

Expected Term Structures

ANDREA BURASCHI, ILARIA PIATTI, and PAUL WHELAN

ANDREA.BURASCHI@IMPERIAL.AC.UK

ILARIA.PIATTI@SBS.OX.AC.UK

PAWH.FI@CBS.DK

ABSTRACT

This paper studies the properties of bond risk premia in the cross-section of *subjective* expectations. We exploit an extensive dataset of yield curve forecasts from financial institutions and document a number of novel findings. First, contrary to evidence presented for stock markets but consistent with rational expectations, the relation between subjective expectations and future realisations is positive, and this result holds for the entire cross-section of beliefs. Second, when predicting short term interest rates, primary dealers display superior forecasting ability when compared to non-primary dealers. Third, we reject the null hypothesis that subjective expected bond returns are constant. When predicting long term rates, however, primary dealers have no information advantage. This suggests that a key source of variation in long-term bonds are risk premia and not short-term rate variation. Fourth, we show that consensus beliefs are not a sufficient statistics to describe the cross-section of beliefs. Moreover, the beliefs of the most accurate agents are those most spanned by a contemporaneous cross-section of bond prices. This supports equilibrium models and Friedman's market selection hypothesis. Finally, we use ex-ante spanned subjective beliefs to study predictions of several reduced-form and structural models and uncover a number of statistically significant relationships in favour of rational expectations.

JEL classification: E3, E4, G10, G12, G14, G17

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

First version: January, 2015.

This version: December, 2016.

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.

I. Introduction

A large asset pricing literature finds compelling evidence of predictability in several asset markets. A stream of the 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 of the literature, 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 features 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 However, [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. [Kojen, Schmeling, and Vrugt \(2015\)](#) find similar results in the context of global equities, currencies and fixed income markets across different countries. Both these studies argue that this result is difficult to reconcile with rational expectation models. In contrast to [Greenwood and Schleifer \(2014\)](#) and [Kojen, Schmeling, and Vrugt \(2015\)](#), 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 the potential source (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 formation of expectations.²

¹See e.g. [Ang, Bekaert, and Wei \(2007\)](#) and [Aioli and Timmermann \(2011\)](#).

²Other studies have looked at the dynamics of private sector expectations about interest rates and at the dynamics of the corresponding forecast errors, see e.g. [Cieslak and Povala \(2012\)](#) for fed fund rate forecasts

We begin by constructing measures of subjective bond risk premia (EBR) from professional 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 expected price changes. 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 median (Q2) forecaster EBRs is 1.06% for 10-year bonds. However, the median of the first quartile (Q1) EBR is -1.66% , which implies that these agents believe long-term bonds are hedges against economic shocks (growth and inflation) while the median of the third quartile (Q3) is $+3.57\%$, which is consistent instead with beliefs of long-term bonds being bets on economic shock. 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 about 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 the important question of whether the marginal investor should be identified by the agent with average (consensus) expectations. Notwithstanding the previous heterogeneity, overall expectations about bond returns display significant elements of rationality. They are positively related to future bond returns and are consistent - at the individual level - with same 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. When we examine in detail predictions conditioning on the identity of the forecaster, we find that banks and broker-dealer that act as primary dealers and trade directly with the Federal Reserve System are more likely to be between 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. We simulate the

and Piazzesi, Salomao, and Schneider (2015) for bond risk premia.

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 account holders as needed, and to bid on a pro-rata basis in all Treasury auctions at reasonably competitive prices.

fictitious trading account of primary dealers if they were trading against non-primary dealers institutions on the basis of their ex-ante forecasts using a simple duration based trading strategy. We find that primary dealers would have been able to persistently accumulate significant profits. We also find that the greatest relative accuracy (profit opportunity) occurs during periods in which the Fed aggressively reduces short-term rates. While this takes by surprise the consensus agent, whose expected excess bond returns are downward biased in these subperiods,⁴ it does not take by surprise primary dealers. 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. The result is quite important given that the top 5 primary dealers hold about 50% of all Treasuries.⁵

Third, we study the properties of long-term expected bond risk premia and 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 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 suggest 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 This suggests that subjective expectations are much less irrational than previously thought.

An important set of questions relates to the properties of the marginal agent who sets

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.

bond prices in equilibrium. While a working hypothesis of several models is that the representative agent holds consensus beliefs, the heterogeneous beliefs literature with short-selling constraints argue that the representative agent has to be an optimist in terms of expected returns (Hong, Sraer, and Yu (2013)). If pessimists cannot sell short, bond prices should reflect the beliefs of optimists. Another set of model, finally, argue that an intrinsic property of competitive markets is market selection. Trading markets eventually punish irrationality and the superior accuracy of rational agents allows them to accumulate economic importance in the Pareto weights. Thus bond prices should reveal (span) more tightly the beliefs of the most accurate agents. We use our rich panel dataset on beliefs to address this question by testing which beliefs are spanned by contemporaneous bond prices. We find that the beliefs of the most accurate agents are on average better spanned by current bond prices. For example, for the 10-year bond, regressions of EBR for portfolios of agents ranked on the basis of past accuracy on the principal components of the yield curve produce an R-squared of around 36% for the most accurate portfolio of agents and only 10% for the least accurate one. This result is consistent with the market selection hypothesis in competitive markets. Indeed, while optimist are on average more accurate in our sample and more spanned, the spanning result is reversed when the pessimists are most accurate. Thus, this result is not supportive of models with short selling constraints (as in Hong, Sraer, and Yu (2013)).

Fourth, an extensive literature in bond markets uses the properties of bond risk premia to propose economic models that are consistent with the data. The empirical evaluation of these models often accepts as approximations to agents expectations econometric projections of future realized returns on lagged state variables. We revisit this approach and instead of using the results of econometric projections, we use subjective expectations as directly revealed by agents to learn the merits of alternative economic models. We find that the out-of-sample performance of the survey-implied bond risk premia are highly competitive in forecasting future realized excess returns relative to some popular reduced form models. Indeed, in some cases subjective bond risk premia significantly outperform projections implied by either Cochrane and Piazzesi (2005) or Ludvigson and Ng (2009) forecasting factors, for all bond maturities. These findings suggests that surveys can indeed be used to build reliable measures of bond risk premia in real time and thus avoid issues related to in-sample versus out-of-sample model fitting. However, instead of the consensus, a better measures of subjective

expectation should build on the beliefs of the most spanned agent. Therefore, we use the spanned measure of *EBR* to evaluate a series of structural and reduced-form models. We find supporting evidence for rational expectation explanations of expected bond returns. In most cases, the empirical sign of the factor loading is consistent with predictions from theory. This result stands in contrast to the findings of [Greenwood and Schleifer \(2014\)](#) in the context of equity markets and suggests that rational expectation models cannot be dismissed so easily.

The paper proceeds as follows. Section [II](#) summarizes the empirical questions we aim to address and presents the data. Section [III](#) discusses the empirical properties of subjective bond risk premia. In Section [IV](#) we study the properties of expected short-term interest rates. Section [V](#) discusses the dynamics of the expected bond excess returns (EBR), the forecasting power of EBR for future realized excess returns and the cross-sectional variations in the forecast accuracy. Section [VI](#) analyses the link between EBR and statistical and structural models of expected bond risk premia proposed in the literature. Section [VII](#) discusses the results and concludes.

II. Framework 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 1 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:

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

Forecasts on future long-term interest rates depend on both expectations on future short-term interest rates and future bond risk premia. 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.

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 tests are set to study this hypothesis:

$H_0^{(1)}$: *Subjective expectations of bond returns are unbiased and the cross-section of individual expectations can be approximated to a reasonable degree of accuracy by the consensus beliefs.*

We test this hypothesis by testing both for the existence of a drift in forecasting errors and whether expectations of bond returns are internally consistent with the same agent expectations about future economic fundamentals (GDP growth and inflation). Since an important question in general equilibrium models is related to beliefs aggregation, we investigate the extent to which consensus beliefs can summarize the cross section of beliefs. Do agents agree about whether bonds are hedges or bets? Indeed, while in the first case their excess bond returns should be negative, in the second case they should be positive.

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 components using data on real time individual expectations. The advantage of our approach is to avoid data aggregation and model-dependent assumptions about the formation of expectations and test:

$H_0^{(2)}$: *Future short-term interest rates are not predictable on the basis of ex-ante expectations in real time.*

Since the dataset provides the identities of each forecasters, we can test this hypothesis directly, without having to assume a specific forecasting model to proxy for agents expectations. Moreover, we can directly investigate which of the forecasters is the most accurate.

For short term rates, for instance, we can distinguish primary dealers from all other banks and institutions and study whether this gives rise to an information advantage.

Third, since *EBR* are direct measures of bond risk premia, we revisit the literature that focus on the link between bond predictability and the dynamics of bond risk premia. We test:

$H_0^{(3)}$: *Long-term bond returns are unpredictable. Agents who seem to forecast short term rates do not have informational advantage in predicting long term bond returns.*

If the hypothesis of absence of long-term bond predictability is rejected even on subjective *EBR*, we can directly investigate the source of this predictability. Is this due to short term interest rates predictability or time varying risk premia?

Fourth, we compare the dynamics of *EBR* to statistical and structural models of risk premia that have been proposed in the literature. Our fourth set of tests investigates whether

$H_0^{(4)}$: *Do existing rational expectation models explain the dynamics of EBR and, if so, which of the models that are known to perform well in fitting bond excess return realizations also fit direct measures of agents risk premia.*

The last part of the paper proposes an alternative assessment of existing fixed income models. While it is tradition to evaluate them on the basis proxies of expected returns, we use direct measures of expected returns.

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.⁷ The contributors are asked

⁷In our analysis we use agent specific forecasts for the Federal Funds rate, Treasury bills with maturities 3-months/6-months/1-year, Treasury notes with maturities 1,2,5,10-years, and the 30-year Treasury bond.

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 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. Second, 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, it never falls below 35, and even considering only the forecasters who contribute to the sample for at least 10 years (120 monthly observations) the number of participants is always above 20. 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 1990.⁸ Third, 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.⁹ Fourth, 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.

Since forecast data consist of yields to maturity of coupon-bearing bonds, we construct curves of *expected* zero coupon discount rates by via 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 func-

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

⁹For a detailed discussion on the issues related to SPF, see [D'Amico and Orphanides \(2008\)](#) and [Giordani and Soderlind \(2003\)](#).

tion.¹⁰ 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 (quarterly horizons out to 1.25-years) zero coupon (continuously compounded) discount rates. 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 this 97 forecasters, 48 participate to the panel for at least 10 years, and on average they contribute to the cross section for about 201 months.

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.

III. The Cross Section of EBRs

A. The cross-sectional distribution

We document a large unconditional heterogeneity in the cross section of EBR point forecasts. Table I provides summary statistics for the median and the first and third quartile of the (1-year) EBR distribution for the 2, 5 and 10-year bonds. 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 there is consensus belief of 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. The top panel of Figure 1 shows the Q1, median and Q3 of

¹⁰Forward rates are given by $f_m = \beta_0 + \beta_1 \exp\left(-\frac{m}{\tau_1}\right) + \beta_2 \frac{m}{\tau_1} \exp\left(-\frac{m}{\tau_1}\right) + \beta_3 \frac{m}{\tau_2} \exp\left(-\frac{m}{\tau_2}\right)$ where m denotes the time to maturity.

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 $\tau_2 > 0$.

the cross-sectional distribution of EBR for 10-year maturity bonds. There exists significant time-varying heterogeneity around the consensus forecast. Given the wide use of the cross-sectional arithmetic mean, i.e. the consensus, in the academic literature dealing with survey data and in the financial industry, it is interesting to test more formally the null hypothesis that the cross-sectional properties of expectations can 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 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 most 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. In other words, 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 all bond maturities. 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.¹²

These findings raise important questions as to whether the assumption that marginal investor has average (consensus) expectation, as often assumed in the literature, is innocuous.

B. Persistence in the cross-section of EBR

An important assumption of bond pricing models is whether bonds hedge consumption risk or they are inflation bets. In the first case, bonds should earn a negative risk premium, in the second expected bond risk premia should be positive. Thus, we ask whether agents show persistence in their beliefs about *EBR*. Are individual forecasters persistently in one particular quartile of the cross-sectional distribution of subjective *EBR*? Figure 2 plots the time series average of seven individual forecasters' positions in the cross-sectional distribution

¹²The counter cyclicity of the dispersion in beliefs is consistent with the empirical evidence in [Patton and Timmermann \(2010\)](#) and [Buraschi, Trojani, and Vedolin \(2014\)](#), among others.

of subjective expected bond returns, for maturities between 2 and 10 years. This plot shows some persistence in individual forecasts. Indeed, in absence of persistence the time series average of the percentiles should be close to 0.5, for all forecasters. Instead, we see in Figure 2 that some institutions, like Goldman Sachs, have been persistent in their optimism about excess bond returns at all maturities; others have been persistent in their pessimism. In order to address this question rigorously, we rank all forecasters according to whether in a given month t their expected bond return 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 that separately for two different bond maturities (2 and 10 years) in Table II. If views are not persistent, all the entries in Table II should be approximately equal to 25%, while we expect the diagonal elements to be significantly higher than 25% in presence of persistent EBRs, and this is exactly what we find, in particular for the most extreme quantiles, Q1 and Q4. 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%. The results suggest that forecasters are persistently optimistic or pessimistic relative to the consensus excess return. When we repeat the same analysis for the transition probabilities for GDP and inflation forecasts (see Table III), we find that also macro forecasts are extremely persistent and the transition probabilities are of the same magnitude as for the EBR forecasts.¹³

The extent of persistence in beliefs about whether EBRs are positive or negative suggests that some agents have strong beliefs that bonds are hedges while others consider them to be inflation bets. This disagreement is persistent.

¹³The evidence of persistence in excess bond returns and macroeconomic forecasts is stronger than what Patton and Timmermann (2010) document for macroeconomic forecasts using data from the Consensus Economics Inc, at a quarterly frequency.

C. Internal consistence of 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 interpretation by investigating whether expected bond returns 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 EBR distribution, as shown in Table IV. If one focuses on the corners of these tables, we find that analysts in the first quartile of the EBR distribution are also in the fourth quartile of the GDP (or CPI) distribution with a probability between 37% and 46%, depending on bond maturity, which is significantly higher than 25%. Macro optimists are thus most likely in the lowest quartile of the cross sectional distribution of EBR forecast, and vice versa. This relation is consistent with the idea that good states of the economy are generally characterised by increasing yields, at least at short maturity, decreasing bond prices and thus lower expected excess returns. At the same time, the pattern is not overwhelming, suggesting that the drivers of beliefs about bond returns and the macroeconomy (GDP and inflation) are not the same.¹⁴

In order to investigate the drivers of this disagreement (being them behavioral or not) one needs to 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.

IV. The Short Rate

We start investigating the predictive accuracy of survey forecasters for the short-term interest rates.

¹⁴Interestingly, unreported results also show that optimism or pessimism about GDP growth is not related to optimism or pessimism about inflation: joint probabilities are close to 25% for all elements of the joint transition matrix.

A. Predictive regressions

We initially explore this question in the context of simple predictive regressions for the three-month Treasury yield. Due to its persistence, we run 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 for each decile $i = 0.10, \dots, 0.90$:

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

where $\Delta y_{t+1}^{3m} = y_{t+1}^{3m} - y_t^{3m}$. Figure 3 shows the cross section of regression coefficients and R^2 of regression (2) for each decile. The intercepts, α_i^n , are monotonically decreasing and insignificant up to the 4th decile; the slope coefficients are positive and significant for all deciles of the distribution. At the same time, we find that they are lower than one for all quantiles and relatively low, ranging between 0.12 and 0.16. The R^2 vary between 11% and 16%, and they are highest for the intermediate deciles. The consensus agent has a slightly larger predictive power but a biased forecast (the alpha is negative), while the low deciles, which correspond to the pessimistic agents in terms of interest rates (optimistic in terms of bond returns) are unbiased but have a slightly lower R-squared. These findings document that expectations of future yields are indeed positively correlated with future realizations across the distribution of beliefs. However, there is a large heterogeneity in the degree of accuracy.

To investigate the characteristics of this heterogeneity, we use the unique advantage of our dataset which provides all forecasters' identities. Then, we revisit the previous regression (2), but where i denotes each single contributor to the BCFF panel. For robustness, we focus on contributors with at least 10 years (120 months) of forecasts. Figure 4 shows the distribution of regression coefficients and R^2 for each forecaster. While the overall results confirm the previous findings of a substantial heterogeneity in predictive performances, two characteristics of these results emerge as striking. First, with the exception of only one forecaster, all estimated slope coefficients are positive and 90% are statistically significant. This suggests that professional forecasters in general do a relatively good job in predicting 3-month yield changes. Second, a few forecasters are extremely accurate, with slope coefficients

larger than 0.5 and R^2 almost equal to 60%. This is contrary to the evidence based on retail individuals and non-professional forecasters.

Finally, we note that the cross-sectional distribution of intercepts in the predictive regressions is largely skewed towards negative values: α_i^{3m} is negative in 45 out of 48 cases, and it is insignificantly different from zero for slightly more than half of the forecasters. This suggests that the average forecasters have been surprised by the extent to which the Fed has taken a dovish stance at the start of recessions.

Figure 5 shows this clearly by plotting the cumulative 3-month yield forecast errors over time for the average forecaster:

$$U_{cons}^{3m}(t) = \sum_{s=0}^t fe_{cons}^{3m}(s), \quad (3)$$

for $t = 1, \dots, T$ and where $fe_{cons}^{3m}(t) = y_{t+1}^{3m} - E_t^{cons}[y_{t+1}^{3m}]$.

$U_{cons}^{3m}(t)$ is not a martingale and has a negative drift, which is then reflected in a negative α in the predictive regression (2). In particular, we see a drastic drop in the cumulative errors in the early 90s, in the early 2000s and during the recent financial crisis.

B. Forecast accuracy

How accurate is the distribution of short rate survey expectations with respect to a credible benchmark, such as a unit root process for the 3-month yield? Due to the significant persistence of short term rates, it has been often argued that the most efficient expectation of the short rate is simply its current value. 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 for the matching period. Given the RMSE of each individual forecaster i , 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 calculate the relative accuracy \mathcal{A}_i of each forecaster as the ratio between the RMSE of each forecaster's expectation and the RMSE of a unit root benchmark:

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

Figure 6 displays the distribution of \mathcal{A}_i for the 48 contributors with at least 10 years of monthly forecasts. Noticeably, each individual forecaster's \mathcal{A}_i is lower than one, suggesting that all agents beat the unit-root benchmark, and the ratios range between 0.34 and 0.92, meaning that the best forecasters have an \mathcal{A}_i that is almost three times smaller than 1. This last finding provides additional evidence that surveys can provide reasonably good measures of expected bond returns.

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 percentiles*, $\mathcal{R}_{i,t}$. Then we compute the time average \mathcal{R}_i of these percentiles. Low percentiles correspond to greater accuracy. As in previous tests, we focus on forecasters with at least 10 years of data. The best forecasters in terms of average percentiles of squared forecast errors are summarized in the following table:

1	Goldman Sachs
2	Nomura Securities Inc.
3	Drexel Burnham Lambert Inc.
4	J.P. Morgan
5	U.S. Chamber of Commerce
6	National Association of Realtors
7	Loomis Sayles & Co.
8	BMO Capital Markets
9	Thredgold Economic Assoc.
10	Chase Manhattan Bank

Interestingly, the first four institutions in this list (and 6 out of the first 10), are currently primary dealers, or have been primary dealers at least once in our sample period, even if overall only 17 of the 48 financial institutions with at least 10 years of forecasts are or

have been primary dealers.¹⁵ Primary dealers are trading counterparties of the Fed in its implementation of monetary policy and they are also expected to make markets for the Fed on behalf of its official account holders, and to bid on a pro-rata basis in all Treasury auctions at reasonably competitive prices. 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.

In order to investigate the null hypothesis that primary dealers have a comparative advantage in forecasting the short rate, we compare the accuracy of this subset of forecasters, i.e. primary dealers, with respect to the other institutions in the panel of survey contributors. The list of primary dealers changes over time, and looking at accuracy percentiles at every time t instead of RMSE allows us to take this into account as well. At each month t , we compute the fraction of primary dealers (who are actually primary dealers and contributors to BCFF during that specific month) that are in the first, second and third tercile of the squared forecast error distribution and then average them over time. On average 40.15% of the primary dealers are in the first tercile, 29.95% in the second and 29.90% in the third.

Overall, the results above seem to show that primary dealers have better predictive performance for the short rate. While this holds unconditionally, it is interesting to understand whether the increased accuracy of primary dealers is generated in specific periods. Figure 7 shows the time series of average accuracy for primary dealers (PD) versus all other contributors (NPD), smoothed by computing a 12-month moving average of the monthly accuracies. It is clear that PD have a comparative advantage, and this advantage seems indeed to be present in specific time periods. The following subsection addresses this issue more formally by analysing the conditional individual forecast accuracy.

C. Conditional forecast accuracy

Figure 8 (upper panel) shows the time series of expected 3-month yield of both PDs and NPDs. The corresponding bottom panels show the forecast errors and the squared forecast errors. A pattern immediately emerges: the average expectations for PDs and NPDs are

¹⁵The list of primary dealers at every point in time can be obtained from the Federal Reserve Bank website.

very similar, but they 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 reduced the short term rate quite aggressively. While these decisions seem to take by surprise the consensus agent, whose expected excess bond returns are biased downward in these subperiods, they do not seem to surprise primary dealers, and this seems especially true during the recent financial crisis.

To investigate these differences rigorously, 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.¹⁶ 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 PDs and NPDs using a Kolmogorov-Smirnov test. The null hypothesis of the Kolmogorov-Smirnov test is that the accuracy percentiles PDs and NPDs 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 PDs and NPDs 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 for PDs and NPDs 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%.

Using the same approach as in the previous subsection (see Equation (3)) Figure 5 shows the cumulative 3-month yield forecast errors over time for the consensus forecaster, the average PD, $U_{PD}^{3m}(t)$, and the average NPD, $U_{NPD}^{3m}(t)$. Cumulative errors display a negative drift for both subgroups of forecasters, but the drift is much smaller for PDs, and it is basically absent during the recent financial crisis. Comparing the cumulative errors against

¹⁶Results are robust to the choice of time periods for the moving average.

a unit walk forecast we find that PD, NPD, and consensus forecasters are competitive but that PDs are significant outperformance over the full sample.

In general we find evidence that primary dealers are much better during inflection points, that are turns of business cycles when the Fed turns dovish by reducing the interest rate. During other period, expectations of the two sets of forecasters, as well as forecast errors, are very similar.

Note that the periods in which the primary dealers are significantly more accurate on their short rate expectations, are exactly the periods in which forecast errors in absolute value are larger and most agents seem to be biased. The consensus agent commits much larger forecast errors in absolute terms in bad times. When the short rate is increasing, the average forecast error for the consensus forecaster is -0.17% . In contrast, the mean of the forecast errors of the consensus forecaster in phases of decreasing interest rates is -0.32% . This is consistent with the findings in [Cieslak and Povala \(2012\)](#) who analyze the survey forecast expectations of the fed fund rate and show that *“most pronounced errors are negative and typically occur during and after NBER recessions as forecasters largely fail in predicting the extent of subsequent monetary easing”*.

This dynamics of the predictive advantage of primary dealers again suggest that they might have better knowledge of the policy function (due to private information or else), better knowledge of the arguments of the policy function, or the information value of observing demand/supply flows.

D. Economic Interpretation

The finding that primary dealers have an advantage in predicting the short term rate in periods of monetary easing has three potential explanations:

First, these sub-period correspond to bad states for the U.S. economy. Primary dealers might have better information about future economic growth. To the extent that interest rate policy is endogenous to economic growth, PDs are more accurate in anticipating monetary policy.

Second, due to their role as intermediaries in the Treasury market, PDs have better knowledge about market demand for Treasury bonds. Thus, they can form more accurate forecasts about the directions of short term interest rates. A potential limit of this hypothesis,

however, is that the superior accuracy of PDs manifests itself mainly during periods of aggressive dovish change in the stance of the monetary policy.

Third, PDs are able to collect information that is not easily available to the market (potentially private) about changes to the stance of the monetary policy.

We test the first hypothesis by comparing the accuracy of PDs and NPDs about future real economic growth and inflation. Figure 9 shows the accuracy percentiles of real GDP and inflation expectations for each individual forecaster and highlights the primary dealers in the sample. It is immediately clear that primary dealers do not perform better than other agents in forecasting the inputs of the Taylor rule, i.e. inflation and GDP growth. In fact, if anything, the accuracy of PDs' inflation expectations is lower than that of NPDs (see also bottom panel of Figure 10).¹⁷ We can formally test the difference between the accuracy distribution of PDs and NPDs as above using a Kolmogorov-Smirnov test. Considering the full sample, the p-value of the tests is 6.5% for inflation and 62.7% for GDP growth, which implies that we cannot reject the null hypothesis in both cases at a level of 5%. However, the distributions of inflation forecast accuracy for PDs and NPDs are significantly different at a level of 10%, and these conclusions do not change if we look at subsamples of increasing and decreasing fed fund rates. Therefore, we cannot reject that the growth forecast accuracy of primary dealers and other institutions come from the same distribution. Actually, the best macro forecasters on average are institutions like Action Economics and ClearView Economics, while big primary dealers as Goldman Sachs, J.P. Morgan and Nomura are consistently in the worst half of growth and inflation forecaster accuracy.

E. Economic significance

The greater accuracy of PDs' expectations on the short rate during periods of decreasing rates is highly statistically significant. Is it also economically significant? In order to test this, we design a fictitious trading strategy based on agents expectations.

To trade their view about about the 3-month yield in 12 months, we assume that agents replicate the forward rate in 12 months for 3 months, using available Treasury bonds with corresponding maturities. Thus, an agent that expects a relatively low short rate with

¹⁷Note that realized GDP growth is available only quarterly. Therefore, the time series of GDP growth accuracy is also quarterly.

respect to consensus would go long the 15-month bond and short the 12-month bond. We approximate this trading strategy by using the 2-year bond as a substitute of the 15-month bond, since the constant maturity 15-month bond yield is not directly available. In other words, we assume that agents expecting a relatively high short term rate in a year will sell the 2-year bond and buy the 1-year bond.

Every month, we stratify agents according to their beliefs relative to the consensus view about the 3 month rate. Then, we compute the return of rolling trading strategy in which agents take positions every month and hold these positions until maturity (i.e. 1 year). We record this fictitious return for every agent and in every month in which the agent is contributing to the panel, and then average over time. The average of the mean returns for primary dealers is 0.13%, and it is -0.026% for non primary dealers. The difference in cumulative returns is summarized in Figure 11.

Even if the difference in expectations and in forecast errors may not appear particularly large between the two categories and is present only in specific periods (see again Figure 8), PDs are able to accumulate (theoretical) profits that are economically very significant.

Notice that the mean return of this strategy across all forecasters is slightly positive but close to zero, at 0.029%. This is suggestive that this cross-section of expectations is representative of the whole population. This also shows the limits of aggregating expectations using consensus beliefs.

V. The Long-term Rates and Bond Risk Premia

In this section we focus on the following questions: First, given a direct subjective measure of expected bond risk premia $erx_{i,t}^n$, we revisit the literature of the time-variation of risk premia which plays an important role in the discussion about the rejection of the expectation hypothesis in bond markets. Second, we quantify the extent of accuracy of professional forecasters. How accurate are agents' expectations with respect to well-cited empirical models? Does the superior predictive ability of primary dealers on short-term rates lead to an advantage long-term bonds? Since long-term bond returns are affected by both changes in short-term interest rates and bond risk premia, if the first component were to be dominant we should find that primary dealers conserved the edge in forecasting long-term returns.

This is, therefore, an indirect test of the importance of the dynamics of bond risk premia for the dynamics of long-term bond returns.

A. *Time-varying risk premia*

An extensive literature in fixed income studies the properties of bond risk premia and argues that these 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 using directly 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:

$$erx_{i,t+1}^n = \alpha_i^n + \beta_i^n erx_{i,t}^n + \epsilon_{i,t+1}^n. \quad (4)$$

The results are summarized in Table V and show that the slope coefficients are significantly different from 0 for all quartiles i at any traditional statistical levels. We can therefore reject the null hypothesis that bond risk premia are constant. The results are very strong and support the hypothesis that expected excess bond returns are indeed 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.

To summarize, these results offer direct evidence in support of the interpretation of the existence of predictability due to time-variation in expected excess bond returns.

B. *Predictive regressions*

To assess the accuracy of these surveys and the extent 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 10 years (120 months)

of forecasts:

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

Figure 12 shows the distribution of regression coefficients and R^2 of regressions (5) for each forecaster. The result show that notwithstanding heterogeneity in accuracy, a few forecasters are extremely accurate with slope coefficients close to one and R^2 almost equal to 20%. The correlation between expectations and future realization of excess bond returns is positive for 40 out of 48 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 Kojien, Schmeling, and Vrugt (2015) find in the context of global equities, currencies and global fixed income returns across countries. This may be due either to issues related to the aggregation in those data sets or to differences between professional and non-professional forecasters. Our results shows that agents beliefs are substantially more rational than otherwise thought.

C. Forecast accuracy

We study forecast accuracy at the level of each individuals forecaster i by computing the root mean squared errors ($RMSE_i^n$) for bond maturity $n = 10$, as

$$RMSE_i^n = \sqrt{\frac{1}{T} \sum_{t=1}^T (rx_{t+1}^n - \text{er}x_{i,t}^n)^2}.$$

They range between 7.6495 and 12.0509. Since individual forecasters may appear in the sample at different times, we assess their accuracy relative to a model. We use two reduced-form predictability factors that are widely used in the literature:

- The Cochrane and Piazzesi (2005) return forecasting factor is a tent-shaped linear combination of forward rates that has been shown to be a powerful predictor of future bond returns. It has been argued to subsume 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 ‘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 display time variation (see, for example, [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 - January 1988. Then, using an expanding window we estimate factor loadings used to construct a date t predicting factor, $CP(t)$, using realized returns available 1-year ago.

- The [Ludvigson and Ng \(2009\)](#) real macro factor is a broad based summary time-series based on a panel of macro economic variables capturing the level of economic activity. However, predictive return regressions based on such panels potentially overstate the information set available to investors in real time. To compare the real time forecast accuracy of macro versus survey based predictability we follow [Ghysels, Horan, and Moench \(2014\)](#) who argue proper tests of macro predictability should be based on vintage first release data. We obtain this data from the Archival Federal Reserve Economic Database (ALFRED) at the Federal Reserve Bank of St. Louis. We build a real-time macro predictability factor in real time computed recursively from the first principle component of a vintage macro panel and denoted this time factor LN .¹⁸

The in-sample $RMSEs$ of these two models over the full sample are 7.3857 and 7.7758, respectively. When we compare these values to those obtained from the surveys, it is evident that these models outperform even some of the best forecasters in-sample. However, comparison is unfair since the model $RMSEs$ are in-sample and affected by a look-ahead bias, as some information is not available to the forecasters in real time. Therefore, we calculate out-of-sample relative performance. The difference is potentially important. In 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. Accordingly, we proceed with an out-of-sample assessment: we initialize both models in January 1998 and obtain model-implied expectations recursively using expanding windows. We compare these to survey forecasts, which are out-of-sample by construction, as agents form their expectations of time $t + 1$ returns only using information available at time t . Then, we compute a measure of relative performance $\mathcal{A}_i^?$:

¹⁸ Our data set broadly covers the same economic categories as [Ghysels, Horan, and Moench \(2014\)](#) which is chosen to match [Ludvigson and Ng \(2009\)](#) as close as possible. The final dataset comprises of a real time panel of 98 economic time series that are transformed into stationary growth rates.

$$\mathcal{A}_{i,t}^n = \frac{RMSE_i^n(Survey)}{RMSE^n(Model)}.$$

Values smaller than one imply better performance under the subjective measure.

Out-of-sample, we find that an important fraction of survey forecasters perform better than both models. For example, relative to both the *CP*-factor model and the *LN*-factor model, the relative accuracy on the 10-year bond of survey forecasters, \mathcal{A}_i^{10} , is less than one for about 21% of the individual agents, and \mathcal{A}_i^{10} is between around 0.8 and 1.5 for all forecasters.¹⁹ These findings suggest that survey-implied bond risk premia are highly competitive in forecasting future realized excess returns relative to popular reduced form models.

In fact, not only there is evidence of accuracy in the cross-section, but this performance 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 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 is not persistent, all the entries in Table VI should be approximately equal to 25%. If, on the other hand, accuracy is persistent, we expect the diagonal elements to be significantly higher than 25%. We find that the accuracy of the most extreme quantiles, Q1 and Q4, are very persistent. For example, a forecaster in the first quartile of the cross-sectional distribution of 10-year EBR has a probability of 56% 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 a bad forecasting performance is more persistent than a good one. In all cases, the probability of remaining in the same quartile is significantly higher than 25% at a level of 5%.

This confirms two conclusions. First, expectations of professional forecasters are far from being irrational. Second, surveys can be used to build reliable measures of bond risk premia.

¹⁹We also find that the out-of-sample RMSE of the models is quite sensitive to the sample period considered and to the choice of the starting date for the out-of-sample period. For robustness, we also require the survey forecasters to have at least 3 years of monthly observations in the out-of-sample period.

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 of the panel and can proxy the beliefs of the marginal agents are not supported by our results.

D. Primary Dealers

The previous section documents that primary dealers 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 the long rate? 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 rate over the life of the bond; a rejection of this hypothesis, on the other hand, would suggest that the dynamics of long-term bond returns are dominated by other components, such as bond risk premia. In this case, knowing the dynamics of short-term rates may not suffice to earn extra returns when trading long-term bonds.

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 , rank them, and average across time periods. Finally, we compare these long-term accuracy percentiles with the corresponding accuracy on the short rate.

Figure 13 shows the average percentiles of each individual forecaster's accuracy for the 10-year excess bond return and the 3-month yield. The red dots correspond to the primary dealers. The two rankings are highly correlated, in fact a regression line fitted on the points in Figure 13 has a significant slope coefficient of 0.78 and an adjusted R-squared of 34%. However, the link is much less strong if we focus on the subsample of primary dealers: the regression coefficient is 0.57 and it is only marginally significant, with an adjusted R-squared of 15%. Thus, the 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. The best forecasters in terms of average percentiles of squared forecast errors for the 10-year bond are summarized below:

1	Thredgold Economic Assoc.
2	UBS
3	Goldman Sachs
4	Fleet Financial Group
5	The Northern Trust Company
6	DePrince & Assoc
7	Fannie Mae
8	J.W. Coons & Associates
9	Nomura Securities Inc.
10	ClearView Economics

At each month t , we compute the fraction of primary dealers (who contribute to BCFF during that specific month) that are in the first, second and third tercile of the squared forecast error distribution and then average them over time. For the 10-year yield, on average 29.11% of the primary dealers are in the first tercile, 31.06% in the second and 39.83% in the third. The results contrast with those for the 3-month yield for which the primary dealers are overrepresented in the best accuracy tercile.

Panel A of Table VII displays the joint distribution of forecast accuracy for the 10-year EBR and 3-month yield, that is the probability of being in a given tercile of the 3-month yield accuracy percentile distribution *and* a given tercile of the 10-year EBR accuracy percentile distribution, at the same time.

The elements on the diagonal show that there is a link between accuracy at the short and at the long end of the term structure, which is not surprising given that, for example, the correlation between realized 3-month and 10-year yield, at the monthly frequency, is around 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 correlation between the accuracy on the 3-month yield and on the 10-year EBR is far from perfect.

When we focus on primary dealers, see Panel B of Table VII, the evidence is different and intriguingly so: the fraction of primary dealers who are accurate in both dimensions is slightly higher than for all forecasters, but there is an asymmetry between the 3-month yield and the

10-year *EBR* accuracies. We test directly the null hypothesis that the accuracy percentiles of PDs and NPDs 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 62.73%. Even after distinguishing periods of increasing and decreasing rates, we cannot reject the null hypothesis with p-values of 75.89% and 22.40%, respectively. Overall, primary dealers have a significantly better predictive performance only for the short rate.

This suggests that the dynamics of expected excess bond returns at longer maturities might indeed be dominated by a bond risk premium component. Moreover this risk premium is time-varying.

Since risk premia are time-varying and accuracy is quite heterogeneous, it is natural to ask whether the most accurate forecasters are also those whose beliefs are more spanned. This question is important in the context of the correct aggregation of beliefs and it is the topic of the following section.

VI. Subjective Risk Premia and Rational Expectation Models

A. *Spanning properties*

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 behavioural finance literature argues that in presence of short-selling constraints marginal agents ought to be those holding optimistic beliefs about expected returns (Scheinkman and Xiong (2003) Hong, Sraer, and Yu (2013)). Since pessimists cannot short-sell, their beliefs are not revealed (spanned) by equilibrium asset prices. The general equilibrium literature with disagreement and speculation argues, on the other hand, that in absence of short-selling constraints irrational agents eventually lose economic weight to the benefits of less biased agents. It is not a matter of optimism but of accuracy. The superior accuracy of rational agents allows them to accumulate economic importance in the Pareto weights of the representative agent (as in Basak (2005), Buraschi and Jiltsov (2006), Jouini and Napp (2006), Xiong and Yan (2010), Chen, Joslin,

and Tran (2012), Buraschi and Whelan (2016), 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 instead 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 the one spanned by bond prices.

We use information on agents beliefs from both the time series and the cross section to address the question of whether the beliefs of the most accurate agents are more spanned by current bond prices. To proceed parsimoniously, we first decompose the yield curve up to 10 years maturity in a small number of (orthogonal) principle components.²⁰ Then, we sort agents according to the level of their accuracy. Namely, at every month t we consider all agents present in the panel in the previous 12 months and compute the average squared forecast errors over a period straddling month t , based on these past year of expectations.²¹ We rank agents by their average accuracy at each time t and form decile portfolios. Then, we compute the average EBR within each decile. This procedure provides us with a cross section of beliefs with different levels of accuracy, that allows us to test the hypothesis that a superior accuracy is correlated with a larger Pareto weight, and therefore a larger degree of spanning.

To test this hypothesis we run regressions of 10-year *EBRs* for different deciles of the accuracy distribution onto the first five principal components of the term structure, which

²⁰The first three factors are often labelled in the literature as level, slope, and curvature, based on how shocks to these factors affect the shape of the yield curve (see, for example, Litterman and Scheinkman (1991), Dai and Singleton (2003), or Joslin, Singleton, and Zhu (2011)). We consider the first five factors, which explain around 99.9999% of the overall variation in yields.

²¹Note that only the forecast errors based on the EBR 12 months before is already realized, but while the others are unrealized they are still likely to affect the accumulation of wealth of the agents up to time t .

efficiently summarize the cross section of bond prices:

$$erx_{i,t}^n = \beta_{i,0}^n + \sum_{j=1}^5 \beta_{i,j}^n PC_{j,t} + \epsilon_{i,t}^n. \quad (6)$$

Table VIII reports the results of these regressions where EBR1 denotes the most accurate and EBR10 the least accurate beliefs. We find a monotonic link between accuracy and degree of spanning, measured as the adjusted R-squared of the regression. Consistent with the general equilibrium literature with disagreement and no frictions, accurate investors' expectations are well spanned by the cross-section of bond prices, while for the least accurate investors the degree of spanning is almost four times lower.

For comparison, we run the spanning regression (6) also for the consensus beliefs, $erx_{c,t}^{10}$ and we find an adjusted R^2 of about 24%, versus 36% of the most accurate agents. As an additional benchmark, we run the same regressions using ex-post realized returns as a proxy for ex-ante bond risk premia. Consistent with the large literature on bond return predictability, we find that the slope of the yield curve (PC2) reveals information about bond risk premia (see Campbell and Shiller (1991) and Fama and Bliss (1987)), but the R^2 is only 18%. On the basis of ex-post realized returns, one might be tempted to conclude that the amount of spanning is somewhat limited. On the other hand, when one considers direct measures of subjective expected returns of accurate agents, there is strong evidence that the variation in subjective bond risk premia is largely spanned by date t yield factors.

Taken together, we conclude that the beliefs of forecasters who have been on average more accurate appear better spanned by contemporaneous prices than the beliefs of the least rational agents. This result is intriguing and consistent with market selection in competitive markets.

B. Rational expectation models vs subjective risk premia

The empirical evaluation of rational expectation models is traditionally conducted by approximating expected risk premia by sample averages of future returns. $E(rx_{t,t+T})$ is often proxied by $\frac{1}{T} \sum_{s=t}^{t+T-1} rx_{s,s+1}$ and conditional expectations $E_t(rx_{t,t+T}|\mathcal{F}_t)$ by sample projections of future realizations $rx_{s,s+1}$ onto observables with respect to 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_t^i 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 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 way the experiment is designed.²²

Direct measures of subjective expectations can address these three problems. They provide a useful alternative to assess alternative structural and reduced-form models of bond risk premia. Under the assumption that erx_t measure expectations of bond excess returns accurately, alternative models of risk premia can be ranked based on their ability to explain the dynamics of erx_t , as opposed to sample averages (or projections) of rx_{t+1} . Indeed, previous results confirm that, out-of-sample, survey-implied bond risk premia are highly competitive in forecasting future realized excess returns relative to some popular reduced form models. Let \mathcal{M}_t^j be a model-implied specification of bond risk premia, we run regressions of the form:

$$erx_{i,t}^n = a_i^n + b_i^n \mathcal{M}_t^j + \epsilon_{i,t}^n. \quad (7)$$

We obtain measures for $erx_{i,t}^n$ both by using the forecasts of the agents with greatest spanning properties (i.e. the most accurate), which should reveal more closely the beliefs of the marginal agent in competitive markets. Risk premium models, \mathcal{M}_t^j , are grouped into three categories: (a) proxies for state-variables that arise in structural models, (b) generalized affine and volatility models, and (c) reduced-form models.

²²Examples include the length of the training period, the start of the out-of-sample period, the use of fixed versus time-varying parameters, the out-of-sample horizon, etc.

- 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 \mathcal{M}_t^1 , 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 ($LRR(g)$) and expected inflation ($LRR(\pi)$).
- In economies with external habit preferences, such as [Campbell and Cochrane \(1999\)](#), 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 a time-varying expected returns. To obtain a proxy of risk premium \mathcal{M}_t^2 , we follow [Wachter \(2006\)](#) and calculate consumption surplus ($Surp$) using a weighted average of 10 years of monthly consumption growth rates: $Surplus = \sum_{j=1}^{120} \phi^j \Delta c_{t-j}$, where the weight is set to $\phi = 0.97^{1/3}$ to match the quarterly autocorrelation of the P/D ratio in the data.²³
- In models where agents agree to disagree, the stochastic discount factor is a direct function of disagreement. Examples of such models include [Buraschi and Whelan \(2016\)](#), who argue that disagreement about real growth rates is a significant determinant of expected excess bond returns, [Ehling, Gallmeyer, Heyerdahl-Larsen, and Illeditsch \(2015\)](#) and [Hong, Sraer, and Yu \(2013\)](#), who argue about the importance of inflation disagreement. We denote their proxies for real and inflation disagreement as $DiB(g)$ and $DiB(\pi)$, respectively.
- Several models argue about the importance of liquidity risk in economies in which financial intermediaries face priced-shocks to funding conditions. An example of this literature is [Fontaine and Garcia \(2012\)](#). We follow their empirical approach and test the significance of their funding liquidity factor (Liq).

²³For consumption data we obtain seasonally adjusted, real per-capita consumption of nondurables and services from the Bureau of Economic Analysis.

VOLATILITY MODELS

Dai and Singleton (2000) provide a detailed study of the completely affine class of term structure models in which elements of the state vector that affect bond volatility also affect expected returns. In an equilibrium context, Le and Singleton (2013) discuss the link with structural models where the state vector follows an affine diffusion and priced volatility risks affects expected returns.²⁴ Motivated by this literature we consider three proxies for volatility risk:

- The intra-month sum of squared returns on a constant maturity 30-day Treasury bill as a proxy for short rate volatility, denoted by $\sigma_y(3m)$.
- The Treasury variance risk premium on 10-year Treasury bond futures as studied by Mueller, Vedolin, Sibtchevsky, and Whelan (2016) which we denote *TVRP*.
- The realized Treasury jump risk proposed by Wright and Zhou (2009) which we denote as *Jump*, updated to the most recent period.

REDUCED-FORM MODELS

Finally, Cochrane and Piazzesi (2005) and Ludvigson and Ng (2009) have proposed two influential factors that are found to explain a significant proportion of realized excess bond returns. The first is based on a combination of forward rates; the second is based on principal components of a large panel data of economic variables. As discussed above we construct real time versions of these return forecasting factors denoting them *CP* and *LN*, respectively.

RESULTS:

Table IX shows the results of regressions of $erx_{i,t}^n$ onto alternative specifications of \mathcal{M}_t for 10-year bond using the most accurate decile of forecasters. To summarize, we find that several of the structural models are indeed consistent with subjective *EBR*. The relationship is positive and statistically significant. This was not granted ex-ante, as the result is contrary to previous studies for equity returns which argue that equilibrium models generate implied risk premia that correlate negatively with empirical risk premia. When we compare different models, several interesting results emerge.

²⁴Such models include the class of long-run risk models (Bansal and Yaron (2004), Bollerslev, Tauchen, and Zhou (2009), or Bansal and Shaliastovich (2013)), habit models (Wachter (2006) or Buraschi and Jiltsov (2007)) or models with heterogeneous agents (Buraschi and Whelan (2012) or Piatti (2014)).

First, when we consider models with disagreement ($DiB(g)$ and $DiB(\pi)$), we find that disagreement about real growth is highly significant and with a positive slope coefficient, while inflation disagreement is not significant. Factor loadings are consistent with the rational disagreement models of [Buraschi and Whelan \(2016\)](#). Moreover, real disagreement is explaining 6% of the variation in 10-year subjective bond risk premia.

Second, the liquidity factor of [Fontaine and Garcia \(2012\)](#) is significantly negatively correlated with $erx_{i,t}^n$, consistent with the interpretation that negative shocks to this factor are bad news for funding conditions, thus raising expected returns. The liquidity factor Liq by itself explains 7% of the variation in 10-year subjective bond risk premia.

Third, for habit models the theoretical relation between consumption surplus ($Surp$) and EBR depends on whether surplus loads positively or negatively on short term interest rates. [Table IX](#) reports a positive and highly statistically significant loading which implies a negative shock to surplus raises expected returns. This is consistent with a setting in which surplus loads positively on short rates in which case returns will be low when marginal utility is high, and thus they command a positive risk premium.

Fourth, models assuming bond risk premia to be proportional to economic uncertainty, in the spirit of [Bansal and Yaron \(2004\)](#), are able to explain a reasonable proportion of the dynamics of subjective bond risk premia $erx_{i,t}^{10}$, with adjusted R^2 of 14%. However, only real uncertainty is significant entering with a positive loading. This is consistent, for instance, with the model discussed in [Bansal and Yaron \(2004\)](#) in which greater real GDP uncertainty raises interest rates, lowers bond prices and thus predicts positive future expected returns.

Fifth, when we examine generalized affine volatility models, we find that the explanatory power of $TVRP$ and $Jump$ factors is statistically large and with slope coefficients consistent with theory: the multivariate adjusted R^2 is 18% and both factors are significant at the 1% level. This is interesting since a well documented puzzle in the term structure literature is the failure of volatility factors to forecast ex-post realized returns. However, we find no relationship between short rate and subjective returns.

Fifth, when we study reduced-form predictive models of bond returns, we find a significant positive relationship between EBR and the real time LN return forecasting factor. The statistical power is very large (significant at well below the 1% level) with an adjusted R^2 of 6%. The explanatory power of real time CP is much weaker. The regression coefficient is

positive, as expected, but only significant at the 10% level.

To summarize, in the context of the equity market, Greenwood and Schleifer (2014) find that several rational expectation models are negatively correlated with survey expectations of stock market returns. 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.*” On the other hand, we find significant positive correlation between proxies of expected excess returns obtained from some of the rational expectation models and expectations of bond excess returns erx_t of the most accurate (and most importantly most spanned) agents. This suggests that, at least in the context of bond markets, rational expectation models cannot be dismissed so quickly.

VII. 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, we find that 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 display significant elements of rationality. In fact, individual expectations of bond returns are consistent with agents’ forecasts about GDP and inflation.

Secondly, we find evidence of predictability in short-term interest rates and we show that 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, and their superior forecast accuracy is both statistically and economically significant. 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. The result is quite important given that the top 5

primary dealers hold about 50% of all Treasuries.

Third, we study the properties of long-term expected bond risk premia and strongly reject the hypothesis that bond risk premia are constant. Moreover, we 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 long-term bond predictability. Overall, results for long-term bond returns strengthen the evidence of rationality in the cross-section of survey forecasters, since 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.

Fourth, expectations of bond risk premia are largely spanned by the current term structure of bonds prices and the degree of spanning is substantially larger than when using sample averages of future excess returns as proxies of bond risk premia. Even more importantly, the degree of spanning greatly differs in the cross-section of agents beliefs. Indeed, there is a strong positive relation between spanning and forecasting accuracy in the cross-section: the beliefs of agents who have been more accurate in their forecasts in the preceding months are more spanned by the term structure of bond yields. This is consistent with the predictions of general equilibrium heterogeneous agents models with speculative trading and no frictions. In these models, the pricing kernel is a stochastic weighted average of agents beliefs, where relative weights depends on the wealth accumulation generated by belief-based trading.

These findings suggests that surveys can indeed be used to build reliable measures of bond risk premia in real time and thus avoid issues related to in-sample versus out-of-sample model fitting, as long as we rely on the beliefs of the most spanned, i.e. most accurate, agents instead of just looking at the consensus. Therefore, we use the spanned measure of *EBR* to evaluate a series of structural and reduced-form models. We find supporting evidence for rational expectation explanations of expected bond returns. In most cases, the empirical sign of the factor loading is consistent with predictions from theory. This result stands in contrast to the findings of Greenwood and Schleifer (2014) in the context of equity markets and suggests that rational expectation models cannot be dismissed so easily.

References

- Aioli, M., C. Capistran, and A. Timmermann, 2011, *The Oxford handbook of economic forecasting* . chap. Forecast combinations (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*.
- Bollerslev, Tim, George Tauchen, and Hao Zhou, 2009, Expected stock returns and variance risk premia, *Review of Financial Studies* 22, 4463–4492.
- 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, A., and P. Whelan, 2012, Term structure models and differences in belief, *Imperial College, Working Paper*.

- Buraschi, Andrea, and Paul Whelan, 2016, Hedging, speculation, and interest rates, *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.
- Campbell, John Y, and Robert J Shiller, 1991, Yield spreads and interest rate movements: A bird's eye view, *Review of Economic Studies* 58, 495–514.
- 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., 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.
- Dai, Q., and K.J. Singleton, 2000, Specification analysis of term structure of interest rates, *Journal of Finance* 55, p. 1943–78.
- Dai, Qiang, and Kenneth Singleton, 2003, Term structure dynamics in theory and reality, *Review of Financial Studies* 16, 631–678.
- 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.
- Ehling, Paul, Michael Gallmeyer, Christian Heyerdahl-Larsen, and Philipp Illieditsch, 2015, Disagreement about inflation and the yield curve, *Working paper. University of Pennsylvania*.
- Fama, Eugene F, and Robert R Bliss, 1987, The information in long-maturity forward rates, *American Economic Review* 77, 680–92.
- Fontaine, Jean-Sébastien, and René Garcia, 2012, Bond liquidity premia, *Review of Financial Studies* 25, 1207–1254.

- Friedman, M., 1953, *Essays in Positive Economics* (University of Chicago Press, Chicago).
- Ghysels, Eric, Casidhe Horan, and Emanuel Moench, 2014, Forecasting through the rear-view mirror: Data revisions and bond return predictability, *FRB of New York Staff Report*.
- 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.
- Joslin, S., K. J. Singleton, and H. Zhu, 2011, A new perspective on gaussian dynamic term structure models, *Review of Financial Studies* 24, 926–970.
- 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.
- Le, A., and K.J. Singleton, 2013, The structure of risks in equilibrium affine models of bond yields, Discussion paper, Working Paper, UNC.
- Litterman, R., and J. Scheinkman, 1991, Common factors affecting bond returns, *The Journal of Fixed Income* 1, 54–61.

- Ludvigson, Sydney C., and Serena Ng, 2009, Macro factors in bond risk premia, *Review of Financial Studies* 22, 5027–5067.
- Mueller, P, A Vedolin, P Sabtchevsky, and P Whelan, 2016, Variance risk premia on stocks and bonds, *Working Paper*.
- 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.
- Piatti, Ilaria, 2014, Heterogeneous beliefs about rare event risk in the lucas orchard, *Working Paper, Oxford University*.
- 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 EO, 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.
- Wright, Jonathan H, and Hao Zhou, 2009, Bond risk premia and realized jump risk, *Journal of Banking & Finance* 33, 2333–2345.
- Xiong, W., and H. Yan, 2010, Heterogeneous expectations and bond markets, *Review of Financial Studies* 23, 1433–1466.

VIII. Tables

Table I. Summary Statistics

Summary statistics of the first (Q1), second (Q2) and third (Q3) quartiles of the distribution of subjective expected excess bond returns, for maturities of 2, 5 and 10 years, and forecast horizon of 1 year. Sample period is January 1988 to July 2015 (331 observations).

Q1	2 Year	5 Year	10 Year
Mean	-0.0283	-0.9473	-1.6574
Std Dev	0.0047	0.0158	0.0330
Min	-0.0133	-0.0600	-0.1047
Max	0.0119	0.0313	0.1087
Skew	-0.0581	-0.0876	0.0062
Kurtosis	2.4914	2.6679	3.1663
AR(1)	0.7949	0.7457	0.7433
Q2	2 Year	5 Year	10 Year
Mean	0.2775	0.3362	1.0629
Std Dev	0.0049	0.0158	0.0333
Min	-0.0110	-0.0437	-0.0823
Max	0.0153	0.0442	0.1217
Skew	0.0502	-0.0610	-0.0281
Kurtosis	2.3404	2.6997	3.0678
AR(1)	0.8174	0.7487	0.7565
Q3	2 Year	5 Year	10 Year
Mean	0.5565	1.4578	3.5662
Std Dev	0.0053	0.0168	0.0359
Min	-0.0063	-0.0280	-0.0532
Max	0.0175	0.0572	0.1539
Skew	0.2350	-0.0158	0.0410
Kurtosis	2.1081	2.5053	2.6772
AR(1)	0.8557	0.7791	0.7857

Table II. Transition Probabilities Returns

This table presents the probability of a forecaster transitioning from a given quartile of the cross-sectional distribution of forecasts to another quartile in the following month, for bond maturities of 2 and 10 years.

	2-year bond				10-year bond				
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	
Q1	0.7542	0.1853	0.0419	0.0186	Q1	0.7433	0.1824	0.0519	0.0224
Q2	0.2012	0.5142	0.2298	0.0548	Q2	0.2084	0.5193	0.2232	0.0491
Q3	0.0437	0.2329	0.5250	0.1984	Q3	0.0480	0.2321	0.5156	0.2042
Q4	0.0129	0.0524	0.2205	0.7142	Q4	0.0119	0.0524	0.2211	0.7146

Table III. Transition Probabilities Macro

This table presents the probability of a forecaster transitioning from a given quartile of the cross-sectional distribution of GDP (left) and CPI (right) forecasts to another quartile in the following month.

	GDP				CPI				
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	
Q1	0.7376	0.1873	0.0542	0.0210	Q1	0.7890	0.1623	0.0387	0.0100
Q2	0.1996	0.5438	0.2071	0.0495	Q2	0.1672	0.6158	0.1891	0.0280
Q3	0.0620	0.2129	0.5520	0.1730	Q3	0.0471	0.1874	0.6080	0.1576
Q4	0.0298	0.0689	0.2051	0.6962	Q4	0.0184	0.0376	0.1883	0.7557

Table IV. Conditional Probabilities Returns vs Macro

This table presents the probability of a forecaster being in a given quartile of the cross-sectional distribution of Macro forecasts (GDP in top panels and CPI in the bottom panels), given that the forecaster is in a particular quartile of the cross-sectional distribution of EBR forecasts, for bond maturities of 2 (left panels) and 10 years (right panels).

		2-year bond				10-year bond				
		Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	
GDP	Q1	0.1947	0.2094	0.2672	0.3286	Q1	0.1897	0.2275	0.2669	0.3158
	Q2	0.2232	0.2780	0.2643	0.2345	Q2	0.2345	0.2589	0.2718	0.2348
	Q3	0.2794	0.2665	0.2715	0.1826	Q3	0.2731	0.2766	0.2555	0.1948
	Q4	0.3687	0.2477	0.2024	0.1813	Q4	0.3652	0.2208	0.2143	0.1997
CPI	Q1	0.1405	0.2143	0.2702	0.3750	Q1	0.1369	0.1930	0.2718	0.3982
	Q2	0.2039	0.2691	0.2990	0.2281	Q2	0.2033	0.2691	0.2993	0.2284
	Q3	0.3153	0.2826	0.2252	0.1769	Q3	0.2907	0.2957	0.2539	0.1597
	Q4	0.4264	0.2326	0.2101	0.1309	Q4	0.4594	0.2318	0.1732	0.1355

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.

Maturity	Q1	Q2	Q3	Q4
2-year	0.41 (4.23)	0.45 (4.55)	0.51 (5.03)	0.55 (5.22)
5-year	0.28 (2.73)	0.35 (3.23)	0.42 (4.01)	0.42 (3.64)
10-year	0.26 (2.91)	0.34 (3.79)	0.41 (4.16)	0.50 (5.06)

Table VI. Transition Probabilities Accuracy

This table 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 bond maturity of 10 years.

	Q1	Q2	Q3	Q4
Q1	56%	28%	12%	4%
Q2	25%	43%	24%	8%
Q3	10%	22%	46%	21%
Q4	5%	7%	19%	70%

Table VII. Joint Accuracy: 10-year vs 3-month

Panel A displays the joint distribution of forecast accuracy for the 10-year EBR and 3-month yield considering all forecasters. Panel B considers only the primary dealers.

Panel A:		10y EBR Acc		
All Forecasters		Good	Average	Bad
	Good	13.67%	11.12%	8.47%
3m Yield Acc	Average	10.96%	13.70%	8.81%
	Bad	8.63%	8.65%	15.99%
Panel B:		10-y EBR Acc		
Primary Dealers		Good	Average	Bad
	Good	18.97%	12.42%	10.28%
3-m yield Acc	Average	10.35%	9.92%	9.20%
	Bad	7.84%	6.63%	14.40%

Table VIII. Spanning of Ex-Ante Accurate Subjective 10-year Bond Return Deciles
Table reports estimates from regressions of spanning regression of deciles of ex-ante accurate subjective expected excess returns on 10-year bonds on the first 5 principle components of the nominal term structure. PC2 is rotated such that a positive shock to this factor implies the slope of the term structure becomes steeper. Deciles are constructed at each point in time based on ranking the sum of the previous years sum of squared forecast errors. EBR_1 denotes the most accurate forecasters while EBR_{10} denotes the least accurate forecasters. 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 January 1989 to January 2014.

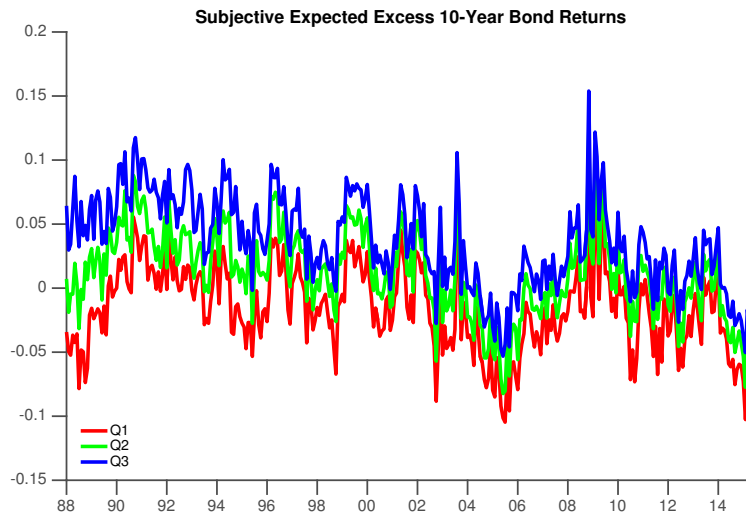
	PC1	PC2	PC3	PC4	PC5	\overline{R}^2
EBR_1	0.00 (3.51)	0.04 (9.42)	-0.00 (-0.04)	-0.29 (-2.62)	-1.91 (-4.00)	36%
EBR_2	0.00 (4.64)	0.03 (8.18)	0.00 (0.05)	-0.27 (-2.81)	-0.60 (-1.59)	30%
EBR_3	0.00 (5.20)	0.03 (6.93)	0.05 (1.58)	-0.20 (-2.20)	-0.66 (-2.03)	28%
EBR_4	0.00 (5.12)	0.03 (7.52)	0.06 (1.80)	-0.15 (-1.48)	-0.21 (-0.64)	25%
EBR_5	0.00 (5.43)	0.02 (5.81)	0.03 (1.04)	-0.10 (-1.10)	-0.37 (-0.94)	22%
EBR_6	0.00 (4.90)	0.02 (4.56)	0.07 (2.11)	-0.00 (-0.03)	-0.35 (-0.86)	17%
EBR_7	0.00 (4.40)	0.02 (4.14)	0.06 (1.74)	-0.02 (-0.15)	0.26 (0.60)	13%
EBR_8	0.00 (4.73)	0.02 (3.64)	0.07 (1.97)	0.04 (0.36)	-0.23 (-0.51)	14%
EBR_9	0.00 (3.98)	0.01 (2.53)	0.08 (2.38)	0.11 (0.96)	0.59 (1.26)	11%
EBR_{10}	0.00 (3.92)	0.01 (2.07)	0.09 (2.12)	0.09 (0.67)	0.06 (0.10)	10%

Table IX. Determinants of Ex-Ante Accurate Subjective 10-year Bond Returns

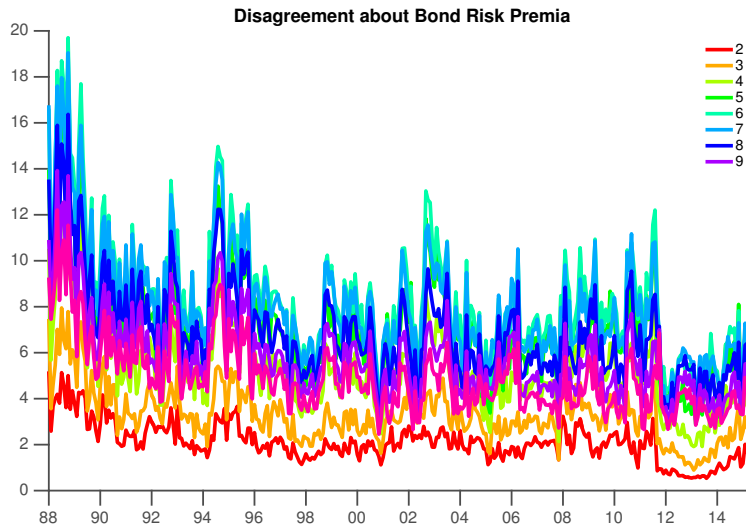
Table reports estimates from regressions of the subjective expected excess returns on 10-year bonds for good forecasters on a set of explanatory variables. These factors are discussed in detail in the main body of the paper. t-statistics, reported in parentheses below the point estimates, are Newey-West corrected. Adjusted R-squared of the regressions are reported in the last column. The sample period is from July 1991 to January 2014.

	<i>DiB(g)</i>	<i>DiB(π)</i>	<i>Liq</i>	<i>Surp</i>	<i>LRR(g)</i>	<i>LRR(π)</i>	<i>TVRP</i>	<i>Jump</i>	$\sigma_y(3m)$	<i>LN</i>	<i>CP</i>	\bar{R}^2
(i)	0.24 (4.23)	0.06 (0.77)										6%
(ii)			-0.27 (-3.00)									7%
(iii)				-0.25 (-3.42)								6%
(iv)					0.40 (5.42)	-0.05 (-0.82)						14%
(v)							-0.24 (-3.75)	0.36 (4.61)	-0.02 (-0.34)			18%
(vi)										0.31 (4.64)	0.14 (1.83)	8%

IX. Figures



(a)



(b)

Figure 1. Cross-Sectional Heterogeneity

The top panel plots the Min, Q1, median, Q3 and max of the cross-sectional distribution of EBR for 5-year maturity bonds. The bottom panel plots the cross-sectional standard deviation of EBR standardized by the full-sample mean EBR.

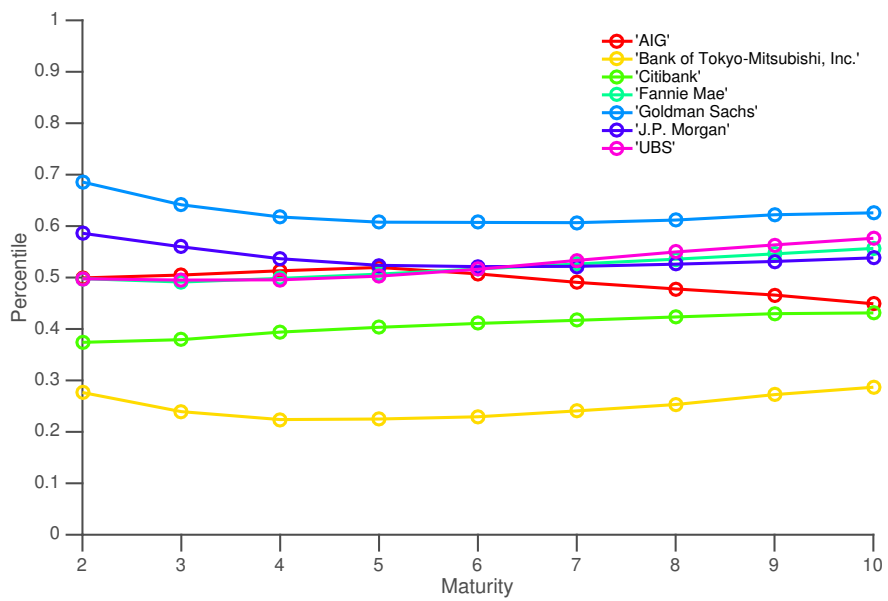


Figure 2. Selected Forecasters' Average Positions

Average position in the cross-sectional distribution of forecasters of four selected forecasters, for bond maturities between 2 and 10 years.

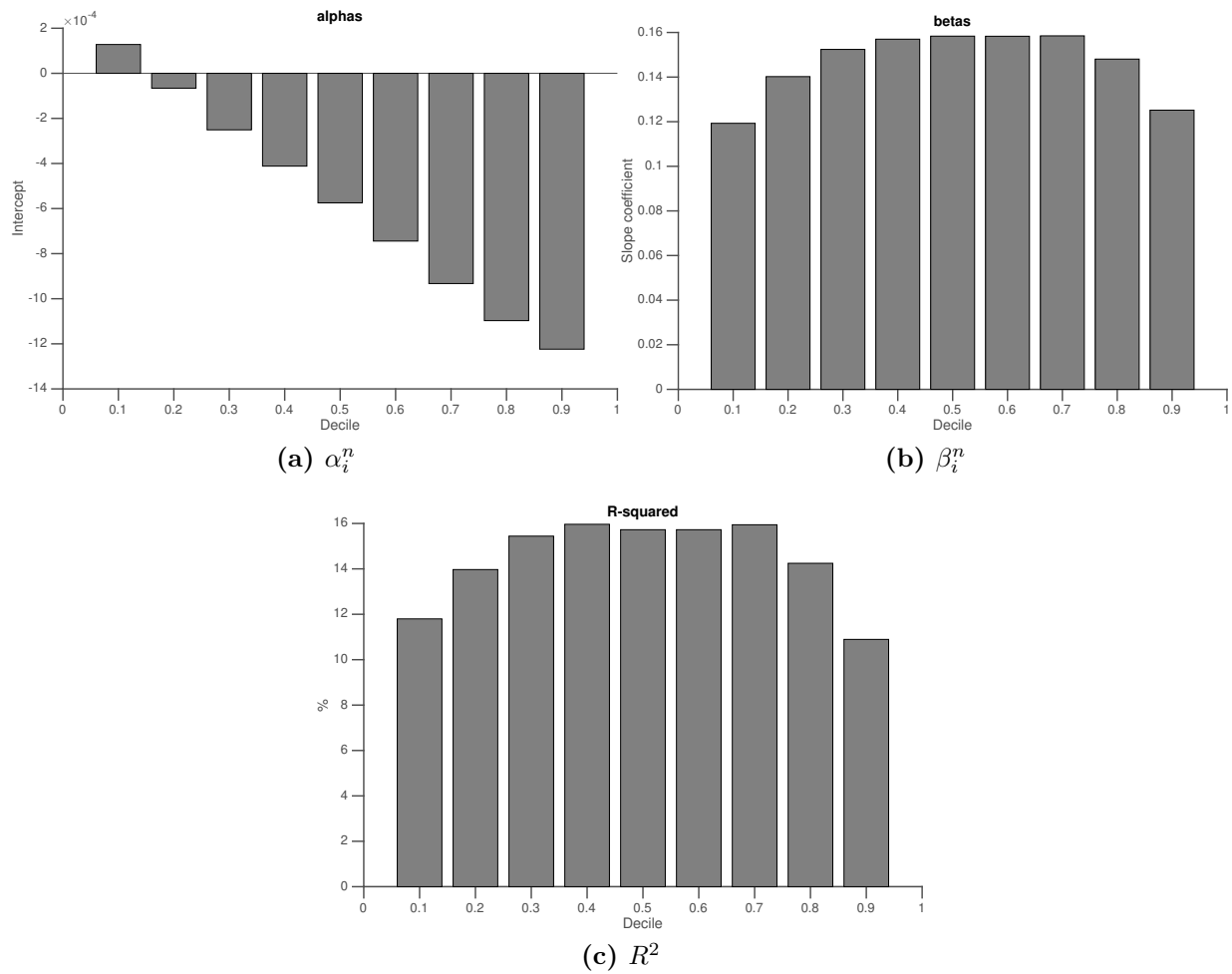


Figure 3. Cross-Section of Short Rate Predictive Regressions

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 percentile i of the cross-sectional distribution of expectations.

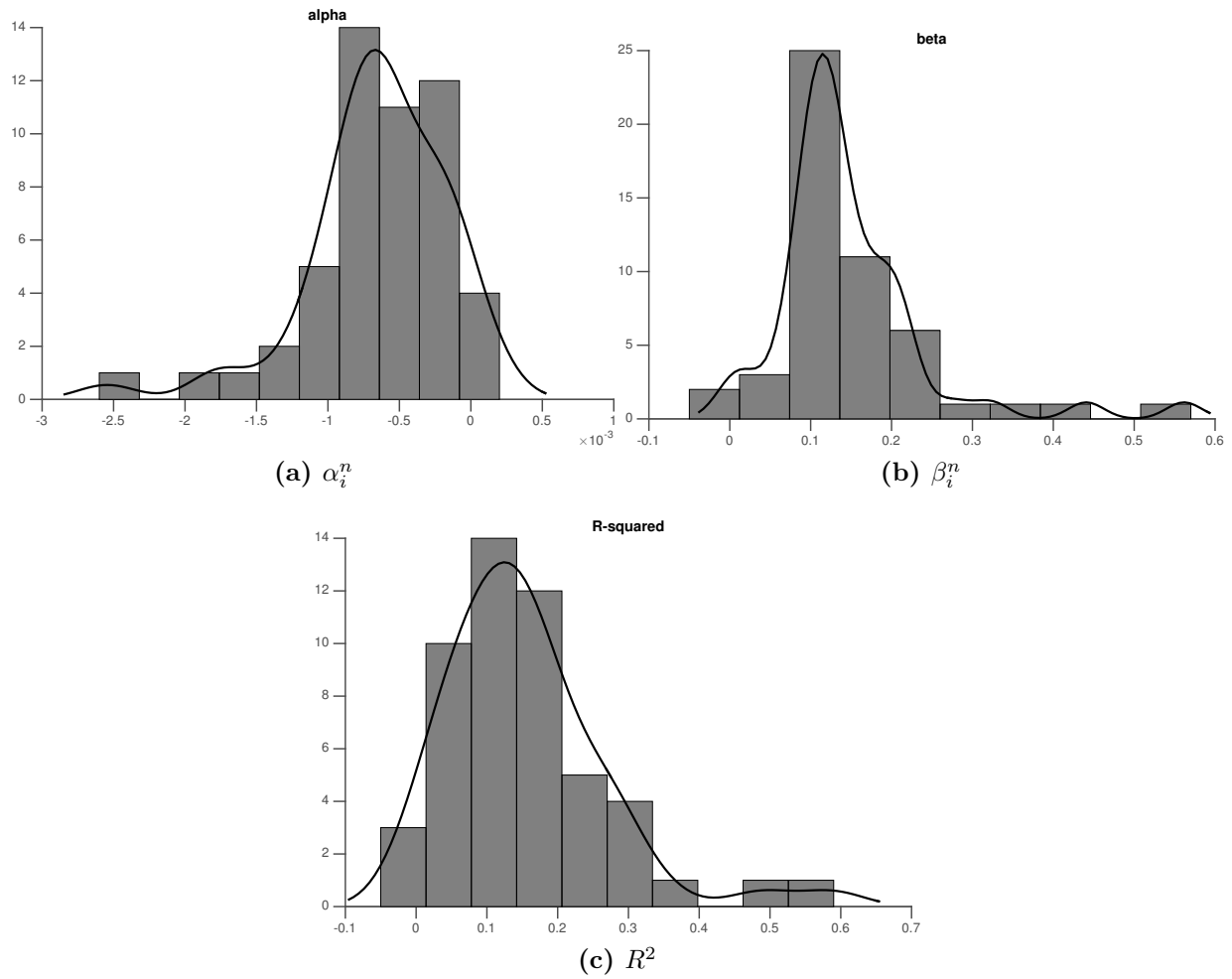


Figure 4. Short Rate Predictive Regressions: Individual Forecasters

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 120 months of forecasts. Solid lines denote kernel density estimates of the cross-sectional distributions.

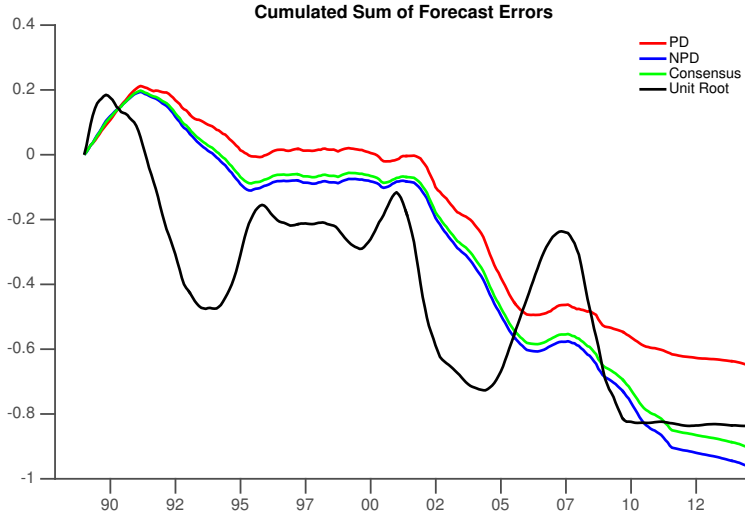


Figure 5. Cumulative 3-month Yield Forecast Errors

Cumulative 3-month yield forecast errors for the average forecaster, i.e. the consensus, primary dealers, non-primary dealers, and a unit root forecast. We consider only contributors with at least 120 months of forecasts.

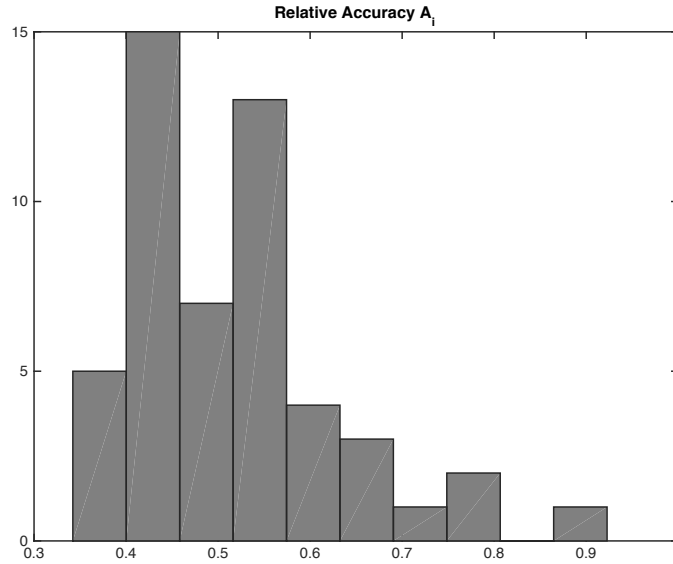


Figure 6. Relative Accuracy

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 unit root benchmark, for the period in which the forecaster is in the panel:

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

We consider only the contributors with at least 120 months of forecasts, for a total of 48 insitutions.

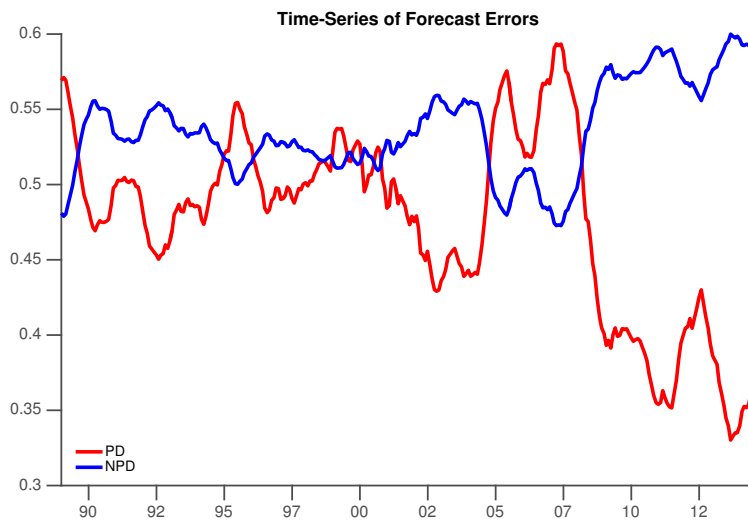


Figure 7. Time Series of 3m Accuracy Percentiles for PD vs NPD

Time series of average accuracy percentiles on the 3-month yield, for primary dealers (PD) and versus all other agents (NPD). The two lines are smoothed using a 12-month moving average.

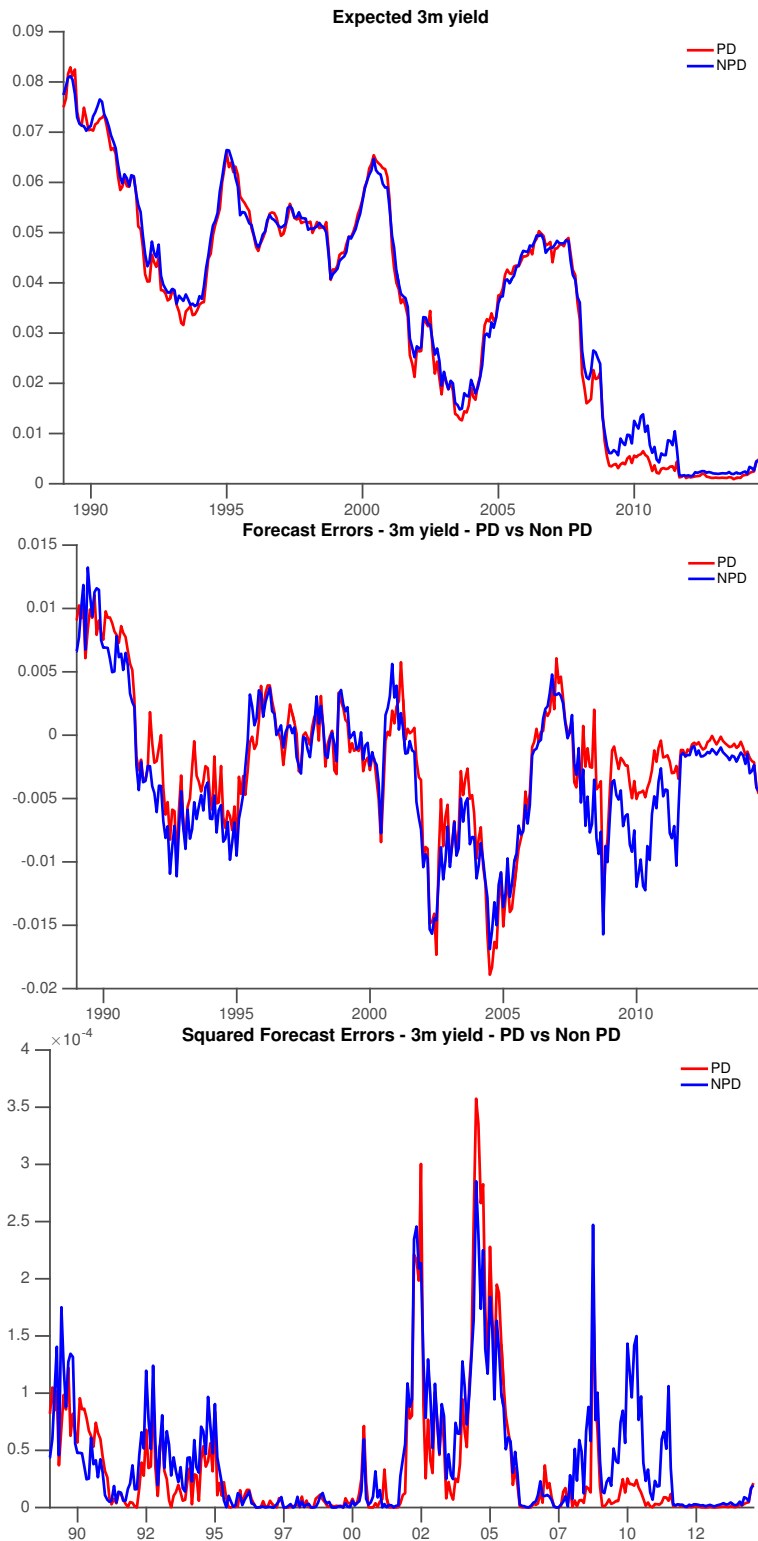


Figure 8. Time Series of expected 3m yield, forecast errors, and squared forecast errors Time series of expected 3-month yield (upper panel) and corresponding forecast errors (middle panel) and squared forecast errors (bottom panel) for primary dealers (PD) and all other agents (NPD).

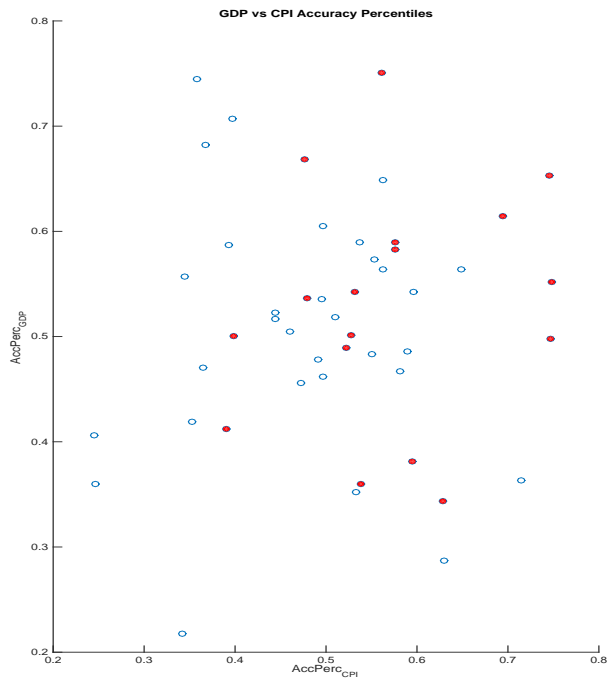


Figure 9. Average Percentiles of Individual Macro Accuracy

Comparison of the average percentiles of individual forecasters accuracy, measured as the squared forecast errors, for real GDP and CPI. The red dots correspond to the primary dealers.

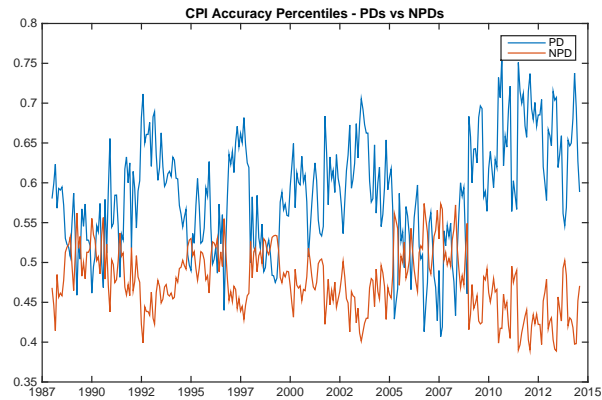
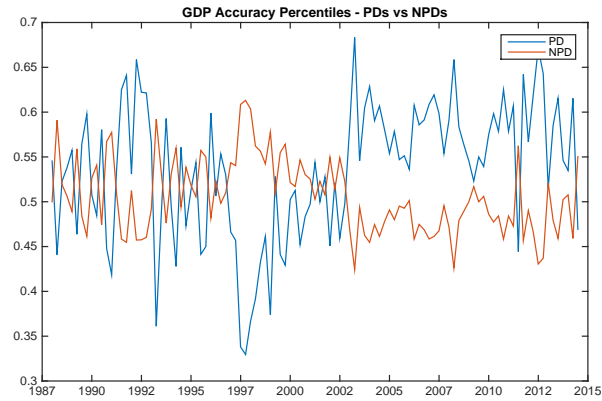


Figure 10. Time Series of Macro Accuracy Percentiles for PD vs NPD
 Time series of average accuracy percentiles on the Real GDP growth (upper panel) and CPI growth (bottom panel), for primary dealers (PD) and all other agents (NPD).

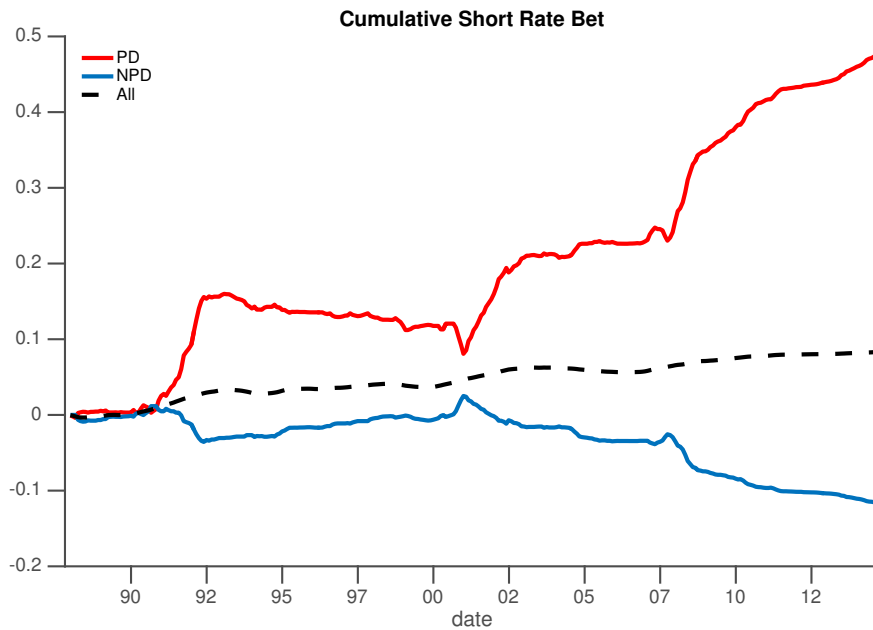


Figure 11. Cumulative Returns on Short Rate Bet for PDs vs NPDs

Cumulative returns on a short rate bet for the average primary dealer (PD) and non primary dealer (NPD). Every month, agents in the left tail of the distribution of 3-month yield expectations go long the 2-year bond and short the 1-year bond, and hold the position for a year. Agents in the right tail of the distribution of 3-month yield expectations do the opposite. We average the returns over PDs and NPDs and plot their cumulative returns assuming a bet is placed every month. The dashed black line denote the cumulative returns on a short rate bet for the average forecaster.

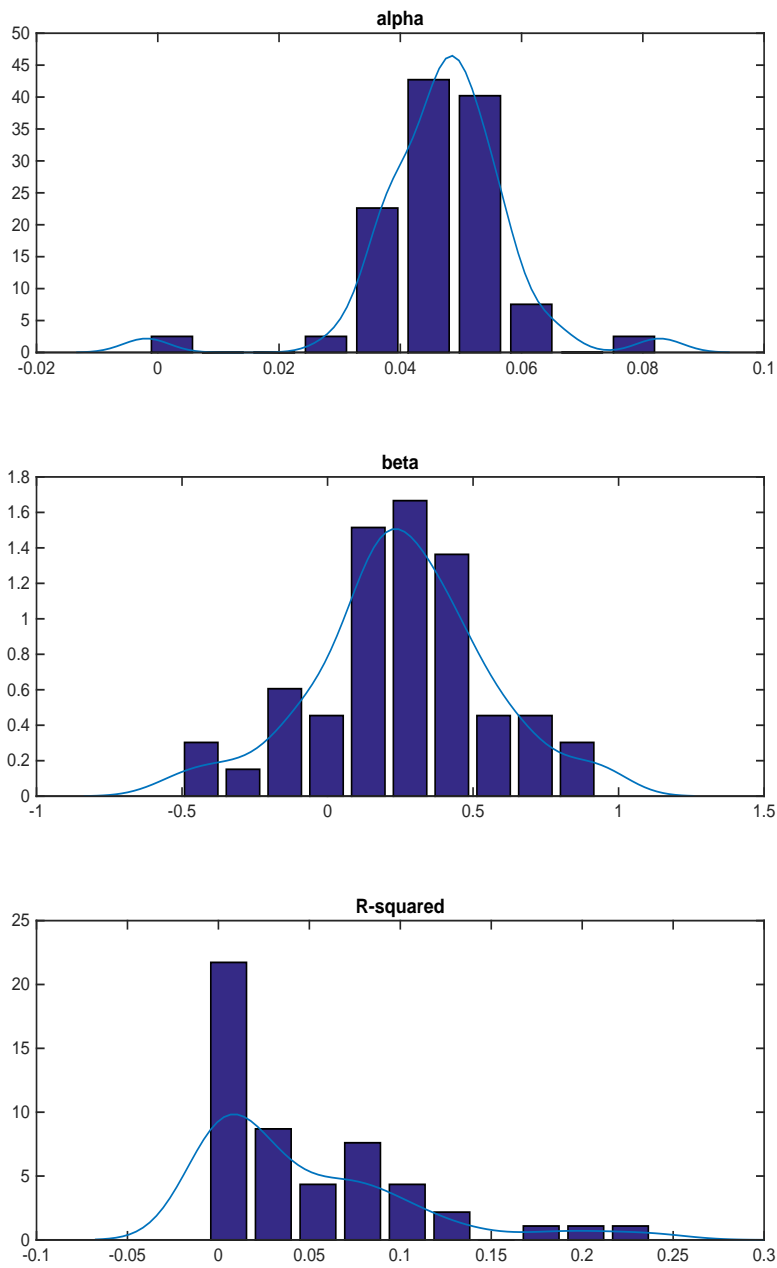


Figure 12. Predictive Regressions Individual Forecasters

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 120 months of forecasts:

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

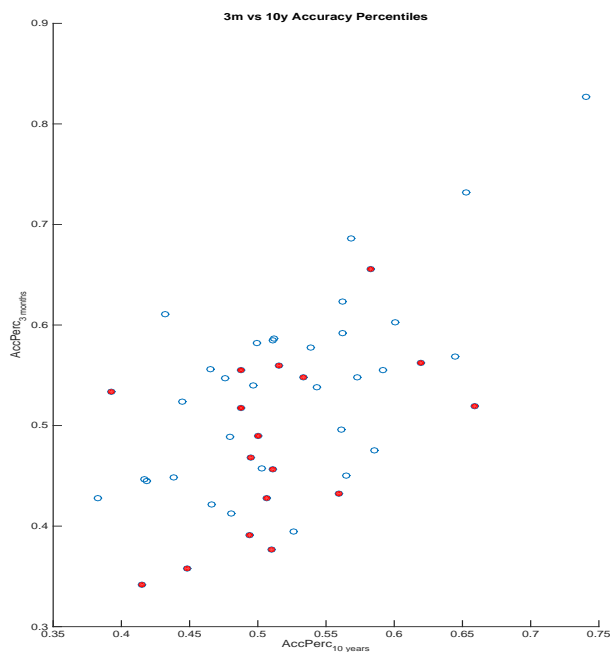


Figure 13. Average Percentiles of Individual Accuracy: 10y EBR vs 3m Yield
 Comparison of the average percentiles of individual forecasters accuracy, measured as the squared forecast errors, for 10-year EBR and 3-month yield. The red dots correspond to the primary dealers.