal, with mean zero.[8]

The forecasts of the yields are constructed as follows:

$$\begin{aligned} E_t \mathbf{y}_{t+h} &= E_t \mathbf{X}_t \hat{\boldsymbol{\beta}}_{t+h} = \mathbf{X}_t E_t \hat{\boldsymbol{\beta}}_{t+h} && \text{(4a)}\\ E_t \hat{\boldsymbol{\beta}}_{t+h} &= \left[I_3 - \hat{\Phi}^h\right]\left[I_3 - \hat{\Phi}\right]^{-1} \boldsymbol{\mu} + \hat{\Phi}^h \boldsymbol{\beta}_t, && \text{(4b)}\end{aligned}$$

where h is the forecast horizon. Here, the second equality in (4a) holds since we use estimated values of the parameters τ_1 and τ_2 at time t, while forming the conditional forecasts.

3.1 Properties of Nominal Yield Curve Factors

Table 1 summarizes the first and second moments of the yield curve factors, and figure 1 plots the first two factors for the full sample. The level and slope factors behave as expected: average interest rates are higher in the 1970s, and the yield curves slopes downwards. The $1995-2006$ period is characterized by an upward sloping yield curve. The persistence of the factors also changes between the two sample periods: the second sample is characterized by larger autocorrelations, at the one-, six- and twelve-month horizons than the 1970s. Figure

---

[7] Since the parameters τ_1 and τ_2 are jointly estimated using the maximum likelihood approach, the X_t vector is time-varying.

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

2 shows the correlation between the yield curve factors and their empirical counterparts. In the first panel of figure 2, we plot β_0 with the yield curve slope, computed as the average of the three-month, two-year and ten-year yields, $(y_t(3) + y_t(24) + y_t(120))/3$. For the sample period $1984 - 2011$, the correlation between the two variables is 0.56. The yield slope, $(y_t(3) - y_t(120))$ along with the second factor β_1 is plotted in the second panel, and the correlation between these variables is 0.61. Aruoba, Diebold and Rudebusch (2006) further interpret these factors in terms of macroeconomic variables. The correlation of β_0 with inflation[9] over between $1984 - 2011$ is 0.22; while this is lower than the estimates of Aruoba, Diebold and Rudebusch, the correlation between β_0 and one-year ahead inflation forecasts reported by the Survey of Professional Forecasters is 0.59.[10] The forecasts are the median forecasts from the SPF. Figure 2 plots the level factor and actual inflation (computed as specified above) and one-year ahead inflation expectations. The second factor, β_1 has been interpreted as approximating capacity utilization in the economy. For the $1984 - 2011$, the correlation between these variables is 0.63.

3.2 Tests of the Forecast Errors

Since the model for factor evolution in (2b), and implied conditional yield forecasts in (4a) 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 out-of-sample forecasts are constructed for the next four years, using a rolling data window. At each step, the one-, three- and six-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

---

[9] This is the inflation based on the annual percent change in the CPI for all Urban Consumers, seasonally adjusted. The data series is obtained using the St. Louis FRED database.

[10] Rudebusch and Wu (2004) emphasize the link between the level factor obtained from their macro-finance model and the actual as well as expected inflation.

conditional expected yields from (4a).

3.2.1 Are the Forecast Errors Unbiased?

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

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

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 panel A of table 2. The null hypothesis of unbiasedness requires $\alpha^n_1 = 0, \forall n$. The coefficients in panel A show that for the one-year yield maturity, as the forecast horizon increases, the implied conditional forecasts of yields overshoot the realized yields. For the five- and ten-year yields, the model undershoots the implied yields, but as the forecast horizon increases, the conditional forecasts are larger than the actual yields.

3.2.2 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:

$$y^n_{t+h} - E_t y^n_{t+h} = \alpha^n + \beta^n E_t y^n_{t+h} + e^n_{t,t+h}. \tag{6}$$

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 panel B of table 2 show that this hypothesis is rejected for the yield maturities considered, across the different forecast horizons.

---

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

3.2.3 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 (the construction of the forecasts is shown in appendix B). That is, in the following regression:

$$y_{t+h}^{n} - E_t y_{t+h}^{n} = \alpha^n + \beta^n \left(E_t y_{t+h}^{n} - E_{t-1} y_{t+h}^{n} \right) + e_{t,t+h}^{n} \tag{7}$$

the intercept and slope coefficients must be statistically not different from zero.[12] The coefficients from the regression in (7) are reported in panel C of table 2. 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.

3.2.4 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 ten-year Treasury yield and three-month Treasury bills are available. We construct the regressions in (5), (6) and (7) using the forecasts at the six- and twelve-month forecast horizons.[13] The results are shown in three panels in table 3. The null of unbiasedness is strongly rejected for the three-month Treasury bills. The median forecasts of the Treasury bills and the ten-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]

---

[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.

[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.

4 Construction of Yield Forecasts under Alternative Learning Models

In this section, investors are assumed to use the term structure model in (2). However, they now update their estimates of the parameters describing the factor evolution process, (μ, Φ), as new information on yields and implied latent factors becomes available. The timing is as follows: at time t, the estimates of $(\beta_0, \beta_1, \beta_2, \beta_3, \tau_1, \tau_2)$ are derived using maximum likelihood estimation. To construct forecasts of the yields at one-, three- and six-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.

In contrast to (2b), this process is represented using a time-varying VAR model (with the coefficients being updated using different learning schemes):

$$\boldsymbol{\beta}_t = \boldsymbol{\mu}_{t-1} + \Phi_{t-1}\boldsymbol{\beta}_{t-1} + \boldsymbol{\eta}_t. \tag{8}$$

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:

$$\begin{aligned} \begin{pmatrix} \mu_{i,t} \\ \phi_{i,t} \end{pmatrix} &= \begin{pmatrix} \mu_{i,t-1} \\ \phi_{i,t-1} \end{pmatrix} + g_i R_{i,t-1}^{-1} q_{i,t-1} \left[\beta_{i,t} - \begin{pmatrix} \mu_{i,t-1} \\ \phi_{i,t-1} \end{pmatrix}' q_{i,t-1} \right] \\ R_{i,t} &= R_{i,t-1} + g_i \left[q_{i,t-1} q_{i,t-1}' - R_{i,t-1} \right] \end{aligned} \tag{9}$$

where $q_{i,t-1} = \left(1, \beta_{i,t}\right)'^{t-1}_{t=0}$, 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:

$$\begin{aligned} E_t \mathbf{y}_{t+h} &= \mathbf{X}_t E_t \hat{\boldsymbol{\beta}}_{t+h} \\ E_t \hat{\boldsymbol{\beta}}_{t+h} &= \left[I_3 - \hat{\Phi}_{t-1}^h \right] \left[I_3 - \hat{\Phi}_{t-1} \right]^{-1} \boldsymbol{\mu}_{t-1} + \hat{\Phi}_{t-1}^h \boldsymbol{\beta}_t. \end{aligned} \tag{10}$$

The only distinction from (4a) 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]

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 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 (8), along with the updating equation in (9). However, the gain is no longer held fixed for the entire sample. In the first variant of endogenous learning, EGL1 hereafter, the gain switches according to the specification below:

$$g = \begin{cases} g_1 \text{ if } \left|\frac{\Omega_t - \bar{\Omega}_k}{\sigma_\Omega}\right| \leq \varepsilon \\ g_2 \text{ if } \left|\frac{\Omega_t - \bar{\Omega}_k}{\sigma_\Omega}\right| > \varepsilon \end{cases} . \tag{11}$$

Here $\bar{\Omega}$ is the average of the k most recent coefficients and σ_Ω is the standard deviation of these k coefficients. The following time line describes the investors' updating mechanism: investors use a baseline time period to estimate gains g_1 and g_2. At time t, they observe new data on β_t, and use the estimated gain g_1 to update there coefficients to Ω_t. They then compare these coefficients to the average of the coefficients for the k most recent periods. If

---

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

the difference is not significant[16], the investors continue to use g_1 to update their coefficients next period. However, if the difference is large, they switch the gain to g_2, and then updates its coefficients to $\Omega_{t.}$.

The novel feature of this learning mechanism is that it allows the investors to endogenously switch their beliefs and permits them to change the weights they place on past data, in response to new information. This does not require the underlying state variable (the endowment process in this simple model) to be regime-dependent.

An alternative to the gain specification in (11) is the following (EGL2 in the following discussions), developed in Gaus (2013):

$$g_t = \bar{g}_{lb} \ + \bar{g}_{sf} \frac{\left| \frac{\Omega_t - \bar{\Omega}_k}{\sigma_\Omega} \right|}{1 + \left| \frac{\Omega_t - \bar{\Omega}_k}{\sigma_\Omega} \right|}, \tag{12}$$

where $\bar{g}_{lb}$ is the lower bound the endogenous gain 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 $0 < \bar{g}_{sf}$, $\bar{g}_{sf} < 1$ and $\bar{g}_{lb} + \bar{g}_{sf} < 1$, g_t will be bounded between zero and one. As times progresses, the investors will increase increase the value of the gain in times when their coefficient estimates are different from the recent past, and decrease the value of the gain when their coefficient estimates are similar.

The process in (12) is different from the gain in (11) the following fundamental way: as the divergence between the recent coefficient estimates and mean increases, investors become more inattentive to the history of data. That is, they begin to weight the more recent observations more heavily. Meanwhile, the gain in (11) implies that the agents may be weighing past observations more ($g_2 < g_1$), or less ($g_2 > g_1$). However, it will not adapt to the the difference between the coefficient estimates and the mean, in the same manner as (12). The comparative numerical results below are presented for the benchmark constant-coefficients case (in which (μ_t, Φ_t) are not updated and $g = 0$), CGL, and gain specifications following (11) and (12).

---

[16] Using the measure defined in (11).

5 Evaluation of the Models and Implications for Investor Expectations

There are two 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? 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 root mean square errors (RMSEs), and then the implications of these results for modeling investor expectations are discussed. The sample period is January 1980 to December 1993, and the out-of-sample forecasts are constructed for the years between 1994-2000. The out-of-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 in section 5.2 below, we describe the mechanism used to compute the optimal gains used in the different learning mechanisms.

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, we must first set the initial values of the gain parameter. We allow the investors to use different gains for the four latent factors, and these initial values are available upon request. The initial sample period is used find the optimal constant gain for the latent factors. These are shown in table 4. The optimal gains (as well as the initial values of the gain parameters) for the remaining forecast horizons are shown in the appendix. 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 (2007a) estimates a gain of 0.02 using a DSGE model for the U.S. economy. To our knowledge, our paper is the first to provide estimates of the gain parameter, using macroeconomic data observed at a daily frequency.

The values of the gain parameter are central to characterizing expectations using these learning models. The values of the gain parameter presented in table 4 show that at the different forecasting horizons, the gain for the factor corresponding to the slope is higher than for the other factors, and it decreases across the forecasting horizon. This implies that while forming conditional expectations at these longer horizons, more weight is being assigned to observations further in the past. Therefore, investors appear to paying more attention to a longer history of data for the yield curve slope, compared to the other factors. The importance of varying gain values is further discussed in the context of the policy experiment simulated below. In the first variant of endogenous learning, EGL1, the optimal values for the gains corresponding to the different factors are specified in the third and fourth columns of 4. The investors are assumed switch between these gains when the difference between the estimated coefficients Ω_t differs from the historical average by more than two standard deviations. In all the following simulations of conditional yield forecasts, investors are assumed to be using $t = 140$ days of data. On observing large deviations from the past coefficients, the investors use g_2. As the coefficients show, large deviations from the historical data on coefficients motivate investors to optimally choose a lower gain; that is, they place greater weight on past observations than before.

5.2 Investor Expectations during the Great Moderation and Model Evaluation

Table 5 presents the comparison of conditional forecasts of the constant-coefficients[17], constant gain and endogenous learning models. In order to compare the forecasting performance, we first fix the yield maturity, and then compare across forecasting horizons. As expected, in general, the constant coefficients model is outperformed by the learning models. For the

---

[17] For the constant-coefficients model, in order to construct the forecast, as additional data becomes available for the latent factors, the coefficients Ω_t are also re-estimated. This strategy is adopted to allow the benchmark model to have the best possible forecasting perfomace against the alternatives. This is in contrast to the methodology of Laubach, Tetlow and Williams (2007)

one-year yield, the learning algorithms outperform the constant coefficients model in a more significant manner as the forecasting horizon increases; the pattern is repeated across the yield curve as well. Also, for all the yield maturities considered, as the forecasting horizon increases, the RMSEs are *lowered* for the learning algorithms, in contrast to the result for the constant coefficients case. For example, at the six-month horizon, for the ten-year yield, the CGL algorithm improves upon the RMSE of the benchmark model by approximately 23%. This improvement in forecasting performance suggests that even during the period of the Great Moderation, there is evidence of investors accounting for structural changes in the factors. Conversely, for a fixed forecasting horizon, the improvement in forecasting horizon for the learning models is larger as the yield maturity increases.

The findings for the learning models can be parsed out further dimensions. For a fixed yield maturity, as the forecasting horizon increases, there are significant improvements in the RMSE, and this is applicable to all the different algorithms. For example, for the one-year yield, the RMSEs at the six-month horizon are lower than the RMSEs at the one-month horizon by 21% and 16% for the constant gain and EGL1 algorithms respectively. The corresponding reductions in RMSEs for the ten-year yield are 18% and 16%. Conversely, when the forecasting horizon is held fixed, as the there are small improvements in forecasting performance across the yield curve.

The predictions of the EGL1 model are not significantly different from those of CGL for the Great Moderation. However, the finding that investors are switching to a lower gain for the different latent factors, when large deviations from the past average of coefficients are observed, is important. Applying the second variant of endogenous learning (EGL2) to the Moderation period, we find that for the one- and three-month horizons, EGL2 implies similar results as EGL1 across the various yield maturities (table 5). At the six-month forecasting horizon, the EGL2 forecasts are better than those implied by EGL1. The fact that EGL2 implies better forecasts at the longer forecast horizons, implies that investors may simply choose to become inattentive to the history of data, when they observe large deviations from the past. However, while they are forecasting over the shorter and medium-term horizons (and the subsequent consumption, savings and investment decisions associated with these horizons), they find it optimal to pay more attention to the past behavior of the data.

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. While this framework does well during the Great Moderation, our analysis for the Great Recession below suggests that during periods of large deviations in the data, from the historical average, it may not be able to capture the belief formation process adequately. Adopting the endogenous learning algorithms proposed above provides an intuitive manner to model investor beliefs during such periods, as well as periods with low volatility. Intuitively, during periods of low volatility, the investors will not need to pay attention to a longer time series of the latent factors. However, as deviations from past averages increase, the investors begin to pay significantly more attention to the historical evolution of the factors. This finding corroborates the findings of Coibion and Gorodnichenko (2012), where the authors find that the degree of information rigidity increases during the Great Moderation, and decreases in periods of high volatility.

5.3 Investor Expectations during the Great Recession

To analyze the implications for the Great Recession period, the baseline period is from July 2006 to June 2009, and the forecasts are constructed for July 2009 to January 2011. The Root Mean Square Error (RMSE) is used to compare the forecasting performance across different models. The results for the different models are presented in table 6. The optimal gains are derived in a similar manner; during the Great Recession, we find that the optimal values of the constant gain are larger than for the Great Moderation by two orders of magnitude . The gains also increase as the forecast horizon is increased to three months. We attribute this partly to the short data sample used to find the optimal gains. Due to the short sample period, we only present the results for the one- and three-month forecasting horizons for this period.

For the recession period, the CGL mechanism improves on the benchmark model by close to 60% for the ten-year yield at the three-month forecasting horizon. During the Great Recession, EGL1 outperforms the constant gain learning algorithm, as the RMSEs show in table 6. Therefore, investors are doing better at predicting yields at all forecasting horizons, when they begin to weight the historical data more heavily. This implies that their conditional forecasts display much more persistence than the CGL model allows for, in periods of large deviations.

In our view, this observation has important implications for considering the effect of different monetary policy actions on investor expectations. If investors place asymmetric weights on recent observations for shorter and longer yields at different forecasting horizons, then the policy of targeting only shorter-term interest rates may not translate into the desired effects on longer-term investor expectations. Both the benchmark model, as well as the constant gain learning approach will be unable to capture this shift in beliefs, and the analysis of monetary policy actions through the lens of these frameworks, may be an incomplete representation of investors' conditional forecasts.

5.4 Other Moments of the Forecasting Errors

Other moments of the forecasting errors are presented in table 7. We find that the mean of the errors for the different yields increases as the forecasting horizon becomes longer. The variance of the errors also increases across the yield spectrum as the forecasting horizon is lengthened, although the autocorrelation of the forecast errors reduces. These findings are similar to the results reported by Diebold and Li (2006). Finally, the mean forecast errors for the recession period are negative: the model consistently produces yield forecasts that overshoot the realized yields during this period.

The properties of the forecast errors implied by the CGL model are further analyzed in table 8. In the latter period, the yield forecasts implied by the CGL algorithm overshoot the realized yields; however, the magnitude of overshooting is smaller compared to the benchmark case. The variance of the forecasting errors is also reduced substantially. Other moments of the forecast errors from the endogenous learning mechanisms during the Moderation and period are shown in tables 9 and 10. The moments of forecasting errors during the Great Recession are shown in the appendix.

5.5 Context in the Literature

The endogenous learning techniques proposed in this paper provide a general mechanism to model the change in investor beliefs, in response to large fluctuations in the data. These can be easily applied in cases where learning for multiple variables is required. It is, however, useful to compare the performance of our learning algorithms, with the endogenous learning process suggested by Marcet and Nicolini (2003). To explain the hyperinflations across different countries, the authors propose a learning algorithm in which the agents switch between a decreasing and constant gain based on the forecasting errors for the variable being forecast (in their case, inflation). While this strategy works well in the case of univariate forecasting, it may be challenged in the case of multiple variables. To test this, we use the Marcet and Nicolini strategy, and the gain parameter now switches between decreasing gain ($g = 1/t$), and the constant gain estimated for the baseline period used above. The results are reported in the appendix.

Using the Marcet and Nicolini (2003) process, denoted as MN1, the gain switches to the constant value when the mean forecasting error exceeds a predetermined value. In the Milani (2007) variation, denoted MN2, the switching occurs when the mean forecasting error exceeds the historical average of forecasting errors. The results show a consistent pattern across the forecasting horizons, for a fixed yield maturity: the MN1 process is outperformed by the remaining learning algorithms. For both the period under consideration, the MN1 and MN2 techniques imply similar RMSEs as EGL1 and EGL2. The results are available upon request.

6 A Policy Experiment: The Effects of an FOMC Announcement

The findings above suggest the following: (a) the constant-coefficients benchmark model is unable to capture the varying persistence in the conditional forecasts of investors, at different forecast horizons and yield maturities; (b) differences in the formation of conditional expectations, at the short- and long-end of the term structure have important implications for analyzing the effects of central bank policies, which influence the short yields, on the term structure of yields.

In order to illustrate the differences in the predictions of the constant-coefficients and the learning models, we consider a policy experiment. In the recent years, following the Great Recession, the communications of the Federal Reserve, through the statements of the Federal Open Markets Committee (FOMC), have provided increasingly explicit guidance about the timing of the monetary policy intervention. The introduction of these statements has been important: the first calendar-based guidance of the FOMC, in August 2011, was found to cause a significant drop in the expectations of the federal funds rate by professional forecasters, as found by Crump, Eusepi and Moench (2013).

The calendar-based guidance in the September 2012 statement shifted the end of the accommodative monetary policy to mid-2015. Following this statement, the October statement made no changes to the policy, and in December 2012, the accommodative stance of policy was made dependent on the state of the economy, with an emphasis on the unemployment rate. Thus, the FOMC statement in September 2012 was important because it was the last explicit change in date-based calendar guidance. We ask the following question: following the release of this statement, what was the evolution of yields in the data, at the one-month horizon, and what were conditional forecasts of yields implied by the constant-coefficient and benchmark models?

In order to construct the conditional forecasts of yields, following the FOMC statement release on September 13, 2012, we adopt the following strategy. The estimates of $(\beta_0, \beta_1, \beta_2, \beta_3, \tau_1, \tau_2)$ for mid-December 2011 to September 13, 2013 are used to derive the AR(1) parameters in (2b). We then construct the one-month ahead forecasts, using the constant-coefficients model, and the different learning algorithms. For these algorithms, the optimal gains found during the Great Moderation are used (since the gains from the Great Recession are found using a significantly smaller sample period). Following the announcement, a one standard deviation fall in β_0 is simulated, and the shock is assumed to last for two days following the announcement. The one-month ahead forecasts of the one-, five- and ten-year yields are shown in figure 3. To show the relative performance of the different models, we show the ratio of the five- and ten-year yields to the one-year yield. At the one-month horizon, the implied yield curve (derived using the estimated β and τ factors, at the one-month ahead horizon) falls. The constant-coefficients model forecasts of yields explains only 19.5% of the total variation in yields, following the shock. However, the constant gain and endogenous gain learning (EGL1) algorithms predict yields that follow the same pattern as the implied yield curve, and explain 34.4% and 29.6% of the total variation in yields following the shock. While the drop in the learning yields is not as large as shown by the actual yield curve, the conditional forecasts are significantly closer than the constant-coefficient benchmark. This experiment suggests that accounting for the varying weights placed on the history of information is important for understanding the forecasts of investors about the term structure.

7 Explaining Expected Excess Returns

Predictable patterns in excess returns for nominal yields in the U.S. have been well documented. Piazzesi, Salomao and Schneider (2013) 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 then investigate the implications of the learning model, along with those of the constant-coefficients model.

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

$$E_t\left[rx_{t,t+h}^{(n+h)}\right] = E_t\left[p_{t+h}^{(n)}\right] - p_t^{(n+h)} - y_t^h, \tag{13}$$

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:

$$E_t\left[rx_{t,t+h}^{(n+h)}\right] = -nE_t\left[y_{t+h}^{(n)}\right] + (n+h)\, y_t^{(n+h)} - y_t^{(h)}. \tag{14}$$

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

$$\begin{aligned} E_t\left[rx_{t,t+h}^{(n+h)}\right] &= -nE_t\left[y_{t+h}^{(n)}\right] + (n+h)\, y_t^{(n+h)} - y_t^{(h)} \\ &= -40E_t\left[y_{t+4}^{(40)}\right] + (40+4)\, y_t^{(40+4)} - y_t^{(4)}. \end{aligned} \tag{15}$$

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 and constant coefficients models; in this case, the $E_t\left[y_{t+4}^{(10)}\right]$ is computed using the conditional forecasts described above in section 4. 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 panel A of table 4).

The evolution of expected excess returns for the constant coefficients, CGL and EGL2 algorithms are shown in the two panels of figure 4. The excess returns from the former are significantly more variable. These are also higher on average for the sample period, than excess returns implied by any of the learning algorithms. The evolution of the expected excess returns over the sample period also reveals an important finding: the excess returns of the constant-coefficients are counter cyclical; in comparison, the excess returns from the learning models are significantly less counter cyclical. This is also the pattern observed in the expected excess returns generated from the SPF data. These findings are similar to the results of Piazzesi, Salomao and Schneider (2013).

Among the learning models, the endogenous learning algorithm matches the patterns observed in survey data better even during the sub-sample of the Great Moderation period. With the constant gain process, there are several sub-period observed during which the implied excess returns are rising, even as the survey expected excess returns are falling. Finally, table 11 shows the moments of the expected excess returns for the ten-year yield, at the one-year forecasting horizon for the SPF survey data, constant coefficients and the different learning models.

8 Conclusion

An empirical analysis of how subjective expectations evolve is useful for both macroeconomists and financial economists. 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. 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 the Nelson-Siegel-Svensson model of characterizing the yield curve can be improved upon by allowing for a process for factor evolution that incorporates time-varying parameters, instead of a constant-coefficient VAR model. Alternative models of expectations formation, the constant-gain learning process and endogenous gain, improve upon the forecasting performance of the spline based method. The improvements in out-of-sample forecasting occurs during periods of low volatility, as well as during the financial crisis period.

Acknowledgements: We thank Bruce Preston, Ricardo Reis and Michael Woodford for detailed comments.

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, pp. 309-338.p. 396-409.*

[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, pp. 856-71.*

[3] 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, pp. 4-20.*

[4] Coibion, O., and Y. Gorodnichenko, (2012), "Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts", *Working paper.*

[5] Crump. R, S. Eusepi and E. Moench, (2013), "Making a Statement: How Did Professional Forecasters React to the August 2011 FOMC Statement?", *Liberty Street Economics Blog.*

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

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

[8] Eusepi, Stefano, and B. Preston, (2011). "Expectations, Learning and Business Cycle Fluctuations". *American Economic Review, forthcoming.*

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

[10] Gaus, E., (2013), "Time-Varying Parameters and Endogenous Learning Algorithms", *Working paper, Ursinus College.*

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

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

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

[14] 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.*

[15] 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.*

[16] Mankiw, N.G., J. Wolfers and R. Reis, (2004), “Disagreement about Inflation Expectations,” *NBER Macroeconomics Annual 2003, pp. 209-248*

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

[18] Milani, F., (2007a), "Learning and Time-Varying Macroeconomic Volatility", *Working paper.*

[19] Milani, F., (2007b), "Expectations, Learning and Macroeconomic Persistence", *Journal of Monetary Economics, 54, pp. 2065-2082.*

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

[21] Piazzesi, M., J. Salomao and M. Schneider, (2013), "Trend and Cycle in Bond Premia", *Working paper.*

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

[23] 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, pp. 395-422.*

[24] Sinha, A., (2013), "Learning and the Yield Curve", *Working paper.*

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

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

Table 1: Properties of Nominal Yield Curve Factors

Factor	1995-2006				
	μ	σ	$\rho(\beta_t, \beta_{t-m})$		
			$m = 1$	$m = 6$	$m = 12$
β_0	1.98	1.99	0.73	0.16	0.23
β_1	2.01	2.26	0.75	0.22	0.25
β_2	1.04	3.28	0.77	0.27	0.14
β_3	12.16	5.72	0.70	-0.14	0.40

Note: The above moments are show for end of month data on the latent factors for the subsamples indicated.

Table 2: Testing Forecast Errors for Nominal Yield Curve Factors

Yield	$h = 1$ month		$h = 3$ months		$h = 6$ months	
Maturity	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)

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.

Table 3: Testing Forecast Errors for SPF Data

Yield	$h = 3$ months		$h = 1$ year	
Maturity	α	β	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)		

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

Table 4: Optimal Values of the Gain Parameter

	Optimal Values of Gain Parameters				
Factors	CGL	EGL1		EGL2	
		g_1	g_2	$\bar{g}_{lb}$	$\bar{g}_{sf}$
	Great Moderation (Forecasting horizon $h = 1$ month)				
β_0	0.0007	0.0030	0.0001	0.0001	0.0009
β_1	0.0022	0.0027	0.0001	0.0011	0.0002
β_2	0.0012	0.0012	0.0001	0.0013	0.0007
β_3	0.0009	0.0010	0.0004	0.0001	0.0006
	Great Recession (Forecasting horizon $h = 1$ month)				
β_0	0.1899	0.0545	0.0001	0.1905	0.0000
β_1	0.1486	0.0378	0.0001	0.1487	0.0000
β_2	0.1132	0.1126	0.0002	0.1132	0.0000
β_3	0.2187	0.1063	0.0010	0.2187	0.0000

Note: These are the optimal gain values for constant gain (CGL), endogenous learning with switching (EGL1) and endogenous gain with the scaling factor (EGL2), at the one-month forecasting horizon, for the two sample periods.

Table 5: Evaluating the Conditional Forecasts

Yield Maturity	RMSE-BM	RMSE-CGL	RMSE-EGL1	RMSE-EGL2
	Forecasting horizon $h = 1$ month			
1 year	3.8583	3.6662	3.6670	3.6651
5 years	3.5326	3.3305	3.3330	3.3319
10 years	3.8077	3.6132	3.6163	3.6157
	Forecasting horizon $h = 3$ months			
1 year	3.7911	3.3629	3.3654	3.3651
5 years	3.5510	3.1529	3.1571	3.1570
10 years	3.8195	3.4432	3.4476	3.4477
	Forecasting horizon $h = 6$ months			
1 year	3.7653	3.0183	3.0675	3.0218
5 years	3.5583	2.6865	2.7541	2.6920
10 years	3.8187	2.9551	3.0272	2.9609

Note: These are the root mean square (RMSE) values for constant coefficients (BM), constant gain (CGL), endogenous learning with switching (EGL1) and endogenous gain with the scaling factor (EGL2) models, at the three forecasting horizons.

Table 6: Evaluating the Conditional Forecasts during the Great Recession

Yield Maturity	RMSE-BM	RMSE-CGL	RMSE-EGL1	RMSE-EGL2
	Forecasting horizon $h = 1$ month			
1 year	21.9134	10.7684	10.6107	10.6438
5 years	29.8131	12.3722	12.2163	12.2476
10 years	29.9448	11.4484	11.3015	11.3330
	Forecasting horizon $h = 3$ months			
1 year	21.9738	11.0417	10.9172	10.9172
5 years	29.8787	12.7111	12.5824	12.5824
10 years	29.9597	11.7318	11.6102	11.6102

Note: These are the root mean square (RMSE) values for constant coefficients (BM), constant gain (CGL), endogenous learning with switching (EGL1) and endogenous gain with the scaling factor (EGL2) models, at the three forecasting horizons.

Table 7: Moments of Forecasting Errors from Benchmark Model for Nominal Yields for the Great Moderation

Yield Maturity	Mean	Std.dev	ρ_1	ρ_{12}
	Forecasting horizon $h = 1$ month			
1 year	1.6691	3.4797	0.7374	0.3113
5 years	2.7581	2.2080	0.5681	0.1074
10 years	3.1550	2.1326	0.5337	0.0375
	Forecasting horizon $h = 3$ months			
1 year	1.7276	3.3757	0.3370	-0.1783
5 years	2.7660	2.2276	0.2381	-0.2964
10 years	3.1415	2.1731	0.2315	-0.2705
	Forecasting horizon $h = 6$ months			
1 year	1.7665	3.3263	0.2644	-0.2467
5 years	2.7319	2.2807	0.1278	-0.2227
10 years	3.0909	2.2432	0.0945	0.0234

Note: These are the mean, standard deviation (Std.dev), and autocorrelations in the forecast errors. ρ_1 denotes the autocorrelation between the forecast errors at the 1 month lag; ρ_{12} is the statistic at the 12 month lag.

Table 8: Moments of Forecasting Errors from Constant Gain Learning Model for Nominal Yields for the Great Moderation

Yield Maturity	Mean	Std.dev	ρ_1	ρ_{12}
	Forecasting horizon $h = 1$ month			
1 year	1.6622	3.2688	0.7764	0.3274
5 years	2.7465	1.8845	0.6244	0.1368
10 years	3.1429	1.7831	0.5818	0.0654
	Forecasting horizon $h = 3$ months			
1 year	1.7340	2.8822	0.5008	-0.1509
5 years	2.7419	1.5570	0.3617	-0.5345
10 years	3.1128	1.4721	0.3317	-0.4897
	Forecasting horizon $h = 6$ months			
1 year	1.4480	2.6491	0.4293	-0.2729
5 years	2.3560	1.2915	0.3475	-0.4472
10 years	2.7072	1.1852	0.2860	-0.2552

Note: These are the mean, standard deviation (Std.dev), and autocorrelations in the forecast errors. ρ_1 denotes the autocorrelation between the forecast errors at the 1 month lag; ρ_{12} is the statistic at the 12 month lag.

Table 9: Moments of Forecasting Errors from EGL1 Model for Nominal Yields for the Great Moderation

Yield Maturity	Mean	Std.dev	ρ_1	ρ_{12}
	Forecasting horizon $h = 1$ month			
1 year	1.6670	3.2672	0.7673	0.3241
5 years	2.7508	1.8827	0.6175	0.1319
10 years	3.1471	1.7821	0.5769	0.0600
	Forecasting horizon $h = 3$ months			
1 year	1.7383	2.8826	0.4987	-0.2708
5 years	2.7456	1.5588	0.3578	-0.2613
10 years	3.1165	1.4748	0.3278	-0.2190
	Forecasting horizon $h = 6$ months			
1 year	1.5437	2.6516	0.4728	-0.6109
5 years	2.4509	1.2566	0.3608	-0.5427
10 years	2.8021	1.1461	0.2959	-0.3122

Note: These are the mean, standard deviation (Std.dev), and autocorrelations in the forecast errors. ρ_1 denotes the autocorrelation between the forecast errors at the 1 month lag; ρ_{12} is the statistic at the 12 month lag.

Table 10: Moments of Forecasting Errors from EGL2 Model for Nominal Yields for the Great Moderation

Yield Maturity	Mean	Std.dev	ρ_1	ρ_{12}
	Forecasting horizon $h = 1$ month			
1 year	1.6659	3.2657	0.7688	0.3266
5 years	2.7503	1.8814	0.6191	0.1321
10 years	3.1468	1.7813	0.5779	0.0599
	Forecasting horizon $h = 3$ months			
1 year	1.7390	2.8818	0.4997	-0.1453
5 years	2.7460	1.5582	0.3581	-0.5284
10 years	3.1169	1.4741	0.3278	-0.4837
	Forecasting horizon $h = 6$ months			
1 year	1.4546	2.6495	0.4312	-0.2879
5 years	2.3604	1.2949	0.3438	-0.3933
10 years	2.7114	1.1900	0.2803	-0.2174

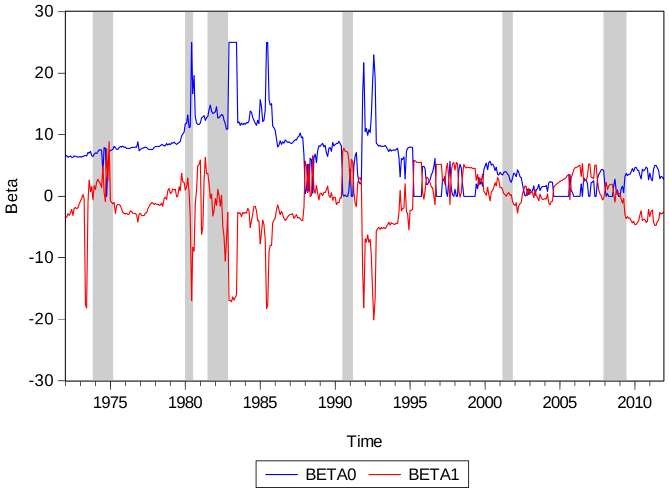
Note: These are the mean, standard deviation (Std.dev), and autocorrelations in the forecast errors. ρ_1 denotes the autocorrelation between the forecast errors at the 1 month lag; ρ_{12} is the statistic at the 12 month lag.

Table 11: Moments of Excess Returns

Data/Model Moment	SPF	BM	CGL	EGL1	EGL2
Mean	0.0858	2.2616	1.5491	2.1674	2.5124
Stdev	0.3624	1.7333	1.0635	0.9621	0.8021

Note: This table reports the mean and standard deviations of the expected excess returns derived using equation (15). The time period used for computing excess returns from SPF data is 1993:Q1 - 2007:Q4. The mean and standard deviation values have been converted to the monthly frequencies.

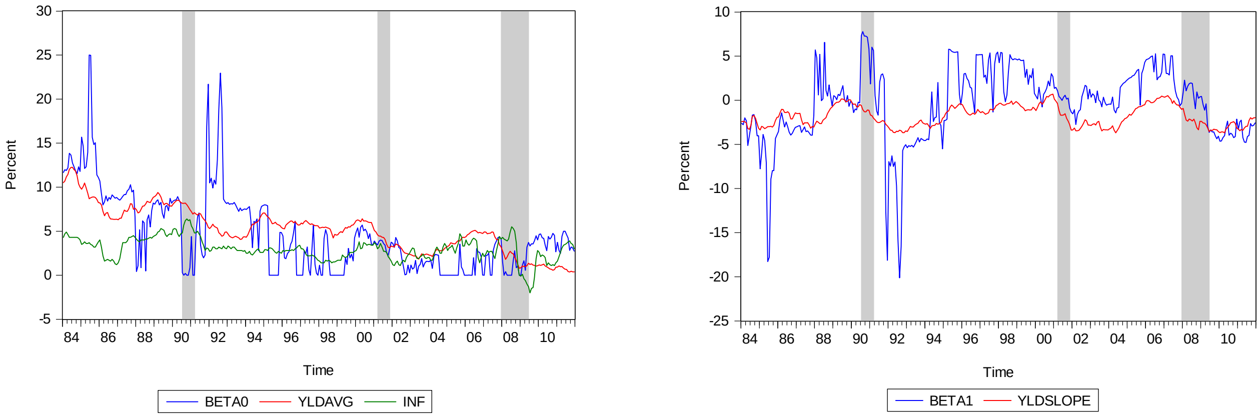
Figure 1: Level and Slope Factors for Nominal Curve



Note: The figure shows the evolution of the first two latent factors of the nominal yield curve. The shaded regions denote the NBER recessions.

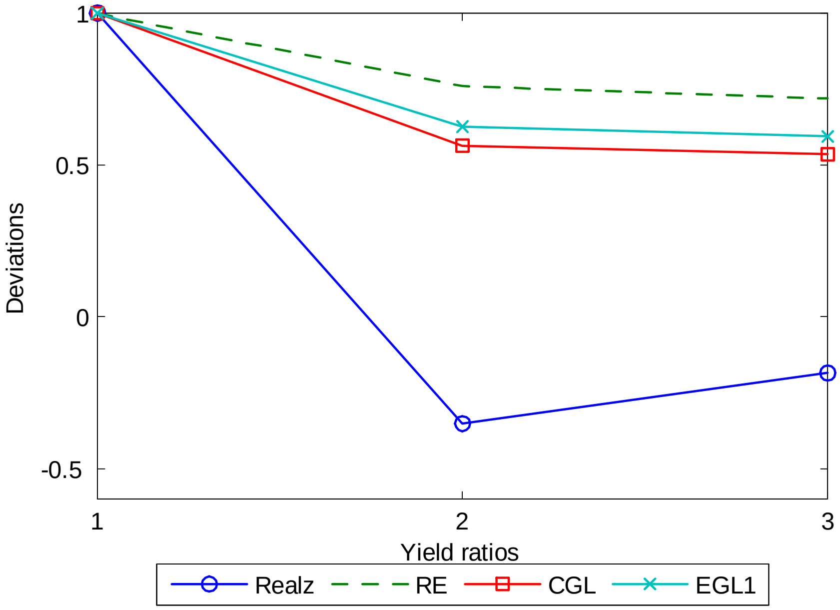
Figure 2: Factors and Empirical Counterparts for Nominal Curve

(*2a*) (*2b*)



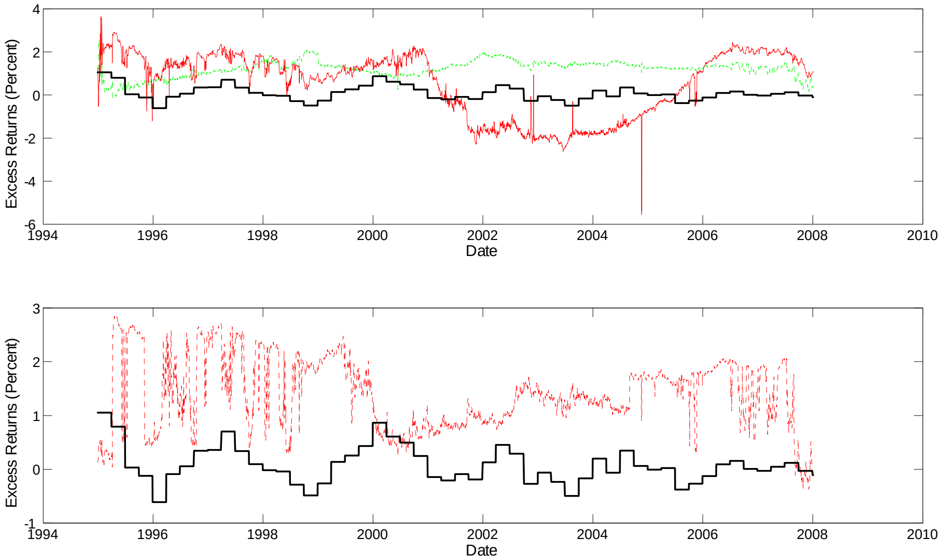
Note: The figure shows the evolution of the first two latent factors of the nominal yield curve, with their obervable empirical counterparts. The shaded regions denote the NBER recessions.

Figure 3: Effects of a Policy Announcement on the Yield Forecasts



Note: The figure shows the one-month ahead conditional forecast of the yield curve, following a one standard deviation fall in β_0. The ratio of the yields are shown here. The graph point corresponding to '2' is the ratio of the one-month ahead five-year yield to the one-year yield, implied by the different models; at point '3', the corresponding ratio of the ten-year yield to the one-year yield is shown. The graph plots the realized yield ratios (Realz), the yields under rational expectations (RE), constant gain learning (CGL) and the endogenous learning mechanism (EGL2).

Figure 4: Evolution of Expected Excess Returns



Note: This figures shows the evolution of expected excess returns for the ten-year yield, derived from the survey data (from the Survey of Professional Forecasters), the constant-coefficients and learning models. In the first panel, the SPF returns are shown by the black line, the CGL returns by the green line, and the EGL1 returns by the red line. In the second panel, the SPF returns are again shown by the black line, and the constant-coefficients returns are shown by the red line.

Appendix

Table 12: Other Moments of Forecasting Errors of Benchmark Model during the Great Recession

Yield Maturity	Mean	Std.dev	ρ_1	ρ_{12}
	Forecasting horizon $h = 1$ month			
1 year	-15.7759	15.2857	-0.1448	-
5 years	-26.0464	14.5786	-0.3960	-
10 years	-26.1070	14.7406	-0.4222	-
	Forecasting horizon $h = 3$ months			
1 year	-15.8152	15.3323	0.1119	-
5 years	-26.0647	14.6807	0.2530	-
10 years	-26.0512	14.8706	0.2685	-

Note: These are the mean, standard deviation (Std.dev), and autocorrelations in the forecast errors. ρ_1 denotes the autocorrelation between the forecast errors at the 1 month lag; ρ_{12} is the statistic at the 12 month lag. For the shorter sample, we do not report the 12 month autocorrelation.

Table 13: Other Moments of Forecasting Errors from CGL model during the Great Recession

Yield Maturity	Mean	Std.dev	ρ_1	ρ_{12}
	Forecasting horizon $h = 1$ month			
1 year	-9.3471	5.3739	-0.0461	-
5 years	-11.6287	4.2456	-0.1031	-
10 years	-10.6908	4.1161	-0.1006	-
	Forecasting horizon $h = 3$ months			
1 year	-9.7222	5.2606	-0.2404	-
5 years	-12.0137	4.1736	-0.2256	-
10 years	-11.0054	4.0846	-0.1713	-

Note: These are the mean, standard deviation (Std.dev), and autocorrelations in the forecast errors. ρ_1 denotes the autocorrelation between the forecast errors at the 1 month lag; ρ_{12} is the statistic at the 12 month lag. For the shorter sample, we do not report the 12 month autocorrelation.

Table 14: Other Moments of Forecasting Errors from EGL1 model during Great Recession

Yield Maturity	Mean	Std.dev	ρ_1	ρ_{12}
	Forecasting horizon $h = 1$ month			
1 year	-9.1565	5.3884	-0.0821	-
5 years	-11.4026	4.4062	-0.1496	-
10 years	-10.4602	4.3002	-0.1497	-
	Forecasting horizon $h = 3$ months			
1 year	-9.5530	5.3110	0.1108	-
5 years	-11.8095	4.3638	0.1314	-
10 years	-10.7969	4.2905	0.1416	-

Note: These are the mean, standard deviation (Std.dev), and autocorrelations in the forecast errors. ρ_1 denotes the autocorrelation between the forecast errors at the 1 month lag; ρ_{12} is the statistic at the 12 month lag. For the shorter sample, we do not report the 12 month autocorrelation.

Table 15: Other Moments of Forecasting Errors from EGL2 model from Great Recession

Yield Maturity	Mean	Std.dev	ρ_1	ρ_{12}
	Forecasting horizon $h = 1$ month			
1 year	-9.1779	5.4175	-0.0905	-
5 years	-11.4246	4.4362	-0.1589	-
10 years	-10.4823	4.3295	-0.1593	-
	Forecasting horizon $h = 3$ months			
1 year	-9.5530	5.3110	0.1108	-
5 years	-11.8095	4.3638	0.1314	-
10 years	-10.7969	4.2905	0.1416	-

Note: These are the mean, standard deviation (Std.dev), and autocorrelations in the forecast errors. ρ_1 denotes the autocorrelation between the forecast errors at the 1 month lag; ρ_{12} is the statistic at the 12 month lag. For the shorter sample, we do not report the 12 month autocorrelation.