Maturity Increasing Over-reaction and Bond Market Puzzles

Daniele d'Arienzo*

May 2020

Abstract

Long-term treasury yields are known to be: i) excessively volatile (Giglio and Kelly (2018)) ii) highly sensitive to short rate movements (Hanson et al. (2018)) as well as iii) highly predictable from non priced factors (Duffee (2013)). I assess the possibility that these puzzles may be due to non-rational investor beliefs. Using survey data as well as data on market beliefs recovered from observed yields, I document that expectations about long rates over-react to news relative to expectations about short rates. I show that introducing diagnostic expectations into an affine term structure model yields such maturity increasing over-reaction and reconciles all three puzzles. When benchmarked to external data on diagnostic distortions, this model accounts for: i) roughly 80% of the excess volatility puzzle ii) for 40% of the excess sensitivity of long rates, and iii) for additional excess bond returns predictability coming from past forecast revisions.

JEL Classification: E43, E71, G41

*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, Max Croce, Stefano Giglio, Massimo Guidolin, Sam Hanson, Christian Skov Jensen, Eben Lazarus, Emile Marin (discussant), Ian Martin, Filip Matějka, Peter Maxted, Florian Nagler, Fulvio Ortu, Stefano Rossi, Andrei Shleifer, Adi Sunderam, Claudio Tebaldi, Yijun Zhou (discussant) and seminar participants at Bocconi University, Harvard University, Boston College, TADC 2019, Sloan-Nomis 2019, EDGE Jamboree 2019, HBS, WU. All errors are my own.

1 Introduction

A large body of empirical work on the term structure of interest rates documents three leading anomalies. First, long term interest rates exhibit excess sensitivity, in that they comove too much with short term interest rates relative to what is implied by the observed persistence of shocks and by the expectations hypothesis (Shiller (1979), Gürkaynak et al. (2005), Hanson et al. (2018), Brooks et al. (2018)). Second, long rates exhibit excess volatility relative to contemporaneous variations in short rates for a large class of rational models (Giglio and Kelly (2018)). This second puzzle is even deeper, for it holds even after allowing for the discount rate variation of conventional term structure models. Third, excess bond returns are highly predictable (Campbell and Shiller (1991)), even after controlling for information embodied in the current yield curve (Cochrane and Piazzesi (2005), Duffee (2013), Cieslak (2018)).

One way to rationalize these findings is to assume that discount rates suitably move over and above what is allowed for by the conventional term structure models considered by Giglio and Kelly (2018). Here I consider instead a second possibility, namely that these puzzles may be primarily due to investors' non-rational expectations within classical term structure models. To analyze this hypothesis, I discipline my analysis by relying on survey data as well as on market expectations about future interest rates, and by relying on the psychologically founded diagnostic expectations model. The latter has been applied to shed light on several phenomena in macro-finance Gennaioli and Shleifer (2018). My results show that a reasonable degree of "diasgosticity" unifies the three puzzles above, offering a good quantitative account of them.

To motivate my approach, consider Blue Chip data on professional forecasters' expectations of one quarter ahead interest rates at maturities of $1y$, $2y$, $5y$, $10y$, $20y$ and $30y$ (monthly frequency). Figure (1) below reports, for each maturity, the pooled OLS coefficient obtained by regressing analysts' forecast errors, defined as realization minus forecasts, on the past forecast revision by the same analysts [1]. Under rational expectations, the forecast

[1] Time t forecast revision of a future variable X is defined as the time t forecast of X minus the time $t-1$ forecast of X.

error should be unpredictable on the basis of past information, so the estimated coefficient should be zero. Figure 1 shows that this is not the case in the data. Two features stand out. First, at any maturity the average analyst over-reacts to news: when she revises interest rates up, the realization falls below the new forecast, leading to a negative coefficient in 1. Second, over-reaction is stronger at longer maturities, namely the regression coefficient is more negative for $30y$ or $20y$ rates than for shorter horizons.

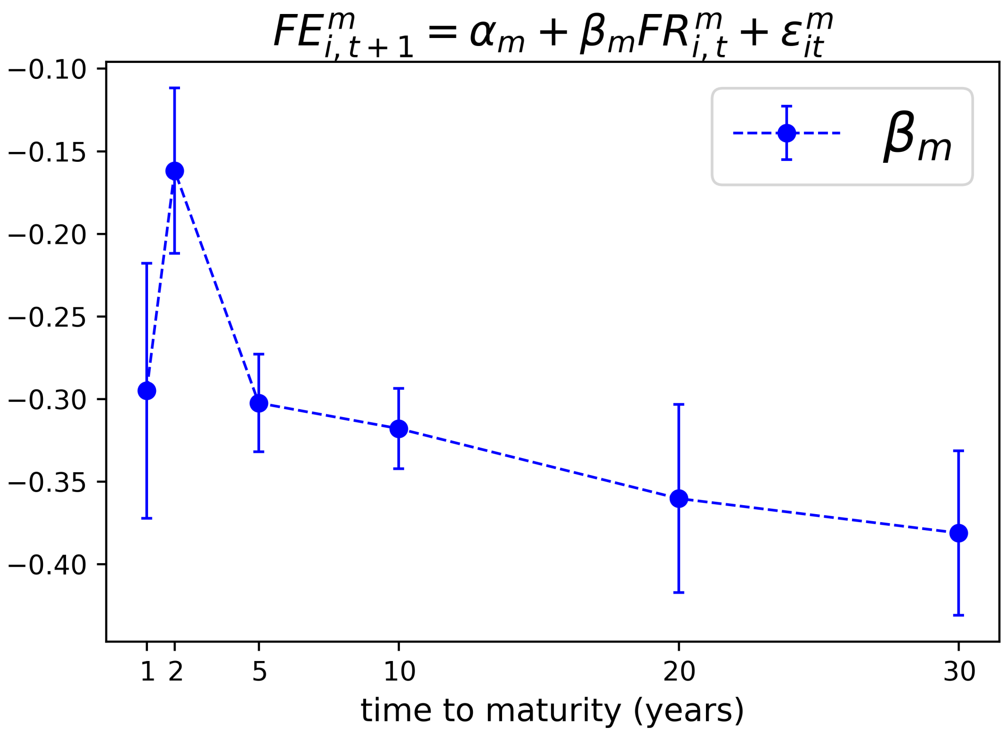


Figure 1: Pooled OLS coefficients obtained by regressing analysts' forecast errors about interest rates to maturity m on analysts' past forecast revisions of rates to the same maturity m. Data are at the monthly frequency. Interest rates are annualized.

This maturity increasing over-reaction intuitively resonates with the three term structure puzzles. If the long run beliefs of traders are more volatile than their short run beliefs, it may be possible that equilibrium long term rates end up being too sensitive, excessively volatile, and predictable. This conjecture raises four questions. First, how does this mechanism precisely work? Second, is maturity increasing over-reaction a robust feature of market beliefs? Third, if so, what is its psychological foundation? And fourth: can such foundation also quantitatively account for the puzzles? My analysis seeks to shed light on these questions.

Section 2 addresses the first question by introducing into an otherwise standard affine term structure setting investors' beliefs that over-weigh recent news, particularly so at longer maturities. I show that when investors have such beliefs, then equilibrium long term yields are too sensitive to short term yields, they are also excessively volatile compared to the latter, and excess bonds returns are predictable on the basis of past forecast revisions, more strongly so at longer maturities. The analysis also connects the severity of these anomalies to the magnitudes of the error predictability coefficients in Figure 1.

Consider the second question next: is maturity increasing over-reaction a robust feature of market beliefs? One potential criticism of Figure 1 is that it is based on survey data that captures only some maturities and is not necessarily representative of the beliefs of the marginal investor. To address this issue, in Sections 3 and 4 I extract from observed yields information about market beliefs using a method proposed by Ross (2015). This allows me to obtain an additional measure of beliefs, directly linked to market investors, that can be cross-validated with survey expectations. When I apply this method, I find that recovered beliefs differ in some ways from professional forecasts beliefs but they are strongly positively correlated with them. In particular, recovered beliefs appear to be tilded toward the beliefs of more accurate forecasters. These results suggest that recovered beliefs are not noise, and that arbitrage may reallocate capital to more rational investors (Buraschi et al. (2018)). As I consider the correlation between forecast revisions and forecast errors, recovered beliefs display some under-reaction at short maturities but crucially they confirm the key features of Figure 1: beliefs about long term interest rates over-react both in absolute terms and of course also relative to those about short term rates. Maturity increasing over-reaction is thus rubust to alternative measures of beliefs.

Ross (2015) methodology relies on restrictive assumptions, in particular it has been criticized for not allowing for persistent shocks to the stochastic discount factor, which are typical of certain asset pricing models (Borovička et al. (2016)). I show that even though this mis-specification may contaminate recovered beliefs, it does not contaminate the pattern of maturity increasing over-reaction. I perform also a large battery of robustness exercises with respect to sample period, frequency of the data, construction of forecast error and revision variables, specification of term structure model as well numerical state space discretization. These results, once again, confirm the pattern of Figure 1: maturity increasing over reaction.

Section 5 then micro-founds beliefs by adapting to a term structure setting the model of diagnostic expectations by Bordalo et al. (2018b). This model captures the psychology of Tversky and Kahneman (1974) representativeness heuristics, and in particular it embodies the "kernel of truth" property: beliefs move excessively in the direction of events that have become relatively more likely in light of the data. I show that in a term structure setting this feature naturally yields maturity increasing over-reaction. Any given signal is in fact relatively more informative for long term events, for which there is a larger baseline uncertainty. Thus, forecast revisions are more aggressive at long maturities. The model departs from the rational benchmark due to a single parameter, the degree of "diagnosticity" θ. Because this parameter has been estimated in previous work, it offers a benchmark for my analysis.

I conclude by calibrating the model and by assessing its ability to match the three puzzles. I find that: i) the diagnosticity distortion parameter calibrated from recovered beliefs is consistent with previous literature (Bordalo et al. (2018a),Bordalo et al. (2018b)), ii) such parameter matches fairly well the documented average over-reaction of analyst forecasts in Figure 1, iii) it accounts for roughly 80% of the excess volatility in interest rates documented by Giglio and Kelly (2018), iv) it account for roughly 40% of the excess sensitivity and v) my model implies that bonds excess returns should be negatively predictable from past revisions, more strongly so for larger maturities, which I successfully test in the data.

Related Literature

My paper contributes to a growing body of research aimed at assessing the explanatory power of non rational expectations using beliefs data. Bordalo et al. (2018a) finds evidence of over-reaction to information in several individual macroeconomic and financial time series. Maturity increasing over-reaction of interest rates is first detected in Bordalo et al. (2018a), where the authors compare the predictability of errors in individual and in consensus macro-financial forecasts. Piazzesi and Schneider (2011) finds that standard bonds market factors (level and slope) are perceived as more persistent by financial analysts relative to statistical analysis 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. Wang (2019) rationalizes maturity increasing reaction within a boundedly rational model. In this model, analysts benchmark the persistence of rates at maturity m to the average persistence across maturities, thus over-reacting at long maturities and under-reacting at short ones. Relative to these paper, I discipline beliefs based on the unifying theory of diagnostic expectations, which offers a valuable external benchmark, and shoe that it accounts, qualitatively and quantitatively, for the three leading yield curve puzzles.

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. Jensen et al. (2019) generalize the recovery theorem to non Markovian and non stationary settings. Qin et al. (2018) propose an empirical test for the degeneracy of the martingale component of the SDF for US treasury bonds and rejects 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 does not entail assumptions on the physical measure (Le et al. (2010)). Close to the intuition of the recovery theorem, Augenblick and Lazarus (2018) provide evidences of excess movements of stock market prices, which they document to be stronger for longer 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 maturity increasing over-reaction into an otherwise standard affine model of the term structure, and I show that in this economy the error predictability of Figure (1) and the excess volatility of Giglio and Kelly (2018), excess sensitivity and excess predictability naturally connect. 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 belief formation and I quantitatively assess the beliefs channel for bond market puzzles. In Section 6, I provide robustness checks. Section 7 concludes. All proofs are in Appendices.

2 Over-reaction to News and the Term Structure of Interest Rates

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 maturity increasing over-reaction and bond market puzzles.

Consider a simple one factor economy, where the factor is AR(1) under both the physical and the risk neutral measure. In appendix A, I discuss how these assumptions can be relaxed, and work out the multi-factor version needed for the empirical implementation.

There is 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 $r_{t,m} := \frac{1}{m} \sum_{i=0}^{m-1} r_{t+i}$. With deterministic interest rates, the yield obtained from investing at maturity m is the interest rate at that maturity: $y_{t,m} = r_{t,m}$. Suppose instead that the short rate r_t is a function of a risk factor X_s (e.g. GDP growth). Then, while at time t the current short rate r_t is known because X_t is also known, longer maturity rates $r_{t,m}$ depend on future values $\{X_{t+k}\}_{k=0}^{m-1}$ of the risk factor, which 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 states and by *risk aversion*. I now characterize these expectations, and in particular their departure from rationality. Next, I show how the same expectations affect – together with risk aversion – the yield curve in equilibrium.

2.1 Maturity Increasing Over-reaction

The short rate is an affine function of the risk factor X_t, which for simplicity I assume to follow a stationary AR(1):

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

where $\varepsilon_t^{\mathbb{P}}$ 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 $\varepsilon_{t,m}^{\mathbb{P}} := \frac{\delta_1}{m}\sum_{k=0}^{m-1}\sum_{i=1}^{k}(\rho^{\mathbb{P}})^{i-1}\sigma^{\mathbb{P}}\varepsilon_{t+i}^{\mathbb{P}}$. 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.

I allow expectations of future interest rates to over-react to the current state in a maturity dependent fashion. Formally, I 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 superscript 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 about future states may display excess sensitivity ($\psi_m^{\theta} \geq 0$) to the current state[2]. I allow the distortion ψ_m^{θ} to be maturity-dependent. For now, I leave such dependence unspecified. $\mathbb{P}^{\theta}$ denotes the distorted data generating process

[2] 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.

measure for the factor and the derivation of Equation (2) from the distorted factor dynamics is performed in Appendix A, Equation 25.

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

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

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

When $\psi_m^\theta = 0$, expectations are rational. When $\psi_m^\theta > 0$ agents over-ract to news. That is, they over-estimate future rates in states that are truly indicative of higher than average future rates: $\mathbb{E}_t^{\mathbb{P}}[r_{t,m}] > \mathbb{E}^{\mathbb{P}}[r_{t,m}]$. By contrast, agents under-estimate future rates in states truly indicative of lower than average future rates $\mathbb{E}_t^{\mathbb{P}}[r_{t,m}] < \mathbb{E}^{\mathbb{P}}[r_{t,m}]$.[3]

In Section 5, I show that under Gaussian noise, Equation (3) follows from the diagnostic expectations model of Bordalo et al. (2018b). Such model founds the distortion parameters ψ_m^θ as functions of a more primitive distortion parameter: the degree of diagnosticity θ. This is why I denote distorted expectations using θ.

2.2 Maturity Increasing Over-reaction and Bond Market Puzzles

Consider a market forming expectations according to Equation (3). What does the yield curve look like? To answer this question, one must also consider investors' risk aversion, which affects – together with beliefs – required rates of return. Risk aversion is captured by a stochastic discount factor (SDF) $M_{t,m}$ that discounts more heavily cash flows received in states in which the investor is poorer. Then, the price of a maturity m zero coupon bond,

[3] 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.

and hence its yield $y_{t,m}$, is pinned down by the discounted expected payoff[4]:

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

where the so called *risk neutral* probability measure $\mathbb{Q}^{\theta}$ over-weights states in which the investor is poor.[5] To use the machinery of affine term structure models, I 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:

$$r^{\theta}_{t,m} = \delta_0 + b^{\mathbb{Q}}_{m} X_t + \psi^{\theta}_{m} b^{\mathbb{P}}_{m} X_t + \sigma^{\theta}_{m} \varepsilon^{\mathbb{Q}}_{t,m},$$

where $b^{\mathbb{Q}}_{m} := \frac{\delta_1}{m}\sum_{i=0}^{m-1}(\rho^{\mathbb{Q}})^{i}$ and $\varepsilon^{\mathbb{Q}}_{t,m} := \frac{\delta_1}{m}\sum_{k=0}^{m-1}\sum_{i=1}^{k}(\rho^{\mathbb{Q}})^{i-1}\sigma^{\mathbb{P}}\varepsilon^{\mathbb{Q}}_{t+i}$. $\rho^{\mathbb{Q}} \neq \rho^{\mathbb{P}}$ is the risk neutral persistence of the factor under rational expectations, while $\varepsilon^{\mathbb{Q}}_{t+1}$ is a zero mean shock with risk neutral variance $\sigma^{\mathbb{Q}} \neq \sigma^{\mathbb{P}}$, again under rational expectations. $b^{\mathbb{Q}}_{m}$ is the risk neutral sensitivity of future rates to X_t and $\varepsilon^{\mathbb{Q}}_{t,m} := \frac{\delta_1}{m}\sum_{k=0}^{m-1}\sum_{i=1}^{k}(\rho^{\mathbb{Q}})^{i-1}\sigma^{\mathbb{Q}}\varepsilon^{\mathbb{Q}}_{t+i}$ is the risk neutral shock at maturity m, both under rational expectations. In sum, risk aversion in captured by a change in model parameters, particularly in the sensitivity of interest rates to X_t.

Importantly, under the belief distortions of Equation (2) the risk adjusted dynamics of interest rates still exhibit over-reaction (if $\psi^{\theta}_{m} > 0$) to the current state X_t.

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

[5]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).

Proposition 1. *Assume homoskedastic $\mathbb{Q}$-shocks. The yield curve under beliefs 2 equals:*

$$y^{\theta}_{t,m} = -\frac{1}{m}\log \mathbb{E}^{\mathbb{Q}^{\theta}}_{t}[e^{-m\cdot r_{t,m}}] = a^{\theta}_{m} + (b^{\mathbb{Q}}_{m} + \psi^{\theta}_{m} b^{\mathbb{P}}_{m})X_t, \tag{5}$$

where:

$$b^{\mathbb{Q}}_{m} = \frac{\delta_1}{m}\sum_{i=0}^{m-1}(\rho^{\ \mathbb{Q}})^i,$$
$$a^{\theta}_{m} = \delta_0 - \frac{1}{m}\log \mathbb{E}^{\mathbb{Q}}[e^{-m\cdot\sigma^{\theta}_{m}\varepsilon^{\mathbb{Q}}_{t,m}}].$$

When expectations are rational, namely $\psi^{\theta}_{m} = 0$, Equation (5) is the signature of *affine term structure models* (see Duffee (2013) for a review). The coefficients a^{θ}_{m} and $b^{\mathbb{Q}}_{m}$ are maturity dependent. Can the model allowing for ψ^{θ}_{m} explain the three bond market puzzles?

2.2.1 Excess Volatility

Interest rates volatility is shaped by the sensitivity $b^{\mathbb{Q}}_{m}$ of yields to changes in the factor X_t. The more persistent is the factor, the higher is $\rho^{\mathbb{Q}}$, the more sensitive is the interest rate at any given maturity to news, the higher is $b^{\mathbb{Q}}_{m}$. On the other hand, because the factor is stationary[6], $|\rho|^{\mathbb{Q}} < 1$, long term rates should be less sensitive than short term rates, namely $b^{\mathbb{Q}}_{m}$ declines with m. The Giglio and Kelly (2018) excess volatility puzzle is rooted in the maturity profile of the $b^{\mathbb{Q}}_{m}$ 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?[7] Consider, for each maturity m, the predictive regression of the forecast error of yields at maturity m, $FE^{\theta}_{t,m} := y^{\theta}_{t,m} - \mathbb{E}^{\theta}_{t-1}[y^{\theta}_{t,m}]$, on the forecast revision

[6] Stationarity of interest rates is an empirically established fact, see Giglio and Kelly (2018).

[7] 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 (2018) show that neither quadratic $\mathbb{Q}$-specifications, nor stationary long memory process nor regime switching models can 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.

$FR^{\theta}_{t-1,m} := \mathbb{E}^{\theta}_{t-1}[y^{\theta}_{t,m}] - \mathbb{E}^{\theta}_{t-2}[y^{\theta}_{t,m}]$:

$$FE^{\theta}_{t,m} = \alpha_m + \beta_m + FR^{\theta}_{t-1,m} + \varepsilon_{t,m}.$$

Coibion and Gorodnichenko (2015) showed that positive (negative) regression coefficients β_m - CG coefficients henceforth - capture under/(over)-reaction to news. The following result connects such error predictability with excess volatility.

Theorem 1. *(Over-reaction and excess volatility). Under beliefs (2) and the affine setting:*

i) (Increasing Over-reaction) the CG coefficients β_m are equal to:

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

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:

$$\frac{\mathbb{V}^{\mathbb{P}}[y^{\theta}_{t,m}]}{\mathbb{V}^{\mathbb{P}}[y^{\theta}_{t,1}]} = \left(\frac{b^{\mathbb{Q}}_m + \psi^{\theta}_m b^{\mathbb{P}}_m}{b^{\mathbb{Q}}_1 + \psi^{\theta}_1 b^{\mathbb{P}}_1}\right)^2,$$

where $\mathbb{V}^{\mathbb{P}}[y^{\theta}_{t,m}]$ 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^{θ}_m is positive and increases in maturity m.

This result conveys two important messages. First, maturity increasing over-reaction, 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) (VAR(1) in the multi-factor setting), 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^θ_m.

2.2.2 Excess Sensitivity

A large literature in macro-finance (Shiller (1979), Mankiw and Summers (1984), Gürkaynak et al. (2005), Hanson et al. (2018)) documents high sensitivity of long-term interest rates relative to short rate movements. In theory, since interest rates are stationary, long rates should barely comove with short rates. In reality, they comove quite a lot. This feature is dubbed excess sensitivity of long-rates, where the assumed benchmark is the expectation hypothesis. Here, I analyze the puzzle through the lens of an affine model[8]. The rational yield curve at time $t+h$ (for any $h \geq 0$) can be rewritten as:

$$y_{t+h,m} = a_m - \delta_0 + \frac{1}{\delta_1} \times \underbrace{(b^{\mathbb{Q}}_m)}_{\text{rational sensitivity}} \times y^\theta_{t,1} + \varepsilon^{\mathbb{P}}_{t,h}. \tag{6}$$

Since the factor is stationary (both under $\mathbb{P}$ and $\mathbb{Q}$), rational models predict a fast decay of the sensitivity of long rates to short rates along maturities, so that $b^{\mathbb{Q}}_m$ should be low for large maturities. Proposition 2 below links this puzzle to maturity increasing over-reaction.

Proposition 2. *Under beliefs (2) and the affine setting:*

$$y^\theta_{t+h,m} = a^\theta_m - \delta_0 + \frac{1}{\delta_1}\left(b^{\mathbb{Q}}_m + \underbrace{\psi^\theta_m b^{\mathbb{P}}_m}_{\textit{excess sensitivity}}\right) y^\theta_{t,1} + \sigma^\theta_m \varepsilon^{\mathbb{P}}_{t,h} \tag{7}$$

Over-reaction to news yields both the pattern in Figure (1) and excess sensitivity of long term rates if and only if ψ^θ_m is positive. Furthermore, maturity increasing over-reaction implies that excess sensitivity should increase with maturity.

The key information in Equation 7 is the term $\psi^\theta_m b^{\mathbb{P}}_m$: maturity increasing over-reaction over-weights the sensitivity of yields to the factor ($\psi^\theta_m > 0$) and more strongly so at larger

[8] Rational affine models are more general than the expectation hypothesis in that they can account for a class of risk premia.

maturities (ψ^θ_m increasing in m). Maturity increasing over-reaction contrasts mean reversion, via the coefficient ψ^θ_m: the observed persistence of long rates will then be larger than the one implied by the persistence of the short rate and a rational affine model.

2.2.3 Excess Bond Returns Predictability

Finally, do predictable forecast errors impact excess bond returns predictability? Consider the one year holding period rational excess bond return defined as:

$$xr_{t,t+h,m} := \log \frac{P_{t+h,m-h}}{P_{t,m}} - \log \frac{1}{P_{t,h}} = my_{t,m} - (m-h)y_{t+h,m-h} - hy_{t,h}.$$

Under rational affine models, the predictability of excess bond returns stems from the factor predictability (Duffee (2013)). Under the beliefs (2) however, the yield curve is also affected by forecast revisions, whose predictability adds up to the rational one.

Proposition 3. *Under the beliefs (2) and the affine setting:*

$$\mathbb{E}^{\mathbb{P}}_t[xr_{t,t+h,m}] - \underbrace{\mathbb{E}^{\mathbb{P}}[xr^\theta_{t,t+h,m}|X_t]}_{\textit{rational predictability}} = -\underbrace{\psi^\theta_{m-h}(m-h)\left(\mathbb{E}^{\mathbb{P}}_t[y^\theta_{t+h,m-h}] - \mathbb{E}^{\mathbb{P}}[y^\theta_{t+h,m-h}]\right)}_{\textit{excess predictability}}. \qquad (8)$$

Over-reaction to news yields negative excess predictability of excess bond returns at maturity m if and only if $\psi^\theta_m > 0$. Moreover, under maturity increasing over-reaction excess predictability grows in magnitude with the maturity m.

Excess predictability arises in that not only the factors, captured by $\mathbb{E}^{\mathbb{P}}[xr^\theta_{t,t+h,m}|X_t]$, but also revisions affect future yield curves. When revisions of future yields go up, then future bonds are under-priced, namely they yields lower returns. To the extent that revisions are predictable, future excess bonds returns will exhibit predictability beyond the one coming from factors only.

The proposition directly links the equilibrium yield curve derived in this section to excess bond return predictability to maturity m from revisions of yields to the same maturity. Excess bond returns predictability as a consequence of systematic errors has been found also in Cieslak (2018) and Wang (2019). Such predictability naturally derives from my theoretical framework. Furthermore, a distinct prediction I derive is the maturity increasing sensitivity of excess bonds returns to maturity m on maturity m revisions of yields.

Thus, maturity increasing over-reaction as suggested from Figure 1 can account for all puzzle. But is the pattern is Figure 1 robust? The next section addresses this question.

3 Beyond Survey Data: Ross Recovery Theorem and the Term Structure of 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[9]. Trough this exercise, I can assess the robustness of maturity increasing over-reaction, and obtain additional useful data from my quantitative evaluation.

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 $M_{t,m}$ is *path independent*, namely there exists a constant δ and a function z of the state variables such that the one period SDF can be written as:

$$M_{t,t+1} = \delta \frac{z(X_t)}{z(X_{t+1})}.$$

Assumption i) is an approximation, but of significant accuracy: 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 (2018) and Martin and Ross (2019)). Nevertheless, as I will discuss later, I perform extensive robustness on relaxations of it.

Assumption ii) is more controversial: Borovička et al. (2016) shows that the path independent assumption is not met in consumption based asset pricing models featuring permanent SDF shocks or long run risks. I later evaluate the robustness of my results to this criticism as well.

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

But consider for now how Ross's method works. Take 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 in my setting correspond to factor values). 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 SDF capturing the marginal rate of substitution between current consumption in i and future consumption in j. Due to risk aversion, M_{ij} over-weights bad states, attaching a higher price to securities that pay out in those state.

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:

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

where $\mathbb{P}^{\theta}_{ij}$ is the distorted transition probability from i to j. An econometrician observing Arrow Debreu prices may recover market beliefs, by performing an "inverse risk adjustment" to the former. Such recovered beliefs could be $\mathbb{P}$ if the market is rational, or $\mathbb{P}^{\theta}$ if it is not.

Ross has shown that such inverse risk adjustment can indeed be done under assumptions i) and ii) above.

Ross's method has been so far used under the assumption of rational expectations. However, as displayed in Figure 2, when market beliefs may not be rational, 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 perform on the recovered market beliefs some statistical tests of rationality, considering in particular the predictability of forecast errors of Figure 1. The outcome of this test tells me if I am in the top or bottom row of Figure (2).

$$\begin{tikzcd} \mathbb{P} \arrow[r, leftrightarrow, "\text{Risk}"] \arrow[d, dashed, leftrightarrow, "\text{Belief distortions}"'] & \mathbb{A} \arrow[d, dashed, leftrightarrow] \\ \mathbb{P}^{\theta} \arrow[r, leftrightarrow] & \mathbb{A}^{\theta} \end{tikzcd}$$

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 possibly distorted Arrow-Debreu prices $\mathbb{A}^{\theta}$.

3.1 Empirical implementation

Of course, to apply Ross' method, I need to know Arrow Debreu prices. These are not directly observed but they can be constructed 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:

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

where $\mathbb{Q}_{ij}^{\theta}$ 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 ($\theta \neq 0$).

Consider a conventional $\mathbb{Q}$-affine three factor model

$$\begin{cases} r_t = \delta_0 + \delta_1^{\top} \mathbf{X}_t \\ \mathbf{X}_t = \rho^{\mathbb{Q}^{\theta}} \mathbf{X}_{t-1} + \Sigma^{\mathbb{Q}^{\theta}} \varepsilon_t^{\mathbb{Q}^{\theta}}, \end{cases} \tag{9}$$

where the $\varepsilon_t^{\mathbb{Q}^{\theta}}$ shocks are i.i.d and bold notation refers to vectors and coefficients are restricted for identification, as carefully discussed in Appendix A, Estimation.

Given this model, parameters estimation proceeds as follow. First, I estimate δ_0 and δ_1 in 9 by OLS regressions of the short rate ($1y$) onto factors (built from the first three principal components of the variance-covariance matrix of the time series of the yield curves). Second, taking into account that beliefs may exhibit maturity dependent distortions, I estimate model

9 at each maturity independently. For each maturity m, I match observed yields to that maturity with the yields predicted by model 9 at the same maturity m. All the details of the procedure are discussed in Appendix A, Estimation.

The crucial conceptual point of the procedure I employ, which differs from the standard estimation of rational $\mathbb{Q}$-parameters (see Cochrane and Piazzesi (2009)), is the maturity by maturity matching of the observed yields. In this way, I do not impose the cross-maturity restrictions entailed by the profile of rational coefficients $b_m^{\mathbb{Q}}$. As shown in appendix A, estimated parameters from model 9 systematically differ across maturities, consistent with my approach.

Finally, in order to implement the recovery theorem, I need to discretize the state space, because the recovery theorem entails an eigenvalue problem for the Arrow-Debrau state price matrix. I use the Rouwenhorst method (Cooley (1995)), which is the state of art in discretizing auto-regressive processes. I can therefore construct, for each maturity m, a state price transition matrix $\mathbb{A}_{ij}^{\theta}$ defined over a finite 3D grid. Such state price transition matrix is the input of the recovery theorem.

3.2 Potential pitfalls

The validity of Ross' method relies on the accuracy of assumptions about i) the stationarity of the state variables and ii) the path independence of the SDF. I perform extensive robustness tests and show that my results survive to relaxations of those assumptions. I first discuss robustness with respect to the path independent SDF assumption. Next, I discuss robustness with respect to stationarity as well as with respect to other model misspecification issues.

3.2.1 The Borovička-Hansen-Scheinkman critique

When the SDF features persistent shocks,, Borovička et al. (2016) showed that the method of Ross (2015) does not recover market beliefs, instead it recovers beliefs adjusted for a martingale component.

Does such misspecification invalidate the detection of maturity increasing over-reacting beliefs that is central here? To answer this question, suppose that the SDF has a martingale component. Then, by applying Ross method I 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.

Theorem 2. *Consider the $\mathbb{Q}$-affine setting, non rational beliefs as in (2) 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 maturity by maturity using Ross recovered beliefs vary with maturity m.

ii) Expectations $\mathbb{E}^\theta[\cdot]$ with respect to the distorted measure $\mathbb{P}^\theta$, violates the law of iterated expectations (LIE).

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

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

where $\tilde{c} > 0$ 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 maturity dependent over-reaction. That is critical for my analysis. The intuition is that maturity dependent distortions entail a strong violation of rationality, namely a violation of the law of iterated expectations[10]. As a result, it cannot be accounted by any rational expectations models, including those featuring permanent SDF shocks. In this respect, the Ross recovery theorem remains useful for spotting maturity dependent beliefs distortions.

Another way of addressing the severity of the Borovička et al. (2016) misspecification is to systematically compare recovered beliefs with survey data. The latter beliefs are not contaminated with SDF of the marginal investor. Thus, they can be used to validate Ross

[10] 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).

recovered beliefs. This is the focus of Section 4, where I find evidences of a remarkable agreement.

3.2.2 Stationarity and Additional Model Misspecification Issues

Martin and Ross (2019) discuss the feasibility of the recovery theorem in the stationary environment of interest rates. An empirically relevant possibility is that the Markov chain driving the yield curve gets ”trapped” in some state with low escaping probability, such a zero or negative lower bound. When this is the case, the sample size - even if large at the daily frequency - may limit the full exploration of the state space. As a consequence, error predictability tests performed with recovered beliefs may be misspecified if the training sample where model 9 is estimated sharply differs from the testing one. A similar concern applies if persistent regime changes occur, perhaps due to changes in monetary policy regimes.

To address these issues, I systematically perform the analysis in different sub-samples and at different time frequencies. Results are shown in Section 6. The main finding is that the maturity increasing over-reaction pattern is remarkably robust. Different samples feature however different degree of maturity increasing over-reaction. In the post-2000 period, where rates are particularly low in level as well as particularly persistent, over-reaction is in magnitude smaller.

Crucially, this is consistent with my model (as discussed in in Section 5), as well as with the evidence of less excess sensitivity in such period discussed in (Hanson et al. (2018)).

Jensen et al. (2019) provide a generalization of Ross (2015) where both the assumptions of Markovianity and stationarity of the state variables are relaxed. Albeit stationarity is a plausible assumption in the context of the term structure of interest rates, Markovianity may fail and this misspecification may potentially lead the econometrician to identify time inconsistent distortions even in a rational expectations world.

Failure of Markovianity means that the yield curve depends on past values of the factors. To address this issue, my factor construction (on the training sample) is based on the use of principal components and as such it is completely agnostic about lag specifications. Moreover, when I validate the ability of factors to capture yields variations on the second half of the sample, I find that the factors still fit the yield curve with R^2 close to one as discussed in

Appendix A. This means that three factors capture the empirically relevant sources of risk.

Finally, It may also be the case that market beliefs are defined also over non-priced variables. Then, when I apply the recovery theorem, I recover the marginal distribution of beliefs over the limited state space of priced risky factors. In this case, systematic predictability of errors may be due to the use of the un-spanned factors as predictors. To address this possibility, in Section 6, I show that forecast errors are also predictable on the basis of past forecasts, which are spanned by the risky factors by construction, and I find once again maturity increasing over-reaction, thus ruling out this channel.

4 Recovered Beliefs, Survey data and Rationality tests

I now study recovered beliefs and compare them with professional forecasters beliefs. This helps me assess whether beliefs measurement capture systematic patterns, because the two sources of data are highly independent

4.1 Data

US treasury yields

Gürkaynak et al. (2007) provides (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[11]. I remove from the sample the forecasters who reply to the survey for a short time period (less then 5 years). At each time t, I remove the 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 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:

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

Bold notation is used for vectors: $\mathbf{X}_t$ contains level, slope and curvature factors. 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*:

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

[11]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.

For each maturity available, I then run the regression:

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

A positive regression coefficient at maturity m, $\beta_m > 0$, capture under-reaction to news: after positive revisions, errors are systematically predicted to be positive, meaning that forecasts reacts too little to news. A negative regression coefficient at maturity m, $\beta_m < 0$ capture over-reaction to news: after positive revisions, errors are systematically negative, meaning that forecasts move too much with news.

The CG coefficients obtained with the Ross recovered beliefs for maturities ranging in $2, 3, \dots 30$ years[12] and by pooled estimation of professional forecasters data are shown in Figure 3.

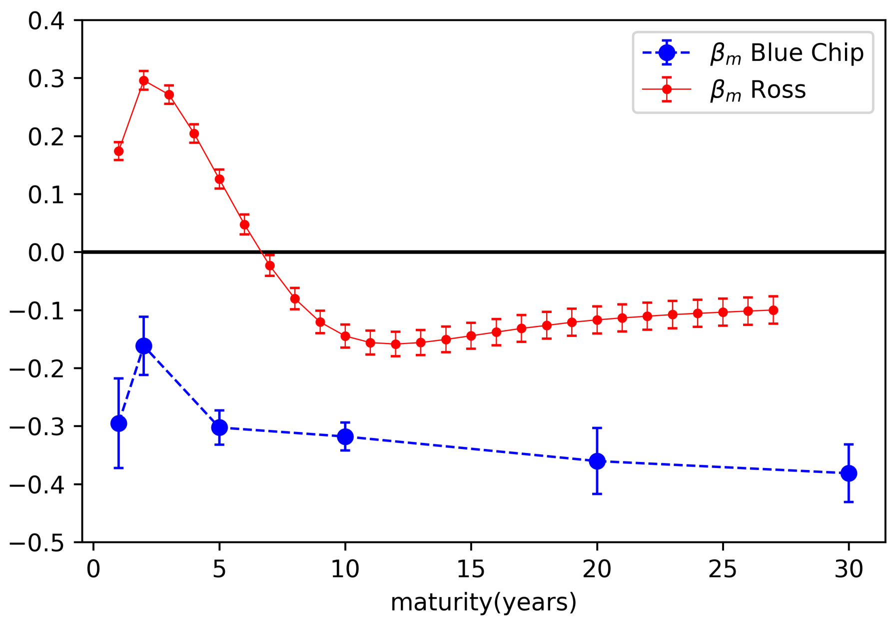


Figure 3: Slope of FE on FR using recovered beliefs (red) and using the Blue Chip dataset (pooled OLS). Confidence intervals are computed at 5%.

The main difference between the two datasets is that Ross recovered beliefs exhibit a

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

pattern of under-reaction to information at the short end of the yield curve. Professional forecaster data, on the contrary, exhibit over-reaction only [13] [14]. It is as if, in the recovered beliefs, the maturity increasing over-reaction is ”shifted up” relative to the Blue Chip forecasts. Crucially, however, both Ross recovered beliefs and the Blue Chip survey display maturity increasing over-reaction to information. Recovered beliefs and beliefs of professional forecasters over-react more for long term yields than for short term ones. Thus, the key property of excess reaction to information at long maturities relative to short ones is robust across measurement methods.

We can further 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 (on average 80% correlation of forecasts and 30% correlation of revisions)[15].

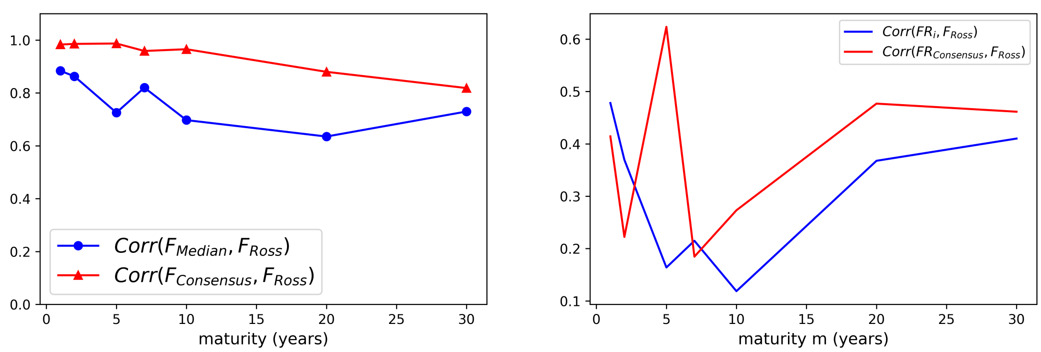


Figure 4: Left panel: correlation between mean forecasts (red triangle), median forecasts (blue circle) and Ross recovered forecasts as a function of the maturity. Right panel: correlation between mean forecast revision (red triangle), median forecast revision (blue circle) and Ross recovered forecast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

The qualitative similarity of predictability patterns across the two datasets is surprising:

[13] 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.

[14] 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.

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

there is no mechanical reason for it, which suggests that both recovered beliefs and Blue Chip data capture systematic patters. This also indicates that maturity increasing over-reaction is robust. 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.

One interesting question arises: where do the difference between the two measured beliefs comes from? One possibility is the misspecification of the SDF in the recovery theorem (Borovička et al. (2016)). This however cannot be the fully story: such misspecification cannot account for short term under-reaction.

Another reason for the difference is that Ross recovered beliefs capture market beliefs, and the marginal investor is potentially different from the average forecaster. In particular, forecasters may exhibit heterogeneous distortions and arbitrage may reallocate capital toward some of them.

To assess these possibilities, Figure (5) plots the estimated distribution of distortions (biases) using a Gaussian kernel density estimator.

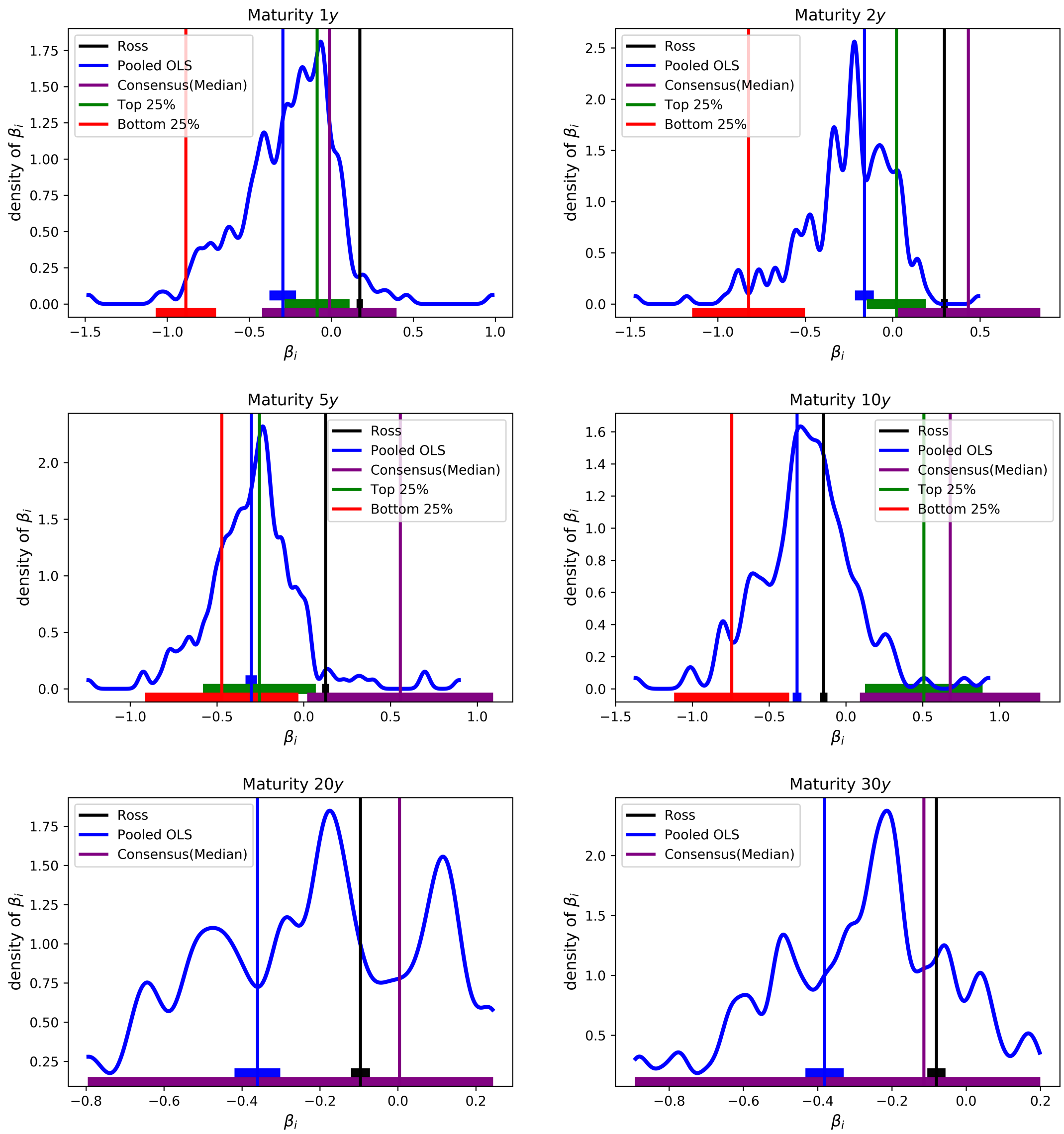


Figure 5: Density of forecasters distortions (CG coefficients) for different maturities. Distortions from recovered beliefs are shown in black, pooled distortions in blue, consensus (median) distortions in purple, best forecasters distortions (top quartile according to the mean square error criterion)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

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 figure also reports as a benchmark the CG coefficients 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[16].

Crucially, 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. This suggests that the Ross recovered beliefs display over-reaction as the average forecaster, yet they weight more heavily on 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.

This can also be seen by looking at the accuracy of forecasters, as measured by the mean square error in prediction. 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.

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).

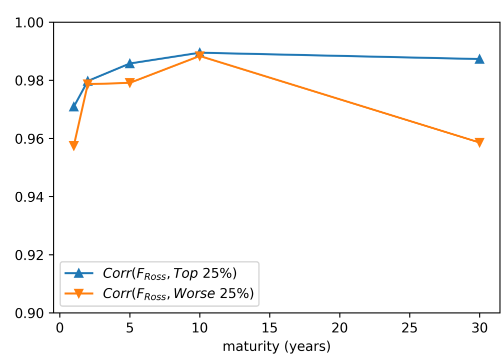


Figure 6: Correlation between mean forecast of top 25% forecasters (ranked according to the MSE criterion) and Ross recovered forecasts (blue triangle), Correlation between mean forecast of worse 25% forecasters (ranked according to the MSE criterion) and Ross recovered forecasts (orange triangle).

[16]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 (6) suggests that the Ross recovered forecasts slightly over-weight more accurate views. This suggests that the discrepancy between recovered beliefs and professional forecaster may then be due to arbitrage capital moving toward better forecaster, impacting more market prices.

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 may help reduce the impact of highly biased forecasters on asset prices, consistent with basic asset pricing theory. Maturity increasing over-reaction is thus robust and, as showed in Section 2, can quantitatively account for the puzzles. But can it so also quantitatively?

5 The Horizon Dependent Diagnostic Expectations Model

I now micro-found ψ_m^θ using the diagnostic expectations model of Bordalo et al. (2018b), I rationalize maturity-increasing over-reaction using a single parsimonious departure of beliefs from rationality, captured by the diagnosticity parameter θ. I then estimate θ and quantify the explanatory power of the model for the bond market puzzles. The benefit of founding ψ_m^θ is that there are external estimates of θ to which my results can be benchmarked.

5.1 Diagnostic Expectations

Diagnostic expectations are based on (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 heuristic has been widely documented in the psychology and cognitive science literature, since the seminal work of Tversky and Kahneman (1974). 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, $r_{t+m} = \frac{1}{m}\sum_{i=1}^{m-1} r_{t+i}$:

$$f_{\mathbb{P}^\theta}(r_{t,m}|\mathbf{X}_t) \propto f_{\mathbb{P}}(r_{t,m}|\mathbf{X}_t)\left(\frac{f_{\mathbb{P}}(r_{t,m}|\mathbf{X}_t)}{f_{\mathbb{P}}(r_{t,m})}\right)^\theta, \tag{10}$$

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 true 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 of 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 over-weighted in beliefs.

Equation 10 differs in two ways from the specification in 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 shown in Gennaioli and Shleifer (2018). I now show that allowing for this role of uncertainty is highly relevant here, because it implies a violation of the law of iterated expectations that is key to account for 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:

$$\begin{cases} r_t = \delta_0 + \delta_1^\top \mathbf{X} \\ \mathbf{X}_t = \rho^{\mathbb{P}} \mathbf{X}_{t-1} + \Sigma^C \varepsilon_t^{\mathbb{P}}, \end{cases} \tag{11}$$

where $\mathbf{X}_t \overset{\mathbb{P}}{\sim} VAR(1)(\rho^{\mathbb{P}}, \Sigma)$, Σ^C is a lower triangular matrix, $\varepsilon_t^{\mathbb{P}}$ 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 $\mathbb{P}$. I has assumed this also in Section 2, but as discussed there and shown in Appendix A, this was not necessary for the results of Section 2, Section 3 and Section 4.

Under this this assumption, 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, $r_{t,m}$, $f_{\mathbb{P}}^{\theta}(r_{t+m}|\mathbf{X}_t)$, *is Gaussian, with mean:*

$$\mathbb{E}_t^{\mathbb{P}^\theta}[r_{t,m}] = \mathbb{E}_t^{\mathbb{P}}[r_{t,m}] + \psi_m^\theta \left(\mathbb{E}_t^{\mathbb{P}}[r_{t,m}] - \mathbb{E}^{\mathbb{P}}[r_{t,m}] \right) \tag{12}$$

and variance:

$$\mathbb{V}_t^{\mathbb{P}^\theta}[r_{t,m}] = \left(\frac{\theta+1}{\mathbb{V}_t^{\mathbb{P}}[r_{t,m}]} - \frac{\theta}{\mathbb{V}^{\mathbb{P}}[r_{t,m}]} \right)^{-1}, \tag{13}$$

where

$$\psi_m^\theta := \frac{\theta \frac{\mathbb{V}_t^{\mathbb{P}}[r_{t,m}]}{\mathbb{V}^{\mathbb{P}}[r_{t,m}]}}{1 + \theta - \theta \frac{\mathbb{V}_t^{\mathbb{P}}[r_{t,m}]}{\mathbb{V}^{\mathbb{P}}[r_{t,m}]}}, \tag{14}$$

$$\mathbb{E}_t^{\mathbb{P}}[r_{t+m}] = \frac{1}{m} \left(\delta_0 + \delta_1^\top \left(\sum_{i=0}^{m-1} \rho^{\mathbb{P}^i} \right) \mathbf{X}_t \right), \tag{15}$$

$$\mathbb{E}^{\mathbb{P}}[r_{t,m}] = \delta_0, \tag{16}$$

$$\mathbb{V}_t^{\mathbb{P}}[r_{t,m}] = \frac{1}{m^2}(\delta_1)^\top \left(\sum_{i=0}^{2} \left(\sum_{i=0}^{m-2} \rho^{\mathbb{P}^{2i}} \Sigma \right) \right) \delta_1 \tag{17}$$

and

$$\mathbb{V}^{\mathbb{P}}[r_{t,m}] = \mathbb{E}[\mathbb{V}_t^{\mathbb{P}}[r_{t,m}]] + \mathbb{V}[\mathbb{E}_t^{\mathbb{P}}[r_{t,m}]] \tag{18}$$

$$= \mathbb{V}_t^{\mathbb{P}}[r_{t,m}] + \frac{1}{m^2}(\delta_1)^\top \left(\sum_{i=0}^{m-1} \rho^{\mathbb{P}^i} \right) \Sigma \left(\sum_{i=0}^{m-1} \rho^{\mathbb{P}^i} \right) \delta_1 \tag{19}$$

As evident from Equation (12), the diagnostic model endogeneizes the maturity increasing over-reaction assumed in reduced from in Section 2. The coefficient ψ_m^θ in formula (14), 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 start 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 more prominent, so they more easily come to mind after information makes them more likely. In this sense, the likelihood ratio in Equation 10 become larger in the tail whose likelihood increases for larger m.

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 over-weighted. 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 Theorem (1):

$$y_{t,m}^\theta = a_m^\theta + (b_m^{\mathbb{Q}})^\top \mathbf{X}_t + \psi_m^\theta (b_m^{\mathbb{P}})^\top \mathbf{X}_t, \tag{20}$$

where however, given the realistic multi-factor setting, $b^{\mathbb{P}}_m$ and $b^{\mathbb{Q}}_m$ are now vectors, with different entries for different factors.

The profile of distortion coefficients ψ^{θ}_m in disciplined in two dimensions. First, coefficients depend on a single scalar parameter θ, which make the diagnostic affine model parsimonious. Second, their shape depend on the underlying data generating process, a three factor Gaussian VAR(1). This discipline provides me with testable implications for the distortion coefficients ψ^{θ}_m.

5.2 Calibration

Theorem (1) links excess volatility to predictable forecast errors, which also suggests two independent routes to estimate the same primitive parameter θ. First, I can retrieve θ using the profile of CG coefficients β_m, obtained with recovered beliefs. To do so, I minimize the square distance between the CG coefficients implied by the affine diagnostic model and the observed ones:

$$\hat{\theta} = \min_{\theta > -1} \sum_m \left(\hat{\beta}_m - \beta_m \right)^2,$$

where $\hat{\beta_m}$ are the CG coefficients obtained from recovered beliefs, while β_m are the diagnostic ones, computed using Theorem 1.

Second, I can retrieve θ using the realized term structure of variance. To do so, I minimize the square distance between the realized term structure of variance $\hat{\mathbb{V}}[y^{\theta}_{t,m}]$ and the one implied by the diagnostic affine model $\mathbb{V}[y^{\theta}_{t,m}]$.

$$\hat{\theta} = \min_{\theta > -1} \sum_m \left(\frac{\hat{\mathbb{V}}[y^{\theta}_{t,m}]}{\mathbb{V}[y^{\theta}_{t,m}]} \right)^2.$$

Results are shown in Table 1.

	Match FE predictability	Match Ex-Volatility
$\hat{\theta}$	0.70	0.47

Table 1: Calibration of the diagnosticity θ, by matching: the observed FE predictability (left); the observed excess volatility (right).

The two estimates differ, but they are close to each other. Most important, they are remarkably close to previous estimates of parameter θ obtained using different data and different methodologies. The following figure show a comparison with Bordalo et al. (2018a), Bordalo et al. (2019) and Bordalo et al. (2018b).

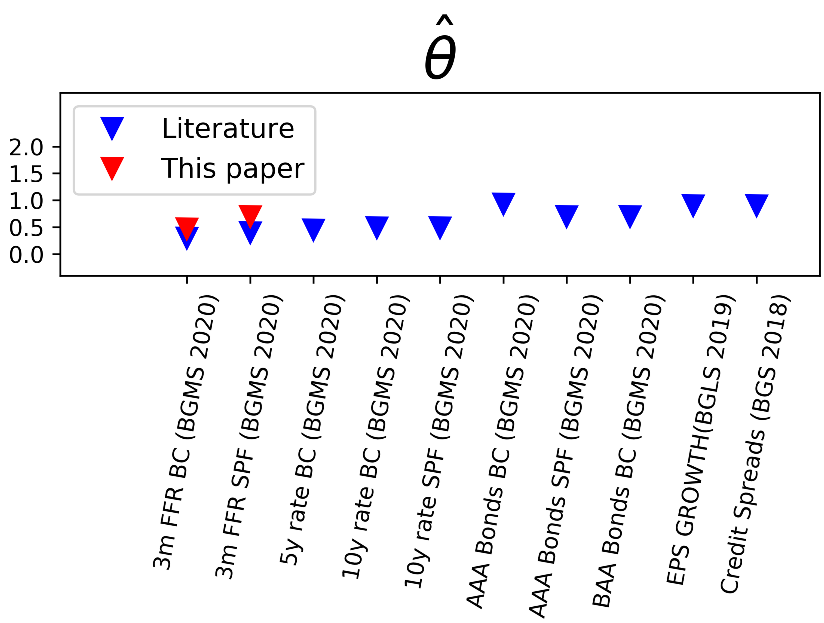


Figure 7: Estimated θ from ex-volatility and from FE predictability (red). Estimated θ in Bordalo et al. (2018a) (blue), which relies in individual time series specific modeling assumptions and simulated method of moments estimation, in Bordalo et al. (2019) and in Bordalo et al. (2018b).

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 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.

So far I estimated θ in sample, matching in one case forecast error predictability and in the other case the excess volatility of long rates. How does the estimate perform out of sample? Can this account for the bond market puzzles? I start with the excess volatility puzzle (Giglio and Kelly (2018)).

5.3 Excess Volatility Puzzle

To assess the ability of the model to account for the excess volatility puzzle (Giglio and Kelly (2018)), I simulate, using both estimated values of θ, the predicted term structure of volatility as well as the predicted CG (error predictability) curve, which are linked by Theorem 1.

Figure (8) reports the term structure of volatilities of yields at different maturities obtained under a three factor affine rational model[17], namely a counter-factual model setting $\theta = 0$, together with the variance of yields implied by the same model in which I plug the estimated $\theta \approx 0.47$ and $\theta \approx 0.7$ (Table 1). Figure 8 also reports the measured term structure of volatilities.

[17]The variance of rational yields is estimated by fitting the short end of the yield curve as in Giglio and Kelly (2018). I used the methodology of Section 3 and imposed consistency across short maturities, namely: $\hat{\rho}^{\mathbb{Q}} := \arg\min_{\rho^{\mathbb{Q}}} \left(\frac{1}{T\bar{M}} \sum_{t,m} (y_{t,m} - \bar{y}_{t,m})^2 \right)$. $\bar{M}$ 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.

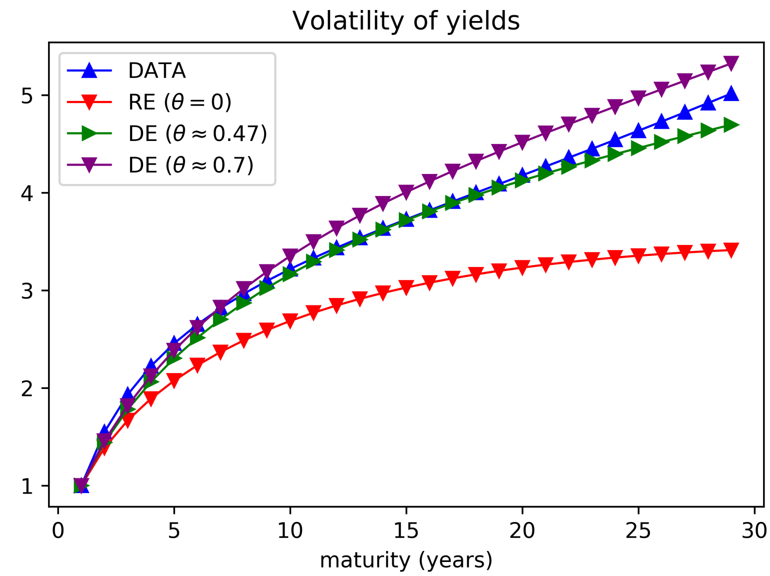


Figure 8: Excess volatility in the data (blue up triangle) versus the fit of a rational expectations affine model (red down triangle) and the diagnostic expectations affine model with $\theta \approx 0.47$ (green right triangle) and $\theta \approx 0.7$ (purple down triangle).

The green curve matches the observed volatility, which is the blue curve. Because $\theta = 0.47$ was obtained by fitting the blue curve, the overlap between the green and the blue curves shows that the fitting ability of the model is good The purple curve shows that the fitting ability is high even if θ is estimated from the CG coefficients. This indicates that there is consistency between excess volatility and CG coefficients, as predicted by Theorem 1. Hence, the error predictability can greatly account for the Giglio and Kelly (2018) fact.

Specifically, considering $\theta \approx 0.7$, obtained by fitting error predictability, shows that more than 80% of the excess volatility puzzle of Giglio and Kelly (2018) can be captured by the maturity increasing over-reaction of the diagnostic affine model. This is sizable, particularly because $\hat{\theta} = 0.7$ is not found by matching the volatility gap.

I also perform the reverse analysis: I use $\hat{\theta} = 0.47$ in Tables 1 obtained by fitting excess volatility and see how it accounts for forecast error predictability both in Blue Chip dataset and for recovered beliefs. Now $\hat{\theta} = 0.70$ is model fitting, while $\hat{\theta} = 0.47$ is used for external validation.

Figure 9 shows the implied CG coefficients from the diagnostic affine model as predicted by Theorem 1, for both values of θ, together with the error predictability patter detected in Blue Chip data as well in with recovered beliefs. While the short end of predictability of recovered beliefs (under-reacting) cannot be matched in this framework, the decreasing profile of error predictability is well captured by the diagnostic affine model.

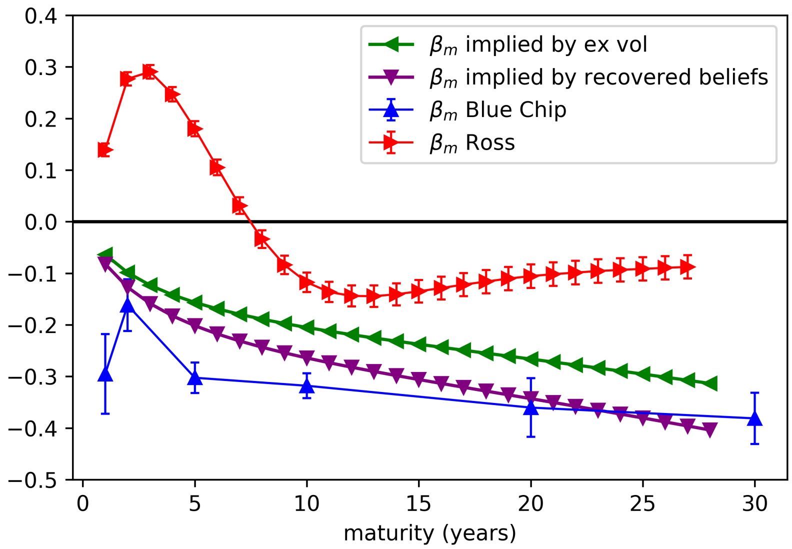


Figure 9: CG coefficients from Blue Chip data (blue up triangle), recovered beliefs (red right triangle) and implied by the calibrated diagnostic expectation model for $\hat{\theta} \approx 0.47$ (green left triangle) and $\hat{\theta} \approx 0.7$ (purple down triangle).

The green curve in Figure 9 show that the fit is poor at the short end, while the slope is well captured. The purple curve is quite aligned with the green, again capturing the consistency of excess volatility, used to fit $\theta = 0.47$ and maturity increasing over-reaction, used to fit $\theta = 0.7$, as predicted by Theorem 1. The model reproduced well quantitative analysts' expectations, as well as excess volatility. I next turn to excess sensitivity and excess bond returns predictability.

5.4 Excess Sensitivity of Long Term Interest Rates

Recall from Proposition 2 the excess sensitivity pattern predicted by the diagnostic affine model:

$$y^{\theta}_{t+h,m} = a^{\theta}_{m} - \delta_0 + \frac{1}{\delta_1}\left(b^{\mathbb{P}}_{m} \quad \underbrace{+\psi^{\theta}_{m} b^{\mathbb{P}}_{m}}_{\text{excess sensitivity}}\right) y^{\theta}_{t,1} + \sigma^{\theta}_{m}\varepsilon^{\mathbb{P}}_{t,h}.$$

I regress for each maturity m the observed yields to maturity m at time $t+h$, $y^{\theta}_{t+h,m}$, on the current short rate $y^{\theta}_{t,1}$, for h equal to one month[18]. Call the OLS coefficients of such regression η_m. In Figure 10, the blue curve plots the coefficients η_m. These coefficients provide an estimate of $b^{\mathbb{Q}}_{m} + \psi^{\theta}_{m} b^{\mathbb{P}}_{m}$, as Proposition 2 implies. The orange curve in Figure 10 shows the rational benchmark, namely the η_m coefficients under a one factor rational model. To be computed, such theoretical coefficients only relies on the estimation of $\rho^{\mathbb{P}}$.

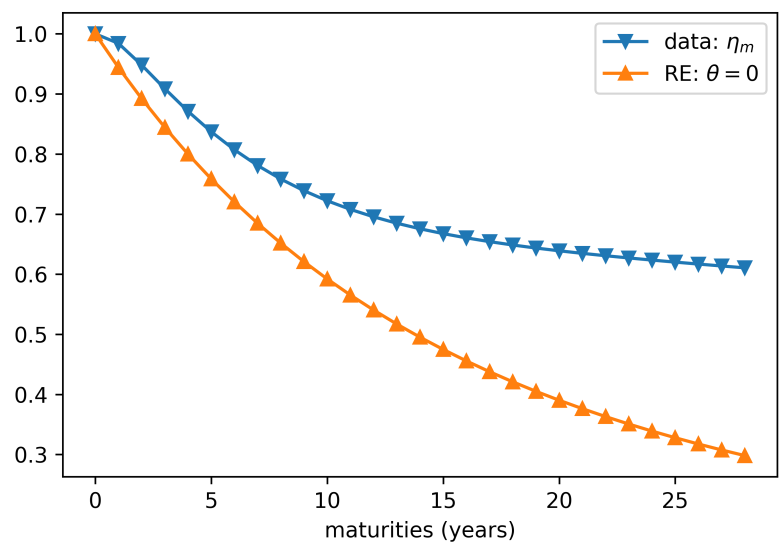


Figure 10: Monthly frequency OLS coefficients of yields to maturity m to the $1y$ yield, as a function of the maturity (blue down triangles). OLS coefficients implied by the measured persistence of the short rate (orange up triangles).

The orange curve lies below the blue curve at all maturities, meaning that interest rates are excessively sensitive to short rate movements. Moreover, the excess sensitivity gap increases

[18]Results are unchanged in the range oh $h = 1d, 1m, \ldots, 3m$.

across maturities. Maturity increasing over-reaction may help bridge the gap between the orange and the blue curve, because the decrease of persistence $(\rho^{\mathbb{P}})^m$ at maturity m may be at least in part offset by maturity increasing over-reaction..

To quantitatively assess this hypothesis, I compute the theoretical OLS coefficients

$$\eta_m = \frac{Cov(y^{\theta}_{t+1,m}, y^{\theta}_{t,m})}{\mathbb{V}[y^{\theta}_{t,1}]}$$

[19] from my estimated three factor affine model. In Figure 11, the orange curve reports the calibration in the rational expectations case, namely $\theta = 0$, showing once again, excess sensitivity relative to the data (blue curve). The green curve reports the calibration under diagnostic expectations, with $\theta = 0.47$, which was used to match excess volatility (Table 1). The red curve is calibrated using $\theta = 0.7$, which was used to match CG coefficients (Table 1).

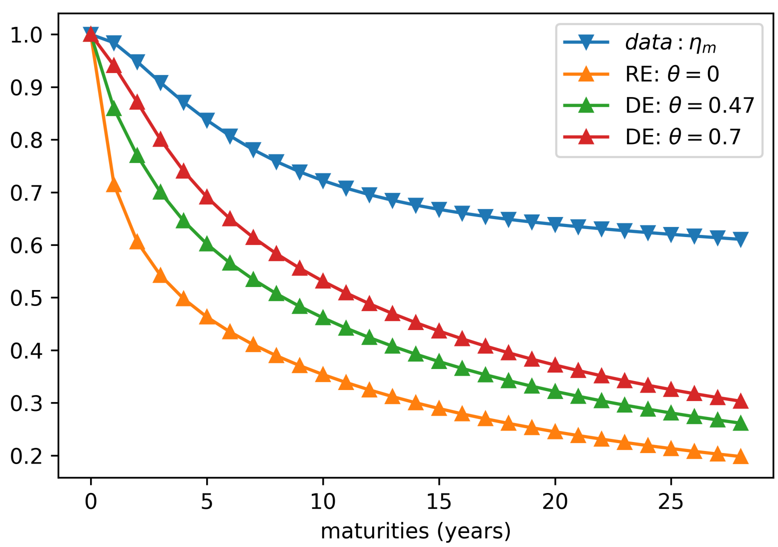


Figure 11: Monthly frequency OLS coefficients of yields to maturity m onto the 1y yield (blue curve). Theoretical counterparts under $\theta = 0$ (orange curve), $\theta = 0.47$ (green curve) and $\theta = 0.7$ (red curve)

Diagnostic expectations help get closer to the actual sensitivity. Using $\theta = 0.7$, which is the same value used to quantify the explanatory power of diagnostic expectations for the

[19]The calendar time $t + 1$ is measured in months, while maturity m in years.

excess volatility puzzle, the model performance is better at short maturities. It explains 60% of the excess sensitivity at $m = 2y$ and 42% at $m = 5y$. Performance is worse at larger maturities: 30% at $m = 30y$ and 26% at $m = 30y$. On average, across maturities, it accounts for roughly 40% of the excess sensitivity.

I also consider the contemporaneous association of yield changes (Hanson et al. (2018)). For each maturity m, I regress changes in yields to maturity m on changes in the short rate:

$$y^{\theta}_{t+h,m} - y^{\theta}_{t,m} = \gamma_{0,h,m} + \gamma_{1,h,m}\left(y^{\theta}_{t+h,1} - y^{\theta}_{t+h,1}\right) + \varepsilon_{t,h,m}. \tag{21}$$

The blue curve in Figure 12 shows the OLS estimated $\gamma_{1,h,m}$ coefficients, for $h = 1m$[20]. The orange curve plots the theoretical OLS coefficients $\gamma_{1,h,m}$, computed under the rational affine term structure model ($\theta = 0$). Calibrations of the diagnostic affine model are plotted for $\theta = 0.47$ (green curve) and $\theta = 0.7$ (red curve). Figure 12 shows again excess sensitivity, particularly strong for large maturities.

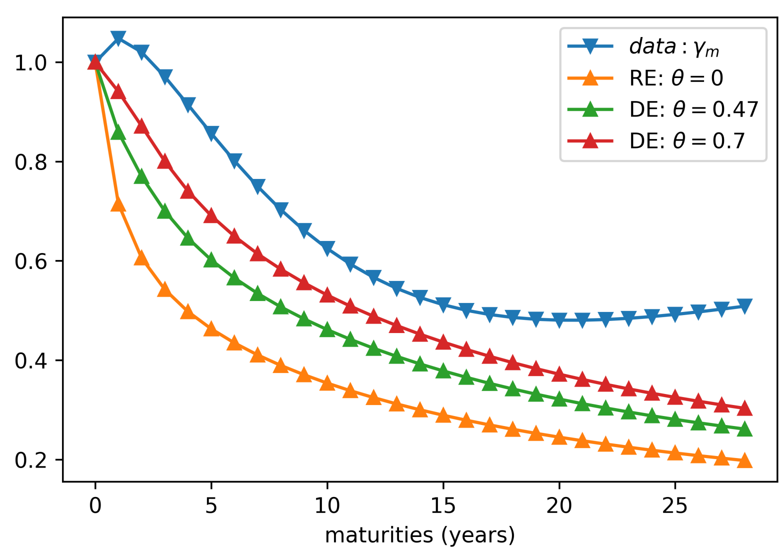


Figure 12: Blue: monthly frequency OLS coefficients of changes in yields to maturity m, $y^{\theta}_{t+h,m} - y^{\theta}_{t,m}$ onto changes in $1y$ yields, $y^{\theta}_{t+h,1} - y^{\theta}_{t,1}$ for $h = 1m$. Theoretical counterparts under $\theta = 0$ (orange curve), $\theta = 0.47$ (green curve), $\theta = 0.7$ (red curve)

Consider for a fixed maturity m, the coefficients $\gamma_{1,h,m}$ as a function of the time window h

[20] Results are similar in the range oh $h = 1m, \ldots, 3m$.

over which yields changes are computed. The diagnostic affine model (as well as the rational) predicts those coefficients to be constant. Hanson et al. (2018), however, shows that this is the case in the pre-2000 sample, while a stark decoupling across frequencies occurs in the post-2000 sample. At high frequency (e.g. $h = 1$ days) Hanson et al. (2018) finds an association between changes is short and long rates much higher then the one in the pre-2000 period and markedly decaying as a function of h. The authors shows that in the post-2000 sample, past changes is short rates are particularly strongly associated with future future long rates, that temporarily over-react. This causes higher sensitivity at high frequencies, followed then by predictable reversals. The authors point to non Markovianity effects at play in the post-2000 sample, which are however ruled out by the specification of beliefs in Equation 2. Theoretical extensions of beliefs in Equation 2 which account for non Markovianity are worked out in Section 6. Those are potentially consistent with the post-2000 fact found in Hanson et al. (2018).

Overall diagnostic expectations of the form of Equation 2 help rationalizing part of the excess sensitivity puzzle.

5.5 Excess bond returns predictability

What are implications of predictable forecast errors for the predictability of bonds excess returns? Consider Proposition 3 of Section 2:

$$\mathbb{E}_t^{\mathbb{P}}[xr_{t,t+h,m}] - \underbrace{\mathbb{E}^{\mathbb{P}}[xr^{\theta}_{t,t+h,m}|X_t]}_{\text{rational predictability}} = -\underbrace{\psi^{\theta}_{m-h}(m-h)FR_t[y^{\theta}_{t+h,m-h}]}_{\text{excess predictability}} \tag{22}$$

To bring this to the data, consider the relation between true and distorted forecast revisions, coming from beliefs in Equation 2:

$$FR^{\theta}_t[y^{\theta}_{t+h,m-h}] = (1+\psi^{\theta}_{m-h})FR_t[y^{\theta}_{t+h,m-h}]. \tag{23}$$

Inserting Equation 23 into 22, I get:

$$\mathbb{E}_t^{\mathbb{P}}[xr_{t,t+h,m}] - \underbrace{\mathbb{E}^{\mathbb{P}}[xr^{\theta}_{t,t+h,m}|X_t]}_{\text{rational predictability}} = \underbrace{-\frac{\psi^{\theta}_{m-h}}{1+\psi^{\theta}_{m-h}}(m-h)FR_t[y^{\theta}_{t+h,m-h}]}_{\text{excess predictability}}. \tag{24}$$

Equation 24 can be by using analysts' beliefs. I regress realized excess bond returns for one year holding period on past forecast revisions, computed from the Blue Chip dataset and from the median forecaster, controlling for past values of the three factors (principal components). Table 2 shows the results of such predictability test.

	$r^{\theta}_{t,t+1y,2}$	$r^{\theta}_{t,t+1y,5}$	$r^{\theta}_{t,t+1y,10}$	$r^{\theta}_{t,t+1y,30}$
FR^2_t	-3.52***			
	(0.65)			
FR^5_t		-10.54***		
		(2.61)		
FR^{10}_t			-13.57*	
			(7.16)	
FR^{30}_t				-18.01
				(22.94)
const	0.05	0.04	0.27	-3.55
	(0.13)	(0.51)	(1.38)	(4.32)
$PC_{1,t}$	-0.06	0.95	19.64	10.65
	(1.43)	(4.88)	(17.92)	(55.97)
$PC_{2,t}$	-1.09***	-2.30***	-1.81	15.30**
	(0.23)	(0.76)	(1.87)	(5.94)
$PC_{3,t}$	-0.30***	-0.29	0.48	0.06
	(0.06)	(0.21)	(0.43)	(1.24)
R^2	0.32	0.10	0.05	0.19
N	336	336	278	146

Table 2: Excess Bond returns predictability. $r^{\theta}_{t,t+1y,2}$ denotes the one year holding excess returns (from time t to time $t+1y$), relative to the one year yield. Excess returns for higher maturities are similarly defined. FR^2_t is the time t revision of next quarter interest rates to maturity 2, relative to the prediction done with information available in the past quarter. Similarly, revisions are defined for larger maturities. HAC standard errors are computed with 5 legs.

The coefficients are all negative and decreasing, which support the maturity increasing over-reaction mechanism of Proposition 3.

To assess the ability of diagnostic expectations to quantitatively capture excess bond returns predictability, I compute the regression coefficients implied by my three factor diagnostic affine model, where I use $\theta = 0.7$, as for the assessment of excess volatility and for that of excess sensitivity. Results are reported in the following Table 3.

maturity m	$2y$	$5y$	$10y$	$30y$
predicted OLS	-1.18	-2.40	-5.40	-17.50

Table 3: Predicted OLS coefficients $\frac{\partial r^{\theta}_{t,t+1,m}}{\partial FR^{m}_{t}}$, from the estimated three factor model and using $\theta = 0.70$.

Calibrated values are smaller but close in size and slope the the observed one. The prediction is good at $m = 30y$, at $m = 10y$ it captures 40% of excess predictability, at $m = 2y$ it captures 34%, worse at $m = 5y$.

In the robustness section, I test that revisions affect yields through forecast errors. I regress excess bond returns predictability on fitted errors from past revisions, and I obtain consistent results. This strongly connects the maturity increasing error predictability of figure 1 and the excess bond returns predictability shown in Table 2: the channel of revisions predicts both.

I also run predictive regressions using recovered beliefs. Here results are smaller but significant, as shown in Section 6. Additional forces may drive the gap. First, factors used in those regressions are estimated from yields principal components, which are an endogenous quantity also capturing some but not all revisions in yields. Second, while Ross forecasts are spanned by current and past factor, forecasters predictions may rely of different predictors. Both mechanisms imply that predictability from recovered beliefs is expected bo the lower in magnitude.

Summing up, the diagnostic affine model, a psychologically founded and parsimonious extension of standard affine pricing models improve our understanding as well as the fitting of the three leading puzzles of excess volatility of long term rates, excess sensitivity of long term rates and excess bond returns predictability.

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 (13) 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.

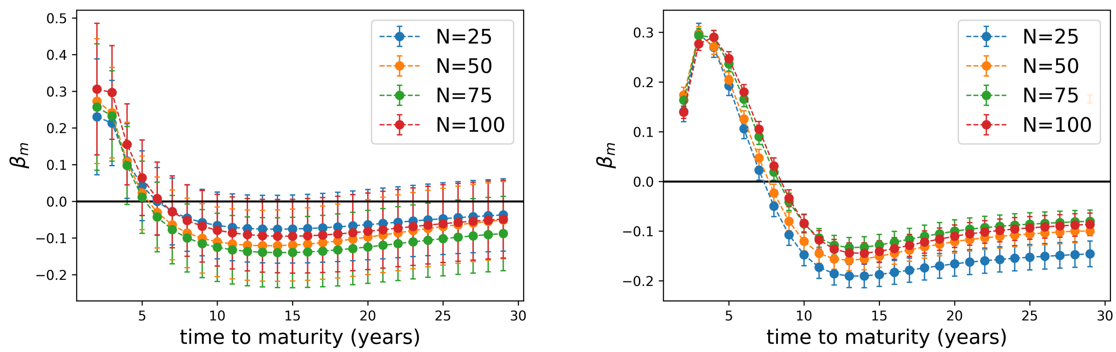
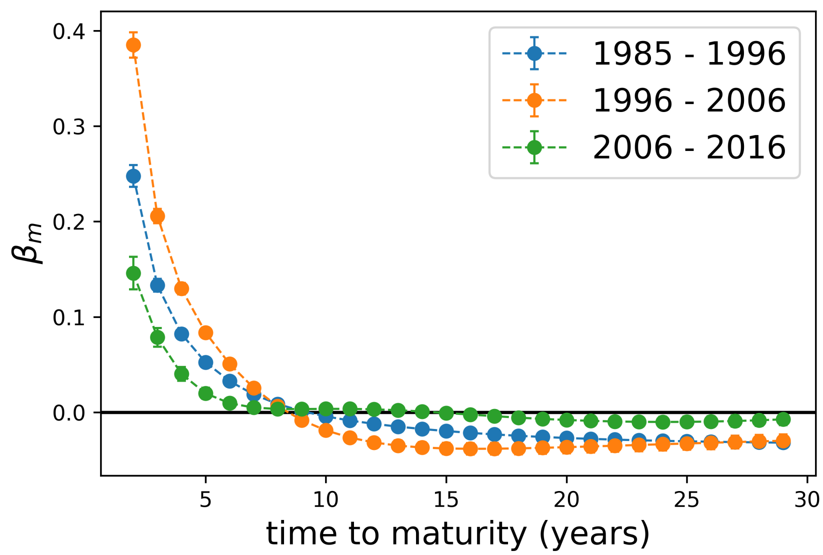


Figure 13: Left panel: CG coefficients at the monthly frequency. Right panel: CG coefficients at the daily frequency.

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 below 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.

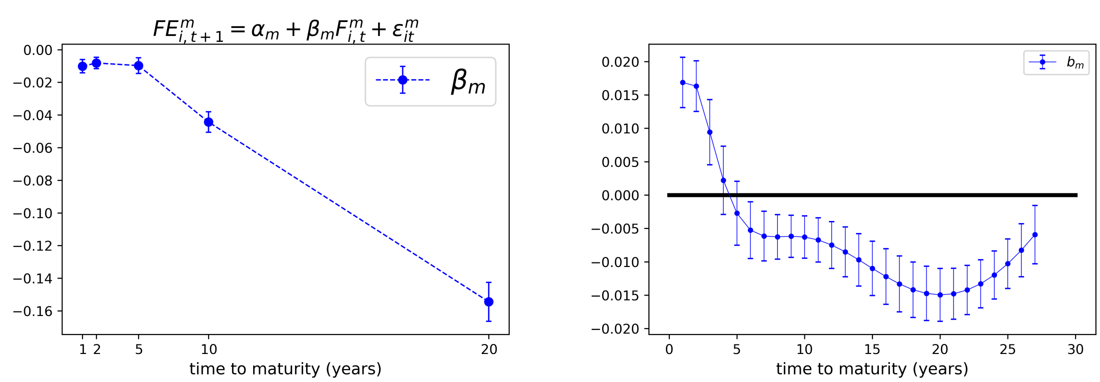


Figure 14: Left panel: Blue Chip Financial. Right panel: Ross recovered beliefs

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 $\mathbf{X}_{t-1}, \mathbf{X}_{t-2}, \ldots$. It is however important to investigate what changes with different benchmark distributions.

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

Consider the following specification, inspired by Bordalo et al. (2018b)[21]:

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

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:

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

[21] 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:

$$y^{\theta}_{t,m} = a^{\theta}_{m} + \psi^{\theta}_{m}\left((b^{\mathbb{Q}}_{m})^{\top}\mathbf{X}_{t+1} - (b^{\mathbb{P}})^{\top}_{m+k}\mathbf{X}_{t-k}\right).$$

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 $0 \le a_k \le 1$ are positive weights on past information such that $\sum_k a_k = 1$ and $1 \le M \le \infty$.[22] The diagnostic distribution of $r_{t,m}$ in this case is still Gaussian, with mean:

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

and variance:

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

where

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

[22] 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.

6.4 Excess bond returns predictability: additional results

	$r^{\theta}_{t,t+1y,2}$	$r^{\theta}_{t,t+1y,5}$	$r^{\theta}_{t,t+1y,10}$	$r^{\theta}_{t,t+1y,30}$
$\hat{FE}^{2}_{t\|t-1}$	-3.54***			
	(0.65)			
$\hat{FE}^{5}_{t\|t-1}$		-10.53***		
		(2.62)		
$\hat{FE}^{10}_{t\|t-1}$			-13.46*	
			(7.19)	
$\hat{FE}^{30}_{t\|t-1}$				-17.71
				(22.98)
const	0.05	0.05	0.27	-3.60
	(0.13)	(0.51)	(1.38)	(4.35)
$PC_{3,t}$	-0.12	0.76	19.03	8.83
	(1.43)	(4.91)	(18.22)	(58.12)
$PC_{3,t}$	-1.10***	-2.32***	-1.84	15.18**
	(0.23)	(0.76)	(1.87)	(6.02)
$PC_{3,t}$	-0.30***	-0.29	0.47	0.06
	(0.06)	(0.21)	(0.43)	(1.23)
R^2	0.31	0.09	0.03	0.17
N	335	335	277	145

Table 4: Excess Bond returns predictability. $r^{\theta}_{t,t+1y,2}$ denotes the one year holding excess returns (from time t to time $t+1y$), relative to the one year yield. Excess returns for higher maturity are similarly defined. $\hat{FE}^{2}_{t}$ is the time t predicted next quarter interest rates to maturity 2 forecast error from time $t-1$ forecast revision relative to the past quarter. Similarly, fitted errors are defined for larger maturities. HAC standard errors are computed with 5 legs.

	xrn 2y	xrn 5y	xrn 10y	xrn 30y
const	1.10***	9.21***	19.92***	48.03***
	(0.23)	(1.89)	(5.39)	(16.14)
PC1	-19.55***	-29.40***	-7.42	55.94
	(1.29)	(9.18)	(26.81)	(71.40)
PC2	1.85*	-20.30**	-51.27*	-122.16
	(1.05)	(9.68)	(28.40)	(82.48)
PC3	-2.11***	-2.62**	2.31	20.18*
	(0.19)	(1.26)	(3.40)	(10.72)
R-squared	0.79	0.43	0.22	0.13
	0.79	0.44	0.23	0.14
ross2y	-0.04***			
	(0.01)			
ross5y		-0.25***		
		(0.07)		
ross10y			-0.59***	
			(0.20)	
ross30y				-1.38**
				(0.60)
R-squared	0.79	0.44	0.23	0.14
No. observations	366	366	366	366

Table 5: Excess Bond returns predictability

7 Conclusions

This paper shows empirically that beliefs about interest rates exhibit maturity increasing over-reaction and that they quantitatively match three leading bond market puzzles: the excess volatility in the term structure of asset prices documented by Giglio and Kelly (2018), the excess sensitivity of long term rates (Gürkaynak et al. (2005), Hanson et al. (2018)), as well as excess bond returns predictability from non priced variables (Duffee (2013)).

Theoretically, I consistently derive these implications from an affine term structure models featuring maturity increasing over-reaction. The crucial property that beliefs fail in the data is the law of iterated expectations. Beliefs over-react more when the objective uncertainty is higher, and as a consequence for larger maturities. This feature is derived from the diagnostic expectation model Bordalo et al. (2018b), adapted to the term structure setting. This poses parsimony on the model's parameters in a way which is disciplined by the data generating process and offers moreover valuable benchmarks for parameter comparisons across the literature. The former feature is key for the empirical tests.

This approach is a first step toward the incorporation of realistic expectation formation processes into realistic macro-financial models,, as well as their empirical assessment.

References

Augenblick, N. and Lazarus, E. (2018). Restrictions on asset-price movements under rational expectations: Theory and evidence. *Available at SSRN 3436384*.

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. (2019). Diagnostic expectations and stock returns. *The Journal of Finance*, 74(6):2839–2874.

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.

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. and Shiller, R. J. (1991). Yield spreads and interest rate movements: A bird's eye view. *The Review of Economic Studies*, 58(3):495–514.

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

Cochrane, J. H. and Piazzesi, M. (2005). Bond risk premia. *American Economic Review*, 95(1):138–160.

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.

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. (2018). Excess volatility: Beyond discount rates. *The Quarterly Journal of Economics*, 133(1):71–127.

Gürkaynak, R. S., Sack, B., and Swanson, E. (2005). The sensitivity of long-term interest rates to economic news: Evidence and implications for macroeconomic models. *American economic review*, 95(1):425–436.

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.

Hanson, S. G., Lucca, D. O., and Wright, J. H. (2018). The excess sensitivity of long-term rates: A tale of two frequencies.

Jensen, C. S., Lando, D., and Pedersen, L. H. (2019). Generalized recovery. *Journal of Financial Economics*, 133(1):154–174.

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.

Mankiw, N. G. and Summers, L. H. (1984). Do long-term interest rates overreact to short-term interest rates?

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. (1979). The volatility of long-term interest rates and expectations models of the term structure. *Journal of political Economy*, 87(6):1190–1219.

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

Wang, C. (2019). Under-and over-reaction in yield curve expectations. *Available at SSRN*.

Appendices

A Affine term structure models with over-reacting beliefs

Consider first the class $\mathbb{Q}$-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. This is ex-post validated as a linear relation between yields and factors holds in the data with high R^2, as shown in Appendix C. Beliefs and the SDF are only restricted by the fact that their product yields a linear $\mathbb{Q}-$VAR(1) dynamics and by regularity conditions, for which I refer to *Leet al*. (2010). The conditions above read:

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

where the $\varepsilon_t^{\mathbb{Q}}$ shocks are i.i.d. 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 $\varepsilon_{t,m}^{\mathbb{Q}} = \sum_{k=1}^{m-1} \sum_{i=1}^{l} (\rho^{\mathbb{Q}})^k \Sigma^{\mathbb{Q}} \varepsilon_{t+i}^{\mathbb{Q}}$. 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 is 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 other than technical regularity conditions, which are worked out in Le et al. (2010).

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 over-reacting beliefs can both generate such distortions, which, in turn, account for the Giglio and Kelly (2018) excess volatility puzzle and explain the predictability of forecast error documented with survey data.
How do beliefs generate such distortions? 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, $\forall\ l \leq m$:

$$\mathbf{X}_{t+l} = (1+\psi^{\theta}_m) f^{(l)}(\mathbf{X}_t)\mathbf{X}_t + \sigma^{\theta}_m \varepsilon^{\mathbb{P}}_{t+l,C}. \tag{25}$$

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

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

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

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

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 independent of beliefs - as it reflects only fundamental risk - and it therefore maps into the believed dynamics above into the risk neutral dynamics:

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

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:

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

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). Similarly, this holds true for higher order Markovian beliefs, such as the ones generated by over-reaction relative to time $t-1$ information (and not relative to the sample average) as discussed in Section 6.

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 1980, with maturities $1y$, $2y$, $\ldots$ 30y

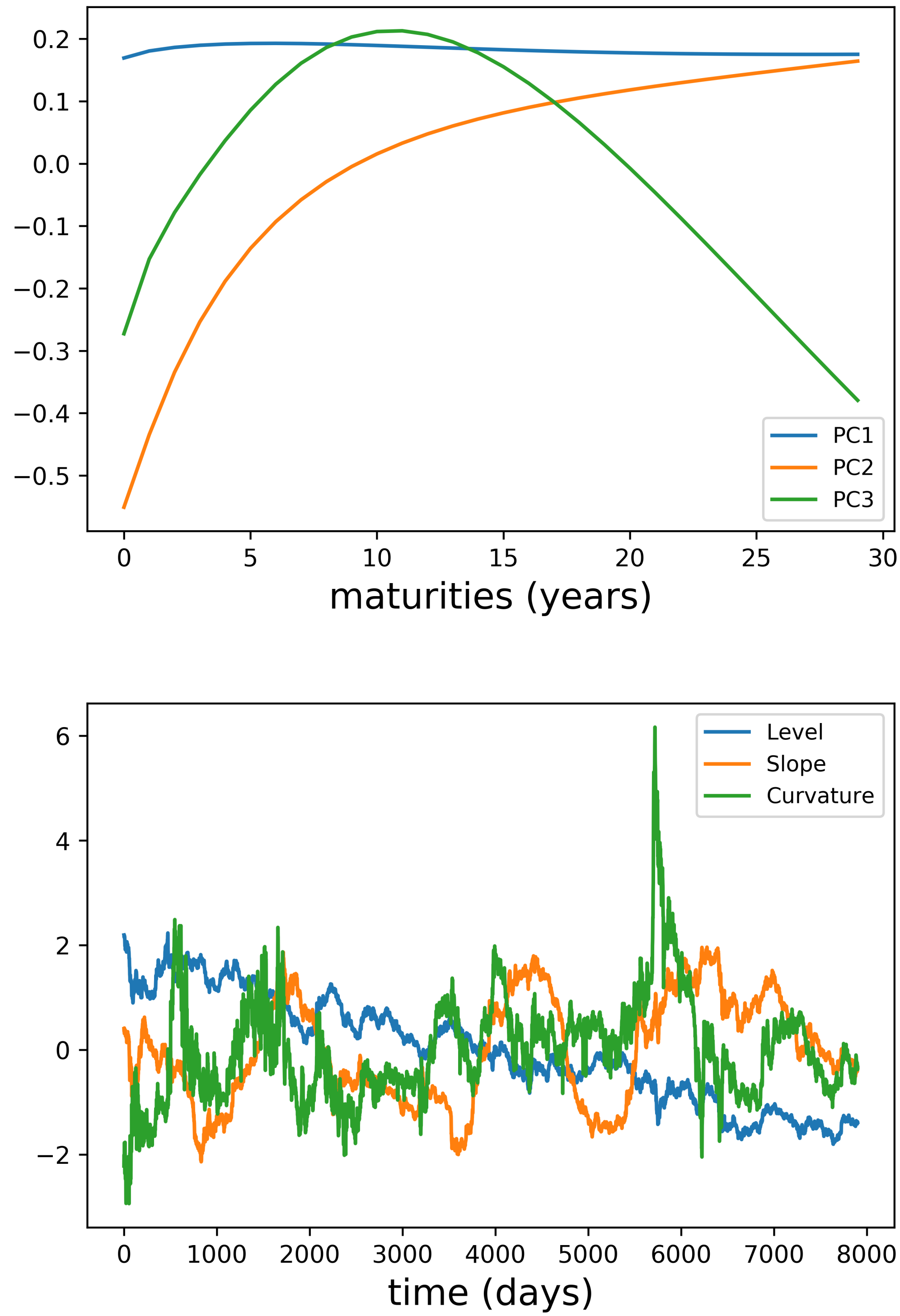
Factors are constructed as follows. First, I estimate the variance covariance matrix of the yield curve using the first half of the sample. In the robustnesses discussed in Section 6, similarly different sub-samples are considered.

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

Then, I consider the first three principal components of the variance covariance matrix of the yield curve, $\underbrace{PC_1}_{30\times 1}$. 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 onto PC_i:

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

Principal components and factors are shown in Figure 15.



.

Figure 15: Top panel: first three principal components of the variance-covariance matrix of the yield curve. Bottom panel: level, slope and curvature factors

The second half of the sample is used for out of sample tests, such as the predictability of forecast errors.

B.2 Parameter Estimation

I estimate the parameters of the risk neutral dynamics:

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

The matrix $\Sigma_C^{\mathbb{Q}}$ assumed upper semi-triangular, namely Cholesky decomposition of the one period variance-covariance matrix of the $^{\mathbb{Q}}$ residuals. This assumption achieve identification of model's parameters (see Giglio and Kelly (2018), Hamilton and Wu (2012)).
Then, I follow Cochrane and Piazzesi (2009) and consider $\Sigma_C^{\mathbb{Q}} \approx \Sigma_C^{\mathbb{P}}$, which I estimate at the daily frequency from the residuals of a $\mathbb{P}$-VAR(1) on the factors. It reads:

$$\hat{\Sigma_C} = \begin{pmatrix} 1.000000 & 0.230580 & 0.214999 \\ 0.230580 & 1.000000 & -0.245012 \\ 0.214999 & -0.245012 & 1.000000 \end{pmatrix}$$

.

I also consider quadratic as well as cubic polynomials in the factor getting no differences in the estimated matrix. Also, at the daily frequency, if factors do not jump and if the possibly non linear persistence also does not jump, daily one would expect non linearity to be not of first order relevance.
Finally, I estimate the remaining parameters $\rho_1^{\mathbb{Q}}$, $\rho_2^{\mathbb{Q}}$ and $\rho_3^{\mathbb{Q}}$ by matching observed yields. Fist, I compute the affine relation for yields. The price at time t of a bond with time to

maturity $m \geq 1$ is:

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

Moving from prices to yields:

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

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 me know discuss how to estimate model's parameters, following Giglio and Kelly (2018) and Hamilton and Wu (2012) but without imposing cross-maturities restrictions. First, I get $\hat{\delta}_0, \hat{\delta}_1^\top$ via OLS estimation of:

$$y_{t,1} := r_t = \delta_0 + \delta_1^\top\mathbf{X}_t.$$

Second, for each maturity $m > 1$, I estimate $\rho_{i,m}^{\mathbb{Q}}$ as:

$$\hat{\rho}^{\mathbb{Q}_{i,m}} := \arg\min_{\rho_{i,m}^{\mathbb{Q}}\in(0,1)}\left(\frac{1}{T_{half}}\sum_{t=1}^{T_{half}}\sum_{m=1}^{30}y_{t,m} - \bar{y}_{m,t}\right)^2,$$

where $y_{t,m}$ is computed via the affine relation, while $\bar{y}_{t,m}$ denote the data. The estimated risk neutral persistence for different maturities sharply differ for the second and third factor. In particular, the estimate persistence increases with maturities.

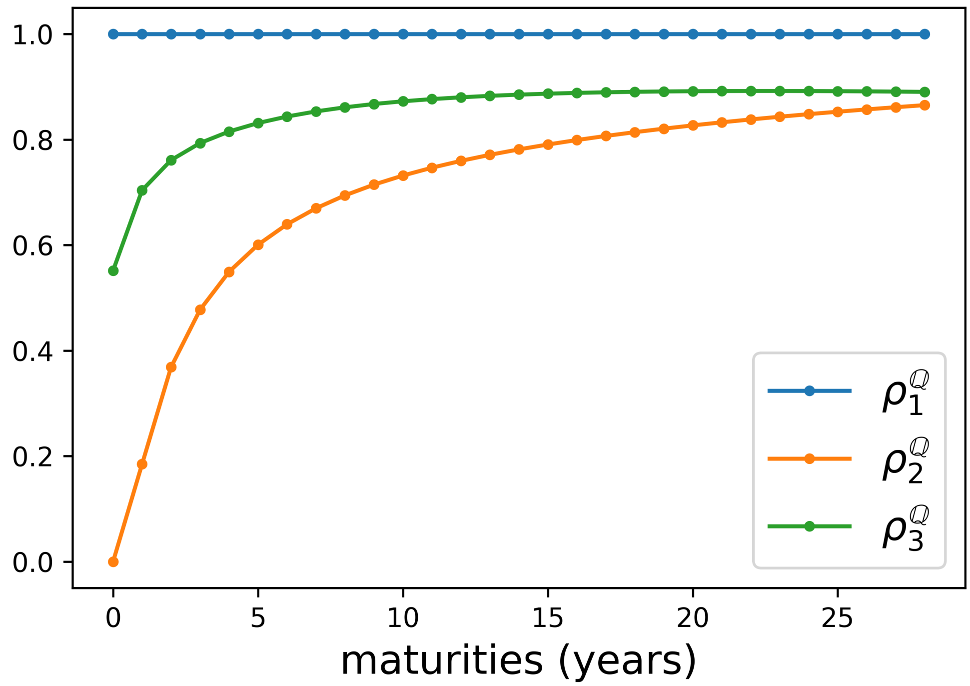


Figure 16: Estimated risk neutral persistence fitting yields to each maturity independently.

The level is highly persistent at all maturities: I consider a single estimated persistence, which will therefore not cause inter-temporal inconsistencies. Despite the wide variation of persistences, a rational affine model provide a reasonable first order matching of the factor loading $b_m^{\mathbb{Q}}$ as shown in the next figure.

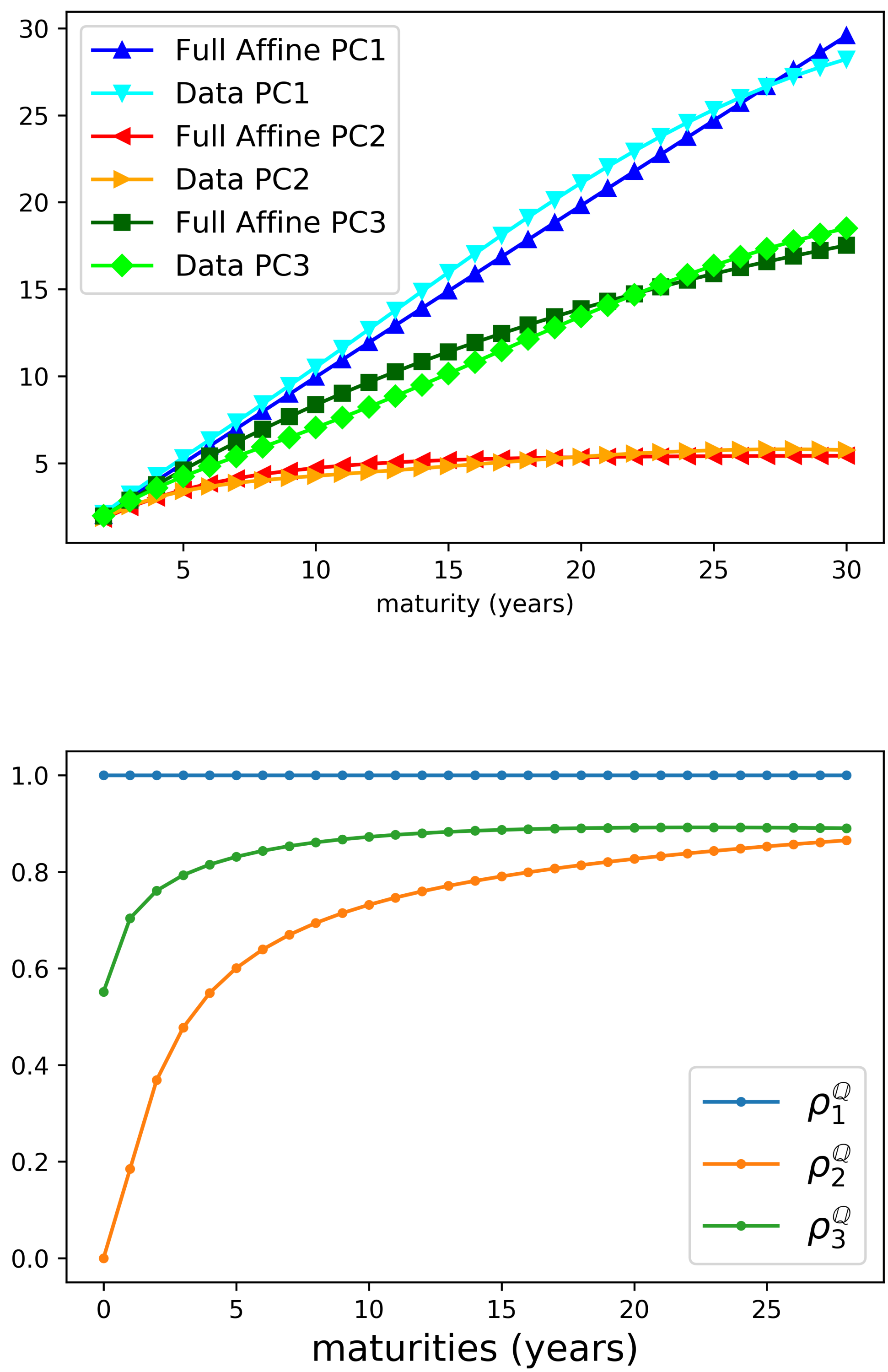


Figure 17: Estimated risk neutral persistence fitting yields to each maturity independently.

B.3 Discretization

Finally, to implement the recovery theorem Ross (2015), I need to discretize the state space in order to compute the Arrow-Debreu price matrix. There are well known methods to efficiently discretize a single $AR(1)$, the state of art of which is the Rouwenhorst method (see Cooley (1995)). I however start from a set of possibly conditionally correlated factor and the single factor recipe is not directly applicable. Ww consider the following transformation provide a set of uncorrelated equation:

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

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

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

Multiplying both sides by Σ_C^{-1}, I get:

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

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

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

the Rouwenhorst discretization method matches conditional and unconditional first and second moments, and is the state of the art among those technique. The grid boundaries ($\bar{X} > 0$ symmetric centered on zero), the step size N and the volatility of a single factor σ are related as: $\bar{X} = \sqrt{N-1}\sigma$ (Cooley (1995)). Spanning the step size from $N = 10$ to $N = 1000$

also allows me to effectively modify the volatility σ, therefore considering deviations from the estimated covariance matrices. Results are shown in Section 6 to be robust to those choices.

Summary of the procedure

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:

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

where $\varepsilon_t^{\mathbb{Q}^\theta}$ are i.i.d. shocks. The matrix $\rho^{\mathbb{Q}^\theta}$ is assumed to be diagonal, $\rho^{\mathbb{Q}^\theta} = diag(\rho_1^{\mathbb{Q}^\theta}, \rho_2^{\mathbb{Q}^\theta}, \rho_3^{\mathbb{Q}^\theta})$.[23] 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. The estimation sample is taken as the first half of the sample, while robustness to sub-samples are discussed in Section 6.

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*, as:

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

[23]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)).

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.

C Tables

In the following Table empirical (OLS) loadings on the factors are shown: R^2 are close to 1 as well known in the literature.

	Y1	Y2	Y3	Y4	Y5	Y6	Y7	Y8	Y9	Y10
const	0.19***	-0.05***	-0.13***	-0.12***	-0.08***	-0.04***	0.00	0.03***	0.06***	0.07***
	(0.02)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)
PC1	0.18***	0.19***	0.19***	0.19***	0.19***	0.19***	0.19***	0.18***	0.18***	0.18***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
PC2	-0.50***	-0.43***	-0.35***	-0.28***	-0.22***	-0.16***	-0.11***	-0.07***	-0.04***	-0.01***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
PC3	-0.28***	-0.20***	-0.11***	-0.02***	0.06***	0.13***	0.18***	0.22***	0.24***	0.25***
	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
R-squared	0.98	1.00	1.00	0.99	0.99	0.99	0.99	1.00	1.00	1.00
	0.98	1.00	1.00	0.99	0.99	0.99	0.99	1.00	1.00	1.00
R-squared	0.98	1.00	1.00	0.99	0.99	0.99	0.99	1.00	1.00	1.00
No. observations	3952	3952	3952	3952	3952	3952	3952	3952	3952	3952

Table 6: Table 1 - OLS Regressions

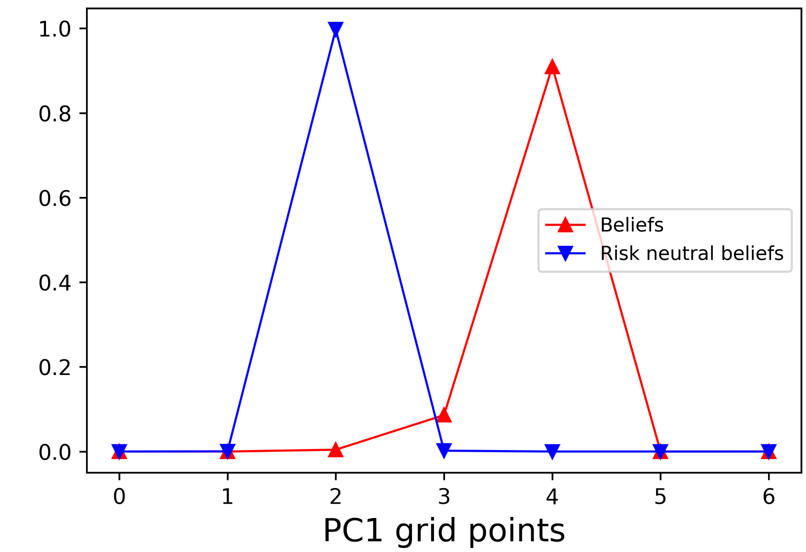
	Y11	Y12	Y13	Y14	Y15	Y16	Y17	Y18	Y19	Y20
const	0.07***	0.06***	0.05***	0.04***	0.02***	0.01	-0.01	-0.02***	-0.03***	-0.04***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
PC1	0.19***	0.19***	0.19***	0.19***	0.19***	0.19***	0.19***	0.19***	0.19***	0.19***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
PC2	0.02***	0.04***	0.05***	0.07***	0.08***	0.09***	0.10***	0.11***	0.12***	0.12***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
PC3	0.25***	0.23***	0.21***	0.19***	0.16***	0.12***	0.09***	0.05***	0.02***	-0.02***
	(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
	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
R-squared	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
No. observations	3952	3952	3952	3952	3952	3952	3952	3952	3952	3952

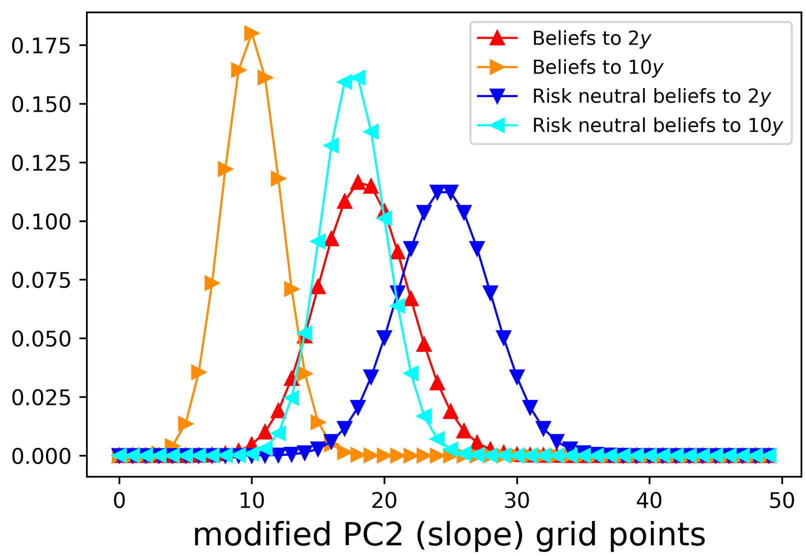
Table 7: Table 1 - OLS Regressions

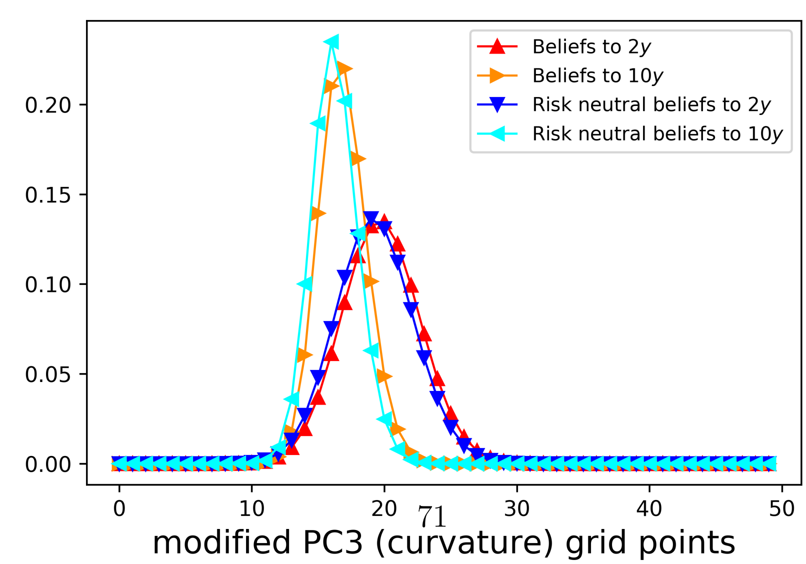
	Y21	Y22	Y23	Y24	Y25	Y26	Y27	Y28	Y29	Y30
const	-0.05***	-0.05***	-0.05***	-0.04***	-0.03***	-0.02***	0.00*	0.03***	0.05***	0.08***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)
PC1	0.18***	0.18***	0.18***	0.18***	0.18***	0.18***	0.17***	0.17***	0.17***	0.17***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
PC2	0.13***	0.13***	0.14***	0.14***	0.14***	0.14***	0.15***	0.15***	0.15***	0.15***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
PC3	-0.06***	-0.09***	-0.12***	-0.16***	-0.19***	-0.21***	-0.24***	-0.27***	-0.29***	-0.31***
	(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	0.99
	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99
R-squared	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99
No. observations	3952	3952	3952	3952	3952	3952	3952	3952	3952	3952

Table 8: Table 1 - OLS Regressions

Finally, recovered beliefs are displayed below.


Beliefs
Risk neutral beliefs
PC1 grid points


Beliefs to 2y
Beliefs to 10y
Risk neutral beliefs to 2y
Risk neutral beliefs to 10y
modified PC2 (slope) grid points


Beliefs to 2y
Beliefs to 10y
Risk neutral beliefs to 2y
Risk neutral beliefs to 10y
modified PC3 (curvature) grid points

The level factor is very persistent hence its transition distributions are very localized, though numerically non zero. Good states for the level are higher values of the factors, whose discretized grid is on the x-axis. On the contrary, the second factor has negative loadings both a maturities $2y$ and $30y$. Hence for the second factor, good states are on the left. This is also true for the third factor, where differences between the recovered and risk neutral distribution are smaller. Overall, the change of measure is smooth and interpretable simply in terms of risk aversion. The change of measure is not trivial and also maturity dependent, especially so for the second factor.

D The Recovery Theorem

Here, I discuss how the recovery theorem works. As discussed in Ross (2015), Markovian discrete state dynamics plus the *path independence* assumption provides me with the dentification of beliefs. Under path independence the fundamental asset pricing equation can be written as:

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

Here, I are considering the possibility that state prices are generated by distorted beliefs. However, as discussed in Section 3, the SDF does not depend on beliefs. 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[24] (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:

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

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

So far, I exploited only *one period* Arrow-Debreu prices. Do prices of *multi-period* Arrow-Debreu securities provide additional information? Under rationality - specifically, If the law of iterated expectation holds[25] - 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. In particular, a testable implication of the law of iterated expectation for affine term structure models is the fact that the term structure of CG coefficients needs to be flat, as shown in Theorem (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 $\mathbb{P}_m^{\theta}$, where $[\mathbb{P}_m^{\theta}]_{ij}$ 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 Theorem 2 (2).

[25]In particular, I need only the Chapman-Kolmogorov equation here.

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. $\square$

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 plugs 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. Then:

$$\frac{b_m^\top Cov(F_t[\mathbf{X}_{t+1}], F_t[\mathbf{X}_{t+1}] - F_{t-k}[\mathbf{X}_{t+1}])b_m}{b_m^\top \mathbb{V}[F_t[\mathbf{X}_{t+1}] - F_{t-k}[\mathbf{X}_{t+1}]]b_m} = 1 - \frac{b_m^\top(\mathbb{V}[F_t[\mathbf{X}_{t+1}]] - Cov(F_t[\mathbf{X}_{t+1}], F_{t-k}[\mathbf{X}_{t+1}]))b_m}{b_m^\top \mathbb{V}[F_t[\mathbf{X}_{t+1}] - F_{t-k}[\mathbf{X}_{t+1}]]b_m}. \tag{26}$$

Therefore, the term is exactly equal to one in the case $k = \infty$, where forecast are revised relative to the sample average. The term is approximately equal to one, for equally persistent processes and also in the case of a single highly persistence factor: both condition are verified in the data, as shown in Appendix B. The argument holds also for non linear processes, considering the local persistence $f(\mathbf{X}_t)$. Moreover, for a fixed persistence matrix, the deviation from one, asymptotically vanishes for $m \to \infty$. This is so because each entry of b_m convergences geometrically for large maturities. There the quantity is asymptotically flat. I now consider the sensitivity of errors on regressions.

$i)$ 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 (26) 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 approximation (26) and assuming $\psi_{m+1} \approx \psi_{m+2}$[26], 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).

□

Proposition 2

[26] Note that m is measure in years while 1 in days.

Proof. Consider the factor distorted dynamics of Appendix and 1, I get:

$$
\begin{aligned}
y_{t+h,m} &= a_m^\theta + \left(b_m^{\mathbb{Q}} + \psi_m^\theta b_m^{\mathbb{P}}(\mathbf{X}_t)\right)\mathbf{X}_{t+h} \\
&= a_m^\theta + \left(b_m^{\mathbb{Q}} + \psi_m^\theta b_m^{\mathbb{P}}(\mathbf{X}_t)\right)(1+\psi_m^\theta)^h \underbrace{f(\dots f(\mathbf{X}_t))}_{\text{h times}}\mathbf{X}_t + \varepsilon_{t,h}^{\mathbb{P}}.
\end{aligned}
$$

This reduces to Proposition 2 in the case of a single $\mathbb{P}$-AR(1) factor. □

Proposition 3

Proof. Under rational expectation, factors span the current yield curve $\{y_{t,m}^\theta\}_m$ but not necessarily all the forecast revisions $\{FR_t[y_{t+h,m-h}]\}_{m>h}$. □

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:

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

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:

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

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 (26), 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[27]. 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. I drop the additive constant $a_m^{\mathbb{Q}^\theta}$ in the following calculations, because it is inessential for the computation of covariances.

$$\begin{aligned}
&\widetilde{FE}^{\theta}_{t+1,m} := \bar{y}^{\theta}_{t+1,m} - \mathbb{E}_t^{\widetilde{\mathbb{P}}^\theta}\left[y^{\theta}_{t+1,m}\right] = \bar{y}^{\theta}_{t+1,m} - (1+\psi^{\theta}_{m+1})\mathbb{E}_t^{\widetilde{\mathbb{P}}}\left[y^{\theta}_{t+1,m}\right] \\
&\bar{y}^{\theta}_{t+1,m} - \mathbb{E}_t^{\widetilde{\mathbb{P}}^\theta}\left[y^{\theta}_{t+1,m}\right] = \bar{y}^{\theta}_{t+1,m} - (1+\psi^{\theta}_{m+1})(\mathbb{E}_t^{\widetilde{\mathbb{P}}}\left[y^{\theta}_{t+1,m}\right] + \mathbb{E}_t^{\mathbb{P}}\left[y^{\theta}_{t+1,m}\right] - \mathbb{E}_t^{\mathbb{P}}\left[y^{\theta}_{t+1,m}\right]) \\
&\ = FE^{\theta}_{t+1,m} + \left(\mathbb{E}_t^{\mathbb{P}}\left[y^{\theta}_{t+1,m}\right] - \mathbb{E}_t^{\widetilde{\mathbb{P}}}\left[y^{\theta}_{t+1,m}\right]\right) - \psi^{\theta}_{m+1}\mathbb{E}_t^{\widetilde{\mathbb{P}}}\left[y^{\theta}_{t+1,m}\right].
\end{aligned}$$

Distorted and misspecified CB coefficients reads:

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

the second term, under one factor model or under approximation (26) does not depend on m. The coefficients $\widetilde{\beta}^{\theta}_m$ are negative and decreasing iff ψ^{θ}_m is positive and increasing and provided that $\frac{(b^{\theta}_m)^\top \mathbb{V}[\widetilde{F}_t[\mathbf{X}_{t+1}]] b^{\theta}_m}{(b^{\theta}_m)^\top Cov(\widetilde{F}_t[\mathbf{X}_{t+1}], \widetilde{F}_{t-1}[\mathbf{X}_{t+1}] b^{\theta}_m} > \frac{\psi^{\theta}_2}{\psi^{\theta}_1}$. Differences of distorting coefficients remain

[27] 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.

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

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

I want to compute the distribution:

$$f_{\mathbb{P}^\theta}(r_{t+m}|\mathbf{X}_t) \propto f_{\mathbb{P}}(r_{t+m}|\mathbf{X}_t)\left(\frac{f_{\mathbb{P}}(r_{t+m}|\mathbf{X}_t)}{f_{\mathbb{P}}(r_{t+m})}\right)^\theta. \tag{27}$$

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 $r_{t,m}$. $\square$

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:

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

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 $i > 1$, $f_i^{\mathbb{P}}(\mathbf{X}_t) = \underbrace{f^{\mathbb{P}}(\dots f^{\mathbb{P}}}_{i \text{ times}}(\mathbf{X}_t))$.

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

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

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:

$$\sqrt{\mathbb{V}_t^{\mathbb{Q}}[r_{t,m}]}\varepsilon_{t,m}^{\mathbb{Q}} = \sqrt{\mathbb{V}_t^{\mathbb{P}}[r_{t,m}]}\varepsilon_{t,m}^{\mathbb{P}} + \mathbb{E}_t^{\mathbb{P}}[r_{t,m}] - \mathbb{E}_t^{\mathbb{Q}}[r_{t,m}], \tag{28}$$

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:

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

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:

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

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).

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 (19) 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[28]. 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.

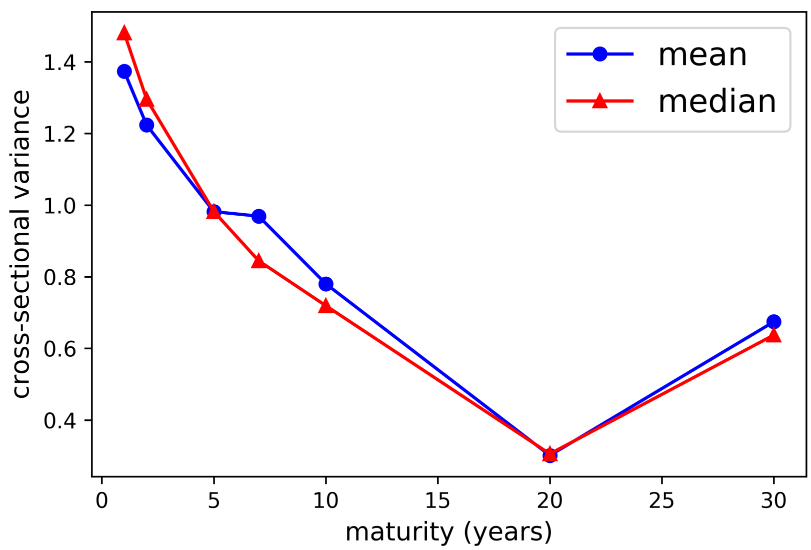


Figure 19: Time mean (blue circle) and median (red triangle) of analysts dispersion (cross sectional standard deviation) as a function of the maturity.

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