Saïd Business School
Research Papers

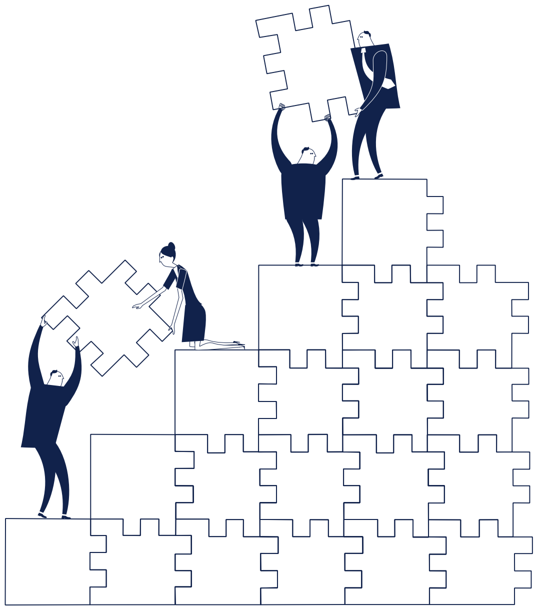
January 2017

Expected Term Structures

Andrea Buraschi
Imperial College Business School

Ilaria Piatti
Saïd Business School, University of Oxford

Paul Whelan
Copenhagen Business School



Saïd Business School RP 2016-36

Expected Term Structures

Andrea Buraschi Ilaria Piatti Paul Whelan

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 realizations 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 evaluate several reduced-form and structural models. We find support for heterogeneous beliefs models and also uncover a number of statistically significant relationships in favour of alternative rational expectations models once the effect of heterogeneous beliefs is taken into account.

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

This version: January 2017

Andrea Buraschi is Chair of Finance at Imperial College Business School; Ilaria Piatti is at Saïd Business School, University of Oxford; and Paul Whelan is at Copenhagen Business School. We thank Christian Eyerdahl-Larsen, Anna Obizhaeva, and the participants of the AFA meeting 2017, Chicago, the CICF meeting 2016, Xiamen, the FMA conference 2016, Helsinki, the 2016 Financial Econometrics and Empirical Asset Pricing Workshop at the University of Lancaster, the 2016 European Summer Symposium in Financial Markets, and the seminar participants of Green Templeton College, Oxford University, Said Business School, University of Porto and Goethe University Frankfurt for valuable comments. Paul Whelan acknowledges financial support from the Center for Financial Frictions (FRIC). The usual disclaimer applies. Emails: `andrea.buraschi@imperial.ac.uk`, `ilaria.piatti@sbs.ox.ac.uk`, `pawh.fi@cbs.dk`.

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. Koijen, 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, 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]

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 expected excess bond returns (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

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

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

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 risky bets on future economic states. We also find clear evidence of persistence in agents expected bond risk premia. For example, a forecaster in the first quartile of the cross-sectional distribution of 2-year EBR has a probability of about 75% to stay in the first quartile the following month, and this probability is about 74% for the 10-year EBR. This is 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 important questions about the common assumption of identifying the marginal investor with the agent with average (consensus) expectations. Notwithstanding the previous heterogeneity, overall expectations about bond returns display significant 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 a 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 changes its stance and aggressively reduces short-term rates. While this takes by surprise all agents, whose expected excess bond returns are downward biased in these subperiods,[4], the bias is smaller for primary dealers. This is consistent either with primary dealers having superior information about Fed's implementation of monetary policy or, more simply, with an information flow advantage originating from their role as market makers in Treasury bonds. The result is quite important given 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

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

[4]This 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.

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

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 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 52% for the most accurate portfolio of agents and only 23% for the least accurate one. This result is consistent with the market selection hypothesis in competitive markets. Indeed, while optimists 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

[6] Koijen, 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.

of future realized returns on lagged state variables. We revisit this approach and instead of using methodologies based on econometric projections, we use subjective expectations as directly revealed in real time 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, thus avoiding the forward looking bias which often affects traditional predicting regressions methods. However, instead of the consensus, a better measure 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, in conjunction with belief heterogeneity. We show that disagreement matters and we find supporting evidence for rational expectation explanations of expected bond returns, once the effect of heterogeneous beliefs is carefully taken into account. 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 expected term structures. In Section IV we study the forecasting properties of expected short-term interest rates. Section V discusses the dynamics of the expected bond excess returns (EBR), the predicting power of EBR for future realized excess returns and the cross-sectional variations in the forecast accuracy. Section VI analyzes 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 one year yield is $rx_{t+1}^n = r_{t+1}^n - y_t^1$, with the gross return being defined as $r_{t+1}^n = p_{t+1}^{n-1} - p_t^n$.

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

$$erx^n_{i,t} = -(n-1) \times \underbrace{E^i_t\left[y^{n-1}_{t+1}\right]}_{\substack{\text{Survey Yield}\\ \text{Forecasts}}} + ny^n_t - y^1_t. \tag{1}$$

Forecasts on future long-term interest rates depend on both expectations on future short-term interest rates $E^i_t\left[y^1_{t+s}\right]$ and future bond risk premia $erx^n_{i,t}$. 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 the following hypothesis:

$H^{(1)}_0$: *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 examining both 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 and tests:

$H^{(2)}_0$: *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, the advantage of our approach is to avoid data aggregation and assumptions about specific forecasting models 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 of their predictive power for future realized returns, we use direct measures of expected returns. We focus on testing the effect on risk premia of belief heterogeneity, and we evaluate the marginal contribution of the factors implied by homogeneous models of risk premia.

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 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

[7] In 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.

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 5 years (60 monthly observations) the number of participants is always above 30. On the other hand, in the SPF the distribution of respondents displays significant variability: the mean number of respondents is around 40, the standard deviation is 13 and in some years the number of contributors is as low as 9. While in the early 70s the number of SPF forecasters was around 60, it decreased in two major steps in the mid 1970s and mid 1980s to as low as 14 forecasters in 1990.[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.

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

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.

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

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

[10] Specifically, we search for the parameters which solve $b_t^j = \arg\min_b \sum_{h=1}^{H_t^j} \left[\left(P^h(b) - P_t^h \right) \times \frac{1}{D_t^h} \right]^2$,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 $h = 1, ..., H_t^j$, 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 $\tau_2 > 0$.

III. The Cross Section of Expected Term Structures

A. Subjective Expectations

Figure 1 gives a first look at the data. Each panel plots quartiles (Q1, Q2(median) and Q3) of the 1-year cross-sectional distribution of expectations.[11]

Consider first short rate, GDP growth and consumption price index growth (inflation) expectations. Casual inspection suggests the time series for all quartiles display rationally anticipated characteristics. For example, GDP growth was expected to be low in precisely the years which the NBER subsequently defined as recessions.[12] Projections for inflation, on the other hand, display a trend over time that was, in fact, subsequently realized.

Moreover, comparing macro versus short rate projections, subjective expectations appear internally consistent, at least visually. For example, between the years 1988 and 1990 inflation was expected to be increasing. At the same time forecasters expected a Federal Reserve policy to combat inflation (high nominal short rates) but that this policy would have a contractionary effect on the real economy (GDP growth). A policy of this type is often called a Taylor rule although such a relationship also arises both in standard Keynesian models or Lucas type consumption based models. Consider now the top left panel, which is the subjective expected excess returns on a 10-year bond. Consistent with the predictions of many structural models, subjective bond risk premia are counter-cyclical, meaning they are negatively correlated with expectations about real growth. For example, expected returns were increasing in the early part of the sample, decreasing in the high growth rate years between the dot-com bubble and the financial crisis, and spiking again around Lehman Brother collapse.

[Insert Figure 1 here.]

B. Subjective Bond Risk Premia

We document a large unconditional heterogeneity in the cross section of EBR point forecasts. Table I provides summary statistics for the median, the first quartile, and the 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.

[11] 1-year average expectations are computed from 4 and 5 quarter ahead projections.

[12] Our sample period covers three recessions: July 1990 - March 1991, March 2001 - November 2001 and December 2007 - June 2009.

The conditional properties of the cross-sectional distribution of EBR display rich dynamics in the time series. The top left panel of Figure 1 shows the Q1, median, and Q3 of 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 indeed be summarized by the consensus. In order to do this, we compute the interquartile range (IQR) of the cross-sectional distribution of EBR, as the difference between Q3 and Q1, for all bond maturities $n = 2, \ldots, 10$, and then regress it on the consensus forecast for the corresponding bond maturity. The slope coefficients of these regressions are positive, and statistically significant for all maturities, but the variations in the consensus forecasts explain only around 3% of the variation in the IQR. Moreover, we can strongly reject the hypothesis that the IQR is constant. In fact, the slope coefficient of a regression of IQR on its 1-year lag is significantly different from zero, for all maturities and at all levels. Therefore, the dispersion in beliefs varies over time and it is not merely a scaled version of the consensus: the mean is not a sufficient statistics for the cross section of expectations.

The top panel of figure 2 highlights the time variation in heterogeneity by plotting the cross-sectional standard deviation of EBR standardized by the full sample mean EBR, for bond maturities 2, 5 and 10-year. The figure also shows that the dispersion in beliefs is state-dependent: it tends to rise at the onset of recessionary periods and drop again as the economy recovers.[13] It is interesting to note that disagreement about long term EBRs is 5 orders of magnitude larger than disagreement about short rates or disagreement about the macro economy (bottom panel of Figure 2). However, disagreement in non-monotonic in maturity displaying a 'hump-shaped' around the 5-year maturity.

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

[Insert Table I and Figure 2 here.]

C. *Belief Persistence*

Figure 2 also demonstrates that disagreement about short rates, bond returns, and the macro economy are all persistent. This raises an interesting question: is disagreement a result of dogmatic beliefs or some alternative behavioural bias or information friction?

In order to address this question we first rank all forecasters according to whether in a given month t their forecast is in the first, second, third or fourth quartile of the cross-sectional

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

distribution. We repeat this exercise for all months in the sample and compute transition probabilities: the probability that forecasters in a given quartile at time t stay in that particular quartile in $t+1$ or move to a different quartile of the distribution.

We do this first for short rates and macro expectations. If views are not persistent, all the entries in these transition matrices should be approximately equal to 25%. Instead, we find that the diagonal elements are significantly higher than 25%, in particular for the most extreme quantiles, Q1 and Q4 where they are always above 70%. This result is striking and even stronger than what Patton and Timmermann (2010) document for macroeconomic forecasts using data from the Consensus Economics Inc, at a quarterly frequency.

[Insert Table II here.]

The question of belief persistence is particularly important in the context of bond pricing models since whether agents are persistently optimistic or pessimistic about returns determines whether bonds hedge or are risky bets on consumption or inflation risk. In the first case, bonds should earn a negative risk premium, in the second expected bond risk premia should be positive. Thus, we ask whether agents show persistence in their beliefs about bond risk premia: are individual forecasters persistently in one particular quartile of the cross-sectional distribution of subjective $EBRs$?

Figure 3 plots the time series average of seven individual forecasters' positions in the cross-sectional distribution of subjective expected bond returns, for maturities between 2 and 10 years. This plot shows that agents are consistently optimistic or pessimistic across maturities. 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 3 that some institutions, like Goldman Sachs, have been persistent in their forecasts about larger than average excess bond returns at all maturities; others have been persistent in their forecasts of negative excess bond returns. Table III addresses this question more formally but computing transition probabilities matrices for subjective excess returns. The results suggest that forecasters have persistent beliefs about bond risk premia, relative to the consensus excess return. 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%.

[Insert Figure 3 and Table III here.]

D. Internally Consistent Beliefs

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

We find that agents who are marginally more optimistic or pessimistic about macroeconomic variables are consistently in one particular quartile of the cross-sectional distribution of short term interest rates, as shown in Table IV. If one focuses on the corners of this table, we find that analysts who forecast lower short-term interest rates are also those forecasting lower GDP growth and, at the same time, lower CPI inflation. For instance, 35% of those who are in the first quartile of the distribution of future short-term interest rate forecasts are also in the first quartile of the distribution for GDP growth forecasts; similarly, 41% of those who are in the first quartile of the distribution of future short-term interest rate forecasts are also in the first quartile of the distribution for CPI inflation forecasts. This relation between forecasts at the individual level is consistent with the idea that good states of the economy are generally characterised by increasing yields, at least at short maturity. At the same time, the pattern is not deterministic, suggesting that beliefs on interest rates and the macroeconomy (GDP and inflation) are not driven by a single factor.

Table V repeats this exercise for long term bond returns. Agents who are in the highest quartiles of the distribution of forecast for inflation and GDP are also in the highest quartiles of the distribution of forecasts of returns. This suggests that agents beliefs are broadly consistent with the rational expectation requirement that agents forecasts interest rates in accordance to the sign of the correlation between short term rates and macro economic variables.

[Insert Tables IV and V here.]

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

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 $i = 0.10, \ldots, 0.90$ of the cross-sectional distribution of three-month yield forecasts:

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

where $\Delta y_{t+1}^{3m} = y_{t+1}^{3m} - y_t^{3m}$. Figure 4 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 3^d decile; the slope coefficients are positive and significant for all deciles of the distribution. The values are very close to one and in five cases are not significantly different from one. The R^2 vary between 8% and 13%, 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 almost 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.

[Insert Figure 4 here.]

To investigate the characteristics of this heterogeneity, we use a unique advantage offered by our dataset which provides 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 5 years (60 months) of forecasts. Figure 5 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 few forecasters, most estimated slope coefficients are positive and statistically significant. This suggests that professional forecasters do a relatively good job in predicting future short rates. Second, a few forecasters are extremely accurate, with slope coefficients larger than 0.5 and R^2 in excess of 10%, with some agents producing an R^2 in excess of 30%. This is contrary to the evidence based on retail individuals and non-professional forecasters.

At the same time, a very intriguing property of the regression coefficients is that the cross-sectional distribution of intercepts is largely skewed towards negative values: α_i^{3m} is negative for 79 of the 84 forecasters and significantly different from zero for slightly less than half of the forecasters. This suggests that the average forecaster has been surprised by the extended decline in short term interest rates over our sample period.

[Insert Figure 2 here.]

Figure 6 shows this bias explicitly by plotting the cumulative 3-month yield forecast errors over time for the average forecaster:

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

for $t = 1, \ldots, T$ and where $fe_{cons}^{3m}(t) = y_{t+1}^{3m} - E_t^{cons}\left[y_{t+1}^{3m}\right]$. Indeed, $U_{cons}^{3m}(t)$ is not a martingale and has a negative drift, which is reflected in the negative α in the predictive regression (2). In particular, we see a drastic increase in cumulative errors in the early 90s, in the early 2000s and during the recent financial crisis.

It is also interesting to compare the drift in cumulative survey forecast errors to alternative predictions. The pink line in Figure 6 corresponds to the forecasts an econometrician would obtain estimating a VAR model in real-time based on a rolling window that includes 10-years of data. In this case cumulative forecast errors are significantly larger than professional forecasters in the first half of the sample but subsequently revert: the VAR makes persistent negative forecast errors followed by persistent positive forecast errors. This suggests that the bias in professional forecasters expectations is fundamentally different than that obtained from a real-time VAR.

The black line in figure 6 plots the cumulative forecast errors one would obtain from the $12M \rightarrow 15M$ forward rate. This shows that forward rates are consistently biased, which is consistent with the existence of a interest rate risk premium. Per se, this is not surprising and it is consistent with an extensive literature that documents deviations from the expectation hypothesis: forward rates are expected future interest rates adjusted for risk. The red line adjusts for this risk at each prediction date t by looking back and subtracting the rolling historical spread between forward rates and three month yields. The cumulative error in this case is close to being a martingale. Comparing risk adjusted versus physical forward rate forecasts we note an interesting observation. The bias in surveys is remarkably correlated and bounded by these measures. It appears as if agents are not forecasting under the physical measure but instead forecasting under a measure which is distorted towards the risk neutral one.

[Insert Figure 6 here.]

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

$$RMSE_i^{3m}(Surv) = \sqrt{\frac{1}{T_i - t_{0,i} + 1} \sum_{t=t_{0,i}}^{T_i} \left(y_{t+1}^{3m} - E_t^i \left[y_{t+1}^{3m} \right] \right)^2},$$

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 7 displays the distribution of $\mathcal{A}_i$ for the 84 contributors with at least 5 years of monthly forecasts. Noticeably, a significant mass of individual forecasters have $\mathcal{A}_i$ between 0.90 and 1.10, suggesting that several agents are as good as the unit-root benchmark (and significantly better than a simple VAR model). Moreover, some agents are extremely precise with $\mathcal{A}_i < 0.90$, suggesting that some prefessional forecasters can provide reasonably good measures of expected bond returns. At the same time, some agents are very poor forecasters with $\mathcal{A}_i > 1.20$.

[Insert Figure 7 here.]

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	J.P. Morgan
3	BMO Capital Markets
4	Nomura Securities Inc.
5	Bank of America
6	Georgia State University
7	Crestar Financial Corp.
8	US Trust Company
9	Chase Manhattan Bank
10	Woodworth Holdings

Interestingly, the first five institutions in this list (and 7 out of the first 10), are currently primary dealers, or have been primary dealers at least once in our sample period, even if overall only 23 of the 84 financial institutions with at least 10 years of forecasts are or have been primary dealers.[14] 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 43% of the primary dealers are in the first tercile, 29% in the second and 28% 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 8 shows the time series of average accuracy percentiles for primary dealers (PD) versus all other contributors (NPD), smoothed by computing a 12-month moving average of the monthly accu-

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

racies. It is clear that PD have a comparative advantage, and this advantage seems indeed to be stronger in specific time periods. The following subsection addresses this issue more formally by analyzing the conditional individual forecast accuracy.

[Insert Figure 8 here.]

C. Conditional forecast accuracy

Figure 9 (upper panel) shows the time series of expected 3-month yield of both PDs and NPDs. The bottom panel shows the corresponding forecast errors. A pattern immediately emerges: the average expectations for PDs and NPDs are 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 short rates are biased upward in these subperiods, they seem to surprise primary dealers less, and this is especially true during the recent financial crisis.

[Insert Figure 9 here.]

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.[15] 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%.

[15]Results are robust to the choice of time periods for the moving average.

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 periods, expectations of the two sets of forecasters, as well as forecast errors, are very similar.

D. 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 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 a rolling trading strategy in which agents take positions every month and hold these positions until maturity (i.e. one 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 10.

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 9), 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.

[Insert Figure 10 here.]

E. 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 11 shows 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.[16] 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.

[Insert Figure 11 here.]

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

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

predictive ability of primary dealers on short-term rates lead to an advantage for 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 directly using the dynamics of expectations that are obtained in a model independent way. Given the time series of subjective bond risk premia $erx^n_{i,t+1}$, 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:

$$erx^n_{i,t+1} = \alpha^n_i + \beta^n_i erx^n_{i,t} + \epsilon^n_{i,t+1}. \tag{4}$$

The results are summarized in Table VI and show that the slope coefficients are significantly different from zero 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.

[Insert Table VI here.]

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

$$rx_{t+1}^n = \alpha_i^n + \beta_i^n erx_{i,t}^n + \epsilon_{i,t+1}^n. \tag{5}$$

Figure 12 shows the distribution of regression coefficients and R^2 of regression (5) for each forecaster. The results show that notwithstanding heterogeneity in accuracy, a few forecasters are extremely accurate with slope coefficients close to one and R^2 larger than 20%. The correlation between expectations and future realization of excess bond returns is positive for 69 out of 84 forecasters.

[Insert Figure 12 here.]

This *positive* relation between expectations and realizations is the opposite to what Greenwood and Schleifer (2014) document in the context of the stock market, and to what Koijen, Schmeling, and Vrugt (2015) find in the context of global equities, currencies and global fixed income returns across countries.[17] 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 show that agents beliefs are substantially more rational than previously 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

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

They range between 7.5013 and 15.8325. Since individual forecasters may appear in the sample at different times, we assess their accuracy relative to a model. We consider 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.

[17] Koijen, 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.

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 displays 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 to 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 recursively from the first principle component of a vintage macro panel and denote 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, this 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 the 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 $\mathcal{A}_i^n$:

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

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

[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.

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 0.55 and 1.5 for all forecasters, and similar results hold for the LN factor.[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 accuracy tends to be persistent. To quantify the persistence, we rank all forecasters according to their accuracy in month t within the distribution of all forecasters 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 VII 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, is very persistent. For example, a forecaster in the first quartile of the cross-sectional distribution of 10-year EBR accuracy has a probability of 58% to stay in the first quartile of accuracy the following month. This probability is 70% for the 4^{th} 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%.

[Insert Table VII here.]

This confirms two conclusions. First, expectations of a significant fraction of professional forecasters are far from being irrational. Second, surveys can be used to build reliable measures of bond risk premia. However, one needs to be mindful of the heterogeneity in the distribution of beliefs. The assumption that consensus can be used as a sufficient statistics 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 was positive one could conclude that long-term bond returns are mainly driven by

[19] We 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.

short rates 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. The two rankings are highly correlated, in fact a regression of the 10-year accuracy percentiles on the 3-month accuracy has a significant slope coefficient of 0.42 and an adjusted R-squared of 21%. However, the link is less strong if we focus on the subsample of primary dealers: the regression coefficient is 0.37 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	Huntington National Bank
5	RidgeWorth Capital Management
6	Fleet Financial Group
7	DePrince & Associates
8	The Northern Trust Company
9	GLC Financial Economics
10	J.W. Coons & Associates

Contrary to our findings for the short rate, only two of the top ten forecasters for the 10-year bond returns are primary dealers. To analyze the performance of primary dealers on the long end of the term structure more formally, we compute, at each month t, 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 36% of the primary dealers are in the first tercile, 30% in the second and 34% 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 VIII 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 VIII, 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 68.2%. Even after distinguishing periods of increasing and decreasing rates or using the Mann-Whitney test, we cannot reject the null hypothesis with p-values larger than 50%. Overall, primary dealers have a significantly better predictive performance only for the short rate.

[Insert Table VIII here.]

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 behavioral finance literature argues that in presence of short-selling constraints marginal agents ought to be those holding optimistic beliefs about expected returns (see e.g. Scheinkman and Xiong (2003) and Hong, Sraer, and Yu (2013)). Since pessimists cannot short-sell, their beliefs are not revealed (spanned) by equilibrium asset prices. The general equilibrium literature 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 (2010), Ehling, Gallmeyer, Heyerdahl-Larsen, and Illeditsch (2015), among others). This argument, consistent with the original "market selection hypothesys" by Friedman (1953) and Alchian (1950), implies that bond prices should span the beliefs of the most accurate agents (i.e. closest to the actual physical probability). As Alchian (1950) argues, "*Realized profits [...] are the mark of success and viability. It does not matter through what process of reasoning or motivation such success was achieved. The fact of its accomplishment is sufficient. This is the criterion by which the economic system selects survivors: those who realize positive profits are the survivors; those who suffer losses disappear.*" If some agents have been consistently more accurate than others, they would have been accumulating more economic weight in the pricing kernel. Thus, these beliefs, rather than the consensus ones, should be 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 tercile portfolios. Then, we compute the average EBR within each tercile. 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 terciles of the accuracy distribution onto the first five principal components of the term structure, which

[20]The 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.

[21]Note 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:

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

Table IX reports the results of this regressions where EBR1 denotes the most accurate and EBR3 the least accurate beliefs. We find a monotonic link between accuracy and degree of spanning, measured as the adjusted R-squared of the regression.[22] 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 much smaller.

[Insert Table IX here.]

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 44%, versus 52% 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 and we find an R^2 of only 31%. 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

[22]The shape of the link between accuracy and spanning is qualitatively robust to the number of accuracy portfolios considered, i.e. if we use quartiles or deciles of the accuracy distribution instead of terciles.

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.[23]

Direct measures of subjective expectations can address these three problems. They provide a useful way 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.

In our setting we can directly address all these issues by running regressions of our direct measure of risk premia on alternative model-implied specifications of risk premia.

Another important practical issue with the evaluation of structural models of bond risk premia relates to the presence of disagreement. If it true that there is heterogeneity, should homogeneous models just be abandoned, or do properties of this heterogeneity help explain bond risk premia above and beyond homogeneous models?

Heterogeneity in standard models with different beliefs has two main dimensions: disagreement and sentiment. The heterogeneous beliefs literature in fact shows how theoretically bond risk premia are affected by the interaction of both, see e.g. Buraschi and Whelan (2010). In particular, Jouini and Napp (2006) show that the market price of risk depends on the level of disagreement in a state dependent way: the impact of heterogeneous beliefs on the market price of risk is positive if the aggregate belief is pessimist, and negative if it is optimist, where optimism means that the aggregate growth rate of the representative agent is lower than under the objective probability. Therefore, sentiment drives the effect of disagreement.

We capture this effect by including aggregate disagreement in the regression of bond risk premia on models and letting its regression coefficient depend directly on sentiment:

[23]Examples 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.

$$
\begin{aligned}
erx_{i,t}^n &= a_i^n + (b_{1,i}^n s_t + b_{2,i}^n) DiB(g)_t + b_{3,i}^n \mathcal{M}_t^j + \epsilon_{i,t}^n \\
&= a_i^n + b_{1,i}^n s_t DiB(g)_t + b_{2,i}^n DiB(g)_t + b_{3,i}^n \mathcal{M}_t^j + \epsilon_{i,t}^n,
\end{aligned} \tag{7}
$$

where $\mathcal{M}_t^j$ denotes a model-implied specification of bond risk premia, and $DiB(g)_t$ and s_t are the two dimensions of heterogeneity in standard models with different beliefs, i.e. disagreement and sentiment. This specification allows us to test directly whether disagreement matters and the potential marginal effect of the risk premium factors dictated by rational expectation models with homogeneous economies.

We obtain measures for $erx_{i,t}^n$ 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. We construct a measure of sentiment (pessimism) as the difference between a physical expectation of GDP growth computed from an econometric projection of realized GDP growth and the growth expectation of our proxy for the marginal agent (see Figure 13).[24] Our proxy for disagreeement about real growth rates $DiB(g)_t$ is from Buraschi and Whelan (2010).

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.

Structural Models

- In models where agents agree to disagree, the stochastic discount factor is a direct function of disagreement. While disagreement about growth and sentiment are considered explicitly in regression (7), consistent with Buraschi and Whelan (2010), other disagreement models such as Ehling, Gallmeyer, Heyerdahl-Larsen, and Illeditsch (2015) and Hong, Sraer, and Yu (2013), argue about the importance of inflation disagreement. We denote their proxies for inflation disagreement as $DiB(\pi)$.

- 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).

[24] Our chosen time series model for the econometrician's measure is a simple AR(4) on quarterly realized GDP growth, at the 1-year horizon. This is considered the benchmark model for forecasting US GDP growth in Marcellino (2008), among others. However, our findings are robust to alternative constructions of the sentiment measure.

- 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 $\mathcal{M}_t$, 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.[25]

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

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.[26] 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, Sabtchevsky, 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

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

[26]Such 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)).

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 X shows the results of regression (7) for 10-year bond using the most accurate tercile of forecasters, and alternative specifications of $\mathcal{M}_t$. To summarize, we find that disagreement matters, independently of the model we consider, but only conditional on sentiment (b_1 is significant while b_2 is not). In fact, the interaction term $s_t \cdot DiB(g)_t$ is always highly significant and with a positive coefficient, consistent with the idea that disagreement increases risk premia in periods of pessimism and viceversa.[27]

Interestingly, several traditional rational expectation measures of risk premia, based for example on liquidity and long run risk, are also significant when we explicitly take into account disagreement. While they do not decrease the significance of the interaction of disagreement and sentiment, their marginal contribution in explaining our direct measure of bond risk premia is important, based on the increase in the regression R-squared. It is also important to note that factor loadings for the significant models are consistent with the theory behind those factors, as discussed above.

Overall, we find that several of the structural models are indeed consistent with subjective EBR once heterogeneity is properly taken into account. 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.

First, the liquidity factor of Fontaine and Garcia (2012) is significantly negatively correlated with $erx^n_{i,t}$, consistent with the interpretation that negative shocks to this factor are bad news for funding conditions, thus raising expected returns. The liquidity factor Liq with sentiment and disagreement explains 32% of the variation in 10-year subjective bond risk premia.

Second, 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 $erx^{10}_{i,t}$, in conjunction with heterogeneity, with adjusted R^2 of 23%. However, only real uncertainty is significant entering with a positive loading. This is

[27] Another way to test this conditional effect of disagreement is to run a regression of $erx^n_{i,t}$ on $DiB(g)_t$ and $\mathcal{M}_t$, separately in periods of positive and negative sentiment. We find that disagreement has a significantly positive effect when our representative agent is pessimistic about consumption growth, and an insignificant, mostly negative effect on risk premia in periods of optimism. However, since realized GDP growth is available only quarterly, these two separate regressions have a very low number of observations. In our sample, there are only around 40 quarters in which the marginal agent is optimistic.

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.

Third, the surplus factor implied by habit formation models is not significantly linked to bond risk premia when disagreement is taken into account. It is interesting to note that surplus appears to be significant when heterogeneity is not included in the regression. This might be due to the fact that both disagreement and surplus are meant to capture changes in the price of risk, while uncertainty in long run risk models is related to the quantity of risk.

Fourth, when we examine generalized affine volatility models, we do not find a significant relationship between volatility and subjective returns.

Finally, when we study reduced-form predictive models of bond returns, we find a significant positive relationship between EBR and the real time CP return forecasting factor. The explanatory power of real time LN is much weaker and has the wrong sign.

[Insert Table X here.]

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, once the properties of belief heterogeneity are carefully taken into account. 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 positive for a large fraction of forecasters, 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 focus on testing the effect on risk premia of belief heterogeneity, and we evaluate the marginal contribution of other factors implied by rational expection models with homogeneous economies. We show that disagreement always matters, but only conditional on sentiment, consistent with the idea that disagreement increases risk premia in periods of pessimism, as predicted by standard models with heterogeneous beliefs. Moreover, we find supporting evidence for several rational expectation explanations of risk premia, when we explicitly take into account the effect of disagreement. 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* (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, Andrea, and Paul Whelan, 2010, Term structure models and differences in beliefs, *Working paper.*

Buraschi, A., and P. Whelan, 2012, Term structure models and differences in belief, *Imperial College, Working Paper.*

Campbell, J.Y., and J.H. Cochrane, 1999, By force of habit: A consumption-based explanation of aggregate stock market behavior, *Journal of political Economy* 107, 205–251.

Chen, H., S. Joslin, and N. Tran, 2012, Rare disasters and risk sharing with heterogeneous beliefs, *Review of Financial Studies* 25, 2189–2224.

Cieslak, A., 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 Illeditsch, 2015, Disagreement about inflation and the yield curve, *Working paper. University of Pennsylvania.*

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.

Marcellino, M., 2008, A linear benchmark for forecasting gdp growth and inflation, *Journal of Forecasting* 27, 305–340.

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, 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

Q1	2 Year	5 Year	10 Year
Mean	−0.03	−0.95	−1.66
Std Dev	0.00	0.02	0.03
Min	−0.01	−0.06	−0.10
Max	0.01	0.03	0.11
Skew	−0.06	−0.09	0.01
Kurtosis	2.49	2.67	3.17
1st Lag Auto	0.79	0.75	0.74
Q2	2 Year	5 Year	10 Year
Mean	0.28	0.34	1.06
Std Dev	0.00	0.02	0.03
Min	−0.01	−0.04	−0.08
Max	0.02	0.04	0.12
Skew	0.05	−0.06	−0.03
Kurtosis	2.34	2.70	3.07
1st Lag Auto	0.82	0.75	0.76
Q3	2 Year	5 Year	10 Year
Mean	0.56	1.46	3.57
Std Dev	0.01	0.02	0.04
Min	−0.01	−0.03	−0.05
Max	0.02	0.06	0.15
Skew	0.23	−0.02	0.04
Kurtosis	2.11	2.51	2.68
1st Lag Auto	0.86	0.78	0.79

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).

	3M				GDP				CPI			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Q1	72%	21%	5%	1%	74%	19%	5%	2%	79%	16%	4%	1%
Q2	22%	51%	23%	4%	20%	54%	21%	5%	17%	62%	19%	3%
Q3	5%	21%	54%	19%	6%	21%	55%	17%	5%	19%	61%	16%
Q4	2%	5%	22%	71%	3%	7%	20%	70%	2%	4%	19%	76%

Table II. Transition Probabilities Short Rates and 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.

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

Table III. Transition Probabilities Subjective Excess 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.

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

Table IV. Conditional Probabilities Short Rates vs Macro

This table presents the probability of a forecaster being in a given quartile of the cross-sectional distribution of Macro forecasts given that the forecaster is in a particular quartile of the cross-sectional distribution of 3 month yield forecasts.

		2-year bond				10-year bond			
		Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
GDP	Q1	19%	21%	27%	33%	19%	23%	27%	32%
	Q2	22%	28%	26%	23%	24%	26%	27%	23%
	Q3	28%	27%	27%	18%	27%	28%	26%	19%
	Q4	37%	25%	20%	18%	37%	22%	21%	20%
		Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
CPI	Q1	14%	21%	27%	38%	14%	19%	27%	40%
	Q2	20%	27%	30%	23%	20%	27%	30%	23%
	Q3	32%	28%	23%	18%	29%	30%	25%	16%
	Q4	43%	23%	21%	13%	46%	23%	17%	14%

Table V. 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).

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

Table VI. 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.

	Q1	**Q2**	**Q3**	**Q4**
Q1	58%	27%	11%	4%
Q2	25%	44%	24%	7%
Q3	9%	22%	47%	21%
Q4	5%	7%	19%	70%

Table VII. 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.

Panel A:		**10y EBR Acc**		
All Forecasters		Good	Average	Bad
	Good	15%	11%	8%
3m Yield Acc	Average	11%	13%	9%
	Bad	8%	9%	16%
Panel B:		**10-y EBR Acc**		
Primary Dealers		Good	Average	Bad
	Good	19%	13%	10%
3-m yield Acc	Average	11%	10%	9%
	Bad	6%	7%	15%

Table VIII. 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.

	PC1	PC2	PC3	PC4	PC5	$\overline{R}^2$
EBR_1	0.00	0.02	-0.02	0.06	-0.15	52%
	(8.52)	(11.94)	(-2.62)	(1.91)	(-1.22)	
EBR_2	0.00	0.01	-0.02	-0.02	0.07	45%
	(10.68)	(8.32)	(-3.43)	(-0.53)	(0.58)	
EBR_3	0.00	0.01	-0.02	-0.07	0.16	23%
	(6.68)	(4.08)	(-2.18)	(-1.71)	(0.92)	

Table IX. Spanning of Ex-Ante Accurate Subjective 10-year Bond Return Terciles

Table reports estimates from regressions of spanning regression of terciles 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. Terciles 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_3 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 July 2015.

	$S \cdot DiB(g)$	$DiB(g)$	$DiB(\pi)$	Liq	$Surp$	$LRR(g)$	$LRR(\pi)$	$TVRP$	$Jump$	$\sigma_y(3m)$	LN	CP	$\overline{R}^2$
(i)	0.32	0.09											12%
	(2.82)	(0.85)											
(ii)	0.35	0.11	−0.26										18%
	(3.08)	(1.05)	(−4.00)										
(iii)	0.39	−0.12		−0.48									30%
	(3.98)	(−1.29)		(−4.50)									
(iv)	0.34	0.08			0.05								11%
	(3.09)	(0.78)			(0.53)								
(v)	0.38	−0.06				0.27	−0.35						21%
	(3.02)	(−0.55)				(2.83)	(−2.92)						
(vi)	0.27	0.08						−0.11	0.18	−0.18			14%
	(2.95)	(0.81)						(−0.62)	(1.37)	(−1.74)			
(vii)	0.36	0.15									−0.14	0.24	19%
	(3.21)	(1.49)									(−1.18)	(2.46)	

Table X. 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 December 2014.

IX. Figures

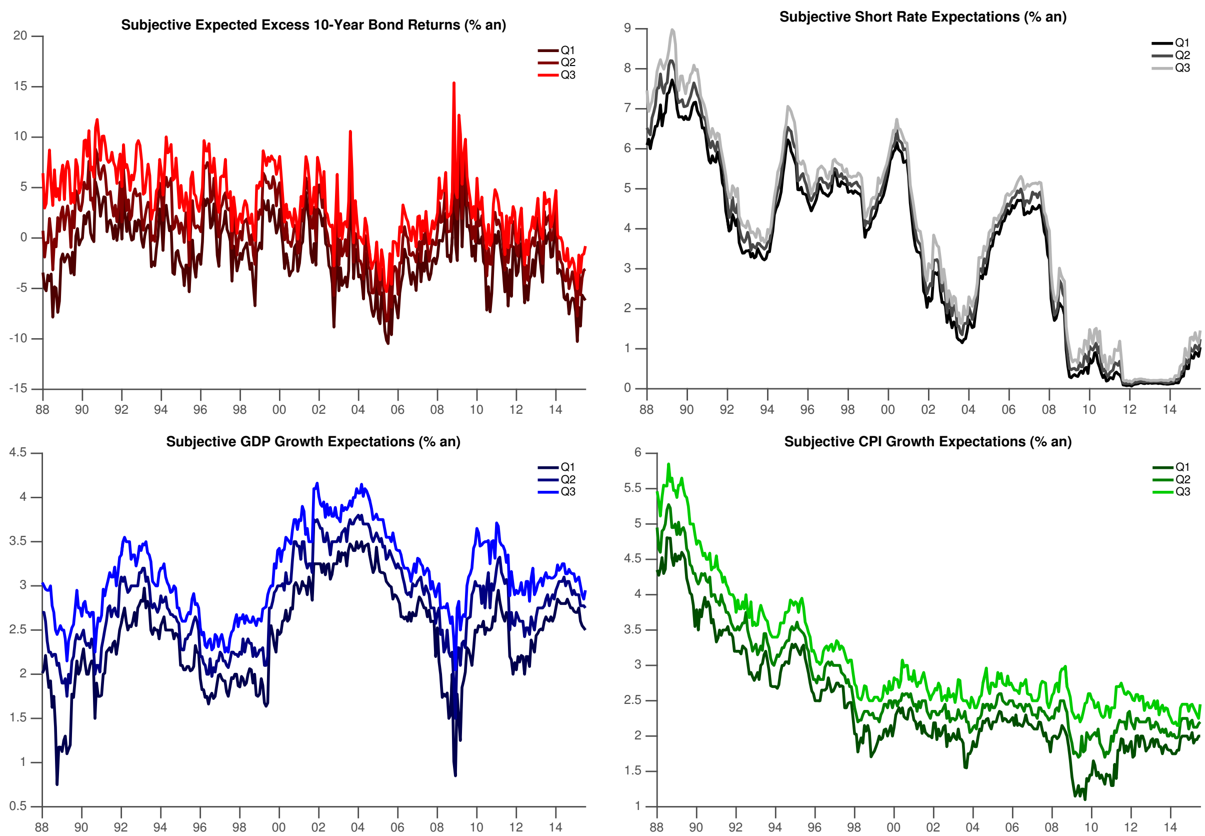


Figure 1. Subjective Expectations

Each panel plots quartiles (Q1, Q2(median) and Q3) of the cross-sectional distribution of expectations. Top Left: 1-year subjective excess returns for 10-year maturity bonds. Top Right: subjective 3-month Treasury yield expectations. Bottom Left: subjective GDP growth expectations. Bottom Right: subjective CPI growth expectations.

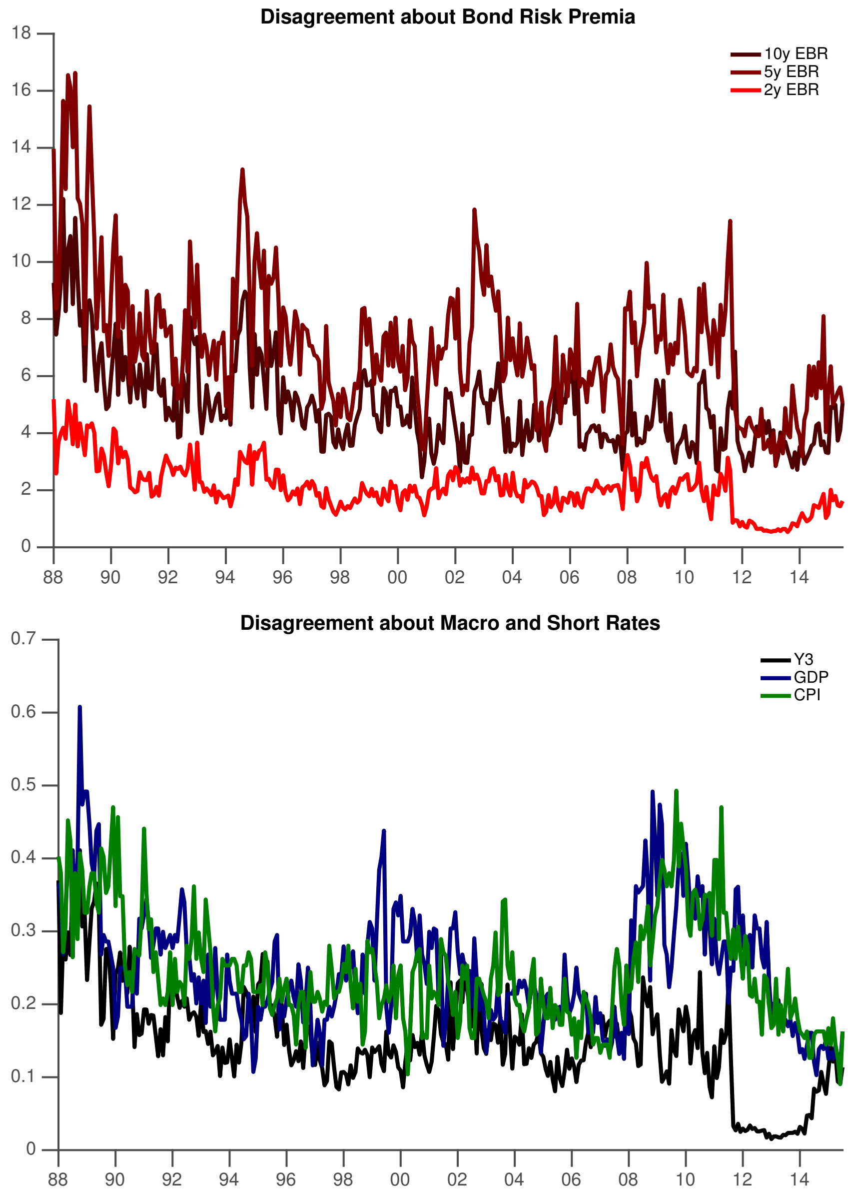


Figure 2. Disagreement about Returns vs Short Rates vs Macro

Top panel plots disagreement about expected bond returns for maturities 2 , 5 and 10-year. Bottom panel plots disagreement about 3-month Treasury yields, GDP and CPI growth. Disagreement is defined as the cross-sectional interquartile range of subjective expectations standardized by the full-sample consensus expectation.

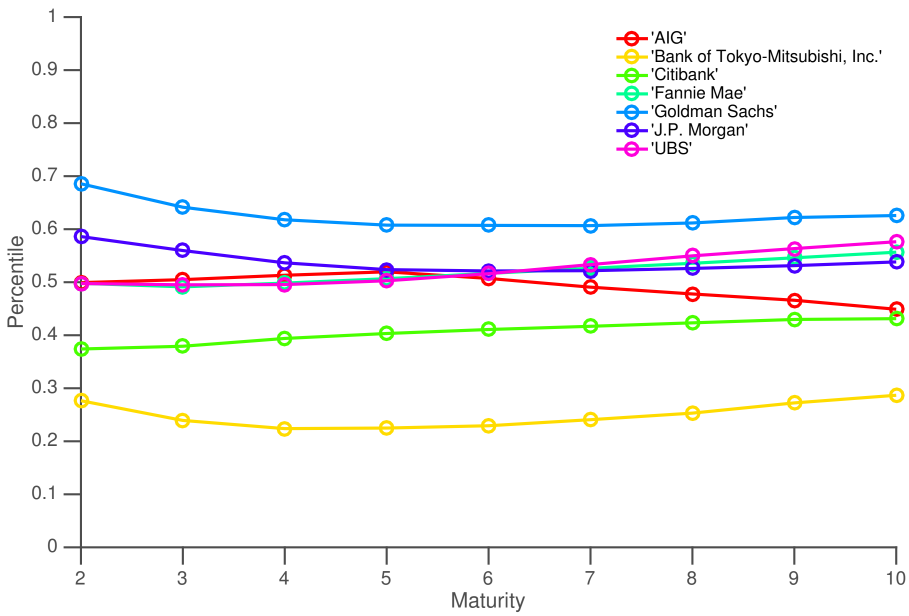
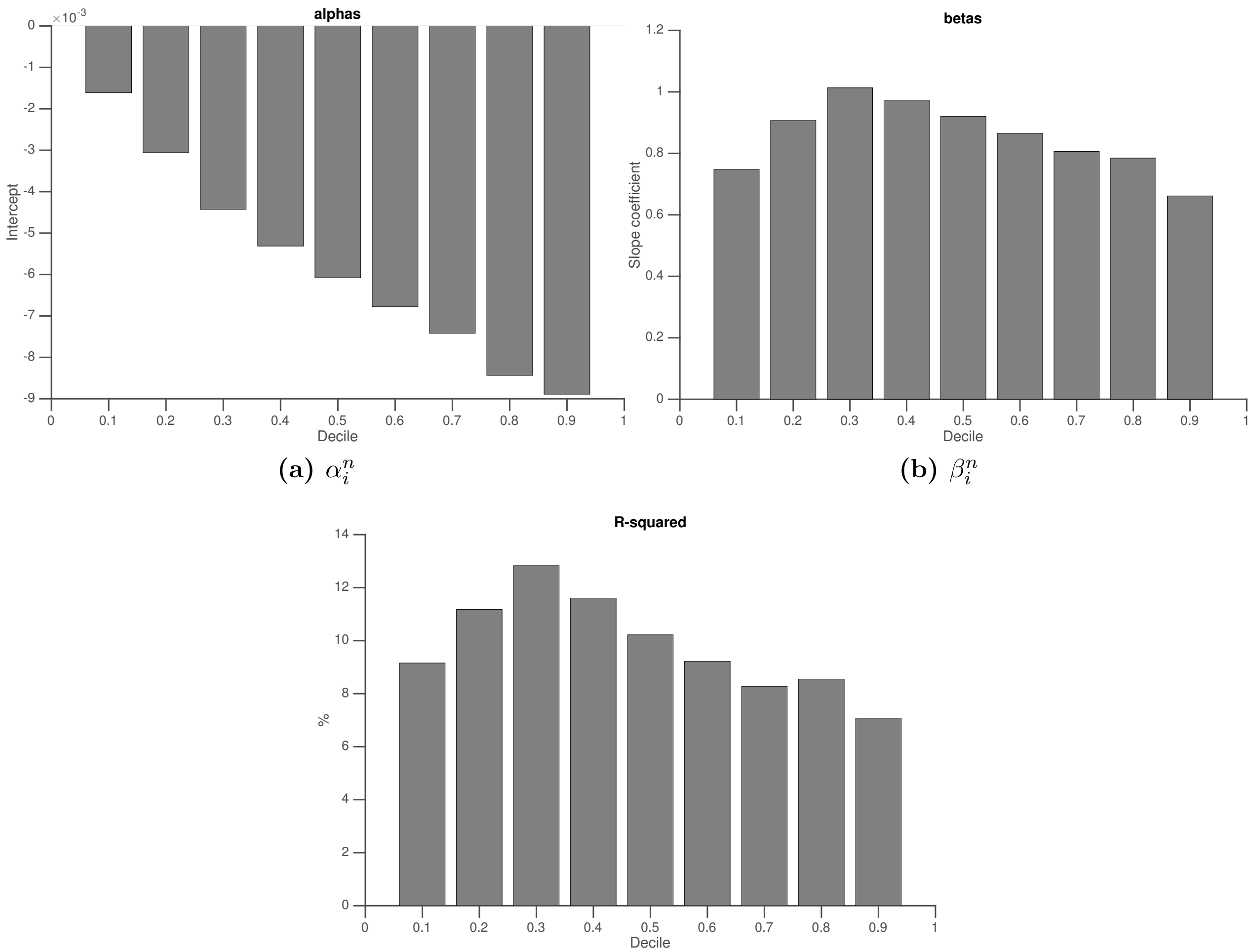


Figure 3. Selected Forecasters' Average Positions
Average position in the cross-sectional distribution of forecasters of seven selected forecasters, for bond maturities between 2 and 10 years.



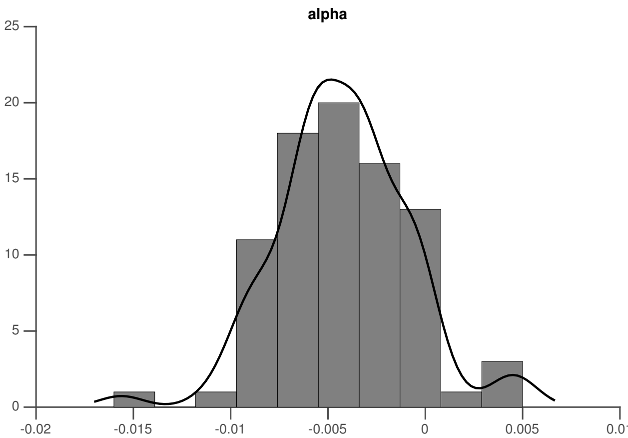
(a) α_i^n

(b) β_i^n

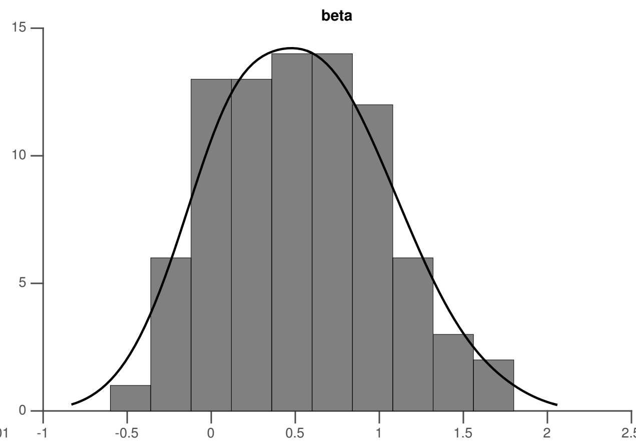
(c) R^2

Figure 4. Cross-Section of Short Rate Predictive Regressions

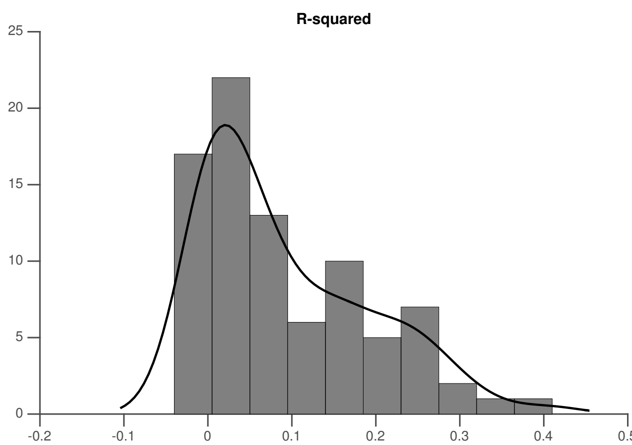
Estimated regression coefficients and adjusted R^2 of regressions of the change in realized 3-month yield on the expected change in 3-month yield for percentile i of the cross-sectional distribution of expectations.



(a) α_i^n



(b) β_i^n



(c) R^2

Figure 5. Short Rate Predictive Regressions: Individual Forecasters
Estimated regression coefficients and adjusted R^2 of regressions of the change in realized 3-month yield on the expected change in 3-month yield for all individual contributors with at least 60 months of forecasts. Solid lines denote kernel density estimates of the cross-sectional distributions.

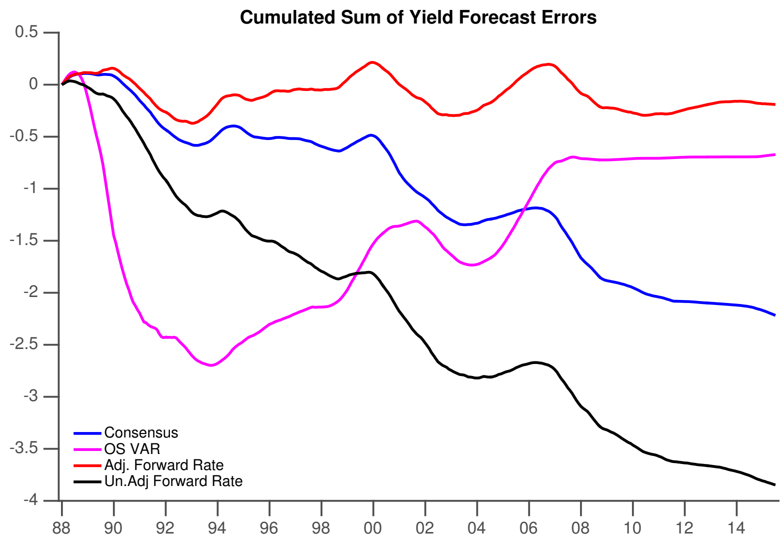


Figure 6. Cumulative 3-month Yield Forecast Errors
Cumulative 3-month yield forecast errors for the average forecaster, i.e. the consensus, an out-of-sample VAR, the forward rate, and the forward rate adjusted by the past average spread between forward rates and realized yields.

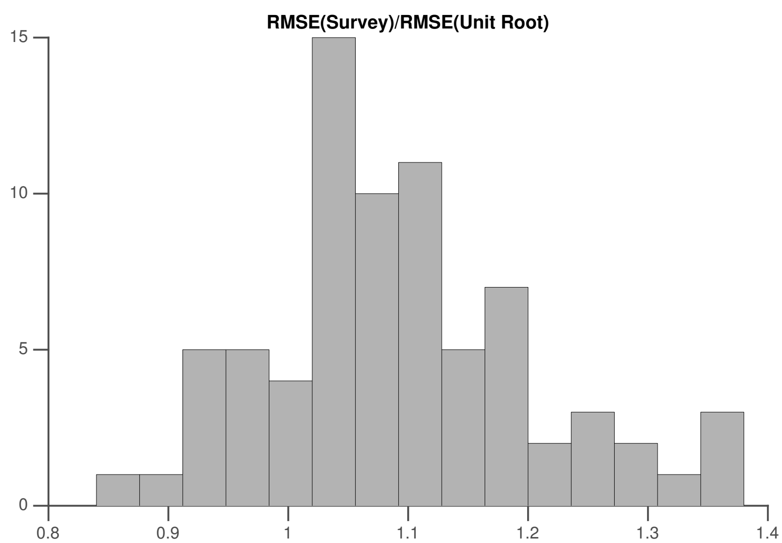


Figure 7. Relative Accuracy
Histogram of the relative accuracy $\mathcal{A}_i$ of each forecaster, that is the ratio between the RMSE of each individual forecaster and the RMSE of a unit root benchmark, for the period in which the forecaster is in the panel:

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

We consider only the contributors with at least 60 months of forecasts, for a total of 84 insitutions.

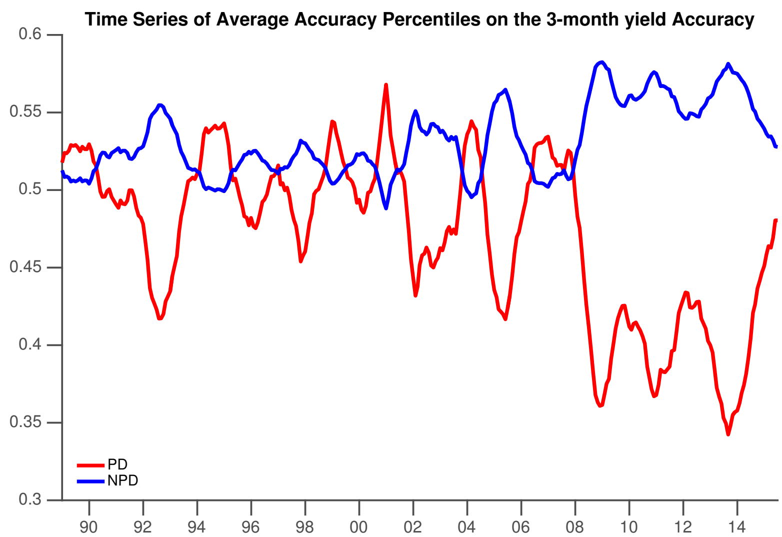


Figure 8. Time Series of Short Rate Accuracy Percentiles for PD vs NPD
Time series of average accuracy percentiles for 3 month Treasury yields for primary dealers (PD) and all other agents (NPD).

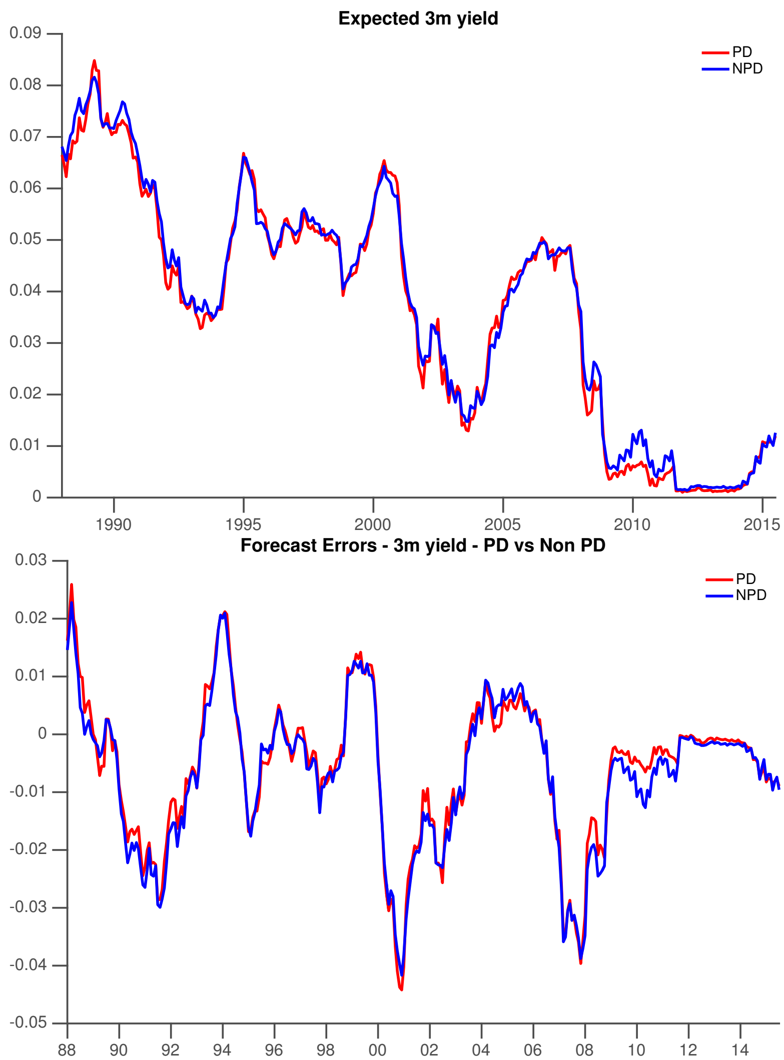


Figure 9. Time Series of expected 3m yield and forecast errors
Time series of expected 3-month yield (top panel) and corresponding forecast errors (bottom panel) for primary dealers (PD) and all other agents (NPD).

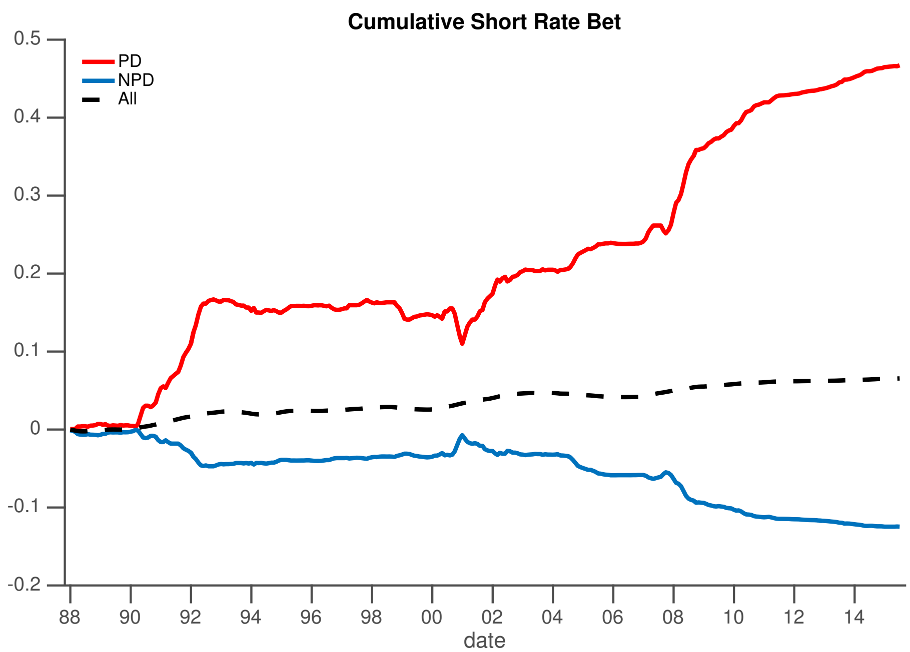
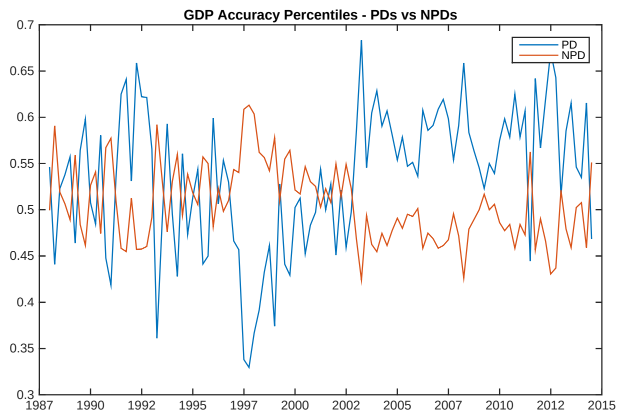


Figure 10. Cumulative Returns on Short Rate Bet for PDs vs NPDs
Cumulative returns on a short rate bet for the average primary dealer (PD) and non primary dealer (NPD). Every month, agents in the left tail of the distribution of 3-month yield expectations go long the 2-year bond and short the 1-year 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.



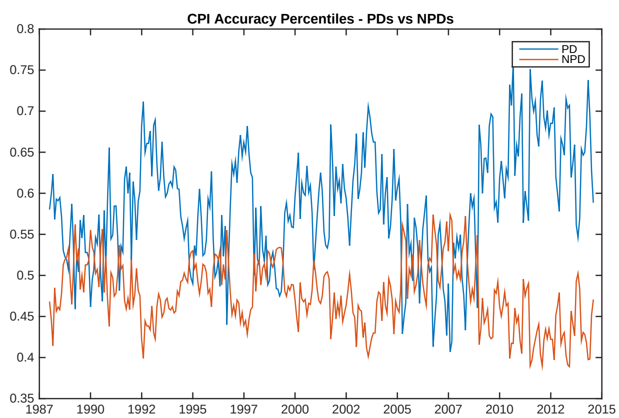
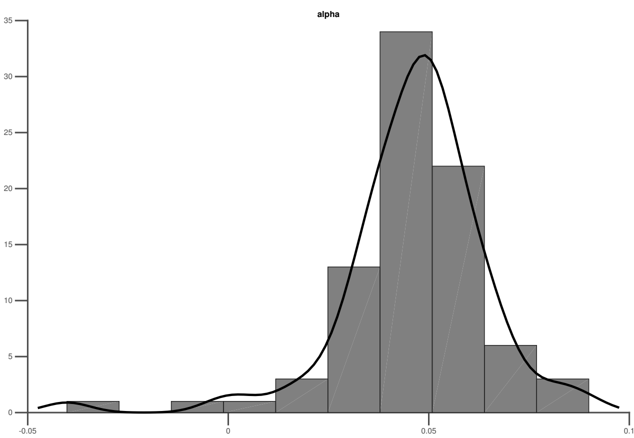
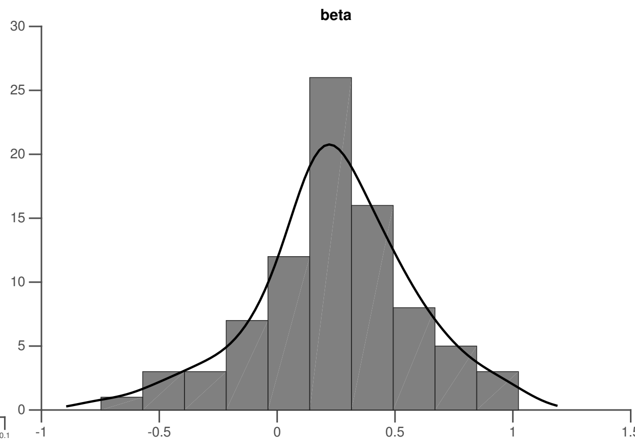


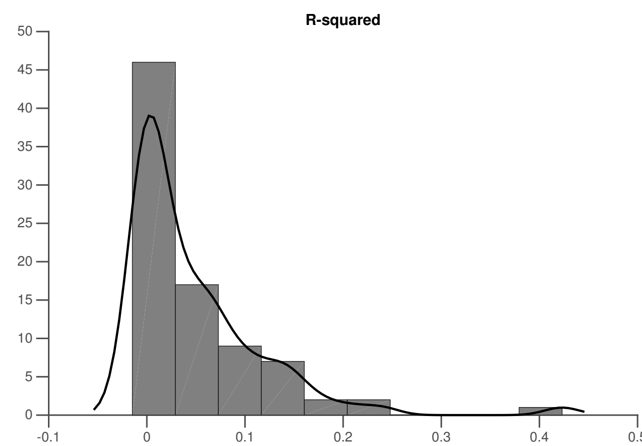
Figure 11. Time Series of Macro Accuracy Percentiles for PD vs NPD
Time series of average accuracy percentiles on the Real GDP growth (upper panel) and CPI growth (bottom panel), for primary dealers (PD) and all other agents (NPD).



(a) α_i^n



(b) β_i^n



(c) R^2

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 60 months of forecasts:

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

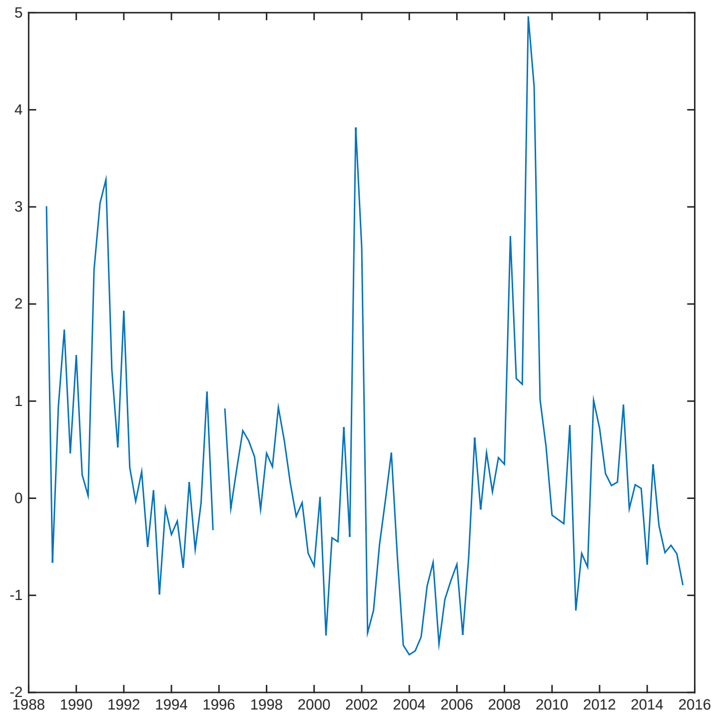


Figure 13. Sentiment Measure

Sentiment (or pessimism) measure, computed as the difference between the expected GDP growth computed from an AR(4) projection of quarterly realized GDP growth at 1-year horizon and the expected GDP growth of our proxy for the representative agent, i.e. the weighted average of all expected GDP growth, where we assign equal weights to all agents with past bond return accuracy in the upper tercile and a weight of zero to the other agents.