Increasing Overreaction and Excess Volatility of Long Rates

Daniele d'Arienzo*

December 9, 2019

Please click here for the latest version. Do not circulate.

Abstract

Giglio and Kelly (2017) find that the volatility of long-term rates is too large relative to that of short-term rates for a large class of rational expectations models. I assess the possibility that such excess volatility may come from investor beliefs. I use survey data on analyst expectations and data on market beliefs recovered from observed yields using the methodology of Ross (2015). I obtain three main findings. First, the two datasets reveal a remarkably similar pattern of *horizon dependent* departures from rationality: expectations about long rates over-react relative to expectations about short rates. Second, a model of diagnostic expectations rationalizes this horizon dependent belief distortions and generates excess volatility of long term rates. Third, when calibrated to the data, this model accounts from roughly 80% of the excess volatility puzzle for a reasonable value of the diagnosticity parameter.

*Bocconi University, Department of Finance, via Röntgen 1, 20136 Milan, Italy, daniele.darienzo@unibocconi.it. I am deeply indebted with Nicola Gennaioli for invaluable guidance. I thank for helpful comments Pedro Bordalo, John Campbell, Francesco Corielli, Sam Hanson, Eben Lazarus, Peter Maxted, Fulvio Ortu, Andrei Shleifer, Claudio Tebaldi, Yijun Zhou (discussant) and seminar participants at Bocconi University, Harvard University, Boston College, TADC 2019 and Sloan-Nomis 2019. All errors are my own.

1 Introduction

Since Shiller (1981) celebrated analysis of the stock market, many studies have documented the so called *excess volatility of asset prices* relative to measures of fundamentals (De Bondt and Thaler (1985), LeRoy and Porter (1981) and Campbell and Shiller (1987) among others). There are two main explanations for this finding. One of them assumes rational expectations and emphasizes the role of discount rate variation (Campbell (2003), Cochrane (2011)). The other one relaxes rationality and views price volatility as reflecting excess volatility of beliefs. Consistent with the latter view, a growing body of work documents excess volatility of beliefs using survey data, e.g. Gennaioli and Shleifer (2018).

A recent paper by Giglio and Kelly (2017) assesses these hypotheses in the realm of the term structure of interest rates. This setting is appropriate because rational term structure models impose precise constraints on the amount of covariation of asset prices at different maturities, even after adjusting for discount rate variation. The key finding of this analysis is that long term interest rates are significantly more volatile with respect to short term interest rates relative to what rational models predict. This finding indicates that beliefs may be a key driver of excess volatility in an important setting such as the bond market where riskless rates are determined. This analysis raises three questions. First, can one go beyond assessing volatility in excess of discount rates and measure beliefs in the bond market? Second, if the answer is affirmative, can one assess which departure of rationality, if any, do these beliefs display? Third, what is the psychological foundation of this departure from rationality, and can it account for the excess volatility of long term rates, not only qualitatively but also quantitatively? This paper seeks to address these questions.

One immediate way to make progress is to proxy the beliefs of the bond market by using expectations data. Figure (1) reports a preliminary analysis using data from Blue Chip on professional forecasts of future rates at maturities of $1y$, $2y$, $5y$, $10y$, $20y$ and $30y$ (monthly frequency). For each maturity, I report the pooled correlation between the forecast revision made by the analyst[1] and his subsequent forecast error. Two features stand out in the

[1]The time t forecast revision of a variable X_{t+m} $(m > 0)$ is defined as the difference between the time t forecast of X_{t+m} and the time $t-1$ forecast of X_{t+m}.

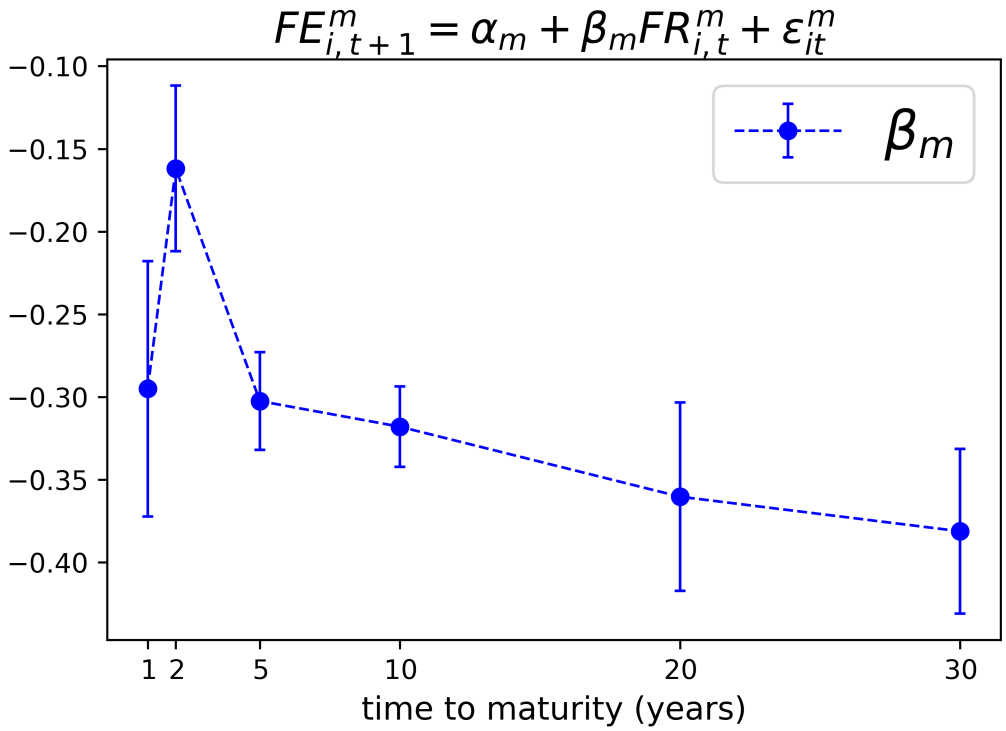


Figure 1: Sensitivity of forecast error to past forecast revision using monthly professional forecasts of (annualized) interest rates.

data. First, at any maturity the average analyst over-reacts to news: when he revises his forecast of future rates up, reality systematically falls below his revised expectation. This is captured by the fact that in Figure (1) all correlations are negative. Second, there is more over-reaction for longer term interest rates, namely the coefficient becomes more negative for longer maturities. This latter aspect especially resonates with the possibility that long-term expectations may move too much with news, causing excess volatility in long-term rates relative to short-rates.

To analyze systematically this possibility, in Section 2 I introduce non-rational beliefs into an otherwise standard affine term structure model. In particular, I allow for distortions in belief formation that are maturity dependent. In this analysis, I show that stronger over-reaction of beliefs for longer maturities, as displayed in Figure (1), indeed gives rise to excess volatility of equilibrium long term rates relative to short term ones. I also show that, within the class of affine models, there is a precise mapping between excess volatility of interest rates at a given maturity and the correlation between forecast revisions and forecast errors

at the same maturity. That is, there is a sort of "duality" between the estimates in Figure (1) and the amount of interest rate volatility measured from equilibrium yields.

The remainder of the paper systematically studies this connection. One limitation in the analysis of Figure (1) concerns the use of survey data. On the one hand, such data are available only for a small subset of maturities, frequencies, and time periods. One would ideally want to have a richer term structure to fully assess the volatility of beliefs. On the other hand, the beliefs of professional forecasters may not be representative about the beliefs of bond traders. Ideally, one would want to measure the beliefs of the marginal investor, and this is clearly not available in conventional datasets.

To address these measurement issues, in Section 3 I use a method developed by Ross (2015) to extract market beliefs from equilibrium yields. This method rests on two assumptions: i) the dynamic of risk factors moving interest rates is stationary, and ii) the stochastic discount factor is path independent. Borovička et al. (2016) criticized assumption ii), showing that it does not hold for asset pricing models that contains long run risks (Bansal and Yaron (2004)) or that more generally include a martingale component in the SDF. Against this critique, I show that even if recovered beliefs are misspecified in the Borovička et al. (2016) sense, the use of Ross' method is still informative with respect to the maturity dependent over- (or under-) reaction that Figure (1) exemplifies. The intuition is that maturity dependent information processing distortions entail a violation of the law of iterated expectations, and such violation cannot be obtained under any rational asset pricing model, including those incorporating long run risks or more generally martingale components in the SDF.

I thus proceed to implement Ross' method for the yield curve, and describe the main patterns of this data in Section 4. The recovered beliefs exhibit, similarly to the professional forecasts, increasing over-reaction at long maturities (greater than $10y$). To validate the use of Ross recovered beliefs, I systematically compare the latter to professional forecasts. The two measured of beliefs display a remarkably strong and positive correlation, which suggests that recovered beliefs are not noise, and instead capture systematic features of market expectations. Interestingly, I find that, compared to the cross section of professional forecasts, Ross recovered beliefs weigh more heavily unbiased and accurate beliefs (relative to equally weighted averages). This property can be traced back to the working of arbitrage

capital (Buraschi et al. (2018)), which partially (but not fully) offsets individual beliefs distortions.

Having validated Ross recovered beliefs, and having confirmed the broad pattern of maturity increasing over-reaction, In Section 5 I ask: what is the psychology of maturity increasing over-reaction? And can this form of belief distortions account quantitatively for excess volatility in long term rates? To answer the first question, I adapt to a term structure setting the diagnostic expectations model of Bordalo et al. (2017). This model features over-reaction and it is grounded on the Tversky and Kahneman (1974) representativeness heuristic. The basic principle of diagnostic beliefs is "kernel of truth": beliefs move in the right direction but exaggerate true features of the data generating process. I show that in a term structure model the kernel of truth logic naturally yields maturity increasing over-reaction. The intuition is that a diagnostic agent will over-react more strongly to a given signal when there is large fundamental uncertainty, which is indeed the case for longer term rates. The model provides therefore a foundation for increasingly over-reacting beliefs and for the violation of the law of iterated expectations, and in a way that is disciplined by the parameters of the data generating process. The latter property proves useful for quantification.

I then conclude the analysis by calibrating the diagnostic expectations model and by quantifying its ability to capture excess volatility in beliefs and in interest rates. I find that: i) the diagnosticity distortion parameter calibrated from recovered beliefs is consistent with previous literature (Bordalo et al. (2018a)), ii) such parameter matches fairly well the documented average over-reaction of analyst forecasts, and iii) it accounts for roughly 80% of the excess volatility in interest rates documented by Giglio and Kelly (2017).

My paper contributes to a growing body of research aimed at *testing* the rational expectation assumption with beliefs data. Bordalo et al. (2018a) finds evidence of over-reaction to information in several *individual* macroeconomic and financial time series. Piazzesi and Schneider (2011) finds that traditional factors (level and slope) are perceived as more persistent by financial analysts and Cieslak (2018) finds that systematic errors in short rate expectations drive bond returns predictability as opposed to time varying risk premia. Brooks et al. (2018) shows that beliefs of professional forecaster over-react to FOMC announcements, generating post-announcement drift and excess sensitivity of long term rates. Relative to these

papers, I use Ross method to recover market beliefs, offer a parsimonious characterization of belief distortions based on diagnostic expectations, and quantify the model's ability to account for the volatility anomaly.

This paper also relates to the debate about the recovery theorem. Martin and Ross (2019) investigates the recovery theorem in fixed income markets, where the assumption of stationary and Markovian state variables may be more plausible, relative to equity markets. My setting is an empirical counterpart of Martin and Ross (2019), where I test for the rationality of recovered beliefs. Qin et al. (2018) propose an empirical test for the degeneracy of the martingale component of the SDF in the context of US treasury bonds and reject it, but the test assumes rational expectations. Similarly to Qin et al. (2018), I implement the Ross recovery theorem using the pricing measure from the estimation of a standard $\mathbb{Q}$-affine term structure model. This class of models is flexible since it does not entail assumptions on the physical measure (Le et al. (2010)). Close to the intuition of the recovery theorem, Augenblick and Lazarus (2017) provide evidences of excess movements of stock market prices, particularly strong for long horizons.

My contribution to this literature is to consider a pattern, maturity increasing over-reaction, that is robust to the Borovička et al. (2016) misspecification. Furthermore, while work on recovery is mostly methodological, I offer a systematic characterization of belief formation and account of the excess volatility pattern.

The paper unfolds as follows: in Section 2, I introduce increasingly over-reacting beliefs into an otherwise standard affine model for the term structure, and I show that in this economy the error predictability of Figure (1) and the excess volatility of Giglio and Kelly (2017) are closely related. In Section 3, I discuss the recovery theorem and its empirical implementation. In Section 4, I compare recovered beliefs with survey data. In Section 5, I introduce a model of beliefs formation and I quantitatively assess the increasingly over-reacting beliefs channel for excess volatility. In Section 6, I provide robustness checks. Section 7 concludes. All proofs are in Appendices.

2 Over-reaction to news and excess volatility

Figure (1) shows that analysts expectations about future interest rates tend to over-react to news, more strongly so for longer maturities. This section formally shows that, within the conventional class of $\mathbb{Q}$-affine models, there is a direct link between such horizon-dependent over reaction in beliefs and the excess volatility of long term interest rates documented by Giglio and Kelly (2017). This connection informs my empirical analysis, aimed at: i) measuring market beliefs, ii) characterizing their departure from rationality, and iii) quantifying the latter's impact on the excess volatility of interest rates. I illustrate the basic logic in a one factor economy, while, for the empirical analysis, I consider multiple factors.

Consider a frictionless market, where zero coupon bonds with different maturities are traded. $P_{t,m}$ denotes the price at time t of a zero coupon bond with time to maturity m. The yield to maturity, $y_{t,m}$, is defined as:

$$P_{t,m} = e^{-m \cdot y_{t,m}},$$

and the yield curve at time t equals the collection of yields $\{y_{t,m}\}_{m \geq 0}$.

Denote the one period interest rate prevailing at time s by r_s (short rate henceforth). Then, the average one period interest rate obtained on an investment at maturity m is equal to $r_{t,m} := \frac{1}{m} \sum_{i=0}^{m-1} r_{t+i}$. In a world with deterministic interest rates, the yield obtained from investing at maturity m is simply equal to the interest rate at that maturity: $y_{t,m} = r_{t,m}$. If instead the short rate is exposed to risk factors (e.g. GDP growth) the interest rate obtained from an investment at maturity m is not known with certainty. Specifically, assume that the short rate r_s is a function of the risk factor at time s, which is denoted by X_s. Then, while at time t the current short rate r_t is known, longer maturity rates $r_{t,m}$ are not known because the dynamics of the risk factor is stochastic. As a result, the price of the bond and hence the yield to maturity $y_{t,m}$ at t is influenced both by *expectations* about future rates and by *risk aversion*. We now characterize these expectations, and in particular their departure from rationality. Next we show how the same expectations affect – together with risk aversion – the yield curve in equilibrium.

Increasingly Over-Reacting Expectations

As in conventional affine models, the short rate is an affine function of the risk factor X_t, which for simplicity we assume to follow a stationary AR(1). Then, the dynamics of the short rate is fully described by:

$$\begin{cases} r_t = \delta_0 + \delta_1 X_t \\ X_t = \rho^{\mathbb{P}} X_{t-1} + \sigma^{\mathbb{P}} \varepsilon_t^{\mathbb{P}}, \end{cases} \tag{1}$$

where $\epsilon_t^{\mathbb{P}}$ is an i.i.d. shock. In this notation, superscript $\mathbb{P}$ captures the so called "physical measure", or data generating process. As a consequence, the dynamics of interest rates at maturity m, $r_{t,m} := \frac{1}{m}\sum_{i=0}^{m-1} r_{t+i}$, reads:

$$r_{t,m} = \delta_0 + b_m^{\mathbb{P}} X_t + \varepsilon_{t,m}^{\mathbb{P}},$$

where $b_m^{\mathbb{P}} := \frac{\delta_1}{m}\sum_{i=0}^{m-1}(\rho^{\mathbb{P}})^i$ and $\varepsilon_{t,m}^{\mathbb{P}} := \frac{\delta_1}{m}\sum_{k=1}^{m-1}\sum_{i=1}^{k}(\rho^{\mathbb{P}})^{k-i}\sigma^{\mathbb{P}}\varepsilon_{t+i}^{\mathbb{P}}$. The coefficients $b_m^{\mathbb{P}}$ capture the sensitivity of interest rates at maturity m, $r_{t,m}$, to current information (X_t) under the physical measure. This sensitivity *geometrically* decays to zero as the maturity increases. $\varepsilon_{t,m}^{\mathbb{P}}$ captures fundamental risk about interest rates at maturity m.

We allow expectations of future interest rates, to depart from rationality. In particular, we allow them to over-react to the current state in a maturity dependent fashion, as suggested by Figure (1). Formally, we assume that at time t the market perceives interest rates at maturity m to be:

$$r_{t,m}^{\theta} = \delta_0 + b_m^{\mathbb{P}}(1+\psi_m^{\theta})X_t + \sigma_m^{\theta}\varepsilon_{t,m}^{\mathbb{P}}, \tag{2}$$

where θ denotes departures from rational expectations. There are two such departures. First, the variance of fundamental shocks is potentially distorted, with $\sigma_m^{\theta} \neq 1$. Second, and more important, beliefs of future rates may display excess sensitivity ($\psi_m^{\theta} \geq 0$) to the current state. The beliefs in Equation (2) arise when agents perceive interest rate shocks $\varepsilon_{t,m}^{\mathbb{P}^{\theta}} := \sigma_m^{\theta}\varepsilon_{t,m}^{\mathbb{P}} + \psi_m^{\theta} b_m^{\mathbb{P}} X_t$ so that perceived news are distorted toward the current state X_t. $\mathbb{P}^{\theta}$ denotes a distorted data generating process measure for the factor, and the derivation of Equation (2) from the distorted factor dynamics is performed in Appendix A.

Equation (2) implies that expectations formed at time t about interest rates at maturity m take the convenient intuitive form:

$$\mathbb{E}_t^{\mathbb{P}}\left[r_{t,m}^{\theta}\right] =: \underbrace{\mathbb{E}_t^{\mathbb{P}^\theta}\left[r_{t,m}\right]}_{\text{distorted expectations}} = \underbrace{\mathbb{E}_t^{\mathbb{P}}\left[r_{t,m}\right]}_{\text{rational expecations}} + \psi_m^{\theta} \underbrace{\left(\mathbb{E}_t^{\mathbb{P}}\left[r_{t,m}\right] - \mathbb{E}^{\mathbb{P}}\left[r_{t,m}\right]\right)}_{\text{news relative to the average}}. \tag{3}$$

The distorted expectation $\mathbb{E}_t^{\mathbb{P}^\theta}[\cdot]$ is equal to the rational expectation plus an adjustment in the direction of the news, where the latter is captured by $\mathbb{E}_t{}^{\mathbb{P}}\left[r_{t,m}\right] - \mathbb{E}^{\mathbb{P}}\left[r_{t,m}\right]$.

Departures from rationality at maturity m are parameterized by the coefficient $\psi_m^{\theta} \geq 0$ When $\psi_m^{\theta} = 0$, expectations are rational. When $\psi_m^{\theta} > 0$ agents overract to news. That is, they over-estimate future rates in states truly indicative of higher than average future rates $\mathbb{E}_t^{\mathbb{P}}\left[r_{t,m}\right] > \mathbb{E}^{\mathbb{P}}\left[r_{t,m}\right]$. By contrast, agents under-estimate future rates in states truly indicative of lower than average future rates $\mathbb{E}_t^{\mathbb{P}}\left[r_{t,m}\right] < \mathbb{E}^{\mathbb{P}}\left[r_{t,m}\right]$.[2]

I allow the distortion parameter ψ_m^{θ} to be maturity-dependent. For now, I leave such dependence unspecified. In Section 5, however, I show that under Gaussian noise, Equation (3) naturally follows from the diagnostic expectation model of Bordalo et al. (2018b). Such model also yields horizon dependent distortion parameters ψ_m^{θ} that are in turn functions of a more primitive distortion parameter θ. This is why I denote distorted expectations using θ.

2.1 Risk Aversion and the Yield Curve

Consider a market forming expectations according to Equation (2). What does the yield curve looks like? To answer this question, one must also consider investors' risk aversion, which affects – together with beliefs – required rates of return. To capture risk aversion, consider a stochastic discount factor (SDF) $M_{t,m}$ discounting more heavily future cash flows occurring in states in which the representative investor is poorer. Then the price of a maturity m zero

[2]The comparison with average information can be generalized to the comparison with a weighted sum of past predictions. This case includes the model of diagnostic expectations of Bordalo et al. (2018b) and it is discussed in Section 6.

coupon bond, and hence yields $y_{t,m}$, is pinned down by the discounted expected payoff[3]:

$$P_{t,m} = \mathbb{E}_t^{\mathbb{P}^\theta}\left[M_{t,m}\right] := \mathbb{E}_t^{\mathbb{Q}^\theta}\left[e^{-m\cdot r_{t,m}}\right], \tag{4}$$

where the so called *risk neutral* probability measure $\mathbb{Q}^\theta$ inflates the perceived probability of states in which the representative investor is poor. By using $\mathbb{Q}^\theta$ the economic analyst can account for risk aversion while using the convenient analytics prevailing under risk neutrality. By the fundamental asset pricing equation, given a stochastic discount factor $M_{t,m}$, the risk neutral measure density $f_{\mathbb{Q}^\theta}(X_{t+m}|X_t)$ associated to it is implicitly defined by the condition:

$$f_{\mathbb{Q}^\theta}(X_{t+m}|X_t)e^{-m\cdot r_{t,m}} = M_{t,m} f_{\ \mathbb{P}^\theta}(X_{t+m}|X_t),$$

which captures the optimality condition for a market with distorted beliefs captured by $\mathbb{P}^\theta$. Here I emphasize that under $\mathbb{P}^\theta$ the factor is still an AR(1) with with maturity dependent distorted persistence and volatility (the full analytics is performed in appendix A). To use the machinery of affine term structure models, we assume that the stochastic discount factor is such that the distorted and risk neutral dynamics $\mathbb{Q}^\theta$ for the factor is AR(1), with maturity dependent persistence and volatility. In this case, the risk adjustment to beliefs $\mathbb{P}^\theta$ yields the following distorted and risk neutral dynamics for interest rates at maturity m:

$$r_{t,m}^\theta = \delta_0 + b_m^{\mathbb{Q}} X_t + \psi_m^\theta b_m^{\mathbb{Q}} X_t + \varepsilon_{t,m}^{\mathbb{Q}},$$

where $b_m^{\mathbb{Q}} := \frac{\delta_1}{m}\sum_{i=0}^{m-1}(\rho^{\mathbb{Q}})^i$ and $\varepsilon_{t,m}^{\mathbb{Q}} := \frac{\delta_1}{m}\sum_{k=2}^{m}\sum_{i=0}^{k-2}(\rho^{\mathbb{Q}})^i\sigma^{\mathbb{P}}\varepsilon_{t+1+i}^{\mathbb{Q}}$. $\rho^{\mathbb{Q}} \neq \rho^{\mathbb{P}}$ is the risk neutral persistence of the factor under rational expectations, while $\varepsilon_{t+1}^{\mathbb{Q}}$ is a zero mean shock with risk neutral variance $\sigma^{\mathbb{Q}} \neq \sigma^{\mathbb{P}}$, again under rational expectations. $b_m^{\mathbb{Q}}$ is the risk neutral sensitivity of future rates to X_t and $\varepsilon_{t,m}^{\mathbb{Q}} := \frac{\delta_1}{m}\sum_{k=1}^{m-1}\sum_{i=1}^{k-1}(\rho^{\mathbb{Q}})^{k-1}\sigma^{\mathbb{Q}}\varepsilon_{t+i}^{\mathbb{Q}}$ is the risk neutral shock at maturity m, both under rational expectations. Importantly, under the belief distor-

[3] Note that under the maturity dependent over-reaction model we have that the law of iterated expectations fails, in the sense that $\mathbb{E}_t^{\mathbb{P}^\theta}\left[M_{t,t+2}\right] \neq \mathbb{E}_t^{\mathbb{P}^\theta}\left[M_{t,t+1}\mathbb{E}_{t+1}^{\mathbb{P}^\theta}[M_{t+1,t+2}]\right]$. We need therefore to take a stance about valuation. In line with the logic of Figure (1), we assume that the market forecasts future rates and price future states with a "buy and hold" valuation approach, namely we set $P_{t,t+2} = \mathbb{E}_t^{\mathbb{P}^\theta}\left[M_{t,t+2}\right]$.

tions of Equation (2) the risk adjusted dynamics of interest rates still exhibit overreaction (ψ_m^θ) to the current state X_t.

Proposition 1. *Assume homoskedastic $\mathbb{Q}$-shocks. The yield curve under over-reacting beliefs is equal to:*

$$y_{t,m}^\theta = -\frac{1}{m}\log \mathbb{E}_t^{\mathbb{Q}^\theta}[e^{-m\cdot r_{t,m}}] = a_m^\theta + (b_m^{\mathbb{Q}} + \psi_m^\theta b_m^{\mathbb{P}})X_t, \tag{5}$$

where:

$$b_m^{\mathbb{Q}} = \frac{\delta_1}{m}\sum_{i=0}^{m-1}(\rho^{\mathbb{Q}})^i,$$
$$a_m^\theta = \delta_0 - \frac{1}{m}\log \mathbb{E}^{\mathbb{Q}}[e^{-m\cdot\sigma_m^\theta \varepsilon_{t,m}^{\mathbb{Q}}}].$$

When expectations are rational, namely $\psi_m^\theta = 0$, Equation (5) is the signature *affine term structure model* (see Duffee (2013) for a review). The coefficients a_m and $b_m^{\mathbb{Q}}$ are maturity dependent. Interest rates volatility is shaped by the sensitivity $b_m^{\mathbb{Q}}$ of yields to changes in the factor X_t. The more persistent is the factor, namely the higher is $\rho^{\mathbb{Q}}$, the more sensitive is the interest rate at any given maturity to news, namely the higher is $b_m^{\mathbb{Q}}$. On the other hand, because the factor is stationary, $|\rho|^{\mathbb{Q}} < 1$, long term rates should be less sensitive than short term rates, namely $b_m^{\mathbb{Q}}$ declines with m. The Giglio and Kelly (2017) excess volatility puzzle is rooted in the maturity profile of the $b_m^{\mathbb{Q}}$ terms: they decay too slowly relative to what is implied by the persistence $\rho^{\mathbb{Q}}$ of the factors. The issue is then: can the horizon dependent over-reaction of Figure 1 rationalize this finding?[4]

With over-reacting beliefs, $\psi_m^\theta > 0$, equilibrium yields retain a tractable form. Overreaction in beliefs enhances the sensitivity of yields to the risk factor relative to a rational world,

[4]A different yet important issue is the extent to which $\mathbb{Q}$-affine models correctly capture the yield curve dynamics. To this regard, it is worth noting that: i) empirically, at each maturity m, the same risk factors explain the variability of yields to that maturity, $y_{t,m}$, with R^2 close to one as shown in Section 3, ii) Giglio and Kelly (2017) show that quadratic $\mathbb{Q}$-specification, stationary long memory process nor regime switching models cannot account for the excess volatility and iii) $\mathbb{Q}$-affine models allows highly non linear $\mathbb{P}$ -dynamics as well as non standard SDFs, provided that their products yields $\mathbb{Q}$ affinity.

namely $b_m^{\mathbb{Q}} \to b_m^{\mathbb{Q}} + \psi_m^\theta b_m^{\mathbb{P}}$. This formula connects over-reaction in beliefs and excess volatility in interest rates.

To see this connection more precisely, consider expectations about future rates, computed using the distorted physical measure $\mathbb{P}^\theta$ and consider the measured volatility of interest rates. I obtain the following result.

Theorem 1. *(Over-reaction and excess volatility).*

Under expectations (3) and the affine setting:

i) (Increasing Overreaction) the CG coefficients β_m obtained by regressing the forecast error made at maturity m using $\mathbb{P}^\theta$ with the forecast revision under $\mathbb{P}^\theta$ about the same maturity are equal to:

$$\beta_m = -c\frac{\psi_m^\theta}{1+\psi_m^\theta},$$

where c is a positive, maturity independent, constant;

ii) (Excess Volatility) the volatility of yields at maturity m relative to the volatility of the short rate is equal to:

$$\frac{\mathbb{V}^{\mathbb{P}}[y_{t,m}^\theta]}{\mathbb{V}^{\mathbb{P}}[y_{t,1}^\theta]} = \left(\frac{b_m^{\mathbb{Q}} + \psi_m^\theta b_m^{\mathbb{P}}}{b_1^{\mathbb{Q}} + \psi_1^\theta b_1^{\mathbb{P}}}\right)^2 > \frac{\mathbb{V}^{\mathbb{P}}[y_{t,m}]}{\mathbb{V}^{\mathbb{P}}[y_{t,1}^\theta]},$$

where $\mathbb{V}^{\mathbb{P}}[y_{t,m}^\theta]$ is the measured volatility of yields while $\mathbb{V}^{\mathbb{P}}[y_{t,m}]$ is the volatility arising under rational expectations. Over reaction to news yields both the pattern in Figure (1) and excess volatility of long term rates if and only if ψ_m^θ is positive and increases in maturity m.

This result conveys two important messages. First over-reaction to news increasing in the horizon m reconciles the observed patterns in forecast errors and excess volatility in long term rates. This is a general result, and holds beyond the more restrictive assumptions of this Section. As I show in the proof of Theorem (1), such connection holds under general $\mathbb{Q}$-affine models, where $\mathbb{P}$ is Markovian and $\mathbb{Q}$ is AR(1), provided expectations about future interest rates over-react according to Equation (3).

Second, and crucially, Theorem (1) says that in the conventional class of $\mathbb{Q}$-affine term structure models, when the data generating process $\mathbb{P}$ for the risk factors is itself AR(1), the maturity dependent over-reacting beliefs in Equation (2) create a precise link between the over-reaction coefficient measured using beliefs data and the excess volatility detected from prices. They are both pinned down by the same maturity increasing distortion ψ_m^θ.

The remainder of the paper assesses empirically these predictions. In Sections 3 and 4 I start tackling this challenge by offering a measure of market beliefs. The survey evidence in Figure (1) has two pitfalls. First, it has only a limited number of maturities. Second, and most important, the beliefs of professional forecasters may not be representative of the beliefs of market participants. To overcome these issues, I use methods developed by Ross to extract information about beliefs from asset prices in conventional affine models[5].

In Section 5 I add the assumption that the physical measure $\mathbb{P}$ is a Gaussian AR(1). In this case, the over-reaction parameters ψ_m^θ can be conveniently founded using the diagnostic expectations model (Bordalo et al. (2018b)). This allows me to obtain an estimate of their distortion parameter θ that can be benchmarked to existing estimates. Furthermore, Theorem (1) highlights a duality between over-reaction in beliefs and excess volatility in interest rates, so that ψ_m^θ can be estimated both from beliefs data and directly from yields. I use this duality to assess which conclusions about ψ_m^θ are most robust.

3 Beyond survey data: Ross Recovery Theorem and the term structure of beliefs

The method of Ross (2015), rests on two assumptions: i) the underlying state of the economy follows a stationary Markovian process, both under $\mathbb{P}$ and under $\mathbb{Q}$, and ii) the SDF $M_{t,t+m}$ is *path independent*, namely it depends only on the final value X_{t+m} and on the initial value X_t of the state variable, and not on the *path* from t to $t+m$. Assumption i) is an approximation, but of significant empirical power: it is widely recognized that few state variables drive the yield curve (see Duffee (2013)) and stationarity is not rejected in the data (see Giglio and Kelly (2017) and Martin and Ross (2019)). Assumption ii) is more controversial: Borovička

[5]I will consider prices of US treasury bonds which are directly related to interest rates and yields.

et al. (2016) shows that path independence is not met in "long run risk" asset pricing model. In this case, the method of Ross (2015) does not recover market beliefs, but beliefs adjusted for a martingale component. This is an important criticism but, as I show below, my analysis is immune to it because the increasingly over-reacting beliefs in (3) can still be recovered from the data.

To grasp how Ross's method works, consider an Arrow-Debreu security that pays one dollar if next period's state is j (assume for simplicity that there is a finite number N of states, which correspond to factor values in my setting). Under rational expectations, if the current state is i, the price of such Arrow Debreu security is equal to:

$$\mathbb{A}_{ij} = M_{ij}\mathbb{P}_{ij},$$

where $\mathbb{P}_{ij} := \mathbb{P}(X_{t+1} = j|X_t = i)$ is the physical probability of transitioning from state i to state j and M_{ij} is the stochastic discount factor (SDF) capturing the marginal rate of substitution between current consumption in i and future consumption in j. Due to risk aversion, M_{ij} overweights bad states, attaching a higher price to a safe bond because it pays out in them.

In the presence of non-rational beliefs, the fundamental asset pricing equation implies that the price of the same Arrow Debreu security is equal to:

$$\mathbb{A}^{\theta}_{ij} = M_{ij}\mathbb{P}^{\theta}_{ij}.$$

An econometrician observing Arrow Debreu prices may recover market beliefs, be they rational $\mathbb{P}_{ij}$ or distorted $\mathbb{P}^{\theta}_{ij}$, by performing an "inverse risk adjustment" to the prices themselves. Ross has shown that such inverse risk adjustment is indeed possible, so that market beliefs can be recovered from prices, under assumptions i) and ii) above. Ross's method has been so far used under the assumption of rational expectations.

As displayed in Figure (2) above, when market beliefs may not be rational, the use of Ross' method entails an ambiguity: the econometrician does not know if Arrow Debreu prices are $\mathbb{A}$ (they reflect rational beliefs) or if they are $\mathbb{A}^{\theta}$ (they reflect non rational beliefs). Here I proceed as follows: I use Ross' method, and then performed on the recovered market beliefs

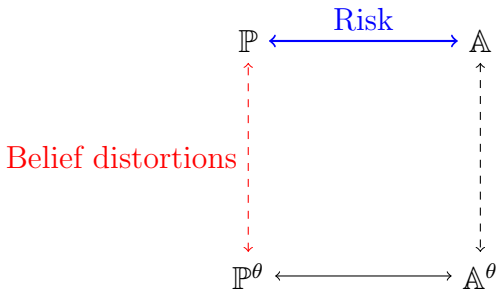


Figure 2: The blue line adjusts the data generating process $\mathbb{P}$ for preferences, which determines Arrow-Debreu prices $\mathbb{A}$. The red line distorts the data generating process $\mathbb{P}$, thus defining beliefs $\mathbb{P}^\theta$, thus generating Arrow-Debreu prices $\mathbb{A}^\theta$.

some statistical tests of rationality, considering in particular the predictability of forecast errors. The outcome of this test tells us if we are in top or bottom row of Figure (2).

Before carrying out the analysis, consider again the critical path independence assumption i). When the SDF is *path independent*, it means that there exists a constant δ and function z of the state such that the one period SDF can be written as:

$$M_{ij} = \delta \frac{z_i}{z_j}.$$

This property is satisfied by conventional CRRA preferences (over stationary variables). However, this assumption has been criticized by Borovička et al. (2016), who show that it assumes away the kind of *long-run* risk adjustments from the SDF that are embedded in many conventional consumption based asset pricing models (CRRA when consumption exhibit stochastic trends, long run risk models as well as habit formation models)[6]. As they show, assuming path independence in these cases is akin to neglecting the fact that the SDF also contains a *martingale* component that changes over time. Does such misspecification invalidate the detection of the horizon dependent over-reacting beliefs that I consider here? To answer this question, suppose that the SDF has a martingale component. Then, by applying Ross method we would not recover the true market beliefs $\mathbb{P}^\theta$ but the beliefs $\tilde{\mathbb{P}}^\theta$ contaminated by misspecification. I then obtain the following result.

[6]However Walden (2017) explicitly showed that relabeling the state variables may lead to the assumption to be met, in the case of CRRA utility with non stationary consumption.

Theorem 2. *Consider the $\mathbb{Q}$-affine setting, non rational expectations as in (3) and suppose that the martingale component of the SDF is non degenerate, so that the econometrician detects $\tilde{\mathbb{P}}^\theta$, not $\mathbb{P}^\theta$. Then, the following are equivalent:*

i) the CG coefficients estimated by regressing forecast errors of Ross recovered beliefs on forecast revisions of Ross recovered beliefs vary with maturity m.

ii) $\mathbb{E}^\theta[\cdot]$ violates LIE.

Moreover, the CG coefficients obtained from Ross recovered beliefs are negative and decreasing:

$$\beta_m^\theta = -c' \frac{\psi_m^\theta}{1+\psi_m^\theta},$$

where $c' > 0$ is a maturity independent constant.

With misspecification à la Borovička et al. (2016), the econometrician applying Ross' method no longer recovers exact beliefs. Crucially, however, rationality tests still allow her to recover the horizon dependent over-reaction, critical for my analysis, and detected in Figure (1) for professional forecasters. The intuition is that horizon-dependent distortions entail a strong violation of rationality, namely a violation of the law of iterated expectations.[7] As such, it cannot be accounted by any rational expectations models, including those featuring long run risk. In this respect, the Ross recovery theorem remains useful for spotting horizon dependent beliefs distortions.

3.1 Empirical implementation

To apply Ross' method, we need Arrow Debreu prices. These are not directly observed but they can be inferred from market prices. To do so, note that one period ahead Arrow-Debreu prices can also be written as state by state discounted risk neutral probabilities:

$$\mathbb{A}_{ij}^\theta = \mathbb{Q}_{ij}^\theta e^{-r(i)},$$

[7]Formally, the LIE is a mathematical theorem which holds true for every probability measure. The inconsistency of $\mathbb{P}^\theta$ at different maturities is mathematically due to the fact that $\mathbb{P}^\theta$ is a different probability measure for each maturity, as defined by condition (3).

where $\mathbb{Q}^{\theta}_{ij}$ is the risk neutral probability of transitioning from state i to state j. The above formula includes both the case of rational expectations ($\theta = 0$) as well as non rational ones ($\theta \neq 0$).

To compute the risk neutral probabilities $\mathbb{Q}^{\theta}_{ij}$ I first estimate the risk neutral parameters of a three factor affine model. Then, I discretize the state space using the Rouwenhorst method (Cooley (1995)), compute the discretized transition probabilities $\mathbb{Q}^{\theta}_{ij}$ and finally I discount them by the known short rate. This procedure yields $\mathbb{A}^{\theta}_{ij}$.

I follow conventional methods and consider as factors the first three principal components of the yield curve (see Duffee (2013)). The construction of the factors is discussed in Appendix B. The three factor $\mathbb{Q}^{\theta}$-dynamics takes the form:

$$\begin{cases} r_t = \delta_0 + \delta_1^{\top} \mathbf{X}_t \\ \mathbf{X}_{t+1} = \rho^{\mathbb{Q}^{\theta}} \mathbf{X_t} + \Sigma^C \epsilon^{\mathbb{Q}^{\theta}}_{t+1}, \end{cases}$$

where $\epsilon^{\mathbb{Q}^{\theta}}_t$ are i.i.d. shocks. The matrix $\rho^{\mathbb{Q}^{\theta}}$ is assumed to be diagonal, $\rho^{\mathbb{Q}^{\theta}} = diag(\rho_1^{\mathbb{Q}^{\theta}}, \rho_2^{\mathbb{Q}^{\theta}}, \rho_3^{\mathbb{Q}^{\theta}})$.[8] Σ^C is the Cholesky decomposition of the variance-covariance matrix of the residuals, Σ. Following Cochrane and Piazzesi (2009), I assume that Σ coincides under $\mathbb{P}^{\theta}$ and $\mathbb{Q}^{\theta}$, so I can estimate it with the variance covariance matrix of residuals in a $VAR(1)$ for $\mathbf{X}_t$. Finally, the entries of the diagonal matrix $\rho^{\mathbb{Q}^{\theta}}$ are estimated by matching observed yields.

Due to my emphasis on maturity dependent over-reaction, I perform this estimation maturity by maturity *independently*, since I do not want to impose restrictions across maturities.

Summarizing the procedure:

1. $\hat{\delta}_0$, $\hat{\delta}_1$ are the OLS estimates of the regression of the short rate on the three factors.

2. $\hat{\Sigma}$ is estimated as the variance covariance matrix of the residuals of a VAR(1) for the factors. $\hat{\Sigma}^C$ is the corresponding Cholesky decomposition.

3. The risk neutral parameters of the factor dynamics are estimated, *maturity by maturity*,

[8] This restriction is necessary to achieve identification of the matrix $\rho^{\mathbb{Q}^{\theta}}$ assuming that the noise is Gaussian. In this case, a Gaussian transition probability density is in fact characterized by 9 independent parameters, which parametrize the mean and the covariance matrix (see Hamilton and Wu (2012)).

as:

$$\hat{\rho}_m^{\mathbb{Q}^\theta} := \arg\min_{\rho^{\mathbb{Q}^\theta}} \frac{1}{T}\sum_t \left(y_{t,m} - \bar{y}_{m,t}\right)^2 = \arg\min_{\rho^{\mathbb{Q}^\theta}} \frac{1}{T}\sum_t \left(\hat{a}_m^\theta + b_m^{\mathbb{Q}^\theta} X_t - \bar{y}_{m,t}\right)^2$$

where $b_m^{\mathbb{Q}^\theta}$ is a function of the the parameters $\rho^{\mathbb{Q}^\theta}$, whose analytic expression in computed in Appendix A.

4. One period ahead Arrow-Debreu prices are finally estimated using yields to maturity m only as fitting the average yield curve:

$$\hat{\mathbb{A}}_{ij}^{(m)} = \hat{\mathbb{Q}}_{ij}^{(m)} e^{-\hat{r}(i)}.$$

A final step before recovering beliefs in needed. The Recovery Theorem entails an eigenvalue problem for the Arrow-Debreu matrix, while the affine specification relies on continuous variables. As I discussed before, to tackle this issue I discretize the continuous state space of $\mathbf{X}_t$ by using the Rouwenhorst method (Cooley (1995)), which represents the state of art in approximating an $AR(1)$ process with a finite state space Markov chain. The method generates a Markov chain which matches mean, variance and autocorrelation of the original $AR(1)$ process. These are the moments I am interested in: the mean of the factor determines the behavior of the average yield curve, the autocorrelation and the variance determine the term structure of volatilities. Further details about the implementation are discussed in Appendix B.

4 Recovered Beliefs, Survey data and Rationality tests

I now study recovered beliefs and compare those with survey data from professional forecasters. The latter step helps me validate the analysis because the two sources of data are highly independent.

4.1 Data

US treasury yields

Gürkaynak et al. (2007) provide (and keep updated) nominal, annualized, zero coupon bond yields with yearly maturities from 1 year to 30 years. Gürkaynak et al. (2007) infer the yield curves time series from observed prices of fixed income instruments. The data are jointly available at all maturities from 11-25-1985 to 12-31-2016, at the daily frequency.

The Blue Chip Survey of Professional forecasters

The Blue chip survey of professional forecasters contains forecasts about yields to maturities $1, 2, 5, 10, 20$ and $30y$ from leading financial institutions, which are flagged in the dataset. Forecasts with maturity $1, 2, 5$ and 10 years are available from January 1984, forecasts with maturity $30y$ are available starting from January 2000, while forecasts with maturity $20y$ are available starting from January 2004. Forecasts about the next quarter yield curve are reported at the monthly frequency, so the prediction horizon oscillates between 2 and 6 months[9]. I remove from the sample those forecasters who reply to the survey for a short time period only (less then 5 years). At each time t, I remove those replies to the survey which contain answers for less than 3 prediction horizons: what I am mostly interested in are indeed forecasts at different horizons. Finally, for each prediction horizon, I remove outliers which are defined as observations in the first and last percentile of the distribution of forecasts.

4.2 Tests of rationality

Tests of rationality involve the predictability of the forecast error, defined, for each available maturity as:

$$FE_{t+1}[y_{t+1,m}] = \bar{y}_{t+1,m} - \hat{y}_{t+1,m|t},$$

where $\bar{y}_{t+1,m}$ is the observed yield and $\hat{y}_{t+1,m|t}$ is the prediction of yields with maturity m at time $t+1$ done at time t. I compute forecast errors using both survey data and recovered

[9] The unpredictability of the forecast error which is implied by the rational expectation hypothesis is in principle unaltered by the moving forecast horizon. However, to ameliorate concerns regarding changes in expectations due to changes of the prediction horizon, I consider time fixed effects in the robustness section.

beliefs. In the former case, each $\hat{y}_{t+1,m|t}$ has multiple observations across different forecasters and the forecast error is forecaster specific. In the latter case, the forecast is computed as:

$$\hat{y}_{t+1,m|t} := \sum_{\mathbf{X}_{t+1} \in \text{Grid}} \underbrace{\mathbb{P}^{\theta}(\mathbf{X}_{t+1}|\mathbf{X}_t)}_{\text{Recovered beliefs}} \times \underbrace{(\hat{a}_m^{\theta} + (\hat{b}_m^{\mathbb{Q}^{\theta}})^{\top}\mathbf{X}_{t+1})}_{\text{Emprical affine mapping}}.$$

Expectations about future yields are conveniently decomposed into distorted expectations about factors (identified using the recovery theorem) and the pricing function (empirical affine mappings). The affine mapping is known because $\hat{a}_m^{\theta}$ and $\hat{b}_m^{\mathbb{Q}^{\theta}}$ have been estimated as discussed in Section 3. To simplify notation, $\hat{y}_{t+1,m|t}$ does not carry superscript θ.
Under the null hypothesis of full information rational expectations, the forecast error should not be predictable on the basis of past information. But what is past information? Following Coibion and Gorodnichenko (2015), I define information at time t by the *forecast revision*:

$$FR_t[y_{t+1,m}] := \hat{y}_{t+1,m|t} - \hat{y}_{t+1,m|t-1}.$$

The logic underlying this definition is that *information moves beliefs*: if no information is observed at time t, then there is no revision in beliefs, i.e. $FR_t[y_{t+1,m}] = 0$.

For each maturity available, I then run the regression:

$$FE_{t+1}[y_{t+1,m}] = \alpha_m + \beta_m FR_t[y_{t+1,m}] + \epsilon_{t+1,m}.$$

I call β_m "CG coefficient" from Coibion and Gorodnichenko (2015). A positive $\beta_m > 0$ is interpreted as *under-reaction* to information at maturity m. In this case, when beliefs about yields at horizon m are revised upward, they systematically under-estimate realized yields. That is, beliefs are not revised enough. On the contrary, $\beta_m < 0$ is interpreted as *over-reaction* to information at maturity m: when beliefs about yields at horizon m are revised upward, they systematically over-estimate realized yields. That is, beliefs are revised too much. The CG coefficients obtained with the Ross recovered beliefs for maturities ranging of $2, 3, \ldots 30$ years[10] and by pooled estimation of professional forecasters data are shown in

[10] The $1y$ yield is used as short rate and therefore predictions to this horizon cannot be computed.

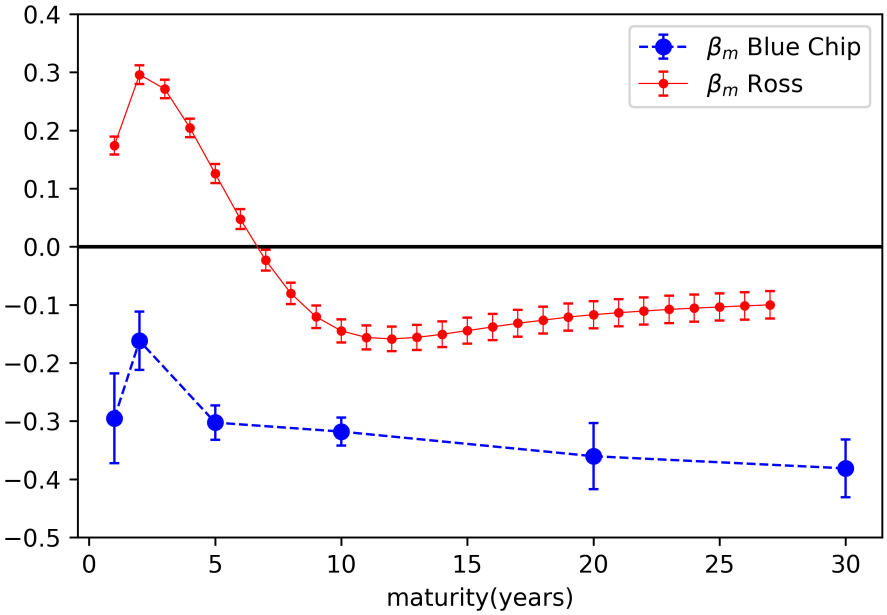


Figure 3: Slope of FE on FR using recovered beliefs (red) and using the Blue Chip dataset (pooled OLS). Confidence intervals are computed at 5%.

Figure 3.

Recovered beliefs also exhibit maturity dependent reaction to information. Just like the beliefs of professional forecasters, they over-react more for long term yields than for short term ones. Thus, the key message of excess reaction to information of long run beliefs relative to short run beliefs is consistent across survey data and Ross recovered beliefs. The main difference between the two datasets arises because Ross recovered beliefs exhibit a pattern of under-reaction to information at the short end of the yield curve and over-reaction to information at the long end. Professional forecaster data, on the contrary, exhibit over-reaction only [11] [12].

Ross recovered beliefs capture market beliefs, and the marginal investor is potentially

[11] In the robustness Section, I consider different specifications of the regression performed with professional forecasters data, including time fixed effects, forecasters fixed effects as well as single and double clustering of standard errors. The results are consistent.

[12] In Bordalo et al. (2018a), the authors find under-reaction at maturities shorter than one year with Blue Chip data, at the quarterly frequency. Here, I do not consider those maturities for the sake of comparison with recovered beliefs.

different from the average forecaster. For this reason the qualitative similarity of predictability patterns across the two datasets is surprising: there is no mechanical reason for it, and this suggests that stronger over-reaction for longer horizons may be a robust feature of beliefs. We can even more directly compare Ross recovered beliefs and professional forecasts by correlating forecast revisions and the level of forecasts across the two datasets (considering the mean as well as the median forecast). Figure (4) shows that the two datasets are quite aligned along both criteria[13].

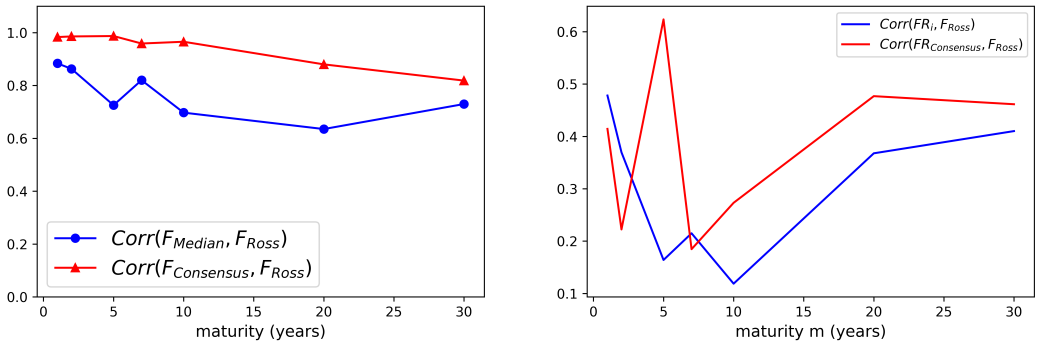


Figure 4: Left panel: correlation between mean forecasts (red triangle), median forecasts (blue circle) and Ross recovered forecasts as a function of the maturity. Right panel: correlation between mean forecast revision (red triangle), median forecast revision (blue circle) and Ross recovered forecast revisions as a function of the maturity.

The strong, positive correlation between Ross recovered beliefs and professional forecasts is important. It indicates that the two types of beliefs data are not noise, and thus they capture significant features of market beliefs. Of course, the benefit of Ross recovered beliefs is that, in addition to being more tightly linked to the marginal investor, they are available for all maturities and frequencies.

Are market beliefs, as measured with the Ross method, better or worse than the beliefs of analysts? There are two possible metrics to asses this: the predictability of the forecast error (which captures biases in forecasting) and the mean square error (which capture *accuracy*

[13] For the comparison, I aggregated recovered beliefs from the daily to the monthly frequency.

in forecasting, i.e. bias plus precision). In Figure (5), I plot the estimated distribution of distortions (biases) using a Gaussian kernel density estimator.

The distribution of forecaster distortions is not symmetric around rationality (i.e. $\beta = 0$): the majority of forecasters over-react to news. This is reflected in the fact that the average bias, captured by the pooled regression in the introduction and reported also in Figure (5) is negative. The above figure also reports as a benchmark the CG coefficient of the consensus forecast, defined as the predictability of errors for the median analyst forecast. Note that the consensus always under-reacts to news, a fact that has been previously documented in Bordalo et al. (2018a), where the authors also reconcile it with it with individual analyst over-reaction[14].

The distortion of the Ross recovered forecast lies in between the median forecaster and the average bias, and it is closer to rationality ($\beta = 0$) than both. Thus, regarding the bias dimension, this analysis suggests that the Ross recovered beliefs weigh more unbiased forecasts, relative to the average professional forecaster bias. This may be due to arbitrage capital moving partly, though not fully, towards less biased investors.

Second, I investigate the accuracy of forecasters, as measured by the mean square error in prediction. In Figure (5), I consider two groups of forecasters. The top 25% most accurate forecasters and the bottom 25% least accurate forecasters. In the data, the former forecasters are rational, namely, their CG coefficients are statistically indistinguishable for zero, while the worst 25% are highly over-reacting. Thus, there is a positive relation between imprecision and bias in forecasting. A direct comparison between the correlations of Ross recovered forecasts and top/worse 25% (ranked according to the MSE criterion) is offered in Figure (6).

Figure (6) suggests that the Ross recovered forecast slightly over-weights more accurate views

Overall this analysis suggests that Ross forecasts weight more both more unbiased forecasters and more accurate forecasters.

[14]The intuition is that when forecasters observe different noisy signals, stemming for instance from heterogeneous information sets, each analysts over-reacts to his own news, but does not react at all to the signal of other analysts. This second effect can be so strong that the consensus forecast under-reacts to the consensus revision even if each analyst over-reacts to his own information.

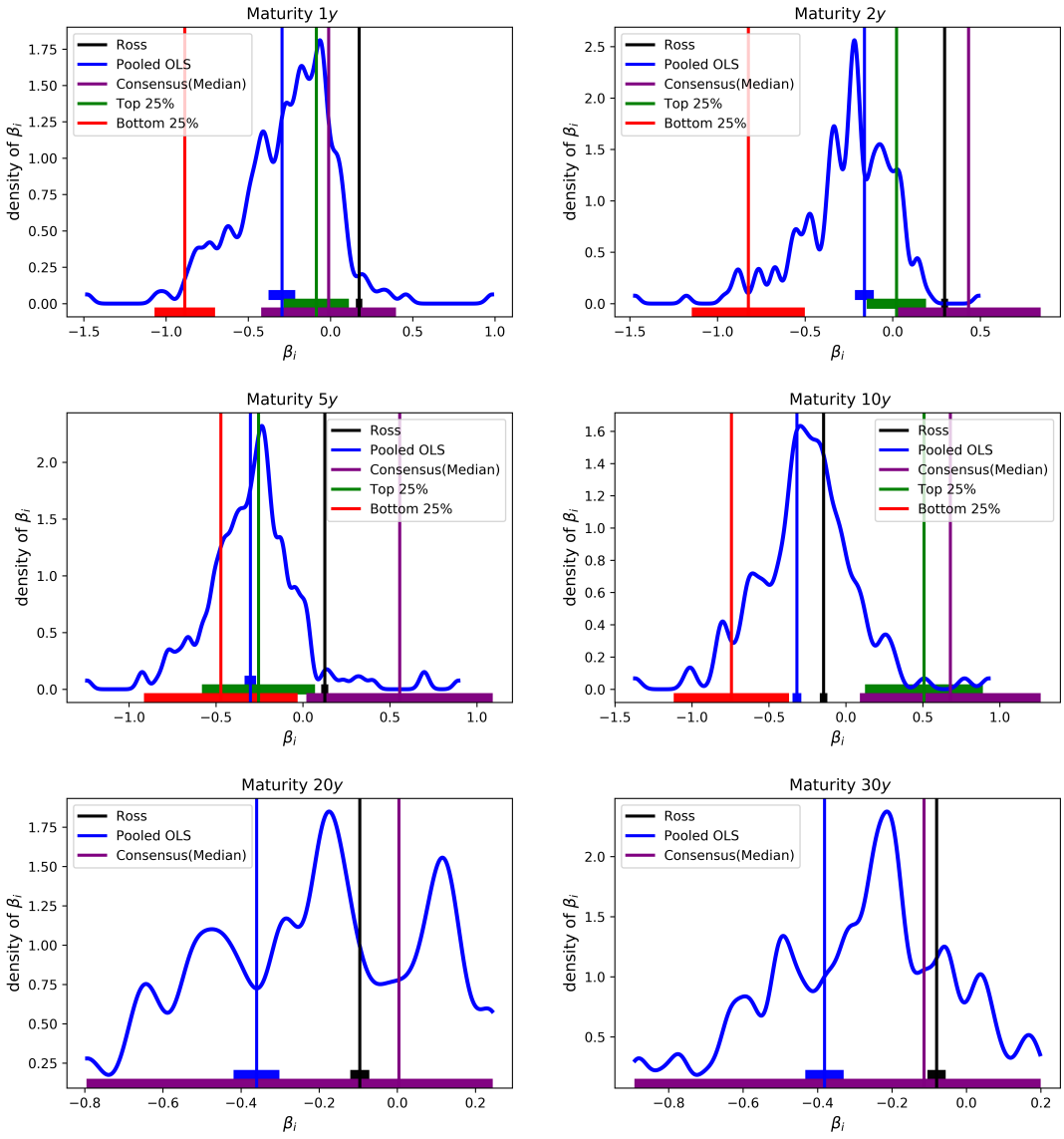


Figure 5: Density of forecasters distortions (CG coefficients) for different maturities. Distortions from recovered beliefs are shown in black, pooled distortions in blue, consensus (median) distortions in purple, best forecasters distortions (top quartile according to the mean square error criterion) in green, worse forecasters distortions (bottom quartile according to the mean square error criterion) in red. Confidence intervals are computed at the 5% level.

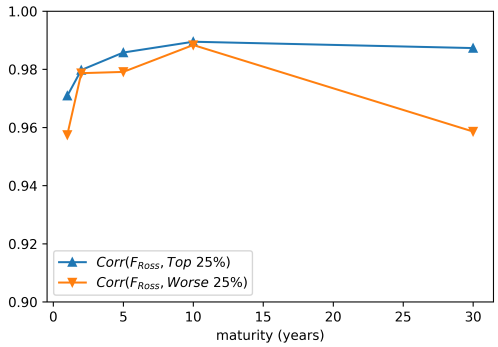


Figure 6: Correlation between mean forecast of top 25% forecasters (ranked according to the MSE criterion) and Ross recovered forecasts (blue triangle), Correlation between mean forecast of worse 25% forecasters (ranked according to the MSE criterion) and Ross recovered forecasts (orange triangle).

Broadly speaking, this Section conveys the following messages. First, market beliefs recovered using Ross method display the same maturity increasing over-reaction displayed by survey data. Second, market beliefs and survey data are highly positively correlated. Third, market beliefs are less biased and more accurate than the average professional forecaster. This indicates that our recovered market beliefs capture systematic patterns in beliefs, and that arbitrage helps reduce the impact of highly biased forecasters on asset prices, consistent with basic asset pricing theory.

Having validated the recovered beliefs, several questions emerge. First, why do forecasts over-react in this maturity dependent way? What is the psychological foundation? Second, can the over-reaction detected from beliefs account for the excess volatility of interest rates along the lines of Theorem (1)? To make progress, I specify a realistic model of belief formation, based on the diagnostic expectation model of Bordalo et al. (2018b), and I show that it offers an answer to both questions: it yields maturity dependent over-reaction, it can *quantitatively* account for belief distortions *and* this in turn captures a big chunck of the excess volatility of interest rates documented by Giglio and Kelly (2017).

5 The Horizon Dependent Diagnostic Expectations Model

Using the diagnostic expectations model of Bordalo et al. (2018b), I now rationalize maturity-increasing over-reaction using a single parsimonious departure of beliefs from rationality, captured by the diagnosticity parameter θ that I estimate and use to quantify the explanatory power of the model for excess volatility of interest rates. As argued in Section 2, the duality between belief distortions and over-reaction in interest rates offers two separate methods for estimating θ, one based on beliefs data, another based on yields. I show that these two independent strategies yield similar estimates for θ that are consistent with previous estimates, and that can quantitatively account for excess volatility of interest rates and for the predictability of the forecast errors.

5.1 Diagnostic Expectations

Diagnostic expectations are based on Kahneman and Tversky's (Tversky and Kahneman (1974)) representativeness heuristics in probability judgments. Representativeness captures the idea that, when making a conditional probabilistic assessment, humans typically over-weight representative (or *diagnostic*) traits, defined as the traits that are objectively more frequent in such group relative to a comparison group. A conventional example is the exaggeration of the probability that an Irish person is red haired, because this hair color is relatively more frequent in Ireland than elsewhere (although even in Ireland it is unlikely in absolute terms). This heuristics has been widely documented in the psychology and cognitive science literature, since the seminal work of Tversky and Kahneman (1974) and it has recently been adapted to a dynamic setting by Bordalo et al. (2018b). When forecasting, economic agents exaggerate objectively positive news relative to a benchmark prediction, shaped by past information.

To capture this idea in my setting, consider the following *diagnostic* distribution at time t of interest rates with maturity m, $r_{t+m} = \frac{1}{m}\sum_{i=1}^{m} r_{t+i}$:

$$f_{\mathbb{P}^\theta}(r_{t,m}|\mathbf{X}_t) \propto f_{\mathbb{P}}(r_{t,m}|\mathbf{X}_t)\left(\frac{f_{\mathbb{P}}(r_{t,m}|\mathbf{X}_t)}{f_{\mathbb{P}}(r_{t,m})}\right)^{\theta},$$

where $f_{\mathbb{P}}(\cdot)$ is the unconditional or long run distribution of $r_{t,m}$.

The diagnostic distribution of $r_{t,m}$ at time t re-weights the correct density $f_{\mathbb{P}}(\cdot|\mathbf{X}_t)$ via the likelihood ratio $\left(\frac{f_{\mathbb{P}}(r_{t+m}|\mathbf{X}_t)}{f_{\mathbb{P}}(r_{t,m})}\right)$ to the power θ. Investors over-weight future values of interest rates that are more likely under current information relative to the average information, where the latter is captured by the long run distribution. The parameter $\theta > 0$ in the diagnostic distribution quantifies the degree of over-reaction. The larger is θ the more outcomes whose likelihood has increased are overweighted in beliefs.

My model differs in two dimensions to the model of Bordalo et al. (2018b). First, in Bordalo et al. (2018b), the authors compare rational forecasts at time t with rational forecast at time $t-1$, while I use the average information as comparison. This is technically convenient because my model preserves Markovianity, which greatly simplifies the identification of recovered beliefs in Section 3. In Section 6, I show, theoretically, that results are qualitatively robust to the specification used in Bordalo et al. (2018b). Second, and more important, the benchmark distribution in Bordalo et al. (2018b) is assumed to have the same volatility as the rational forecast. This assumption is innocuous when considering a fixed horizon as in Bordalo et al. (2018b) and it greatly simplifies the math. However, this assumption misses an important intuition: namely that diagnostic distortions should depend on the underlying uncertainty about the economic environment, as discussed in **?**. I now show that allowing for this role of uncertainty is highly relevant here, because it implies the violation of the law of iterated expectations taking the form of maturity increasing over-reaction.

The diagnostic distribution can be conveniently computed for linear and Gaussian dynamics. Assume that the $\mathbb{P}$ dynamics of the factor is:

$$\begin{cases} r_t = \delta_0 + \delta_1^\top \mathbf{X} \\ \mathbf{X}_t = \rho^{\mathbb{P}} \mathbf{X}_{t-1} + \Sigma^C \epsilon_t^{\mathbb{P}}, \end{cases}$$

where $\mathbf{X}_t \overset{\mathbb{P}}{\sim} VAR(1)(\rho^{\mathbb{P}}, \Sigma)$, Σ^C is a lower triangular matrix, $\epsilon_t^{\mathbb{P}}$ are i.i.d. Gaussian shocks and $\Sigma := \Sigma^C {\Sigma^C}^\top$ is the one period variance-covariance matrix. Here, in order to derive a parsimonious expression for increasing over-reaction, I impose additional assumptions relative

to $\mathbb{Q}$-affine setting (the latter is the only assumption I needed so far). Specifically, I assume a VAR(1) Gaussian dynamics for the factor, under the physical measure $\mathbb{P}$. Then, the diagnostic distribution of interest rates at maturity m is characterized as follows.

Theorem 3. *(Diagnostic distribution) Given* $\mathbf{X}_t \overset{\mathbb{P}}{\sim} VAR(1)(\rho^{\mathbb{P}}, \Sigma)$*, under Gaussian noise, the diagnostic distribution of interest rates to maturity* m*,* $r_{t,m}$*,* $f^{\theta}_{\mathbb{P}}(r_{t+m}|\mathbf{X}_t)$*, is Gaussian, with mean:*

$$\mathbb{E}^{\mathbb{P}^\theta}_t[r_{t,m}] = \mathbb{E}^{\mathbb{P}}_t[r_{t,m}] + \psi^{\theta}_m \left(\mathbb{E}^{\mathbb{P}}_t[r_{t,m}] - \mathbb{E}^{\mathbb{P}}[r_{t,m}]\right) \tag{6}$$

and variance:

$$\mathbb{V}^{\mathbb{P}^\theta}_t[r_{t,m}] = \left(\frac{\theta+1}{\mathbb{V}^{\mathbb{P}}_t[r_{t,m}]} - \frac{\theta}{\mathbb{V}^{\mathbb{P}}[r_{t,m}]}\right)^{-1}, \tag{7}$$

where

$$\psi^{\theta}_m := \frac{\theta \frac{\mathbb{V}^{\mathbb{P}}_t[r_{t,m}]}{\mathbb{V}^{\mathbb{P}}[r_{t,m}]}}{1+\theta-\theta\frac{\mathbb{V}^{\mathbb{P}}_t[r_{t,m}]}{\mathbb{V}^{\mathbb{P}}[r_{t,m}]}}, \tag{8}$$

$$\mathbb{E}^{\mathbb{P}}_t[r_{t+m}] = \frac{1}{m}\left(\delta_0 + \delta_1^{\top}\left(\sum_{i=0}^{m-1}\rho^{\mathbb{P}^i}\right)\mathbf{X}_t\right), \tag{9}$$

$$\mathbb{E}^{\mathbb{P}}[r_{t,m}] = \delta_0, \tag{10}$$

$$\mathbb{V}^{\mathbb{P}}_t[r_{t,m}] = \frac{1}{m^2}(\delta_1)^{\top}\left(\sum_{i=0}^{2}\left(\sum_{i=0}^{m-2}\rho^{\mathbb{P}^{2i}}\Sigma\right)\right)\delta_1 \tag{11}$$

and

$$\mathbb{V}^{\mathbb{P}}[r_{t,m}] = \mathbb{E}[\mathbb{V}^{\mathbb{P}}_t[r_{t,m}]] + \mathbb{V}[\mathbb{E}^{\mathbb{P}}_t[r_{t,m}]] \tag{12}$$

$$= \mathbb{V}^{\mathbb{P}}_t[r_{t,m}] + \frac{1}{m^2}(\delta_1)^{\top}\left(\sum_{i=0}^{m-1}\rho^{\mathbb{P}^i}\right)\Sigma\left(\sum_{i=0}^{m-1}\rho^{\mathbb{P}^i}\right)\delta_1 \tag{13}$$

As evident from Equation (6), the diagnostic model endogeneizes the maturity increasing over-reaction assumed in reduced from in Section 2. The coefficient ψ^{θ}_m in formula (8), that characterizes the departures from rationality in Section 2, it now pinned down by: i) the scalar parameter θ, ii) the maturity m, and iii) the true conditional and unconditional variances

$\mathbb{V}_t^{\mathbb{P}}[r_{t,m}]$ and $\mathbb{V}^{\mathbb{P}}[r_{t,m}]$. Belief distortions starts from zero at $m = 0$, then become positive at $m = 1$ and increase monotonically as $m \to \infty$, approaching a finite limiting value equal to θ.

The intuition for this result is simple: as the maturity m increases, fundamental uncertainty is higher. As a result, the tails of the distribution are prominent, so they more easily come to mind after information makes them more likely. That is, after good news, the right tail becomes very representative for long maturities and it is highly over-weighted. As a result, expectations are too optimistic. After bad news, the left tail becomes very representative for long maturities and is highly overweighted. As a result, expectations become too pessimistic. In both cases, beliefs over-react and they do so more for longer maturities.

Theorem (2) also shows that in reduced form, for a given value of ψ_m^θ, the diagnostic model yields the same expectations for interest rates assumed in Section 2 and it therefore yields the same rule for equilibrium yields as in Theorem (1):

$$y_{t,m}^\theta = a_m^\theta + (b_m^{\mathbb{Q}})^\top \mathbf{X}_t + \psi_m^\theta (b_m^{\mathbb{P}})^\top \mathbf{X}_t,$$

where however we consider the realistic multi-factor setting and the coefficients $b_m^{\mathbb{P}}$ and $b_m^{\mathbb{Q}}$ are now vectors, with different entries for different factors. Of course, the distortion ψ_m^θ now depends on the data generating process, on maturity m, and on diagnosticity θ. This aspect places restrictions in the quantitative analysis.

5.2 Calibration

Theorem (1) suggests two independent routes to estimate the same primitive parameter θ. First, I retrieve θ using the profile of CG coefficients β_m, obtained with recovered beliefs. I match β_m with their theoretical counterparts, as derived in Theorem (1). I can then assess the fitting ability of such estimated θ on excess volatility. Second, I follow the mirror route: I estimate θ by matching excess volatility, and then assess its ability in fitting the profile of CG coefficients.

For the first exercise, I estimate θ that best matches the profile of CG coefficients. Note that the fit here cannot be perfect: in the short end of the curve Ross recovered beliefs

under-react, a pattern which cannot be rationalized under diagnostic expectations, namely $\theta > 0$. The estimated value obtained when using this method fulfills:

$$\hat{\theta} = \arg\min_{\theta} \sum_{i=2}^{30} \left(\hat{\beta}_m - \beta_m^{\theta} \right)^2 \approx 0.7.$$

Second, I consider excess volatility. The diagnostic parameter θ disciplines excess volatility: the excess sensitivity of yields to factors is controlled by ψ_m^{θ}, which is a function of the rational conditional variance of interest rates with maturity m and the rational unconditional variance of interest rates with the same maturity. The variance of yields in a diagnostic world relates to the variance of yields in a rational world as:

$$\underbrace{\mathbb{V}^{\mathbb{P}}[y_{m,t}^{\theta}]}_{\text{data}} = \underbrace{\mathbb{V}^{\mathbb{P}}[y_{t,m}]}_{\text{RE model}} + \psi_m^{\theta} (b_m^{\mathbb{P}})^{\top} \Sigma b_m^{\mathbb{P}},$$

From this observation, I can calibrate the diagnostic parameter θ from the excess volatility of yields itself. Indeed, the variance of rational yields of the right hand side can be estimated by fitting the short end of the yield curve as in Giglio and Kelly (2017), while $b_m^{\mathbb{P}}$ is can be computed using the estimated persistence $\rho^{\mathbb{P}}$ obtained from a VAR(1) estimation of the $\mathbb{P}$-factor dynamics[15]. Therefore, I estimate θ as:

$$\hat{\theta} = \arg\min_{\theta} \sum_{i=1}^{30} \left(\mathbb{V}^{\mathbb{P}}[y_{m,t}^{\theta}] - \mathbb{V}^{\mathbb{P}}[y_{t,m}^{\theta=0}] - \psi_m^{\theta} (b_m^{\mathbb{P}})^{\top} \Sigma b_m^{\mathbb{P}}) \right)^2 \approx 0.47.$$

The two estimates differ, but remarkably they are close to previous estimates of parameter θ obtained using different data and different methodologies. For instance, Bordalo et al. (2018a) use survey forecasts of professional forecasters for many macro-financial variables and shows that typically there is over-reaction with magnitude $\theta \sim 0.6$, and equal to $\theta \sim 0.48$ for medium to long term interest rates. This is an additional confirmation that in my setting Ross recovery captures robust patterns of beliefs. To grasp the quantitative meaning of the estimated values of θ, consider the benchmark $\theta = 1$. In this case, distorted forecasts of

[15] I used the methodology of Section 3 and imposed consistency across short maturities, namely: $\hat{\rho}^{\mathbb{Q}} := \arg\min_{\rho^{\mathbb{Q}}} \left(\frac{1}{T\bar{M}} \sum_{t,m} \left(y_{t,m} - \bar{y}_{t,m} \right)^2 \right)$. $\bar{M}$ is set as equal to 5: it quantifies how short the short end of the yield curve is. Robustnesses relative to the parameters to be included in the short end show consistency of the results.

long maturity rates are equal to the rational forecast plus the revision. Assuming that the baseline rational forecast for the long run rate is around 2%, after the arrival of news indicative of a higher rational forecast of 3%, the diagnostic forecast will be then 4%. Distortions are thus sizable. This back-of-the-envelope calculation shows that the numbers at play are economically relevant.

Having estimated values for the distortion parameter θ, we can now evaluate the accuracy of the diagnostic model for excess volatility in interest rates. Figure (7) reports the volatilities of yields at different maturities obtained under a three factor affine rational model, namely a counter-factual model setting $\theta = 0$, together with the variance of yields obtained from the same model in which we plug the estimated $\theta \approx 0.47$ and $\theta \approx 0.7$.

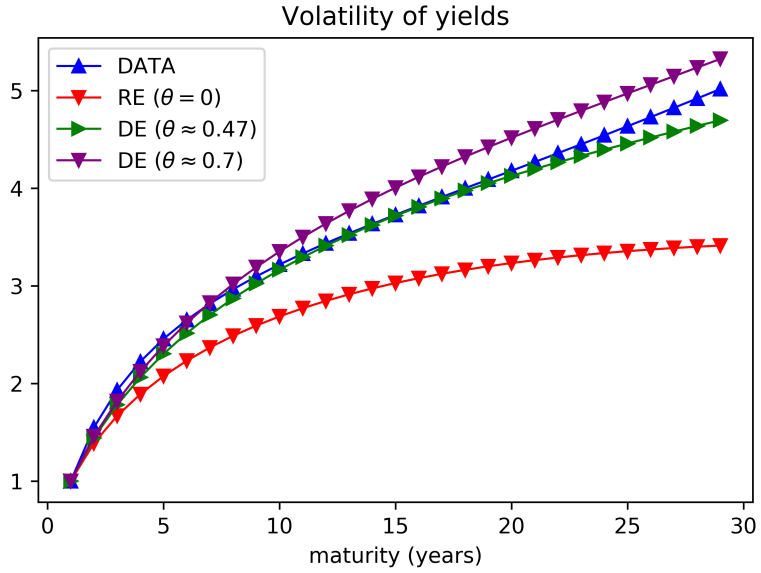


Figure 7: Excess volatility in the data (blue up triangle) versus the fit of a rational expectations affine model (red down triangle) and the diagnostic expectations affine model with $\theta \approx 0.47$ (green right triangle) and $\theta \approx 0.7$ (purple left triangle).

Figure (7) shows that diagnostic expectations capture much of the variation of the yield curve. At the longest maturity available, diagnostic expectations fit roughly $\sqrt{\frac{\mathbb{V}^{\mathbb{P}}[y^{\hat{\theta}}_{t,30}]}{\mathbb{V}^{\mathbb{P}}[y^{\theta}_{t,30}]}} \approx 82\%$ of excess volatility. Considering the distortion parameter estimated from the profile of CG coefficients, $\theta \approx 0.7$, provide an explanation of ≈ 55 % of the excess volatility. This latter

case shows that information implied from CG coefficients actually does help explaining excess volatility.

The symmetric question is, as suggested by Theorem (1), how much of the predictability of the forecast error can be explained from a distorting parameter θ which is inferred from excess volatility? The link between the forecast error predictability (CG coefficients) and the over-reaction is the one in Theorem (2), where the coefficients ψ_m^θ have now been derived from the diagnostic expectations model.

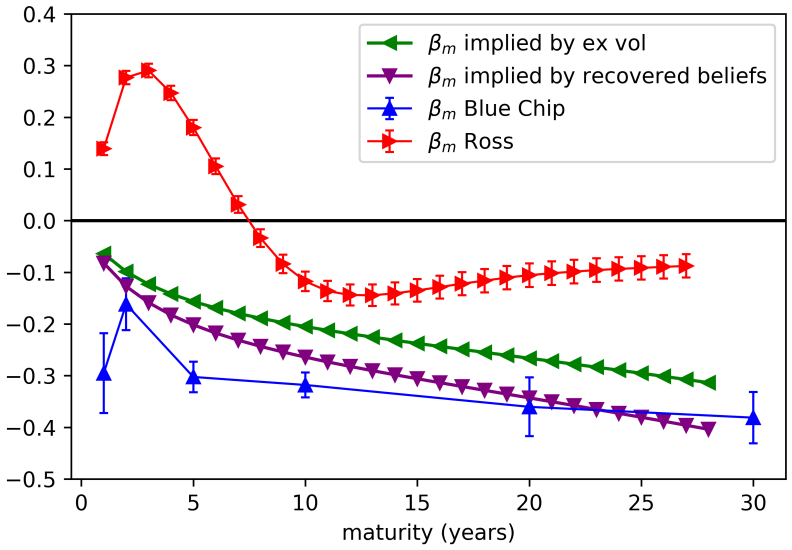


Figure 8: CG coefficients from Blue Chip data (blue up triangle), recovered beliefs (red right triangle) and implied by the calibrated diagnostic expectation model for $\hat{\theta} \approx 0.47$ (green left triangle) and $\hat{\theta} \approx 0.7$ (purple down triangle).

Figure (8) shows the term structure of CG coefficients, in the data (survey data in Blue, Ross recovered beliefs in red) as well as the β_m profiles implied by the calibrated parameters $\hat{\theta} \approx 0.47$ from excess volatility (green) and the calibrated distorted parameter $\hat{\theta} \approx 0.7$ from CG coefficients (purple). The fit is not expected to be perfect because under-reaction cannot be reconciled directly with the diagnostic expectation model. Indeed, within this model, the CG coefficients should be negative. The presence of under-reaction, particularly at short maturities, has been previously documented by different authors (Bouchaud et al. (2019), Bordalo et al. (2018a)). Other forces may be at play other than representativeness, such

as disagreement and informational frictions, which needs to be further investigate. When considering the whole term structure, however, there is a clear pattern of increasing *reaction* to news, and, in particular, or over-reaction at long maturities.

6 Robustness

In this Section, I discuss theoretical and empirical robustness to the recovery theorem. Then, I consider different specifications of the forecast error predictability tests. Finally, I consider alternative specifications of the diagnostic expectations model.

6.1 Empirical implementation of the recovery theorem

6.1.1 Discretization and sampling frequency

Figure (9) shows that the number of states chosen for the discretization for each factor ($N = 25, 50, 75, 100$) and the sampling frequency (daily versus monthly) do not qualitatively affect the CG coefficients curve.

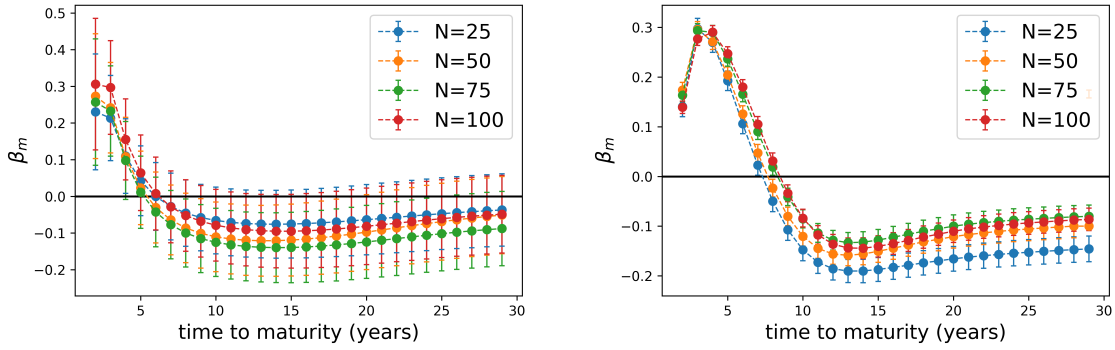
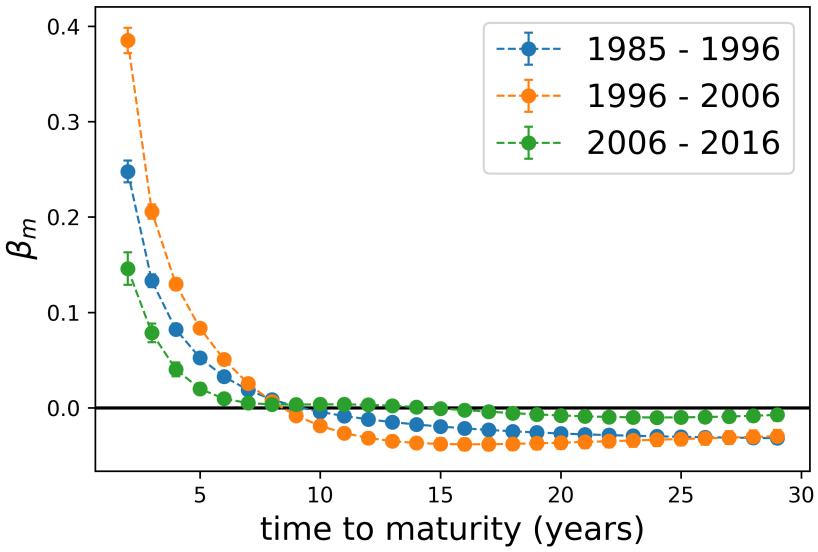


Figure 9: Left panel: CG coefficients at the monthly frequency. Right panel: CG coefficients at the daily frequency.

6.1.2 Error predictability in different sub-samples

The level of the nominal yield curve is close to unit root process, especially at high frequency (e.g. daily). Does possible non stationarity relates to under/over-reaction to news? Figure

(??) shows that the decreasing pattern in CG coefficients survives in different subsamples.



6.2 Predictability of the forecast error, different tests

News have been defined as the difference between the rational forecast and the average forecast, in the diagnostic expectation model used. Here I consider CG like tests, where the forecast revision is defined as the difference between the rational forecast and the long run forecast. The figure shows qualitative agreement, both using survey data and using recovered beliefs.

6.3 Diagnostic Expectations: different benchmark distributions

One important degree of freedom in the specification of the diagnostic expectation model is the choice of the *comparison* distribution. Here, I have chosen the unconditional or long run distribution of future rates. This is a convenient choice because the diagnostic distribution remains Markovian, namely $r_{t,m}$ depends on $\mathbf{X}_t$ but not on $\mathbf{X}_{t-1}, \mathbf{X}_{t-2}, \dots$. It is however important to investigate what changes with different benchmark distributions.

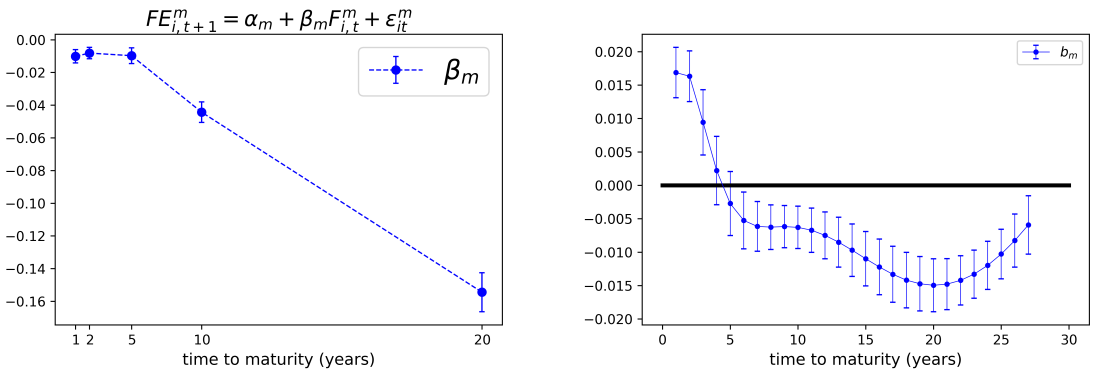


Figure 10: Left panel: Blue Chip Financial. Right panel: Ross recovered beliefs

Over-reaction relative to time $t-1$ prediction

Consider the following specification, inspired by Bordalo et al. (2018b)[16]:

$$f_{\mathbb{P}^\theta}(r_{t,m}|\mathbf{X}_t) \propto f_{\mathbb{P}}(r_{t,m}|\mathbf{X}_t)\left(\frac{f_{\mathbb{P}}(r_{t,m}|\mathbf{X}_t)}{f_{\mathbb{P}}(r_{t,m}|\mathbf{X}_{t-1})}\right)^\theta.$$

The diagnostic distribution of $r_{t,m}$ in this case is still Gaussian, with mean:

$$\mathbb{E}_t^{\mathbb{P}^\theta}[r_{t,m}] = \mathbb{E}_t^{\mathbb{P}}[r_{t,m}] + \psi_m^\theta\left(\mathbb{E}_t^{\mathbb{P}}[r_{t,m}] - \mathbb{E}_{t-1}^{\mathbb{P}}[r_{t,m}]\right)$$

and variance:

$$\mathbb{V}_t^{\mathbb{P}^\theta}[r_{t,m}] = \left(\frac{\theta+1}{\mathbb{V}_t^{\mathbb{P}}[r_{t,m}]} - \frac{\theta}{\mathbb{V}_{t-1}^{\mathbb{P}}[r_{t,m}]}\right)^{-1},$$

[16]In Bordalo et al. (2018b), the authors consider the convenient benchmark distribution as $f_{\mathbb{P}}(r_{t,m}|\mathbf{X}_t := \mathbf{X}_{t-1})$. This simplifies the algebra of diagnostic expectations: only the first moment is distorted and the distortion is independent of the maturity. This assumption is innocuous from a single horizon perspective, which is the setting of Bordalo et al. (2018b), yet it forces the law of iterated expectations to hold, differently from slight perturbations of the benchmark distribution. Therefore, this simplifying assumption is not appropriate for my setting.

where

$$\psi_m^\theta := \frac{\theta \frac{\mathbb{V}_t^{\mathbb{P}}[r_{t,m}]}{\mathbb{V}_{t-1}^{\mathbb{P}}[r_{t,m}]}}{1+\theta-\theta \frac{\mathbb{V}_t{}^{\mathbb{P}}[r_{t,m}]}{\mathbb{V}_{t-1}^{\mathbb{P}}[r_{t,m}]}}.$$

Also in this case, the distortion coefficients ψ_m^θ starts at zero (namely $\lim_{m\to 0} \psi_m^\theta = 0$), asymptotically approaches θ (namely $\lim_{m\to\infty} \psi_m^\theta = \theta$) and they are increasing. Moving to risk neutral distorted expectations, the diagnostic yield curve reads:

$$y_{t,m}^\theta = a_m^{\theta} + (b_m^{\mathbb{Q}})^\top \mathbf{X}_t + \psi_m^\theta \left((b_m^{\mathbb{P}})^\top \mathbf{X}_t - (b_{m+1}^{\mathbb{P}})^\top \mathbf{X}_{t-1} \right)$$

In this case, the relation between the volatility of the yields in a diagnostic world, relative to the rational one is modified by forecast revision of interest rates. After good news yields are higher while after bad news are lower, relative to the RE case. Unconditionally, yields display higher variance in the diagnostic case, since the extra term $\psi_m^\theta (b_m^{\mathbb{P}})^\top \mathbf{X}_t - (b_{m+1}^{\mathbb{P}})^\top \mathbf{X}_{t-1}$ is positively correlated with $(b_m^{\mathbb{Q}})^\top \mathbf{X}_t$. This is so because $b_{m+1}^{\mathbb{P}} < b_m^{\mathbb{P}}$ and the $\mathbb{P}$ persistence is on average smaller than one (or in the on linear case it is assumed the be smaller than one on average). they display higher volatility.

Over-reaction relative to time $t-k$ prediction ($k>1$)

In this case the logic of the case $k=1$ still goes through with:

$$\psi_m^\theta := \frac{\theta \frac{\mathbb{V}_t^{\mathbb{P}}[r_{t,m}]}{\mathbb{V}_{t-k}^{\mathbb{P}}[r_{t,m}]}}{1+\theta-\theta \frac{\mathbb{V}_t{}^{\mathbb{P}}[r_{t,m}]}{\mathbb{V}_{t-k}^{\mathbb{P}}[r_{t,m}]}}.$$

and:

$$y_{t,m}^{\theta} = a_m^{\theta} + \psi_m^{\theta} \left((b_m^{\mathbb{Q}})^{\top} \mathbf{X}_{t+1} - (b^{\mathbb{P}})_{m+k}^{\top} \mathbf{X}_{t-k} \right).$$

Over-reaction relative to a weighted average of past predictions

Yesterday information and average information are two useful benchmark, yet one may consider a more "colorful" memory. I consider the following specification, inspired by Bordalo et al. (2018b), Internet Appendix):

$$f_{\mathbb{P}^{\theta}}(r_{t,m}|\mathbf{X}_t) \propto f_{\mathbb{P}}(r_{t,m}|\mathbf{X}_t) \prod_{k=1}^{M} \left(\frac{f_{\mathbb{P}}(r_{t,m}|\mathbf{X}_t)}{f_{\mathbb{P}}(r_{t,m}|\mathbf{X}_{t-k})} \right)^{\theta a_k}.$$

where $0 \leq a_k \leq 1$ are positive weights on past information such that $\sum_k a_k = 1$ and $1 \leq M \leq \infty$.[17] The diagnostic distribution of $r_{t,m}$ in this case is still Gaussian, with mean:

$$\mathbb{E}_t^{\mathbb{P}^{\theta}}[r_{t,m}] = \mathbb{E}_t^{\mathbb{P}}[r_{t,m}] + \psi_m^{\theta} \left(\mathbb{E}_t^{\mathbb{P}}[r_{t,m}] - \sum_{k=1}^{M} a_k \mathbb{E}_{t-k}^{\mathbb{P}}[r_{t,m}] \right)$$

and variance:

$$\mathbb{V}_t^{\mathbb{P}^{\theta}}[r_{t,m}] = \left(\frac{\theta + 1}{\mathbb{V}_t^{\mathbb{P}}[r_{t,m}]} - \theta \sum_{k=1}^{M} \frac{a_k}{\mathbb{V}_{t-k}^{\mathbb{P}}[r_{t,m}]} \right)^{-1},$$

where

$$\psi_m^{\theta} := \frac{\theta \mathbb{V}^{\mathbb{P}}[r_{t,m}] \sum_{k=1}^{M} \frac{a_k}{\mathbb{V}_{t-k}^{\mathbb{P}}[r_{t,m}]}}{1 + \theta - \theta \mathbb{V}^{\mathbb{P}}[r_{t,m}] \sum_{k=1}^{M} \frac{a_k}{\mathbb{V}_{t-k}^{\mathbb{P}}[r_{t,m}]}}.$$

[17] When $M = \infty$ suitable regularity conditions need to be assumed for convergence.

Also in this case, the distortion coefficients ψ_m^θ starts at zero (namely $\lim_{m\to 0}\psi_m^\theta = 0$), asymptotically approaches θ (namely $\lim_{m\to\infty}\psi_m^\theta = \theta$) and they are increasing. Moving to risk neutral distorted expectations,the diagnostic yield curve reads:

$$y_{t,m}^\theta = a_m^\theta + \psi_m^\theta\left((b_m^{\mathbb{Q}})^\top \mathbf{X}_{t+1} - \sum_{k=1}^{M} a_k (b^{\mathbb{P}})_{m+k}^\top \mathbf{X}_{t-k}\right).$$

The diagnostic yield curve is highly non Markovian in this case, yet it still feature the excess volatility pattern.

7 Conclusions

This paper shows empirically that beliefs distortions increases with maturity (decreasing CG coefficients) and that this is tightly linked to the excess volatility in the term structure of asset prices documented by Giglio and Kelly (2017). The crucial property that beliefs fail in the data and that accounts for excess volatility is the law of iterated expectations. I show that a diagnostic affine model, can quantitatively capture the variation in excess volatility and the distortions in the individual beliefs. In the model agents over-react differently for different levels of the fundamental uncertainty, which naturally varies in the context of term structure. This approach is a first step toward two research direction: the investigation of the effects of higher moments in beliefs distortions and the incorporations of non rational beliefs into quantitative finance models. At the aggregate level, I show the beliefs dynamics mixes under-reaction at short maturities and over-reaction at long-maturities. A foundation of such cross-over needs further investigations. Methodologically, this paper implements empirically the Ross Recovery theorem in the context of the term structure of interest rates, and show that, despite the identification problem raised by Borovička et al. (2016), recovered beliefs can spot inter-temporal belief inconsistencies. This is important to augment survey data with asset prices information and it is portable to different domains, such as the study of the equity term structure.

References

Augenblick, N. and Lazarus, E. (2017). Restrictions on asset-price movements under rational expectations: Theory and evidence. Technical report, Working Paper.

Bansal, R. and Yaron, A. (2004). Risks for the long run: A potential resolution of asset pricing puzzles. *The journal of Finance*, 59(4):1481–1509.

Bordalo, P., Gennaioli, N., Ma, Y., and Shleifer, A. (2018a). Overreaction in macroeconomic expectations. Technical report, Working Paper.

Bordalo, P., Gennaioli, N., Porta, R. L., and Shleifer, A. (2017). Diagnostic expectations and stock returns. Technical report, National Bureau of Economic Research.

Bordalo, P., Gennaioli, N., and Shleifer, A. (2018b). Diagnostic expectations and credit cycles. *The Journal of Finance*, 73(1):199–227.

Borovička, J., Hansen, L. P., and Scheinkman, J. A. (2016). Misspecified recovery. *The Journal of Finance*, 71(6):2493–2544.

Bouchaud, J.-p., Krueger, P., Landier, A., and Thesmar, D. (2019). Sticky expectations and the profitability anomaly. *The Journal of Finance*, 74(2):639–674.

Brooks, J., Katz, M., and Lustig, H. (2018). Post-fomc announcement drift in us bond markets. Technical report, National Bureau of Economic Research.

Buraschi, A., Piatti, I., and Whelan, P. (2018). Rationality and subjective bond risk premia.

Campbell, J. Y. (2003). Consumption-based asset pricing. *Handbook of the Economics of Finance*, 1:803–887.

Campbell, J. Y. and Shiller, R. J. (1987). Cointegration and tests of present value models. *Journal of political economy*, 95(5):1062–1088.

Cieslak, A. (2018). Short-rate expectations and unexpected returns in treasury bonds. *The Review of Financial Studies*, 31(9):3265–3306.

Cochrane, J. H. (2011). Presidential address: Discount rates. *The Journal of finance*, 66(4):1047–1108.

Cochrane, J. H. and Piazzesi, M. (2009). Decomposing the yield curve. In *AFA 2010 Atlanta Meetings Paper*.

Coibion, O. and Gorodnichenko, Y. (2015). Information rigidity and the expectations formation process: A simple framework and new facts. *American Economic Review*, 105(8):2644–78.

Cooley, T. F. (1995). *Frontiers of business cycle research*. Princeton University Press.

De Bondt, W. F. and Thaler, R. (1985). Does the stock market overreact? *The Journal of finance*, 40(3):793–805.

Duffee, G. (2013). Forecasting interest rates. In *Handbook of economic forecasting*, volume 2, pages 385–426. Elsevier.

Gennaioli, N. and Shleifer, A. (2018). *A Crisis of Beliefs: Investor Psychology and Financial Fragility*. Princeton University Press.

Giglio, S. and Kelly, B. (2017). Excess volatility: Beyond discount rates. *The Quarterly Journal of Economics*, 133(1):71–127.

Gürkaynak, R. S., Sack, B., and Wright, J. H. (2007). The us treasury yield curve: 1961 to the present. *Journal of monetary Economics*, 54(8):2291–2304.

Hamilton, J. D. and Wu, J. C. (2012). Identification and estimation of gaussian affine term structure models. *Journal of Econometrics*, 168(2):315–331.

Le, A., Singleton, K. J., and Dai, Q. (2010). Discrete-time affine-q term structure models with generalized market prices of risk. *The Review of Financial Studies*, 23(5):2184–2227.

LeRoy, S. F. and Porter, R. D. (1981). The present-value relation: Tests based on implied variance bounds. *Econometrica: Journal of the Econometric Society*, pages 555–574.

Martin, I. W. and Ross, S. A. (2019). Notes on the yield curve. *Journal of Financial Economics*.

Meyer, C. D. (2000). *Matrix analysis and applied linear algebra*, volume 71. Siam.

Piazzesi, M. and Schneider, M. (2011). Trend and cycle in bond premia. *Manuscript, Stanford Univ., http://www. stanford. edu/ piazzesi/trend cycle. pdf.*

Qin, L., Linetsky, V., and Nie, Y. (2018). Long forward probabilities, recovery, and the term structure of bond risk premiums. *The Review of Financial Studies*, 31(12):4863–4883.

Ross, S. (2015). The recovery theorem. *The Journal of Finance*, 70(2):615–648.

Shiller, R. J. (1981). The use of volatility measures in assessing market efficiency. *The Journal of Finance*, 36(2):291–304.

Tversky, A. and Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *science*, 185(4157):1124–1131.

Walden, J. (2017). Recovery with unbounded diffusion processes. *Review of Finance*, 21(4):1403–1444.

A Affine term structure models with overreacting beliefs

.

Consider first the standard $\mathbb{Q}$-affine term structure models, which are defined by the two following ingredients. First, few factors (or state variables) $\mathbf{X}_t = (X_{1,t}, X_{2,t}, \cdots)$ drive the short rate in an affine fashion and, second, the $\mathbb{Q}$-dynamics (assuming no arbitrage) of the factors is a VAR(1) with homoskedastic shocks.

$$\begin{cases} r_t = \delta_0 + \delta_1^\top \mathbf{X}_t \\ \mathbf{X}_t = \rho^{\mathbb{Q}} \mathbf{X}_{t-1} + \Sigma^{\mathbb{Q}} \varepsilon_t^{\mathbb{Q}}. \end{cases}$$

This defines the class of $\mathbb{Q}$-affine models I consider, together with sufficient regularity conditions for the quantities computed to be well defined. The convenient affine specification and linear dynamics implies that prices are exponentially affine in the factors:

$$\begin{aligned} y_{t,m} &= -\frac{1}{m} \log P_{t,m} = -\frac{1}{m} \mathbb{E}_t^{\mathbb{Q}} \left[e^{m \cdot r_{t,m}} \right] \\ &= \delta_0 - \log \mathbb{E}_t^{\mathbb{Q}} \left[e^{-\varepsilon_{t,m}^{\mathbb{Q}}} \right] + \frac{\delta_1^\top}{m} \sum_{i=0}^{m-1} (\rho^{\mathbb{Q}})^i \mathbf{X}_t, \end{aligned}$$

where $b_m^{\mathbb{Q}} = \frac{\delta_1^\top}{m} \sum_{i=0}^{m-1} (\rho^{\mathbb{Q}})^i$ and $\varepsilon_{t,m}^{\mathbb{Q}} = \sum_{k=1}^{m-1} \sum_{i=1}^{l} (\rho^{\mathbb{Q}})^{l-1} \Sigma^{\mathbb{Q}} \varepsilon_{t+i}$. The previous expression is affine since $\mathbb{Q}$ shocks are independent of time t information and therefore the cumulant generating function in the previous expression in maturity dependent by not state dependent. This setting is quite general since I do not have assumptions about the physical measure nor about the SDF and is spirit close to the one of Le et al. (2010). There, however, the authors impose some structure of $\mathbb{P}$ and of the SDF such that one can have analytical expressions in the $\mathbb{P}$ domain, needed for estimation of $\mathbb{P}$-moments.

First, I discuss how this setting relates to the empirical analysis of Section 3 and Section 4 and then how this setting relates to the theoretical model of beliefs formation developed in Section 2. I do so by incrementally adding structure and assumptions needed, relative to the $\mathbb{Q}$-affine benchmark so far discussed.

In section 3, I apply the recovery theorem independently estimating the $\mathbb{Q}$ measure at different maturities. This amounts to detect distortions in the sensitivity coefficients $b_m^{\mathbb{Q}} = \frac{\delta_1^\top}{m}\sum_{i=0}^{m-1}(\rho^{\mathbb{Q}})^i$. Theorem (1) shows that overreacting beliefs can both generate such distortions, which, in turn, account for the Giglio and Kelly (2017) excess volatility puzzle and explain the predictability of forecast error documented with survey data.

How do beliefs generate such distortion? Assume that, under the physical measure $\mathbb{P}$, factors evolves in a Markovian fashion:

$$\mathbf{X}_{t+m} = f^{(m)}(\mathbf{X}_t)\mathbf{X}_t + \Sigma^{\mathbb{P}}\varepsilon^{\mathbb{P}}_{t+m,C},$$

where $f^{(m)}(\mathbf{X}_{t-1}) = \underbrace{f(f(\cdots}_{\text{m times}}(\mathbf{X}_{t-1})$ denotes a Markovian, yet possibly non linear dynamics for the factors and $\varepsilon^{\mathbb{P}}_{t+m,C} = \sum_{i=0}^{m-2} f^{(i+1)}(\mathbf{X}_t)\Sigma^{\mathbb{P}}\varepsilon^{\mathbb{P}}_{t+1+i}$ denotes the cumulated shock to maturity m, which is heteroskedastic if the $\mathbb{P}$-dynamics is non linear. Then, assume that, when considering horizon m, the market distorts the dynamics of the factors in a maturity m dependent fashion. Formally, $\forall\ l \leq m$:

$$\mathbf{X}_{t+l} = (1+\psi_m^\theta)f^{(l)}(\mathbf{X}_t)\mathbf{X}_t + \sigma_m^\theta\varepsilon^{\mathbb{P}}_{t+l,C}.$$

The local persistences of all fundamentals up to time $t+m$ are inflated if $\psi_m^\theta > 0$. Equivalently, cumulated shocks on the factors at time $t+l$ when forming expectations at maturity $m \geq l$ are distorted as:

$$\varepsilon^{\mathbb{P}^\theta}_{t+l,C} = \sigma_m^\theta\varepsilon^{\mathbb{P}}_{t+l,C} + \psi_m^\theta f^{(l)}(\mathbf{X}_t)\mathbf{X}_t.$$

Then, the distorted interest rates to maturity m at time t reads:

$$r_{t,m} = \frac{1}{m}\sum_{i=0}^{m} r_{t+i} = \delta_0 + (1+\psi_m^\theta)b_m^{\mathbb{P}}(\mathbf{X}_t)\mathbf{X}_t + \sigma_m^\theta\varepsilon^{\mathbb{P}}_{t+m,C}.$$

where $b_m^{\mathbb{P}}$ as well as cumulated shocks to interest rates are defined as in Section 2. Then,

I assume that the SDF is such that:

$$\mathbf{X}_{t+l} = (\rho^{\mathbb{Q}})^l \mathbf{X}_t + \psi_m^\theta f^{(l)}(\mathbf{X}_t)\mathbf{X}_t + \sigma_m^\theta \varepsilon_{t+l,C}^{\mathbb{Q}}.$$

Equivalently, shocks on the factor at time $t+l$ when forming risk adjusted expectations at maturity $m \geq l$ are distorted as:

$$\varepsilon_{t+l,C}^{\mathbb{Q}^\theta} = \sigma_m^\theta \varepsilon_{t+l,C}^{\mathbb{Q}} + \psi_m^\theta b_m^{\mathbb{P}}(\mathbf{X}_t)\mathbf{X}_t.$$

which implies that:

$$r_{t,m} = \frac{1}{m}\sum_{i=0}^{m} r_{t+i} = \delta_0 + b_m^{\mathbb{Q}}\mathbf{X}_t + b_m^{\mathbb{P}}(\mathbf{X}_t)\mathbf{X}_t + \sigma_m^\theta \varepsilon_{t,m}^{\mathbb{Q}}.$$

This expression settles the ground to test both excess volatility of yields ($b_m^{\mathbb{P}}(\mathbf{X}_t)$ positively and increasingly in m contributing to the volatility of yields) and increasingly over-reacting beliefs ($\psi_m^\theta > 0$ and increasing in m).

B Estimation

B.1 Factor construction

I use the data available from Gürkaynak et al. (2007), who provide a daily estimate of the yield curve, from ??, with maturities $1y$, $2y$, $\dots 30$y
Factors are constructed as follows. First estimate of the variance covariance matrix as:

$$\frac{1}{T}\sum_{t=1}^{T}\underbrace{y_t y_t^\top}_{30\times 30}.$$

Then, I consider the first three principal components of the variance covariance matrix, $\underbrace{PC_1}_{30\times 1}$. PC_1 and PC_3. Principal components are unconditionally orthogonal. For $i = 1, 2, 3$, I build factor X_i as the projection of the yield curve into PC_i:

$$X_{t,i} := \langle PC_i, y_t \rangle.$$

Principal components and factors are plotted in Figure **??**.

B.2 Parameter Estimation

We want to estimate the parameters of the risk neutral dynamics:

$$y_{t+1} = \delta_0 + \delta_1^\top X_t$$
$$\mathbf{X}_t = \rho \mathbf{X}_{t-1} + \Sigma^C \epsilon_t^{\mathbb{Q}}$$

where is the upper triangular Cholesky decomposition of the one period variance-covariance Σ. This is assumed not to be stochastic and to be the same under $\mathbb{Q}$ and under $\mathbb{P}$. Therefore is can be estimated via OLS.

In order to estimate the remaining parameters ρ_1, ρ_2 and ρ_3, let me compute the affine relation for yields. The price at time t of a bond with time to maturity $m \geq 1$ is:

$$P_{t,m} = e^{-y_{t,m}m} = \mathbb{E}_t^{\mathbb{Q}}\left[e^{-(r_{t+1}+\cdots+r_{t+m})}\right] =$$
$$= \exp\left\{-\delta_0 m - \left(\sum_{i=1}^{3}\delta_{1,i}\sum_{k=0}^{m-1}\rho_i^k\right)X_{t,i}\right\}\mathbb{E}_t^{\mathbb{Q}}\left[\exp\left\{-\sum_{i=0}^{m-2}\delta_1^\top\Sigma^C\rho^i\right\}\right].$$

Moving from prices to yields:

$$y_{t,m} = -\frac{\log P_{t,m}}{m} = \delta_0 + \frac{\delta_1^\top\sum_{k=0}^{m-1}\rho^k}{m}\mathbf{X}_t - \frac{\log\mathbb{E}_t^{\mathbb{Q}}\left[\exp\left\{-\sum_{i=0}^{m-2}\delta_1^\top\Sigma^C\rho^i\right\}\right]}{m}$$

Note that the maturity m need to be expressed in the same units of the sampling frequencies. For example, for monthly data and yearly maturities for $1y$ to $30y$, the maturities should be expressed in months, $m = 12\times 1, \ldots 12\times 30$.
Let us know discuss hot to estimate model's parameters via OLS. First, we get $\hat{\delta}_0, \hat{\delta}_1^\top$ via OLS estimation of:

$$y_t^1 := r_t = \delta_0 + \delta_1^\top X_t.$$

Thus we can estimate $\rho_{i,\tau}$ as:

$$\hat{\rho} := \arg\min_{\rho_{i,\tau}}\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{m=1}^{30}y_{t,m}(\rho) - \bar{y}_{m,t}\right)^2,$$

where $y_{t,m}(\rho)$ is computed via the affine relation, while $\bar{y}_{m,t}$ denote the data. Similarly, for computing, maturity by maturity parameters, I compute:

$$\hat{\rho}_m := \arg\min_{\rho_{i,\tau}}\left(\frac{1}{T}\sum_{t=1}^{T}y_{t,m}(\rho) - \bar{y}_{m,t}\right)^2.$$

B.3 Discretization

I need to discretize the state space in order to compute the Arrow-Debreu price matrix. There are well known methods to efficiently discretize an $AR(1)$. In the case of multivariate processes with not trivial correlations, the following transformation provide a set of uncorrelated equation:

$$\mathbf{X}_t \longrightarrow \mathbf{Z}_t := {\Sigma^C}^{-1}\mathbf{X}_t.$$

Now, consider the VAR(1) $\mathbb{Q}$-dynamics:

$$\mathbf{X}_t = \rho\mathbf{X}_{t-1} + \Sigma^C\epsilon_t^{\mathbb{Q}}.$$

Multiplying both sides by ${\Sigma^C}^{-1}$, I get:

$$\mathbf{Z}_t = \rho\mathbf{Z}_{t-1} + \epsilon_t^{\mathbb{Q}}.$$

This is a system of three independent $AR(1)$ that can be independently discretized. After the discretization I come back to the real factor with the inverse transformation to the discretized version of $\mathbf{Z}_t$, say $\mathbf{Z}_t^D$:

$$\mathbf{X}_t^D \longleftarrow \Sigma^C\mathbf{Z}_t^D.$$

I rely on the Rouwenhorst discretization method, which matches conditional and unconditional first and second moments, and is the state of the art among those technique. I use 10 levels for each factor, thus having overall 1000 states. This choice does not seems to affect the results.

C Tables

	Y1	Y2	Y3	Y4	Y5	Y6	Y7	Y8	Y9	Y10
const	3.71***	3.97***	4.21***	4.43***	4.63***	4.81***	4.97***	5.11***	5.23***	5.80***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$X_{1,t}$	0.21***	0.22***	0.22***	0.21***	0.21***	0.20***	0.20***	0.19***	0.19***	0.16***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$X_{2,t}$	0.52***	0.42***	0.33***	0.25***	0.19***	0.13***	0.08***	0.04***	0.01***	-0.16***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$X_{3,t}$	0.25***	0.16***	0.08***	0.01***	-0.05***	-0.11***	-0.15***	-0.18***	-0.21***	0.36***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
R-squared	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
No. observations	7903	7903	7903	7903	7903	7903	7903	7903	7903	7903

Table 1: Regressions of yields on factors

	Y11	Y12	Y13	Y14	Y15	Y16	Y17	Y18	Y19	Y20
const	5.43***	5.51***	5.58***	5.64***	5.69***	5.73***	5.76***	5.79***	5.81***	5.83***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$X_{1,t}$	0.18***	0.18***	0.18***	0.18***	0.18***	0.17***	0.17***	0.17***	0.17***	0.17***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$X_{2,t}$	-0.04***	-0.06***	-0.07***	-0.09***	-0.10***	-0.11***	-0.12***	-0.12***	-0.13***	-0.13***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$X_{3,t}$	-0.22***	-0.22***	-0.21***	-0.19***	-0.17***	-0.14***	-0.11***	-0.08***	-0.05***	-0.01***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
R-squared	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
No. observations	7903	7903	7903	7903	7903	7903	7903	7903	7903	7903

	Y21	Y22	Y23	Y24	Y25	Y26	Y27	Y28	Y29	Y30
const	5.84***	5.84***	5.85***	5.85***	5.84***	5.84***	5.83***	5.82***	5.81***	5.80***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$X_{1,t}$	0.17***	0.17***	0.17***	0.17***	0.17***	0.17***	0.16***	0.16***	0.16***	0.16***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$X_{2,t}$	-0.14***	-0.14***	-0.14***	-0.15***	-0.15***	-0.15***	-0.15***	-0.15***	-0.16***	-0.16***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$X_{3,t}$	0.03***	0.06***	0.10***	0.14***	0.18***	0.22***	0.25***	0.29***	0.33***	0.36***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
R-squared	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
No. observations	7903	7903	7903	7903	7903	7903	7903	7903	7903	7903

The estimated covariance matrix of the factors is:

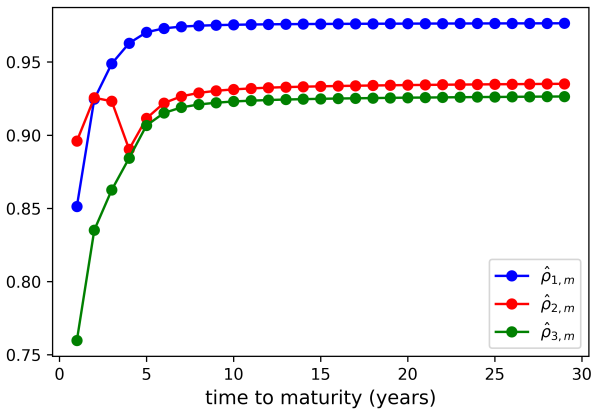
$$\Sigma = \begin{pmatrix} 1. & -0.333506 & -0.221877 \\ -0.333506 & 1 & 0.033376 \\ -0.221877 & -0.333506 & 1 \end{pmatrix}.$$

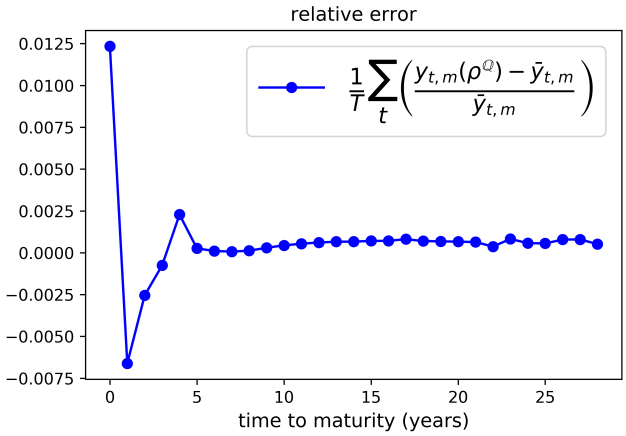
The Cholesky decomposition of the estimated covariance matrix of the factors is:

$$\Sigma^C = \begin{pmatrix} 1. & 0. & 0. \\ -0.333506 & 0.94274798 & 0 \\ -0.221877 & -0.0430882 & 0.974122171 \end{pmatrix}.$$

Average estimated ρ :

$$\rho^{\mathbb{Q}} = \begin{pmatrix} 0.968190 & 0 & 0 \\ 0 & 0.92868686 & 0 \\ 0 & 0 & 0.911450 \end{pmatrix}$$





D The Recovery Theorem

Let me now show how the *path independence* assumption helps with the identification of beliefs. Under path independence the fundamental asset pricing equation can be written as:

$$\mathbb{A}^{\theta}_{ij} z_j = \delta z_i \mathbb{P}^{\theta}_{ij}.$$

Noticing that probabilities add up to one, one gets:

$$\sum_j \mathbb{A}^{\theta}_{ij} z_j = \delta z_i.$$

This is an eigenvalue problem from the Arrow-Debreu matrix. By Perron-Frobenious theorem[18] (see Meyer (2000)), the Arrow-Debreu matrix has unique positive eigenvector (which can be identified with z) corresponding to the largest eigenvalue (which can be identified with δ). Therefore, under *path independent SDF* it is possible to identify beliefs as:

$$\mathbb{P}^{\theta}_{ij} = \mathbb{A}^{\theta}_{ij} \frac{1}{\delta} \frac{z_j}{z_i}.$$

[18]The Perron-Frobenious theorem assumes a positive and irreducible matrix. The Arrow-Debreu matrix is positive and irreducible under no arbitrage.

So far, I exploited only *one period* Arrow-Debreu prices. Do prices of *multi-period* Arrow-Debreu secutiries provide additional information? If the law of iterated expectation holds, then long term beliefs are *iterations* of short term belies:

$$\mathbb{P}_m = \underbrace{\mathbb{P} \times \cdots \times \mathbb{P}}_{\text{m times}} = \mathbb{P}^m,$$

where $\mathbb{P}^m$ is the m-th power of the one period transition probability matrix $\mathbb{P}$. In this case one period Arrow-Debreu prices are sufficient to identify the term structure of beliefs. However, if the law holds true or not is ultimately an empirical question: moreover, a testable implication of the law of iterated expectation is the fact that the term structure of CG coefficients needs to be flat (2). Given a panel of bond prices $\{P_{t,m}\}_{t,m}$, Arrow-Debreu prices to maturity m can be therefore estimated by relying of the time series $\{P_{m,t}\}_t$ only. Using the recovery theorem, the econometrician can access the term structure of beliefs $\mathbb{P}^\theta_m$, where $[\mathbb{P}^\theta_m]_{ij}$ is the *believed* probability of transitioning from state i to state j in m steps and test the *horizon dependence* of beliefs as discussed in (2).

E Proofs

Proposition 1

Proof. From the last equation in Appendix A, it is straightforward to compute:

$$y_{t,m} = -\frac{1}{m}\mathbb{E}_t^{\mathbb{Q}^\theta}\left[e^{-m\cdot r_{t,m}}\right] = \delta_0 - \frac{1}{m}\log\mathbb{E}^{\mathbb{Q}}\left[e^{-\sigma_m^\theta \varepsilon_{t,m}^{\mathbb{Q}}}\right] + b_m^{\mathbb{Q}}\mathbf{X}_t + \psi_m^\theta b_m^{\mathbb{P}}(\mathbf{X}_t)\mathbf{X}_t.$$

The expectations is the cumulant generating function is independent of time t information because $\mathbb{Q}$ shocks are homoskedastic.

□

Theorem 1 (Increasing over-reaction and excess volatility)

Proof. First, in order to compute the forecast of future yields (which are an endogenous variable) at maturity m, note that the agent first compute distorted forecasts of factors and then she plug those in into the pricing equation which relates yields to factors. Therefore, there is a compounding effect of distortion coefficients, which will appear in the following calculations. The following approximation will be helpful. Consider the term:

$$\frac{b_m^\top Cov(F_t[\mathbf{X}_{t+1}], F_t[\mathbf{X}_{t+1}] - F_{t-k}[\mathbf{X}_{t+1}])b_m}{b_m^\top \mathbb{V}[F_t[\mathbf{X}_{t+1}] - F_{t-k}[\mathbf{X}_{t+1}]]b_m},$$

where $F_t[\mathbf{X}_{t+1}] - F_{t-k}[\mathbf{X}_{t+1}]$ is the forecast revision, when the reference distribution in the expectation model is the $k-$lagged one. The case $k = \infty$ corresponds to taking the unconditional (or historical) average as benchmark. Convex combination of the past forecasts can be easily handled as well. In this proof, $b_m := b_m^{\mathbb{P}}(\mathbf{X}_t)$ for convenience of notation. If the DGP is a $VAR(1)$ with diagonal persistence matrix ρ, then:

$$\frac{b_m^\top Cov(F_t[\mathbf{X}_{t+1}], F_t[\mathbf{X}_{t+1}] - F_{t-k}[\mathbf{X}_{t+1}])b_m}{b_m^\top \mathbb{V}[F_t[\mathbf{X}_{t+1}] - F_{t-k}[\mathbf{X}_{t+1}]]b_m} = 1 - \frac{b_m^\top(\rho^2 - \rho^{1+k})\mathbb{V}[\mathbf{X}_t]b_m}{b_m^\top \mathbb{V}[F_t[\mathbf{X}_{t+1}] - F_{t-k}[\mathbf{X}_{t+1}]]b_m}. \tag{14}$$

Therefore, the term is approximately equal to one, for highly persistent processes. The argument goes through also for non linear processes, provided that the local persistence $\rho(\mathbf{X}_t)$ is, on average, close to one. Moreover, for a fixed persistence matrix ρ, the deviation from one, asymptotically vanishes for $m \to \infty$. This is so because each entry of b_m convergences geometrically for large maturities. The approximation is exact also in the special case of exactly equally persistent factors or in the special case of a single factor model.

i) Under rational expectations. $\beta_m = 0$ because the forecast error is unpredictable on the basis of past information.
Under maturity independent distortion ψ^θ:

$$\beta_m = \frac{Cov\left(FE^\theta_{t+1}[y^\theta_{t+1,m}], FR^\theta_t[y^\theta_{t+1,m}]\right)}{\mathbb{V}\left[FR^\theta_t[y^\theta_{t+1,m}]\right]} = \frac{-\psi^\theta(1+\psi^\theta)}{(1+\psi^\theta)^2} \frac{(b^\theta_m)^\top Cov\left(F_t[\mathbf{X}_{t+1}], FR_t[\mathbf{X}_{t+1}]\right) b^\theta_m}{(b^\theta_m)^\top \mathbb{V}[FR_t[\mathbf{X}_{t+1}]] b^\theta_m}.$$

Under average reference forecast or under one factor model or under approximation (14) the second fraction reduces to a positive constant. Therefore β_m is negative $\iff$ $\psi^\theta > 0$.
Under maturity dependent distortion ψ^θ_m:

$$\begin{aligned}\beta_m &= \frac{Cov\left(FE^\theta_{t+1}[y^\theta_{t+1,m}], FR^\theta_t[y^\theta_{t+1,m}]\right)}{\mathbb{V}\left[FR^\theta[y^\theta_{t+1,m}]\right]} \\ &= \frac{-\psi^\theta_{m+1}(1+\psi^\theta_m)}{(1+\psi^\theta_m)^2} \frac{(b^\theta_m)^\top Cov\left(F_t[\mathbf{X}_{t+1}], F_t[\mathbf{X}_{t+1}] - \frac{\psi^\theta_{m+2}}{\psi^\theta_{m+1}} F_{t-1}[\mathbf{X}_{t+1}]\right) b^\theta_m}{(b^\theta_m)^\top \mathbb{V}[F_t[\mathbf{X}_{t+1}] - \frac{\psi^\theta_{m+2}}{\psi^\theta_{m+1}} F_{t-1}[\mathbf{X}_{t+1}]] b^\theta_m}.\end{aligned}$$

Under average reference forecast or under one factor model the second fraction reduces to a positive constant. Under approximation (14) and assuming $\frac{\psi_{m+2}}{\psi_{m+1}}\rho_i < 1$ for each entry of the persistence matrix ρ, the second fraction reduces to a positive constant. β^θ_m is thus negative and increasing $\iff$ $\psi^\theta_m > 0$ and ψ^θ_m is increasing in m.
ii) It follows from the expression in Proposition (1).

□

Theorem 2

Proof. *i*) Calculations

ii)

1.) Under rational expectations, the misspecified CG coefficients, computed with forecasts identified by the recovery theorem and distorted by the martingale component of the SDF, are:

$$\widetilde{\beta}_m = \frac{Cov\left(\widetilde{FE}_{t+1}[y_{t+1,m}], \widetilde{FR}_t[y_{t+1,m}]\right)}{\mathbb{V}\left[\widetilde{FR}_t[y_{t+1,m}]\right]}.$$

First, consider the one period ahead misspecified forecast. I drop the additive constant $a_m^{\mathbb{Q}}$ in the following calculations, because it is inessential for the computation of covariances.

$$\begin{aligned}\widetilde{F}_t[\mathbf{X}_{t+1}] &= b_m^\top \mathbb{E}_t\left[h_{t+1}\mathbf{X}_{t+1}\right] = b_m^\top \mathbb{E}_t\left[\mathbf{X}_{t+1}\right]\mathbb{E}_t\left[h_{t+1}\right] + b_m^\top \mathrm{Cov}_t\left(h_{t+1}, \mathbf{X}_{t+1}\right) \\ &= b_m^\top \mathbb{E}_t\left[\mathbf{X}_{t+1}\right] + b_m^\top \mathrm{Cov}_t\left(h_{t+1}, \mathbf{X}_{t+1}\right).\end{aligned}$$

Therefore, the misspecified forecast error reads:

$$\widetilde{FE}_{t+1}[y_{t+1,m}] = y_{t+1,m} - \widetilde{\mathbb{E}}_t\left[y_{t+1,m}\right] = FE_{t+1}[y_{t+1,m}] - b_m^\top Cov_t\left(h_{t+1}, \mathbf{X}_{t+1}\right).$$

Similarly, the two period ahead misspecified forecast reads:

$$\begin{aligned}\widetilde{F}_{t-1}[y_{t+1,1}] &= b_m^\top \mathbb{E}_{t-1}\left[h_{t+1}\mathbf{X}_{t+1}\right] = b_m^\top \mathbb{E}_{t-1}\left[\mathbf{X}_{t+1}\right]\mathbb{E}_{t-1}\left[h_{t+1}\right] + b_m^\top \mathrm{Cov}_{t-1}\left(h_{t+1}, \mathbf{X}_{t+1}\right) \\ &= b_m^\top \mathbb{E}_{t-1}\left[\mathbf{X}_{t+1}\right] + b_m^\top \mathrm{Cov}_{t-1}\left(h_{t+1}, \mathbf{X}_{t+1}\right).\end{aligned}$$

The forecast revision of yields with maturity m reads:

$$\widetilde{FR}_t[y_{t+1,m}] = FR_t[y_{t+1,m}] + b_m^\top\left(Cov_t\left(h_{t+1}, \mathbf{X}_{t+1}\right) - Cov_{t-1}\left(h_{t+1}, \mathbf{X}_{t+1}\right)\right).$$

The covariance between forecast error and forecast revision reads:

$$Cov\left(\widetilde{FE}_{t+1}[y_{t+1}], \widetilde{FR}_t[y_{t+1,m}]\right) = b_m^\top B_1 b_m,$$

where:

$$B_1 := -\text{Cov}\left(\text{Cov}_t\left(h_{t+1}, \mathbf{X}_{t+1}\right), \widetilde{FR}_t[\mathbf{X}_{t+1}]\right).$$

I used the fact that under rational expectations the forecast error and the forecast revision are orthogonal. Note that the term B_1 does not depend on the maturity m. The variance of the forecast revision reads:

$$\mathbb{V}\left[\widetilde{FR}_t[y_{t+1,m}]\right] = b_m^\top \left(FR_t[\mathbf{X}_{t+1}] + B_2\right) b_m,$$

where:

$$\begin{aligned} B_2 = &\mathbb{V}\left[\text{Cov}_t\left(h_{t+1}, \mathbf{X}_{t+1}\right) - Cov_{t-1}\left(h_{t+1}, \mathbf{X}_{t+1}\right)\right] \\ &+ 2Cov\left(FR_t[\mathbf{X}_{t+1}], Cov_t\left(h_{t+1}, \mathbf{X}_{t+1}\right) - Cov_{t-1}\left(h_{t+1}, \mathbf{X}_{t+1}\right)\right). \end{aligned}$$

The misspecified CG coefficients therefore read:

$$\widetilde{\beta}_m = \frac{b_m^\top \left(Cov\left(FE_{t+1}[\mathbf{X}_{t+1}], FR_t[\mathbf{X}_t]\right) + B_1\right) b_m}{b_m^\top \left(\mathbb{V}[FR_t[\mathbf{X}_{t+1}]] + B_2\right) b_m} = \frac{b_m^\top B_1 b_m}{b_m^\top \left(\mathbb{V}[FR_t[\mathbf{X}_{t+1}]] + B_2\right) b_m},$$

which, under approximation (14), does not depend on m. The denominator is positive, since it is a variance. The numerator is positive in the case of standard consumption based asset pricing models[19]. In this case, the econometrician detects a flat term structure of CG coefficients, even though there are not departures from RE. Conversely, a non trivial term structure of CG regression coefficients reveal to the econometrician that beliefs are not rational.

2. Non rational expectations. First, consider the one period ahead misspecified forecast.

[19] In standard consumption based asset pricing models, the SDF is negatively correlated with consumption growth, which corresponds to a linear combination of the factors. Therefore, when positive news arrives, the first term in the unconditional covariance in B_1 decreases, while the forecast revision increases. The unconditional covariance is therefore negative and thus B_1 is positive.

I drop the additive constant $a_m^{\mathbb{Q}^\theta}$ in the following calculations, because it is inessential for the computation of covariances.

$$\widetilde{FE}_{t+1,m}^{\theta} := \bar{y}_{t+1,m}^{\theta} - \mathbb{E}_t^{\widetilde{\mathbb{P}}^\theta}\left[y_{t+1,m}^{\theta}\right] = \bar{y}_{t+1,m}^{\theta} - (1+\psi_{m+1}^{\theta})\mathbb{E}_t^{\widetilde{\mathbb{P}}}\left[y_{t+1,m}^{\theta}\right]$$
$$\bar{y}_{t+1,m}^{\theta} - \mathbb{E}_t^{\widetilde{\mathbb{P}}^\theta}\left[y_{t+1,m}^{\theta}\right] = \bar{y}_{t+1,m}^{\theta} - (1+\psi_{m+1}^{\theta})(\mathbb{E}_t^{\widetilde{\mathbb{P}}}\left[y_{t+1,m}^{\theta}\right] + \mathbb{E}_t^{\mathbb{P}}\left[y_{t+1,m}^{\theta}\right] - \mathbb{E}_t^{\mathbb{P}}\left[y_{t+1,m}^{\theta}\right])$$
$$= FE_{t+1,m}^{\theta} + \left(\mathbb{E}_t^{\mathbb{P}}\left[y_{t+1,m}^{\theta}\right] - \mathbb{E}_t^{\widetilde{\mathbb{P}}}\left[y_{t+1,m}^{\theta}\right]\right) - \psi_{m+1}^{\theta}\mathbb{E}_t^{\widetilde{\mathbb{P}}}\left[y_{t+1,m}^{\theta}\right].$$

Distorted and misspecified CB coefficients reads:

$$\widetilde{\beta}_m^{\theta} = \beta_m^{\theta} + \frac{{b_m^{\theta}}^{\top}\left(Cov\left(F_{t+1}[\mathbf{X}_{t+1}] - \widetilde{F}_t[\mathbf{X}_{t+1}], \widetilde{F}_t[\mathbf{X}_{t+1}] - \frac{\psi_{m+2}}{\psi_{m+1}}\widetilde{F}_{t-1}[\mathbf{X}_{t+1}]\right) + B_1\right)b_m^{\theta}}{{b_m^{\theta}}^{\top}\left(\mathbb{V}[\widetilde{FR}_t[\mathbf{X}_{t+1}]] + B_2\right)b_m^{\theta}}$$

the second term, under one factor model or under approximation (14) does not depend on m. The coefficients $\widetilde{\beta}_m^{\theta}$ are negative and decreasing iff ψ_m^{θ} is positive and increasing and provided that $\frac{(b_m^{\theta})^{\top}\mathbb{V}[\widetilde{F}_t[\mathbf{X}_{t+1}]]b_m^{\theta}}{(b_m^{\theta})^{\top}Cov(\widetilde{F}_t[\mathbf{X}_{t+1}],\widetilde{F}_{t-1}[\mathbf{X}_{t+1}]b_m^{\theta}} > \frac{\psi_2^{\theta}}{\psi_1^{\theta}}$. Differences of distorting coefficients remain identified.

□

Theorem (3 ($\mathbb{P}$-diagnostic expectations)

Proof.

$$r_{t,m} \overset{\mathbb{P}}{\sim} \mathcal{N}\left(\delta_0, \frac{1}{m^2}(\delta_1)^{\top}\Sigma(1-(\rho^{\mathbb{P}})^2))^{-1}\delta_1\right)$$

and

$$r_{t+m}|\mathbf{X}_t \overset{\mathbb{P}}{\sim} \mathcal{N}\left(\delta_0 + (\delta_1)^{\top}\left(\sum_{i=0}^{m-1}(\rho^{\mathbb{P}})^i\right)\mathbf{X}_t, \frac{1}{m^2}(\delta_1)^{\top}\Sigma\left(\sum_{i=0}^{m-2}(\rho^{\mathbb{P}})^{2i}\right)\delta_1\right).$$

We want to compute the distribution:

$$f_{\mathbb{P}^\theta}(r_{t+m}|\mathbf{X}_t) \propto f_{\mathbb{P}}(r_{t+m}|\mathbf{X}_t)\left(\frac{f_{\mathbb{P}}(r_{t+m}|\mathbf{X}_t)}{f_{\mathbb{P}}(r_{t+m})}\right)^\theta. \tag{15}$$

First observe that, given $G_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $G_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ with $\sigma_2^2 > \sigma_1^2$:

$$\frac{1}{\int_{\mathbb{R}} f_{G_2}(x)\left(\frac{f_{G_1}(x)}{f_{G_2(x)}}\right)^\theta dx} f_{G_2}(x)\left(\frac{f_{G_1}(x)}{f_{G_2(x)}}\right)^\theta$$

is a Gaussian pdf with mean:

$$\mu_1 + \frac{\theta\frac{\sigma_1^2}{\sigma_2^2}}{1+\theta-\theta\frac{\sigma_1^2}{\sigma_2^2}}(\mu_1-\mu_2),$$

and variance:

$$\left(\frac{1+\theta}{\sigma_1^2} - \frac{\theta}{\sigma_2^2}\right)^{-1}.$$

Then, apply the previous computation to $r_{t,m}$. □

Rational pricing of zero coupon bonds

Consider the arbitrage free price of zero coupon bonds (ZCB) at time t with time to maturity $t+m$:

$$P_{t,m} = \mathbb{E}_t^{\mathbb{Q}}\left[e^{-r_{t,m}}\right],$$

where, as usual, $r_{t,m} = \frac{\sum_{i=0}^{m-1} r_{t+i}}{m}$ and r_t denotes the short rate at time t. Assume that the short rate is an affine function of a set of factors, $r_t = \delta_0 + \delta_1^\top \mathbf{X}_t$, which evolve as a VAR(1) under the risk neutral measure:

$$\mathbf{X}_{t+1} = \rho^{\mathbb{Q}}\mathbf{X}_t + \Sigma^{\mathbb{Q}^C}\varepsilon_{t+1}^{\mathbb{Q}}.$$

$\Sigma^{\mathbb{Q}^C}$ is upper triangular, $\rho^{\mathbb{Q}}$ is diagonal and $\Sigma^{\mathbb{Q}} := (\Sigma^{\mathbb{Q}^C})^\top \Sigma^{\mathbb{Q}^C}$ is the variance covariance matrix of the residuals. Then, $r_{t,m}$ has conditional risk neutral mean and variances given by:

$$\mathbb{E}_t^{\mathbb{Q}}[r_{t,m}] = \delta_0 + \frac{\delta_1^\top \sum_{i=0}^{m-1} (\rho^{\mathbb{Q}})^i \mathbf{X}_t}{m}$$
$$\mathbb{V}_t^{\mathbb{Q}}[r_{t,m}] = \frac{\delta_1^\top \sum_{i=0}^{m-2} \sum_{j=0}^{i} (\rho^{\mathbb{Q}})^{2j} \Sigma^{\mathbb{Q}} \delta_1}{m^2}$$

Consider the following class of additive $\mathbb{P}$-dynamics for the factors.

$$\mathbf{X}_{t+1} = f^{\mathbb{P}}(\mathbf{X}_t) + \Sigma^{\mathbb{P}^C} \varepsilon_{t+1}^{\mathbb{P}},$$

where $\Sigma^{\mathbb{P}^C}$ is upper triangular, $f^{\mathbb{P}}(\mathbf{X_t})$ is diagonal and $\Sigma^{\mathbb{P}} := (\Sigma^{\mathbb{P}^C})^\top \Sigma^{\mathbb{P}^C}$ is the variance covariance matrix of the residuals. The $\mathbb{P}$ dynamics is assumed to be Markovian as the risk neutral one and with additive noise. It not assumed to be linear. The linear case is simply recovered imposing $f^{\mathbb{P}}(\mathbf{X}_t) = \rho^{\mathbb{P}} \mathbf{X}_t$. Then, $r_{t,m}$ has conditional physical mean and variances given by:

$$\mathbb{E}_t^{\mathbb{P}}[r_{t,m}] = \delta_0 + \frac{\delta_1^\top \sum_{i=0}^{m-1} (f_i^{\mathbb{P}}(\mathbf{X}_t)}{m}$$
$$\mathbb{V}_t^{\mathbb{P}}[r_{t,m}] = \frac{\delta_1^\top \sum_{i=0}^{m-2} \sum_{j=0}^{i} (f_j^{\mathbb{P}})^2 \Sigma^{\mathbb{P}} \delta_1}{m^2},$$

where $f_0^{\mathbb{P}}(\mathbf{X}_t) = \mathbf{X}_t$, $f_1^{\mathbb{P}}(\mathbf{X}_t) = f^{\mathbb{P}}(\mathbf{X}_t)$ and, for $i > 1$, $f_i^{\mathbb{P}}(\mathbf{X}_t) = \underbrace{f^{\mathbb{P}}(\ldots f^{\mathbb{P}}}_{i \text{ times}}(\mathbf{X}_t))$.

Thus, under this class of asset pricing models the following change of measure applies:

$$\Sigma^{\mathbb{Q}^C} \varepsilon_{t+1}^{\mathbb{Q}} = \Sigma^{\mathbb{P}^C} \varepsilon_{t+1}^{\mathbb{P}} + (f^{\mathbb{P}}(\mathbf{X}_t) - \rho^{\mathbb{Q}} \mathbf{X}_t)$$
$$= \Sigma^{\mathbb{P}^C} \epsilon_{t+1}^{\mathbb{P}} + \mathbb{E}_t^{\mathbb{P}}[\mathbf{X}_{t+1}] - \mathbb{E}_t^{\mathbb{Q}}[\mathbf{X}_{t+1}].$$

Effective shocks to interest rates with maturity m, $r_{t,m} - \mathbb{E}_t^k[r_{t,m}]$, for $k = \mathbb{P}, \mathbb{Q}$ are given by:

$$\delta_1^\top \sum_{i=0}^{m-2}\sum_{j=0}^{i}(\rho^{\mathbb{Q}})^j \Sigma^{\mathbb{Q}}\varepsilon_{t+i}^{\mathbb{P}} = \delta_1^\top \sum_{i=0}^{m-2}\sum_{j=0}^{i} f_j^{\mathbb{P}}(\mathbf{X}_t)\Sigma^{\mathbb{P}}\varepsilon_{t+i}^{\mathbb{P}} + m\left(\delta_1^\top \sum_{i=0}^{m-1} f_i^{\mathbb{P}}(\mathbf{X}_t) - \delta_1^\top \sum_{i=0}^{m-1}(\rho^{\mathbb{Q}})^i \mathbf{X}_t\right),$$

or:

$$\sqrt{\mathbb{V}_t^{\mathbb{Q}}[r_{t,m}]}\varepsilon_{t,m}^{\mathbb{Q}} = \sqrt{\mathbb{V}_t^{\mathbb{P}}[r_{t,m}]}\varepsilon_{t,m}^{\mathbb{P}} + \mathbb{E}_t^{\mathbb{P}}[r_{t,m}] - \mathbb{E}_t^{\mathbb{Q}}[r_{t,m}], \tag{16}$$

where: $\varepsilon_{t,m}^k := \frac{\sum_{i=1}^{m-1}\sum_{j=0}^{i}\varepsilon_{t+j}^k}{m}$, for $k = \mathbb{P}, \mathbb{Q}$.

Pricing of zero coupon bonds with over-reacting beliefs

First consider the case in which beliefs distortions only affects linearly the first moment: $\mathbb{E}_t^{\mathbb{P}^\theta}[r_{t,m}] := a_{t,m}\mathbb{E}^{\mathbb{P}}[r_{t,m}] + b_{t,m}$, where $a_{t,m}, b_{t,m}$ and known at time t. Then, an agent which believes $\mathbb{E}_t^{\mathbb{P}^\theta}[r_{t,m}]$ is the correct mean, while having the same preferences as in the rational case, will adjust beliefs as:

$$\sqrt{\mathbb{V}_t^{\mathbb{Q}}[r_{t,m}]}\epsilon_{t,m}^{\mathbb{Q}} = \sqrt{\mathbb{V}_t^{\mathbb{P}}[r_{t,m}]}\epsilon_{t,m}^{\mathbb{P}} + a_{t,m}\mathbb{E}_t^{\mathbb{P}}[r_{t,m}] + b_{t,m} - \mathbb{E}_t^{\mathbb{Q}^\theta}[r_{t,m}].$$

The risk neutral distorted expectation $\mathbb{E}_t^{\mathbb{Q}^\theta}[r_{t,m}]$ is determined from the condition that the change of measure (preference adjustment) is unchanged:

$$\mathbb{E}_t^{\mathbb{Q}^\theta}[r_{t,m}] = \mathbb{E}_t^{\mathbb{Q}}[r_{t,m}] + (a_{t,m} - 1)\mathbb{E}_t^{\mathbb{P}}[r_{t,m}] + b_{t,m}.$$

Consider now the case in which the variance is also distorted, in particular it is scaled as: $\mathbb{V}_t^{\mathbb{P}^\theta}[r_{t,m}] = c_{t,m}^2\mathbb{V}_t^{\mathbb{P}}[r_{t,m}]$, where $c_{t,m}^2$ is known at time t. Then:

$$\sqrt{\mathbb{V}_t^{\mathbb{Q}^\theta}[r_{t,m}]}\epsilon_{t,m}^{\mathbb{Q}} = c_{t,m}\sqrt{\mathbb{V}_t^{\mathbb{P}}[r_{t,m}]}\epsilon_{t,m}^{\mathbb{P}} + a_{t,m}\mathbb{E}_t^{\mathbb{P}}[r_{t,m}] + b_{t,m} - \mathbb{E}_t^{\mathbb{Q}^\theta}[r_{t,m}].$$

Then:

$$\mathbb{V}_t^{\mathbb{Q}^\theta}[r_{t,m}] = \mathbb{V}_t^{\mathbb{Q}}[r_{t,m}]c_{t,m},$$

and:

$$\mathbb{E}_t^{\mathbb{Q}^\theta}[r_{t,m}] = \mathbb{E}_t^{\mathbb{Q}}[r_{t,m}] + (a_{t,m} - 1)\mathbb{E}_t^{\mathbb{P}}[r_{t,m}] + b_{t,m}.$$

Note that if $\mathbb{P}^\theta$ satisfies the law of iterated expectations, then:

$$\mathbb{E}_{t-k}^{\mathbb{P}}[\mathbb{E}_t^{\mathbb{P}^\theta}[r_{t,m}] - \mathbb{E}_t^{\mathbb{P}}[r_{t,m}]] = (a_{t-k,m} - 1)\mathbb{E}_{t-k}^{\mathbb{P}}[r_{t-k,m}] + b_{t-k,m} = \mathbb{E}_{t-k}^{\mathbb{P}^\theta}[r_{t-k,m}] - \mathbb{E}_{t-k}^{\mathbb{P}}[r_{t-k,m}].$$

This means that the conditional bias $\mathbb{E}_t^{\mathbb{P}^\theta}[r_{t,m}] - \mathbb{E}_t^{\mathbb{P}}[r_{t,m}]$ is a $\mathbb{P}-$martingale and it is therefore unpredictable. Similarly, temporary misspricing is not predictable. In the case the law is violated instead there is, in principle, predictable misspricing (arbitrage opportunities).

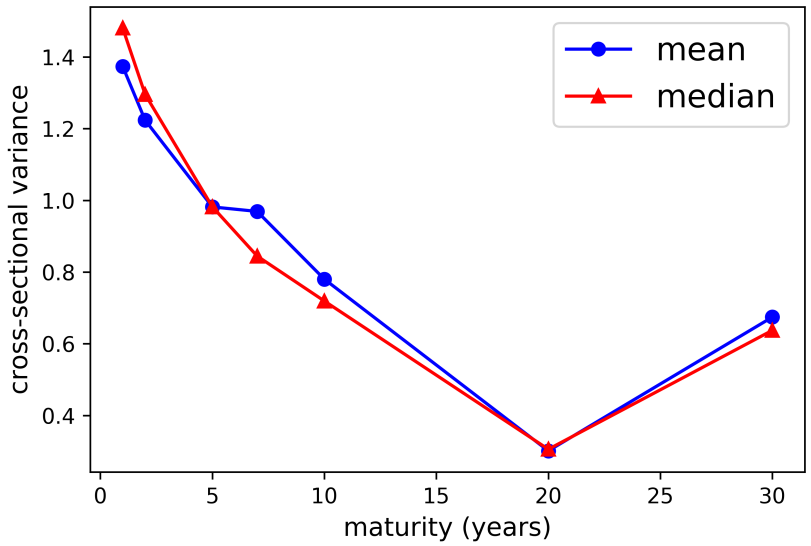


Figure 11: Time mean (blue circle) and median (red triangle) of analysts dispersion (cross sectional standard deviation) as a function of the maturity.

F Additional Results

Figure (5) shows that distortions from recovered beliefs are a combination of the "average" distortions (i.e. the distortion from the pooled regression) and of the distortion of the "average" (i.e. the consensus forecast), at least statistically. To quantitatively predict the aggregate outcome, one would need a specific model of aggregation, which is left for future work. The aforementioned mechanism proposed by Bordalo et al. (2018a) is however consistent with some evidence from the cross-sectional dispersion of forecasters: Figure (11) shows that the average cross-sectional dispersions decrease with maturity. Views are more aligned for predictions at longer horizons than for predictions at shorter horizons[20]. However, at longer horizon, data are much less rich, which may weaken the claim. In fact, as shown in Figure (5), the maturities $20y$ and $30y$, it is not possible to categorize forecasters along the mean square error dimension; also, for the consensus forecast, confidence intervals are huge.

[20]This may be intuitive thinking that a sharp benchmark for long run interest rates forecasts may be the one set by central banks.