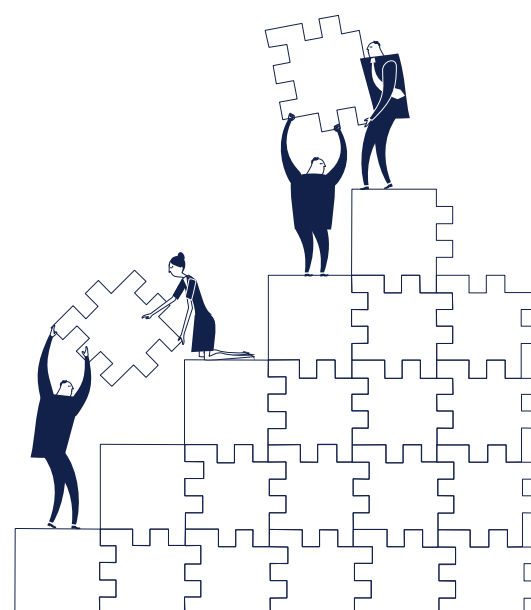Saïd Business School
Research Papers

May 2018

Rationality and Subjective Bond Risk Premia

Andrea Buraschi
Imperial College Business School

Ilaria Piatti
Saïd Business School, University of Oxford

Paul Whelan
Copenhagen Business School

Rationality and Subjective Bond Risk Premia

Andrea Buraschi   Ilaria Piatti   Paul Whelan

May 2018

Abstract

This paper documents large micro-heterogeneity and forecasting skill in the cross-section of survey based bond risk premia. We reject informationally constrained rational expectations but show a learning model distorted by sentiment is consistent with the data. Aggregating, we propose a belief measure for the marginal agent that is consistent with Friedman's market selection hypothesis. This measure is available in real-time and compares favourably to popular statistical models. Moreover, forecast errors from this measure, while predictable, are not easily corrected in real-time. Finally, we re-assess structural models and find support for both sentiment and time-varying quantity of risk channels.

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

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 argues that this predictability is driven by behavioural biases affecting the dynamics of subjective beliefs, informational frictions, or both.

This paper sheds lights on the debate by exploiting a dataset that provides us with forecaster specific expectations about U.S. Treasury bond yields, GDP and inflation to study the properties of subjective bond risk premia. The survey-based approach has a number of economic and econometric advantages compared to the standard approach of inferring risk premia from projections of future return realizations on lagged state variables. Most importantly, survey-based expectations of future excess returns are by construction model-free and forward-looking, and thus provide us with a direct and real-time representation of investor expectations. This allows us to address a series of questions related to agents rationality and the efficiency of markets in aggregating beliefs.

We begin by constructing measures of subjective bond risk premia (EBR) from professional market participants' expectations of future yields. Using forecaster specific data, we study the cross-sectional properties of agents' beliefs on bond risk premia and study the extent to which beliefs satisfy full-information rational expectations (FIRE) restrictions. The availability of individual level data is especially important in this context since it is well known that consensus regressions are inconsistent in presence of micro-heterogeneity. For example, if individual agents are rational but have access to different signals, rational expectation tests based on consensus beliefs can lead to over-rejection of the null of rationality (Figlewski and Wachtel (1983)). Inconsistency can also go in the opposite direction if agents form heterogeneous irrational forecasts since errors would tend to be averaged out in a consensus regression, thus leading to excessive acceptance of the rational expectation hypothesis (Zellner (1962), Keane and Runkle (1990), and Bonham and Cohen (2001)). We also use the properties of the link between individual forecast errors and forecast revisions to learn about alternative expectation models, such as noisy rational expectations, sticky information, and information neglect.[1]

_____

[1]Many authors have employed consensus forecasts to rational expectations. A non-exhaustive list includes Ang, Bekaert, and Wei (2007) and Aiolfi, Capistran, and Timmermann (2011) for macro quantities, Frankel and Froot (1987) for exchange rates, Froot (1989) and Cieslak (2017) for interest rates, Bacchetta, Mertens, and Van Wincoop (2009) and Greenwood and Schleifer (2014) for equity, and Koijen, Schmeling, and Vrugt (2015) for multiple asset classes.

When agents hold heterogeneous beliefs, an important issue is how competitive markets aggregate individual beliefs in those of the marginal investor.[2] In the empirical literature, a common approach is to proxy subjective beliefs with consensus expectations, in which all agents have the same weight. In some cases, this choice is imposed by data limitations, but in the context of asset pricing this is tantamount to assuming that all agents affect prices with equal weights, including those with irrational beliefs. Friedman (1953) and Alchian (1950) argue instead that market selection in competitive markets is a powerful force affecting the characteristics of the representative agent. Those agents that are consistently more accurate than others should accumulate more economic weight in the pricing kernel. Their argument is a compelling reminder to many behavioural studies: instances of irrationality at the individual level do not necessarily imply that the representative agent is irrational and market prices inefficient. We proxy for the beliefs of the representative agent by ranking agents according to their past accuracy and studying the spanning properties of this measure versus measures generated by alternative aggregators. Five sets of empirical results emerge.

First, we document a large unconditional heterogeneity in the cross-section of $EBRs$. The mean forecaster EBR is 1.06% for 10-year bonds, while the mean of the first quartile is negative at -1.66%, which implies that these agents believe long-term bonds are hedges against economic shocks rather than risky bets on future economic states. We also find clear evidence of persistence in agent-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.

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 stock markets. However, most forecasters tend to under-predict excess bond returns, which is an indication of a rejection of the null hypothesis of FIRE at the individual level. Despite this bias, a

---

[2]Important empirical contributions to the literature on belief heterogeneity include Malmendier and Nagel (2011), Das, Kuhnen, and Nagel (2017), Patton and Timmermann (2010) and Andrade, Crump, Eusepi, and Moench (2014). Structural models that include a role of belief heterogeneity 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), Jouini, Marin, and Napp (2010), Ehling, Gallmeyer, Heyerdahl-Larsen, and Illeditsch (2018), and Borovicka (2011).

large fraction of professional forecasters are reasonably good relative 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 primary dealers are more likely to be among the top forecasters of the short-term interest rate but they are not significantly better than other institutions in forecasting long-term bonds returns.[3] 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 models of expectation formation that might explain the rejection of FIRE at the individual level and the persistent heterogeneity in beliefs with different levels of accuracy.[4] An important set of models assume that agents hold rational expectations but are subject to information frictions (Coibion and Gorodnichenko (2015)). A testable implication of these models is that individuals forecast errors are not correlated with their contemporaneous forecast revisions. Empirically, we find that the slope coefficients of regressions of forecast errors on forecast revisions are significantly negative for almost all agents. This is inconsistent with informationally-constrained rationality, including both noisy rational expectations and sticky information models. On the other hand, we show that the negative sign is consistent with models in which agents neglect information in forming their beliefs. This bias induces a deviation between the beliefs of agents and those of the econometrician, which Dumas, Kurshev, and Uppal (2009) define as sentiment. We propose an empirical proxy for sentiment and find evidence of a systematic component that is correlated across agents and affects both expectations of bond yields and economic growth.

Fourth, we study the conjecture of Friedman (1953) and Alchian (1950) that competition in financial markets washes away the impact of irrational beliefs in the pricing measure. We construct, at any point in time t, the average beliefs of the agents who were most accurate up to time t, which we denote $EBR_t^\star$. Consistent with Friedman (1953), we find that $EBR_t^\star$ has better spanning properties of contemporaneous bond prices than other aggregators, which highlights

---

[3]Primary dealers are trading counter-parties of the New York Fed in its implementation of monetary policy and are also expected to make markets for the New York Fed on behalf of its official accountholders as needed.

[4]Contributions studying deviations from full information rational expectations include Mankiw and Reis (2002), Woodford (2002), Sims (2003), Coibion and Gorodnichenko (2015), Dovern, Fritsche, Loungani, and Tamirisa (2015), and Mackowiak and Wiederholt (2009).

the important role of competitive markets in aggregating beliefs. If market competition were able to filter out irrational beliefs perfectly, prediction errors associated to $EBR_t^\star$ should be orthogonal to any public information available at time t. We find instead statistical evidence against orthogonality, which suggests that market competition is not sufficiently effective in penalizing biases. Interestingly, however, when we investigate the economic significance of this predictability, we find that most attempts to use public information to correct the biases in real time are not economically significant. The most effective correction is obtained from the average forecast errors over rolling windows and reduces the root mean squared error of the forecast by about 13% for the 10-year bond. This is consistent with a dogmatic persistent deviation from the econometrician's beliefs even at the level of the marginal agent.

Finally, we propose an application of $EBR^\star$ that relates to an extensive macro-finance literature that studies the dynamics of bond risk premia.[5] We revisit the literature by replacing future average returns with $EBR_t^\star$, which is forward looking and available in real-time. 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 generate large R^2 with loadings on risk factor proxies that are consistent with predictions from theory. Moreover, 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.

The paper proceeds as follows. Section I presents the data. Section II investigates the accuracy of forecasters. Section III considers alternative channels of expectation formation. Section IV addresses the question of beliefs aggregation. Section V conducts tests of rationality on our proxy for the beliefs of the representative agent. Section VI uses $EBR^\star$ to evaluate the performance of structural models of expected bond risk premia. Section VII concludes.

---

[5]Duffee (2013) surveys the literature on the determinants of bond risk premia and concludes that most macro models are not very successful empirically. Structural contributions on the dynamics of risk premia include Campbell and Cochrane (1999) and Wachter (2006) for habit preferences; Bansal and Yaron (2004), Bansal and Shaliastovich (2013) for preference for resolution of uncertainty and long-run risks; Dumas, Kurshev, and Uppal (2009), Buraschi and Whelan (2017)) for belief heterogeneity. General predictions that links quantities of risk to compensation for risk date back to Merton (1980), French, Schwert, and Stambaugh (1987), and Duffee (2002).

I. The Cross Section of Subjective Bond Risk Premia

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.[6] It is unique with respect to alternative commonly studied surveys along at least four dimensions. First, the dataset provides the identity of the forecasters. This allows us to track each individual in the cross section and time series.[7] 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.[8] Fourth, Bluechip has always been administered by the same agency,

---

[6] 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 GDP expectations from BCFF.

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

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

while other surveys, such as SPF, have been administered by different agencies over the years. Moreover, SPF changed some of the questions in the survey, and some of these changes crucially affected the forecasting horizon.[9] 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 estimate a Nelson-Siegel model on individual agent subjective par-yield forecasts. The Nelson-Siegel model assumes that the instantaneous forward rate is given by a 3-factor parametric function.[10] We calculate the term structures using all available maturities (including 30-year Treasury yield forecasts) and obtain a monthly panel data of expected constant time-to-maturity zero coupon (continuously compounded) discount rates. 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.

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

---

points considerably.

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

[10] 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.

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 compute the implied cross-section of EBR as $erx_{i,t}^n = E_t^i\left[p_{t+1}^{n-1}\right] - p_t^n - y_t^1$ since from the surveys we directly observe $E_t^i\left[y_{t+1}^{n-1}\right]$

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

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. Belief Heterogeneity

Figure 1 takes 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.

[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]

---

[11]Table I in the Supplemental Appendix provides summary statistics for the median, the first quartile, and the third

7

The conditional properties of the cross-sectional distribution of EBR display rich dynamics in the time series and there exists significant time-varying heterogeneity around the consensus forecast. 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 observations demonstrate the existence of micro-heterogeneity, which implies that FIRE cannot be tested using aggregate data or panel regressions with common coefficients.

C.   Belief Persistence

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

---

quartile of the (1-year) EBR distribution for the 2, 5 and 10-year bonds.

[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 1 in the Supplemental Appendix.

[14]Supplemental Appendix B 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.

[Insert Table I here.]

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 show that forecasters have highly persistent beliefs about bond risk premia. 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 significance level of 5%.[16]

The question of belief persistence is particularly interesting 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.

II. 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.

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:

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

---

[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.

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. 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 related to the measurement error and aggregation in 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.

[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.

---

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

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 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.[18] Therefore, we construct a real-time version of the CP factor: we initialise a CP with 10-years of data from January 1978 to January 1988 via a two stage projection of *forward spreads*.[19] 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: $CP(t)$.[20] 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.[21]

Table II presents root mean square prediction errors for 10-year bond excess returns, based on forecasts from the unconditional top 10 forecasters (BEST), the bottom 10 forecasters (WORST), and the arithmetic average of the cross-section of forecasters for each prediction date t (CONS), i.e. the consensus.[22] 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

---

[18] 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 right hand variables have a root very close to unity, such as the first stage Cochrane-Piazzesi regression.

[19] 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.

[20] 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.

[21] 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.

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

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 excluding the ZLB period. However, we also find that real-time forecasts implied by $Slope$ always out-performing CP. In the full sample, the RMSEs from the $Slope$ forecasts are almost identical to BEST and better than the consensus. 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 $Slope$.

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 3 presents the histogram for the cross section of $\mathcal{A}_i^n$, for bond maturity n equal to 10 years. We find that an important fraction of survey forecasters outperform 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, a large fraction of the distribution lies between 1.00 and 1.20 for the full sample, indicating that most agents are slightly worse. However, consistent with the discussion above this is driven by large errors during the ZLB period. Excluding this period around 40% of the agents in the cross section outperform a $Slope$ forecast.

[Insert Figure 3 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,

12

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 4th 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. 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).

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 3-month yield, denoted $\mathcal{R}_{i,t}^{3m}$. First, the left panel of Table III shows that, similar to long-term bonds, accuracy (good and bad) in predicting short rates tends to be persistent. 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. The elements on the diagonal show the existence of a link between the two accuracies. However, the rank correlation of the accuracy percentiles on the 3-month yield and 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. Which institutions 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 RidgeWorth, 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

---

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

better predictive performance only for the short rate.[24] This might be explained by primary dealers having specific information about (or better models to interpret) monetary policy.[25] 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.

III. Beyond the Line of FIRE

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. In this section we consider alternative channels of expectation formation.

Two important stream of the literature investigate economies with rational but sticky-information (Mankiw and Reis (2002)) and with rational but noisy-information (Woodford (2002), Sims (2003), and Mackowiak and Wiederholt (2009)). In the first case, 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. When agents are subject to noisy information, they rationally update their beliefs but, since they can never fully observe the true state, they use an optimal signal-extraction filter. If one were to aggregate these expectations, Coibion and Gorodnichenko (2015) show that average forecast errors are predictable by forecast revisions. They use data on consensus inflation expectations and find that, at the aggregate level, forecast revisions (positively) predict forecast errors. As they highlight, however, "the predictability of the *average* ex-post forecast errors across agents using ex-ante forecast revisions is an emergent property of the aggregation across individuals, not a property of the individual forecasts." Indeed, in both cases, to the extent that agents expectations are rational, even in the presence of information constraints ex-post forecast errors should continue to be unpredictable at the individual level. To avoid the potential confounding effect of aggregation, we study informationally constrained rationality by directly using individual level data. For each single contributor i to the BCFF panel, we

---

[24]See also Section C of the Supplemental Appendix for an analysis of the dynamics of 3-month yield forecast accuracy.

[25]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 D of the Supplemental Appendix.

run the following regression and test $H_0 : b^{i,n} = 0$:

$$FE^{i,n}_{t,t+1} = a^{i,n} + b^{i,n} FR^{i,n}_t + \epsilon^{i,n}_{t+1}, \qquad (4)$$

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 $FR^n_{i,t}$ is its forecast revision, computed as the difference between its time-t forecast and the same forecast made the previous month.[26]

Panel (a) of Figure 4 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 4 here.]

At the individual level, models of information frictions imply orthogonality between forecast updates and forecast errors. This is inconsistent with our finding of a mostly negative and significant slope coefficient in regression (4). These results clearly suggest that not only we can reject a full-information version of the rational expectation hypothesis, but also explanations based on the role of information rigidities, such as sticky or noisy information.[27]

The finding 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)).[28] 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 of their dynamics. To directly explore this third explanation, we investigate the implications of a model with noisy information when expectations are biased by the statistical neglect of useful information. Suppose that x_t is the (hidden) state of the economy and y_t is the variable that

---

[26]Results are very similar when using forecast revisions of long-term yields at quarterly frequency.

[27]We also find that the negative slope coefficient is mainly due to the dynamics of bond risk premia, as opposed to the dynamics of short term interest rates.

[28]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.

the agent is trying to forecast, such as the bond yield. The true yield depends on two factors, i.e. x_t and z_t, but the agent ignores z_t and forms expectations only based on x_t. If $z_t \neq 0$, agents expectation are biased. Assume the following simple $AR(1)$ dynamics for x_t and z_t:

$$
\begin{aligned}
x_{t+1} &= a_0 + a_1 x_t + \varepsilon_{t+1}^x, \qquad (5) \\
z_{t+1} &= b_0 + b_1 z_t + \varepsilon_{t+1}^z, \\
y_{t+1} &= x_{t+1} + z_{t+1} + w_{t+1},
\end{aligned}
$$

with x_{t+1} independent of z_{t+1} and $[\varepsilon_{t+1}^x, \varepsilon_{t+1}^z, w_{t+1}]$ being orthogonal zero-mean innovations. In our context, z_t is the difference between the econometrician and the agent's forecasts of the state of the economy, and can therefore be interpreted as a measure of sentiment.

Proposition 1. *Let K^i be the individual-specific Kalman gain associated with the state-space (5), and let $FE_{t,t+1}^i$ and FR_t^i denote the forecast error and forecast revision of agent i about y_{t+1}, conditional on his information at time t. Then,*

(i) Average forecast errors and average forecast revisions are proportional to average z_t:

$$
E\left(FE_{t,t+1}^i\right) = \frac{1 - a_1}{1 - a_1(1 - K^i)} E\left(z_t\right) \quad and \quad E(FR_t^i) = \frac{a_1(1 - a_1)K^i}{1 - a_1(1 - K^i)} E(z_t). \quad (6)
$$

(ii) The projection coefficient of $FE_{t,t+1}^i$ on FR_t^i is $\frac{1}{K^i}\left[\frac{b_1}{a_1} - K^i\right]$;

(iii) At time t, if $a_1 > 0$ then FR_t^i is positively related to z_{t-1}, while $FE_{t,t+1}^i$ is positively related to z_t and negatively to z_{t-1}.

Proof: *See Section E of the Supplemental Appendix.*

The proposition implies that the neglect of important information can introduce rich implications for both the joint and marginal dynamics of forecast errors and forecast revisions.[29] Forecast errors are predictable and the sign depends on $\frac{b_1}{a_1}$ relative to the Kalman gain K^i. Since $K^i < 1$, it is positive when the persistence b_1 of z_t is larger than the persistence a_1 of the state

---

[29] The model-implied expressions for the Bayesian update of the latent state (See Section E of the Supplemental Appendix) resemble the case of diagnostic expectations in Bordalo, Gennaioli, Ma, and Shleifer (2018), in which agents overstate the impact of news by the multiplicative factor. However, in our setting there is also a bias in the perceived prediction error, that includes the current level of sentiment and the difference between the agent's estimated state and the econometrician's, which is a function of past levels of sentiment and past shocks.

variable x_t. The opposite occurs when the state variable x_t is significantly more persistent than z_t. We assume that individual agents are heterogeneous in their perceived persistence of the latent state, a_1^i. This assumption implies a cross-section of regression coefficients in Equation (4) which are negatively linked to a_1^i.

Figure 4 shows that the slope coefficient of $FE_{t,t+1}^i$ on FR_t^i are negative and $FE_{t,t+1}^i$ are positively autocorrelated. These two results are potentially consistent with the existence of neglected information at the individual level, with a dynamics characterized by a small b_1/a_1 ratio on average but a distribution of coefficients that is driven by the agent-specific state persistence a_1^i.

Is the evidence of information neglect also systematic across agents? To investigate directly this question, we construct a direct measure of forecaster's sentiment. A growing asset pricing literature that models heterogeneous beliefs in general equilibrium allows for a deviation of agents beliefs about fundamental growth from the expectation of an unbiased econometrician. Dumas, Kurshev, and Uppal (2009) refer to this process as sentiment. To study the link between agent specific and systematic sentiment, we use agent-specific expectations on one-year GDP growth, $E_t^i(gdp_{t+1})$, and compare them to the expectation implied by a benchmark model, $E(gdp_{t+1}|\mathcal{M}_t)$. 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: $\mathcal{S}_t^{gdp} = E_t^i(gdp_{t+1}) - E(gdp_{t+1}|\mathcal{M}_t)$, 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. Note that a positive sentiment $\mathcal{S}_t^{gdp}$ corresponds to a negative z_t in the model above, whereby agents' forecast of yields is higher than the rational forecast $x_t + z_t$, or a positive z_t if the observable variable y_t is the bond return instead of the level of yields, as in our regression (4), since if agents over-predict yields they under-predict bond returns. Our sentiment proxy $\mathcal{S}_t^{gdp}$ is significantly positive on average, with a sample mean of around 0.37% and has an autocorrelation of about 0.58, which is lower than the persistence of individual expectations for the large majority of agents, consistent with a small b_1/a_1^i ratio on average.

Does aggregate sentiment explain the bias in the forecast errors for long-term bond excess

returns at the level of the individual forecasters? Panel (d) of Figure 4 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:

$$FE_{t,t+1}^{i,10} = a_s^{i,10} + b_s^{i,10} \mathcal{S}_t^{gdp} + \epsilon_{i,t+1}^{10}. \tag{7}$$

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. The result is consistent with Proposition 1.[30] Moreover, as implied by (6) with $E(z_t) > 0$, the time series average forecast error is positive for the large majority (around 92%) of the agents, and we find a strong rank correlation of about 42% between the agents' mean forecast errors and their regression coefficient in a projection of forecast errors onto forecast revisions. This result is consistent with Proposition 1, which implies that an agent with a large perceived persistence a_1^i will have both a smaller $E\left(FE_{t,t+1}^i\right)$ and a more negative projection coefficient of $FE_{t,t+1}^i$ on FR_t^i.

To summarize, the results suggest the existence of a *systematic* behavioral component in agents beliefs, that is at least partially responsible for the statistical rejection of both FIRE and noisy rational expectations hypotheses. Agents are optimistic on average in our sample (i.e. sentiment tends to be positive) and this drives average positive forecast revisions and forecast errors on bond excess returns. Moreover, the signs of the regression coefficients of forecast error on forecast revisions, forecast error on sentiment, and forecast error on lagged forecast error are mutually consistent with the existence of ignored information (sentiment) with positive (albeit small) persistence.

IV. Belief aggregation and subjective bond risk premia

The extent of cross-sectional heterogeneity in beliefs raise a series of questions about how beliefs should be aggregated to construct an empirical proxy of subjective bond risk premia for the

---

[30]Results are robust to the use of 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.

representative agent. The solution is far from trivial. It is common in the empirical literature to use consensus expectations as a proxy of subjective beliefs. In some cases, the choice is forced by data limitations. In the context of asset pricing, this is tantamount to assuming that the marginal agent holds consensus beliefs. However, this is often inconsistent with the implications of heterogeneous general equilibrium models, in which the beliefs of the marginal agent deviate from consensus.

The general equilibrium literature that studies economies where trade is generated because of heterogeneous beliefs argues that in absence of short-selling constraints irrational agents eventually lose economic weight to the benefit of less biased agents.[31] This argument, consistent with the original "market selection hypothesis" by Alchian (1950) and Friedman (1953) implies that bond prices should span the beliefs of the most accurate agents (i.e. closest to the actual physical probability).[32] 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. The "market selection hypothesis" has been used as a powerful theoretical arguments in support of the efficient markets paradigm as this does not require that all agents are rational in order for prices to reflect the true data generating process. While the argument is made asymptotically, Kogan, Ross, Wang, and Westerfield (2006) show that even when agents with inferior beliefs do not survive in the long run, their impact on prices can persist. Indeed, although they are linked, survival and price impact are two distinct concepts. Kogan, Ross, Wang, and Westerfield (2016) discuss the conditions on agent's preferences in which agents who fail to survive in the long run can still affect the prices of low-aggregate consumption states.[33]

To account for the potential link between market selection and price impact, 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 ac-

---

[31] See, for example, 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)

[32] 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. "*

[33] Additional important contributions on this topic include Yan (2008) who argue that it may take a long time to effectively eliminate the impact of irrational investors. Moreover, even a slightly smaller time discount rates may lead an irrational investor to dominate the market in the long run, even if his belief substantially and persistently deviates from the truth.

cording 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 portfolios of agents based on this past accuracy. Note that the realized forecast errors up to time t will be based on expectations formed up to time $t - 12$ months, since we consider forecasts with a fixed one year horizon. We use an initial window of 5 years. Then, we compute the average EBR for different bond maturities for the agents in the top half of the accuracy percentiles distribution at each time t. This is our preferred measure of aggregate subjective bond risk premia, which we denote $EBR^{\star}$. 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. We also tried alternative simple aggregation schemes that place weight on the most accurate agents, such as linear weights or a simple average of the top quartile, which generate quantitively similar results. One could attempt a more sophisticated non-linear aggregation scheme to recover the unobservable weights of the agents in the pricing kernel and test ex-post the difference in the price impact of the most versus the least accurate agents. We decided instead to rely on an aggregation scheme that is less arbitrary and sample dependent as possible and study its properties.

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.[34]

[Insert Table V here.]

We find that $EBR^{\star}$ is positively correlated to future realized excess returns and has lower RMSE than the consensus. The correlation between $EBR^{\star}$ and future realized returns is 15%

---

[34]$EBR^{\star}$ can be computed for the sample that spans the period December 1993 - July 2015. In this period, Huntington National Bank and RidgeWorth Capital Management contribute to the panel for a total of 85 and 76 months, respectively, while for example Fannie Mae has 234 monthly contributions.

and 13% for the 5 years and 10 years bond, respectively, and it increases to 40% and 33% 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 particularly important to account for belief aggregation and select the appropriate agents rather than simply taking the simple average of all agents' expectations.

Figure 5 shows the time series of our $EBR^\star$ measure against the corresponding realized excess returns, for a 5-year (left panels) and a 10-year (right panels) bond.

[Insert Figure 5 here.]

It is clear from these figures that $EBR^\star$ has a downward bias. The representative agent tends to underestimate future excess bond returns as 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. This observation is consistent with the sentiment biased Kalman filter results in section III.

A.   Alternative Aggregators

The existence of frictions can affect the properties of the marginal agent's SDF rather differently than suggested by Friedman's market selection hypothesis. Indeed, Harrison and Kreps (1978), Scheinkman and Xiong (2003) and Hong, Sraer, and Yu (2017) discuss economies in which the equilibrium SDF is biased toward the beliefs of the optimist agent. Agents with negative expectations about future returns are restricted from selling short, thus their beliefs do not affect the properties of the equilibrium SDF. Individual level data allows us to investigate this alternative aggregator as follows. We identify the beliefs of the unrestricted agents as those with positive expected returns at any time t, and denote the average of their beliefs EBR^+. Then, we can directly compare the properties of EBR^+ to those of $EBR^\star$. Depending on the dynamics of agents beliefs, the number of agents used to obtain EBR^+ changes over time as well as their identities.

Figure 6 compares the five-years rolling average R^2 of a contemporaneous (spanning) regression of $EBR^\star$ and EBR^+ on the term structure of bond prices. If short-selling constraints are important, then EBR^+ should be better explained by contemporaneous bond prices than $EBR^\star$. In the period 2002-2009, until the start of quantitative easing and forward guidance, we find that the rolling R^2 of $EBR^\star$ is significantly greater that the R^2 produced by EBR^+. Interestingly, prior to 2002, $EBR^\star$ and EBR^+ are broadly equivalent. On the one hand, this may support the interpretation of the existence of binding short-selling constraints during that period. However, one may also notice that in the five year period ending in 2002, Treasury bond market optimists should have accumulated significant capital gains as the effective Fed fund rate dropped from 5.50% (October 1997) to 1.74% (February 2001). A similar observation applies in the period after 2009, when the two R^2 are not very different. Even during this period, the Fed aggressively reduced interest rates using both conventional and unconventional policies.

[Insert Figure 6 here.]

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 penalises 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 (and is correlated with) popular forecasting models purely based on price information.

V.   Marginal Agent's Beliefs

In the first part of the paper, we show that individual agents expectations deviates from FIRE. Nonetheless, Friedman's conjecture is that markets may still be able to aggregate information and beliefs efficiently by allocating dynamically resources to the least biased agents. In what follows, we use our measure of $EBR^\star_t$ as a proxy for the subjective bond risk premium of the representative agent to investigate the extent to which markets are able to filter out individual biases that might be present at the individual level.

First, we form forecast errors on bond excess returns from the difference between future realized and expected excess returns, $FE^n_{t,t+1} = rx^n_{t,t+1} - EBR^\star_{n,t}$. Then, we investigate the

extent to which $FE_{t,t+1}^n$ are predictable by information available at time t. If markets are able to aggregate all information at time t and cure the existence of individual biases in beliefs reported earlier, then any public information at time t should be orthogonal to $FE_{t,t+1}^n$.

Following Cochrane and Piazzesi (2005), we summarize the information in the cross section of date t yields using five principal components and we estimate the following set of complimentary regressions:

$$EBR_{n,t}^\star = \alpha_1 + \beta_1^\top PCs + \varepsilon_t \tag{8}$$

$$rx_{t,t+1}^n = \alpha_2 + \beta_2^\top PCs + \eta_{t+1} \tag{9}$$

$$FE_{t,t+1}^n = \alpha_3 + \beta_3^\top PCs + \nu_{t+1} \tag{10}$$

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 10 years. Adjusted R-squared of the regressions that includes 5 principal 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 first line of Table VI shows that a low number of principal components explain a substantial proportion of the variation in subjective bond risk premia, suggesting the existence of a strong link between $EBR_{n,t}^\star$ beliefs and current bond prices. The degree of spanning of realized returns is smaller and the loadings on the PCs, i.e. β_2 in equation (9), are significantly different from β_1 in Equation (9), which implies significant degree of predictability in forecast errors in the last line of Table VI.

This suggests that even after allowing markets to aggregate beliefs according to Friedman's selection argument, bond return expectations do not exploit all available information and prediction errors are not orthogonal even to rather simple type of public information. In order to address this argument one needs to investigate the economic significance of the predictable component of these forecast errors.

A. 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. In this real-time experiment, we initialise a rolling regression with a window of 5 years of data and recursively estimate a projection of realised errors on the cross-section of forward spreads.[35] The loadings available in the forecast error regression at date t can only be learned from errors realised one year earlier. These loadings are then applied to date t forward spreads (F) in order to build a 'corrected' $EBR^\star$ from the following system

$$\widehat{FE}^n_{t-1,t} = \hat{\alpha}_{t-1} + \hat{\beta}^\top_{t-1} F_{t-1} \tag{11}$$

$$\xi_t = \hat{\alpha}_{t-1} + \hat{\beta}^\top_{t-1} F_t \tag{12}$$

$$\widehat{EBR}^\star_t = EBR^\star_t + \xi_t \tag{13}$$

The subscript t in the parameters $\hat{\alpha}_{t-1}$ and $\hat{\beta}^\top_{t-1}$ 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 around 17% for the 10-year bond.[36] This shows that the expectations embedded in $EBR^\star$ 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 $EBR^\star$ dominates its corrected counterpart, mainly in terms of variability, meaning that the apparent bias in agents beliefs is not easy to

---

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

[36]Figure 4 in the Supplemental Appendix shows the robustness of this finding to alternative window lengths and different bond maturities. 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.

correct using information available in real-time. This provides a possible explanation for why subjective expectation can be persistently different from full information rational expectations in the long run.

Finally, we note that if one were to correct $EBR^\star$ for a simple constant bias, obtained from the mean of the forecast errors over a five year rolling window, the RMSE of the forecast would decrease by about 13% for the 10-year bond. This could be consistent with either agents having asymmetric loss functions, which penalize more severely optimistic bond return forecast errors, or with dogmatic persistent beliefs that imply a sentiment bias, as illustrated in Section III. Consistent with the last interpretation, we find that the bias correction is highly correlated with a rolling average of the monthly sentiment proxy over the same rolling windows. Strikingly, for 10-year bond expected excess returns, this correlation is $\sim 89\%$ over the full sample.

VI. Understanding Subjective Bond Risk Premia

In this section we evaluate alternative models for risk premia based on their ability to explain the dynamics of $EBR^\star_{n,t}$, as opposed to projections on future excess returns:

$$EBR^\star_{n,t} = a + b^\top \mathbf{X}_t + \varepsilon_{n,t},\tag{14}$$

and consider alternative risk factors, $\mathbf{X}_t$, that have been proposed by the 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.

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 revisit 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 $\widehat{\sigma}(n)$.

Finally, 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. Therefore, we ask whether the risk factor proxies discussed above are linked to subjective bond risk premia, while allowing for a contemporaneous

---

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

biases in 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.[38] When controlling for sentiment, we are explicitly taking into account that subjective risk premia can be driven by both behavioural and rational elements.

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 1993 and July 2015.[39] Table VII reports the regression results for 10-year maturity bonds.[40]

[Insert Table VII here.]

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 4.6%. 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 non significant at the 5% level and the R-squared of the regression is around 10%. Interestingly, controlling for $\mathcal{S}_t$ the estimate of $DiB(g)$ is unaltered but that of $DiB(\pi)$ rises slightly and the $\overline{R}^2$ rises to 16%.

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 -1.17 and the R-squared is 0.85%. Interestingly, however, after we take into account the effect of sentiment, the t-statistic on $Surp$ rises to -1.57 with an $\overline{R}^2$ of 6.4%.

---

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

[39]This is the time period for which $EBR^\star_{n,t}$ is available, since we used an initial window of 5 years for the computation of past average accuracy percentiles, and there is a lag of 12 months between expectations and realizations.

[40]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.

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 $EBR_t^\star$ as dependent variable the statistical significance of inflation uncertainty is quite remarkable, with a t-statistic equal to 4 for the 10 year bond and an $\overline{R}^2$ equal to 19%. 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 29%. 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 $EBR_{10,t}^\star$ rather than realisations. The R-squared is quite high, around 18%, and the t-statistic of 7.18 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 29% 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.

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 $EBR_{10,t}^\star$ and bond volatility. 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 dis-

tinguishing 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.

VII.   Conclusion

This paper studies the expectations of bond returns reveiled by 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. Agents disagree on whether bonds provide insurance or are risky bets on the state of economy. Moreover, we document significant persistence in agents beliefs about bond expected returns. This translates into significant persistence in the rankings of different agents for their accuracy in predicting long-term bond returns.

Second, when we compare beliefs about short-term and long-term bonds, we find that the most accurate agents at forecasting the long end of the term structure are not the best in predicting short-term rates. 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 formally test and clearly reject both full-information rational expectations and weaker versions of rational expectations with information frictions. We show that agents display a systematically optimistic behaviour a that a distorted Kalman filter matches well their expectation formation. These tests take advantage of individual specific beliefs information that allows us to avoid the biases and confounding effects of consensus data.

Fourth, we explicitly address Friedman (1953) conjecture that market competition drives resources to agents with (more) rational beliefs. We construct an aggregate measure of subjective bond risk premia based on past accuracy which we denote as $EBR^{\star}$. Using this measure we show that while forecast errors are predictable in a statistical sense they are not easy to correct economically in real time.

Finally, we give an application of $EBR^{\star}$ in the context of the evaluation of structural models. Instead of using average future realized returns as the dependent variable, we use $EBR^{\star}$ and

compare a series of alternative models proposed in the literature. We find support for both behavioural and rational determinants of bond risk premia. In particular, we demonstrate the importance of controlling for agents sentiment in models in which the time variation of risk premia depend on the interaction of the prices and quantities of risks.

References

Aiolfi, 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, P., E. Mertens, and E. 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, H. S., and R. 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.

Bordalo, P., N. Gennaioli, Y. Ma, and A. Shleifer, 2018, Overreaction in macroeconomic expectations, *working paper.*

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.

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, M., and P. 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.

———, 2008, Decomposing the yield curve, *Graduate School of Business, University of Chicago, 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.

Duffee, G. 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, P., M. Gallmeyer, C. Heyerdahl-Larsen, and P. Illeditsch, 2018, Disagreement about inflation and the yield curve, *Journal of Financial Economics*.

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

Frankel, J. A., and K. 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, K. R., G. W. Schwert, and R. 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, R. S., B. Sack, and J. H. Wright, 2006, The u.s. treasury yield curve: 1961 to the present, *Federal Reserve Board Working Paper Series*.

Harrison, J.M., and D.M. Kreps, 1978, Speculative investor behavior in a stock market with heterogeneous expectations, *The Quarterly Journal of Economics* 92, 323–336.

Hong, H., D. Sraer, and J. 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.

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.

Kogan, Leonid, Stephen A Ross, Jiang Wang, and Mark M Westerfield, 2006, The price impact and survival of irrational traders, *The Journal of Finance* 61, 195–229.

———— , 2016, Market selection, *Working paper, Sloan School of Management.*

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, R. 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.

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.

Yan, Hongjun, 2008, Natural selection in financial markets: Does it work?, *Management Science* 54, 1935–1950.

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

VIII. 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). 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 (37 observations in the top panel, 18 in the bottom panel) 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)	Nomura Securities Inc.
(3)	DePrince & Assoc
(4)	Cycledata Corp
(5)	Loomis Sayles & Co
(6)	Bank of America Securities
(7)	Keller Economic Advisers
(8)	Thredgold Economic Assoc.
(9)	Goldman Sachs
(10)	The Northern Trust Company

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 half of forecasters based on past accuracy percentiles on the 10-year bond excess returns. Sample period is December 1993 - July 2015.

Maturity	constant	PC1	PC2	PC3	PC4	PC5	$\overline{R}^2$	$\Delta\overline{R}^2$
$EBR^{\star}_{10,t}$	0.92	1.50	1.40	1.05	-0.45	-0.56	39.30	3.99
	(2.51)	(4.36)	(4.97)	(3.97)	(1.32)	(-2.20)		
rx^{10}_{t+1}	4.76	1.38	3.61	-0.15	1.51	-2.24	32.66	13.64
	(5.39)	(1.74)	(4.29)	(-0.18)	(2.94)	(-3.40)		
FE^{10}_{t+1}	3.84	-0.12	2.21	-1.19	1.96	-1.68	21.15	11.47
	(3.77)	(-0.13)	(2.42)	(-1.38)	(3.53)	(-2.39)		

Table VI. Expected Returns, Realized Returns, and Forecast Error Predictability
Estimates from regressions of subjective expected returns ($EBR^{\star}_{10,t}$), realized returns (rx^{10}_{t+1}) and forecast errors (FE^{10}_{t+1}) on the principal components of yields (PCs), for a 10-year bond. 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 1993 to July 2015.

	$\mathcal{S}$	$DiB(g)$	$DiB(\pi)$	$Surp$	$LRR(g)$	$LRR(\pi)$	$\widehat{\sigma}$	$\overline{R}^2$
	−0.22							4.60
	(−2.66)							
(I)		0.24	0.11					9.85
		(2.51)	(1.30)					
	−0.26	0.24	0.15					16.35
	(−3.35)	(2.49)	(1.76)					
(II)				−0.11				0.85
				(−1.17)				
	−0.25			−0.15				6.41
	(−3.01)			(−1.57)				
(III)					0.07	0.41		19.20
					(0.95)	(4.0)		
	−0.33				−0.02	0.52		29.02
	(−4.87)				(−0.31)	(5.08)		
(IV)							0.43	18.39
							(7.18)	
	−0.33						0.50	28.89
	(−4.43)						(7.30)	

Table VII. Determinants of Ex-Ante Subjective Bond Returns

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 1993 to July 2015.
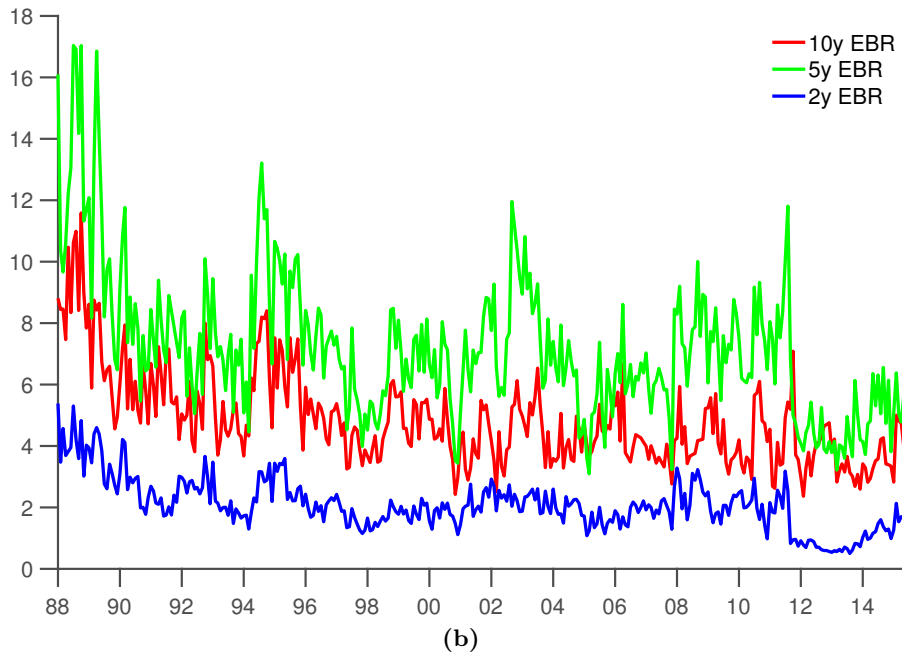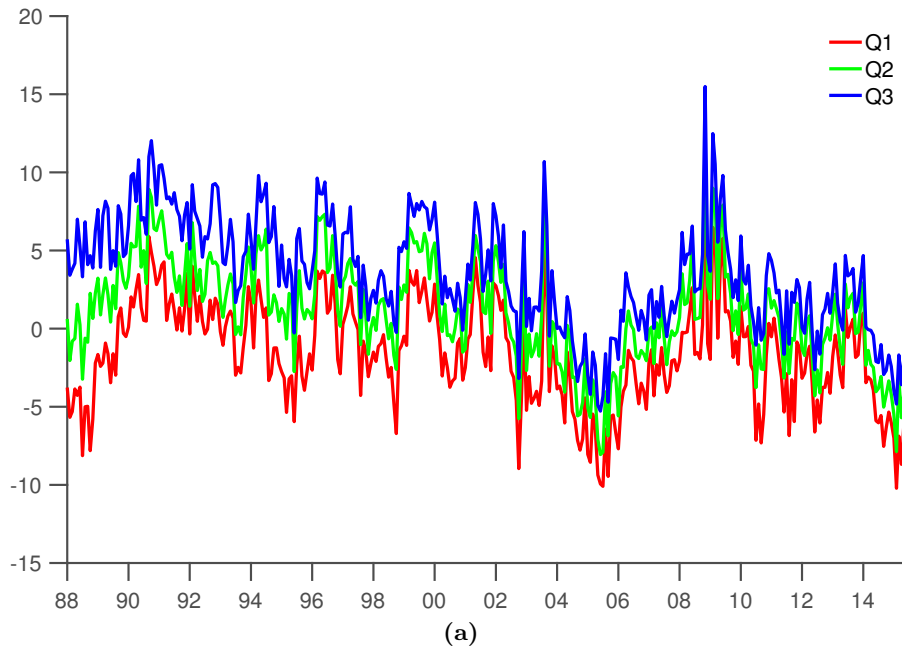
IX. Figures

(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.
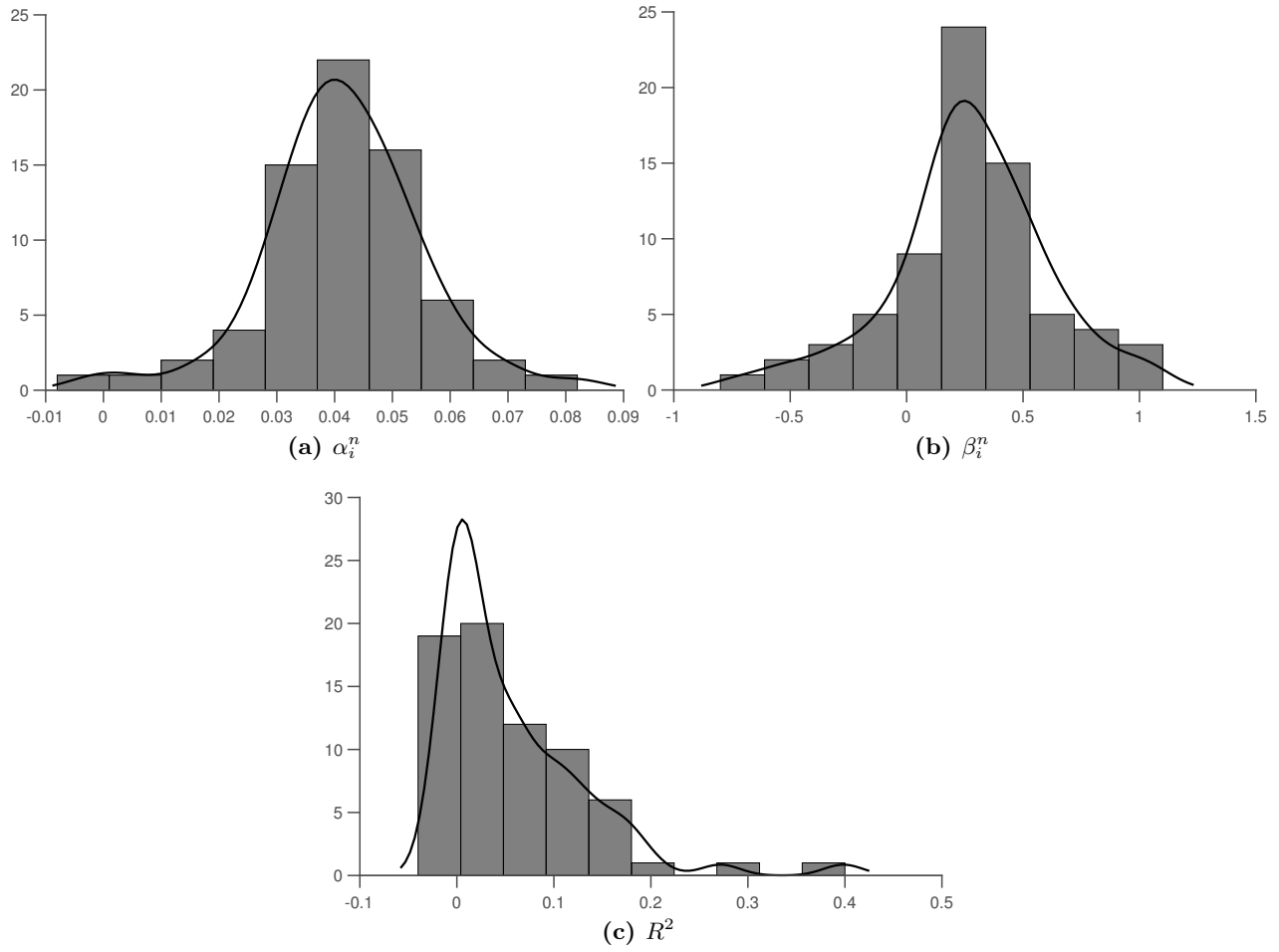
(a) α_i^n

(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:

$$rx_{t+1}^{10} = \alpha_i^{10} + \beta_i^{10} erx_{i,t}^{10} + \epsilon_{i,t+1}^{10},$$
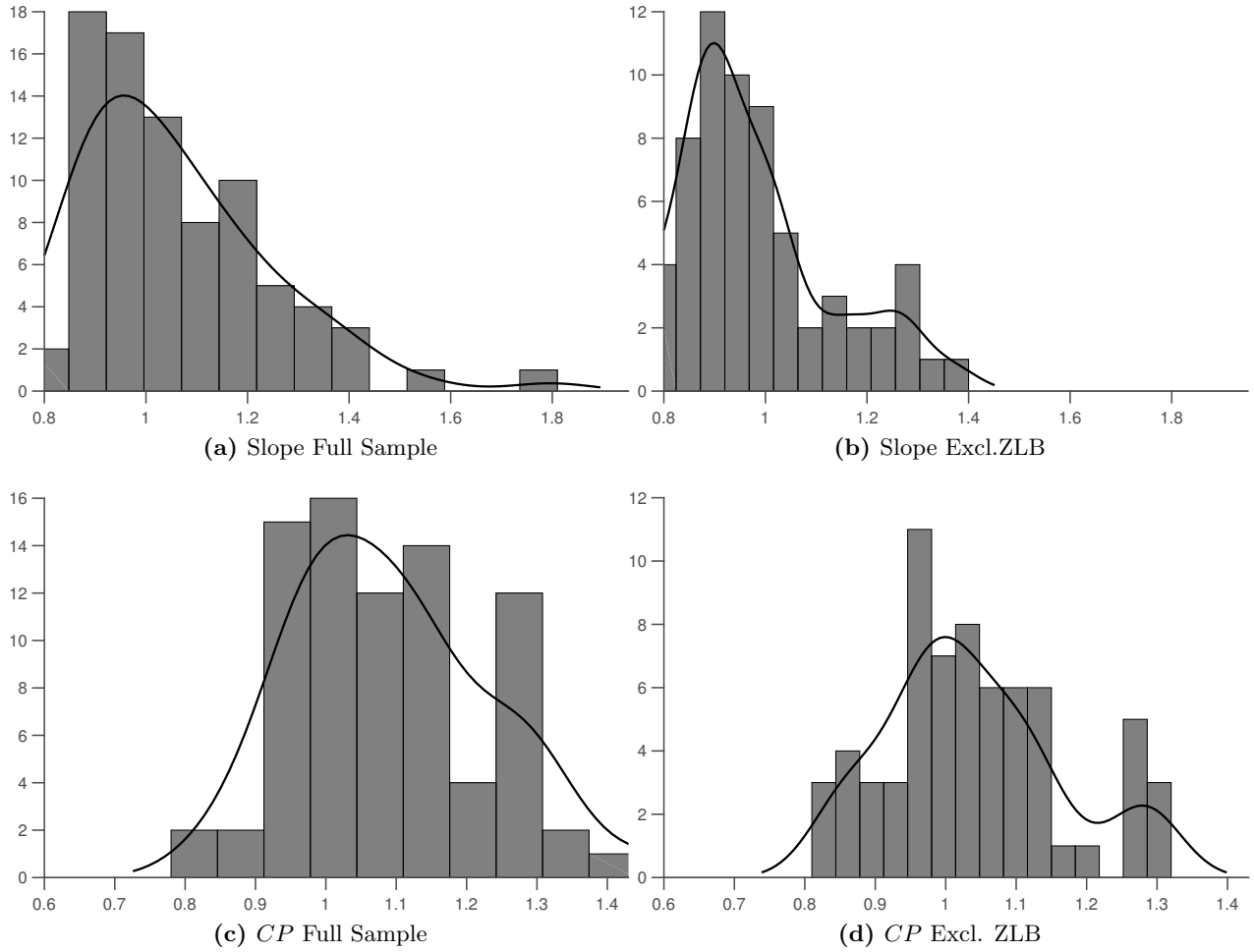
44

Figure 3. 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.
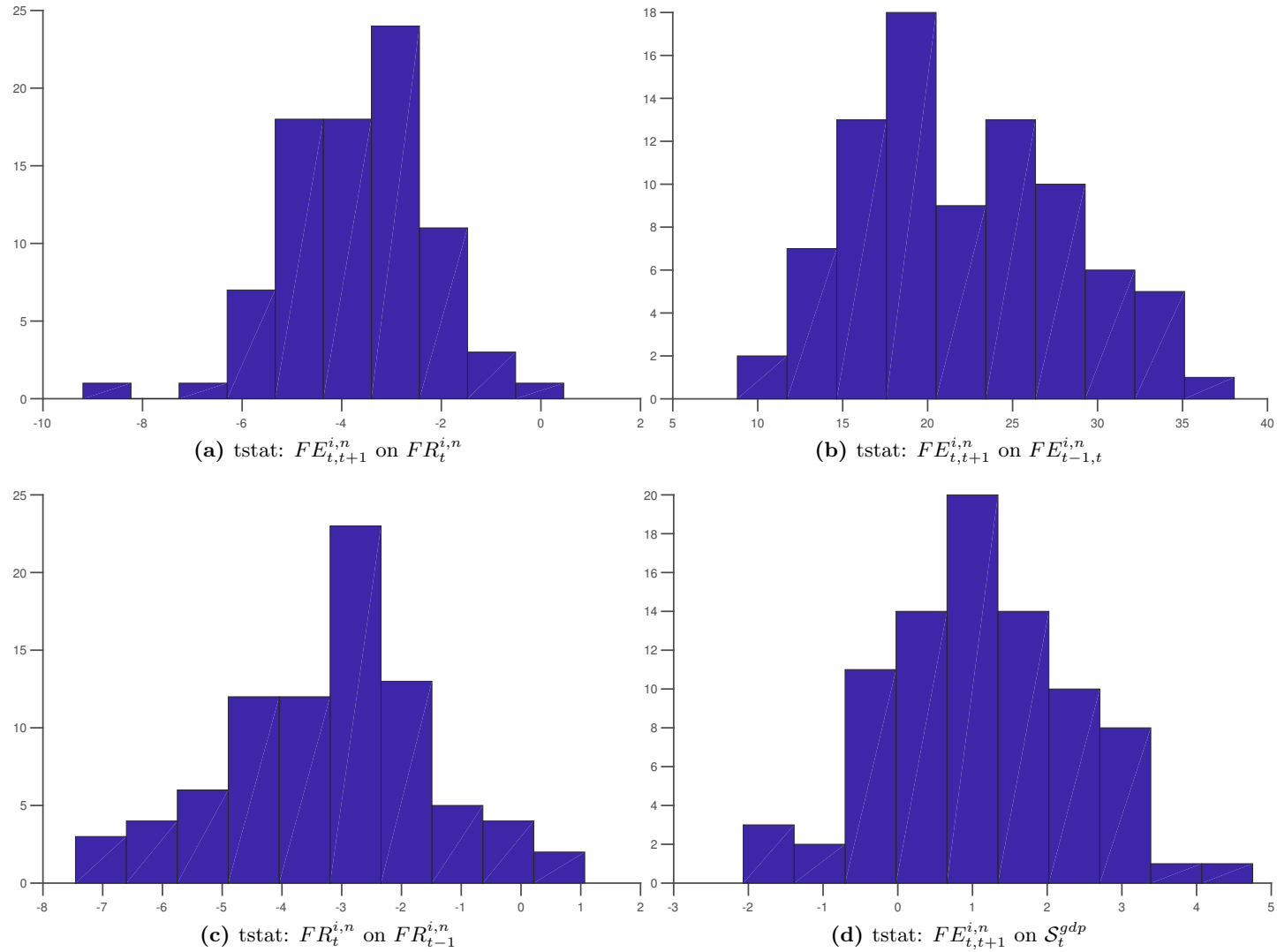
45

Figure 4. Full Information and Noisy RE Tests

The top left panel shows the t-statistics for the slope coefficient of a 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. The top right (bottom left) panel shows the t-statistics for the slope coefficient of a regression of individual forecast errors (forecast revisions) on their lagged values. The bottom right panel shows histograms of t-statistics of the slope coefficients of a regression of forecast errors on sentiment. Sentiment is estimated on growth sentiment, $\mathcal{S}_t^{gdp}$, where t is measured in quarters rather than months since GDP growth data are available only at quarterly frequency.
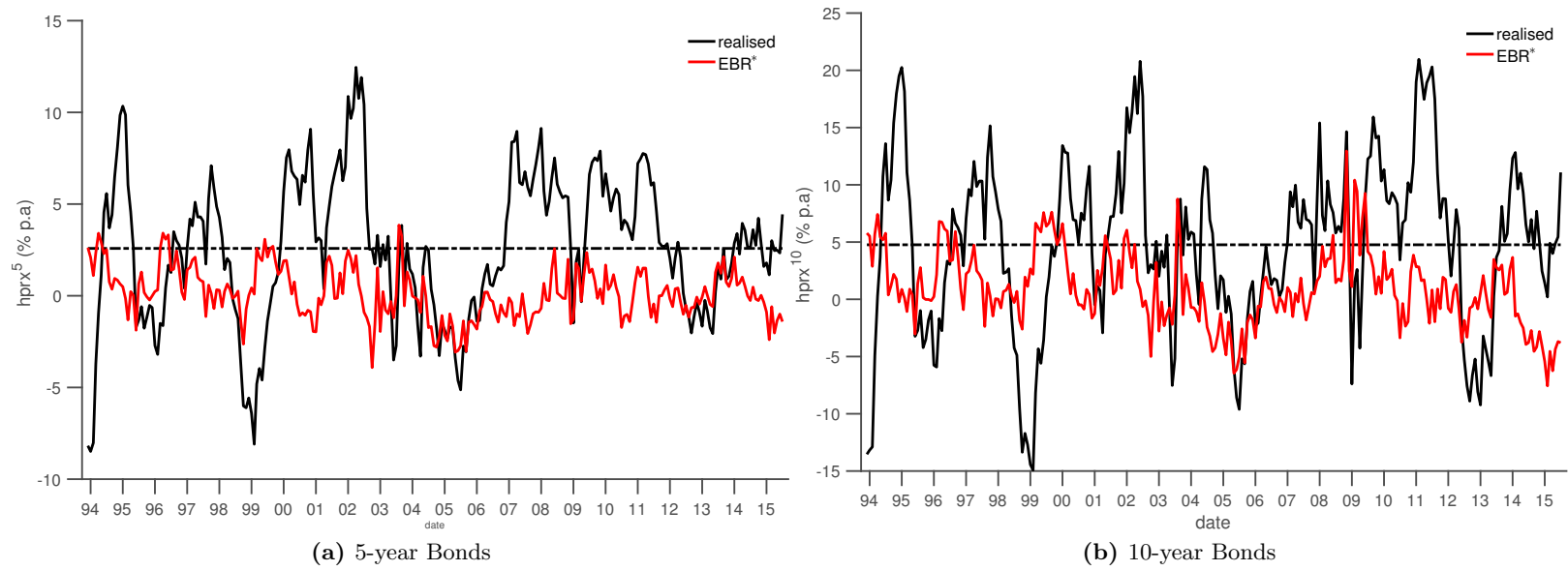
Figure 5. Realized vs Expected Returns

Realized versus expected excess returns for a 5-year (left panel) and a 10-year (right panel) bond. Realized returns are lined up with expectations, i.e., the black line indicates excess returns that will be realised in 1 year.
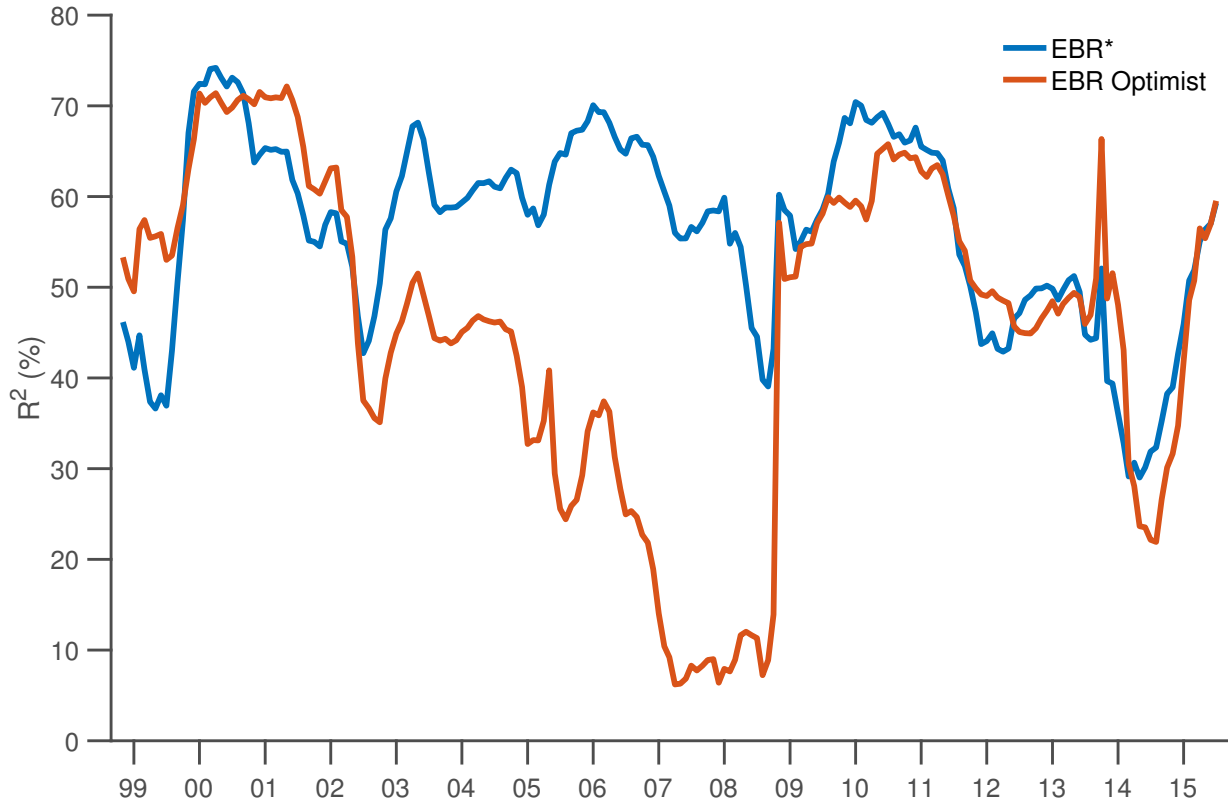
Figure 6. Spanning Tests of Alternative Aggregate Measures.
Differences in R^2 of contemporaneous 60-months rolling regressions of EBR^+ and $EBR^\star$ on the cross-section of bond prices, summarized by the principal components of the term structure of interest rates. EBR^+ is obtained using the beliefs of the unrestricted agents with positive expected bond excess returns; $EBR^\star$ is based on the beliefs of the top half forecasters based on past accuracy.

Rationality and Subjective Bond Risk Premia

Andrea Buraschi, Ilaria Piatti and Paul Whelan

Supplemental Appendix

A. Summary Statistics

Table I provides summary statistics for the median, the first quartile and the third quartile of the cross-sectional distribution of 1-year expected excess bond returns (EBR) for maturities of 2, 5 and 10 years.

Figure 1 shows the dynamics of the quartiles of the cross-sectional distribution of macroeconomic and short rate forecasts and the level of disagreement, defined as the cross-sectional interquartile range of subjective expectations standardized by the full-sample consensus expectation. 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).

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).
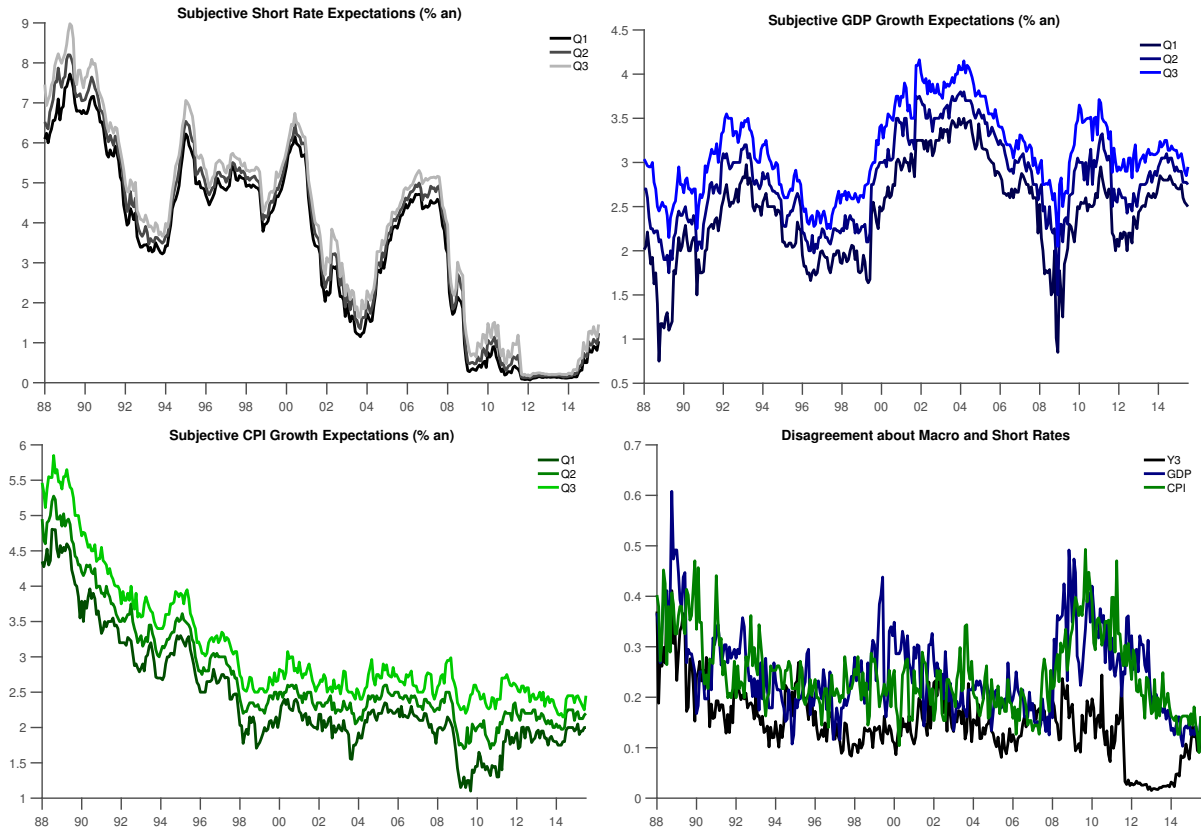
Figure 1. Subjective Expectations

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.

B.   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.

C.   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.
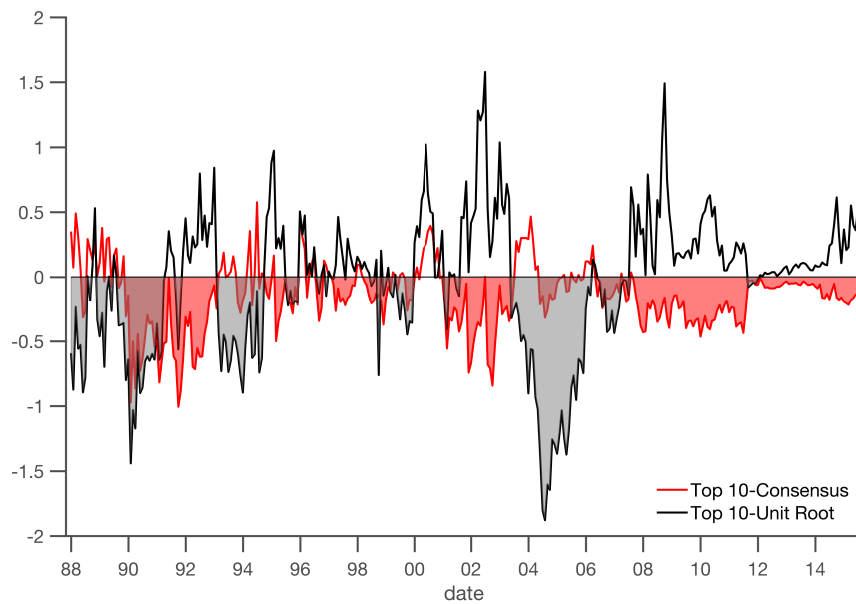

Figure 2. Absolute Short Rate Error Differences
Differences in absolute forecast errors between the the top decile of forecasters and the consensus, and the top decile and a unit root forecast.

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.

D. 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.
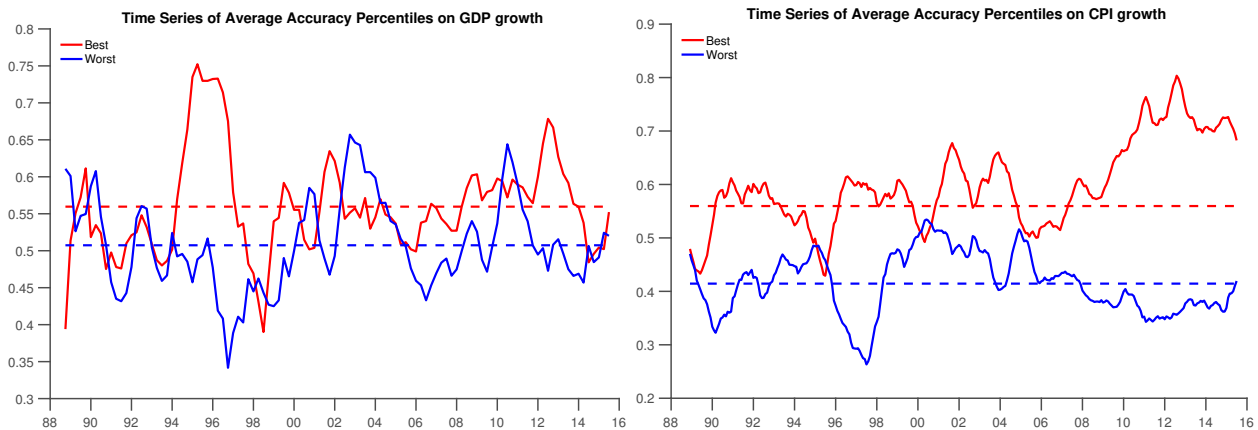

Figure 3. Macro Accuracy Percentiles
Time series of average accuracy percentiles on the Real GDP growth (left) and CPI growth (right), for the top and bottom decile short rate forecasters.

---

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

E. Kalman Filter with Sentiment

Assume the state variable x_t is a measure of economic growth or of the state of the economy and is latent. The observable variable y_t in the filter represents the level of yields or another variable that the agents are trying to forecast.

Assume also that the true expected yield is $x_t + z_t$ while the agents ignore z_t and think the expected yield is just x_t. z_t represents a measure of *sentiment*, which is common to all agents i.[3]

The transition and measurement equations under the measure of the econometrician are as follows:

$$
\begin{aligned}
x_{t+1} &= a_0 + a_1 x_t + \varepsilon^x_{t+1}, & (1)\\
z_{t+1} &= b_0 + b_1 z^i_t + \varepsilon^z_{t+1}, & (2)\\
y_t &= x_t + z_t + w_t, &
\end{aligned}
$$

$$(3)$$

while agents ignore z, i.e. their expectations are biased, and believe

$$y_t = x_t + w_t.$$

Assume for simplicity the moment that the two state variables x and z are independent.[4]

A. *Steps of the filter*

- Initialize at the unconditional mean and variance of the state:

$$
\begin{aligned}
x_{0|0} &= \frac{a_0}{1 - a_1}\\
P^x_{0|0} &= \sigma^2_{\varepsilon^x}\\
z_{0|0} &= \frac{b_0}{1 - b_1}\\
P^z_{0|0} &= \sigma^2_{\varepsilon^z}
\end{aligned}
$$

---

[3]A positive value of z_t implies that the agents expect a lower yield than the econometrician and be interpreted as pessimism.

[4]Results can be easily extended to the case of correlated state innovations.

The initial value of the x state for the agent is also $x^i_{0|0} = \frac{a_0}{1-a_1}$.

- Time-update equations:

    For the agents:

$$x^i_{t|t-1} = a_0 + a_1 x^i_{t-1|t-1}. \tag{4}$$

$$P^x_{t|t-1} = a_1^2 P^x_{t-1|t-1} + \sigma^2_{\varepsilon^x}. \tag{5}$$

    For the econometrician:

$$x_{t|t-1} = a_0 + a_1 x_{t-1|t-1}. \tag{6}$$

$$z_{t|t-1} = a_0 + a_1 z_{t-1|t-1} \tag{7}$$

$$P^x_{t|t-1} = a_1^2 P^x_{t-1|t-1} + \sigma^2_{\varepsilon^x} \tag{8}$$

$$P^z_{t|t-1} = a_1^2 P^z_{t-1|t-1} + \sigma^2_{\varepsilon^z}. \tag{9}$$

- Prediction error and variance of the measurement equation:

    The yield forecast of the agent is $\hat{y}^i_{t|t-1} - x^i_{t|t-1}$ while the econometrician's forecast is $\tilde{y}_{t|t-1} - x_{t|t-1} - z_{t|t-1}$. Therefore, their prediction errors are given by:

$$\eta^i_t = y_t - x^i_{t|t-1}, \tag{10}$$

$$\tilde{\eta}_t = y_t - x_{t|t-1} - z_{t|t-1}, \tag{11}$$

    with variances

$$S_t = P^x_{t|t-1} + \sigma^2_w, \tag{12}$$

$$\tilde{S}_t = P^x_{t|t-1} + P^z_{t|t-1} + \sigma^2_w, \tag{13}$$

    respectively.

- Update filtering:

The Kalman gain for the agent is

$$K_t^i = P_{t|t-1}^x(S_t)^{-1} = \frac{P_{t|t-1}^x}{P_{t|t-1}^x + \sigma_w^2}. \tag{14}$$

The Kalman gain for the econometrician has two components, corresponding to the signal to noise ratios of the two states:

$$\tilde{K}_{1,t} = \frac{P_{t|t-1}^x}{P_{t|t-1}^x + P_{t|t-1}^z + \sigma_w^2}, \tag{15}$$

$$\tilde{K}_{2,t} = \frac{P_{t|t-1}^z}{P_{t|t-1}^x + P_{t|t-1}^z + \sigma_w^2}. \tag{16}$$

Note that

$$K_t^i = \tilde{K}_{1,t}\left(1 + \frac{P_{t|t-1}^z}{P_{t|t-1}^x + \sigma_w^2}\right) = \tilde{K}_{1,t}(1 + \theta_{t|t-1}), \tag{17}$$

which means that the agent has an inflated Kalman gain since he interprets the divergence between yields and x due to sentiment as an additional prediction error in x that he tries to correct.[5]

The updated state for the agent is thus:

$$x_{t|t}^i = x_{t|t-1}^i + K_t^i \eta_t^i, \tag{18}$$

while for the econometrician the update is:

$$x_{t|t} = x_{t|t-1} + \tilde{K}_{1,t}\tilde{\eta}_t, \tag{19}$$

$$s_{t|t} = s_{t|t-1} + \tilde{K}_{2,t}\tilde{\eta}_t, \tag{20}$$

Combining (18) and (4) we can also write:

$$x_{t+1|t}^i = a_0 + a_1 x_{t|t-1}^i + a_1^i K_t^i \eta_t^i. \tag{21}$$

---

[5]Note that the effect could be reversed if the states x and z are positively correlated and this correlation is strong enough to offset the effect of θ. In this case, the agent would observe a smaller prediction error when x is large and therefore under-react to the shock.

Combining (18) with (17) we get:

$$
\begin{aligned}
x^i_{t|t} &= x^i_{t|t-1} + \tilde{K}_{1,t}(1 + \theta_{t|t-1})\eta^i_t \\
x^i_{t|t} &= x^i_{t|t-1} + \tilde{K}_{1,t}(1 + \theta_{t|t-1})\left[\tilde{\eta}_t - (x^i_{t|t-1} - x_{t|t-1}) + s_{t|t-1}\right].
\end{aligned}
\tag{22}
$$

This expression resembles the case of diagnostic expectations in Bordalo, Gennaioli, Ma and Shleifer (2018), in which agents overstate the impact of news by the multiplicative factor θ. However, in this setting we also have a bias in the perceived prediction error, that includes the current level of sentiment and the difference between the agent's estimated state and the econometrician's, which is a function of past levels of sentiment and past shocks.

The difference between the agent-specific and the rational forecasts of the state can be written as:

$$
x^i_{t+1|t} - x_{t+1|t} = a_1 K^i_t \tilde{\eta}_t + a_1 K^i_t s_{t|t-1} + a_1\left(1 - K^i_t\right)(x^i_{t|t-1} - x_{t|t-1}).
\tag{23}
$$

Since this difference appears also in the forecast errors and forecast revisions, for convenience let us denote it by $d_{t+1|t} \equiv x^i_{t+1|t} - x_{t+1|t}$. Its unconditional mean is

$$
E(d) = \frac{a_1 K^i}{1 - a_1(1 - K^i)} E(z),
$$

where $E(z) = \frac{b_0}{1-b_1}$ and K^i is the steady-state Kalman gain of the agent.

B. Forecast Revisions

Forecast revisions can be written as follows:

$$
\begin{aligned}
FR^i_t &= x^i_{t+1|t} - x^i_{t+1|t-1} \\
&= a_0 + a_1 x^i_{t|t-1} + a_1 K^i_t \eta^i_t - a_0 - a_1 x^i_{t|t-1} \\
&= a_1 K^i_t \eta^i_t \\
&= a_1 \tilde{K}_{1,t}(1 + \theta_{t|t-1})\left(\tilde{\eta}_t - d_{t|t-1} + z_{t|t-1}\right).
\end{aligned}
\tag{24}
\tag{25}
$$

11

Note that since the agents are biased the prediction errors η_t^i do not have mean zero.

In fact, the unconditional mean of the forecast errors is:

$$E(FR) = \frac{a_1(1-a_1)K^i}{1-a_1(1-K^i)}E(z),$$

which has the same sign of the mean sentiment if $0 < a_1 < 1$. In the data, agents seem to overestimate yields, which implies a negative z and therefore the mean yield forecast error should be negative and the mean excess return forecast error should be positive.

Equation (25) implies that forecast revisions FR_t^i are positively linked to sentiment $z_{t|t-1}$, with regression coefficient $a_1 \tilde{K}_{1,t}(1+\theta_{t|t-1})$, assuming that the persistence of the state variable a_1 is positive and noting that the Kalman gain is a number between zero and one.

Equation (25) also implies that forecast revisions are persistent, given the persistence of d and z.

C. Forecast errors

The forecast error for agent i is given by:

$$
\begin{aligned}
FE_{t,t+1}^i &= y_{t+1} - x_{t+1|t}^i \\
&= \tilde{\eta}_{t+1} - d_{t+1|t} + z_{t+1|t}
\end{aligned}
\tag{26}
$$

$$\tag{27}$$

The coefficient on current sentiment $z_{t+1|t}$ is positive but FE depends negatively on past values of sentiment through $d_{t+1|t}$.

The unconditional mean of the forecast errors is given by:

$$E(FE) = \frac{1-a_1}{1-a_1(1-K^i)}E(z) = \frac{E(FR)}{a_1 K^i}. \tag{28}$$

As for the forecast revisions, assuming $b_0 < 0$ and $0 < a_1 < 1$, the (time series) average yield forecast error will be negative for all agents, and (26) implies a persistent forecast error.

D. *Forecast error on forecast revision*

Inserting the expression for the forecast revision in the forecast error (26) and using the dynamics of z and d we get:

$$FE_{t,t+1}^i = b_0 + \left(\frac{b_1}{a_1 K_t^i} - 1 \right) FR_t^i + f(\text{lagged sentiment}) + f(\text{current and past shocks}) \quad (29)$$

Therefore, the coefficient of the forecast error on the forecast revision is negative, as in the data, if sentiment is not too persistent, i.e. $b_1 < a_1 K_t^i$. However, this regression in the data might suffer from endogeneity since the independent variable FR_t^i is correlated with the shocks.

F.   Empirical Measures of 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, $E_t^i(gdp_{t+1})$, and compare them to the expectation implied by a benchmark model, $E(gdp_{t+1}|\mathcal{M}_t)$. 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. Our measure of growth sentiment is then the difference between the survey and model-implied expectation of GDP growth for the average agent: $\mathcal{S}_t^{gdp} = E_t^i(gdp_{t+1}) - E(gdp_{t+1}|\mathcal{M}_t)$, 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.

This measure of GDP growth sentiment $\mathcal{S}_t^{gdp}$ however is only available at quarterly frequency. Therefore, we also construct 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 4 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.[6] 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.[7]

Our results suggest that there is a *systematic* behavioral 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, as shown in the paper.

---

[6]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 VI of the paper, since it is available monthly instead of quarterly.

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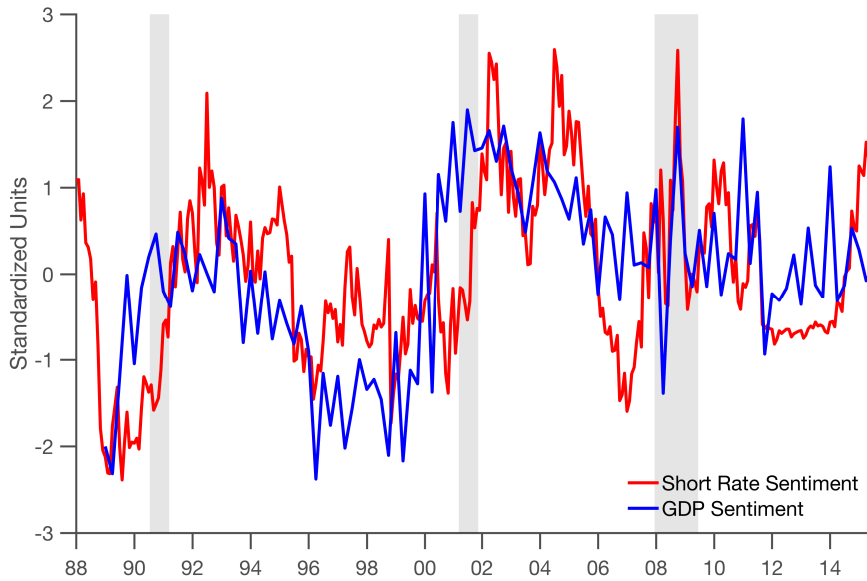
(a)
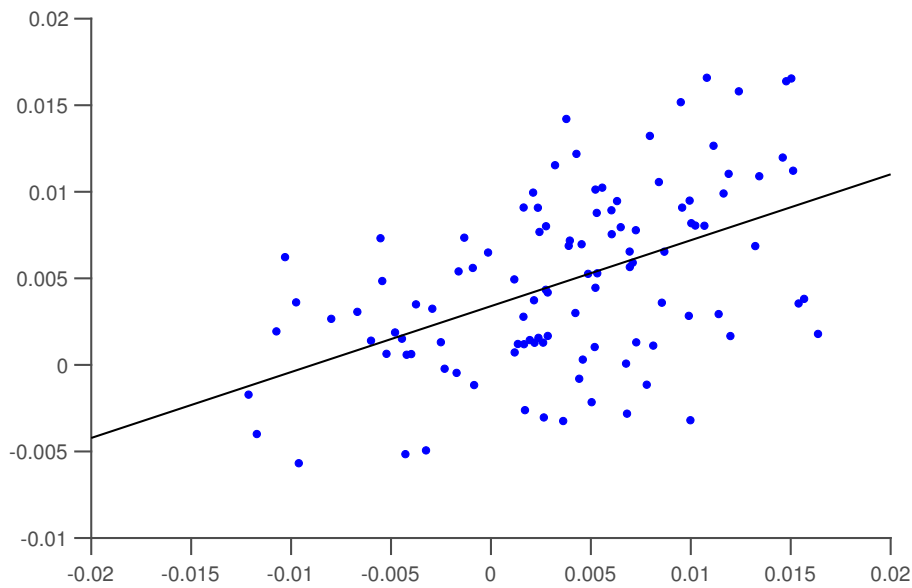
(b)

Figure 4. 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.
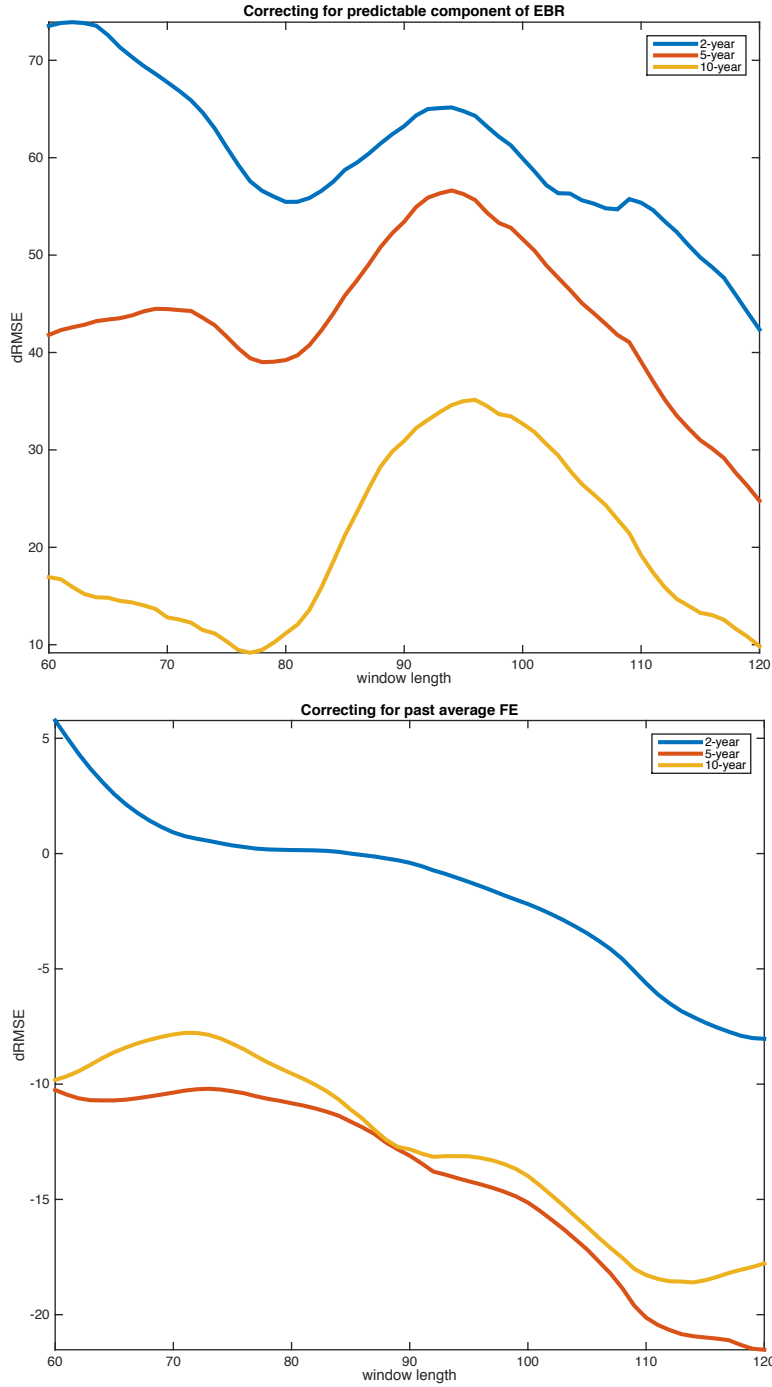
G. Additional Results

Figure 5. 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. In the top panel, the correction is based on a projection of the forecast errors realized over the rolling window on forward spreads. In the bottom panel, we correct $EBR^\star$ only by the average realized forecast errors over the rolling windows.

16

	$\mathcal{S}$	$DiB(g)$	$DiB(\pi)$	$Surp$	$LRR(g)$	$LRR(\pi)$	$\widehat{\sigma}$	$\overline{R}^2$
	−0.26							6.46
	(−3.10)							
(I)		0.05	0.12					2.44
		(0.60)	(1.51)					
	−0.29	0.05	0.17					10.54
	(−3.49)	(0.58)	(2.03)					
(II)				−0.04				−0.25
				(−0.56)				
	−0.27			−0.08				6.74
	(−3.27)			(−1.15)				
(III)					−0.04	0.29		6.88
					(−0.61)	(2.39)		
	−0.36				−0.14	0.41		18.54
	(−4.81)				(−2.04)	(3.35)		
(IV)							0.23	5.21
							(3.46)	
	−0.38						0.35	17.98
	(−4.24)						(4.84)	

Table III. Determinants of Ex-Ante Subjective Bond Returns

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 1993 to July 2015.