

Accepted Manuscript

Can behavioral biases explain the rejections of the expectation hypothesis of the term structure of interest rates?

George Bulkley, Richard D.F. Harris, Vivekanand Nawosah

PII: S0378-4266(15)00095-3

DOI: <http://dx.doi.org/10.1016/j.jbankfin.2015.03.018>

Reference: JBF 4703

To appear in: *Journal of Banking & Finance*

Received Date: 12 July 2013

Accepted Date: 21 March 2015



Please cite this article as: Bulkley, G., Harris, R.D.F., Nawosah, V., Can behavioral biases explain the rejections of the expectation hypothesis of the term structure of interest rates?, *Journal of Banking & Finance* (2015), doi: <http://dx.doi.org/10.1016/j.jbankfin.2015.03.018>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Can behavioral biases explain the rejections of the expectation hypothesis of the term structure of interest rates?

George Bulkley¹, Richard D. F. Harris² and Vivekanand Nawosah^{3,*}

March 2015

Abstract

We test whether the rejections of the expectations hypothesis can be explained by two behavioral biases: the law of small numbers and conservatism. We use the term structure to decompose excess bond returns into components related to expectation errors and expectation revisions, enabling a direct test of behavioral models using the expectations of market participants. We find systematic patterns in expectation errors, and expectation revisions, which are consistent with these two biases. We show that a trading strategy that exploits these biases delivers significant economic profits and that our results are unlikely to be driven by a time-varying risk premium.

JEL classification: G02; G12; G14

Keywords: Behavioral bias; Expectations hypothesis of the term structure of interest rates; Representativeness; Law of small numbers; Conservatism.

¹ Department of Accounting and Finance, University of Bristol, 8 Woodland Road, Bristol BS8 1TN, UK. Email: George.Bulkley@bristol.ac.uk. Tel: +44 117 9289049.

² Xfi Centre for Finance and Investment, University of Exeter, Streatham Court, Rennes Drive, Exeter EX4 4ST, UK. Email: R.D.F.Harris@exeter.ac.uk. Tel: +44 1392 723215.

³ Essex Business School, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, UK. Email: vnawo@essex.ac.uk. Tel: +44 1206 873930. Fax: +44 (0) 1206 873429.

* Corresponding author.

We gratefully acknowledge financial support from the ESRC under research grant ACRR2784. We would also like to thank seminar participants at the University of Exeter, the University of Essex and Queen's University of Belfast.

1. Introduction

The expectations hypothesis (EH) of the term structure of interest rates states that the yield on a bond is determined by the expected short yield over the life of the bond plus a constant risk premium. The EH is usually tested by examining whether the market's expectations of future changes in bond yields, which are implicit in the term structure of interest rates, are unbiased. Empirical evidence from a large number of studies for different countries, different time periods and different bond maturities overwhelmingly rejects the EH.¹

In this paper, we examine whether this failure of the EH might be accounted for by biases in investors' expectations that arise from two well known behavioral models. The first is the 'law of small numbers' (LSN), which is a type of representativeness bias (see, for example, Tversky and Kahneman, 1971). The LSN describes the way in which individuals have a tendency to expect the moments of a population to be reflected even in short samples of data that are drawn from that population. It is motivated by experimental evidence that individuals tend to over-extrapolate from short runs of data. The second behavioral bias that we examine is 'conservatism', which builds on the widespread finding that individuals tend to be too conservative when reacting to new information. In particular, agents attach too much weight to their current beliefs and too little weight to recent news. Daniel et al. (1998) show that overconfidence in prior judgments about stocks can lead to investors giving too little weight to new public information compared to the weights that are specified by Bayes' rule. This leads to initial underreaction to new information but, over time, agents learn of their mistake and so there are subsequent revisions in expectations that are of the same sign as the initial response to the news.

¹ See, for example, Shiller (1979), Shiller et al. (1983), Campbell and Shiller (1984), Mankiw and Summers (1984), Mankiw (1986), Campbell and Shiller (1991) and Campbell (1995). Hardouvelis (1994) demonstrates that the rejection of the REH is not confined to the US.

The bond market offers an opportunity to directly test for the existence of these biases because the market's (risk neutral) expectation at any date for the short yield at any future date can be inferred from the term structure. We introduce a decomposition that allows us to construct both a series of expectational errors for forecasts at different horizons and a series of revisions in those expectations. The first of these components is used to test the short and long run implications of the LSN, while the second is used to test the implications of the conservatism bias.

We find systematic patterns in expectation errors and expectation revisions of the short yield for US zero-coupon Treasury securities that are consistent with both the LSN and conservatism biases. We investigate whether these biases are economically significant by examining whether a rational risk-averse investor could profitably exploit these patterns in the data. We report Sharpe ratios and Alphas from trading strategies that employ real time out-of-sample predictions of forecast errors implied by the behavioral models. We find that these strategies delivers significant risk-adjusted returns.

An alternative to the behavioral explanation investigated here, and one that has received much more attention, is that the failure of the EH is due to the assumption of a constant risk premium. If the risk premium is in fact time-varying it would not be surprising to find that the yield spread, which incorporates the risk premium, forecasts excess returns. However, the challenge is to develop an economic model that can explain the scale of the rejection of the EH. Dai and Singleton (2000) develop a statistical model of risk pricing that can explain the findings of Campbell and Shiller (1991) but they do not ground their model in economic fundamentals. Similarly, Cochrane and Piazzesi (2005) interpret their evidence that lagged yield spreads forecast excess returns as a statistical model of a time-varying risk premium but

acknowledge that their results are not necessarily consistent with an economic model of risk pricing.²

A problem for any risk-based explanation is that the volatility of the risk premium would have to be considerably higher than could be obtained under plausible levels of risk aversion. Backus, Gregory and Zin (1989) use a calibrated representative agent model to show that in a standard expected utility framework there cannot be sufficient time-variation in the risk premium to explain the scale of the rejection of the EH in the Campbell and Shiller (1991) tests. Similarly, Rudebusch and Swanson (2008; page 112) conclude from an investigation of the term premium in a dynamic stochastic general equilibrium framework that “*these models are very far from matching the level and variability of the term premium [...] we see in the data*”. It may be possible to rationalize the size and volatility of the term premium under the standard expected utility paradigm if we allow for a more complex specification of risk preferences. For example, Wachter (2006) shows that, if external habit persistence is introduced in a consumption-based model, the volatility of the term premium is significantly higher and many puzzling features of the empirical evidence can be explained. Piazzesi and Schneider (2006) calibrate term premia close to those observed in practice by assuming Epstein-Zin preferences, combined with the assumption that inflation shocks are negatively correlated with consumption shocks.

Given the attention that a time-varying risk premium has received in the literature, and notwithstanding these reservations, we ask whether the results that we report, and which we interpret as evidence of behavioral biases, could be explained in this way. We examine whether variables that are known to be correlated with the risk premium could be driving the explanatory power of our measured expectation errors and revisions. We show that, while

² Ludvigson and Ng (2009) extend the model of excess returns estimated by Cochrane and Piazzesi (2005) to include real factors, with the same interpretation.

lagged yield spreads and some macroeconomic variables do indeed have significant explanatory power for excess bond returns, they are almost orthogonal to the expectation errors and revisions that we infer from the term structure. This suggests that investors are subject to behavioral biases, but at the same time are risk averse, where the risk premium is time-varying.

In the following section, we summarize the theoretical background of the EH and the decomposition of excess returns into expectations error and expectations revision components. In Section 3, we describe the LSN and conservatism biases and their testable implications for expectation errors and expectation revisions. In Section 4, we report the empirical results. In Section 5, we investigate the role of a time-varying risk premium. In Section 6, we conduct out-of-sample tests of predictability. In Section 7, we assess the economic value of the predictive power of the behavioral variables for trading strategies that exploit these biases. In Section 8 we allow for a time varying-risk premium and test whether, after controlling for risk, behavioral trading strategies earn positive excess returns. Section 9 concludes.

2. Theoretical Background

Consider an n -period zero coupon bond with unit face value, whose price at time t is P_t^n . The yield to maturity of the bond, Y_t^n , satisfies the relation

$$P_t^n = \frac{1}{(1 + Y_t^n)^n} \quad (1)$$

or, in natural logarithms,

$$p_t^n = -ny_t^n \quad (2)$$

where $p_t^n = \ln P_t^n$ and $y_t^n = \ln(1 + Y_t^n)$. If the bond is sold before maturity then the log m -period holding period return, $r_{t,t+m}^n$, where $m < n$, is defined as the change in log price, $p_{t+m}^{n-m} - p_t^n$, which using (2) can be written as

$$\begin{aligned} r_{t,t+m}^n &= p_{t+m}^{n-m} - p_t^n \\ &= ny_t^n - (n-m)y_{t+m}^{n-m} \end{aligned} \quad (3)$$

The EH states that, conditional on the current information set, the expected m -period return for two bonds of different maturities, n_1 and n_2 , should be equal for all m except for the difference in time-invariant risk premia:

$$E[r_{t,t+m}^{n_1} | \Omega_t] = E[r_{t,t+m}^{n_2} | \Omega_t] + \phi_m^{n_1} - \phi_m^{n_2} \quad (4)$$

where $E[\cdot | \Omega_t]$ is the expectation conditional on the time- t information set, Ω_t , and $\phi_m^{n_1}$ and $\phi_m^{n_2}$ are the constant m -period risk premia on the two bonds. This gives rise to a number of implications concerning the relationship between the current yield spread (the difference between the yields of long and short maturity bonds) and (a) the change in the long yield over the life of the short bond and (b) the cumulative change in the short yield over the life of the long bond. Empirical tests of these implications invariably lead to a strong rejection of the EH (see, for example, Campbell and Shiller, 1991).

To explore the reasons behind the rejection of the EH, it is useful to recast the EH in terms of the holding period return relative to a risk free investment. By setting $n_1 = n$ and $n_2 = m$ in (4), we can define the excess return as the difference between the uncertain m -period return on an n -period bond (the long bond) and the certain m -period return on an m -period bond (the short bond):

$$\begin{aligned}\rho_{t,t+m}^n &= r_{t,t+m}^n - r_{t,t+m}^m \\ &= ny_t^n - (n-m)y_{t+m}^{n-m} - my_t^m\end{aligned}\quad (5)$$

Under the EH, this excess return should be unforecastable. In the equity market, tests of behavioral biases have focused on the time series properties of abnormal returns. In particular, it has been widely reported that unexpected equity returns display positive serial correlation at short horizons (momentum) and negative serial correlation at longer horizons (reversals) (see, for example, Jegadeesh and Titman, 1993).

The evidence of momentum and reversals in equity returns is, in principle, consistent with a behavioral explanation of expectations formation. However, without further assumptions, it is not possible to directly test such an explanation in the equity market using expectation errors and expectation revisions since these are not separately identifiable in equity returns. In the bond market, however, we have the advantage that the cash flows (i.e. the coupons and the face value of the bond) are known with certainty and so we are able to decompose excess bond returns into the part due to errors in expectations about future short yields and the part due to expectation revisions about future short yields. This allows us to directly test the implications of behavioral biases for the time series properties of realized expectation errors and for the time series of revisions in expectations.

To do so, we first write the excess return in terms of expectations about the ‘fundamental’ short yield. By setting $n_1 = m = n$ and $n_2 = 1$ in (3) and (4), we can write the yield on an n -period bond as the sum of current and expected future short yields over the n -period life of the long bond:

$$y_t^n = \frac{1}{n} \sum_{i=0}^{n-1} E_t y_{t+i}^1 + \phi_n^n \quad (6)$$

Substituting into (5) then leads to the following decomposition of excess returns, which is the basis for the tests of the behavioral models set out in Section 4:

$$\begin{aligned}\rho_{t,t+m}^n &= \left(y_t^1 + \sum_{i=1}^{n-1} E_t y_{t+i}^1\right) - \left(y_{t+m}^1 + \sum_{i=1}^{n-m-1} E_{t+m} y_{t+m+i}^1\right) - \left(y_t^1 + \sum_{i=1}^{m-1} E_t y_{t+i}^1\right) + \phi_m^n - \phi_m^m \\ &= -(y_{t+m}^1 - E_t y_{t+m}^1) - \sum_{i=1}^{n-m-1} (E_{t+m} y_{t+m+i}^1 - E_t y_{t+m+i}^1) + \phi_m^n - \phi_m^m\end{aligned}\quad (7)$$

Equation (7) states that, under the EH, over the life of an m -period bond, the difference between the uncertain return on an n -period bond and the certain return on the m -period bond can be decomposed into (i) the difference between the short yield at time $t + m$ and the market's expectation at time t of that short yield (i.e. an expectation error term), (ii) the revision in expectations between time t and time $t + m$ of the short yield between time $t + m + 1$ and time n , the maturity of the long bond (i.e. an expectation revision term) and (iii) a constant risk premium term. Intuitively, (i) implies that, if the short yield at time $t + m$ is higher than expected at time t , this will result in a lower holding period return for the n period bond while (ii) implies that, if the expectation of short yields between time $t + m + 1$ and time n are higher at time $t + m$ than at time t , then this too will depress holding period returns on the n period bond bought at time t and held until time m . Under the EH, the expected value of both the expectation error term and the expectation revision term should be zero. The evidence against the EH, which is a joint hypothesis of rational expectations and constant risk premia, can therefore be thought of as potentially arising from systematically biased expectations that give rise to predictability in the expectation errors and expectation revisions of the short yield. In the following section, we describe two behavioral models that have been used to explain such biases in the equity market.

3. Behavioral Models of Expectations Formation

A. The Law of Small Numbers

The Law of Small Numbers (LSN) describes the belief that a randomly drawn sample of data will reflect the characteristics of the population from which it is drawn more closely than sampling theory would predict. The LSN is related to two specific behavioral biases that have been documented in the psychology literature. The first is ‘base rate neglect’, which describes the finding that subjects put too little weight on the unconditional probability of observing a particular sample. The second is ‘sample size neglect’, which describes the finding that subjects overestimate the statistical relevance of information that is contained in the sample (see Tversky and Kahneman, 1971). Both base rate neglect and sample size neglect cause subjects to overweight (compared to a Bayesian) the importance of a given sample of data when making inferences about the population from which it is drawn. Barberis et al. (1998) and Rabin (2002) derive the implications of the LSN for returns in equity markets and show that it results in momentum in abnormal returns in the short run, an empirical feature of equity returns that is well documented. The LSN has similar implications for the bond market. Assume that the short yield follows an autoregressive process with i.i.d. shocks and that this model is known. Under the LSN, agents will be too confident (compared to a Bayesian) that they will see equal numbers of positive and negative shocks in short samples of data. This implies that, if the shock is negative in one period so that the forecast error is negative, investors will expect the following period’s shock to be positive with probability greater than 50%. Thus, their short yield forecast will be higher than is implied by the autoregressive model. However, under the true model, the next period’s innovation is positive with 50 percent probability and hence investors will experience a second negative surprise

with more than 50 percent probability. The LSN therefore predicts that there will be positive short run serial correlation in one-step ahead forecast errors for the short yield.

In terms of the decomposition given by (7), this bias leads to positive serial correlation in the first component of the excess return (the expectations error component), contributing to positive serial correlation in excess returns over short horizons. This leads to the following testable hypothesis:

H_1 : $y_{t+m}^1 - E_t y_{t+m}^1$ is positively serially correlated for small values of m .

So far we have described the implications of the LSN for one-step ahead forecast errors given agents' beliefs about the model that generated the sample. But the fact that subjects tend to overweight (compared to a Bayesian) the importance of a given sample of data also has implications for how subjects revise their beliefs about the model in the light of runs of data. If agents expect relatively small samples to closely reflect population moments then this implies that, when they observe a series of observations that do not accord with their original beliefs, they too readily interpret this as evidence that their original beliefs were incorrect. They therefore update their beliefs about the model too quickly relative to a Bayesian. Over time, this leads to fluctuation in agents' beliefs about the model parameters around their true values. Assuming that successive samples are drawn randomly, beliefs about the model parameter values will therefore exhibit negative serial correlation.

In order to test this long term implication of the LSN in the bond market, we note that, if beliefs about the model parameters exhibit negative serial correlation, forecast errors (which in part reflect model errors) will inherit this negative serial correlation. We cannot expect to detect this using short term forecasts because, although they reflect beliefs about the model, short term forecast errors will be dominated by the short term implications of the LSN

discussed above. In order to detect negative serial correlation in forecast errors that result from model error, we focus on errors in long term forecasts since it is these that will more clearly reflect mistaken beliefs about the systematic part of the model. The LSN therefore predicts that there will be negative serial correlation in the long term forecast and so we test the following hypothesis:

$H_2: y_{t+m}^1 - E_t y_{t+m}^1$ is negatively serially correlated for large values of m .

B. The Conservatism Bias

‘Conservatism’ describes a subject’s response to new information. It describes the possibility that individuals are too slow to revise their beliefs, effectively attaching too much weight to their prior beliefs about the true model and too little weight to new information. Daniel et al. (1998) build on the closely related ‘overconfidence bias’, which has similar testable implications. They show that this bias can lead to underreaction to news as agents’ expectations following the news are not immediately revised to the full extent that would be justified by Bayesian updating. However, over time agents learn of their initial underreaction and so there are subsequent revisions in agents’ expectations that are of the same sign as the initial response to the news announcement. This process is consistent with evidence of momentum in returns and is further confirmed in the equity market by evidence of underreaction to public news such as earnings announcements.

The existence of conservatism implies that the revisions in expectations of future short yields that we observe each period will typically be too small resulting in further revisions of the same sign in subsequent periods.³ This leads to a third testable hypothesis:

³ If investors learn of their initial underreaction immediately, then they will adjust their expectations the following period, leading to serial correlation in one-period expectation revisions. More generally, however, it is possible that this process could extend for more than

H_3 : $E_{t+m}y_{t+m+i}^1 - E_t y_{t+m+i}^1$ is positively serially correlated for small values of m and i .

In the following section, we examine the evidence for momentum and return reversals in the bond market and report the results of testing hypotheses H_1 , H_2 and H_3 .

4. Empirical Evidence

A. Data

We use the synthetic monthly zero-coupon bond yields on US Treasury securities for the period January 1952 to December 2012. We update the zero coupon bond yield data estimated by Bulkley, Harris and Nawosah (2011), which ended in December 2009, to December 2012 and use the extended data set in this paper. The data are continuously compounded and recorded as annualized percentages.

Following Fama (2006) and Bulkley, Harris and Nawosah (2011), we include a dummy variable in all the regressions that we estimate to capture the significant structural break in bond yields of all maturities that occurred in 1980-81. Structural stability tests suggest a breakpoint between June 1981 and June 1982 depending on the regression estimated. For consistency, we assume a common breakpoint at December 1981. To investigate the robustness of our findings, we report results for the post-break sample from January 1982 to December 2012. We additionally consider the much shorter sample from January 2008 to December 2012, which follows the recent financial crisis.

B. Evidence on Momentum and Return Reversals

We first investigate whether our sample exhibits the stylized features of short term momentum and long term return reversals that have been documented in many asset markets

one period and so we test for serial correlation in m -period expectation revisions.

and many countries (see for example Cutler, Poterba, and Summers, 1991; Asness, Moskowitz, and Pedersen, 2013). In particular, we estimate the degree of serial correlation in excess holding period returns using the following regression:

$$\rho_{t,t+m}^n = \alpha_1 + \gamma_1 D_t + \beta_1 \rho_{t-m,t}^n + \varepsilon_{1,t+m} \quad (8)$$

where the m -period excess holding period return for an n -period bond, $\rho_{t,t+m}^n$, is defined by equation (7) and D_t is a dummy variable that is set to one after December 1981 and zero otherwise. Table 1 reports the results of estimating regression (8) for $n = 2, 3, 6, 9, 12, 24, 36, 48, 60$ and 120 months and $m = 1, 2, 3, 6, 9, 12, 24, 36, 48$ and 60 months for the full 1952-2012 sample.⁴ The regression is estimated by OLS and standard errors are computed using the Newey and West (1987) estimator to allow for the fact that the dependent variable is overlapping. To limit the size of the table, we only report the estimated slope coefficients and omit the intercept and dummy coefficients.⁵ Newey-West corrected t -statistics are reported in parentheses. It is worth noting at the outset that Table 1 has a banding structure that clearly highlights the pattern of results expected: the first column has elements that are significantly positive (consistent with short-run momentum), whereas the bottom-right triangle groups coefficients that are significantly negative (consistent with long-term reversals). In particular, for the shortest holding period of one month (in the first column), we find very significant positive serial correlation in excess holding period returns for all but the longest bond maturity. This evidence complements that of Asness, Moskowitz and Pedersen (2013), who report strong evidence of momentum in real bond yields for one month holding periods for 10 countries. However, in contrast with Cutler, Poterba and Summers (1991), who find

⁴ Note that y_{t+m}^{n-m} needs to be approximated in a number of cases due to the unavailability of the certain maturities in our dataset. Here and elsewhere in the paper, we linearly interpolate between the yields of adjacent maturities to approximate the missing yields.

⁵ We note that the coefficient on the dummy variable is strongly significant in most cases, suggesting the importance of the structural break.

momentum in government bond returns for holding horizons of up to one year, we do not find significant evidence of momentum at horizons of longer than one month.

For longer holding periods between 24 and 120 months, there is very significant negative serial correlation in excess holding period returns for longer maturity bonds, suggesting that there are return reversals in excess holding period returns in the bond market. This is stronger evidence than reported by Cutler, Poterba and Summers (1991), who find using US data 1960-1988 only weak evidence of negative autocorrelation at longer lags (although they do find negative autocorrelation at longer horizons for a sample of 12 other countries). The pattern of momentum and return reversals in our data is similar to that reported in U.S. equity data (see, for example, Jegadeesh and Titman, 1993), although the horizon over which there is significant momentum in excess returns in our sample is shorter than the six to twelve months typically found in the equity market.

To conserve space, the results of estimating regression (8) for the two sub-samples are reported in the web appendix. Table A.1 reports the results for the 1982-2012 sample, while Table A.2 reports the results for the 2008-2012 sample. As documented in the web appendix, the evidence largely confirms the existence of momentum and reversals in the two shorter periods.

[Table 1]

C. Expectation Errors and the LSN

We next test the implications of the LSN for expectation errors. In Tables 2 and 3 we report the results of estimating the following regression:

$$y_{t+m}^1 - E_t y_{t+m}^1 = \alpha_2 + \gamma_2 D_t + \beta_2 (y_t^1 - E_{t-m} y_t^1) + \varepsilon_{2,t+m} \quad (9)$$

where $E_t y_{t+m}^1 = (m+1)y_t^{m+1} - my_t^m$ is the forecast of y_{t+m}^1 that is implicit in the current term structure of interest rates. The evidence for the short-run implications of the LSN (hypothesis H_1) comes from small values of m , while for the long-run implications (hypothesis H_2), we are interested in larger values of m . Table 2 reports the results of estimating (9) for $m = 1, 2, 3, 4, 5$ and 6 months. For the full 1952-2012 sample, there is very significant positive serial correlation in one-step ahead expectation errors for the short yield for horizons of one and two months. For longer horizons, there is no significant serial correlation. In the 1982-2012 sample, the evidence of positive serial correlation in one-step ahead expectation errors is even stronger and is statistically significant up to the 6-month horizon. In the 2008-2012 sample, there is evidence of positive serial correlation in one-step ahead expectation errors up to the 6-month horizon, although it is generally weaker than in the 1982-2012 sample.

[Table 2]

Table 3 reports the results of testing the long run implications of the LSN using regression (9) with lags of $m = 9, 12, 18, 24, 36, 48, 60$ and 120 months. For the full 1952-2012 sample, the estimated slope coefficient is insignificant for lags of 9, 12 and 18 months, becoming significantly negative for horizons of 24, 36, 48, 60 months. It is not surprising that there is an interval in which the slope coefficient is insignificant since at intermediate horizons the short and long term implications of the LSN act in opposite directions. The evidence suggests that it is at the 4-year horizon that the impact of fluctuating beliefs about model parameters is strongest. Under the LSN we explain this negative serial correlation as a result of agents revising their model too much in response to recent data, resulting in beliefs about parameters that fluctuate about their true value. Long run forecast errors will, in part, reflect model error and hence inherit the negative serial correlation. The effect becomes attenuated at longer horizons because, as the sample size increases, the sample moments more closely match the

population moments and so agents do not see a need to revise their beliefs about the true model. The pattern of evidence in the 1982-2012 sample is generally similar, although the slope coefficient becomes significantly positive for the 120-month bond. We do not report results for the 2008-2012 sample because it is too short for reliable inference given the long horizons considered.

[Table 3]

D. Expectation Revisions and the Conservatism Bias

Table 4 reports the results of the test of the conservatism bias, hypothesis H_3 , which is that expectation revisions are positively serially correlated at short lags. To test this hypothesis, we estimate the following regression:

$$E_{t+m}y_{t+m+i}^1 - E_t y_{t+m+i}^1 = \alpha_3 + \gamma_3 D_t + \beta_3 (E_t y_{t+i}^1 - E_{t-m} y_{t+i}^1) + \varepsilon_{3,t+m} \quad (10)$$

where $E_t y_{t+k}^1 = (k+1)y_t^{k+1} - ky_t^k$. The regression is estimated for horizons $i = 1$ to 12 months and lag $m = 1$ month. For the 1952-2012 sample, consistent with hypothesis H_3 , there is positive serial correlation in the one-step ahead expectation revisions for the short yield at all horizons except one month. The pattern of serial correlation increases with the time horizon up to three months and then generally declines. In all cases, there is statistically significant positive serial correlation in expectation revisions, strongly supporting the prediction of the conservatism bias. The results for the 1982-2012 sample also support the conservatism bias, although with an even greater degree of positive serial correlation for the 1-month and 2-month horizons. For the 2008-2012 sample, positive serial correlation is observed for horizons of one to five months. For the remaining horizons, the serial correlation is not significantly different from zero.

[Table 4]

5. Time-Varying Risk Premia

The measures of expectation errors and revisions employed in the tests reported above are inferred from the term structure under the assumption of a constant risk premium. However, if the risk premium is in fact time-varying, this will introduce measurement error into the inferred expectation errors and revisions. In this section, we examine whether the dynamic properties of expectation errors and revisions can be explained by time-variation in the risk premium. We re-estimate equations (9) and (10) adding a range of financial and macroeconomic risk factors that have been suggested as proxies for the risk premium. In particular, we consider Cochrane and Piazzesi's (2005) five forward rates, which are widely used in the recent empirical literature, and a set of standard macroeconomic variables used in prior research. These include the inflation rate (π) as measured by the change in the log of the US CPI index, a measure of business cycle activity or output gap (gap) due to Cooper and Priestley (2009), a measure of economic growth (g) defined as the change in the log of the industrial production index, the change in the unemployment rate (ΔU), and a measure of bond market volatility (Vol) constructed using the rolling 12-month standard deviation of the log change in the 10-year Treasury yield.^{6,7} The changes are measured over the previous

⁶ See for example Cochrane and Piazzasi (2002), Kim and Moon (2005), Cooper and Priestley (2009), Ludvigson and Ng (2009), Duffee (2012) and the references therein. The recent macro-finance literature also recognizes the importance of macroeconomic variables related to inflation and real activity as determinants of bond risk premia, see for example Ang and Piazzesi (2003), Rudebusch and Wu (2004), Ang, Bekaert, and Wei (2008), Christensen, Lopez, and Rudebusch (2010) and Joslin, Pribsch, and Singleton (2012) amongst others.

⁷ Following Cooper and Priestley (2009), gap is measured by the deviation of the log of the industrial production index from a trend that includes both a linear and a quadratic component.

month.⁸

The augmented versions of (9) and (10) take the following form:

$$FE_{t+m} = \alpha_2 + \gamma_2 D_t + \beta_2 FE_t + \delta_2' CP_t + \theta_2' Macro_t + \varepsilon_{2,t+m} \quad (11)$$

$$ER_{t+m,i} = \alpha_3 + \gamma_3 D_t + \beta_3 ER_{t,i} + \delta_3' CP_t + \theta_3' Macro_t + \varepsilon_{3,t+m} \quad (12)$$

where $FE_{t+m} = y_{t+m}^1 - E_t y_{t+m}^1$ and $ER_{t+m,i} = E_{t+m} y_{t+m+i}^1 - E_t y_{t+m+i}^1$ are the dependent variables from equations (9) and (10), respectively. $FE_t = y_t^1 - E_{t-m} y_t^1$ and $ER_t = E_t y_{t+i}^1 - E_{t-m} y_{t+i}^1$ represent the behavioral predictors from equations (9) and (10), respectively. Henceforth, the acronyms FE and ER are used to refer to forecast error and expectation revision, respectively. CP is a vector containing Cochrane and Piazzesi's (2005) one-year yield and 2-5 year forward rates, $CP = [f_{CP}^1 \ f_{CP}^2 \ f_{CP}^3 \ f_{CP}^4 \ f_{CP}^5]'$. $Macro$ is a vector that groups together the macroeconomic variables, $Macro = [\pi \ gap \ g \ \Delta U \ Vol]'$. δ_2 , θ_2 , δ_3 and θ_3 are 5×1 vectors of regression coefficients.⁹

To save space, we report the results only for the full 1952-2012 sample. We also restrict attention to selected horizons and refer to the web appendix for full-length tables containing results for the full set of horizons considered earlier in the paper. Tables 5 and 6 report the results from estimating equation (11) for a variety of specifications for short and long horizons, respectively. Panel A of each table presents results for the specification including the CP forward rates only, Panel B for the specification including the macro variables only, and Panel C for the specification including both the CP forward rates and the macro

⁸ The results that follow are similar when changes are measured on a year-on-year basis. Also, it makes little difference to the conclusions when we lag macro variables by a month to account for the delay in the release of macroeconomic data or use annual averages of the forward rates and selected macro variables instead in our regressions.

⁹ We tested specifications of equations (11) and (12) that include Ludvigson and Ng's (2009) latent factors in place of the observed macroeconomic risk factors for the 1964-2007 period for which these factors are available, and the conclusions that we draw are generally similar.

variables. Comparing Table 2 with Table 5 (and Table B.1 in the web appendix), we see that at short horizons the risk proxies substantially improve the \bar{R}^2 in all three specifications and at almost all horizons. Although their significance varies between horizons their importance is nevertheless evident across the three specifications. However, it is notable that while the risk proxies are statistically significant, they appear to be approximately orthogonal to the behavioral variable. In particular the size and statistical significance of the coefficient on FE at the 1-month horizon is only very marginally affected, irrespective of the risk proxies used.¹⁰

[Table 5]

In Table 6 (and Table B.2 in the web appendix), we see that the predictive power of the regressions improves substantially, suggesting that the risk variables do matter at long horizons. However, the significance of the FE variable is again essentially unchanged at horizons of between two and five years in both samples, irrespective of the risk proxies used.

[Table 6]

The results of estimating equation (12) for horizons $i = 1, 2, 3, 6, 9$ and 12 months are reported in Table 7. In Table B.3 in the web appendix, we report the results for all horizons from 1 to 12 months. The risk premium variables also appear to be important determinants of expectation revisions. However, the slope coefficient on the ER variable remains significant at all horizons except one month and across all three specifications. These results are similar to those reported in Table 4.

[Table 7]

¹⁰ We also note that in the post-1981 sample, the coefficient on FE is found to be strongly significant at all horizons up to four months. The results for the post-2007 sample are weaker but the coefficient on FE remains significant in several cases.

To summarize, the explanatory power of the behavioral variables is only marginally affected by the introduction of proxies for a time-varying risk premium. Although a time-varying risk premium and behavioral models are sometimes viewed as competing explanations, there is no reason for these two models to be mutually exclusive. In particular, there is no reason why agents who exhibit behavioral biases in forming expectations should not also be risk averse in an environment where risk is time-varying.

6. Out-of-Sample Forecasting Performance

In this section we examine whether an investor trading in real time, with only past data available to estimate a model of expectation errors, could have exploited these behavioral biases to predict forecast errors. In this out-of-sample analysis, forecasts are generated using a recursive estimation scheme with an initial window size of 60 months. Model parameters are estimated every month using data up to month t , that is, forecast errors and revisions in month t are regressed on lagged predictors, measured in month $t-m$ or month $t-i$ depending on the regression used, and then out-of-sample m -step, or i -step, ahead predictions are generated using the estimated parameters and predictors measured in month t . Again, a variety of regression specifications for equations (11) and (12) are used to construct the out-of-sample forecasts.

Since our main concern is the relative performance of the behavioral variables versus the risk premium proxies, we apply two out-of-sample forecast encompassing tests to examine the incremental predictive ability of the two sets of variables. The first test is based on Mincer-Zarnowitz (1969) regressions of the following form:

$$FE_{t+m} = \vartheta_1 + \vartheta_2 \widehat{FE}_{t+m}^{FE} + \vartheta_3 \widehat{FE}_{t+m}^{Risk} + \varepsilon_{t+m} \quad (13)$$

$$ER_{t+m,i} = \kappa_1 + \kappa_2 \widehat{ER}_{t+m,i}^{ER} + \kappa_3 \widehat{ER}_{t+m,i}^{Risk} + \varepsilon_{t+m,i} \quad (14)$$

where \widehat{FE}_{t+m}^{FE} is the time $t+m$ forecast of the forecast error based on equation (9) and $\widehat{FE}_{t+m}^{Risk}$ is the forecast based on equation (11) including only the risk factors. $\widehat{FE}_{t+m}^{Risk} = \widehat{FE}_{t+m}^{CP}$ when only the CP forward rates are used as predictors, $\widehat{FE}_{t+m}^{Risk} = \widehat{FE}_{t+m}^{Macro}$ when only the macro variables are used and $\widehat{FE}_{t+m}^{Risk} = \widehat{FE}_{t+m}^{CP+Macro}$ when both the CP forward rates and the macro variables are used. Similarly, $\widehat{ER}_{t+m,i}^{ER}$ is the predicted forecast revision based on equation (10) and $\widehat{ER}_{t+m,i}^{Risk}$ is the prediction based on equation (12) including only the risk factors. $\widehat{ER}_{t+m,i}^{Risk} = \widehat{ER}_{t+m,i}^{CP}$ when only the CP forward rates are used as predictors, $\widehat{ER}_{t+m,i}^{Risk} = \widehat{ER}_{t+m,i}^{Macro}$ when only macro variables are used, and $\widehat{ER}_{t+m,i}^{Risk} = \widehat{ER}_{t+m,i}^{CP+Macro}$ when both the CP forward rates and macro variables are used. We make inferences about the predictive ability of the behavioral variables and the risk proxies by comparing the relative size and significance of the slope coefficients of the competing forecasts entering equations (13) and (14).

The second test is based on the ENC-NEW statistic suggested by Clark and McCracken (2001). The null hypothesis of the test is that the restricted model forecasts encompass the unrestricted model forecasts. To implement this test, we compare out-of-sample forecasts from an unrestricted specification of equation (11) (equation (12)) that includes both FE (ER) and a combination of the risk factors to those based on a restricted model that includes only the risk factors. Rejection of the null hypothesis suggests that the unrestricted model contains incremental information or, in other words, the behavioral variables contain additional information that is not contained in the risk factors. Bootstrapped critical values provided in Clark and McCracken (2001, 2005) are used to evaluate the statistical significance of the test statistic. We also report the out of sample \bar{R}^2 from the above encompassing regressions.

The results of the encompassing tests are summarized in Tables 8, 9 and 10. The results reported in Table 8 show that both FE -based and risk-based forecasts have predictive power. For both sets of forecasts, the slope coefficient is positive and significantly greater than zero

at the 1-month horizon, whichever set of risk factors is used. The ENC-NEW statistic is also strongly statistically significant at the 1-month horizon for all three specifications. These results confirm that the in-sample short run LSN evidence reported in Table 5 also holds out-of-sample suggesting that the risk-based models do not encompass the *FE*-based model.

[Table 8]

The results in Table 9 confirm the evidence in favor of the long run implications of the LSN for all horizons found in Table 6, except 60 months. The slope coefficient of the risk-based forecasts is significantly greater than zero in only one case and even turns negative in many cases, whereas the slope coefficient of the behavioral forecasts is generally significant (and positive) for horizons of 24, 36 and 48 months and, to some extent, 18 months. This suggests that the behavioral forecasts dominate the CP forecasts for these horizons. The ENC-NEW test statistic is strongly significant for horizons of 18 months and longer, providing evidence that there is incremental information in the *FE* variable.¹¹

[Table 9]

Table 10 reports the out-of-sample results for the *ER*-based forecasts relative to the risk-based forecasts. The encompassing regression results in Panel A confirm the in-sample findings in Panel A of Table 7. Although the slope coefficient on the CP forecasts is significantly positive in several cases, we see that the risk-based forecasts do not encompass the *ER*-based forecasts for most horizons. This is strongly supported by the ENC-NEW test. The results in Panel B and Panel C are slightly weaker but they nevertheless overwhelmingly reject the hypothesis that the risk-based forecasts encompass the *ER*-based forecasts.

¹¹ Comparing the encompassing \bar{R}^2 to the \bar{R}^2 from the Mincer-Zarnowitz regression that includes only risk-based forecasts (not reported), we observe an increase in the forecasting performance generally, again indicating that the *FE* variable contains incremental information.

[Table 10]

7. The Economic Value of Predictability

Given that the behavioral variables forecast expectation errors and expectation revisions, both in-sample and out-of-sample, it is natural to ask whether bond market investors could have profitably exploited this predictability. In this section we report Sharpe ratios for trading rules based on the behavioral variables and risk proxies. We also assess the significance of the difference in Sharpe ratios between the different strategies using the Jobson and Korkie (1981) methodology with the correction proposed in Memmel (2003). We start by reporting real time trading profits for equally-weighted portfolios of selected bonds.¹² We use two specific trading rules, both of which are self-financing. The first rule (Strategy A) is a long/short strategy that involves holding one unit of a portfolio of long bonds and shorting one unit of the short bond if the portfolio's predicted excess return is positive, and vice versa. The second rule (Strategy B) involves taking a long or short position in a portfolio of long bonds (and an opposite position of equal size in the short bond), with the size of the position suggested by the portfolio's predicted excess return, following Cochrane and Piazzesi (2005). We produce out-of-sample forecasts of excess returns based on the same recursive procedure with a 5-year initial estimation window as in Section 6. The m -step ahead forecasts are generated recursively from the following excess return regression:

$$\frac{1}{N} \sum_{n \in S} \rho_{t,t+m}^n = \alpha + \gamma D_t + \nu X_t + \varepsilon_{t+m} \quad (15)$$

where the dependent variable is the difference between the average m -month return across N n -month long bonds and the certain return on the short m -month bond and represents the excess return on the portfolio of these N bonds. m is the holding horizon and we consider

¹² The analysis is also performed on individual maturity bonds and, as expected, the pattern of the results is similar.

horizons $m = 1, \dots, 12, 24, 36, 48$ and 60 months. S collects the maturity values (in months) for the full set of long bonds used to form the portfolios: $S = [2\ 3\ 6\ 9\ 12\ 24\ 36\ 48\ 60\ 120]$. The values taken by n depend on the horizon under consideration. The number of bonds used decreases with horizon: all 10 bonds are used for $m = 1$, the 9 bonds with maturities of 3 to 120 months for $m = 2$, and so on. The right hand side variables, X_t , are either the behavioral variables or the risk proxies. The strategy that exploits expectation errors (the *FE* strategy) is based on forecasts from equation (15) when $X_t = FE_t$, as defined above, and the strategy that trades on expectation revisions (the *ER* strategy) is based on forecasts when $X_t = ER_t$. Similarly, risk-based *CP* and *CP+Macro* strategies are based on forecasts when $X_t = CP_t$ and $X_t = [CP_t\ Macro_t]'$, respectively.

Once the forecasts of excess returns have been produced, we proceed to calculate the trading strategy profits. Under Strategy A, the profits are $\left(\frac{1}{N}\sum_{n \in S} \rho_{t,t+m}^n\right)$ when the portfolio's predicted excess return is positive, and $\left[\left(\frac{1}{N}\sum_{n \in S} \rho_{t,t+m}^n\right) \times (-1)\right]$ when the portfolio's predicted excess return is negative. Under Strategy B, the profits are calculated as:

$$\left(\frac{1}{N}\sum_{n \in S} \rho_{t,t+m}^n\right) \times E_t\left(\frac{1}{N}\sum_{n \in S} \rho_{t,t+m}^n\right) \quad (16)$$

where the size of the position is equal to $E_t\left(\frac{1}{N}\sum_{n \in S} \rho_{t,t+m}^n\right)$, expressed as an annual percentage to be consistent with Cochrane and Piazzesi (2005). If the sign of the portfolio's excess return is correctly predicted, Strategy B results in a profit, otherwise it results in a loss.

The trading rule profits from the two strategies are thus a multiple of \$1 each month.

After calculating the time series of the out-of-sample trading rule returns for the various strategies, we evaluate their performance using Sharpe ratios. We also assess the significance of the difference in Sharpe ratios between the behavioral strategy and the two risk-based

strategies using the corrected Jobson and Korkie (1981) test.¹³ Given that our trading rule returns are serially correlated, we pre-whiten the time series by fitting appropriate ARMA models prior to performing the test.

Table 11, Panel A, compares the performance of the *FE* strategy to the *CP* and *CP+Macro* strategies for holding horizons $m = 1, 2, 3, 6, 9, 12, 24, 36, 48$, and 60 months. For Strategy A, we see that the *FE* strategy attains a higher Sharpe ratio than the risk-based strategies in all cases and the difference is statistically significant at the 5% level in all but two cases. For Strategy B, again, the *FE* strategy always achieves a higher Sharpe ratio than the risk-based benchmarks but the differences in performance are generally larger than under Strategy A. The difference in Sharpe ratios between the *FE* strategy and both the *CP* and *CP+Macro* strategies is statistically significant for all horizons.

[Table 11]

Panel B of Table 11 presents the results for the *ER* strategy against the risk-based strategies for horizons $m = 2$ to 12 months. We can see from both panels that the Sharpe ratios of the *ER* strategy exceed those of the risk-based strategies in most cases. For Strategy A, the results are mixed for horizons up to 6 months. Comparing the results of the *ER* strategy with those of the *CP* strategy, we see that the *ER* strategy has higher Sharpe ratios for all horizons, although the differences are significant only at horizons longer than 6 months. However, the Sharpe ratios of the *ER* strategy are significantly higher than the *CP+Macro* strategy in all cases except 2 and 4 months. The results for Strategy B present a stronger picture: the *ER* strategy

¹³ We also employ the robust studentized time series bootstrap procedure of Ledoit and Wolf (2008) using a grid of block sizes that includes the value of the overlap of the data. The results are qualitatively similar to those of the Jobson and Korkie (1981) test.

achieves higher Sharpe ratios in all cases and the differences are all statistically significant.¹⁴

Overall, the superior performance of the behavioral strategies suggests that the predictive ability of the behavioral variables is robust and that trading on the basis of expectation errors and revisions would add economic value beyond that which would be expected in compensation for risk.

8. Factor Mimicking Portfolios and Strategy Alphas

To further check the robustness of our results, we construct portfolios mimicking the risk factors and directly test whether the alphas of our behavioral strategies with respect to these mimicking portfolios are significantly different from zero. Factor mimicking portfolios are estimated by projecting each original (non-traded) risk factor onto the span of tradable asset returns. We construct six mimicking portfolios, one for each of our five macro variables and one for the *CP* factor. Only one portfolio is constructed to mimic the information in the *CP* forward rates. The *CP* single forward-rate factor (see Cochrane and Piazzesi (2005)), which is a linear combination of the forward rates, is used as the original risk factor in the construction. The single factor, denoted *CPF*, is estimated out-of-sample recursively and measured at the one-month horizon. Regarding the choice of base assets, we consider six equity portfolios and two bond portfolios from the wide range of assets commonly used in the related literature. The equity portfolios are the six Fama-French benchmark portfolios sorted on size and book-to-market and the bond portfolios include a portfolio of intermediate

¹⁴ We also extended the analysis to calculate the utility-based certainty equivalent (*CE*) return gains for each strategy (see, for example, Brennan and Xia (2004) and Campbell and Thompson (2008)). Using a range of values for the coefficient of relative risk aversion, we find that the *CE* gains for the behavioral strategies are almost always positive and considerable in size (well over 100 basis points in most cases) compared to the (often negative) values obtained for the risk-based strategies.

Treasury bonds and a portfolio of long Treasury bonds.^{15, 16}

Following Breeden, Gibbons and Litzenburger (1989) and Adrian, Etula and Muir (2014), we estimate the asset weights for the mimicking portfolios by regressing each risk variable on the excess returns (over the risk free rate) of each of the base assets.¹⁷ For each risk variable $Y = CPF, \pi, gap, g, \Delta U$ or Vol , we run the following regression:

$$Y_t = \alpha_Y + B_Y' R_{p,t} + \varepsilon_{Y,t} \quad (17)$$

where R_p is a vector that collects the monthly excess returns of the base assets, α_Y is the intercept and B_Y is an 8×1 vector of slope coefficients. We normalize the sum of the portfolio weights to one in each case. The monthly return on the mimicking portfolio for risk variable Y is then given by:

$$M_{Y,t} = B_Y' R_{p,t}. \quad (18)$$

Once we have the factor mimicking portfolio returns, we evaluate the performance of our behavioral strategies using the alphas from time-series regressions of the following form:

$$R_{s,m,t} = \alpha_{s,m} + \zeta_{s,m} RXB_{m,t} + \Xi_{s,m}' M_t + \varepsilon_{s,m,t} \quad (19)$$

where $R_{s,m}$ denotes the out-of-sample trading rule returns for $s = FE$ or ER strategy for a horizon of m months, RXB_m denotes the excess returns on a portfolio of long-term Treasury

¹⁵ The six Fama-French portfolio returns are downloaded from Kenneth French's data library. In unreported results, we also incorporate industry sorted portfolios, a momentum portfolio and the US stock market portfolio (the CRSP value-weighted index of all stocks). The results are similar to those reported below when we use different permutations of the expanded set of base assets.

¹⁶ Note that, in estimating the portfolio that mimics CPF , we drop the intermediate bond portfolio that includes 2-5 year bonds which also enter the calculation of CPF . Excluding the long bond portfolio as well generates similar results.

¹⁷ Including a set of control variables (the term spread, default spread, short rate, dividend yield, lagged inflation, lagged growth and lagged stock market return) has little effect on the results.

bonds measured over the same m -month horizon (used as a proxy for the bond market portfolio) and \mathbf{M} is a vector containing the monthly mimicking portfolio returns, $\mathbf{M} = [M_{CPF} \ M_{\pi} \ M_{gap} \ M_g \ M_{\Delta U} \ M_{Vol}]'$. $\alpha_{s,m}$ is the intercept (alpha), $\zeta_{s,m}$ is the slope coefficient on RXB_m and $\Xi_{s,m}$ is a 6×1 vector of slope coefficients for the mimicking portfolios.¹⁸ In unreported results, we analyze the correlations between the mimicking portfolios and the original risk variables. The absolute correlations range from 0.10 to 0.25, which are consistent with the values reported in the literature. We also examine the correlations between the returns on the six mimicking portfolios. The absolute correlations range from 0.45 to 0.94. In order to avoid multicollinearity, we ensure that no correlation coefficient between the mimicking portfolios included in the regressions exceed a value of 0.8 and this is achieved by dropping M_{gap} and $M_{\Delta U}$. Furthermore, an examination of the mean-variance properties of the mimicking portfolios reveals that the mimicking portfolios are located much closer to the tangency portfolio than are the base assets. This is reflected in the higher Sharpe ratios of the mimicking portfolios.

[Table 12]

Owing to space constraints, we report results only for the long/short strategy that involves holding one unit of a portfolio of long bonds and shorting one unit of the short bond if the portfolio's predicted excess return is positive, and vice versa (Strategy A). It is worth noting that the results are much stronger for both *FE*- and *ER*-based strategies when the size of the

¹⁸ We perform further sensitivity checks on our results. While the mimicking portfolios in the paper use fixed weights, we also construct mimicking portfolios with time-varying weights (see, for example, Ferson and Harvey (1991)). We also experimented with alternative horizons over which the risk variables and the base asset returns are measured, in particular by matching the holding horizons of our trading strategies. We also estimated a variety of specifications for equation (19), with and without the excess returns on long bonds: for example, we introduce each mimicking factor by itself as well as evaluate the role of the macro portfolios separately from the *CP* portfolio. The main conclusions that we draw from the wide array of sensitivity analyses are unaltered. These results are available from the authors.

long or short position in a portfolio of long bonds varies with the portfolio's predicted excess return (Strategy B). Panel A of Table 12 presents the results of estimating equation (19) for the *FE*-based strategy for holding horizons $m = 1, 2, 3, 6, 9, 12, 24, 36, 48$, and 60 months. The bond market portfolio does very well in capturing the variation in the strategy returns. The beta estimates are positive and highly significant for horizons up to 24 months. Regarding the contribution of the portfolios constructed to mimic the risk factors related to *CP* and macroeconomic factors, a similar picture to the results reported in Tables 5-6 emerges. The loadings on the mimicking portfolios are significant in several cases, at least at the 10% level. M_π , M_g and M_{Vol} appear to be the most relevant return factors. Despite their importance, the bond market factor and the mimicking portfolios leave a significant fraction of the *FE*-strategy returns unexplained. The estimated alphas are strongly significant in all cases and in the range of 1.0% to 1.6% per annum in the majority of cases.¹⁹

In Panel B of Table 12, we present the results for the *ER*-based strategy for horizons $m = 2$ to 12 months. Again, the bond market factor proves to be important and the evidence on the explanatory power of the mimicking portfolios resembles that evinced in Table 7. The alphas are positive for all horizons, although statistically significant only for horizons of five to 12 months.²⁰

9. Conclusion

There is overwhelming evidence that the expectations hypothesis (EH) does not describe how long yields are determined in practice. In this paper, we explore the possibility that the EH fails because short yield expectations are subject to behavioral biases. To explore this idea, we test the specific biases that have been invoked to explain the stylized features of short-

¹⁹ Strategy B produces alphas ranging between 2.5% and 3.5% per annum in most cases.

²⁰ Again, Strategy B generates much larger alphas (between 0.9% and 1.5% per annum) and strongly significant across all horizons.

term momentum and long-term return reversals in equity returns. We focus on the LSN and the conservatism biases and derive the testable implications of these biases for expectations in the bond market. In contrast with the equity market, where the markets' expectations of earnings at specific dates cannot be inferred directly from stock prices, investor's expectations of the short yield can be inferred from the term structure of interest rates. The bond market therefore offers a valuable opportunity to directly test the implications of behavioral biases for expectation errors.

We find evidence for both of these biases, both in the full sample from 1952 to 2012 and in shorter, more recent samples following the structural break in bond yields in 1981, and following the financial crisis of 2007. Widely accepted proxies for the time-varying risk premium do have explanatory power in our model but they do not alter the verdict on the presence of behavioral biases. It appears that investors have expectations that are subject to behavioral biases but they are also risk averse and the risk premium they require is time-varying.

We also show that the biases could have been profitably exploited. We find that trading strategies based on behavioral variables deliver higher Sharpe ratios than strategies based on widely used proxies for a time varying risk premium. We also find positive alphas with respect to mimicking portfolios that are based on these risk proxies.

There is still much work to be done in the ongoing debate between risk-based and behavioral interpretations of anomalies. One interesting direction that has recently received attention is to examine whether the scale of anomalies is reduced where there is evidence of participation by more sophisticated investors. For example, are anomalies stronger in less developed markets or when risk capital mobility is lower, assuming mobile capital is in the hands of more sophisticated investors? Potì and Siddique (2013) study the impact of risk capital

mobility on anomalies and Pukthuanthong-Le et al. (2007) examine whether technical trading rules are more successful for developing market currencies. It would be an interesting direction for future work to develop these ideas in the context of the bond market and ask whether evidence for the behavioral models systematically varies across markets that are differentiated by similar characteristics.

Identifying that the failure of the EH is due to behavioral biases implies excess volatility in bond returns relative to the volatility of the short yield. Since such excess volatility can have real welfare costs, it is of interest to policy makers to consider how this volatility might be damped. The role of expectation errors that we identify suggests an important role for managing expectations so that these errors are minimized. If “forward guidance” by central banks reduces expectation errors then it will serve to reduce the importance of this source of excess volatility in returns. Once sufficient data has accumulated from regimes where “forward guidance” has been active, it will be interesting to consider its impact on expectation errors and the attendant behavioral biases.

References

- Adrian, T., E. Etula, and T. Muir, 2014, "Financial Intermediaries and the Cross-Section of Asset Returns," *Journal of Finance* 69, 2557–2596.
- Ang, A., and M. Piazzesi, 2003, "A No-Arbitrage Vector Autoregression of Term Structure Dynamics With Macroeconomic and Latent Variables," *Journal of Monetary Economics* 50, 745–787.
- Ang, A., G. Bekaert, and M. Wei, 2008, "The Term Structure of Real Rates and Expected Inflation," *Journal of Finance* 63, 797–849.
- Asness, C. S., T. J. Moskowitz, and L.H. Pedersen, 2013 "Value and Momentum Everywhere," *Journal of Finance*, 68, 929–985.
- Backus, D., A. Gregory and S. Zin, 1989, "Risk Premiums in the Term Structure: Evidence from Artificial Economies," *Journal of Monetary Economics* 24, 371–399.
- Barberis, N., A. Shliefer, and R. Vishny, 1998, "A Model of Investor Sentiment," *Journal of Financial Economics* 49, pp. 307–343.
- Bloomfield, R. J., and J. Hales, 2002, "Predicting the Next Step of a Random Walk: Experimental Evidence of Regime-Shifting Beliefs," *Journal of Financial Economics* 65, 397–414.
- Breeden, D. T., M. R. Gibbons, and R. Litzenberger, 1989, "Empirical Tests of the Consumption-Oriented CAPM," *Journal of Finance* 44, 231–262.
- Brennan, M. J., and Y. Xia, 2004, "Persistence, Predictability, and Portfolio Planning," Working Paper, UCLA and Wharton.
- Bulkley, G., R. Harris, and V. Nawosah, 2011, "Revisiting the Expectations Hypothesis of the Term Structure of Interest Rates," *Journal of Banking and Finance* 35, 1202–1212.
- Campbell, J., 1995, "Some Lessons from the Yield Curve," *Journal of Economic Perspectives* 9, 125–152.
- Campbell, J., and R. Shiller, 1984, "A Simple Account of the Behavior of Long Term Interest Rates," *American Economic Review, Papers and Proceedings* 74, 44–48.
- Campbell, J., and R. Shiller, 1991, "Yield Spreads and Interest Rate Movements: A Bird's Eye View," *Review of Economic Studies* 58, 495–514.
- Campbell, J. Y., and S. B. Thompson, 2008, "Predicting Excess Stock Returns out of Sample: Can Anything Beat the Historical Average?," *Review of Financial Studies* 21, 1509–1531.
- Christensen, J. H. E., J. A. Lopez, and G. D. Rudebusch, 2010, "Inflation Expectations and Risk Premiums in an Arbitrage-Free Model of Nominal and Real Bond Yields," *Journal of Money, Credit, and Banking* 42, 143–178.
- Clark, T. E., and M. W. McCracken. 2001, "Tests of Forecast Accuracy and Encompassing for Nested Models," *Journal of Econometrics* 105, 85–110.

- Clark, T. E., and M. W. McCracken, 2005, "Evaluating Direct Multi-Step Forecasts," *Econometric Reviews* 24, 369-404.
- Cochrane, J., and M. Piazzesi, 2002, "Bond Risk Premia," NBER Working Paper No. 9178.
- Cochrane, J., and M. Piazzesi, 2005, "Bond Risk Premia," *American Economic Review* 95, 138-160.
- Cooper, I., and R. Priestly, 2009, "Time-Varying Risk Premiums and the Output Gap," *Review of Financial Studies* 22, 2801-2833.
- Cutler, D., J. Poterba, and L. Summers, 1991, "Speculative dynamics," *Review of Economic Studies* 58, 529-546.
- Dai, Q., and K. Singleton, 2000, "Specification Analysis of Affine Term Structure Models," *Journal of Finance* 50, 1943-1978.
- Daniel, K., D. Hirshleifer, and A. Subrahmanyam, 1998, "Investor Psychology and Security Market Under-and Overreactions," *Journal of Finance* 53, 1839-1885.
- Duffee G. R., 2012, "Bond Pricing and the Macroeconomy," *Handbook of Economics of Finance*, G. Constantinides, M. Harris, and R. Stulz, Eds., Forthcoming.
- Fama, E. F., 2006, "The behavior of interest rates," *Review of Financial Studies* 19, 359-379.
- Ferson, W. E., and C. R. Harvey, 1991, "The Variation of Economic Risk Premiums," *Journal of Political Economy* 99, 385-415.
- Hansen, L. P., and R. J. Hodrick, 1980, "Forward Exchange Rates as Optimal Predictors of Future Spot Rates: An Econometric Analysis," *Journal of Political Economy* 88, 829-853.
- Hardouvelis, G., 1994, "The Term Structure Spread and Future Changes in Long and Short Rates in the G7 Countries: Is There a Puzzle?," *Journal of Monetary Economics* 33, 255-283.
- Jegadeesh, N., and S. Titman, 1993, "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency," *Journal of Finance* 48, 65-91.
- Jobson, J. D., and R. Korkie, 1981, "Performance Hypothesis Testing with the Sharpe and Treynor Measures," *Journal of Finance* 36, 889-908.
- Joslin, S., Pribsch, M., and Singleton, K., 2012, "Risk Premiums in Dynamic Term Structure Models with Unspanned Macro Risks," *Journal of Finance*, Forthcoming.
- Kim, H., and Moon, J., 2005, "Do Macroeconomic Variables Forecast Bond Returns?," SUNY at Buffalo Working Paper.
- Ledoit, O., and M. Wolf, 2008, "Robust performance hypothesis testing with the sharpe ratio," *Journal of Empirical Finance* 15, 850-9.
- Ludvigson, S., and Ng, S., 2009, "Macro Factors in Bond Risk Premia," *Review of Financial Studies* 22, 5027-5067.

- Mankiw, N., 1986, "The Term Structure of Interest Rates Revisited," *Brookings Papers on Economic Activity* 1, 61-96.
- Mankiw, N., and L. Summers, 1984, "Do Long Term Interest Rates Overreact to Short Term Interest Rates?," *Brookings Papers on Economic Activity* 1, 223-242.
- Mommel, C., 2003, "Performance Hypothesis Testing with the Sharpe Ratio," *Finance Letters* 1, 21-23.
- Mincer, J., and V. Zarnowitz, 1969, "The Evaluation of Economic Forecasts," in J. Mincer, Ed., *Economic Forecasts and Expectations*, New York: National Bureau of Economic Research, 3-46.
- Newey W. K., and K. D. West, 1987, "A Simple, Positive Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica* 55, 3, 703-708.
- Piazzesi, M., and Schneider, M. 2006, "Equilibrium Yield Curves" *NBER Macroeconomics Annual* 2006, 21, 389 – 472.
- Poti, V., and Siddique, A., 2013, "What Drives Currency Predictability?," *Journal of International Money and Finance* 36, 86-106.
- Pukthuanthong-Le, K., F.A. Elayan, and L.C. Rose, 2007. "Equity and Debt Market Responses to Sovereign Credit Ratings Announcement," *Global Finance Journal*, 18, 47-83.
- Rabin, M., 2002, "Inference by Believers in the Law of Small Numbers," *Quarterly Journal of Economics* 117, 775-816.
- Rudebusch, G.D., and Swanson, E.T., 2008 "Examining the Bond Premium Puzzle with a DSGE Model" *Journal Of Monetary Economics* 55, October, S111-S126.
- Rudebusch, G. D., and Wu, T., 2008, "A Macro-Finance Model of the Term Structure, Monetary Policy and the Economy," *Economic Journal* 118, 906-926.
- Shiller, R., 1979, "The Volatility of Long Term Interest Rates and Expectations Models of the Term Structure," *Journal of Political Economy* 82, 1190-1219.
- Shiller, R., J. Campbell and K. Schoenholtz, 1983, "Forward Rates and Future Policy: Interpreting the Term Structure of Interest Rates," *Brookings Papers on Economic Activity* 1, 173-217.
- Tversky, A., and D. Kahneman, 1971, "Belief in the Law of Small Numbers," *Psychological Bulletin* 76, 105-110.
- Wachter, J., 2006, "A Consumption Based Model of the Term Structure of Interest Rates," *Journal of Financial Economics* 79, 365-399.

Table 1 Momentum and Return Reversals

<i>n</i>	<i>m</i>									
	1	2	3	6	9	12	24	36	48	60
2	0.279** (4.016)	-	-	-	-	-	-	-	-	-
3	0.186** (2.100)	0.172** (2.167)	-	-	-	-	-	-	-	-
6	0.191** (2.497)	0.080 (0.842)	0.041 (0.312)	-	-	-	-	-	-	-
9	0.174** (2.683)	0.017 (0.161)	-0.043 (-0.321)	-0.006 (-0.060)	-	-	-	-	-	-
12	0.180** (3.027)	0.023 (0.207)	-0.033 (-0.245)	0.003 (0.034)	0.175** (2.934)	-	-	-	-	-
24	0.155** (3.388)	-0.017 (-0.173)	-0.061 (-0.526)	-0.005 (-0.064)	0.132 (2.017)	0.079 (0.779)	-	-	-	-
36	0.123 (2.676)	-0.022 (-0.256)	-0.064 (-0.626)	-0.007 (-0.101)	0.095 (1.415)	0.025 (0.239)	-0.201** (-2.171)	-	-	-
48	0.107** (2.502)	-0.035 (-0.456)	-0.064 (-0.694)	-0.011 (-0.159)	0.064 (0.910)	-0.020 (-0.180)	-0.204** (-2.279)	-0.369** (-3.569)	-	-
60	0.095** (2.280)	-0.040 (-0.569)	-0.053 (-0.625)	-0.011 (-0.156)	0.043 (0.565)	-0.052 (-0.460)	-0.207** (-2.297)	-0.329** (-2.938)	-0.388** (-5.741)	-
120	0.069 (1.583)	-0.055 (-0.925)	-0.043 (-0.621)	-0.015 (-0.200)	-0.007 (-0.079)	-0.135 (-1.039)	-0.203** (-1.981)	-0.237** (-2.042)	-0.304** (-5.113)	-0.351** (-7.828)

Notes: The table reports the results of estimating regression (8) in the main text. Results are reported for bond maturities $n = 2, 3, 6, 9, 12, 24, 36, 48, 60$ and 120 months, and holding periods $m = 1, 2, 3, 6, 9, 12, 24, 36, 48$ and 60 months. The first number in each set is the estimate of the slope coefficient for the corresponding maturity (n) and horizon (m) combination. The constant term and the coefficient on the dummy are not reported. The figure in parentheses below each coefficient estimate is the Newey and West (1987) corrected t-statistic. ** and * indicate significance at the 5% and 10% levels, respectively. The sample period is from 01/1952 to 12/2012.

Table 2 Short Term Predictions of the LSN

m	Panel A: Full sample				Panel B: Post-1981 sub-sample				Panel C: Post-2007 sub-sample			
	α_2	γ_2	β_2	\bar{R}^2	α_2	γ_2	β_2	\bar{R}^2	α_2	γ_2	β_2	\bar{R}^2
1	-0.020** (-6.092)	0.000 (0.142)	0.280** (4.000)	0.076	-0.015** (-6.418)	-	0.424** (6.257)	0.178	-0.004* (-1.766)	-	0.478** (3.410)	0.216
2	-0.032** (-5.237)	-0.002 (-0.368)	0.174** (2.166)	0.028	-0.024** (-7.023)	-	0.413** (5.920)	0.170	-0.008** (-2.281)	-	0.294 (1.422)	0.085
3	-0.043** (-4.438)	-0.002 (-0.223)	0.095 (0.728)	0.006	-0.028** (-5.895)	-	0.410** (3.545)	0.172	-0.016** (-2.002)	-	0.249** (2.254)	0.050
4	-0.055** (-4.543)	-0.004 (-0.302)	0.006 (0.041)	-0.002	-0.035** (-5.437)	-	0.349** (3.208)	0.143	-0.023* (-1.806)	-	0.256** (2.698)	0.047
5	-0.060** (-4.558)	-0.010 (-0.726)	-0.029 (-0.226)	0.000	-0.046** (-5.368)	-	0.250** (2.751)	0.081	-0.024* (-1.846)	-	0.301** (3.129)	0.085
6	-0.056** (-4.144)	-0.021 (-1.327)	0.001 (0.004)	0.006	-0.053** (-5.155)	-	0.210** (2.366)	0.062	-0.026* (-1.993)	-	0.280** (2.961)	0.097

Notes: The table reports the results of estimating regression (9) in the main text. Results are reported for horizons $m = 1, 2, 3, 4, 5$, and 6 months. Panel A reports results for the full sample, 01/1952–12/2012, Panel B for the post-1981 sub-sample, 01/1982–12/2012, and Panel C for the post-2007 sub-sample, 01/2008–12/2012. The constant term and the coefficient on the dummy are not reported. The figure in parentheses below each coefficient estimate is the Newey and West (1987) corrected t-statistic. \bar{R}^2 is the adjusted R -squared. ** and * indicate significance at the 5% and 10% levels, respectively.

Table 3 Long Term Predictions of the LSN

m	Panel A: Full sample				Panel B: Post-1981 sub-sample			
	α_2	γ_2	β_2	\bar{R}^2	α_2	γ_2	β_2	\bar{R}^2
9	-0.040** (-2.532)	-0.047** (-2.221)	0.146* (1.706)	0.068	-0.071** (-4.134)	-	0.214** (2.148)	0.060
12	-0.039* (-1.739)	-0.073** (-2.608)	0.128 (1.280)	0.100	-0.087** (-3.479)	-	0.245** (2.152)	0.069
18	-0.042 (-1.285)	-0.129** (-3.117)	-0.100 (-0.796)	0.112	-0.142** (-3.693)	-	0.055 (0.427)	0.000
24	-0.037 (-0.958)	-0.195** (-4.160)	-0.245** (-2.284)	0.195	-0.208** (-6.706)	-	-0.180** (-2.165)	0.031
36	-0.023 (-0.603)	-0.335** (-5.874)	-0.490** (-5.486)	0.434	-0.332** (-8.268)	-	-0.495** (-4.655)	0.332
48	-0.003 (-0.076)	-0.409** (-6.170)	-0.523** (-4.062)	0.527	-0.337** (-12.559)	-	-0.328** (-2.694)	0.151
60	0.014 (0.351)	-0.451** (-7.082)	-0.444** (-2.386)	0.538	-0.303** (-6.051)	-	-0.033 (-0.164)	-0.003
120	0.080 (0.918)	-0.633** (-6.452)	-0.304 (-1.553)	0.578	-0.242** (-2.551)	-	0.427** (3.179)	0.236

Notes: The table reports the results of estimating regression (9) in the main text. Results are reported for horizons $m = 6, 9, 12, 18, 24, 36, 48, 60$ and 120 months. Panel A reports results for the full sample, 01/1952–12/2012 and Panel B for the post-1981 sub-sample, 01/1982–12/2012. The figure in parentheses below each coefficient estimate is the Newey and West (1987) corrected t-statistic. \bar{R}^2 is the adjusted R -squared. ** and * indicate significance at the 5% and 10% levels, respectively.

Table 4 Predictions of the Conservatism Bias

<i>i</i>	Panel A: Full sample				Panel B: Post-1981 sub-sample				Panel C: Post-2007 sub-sample			
	α_3	γ_3	β_3	\bar{R}^2	α_3	γ_3	β_3	\bar{R}^2	α_3	γ_3	β_3	\bar{R}^2
1	-0.010** (-3.454)	-0.003 (-0.855)	0.121 (1.323)	0.013	-0.011** (-5.324)	-	0.259** (2.110)	0.065	-0.003** (-2.403)	-	0.585** (4.803)	0.345
2	-0.007** (-2.669)	0.000 (-0.056)	0.171** (2.006)	0.027	-0.006** (-3.433)	-	0.274** (2.751)	0.073	-0.003** (-2.431)	-	0.524** (5.543)	0.270
3	-0.006** (-2.837)	-0.001 (-0.305)	0.173** (2.315)	0.027	-0.008** (-4.496)	-	0.110 (1.569)	0.009	-0.004** (-2.246)	-	0.466** (4.126)	0.208
4	-0.002 (-1.030)	-0.006** (-2.071)	0.129** (2.171)	0.019	-0.008** (-4.540)	-	0.114* (1.760)	0.010	-0.005** (-2.162)	-	0.451** (3.818)	0.193
5	0.002 (0.994)	-0.010** (-3.577)	0.098* (1.922)	0.021	-0.008** (-4.510)	-	0.090 (1.568)	0.005	-0.005** (-2.103)	-	0.347** (2.854)	0.109
6	0.004* (1.703)	-0.012** (-3.900)	0.102** (2.017)	0.029	-0.008** (-4.232)	-	0.115** (2.209)	0.011	-0.005** (-2.206)	-	0.159* (1.906)	0.010
7	0.004* (1.681)	-0.012** (-3.708)	0.121** (2.703)	0.036	-0.007** (-3.968)	-	0.200** (3.230)	0.038	-0.004* (-1.971)	-	0.028 (0.311)	-0.017
8	0.004 (1.481)	-0.012** (-3.716)	0.128** (2.993)	0.040	-0.008** (-4.286)	-	0.224** (2.743)	0.048	-0.003 (-1.303)	-	-0.009 (-0.073)	-0.018
9	0.003 (1.281)	-0.012** (-3.678)	0.108** (2.607)	0.033	-0.008** (-4.340)	-	0.160** (1.972)	0.023	-0.002 (-0.922)	-	-0.052 (-0.340)	-0.015
10	0.002 (0.982)	-0.011** (-3.679)	0.095** (2.515)	0.028	-0.009** (-4.365)	-	0.108 (1.569)	0.009	-0.003 (-1.096)	-	-0.120 (-0.829)	-0.002
11	0.001 (0.738)	-0.010** (-3.609)	0.094** (2.849)	0.024	-0.009** (-4.246)	-	0.097* (1.764)	0.007	-0.004 (-1.488)	-	-0.146 (-1.122)	0.006
12	0.001 (0.646)	-0.009** (-3.341)	0.090** (2.813)	0.020	-0.008** (-3.914)	-	0.097** (2.075)	0.007	-0.005* (-1.904)	-	-0.128 (-1.007)	0.000

Notes: The table reports the results of estimating regression (10) in the main text. Results are reported for horizons $i = 1$ to 12 months and lag $m = 1$ month. Panel A reports results for the full sample, 01/1952–12/2012, Panel B for the post-1981 sub-sample, 01/1982–12/2012, and Panel C for the post-2007 sub-sample, 01/2008–12/2012. The figure in parentheses below each coefficient estimate is the Newey and West (1987) corrected t-statistic. \bar{R}^2 is the adjusted R -squared. ** and * indicate significance at the 5% and 10% levels, respectively.

Table 5 Short Term Predictions of the LSN after Controlling for Risk

<i>m</i>	<i>Intercept</i>	<i>D</i>	<i>FE</i>	f_{CP}^1	f_{CP}^2	f_{CP}^3	f_{CP}^4	f_{CP}^5	π	<i>gap</i>	<i>g</i>	ΔU	<i>Vol</i>	\bar{R}^2
Panel A: Cochrane-Piazzessi forward rates (CP) as risk factors														
1	-0.003 (-0.665)	-0.004 (-1.033)	0.229** (3.514)	-0.009* (-1.910)	0.005 (0.575)	0.001 (0.119)	-0.005 (-0.470)	0.004 (0.640)	-	-	-	-	-	0.150
2	-0.002 (-0.188)	-0.008 (-1.099)	0.136** (1.997)	-0.010 (-1.278)	0.003 (0.184)	0.001 (0.031)	-0.014 (-0.649)	0.014 (1.231)	-	-	-	-	-	0.129
3	-0.005 (-0.317)	-0.008 (-0.733)	0.064 (0.537)	-0.003 (-0.295)	-0.012 (-0.467)	0.005 (0.115)	-0.013 (-0.403)	0.017 (1.052)	-	-	-	-	-	0.116
6	-0.001 (-0.026)	-0.031* (-1.658)	-0.015 (-0.121)	-0.002 (-0.108)	-0.011 (-0.285)	-0.025 (-0.491)	0.012 (0.263)	0.016 (0.620)	-	-	-	-	-	0.132
Panel B: Macroeconomic variables (Macro) as risk factors														
1	-0.007* (-1.619)	0.006 (1.396)	0.218** (3.006)	-	-	-	-	-	-0.015* (-1.999)	0.018 (0.772)	0.005** (2.138)	0.009 (0.866)	-0.640** (-2.788)	0.121
2	-0.013 (-1.364)	0.004 (0.482)	0.105 (1.156)	-	-	-	-	-	-0.029** (-2.221)	0.071* (1.809)	0.006 (1.491)	-0.009 (-0.757)	-0.826 (-1.423)	0.083
3	-0.019 (-1.238)	0.005 (0.391)	0.021 (0.153)	-	-	-	-	-	-0.037** (-2.103)	0.098* (1.738)	0.005 (1.146)	-0.026* (-1.778)	-0.970 (-1.070)	0.065
6	-0.030 (-1.264)	-0.009 (-0.404)	-0.084 (-0.690)	-	-	-	-	-	-0.020 (-0.823)	0.109 (1.064)	0.006 (1.014)	-0.044** (-2.221)	-1.536 (-1.016)	0.052
Panel C: Both Cochrane-Piazzessi forward rates and macroeconomic variables (CP+Macro) as risk factors														
1	-0.010** (-2.056)	-0.002 (-0.345)	0.199** (3.030)	-0.012** (-2.267)	0.004 (0.469)	0.002 (0.160)	-0.011 (-0.976)	0.015** (2.138)	0.001 (0.088)	0.091** (3.060)	0.005** (2.006)	0.006 (0.582)	-0.521* (-1.708)	0.176
2	-0.014 (-1.453)	-0.007 (-0.754)	0.096 (1.272)	-0.012 (-1.597)	0.001 (0.058)	0.001 (0.027)	-0.022 (-0.965)	0.029** (2.293)	-0.003 (-0.220)	0.170** (3.133)	0.005 (1.486)	-0.015 (-1.417)	-0.429 (-0.621)	0.159
3	-0.019 (-1.323)	-0.009 (-0.584)	0.009 (0.071)	-0.004 (-0.325)	-0.017 (-0.626)	0.003 (0.070)	-0.019 (-0.552)	0.032* (1.817)	-0.007 (-0.490)	0.193** (2.455)	0.006 (1.332)	-0.036** (-2.567)	-0.304 (-0.294)	0.147
6	-0.018 (-0.811)	-0.029 (-1.264)	-0.091 (-0.721)	-0.006 (-0.298)	-0.012 (-0.292)	-0.024 (-0.493)	0.001 (0.020)	0.030 (1.093)	0.027 (1.143)	0.229* (1.935)	0.008 (1.218)	-0.061** (-3.276)	-0.076 (-0.047)	0.166

Notes: The table reports the results of estimating regression (11) in the main text. Results are reported for horizons $m = 1, 2, 3$ and 6 months. Panel A reports results for the specification including the five CP forward rates only ($f_{CP}^1, f_{CP}^2, f_{CP}^3, f_{CP}^4, f_{CP}^5$). Panel B reports results for the specification including the macro variables only: the inflation rate (π), a measure of output gap (*gap*), a measure of economic growth (*g*), the change in the unemployment rate (ΔU), and a measure of bond market volatility (*Vol*). Panel C reports results for the specification including both the CP forward rates and the macro variables. *D* is a dummy variable that is set to one after December 1981 and zero otherwise. The figure in parentheses below each coefficient estimate is the Newey and West (1987) corrected t-statistic. \bar{R}^2 is the adjusted *R*-squared. ** and * indicate significance at the 5% and 10% levels, respectively. The sample period is from 01/1952 to 12/2012.

Table 6 Long Term Predictions of the LSN after Controlling for Risk

<i>m</i>	<i>Intercept</i>	<i>D</i>	<i>FE</i>	f_{CP}^1	f_{CP}^2	f_{CP}^3	f_{CP}^4	f_{CP}^5	π	<i>gap</i>	<i>g</i>	ΔU	<i>Vol</i>	\bar{R}^2
Panel A: Cochrane-Piazzesi forward rates (CP) as risk factors														
18	0.077 (1.630)	-0.109** (-2.473)	-0.218* (-1.731)	0.053 (1.364)	0.005 (0.114)	-0.095 (-1.448)	-0.095 (-1.270)	0.112* (1.763)	- -	- -	- -	- -	- -	0.274
24	0.087* (1.671)	-0.173** (-3.399)	-0.464** (-3.101)	0.075 (1.629)	0.061 (1.148)	-0.217** (-2.847)	-0.056 (-0.707)	0.119 (1.403)	- -	- -	- -	- -	- -	0.365
36	0.068 (0.998)	-0.315** (-4.639)	-0.493** (-2.895)	0.033 (0.773)	0.002 (0.029)	-0.083 (-1.448)	-0.028 (-0.519)	0.061 (0.945)	- -	- -	- -	- -	- -	0.484
48	0.081 (1.247)	-0.379** (-5.041)	-0.346** (-2.192)	-0.006 (-0.181)	-0.050 (-0.915)	0.019 (0.341)	0.006 (0.135)	0.016 (0.289)	- -	- -	- -	- -	- -	0.563
60	0.119* (1.702)	-0.407** (-4.815)	-0.298 (-1.556)	0.010 (0.296)	-0.079 (-1.171)	0.058 (0.835)	0.030 (0.330)	-0.037 (-0.489)	- -	- -	- -	- -	- -	0.575
Panel B: Macroeconomic variables (Macro) as risk factors														
18	-0.018 (-0.584)	-0.107** (-2.534)	-0.208** (-2.042)	- -	- -	- -	- -	- -	0.025 (0.391)	0.427 (1.467)	0.008 (1.338)	-0.034 (-0.793)	-2.490 (-1.267)	0.168
24	-0.013 (-0.408)	-0.171** (-3.799)	-0.354** (-3.393)	- -	- -	- -	- -	- -	0.041 (0.503)	0.515 (1.603)	0.003 (0.328)	-0.034 (-0.654)	-2.716* (-1.660)	0.254
36	-0.013 (-0.329)	-0.319** (-6.472)	-0.546** (-4.632)	- -	- -	- -	- -	- -	0.023 (0.283)	0.490 (1.595)	0.000 (0.018)	-0.017 (-0.337)	-1.531 (-0.959)	0.465
48	0.039 (1.041)	-0.380** (-6.318)	-0.531** (-4.468)	- -	- -	- -	- -	- -	0.016 (0.359)	0.281 (1.024)	0.001 (0.207)	0.034 (0.864)	-2.913** (-2.933)	0.555
60	0.125** (3.573)	-0.368** (-6.257)	-0.385** (-3.094)	- -	- -	- -	- -	- -	0.064 (1.052)	0.112 (0.538)	-0.009 (-0.946)	0.025 (0.700)	-7.335** (-7.055)	0.642
Panel C: Both Cochrane-Piazzesi forward rates and macroeconomic variables (CP+Macro) as risk factors														
18	0.050 (1.223)	-0.103** (-2.537)	-0.275** (-2.400)	0.037 (1.032)	0.018 (0.389)	-0.083 (-1.286)	-0.130** (-1.992)	0.135** (2.445)	0.099** (2.191)	0.359 (1.475)	0.012* (1.915)	-0.084** (-2.406)	0.252 (0.124)	0.326
24	0.057 (1.158)	-0.164** (-3.684)	-0.485** (-3.355)	0.051 (1.201)	0.077 (1.447)	-0.203** (-2.755)	-0.099 (-1.374)	0.153* (1.803)	0.099* (1.993)	0.409 (1.451)	0.004 (0.599)	-0.080* (-1.772)	-0.401 (-0.175)	0.406
36	0.038 (0.529)	-0.306** (-5.649)	-0.460** (-3.027)	0.004 (0.106)	0.020 (0.297)	-0.070 (-1.279)	-0.060 (-0.817)	0.089 (0.969)	0.070 (1.342)	0.403 (1.175)	0.003 (0.488)	-0.059* (-1.707)	0.091 (0.029)	0.505
48	0.052 (0.817)	-0.360** (-5.483)	-0.325** (-2.276)	-0.032 (-0.852)	-0.035 (-0.645)	0.034 (0.724)	-0.030 (-0.518)	0.053 (0.753)	0.068* (1.819)	0.315 (0.909)	0.003 (0.463)	0.012 (0.371)	-2.428 (-1.037)	0.581
60	0.084* (1.695)	-0.365** (-6.187)	-0.276** (-2.446)	-0.026 (-0.918)	-0.066 (-1.264)	0.086 (1.427)	-0.011 (-0.130)	0.023 (0.395)	0.078 (1.598)	0.293 (1.260)	-0.010 (-1.033)	0.026 (0.954)	-8.688** (-5.654)	0.664

Notes: The table reports the results of estimating regression (11) in the main text. Results are reported for horizons $m = 18, 24, 36, 48$ and 60 months. Panel A reports results for the specification including the five CP forward rates only ($f_{CP}^1, f_{CP}^2, f_{CP}^3, f_{CP}^4, f_{CP}^5$). Panel B reports results for the specification including the macro variables only: the inflation rate (π), a measure of output gap (*gap*), a measure of economic growth (*g*), the change in the unemployment rate (ΔU), and a measure of bond market volatility (*Vol*). Panel C reports results for the specification including both the CP forward rates and the macro variables. *D* is a dummy variable that is set to one after December 1981 and zero otherwise. The figure in parentheses below each coefficient estimate is the Newey and West (1987) corrected t-statistic. \bar{R}^2 is the adjusted *R*-squared. ** and * indicate significance at the 5% and 10% levels, respectively. The sample period is from 01/1952 to 12/2012.

Table 7 Predictions of the Conservatism Bias after Controlling for Risk

<i>i</i>	<i>Intercept</i>	<i>D</i>	<i>ER</i>	f_{CP}^1	f_{CP}^2	f_{CP}^3	f_{CP}^4	f_{CP}^5	π	<i>gap</i>	<i>g</i>	ΔU	<i>Vol</i>	\bar{R}^2
Panel A: Cochrane-Piazzesi forward rates (CP) as risk factors														
1	0.004 (0.619)	-0.003 (-0.593)	0.107 (1.208)	-0.004 (-0.728)	0.012 (1.012)	-0.015 (-0.869)	0.004 (0.306)	0.001 (0.086)	- -	- -	- -	- -	- -	0.033
2	0.001 (0.214)	-0.001 (-0.227)	0.170* (1.906)	-0.004 (-0.635)	0.010 (0.912)	-0.014 (-0.826)	0.005 (0.358)	0.001 (0.117)	- -	- -	- -	- -	- -	0.035
3	0.000 (-0.023)	-0.004 (-0.986)	0.182** (2.234)	-0.006 (-0.998)	0.006 (0.703)	-0.008 (-0.521)	0.005 (0.368)	0.000 (0.056)	- -	- -	- -	- -	- -	0.041
6	0.005 (1.021)	-0.014** (-3.205)	0.104** (2.017)	-0.001 (-0.138)	0.003 (0.350)	-0.009 (-0.839)	0.004 (0.368)	0.002 (0.254)	- -	- -	- -	- -	- -	0.025
9	0.005 (1.180)	-0.012** (-3.117)	0.098** (2.276)	0.007** (2.180)	-0.009 (-1.292)	-0.009 (-0.898)	0.009 (1.032)	0.002 (0.298)	- -	- -	- -	- -	- -	0.043
12	0.002 (0.677)	-0.009** (-2.576)	0.103** (2.760)	0.011** (4.157)	-0.025** (-3.752)	0.009 (0.964)	0.004 (0.400)	0.000 (0.047)	- -	- -	- -	- -	- -	0.042
Panel B: Macroeconomic variables (Macro) as risk factors														
1	-0.008 (-1.312)	0.000 (0.068)	0.089 (0.968)	- -	- -	- -	- -	- -	-0.005 (-0.724)	0.057** (2.699)	0.006** (2.026)	-0.001 (-0.165)	-0.180 (-0.519)	0.031
2	-0.006 (-1.067)	0.002 (0.323)	0.141* (1.870)	- -	- -	- -	- -	- -	-0.002 (-0.279)	0.023 (1.184)	0.004 (1.524)	-0.010 (-1.210)	-0.105 (-0.265)	0.035
3	-0.003 (-0.626)	0.000 (0.065)	0.142* (1.903)	- -	- -	- -	- -	- -	-0.004 (-0.482)	-0.004 (-0.215)	0.003 (1.314)	-0.014* (-1.649)	-0.174 (-0.517)	0.036
6	0.000 (-0.089)	-0.011** (-2.717)	0.063* (1.644)	- -	- -	- -	- -	- -	0.008 (1.183)	0.043** (2.140)	0.004* (1.852)	-0.014 (-1.489)	0.018 (0.071)	0.041
9	0.002 (0.693)	-0.010** (-2.330)	0.081** (2.037)	- -	- -	- -	- -	- -	0.006 (1.027)	0.045** (2.149)	0.001 (0.409)	-0.013* (-1.655)	-0.111 (-0.423)	0.040
12	0.003 (1.036)	-0.007** (-2.222)	0.076** (2.292)	- -	- -	- -	- -	- -	0.001 (0.208)	0.033* (1.770)	0.000 (-0.142)	-0.009 (-1.047)	-0.117 (-0.623)	0.020
Panel C: Both Cochrane-Piazzesi forward rates and macroeconomic variables (CP+Macro) as risk factors														
1	-0.003 (-0.542)	-0.003 (-0.471)	0.083 (0.931)	-0.006 (-0.922)	0.012 (0.946)	-0.014 (-0.798)	0.002 (0.113)	0.004 (0.499)	0.006 (0.741)	0.081** (2.390)	0.005* (1.946)	-0.004 (-0.443)	0.144 (0.364)	0.049
2	-0.003 (-0.562)	-0.001 (-0.147)	0.144* (1.946)	-0.005 (-0.636)	0.009 (0.766)	-0.012 (-0.736)	0.003 (0.230)	0.002 (0.300)	0.006 (0.918)	0.042 (1.279)	0.004 (1.393)	-0.011 (-1.333)	0.134 (0.294)	0.043
3	-0.003 (-0.705)	-0.004 (-0.686)	0.156* (1.853)	-0.006 (-0.925)	0.004 (0.496)	-0.006 (-0.422)	0.003 (0.202)	0.003 (0.426)	0.003 (0.473)	0.031 (0.983)	0.003 (1.112)	-0.014 (-1.621)	-0.031 (-0.075)	0.046
6	0.000 (-0.011)	-0.013** (-2.774)	0.074* (1.813)	-0.004 (-0.688)	0.004 (0.451)	-0.007 (-0.648)	0.000 (-0.035)	0.005 (0.599)	0.013* (1.722)	0.066** (2.419)	0.004* (1.824)	-0.015 (-1.516)	0.128 (0.430)	0.043
9	0.002 (0.444)	-0.012** (-2.784)	0.079** (2.453)	0.005 (1.402)	-0.009 (-1.194)	-0.007 (-0.819)	0.005 (0.664)	0.005 (0.735)	0.006 (1.120)	0.052** (2.105)	0.001 (0.769)	-0.015* (-1.902)	0.026 (0.097)	0.051
12	0.001 (0.203)	-0.009** (-2.592)	0.093** (2.443)	0.011** (3.719)	-0.026** (-3.818)	0.009 (1.080)	0.003 (0.275)	0.002 (0.334)	-0.001 (-0.219)	0.025 (1.032)	0.001 (0.515)	-0.011 (-1.228)	-0.017 (-0.076)	0.041

Notes: The table reports the results of estimating regression (12) in the main text. Results are reported for horizons $i = 1, 2, 3, 6, 9$ and 12 months and lag $m = 1$ month. Panel A reports results for the specification including the five CP forward rates only ($f_{CP}^1, f_{CP}^2, f_{CP}^3, f_{CP}^4, f_{CP}^5$). Panel B reports results for the specification including the macro variables only: the inflation rate (π), a measure of output gap (gap), a measure of economic growth (g), the change in the unemployment rate (ΔU), and a measure of bond market volatility (Vol). Panel C reports results for the specification including both the CP forward rates and the macro variables. D is a dummy variable that is set to one after December 1981 and zero otherwise. The figure in parentheses below each coefficient estimate is the Newey and West (1987) corrected t-statistic. \bar{R}^2 is the adjusted R -squared. ** and * indicate significance at the 5% and 10% levels, respectively. The sample period is from 01/1952 to 12/2012.

Table 8 Out of sample Forecast Encompassing Tests Results: Short Run LSN

<i>m</i>	Panel A: <i>FE</i> vs. <i>CP</i>					Panel B: <i>FE</i> vs. <i>Macro</i>					Panel C: <i>FE</i> vs. <i>CP+Macro</i>				
	<i>Intercept</i>	\widehat{FE}_{t+m}^{FE}	\widehat{FE}_{t+m}^{CP}	\bar{R}^2	<i>ENC-NEW</i>	<i>Intercept</i>	\widehat{FE}_{t+m}^{FE}	$\widehat{FE}_{t+m}^{Macro}$	\bar{R}^2	<i>ENC-NEW</i>	<i>Intercept</i>	\widehat{FE}_{t+m}^{FE}	$\widehat{FE}_{t+m}^{CP+Macro}$	\bar{R}^2	<i>ENC-NEW</i>
1	0.000 (-0.034)	0.494** (2.404)	0.410** (2.682)	0.049	27.16	-0.001 (-0.210)	0.460** (1.976)	0.408** (2.070)	0.050	15.44	0.001 (0.103)	0.493** (2.284)	0.475** (3.147)	0.076	14.79
2	-0.014* (-1.694)	0.237 (1.122)	0.320** (2.081)	0.035	12.23	-0.018** (-2.016)	0.184 (0.670)	0.295 (1.019)	0.027	-3.36	-0.014* (-1.803)	0.261 (1.260)	0.321* (1.757)	0.044	-0.77
3	-0.031** (-2.727)	0.130 (0.640)	0.194 (1.413)	0.018	2.69	-0.033** (-2.891)	0.061 (0.166)	0.224 (0.552)	0.015	-11.15	-0.032** (-2.922)	0.136 (0.643)	0.189 (1.001)	0.021	-10.53
4	-0.036** (-2.564)	0.095 (0.469)	0.218* (1.990)	0.022	-4.72	-0.040** (-2.715)	0.077 (0.212)	0.172 (0.422)	0.013	-7.14	-0.038** (-2.742)	0.103 (0.489)	0.196 (1.075)	0.025	-10.27
5	-0.046** (-3.160)	0.040 (0.229)	0.198 (1.568)	0.014	-10.24	-0.051** (-3.340)	0.084 (0.244)	0.078 (0.190)	0.004	-0.76	-0.046** (-3.320)	0.045 (0.241)	0.204 (1.051)	0.022	-7.24
6	-0.053** (-3.313)	0.002 (0.012)	0.180 (1.322)	0.010	-11.22	-0.059** (-3.600)	0.099 (0.297)	0.001 (0.002)	0.000	-0.96	-0.053** (-3.464)	0.021 (0.114)	0.172 (0.873)	0.015	-8.16

Notes: The table reports the results of out-of-sample forecast encompassing tests. Results are reported for horizons $m = 1, 2, 3, 4, 5$, and 6 months. Panel A reports results of forecast comparisons when only the CP forward rates are used as risk factors, Panel B when only the macro variables are used, and Panel C when both the CP forward rates and the macro variables are used. The first four columns of each panel reports the results of estimating the Mincer-Zarnowitz regression (13) in the main text. The test compares the out-of-sample forecasts of forecast errors based on the estimation of equation (9), \widehat{FE}_{t+m}^{FE} , with those based on the estimation of equation (11) including only the risk factors, $\widehat{FE}_{t+m}^{Risk}$. $\widehat{FE}_{t+m}^{Risk} = \widehat{FE}_{t+m}^{CP}$ when only the CP forward rates are used as risk factors, $\widehat{FE}_{t+m}^{Risk} = \widehat{FE}_{t+m}^{Macro}$ when only the macro variables are used and $\widehat{FE}_{t+m}^{Risk} = \widehat{FE}_{t+m}^{CP+Macro}$ when both the CP forward rates and the macro variables are used. The figure in parentheses below each coefficient estimate is the Newey and West (1987) corrected t-statistic. \bar{R}^2 is the adjusted R-squared. ** and * indicate significance at the 5% and 10% levels, respectively. The column labeled "ENC-NEW" reports the encompassing test statistic of Clark and McCracken (2001). The test compares out-of-sample forecasts from an unrestricted specification of equation (11) in the main text that includes both *FE* and risk factors, to those based on a restricted model that includes only the risk factors. ENC-NEW statistics that are statistically significant at the 5% level on the basis of bootstrapped critical values provided in Clark and McCracken (2001, 2005) are in bold. The sample period is from 01/1952 to 12/2012.

Table 9 Out of sample Forecast Encompassing Tests Results: Long Run LSN

<i>m</i>	Panel A: <i>FE</i> vs. <i>CP</i>					Panel B: <i>FE</i> vs. <i>Macro</i>					Panel C: <i>FE</i> vs. <i>CP+Macro</i>				
	<i>Intercept</i>	\widehat{FE}_{t+m}^{FE}	\widehat{FE}_{t+m}^{CP}	\bar{R}^2	<i>ENC-NEW</i>	<i>Intercept</i>	\widehat{FE}_{t+m}^{FE}	$\widehat{FE}_{t+m}^{Macro}$	\bar{R}^2	<i>ENC-NEW</i>	<i>Intercept</i>	\widehat{FE}_{t+m}^{FE}	$\widehat{FE}_{t+m}^{CP+Macro}$	\bar{R}^2	<i>ENC-NEW</i>
9	-0.051** (-2.211)	0.202 (1.169)	0.030 (0.265)	0.019	9.19	-0.062** (-2.649)	0.546* (1.861)	-0.395 (-1.246)	0.041	-8.42	-0.051** (-2.312)	0.207 (1.079)	0.033 (0.184)	0.019	3.96
12	-0.047 (-1.375)	0.231 (1.116)	0.123 (1.241)	0.044	-0.34	-0.062* (-1.740)	0.522* (1.981)	-0.299 (-1.059)	0.053	-14.21	-0.047 (-1.419)	0.295 (1.397)	0.047 (0.321)	0.039	-11.65
18	-0.040 (-0.948)	0.331 (1.411)	0.141 (1.099)	0.058	15.43	-0.058 (-1.254)	0.490* (2.202)	-0.199 (-0.936)	0.055	2.82	-0.038 (-1.001)	0.384* (1.689)	0.073 (0.506)	0.053	15.57
24	-0.046 (-0.973)	0.355* (1.742)	0.093 (0.596)	0.065	37.07	-0.074 (-1.376)	0.515** (2.058)	-0.301 (-0.743)	0.071	54.24	-0.047 (-1.041)	0.395* (1.926)	0.018 (0.107)	0.062	38.88
36	-0.036 (-0.360)	0.774** (2.212)	-0.338** (-2.728)	0.165	18.66	-0.072 (-0.742)	0.603 (1.546)	-0.252 (-1.008)	0.135	120.79	-0.043 (-0.446)	0.817** (2.596)	-0.469** (-2.780)	0.187	11.71
48	-0.084 (-0.905)	0.693** (2.079)	-0.311 (-0.909)	0.151	33.47	-0.090 (-0.625)	0.420** (2.024)	0.006 (0.015)	0.099	124.92	-0.094 (-0.976)	0.620** (2.496)	-0.295 (-0.810)	0.136	16.40
60	-0.161 (-1.588)	0.363 (1.026)	-0.123 (-0.344)	0.062	49.55	-0.012 (-0.094)	0.076 (0.560)	0.924** (3.021)	0.237	39.08	-0.166 (-1.376)	0.311 (1.346)	-0.080 (-0.280)	0.057	28.84
120	-0.414** (15.02)	0.165 (1.016)	-0.265** (-2.703)	0.279	8.15	-0.429** (-18.30)	0.044 (0.249)	-0.128 (-0.753)	0.054	-24.38	-0.390** (-10.15)	0.094 (0.649)	-0.168** (-2.217)	0.180	5.13

Notes: The table reports the results of out-of-sample forecast encompassing tests. Results are reported for horizons $m = 6, 9, 12, 18, 24, 36, 48, 60$ and 120 months. Panel A reports results of forecast comparisons when only the CP forward rates are used as risk factors, Panel B when only the macro variables are used, and Panel C when both the CP forward rates and the macro variables are used. The first four columns of each panel reports the results of estimating the Mincer-Zarnowitz regression (13) in the main text. The test compares the out-of-sample forecasts of forecast errors based on the estimation of equation (9), \widehat{FE}_{t+m}^{FE} , with those based on the estimation of equation (11) including only the risk factors, $\widehat{FE}_{t+m}^{Risk}$. $\widehat{FE}_{t+m}^{Risk} = \widehat{FE}_{t+m}^{CP}$ when only the CP forward rates are used as risk factors, $\widehat{FE}_{t+m}^{Risk} = \widehat{FE}_{t+m}^{Macro}$ when only the macro variables are used and $\widehat{FE}_{t+m}^{Risk} = \widehat{FE}_{t+m}^{CP+Macro}$ when both the CP forward rates and the macro variables are used. The figure in parentheses below each coefficient estimate is the Newey and West (1987) corrected t-statistic. \bar{R}^2 is the adjusted *R*-squared. ** and * indicate significance at the 5% and 10% levels, respectively. The column labeled "ENC-NEW" reports the encompassing test statistic of Clark and McCracken (2001). The test compares out-of-sample forecasts from an unrestricted specification of equation (11) in the main text that includes both *FE* and risk factors, to those based on a restricted model that includes only the risk factors. ENC-NEW statistics that are statistically significant at the 5% level on the basis of bootstrapped critical values provided in Clark and McCracken (2001, 2005) are in bold. The sample period is from 01/1952 to 12/2012.

Table 10 Out of sample Forecast Encompassing Tests Results: Conservatism Bias

<i>i</i>	Panel A: FE vs. CP					Panel B: FE vs. Macro					Panel C: FE vs. CP+Macro				
	Intercept	$\bar{ER}_{t+m,i}^{ER}$	$\bar{ER}_{t+m,i}^{CP}$	\bar{R}^2	ENC-NEW	Intercept	$\bar{ER}_{t+m,i}^{ER}$	$\bar{ER}_{t+m,i}^{Macro}$	\bar{R}^2	ENC-NEW	Intercept	$\bar{ER}_{t+m,i}^{ER}$	$\bar{ER}_{t+m,i}^{CP+Macro}$	\bar{R}^2	ENC-NEW
1	-0.006 (-1.461)	0.290 (1.111)	0.087 (0.446)	0.010	0.45	-0.006 (-1.348)	0.082 (0.263)	0.336 (1.259)	0.014	-3.98	-0.007 (-1.544)	0.235 (0.945)	0.164 (0.907)	0.013	-2.87
2	-0.004 (-1.260)	0.487* (1.699)	-0.083 (-0.455)	0.013	11.03	-0.003 (-1.050)	0.329 (0.827)	0.129 (0.397)	0.013	2.90	-0.003 (-1.131)	0.392 (1.359)	0.056 (0.320)	0.013	4.55
3	-0.007** (-2.313)	0.418 (1.636)	-0.232 (-1.297)	0.008	19.78	-0.005 (-1.624)	0.372 (1.120)	0.037 (0.113)	0.005	6.80	-0.005* (-1.728)	0.387 (1.495)	0.011 (0.056)	0.005	12.06
4	-0.007** (-2.588)	0.316 (1.266)	-0.635** (-2.149)	0.012	8.54	-0.005* (-1.993)	0.163 (0.554)	-0.049 (-0.160)	-0.002	-0.73	-0.005* (-2.140)	0.190 (0.753)	-0.169 (-0.903)	0.000	4.15
5	-0.001 (-0.578)	0.592* (1.945)	-0.348 (-1.179)	0.006	3.07	-0.001 (-0.380)	0.232 (0.743)	0.178 (0.468)	0.001	-3.64	-0.001 (-0.457)	0.450* (1.683)	-0.174 (-0.619)	0.003	0.32
6	0.002 (0.779)	0.742** (2.181)	-0.163 (-0.603)	0.016	3.71	0.002 (1.063)	0.414 (1.027)	0.260 (0.618)	0.016	-1.11	0.002 (0.750)	0.698** (2.058)	-0.125 (-0.423)	0.015	1.10
7	0.003 (1.452)	0.797** (2.803)	-0.093 (-0.416)	0.031	3.93	0.003* (1.731)	0.581 (1.582)	0.167 (0.459)	0.031	0.76	0.003 (1.353)	0.782** (2.744)	-0.083 (-0.357)	0.031	1.78
8	0.003 (1.498)	0.723** (3.084)	0.018 (0.107)	0.039	4.68	0.003* (1.725)	0.675** (2.173)	0.080 (0.264)	0.039	1.87	0.003 (1.423)	0.748** (3.171)	-0.012 (-0.065)	0.039	2.55
9	0.002 (1.112)	0.589** (2.805)	0.144 (1.074)	0.032	3.16	0.003 (1.261)	0.662** (2.273)	0.054 (0.183)	0.031	1.99	0.002 (1.055)	0.650** (3.080)	0.075 (0.518)	0.031	1.74
10	0.001 (0.548)	0.488** (2.450)	0.228* (1.903)	0.025	2.92	0.001 (0.471)	0.651** (2.378)	-0.016 (-0.054)	-0.020	0.78	0.001 (0.473)	0.556** (2.808)	0.146 (1.158)	0.022	1.86
11	0.000 (0.113)	0.443** (2.283)	0.277** (2.352)	0.020	4.64	0.000 (-0.224)	0.660** (2.565)	-0.122 (-0.418)	0.012	0.28	0.000 (-0.013)	0.503** (2.608)	0.184 (1.585)	0.016	3.25
12	0.000 (-0.221)	0.390* (1.794)	0.301** (2.697)	0.015	5.26	-0.001 (-0.674)	0.664** (2.447)	-0.259 (-0.860)	0.006	0.42	-0.001 (-0.382)	0.440** (2.048)	0.192* (1.743)	0.009	3.98

Notes: The table reports the results of out-of-sample forecast encompassing tests. Results are reported for horizons $i = 1$ to 12 months and lag $m = 1$ month. Panel A reports results of forecast comparisons when only the CP forward rates are used as risk factors, Panel B when only the macro variables are used, and Panel C when both the CP forward rates and the macro variables are used. The first four columns of each panel reports the results of estimating the Mincer-Zarnowitz regression (14) in the main text. The test compares the out-of-sample forecasts of forecast revisions based on the estimation of equation (10), $\bar{ER}_{t+m,i}^{ER}$, with those based on the estimation of equation (12) including only the risk factors, $\bar{ER}_{t+m,i}^{Risk}$. $\bar{ER}_{t+m,i}^{Risk} = \bar{ER}_{t+m,i}^{CP}$ when only the CP forward rates are used as risk factors, $\bar{ER}_{t+m,i}^{Risk} = \bar{ER}_{t+m,i}^{Macro}$ when only the macro variables are used and $\bar{ER}_{t+m,i}^{Risk} = \bar{ER}_{t+m,i}^{CP+Macro}$ when both the CP forward rates and the macro variables are used. The figure in parentheses below each coefficient estimate is the Newey and West (1987) corrected t-statistic. \bar{R}^2 is the adjusted R -squared. ** and * indicate significance at the 5% and 10% levels, respectively. The column labeled "ENC-NEW" reports the encompassing test statistic of Clark and McCracken (2001). The test compares out-of-sample forecasts from an unrestricted specification of equation (12) in the main text that includes both ER and risk factors, to those based on a restricted model that includes only the risk factors. ENC-NEW statistics that are statistically significant at the 5% level on the basis of bootstrapped critical values provided in Clark and McCracken (2001, 2005) are in bold. The sample period is from 01/1952 to 12/2012.

Table 11 Sharpe Ratios: Behavioral Strategy versus *CP* and *CP+Macro* Strategies

Strategy	Panel A: <i>FE</i> Strategy					Panel B: <i>ER</i> Strategy				
	<i>m</i>	Strategy A		Strategy B		<i>m</i>	Strategy A		Strategy B	
		SR	<i>p</i> -value	SR	<i>p</i> -value		SR	<i>p</i> -value	SR	<i>p</i> -value
<i>FE</i>	1	0.140	-	0.125	-	2	0.059	-	0.143	-
<i>CP</i>		0.106	0.220	0.059	0.016		0.058	0.497	-0.001	0.000
<i>CP+Macro</i>		0.045	0.016	0.030	0.003		-0.007	0.068	-0.053	0.000
<i>FE</i>	2	0.164	-	0.165	-	3	0.108	-	0.153	-
<i>CP</i>		0.084	0.038	0.025	0.000		0.060	0.148	0.027	0.000
<i>CP+Macro</i>		0.030	0.002	-0.025	0.000		0.041	0.034	-0.038	0.000
<i>FE</i>	3	0.179	-	0.154	-	4	0.127	-	0.193	-
<i>CP</i>		0.068	0.003	0.035	0.000		0.111	0.296	0.088	0.000
<i>CP+Macro</i>		0.053	0.001	-0.024	0.000		0.083	0.109	0.014	0.000
<i>FE</i>	6	0.162	-	0.174	-	5	0.172	-	0.231	-
<i>CP</i>		0.142	0.197	0.096	0.003		0.165	0.294	0.133	0.000
<i>CP+Macro</i>		0.057	0.001	0.036	0.000		0.110	0.024	0.052	0.000
<i>FE</i>	9	0.367	-	0.255	-	6	0.218	-	0.257	-
<i>CP</i>		0.123	0.000	0.147	0.000		0.196	0.493	0.166	0.000
<i>CP+Macro</i>		0.071	0.000	0.081	0.000		0.112	0.002	0.084	0.000
<i>FE</i>	12	0.350	-	0.236	-	7	0.279	-	0.274	-
<i>CP</i>		0.137	0.000	0.141	0.001		0.184	0.005	0.173	0.000
<i>CP+Macro</i>		0.059	0.000	0.055	0.000		0.122	0.000	0.069	0.000
<i>FE</i>	24	0.407	-	0.509	-	8	0.282	-	0.270	-
<i>CP</i>		0.143	0.000	0.236	0.000		0.149	0.000	0.188	0.000
<i>CP+Macro</i>		0.081	0.000	0.111	0.000		0.088	0.000	0.086	0.000
<i>FE</i>	36	0.595	-	0.576	-	9	0.301	-	0.286	-
<i>CP</i>		0.256	0.000	0.234	0.000		0.191	0.000	0.215	0.000
<i>CP+Macro</i>		0.160	0.000	0.208	0.000		0.120	0.000	0.115	0.000
<i>FE</i>	48	0.669	-	0.624	-	10	0.296	-	0.279	-
<i>CP</i>		0.371	0.018	0.151	0.000		0.177	0.000	0.209	0.000
<i>CP+Macro</i>		0.362	0.000	0.209	0.000		0.122	0.000	0.115	0.000
<i>FE</i>	60	0.526	-	0.588	-	11	0.292	-	0.282	-
<i>CP</i>		0.377	0.003	0.098	0.000		0.185	0.000	0.198	0.000
<i>CP+Macro</i>		0.383	0.047	0.045	0.000		0.108	0.000	0.104	0.000
<i>FE</i>						12	0.295	-	0.266	-
<i>CP</i>							0.171	0.000	0.191	0.000
<i>CP+Macro</i>							0.065	0.000	0.091	0.000

Notes: This table compares the out-of-sample performance of the behavioral strategies to that of the *CP* and *CP+Macro* strategies. Panel A reports results for *FE*-based strategy for holding horizons $m = 1, 2, 3, 6, 9, 12, 24, 36, 48$, and 60 months and Panel B for *ER*-based strategy for holding horizons $m = 2$ to 12 months. Results are reported under both Strategy A and Strategy B. The column labeled “SR” reports the Sharpe ratios for the different strategies. The column labeled “*p*-value” reports the *p*-value of the difference between the Sharpe ratio of the *FE* strategy from that of the risk-based strategy indicated by the row label. The *p*-values are computed using the methodology in Jobson and Korkie (1981) with the Memmel (2003) correction. ARMA-based pre-whitened trading rule returns are used in the test. The sample period is from 01/1952 to 12/2012.

Table 12 Strategy Alphas

Panel A: FE Strategy								Panel B: ER Strategy							
<i>m</i>	<i>Intercept</i>	<i>RXB</i>	<i>M_{CPF}</i>	<i>M_π</i>	<i>M_g</i>	<i>M_{Vol}</i>	\bar{R}^2	<i>m</i>	<i>Intercept</i>	<i>RXB</i>	<i>M_{CPF}</i>	<i>M_π</i>	<i>M_g</i>	<i>M_{Vol}</i>	\bar{R}^2
1	1.025** (2.869)	0.320** (6.433)	-0.101 (-0.669)	0.155 (0.701)	-0.923* (-1.727)	0.572 (1.479)	0.403	2	0.052 (0.193)	0.270** (2.965)	0.083 (0.765)	-0.161 (-0.977)	-0.628 (-1.315)	0.573 (1.356)	0.224
2	1.052** (3.449)	0.246** (2.588)	-0.009 (-0.072)	-0.080 (-0.531)	-0.631 (-1.242)	0.586 (1.434)	0.198	3	0.375 (1.193)	0.309** (4.134)	0.012 (0.117)	-0.030 (-0.228)	-0.176 (-0.481)	0.092 (0.296)	0.266
3	0.783** (2.048)	0.313** (3.343)	-0.125 (-1.163)	0.068 (0.469)	-0.083 (-0.236)	0.421 (1.490)	0.342	4	0.468 (1.546)	0.310** (4.165)	-0.136 (-1.247)	0.103 (0.715)	-0.229 (-0.827)	0.099 (0.405)	0.286
6	0.755* (1.722)	0.247** (2.831)	0.087 (1.518)	-0.231** (-2.870)	-0.872** (-3.221)	0.449* (1.925)	0.149	5	0.671* (1.810)	0.300** (3.993)	0.044 (0.577)	-0.152** (-1.985)	-0.701** (-2.665)	0.398 (1.635)	0.252
9	1.453** (2.799)	0.259** (2.659)	-0.026 (-0.398)	-0.110 (-1.516)	-0.168 (-0.839)	0.190 (1.060)	0.179	6	0.915** (2.279)	0.292** (3.733)	0.059 (0.906)	-0.115 (-1.487)	-0.503* (-1.651)	0.203 (0.674)	0.218
12	1.458** (2.126)	0.188 (1.137)	0.083 (1.440)	-0.163** (-2.103)	-0.477* (-1.947)	0.280 (1.273)	0.076	7	1.320** (2.645)	0.290** (3.075)	0.096 (1.396)	-0.249** (-2.962)	-0.578** (-2.430)	0.340 (1.491)	0.190
24	0.928** (3.149)	0.311** (3.911)	-0.013 (-0.369)	-0.064 (-1.462)	-0.070 (-0.449)	0.070 (0.433)	0.242	8	1.301** (2.509)	0.203* (1.793)	0.097 (1.084)	-0.244** (-2.545)	-0.693** (-2.878)	0.492** (2.405)	0.096
36	1.456** (4.746)	-0.157 (-1.013)	-0.014 (-0.533)	-0.016 (-0.489)	-0.077 (-0.705)	0.118** (2.196)	0.067	9	1.280** (2.406)	0.162 (1.258)	0.034 (0.493)	-0.171** (-2.512)	-0.528** (-2.153)	0.502** (2.653)	0.068
48	1.382** (5.101)	-0.154 (-1.002)	0.002 (0.098)	-0.005 (-0.265)	-0.133* (-1.755)	0.155** (2.390)	0.055	10	1.250** (2.356)	0.124 (0.903)	0.053 (0.981)	-0.173** (-2.721)	-0.643** (-2.835)	0.484** (2.424)	0.042
60	1.569** (3.806)	-0.246 (-1.169)	0.055* (1.937)	-0.074** (-2.307)	-0.031 (-0.307)	0.047 (0.541)	0.051	11	1.144** (2.263)	0.106 (0.795)	-0.053 (-0.932)	-0.015 (-0.191)	-0.379* (-1.859)	0.310 (1.604)	0.040
								12	1.289** (2.238)	0.143 (0.918)	0.019 (0.255)	-0.115 (-1.171)	-0.608** (-2.490)	0.461** (2.160)	0.049

Notes: The table presents the results of estimating equation (19) in the main text. Results are reported under Strategy A only. Panel A reports results for *FE*-based strategy for holding horizons $m = 1, 2, 3, 6, 9, 12, 24, 36, 48$, and 60 months and Panel B for *ER*-based strategy for holding horizons $m = 2$ to 12 months. *RXB* denotes the excess returns on a portfolio of long-term Treasury bonds measured over m -month horizon. M_Y represents the monthly return on the mimicking portfolio for risk variable $Y = CPF, \pi, g, \text{ or } Vol$, where *CPF* is the *CP* forward-rate factor, π is the inflation rate, g is a measure of economic growth and *Vol* is a measure of bond market volatility. Intercept (the regression constant) is the time-series alpha, expressed in percent per year. The figure in parentheses below each coefficient estimate is the Newey and West (1987) corrected t-statistic. \bar{R}^2 is the adjusted *R*-squared. ** and * indicate significance at the 5% and 10% levels, respectively. The sample period is from 01/1952 to 12/2012.