

Expected Term Structures

Andrea Buraschi

Ilaria Piatti

Paul Whelan

Abstract

We construct and study the cross-sectional properties of survey-based bond risk premia and compare them to their traditional statistical counterparts. We document large heterogeneity in skill, identify top forecasters, and learn about the importance of subjective risk premia in long-term bonds dynamics. The consensus is not a sufficient statistics of the cross-section of expectations and we propose an alternative real-time aggregate measure of risk premia consistent with Friedmans market selection hypothesis. We then use this measure to evaluate structural models and find support for heterogeneous belief economies generating time-varying risk premia via both a quantity and price of risk channel.

Keywords: Cross-Section of Beliefs, Bond Risk Premia, Expectation Formation, Rational Expectations

First version: September 2015

This version: September 2017

Andrea Buraschi is Chair of Finance at Imperial College Business School; Ilaria Piatti is at Saïd Business School, University of Oxford; and Paul Whelan is at Copenhagen Business School. This paper was awarded the 2016 CIBC Best Paper Prize (July 7, 2016), sponsored by Sloan School of Management at MIT and Tsinghua University. We thank Marco Giacomini (EFA Discussant), Anna Cieslak (Adam Smith Discussant) Christian Eyerdahl-Larsen (AFA Discussant), Anna Obizhaeva, and the participants of the AFA meeting 2017, Chicago, the CIBC meeting 2016, Xiamen, the FMA conference 2016, Helsinki, the 2016 Financial Econometrics and Empirical Asset Pricing Workshop at the University of Lancaster, the 2016 European Summer Symposium in Financial Markets, and the seminar participants of Green Templeton College, Oxford University, Saïd Business School, University of Porto, Goethe University Frankfurt, Paris Dauphine University, Bocconi University and the 2017 Adam Smith at HEC Paris for valuable comments. Paul Whelan gratefully acknowledges financial support from the Center for Financial Frictions (FRIC), grant no. DNRF102. The usual disclaimer applies. Emails: andrea.buraschi@imperial.ac.uk, ilaria.piatti@sbs.ox.ac.uk, pawh.fi@cbs.dk.

A large literature finds compelling evidence of predictability in several asset markets. A stream of this literature interprets this result as evidence of a time-varying risk premium that can be understood in the context of rational general equilibrium models. A second stream, on the other hand, argues that several characteristics of this predictability are more likely due to the existence of behavioral biases affecting the dynamics of subjective beliefs, informational frictions, or both. In this paper, we use a detailed data set of investors' forecasts about future interest rates to obtain a direct measure of subjective expectations on long-term bond returns and short-term interest rates. We use their time-series and cross-sectional dynamics to study the properties of bond risk premia as revealed by agents, as opposed to infer bond risk premia from projections of future return realizations on lagged state variables.

The existing literature that uses macroeconomic survey expectations argues that survey data indeed contain useful information about future GDP and inflation.¹ At the same time, Malmendier and Nagel (2011) argue that agents update their expectations heterogeneously, depending on their personal experiences. Indeed they find evidence of substantial disagreement between young and old individuals in periods of highly volatile inflation (e.g. in the 1970s). In addition, Greenwood and Schleifer (2014) report that forecasts about tradeable market variables, such as stock returns, not only are inaccurate but they are even negatively correlated with future actual realizations. Kojien, Schmeling, and Vrugt (2015) find similar results in the context of global equities, currencies and fixed income markets across different countries. In contrast to these studies, we focus on a dataset that provides us with the forecasters identity. This unique feature allows us to examine several new questions that cannot be addressed when data are available only at the aggregate level. We show that the use of consensus expectations to proxy for the expectations of the marginal investor is misleading and does not reveal important properties. Moreover, we focus on bond markets to explore the time dimension of predictability (short-term versus long-term yields). This allows us to study potential sources (if any) of bond return predictability, which could originate either from short-term interest rate predictability or time-variation in bond risk premia, and alternative models of expectation formation.²

We begin by constructing measures of subjective bond risk premia (*EBR*) from professional

¹See e.g. Ang, Bekaert, and Wei (2007) and Aioli, Capistran, and Timmermann (2011).

²Other studies that investigate the dynamics of private sector expectations about interest rates and the corresponding forecast errors include Cieslak and Povala (2012) for fed fund rate forecasts and Piazzesi, Salomao, and Schneider (2015) for bond risk premia.

market participants' expectations regarding future yields. Specifically, we use Treasury coupon bond yield forecasts at the agent specific level to obtain a set of constant maturity 1-year zero-coupon bond yield expectations. Individual agent *EBRs* are then obtained by subtracting the date t observable risk free rate from their price change individual expectations. With these measures at hand we document a number of novel findings.

First, we document a large unconditional heterogeneity in the cross-section of *EBR* point forecasts. The mean forecaster *EBRs* is 1.06% for 10-year bonds. However, the mean of the first quartile *EBR* is negative and equal to -1.66% , which implies that these agents believe long-term bonds are hedges against economic shocks while the mean of the third quartile is positive and equal to $+3.57\%$, which is instead consistent with beliefs of long-term bonds being risky bets on future economic states. We also find clear evidence of persistence in agents expected bond risk premia. For example, a forecaster in the first quartile of the cross-sectional distribution of 2-year *EBR* has a probability of about 75% to stay in the first quartile the following month, and this probability is about 74% for the 10-year *EBR*. This is three times what it should be under the null hypothesis of no persistence. Finally, we find evidence against the null hypothesis that the cross-sectional properties of expectations can be summarized by the consensus value. This raises important questions about the common assumption of identifying the marginal investor with the agent with average (consensus) expectations. Notwithstanding the previous heterogeneity, overall expectations about bond returns display significant accuracy. They are positively related to future bond returns and are internally consistent - at the individual level - with same agents' forecasts about GDP and inflation.

Second, we find evidence of predictability in short-term interest rates. Indeed, consensus forecasts are reasonably accurate with respect to several econometric benchmark models estimated in real-time, including a Taylor rule model and a factor VAR model using the three term structure characteristics (level, slope, and curvature) with dynamically adjusted lag structure. Moreover, there is significant dispersion in accuracy in the cross-section and the ranking in the accuracy is rather persistent over time. Finally, the relative performance of top and bottom forecasters is state-dependent. During recessions, the spread in RMSE between top and bottom forecasters widens significantly and top forecasters do much better than market-based forecasts such as the forward rate. When we examine predictions after conditioning on the identity of the forecaster, we find that banks and brokers acting as primary dealers and trade directly with

the Federal Reserve are more likely to be among the top forecasters of the short-term interest rate.³ The superior forecasting ability of primary dealers is not only statistically but also economically significant. Interestingly, however, primary dealers do not perform better than other agents when asked to predict either gdp or inflation (two ingredients of the Taylor rule). We also find that the greatest relative accuracy (profit opportunity) occurs during periods in which the Fed changes its stance and aggressively reduces short-term rates. While most of these events are unanticipated by most agents, whose expected excess bond returns are downward biased in these subperiods,⁴ the bias is smaller for primary dealers. This is consistent either with primary dealers having superior information about Fed’s implementation of monetary policy or, more simply, with an information flow advantage originating from their role as market makers in Treasury bonds. The result is quite important given the large transaction volume of Treasury bonds managed by primary dealers.⁵

When we study the properties of long-term expected bond risk premia, we strongly reject the hypothesis that bond risk premia are constant. We find that expected bond excess returns are time varying across all deciles of the cross-sectional distribution of forecasters. However, agents who have an edge in forecasting short term rates do not have a persistent edge in predicting long term bond returns. Banks that act as primary dealers are not significantly better than others in forecasting long-term bonds returns. This is interesting since it shows that the main determinant of long-term bond returns predictability is not the predictability of short-term interest rates. Rather, the results highlight the importance of time variation in bond risk premia. In the context of these results, we also find that the slope coefficient of predictive regressions of bond excess returns on their ex-ante subjective expectations is always positive, contrary to what Greenwood and Schleifer (2014) document in the context of the stock market.⁶

³Primary dealers are trading counterparties of the New York Fed in its implementation of monetary policy. They are also expected to make markets for the New York Fed on behalf of its official accountholders as needed, and to bid on a pro-rata basis in all Treasury auctions at reasonably competitive prices.

⁴This is consistent with the findings in Cieslak and Povala (2012) who analyze survey forecast expectations of the fed fund rate and show that the largest errors are negative and occur during and after NBER recessions.

⁵Statistics are available in the Primary Dealers section of the New York Fed website: www.newyorkfed.org/markets/primarydealers.

⁶Koijen, Schmeling, and Vrugt (2015) also find that survey expectations of returns negatively predict future returns in the time series in three major asset classes: global equities, currencies, and global fixed income. However, instead of looking at the slope coefficient of predictive regressions, they show this by building a survey-based portfolio strategy. The strategy goes a dollar long or short in country i in month t when the consensus forecast is above or below a certain threshold, which is set to be equal to the middle value World Economic Survey respondents can select.

Third, an extensive literature in bond markets uses the dynamics of interest rates and bond risk premia to help specifying macro-finance models able to capture salient features of the data. The empirical evaluation of these models is often carried out by approximating agents expectations with econometric projections of future realized returns on lagged state variables. We revisit this approach and investigate the use of a cross-sectional information on subjective expectations, as directly revealed in real-time by agents, to compare the performance of alternative macro-finance models in explaining bond risk premia. To summarize information from the cross-section, instead of using reduced-form panel regression techniques we construct a time series of EBR_t^* which is based on the beliefs of the agents who have been most accurate up to time t . Friedman (1953) and Alchian (1950) argue that market selection in competitive markets is a powerful force affecting the characteristics of the representative agent. Trading markets eventually punish irrationality and those agents that are consistently more accurate than others accumulate more economic weight in the pricing kernel. Thus, their beliefs, rather than the consensus ones, should be the one more tightly revealed (spanned) by bond prices. Since our data set provides us with the names of the forecaster, at any point of time t we are able to construct the average beliefs of the agents who were most accurate up to time t , which we denote as EBR_t^* . After discussing about the specific properties of this time series, we use EBR_t^* as a dependent variable in model specification tests, instead of future average returns, to compare the relative performance of alternative models. The use of this alternative metrics is also motivated by the fact that they are available in real-time, thus allowing us to avoid the look-ahead bias which often plagues methodologies that require model-dependent assumptions about agents expectations. An important finding is that several traditional structural models perform quite well under this metric even if they are poor in forecasting actual realized returns. In particular, models that argue about the importance of the quantity of risk channel as a source of time-variation in bond risk premia generate large R^2 and standard statistical tests cannot reject their risk premia specifications. We also show that two distinct characteristics of the cross-section of beliefs, namely disagreement and sentiment, matter for the dynamics of bond risk premia. Indeed, as predicted by heterogeneous models when bonds are risky assets, negative sentiment amplifies the marginal impact of disagreement on bond risk premia. In periods with negative sentiment (aggregate relative pessimism), large disagreement can generate large expected bond risk premia. The opposite is true in periods when sentiment is positive.

Even after taking into account the effects of heterogeneity, we find supporting evidence for some traditional macro-finance models of expected bond returns. In most cases, the empirical sign of the factor loading is consistent with predictions from theory.

Our fourth set of questions relates to an important stream of the asset pricing literature that studies the role played by behavioral biases and frictions for the dynamics of asset prices. We use our measure of expected bond returns EBR_t^* to compute agents' prediction errors. If expectations are rational these prediction errors should be orthogonal to any public information available at time t . Moreover, in a competitive and frictionless market, asset prices at time t should be able to aggregate any relevant publicly available information. We consider market-based information and run a series of orthogonality tests. In the context of bond markets, Cochrane and Piazzesi (2005) argue that the cross-section of forward rates contains important information that is relevant to construct a single factor that reveals important properties of bond risk premia. Statistically, we find strong evidence of violation of orthogonality, using information from the cross-section of forward rates. This would be consistent with a behavioral component in the formation of agents expectations as they seem to disregard statistically significant public information. However, the clear statistical rejection of conditional orthogonality does not translate in economic significance. Indeed, any attempt to use this information to correct agents forecasts and reduce their ex-post forecast errors fails when we restrict such a correction to use exclusively real-time data.

The paper proceeds as follows. Section I summarizes the empirical questions we aim to address and presents the data. Section II discusses the empirical properties of subjective expected term structures. In Section III we study the forecasting properties of expected short-term interest rates. Section IV discusses the dynamics of the expected long-term rates and the link between short and long-term rate prediction accuracy. Section V uses EBR to revisit the performance of structural models of expected bond risk premia proposed in the literature and the spanning properties of the forward curve. Section VI discusses the results and concludes.

I. Framework, Questions and Data

Given information on individual expectations about future interest rates, we compute individual subjective risk premia as follows. Let p_t^n be the logarithm of the time- t price of a risk-free zero-

coupon bond that pays one unit of the numeraire n -years in the future. Spot yields and forward rates are then defined as $y_t^n = -\frac{p_t^n}{n}$ and $f_t^n = p_t^n - p_t^{n-1}$, respectively. The realized holding period bond return in excess of the one year yield is $rx_{t+1}^n = r_{t+1}^n - y_t^1$, with the gross return being defined as $r_{t+1}^n = p_{t+1}^{n-1} - p_t^n$.

The individual expected bond excess return (EBR) of agent i at one-year horizon for a bond maturity n is defined as $erx_{i,t}^n \equiv E_t^i [rx_{t+1}^n]$. Using survey forecasts on $E_t^i [y_{t+1}^{n-1}]$ we can compute the implied cross-section of EBR as $erx_{i,t}^n = E_t^i [p_{t+1}^{n-1}] - p_t^n - y_t^1$. Indeed, from the surveys we directly observe $E_t^i [y_{t+1}^{n-1}]$, so that:

$$erx_{i,t}^n = -(n-1) \times \underbrace{E_t^i [y_{t+1}^{n-1}]}_{\substack{\text{Survey Yield} \\ \text{Forecasts}}} + ny_t^n - y_t^1. \quad (1)$$

Forecasts on future long-term interest rates depend on both expectations on future short-term interest rates $E_t^i [y_{t+s}^1]$ and future bond risk premia $erx_{i,t}^n$. We use a panel data of named forecasts on both short-term and long-term yields to address a number of questions that have been of great relevance in the financial economics literature, which we formalize in the following four questions.

First, a common assumption in the literature is the existence of a representative agent with rational expectations. While agents' expectations may be wrong, this assumption implies that they are not systematically biased and are internally consistent. Our first set of tests study the following question:

Q₁ : Are subjective expectations of bond returns unbiased and what is the extent to which the cross section of individual expectations can be approximated by the consensus belief?

We start by investigating the extent to which consensus beliefs summarize the cross section of expectations. Then, we test for the existence of a drift in forecasting errors and whether it is time-varying in a systematic way. These two properties have direct implications for a large literature that studies equilibrium models with heterogeneous beliefs and speculation. Indeed, the empirical predictions of these models depend on the significance of biases in beliefs aggregation. In these economies the equilibrium stochastic discount factor is affected by both sentiment and disagreement,⁷ thus equilibrium interest rates and bond risk premia can potentially deviate

⁷Sentiment relates to the difference between the (potentially biased) wealth-weighted average of beliefs and

from those implied by traditional Lucas tree models by amounts that depend directly on these two components of the distribution of beliefs (see Buraschi and Whelan (2010) for a detailed discussion in the context of bond markets). Finally, we investigate whether expectations on interest rates are internally consistent with the same agent expectations on future economic fundamentals (i.e. GDP growth and inflation). This is of relevance in the context of the current debate between rational and behavioral models about whether agents expectations are (un)correlated with prime properties of fundamentals under the physical measure.

Second, an extensive empirical literature argues about the existence of bond returns predictability. This may originate from either predictability of future short-term interest rates or time-variation in bond risk premia. Our second set of tests studies these two channels using data on real time individual expectations and addresses the following question:

Q₂: Can professional forecasters predict short-term interest rates? Are those agents who are better in predicting future short-term rates also better in predicting long-term bond returns?

Since our dataset provides detailed information about the identity of each forecasters, we can avoid issues related to data aggregation and can conduct an agent-specific study. Moreover, since we are already provided with their expectations, we can avoid assumptions about the models they use to form their expectations. We first investigate and rank the accuracy of each forecasters over time. For short-term rates, we distinguish between primary dealers and other financial institutions to study whether their status with the U.S. Federal Reserve Bank gives them an information advantage. We use the answer to the first question to address a second important question in the literature. Long-term bond returns can be predictable either because short term interest rates are predictable or because of time variation in risk premia. Thus, we compare the ranking in accuracy of short and long-term interest rates and study the extent to which agents who appear to do better in forecasting short-term rates also have an advantage in forecasting long-term bond returns.

Third, we investigate an alternative metrics of model misspecification to compare existing fixed-income models. Indeed, it is tradition to evaluate fixed-income models on the basis of the (unbiased) expectation under the true physical measure; disagreement relates to the cross-sectional dispersion in beliefs.

their predictive power for future *realized* average returns. On the basis of this metrics, a wide consensus has emerged in the literature that many of these macro models are not very successful empirically. Thus, we decide to use a metrics based on the use of a directly observed measure of subjective expected returns instead of indirect empirical projections. We study if the conclusions about these models are different once we use an alternative proxy for expected bond risk premia on the left hand side of these regressions. We revisit several popular models, among which those assuming (a) habit preferences, (b) early resolution of uncertainty and the existence of long-run risks, (c) belief disagreement, (d) affine factor dynamics. It is well-known, for instance, that completely affine models and more generally models that predict a tight link between the quantity of bond risk and expected bond returns find it very difficult to explain realized future bond returns. Does this conclusion change if one uses *EBR*? We ask:

Q₃: Do existing rational expectation equilibrium models explain the dynamics of EBR with the correct slope coefficient? Are properties of beliefs, such as sentiment and disagreement, important state variables in the dynamics of EBR?

Finally, an important current debate in the asset pricing literature relates to role played by the interaction of behavioral biases in agents beliefs and market frictions for the dynamics of asset prices. We use our measure of expected bond returns to compute the forecasts errors computed by professional agents. A direct test of rationality relates to the orthogonality of these forecast errors with respect to any public information available at the time these forecasts are made. If markets are competitive and frictionless, moreover, most of these information should be embedded in a cross-section of asset prices (see, e.g. Cochrane and Piazzesi (2005)).

We ask:

Q₄: What is the extent of rationality in agents beliefs? Are forecast errors orthogonal to public information available in the term structure of bond prices? Is there evidence of relevant information that can explain the forecast errors committed by agents, unspanned by financial markets at the time these predictions are made?

The last two questions are addressed after paying explicit attention to the belief aggregation issue. Indeed, a common working hypothesis of single agent models is that the representative

agent holds consensus beliefs. However, in economies with frictions (such as short-selling constraints) markets prices may temporarily span the characteristics of those agents with are less affected by these frictions, even if they hold irrational beliefs. In frictionless and competitive markets, on the other hand, the market selection hypothesis (see, e.g., Friedman (1953) and Alchian (1950)) suggests that inaccurate agents eventually lose economic weight and their influence on the stochastic discount factor. Thus, we explicitly distinguish two sets of beliefs: consensus beliefs (EBR^c) and those of the most accurate agents (EBR^*). According to Friedman (1953) and Alchian (1950), EBR^* , rather than consensus expectations, should be those more tightly linked to (spanned by) bond prices.

A. *The Data*

This section briefly introduces the data and provides a description of subjective bond excess returns. All data are monthly, from January 1988 to July 2015.

We construct measures of expected bond risk premia (EBR) directly from professional market participants' expectations regarding future yields. The BlueChip Financial Forecasts (BCFF) is a monthly survey providing extensive panel data on the expectations of professional economists working at leading financial institutions about all maturities of the yield curve and economic fundamentals, such as GDP and inflation.⁸ The contributors are asked to provide point forecasts at horizons that range from the end of the current quarter to 5 quarters ahead (6 from January 1997).

BCFF represents the most extensive dataset currently available to investigate the role of expectations formation in asset pricing. It is unique with respect to alternative commonly studied surveys along at least four dimensions. First, the dataset provides the identity of the forecasters. This allows us to track each individual in the cross-section and time-series. Indeed, most of our questions could not be addressed with datasets with pre-aggregated data. Second, the dataset is available at a monthly frequency, while other surveys, such as the Survey of Professional Forecasters' (SPF) is available only at quarterly frequency. This increases the power of asset pricing tests. Third, the number of participants in the survey is large and stable over time. In our sample it is 42 on average, with a standard deviation of about 2.3. Moreover,

⁸In our analysis we use agent specific forecasts for the Federal Funds rate, Treasury bills with maturities 3-months/6-months, Treasury notes with maturities 1, 2, 5, 10-years, and the 30-year Treasury bond.

it never falls below 35, and even considering only the forecasters who contribute to the sample for at least 5 years (60 monthly observations) the number of participants is always above 30. On the other hand, in the SPF the distribution of respondents displays significant variability: the mean number of respondents is around 40, the standard deviation is 13 and in some years the number of contributors is as low as 9. While in the early 70s the number of SPF forecasters was around 60, it decreased in two major steps in the mid 1970s and mid 1980s to as low as 14 forecasters in 1990.⁹ Fourth, Bluechip has always been administered by the same agency, while other surveys, such as SPF, have been administered by different agencies over the years. Moreover, SPF changed some of the questions in the survey, and some of these changes crucially affected the forecasting horizon.¹⁰ Finally, the survey is conducted in a short window of time, between the 25th and 27th of the month and mailed to subscribers within the first 5 days of the subsequent month. This allows the empirical analysis to be unaffected by biases induced by staleness or overlapping observations between returns and responses.

To obtain curves of expected zero coupon discount rates we use the Svensson (1994) method, which is widely used in the estimation of realized zero coupon discount rates. The Svensson (1994) model assumes that the instantaneous forward rate is given by a 5-factor parametric function. To estimate the set of parameters we minimize the weighted sum of the squared deviations between actual and model-implied prices.¹¹ We calculate the term structures using all available maturities (including 30-year Treasury yield forecasts) and obtain a monthly panel data of expected constant time-to-maturity zero coupon (continuously compounded) discount rates. The holding period is quarterly up to 1.25-years and the maturities are evenly spaced between 1 and 10-years (we disregard maturities greater than 10-years). Over the whole sample there are 97 forecasters for which we can compute the whole expected term structure of zero-coupon yields and on average they contribute to the cross-section for about 138 months. Of these 97 forecasters, 84 participate to the panel for at least 5 years, and on average they contribute to the cross section for about 154 months.

⁹If one restricts the attention to forecasters who participated to at least 8 surveys, this limits the number of data points considerably.

¹⁰For a detailed discussion on the issues related to SPF, see D'Amico and Orphanides (2008) and Giordani and Soderlind (2003).

¹¹Specifically, we search for the parameters which solve $b_t^j = \arg \min_b \sum_{h=1}^{H_t^j} \left[(P^h(b) - P_t^h) \times \frac{1}{D_t^h} \right]^2$, where H_t^j denotes the number of bonds available by forecaster j in month t , $P^h(b)$ is the model-implied price for bond $h = 1, \dots, H_t^j$, P_t^h is its expected bond price, and D_t^h is the corresponding Macaulay duration. We also impose the following set of parameter restrictions: $\beta_0 > 0$, $\beta_0 + \beta_1 > 0$, $\tau_1 > 0$, and $\tau_2 > 0$.

For realized bond data we use zero-coupon bond yields provided by Gürkaynak, Sack, and Wright (2006) which are available from the Federal Reserve website.

II. The Cross Section of Expected Term Structures

A. Subjective Expectations and Bond Risk Premia

Figure 1 gives a first look at the data. The top panel plots quartiles (Q1, Q2(median) and Q3) of the 1-year cross-sectional distribution of subjective expected excess returns on a 10-year bonds.¹² We find that, consistent with the predictions of many structural models, subjective bond risk premia are counter-cyclical: they are negatively correlated with expectations about real growth. For example, expected returns are increasing in the early part of the sample, decreasing in the high growth rate years between the dot-com bubble and the financial crisis, and spiking again around Lehman Brother collapse. Moreover, as we compare macro versus short rate expectations, subjective expectations appear consistent with a Taylor rule relationship. For example, between the years 1988 and 1990 agents expected inflation to increase. At the same time forecasters expected the Federal Reserve to increase short term rates and that this policy would have a contractionary effect on the real economy (GDP growth).

At the same time, we document large unconditional heterogeneity in the cross section of EBR forecasts. The median (Q2) forecaster EBR is 1.06% for 10-year bonds. However, the first and third quartiles (Q1 and Q3) are -1.66% and +3.57% for the same maturity, respectively. This implies that while the consensus believes in a positive risk premium, a significant fraction of investors believe in a negative bond risk premium. Moreover, the spread between the Q1 and Q3 unconditional expected excess bond returns is increasing with the bond maturity.

The conditional properties of the cross-sectional distribution of EBR display rich dynamics in the time series (see again the top panel of Figure 1). There exists significant time-varying heterogeneity around the consensus forecast. Given the wide use of consensus (average) expectations both in the literature and in the financial industry, it is interesting to test more formally the null hypothesis that the cross-sectional properties of expectations can indeed be summarized by the consensus. In order to do this, we compute the interquartile range (IQR) of the cross-sectional distribution of EBR, as the difference between Q3 and Q1, for all bond

¹²1-year average expectations are computed from 4 and 5 quarter ahead projections.

maturities $n = 2, \dots, 10$, and then regress it on the consensus forecast for the corresponding bond maturity. The slope coefficients of these regressions are positive, and statistically significant for all maturities, but the variations in the consensus forecasts explain only around 3% of the variation in the IQR. Moreover, we can strongly reject the hypothesis that the IQR is constant. In fact, the slope coefficient of a regression of IQR on its 1-year lag is significantly different from zero, for all maturities and at all levels. Therefore, the dispersion in beliefs varies over time and it is not merely a scaled version of the consensus: the mean is not a sufficient statistics for the cross section of expectations.

The bottom panel of Figure 1 highlights the time variation in heterogeneity by plotting the cross-sectional standard deviation of EBR standardized by the full sample mean EBR, for bond maturities 2, 5 and 10-year. The figure also shows that the dispersion in beliefs is state-dependent: it tends to rise at the onset of recessionary periods and drop again as the economy recovers.¹³ Also, disagreement is non-monotonic in maturity displaying a ‘hump shape’ around the 5-year maturity. These findings motivate a rigorous study of whether the assumption that the marginal investor has average (consensus) expectations, as often assumed in the literature, is innocuous.

B. Belief Persistence

Disagreement about short rates, bond returns, and the macro economy are all persistent. This raises an interesting question: is disagreement a result of dogmatic beliefs, information frictions, or both? In order to address this question we first rank all forecasters according to whether in a given month t their forecast is in the first, second, third or fourth quartile of the cross-sectional distribution. We repeat this exercise for all months in the sample and compute transition probabilities: the probability that forecasters in a given quartile at time t stay in that particular quartile in $t+1$ or move to a different quartile of the distribution. We do this first for short rates in the top left panel of Table I. If views are not persistent, all the entries in these transition matrix should be approximately equal to 25%. Instead, we find that the diagonal elements are significantly higher than 25%, in particular for the most extreme quantiles, Q1 and Q4 where

¹³The counter-cyclicality of the dispersion in beliefs is consistent with the empirical evidence in Patton and Timmermann (2010) and Buraschi, Trojani, and Vedolin (2014), among others.

they are always above 70%.¹⁴

The question of belief persistence is particularly important in the context of bond pricing models since whether agents are persistently optimistic or pessimistic about bond risk premia is related to agents' perception about bonds being hedging assets or rather risky bets on consumption (inflation) risk. In the first case, bonds should earn a negative risk premium, in the second expected bond risk premia should be positive. Thus, we estimate the extent to which individual forecasters are persistently in one particular quartile of the cross-sectional distribution of subjective *EBRs*. The top right panels of Table I show the transition matrices for subjective excess returns. The results suggest that forecasters have persistent beliefs about bond risk premia, relative to consensus excess returns. For example, a forecaster in the first quartile of the cross-sectional distribution of 2-year EBR has a probability of 75% to stay in the first quartile the following month, and this probability is 74% for the 10-year EBR, which is about three times what it should be under the null hypothesis of no persistence. In all cases, the probability of remaining in the same quartile is significantly higher than 25% at a level of 5%.

C. Internally Consistent Beliefs

Some readers may interpret the previous results as prima-facie evidence of either irrationality in the formation of beliefs or of dogmatic priors in agents' models. We address this conjecture by investigating whether expected term structures are consistent with agents' expectations about future economic fundamentals. Since we know the identity of each forecaster on both future interest rates and future state of the economy (GDP growth and inflation), we can ask whether these are mutually consistent.

We find that agents who are marginally more optimistic or pessimistic about macroeconomic variables are consistently in one particular quartile of the cross-sectional distribution of short term interest rates, as shown in Panel B of Table I. If one focuses on the corners of this table, we find that analysts who forecast lower short-term interest rates are also those forecasting lower GDP growth and, at the same time, lower CPI inflation. For instance, 35% of those who are in the first quartile of the distribution of future short-term interest rate forecasts are also

¹⁴This result is striking and even stronger than what Patton and Timmermann (2010) document for macroeconomic forecasts using data from the Consensus Economics Inc, at a quarterly frequency. Transition matrices for GDP and CPI growth expectations show very similar results and are available from the authors upon request.

in the first quartile of the distribution for GDP growth forecasts; similarly, 41% of those who are in the first quartile of the distribution of future short-term interest rate forecasts are also in the first quartile of the distribution for CPI inflation forecasts. This relation between forecasts at the individual level is consistent with the idea that good states of the economy are generally characterised by increasing yields, at least at short maturity. At the same time, the pattern is not deterministic, suggesting that beliefs on interest rates and the macroeconomy (GDP and inflation) are not driven by a single factor.

In order to investigate the drivers of this disagreement we directly study the dynamics and accuracy of these beliefs. In this context, it is useful to distinguish between beliefs about short-term interest rates and bond risk premia. This is the topic of the next two sections, which are cast in the predictability regression framework used in the classical bond literature.

III. The Short Rate

A. Accuracy of Professional Forecasters

We first study the degree of accuracy of individual forecasts for future short term rates and the degree of heterogeneity in the accuracy. We initially explore this question in the context of simple predictive regressions for the three-month Treasury yield. Due to its persistence, we run the predictive regression in differences where the dependent variable is specified as future realized monthly changes in 3-month rate and the independent variables are the corresponding expected changes according to survey beliefs:

$$\Delta y_{t+1}^{3m} = \alpha_i^{3m} + \beta_i^{3m} [E_t^i(y_{t+1}^{3m}) - y_t^{3m}] + \epsilon_{i,t+1}^{3m}, \quad (2)$$

where $\Delta y_{t+1}^{3m} = y_{t+1}^{3m} - y_t^{3m}$. Regression (2) is run at the individual level, where i denotes each single contributor to the BCFF panel. This avoids several problems related to data aggregation that affect retail surveys in which the identity of the forecaster is not known. For robustness, we focus on contributors with at least 5 years (60 months) of forecasts. Figure 2 shows the distribution of regression coefficients and R^2 for each forecaster.

With the exception of few forecasters, most of the estimated β_i slope coefficients are positive and statistically significant. This suggests that professional forecasters do a relatively good job

in predicting future short rates. Moreover, a few forecasters are extremely accurate, with slope coefficients larger than 0.5 and R^2 in excess of 10%, with some agents producing an R^2 in excess of 30%.

To study in further details the properties of the distribution of short rate survey expectations, we compare them to five models that have been widely used in the literature: (i) a unit root forecast (UR), (ii) $VAR(1)$, (iii) $VAR_D(p)$, (iv) the 12m-to-15m forward rate (FR), and (v) a Taylor rule ($Taylor$). All these benchmark models are estimated in real-time using current and lagged information to avoid any look-ahead bias. Table II shows the root mean square prediction errors for the 3-month interest rate in 12-months ahead. The two VAR models use 10-years of rolling monthly data on 3-month rates, the slope (10-year minus 1-year rate), and curvature (two times the 5-year yield minus the 1- and 10-year yields) of the term structure. $VAR(1)$ uses a one month lag, $VAR_D(p)$ has a variable lag dimension which is chosen dynamically using the Akaike information criterion. The Taylor-rule model is a vector autoregression estimated on 10 years of rolling monthly data using the 3-month rates, the annual inflation rate of the All-Urban consumer price index excluding food and energy, and the industrial production output gap estimated using a recursive Hodrick-Prescott Filter.

Two main results emerge. First, consensus forecasts are reasonably accurate with respect to several econometric benchmark models estimated in real-time. Both the forward rate and a Taylor rule model generate higher RMSE than consensus forecasts (1.71 and 1.51, respectively, versus 1.41). Also, a factor VAR model using the three term structure characteristics (level, slope, and curvature) with dynamically adjusted lag structure does not perform better out-of-sample.

Second, there is significant dispersion in accuracy around the consensus forecasts. Indeed, some forecasters are remarkably better than real-time econometric models. We sort forecasters according to their rankings in terms of accuracy over the whole sample.¹⁵ Table II shows that forecasters in the top-decile (Top) are significantly better than any benchmark. On the other hand, the bottom-decile ($Bottom$) forecasters underperform all benchmarks with the exception of predictions implied by the current forward rate. The performance of top forecasters implies that it should be indeed possible to beat a simple aggregation of beliefs, such as the consensus. To some extent, this is due to the fact that some forecasters perform rather poorly. We also

¹⁵Details on the sorting procedure are provided in subsection B.

expect that it should be possible to estimate more sophisticated dynamic econometric models with better forecasting performance than those specified in (i)-(v).¹⁶ The surprising take-away of these results, however, is that this is not an effortless task, if one uses out-of-sample metrics.

Finally, the relative performance of top and bottom forecasters is state-dependent. During recessions, the spread in RMSE between top and bottom forecasters widens significantly. Moreover, during recessions top forecasters seem to do much better than market-based forecasts such as the forward rate.

To document the cross-sectional differences, let the root-mean squared error of each individual forecaster i be defined as $RMSE_i^{3m}(Surv) = \sqrt{\frac{1}{T_i - t_{0,i} + 1} \sum_{t=t_{0,i}}^{T_i} (y_{t+1}^{3m} - E_t^i [y_{t+1}^{3m}])^2}$. We define the relative accuracy \mathcal{A}_i^{3m} of each forecaster as the ratio between the $RMSE$ of each forecaster's expectation and the $RMSE$ of the unit-root benchmark. This benchmark is chosen both because of its performance and because it does not rely on any parameter estimation.

$$\mathcal{A}_i^{3m} = \frac{RMSE_i^{3m}(Surv)}{RMSE^{3m}(UnitRoot)}.$$

Since the panel is unbalanced, as forecasters do not participate in the same periods, we compare the relative performance of each forecaster with respect to the naive benchmark during the period in which the forecaster is present. The left panel of Figure 3 displays the distribution of \mathcal{A}_i^{3m} for the 84 contributors with at least 5 years of monthly forecasts. Noticeably, a significant mass of individual forecasters have \mathcal{A}_i between 0.90 and 1.10, suggesting that several agents are as good as the unit-root benchmark (and significantly better than a simple VAR model). Moreover, some agents are extremely precise with $\mathcal{A}_i < 0.90$, suggesting that some professional forecasters can provide reasonably good measures of expected bond returns. At the same time, some agents are very poor forecasters with $\mathcal{A}_i > 1.20$.

We find these differences in short rate forecast accuracy to be persistent. The left panel of Table III shows the probability of a forecaster transitioning from a given quartile of the cross-sectional distribution of forecasts' accuracy to another quartile in the following month, for 3-month interest rates. The results show that agents who are in the top or bottom quartiles by forecast accuracy tend to remain in the same quartile the following periods.

An intriguing property of the results of regression (2) is that the cross-sectional distri-

¹⁶See, for instance, the work by Giacomelli, Laursen, and Singleton (2016).

bution of intercepts is largely skewed towards negative values: α_i^{3m} is negative for 79 of the 84 forecasters and significantly different from zero for slightly less than half of the forecasters. This suggests that the average forecaster has been surprised by the extended decline in short term interest rates over our sample period. To investigate the time series property of these surprises, let us define $E_t^i(y_{t+1}^{3m})$ and $E(y_{t+1}^{3m}|\mathcal{M}_t)$ the individual forecasts of agent i and those implied by the benchmark model \mathcal{M}_t , respectively. The one-year ahead forecast error is $u_i(t+1) = y_{t+1}^{3m} - E_t^i(y_{t+1}^{3m})$ and let $s_i(t)$ be the difference between the model-implied expectation and the subjective expectation, a measure of the relative sentiment of individual forecaster relative to the model \mathcal{M}_t . Finally, let $U(t)$ be the cumulative forecasts error:

$$s_i(t) = E[u_i(t+1)|\mathcal{M}_t] \quad \text{and} \quad U_i(t) = \sum_{s=0}^t u_i(s). \quad (3)$$

Figure 4 summarizes the dynamics of $s_i(t)$ and $U_i(t)$ for the consensus agent. Three observations emerge: First, $U_i(t)$ is not a martingale and has a negative drift, which is reflected in the negative α in the predictive regression (2). Second, cumulative errors increase quite significantly in the early 90s, in the early 2000s, and during the recent financial crisis. These three periods happen to coincide with the largest GDP contractions in our sample. Third, the plot for $s_i(t)$ shows that during these periods individual forecasts deviate from the model-based ones.¹⁷

This ex-ante bias in subjective expectations, $s_i(t)$, displays a striking correlation with the in-12m for-3m forward spread (the forward bond risk premium), with a correlation equal to -60% .¹⁸ When the consensus agent is more optimistic about short term rates, the forward spread is also high. This is consistent with a large literature that studies economies with agents holding heterogeneous subjective beliefs in which the equilibrium stochastic discount factor and interest rates are affected by a belief aggregation bias, as discussed in Buraschi and Whelan (2010), Jouini and Napp (2006), and references therein. Indeed, in these models average optimism/pessimism enters directly the stochastic discount factor. To investigate this

¹⁷These results are related to the work of Cieslak (2016), who shows that “entering recessions, agents systematically overestimate the future real rate and underestimate unemployment. These forecast errors induce a predictable component in realized bond excess returns”.

¹⁸The forward spread is computed as the difference between the 12M \rightarrow 15M forward rate and the spot 3-month yield.

property, we construct a measure of *Sentiment* in survey forecasts,

$$\mathcal{S}_t = -s_i(t), \quad i = \text{consensus}. \quad (4)$$

We take the negative value of the expected forecast error so that positive values of \mathcal{S}_t correspond to an average *optimism* of the survey forecasters. Is sentiment \mathcal{S}_t on bond yields consistent with sentiment about economic fundamentals? To address this question, we use expectations on one year GDP growth and calculate $\mathcal{S}_t^{gdp} = E_t^i(gdp_{t+1}) - E(gdp_{t+1}|\mathcal{M}_t)$.¹⁹ To compute one-year model forecasts under the econometrician measure \mathcal{M}_t , we borrow from the empirical macro-finance literature and use a time-series $AR(4)$ model using quarterly realized *GDP* growth (see, for instance, Marcellino (2008) and references therein). We find that the two measures of sentiment, \mathcal{S}_t and \mathcal{S}_t^{gdp} , are highly correlated, which is consistent with sentiment in bond markets to be related to agents' beliefs about the dynamics of fundamentals.²⁰

B. Who are the top forecasters?

Is it possible to identify a subset of forecasters who are especially good at predicting short-term interest rates? Since forecasters contribution to the survey can occur at different time periods, we compute the squared forecast error at each time t , and the percentiles of these squared errors for each forecaster, that we call *accuracy ranking* percentiles, $\mathcal{R}_{i,t}$. Then we compute the time average \mathcal{R}_i of these percentiles. Low percentiles correspond to greater accuracy. The best forecasters in terms of average percentiles of squared forecast errors are summarized in Table IV.

Interestingly, all of the top five institutions in this list (and 7 out of the first 10) are currently primary dealers, or have been primary dealers at least once in our sample period.²¹ This is remarkable given that only 17 of the 51 institutions in the survey with at least 10 years of contributions are or have been primary dealers.²² Primary dealers are trading counterparties of the Fed in its implementation of monetary policy and they are also expected to make

¹⁹GDP growth sentiment is available at a quarterly frequency since realized GDP growth is quarterly.

²⁰This is consistent with theoretical models in which the source of the ex-ante expectation bias is driven by sentiment in the endowment growth. While GDP growth sentiment might be more intuitively linked to the consumption growth sentiment of these models, we will focus on the short-rate sentiment \mathcal{S}_t as an explanatory variable for bond risk premia in section V, since it is available monthly instead of quarterly.

²¹We consider forecasters with at least 10 years of data.

²²The list of primary dealers at every point in time can be obtained from the Federal Reserve Bank website.

markets for the Fed on behalf of its official account holders, and to bid on a pro-rata basis in all Treasury auctions. Their superior performance is consistent either with primary dealers superior information about the Fed's implementation of monetary policy or, more simply, with an information flow advantage originating from their role as market makers in Treasury bonds. In either case, the result is quite important given that the top 5 primary dealers hold about 50% of outstanding Treasuries.

B.1. State dependence in the difference in accuracy

Are the differences in forecast accuracy and the bias between top and bottom forecasters state dependent? The top panel in Figure 6 shows the difference in the absolute error between forecasters in the top decile and two benchmarks, namely the consensus forecast and the expectation from the unit root model. Negative values occur when the top forecasters are more accurate than the benchmark. Independent of the benchmark used, we can statistically reject at the 5% confidence level the null hypothesis that the difference is constant. Moreover, we find that the top forecasters are more accurate than the consensus in 1990-1993, 2001-2003, and 2008-2011. All these periods are recessions and are characterized by important changes in the stance of the monetary policy.

To analyze in further detail this result and whether it is due to the use of ex-post ranking information, we compare the time series of average accuracy percentiles for primary dealers (PDs) versus all other contributors (NPDs). As we expected, we find clear evidence that the comparative advantage of PDs is stronger in recession periods. Namely, the average expectation errors for PDs and NPDs diverge significantly in the early 90s, in the early 2000s and during the recent financial crisis. These periods are all characterized by a change of monetary policy in which the Fed has aggressively reduced the short term rate. While these decisions seem to take by surprise the consensus agent, whose expected short rates are biased upward in these subperiods, primary dealers are significantly more accurate, and this is especially true during the recent financial crisis.

To investigate these differences formally, we split the sample in two parts to capture persistent periods of increasing and decreasing interest rates, respectively. We compute the exponential moving average of the monthly change in the fed fund rate over the previous 12 months.²³

²³Results are robust to the choice of time periods for the moving average.

Considering the whole sample, there are 195 months in which this exponential moving average of changes is negative and 113 in which it is positive. We then recompute the average accuracy percentiles for each individual forecaster explicitly distinguishing these two time periods and we compare the distribution of accuracy percentiles for top and bottom forecasters using a Kolmogorov-Smirnov test. The null hypothesis of the Kolmogorov-Smirnov test is that the accuracy percentiles of top and bottom are drawn from the same distribution. Unconditionally (considering the full sample), the p-value of the test is 15%, which implies that we cannot reject the null hypothesis. However, in the subperiod in which the Fed has been more active in conducting a dovish policy on the short term rate, the p-value of the test is 1.61%. In these sub-periods we can strongly reject the hypothesis that accuracy percentiles of top and bottom forecasters are drawn from the same distribution. On the other hand, the p-value of the test in periods of increasing fed fund rate is 47.75%, suggesting that the distribution of accuracy is very similar in these periods. A Mann-Whitney U-test for the difference in medians between the accuracy percentile distributions yields similar results: unconditionally the p-value is 4.98%, in periods of increasing rates it is 58.99%, and in periods of decreasing rates it is 0.80%. We repeat these test using the ex-ante classification of PDs versus NPDs and we obtain the same conclusions.

The evidence that some financial institutions are much better during inflection points, e.g. turns of business cycles when the Fed turns dovish by reducing the interest rate, is intriguing. It is potentially consistent with these institutions either being better in forecasting economic fundamentals, having specific information about the stance of the monetary policy, or having useful order flow information on short-term bonds. Since we have named forecasts also on economic fundamentals, we test the first hypothesis by comparing the accuracy of top versus bottom forecasters about future real economic growth and inflation. We find that top forecasters do not perform better than other agents in forecasting the inputs of the Taylor rule, i.e. inflation and GDP growth.²⁴ In fact, if anything, the accuracy of top forecasters' inflation expectations is lower, with an average accuracy ranking of 0.57 versus 0.42 for bottom forecasters (see bottom panels of Figure 6). Similarly, the GDP growth accuracy is worse for the top short rate forecasters, at 0.56 versus 0.51, respectively. Moreover, despite the time variation, the

²⁴Note that realized GDP growth is available only quarterly. Therefore, the time series of GDP growth accuracy is also quarterly.

top short-rate forecasters (the great majority of which are primary dealers) are virtually never more accurate than the worse short-rate forecasters in predicting either inflation or real GDP growth. Indeed, the best macro forecasters on average are institutions like Action Economics and ClearView Economics; primary dealers such as Goldman Sachs, J.P. Morgan and Nomura are consistently in the worst half of growth and inflation forecast accuracy.

This suggests that either order flow information in fixed income markets plays an important role for the precision of interest rate forecasts, or primary dealers have specific information about (or better models to interpret) monetary policy.

IV. The Long Rate

In this section we focus on two questions: First, given a direct subjective measure of expected excess bond returns $erx_{i,t}^n$, we revisit the literature on the time-variation of risk premia. This literature plays an important role in the interpretation of the rejection of the expectation hypothesis in bond markets. Second, we quantify the accuracy of professional forecasters and ask whether the superior predictive ability displayed by some financial institutions (among which primary dealers) for short-term rates leads them to a competitive advantage also in predicting long-term bond returns. Since long-term bond returns are affected by both changes in short-term interest rates and bond risk premia, if bond risk premia are small and slowly time-varying we should find that primary dealers have an edge in forecasting also long-term bond returns. This is, therefore, an indirect test of the importance of the risk premium channel in the predictability of long-term bond returns.

A. Time-varying risk premia

An extensive literature studies the properties of bond risk premia and argues that they are time varying. Empirical proxies of conditional bond risk premia usually either require the specification of a model or they use ex-post data on bond returns. The limit of arguments based on the central limit theorem is of course the lack of sufficiently long data samples. For this reason, some studies have argued that the results are not statistically convincing. Our data allows us to study bond risk premia directly using the dynamics of expectations that are obtained in a model independent way. Given the time series of subjective bond risk premia

$erx_{i,t+1}^n$, we run regressions for different quartiles of the cross-sectional distribution for 2, 5 and 10-year zero-coupon bonds on a constant and their own lag at the 1-year horizon:

$$erx_{i,t+1}^n = \alpha_i^n + \beta_i^n erx_{i,t}^n + \epsilon_{i,t+1}^n. \quad (5)$$

The results are summarized in Table V and show that the slope coefficients are significantly different from zero for all quartiles i at any traditional statistical levels. The results are very strong and support the hypothesis that expected excess bond returns are time varying. Moving from the first to the fourth quartile, for all bond maturities, the autocorrelation coefficient is monotonically increasing. Those agents who believe bonds are hedges (e.g. *EBR* pessimists) have less persistent and less predictable (in the R^2 sense) expected bond returns. However, even for these agents the t-statistics of the slope coefficient is close to 3. These results offer, for the first time, direct evidence supporting the interpretation that predictability is likely due to time variation in expected excess bond returns.

B. Accuracy of Professional Forecasters

To assess the accuracy of these surveys and the degree of heterogeneity, we first run a simple predictive regression of realized excess returns on the subjective EBR, for each single contributor to the BCFF panel, focusing on the contributors with at least 5 years (60 months) of forecasts:

$$rx_{t+1}^n = \alpha_i^n + \beta_i^n erx_{i,t}^n + \epsilon_{i,t+1}^n. \quad (6)$$

Figure 7 shows the distribution of individual regression coefficients and R^2 of regression (6). Despite the heterogeneity in accuracy, a few forecasters are extremely accurate with slope coefficients close to one and R^2 larger than 20%. The correlation between expectations and future realization of excess bond returns is positive for 69 out of 84 forecasters. This *positive* relation between expectations and realizations is the opposite to what Greenwood and Schleifer (2014) document in the context of the stock market, and to what Koijen, Schmeling, and Vrugt (2015) find in the context of global equities, currencies and global fixed income returns across countries.²⁵ This may be due either to issues related to the aggregation in those data sets or to

²⁵Koijen, Schmeling, and Vrugt (2015) also find that survey expectations of returns negatively predict future returns in the time series in three major asset classes: global equities, currencies, and global fixed income.

the fact that bond returns are easier to predict than equities, due to the absence of uncertainty about future nominal cash-flows. Alternatively, it is possible that professional forecasters are simply better than retail investors and consumers, due to professional incentives and access to superior information. Our results show that agents beliefs are substantially more accurate than previously thought.

We then study forecast accuracy at the level of each individual forecaster i by computing their root mean squared errors ($RMSE_i^n$) as

$$RMSE_i^n(Surv) = \sqrt{\frac{1}{T_i - t_{0,i} + 1} \sum_{t=t_{0,i}}^{T_i} (rx_{t+1}^n - erx_{i,t}^n)^2},$$

for bond maturity $n = 10$. They range between 5.2997 and 15.8325. Since individual forecasters may appear in the sample at different times, we assess their accuracy relative to a model. In the case of long term bonds, the model that is most commonly used in the literature is Cochrane and Piazzesi (2005). The Cochrane and Piazzesi (2005) return forecasting factor is a tent-shaped linear combination of forward rates and it subsumes the information contained in the level, slope and curvature of the term structure. However, the in-sample predictive content of the Cochrane-Piazzesi factor relies on estimates of factor loadings that were not available in real time. For example, the coefficients of the ‘tent-shaped’ factor used to forecast returns in the 1990s uses information available during the 2000s. In real time the shape of the factor loadings on the forward curve displays time variation (see, e.g., Bauer and Hamilton (2015)). We construct a real-time version of the CP factor as follows. We initialise the factor loadings with 5-years of data from January 1983 to January 1988. Then, using a rolling window we estimate factor loadings used to construct a date t predicting factor, $CP(t)$, using realized returns available 1-year ago.

Using this real-time $CP(t)$ factor we proceed with an out-of-sample assessment:²⁶ we initialize the model in January 1998 and obtain model-implied expectations recursively using rolling windows. Then, we compare this to survey forecasts, which are out-of-sample by construction

However, instead of looking at the slope coefficient of predictive regressions, they show this by building a survey-based portfolio strategy. The strategy goes a dollar long or short in country i in month t when the consensus forecast is above or below a certain threshold, which is set to be equal to the middle value World Economic Survey respondents can select.

²⁶In the context of equity returns, Goyal and Welch (2008) document significant differences of in-sample versus out-of-sample performances of several well-known models.

by computing a measure of relative performance \mathcal{A}_i for each forecaster with respect to the *CP* model:

$$\mathcal{A}_i^n = \frac{RMSE_i^n(Survey)}{RMSE^n(CP)}.$$

Values smaller than one imply better performance under the subjective measure. The right panel of Figure 3 presents the histogram for the cross-section of \mathcal{A}_i^n , for bond maturity n equal to 10 years. We find that an important fraction of survey forecasters perform better than the *CP* model estimated out-of-sample. About 40% of the forecasters have $\mathcal{A}_i^n < 1$. At the same time, we find that about a fifth of forecasters have $\mathcal{A}_i^n > 1.2$, thus producing forecasts significantly worse than the *CP* model. These findings confirm the large degree of heterogeneity in performance documented above, but they also suggest that professional forecasters expected bond returns are competitive relative to popular reduced form models. In fact, not only there is evidence of accuracy in the cross-section, but this accuracy tends to be persistent. To quantify the persistence, we rank all forecasters according to their accuracy in month t within the distribution of all forecasters in the panel at that moment. Namely, we calculate the percentile of squared forecast errors of bond excess returns. We repeat this exercise for all months in the sample and compute transition probabilities, defined as the probability that forecasters in a given quartile at time t stay in that particular quartile in $t + 1$ or move to a different quartile of the distribution. If accuracy were not persistent, all the entries in the middle panel of Table III should be approximately equal to 25%. We find instead that the diagonal elements are significantly higher than 25%, especially for the most extreme quantiles, Q1 and Q4. For example, a forecaster in the first quartile of the cross-sectional distribution of 10-year EBR accuracy has a probability of 58% to stay in the first quartile of accuracy the following month. This probability is 70% for the 4th quartile, which contains the worst forecasters, suggesting that bad forecasters produce more persistently poor forecasts, even more so than good forecasters. Two conclusions emerge. First, expectations of a significant fraction of professional forecasters are rather competitive with respect to popular fixed income models, and not as bad as previously reported. Second, surveys can be used to build reliable measures of bond risk premia. However, one needs to be mindful of the heterogeneity in the distribution of beliefs. The assumption that consensus can be used as a sufficient statistics for the whole panel and can proxy the beliefs of the marginal agent are not supported by our results.

C. Link between short and long rate accuracy

The previous section documents that some institutions have a comparative advantage in predicting the short rate. Does their superior predictive power for the short rate lead also to superior predictive power on long rates? This question is important for several reasons. First, if the answer were positive one could conclude that long-term bond returns are mainly driven by short rates over the life of the bond. A rejection of this hypothesis, on the other hand, would suggest that a significant fraction of the dynamics of long-term bond returns is driven by other components, such as bond risk premia. In this case, knowledge of the dynamics of short-term rates is not sufficient to generate accurate forecasts of long-term bond returns.

To test this hypothesis, we compute the accuracy percentiles on the 10-year excess bond returns for each individual forecaster by squaring forecast errors at each month t , ranking them, and computing averages across time periods. Finally, we compare these long-term accuracy percentiles with the corresponding accuracy percentiles for the short rate. The two rankings are correlated: A regression of the 10-year accuracy percentiles on the 3-month accuracy has a significant slope coefficient of 0.42 and an adjusted R-squared of 21%. However, this correlation is rather weak. The right panel of Table III summarizes the joint distribution of forecast accuracy for the 10-year EBR and 3-month yield, that is the probability of being in a given quartile of the 3-month yield accuracy percentile distribution *and* a given quartile of the 10-year EBR accuracy percentile distribution, at the same time. The elements on the diagonal show the existence of a link between the two accuracies, which is not surprising given that the correlation between realized 3-month and 10-year yield, at the monthly frequency, is about 86%, and the correlation between the 3 month yield and the slope of the term structure (computed as the difference between the 10 and the 1-year yield) is -74%. However, the rank correlation in terms of accuracy on the 3-month yield and on the 10-year EBR is far from perfect. Only 34% of the top short-rate forecasters are also top long-term yield forecasters.

The best forecasters in terms of average percentiles of squared forecast errors for the 10-year bond are summarized in the left panel of Table VI. Most of the top-ten forecasters for the 10 year bond yield are not in the top-ten list for the short-term rate. Indeed, only two are also in the top-ten list for the short-term rate. Moreover, only three (UBS, Goldman Sachs, and Nomura) of the top ten forecasters for the 10-year bond returns are primary dealers. At the

same time, financial institutions such as J.P. Morgan, BMO Capital Markets, Bank of America, Chase Manhattan Bank who can forecast the short rate rather well, are poor performers at the long end of the term structure. To formally analyze the relative performance of PDs and NPDs at the long end of the term structure, we statistically test the null hypothesis that their accuracy percentiles for the 10-year excess return are drawn from the same distribution using a Kolmogorov-Smirnov test. Unconditionally (considering the full sample), the p-value of the test is 68.2%. Even after distinguishing periods of increasing and decreasing rates or using the Mann-Whitney test, we cannot reject the null hypothesis with p-values larger than 50%. Thus, primary dealers have a significantly better predictive performance only for the short rate. Greater accuracy of primary dealers on the short-end of the term structure is not reflected in a greater accuracy on long-term bond excess returns.

Finally, we investigate the conditional properties of these forecasts. Figure 8 summarizes the difference in absolute errors between the top decile of forecasters (“Top”), the consensus and a benchmark (CP) model for long-term bond returns. The outperformance of the Top with respect to consensus forecasts is persistent as highlighted by the persistent negative red shaded area in Figure 8. Moreover, and perhaps even more interestingly, the periods in which the top long-rate forecasters do well relative to the benchmark model are not the same as those in which the top short-rate forecasters outperform their benchmark. This can be seen by comparing the shaded areas in Figures 8 and Figure 6. For example, in the periods January 2005 - February 2006, September 2008 - April 2009, and January 2012 - October 2015 the outperformance for the long-term bond returns is not correlated with a similar outperformance for the short-term rate.

We conclude that an important component of the dynamics of expected bond returns at longer maturities is a risk premium term, which is not completely revealed by the dynamics of short term interest rates. Given the heterogeneity in accuracy, it is then natural to ask how beliefs should be aggregated to construct an empirical proxy of subjective bond risk premia and this is the topic of the following subsection.

D. Belief aggregation and subjective bond risk premia

It is common in the empirical literature to use consensus expectations as a proxy of subjective beliefs. In some cases, the choice is forced by data limitations. In the context of asset

pricing, this is tantamount to assuming that the marginal agent holds consensus beliefs. Different streams of the literature, however, study equilibrium models in which the beliefs of the marginal agent deviate from consensus. For instance, the behavioral finance literature argues that in presence of short-selling constraints marginal agents ought to be those holding optimistic beliefs about expected returns (see e.g. Scheinkman and Xiong (2003) and Hong, Sraer, and Yu (2013)). Since pessimists cannot short-sell, their beliefs are not revealed (spanned) by equilibrium asset prices. The general equilibrium literature that studies economies where agents speculate on their (heterogeneous) beliefs argues, on the other hand, that in absence of short-selling constraints irrational agents eventually lose economic weight to the benefit of less biased agents. What matters is not agents' optimism but rather their accuracy. The superior accuracy of rational agents allows them to accumulate economic importance in the Pareto weights of the representative agent (see, e.g., Basak (2005), Buraschi and Jiltsov (2006), Jouini and Napp (2006), Xiong and Yan (2010), Chen, Joslin, and Tran (2012), Buraschi and Whelan (2010), Ehling, Gallmeyer, Heyerdahl-Larsen, and Illeditsch (2015), among others). This argument, consistent with the original "market selection hypothesis" by Friedman (1953) and Alchian (1950), implies that bond prices should span the beliefs of the most accurate agents (i.e. closest to the actual physical probability). As Alchian (1950) argues, "*Realized profits [...] are the mark of success and viability. It does not matter through what process of reasoning or motivation such success was achieved. The fact of its accomplishment is sufficient. This is the criterion by which the economic system selects survivors: those who realize positive profits are the survivors; those who suffer losses disappear.*" If some agents have been consistently more accurate than others, they would have been accumulating more economic weight in the pricing kernel. Thus, these beliefs, rather than the consensus ones, should be a better proxy of the beliefs of the marginal agent.

In light of this argument, we use information on agents beliefs from both the time series and the cross section to build an alternative aggregate measure of subjective bond risk premia.

First, at every month t we sort agents according to the level of their accuracy up to that date. Specifically, we compute the average accuracy ranking percentiles up to t of all agents present in the panel at time t and form quartile portfolios of agents based in this past accuracy.²⁷

²⁷Note that the realized forecast errors up to time t will be based on expectations formed up to time $t - 12$ months, since we consider forecasts with a fixed one year horizon. We use an initial window of 10 years.

According to Friedman (1953) and Alchian (1950), agents relying on the predictions of the top quartile agents have accumulated greater economic wealth as of time t . Then, we compute the average EBR for different bond maturities within each quartile. The EBR formed on the quartile with greatest past accuracy is our preferred measure of aggregate subjective bond risk premia, which we denote EBR^* . Indeed, this selection procedure has the advantage of being in real time (it uses only past information) and is not affected by look-ahead bias. When we compare the top 10 forecasters based on the average ex-ante and unconditional performance rankings (see right panel of Table VI), we find that the intersection of the two rankings is made of seven out of ten forecasters. On the one hand, this shows that aggregating beliefs using an ex-ante approach does matter; on the other hand, the persistence in accuracy is such that the two rankings are highly correlated. We find that EBR^* is significantly positively correlated to future realized excess returns, for all bond maturities, and it has higher predictive regression R-squared and lower RMSE than the consensus. The correlation between EBR^* and future realized returns is 18% and 21% for the 5 years and 10 years bond, respectively. The slope coefficient of a regression of future realized returns onto EBR^* is positive and equal to 0.48 and 0.44 for the 5 and 10 years bond, respectively. We compare the $RMSE$ for 10-year bond excess returns of our EBR^* measure with the consensus and the real-time version of the Cochrane and Piazzesi factor based on parameters estimated using available information at time t . We find that at the 10 year maturity, EBR^* generates a $RMSE$ equal to 8.37 versus 8.69 and 8.97 for the consensus and CP, respectively.

These results suggest that EBR^* is an attractive empirical proxy of expected bond risk premia, for several reasons. It is genuinely ex-ante, contrary to other ex-post measures. Moreover, it penalizes inaccurate agents which is consistent with the spirit of the market selection hypothesis in competitive markets by Friedman (1953) and Alchian (1950). Finally, this performs well against popular forecasting models purely based on price information.

V. Learning about the Determinants of Bond Risk Premia

The empirical evaluation of structural asset pricing models is traditionally conducted by approximating conditional expectations in the specification of risk premia with sample projections of future realizations $rx_{t,t+T}$ onto observables variables which are part of the information set

\mathcal{F}_t . This is potentially problematic for at least three reasons. First, sample projections based on future realizations can be quite different from true investors expectations. We have a clear example of this in the context of our data when we find that, at the individual level, $erx_{i,t}$ are more persistent than what a pure rational model would imply. Second, long horizon predictability regressions give rise to overlapping errors which affect the estimators properties. While it is possible to cure the asymptotic properties of projection coefficients using well-known correction methods, these solutions do not address the inevitable challenge of the reduced number of genuinely independent observations. A regression of 5 year holding period returns on a 10 year sample has only two truly independent observations, even when the data is sampled daily. Finally, traditional predicting regressions with dependent variables constructed from future return realizations always raise the question of the extent to which in-sample results can be extended out-of-sample. At the same time, if in-sample regressions are plagued by look-ahead bias, out-of-sample regressions are typically exposed to the excess flexibility critique: the results are sensitive to the specific assumptions made to design the experiment.²⁸ For these reasons, as an alternative to future return realizations, we use as an ex-ante proxy for bond risk premia EBR_t^* , which is based on the forecasts of the agents with greatest past accuracy, as described in the previous section. In addition to being unaffected by forward looking biases, previous results confirm that, out-of-sample, EBR_t^* are competitive in forecasting future realized excess returns relative to some popular reduced form models. Thus, we evaluate alternative models of risk premia based on their ability to explain the dynamics of EBR_t^* , as opposed to sample averages (or projections) of future excess returns. Figure 9 displays our EBR_t^* measure for the 10-year bond, compared to the ex-post realized returns.

In this section we investigate regressions of the type

$$EBR_t^* = a_0 + b_i \mathcal{M}_{i,t} + \varepsilon_{i,t},$$

and consider three alternative specification for the dynamics of bond risk premia \mathcal{M}_t that have been widely studied in the literature.

²⁸Examples include the length of the training period, the start of the out-of-sample period, the use of fixed versus time-varying parameters, and the out-of-sample horizon.

A. Model Specifications of Bond Risk Premia: \mathcal{M}_t

In economies with external habit preferences, e.g. Campbell and Cochrane (1999), Wachter (2006), and Buraschi and Jiltsov (2007), time variation in risk compensation arises because of an endogenously time-varying price of risk. Shocks to the current endowment affect the wedge between consumption and habit, i.e. the consumption surplus, which induces time-varying expected returns. To obtain a proxy of risk premium \mathcal{M}_t , we follow Wachter (2006) and calculate consumption surplus (*Surp*) using a weighted average of 10 years of monthly consumption growth rates: $Surp = \sum_{j=1}^{120} \phi^j \Delta c_{t-j}$, where the weight is set to $\phi = 0.97^{1/3}$ to match the quarterly autocorrelation of the price-dividend ratio in the data.²⁹

In long-run risk economies with recursive preferences (see e.g. Bansal and Yaron (2004)), time variation in risk compensation arises from economic uncertainty (second moments) of the conditional growth rate of fundamentals. To obtain a proxy for economic uncertainty we adapt the procedure of Bansal and Shaliastovich (2013). First, we use our survey data on consensus expectation of *GDP* growth and inflation and fit a bivariate *VAR*(1). In a second step, we compute a GARCH(1,1) process on the VAR residuals to estimate the conditional variance of expected real growth $\sigma(g)$ and expected inflation $\sigma(\pi)$.

The literature on heterogeneous agents argues that in presence of heterogeneous beliefs, risk premia could be affected by the interaction of sentiment and disagreement, see e.g. Buraschi and Whelan (2010) and Jouini and Napp (2006). Buraschi and Whelan (2010) derive expressions for expected bond risk premia under the measure of an unbiased econometrician endowed with the knowledge of the true data generating process. They show that “bond risk premia are equal to the product of two terms. The first term is given by the price of risk, which is equal to the traditional risk premium emerging in an homogeneous Lucas economy plus a term equal to the wealth-weighted belief sentiment bias. The second term is the sensitivity of bond prices to consumption shocks, i.e. the quantity of risk.” Formally, in their economy if g_t is the true endowment growth rate and $E^i(g_t)$ is the subjective expectation of agent $i = a, b$, the

²⁹We obtain seasonally adjusted, real per-capita consumption of nondurables and services from the Bureau of Economic Analysis.

instantaneous bond risk premium EBR_t is given by

$$\begin{aligned}
 EBR_t &= \underbrace{\left[\gamma\sigma_c - \frac{1}{\sigma_c}\mathcal{S}_t \right]}_{\text{Price of Risk}} \times \underbrace{\sigma_B(DiB_t, \mathcal{S}_t)}_{\text{Quantity of Risk}} \\
 \mathcal{S}_t &= \textit{Sentiment Bias} = (\omega_t^a E^a(g_t) + \omega_t^b E^b(g_t)) - g_t \\
 DiB_t &= \textit{Disagreement} = (E^a(g_t) - E^b(g_t))/\sigma_g
 \end{aligned} \tag{7}$$

where σ_c and σ_B are the volatility of endowment and bond returns, respectively, and ω_t^i are the relative wealth weights of the two agents. When long-term bonds are risky assets, bond volatility σ_B is increasing in disagreement DiB_t .³⁰ Moreover, bond risk premia can switch sign depending on the level of sentiment \mathcal{S}_t and are larger when sentiment \mathcal{S}_t is negative. To study the relative merit of these models, first we construct an empirical proxy of sentiment \mathcal{S}_t and DiB_t . For \mathcal{S}_t , we use the spread between subjective and model-based forecasts, as defined in equation (4). To obtain a proxy of disagreement DiB_t , we compute the interquartile range of the distribution of short rate forecasts.

B. Empirical Results

We run a series of multivariate regressions of EBR_t^* on the state variables implied by the model specification above, based on a sample that ranges between December 1990 and July 2015.³¹ Table VIII summarizes regression results for bond maturities of 5 and 10 years.

Specification (i) of Table VIII investigates the statistical significance of consumption surplus. When agents have habit preferences, the price of risk is state-dependent and negatively related to the consumption surplus ratio. We find that the slope coefficient has the correct sign. However, both the t-statistic and R^2 are rather small: the t-statistics for the coefficient on *Surp* is -1.36 (-1.27) and the R^2 is 0.72% (1.00%) for the 5 (10) year bond maturity. Interestingly, however, after we control for heterogeneous beliefs (see below), the consumption surplus ratio becomes significant and able to explain a significant part of the total variance of EBR_t^* . Indeed, in specification (iv) the R^2 is 17.05% (23.61%) for the 5 (10) year bond, respectively.

³⁰Long term bonds are risky assets when their price dynamics is positively correlated with endowment shocks. This occurs when bond factor loading $\sigma_B(DiB_t)$ is positive.

³¹In this section we use an initial window of 2 years only in the construction of EBR_t^* in order to maximize the sample size in the regressions.

Specification (ii) of Table VIII focuses on the significance of proxies of economic growth and inflation uncertainty, $\sigma(g)$ and $\sigma(\pi)$, as suggested by long-run risk models. When using future realized average bond returns, these proxies are notoriously rather weak. On the other hand, we find that using EBR_t^* as dependent variable the statistical significance of inflation uncertainty is quite remarkable, with a t-statistic equal to 4.08 for the 10 year bond and an R^2 equal to 26.61%. Larger values of long-run economic uncertainty are correlated with greater subjective expected bond risk premia. This is consistent, for instance, with the model discussed in Bansal and Yaron (2004) in which greater uncertainty raises interest rates, lowers bond prices and increases future expected bond returns. However, for the 5-year bond only inflation uncertainty enters with a statistically significant loading.

Specification (iii) of Table VIII investigates the role played by the belief structure. Consistent with the spirit of heterogeneous beliefs models (see equation (7)), we run a two-stage regression. First, we compute realized bond return volatility $\sigma_{t,t+1}^{(n)}$ for different bond maturities n as the sum of squared yield changes (returns) within the month. We regress realized bond volatility $\sigma_{t,t+1}^{(n)}$ onto DiB_t and \mathcal{S}_t , to obtain $\hat{\sigma}^{(n)}(DiB_t, \mathcal{S}_t)$. Then, in the second stage, we follow the model-implied specification (7) and run the regression:

$$EBR_{n,t}^* = a_0 + (a + b\mathcal{S}_t)\hat{\sigma}^{(n)}(DiB_t, \mathcal{S}_t) + \epsilon_t^{(n)}. \quad (8)$$

The first stage regressions in Table VII show that the belief structure is significant in affecting the quantity of risk, as in (7). Indeed, both sentiment and disagreement increase bond volatility. The R^2 of the regression for 5 year bonds volatility is 15% and the t-statistics for DiB_t and \mathcal{S}_t are 4.27 and 4.06, respectively. For 10 year bonds, the R^2 is lower but sentiment continues to be highly statistically significant. In the second stage, we find that the coefficients a and b are both strongly statistically significant, with t-statistics equal to 4.71 (6.88) and -5.43 (-7.43) for the 5 (10) year bond, respectively. The R^2 of the regression for the 10 year bond is rather large and equal to 19.67%. Moreover, the signs of the coefficients are consistent with the theory. The coefficient a is positive, so that a marginal increase in the quantity of risk $\hat{\sigma}$ increases EBR_t^* . The coefficient b is negative, so that negative sentiment amplifies the previous effect. At the same time, a very positive level of sentiment can induce a switch in the sign of EBR_t^* , which can become negative.

Taken together, the first and second stage results suggest that sentiment affects, at the same time, both the quantity of risk $\hat{\sigma}$ and the price of risk.³² The overall effect is non-linear, since sentiment \mathcal{S}_t affects both the quantity and price of risk, and with opposite signs. Large bond risk premia are obtained for mildly negative sentiment \mathcal{S}_t and large disagreement DiB_t .

An important result that emerges from specifications (iv)-(v) is that the dynamics of beliefs clearly matters for bond risk premia, both for bonds with 5 and 10 years maturity. Indeed, controlling for alternative risk premia specifications never decreases the significance of the interaction of disagreement and sentiment found in specification (iii). The t-statistics of the interaction term $\mathcal{S}_t\hat{\sigma}$ are between -3.95 and -8.29 . The slope coefficient is negative, confirming the previous result that negative sentiment reinforces the marginal impact of the quantity of risk on bond risk premia.

To summarize, in most cases the empirical proxies implied by equilibrium models significantly explain the dynamics of subjective EBR^* . This was not obvious ex-ante to us and the results contrast with previous studies for equity returns which argue that equilibrium models generate implied risk premia that correlate negatively with survey-implied risk premia. For instance, Greenwood and Schleifer (2014) use equity market data and find a negative correlation between model-implied equity risk premia and survey expectations. They interpret their result as clear evidence of a rejection of rational expectations models: “*We can reject this hypothesis with considerable confidence. This evidence is inconsistent with the view that expectations of stock market returns reflect the beliefs or requirements of a representative investor in a rational expectations model.*” We find instead that in bond markets there is a significant positive correlation between most equilibrium models and empirical proxies of EBR^* , based on the most accurate agents. Both models of economic uncertainty and habit formation capture different characteristics of bond risk premia, once the properties of belief heterogeneity are carefully taken into account.

³²Disagreement matters on its own with a positive slope coefficient. The larger the disagreement the larger the expected bond risk premium EBR_t^* . When we run separate regressions of EBR_t^* on DiB_t , in periods of positive and negative sentiment, we confirm the previous result.

C. *The Quantity of Risk Channel Revisited.*

A well established result in the fixed income literature highlights the tenuous link between interest rate volatility and future average bond excess return (Duffee (2002)). This has motivated a significant discussion challenging the ability of completely affine term structure models to explain bond risk premia and calling for extensions of these specifications. To learn more about this link, Table VIII (specification (vi)) shows the results of a regression of EBR^* onto the realized quantity of risk $\sigma_{t-1,t}^{(n)}$:

$$EBR_t^* = a_0 + b\sigma_{t-1,t}^{(n)} + \epsilon_t. \quad (9)$$

The regression results show that the quantity of risk channel is significant. We obtain a R^2 of 1.36% and 8.53% for the 5 and 10-year bonds, respectively, and the t-statistics of b are 1.83 and 5.03. This is very interesting and rather surprising. Indeed, contrary to the previous literature we find that the quantity of risk is statistically correlated with our measure of risk premia EBR_t^* . This highlights the importance of distinguishing between expected bond risk premia and average future return realizations. Moreover, it suggests that models in which the quantity of risk plays an important role for bond risk premia should not be dismissed. Moreover, Table VII shows that the quantity of risk is affected by the belief structure.

D. *Behavioral Components in Agents Expectations and the Formation of Agents Beliefs*

Forecast errors provide an interesting source of information to learn about the process of formation of agents beliefs. For instance, what is the extent to which agents forecasts errors are predictable ex-ante? Can they be easily corrected using public information available at the time these predictions are made? This is an important question that is related to the debate about the significance of behavioral biases in agents expectations and the empirical plausibility of the assumption of rational expectation in asset pricing models.

In order to investigate these questions, we run a series of regressions in which we project future forecasts errors $FE_{t,t+1}^n$ onto information available at time t . We define FE^n as the difference between future realized and expected excess returns, $FE_{t,t+1}^n = rx_{t,t+1}^n - EBR_{n,t}^*$.

We start by investigating the extent to which one could use information contained in the term structure of bond yields at the time agents form their predictions. This is motivated

by the results in Cochrane and Piazzesi (2005) who argue about the existence of a significant predicting factor of bond returns that is spanned by the cross-section of bond yields. Since bond yields are strongly correlated, we use a simple principal component (*PC*) decomposition to orthogonalize bond yields over the whole sample. Table IX (Panel A) summarizes the results and shows that a combination of five principal components can explain between 23% (2-year bonds) and 28% (10-year bonds) of the variance of the forecast errors. The fourth principal component is strongly significant across all bond maturities. This may suggest that forecast errors are predictable and could easily be corrected using information from the term structure. In the case of 10-year bonds, the two most significant sources of FE_t^{10} predictability are the second and fourth principal components. In separate regressions, we find that the second principal component (slope of the yield curve) is highly correlated with sentiment \mathcal{S}_t . This suggests that \mathcal{S}_t is indeed correlated with the expectation bias and if agents were to be aware of the magnitude of \mathcal{S}_t they could improve their expectations. The fourth principal component is correlated with a combination of both *DiB* and \mathcal{S}_t , while the level of *DiB* is correlated with the first principal component (level of the term structure).³³ To check the robustness of the previous result, we run a regression of forecast errors on the term structure of forward rates f_t available at time t with maturities between one and five years:

$$FE_{t,t+1}^n = \gamma^{n'} [1 \quad f_t]' + \varepsilon_{t+1}^n$$

Indeed, while the previous principal components are estimated using using full-sample information, the term structure of forward rates is available to the agents without ex-post information. The results are reported in Panel B of Table IX and show that, although the R^2 is generally lower than in the previous regressions, the five forward rates have some predictive power and can explain between 12% (5-year bond) and 24% (10-year bond) of the total variance of the forecast errors FE_t^n . We formally test the existence of a single factor ξ_t that can be identified from the forward curve and that could be used to improve agents expectations. We follow the approach of Cochrane and Piazzesi (2005) and estimate the combination of five forward rates $\xi_t = \hat{\gamma}' f_t$ as the fitted value of a regression of the average forecast errors across all bond maturities n on the cross-section of forward rates. This is equivalent to impose the restriction that

³³The correlation between the five principal components and $S(t)$ and *DiB* are $Corr(PCs, \mathcal{S}) = [-0.06 \quad 0.54 \quad -0.32 \quad -0.09 \quad -0.04]'$, while $Corr(PCs, DiB) = [0.50 \quad 0.26 \quad -0.05 \quad 0.12 \quad -0.05]'$.

the vector γ^n differs across maturities n only up to a scale parameter β_n , namely $\gamma^{n'} = \beta_n \gamma'$. Table X (Panel A) shows that while ξ_t is not significant in explaining the forecast errors on short term bonds, it is significant for the 5 and 10-year bonds. The R^2 ranges between 11% and 23%, and the t-statistics of the slope coefficient of FE^n onto ξ_t ranges between 3.08 and 4.93. This suggests that agents would be able, if they so wished, to correct their long term bonds forecasts using a simple combination of publicly available forward rates. To examine the source of this predictability, we study whether this single forecasting factor helps predicting FE^n because it helps explaining EBR^* or because it helps predicting future realized returns, namely the part of returns that is seemingly unpredictable by the agents. Table X (Panel B) shows that the single factor ξ_t is mildly correlated with contemporaneous EBR_t^* but it is strongly significant in predicting future realized returns. This seems to confirm that this factor is indeed unspanned by agents beliefs and they are not using important public information to improve their forecasts.

Statistical significance does not necessarily imply economic significance. To study the economic significance of behavioral components in agents expectations, we design an experiment in which we construct fictitious bond return expectations by correcting EBR^* using the single factor ξ_t which seems to explain agents beliefs bias. The corrected EBR_c^* is defined as $EBR^* + \hat{\beta}_{t,n} \hat{\gamma}'_t [1 \ f_t]'$. The subscript t in the parameters vector $\hat{\beta}'_{t,n}$ and $\hat{\gamma}'_t$ indicates that the correction is restricted to use only real-time information which is available at time t , so that the $\hat{\beta}_t$ is estimated using a rolling window to replicate real-life conditions of a trader. We compute and compare the $RMSE$ implied by both the original EBR^* and the corrected EBR_c^* . We find that, although the initial regressions indicate the existence of predictability in the forecast errors, the $RMSE$ of the corrected forecasts are not unambiguously lower than the uncorrected ones. For instance, using a window of 60 months in the estimation of the correction parameters, while the $RMSE$ for the 5-year bond decreases by 10.53%, the $RMSE$ of the 10-year bond increases by 15.9%.³⁴ This shows that the expectations embedded in EBR^* cannot be easily improved using market based information. The uncorrected version of EBR^* seems to dominate its corrected counterpart, mainly in terms of variability, suggesting that the apparent bias in survey-based expectations is not easy to correct using the information available in real-time.

³⁴Using a rolling window of 8 instead of 5 years, the $RMSE$ of the EBR^* corrected for the predictable component of forecast errors computed in real-time increase by 32%, 6% and 2%, for the 2, 5 and 10-year bond, respectively.

VI. Conclusion

This paper studies the expectations of bond returns taken directly from survey data and compares them to traditional measures of bond risk premia measured from ex-post realizations. Our analysis reveals a number of interesting results.

First, individual risk premia are largely heterogeneous and the consensus does not subsume the information contained in the distribution of forecasts. We find a significant amount of persistence in agents beliefs on bond excess returns and in the degree of optimism/pessimism relative to consensus. However, overall expectations about bond returns are remarkably precise with respect to statistical models. Moreover, individual expectations of bond returns are internally consistent with agents' forecasts about GDP and inflation.

Second, we find evidence of predictability in short-term interest rates and the accuracy of the best forecasters is persistent over time. In particular, we find that primary dealers are more likely to be between the top forecasters of the short-term interest rate. This is consistent either with primary dealers superior information about Fed's implementation of monetary policy or, more simply, with an information flow advantage originating from their role as market maker in Treasury bonds. We then show that agents who are more accurate in forecasting short term rates do not have a persistent edge in predicting long term bond returns. This finding supports the idea that time variation in bond risk premia plays an important role in the predictability of long-term bond returns. Indeed, we directly study expectations of long-term bond excess returns and strongly reject the hypothesis that bond risk premia are constant.

These findings suggests that surveys can be used to build reliable measures of bond risk premia as an alternative to empirical projections of future return realizations on lagged state variables. Survey expectations are available in real time and can sidestep the issue of in-sample versus out-of-sample model fitting. We take into account the persistent heterogeneity in the cross-section of beliefs and build an aggregate measure of subjective bond risk premia based on the agents past accuracy, consistent with the market selection hypothesis that accurate agents accumulate weight in the pricing kernel. We then use our measure, that we denote as EBR^* , to evaluate and compare alternative macro-finance models. As we substitute EBR^* as the dependent variable, we find surprisingly supportive evidence for several macro-finance models of bond risk premia. In particular, models that argue about the importance of the quantity of

risk as a source of time variation in bond risk premia perform significantly better in explaining EBR^* than future realized average returns. We also show that, consistent with heterogenous beliefs models, two characteristics of the cross-section of beliefs, namely disagreement and sentiment, significantly affect the risk premium dynamics in a non-linear way. This highlights the importance of the specification of agents expectations.

Finally, we use our measure of expected bond returns EBR_t^* to compute forecast errors and analyze their predictability by information included in bond prices at time t . We find statistical evidence of predictability in forecast errors by a combination of forward rates, in violation of rational expectations. However, when we try to correct EBR^* by the predictable component of forecast errors we find that the $RMSE$ of the corrected forecasts are not unambiguously lower than the uncorrected ones, suggesting that the expectations embedded in EBR^* cannot be easily improved using market based information in real-time.

References

- Aioli, M., C. Capistran, and A. Timmermann, 2011, *The Oxford handbook of economic forecasting* (Oxford: Oxford University Press).
- Alchian, A., 1950, Uncertainty, evolution and economic theory, *Journal of Political Economy* 58, 211–221.
- Ang, A., G. Bekaert, and M. Wei, 2007, Do macro variables, asset markets, or surveys forecast inflation better?, *Journal of Monetary Economics* 54, 1163–1212.
- Bansal, R., and I. Shaliastovich, 2013, A long-run risks explanation of predictability puzzles in bond and currency markets, *Review of Financial Studies* 26, 1–33.
- Bansal, R., and A. Yaron, 2004, Risks for the long run: A potential resolution of asset pricing puzzles, *The Journal of Finance* 59, 1481–1509.
- Basak, S., 2005, Asset pricing with heterogeneous beliefs, *Journal of Banking and Finance*, 29, 2849–2881.
- Bauer, Michael D, and James D Hamilton, 2015, Robust bond risk premia, *Available at SSRN 2666320*.
- Buraschi, A., and A. Jiltsov, 2006, Model uncertainty and option markets with heterogeneous beliefs, *The Journal of Finance* 61, 2841–2897.
- , 2007, Habit formation and macroeconomic models of the term structure of interest rates, *Journal of Finance* 62, 3009 – 3063.
- Buraschi, A., F. Trojani, and A. Vedolin, 2014, Economic Uncertainty, Disagreement, and Credit Markets, *Management Science* 60, 1281–1296.
- Buraschi, Andrea, and Paul Whelan, 2010, Term structure models and differences in beliefs, *Working paper*.
- Campbell, J.Y., and J.H. Cochrane, 1999, By force of habit: A consumption-based explanation of aggregate stock market behavior, *Journal of political Economy* 107, 205–251.

- Chen, H., S. Joslin, and N. Tran, 2012, Rare disasters and risk sharing with heterogeneous beliefs, *Review of Financial Studies* 25, 2189–2224.
- Cieslak, A., 2016, Short-rate expectations and unexpected returns in treasury bonds, *Working paper, Duke University, Fuqua School of Business*.
- , and P. Povala, 2012, Expecting the Fed, *Working paper, Kellogg School of Management*.
- Cochrane, J.H., and M. Piazzesi, 2005, Bond risk premia, *American Economic Review* 95, 138–160.
- D’Amico, S., and A Orphanides, 2008, Uncertainty and Disagreement in Economic Forecasting, Discussion paper, Divisions of Research & Statistics and Monetary Affairs, Federal Reserve Board.
- Duffee, G.R., 2002, Term premia and interest rate forecasts in affine models, *The Journal of Finance* 57, 405–443.
- Ehling, Paul, Michael Gallmeyer, Christian Heyerdahl-Larsen, and Philipp Illeditsch, 2015, Disagreement about inflation and the yield curve, *Working paper, University of Pennsylvania*.
- Friedman, M., 1953, *Essays in Positive Economics* (University of Chicago Press, Chicago).
- Giacoletti, Marco, Kristoffer Laursen, and Kenneth Singleton, 2016, Learning, dispersion of beliefs, and risk premiums in an arbitrage-free term structure model, *Working paper, Stanford University*.
- Giordani, P., and P Soderlind, 2003, Inflation forecast uncertainty, *European Economic Review* 47, 1037–1059.
- Goyal, A., and I. Welch, 2008, A comprehensive look at the empirical performance of equity premium prediction, *Review of Financial Studies* 21, 1455–1508.
- Greenwood, R., and A. Schleifer, 2014, Expectations of returns and expected returns, *Review of Financial Studies* 27, 714–746.

- Gürkaynak, Sack, and Wright, 2006, The u.s. treasury yield curve: 1961 to the present, *Federal Reserve Board Working Paper Series*.
- Hong, H., D. Sraer, and J. Yu, 2013, Reaching for maturity, Discussion paper, Princeton University.
- Jouini, E., and C. Napp, 2006, Heterogeneous beliefs and asset pricing in discrete time: An analysis of pessimism and doubt, *Journal of Economic Dynamics and Control* 30, 1233–1260.
- Koijen, R., M. Schmeling, and E. Vrugt, 2015, Survey expectations of returns and asset pricing puzzles, Discussion paper, London Business School, Cass Business School and VU University Amsterdam.
- Malmendier, U., and S. Nagel, 2011, Depression Babies: Do Macroeconomic Experiences Affect Risk-Taking?, *Quarterly Journal of Economics* 126, 373–416.
- Marcellino, M., 2008, A linear benchmark for forecasting gdp growth and inflation, *Journal of Forecasting* 27, 305–340.
- Patton, A. J., and A. Timmermann, 2010, Why do Forecasters Disagree? Lessons from the Term Structure of Cross-Sectional Dispersion, *Journal of Monetary Economics* 57, 803–820.
- Piazzesi, M., J. Salomao, and M. Schneider, 2015, Trend and cycle in bond premia, Discussion paper, Stanford, NBER and University of Minnesota.
- Scheinkman, J.A., and W. Xiong, 2003, Overconfidence and speculative bubbles, *Journal of Political Economy* 111, 1183–1220.
- Svensson, Lars, 1994, Estimating and interpreting forward interest rates: Sweden 1992-1994, Discussion paper, National bureau of economic research.
- Wachter, J.A., 2006, A consumption-based model of the term structure of interest rates, *Journal of Financial Economics* 79, 365–399.
- Xiong, W., and H. Yan, 2010, Heterogeneous expectations and bond markets, *Review of Financial Studies* 23, 1433–1466.

VII. Tables

Panel A: Transition Probabilities

	3-month rate				2-year bond				10-year bond			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Q1	72%	21%	5%	1%	75%	19%	4%	2%	74%	18%	5%	2%
Q2	22%	51%	23%	4%	20%	51%	23%	5%	21%	52%	22%	5%
Q3	5%	21%	54%	19%	4%	23%	52%	20%	5%	23%	52%	20%
Q4	2%	5%	22%	71%	1%	5%	22%	71%	1%	5%	22%	71%

Panel B: Conditional Probabilities Short Rates vs Macro

	GDP				CPI			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Q1	35%	26%	22%	16%	41%	26%	21%	11%
Q2	26%	27%	28%	19%	26%	30%	28%	16%
Q3	22%	26%	29%	23%	20%	23%	31%	27%
Q4	20%	22%	25%	33%	15%	20%	26%	39%

Table I. Transition and Conditional Probabilities

Panel A presents the probability of a forecaster transitioning from a given quartile of the cross-sectional distribution of 3-month yield (left), 2-year bond excess returns (middle) and 10-year bond excess returns (right) forecasts to another quartile in the following month. Panel B presents the probability of a forecaster being in a given quartile of the cross-sectional distribution of Macro forecasts given that the forecaster is in a particular quartile of the cross-sectional distribution of 3 month yield forecasts.

	Cons	Top	Bottom	UR	$VAR(1)$	$VAR_D(p)$	FR	Taylor
All	1.41	1.28	1.69	1.34	1.48	1.62	1.71	1.51
Expansions	1.34	1.24	1.58	1.32	1.43	1.54	1.59	1.46
Recessions	1.89	1.57	2.41	1.54	1.79	2.13	2.44	1.88

Table II. Root Mean Square Errors

This table presents root mean square prediction errors for the 3-month interest rate in 12-months ahead. The first three columns correspond to consensus, top decile, and bottom decile forecaster mean averages from BCFF. The remaining columns correspond to real-time statistical models. UR is a unit root forecast. $VAR(1)$ and $VAR_D(p)$ are vector autoregressions estimated on 10-years of rolling monthly data 3-month rates, the slope, and curvature of the term structure. The lag length in $VAR_D(p)$ is chosen dynamically using the Akaike information criterion. FR is 12 \rightarrow 15 forward rate, i.e., the risk neutral 3-month rate expected in 12-months time. Taylor is a vector autoregressions estimated on 10-years of rolling monthly data 3-month rates, the annual inflation rate of the All-Urban consumer price index excluding food and energy, and the industrial production output gap estimated using a recursive Hodrick-Prescott Filter. The rows ‘All’ reports RMSEs for all months in our sample period which is January 1988 to July 2015 (331 observations). Recessions reports RMSEs (37 observations) coming from forecasts made during NBER recessions and Expansions reports RMSEs for all other dates (294 observations).

	3M				10Y				3M vs 10Y			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Q1	64%	25%	7%	2%	58%	27%	11%	4%	34%	26%	22%	17%
Q2	22%	48%	24%	5%	25%	44%	24%	7%	25%	32%	26%	17%
Q3	8%	21%	48%	22%	9%	22%	47%	21%	19%	25%	32%	24%
Q4	3%	5%	19%	72%	5%	7%	19%	70%	17%	16%	21%	47%

Table III. Accuracy Transition Probabilities

The left panel presents the probability of a forecaster transitioning from a given quartile of the cross-sectional distribution of forecasts’ accuracy to another quartile in the following month, for 3-month interest rates. The middle panel presents the same transition probabilities of 10-year bond excess return accuracy. The right panel presents the joint distribution of forecast accuracy for the 10-year EBR and 3-month yield, that is the probability of being in a given quartile of the 3-month yield accuracy percentile distribution and a given quartile of the 10-year EBR accuracy percentile distribution, at the same time.

Top Short Rate Forecasters	
(1)	Goldman Sachs
(2)	J.P. Morgan
(3)	BMO Capital Markets
(4)	Nomura Securities Inc.
(5)	Bank of America
(6)	Georgia State University
(7)	Crestar Financial Corp.
(8)	US Trust Company
(9)	Chase Manhattan Bank
(10)	Woodworth Holdings

Table IV. Top 10 Short Rate Forecasters

This table presents the top 10 short rate forecasters in terms of average accuracy percentile ranking over the full sample, for the 3-month yield. We consider only forecasters who contribute to the panel for at least 10 years.

Maturity	Q1	Q2	Q3	Q4
2-year	0.33 (3.22)	0.41 (4.17)	0.48 (4.84)	0.50 (4.51)
5-year	0.29 (2.81)	0.35 (3.42)	0.42 (4.29)	0.40 (3.25)
10-year	0.26 (2.72)	0.33 (3.41)	0.41 (4.22)	0.43 (3.79)

Table V. Autoregressive Regression

Slope coefficients of the regressions of the quartiles (Q1 to Q4) of the cross-sectional distribution of subjective excess returns of 2, 5, and 10-year zero-coupon bonds on a constant and their own lag at the 1-year horizon. t-statistics, reported in parentheses below the point estimates, are Newey-West corrected.

	Unconditional Best	Most Represented in EBR*
(1)	Thredgold Economic Assoc.	Fannie Mae
(2)	UBS	UBS
(3)	Goldman Sachs	Thredgold Economic Assoc.
(4)	Fleet Financial Group	The Northern Trust Company
(5)	Fannie Mae	Goldman Sachs
(6)	DePrince & Assoc	Loomis Sayles & Co.
(7)	J.W. Coons & Associates	Banc of Amercia Securities
(8)	Nomura Securities Inc.	BMO Capital Markets
(9)	The Northern Trust Company	DePrince & Assoc
(10)	ClearView Economics	J.W. Coons & Associates

Table VI. Top 10 Bond Forecasters

The left panel of this table presents the top 10 forecasters in terms of average accuracy percentile ranking for the 10-year bond excess returns, over the full sample. The right panel shows the 10 agents who are more often (in terms of number of months) present in our *EBR** index, which conditionally selects the top quartile of forecasters based on past accuracy percentiles on the 10-year bond excess returns.

	const	<i>DiB</i>	\mathcal{S}	$\overline{R}^2(\%)$
$\sigma_{t,t+1}^{(2)}$	0.02	4.79	1.25	27
	(5.96)	(8.92)	(3.82)	
$\sigma_{t,t+1}^{(5)}$	0.04	2.27	1.25	15
	(11.89)	(4.27)	(4.06)	
$\sigma_{t,t+1}^{(10)}$	0.05	0.56	0.84	5
	(15.79)	(1.11)	(3.04)	

Table VII. Volatility Regressions

This table reports estimates from a forecasting regression of monthly realised bond volatility, for 2, 5, and 10 year bonds, onto disagreement and sentiment factors:

$$\sigma_{t,t+1}^{(n)} = a^{(n)} + b^{(n)} DiB_t + c^{(n)} \mathcal{S}_t + \epsilon_{t,t+1}^{(n)}$$

where intra-month volatility is computed as sum of squared yield changes (returns) within the month. Disagreement and sentiment factors are discussed in detail in the main body of the paper. We denote the fitted value from this regression as $\hat{\sigma}^{(n)}(DiB, \mathcal{S})$. t-statistics, reported in parentheses below the point estimates, are Newey-West corrected. Adjusted R-squared of the regressions are reported in the last column. The sample period is from December 1990 to July 2015.

	$\hat{\sigma}^{(n)}(DiB, \mathcal{S})$	$\mathcal{S} \times \hat{\sigma}^{(n)}(DiB, \mathcal{S})$	<i>Surp</i>	$\sigma(g)$	$\sigma(\pi)$	$\sigma_{t-1,t}^{(n)}$	\overline{R}^2 (%)
5-year bond maturity							
(i)			-0.08 (-1.36)				0.72
(ii)				0.08 (1.04)	0.42 (2.51)		20.95
(iii)	0.54 (4.71)	-0.72 (-5.43)					13.89
(iv)	0.60 (5.22)	-0.81 (-6.27)	-0.19 (-3.07)				17.05
(v)	0.35 (2.06)	-0.68 (-3.95)		-0.17 (-1.84)	0.56 (3.94)		34.18
(vi)						0.12 (1.83)	1.36
10-year bond maturity							
(i)			-0.10 (-1.27)				1.00
(ii)				0.16 (2.11)	0.41 (4.08)		26.61
(iii)	1.46 (6.88)	-1.67 (-7.43)					19.67
(iv)	1.58 (7.28)	-1.83 (-8.29)	-0.21 (-2.94)				23.61
(v)	0.93 (3.94)	-1.26 (-5.23)		-0.11 (-1.24)	0.54 (6.74)		39.96
(vi)						0.29 (5.03)	8.53

Table VIII. Determinants of Ex-Ante Subjective Bond Returns

This table reports estimates from regressions of the subjective expected excess returns on 5-year bonds (Panel A) 10-year bonds (Panel B) and for good forecasters on a set of explanatory variables:

$$EBR_{n,t}^* = a_0 + b_i^n \mathcal{M}_t^j + \epsilon_{i,t}^n.$$

These factors are discussed in detail in the main body of the paper, and all variables are standardized. t-statistics, reported in parentheses below the point estimates, are Newey-West corrected. Adjusted R-squared of the regressions are reported in the last column. The sample period is from December 1990 to July 2015.

Panel A: Forecast errors on PCs							
Maturity	const	$PC(1)$	$PC(2)$	$PC(3)$	$PC(4)$	$PC(5)$	$\overline{R}^2(\%)$
2-year	0.01	0.00	0.00	0.02	0.19	-0.21	23
	(2.34)	(1.60)	(0.06)	(1.20)	(2.28)	(-1.01)	
5-year	0.04	0.00	0.02	-0.00	0.53	-1.55	24
	(3.39)	(1.22)	(1.29)	(-0.03)	(2.59)	(-2.26)	
10-year	0.04	-0.00	0.06	-0.12	0.90	-2.43	28
	(3.26)	(-0.30)	(2.90)	(-1.06)	(3.46)	(-1.58)	

Panel B: Forecast errors on forwards							
Maturity	const	$f(1)$	$f(2)$	$f(3)$	$f(4)$	$f(5)$	$\overline{R}^2(\%)$
2-year	-0.01	1.43	-2.16	-1.52	4.60	-1.90	18
	(-1.63)	(1.80)	(-0.54)	(-0.14)	(0.35)	(-0.35)	
5-year	-0.01	0.08	8.27	-36.01	48.24	-19.58	12
	(-0.80)	(0.03)	(0.65)	(-1.06)	(1.25)	(-1.25)	
10-year	0.01	-7.43	44.74	-137.06	171.94	-71.80	24
	(0.49)	(-2.02)	(2.56)	(-3.08)	(3.37)	(-3.33)	

Table IX. Forecast errors on PCs and forwards

This table reports estimates from regressions of forecast errors, i.e. realized excess returns minus subjective expected excess returns on the first five principal components of the term structure of interest rates (Panel A) and on the cross-section of forward rates with maturities between one and five years (Panel B). t-statistics, reported in parentheses below the point estimates, are Newey-West corrected. Adjusted R-squared of the regressions are reported in the last column. The sample period is from December 1998 to July 2015.

Maturity	const	F	$\bar{R}^2(\%)$	const	F	$\bar{R}^2(\%)$	const	F	$\bar{R}^2(\%)$
	Forecast Errors			Realized Returns			Expected Returns		
2-year	0.00 (0.72)	0.14 (1.53)	3	0.00 (0.38)	0.21 (2.39)	8	-0.00 (-1.05)	0.07 (2.23)	7
5-year	0.01 (0.86)	0.79 (3.08)	11	-0.00 (-0.11)	1.08 (4.20)	21	-0.01 (-2.15)	0.29 (2.54)	11
10-year	-0.01 (-0.67)	2.07 (4.93)	23	-0.02 (-1.09)	2.47 (5.99)	33	-0.01 (-0.74)	0.40 (1.55)	4

Table X. Forecast errors, realized and expected returns on single factor

This table reports estimates from regressions of forecast errors (left panel), realized (middle panel) and subjective expected returns (right panel) on a single factor of forward rates, generated by regressing the average forecast error across maturities on the cross-section of forward rates with maturities between one and five years. t-statistics, reported in parentheses below the point estimates, are Newey-West corrected. Adjusted R-squared of the regressions are reported in the last column. The sample period is from December 1998 to July 2015.

VIII. Figures

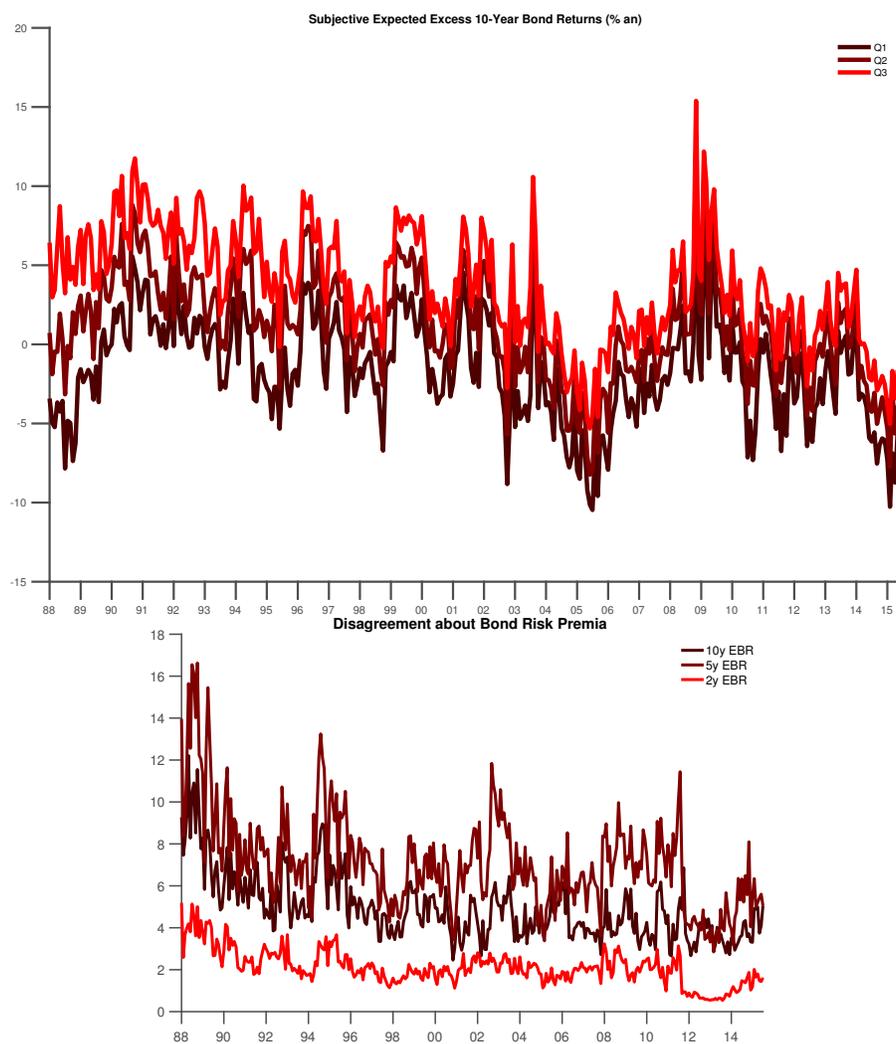


Figure 1. Subjective Expectations

The top panel plots quartiles (Q1, Q2(median) and Q3) of the cross-sectional distribution of 1-year subjective expected excess returns for 10-year maturity bonds. The bottom panel plots disagreement about expected bond returns for maturities 2 , 5 and 10-year, defined as the cross-sectional interquartile range of subjective expectations standardized by the full-sample consensus expectation.

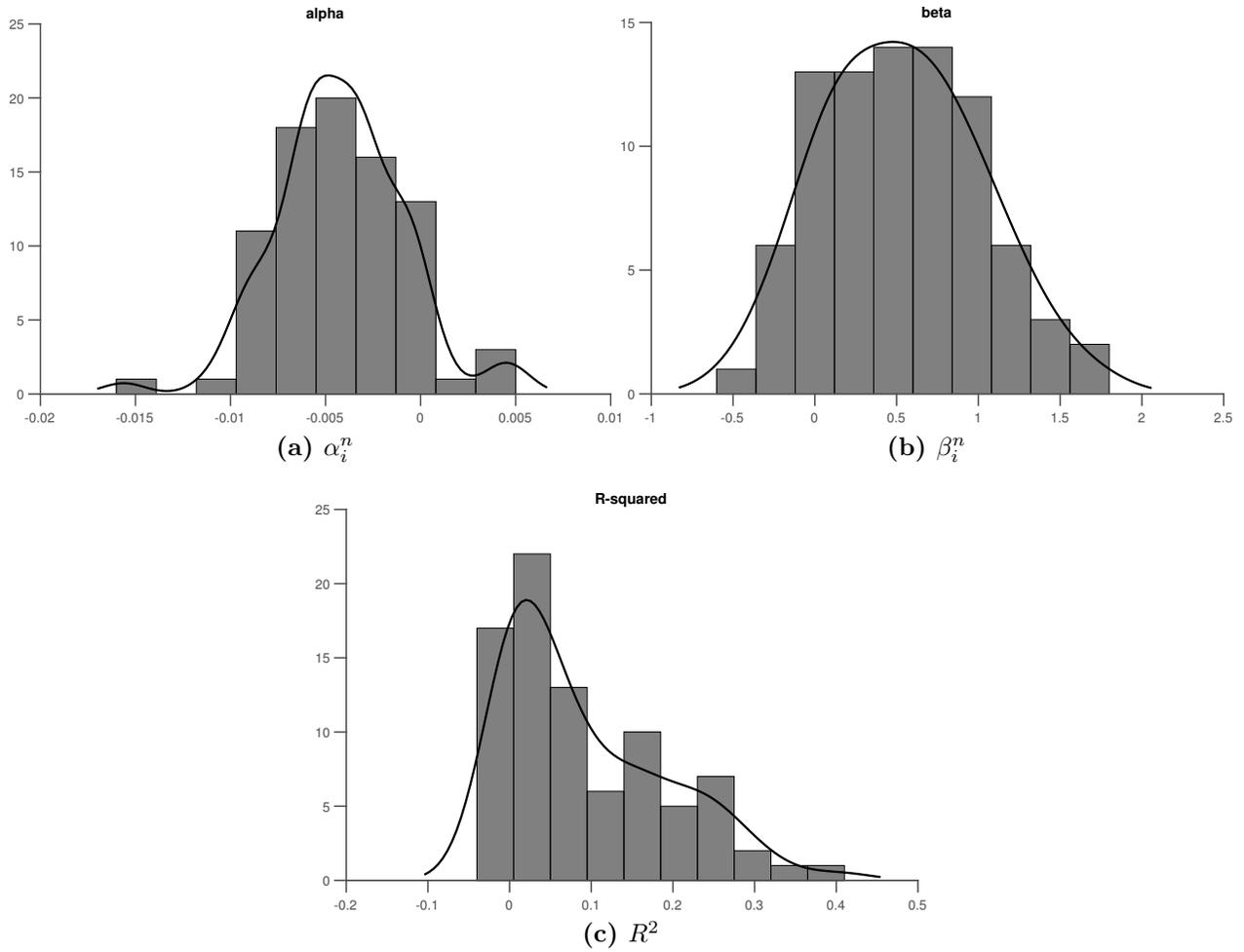


Figure 2. Short Rate Predictive Regressions: Individual Forecasters

Estimated regression coefficients and adjusted R^2 of regressions of the change in realized 3-month yield on the expected change in 3-month yield for all individual contributors with at least 60 months of forecasts. Solid lines denote kernel density estimates of the cross-sectional distributions.

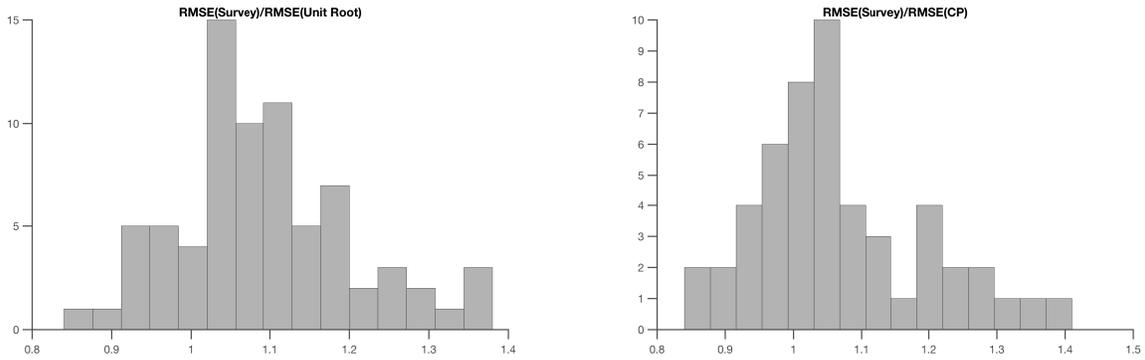


Figure 3. Relative Accuracy

Histogram of the relative accuracy \mathcal{A}_i of each forecaster, that is the ratio between the RMSE of each individual forecaster and the RMSE of a benchmark, that is the unit root model for short rate forecasts (left panel) and a real time bond risk premium forecast implied by the CP factor for the 10-year excess bond returns (right panel), for the period in which the forecaster is in the panel. We consider only the contributors with at least 60 months of forecasts, for a total of 84 institutions.

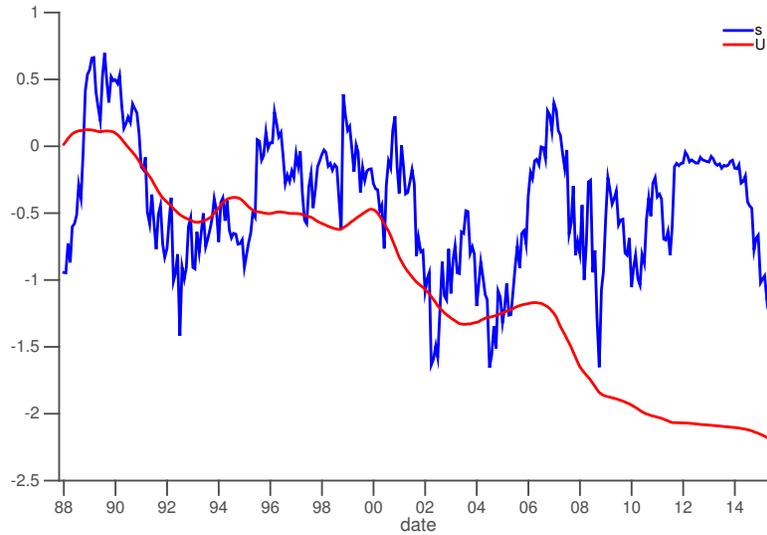


Figure 4. Cumulative Errors and Sentiment

Cumulative short rate forecast errors from the average forecaster (the consensus), $U(t)$, and the difference between model-implied and survey-based short rate expectation, $s(t)$.

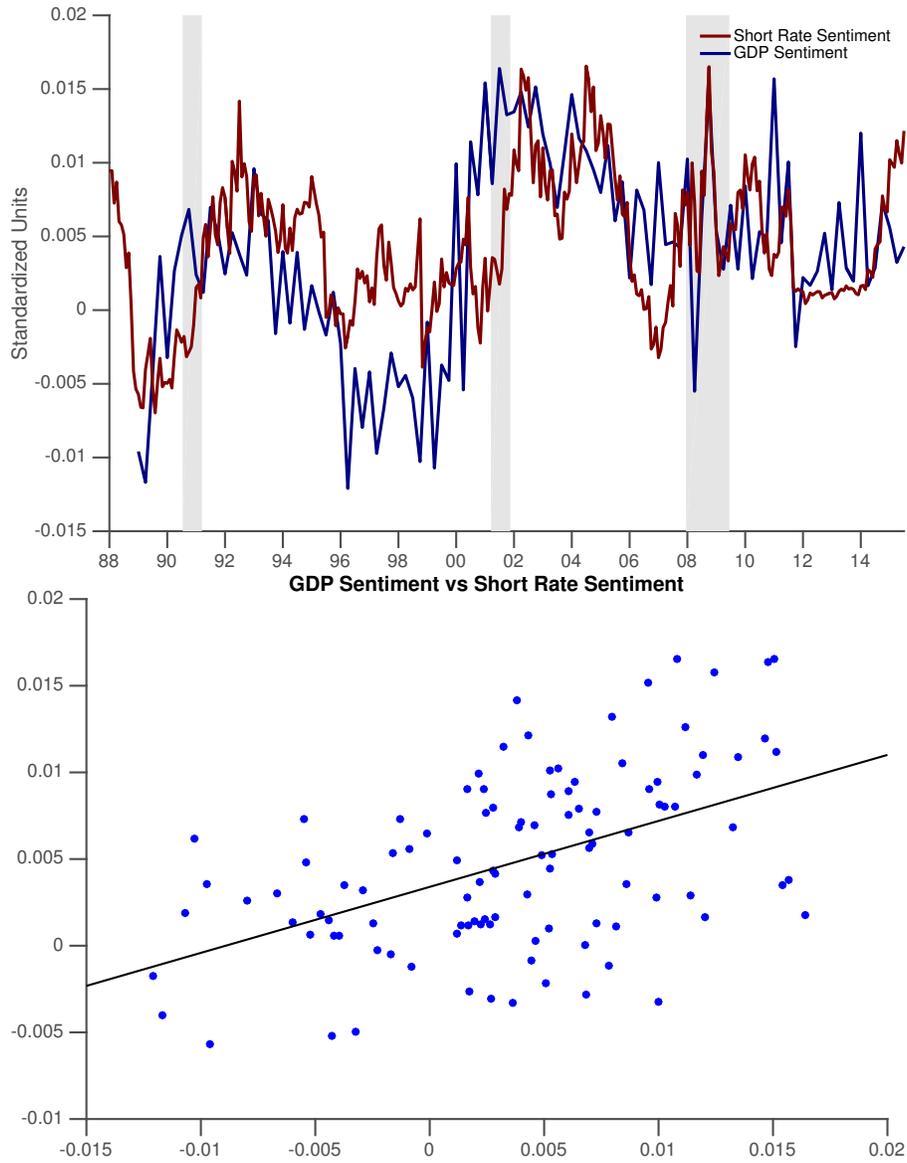


Figure 5. Sentiment Measures

The red line denotes the *Sentiment* measure, computed as the difference between the simple average of expected short-rate from surveys and the expected short-rate implied by a unit-root forecast at 1-year horizon. The blue line is an equivalent measure of sentiment on GDP growth expectations, where the physical expectation is computed from an AR(4) projection of quarterly realized GDP growth. Bottom is a scatter plot corresponding to the time series in the top plot.

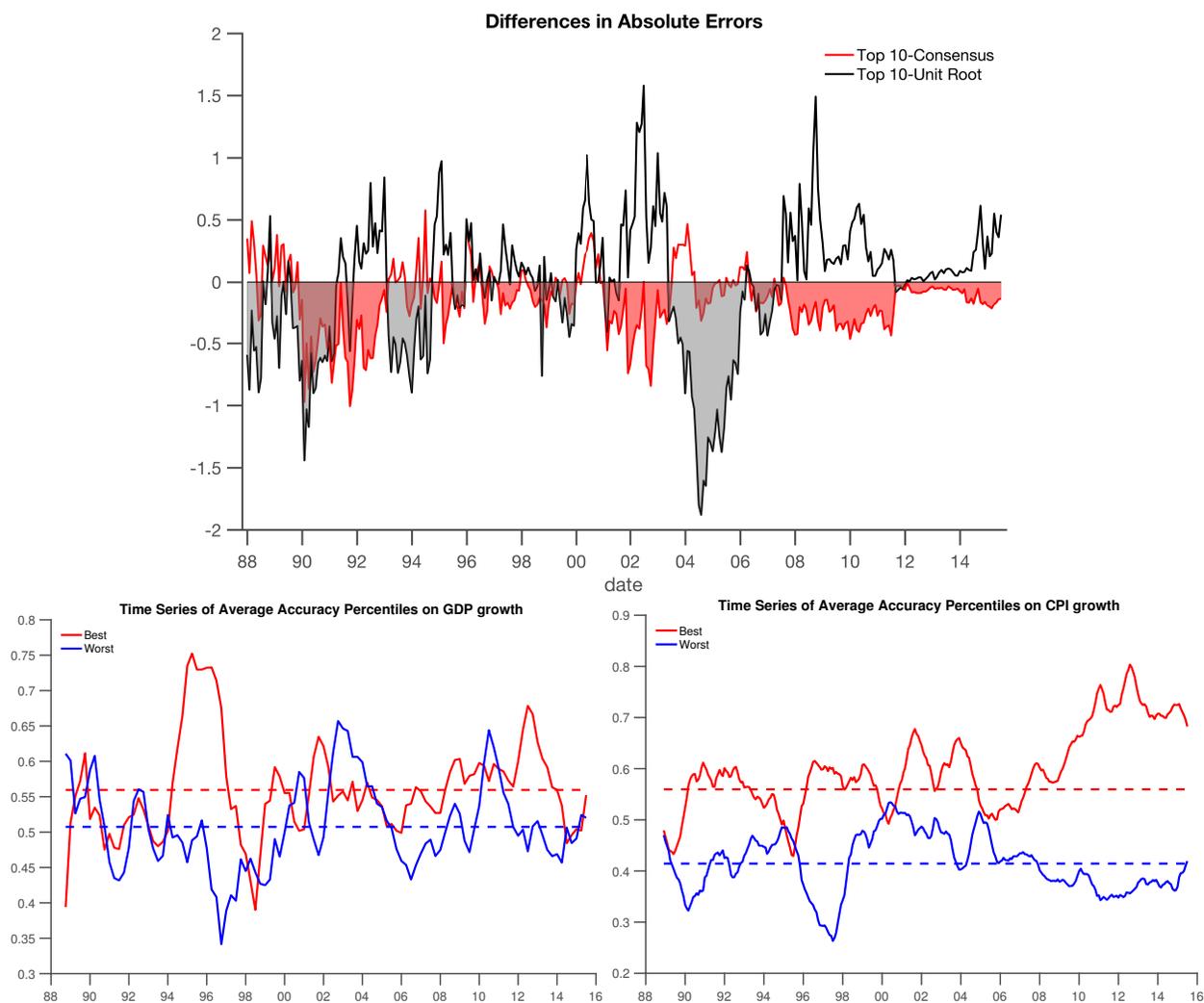


Figure 6. Absolute Short Rate Error Differences and Macro Accuracy Percentiles
 The top panel plots differences in absolute forecast errors between the the top 10 forecasters and the consensus, and the top 10 forecasters and a unit root forecast. The bottom panels show the time series of average accuracy percentiles on the Real GDP growth (left) and CPI growth (right), for the top 10 and bottom 10 short rate forecasters.

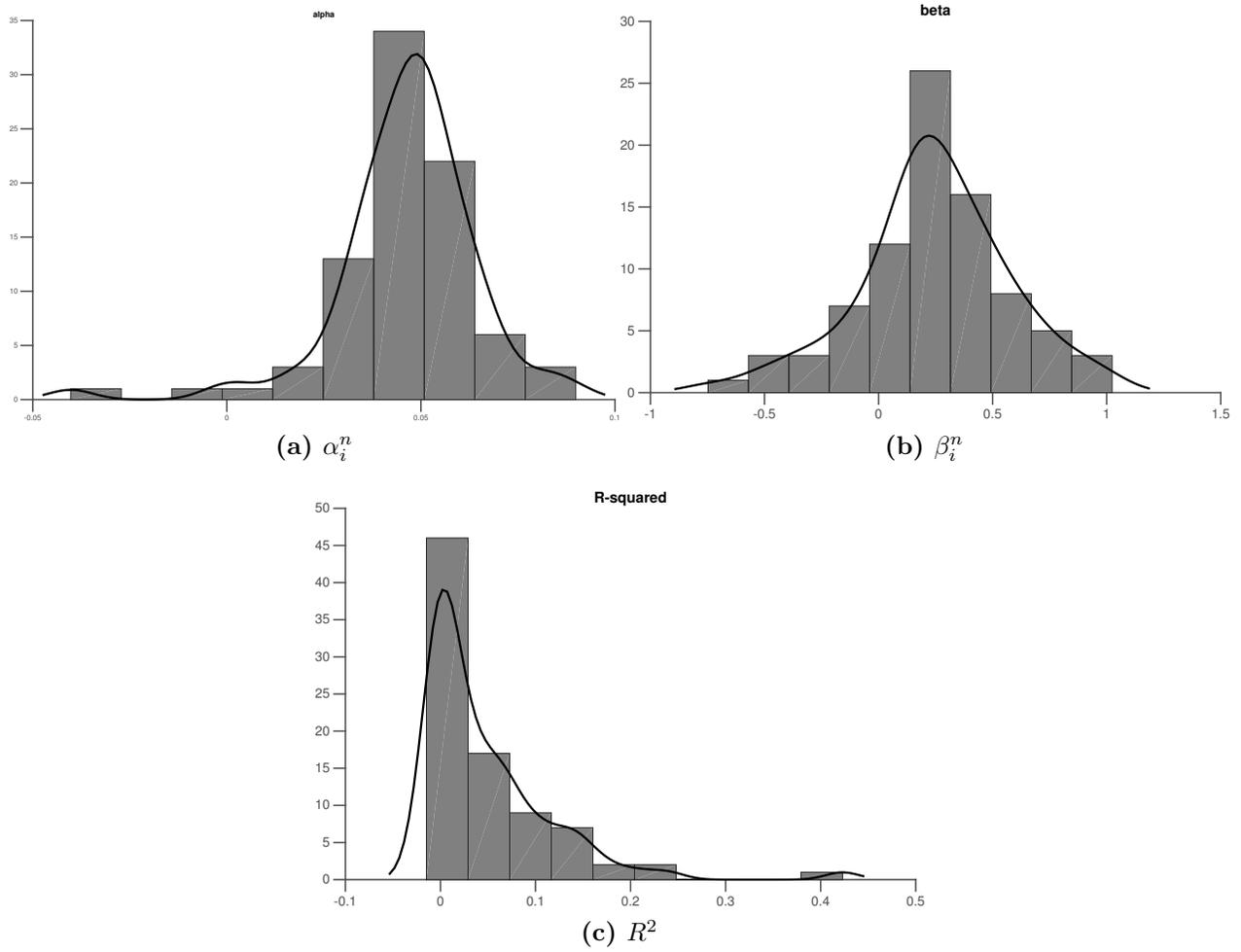


Figure 7. Predictive Regressions Individual Forecasters

Estimated regression coefficients and adjusted R^2 of regressions of the realized excess 10-year bond returns on the expected excess bond returns for all individual contributors with at least 60 months of forecasts:

$$rx_{t+1}^{10} = \alpha_i^{10} + \beta_i^{10} erx_{i,t}^{10} + \epsilon_{i,t+1}^{10}.$$

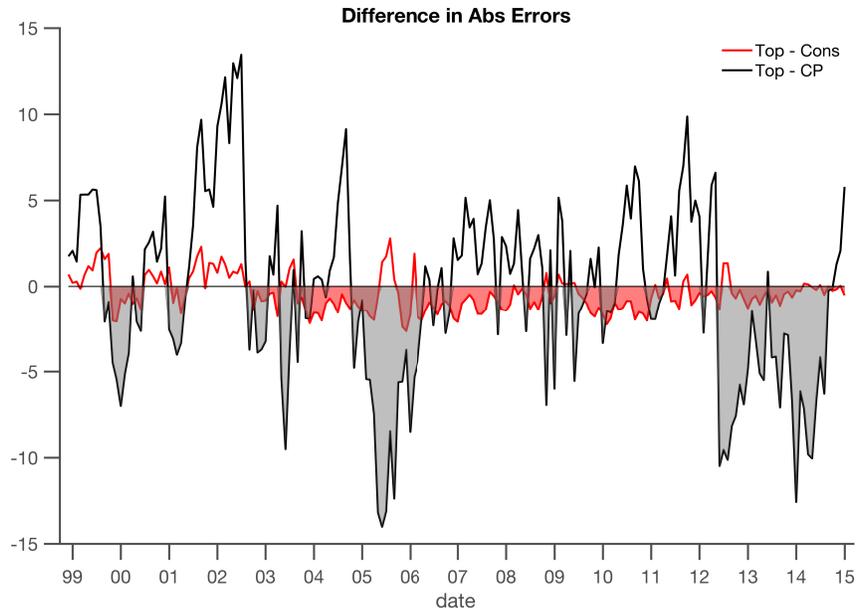


Figure 8. Absolute Bond Return Error Differences

Differences in absolute 10 year bond return forecast errors between the top 10 forecasters and the consensus, and the top 10 forecasters and the Cochrane and Piazzesi return forecasting factor.

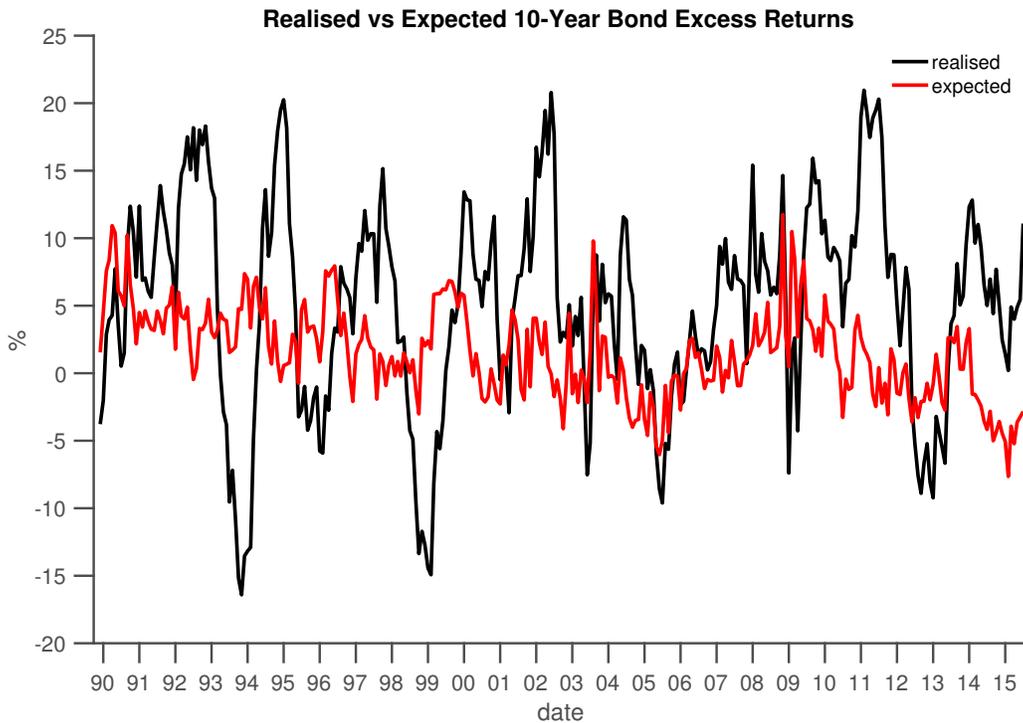


Figure 9. Realised vs Expected Returns

The black line denotes the realised excess return on a 10-year bond, while the red line denotes our measure of expected excess 10-year bond returns, EBR^* , measured at the time in which the expectation is realized, i.e. one year later, for comparison with the realised series.