Rationality and Subjective Bond Risk Premia

Andrea Buraschi Ilaria Piatti Paul Whelan

February 2018

Abstract

We construct and study the cross-sectional properties of survey-based bond risk premia and compare them to traditional statistical counterparts. We document large heterogeneity in skill, identify top forecasters, and learn about the importance of subjective risk premia in long-term bonds dynamics. Next, we propose a new real-time aggregate measure of bond risk premia consistent with Friedman's market selection hypothesis. Finally, we use this measure to evaluate behavioural versus rational explanations of subjective risk premia and find support for models that include both sentiment and time-varying quantity of risk channels.

Keywords: Rational Expectations, Beliefs, Bond Risk Premia

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 Marco Giacoletti (EFA Discussant), Anna Cieslak (Adam Smith Discussant) Christian Eyerdahl-Larsen (AFA Discussant), Anna Obizhaeva (FMA Discussant), and the participants of the AFA meeting 2017, Chicago, the CICF meeting 2016, Xiamen, (Best Paper Award), 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, Goethe University Frankfurt, Paris Dauphine University, Bocconi University and the 2017 Adam Smith at HEC Paris for valuable comments. Paul Whelan gratefully acknowledges financial support from the Center for Financial Frictions (FRIC), grant no. DNRF102. The usual disclaimer applies. Emails: andrea.buraschi@imperial.ac.uk, ilaria.piatti@sbs.ox.ac.uk, pawh.fi@cbs.dk.

A large literature finds compelling evidence of predictability in several asset markets. A stream of this literature interprets this result as evidence of a time-varying risk premium that can be understood in the context of rational general equilibrium models. A second stream, on the other hand, argues that several characteristics of this predictability are more likely due to the existence of behavioural biases affecting the dynamics of subjective beliefs, informational frictions, or both. In this paper, we focus on a dataset that provides us with forecaster specific expectations about short and long-term U.S. Treasury bond yields, and corresponding projections for future GDP and inflation.

We use the time-series and cross-sectional dynamics of this data to study the properties of subjective bond risk premia as revealed by agents, as opposed to the traditional approach of inferring bond risk premia from projections of future return realizations on lagged state variables. The survey-based approach has a number of economic and econometric advantages compared to the standard projections, including the lack of small sample biases, potential overfitting and model misspecification. Most importantly, the survey-based expectations of future excess returns are by construction model-free and forward-looking, and thus provide us with a truthful and real-time representation of the expectations of economic agents, which are then reflected into bond prices. Since we observe expectations at the level of individual forecasters we can obtain and study a cross section of subjective risk premia in addressing questions related to rationality and aggregation.

We begin by constructing measures of subjective bond risk premia (EBR) from professional market participants' expectations regarding future yields. Specifically, we use Treasury coupon bond yield forecasts at the agent specific level to obtain a set of constant maturity 1-year zero-coupon bond yield expectations. Individual agent $EBRs$ are then obtained by subtracting the date t observable risk free rate from their price change individual expectations. Using individual specific data, we can directly study the characteristics of those agents who show persistent skill and the extent to which agents who can persistently predict long-term bond returns are also those who can predict variations in short-term interest rates. Moreover, individual level data allows us to test alternative models of beliefs formation and rationality, which would not be possible using aggregated data. Given the large and persistent heterogeneity in expectations and accuracy that we document, we propose an alternative belief aggregation mechanism that results in a better proxy of the beliefs of the marginal agent than consensus beliefs. This allows us to address a number of questions related to expectation formation in the context of asset pricing.

First, we document a large unconditional heterogeneity in the cross-section of EBR point forecasts. The mean forecaster $EBRs$ is 1.06% for 10-year bonds. However, the mean of the first quartile EBR is negative and equal to -1.66%, which implies that these agents believe long-term bonds are hedges against economic shocks while the mean of the third quartile is positive and equal to $+3.57\%$, which is instead consistent with beliefs of long-term bonds being risky bets on future economic states. We also find clear evidence of persistence in agents-specific bond risk premia. For example, a forecaster in the first quartile of the cross-sectional distribution of 2-year EBR has a probability of about 75% to stay in the first quartile the following month, and this probability is about 74% for the 10-year EBR. This is three times what it should be under the null hypothesis of no persistence. This is important in the context of the literature on rational expectations since it is well known that in the presence of heterogeneity, tests based on consensus proxies are affected by aggregation biases (see, e.g. Bonham and Cohen (2001), Zellner (1962), and Keane and Runkle (1990)). This highlights an important concern about traditional empirical work that assumes the marginal investor has consensus (arithmetic average) expectations, which we attempt to address.

Second, professional forecasters are reasonably accurate. The slope coefficient of predictive regressions of bond excess returns on their ex-ante subjective expectations is on average positive, contrary to what Greenwood and Schleifer (2014) document in the context of the stock markets. However, forecasters tend to underpredict excess bond returns since the intercept in the predictive regressions is positive and significant for most agents. Despite this bias, which is an indication of a rejection of the null hypothesis of full-information rational expectations (FIRE) at the individual level, agents beliefs about returns are substantially more accurate than previously thought. In fact, a large fraction of professional forecasters are reasonably good with respect to several econometric benchmark models estimated in real-time and the ranking in their accuracy is persistent over time. When we investigate in further detail the identities of the top forecasters, we find that banks and brokers acting as primary dealers and trading directly with the Federal Reserve are more likely to be among the top forecasters of the short-term interest rate.[1] This is consistent either with primary dealers having superior

[1] Primary dealers are trading counterparties of the New York Fed in its implementation of monetary policy

information about Fed's implementation of monetary policy, better ability in forecasting the future state of the economy, or with an information flow advantage originating from their role as market makers in Treasury bonds. Interestingly, however, primary dealers are not significantly better than other institutions in forecasting long-term bonds returns. The lack of a strong rank correlation in the distribution of forecasters for long and short-term bonds is a strong indication that the main determinant of long-term bond returns predictability is not the predictability of short-term interest rates but the time variation in bond risk premia.

Third, we consider alternative channels of expectation formation that could explain the rejection of FIRE at the level of individual forecasters and the formation of persistent heterogeneous beliefs with different levels of accuracy. One possibility could be that agents form rational expectations but subject to heterogeneous information frictions. Models with sticky-information or noisy-information with heterogeneous degrees of information rigidity could potentially be consistent with both the rejection of FIRE and the cross section of accuracy. A robust implication of models with information rigidities is that forecast errors are positively correlated with contemporaneous forecast revisions (Coibion and Gorodnichenko (2015)). We estimate regression coefficients at the individual level and report the distribution of individual slope coefficients. For almost all the agents, the slope coefficient is negative. This implies a strong rejection of informationally-constrained rationality and supports more behavioural explanations of the predictability of forecast errors. Therefore, we investigate the biases in beliefs and their dynamics, using a measure of sentiment, defined as the deviation between agents beliefs about fundamental growth and the expectation of an unbiased econometrician. We show that sentiment in the cross section of forecasts at least partially explains the bias in the forecast errors for long-term bond excess returns, suggesting that there is a systematic behavioural component in agents beliefs. Agents are optimistic on average in our sample (i.e. sentiment tends to be positive) and this drives an average positive forecast errors on bond excess returns.

Fourth, we address the question of how beliefs should be aggregated to construct an empirical proxy of subjective bond risk premia for the representative agent. A common approach in the empirical literature is to proxy subjective beliefs with consensus expectations, in which both accurate and inaccurate agents have the same arithmetic weight. In some cases, this choice is imposed by data limitations. In the context of asset pricing, however, this is tantamount to

and are also expected to make markets for the New York Fed on behalf of its official accountholders as needed.

assuming that the marginal agent holds consensus beliefs, but Friedman (1953) and Alchian (1950) argue that market selection in competitive markets is a powerful force affecting the characteristics of the representative agent. Trading markets eventually punish irrationality and those agents that are consistently more accurate than others accumulate more economic weight in the pricing kernel. Thus, their beliefs, rather than the consensus ones, should be the one more tightly revealed (spanned) by bond prices. Since our data set provides us with the forecasters' identity, at any point in time t we can construct the average beliefs of the agents who were most accurate up to time t, which we denote as $EBR_t^{\star}$. This new belief aggregation is consistent with the spirit of Friedman (1953) and Alchian (1950) and, after discussing the specific properties of this time series, we use $EBR_t^{\star}$ as a dependent variable in a series of rational expectation tests, to investigate the rationality of our proxy for the marginal agent's beliefs. If expectations are rational, prediction errors should be orthogonal to any public information available at time t. We find statistical evidence against orthogonality using contemporaneous information in the term structure of bond yields. In particular, the 4th and 5th principal components of bond yields are significantly linked to forecast errors since they are not correlated to $EBR_t^{\star}$ but explain around half of the predictable variation in realized returns across maturities. While this result seems to suggests that agents are not exploiting all available information when forming beliefs, it could also be the case that the predictive power of higher order principal components is only a statistical feature of the data, detectable ex-post, but that should not be considered as agents form expectations in real time. To address this issue we investigate how the economic significance of the predictable component of the forecast errors maps into economic significance. This question is central in the debate about the significance of behavioural biases in agents expectations and the empirical plausibility of the assumption of rational expectation in asset pricing models.

Summarising, we find that statistical rejection of conditional orthogonality does not translate into economic significance. Indeed, any attempt to use publicly available information to correct the ex-ante bias in agents forecasts and reduce their ex-post forecast errors fails when we restrict such a correction to use exclusively real-time data. However, if we correct $EBR_t^{\star}$ only for the constant bias, i.e. the intercept in the regression of forecast errors on the forward spread over rolling windows, the root mean squared error of the forecast decreases by about 10% for the 10-year bond, which is consistent with dogmatic persistent beliefs, with the average positive alpha in the predictive regressions and with the positive correlation between individual forecast errors and sentiment.

Our last set of questions relates to an extensive macro-finance literature on the dynamics of interest rates and bond risk premia. The empirical evaluation of these models is often carried out by constructing bond risk premia after approximating agents expectations with econometric projections of future realized returns on lagged state variables. We revisit this approach and use the cross-sectional information on subjective expectations, as directly revealed in real-time by agents, to compare the performance of alternative models in explaining the dynamics of risk premia. Instead of future average returns, we directly use the time series of $EBR_t^{\star}$ which is based on the beliefs of the agents who have been most accurate up to time t. The use of this alternative metrics is also motivated by the fact that they are available in real-time, thus allowing us to avoid the look-ahead bias which often plagues methodologies that require model-dependent assumptions about agents expectations. An important finding is that several traditional structural models perform quite well under this metric. In particular, models that argue about the importance of the quantity of risk channel as a source of time variation in bond risk premia generate large R^2 with loadings on risk factor proxies that are consistent with predictions from theory. Moreover, in the regressions of $EBR_t^{\star}$ on the state variables implied by structural models, we also include our measure of sentiment, to control for the bias in the marginal agent's expectations, and we show that sentiment is always statistically significantly linked to subjective risk premia and largely increases the predictive power of structural models. This results highlight the importance of both behavioural and rational elements in driving bond risk premia. Finally, considering the link between bond volatility and bond risk premia we document a surprising result. The empirical sign of bond volatility on subjective bond risk premia is positive and statistically significant, consistent with a very general prediction from structural models. This is interesting given the long standing tension that fails to find a link between the volatility channel and future realized excess returns.

The paper proceeds as follows. Section I presents the data and summarizes the empirical questions we aim to address. Section II discusses the cross-sectional properties of subjective bond risk premia. Section III investigates the accuracy of professional forecasters and documents the extent of persistence in forecasting skill. Section IV considers alternative channels of expectation formation to rationalize the existence of persistent heterogeneous beliefs with different levels of accuracy. Section V addresses the question of beliefs aggregation given the large extent of heterogeneity in the cross section. Section VI conducts direct tests of rationality on the beliefs of our proxy for the representative agent. Section VII uses $EBR^{\star}$ to evaluate the performance of structural models of expected bond risk premia proposed in the literature and Section VIII concludes.

Related Literature: We contribute to four strands of the macro-finance literature.

First, we contribute to a literature that uses survey data to learn about expectation formation that makes two conflicting arguments. A first set of authors have argued that subjective forecasts contain valuable information about future GDP and inflation (Ang, Bekaert, and Wei (2007) and Aioli, Capistran, and Timmermann (2011), in support of rational expectations. A second set has argued that survey forecasts about financial returns provide evidence against the rational expectations assumption (Frankel and Froot (1987) for exchange rates, Froot (1989) for interest rates, Bacchetta, Mertens, and Van Wincoop (2009) and Greenwood and Schleifer (2014) for equity). Koijen, Schmeling, and Vrugt (2015) find that a trading strategy that uses survey expectations as a signal negatively predict future returns in the time series in three major asset classes: global equities, currencies, and global fixed income. Focusing on short rate expectations Cieslak (2017) argue that forecast errors could be predicted ex-post with proxies of real activity which provides an explanation for so-called 'unspanned' return predictability (Duffee (2011a) and Joslin, Priebsch, and Singleton (2014)). The closest paper to ours is Piazzesi, Salomao, and Schneider (2015) who study a consensus measure of subjective bond risk premia and argue yield expectations are more persistent and less volatile than those obtained by common statistical measures.

Different than these papers, we study the link between rationality and bond risk premia at the individual forecaster level. Indeed, the existence of subjective expectations plays an important role since what matters for asset pricing is not consensus expectations, but the expectation of the marginal agent. Moreover, depending on the level of underlying micro-heterogeneity, aggregation of beliefs to a consensus proxy may lead to inconsistent parameter estimates and conclusions in commonly employed tests of rational expectations (Figlewski and Wachtel (1983), Keane and Runkle (1990), Bonham and Cohen (2001)). Our study attempts to address issues related to data aggregation and the notion of a consensus investor rationality by working directly at the individual forecaster level.

Second, considering the cross-section of expectations we also contribute to a literature which studies belief heterogeneity. For example, Malmendier and Nagel (2011) and Das, Kuhnen, and Nagel (2017) argue that agents update their expectations heterogeneously, depending on their personal experiences and social status, and find evidence of substantial disagreement between young and old individuals in periods of highly volatile inflation (e.g. in the 1970s). Patton and Timmermann (2010) document that dispersion in beliefs about macro fundamentals is countercyclical and that forecaster expectations are persistently optimistic or pessimistic. Andrade, Crump, Eusepi, and Moench (2014) study the cross-section of GDP, inflation, and short rate expectations showing that agents disagree at all horizons, including the long run, and that the term structure of disagreement is downward sloping for macro expectations while it is upward sloping for short rate expectations.[2] The novel contribution of this paper is to investigate the extent to which agents agree about whether bonds are risky bets or hedges, namely whether agents expect bonds to generate positive or negative risk premia, and document the extent of heterogeneity and persistence in agents beliefs about future bond excess returns.[3]

Third, we study a set of questions related to a macroeconomic literature that emphasizes the role of belief formation by relaxing the assumption of full information rational expectations (FIRE). These questions are addressed after noticing that models with either information rigidities (Mankiw and Reis (2002)) or noisy-information (Woodford (2002) and Sims (2003)) have specific implications for the link between forecast errors and forecast revisions (Coibion and Gorodnichenko (2015), Dovern, Fritsche, Loungani, and Tamirisa (2015), and Mackowiak and Wiederholt (2009)). This paper contributes to this strand of literature by investigating the cross-section of FIRE deviations. Moreover, we study information frictions embedded in interest rate expectations and study a set of alternative channels that can explain deviations from FIRE.

Fourth, it is tradition to evaluate fixed-income models on the basis of their predictive power for future *realized* returns. Based on using realised returns as a proxy for expected returns, a

[2]Models that include a role of belief heterogeneity in equilibrium asset pricing include Detemple and Murthy (1994), Hong and Stein (2003), Scheinkman and Xiong (2003), Basak (2005), Buraschi and Jiltsov (2006), David (2008), Gallmeyer and Hollifield (2008), Dumas, Kurshev, and Uppal (2009), Bhamra and Uppal (2014), Xiong and Yan (2010), Ehling, Gallmeyer, Heyerdahl-Larsen, and Illeditsch (2018), Borovicka (2011), and Jouini, Marin, and Napp (2010).

[3]For a discussion of bond as risky bets versus deflation hedges, see Campbell, Sunderam, and Viceira (2017).

consensus has emerged in the literature that many of these macro models are not very successful empirically (Duffee (2013)). We revisit these tests using a measure of bond risk premia built from the cross-section of past performance weighted expectations as a proxy for the wealth weighted belief of the marginal agent. We consider several popular models, among which those assuming (a) habit preferences (Campbell and Cochrane (1999), Wachter (2006)), (b) early resolution of uncertainty and long-run risks (Bansal and Yaron (2004) , Bansal and Shaliastovich (2013)), (c) belief disagreement (Dumas, Kurshev, and Uppal (2009), Buraschi and Whelan (2017)); (d) a general prediction that links quantities of risk to compensation for risk (Merton (1980), French, Schwert, and Stambaugh (1987), Duffee (2002)). To summarise, using this real-time subjective expected return metric, we find support for economies generating time-varying bond risk premia via an interaction between a quantity and a price of risk channel.

I. Empirical Design

This Section briefly introduces the data to then illustrate the empirical questions we aim to address.

A. Data

We construct real-time measures of subjective bond risk premia directly from professional market participants' expectations regarding future yields for the sample period January 1988 to July 2015. 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. In our analysis we use agent specific forecasts for the Federal Funds rate, Treasury bills with maturities 3-months/6-months, Treasury notes with maturities 1, 2, 5, 10-years, and the 30-year Treasury bond. 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.[4] It is unique with respect to alternative commonly

[4]Other studies that use BlueChip include Piazzesi, Salomao, and Schneider (2015) and Cieslak (2017), who study consensus bond risk premia and fed fund rates, respectively. Chernov and Mueller (2012), Andrade, Crump, Eusepi, and Moench (2014) and Buraschi and Whelan (2017) use inflation expectations from BCFF.

studied surveys along at least four dimensions. First, the dataset provides the identity of the forecasters. This allows us to track each individual in the cross section and time series. [5] Indeed, most of our questions could not be addressed using datasets with pre-aggregated data. Second, the dataset is available at a monthly frequency, while other surveys, such as the Survey of Professional Forecasters' (SPF) is available only at quarterly frequency. This increases the power of asset pricing tests. Third, the number of participants in the survey is large and stable over time. In our sample it is 42 on average, with a standard deviation of about 2.3. Moreover, 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.[6] Fourth, Bluechip has always been administered by the same agency, while other surveys, such as SPF, have been administered by different agencies over the years. Moreover, SPF changed some of the questions in the survey, and some of these changes crucially affected the forecasting horizon.[7] Finally, the survey is conducted in a short window of time, between the 25th and 27th of the month and mailed to subscribers within the first 5 days of the subsequent month. This allows the empirical analysis to be unaffected by biases induced by staleness or overlapping observations between returns and responses.

One complication of BCFF forecasts is that while surveys are conducted on a monthly basis the projections are reports on a future quarter calendar cycle so that the forecast horizon varies each month. For example, in January, April, July and October a 12-month ahead and 15-month ahead forecast is reported, whereas in March, June, September and December a 10-month ahead and 13-month ahead forecast is reported. To construct a j-quarter ahead constant maturity forecast we linearly interpolate along adjacent horizons for 2nd and 3rd months in the cycle.

To obtain curves of expected zero coupon discount rates we uses the Svensson (1994) method,

[5] Forecasters are identified by institution's name. For example, 'J.P. Morgan' or 'Goldman Sachs' or 'Fannie Mae'.

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

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

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.[8] 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. Holding periods are quarterly up to 1.25-years (in what follows we focus on 1-year excess returns) for bond maturities evenly spaced between 1 and 10-years (we disregard maturities greater than 10-years). Over the whole sample there are 97 forecasters for which we can compute the whole expected term structure of zero-coupon yields and on average they contribute to the cross-section for about 138 months. Of these 97 forecasters, 84 participate to the panel for at least 5 years, and on average they contribute to the cross section for about 154 months.

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.

B. *Framework and Questions*

Equilibrium term structures that allow for belief heterogeneity depend explicitly on the aggregation properties of subjective expectations about future short-term interest rates and future bond risk premia. The data set described above allows us to observe both sets of expectations. Specifically, given information on individual expectations about the cross-section of future interest rates, BCFF allows us to 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 are then defined as $y_t^n = -\frac{p_t^n}{n}$. The realized excess log return from holding the n-years bond from date t to $t+h$ is $rx_{t+h}^n = r_{t+h}^n - hy_t^h$, with the gross return being defined as $r_{t+h}^n = p_{t+h}^{n-h} - p_t^n$. In the analysis the follows, we focus on a 1-year holding periods indicated by $h = 1$.

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

[8] To estimate the set of parameters we minimize the weighted sum of the squared deviations between actual and model-implied prices. 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 $\tau_2 > 0$.

compute the implied cross-section of EBR as $erx^n_{i,t} = E^i_t\left[p^{n-1}_{t+1}\right] - p^n_t - y^1_t$ since from the surveys we directly observe $E^i_t\left[y^{n-1}_{t+1}\right]$

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

Given a panel of survey expectations with this structure, the first question that arises is how one should conduct tests in aggregate. Indeed, belief aggregation poses a number of well known difficulties. For example, if agents have access to different signals, rational expectation tests based on consensus (arithmetic average) beliefs are inconsistent and can lead to overrejection of the null of rationality (Figlewski and Wachtel (1983)). Inconsistency can also go in the opposite direction since belief averaging can lead to false acceptance of the null (Keane and Runkle (1990)). Avoiding consensus aggregation by pooled estimation is also not a viable alternative since it can, depending on the degree of belief heterogeneity, also lead to inconsistency (Bonham and Cohen (2001)). Therefore, we study the properties of the cross section of individual expectations, and the extent to which it can be efficiently summarized by the consensus. Second, the understanding of the dynamics of the distribution of beliefs requires an analysis of the accuracy of the subjective expectations. Thus we question the degree of predictability implied by survey-based expectations, with a focus on both the time series and cross section of accuracy, and the link to models of expectations formation. A further question that arises naturally is how to think about the beliefs of the marginal investor. In frictionless and competitive markets, heterogeneous agent economies suggest that inaccurate agents eventually lose economic weight and their influence on the stochastic discount factor.[9] Taking explicitly into account that what matters for equilibrium prices is not consensus but wealth weighted (performance adjusted) beliefs, we build a new aggregate measure of subjective expectations. This proxy of the beliefs of the marginal agent is then used to address questions related to its rationality, and to the determinants of subjective bond risk premia.

Summarising, our empirical design is focused on the following set of questions:

Q_1 : *What are the cross-sectional properties of agents' beliefs on bond risk premia in terms of their heterogeneity, persistence, and state-dependence?*

[9]The natural selection hypothesis as advocated by Alchian (1950) and Friedman (1953) explicitly make this argument as the mechanism through which economy eventually tends towards rationality.

Q_2: *Are agents 'good' at forecasting long term interest rates with respect to benchmark statistical models? Is there persistent skill in the cross-section of forecasters?*

Q_3: *What is the extent of full-information rationality? If agents are not fully rational, is it possible to exploit their forecasts errors using information by financial markets?*

Q_4: *Do state variable proxies arising in benchmark structural models explain the dynamics of* EBR. If so, do they explain EBR *with the correct slope coefficient?*

II. The Cross Section of Expected Term Structures

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

[Insert Figure 1 here.]

We also document large unconditional heterogeneity in the cross section of EBR forecasts. The median (Q2) forecaster EBR is 1.06% for 10-year bonds. However, the first and third quartiles (Q1 and Q3) are -1.66% and +3.57% for the same maturity, respectively. This implies that while the consensus believes in a positive risk premium, a significant fraction of investors believe in a negative bond risk premium.[11]

The conditional properties of the cross-sectional distribution of EBR display rich dynamics in the time series (see again the top panel of Figure 1). There exists significant time-varying

[10]See Figure 2 in the Supplemental Appendix.

[11]Table II in the Supplemental Appendix 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.

heterogeneity around the consensus forecast. Given the wide use of consensus (average) expectations both in the literature and in the financial industry, it is interesting to test more formally the null hypothesis that the cross-sectional properties of subjective 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 bottom panel of Figure 1 highlights the time variation in heterogeneity by plotting the cross-sectional standard deviation of EBR standardized by the full sample mean EBR, for bond maturities 2, 5 and 10-year. The figure also shows that the dispersion in beliefs is state-dependent: it tends to rise at the onset of recessionary periods and drop again as the economy recovers.[12] Also, disagreement is non-monotonic in maturity displaying a 'hump shape' around the 5-year maturity.[13]

These findings motivate a rigorous study of whether the assumption that the marginal investor has average (consensus) expectations, often used in the literature, is innocuous.

A. *Belief Persistence*

Disagreement about short rates, bond returns, and the macro economy are all persistent. This raises an interesting question: is disagreement a result of dogmatic beliefs, information frictions, or both? In order to address this question we first rank all forecasters according to whether in a given month t their forecast is in the first, second, third or fourth quartile of the cross-sectional distribution. We repeat this exercise for all months in the sample and compute transition

[12] 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.

[13] We also find that disagreement about long term bond excess returns is more than ten times larger than disagreement about short rates or disagreement about the macro economy, see bottom right panel of Figure 2 in the Supplemental Appendix.

probabilities, i.e. the probability that forecasters in a given quartile at time t stay in that particular quartile the following month or move to a different quartile of the distribution. We do this first for short rate forecasts in the left panel of Table I.[14] If views are not persistent, all the entries in these transition matrix should be approximately equal to 25%. Instead, we find that the diagonal elements are significantly higher than 25%, in particular for the most extreme quantiles, Q1 and Q4 where they are always above 70%.[15]

[Insert Table I 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 bond risk premia is related to agents' perception about bonds being hedging assets or rather risky bets on consumption (inflation) risk. In the first case, bonds should earn a negative risk premium, in the second expected bond risk premia should be positive. Thus, we estimate the extent to which individual forecasters are persistently in one particular quartile of the cross-sectional distribution of subjective $EBRs$. The middle and right panels of Table I show the transition matrices for subjective excess returns on 2 and 10-year bonds, respectively. The results suggest that forecasters have persistent beliefs about bond risk premia, relative to consensus excess returns. For example, a forecaster in the first quartile of the cross-sectional distribution of 2-year EBR has a probability of 75% to stay in the first quartile the following month, and this probability is 74% for the 10-year EBR, which is about three times what it should be under the null hypothesis of no persistence. In all cases, the probability of remaining in the same quartile is significantly higher than 25% at a level of 5%.[16]

In order to investigate the drivers of this disagreement we directly study the dynamics and accuracy of these beliefs. This is the topic of the next section, which is cast in the predictability regression framework used in the classical bond literature.

[14]Supplemental Appendix A then investigates whether expected term structures are consistent with agents' expectations about future economic fundamentals. We show 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.

[15]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. Transition matrices for GDP and CPI growth expectations show very similar results and are available from the authors upon request.

[16]Even at an annual instead of monthly frequency, the probability of remaining in the same quartile is significantly higher than 25%. For example, a forecaster in the first (fourth) quartile of the cross-sectional distribution of 10-year EBR has a probability of 45% (42%) to be in the first quartile the following year.

III. Long-Term Bond Predictability

In this section, first we investigate the accuracy of professional forecasters and document the extent of persistence (if any) in the forecasting ability for long-term bonds returns. Since long-term bond returns are affected by both changes in short-term interest rates and bond risk premia, we investigate the extent to which skill in forecasting short-term rates can translate into superior ability to forecast long-term bond returns. If bond risk premia were small and slowly time-varying, one should find the two skills to be highly correlated. If bond risk premia were large and time-varying, on the other hand, one should find lack of correlation between the two.

A. *Accuracy of professional forecasters*

To assess the accuracy and the degree of heterogeneity in the cross section, we first run a simple predictive regression of realized excess returns on subjective $EBRs$, for each single contributor i to the BCFF panel, focusing on the contributors with at least 5 years (60 months) of forecasts for bonds with maturity n and forecast horizon $h = 1$ year:

$$rx_{t+1}^{n} = \alpha_i^n + \beta_i^n erx_{i,t}^n + \epsilon_{i,t+1}^n. \tag{2}$$

Figure 2 shows the distribution of individual regression coefficients and R^2 of regression (2) for a 10-year bond. Despite the heterogeneity in accuracy, a few forecasters are extremely accurate with slope coefficients close to one and R^2 larger than 20%. The correlation between expectations and future realization of excess bond returns is positive for 69 out of 84 forecasters. The coefficient β_i^{10} is significantly positive at 5% level for 30 agents and significantly negative for only two of them. This mostly *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

[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

related to the measurement error and aggregation in consumer survey data or to the fact that bond returns are easier to predict than equities, due to the absence of uncertainty about future nominal cash-flows. Alternatively, it is possible that professional forecasters are simply better than retail investors and consumers, due to professional incentives and access to superior information.

Despite the positive correlation with future realized returns, forecasters' beliefs still display some potential signs of irrationality in their level. In fact, the α_i^{10} coefficient in regression (2) is positive for 82 out of 84 agents and significant at 5% level for 71 agents, meaning that forecasters tend to underpredict excess bond returns. However, even taking into account this bias, our results show that agents beliefs about returns are substantially more accurate than previously thought.

[Insert Figure 2 here.]

We further study forecast accuracy at the level of each individual forecaster i by computing their root mean squared errors

$$RMSE_i^n(Survey) = \sqrt{\frac{1}{T_i}\sum_{t\in\tau_i}\left(rx_{t+1}^n - erx_{i,t}^n\right)^2},$$

for bond maturity $n = 10$ years, where τ_i is the set of T_i dates in which forecaster i contributes to the panel. The individual RMSEs range between 5.30 and 15.83. Since individual forecasters appear in the sample at different times, we assess their accuracy relative to two benchmark models.

Our first benchmark model is the Cochrane and Piazzesi (2005) return forecasting factor, which is a tent-shaped linear combination of forward rates that has been claimed to subsume the information contained in the level, slope and curvature of the term structure. However, the in-sample predictive content of the Cochrane-Piazzesi factor relies on estimates for factor loadings that were not available in real time. For example, the coefficients of the 'tent-shaped' factor used to forecast returns in the 1990s uses information available during the 2000s. In real time the shape of the factor loadings on the forward curve displays time variation (see, e.g., Bauer and Hamilton (2015)). One should also be concerned with any regression where the

Economic Survey respondents can select.

right hand variables have a root very close to unity, such as the first stage Cochrane-Piazzesi regression. With these concerns in mind, we construct a real-time version of the CP factor as follows: We initialise a CP with 10-years of data from January 1978 to January 1988 via a two stage projection of *forward spreads*.[18] Then, using an expanding window we estimate factor loadings using realized returns available 1-year ago, and apply these to date t forward spreads to construct a date t predicting factor: $CP(t)$.[19] Using this real-time $CP(t)$ factor we proceed with an out-of-sample assessment. Our second, more parsimonious, benchmark is the implied forecast by the slope of the yield curve, defined as the 10-year yield minus the 1-year yield. This forecast requires computing a single factor loading (the beta on the slope), which we estimate in real-time in the same way that we estimate factor loadings on $CP(t)$, using the same initialisation period and same rolling window length.[20]

Table II presents root mean square prediction errors for 10-year bond excess returns, based on forecasts from the unconditional top 10 forecasters (WORST), the bottom 10 forecasters (BEST), and the arithmetic average of the cross-section of forecasters for each prediction date t (CONS), i.e. the consensus.[21] The top panel reports RMSEs for all months in the sample period while the bottom panel excludes the zero lower bound period, i.e. January 2009 to July 2015. Focusing on the performance of surveys, we see large heterogeneity between the best and worst forecasters and this dispersion in skill persists when excluding the zero lower bound (ZLB) period. Also, we find evidence of state dependence: when we distinguish between recessions and expansions we show that all agents are much better at forecasting returns in bad times, consistent with the idea that the dynamics of interest rates are dominated by risk premia in these periods. For example, in the full sample the RMSE for the top ten forecasters drops from 7.75 to 5.15 but also the RMSE for the worst forecasters drops, from 10.56 to 8.19. Next, compare surveys to our benchmark models. In the full sample, both CONS and BEST outperform real-time CP and this holds in recessions and expansions, including and

[18]Cochrane and Piazzesi (2008) advocate the use of forward spreads (n-year forward rates minus the 1-year yield) instead of forward rates to remove spurious level effects. Indeed, in unreported results we find a real time CP constructed from the level of forward rates contains a bias, presumably because of the trend in interest rates in our sample. Constructing CP from spreads significantly improves its out-of-sample performance.

[19]The out of sample projection also requires estimating the relationship between $CP(t)$ and expected returns on n-year bonds, which we estimate using only information that was available in real-time.

[20]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.

[21]Top and bottom 10 forecasters are identified by looking at the average accuracy ranking percentiles, $\mathcal{R}_i$ defined below, over the full sample.

excluding the ZLB period. However, we also find that real-time forecasts implied by *Slope* always outperforming CP. In the full sample, the RMSEs from the *Slope* forecasts are almost identical to BEST and better than the CONS. However, the relatively strong performance of the *Slope* is mainly driven by large forecast errors made by surveys in the ZLB period. Excluding this period, even the consensus is beating *Slope* with RMSE equal to 7.98 compared to 8.11 for the *Slope*.

Figure 3 investigates forecast evaluation in the time-series by plotting the differences in absolute 10-year bond excess return forecast errors between BEST and the *Slope*, and between BEST and CP. Errors are plotted for the dates when they are realised. Shaded areas below zero are periods in which BEST is more accurate (in an absolute sense) than the benchmark models. The time series of absolute error differences reveals that relative forecast accuracy is changing over time in a persistent manner. For example, between 1992-1994, 2001-2003, 2003-2006, 2012-2013 surveys produce superior predictions. Interestingly, in the period that follows hitting the zero lower bound (December 2008), survey forecasts perform badly. This poor performance is again repeated in the final years of our sample in what is popularly known as the 'lift off' period.[22]

[Insert Table II and Figure 3 here.]

Next, we compare individual agent survey forecasts, out-of-sample by construction, to these benchmark models by computing a measure of relative performance $\mathcal{A}_i^n$ for the periods in which agents are present in the panel:

$$\mathcal{A}_i^n = \frac{RMSE_i^n(Survey)}{RMSE^n(Model)}. \tag{3}$$

Values smaller than one imply better performance under the subjective measure. Figure 4 presents the histogram for the cross section of $\mathcal{A}_i^n$, for bond maturity n equal to 10 years. We find that an important fraction of survey forecasters out-perform these benchmark models. For the CP model in the full-sample around 40% of the forecasters have $\mathcal{A}_i^n < 1$. At the same time, we find that about a fifth of forecasters have $\mathcal{A}_i^n > 1.2$, thus producing forecasts significantly worse than the CP model. Compared to the *Slope* forecast the mass off the distribution lies between 1.00 and 1.20 for the full sample, indicating that most agents are slightly worse.

[22]`https://www.wsj.com/articles/fed-leaves-interest-rates-unchanged-1442512974`

However, consistent with the discussion above this is driven by large errors since hitting the ZLB. Excluding this period around 40% of the agents in the cross section outperform a *Slope* forecast.

[Insert Figure 4 here.]

Given the large heterogeneity in survey expectations it might not be surprising that there is evidence of accuracy in the cross section, but what is surprising is that this accuracy tends to be persistent. Since forecasters contribution to the survey can occur at different time periods, to quantify the persistence we first compute the squared forecast error at each time t. Then we calculate the percentiles of these squared errors for each forecaster, that we call *accuracy ranking* percentiles, $\mathcal{R}_{i,t}$, and we compute the time average $\mathcal{R}_i$ of these percentiles. Low percentiles correspond to greater accuracy. We repeat this exercise for all months t 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 the following month or move to a different quartile of the distribution. If accuracy were not persistent, all the entries in the middle panel of Table III should be approximately equal to 25%. We find instead that the diagonal elements are significantly higher than 25%, especially for the most extreme quantiles, Q1 and Q4. For example, a forecaster in the first quartile of the cross-sectional distribution of 10-year EBR accuracy has a probability of 58% to stay in the first quartile of accuracy the following month. This probability is 70% for the 4^{th} quartile, which contains the worst forecasters, suggesting that bad forecasters produce more persistently poor forecasts, even more so than good forecasters.

[Insert Table III here.]

Summarising, two conclusions emerge. First, expectations of a significant fraction of professional forecasters are rather good even with respect to popular fixed income models, and not as bad as previously reported. At the same time, while professional forecasts can be used to build reliable measures of bond risk premia, one needs to be mindful of the heterogeneity in the distribution of these beliefs. The assumption that consensus can be used as a sufficient statistics for the whole panel to proxy the beliefs of the marginal agent is certainly not supported by the empirical evidence. Second, both good and bad accuracy is persistent.

B. Short-rate vs long-rate accuracy

A natural question to ask is 'does the superior predictive power that some forecasters display for long-term predictions originate from their ability to predict short-term rates?'. This question relates to the important issue of whether the dynamics of long term interest rates is driven by variation in expected short rates (cash-flow channel) or risk premia (discount rate channel).

Similar to the analysis we have conducted for long-term bond returns, we rank all forecasters according to their accuracy in each month t in predicting the short-term rate and denote their accuracy ranking percentiles for the 3-month yield $\mathcal{R}_{i,t}^{3m}$. The left panel of Table III shows the probability of a forecaster transitioning from a given quartile of the cross-sectional distribution of $\mathcal{R}_{i,t}^{3m}$ to another quartile in the following month. The results show that agents who are in the top or bottom quartiles by forecast accuracy tend to remain in the same quartile the following periods. So, similar to long-term bonds, accuracy (good and bad) in predicting short rates tends to be persistent.

To test the hypothesis that the ability to predict long-term bond returns comes from skill in forecasting short rates, we compare the long-term accuracy percentiles for the 10-year bond with the corresponding accuracy percentiles for the short rate. Indeed, we find that the two rankings are correlated: a regression of the 10-year accuracy percentiles on the 3-month accuracy has a significant slope coefficient of 0.42 and an adjusted R-squared of 21%. The right panel of Table III summarizes the conditional distribution of forecast accuracy for the 10-year EBR given the 3-month yield accuracy, that is, the probability that a forecaster is in a given quartile of the 10-year EBR accuracy percentile distribution knowing that the forecaster is in a given quartile of the 3-month yield accuracy percentile distribution. The elements on the diagonal show the existence of a link between the two accuracies. However, the rank correlation in terms of accuracy on the 3-month yield and on the 10-year EBR is far from perfect. Only 34% of the top short-rate forecasters are also top long-term yield forecasters.

To investigate in further detail the link between short-rate and long-term bond predictability, we take advantage of the knowledge of the specific identity of each individual forecaster. Who are the institutions that are especially good at predicting short-term interest rates and long-term bond returns? The answer is summarized in Table IV, which shows the top ten forecasters in terms of average percentiles of squared forecast errors for 10-year bond excess returns (left panel) versus short rates (right panel). These lists show that most of the best forecasters for the 10-year bond excess returns are not in the top-ten list for the short-term rate. In fact, only three institutions, i.e. Goldman Sachs, Nomura and Thredgold, are in both lists.

[Insert Table IV here.]

Interestingly, 6 out of the first 10 institutions in the list of top short-rate forecasters are currently primary dealers, or have been primary dealers at least once in our sample period. This is remarkable given that only 24 of the 84 institutions in the survey with at least 5 years of contributions are or have been primary dealers.[23] However, only three (UBS, Goldman Sachs, and Nomura) of the top ten forecasters for the 10-year bond excess returns are primary dealers. At the same time, financial institutions such as J.P. Morgan and BMO Capital Markets who can forecast the short rate rather well, are poor performers at the long end of the term structure. To formally analyze the relative performance of primary dealers versus non-primary dealers, for short versus long-rate predictions, we test the null hypothesis that their accuracy percentiles are drawn from the same distribution using a Kolmogorov-Smirnov test. Unconditionally, for 10-year bond excess returns forecasts 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%. Conducting the same test for short-rate forecasts we always reject the null at the 5% level. Thus, primary dealers have a significantly better predictive performance only for the short rate. This might be explained by primary dealers having specific information about (or better models to interpret) monetary policy.[24] Nonetheless, even this advantage does not easily translate into an advantage to forecast long-term bond returns. We conclude that an important component of the dynamics of expected bond returns at longer maturities is a risk premium term, which is not completely revealed by the dynamics of short-term interest rates.

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

[24]Alternative explanations are linked to accuracy about economic fundamentals and order flow information on short-term bonds. We describe them and test the first of these hypotheses in Section C of the Supplemental Appendix.

IV. Understanding the Cross Section of Accuracy

The large and persistent differences in predictive performance highlighted in the previous section raise the question of what drives the heterogeneity in forecasters' accuracy and beliefs, as well as the average bias reflected in the cross section of alphas in the predictive regression (2), which was suggestive of a deviation from the full-information version of the rational expectation hypothesis (FIRE) at the level of the individual forecasters. In this section we consider alternative channels of expectation formation that could explain the rejection of FIRE and the formation of persistent heterogeneous beliefs with different levels of accuracy.

A. Information frictions

First, agents could be forming rational expectations but subject to heterogeneous information. Indeed, when agents form rational expectations subject to information constraints ex-post forecast errors are predictable. Two important examples include models with sticky-information as in Mankiw and Reis (2002), and economies with noisy-information (Woodford (2002), Sims (2003), and Mackowiak and Wiederholt (2009)), in which agents rationally update their beliefs but, since they can never fully observe the true state, they use an optimal signal-extraction filter. In this case, agent's predictions are a weighted average of their prior beliefs and the new information received where the degree of information rigidity affects the relative weight of prior beliefs. In the case of sticky information, agents update their information sets infrequently as a result of a fixed costs of information acquisition and the degree of information rigidity is the probability of not acquiring new information each period.

Coibion and Gorodnichenko (2015) show that in models with rational agents subject to information constraints, future forecast errors are positively correlated with forecast revisions. Thus, for each single contributor i to the BCFF panel, we run the following regression and study informationally constrained rationality by testing $H_0 : b^{i,n} > 0$:

$$FE^{i,n}_{t,t+1} = a^{i,n} + b^{i,n} \Delta erx^{n}_{i,t} + \epsilon^{n}_{i,t+1}, \tag{4}$$

where $FE^{i,n}_{t,t+1} = rx^{n}_{t+1} - erx^{n}_{i,t}$ denotes the forecast error of forecaster i for the excess return of a bond with maturity n and $\Delta erx^{n}_{i,t}$ is its forecast revision, computed as the difference between its time-t forecast and the same forecast made the previous month.[25]

The top panel of Figure 5 shows the distribution of individual t-statistics of the slope coefficients $b^{i,n}$ in regression (4), for the 10-year excess bond returns ($n = 10$ years), focusing on the 84 contributors with at least 60 monthly forecasts. The forecast revision coefficient is negative for all but one agent, and significantly so for 80 out of 84 forecasters, at the 5% level.

[Insert Figure 5 here.]

Models of information frictions imply a *positive* relation between forecast updates and forecast errors, which is inconsistent with our finding of a mostly negative and significant slope coefficient in regression (4). Moreover, in models of sticky information, the predictability in forecast errors follows from the aggregation of forecasts across agents, even if no such predictability exists at the individual level, while we find a strong negative relation between forecast errors and forecast revisions also for the individual forecasters.

These results clearly suggest that not only we can reject a full-information version of the rational expectation hypothesis, but also that information rigidities, such as sticky or noisy information models, play a limited role in the formation of expectation about tradable securities such as US treasury bonds.[26]

The finding of a negative relation between forecast errors and forecast revisions in the cross section of forecasts is potentially consistent with more behavioural models of expectation formation. One example are models in which forecasters have asymmetric loss functions and are heterogeneous in their degree of loss-aversion (Capistran and Timmermann (2009)).[27] A second example are models that induce forecasters to smooth their predictions due to overconfidence or to avoid giving the impression of lack of confidence in their beliefs. Similar to heterogeneity in loss aversion, forecast smoothing makes forecast errors predictable even in absence of information frictions. Finally, the result could be due to biases in beliefs and the specific properties

[25]Note that this is not exactly a forecast revision because the horizon of the forecast is fixed to 1 year, so we are not taking the difference between two consecutive forecasts of exactly the same object, but given the monthly frequency and the yearly horizon the approximation error is likely to be very small. Moreover, we find very similar results using exact forecast revisions of long-term yields at quarterly frequency.

[26]We can also show that the negative slope coefficient is mainly coming from forecasts and forecast errors on long-term yields, while for short-term yields we find a mostly positive but insignificant link between individual forecast errors and forecast revisions.

[27]A simple model with asymmetries in the forecasters' cost of over- and under-predictions on the lines of Capistran and Timmermann (2009) is also consistent with the systematic bias in forecast errors and the persistence in the ranking of forecasts and forecast errors in the cross section.

of their dynamics.

B. Sentiment

To investigate the biases in beliefs and their dynamics, we construct a measure of forecaster's sentiment. In the general equilibrium literature with heterogeneous beliefs, *Sentiment* is defined as the deviation between agents beliefs about fundamental growth and the expectation of an unbiased econometrician. We use agent-specific expectations on one-year GDP growth, $E_t^i\left(gdp_{t+1}\right)$, and compare them to the expectation implied by a benchmark model, $E\left(gdp_{t+1}|\mathcal{M}_t\right)$. To compute one-year model forecasts under the econometrician measure $\mathcal{M}_t$, we borrow from the empirical macro-finance literature and use a time-series $AR(4)$ model using quarterly realized GDP growth (see, for instance, Marcellino (2008) and references therein). Our measure of growth sentiment is then the difference between the survey and model-implied expectation of GDP growth for the average agent: $\mathcal{S}_t^{gdp} = E_t^i\left(gdp_{t+1}\right) - E\left(gdp_{t+1}|\mathcal{M}_t\right)$, for i equal to the consensus. Positive values of sentiment $\mathcal{S}_t^{gdp}$ correspond to an average *optimism* in subjective beliefs about growth with respect to an econometrician.

Does this sentiment in the cross section of forecasts explain the bias in the forecast errors for long-term bond excess returns? Panel (b) of Figure 5 shows the distribution of individual t-statistics of the slope coefficients in regressions of the individual forecast errors for the 10-year bond on our sentiment proxy:

$$FE_{t,t+1}^{i,10} = a_s^{i,10} + b_s^{i,10}\mathcal{S}_t^{gdp} + \epsilon_{i,t+1}^{10}. \tag{5}$$

The coefficient of sentiment is positive for 84% of the agents, and significant for around one fourth of them at the 5% level. This suggests that in periods of positive sentiment, i.e. optimism about fundamentals, most agents tend to under predict excess long-term bond returns and commit larger (positive) forecast errors.

Note however that t in the regression above is measured in quarters rather than months since GDP growth data are available only at quarterly frequency. Therefore, we check the robustness of the results to an alternative measure of sentiment $\mathcal{S}_t$ that is available monthly, based on 3-month rate forecasts: $\mathcal{S}_t = E_t^i\left(y_{t+1}^{3m}\right) - E\left(y_{t+1}^{3m}|\mathcal{M}_t\right)$ for i equal to the consensus agent. This measure captures the average optimism in subjective beliefs about short rates with respect to an econometrician, whose expectations are assumed to follow a benchmark unit-root model. Figure 6 shows that the two measures of sentiment, $\mathcal{S}_t$ and $\mathcal{S}_t^{gdp}$, are highly correlated, which is consistent with sentiment in bond markets to be related to agents' beliefs about the dynamics of fundamentals.[28] The largest deviations between individual forecasts and model-implied ones, i.e. the highest values of sentiment, seem to occur in the early 90s, in the early 2000s, and during the recent financial crisis, and these periods coincide with the largest GDP contractions in our sample.[29]

[Insert Figure 6 here.]

When using $\mathcal{S}_t$ as a measure of sentiment in regression (5), the coefficient of sentiment is positive for 81 out of 84 agents, and significant at the 5% level for 50 agents (see Panel (c) in Figure 5). Moreover, sentiment is significantly linked to the aggregate forecast error FE^n and explains around 15% of its variation, for the 10-year bond. The relationship between subjective risk premia and sentiment is further analysed in Section VII in the context of equilibrium models with heterogeneous beliefs.

The results above suggest that there is a *systematic* behavioral component in agents beliefs, that is at least partially responsible for the statistical rejection of the rational expectation hypothesis. Agents are optimistic on average in our sample (i.e. sentiment tends to be positive) and this drives an average positive forecast errors on bond excess returns.

Given the results in this section and the extent of cross-sectional heterogeneity in beliefs, the question of how beliefs should be aggregated to construct an empirical proxy of subjective bond risk premia for the representative agent is far from trivial. This is the topic of the next section.

V. Belief aggregation and subjective bond risk premia

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

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

[29] These results are related to the work of Cieslak (2017), who shows that 'entering recessions, agents systematically overestimate the future real rate and underestimate unemployment. These forecast errors induce a predictable component in realized bond excess returns' .

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 (2017)). Since pessimists cannot short-sell, their beliefs are not revealed (spanned) by equilibrium asset prices. The general equilibrium literature that studies economies where agents speculate on their (heterogeneous) beliefs argues, on the other hand, that in absence of short-selling constraints irrational agents eventually lose economic weight to the benefit of less biased agents. What matters is not agents' optimism but rather their accuracy. The superior accuracy of rational agents allows them to accumulate economic importance in the Pareto weights of the representative agent (see, e.g., Basak (2005), Buraschi and Jiltsov (2006), Jouini and Napp (2006), Xiong and Yan (2010), Chen, Joslin, and Tran (2012), Buraschi and Whelan (2017), Ehling, Gallmeyer, Heyerdahl-Larsen, and Illeditsch (2018), among others). This argument, consistent with the original "market selection hypothesis" by Friedman (1953) and Alchian (1950), implies that bond prices should span the beliefs of the most accurate agents (i.e. closest to the actual physical probability). As Alchian (1950) argues, "*Realized profits [...] are the mark of success and viability. It does not matter through what process of reasoning or motivation such success was achieved. The fact of its accomplishment is sufficient. This is the criterion by which the economic system selects survivors: those who realize positive profits are the survivors; those who suffer losses disappear.*" If some agents have been consistently more accurate than others, they would have been accumulating more economic weight in the pricing kernel. Thus, these beliefs, rather than the consensus ones, should be a better proxy of the expectation of the marginal agent.

In light of this argument, we use information on agents beliefs from both the time series and the cross section to build an alternative aggregate measure of subjective bond risk premia. First, at every month t we sort agents according to the level of their accuracy up to that date. Specifically, we compute the average of the accuracy ranking percentiles $\mathcal{R}_{i,t}$ up to t of all agents present in the panel at time t and form quartile portfolios of agents based on this past accuracy.[30] According to Friedman (1953) and Alchian (1950), agents relying on the

[30]Note that the realized forecast errors up to time t will be based on expectations formed up to time $t-12$

predictions of the top quartile agents have accumulated greater economic wealth as of time t. Then, we compute the average EBR for different bond maturities within each quartile. The EBR formed on the quartile with greatest past accuracy is our preferred measure of aggregate subjective bond risk premia, which we denote $EBR^{\star}$. Indeed, this selection procedure has the advantage of being in real time (it uses only past information), so it is not affected by look-ahead bias. When we compare the top 10 forecasters based on the average ex-ante (see Table V) and unconditional performance rankings (see left panel of Table IV), we find that the intersection of the two rankings is made of five out of ten forecasters. On the one hand, this shows that aggregating beliefs using an ex-ante approach does matter; on the other hand, the persistence in accuracy is such that the two rankings are highly correlated. It is also worth noting that the institutions which are in the top ten ranking based on average accuracy percentile but not in the top ten ranking based on the number of appearances in $EBR^{\star}$, like Huntington National Bank and RidgeWorth Capital Management, tend to have a smaller number of observations and are therefore naturally less likely to be selected on an ex-ante basis.[31]

[Insert Table V here.]

We find that $EBR^{\star}$ is significantly positively correlated to future realized excess returns, for all bond maturities, and it has higher predictive regression R-squared and lower RMSE than the consensus. The correlation between $EBR^{\star}$ and future realized returns is 15% and 13% for the 5 years and 10 years bond, respectively, and it increases to 40% and 32% in the second half of the sample, i.e. from April 2003. The improvement in the predictive performance of $EBR^{\star}$ over time can be due to the fact that most of the top performers unconditionally (see again the left panel of Table IV), like Thredgold, UBS and Goldman Sachs, start to contribute to the BCFF panel only after 2000. In fact, in the last part of the sample we also observe larger differences between the performances of $EBR^{\star}$ and the consensus, suggesting that since the beginning of the century it is even more important to select the appropriate agents rather than simply taking the simple average of all agents' expectations.

months, since we consider forecasts with a fixed one year horizon. We use an initial window of 2 years.

[31] $EBR^{\star}$ can be computed for the sample that spans the period December 1990 - July 2015. In this period, Huntington National Bank and RidgeWorth Capital Management contribute to the panel for a total of 105 and 41 months, respectively, while for example UBS and Fannie Mae have 183 and 266 monthly contributions, respectively.

We then compare the $RMSE$ for 10-year bond excess returns of our $EBR^{\star}$ measure with the real-time version of the standard Cochrane and Piazzesi factor based on parameters estimated using available information at time t. We find that $EBR^{\star}$ generates a $RMSE$ equal to 8.83 versus 9.35 for CP. Figure 7 shows the time series of our $EBR^{\star}$ measure against the corresponding realized excess returns and the real-time expectations based on the CP (bottom panels) and Slope (top panels) factors, for a 5-year (left panels) and a 10-year (right panels) bond.

[Insert Figure 7 here.]

It is clear from these figure that $EBR^{\star}$ has a small downward bias, meaning that our proxy for the representative agents tends to underestimate future excess bond returns. In fact, the subjective risk premium is almost always below the unconditional mean of the realized excess return, denoted by the horizontal dashed line in the graphs. However, the same bias is visible in the real-time CP estimate in the bottom panels, which is actually also highly correlated with $EBR^{\star}$, in particular in the first part of the sample. Notably, in the periods in which $EBR^{\star}$ and the CP-based expectation diverge (e.g. 2003-2009) are characterized by a larger accuracy of the survey-based estimate, leading to the overall lower RMSE. The slope-based excess return expectation in the top panels is much smoother and it does not seem to display a downward bias. However, it is still highly correlated with $EBR^{\star}$, with correlations of about 46% and 29% for the 5 and 10-year bonds, respectively. The correlation of $EBR^{\star}$ with the slope factor is higher than with CP and more persistent over the sample period.

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

VI. Rationality of the Marginal Agent

The beliefs of the average and marginal agent may differ in certain economic periods and in equilibrium what matters for asset prices are those of the marginal agent. In what follows, we use our measure of $EBR_t^\star$ to design a direct test of rational expectations that can address several interesting questions. For instance, what is the extent to which agents forecast errors are predictable ex-ante? Can they be easily corrected using public information available at the time these predictions are made? How does statistical significance in these tests map into economic significance? These questions relate to the debate about the significance of behavioural biases in agents expectations and the empirical plausibility of the assumption of rational expectation in asset pricing models.

A. Rational Expectations Tests

In order to investigate the first question, we test whether future forecasts errors $FE_{t,t+1}^n$ are predictable by information available at time t. Forecast errors are defined as the difference between future realized and expected excess returns, $FE_{t,t+1}^n = rx_{t,t+1}^n - EBR_{n,t}^\star$. We investigate the extent to which one could use information contained in the term structure of bond yields at the time agents form their predictions. The term structure literature typically argues 3 principle components extracted from the cross section of date t yields can serve as orthogonal factors spanning the primitive shocks to the economy. More recent term structure models have argued for the existence of hidden factors which have negligible impact in the cross section but contain valuable information about future expected returns (Joslin, Priebsch, and Singleton (2014) and Duffee (2011b)). Indeed, Cochrane and Piazzesi (2005) argue for the existence of significant information about bond returns contained in the low order 4th and 5th principle components of yields. Motivated by these findings we assume that the date t information set is summarised by five principle components (PCs) and estimate the set of complimentary regressions

$$EBR_{n,t}^\star = \alpha_1 + \beta_1^\top PCs + \varepsilon_t \tag{6}$$

$$rx_{t,t+1}^n = \alpha_2 + \beta_2^\top PCs + \eta_{t+1} \tag{7}$$

$$FE_{t,t+1}^n = \alpha_3 + \beta_3^\top PCs + \nu_{t+1} \tag{8}$$

where the factor loadings on the forecast errors are mechanically linked by $\beta_3^\top = \beta_2^\top - \beta_1^\top$. Table VI reports the results for bond maturity n of 2, 5 and 10 years. Adjusted R-squared of the regressions that includes 5 principle components is reported in the $\overline{R}^2$ column, while the last column reports the change in the R-squared when moving from the first 3 PCs to all 5

PCs.

[Insert Table VI here.]

The top panel of table VI shows that the combination of 5 principle components explains a substantial proportion of the variation subjective bond risk premia, suggesting that beliefs are highly spanned by current bond prices. For all maturities loadings on PC1 (level) and PC2 (slope) are positive, economically large, and highly significant, and the $\overline{R}^2$ of the regressions range between 38% and 45%. The 4th and 5th principle components, however, are statistical insignificant (expect PC5 for the 10-year bond) and economically small. The middle panel repeats this exercise with realised returns on the left hand side. Again, we find loadings on PC1 (level) and PC2 (slope) which are are positive, economically large, and significant. However, different than for subjective expected returns but consistent with the findings of Cochrane and Piazzesi (2005) we find economically meaningful information about bond risk premia (estimated ex-post) contained in higher order 4th and 5th PCs. The final column ($\Delta\overline{R}^2$) reports the additional contribution in R-squared when moving from the first 3 to 5 principle components and shows the low order PCs are contributing to around half of the predictable variation of realized returns across maturities. The bottom panel reports predictability in forecast errors. In the case of 2-year bonds PC1 is significant at the 5% level, and in the case of 10-year bonds PC2 is significant at the 5% level, which is related to Piazzesi, Salomao, and Schneider (2015) who show that BCFF consensus forecast errors on long-term rates are positively linked to the slope of the term structure. However, the bulk of the predictable variation in errors in coming from the 4th and 5th principle components, consistent with their differential loading in the realized and expected returns regressions in the top two panels. The increase in explanatory power $\Delta\overline{R}^2$ shows that almost all predictable variation is coming from PC4 and PC5 for maturities $n = 2$ and $n = 5$, while for maturities $n = 10$ around half is coming from PC4/PC5 and half from the slope of the yield curve.

This result suggests that when forming beliefs, agents are not exploiting all available information. The low order PCs are 'unspanned' by agents expectations and thus predict realized returns since they are correlated with the 'unexpected' ex-post component. This may point towards a degree of irrationality in agents expectations and suggest they may be able to improve their long term bond forecasts by learning about their mistakes. Alternatively, one could argue that the predictive power of PC4 and PC5 is only a statistical feature of these data, detectable ex-post, but that should not be considered as agents form expectations in real time. In order to address this in more detail we investigate the economic significance of the predictable component of the forecast errors.

B. *Economic Significance of Deviation from Rational Expectations*

Statistical significance does not necessarily imply economic significance. To study the economic significance of behavioural components in agents expectations, we design an experiment in which we construct fictitious bond return expectations by correcting the predictable errors generated by $EBR^\star$ using information available in cross-section of date t yields.

We conduct a real-time experiment by initialising a rolling regression with a window of 5 years of data and recursively estimating a projection of realised errors on the cross-section of forward spreads.[32] The loadings available in the forecast error regression at date t can only be learned from errors realised 1-year ago. These loadings are then applied to date t forward spreads (F) in order to build a 'corrected' $EBR^\star$ from the following system

$$\widehat{FE}^{n}_{t-1,t} = \hat{\alpha}_{t-1} + \hat{\beta}^{\top}_{t-1} F_{t-1} \tag{9}$$

$$\xi_t = \hat{\alpha}_{t-1} + \hat{\beta}^{\top}_{t-1} F_t \tag{10}$$

$$\widehat{EBR}^{\star}_t = EBR^\star_t + \xi_t \tag{11}$$

The subscript t in the parameters $\hat{\alpha}_{t-12}$ and $\hat{\beta}^{\top}_{t-12}$ indicates that the correction is restricted to use only real-time information which is available at time t. The predictable component of the forecast errors is estimated using a rolling window to replicate real-life conditions of a trader.

We compute and compare the $RMSE$ implied by both the original $EBR^\star$ and the corrected $\widehat{EBR}^{\star}$. We find that, although the initial regressions indicate the existence of predictability in the forecast errors, the $RMSE$ of the corrected forecasts are unambiguously *higher* than the uncorrected ones. For instance, using a rolling window of 5 years in the estimation of the correction parameters, the RMSE increase by 74% for the 2-year bond and 42% for the 5-year bond and 17% for the 10-year bond.[33] This shows that the expectations embedded in $EBR^\star$

[32] The space of forward rates or yields spans the space of principle components but avoids computation of the PCs at each date which introduces unnecessary measurement error in the estimation.

[33] Figure 2 in the Supplemental Appendix shows the robustness of this finding to alternative window lengths.

cannot be easily improved using market based information. The empirical results of this section suggest that even when agents deviate from full information rationality, the uncorrected version of $EBR^{\star}$ seems to dominate its corrected counterpart, mainly in terms of variability, meaning that the apparent bias in agents beliefs is not easy to correct using the information available in real-time. This provides one explanation for why subjective expectation can be persistently different from full information rational expectations in the long run.

However, if we correct $EBR^{\star}$ only for the constant bias, i.e. the mean of the forecast errors over 5-year rolling windows, the RMSE of the forecast decreases by about 10% for the 10-year bond.[34] This is consistent with dogmatic persistent beliefs, with the average positive alpha in the predictive regression (2) and with the mass of positive betas in the regression (5) of forecast errors on sentiment. Indeed, the bias correction is highly correlated with a rolling average of short rate sentiment over the same rolling windows. For a 5-year bond, this correlation is about 37% over the full sample and 65% when we exclude the ZLB period.[35]

VII. Learning about the Determinants of Bond Risk Premia

The empirical evaluation of structural asset pricing models is traditionally conducted by approximating conditional expectations in the specification of risk premia with sample projections of future realizations rx_{t+h} onto observables variables which are part of the information set $\mathcal{F}_t$. This is potentially problematic for at least three reasons. First, sample projections based on future realizations can be quite different from true investors expectations. We have a clear example of this in the context of our data when we find that, at the individual level, $erx_{i,t}$ are more persistent than what a pure rational model would imply. Second, long-horizon predictability regressions give rise to overlapping errors which affect the estimators' properties. While it is possible to adjust the asymptotic properties of projection coefficients using well-known correction methods, these solutions do not address the inevitable challenge of the reduced number of genuinely independent observations. A regression of 5 year holding period returns on a 10 year sample has only two truly independent observations, even when the data is sampled daily.

While the magnitude of the corrections do change when varying the size of the window, the sign of the change and thus the message is robust. Correcting for the predictable component of FEs in real time always increases the RMSEs while correcting for a constant bias reduces RMSEs.

[34]The decrease in the RMSE correcting for the constant bias is larger, i.e. up to almost 20%, if we increase the length of the rolling window to 10 years.

[35]Short rate sentiment is mechanically close to zero during the ZLB period.

Finally, traditional predictive regressions with dependent variables constructed from future return realizations always raise the question of the extent to which in-sample results can be extended out-of-sample. At the same time, if in-sample regressions are plagued by look-ahead bias, out-of-sample regressions are typically exposed to the excess flexibility critique: the results are sensitive to the specific assumptions made to design the experiment.[36] For these reasons, as an alternative to future return realizations, we use as an ex-ante proxy for bond risk premia $EBR^{\star}_{n,t}$, which is based on the forecasts of the agents with greatest past accuracy, as described in section V. In addition to being unaffected by forward looking biases, previous results confirm that, out-of-sample, $EBR^{\star}_{n,t}$ are competitive in forecasting future realized excess returns relative to some popular reduced form models. Thus, we evaluate alternative models of risk premia based on their ability to explain the dynamics of $EBR^{\star}_{n,t}$, as opposed to sample averages (or projections) of future excess returns.

In this section we investigate regressions of the type

$$EBR^{\star}_{n,t} = a + b^{\top}\mathbf{X}_t + \varepsilon_{n,t}, \tag{12}$$

and consider three alternative specification for the dynamics of bond risk premia $\mathbf{X}_t$ that have been proposed by the equilibrium term structure literature.

A. Model Specifications

The literature on heterogeneous agents argues risk premia are affected by disagreement which operates primarily through the quantities of risk channel. Considering the role of disagreement about real shocks, Buraschi and Whelan (2017) derive bond pricing expressions showing that, if agents are sufficiently risk tolerant, speculation (disagreement) increases bond risk premia via a quantity of risk channel. Disagreement about nominal quantities may also matter for bond risk premia as Ehling, Gallmeyer, Heyerdahl-Larsen, and Illeditsch (2018) show in the context of a nominal exchange economy. In this case, disagreement about inflation can have real effects if agents are willing to trade on their beliefs. We proxy for real disagreement ($DiB(g)$) and nominal disagreement ($DiB(\pi)$) using the 4-quarter ahead cross-sectional inter-quartile range in GDP and CPI forecasts from our survey dataset.

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

In economies with external habit preferences, e.g. Campbell and Cochrane (1999), Wachter (2006), and Buraschi and Jiltsov (2007), time variation in risk compensation arises because of an endogenously time-varying price of risk. Shocks to the current endowment affect the wedge between consumption and habit, i.e. the consumption surplus, which induces time-varying expected returns. To obtain a proxy of risk premium $\mathcal{M}_t$, we follow Wachter (2006) and calculate consumption surplus ($Surp$) using a weighted average of 10 years of monthly consumption growth rates: $Surp = \sum_{j=1}^{120} \phi^j \Delta c_{t-j}$, where the weight is set to $\phi = 0.97^{1/3}$ to match the quarterly autocorrelation of the price-dividend ratio in the data.[37]

In long-run risk economies with recursive preferences (see e.g. Bansal and Yaron (2004)), time-varying risk premia are driven by 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 4-quarter 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 $LRR(\pi)$.

Le and Singleton (2013) discuss the common link between structural models where priced volatility risks impact expected bond returns. Indeed, the link between volatility risk and expected returns is a general statement (Merton (1980), French, Schwert, and Stambaugh (1987)). However, a well established puzzle in the fixed income literature is that this link is not born out in the data (Duffee (2002)). This has motivated a significant discussion challenging the ability of completely affine term structure models to explain bond risk premia and calling for extensions of these specifications. We re-visit this link using survey expectations of bond risk premia and proxy for interest rate volatility using intra-month sum of squared yield changes (returns) on a constant maturity n-year zero-coupon bonds, which we denote $\widehat{\sigma}(n)$.

In the context of equity markets, questions about return variation are often cast in terms of the log-linearised present value identity (Campbell and Shiller (1988)) which states that time variation in the price-dividend ratio is mechanically linked to discount rates or cash flow news. Since the price-dividend ratio is time-varying, either stock returns, dividend growth, or both are predictable. While the debate about the importance of the discount rate and cash flow channels

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

in return predictability is a long standing one (Cochrane (2007)) there is a general consensus that risk premia are time varying. Indeed, many equilibrium models have been designed to generate this feature. Motivated by this literature, we study the prediction that price-dividend ratios vary with discount rates using $EBR^{\star}_{n,t}$.[38] In fact, if $EBR^{\star}_{n,t}$ actually captures the pricing kernel of the representative agent it should be linked to the pricing of risk in the stock market as well. Following Goyal and Welch (2008), dividends in the computation of the log price-dividend ratio are seasonally-adjusted by taking the sum of the dividends over the previous 12 months.[39]

Finally, summarising the evidence presented in the sections above, we have shown that there is a systematic behavioural component in agents beliefs. On average agents have positive forecast errors on excess bond returns which results in a downward bias in real-time expectations. We linked this bias to optimism about GDP growth and an upward bias in short rate expectations. In the following we ask whether the structural proxies discussed here are linked to subjective bond risk premia, while allowing for a contemporaneous behavioural driver of agents' beliefs. We do this by controlling for short rate sentiment $\mathcal{S}_t$ in the regressions of $EBR^{\star}_{n,t}$ on the structural risk premium proxies.[40] In doing so, we explicitly take into account that subjective risk premia can be driven by both behavioural and rational elements, and we ask what fraction of the variation in risk premia do rational expectations risk factor proxies explain after controlling for an observable proxy of the behavioural bias in agents expectations.

B. Empirical Results

We run a series of multivariate regressions of $EBR^{\star}_t$ on the state variables implied by the model specifications above, based on a sample that ranges between December 1990 and July 2015.[41] Table VII reports the regression results for 10-year maturity bonds.[42]

[Insert Table VII here.]

[38]The log price-dividend ratio is used also by Greenwood and Schleifer (2014) as a rational measure of risk premia in the stock market.

[39]We obtain the with- and without-dividend monthly returns on the value-weighted portfolio of all NYSE, Amex and Nasdaq stocks from CRSP.

[40]GDP sentiment ($\mathcal{S}^{gdp}_t$) and short rate sentiment ($\mathcal{S}_t$) are highly correlated at quarterly frequencies but we choose to $\mathcal{S}_t$ which is available at the monthly frequency.

[41]This is the time period for which $EBR^{\star}_{n,t}$ is available, since we used an initial window of 2 years for the computation of past average accuracy percentiles, and there is a lag of 12 months between expectations and realizations.

[42]Table III in the Supplemental Appendix reports the results for the 5-year maturity bond which are close to the results for the 10-year bond.

The top two rows report estimates from a regression of $EBR^{\star}_{10,t}$ on sentiment: the point estimate is negative and significant at the 5% level, with a small R-squared of about 3%. This result is consistent with our findings above that there is a slowly moving predictable component in agents errors resulting from a downward bias in the expectations about bond returns.

The role played by differences in beliefs is reported in specification (I). Consistent with the prediction of heterogeneous beliefs models both disagreement about real growth and inflation are positive and significant. In the case of $DiB(g)$ the point estimate is highly significant at the 1% level while $DiB(\pi)$ is significant at the 5% level and the R-squared of the regression is around 11%. Interestingly, controlling for $\mathcal{S}_t$ the estimate of $DiB(g)$ is unaltered but the significance of $DiB(\pi)$ rises to the 1% level and the $\overline{R}^2$ rises to 15%.

When agents have habit preferences, the price of risk is state-dependent and negatively related to the consumption surplus ratio. Specification (II) shows that the slope coefficient in this regression does have the correct sign. However, both the t-statistic and $\overline{R}^2$ are rather small: the t-statistics for the coefficient on $Surp$ is -1.46 and the R-squared is 1.18%. Interestingly, however, after we take into account the effect of sentiment, the t-statistic on $Surp$ rises to -1.81 with an $\overline{R}^2$ of 4.55%.

Specification (III) focuses on the significance of proxies of economic growth and inflation uncertainty, $LRR(g)$ and $LRR(\pi)$, as suggested by long-run risk models. We find that using $EBR^{\star}_t$ as dependent variable the statistical significance of inflation uncertainty is quite remarkable, with a t-statistic equal to 4.34 for the 10 year bond and an $\overline{R}^2$ equal to 23.09%. Controlling for $\mathcal{S}_t$ increases the strength of this result both in terms of statistical significant and the regression $\overline{R}^2$ which rises to 31.05%. Larger values of long-run economic uncertainty about inflation are correlated with greater subjective expected bond risk premia. This is consistent, for instance, with the model discussed in Bansal and Yaron (2004) in which greater uncertainty raises interest rates, lowers bond prices and increases future expected bond returns. At the same time, the loading on real uncertainty is economically small and statistically insignificant, regardless of whether we control for $\mathcal{S}_t$ or not.

Specification (IV) re-examines the link between bond volatility and expected returns. The regression results show that the quantity of risk channel is significant when tested on $EBR^{\star}_{10,t}$ rather than realisations. The R-squared is quite high, around 12%, and the t-statistic of 5.85 is highly significant. Importantly, the point estimate is also positive consistent with theory that predicts investors demand compensation for holding volatility risk. Controlling for sentiment raises the R^2 to about 17% while both factor loadings remain highly statistically significant. This result is interesting in its own right since it suggests that term structure models in which the quantity of risk plays a role should not be dismissed.

Finally, specification (V) tests for a statistical correlation between $EBR^{\star}_{10,t}$ and the PD ratio. Remarkably, in a univariate regression, we obtain a negative factor loading with a t-statistic of -3.30 and an $\overline{R}^2$ of 6.50%. Consistent with the discount rate channel explanation for PD predictability, we find a drop in equity prices relative to past dividend payouts is positively correlated with subjective bond risk premia. As with volatility, controlling for sentiment raises the R^2 and both factor loadings are statistically significant.

To summarize, these findings show empirical proxies implied by equilibrium models explain the dynamics of subjective bond risk premia. Moreover, consistent with rational pricing of risk, we document a positive significant link between $EBR^{\star}_{10,t}$ and bond volatility, and a negative statistical link with the price-dividend ratio. This result contrasts with previous studies for equity returns which argue that equilibrium models generate implied risk premia that correlate negatively with survey-implied risk premia. For instance, Greenwood and Schleifer (2014) use equity market data and find a negative correlation between model-implied equity risk premia and survey expectations. They interpret their result as clear evidence of a rejection of rational expectations models: '*We can reject this hypothesis with considerable confidence. This evidence is inconsistent with the view that expectations of stock market returns reflect the beliefs or requirements of a representative investor in a rational expectations model.*' However, we also find a significant role for a behavioural component in investors beliefs, as captured by a real-time proxy for sentiment. This is an important point: when distinguishing between behavioural and rational theories of financial economics, one should not dismiss these channels as being mutually exclusive. Indeed, the findings here suggest that both are driving subjective bond risk premia.

VIII. Conclusion

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

First, we show that individual subjective bond risk premia are heterogeneous and that forming consensus average beliefs disregards important information in the cross-section of beliefs. For example, valuable information regarding motives for trade is lost when taking arithmetic averages: at the individual level we document large individual persistence in agents beliefs about whether bonds are insurances or risky bets.

Second, we study predictability in long-term bond excess returns and show the accuracy of the best forecasters is persistent over time. Interestingly, the most accurate agents at forecasting the long end of the term structure are not the best in predicting short-term rates. In particular, we find that primary dealers are more likely to be between the top forecasters of the short-term interest rate but they do not have a persistent edge in predicting long-term bond returns. This finding supports the idea that time variation in bond risk premia plays an important role in the predictability of long-term bond returns.

Third, we take into account the persistent heterogeneity and forecasting skill in the cross section of beliefs by constructing an aggregate measure of subjective bond risk premia based on past accuracy, consistent with the market selection hypothesis that accurate agents accumulate weight in the pricing kernel. We then use our measure, that we denote as $EBR^{\star}$ to test the rationality of the marginal agent and to evaluate and compare alternative macro-finance models.

In order to test the rationality of $EBR^{\star}_t$, we compute forecast errors and analyze their predictability by information included in bond prices at time t. We find statistical evidence of predictability in forecast errors by a combination of forward rates, in violation of rational expectations. However, when we try to correct $EBR^{\star}$ by the predictable component of forecast errors we find that the $RMSE$ of the corrected forecasts are not unambiguously lower than the uncorrected ones, suggesting that the expectations embedded in $EBR^{\star}$ cannot be easily improved in real-time.

Finally, we substitute $EBR^{\star}$ for future realized excess returns as the dependent variable to evaluate alternative models. Surprisingly, we find support for both behavioural and rational determinants of bond risk premia. In particular, we demonstrate the importance of a sentiment bias in agents beliefs in addition to structural models in which risk premia vary with the dynamics of the quantities of risk channel.

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.

Andrade, P., R. K. Crump, S. Eusepi, and E. Moench, 2014, Fundamental disagreement, *Staff Report: Federal Reserve Bank of New York*.

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.

Bacchetta, Philippe, Elmar Mertens, and Eric Van Wincoop, 2009, Predictability in financial markets: What do survey expectations tell us?, *Journal of International Money and Finance* 28, 406–426.

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, M. D., and J. D. Hamilton, 2015, Robust bond risk premia, *working paper*.

Bhamra, Harjoat S, and Raman Uppal, 2014, Asset prices with heterogeneity in preferences and beliefs, *Review of Financial Studies* 27, 519–580.

Bonham, C., and R. Cohen, 2001, To aggregate, pool, or neither: Testing the rational-expectations hypothesis using survey data, *Journal of Business & Economics Statistics* 19, 278–291.

Borovicka, J., 2011, Survival and long-run dynamics with heterogeneous beliefs under recursive preferences, Discussion paper, Working Paper.

Buraschi, A., and A. Jiltsov, 2006, Model uncertainty and option markets with heterogeneous beliefs, *The Journal of Finance* 61, 2841–2897.

——— , 2007, Habit formation and macroeconomic models of the term structure of interest rates, *Journal of Finance* 62, 3009 – 3063.

Buraschi, A., F. Trojani, and A. Vedolin, 2014, Economic Uncertainty, Disagreement, and Credit Markets, *Management Science* 60, 1281–1296.

Buraschi, A., and P. Whelan, 2017, Sentiment, speculation, and interest rates, *Working paper*.

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

Campbell, John Y, and Robert J Shiller, 1988, The dividend-price ratio and expectations of future dividends and discount factors, *The Review of Financial Studies* 1, 195–228.

Campbell, J. Y., A. Sunderam, and L. M. Viceira, 2017, Inflation bets or deflation hedges? the changing risks of nominal bonds, *Critical Finance Review*.

Capistran, C., and A. Timmermann, 2009, Disagreement and biases in inflation expectations, *Journal of Money, Credit and Banking* 41, 365–396.

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

Chernov, Mikhail, and Philippe Mueller, 2012, The term structure of inflation expectations, *Journal of financial economics* 106, 367–394.

Cieslak, A., 2017, Short-rate expectations and unexpected returns in treasury bonds, *Working paper*.

Cochrane, J.H., and M. Piazzesi, 2005, Bond risk premia, *American Economic Review* 95, 138–160.

Cochrane, John H, 2007, The dog that did not bark: A defense of return predictability, *The Review of Financial Studies* 21, 1533–1575.

Cochrane, John H., and Monika Piazzesi, 2008, Decomposing the yield curve, *Working paper.*

Coibion, O., and Y. Gorodnichenko, 2015, Information rigidity and the expectation formation process: A simple framework and new facts, *American Economic Review* 105, 2644–2678.

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.

Das, S, C. Kuhnen, and S. Nagel, 2017, Socioeconomic status and macroeconomic expectations, *Working paper.*

David, A., 2008, Heterogeneous beliefs, speculation, and the equity premium, *The Journal of Finance* 63, 41–83.

Detemple, J., and S. Murthy, 1994, Intertemporal asset pricing with heterogeneous beliefs, *Journal of Economic Theory* 62, 294–320.

Dovern, J., U. Fritsche, P. Loungani, and N. Tamirisa, 2015, Information rigidities: Comparing average and individual forecasts for a large international panel, *International Journal of Forecasting* 31, 144–154.

Duffee, G.R., 2002, Term premia and interest rate forecasts in affine models, *The Journal of Finance* 57, 405–443.

——— , 2011a, Information in (and not in) the term structure, *Review of Financial Studies* 24(9), 2895–2934.

Duffee, Gregory R., 2011b, Information in (and not in) the term structure, *Review of Financial Studies* 24, 2895–2934.

Duffee, Gregory R, 2013, Bond pricing and the macroeconomy, in *Handbook of the Economics of Finance*vol. 2 . pp. 907–967 (Elsevier).

Dumas, B., A. Kurshev, and R. Uppal, 2009, Equilibrium portfolio strategies in the presence of sentiment risk and excess volatility, *The Journal of Finance* 64, 579–629.

Ehling, Paul, Michael Gallmeyer, Christian Heyerdahl-Larsen, and Philipp Illeditsch, 2018, Disagreement about inflation and the yield curve, *Journal of Financial Economics.*

Figlewski, Stephen, and Paul Wachtel, 1983, Rational expectations, informational efficiency, and tests using survey data: A reply, *The Review of Economics and Statistics* pp. 529–531.

Frankel, Jeffrey A, and Kenneth A Froot, 1987, Short-term and long-term expectations of the yen/dollar exchange rate: evidence from survey data, *Journal of the Japanese and International Economies* 1, 249–274.

French, Kenneth R, G William Schwert, and Robert F Stambaugh, 1987, Expected stock returns and volatility, *Journal of financial Economics* 19, 3–29.

Friedman, M., 1953, *Essays in Positive Economics* (University of Chicago Press, Chicago).

Froot, K., 1989, New hope for the expectations hypothesis of the term structure of interest rates, *Journal of Finance* 44, 283–305.

Gallmeyer, M., and B. Hollifield, 2008, An examination of heterogeneous beliefs with a short-sale constraint in a dynamic economy, *Review of Finance* 12, 323–364.

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, Harrison, David Sraer, and Jialin Yu, 2017, Inflation bets on the long bond, *The Review of Financial Studies* 30, 900–947.

Hong, H., and J.C. Stein, 2003, Differences of opinion, short-sales constraints, and market crashes, *Review of Financial Studies* 16, 487–525.

Joslin, Scott, Marcel Priebsch, and Kenneth J Singleton, 2014, Risk premiums in dynamic term structure models with unspanned macro risks, *The Journal of Finance* 69, 1197–1233.

Jouini, E., J.M. Marin, and C. Napp, 2010, Discounting and divergence of opinion, *Journal of Economic Theory* 145, 830–859.

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.

Keane, M. P., and D. E. Runkle, 1990, Testing the rationality of price forecasts: New evidence from panel data, *American Economic Review* 80, 714–735.

Koijen, R., M. Schmeling, and E. Vrugt, 2015, Survey expectations of returns and asset pricing puzzles, *Working paper.*

Le, A., and K.J. Singleton, 2013, The structure of risks in equilibrium affine models of bond yields, Discussion paper, Working Paper.

Mackowiak, B., and M. Wiederholt, 2009, Optimal sticky prices under rational inattention, *American Economic Review* 99, 769–803.

Malmendier, U., and S. Nagel, 2011, Depression Babies: Do Macroeconomic Experiences Affect Risk-Taking?, *Quarterly Journal of Economics* 126, 373–416.

Mankiw, N.G., and R. Reis, 2002, Sticky information versus sticky prices: A proposal to replace the new keynesian phillips curve, *Quarterly Journal of Economics* 117, 1295–1328.

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

Merton, Robert C, 1980, On estimating the expected return on the market: An exploratory investigation, *Journal of financial economics* 8, 323–361.

Patton, A. J., and A. Timmermann, 2010, Why do Forecasters Disagree? Lessons from the Term Structure of Cross-Sectional Dispersion, *Journal of Monetary Economics* 57, 803–820.

Piazzesi, M., J. Salomao, and M. Schneider, 2015, Trend and cycle in bond premia, Discussion paper, Working paper.

Scheinkman, J.A., and W. Xiong, 2003, Overconfidence and speculative bubbles, *Journal of Political Economy* 111, 1183–1220.

Sims, C. A., 2003, Implications of rational inattention, *Journal of Monetary Economics* 50, 665–690.

Svensson, L., 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.

Woodford, M., 2002, *Knowledge, Information, and Expectations in Modern MacroeconMacro: In Honor of Edmund S. Phelps* . chap. Imperfect Common Knowledge and the Effects of Monetary Policy (Princeton University Press).

Xiong, W., and H. Yan, 2010, Heterogeneous expectations and bond markets, *Review of Financial Studies* 23, 1433–1466.

Zellner, A., 1962, Estimation of cross-section relations: Analysis of a common specification error, *Metroeconomica* 14, 111–117.

IX. Tables

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

Table I. Transition Probabilities
Probability of a forecaster transitioning from a given quartile of the cross-sectional distribution of 3-month yield (left), 2-year bond excess returns (middle) and 10-year bond excess returns (right) forecasts to another quartile in the following month.

	WORST	CONS	BEST	*Slope*	*CP*
			Panel A: Full Sample		
ALL	10.56	8.60	7.75	7.71	9.37
Expansions	10.82	8.89	8.01	7.99	9.09
Recessions	8.19	5.73	5.15	5.04	11.37
			Panel B: Excluding ZLB		
ALL	9.77	8.03	7.28	8.06	9.63
Expansions	9.88	8.26	7.51	8.38	9.37
Recessions	7.95	4.83	3.99	4.18	10.58

Table II. Accuracy of Survey Forecasts vs Slope vs CP
Root mean square prediction errors for 10-year bond excess returns, based on forecasts from the unconditional worst 10 forecasters (WORST), the best 10 forecasters (BEST), and the simple average of survey expectations (CONS), i.e. the consensus. The top panel reports RMSEs for all months in the sample period covering January 1988 to July 2015 (331 observations), and the bottom panel excludes the zero lower bound subsample January 2009 to July 2015 (leaving 253 observations). *Recessions* report RMSEs coming from forecasts made during NBER recessions (top panel = 37 , bottom panel = 18) and *Expansions* report RMSEs for all other dates.

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

Table III. Accuracy Transition Probabilities

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

	Top 10-year Bond Forecasters	Top Short Rate Forecasters
(1)	Thredgold Economic Assoc.	Goldman Sachs
(2)	UBS	GLC Financial Economics
(3)	Goldman Sachs	RidgeWorth Capital Management
(4)	Huntington National Bank	Economist Intelligence Unit
(5)	Fleet Financial Group	Tucker Anthony, Inc.
(6)	RidgeWorth Capital Management	J.P. Morgan
(7)	Fannie Mae	BMO Capital Markets
(8)	DePrince & Assoc	Societé Generale
(9)	J.W. Coons & Assoc	Aubrey G. Lanston & Co.
(10)	Nomura Securities Inc.	Nomura Securities Inc.

Table IV. Top 10 Forecasters

The left panel of this table presents the top 10 forecasters in terms of average accuracy percentile ranking for the 10-year bond excess returns, over the full sample. The right panel shows the top 10 short rate forecasters in terms of average accuracy percentile ranking over the full sample, for the 3-month yield. We consider only forecasters who contribute to the panel for at least 5 years.

	Most Represented in EBR*
(1)	Fannie Mae
(2)	Thredgold Economic Assoc.
(3)	UBS
(4)	Bank of America Securities
(5)	Goldman Sachs
(6)	The Northern Trust Company
(7)	Loomis Sayles & Co.
(8)	National Association of Realtors
(9)	DePrince & Assoc
(10)	BMO Capital Markets

Table V. Who is in $EBR^{\star}$?

This table shows the 10 agents who are more often (in terms of number of months) present in our $EBR^{\star}$ index, which conditionally selects the top quartile of forecasters based on past accuracy percentiles on the 10-year bond excess returns.

Maturity	constant	PC1	PC2	PC3	PC4	PC5	$\overline{R}^2$	$\Delta\overline{R}^2$
$EBR^{\star}_{2,t}$	0.24	0.18	0.24	-0.12	-0.05	-0.01	48.38	0.92
	(4.19)	(3.51)	(5.20)	(-3.25)	(-0.93)	(-0.37)		
$EBR^{\star}_{5,t}$	0.41	0.54	0.84	-0.07	-0.02	0.01	37.72	-0.41
	(2.11)	(3.20)	(5.75)	(-0.60)	(-0.10)	(0.05)		
$EBR^{\star}_{10,t}$	1.36	1.55	1.37	0.96	-0.50	-0.50	44.99	3.86
	(4.32)	(6.69)	(5.74)	(4.34)	(-1.60)	(-2.40)		
rx^{2}_{t+1}	0.77	0.46	0.18	0.07	0.37	-0.27	26.50	13.72
	(4.22)	(3.15)	(1.05)	(0.47)	(2.44)	(-3.09)		
rx^{5}_{t+1}	2.83	0.83	1.26	-0.10	1.11	-1.28	29.43	16.91
	(4.00)	(2.21)	(2.36)	(-0.22)	(2.81)	(-4.24)		
rx^{10}_{t+1}	5.08	0.95	3.35	-0.20	1.31	-2.34	32.36	12.29
	(5.36)	(1.34)	(4.09)	(-0.26)	(2.60)	(-3.97)		
FE^{2}_{t+1}	0.53	0.27	-0.05	0.19	0.43	-0.25	18.68	15.59
	(2.69)	(1.97)	(-0.27)	(1.16)	(2.71)	(-2.69)		
FE^{5}_{t+1}	2.42	0.29	0.41	-0.03	1.13	-1.28	17.16	16.64
	(3.81)	(0.74)	(0.69)	(-0.06)	(2.67)	(-4.05)		
FE^{10}_{t+1}	3.72	-0.61	1.98	-1.16	1.81	-1.83	18.97	10.49
	(3.42)	(-0.73)	(2.15)	(-1.37)	(3.09)	(-2.68)		

Table VI. Expected Returns, Realised Returns, and Forecast Error Predictability
Estimates from regressions of subjective expected returns (top panel), realized returns (middle panel) and forecast errors (bottom panel) on the principle components of yields (PCs). t-statistics, reported in parentheses below the point estimates, are Newey-West corrected. Adjusted R-squared of the regressions that includes 5 principle components is reported in the $\overline{R}^2$ column. The final column reports the change in the R-squared when moving from the first 3 PCs to all 5 PCs. The sample period is from December 1990 to July 2015.

	$\mathcal{S}$	$DiB(g)$	$DiB(\pi)$	$Surp$	$LRR(g)$	$LRR(\pi)$	$\widehat{\sigma}$	PD	$\overline{R}^2$
	−0.17								2.56
	(−2.20)								
(I)		0.23	0.15						10.57
		(2.80)	(1.96)						
	−0.20	0.23	0.17						14.68
	(−3.02)	(2.82)	(2.39)						
(II)				−0.12					1.18
				(−1.46)					
	−0.19			−0.15					4.55
	(−2.58)			(−1.81)					
(III)					0.10	0.43			23.09
					(1.36)	(4.34)			
	−0.30				0.01	0.55			31.05
	(−5.06)				(0.12)	(5.34)			
(IV)							0.35		12.14
							(5.85)		
	−0.23						0.39		17.40
	(−3.32)						(5.79)		
(V)								−0.25	6.50
								(−3.30)	
	−0.20							−0.28	10.53
	(−2.75)							(−3.66)	

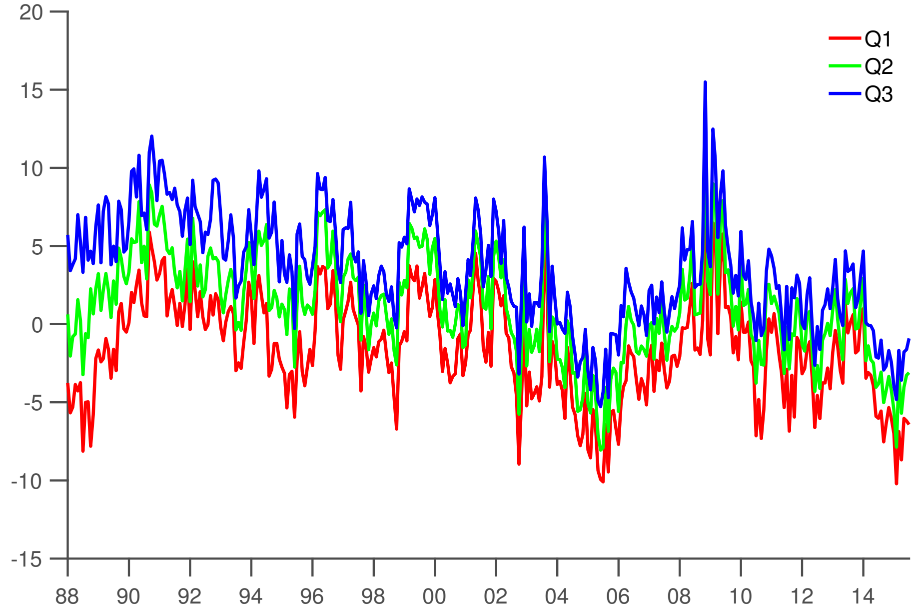
Table VII. Determinants of Ex-Ante Subjective Bond Returns

This table reports estimates from regressions of the subjective expected excess returns on 10-year bonds on a set of explanatory variables:

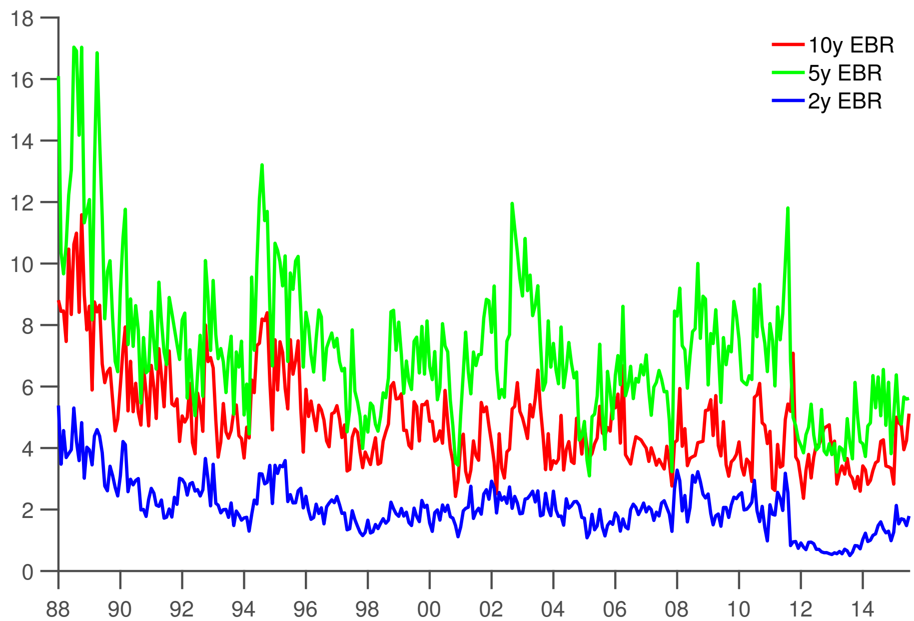
$$EBR^{\star}_{10,t} = a + b^{\top}\mathbf{X}_t + \epsilon_{10,t}.$$

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

X. Figures



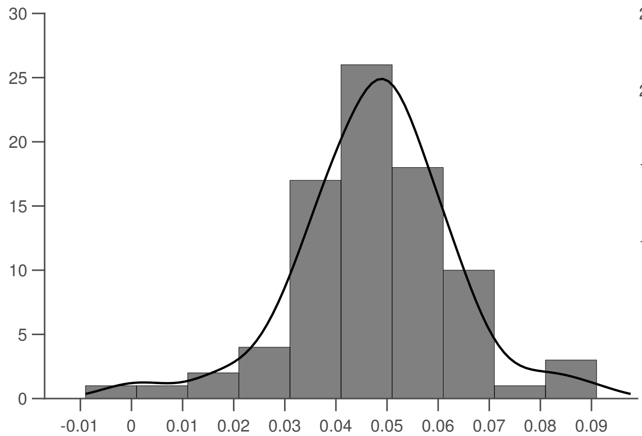
(a)



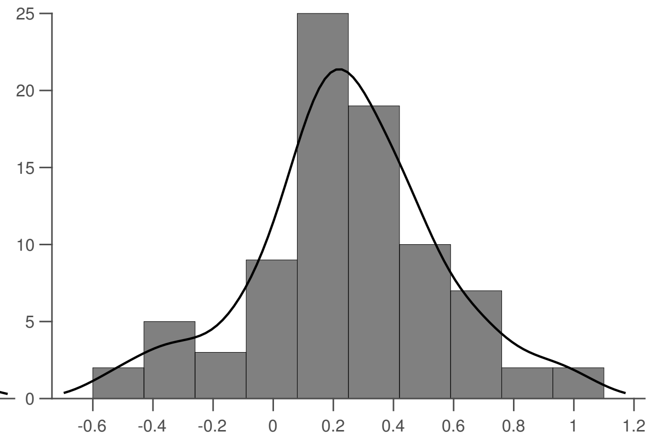
(b)

Figure 1. Subjective Expectations

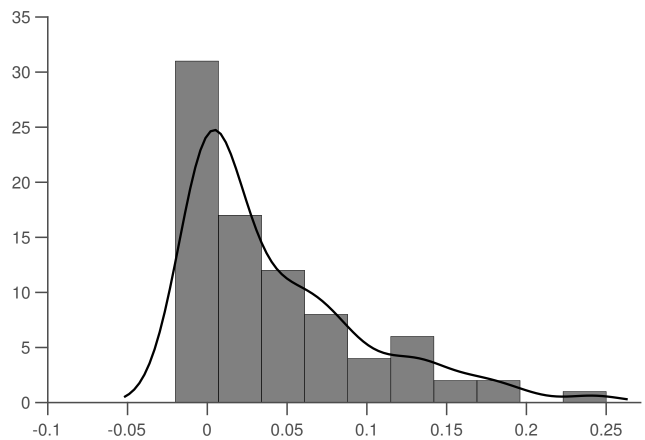
The top panel plots quartiles (Q1, Q2(median) and Q3) of the cross-sectional distribution of 1-year subjective expected excess returns for 10-year maturity bonds. The bottom panel plots disagreement about expected bond returns for maturities 2 , 5 and 10-year, defined as the cross-sectional interquartile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



(a) α_i^n



(b) β_i^n



(c) R^2

Figure 2. 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:

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

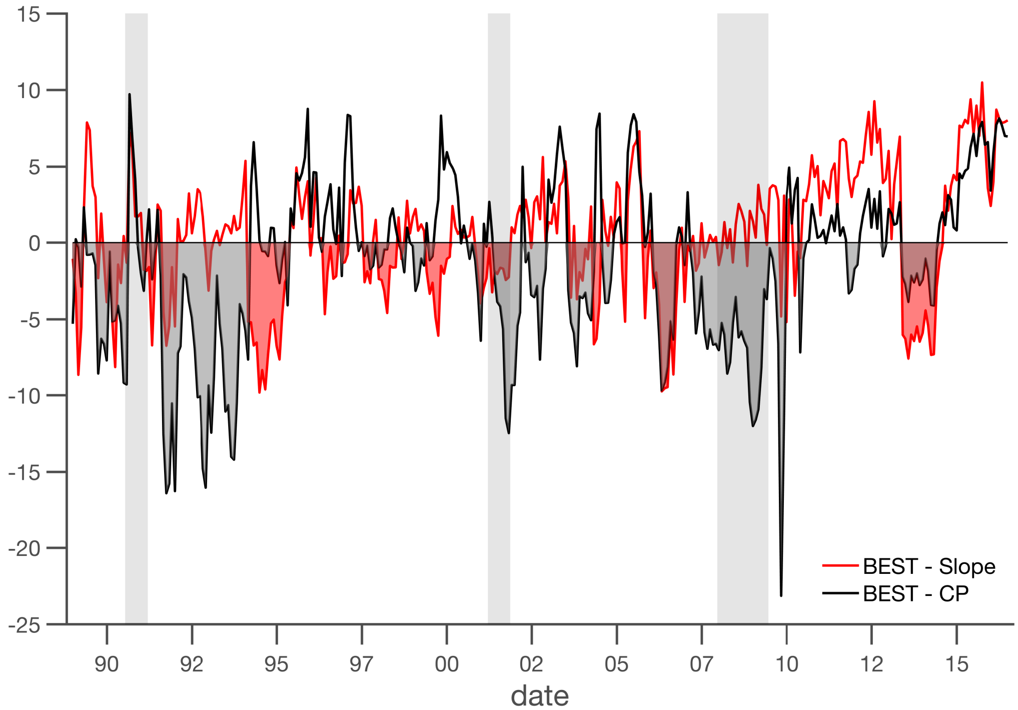
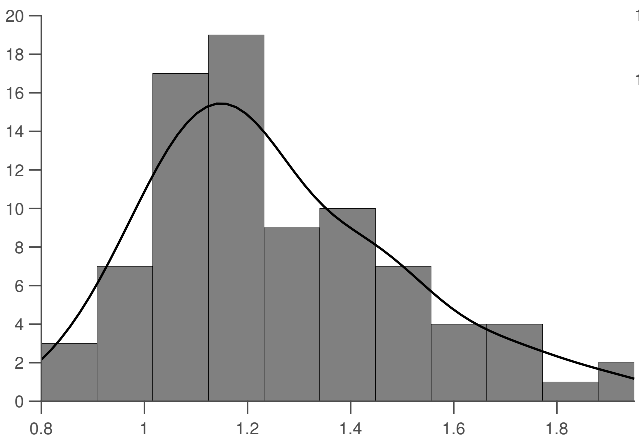
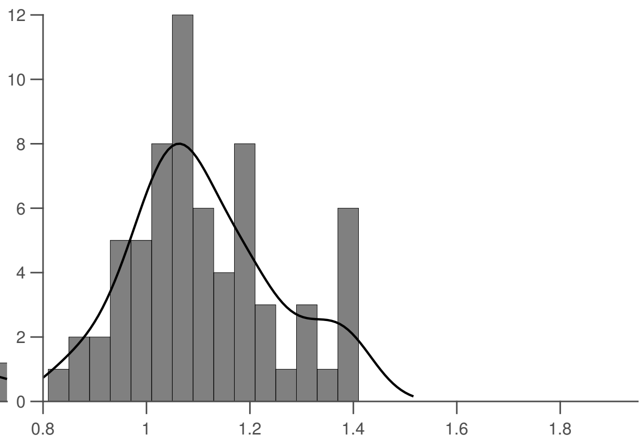


Figure 3. Absolute Error Differences

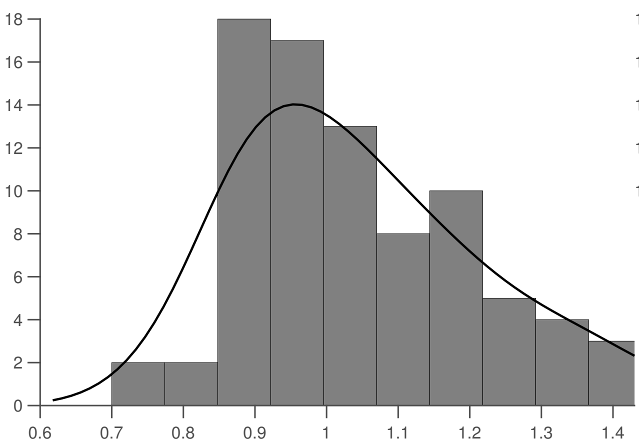
Differences in absolute 10-year bond return forecast errors between the top 10 forecasters and the errors from a real-time forecast implied by the slope of the term structure, and the top 10 forecasters and a real time forecast implied by the CP return forecasting factor.



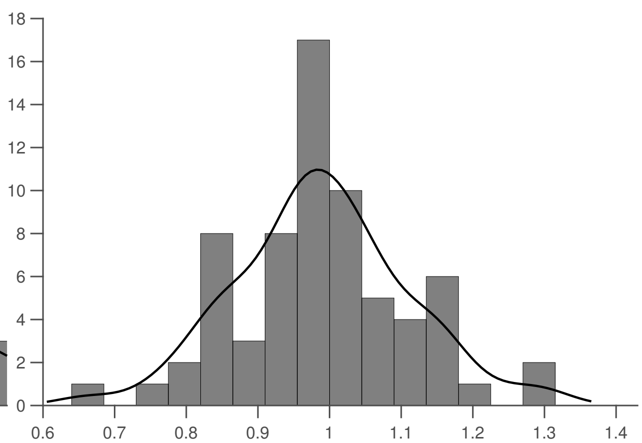
(a) Slope Full Sample



(b) Slope Excl. ZLB



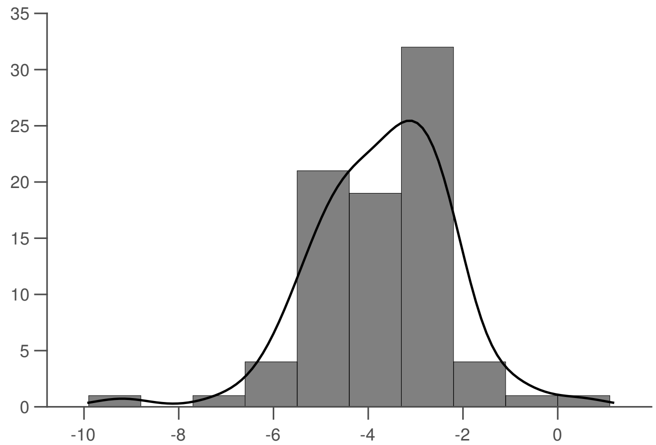
(c) CP Full Sample



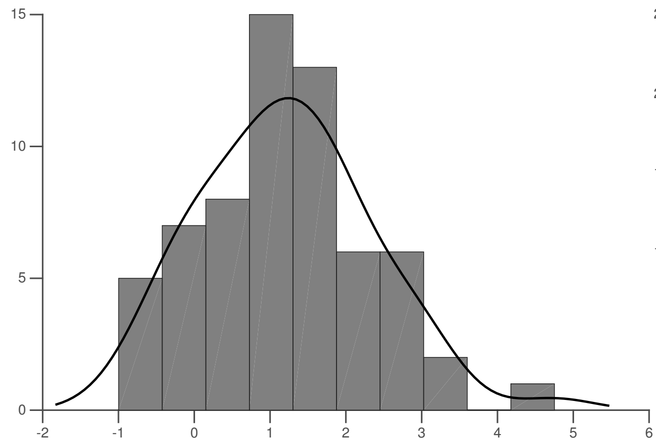
(d) CP Excl. ZLB

Figure 4. Relative Accuracy

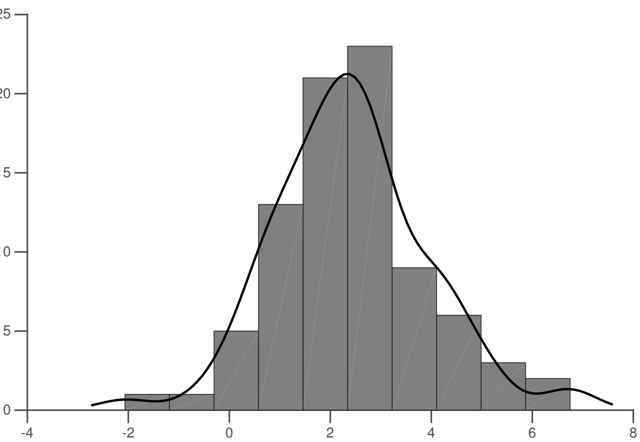
Histograms of the relative accuracy $\mathcal{A}_i^{10}$ of each forecaster, that is, the ratio between the RMSE of each individual forecaster and the RMSE of a benchmark, for 10-year bond excess returns. The benchmark models we consider are a real-time forecast implied the slope (10-year yield minus 1-year yield) or the CP factor, for the period in which the forecaster is in the panel. We consider only the contributors with at least 60 months of forecasts, for a total of 84 institutions.



(a) $FE^{i,n}_{t,t+1}$ on $\Delta erx^{n}_{i,t}$

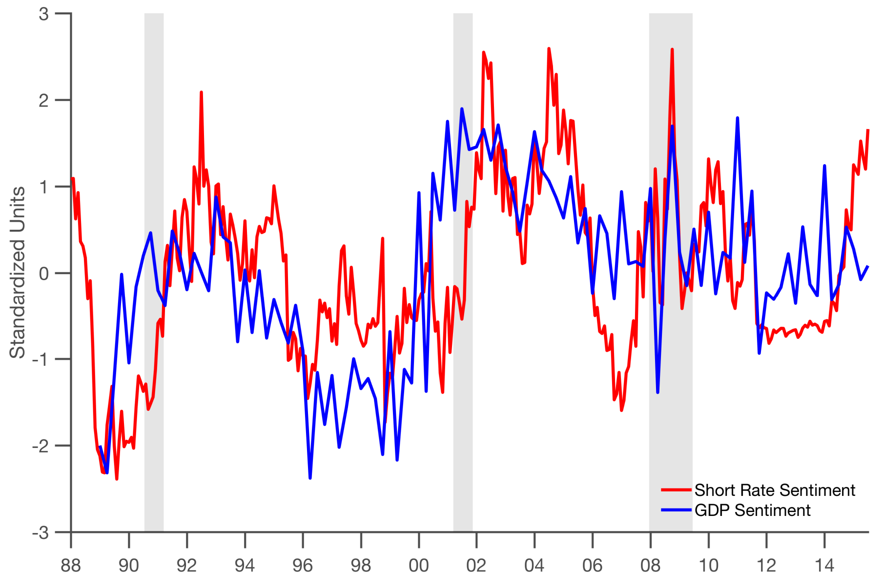


(b) $FE^{i,n}_{t,t+1}$ on $\mathcal{S}^{gdp}_{t}$

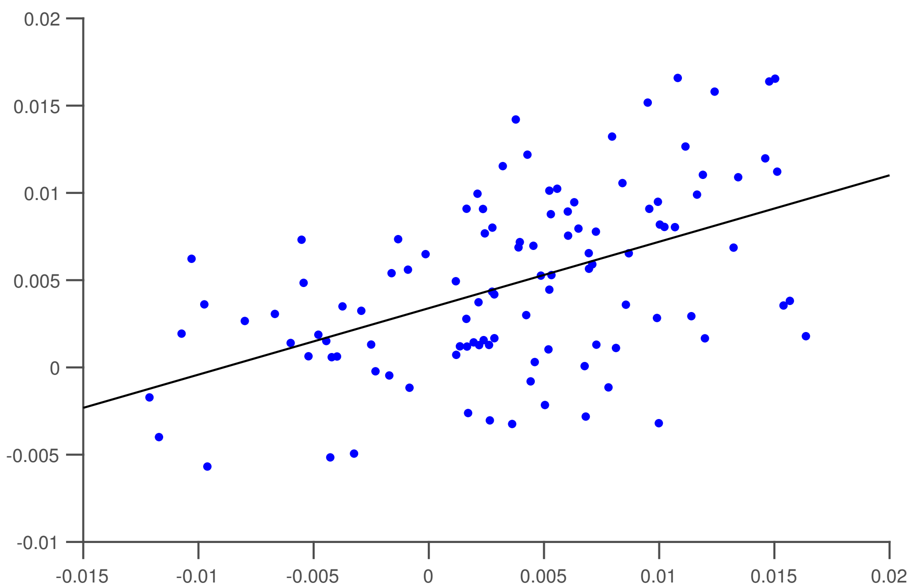


(c) $FE^{i,n}_{t,t+1}$ on $\mathcal{S}_{t}$

Figure 5. Individual Forecast Errors on Forecast Revisions and Sentiment
Panel (a) shows a histogram of the t-statistics of the slope coefficient in the regression of individual forecast errors on forecast revisions, for the 10-year excess bond return, focusing on the 84 agents with at least 60 monthly forecasts. Panel (b) shows a histogram of the t-statistics of the slope coefficient in the regression of the same individual forecast errors on growth sentiment, $\mathcal{S}^{gdp}_{t}$, where t is measured in quarters rather than months since GDP growth data are available only at quarterly frequency. Panel (c) uses an alternative measure of sentiment $\mathcal{S}_{t}$ based on short rate forecasts, which is available monthly.



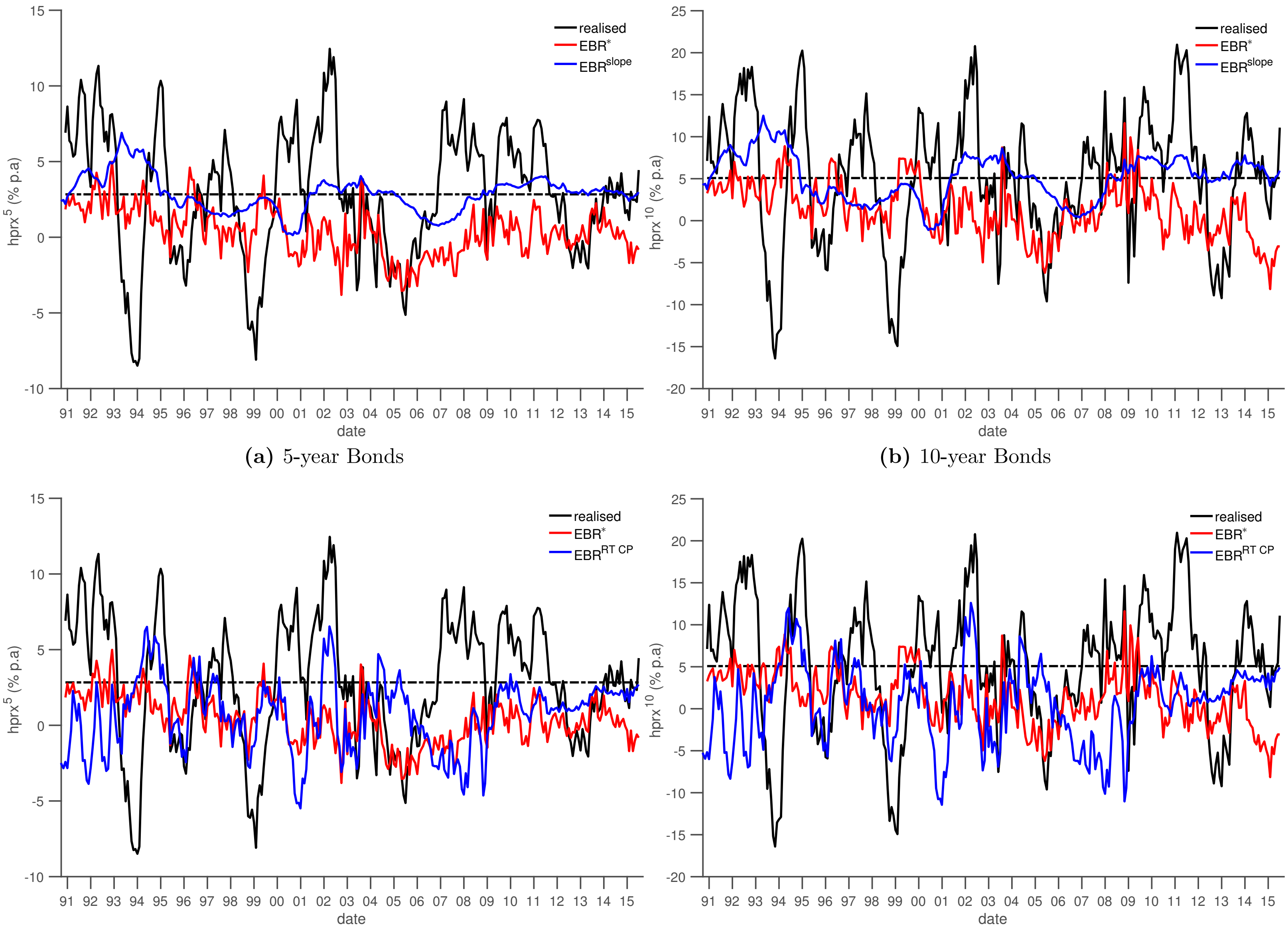
(a)



(b)

Figure 6. Sentiment Measures

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



(a) 5-year Bonds

(b) 10-year Bonds

(c) 5-year Bonds

(d) 10-year Bonds

Figure 7. Realized vs Expected Returns

This figure plots realised versus expected excess returns for a 5-year (left panels) and a 10-year (right panels) bond. Realised returns are lined up with expectations, i.e., the black line indicates excess returns that will be realised in 1-years time. The top panels compares realised returns with $EBR^{\star}$ versus the implied expected return from a real time slope forecast. The bottom panels compares realised returns with $EBR^{\star}$ versus the implied expected return from a real time CP forecast.

Rationality and Subjective Bond Risk Premia

Andrea Buraschi, Ilaria Piatti and Paul Whelan

Supplemental Appendix

I. Summary Statistics

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).

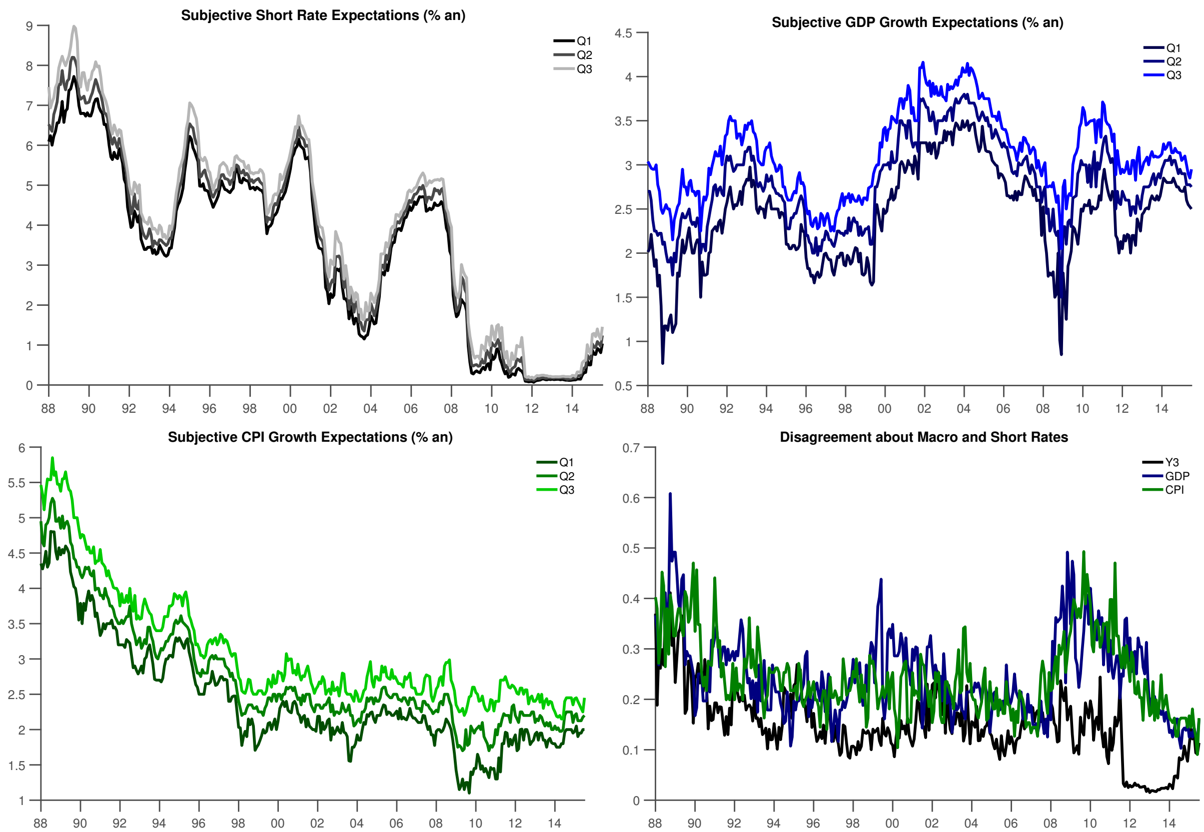


Figure 1. Subjective Expectations

The first three panels plot quartiles (Q1, Q2(median) and Q3) of the cross-sectional distribution of 1-year expectations of 3-month treasury yield forecasts (top left panel), GDP growth (top right panel) and CPI growth (bottom panel). The bottom right panel plots disagreement about 3-month Treasury yields, GDP and CPI growth, defined as the cross-sectional interquartile range of subjective expectations standardized by the full-sample consensus expectation.

II. Internally Consistent Beliefs

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 II. 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.

	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 II. 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 (GDP growth on the left and Inflation on the right) given that the forecaster is in a particular quartile of the cross-sectional distribution of 3 month yield forecasts.

III. State Dependence in the Difference in Accuracy

This appendix studies whether the differences in forecast accuracy for the 3-month yield, and the bias between top and bottom forecasters, discussed in Section III of the main text, are state dependent. Figure 2 shows the difference in the absolute error between forecasters in the top decile and two benchmarks, namely the consensus forecast and the expectation from the unit root model. Negative values occur when the top forecasters are more accurate than the benchmark. Independent of the benchmark used, we can statistically reject at the 5% confidence level the null hypothesis that the difference is constant. Moreover, we find that the top forecasters are more accurate than the consensus in 1990-1993, 2001-2003, and 2008-2011. All these periods are recessions and are characterized by important changes in the stance of the monetary policy.

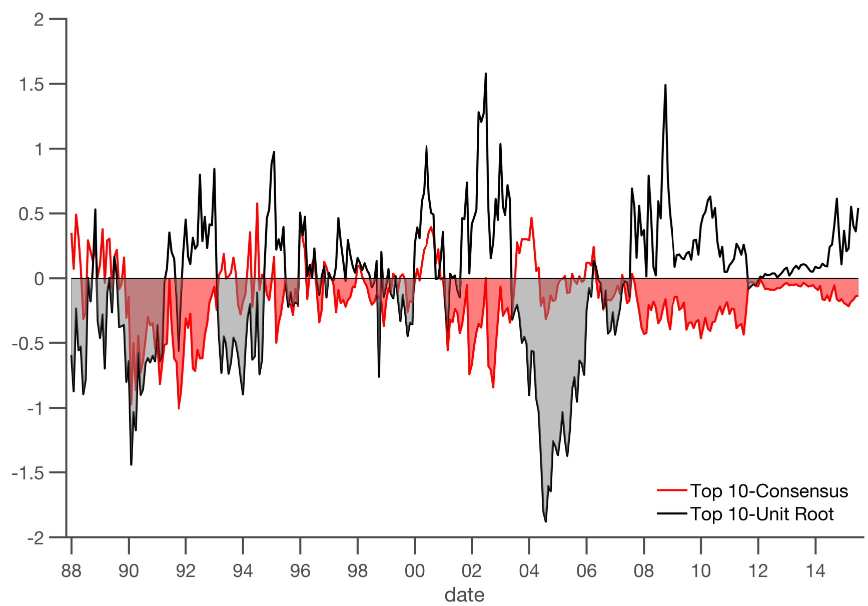


Figure 2. Absolute Short Rate Error Differences
Differences in absolute forecast errors between the the top decile of forecasters and the consensus, and the top decile and a unit root forecast.

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

To investigate these differences formally, we split the sample in two parts to capture persistent periods of increasing and decreasing interest rates, respectively. We compute the exponential moving average of the monthly change in the fed fund rate over the previous 12 months.[1] Considering the whole sample, there are 195 months in which this exponential moving average of changes is negative and 113 in which it is positive. We then recompute the average accuracy percentiles for each individual forecaster explicitly distinguishing these two time periods and we compare the distribution of accuracy percentiles for top and bottom forecasters using a Kolmogorov-Smirnov test. The null hypothesis of the Kolmogorov-Smirnov test is that the accuracy percentiles of top and bottom are drawn from the same distribution. Unconditionally (considering the full sample), the p-value of the test is 15%, which implies that we cannot reject the null hypothesis. However, in the subperiod in which the Fed has been more active in conducting a dovish policy on the short term rate, the p-value of the test is 1.61%. In these sub-periods we can strongly reject the hypothesis that accuracy percentiles of top and bottom forecasters are drawn from the same distribution. On the other hand, the p-value of the test in periods of increasing fed fund rate is 47.75%, suggesting that the distribution of accuracy is very similar in these periods. A Mann-Whitney U-test for the difference in medians between the accuracy percentile distributions yields similar results: unconditionally the p-value is 4.98%, in periods of increasing rates it is 58.99%, and in periods of decreasing rates it is 0.80%. We repeat these test using the ex-ante classification of PDs versus NPDs and we obtain the same conclusions.

IV. Why are Primary Dealers better at Predicting the Short Rate?

The fact that PDs are much better than other institutions at predicting the short rate precisely during inflection points, e.g. turns of business cycles when the Fed turns dovish by reducing the interest rate, is rather intriguing. It is potentially consistent with these institutions either being better in forecasting economic fundamentals, having specific information about the stance of the monetary policy, or having useful order flow information on short-term bonds. Since we have named forecasts also on economic fundamentals, we test the first hypothesis by comparing

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

the accuracy of top versus bottom short-rate forecasters about future real economic growth and inflation (see Figure 3). We find that top forecasters do not perform better than other agents in forecasting the inputs of the Taylor rule, i.e. inflation and GDP growth.[2] In fact, if anything, the accuracy of top forecasters' inflation expectations is lower, with an average accuracy ranking of 0.57 versus 0.42 for bottom forecasters (see bottom panels of Figure 2). Similarly, the GDP growth accuracy is worse for the top short rate forecasters, at 0.56 versus 0.51, respectively. Moreover, despite the time variation, the top short-rate forecasters (the great majority of which are primary dealers) are virtually never more accurate than the worse short-rate forecasters in predicting either inflation or real GDP growth. Indeed, the best macro forecasters on average are institutions like Action Economics and ClearView Economics; primary dealers such as Goldman Sachs, J.P. Morgan and Nomura are consistently in the worst half of growth and inflation forecast accuracy.

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

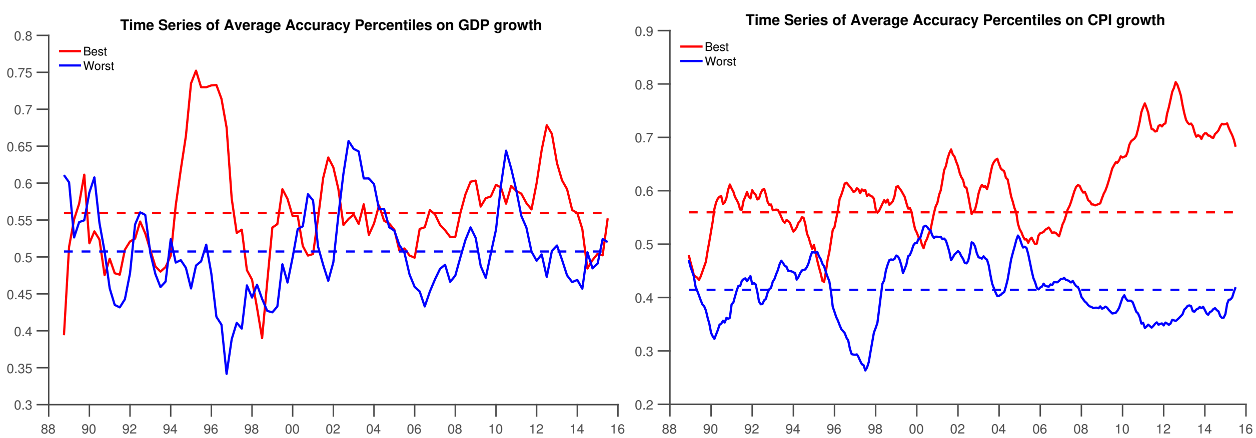


Figure 3. Macro Accuracy Percentiles
Time series of average accuracy percentiles on the Real GDP growth (left) and CPI growth (right), for the top and bottom decile short rate forecasters.

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

V. Additional Results

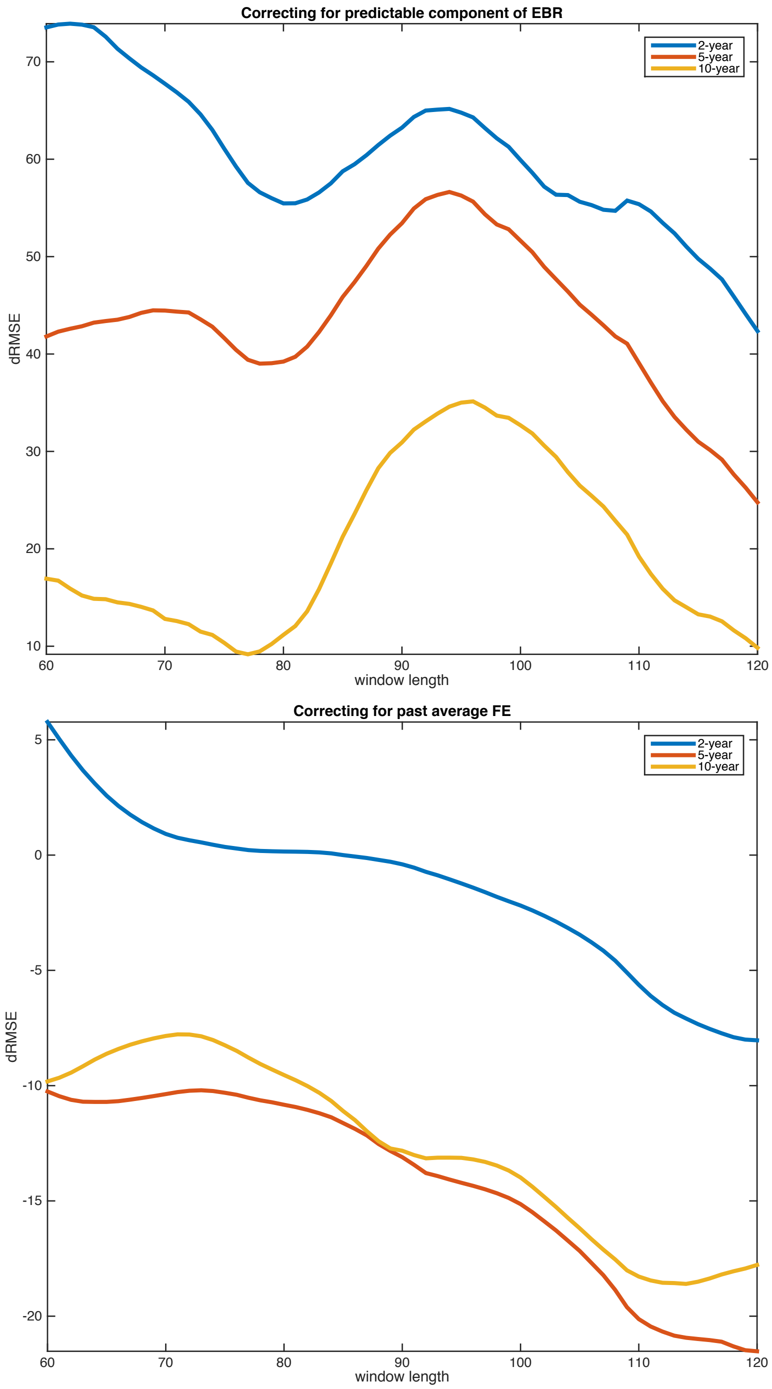


Figure 4. Change in RMSE correcting $EBR^{\star}$

Differences in root mean square errors between the corrected $\widehat{EBR}^{\star}$ and the original $EBR^{\star}$, as a function of the rolling window length used in the correction estimation, for bond maturities of 2, 5 and 10 years. On the left, the correction is based on a projection of the forecast errors realized over the rolling window on forward spreads. On the right, we correct $EBR^{\star}$ only by the average realized forecast errors over the rolling windows.

	$\mathcal{S}$	$DiB(g)$	$DiB(\pi)$	$Surp$	$LRR(g)$	$LRR(\pi)$	$\widehat{\sigma}$	PD	$\overline{R}^2$
(1)	−0.17								2.68
	(−2.21)								
(2)		0.10	0.17						5.67
		(1.32)	(2.38)						
(3)	−0.20	0.10	0.20						9.79
	(−2.86)	(1.35)	(2.80)						
(4)				−0.12					1.10
				(−1.82)					
(5)	−0.20			−0.15					4.59
	(−2.60)			(−2.30)					
(6)					−0.03	0.45			18.12
					(−0.41)	(3.01)			
(7)	−0.32				−0.12	0.58			27.13
	(−5.39)				(−1.81)	(3.69)			
(8)							0.18		3.07
							(2.77)		
(9)	−0.25						0.25		8.74
	(−3.03)						(3.58)		
(10)								−0.30	9.28
								(−4.34)	
(11)	−0.21							−0.33	13.70
	(−2.88)							(−4.69)	

Table III. Determinants of Ex-Ante Subjective Bond Returns

This table reports estimates from regressions of the subjective expected excess returns on 5-year bonds on a set of explanatory variables:

$$EBR^{\star}_{5,t} = a + b^{\top}\mathbf{X}_t + \epsilon_{5,t}.$$

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