

Speculation, Risk Premia and Expectations in the Yield Curve

FRANCISCO BARILLAS and KRISTOFFER NIMARK*

Abstract

An affine asset pricing model in which agents have rational but heterogeneous expectations about future asset prices is developed. We estimate the model using data on bond yields and individual survey responses from the *Survey of Professional Forecasters* and perform a novel three-way decomposition of bond yields into (i) average expectations about short rates (ii) risk premia and (iii) a speculative component due to heterogeneous expectations about the resale value of a bond. We prove that the speculative term must be orthogonal to public information in real time and therefore statistically distinct from risk premia. Empirically, the speculative component is quantitatively important, accounting for up to one percentage point of US yields. Furthermore, estimates of historical risk premia from the heterogeneous information model are less volatile than, and negatively correlated with, risk premia estimated using a standard Affine Gaussian Term Structure model.

JEL classification: G12, G14.

*Barillas is with the Goizueta Business School at Emory University. Nimark is with CREI, Universitat Pompeu Fabra and Barcelona GSE. The authors thank Tarun Chordia, Thomas Dangl, Alex Kohlhas, Christian Matthes, Joonki Noh, Stefan Pitschner, Cesare Robotti, Jay Shanken, Chris Telmer, Yimei Zou and seminar participants at the NBER Summer Institute 2012, Duke University, Chicago Finance Junior Symposium, Down Under Finance Conference at Melbourne University, European Winter Finance Summit, Universitat Pompeu Fabra and the Toulouse TIGER forum 2013 for useful comments and suggestions.

A large part of the existing empirical literature analyzing the term structure of interest rates implicitly or explicitly decomposes bond yields into risk premia and expectations about future risk-free interest rates.¹ However, both casual observation and survey evidence suggest that there is a lot of disagreement about future interest rates. In this paper, we ask how this fact should change our view about what components make up bond yields. We present a flexible affine asset pricing framework in which agents have rational but heterogeneous expectations about future bond yields. This framework is then used to argue that heterogeneous expectations about the resale value of a bond give rise to an empirically important third bond yield component due to speculative behavior. We prove that the speculative component must be orthogonal to public information and is therefore statistically distinct from both risk premia and expectations about future risk-free interest rates.

In markets where assets are traded among agents who may not want to hold an asset until it is liquidated, expectations about the resale value of an asset will matter for its current price. When rational agents have access to different information about future fundamentals, the price of the asset deviates systematically from the “consensus value” defined as the hypothetical price that would reflect the average opinion of the (appropriately discounted) fundamental value of the asset (e.g. Allen, Morris and Shin 2006, Bacchetta and van Wincoop 2006, Nimark 2012). These deviations from the consensus price occur because an individual agent’s expectation about the resale value of an asset can with heterogeneous expectations be different from what the individual agent would be willing to pay for the asset if he were to hold it until maturity. Heterogeneous expectations then give rise to speculative behavior in the sense of Harrison and Kreps (1978), who defined investors as exhibiting “*speculative behavior if the right to resell a stock makes them willing to pay more for it than they would pay if obliged to hold it forever (p.323)*”.²

In this paper, we derive an affine pricing framework for empirically quantifying the type of speculative behavior described above. The framework differs from most of the previous theoretical literature on asset pricing with heterogeneously informed agents in that we do not specify utility functions, nor do we model the portfolio decisions of agents explicitly.³ Instead, we make an effort to stay as close as possible to the

¹Some examples are Joslin, Singleton and Zhu (2011), Duffee (2002), Cochrane and Piazzesi (2008), Bauer, Rudebusch and Wu (2012), Joslin, Priebsch and Singleton (2012).

²Harrison and Kreps (1978) impose a short sales constraint on the agents in their model that implies that speculative behavior always increases the price of the asset. In our framework, speculation can either increase or decrease the price of a bond.

³Some early examples of papers studying the theoretical implications of heterogeneous information on asset prices in a rational setting are Grossman (1976), Hellwig (1980), Grossman and Stiglitz (1980), Admati (1985), Singleton (1987). More recent examples include Allen, Morris and Shin (2006), Bacchetta and van Wincoop (2006) and Nimark (2012). As we show in the Internet Appendix associated with the paper, the affine model presented here nests the equilibrium model in Nimark (2012), in which agents make explicit trading and portfolio decisions, as a special case.

large empirical literature that uses affine models to study asset prices. In the standard full information affine no-arbitrage framework, variation across time in expected excess returns is explained by variation across time in either the amount of risk or in the required compensation for a given amount of risk. Gaussian models such as the $A_0(N)$ models of Dai and Singleton (2000) or the model in Joslin, Singleton and Zhu (2011) focus on the latter and identify the price of risk as an affine function of a small number of factors that also determine the dynamics of the risk-free short rate. Similar to this approach, we specify a model in which variation in expected excess returns across individual agents, in the absence of arbitrage, must be accompanied by variation across agents in the required compensation for risk. The framework is flexible and nests a standard affine Gaussian term structure model if the signals observed by agents are perfectly informative about the state. This facilitates comparison of our results to the large existing literature on affine term structure models. However, the framework is general and can also be used to price other classes of assets.

The main empirical contribution of the paper is a novel three-way decomposition of bond yields. We show that in addition to the classic components due to risk premia and expectations about future risk-free short rates, heterogeneous information introduces a third term due to speculation. The speculative component in bond yields is quantitatively important, accounting for up to a full percentage point of medium- to long-maturity yields in the 1980s and up to 60 basis point of yields in the low nominal yield environment of the last decade. The speculative term arises when individual agents' expectations about the average expectations about the resale value of a bond is different from their own best estimate. This difference between an agent's expectation and his expectation about the average expectation can equivalently be thought of as a higher order prediction error, i.e. a prediction about the error in other agents' forecasts. In the model, all agents form rational, or model consistent, expectations and it is thus not possible for individual agents to predict the error in the average expectation based on information that is also available to all other agents. The speculative component must therefore be orthogonal to all public information available in real time which makes it statistically distinct from traditional risk premia, which can be predicted conditional on publicly available information such as bond prices.

Allowing for heterogeneous information also changes the cyclical properties of the common component of risk premia, as compared to a full information model. Risk premia estimated from the model with heterogeneous information is less volatile than, and negatively correlated with risk premia extracted using the nested full information model of Joslin, Singleton and Zhu (2011).

When implementing the framework empirically, we treat the individual responses in the Survey of Professional Forecasters as being representative of the bond yield expectations of agents randomly drawn from the population of agents in the model. There is substantial dispersion of survey responses and the average cross-sectional standard deviation of the one-year-ahead forecasts of the Federal Funds Rate is approximately 40 basis points. In reality, bond prices depend on many things, including monetary policy, current attitudes towards risk and political, macroeconomic and financial market developments. There is thus a vast amount of information available that could help predict future bond prices.

In the set-up presented here, different agents observe different signals with idiosyncratic noise components about a vector of common latent factors. The agents use these signals to form rational, or model consistent, expectations about future risk-free rates and risk premia. This set-up is a simple way to capture the fact that, in practice, it is too costly for agents to pay attention to all available information that could potentially help predict bond prices. With slightly different vantage points and historical experiences, agents instead tend to observe different subsets of all available information. Since the signals contain information about a common vector of latent factors, information sets will be highly correlated across agents, but not perfectly so. Formally, the set-up is similar to the information structure in Diamond and Verrecchia (1981), Admati (1985), Singleton (1987), Allen, Morris and Shin (2006) and Bacchetta and van Wincoop (2006). Because the model is populated by a continuum of agents who have heterogeneous expectations about future bond yields, it is possible to use individual survey responses of interest rate forecasts in combination with likelihood based methods to estimate the parameters of the model. While other papers have used survey data to estimate term structure models, we believe our paper is the first to use a model that can explain the observed dispersion of survey forecasts.⁴ Unlike papers that treat individual survey responses, or a measure of central tendency such as the mean or median survey, as a noisy measure of a single representative agent's forecast, we can use the full cross-section of survey responses to get sharper estimates of how much information that is available to the agents that populate our model.

The information in the cross-section of survey forecasts clearly disciplines the parameters that directly govern the precision of agents' information. If the agent-specific signals are too precise or so noisy that they will be disregarded, the model will fail to fit the cross-sectional dispersion of forecasts in the Survey of Professional Forecasters. Less obviously, using the full cross-section of individual survey responses also

⁴See D'Amico, Kim and Wei (2008), Chun (2011) and Piazzesi and Schneider (2011) for examples of studies who have used survey data to estimate term structure dynamics.

restricts the dynamics of bond yields. If observing bond yields reveals the latent factors perfectly, agents will also disregard their agent-specific signals. Parameterizations of the model that make the latent factors an invertible function of bond yields will thus be rejected by the data, since too informative bond prices would imply a counter-factually degenerate cross-sectional distribution of expectations. That the model is forced to match the observed dispersion of yield forecasts thus empirically imposes restrictions that are similar to the theoretical restrictions imposed in models with unspanned factors, e.g. Duffee (2011), Joslin, Priebsch and Singleton (2010) and Barillas (2013).

The restrictions that the speculative term must be orthogonal to bond prices in real time and that bond prices cannot be too revealing if we are to fit the cross-sectional dispersion of survey forecasts are both consequences of that agents use the information in the endogenous bond prices to form model consistent expectations. In models employing alternative assumptions about agents' beliefs and where agents do not need to filter from endogenous bond prices, such as the models in Xiong and Yan (2010) and Chen, Joslin and Tran (2010, 2012), these restrictions would be absent. For example, in the model of Xiong and Yan (2010), two groups of agents with heterogeneous beliefs take on speculative positions against each other. The interaction between heterogeneous expectations and relative wealth dynamics in that model would to an outside econometrician be indistinguishable from traditional time varying risk premia and not be orthogonal to bond prices. Unlike our model, Xiong and Yan (2010) thus proposes an alternative explanation to the well-documented failure of the expectations hypothesis. Another difference between the two papers is that in the equilibrium model of Xiong and Yan (2010), the interaction between heterogeneous beliefs and relative wealth dynamics are important. Since we do not model the trading and portfolio decisions of agents explicitly, our framework is silent on the empirical importance of this channel.

The next section derives a number of implications of heterogeneous information for stochastic discount factor based asset pricing and defines the speculative component of a bond's price in a general setting. There, we formally prove that excess returns that are predictable based on agent-specific information, as well as the aggregate speculative component in bond prices, must be orthogonal to public information in real time. Section II then presents an operational affine framework for empirically modeling the term structure of interest rates when agents observe different information relevant for predicting future bond prices. Section III shows how the affine model can be used to decompose the term structure into the standard components, i.e. risk premia and expectations about future short rates, as well as a speculative component driven by information heterogeneity. Section IV describes in more detail the empirical specification and how the

model can be estimated. Section V presents the empirical results. Section VI concludes and the Appendix contains an additional proof and detailed derivations of the affine model’s pricing equations.

I. Stochastic discount factors and heterogeneous information

A number of implications of information heterogeneity for asset pricing can be understood without reference to a fully specified model. In this section we derive some results that can be framed simply in terms of stochastic discount factors and excess returns. We first show that if agents disagree about future bond prices, agents’ stochastic discount factors must also differ. We then demonstrate that the component of expected excess returns that is due purely to information heterogeneity is statistically distinct from standard sources of time varying risk premia because it must be orthogonal to public information in real time. In this section we also define the speculative component of bond prices as the difference between the actual price of a bond and the counterfactual price the bond would have if all agents shared the expectations of the “average” agent and this fact was common knowledge. The section following this one presents an explicit affine no-arbitrage model, featuring the properties discussed and derived here, that we will later use to quantify the speculative term in bond prices. However, the results derived here are general and apply also to equilibrium models of the term structure of interest rates where agents are heterogeneously informed, such as the model in Nimark (2012).

A. Stochastic discount factors and expectations about future bond prices

In standard common information models, the price P_t^n of a zero-coupon, no-default bond with n periods to maturity is given by

$$P_t^n = E [M_{t+1} P_{t+1}^{n-1} | \Omega_t] \quad (1)$$

where Ω_t is the common information set in period t and M_{t+1} is the stochastic discount factor. In the absence of arbitrage, this relationship has to hold for all maturities n . In a model with heterogeneous information a similar relation holds, except that the SDF is now agent-specific so that for all agents $j \in (0, 1)$ and all maturities n the relationship

$$P_t^n = E [M_{t+1}^j P_{t+1}^{n-1} | \Omega_t^j] \quad (2)$$

must hold. Here Ω_t^j is the information set of agent j in period t . All agents observe the current price for bonds so the left hand side of (2) is common to all agents. However, agent-specific information sets introduce heterogeneity in expectations of P_{t+1}^{n-1} . For (2) to continue to hold when expectations about P_{t+1}^{n-1} differ across agents, the stochastic discount factor M_{t+1}^j must also be agent-specific. There is thus a close relationship between heterogeneity in expectations about future prices and heterogeneity in stochastic discount factors. Any SDF based framework that incorporates heterogeneity in expected returns must therefore allow for heterogeneity in stochastic discount factors as well. As a consequence, the model presented in Section II features agent-specific state variables. These state variables determine both an individual agents' expectation about future bond prices as well as the agent's required compensation for risk.

B. Excess returns and public information

In equilibrium models with heterogeneously informed agents such as those of Hellwig (1980), Admati (1985) or Singleton (1987), agents with more optimistic views about the return on a risky asset will hold more of it in their portfolios. In equilibrium, the higher excess return that optimistic agents expect to earn relative to pessimistic agents is compensation for holding a riskier portfolio. A positive excess return that is predictable based on agent-specific information is thus similar to a positive expected excess return that arise from standard sources in that it can only be earned as compensation for risk. However, it is possible to demonstrate that the component of expected excess return that is due to heterogeneous information has a distinct characteristic: It must be orthogonal to public information in real time.

Start by defining agent j 's expected one-period excess return $rx_{t,j}^n$ on a zero-coupon bond with n periods to maturity as

$$rx_{t,j}^n \equiv E \left[p_{t+1}^{n-1} \mid \Omega_t^j \right] - p_t^n - r_t \quad (3)$$

where p_t^n is the log price of an n -period bond and r_t is the one period risk-free rate. If agents have heterogeneous expectations about future bond prices, definition (3) implies that there will also be heterogeneity in expected excess returns. The difference between agent j 's expected excess return and the cross-sectional average expected excess return is given by

$$rx_{t,j}^n - \int rx_{t,i}^n di = E \left[p_{t+1}^{n-1} \mid \Omega_t^j \right] - \int E \left[p_{t+1}^{n-1} \mid \Omega_t^i \right] di \quad (4)$$

In general, this quantity will not coincide with agent j 's subjective view of the same quantity. Taking the expectation of (4) conditional on agent j 's information set and denoting this quantity by $s_{t,j}^n$ gives

$$s_{t,j}^n \equiv rx_{t,j}^n - E \left[\int rx_{t,i}^n di \mid \Omega_t^j \right] \quad (5)$$

which by (3) can be expressed as

$$s_{t,j}^n = E \left[\left(p_{t+1}^{n-1} - \int E [p_{t+1}^{n-1} \mid \Omega_t^i] di \right) \mid \Omega_t^j \right]. \quad (6)$$

so that $s_{t,j}^n$ is the difference between the return that agent j 's expects to earn on holding an n -period bond for one period and what he thinks the average agent expects to earn. That is, $s_{t,j}^n$ can be understood as a second order prediction error, i.e. agent j 's prediction about the error other agents are making in period t when predicting what the price of the bond will be in the next period. When all agents form rational expectations, it is not possible for individual agents to predict the errors that other agents are making by conditioning on public information, which by definition is available to all agents. The term $s_{t,j}^n$ must thus be orthogonal to public information in real time. We now prove this more formally.

PROPOSITION 1: *The term $s_{t,j}^n$ is orthogonal to public information in real time, i.e.*

$$E (s_{t,j}^n \omega_t) = 0 : \forall \omega_t \in \Omega_t \quad (7)$$

where Ω_t is the public information set at time t defined as the intersection of agents' period t information sets

$$\Omega_t \equiv \bigcap_{j \in (0,1)} \Omega_t^j. \quad (8)$$

Proof. The law of iterated expectations states that for a random variable X

$$E [E [X \mid \Omega] \mid \Omega'] = E [X \mid \Omega']$$

if $\Omega' \subseteq \Omega$. Taking expectations of $s_{t,j}^n$ conditional on the public information set Ω_t gives

$$E [s_{t,j}^n \mid \Omega_t] = E \left(E \left[\left(p_{t+1}^{n-1} - \int E [p_{t+1}^{n-1} \mid \Omega_t^i] di \right) \mid \Omega_t^j \right] \mid \Omega_t \right) \quad (9)$$

The definition of the public information set (8) implies that $\Omega_t \subseteq \Omega_t^j$ for all j . Applying the law of iterated expectations to the right hand side of (9) then gives

$$\begin{aligned} E [s_{t,j}^n | \Omega_t] &= E (p_{t+1}^{n-1} | \Omega_t) - E (p_{t+1}^{n-1} | \Omega_t) \\ &= 0 \end{aligned} \tag{10}$$

which completes the proof. □

That expected excess returns due to heterogeneous information must be orthogonal to public information in real time makes $s_{t,j}^n$ statistically distinct from classical time varying risk premia. There is a large literature that documents that excess returns on nominal bonds can be predicted by conditioning on public information. Notable examples using term structure variables include Fama and Bliss (1987), Duffee (2002), Dai and Singleton (2002, 2003) and Cochrane and Piazzesi (2005). There are also papers that document that non-term structure variables, such as macroeconomic indicators, may help predict excess returns, e.g. Ludvigson and Ng (2009) and Joslin, Priebsch and Singleton (2012).

Even though excess returns due to information heterogeneity are statistically distinct from the predictable excess returns documented in these papers, introducing heterogeneous information into an SDF based asset pricing model does not require modifications to the specification of agents' required compensation for risk. In the next section we will demonstrate how the risk premia associated with heterogeneous expectations about future bond prices can be parsimoniously parameterized in an affine model that nests the standard full-information model as a special case.

C. Bond prices and higher order expectations

Harrison and Kreps (1978) defined speculative behavior as when the right of reselling an asset before maturity changes its equilibrium price. In the model presented here, we implicitly assume that long maturity bonds are traded in every period and that agents are price takers. Individual agents therefore need to predict what other agents will be willing to pay for a bond at the next trading opportunity. Since the price other agents will be willing to pay for a bond in the future depends on their future risk adjusted and discounted expectations of bond prices further into the future, individual agents need to form higher order expectations about these quantities to predict the next period bond price.

It is well known that higher order expectations are distinct from first order expectations when agents have heterogeneous information (e.g. Allen, Morris and Shin 2006 and Bacchetta and van Wincoop 2006). It is thus possible that individual agents believe that other agents will be willing to pay more (or less) for a bond in the future than they would be willing to pay themselves if they were to hold on to the bond until it matures. Heterogeneous information thus introduces what Allen, Morris and Shin (2006) call a “Keynesian beauty contest” into asset markets, where heterogeneously informed agents need to “forecast the forecast of others”.

To make this argument more formally, it will be useful to restrict our attention to jointly log-normal processes for prices and stochastic discount factors. We will also assume that conditional variances are deterministic and common across agents. (The latter assumption is not crucial for any of the results, but helps simplifying the notation.) We can then write the no-arbitrage condition (2) in log terms as

$$p_t^n = E \left[m_{t+1}^j \mid \Omega_t^j \right] + E \left[p_{t+1}^{n-1} \mid \Omega_t^j \right] + \frac{1}{2} Var \left(m_{t+1}^j + p_{t+1}^{n-1} \mid \Omega_t^j \right) \quad (11)$$

The assumption that individual agents are price takers means that when we evaluate the no-arbitrage condition (11) for agent j , we replace the expectation of the next period price p_{t+1}^{n-1} by agent j 's expectation of what other agents will be willing to pay for the bond in the next period. Leading the no-arbitrage condition (11) and using it to substitute out p_{t+1}^{n-1} from agent j 's expectation above gives

$$\begin{aligned} p_t^n &= E \left[m_{t+1}^j \mid \Omega_t^j \right] \\ &+ E \left[E \left(m_{t+2}^i \mid \Omega_{t+1}^i \right) \mid \Omega_t^j \right] + E \left[E \left(p_{t+2}^{n-2} \mid \Omega_{t+1}^i \right) \mid \Omega_t^j \right] \\ &+ \frac{1}{2} Var \left(m_{t+1}^j + p_{t+1}^{n-1} \mid \Omega_t^j \right) + \frac{1}{2} Var \left(m_{t+2}^i + p_{t+2}^{n-2} \mid \Omega_{t+1}^i \right) \end{aligned} \quad (12)$$

where the superscript i is used to indicate any agent i such that $i \neq j$. In this paper, agent j does not have any information relevant for predicting the expectations and stochastic discount factors of any other particular agent. Agent j 's expectations of any other agent's expectation then coincide with agent j 's expectation about the average expectation. That is, agent j 's expectation about agent i 's future expectation about bond prices and discount factors coincide with agent j 's expectations about the future cross-sectional average

expectation about the same quantities. We can thus substitute in

$$E \left(E \left[p_{t+2}^{n-2} \mid \Omega_{t+1}^i \right] \mid \Omega_t^j \right) = E \left(\int E \left[p_{t+2}^{n-2} \mid \Omega_{t+1}^i \right] di \mid \Omega_t^j \right) \quad (13)$$

and

$$E \left(E \left[m_{t+2}^i \mid \Omega_{t+1}^i \right] \mid \Omega_t^j \right) = E \left(\int E \left[m_{t+2}^i \mid \Omega_{t+1}^i \right] di \mid \Omega_t^j \right)$$

into (12). Since the resulting expressions hold for each agent j , and thus also for the average agent, the log price of an n period bond can be written as

$$\begin{aligned} p_t^n &= \int E \left[m_{t+1}^j \mid \Omega_t^j \right] dj \\ &+ \int E \left[\left(\int E \left[m_{t+2}^i \mid \Omega_{t+1}^i \right] di + \int E \left[p_{t+2}^{n-2} \mid \Omega_{t+1}^i \right] di \right) \mid \Omega_t^j \right] dj \\ &+ \frac{1}{2} Var \left(m_{t+1}^j + p_{t+1}^{n-1} \mid \Omega_t^j \right) + \frac{1}{2} Var \left(m_{t+2}^i + p_{t+2}^{n-2} \mid \Omega_{t+1}^i \right) \end{aligned} \quad (14)$$

The price of a bond with n periods to maturity in period t is thus partly a function of the average expectation in period t of the average expectation in period $t + 1$ of the price of the same bond in period $t + 2$ when it has $n - 2$ periods left to maturity. If all agents shared the same information set, we could apply the law of iterated expectations and replace this second order expectation of the price p_{t+2}^{n-2} with the common expectation. However, with heterogeneous information sets, the law of iterated expectations does not apply and agents second order expectations, i.e. their expectations about other agents' expectations about p_{t+2}^{n-2} may differ from their own expectations about the price.

We can continue to recursively substitute out expectations about future prices from equation (11) and express the log price of a bond as the sum of higher order expectations about future stochastic discount factors

$$\begin{aligned} p_t^n &= \int E \left[m_{t+1}^j \mid \Omega_t^j \right] dj \\ &+ \int E \left[\int E \left[m_{t+2}^i \mid \Omega_{t+1}^i \right] di \mid \Omega_t^j \right] dj + \dots \\ &\dots + \int E \left[\int E \left[\dots \int E \left[m_{t+n}^i \mid \Omega_{t+n-1}^{i'} \right] di' \dots \right] di \mid \Omega_t^j \right] dj \\ &+ \frac{1}{2} \sum_{s=0}^{n-1} Var \left(m_{t+1+s}^j + p_{t+1+s}^{n-1-s} \mid \Omega_{t+s}^j \right) \end{aligned} \quad (15)$$

where we used that the conditional variance terms are the same for every agent to simplify the sum of the conditional variance terms. In the equilibrium model of Nimark (2012), the price of an n -period bond can be written as a function of a random supply shock and higher order expectations about future risk-free short rates. That model is a restricted special case of the more general framework presented here where we allow for both time-discounting and risk-adjustment to influence the rate at which future pay-offs are discounted.⁵

D. *The speculative component in bond prices*

We will define the speculative component in the price of a bond as the difference between the actual price p_t^n and the counterfactual “consensus” price \bar{p}_t^n . As in Allen, Morris and Shin (2006), the consensus price is the hypothetical price a bond would have if by chance, all agents’ higher order expectations about future discount rates coincided with the current first order expectations of the average agent (while holding conditional variances fixed). Iterating the price equation (11) forward under this assumption gives the consensus price of a bond as a sum of the average agent’s first order expectations about future discount rates

$$\bar{p}_t^n = \sum_{s=1}^n \int E \left[m_{t+s}^j \mid \Omega_t^j \right] dj + \frac{1}{2} \sum_{s=0}^{n-1} Var \left(m_{t+1+s}^j + p_{t+1+s}^{n-1-s} \mid \Omega_{t+s}^j \right) \quad (16)$$

By subtracting the consensus price (16) from the actual price (15) we can write the speculative term $p_t^n - \bar{p}_t^n$ as a sum of higher order prediction errors about future discount factors

$$\begin{aligned} p_t^n - \bar{p}_t^n &= - \int E \left[m_{t+1}^j - m_{t+1}^j \mid \Omega_t^j \right] dj \\ &\quad - \int E \left[m_{t+2}^j - \int E \left[m_{t+2}^i \mid \Omega_{t+1}^i \right] di \mid \Omega_t^j \right] dj - \dots \\ &\quad \dots - \int E \left[m_{t+n}^j - \int E \left[\dots \int E \left[m_{t+n}^i \mid \Omega_{t+n-1}^i \right] di' \dots \right] di \mid \Omega_t^j \right] dj \end{aligned} \quad (17)$$

Stochastic discount factors generally have a time discount and a risk adjustment component and the price of the asset depends negatively on both. This means that the speculative component (17) will be positive if individual agents, on average, think that other agents underestimate either future risk-free interest rates or future risk premia.

Since the speculative term is caused by individual agents believing that other agents either over- or

⁵The Internet Appendix contains a brief description of the equilibrium model of Nimark (2012) and exactly how that model maps into the affine framework presented in Section II.

underestimate future discount rates, it must also be orthogonal to public information in real time. That is, it is not possible for individual agents to predict other agents' forecast errors using public information. This result is stated more formally in Proposition 2.

PROPOSITION 2: *The speculative term $p_t^n - \bar{p}_t^n$ is orthogonal to public information in real time, i.e.*

$$E([p_t^n - \bar{p}_t^n] \omega_t) = 0 : \forall \omega_t \in \Omega_t \quad (18)$$

where Ω_t is the public information set defined as in Proposition 1.

Proof. In the Appendix. □

As in Proposition 1, the proof follows directly from taking expectations of (17) conditional on the public information set Ω_t and using that $\Omega_t \subseteq \Omega_{t+s}^j$ for every j and $s > 0$.

So far, we have derived some general results that should hold in any SDF based asset pricing model with heterogeneously informed agents that form model consistent expectations. In particular, we have shown that both the component of an agent's expected excess return that is due to information heterogeneity and the speculative component in bond prices must be orthogonal to public information in real time. The next section develops an explicit affine no-arbitrage model with these properties.

II. An affine term structure model with heterogeneous information

This section describes an operational framework for arbitrage-free asset pricing where agents have heterogeneous information relevant for predicting future bond returns. The basic set-up follows a large part of the affine term structure literature (see Duffie and Kan 1996 and Dai and Singleton 2000) and posits that the short rate r_t is an affine function of a vector of exogenous state variables.

Allowing for heterogeneously informed agents necessitates two changes in terms of how the model is specified and solved relative to the standard full information set-up. The first is that we need to specify a functional form for the individual agents' SDFs that allows for heterogeneity in expected returns. Below we propose a form that is analogous to the standard formulation under full information and, indeed, nests the standard formulation when signals reveal the exogenous state perfectly. This strategy allows for a flexible empirical specification while nesting more restricted equilibrium models such as the model in Nimark (2012)

as a special case.

As explained in Section I, heterogeneous information sets make it necessary for agents to “forecast the forecasts of others”, e.g. Townsend (1983). The exogenous factors are then no longer a complete description of the state of the model. Instead, we need to expand the state vector to also include higher order expectations of the factors, i.e. expectations about other agents’ expectations about the factors. The law of motion for the higher order expectations of the factors has to be determined jointly with bond prices since agents use the information contained in bond prices to form expectations about the unobservable factors and about the expectations of other agents. Heterogeneous information thus introduces an additional step in deriving a process for bond prices that is not present in the full information set-up with only exogenous state variables. To solve the model, we employ the approximation method proposed in Nimark (2011).

The end product of this section is an equation that describes the price of an n -period bond as a function of the state of the model. To get there, we start by defining the state and by conjecturing a functional form for the bond price equation. We then describe the law of motion of the state and how it is determined partly by the information sets available to agents. Taking the law of motion of the state and the conjectured bond price equation as given, it is straightforward to determine the risk associated with holding bonds. This risk can then be priced by the specified stochastic discount factor.

A. *The conjectured processes for bond prices and the state*

Following the affine term structure literature, the one period risk-free rate r_t is an affine function of the state variables x_t

$$r_t = \delta_0 + \delta'_x x_t. \quad (19)$$

The d -dimensional vector x_t follows a first order vector auto regression

$$x_{t+1} = \mu^P + F^P x_t + C \varepsilon_{t+1} : \varepsilon_{t+1} \sim N(0, I). \quad (20)$$

In a full information setting, we would normally proceed by specifying a functional form for the stochastic discount factor that would allow us to derive the price of a bond of any maturity as an affine function of the factors x_t . In our heterogeneous information set-up the factors determining the short rate are not directly observable by the agents. Instead, agents observe the signal vector x_t^j , which is the sum of the true vector

x_t and an idiosyncratic noise component

$$x_t^j = x_t + Q\eta_t^j : \eta_t^j \sim N(0, I) \quad (21)$$

where the noise shocks η_t^j are independent across agents. The vector x_t^j is the source of agent-specific information about the unobservable exogenous state x_t . The precision of the signals x_t^j is common across agents and determined by the matrix Q . Apart from the dimensionality of the vector x_t , this specification of agents' private signals is completely analogous to those in for instance Admati (1985), Singleton (1987) and Allen, Morris and Shin (2006).

Agents cannot by direct observation distinguish between idiosyncratic noise shocks η_t^j and the common factors x_t . This implies that an innovation to x_t is partly attributed to idiosyncratic sources so that on average, agents under-react to innovations to the factors. The presence of the idiosyncratic shocks thus changes the responses of expectations and bond prices to innovations in x_t , even though they average to zero in the cross-section.

As explained in Section I, agents' expectations about future bond prices depend on their expectations about other agents' future expectations about other agents' discount rates further into the future. These higher order expectations of future discount rates can be reduced to higher order expectations about the current latent factors x_t . The relevant state of the model can be shown to be the hierarchy of higher order expectations X_t defined as

$$X_t \equiv \begin{bmatrix} x_t \\ x_t^{(1)} \\ \vdots \\ x_t^{(\bar{k})} \end{bmatrix} \quad (22)$$

where the average k order expectations $x_t^{(k)}$ is defined recursively as

$$x_t^{(k)} \equiv \int E \left[x_t^{(k-1)} \mid \Omega_t^j \right] dj. \quad (23)$$

The integer \bar{k} is the maximum order of expectations considered. Nimark (2011) demonstrates that a finite \bar{k} is sufficient to approximate the equilibrium dynamics accurately.

We will conjecture (and later verify) that the state X_t follows a first order vector auto regression

$$X_{t+1} = \mu_X + \mathcal{F}X_t + \mathcal{C}u_{t+1} : u_{t+1} \sim N(0, I) \quad (24)$$

where u_{t+1} is a vector containing all aggregate shocks, i.e. those shocks that either affect the true state x_t or the average (higher order) expectations of x_t . If the variance of the idiosyncratic noise shocks η_t^j is zero, the signal vector x_t^j will reveal the factors x_t perfectly. The higher order expectations in the hierarchy X_t then coincide with the true factors x_t and each element of X_t then follows the law of motion (20) of the exogenous factors.

The price of a bond with maturity n is conjectured (and later verified) to be an affine function of the state X_t plus a maturity specific disturbance v_t^n

$$p_t^n = A_n + B_n' X_t + v_t^n. \quad (25)$$

That is, bond prices depend on the exogenous factors x_t as well as on the average higher order expectations of these factors. The maturity specific disturbance v_t^n prevents bond prices from revealing the expectations of other agents. The shocks v_t^n thus play a similar role as the random supply shocks arising from noise traders in Admati (1985). Since all agents use prices to extract information about the state X_t , the maturity specific shocks affect the average expectation about the state vector. The maturity specific shocks are thus part of the vector u_{t+1} in (24).

In equilibrium models where agents solve an explicit portfolio problem, a positive supply shock makes the price of an asset fall since a higher expected excess return is necessary to convince risk averse agents to absorb the additional supply into their portfolios (e.g. Admati 1985 and Singleton 1987). The higher expected excess return due to the increased supply is thus compensation that agents require for holding a riskier portfolio with a larger share of the risky asset. The SDF based framework presented here is consistent with this interpretation of the maturity specific disturbances v_t^n , though we do not model the portfolio decisions of agents explicitly.⁶

It is perhaps worth pointing out here that even though the state vector is high dimensional, this by itself will not increase our degrees of freedom in terms of fitting bond yields. The fact that the endogenous

⁶In the Internet Appendix we demonstrate that the equilibrium model in Nimark (2012) is a special case of the affine framework derived here in which supply shocks enter the equilibrium price in exactly the same way as the maturity specific shocks v_t^n in (25).

state variables $x_t^{(k)}$ are rational expectations of the lower order expectations in $x_t^{(k-1)}$ disciplines the law of motion (24) and the matrices \mathcal{F} and \mathcal{C} are completely pinned down by the parameters of the process governing the true exogenous factors x_t and the parameters that govern how precise agents' signals about x_t are. How to find the matrices \mathcal{F} and \mathcal{C} in practice is described in Appendix C.

B. Agents' filtering problem

To form expectations about future bond prices, agents need to form an estimate of the current aggregate state X_t . Agents know the law of motion of the state X_t as well as how the state maps into the vector of observable variables and since the model is linear with Gaussian shocks, the Kalman filter delivers optimal state estimates. In each period agents observe the short rate r_t , a vector of current bond yields with maturity up to \bar{n}

$$y_t \equiv \left[\frac{1}{2}p_t^2 \quad \dots \quad n^{-1}p_t^n \quad \dots \quad \bar{n}^{-1}p_t^{\bar{n}} \right]' \quad (26)$$

as well as the agent-specific signals x_t^j . The variables observable to agent j can be collected in the vector z_t^j defined as

$$z_t^j \equiv \begin{bmatrix} x_t^j \\ r_t \\ y_t \end{bmatrix} \quad (27)$$

Through observing equilibrium bond yields, agents extract information about the unobservable state of the economy which partly consists of the expectations of other agents. This contrasts with difference-in-beliefs models where agents “agree to disagree”. When agents agree to disagree, the beliefs of all agents are common knowledge and from the agents' perspective, there is no additional information about other agents' beliefs in the endogenous yields.

Agents do not forget and the information set of agent j is the filtration defined by

$$\Omega_t^j = \left\{ z_t^j, \Omega_{t-1}^j \right\} \quad (28)$$

The law of motion of the state (24) and the definition of the observables (27) then let us describe the filtering

problem of agent j as a standard state space system

$$X_{t+1} = \mu_X + \mathcal{F}X_t + \mathcal{C}u_{t+1} \quad (29)$$

$$z_t^j = \mu_z + DX_t + R \begin{bmatrix} u_t \\ \eta_t^j \end{bmatrix} \quad (30)$$

The vector of constants μ_z and the matrices D and R in the measurement equation (30) are defined in Appendix C. Given the state space system (29) - (30), agent j 's state estimate evolves according to the Kalman filter updating equation

$$E \left[X_t \mid \Omega_t^j \right] = (I - KD) \mathcal{F} E \left[X_{t-1} \mid \Omega_{t-1}^j \right] + K z_t^j \quad (31)$$

where K is the Kalman gain. Since bond yields are part of the observation vector z_t^j , the matrix D in the measurement equation (30) is partly a function of the vectors B_n in the conjectured bond price equation (25). This implies that we have to solve simultaneously for the filtering problem and the pricing equation (25).

By the definition (22) of the state X_t , the state is partly made up of the cross-sectional average of the expectation of the state, i.e. by the cross-sectional average of the update equation (31). The Kalman filter thus plays a dual role in the model: It determines both agents' expectations about the state as well as the law of motion (24) of the very same state that the agents form expectations about.

C. *The stochastic discount factor of agent j*

Agents want to be compensated for the risk associated with holding bonds. The conjectured bond price equation (25) implies that this risk arises from uncertainty about future states X_t and from future realizations of the maturity specific shocks v_t^n . Here we specify the agents' stochastic discount factors that will be used to price this risk.

The stochastic discount factor of agent j is denoted M_{t+1}^j . In the absence of arbitrage, the relationship

$$p_t^{n+1} = \log E \left[M_{t+1}^j P_{t+1}^n \mid \Omega_t^j \right] \quad (32)$$

must be satisfied for each agent j and maturity n . Following the full information affine asset pricing literature closely, we specify the logarithm of agent j 's stochastic discount factor to follow

$$m_{t+1}^j = -r_t - \frac{1}{2}\Lambda_t^{j'\Sigma_a\Lambda_t^j} - \Lambda_t^{j'}a_{t+1}^j : a_{t+1}^j \sim N(0, \Sigma_a) \quad (33)$$

The vector a_{t+1}^j is defined to span the conditional one-period-ahead forecast errors of agent j for each maturity n . That is, for each maturity n , agent j 's forecast error can be written as a linear function of the vector a_{t+1}^j

$$p_{t+1}^{n-1} - E[p_{t+1}^{n-1}|\Omega_t^j] = \psi_{n-1}a_{t+1}^j \quad (34)$$

and the vector a_{t+1}^j thus spans the risk that agent j requires compensation for. Given the conjectured bond price equation (25), agent j 's forecast in period t of a bond price in period $t + 1$ can be incorrect ex post either because his forecast of the state X_{t+1} was incorrect or because of a maturity specific shock v_{t+1}^n (which by construction is unpredictable based on period t information). Both the state forecast error and the maturity specific shocks thus need to be included in the vector a_{t+1}^j

$$a_{t+1}^j \equiv \begin{bmatrix} X_{t+1} - E(X_{t+1}|\Omega_t^j) \\ v_{t+1} \end{bmatrix} \quad (35)$$

where

$$v_t \equiv \begin{bmatrix} v_t^2 & \dots & v_t^{\bar{n}} \end{bmatrix}'. \quad (36)$$

Given these definitions, the row vector ψ_{n-1} that maps the vector a_{t+1}^j into agent j 's one period ahead forecast error of the price of an $n - 1$ period bond is given by

$$\psi_1 \equiv \begin{bmatrix} B_1' & 0 \end{bmatrix} \quad (37)$$

if $n = 2$ (since there is no maturity-specific shock in the risk-free one period bond's price) and

$$\psi_{n-1} \equiv \begin{bmatrix} B_{n-1}' & e'_{n-2} \end{bmatrix} \quad (38)$$

if $n > 2$. The vector e_n has a one in the n^{th} element and zeros elsewhere.

The vector of risk prices Λ_t^j in agent j 's stochastic discount factor (33) is assumed to be an affine function of the agent-specific state X_t^j and the maturity specific shocks v_t

$$\Lambda_t^j = \Lambda_0 + \Lambda_X X_t^j + \Lambda_v v_t. \quad (39)$$

The agent-specific state X_t^j consists of the vector of agent j specific exogenous factors x_t^j as well as of agent j 's expectations (up to order \bar{k}) about the latent vector x_t

$$X_t^j \equiv \begin{bmatrix} x_t^j \\ E[x_t | \Omega_t^j] \\ \vdots \\ E[x_t^{(\bar{k}-1)} | \Omega_t^j] \end{bmatrix}. \quad (40)$$

The vector X_t^j determines both agent j 's required compensation for risk as well as his expectations about future bond prices.

D. The bond price recursions

We now have all the ingredients needed to find A_n and B_n in the conjectured bond price equation (25). Start by substituting in the expressions (33) for the SDF into the no-arbitrage condition (32) to get

$$p_t^n = \log E \left[\exp \left(-r_t - \frac{1}{2} \Lambda_t^{j'} \Sigma_a \Lambda_t^j - \Lambda_t^{j'} a_{t+1}^j + p_{t+1}^{n-1} \right) | \Omega_t^j \right] \quad (41)$$

We will substitute out the price p_{t+1}^{n-1} from (41) via three intermediate steps. First, use the definition (34) to write p_{t+1}^{n-1} as the sum of agent j 's expectations about the price and his forecast error

$$p_{t+1}^{n-1} = E \left[p_{t+1}^{n-1} | \Omega_t^j \right] + \psi_{n-1} a_{t+1}^j \quad (42)$$

Second, note that the conjectured price equation (25) and the law of motion of the state (24) together with rational expectations imply that agent j 's expectation of the next period price can be expressed as a function

of his expectations about the current state, i.e.

$$E \left[p_{t+1}^{n-1} | \Omega_t^j \right] = A_{n-1} + B'_{n-1} \mu_X + B'_{n-1} \mathcal{F} E \left[X_t | \Omega_t^j \right]. \quad (43)$$

Third, agent j 's expectation of the current state can by definition (40) be written as

$$E \left[X_t | \Omega_t^j \right] = H X_t^j \quad (44)$$

where H is the average expectations operator $H : \mathbb{R}^{d(\bar{k}+1)} \rightarrow \mathbb{R}^{d(\bar{k}+1)}$

$$H \equiv \begin{bmatrix} \mathbf{0} & I_{d\bar{k}} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \quad (45)$$

The matrix H increases each order of expectation in a hierarchy by annihilating the zero order expectation and replacing it with the first order expectation and by replacing the first order expectation with the second order expectation, and so on.

The expressions (42) - (44) can then be used in reverse order to substitute out p_{t+1}^{n-1} from (41). After simplifying the resulting expression we get

$$\begin{aligned} p_t^n &= -r_t + A_{n-1} + B'_{n-1} \mu_X + B'_{n-1} \mathcal{F} H X_t^j \\ &\quad + \frac{1}{2} \psi_{n-1} \Sigma_a \psi'_{n-1} - \psi_{n-1} \Sigma_a \Lambda_t^j. \end{aligned} \quad (46)$$

That is, the price of an n -period bond is a function of the risk-free interest rate r_t , a number of constants and terms specific to agent j . The no-arbitrage condition (32) has to hold for all agents at all times. This implies that we could choose any agent j 's state X_t^j as being the state variable that bond prices are a function of. However, the most convenient choice from a modeling perspective is to let bonds be priced by the SDF of the fictional agent whose state X_t is defined to coincide with the cross-sectional average state so that

$$X_t \equiv \int X_t^j dj \quad (47)$$

The identity of the average agent will change over time as idiosyncratic shocks change an individual agent's relative position in the cross-sectional distribution. However, the identity of the average agent is of no

consequence and the advantage of letting the average agent's SDF price bonds is that it allows us to write log bond prices in the conjectured form (25), i.e. as a function of the average state X_t . We can thus substitute in X_t for X_t^j in (39) and (46). The last step required to find the conjectured form of the bond price equation is to use the process for the short rate (19) to replace r_t in (46). After simplifying, we get

$$\begin{aligned} p_t^n &= -\delta_0 + A_{n-1} + B'_{n-1}\mu_X + \frac{1}{2}\psi_{n-1}\Sigma_a\psi'_{n-1} - \psi_{n-1}\Sigma_a\Lambda_0 \\ &\quad -\delta_X X_t + B'_{n-1}\mathcal{F}H X_t - \psi_{n-1}\Sigma_a\Lambda_x X_t \\ &\quad -\psi_{n-1}\Sigma_a\Lambda_v v_t \end{aligned} \quad (48)$$

where

$$\delta'_X = \begin{bmatrix} \delta'_x & \mathbf{0} \end{bmatrix}$$

The bond price recursions for A_{n+1} and B_{n+1} in the bond price equation (25) are thus given by

$$A_{n+1} = -\delta_0 + A_n + B'_n\mu_X + \frac{1}{2}\psi_n\Sigma_a\psi'_n - \psi_n\Sigma_a\Lambda_0 \quad (49)$$

and

$$B'_{n+1} = -\delta_X + B'_n\mathcal{F}H - \psi_n\Sigma_a\Lambda_x \quad (50)$$

As in a full information set-up, the recursions (49) and (50) can be started from

$$A_1 = -\delta_0 \quad (51)$$

$$B_1 = -\delta'_X \quad (52)$$

where $p_t^1 = -r_t$.⁷

Readers familiar with the standard affine model will recognize that the recursive expressions for A_n and B'_n above are completely analogous to the corresponding expressions in the standard full information model. Replacing Σ_a by CC' , $\mathcal{F}H$ by F^P and ψ_{n-1} by B'_{n-1} delivers the standard expressions. The interpretation

⁷Appendix B contains a step-by-step derivation of the bond price recursions. By recursive substitution of (50) it is possible to also express the price of an n period bond as an explicit function of higher order expectations, i.e. as $p_t^n = A_n - \delta_X \sum_{s=0}^{n-1} (\mathcal{F}H)^s X_t - \psi_{n-1-s}\Sigma_a\Lambda_x \sum_{s=0}^{n-2} (\mathcal{F}H)^s X_t$. When multiplied by the state vector X_t , the matrix $\mathcal{F}H$ moves expectations one period forward in time and one step up in order of expectations so that the term $-\delta_X \sum_{s=0}^{n-1} (\mathcal{F}H)^s X_t$ is the cumulative sum of higher order expectations about future short rates and the term $\psi_{n-1-s}\Sigma_a\Lambda_x \sum_{s=0}^{n-2} (\mathcal{F}H)^s X_t$ is the cumulative sum of higher order expectations about future risk premia.

of the corresponding matrices are also the same. Both Σ_a and CC' are the covariance of the vector of risk that agents require compensation for. The only difference is that in the model presented here, the risk of holding bonds arises not only from innovations to the true factors x_t but also from current state uncertainty, future innovations to higher order expectations and maturity specific shocks. Similarly, both $\mathcal{F}H$ and F^P are matrices that agents use to form expectations about the next period's state, conditional on the current state. Finally, both ψ_{n-1} and B'_{n-1} are vectors that translate innovations to the respective risk vector a_{t+1}^j and $C\varepsilon_{t+1}$ into innovations to bond prices.

E. Restricting Λ_t^j

In the conjectured bond price equation (25) the maturity specific shock v_t^n enters the price function with a unit coefficient. In Appendix B we show that setting

$$\Lambda_v = -(\Psi\Sigma_a)_{right}^{-1} \quad (53)$$

where $(\cdot)_{right}^{-1}$ denotes the (right) one-sided inverse of a matrix and

$$\Psi = \begin{bmatrix} \psi_1 \\ \vdots \\ \psi_{n-1} \end{bmatrix} \quad (54)$$

ensures that the price equation (48) above is consistent with the conjectured price equation. Specifying Λ_v as in (53) has the additional advantage that letting Λ_t^j depend on v_t^n does not introduce additional free parameters relative to the standard model without maturity specific shocks.

F. Solving the model

Solving the model implies finding the matrices \mathcal{F} and \mathcal{C} in (24) and A_n and B'_n in (25). Since the law of motion of the state depends on the bond price equation through the filtering problem of the agents and because the bond price equation in turn depends on the law of motion of the state, it is necessary to solve for (24) and (25) simultaneously. Appendix C describes how the method proposed in Nimark (2011) can be

adapted to find a fixed point of this mapping.

III. Speculation in the affine model

Section I derived some general implications of heterogeneous information for the relationship between stochastic discount factors, agents' expectations and asset prices. There, the speculative component in a bond's price was defined as the difference between the actual price, which depends on higher order expectations about future bond prices (or discount factors), and the counterfactual price a bond would have if these higher order expectations coincided with the average first order expectation. In order to quantify the speculative component using the affine model, we thus need to operationalize the counterfactual consensus price.

A. *The counterfactual consensus price*

In the affine model presented above, the forecasting problem of predicting other agents' future expectations about bond prices and discount factors can be reduced to forming expectations about other agents' expectations about the current state X_t . The state summarizes all information that is possible to know about future states, so perceived agreement about the current state implies perceived agreement about expected future states. This means that if, by chance, an individual agent's first and higher order expectations about the state x_t coincide, the agent believes that other agents share his predictions about future bond prices. There will then be no speculative motive for trade, since the agent then believes that other agents will only be willing to pay as much for the bond in the future as he expects himself to be willing to pay, were he to hold on to the bond until maturity.

We can thus specify the counterfactual consensus price \bar{p}_t^n as the price of an n period bond that would prevail if average first and higher order expectations about the latent state x_t coincided. It can be computed as

$$\bar{p}_t^n = A_n + B_n' \bar{H} X_t + v_t^n \tag{55}$$

where the matrix \overline{H} is what we call the *consensus operator* $\overline{H} : \mathbb{R}^{d(\bar{k}+1)} \rightarrow \mathbb{R}^{d(\bar{k}+1)}$ defined so that

$$\begin{bmatrix} x_t \\ x_t^{(1)} \\ \vdots \\ x_t^{(1)} \end{bmatrix} = \overline{H} \begin{bmatrix} x_t \\ x_t^{(1)} \\ \vdots \\ x_t^{(\bar{k})} \end{bmatrix} \quad (56)$$

That is, \overline{H} is a matrix that takes a hierarchy of expectations about x_t and equates higher order expectations with the first order expectation.

B. The speculative component in bond prices

We can use \overline{H} to decompose the current n period bond price into a component that depends only on the average first order expectation and a speculative component that is the difference between the actual price and the counterfactual consensus price \overline{p}_t^n . By adding and subtracting the consensus price (55) to the right hand side of the bond price equation (25) we get the expression

$$p_t^n = A_n + B_n' \overline{H} X_t + \underbrace{B_n' (I - \overline{H}) X_t}_{\text{speculative term}} + v_t^n \quad (57)$$

since

$$p_t^n - \overline{p}_t^n = B_n' (I - \overline{H}) X_t \quad (58)$$

The price of an n -period bond can thus be written as a sum of commonly known components and a simple expressions capturing the difference between the actual price and the answer you would get if you asked the “average” agent what he thinks the price would be if all agents, by chance, had the same state estimate as he did (while holding conditional uncertainty constant).⁸

It follows from Proposition 2 that the speculative component must be orthogonal to public information in real time and the intuition is straightforward: By construction, $(I - \overline{H}) X_t$ is a vector of higher order predictions errors, i.e. a vector of differences between first and higher order expectations about the latent state x_t . Since it is not possible to predict other agents’ errors using publicly available information, any

⁸Bacchetta and van Wincoop (2006) refers to the equivalent object in their model as the “higher order wedge”.

linear function of $(I - \bar{H}) X_t$ must be orthogonal to public information in real time.

If signals are noisy, differences between agents' first and higher order expectations about the current state translate into differences between first and higher order expectations about future states. That is, if an individual agent believes that other agents have a different estimate of the current state than he does, then it is rational to believe that other agents will also have different expectations from himself in the future, unless future signals are expected to perfectly reveal the state. It is also rational for an individual agent to expect other rational agents' expectations to be revised in the future when more signals are observed towards what the individual agent thinks is a better prediction about the future states. That is, second (and higher) order expectations are not martingales, but are predicted to be revised towards an agent's best prediction, i.e. his first order expectation, which for the usual reasons is a martingale.

It is the fact that differences between first and higher order expectations are expected to be persistent that induces speculative behavior in the model. If, for instance, it was common knowledge that everybody would observe a perfect signal about the state in the next period, there would be no motive to speculate since it would also be common knowledge that all agents would share the same valuation of the asset in the next period.

C. A three-way decomposition of the yield curve

Below, we will quantify the importance for bond yield dynamics of the speculative component derived above. While much of the focus in this paper is on the speculative component itself, it is also of interest to investigate how allowing for heterogeneous information may change our estimates of the classical components of the yield curve, i.e. short rate expectations and risk premia. To compare the implied estimates of (first order) short rate expectations and risk premia from our model to those produced by a standard affine common information model, we need to decompose the non-speculative component in (57) further. What we want is a decomposition of the form

$$p_t^n = \underbrace{A_n^{rp} + B_n^{rp'} X_t}_{\text{classic risk premia}} + \underbrace{A_n^r + B_n^{r'} X_t}_{\text{short rate expectations}} + \underbrace{B_n' (I - \bar{H}) X_t}_{\text{speculative term}} + v_t^n. \quad (59)$$

The classic risk premia terms can be found by subtracting the average first order expectations about future short rate expectations from the non-speculative component in (57). This implies that the scalar A_n^{rp} and the

vector B_n^{rp} in (59) are given by

$$A_n^{rp} = A_n - A_n^r, \quad B_n^{rp'} = B_n' \bar{H} - B_n^{r'}$$

where A_n^r and B_n^r that determine the average first order expectations about future short rates are given by

$$A_n^r = -n(\delta_0 + \delta_X \mu_X) - \delta_X \sum_{s=0}^{n-1} \mathcal{F}^s \mu_X, \quad B_n^{r'} = -\delta_X \sum_{s=0}^{n-1} \mathcal{F}^s H.$$

The first two terms in (59) thus corresponds to the classic terms of the yield curve decomposition in Cochrane and Piazzesi (2008) and Joslin, Singleton and Zhu (2011) and are independent of any discrepancy between first and higher order expectations. In the limit with perfectly precise signals, the speculative term tend to zero since both first and higher order expectations about x_t then coincide with the true factors. The two classical terms, together with the maturity specific shocks, would then determine bond yields completely.

IV. Empirical specification

In order to make the model presented in Section II operational we will need to be specific about some of the details that up until this point have been presented at a more general level. Here, we describe how the factor processes are normalized and how the prices of risk can be parameterized parsimoniously when higher order expectations enter as state variables. In this section we also describe how the cross-sectional dispersion of the individual responses in the Survey of Professional Forecasters can be exploited in likelihood based estimation of the model's parameters.

A. Exogenous factor dynamics and the risk-free interest rate

The first choice to be made is to decide how many factors to include in the exogenous vector x_t . In the estimated specification, x_t is a three dimensional vector so that in the special case with perfectly informed agents and no maturity specific shocks, the model collapses to a standard three factor affine Gaussian no-arbitrage model. Since the factors are latent we need to normalize their law of motion. We follow Joslin, Singleton and Zhu (2011) and let the risk neutral dynamics of the factors follow a first order vector autore-

gressive process

$$x_{t+1} = \mu^Q + F^Q x_t + C \varepsilon_{t+1}^Q \quad (60)$$

with the restrictions that $\mu^Q = 0$ and that the matrix F^Q is diagonal with the factors ordered in descending degree of persistence under the risk neutral dynamics. Furthermore, C is restricted to be lower triangular. Finally, δ_x in the short rate equation (19) is a vector of ones. These restrictions ensure that all parameters are identified in the special case of perfectly informed agents and no maturity specific shocks.

B. Parameterizing the prices of risk

Following the full information affine literature as closely as possible, we specify agent j 's vector of risk prices as an affine function of the agent-specific state X_t^j

$$\Lambda_t^j = \Lambda_0 + \Lambda_X X_t^j + \Lambda_v v_t \quad (61)$$

The state vector X_t^j is high dimensional and, as a consequence, leaving Λ_0 and Λ_X completely unrestricted would result in a very large number of free parameters. To avoid an over-parameterized model we therefore restrict Λ_0 and Λ_X as follows

$$\Lambda_t^j = \begin{bmatrix} \lambda_0 \\ \mathbf{0} \end{bmatrix} + \begin{bmatrix} \lambda_x & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} X_t^j + \Lambda_v v_t \quad (62)$$

where λ_0 is a 3×1 vector and λ_x is a 3×3 matrix. Restricting Λ_t^j this way also implies that the model nests the standard specification if agents' signals are perfectly precise and the variance of v_t equals zero. The matrix Λ_v does not contain any freely estimated parameters and is given by (53). The number of estimated parameters in Λ_t^j is thus the same as in the price of risk specification in a standard gaussian full information three-factor model, e.g. Duffee (2002) and Joslin, Singleton and Zhu (2011).

The empirical specification is parameterized by the matrices F^Q and C which govern the processes of the latent factors x_t , the diagonal matrix Q which specifies the standard deviation of the idiosyncratic noise in the agent-specific signals about x_t , the constant δ_0 in the risk-free short rate equation (19), σ_v the standard deviation of the maturity specific disturbances v_t^n (specified so that $\sqrt{\text{var}(v_t^n)} = n\sigma_v$, i.e. so that the standard deviation of the impact on *yields* is constant across maturities) and the vector λ_0 and matrix λ_x

which govern risk premia. The model has 28 parameters in total and relative to a canonical full information three-factor affine model, the only additional parameters are the three diagonal elements of Q that govern the precision of the agent-specific signals.

C. *Implied physical dynamics*

The physical dynamics of the factors are implicitly defined by the combination of the risk neutral dynamics (60) and the prices of risk vector (61). The vector μ^P and the coefficient matrix F^P in (20) are given by

$$\mu^P = \mu^Q + CC'\lambda_0, \quad F^P = F^Q + CC'\lambda_x \quad (63)$$

In the limit with perfectly precise signals, the risk neutral and physical dynamics of the affine model have the usual interpretation: While the latent factors follow the physical dynamics, bonds can be priced *as if* agents were risk neutral and the factors followed the risk neutral dynamics. The physical dynamics then also completely determine the law of motion of the extended state X_t .

D. *Agents' information sets*

Agent j observes the factors x_t^j as defined in (21) which is the source of agent j 's heterogeneous information about the common factors x_t . Each agent also observes the risk-free short rate r_t . In addition to these exogenous signals, all agents can observe all bond yields up to maturity \bar{n} , where \bar{n} is the largest maturity used in the estimation of the model. Here, the longest maturity yield that we will use in estimation is a 10 year bond implying that $\bar{n} = 40$ with quarterly data.

E. *Choosing the maximum order of expectation \bar{k}*

In Nimark (2011) it is demonstrated that, under quite general conditions, it is possible to accurately represent the dynamics of an infinite horizon model with heterogeneously informed agents by a finite dimensional state vector, despite of the infinite regress of “forecasting the forecasts of others” that arises in models where heterogeneously informed agents need to predict the future actions of other agents (e.g. Townsend 1983). In our set-up, bonds are finitely lived so the price of a bond depends only on a finite

number of higher order expectations about future discount rates. It is possible to use this fact to write down an exact representation of the equilibrium dynamics in which agents' higher order expectations about future discount rates make up the state of the model. However, it is more tractable to let the state be a hierarchy of a finite number of higher order expectations about the current exogenous factors x_t as in (22). What "finite" means in practice has to be checked on a case by case basis. In the final specification used for estimation, we set the maximum order of expectation \bar{k} equal to 40. This is more than sufficient as most of the dynamics are captured by the first five orders of expectations.

In this paper, we model agents as explicitly forming higher order expectations, i.e. expectations about other agents' expectations and the equilibrium representation can be interpreted as a literal description of agents' behavior. Given the prevalence of quotes of Keynes' beauty contest metaphor in the finance literature it appears that many people find the related intuition appealing. However, it may strain credulity to think that agents form expectations beyond two or three orders and here we solve the model by assuming that agents form up to the 40th order of expectations. An alternative interpretation is to view the equilibrium representation simply as a convenient recursive functional form to model agents who have access to heterogeneous information and condition on the entire history of observables to predict next period bond yields. The main advantage of the method is then to deliver a tractable and time invariant recursive representation of the equilibrium dynamics of the model.

F. Estimating the model using bond yields and survey data

The parameters of the model can be estimated by likelihood based methods. We use quarterly data on bond yields with one, five and ten years to maturity with the sample spanning the period 1971:Q4 to 2011:Q4. The zero-coupon yield data is taken from the Gurkaynak, Sack and Wright (2007) data set available from the Federal Reserve Board. In addition to bond yields we also use one quarter ahead forecasts of the T-Bill rate and the 1 quarter ahead forecasts of the 10 year bond rate from the Survey of Professional Forecasters (SPF). The individual survey responses are collected in the vectors $y_{t+1|t}^1$ and $y_{t+1|t}^{40}$. In the model, the cross-sectional distribution of agents' one-period-ahead forecasts of the risk-free short rate is Gaussian with mean and variance given by

$$E \left[r_{t+1} \mid \Omega_t^j \right] \sim N \left(A_1 + B_1' \mu_X + B_1' \mathcal{F} H X_t, B_1' \mathcal{F} \Sigma_j \mathcal{F}' B_1 \right) \quad (64)$$

where Σ_j is the cross-sectional covariance of expectations about the current state, i.e.

$$\Sigma_j \equiv E \left[H \left(X_t^j - X_t \right) \left(X_t^j - X_t \right)' H' \right] \quad (65)$$

As econometricians, we can thus treat the individual survey responses of T-Bill rate forecasts as noisy measures of the average expectation of the short rate r_t where the variance of the “noise” is determined by the model implied cross-sectional variance of short rate expectations.⁹ The corresponding distribution for the one-period-ahead forecast of the 10 year yield is

$$E \left[y_{t+1}^{40} \mid \Omega_t^j \right] \sim N \left(\frac{1}{40} A_{40} + \frac{1}{40} B'_{40} \mu_X + \frac{1}{40} B'_{40} \mathcal{F} H X_t, \frac{1}{40} B'_{40} \mathcal{F} \Sigma_j \mathcal{F}' B_{40} \frac{1}{40} \right) \quad (66)$$

The deviations of individual agents’ forecasts from the average forecasts are caused by idiosyncratic shocks that are independent across agents. The covariance of the cross-sectional “measurement errors” in $y_{t+1|t}^1$ and $y_{t+1|t}^{40}$ can thus be specified as the scalars $B'_1 \mathcal{F} \Sigma_j \mathcal{F}' B_1$ and $\frac{1}{40} B'_{40} \mathcal{F} \Sigma_j \mathcal{F}' B_{40} \frac{1}{40}$ multiplied by an identity matrix.

G. The likelihood function

Given the model and the data, the log likelihood function

$$\log L(\bar{z}^T) = -\frac{1}{2} \left\{ \sum_{t=1}^T 2\pi \dim(\tilde{z}_t) + \log \left| \tilde{\Sigma}_{t|t-1} \right| + \tilde{z}'_t \tilde{\Sigma}_{t|t-1}^{-1} \tilde{z}_t \right\} \quad (67)$$

can be evaluated by computing the Kalman filter innovations

$$\tilde{z}_t = \bar{z}_t - E \left[\bar{z}_t \mid \bar{z}^{t-1} \right] \quad (68)$$

from the state space system

$$X_t = \mu_X + \mathcal{F} X_{t-1} + \mathcal{C} \bar{u}_t \quad (69)$$

$$\bar{z}_t = \bar{\mu}_t^z + \bar{D}_t X_t + \bar{R}_t \bar{u}_t : \bar{u}_t \sim N(0, I) \quad (70)$$

⁹The Appendix contains details of how to compute the cross-sectional variance Σ_j in practice.

where \bar{z}_t is the vector of observables

$$\bar{z}_t = \left[y_t^A \quad y_t^{20} \quad y_t^{40} \quad y_{t+1|t}^1 \quad y_{t+1|t}^{40'} \right]' \quad (71)$$

and $\tilde{\Sigma}_{t|t-1}$ is the covariance of the innovations \tilde{z}_t . The vector $\bar{\mu}_t^z$ and the matrices \bar{D}_t and \bar{R}_t in the measurement equation (70) are defined in the Appendix.

The number of survey responses varies over time and surveys are not available at all for the period before 1981:Q3. Therefore, the dimensions of $\bar{\mu}_t^z$, \bar{D}_t and \bar{R}_t are also time-varying. This fact may influence the precision of our estimates of the state, i.e. we will have more precise estimates of the latent state X_t when there is a large number of survey responses available. Using individual survey responses and likelihood based methods also naturally incorporates that we have more precise information about the cross-sectional average expectations of agents when there are 50 responses (the sample maximum) compared to when there are only 9 responses (the sample minimum). This information is lost when using measures of central tendency like a mean or median forecast.

H. Estimation procedure

The next section will present empirical results based on the heterogeneous information model described above as well as the nested full information model without maturity specific shocks of Joslin, Singleton and Zhu (2011). To obtain parameter estimates for the full information model we first estimate a model without using survey data following the procedure in Singleton *et al* (2011) which reliably finds the maximum of the likelihood in a model estimated using only yields. Subsequently, we proceed to estimate the same full information model with both survey data and bond yields, taking the estimates from the yields-only estimates as starting values. The maximum likelihood estimates of the surveys plus yields model are found by first using a numerical optimizer and then the Metropolis-Hastings algorithm.¹⁰ To incorporate the surveys into the full information model we treat the individual responses as noisy measures of the model implied common expectation about future yield as in Kim and Orphanides (2005) and Chernov and Muller (2012).

The parameters of the heterogeneous information model is estimated using Bayesian methods with (improper) uniform priors so that the posterior mode coincides with the maximum likelihood estimates and the

¹⁰We experimented with a number of alternative starting values and optimization routines. Of these alternatives, the procedure reported here resulted in the highest posterior likelihood.

posterior density is proportional to the likelihood function. We take 1,000,000 draws from a Metropolis-Hastings algorithm (e.g. Geweke 2005) with the maximum likelihood estimates of the full information model as starting values for the parameter vector. The modes and the probability intervals presented in the next section are based on the last 500 000 draws.

V. Empirical Results

This section contains the main empirical results of the paper. Here, we first present the parameter estimates and discuss how these are influenced by the fact that individual survey responses are used in estimation. This is followed by a decomposition of historical bond yields into risk premia, first order expectations about the risk-free short interest rate and a speculative component driven by differences between first and higher order expectations. In this section we also compare estimates of historical risk premia and short rate expectations from our heterogeneous information model with estimates of the same quantities from the nested full information model of Joslin, Singleton and Zhu (2011). The section ends by quantifying the total effect of information imperfections and an assessment of how useful the agent-specific signals are for predicting excess returns.

A. *Parameter estimates and the dispersion of survey responses*

Table 1 presents the posterior modes along with 95% probability intervals (in square brackets). Since the factors are latent and have no particular economic interpretation, most of the estimated parameter values are of no particular interest when viewed in isolation. However, using the individual survey responses in estimation has interesting implications for those parameters in the model that govern how informed agents are about the latent factors. Before discussing these implications, we first note that the model does a good job of fitting the cross-sectional dispersion in the survey data. The model implied dispersion of the one-quarter-ahead forecasts of the short interest rate has a cross-sectional standard deviation of 43 basis points, compared to the 40 basis points sample average in the survey data. The model implied dispersion of the one-quarter-ahead forecasts of the 10 year yield is 27 basis points versus 40 basis points in the data. The model thus fits the cross-sectional dispersion in the survey data well.

The model also provides a good fit of bond prices. The standard deviation of the maturity specific shocks is 50 basis points which is comparable to the 37 basis points standard deviation of the pricing errors in the

full information model estimated using the same yields and survey data. The model implied unconditional yields are within a few basis points of the sample averages. The model is thus sufficiently flexible to explain both the dispersion in survey forecasts and bond yields well.

Using the full cross-section of surveys to estimate the parameters of the model will clearly influence the estimates of the parameters that determine the precision of the agent-specific signals. However, the relationship between the precision of the agent-specific signals and the model implied dispersion is non-monotonic. When the agent-specific signals are very precise, the cross-sectional dispersion is close to zero and the dynamics of bond yields will be close to those of the full information model. When the agent-specific signals are very imprecise, agents attach little weight to them, and again, the cross-sectional dispersion will be close to zero. To match the substantial dispersion observed in the survey responses, intermediate values for the parameters that govern the precision of the agent-specific signals are required. The estimates of Q_1 , Q_2 and Q_3 that govern the precision of the private signals thus cannot be neither too large, nor too small relative to the estimates of C_1 , C_2 and C_3 that govern the standard deviation of the innovations to the latent factors. At the posterior mode, the variance of the idiosyncratic noise in the agent-specific signals are between 2.5 to 4 times as large as the innovations to the true factors.

Less obviously, the cross-sectional dispersion in surveys will also discipline the dynamics of bond prices more generally. In our model, depending on the parameters, bond prices may or may not reveal the state perfectly. If bond prices are too revealing about the latent exogenous factors x_t , the cross-sectional dispersion will be too low relative to the dispersion in the survey data, regardless of the precision of the agent-specific signals. How informative bond yields are depends on how different the persistence of each factor is under the risk-neutral dynamics. The intuition is straightforward: If there is only one factor that is persistent enough under the risk neutral dynamics to move the long end of the yield curve, a change in long maturity yields can only be caused by a change in the high persistence factor. Similarly, if there is only one factor with very low persistence under the risk neutral measure, a change that is exclusive to the short end of the yield curve can only be caused by a change to the low persistence factor. More generally, observing bond yields will be very informative about the latent factors if each factor has a very different implication for the shape of the yield curve.

If simply observing the yield curve would be enough to get very precise estimates of the latent factors, agents would put little or no weight on their private signals. The cross-sectional dispersion of expecta-

Table I
Parameter estimates for the heterogeneous information model

The table reports parameter estimates for the affine heterogeneous information model obtained from the last 500,000 draws of a Metropolis Hastings algorithm. The data are a panel of yields with maturities of one- five- and ten-year as well as one quarter ahead individual survey forecasts of the 3 month treasury bill rate and the 10-year yield. The sample period is from 1971:Q2 to 2011Q4. The table reports the posterior mode for each parameter along with 95% probability intervals (in square brackets)

Factor processes and private signal standard deviations			
$F_{1,1}^Q$	0.9984 [0.9969 0.9986]	C_{21}	-0.0014 [-0.0032 0.0020]
$F_{2,2}^Q$	0.75629 [0.7483 0.7669]	C_{31}	-0.0134 [-0.0161 -0.0104]
$F_{3,3}^Q$	0.75624 [0.7479 0.7664]	C_{32}	-0.0935 [-0.1010 -0.0874]
C_1	0.0080 [0.0078 0.0086]	$Q_{1,1}$	0.0321 [0.0317 0.0333]
C_2	0.1088 [0.1031 0.1196]	$Q_{2,2}$	0.2661 [0.2087 0.3167]
C_3	0.0128 [0.0119 0.0140]	$Q_{3,3}$	0.0397 [0.0307 0.0440]
Maturity specific disturbances		Short rate constant r_t	
σ_v	0.0050 [0.0047 0.0055]	δ_0	0.2415 [0.0369 0.7846]
Risk Premia Parameters			
$\lambda_{0,1}$	-221.3 [-234.8 -122.3]	$\lambda_{x,21}$	41.98 [31.46 51.21]
$\lambda_{0,2}$	-15.12 [-25.95 22.23]	$\lambda_{x,22}$	-697.5 [-713.6 -630.0]
$\lambda_{0,3}$	-77.11 [-85.35 -44.62]	$\lambda_{x,23}$	-664.4 [-674.6 -602.4]
$\lambda_{x,11}$	97.78 [66.92 123.5]	$\lambda_{x,31}$	48.84 [36.79 60.54]
$\lambda_{x,12}$	-1415 [-1457 -1065]	$\lambda_{x,32}$	-796.8 [-829.6 -735.0]
$\lambda_{x,13}$	-1431 [-1468 -1073]	$\lambda_{x,33}$	-774.8 [-799.9 -720.2]

tions would be then be too concentrated relative to the survey data. Estimating the model using the full cross-section of survey responses thus imposes strong restrictions on the risk-neutral dynamics and these restrictions are binding in practise: The posterior estimates of the second and third eigenvalue of the factor process under the risk neutral dynamics, i.e. $F_{2,2}^Q$ and $F_{3,3}^Q$, are very similar, i.e. 0.75629 and 0.75624 . This means that it is virtually impossible for the agents in the model to disentangle the individual effects of the second and third factor on the yield curve, even when the variance of the maturity specific shocks is small.

The restrictions on the risk neutral dynamics that ensures that there is room for disagreement is similar to those that ensures that some factors are unspanned, e.g. Joslin, Priebsch and Singleton (2012), Duffee (2011) and Barillas (2013). An unspanned factor is by definition not priced, i.e. does not affect current bond yields. The flip side of this definition is that an unspanned factor cannot be extracted from the yield curve by inverting the bond price function. Yet, Joslin *et al* (2012) and Duffee (2011) demonstrate that unspanned factors can help forecast future interest rates even after conditioning on the current yield curve. The restriction imposed on our model by fitting the cross-section of survey forecasts is thus similar to imposing an unspanned factor structure: Only parameterizations that ensures that the state is not an invertible function of bond yields will leave room for the agent-specific signals to play a role. Parameterizations that imply that the state is almost perfectly revealed by bond yields will be rejected by the data.

An alternative strategy to use the survey data is to treat individual responses as noisy measures of a common expectation held by all agents as in Kim and Orphanides (2005) and Chernov and Mueller (2012). This is also the strategy that we follow when estimating the full information model with survey data. Others have used a measure of central tendency from the surveys, like a mean or median, to represent a noisy measure of the expectations of a representative agent (see Piazzesi and Schneider (2011)). In these alternative strategies, the information in the cross-sectional variance of survey responses does not directly restrict the dynamics of bond prices.

B. Historical decompositions

We can use the estimated model to measure how large the speculative term has been historically. From Proposition 2 we know that the speculative term must be orthogonal to public information available to all agents in real time, such as bond prices. However, as econometricians we have access to the full sample of data and can use information from period $t + 1, t + 2, \dots, T$, to form an estimate of the speculative term in period t . The Kalman smoother (see for instance Durbin and Koopman 2002) can be used to back out an

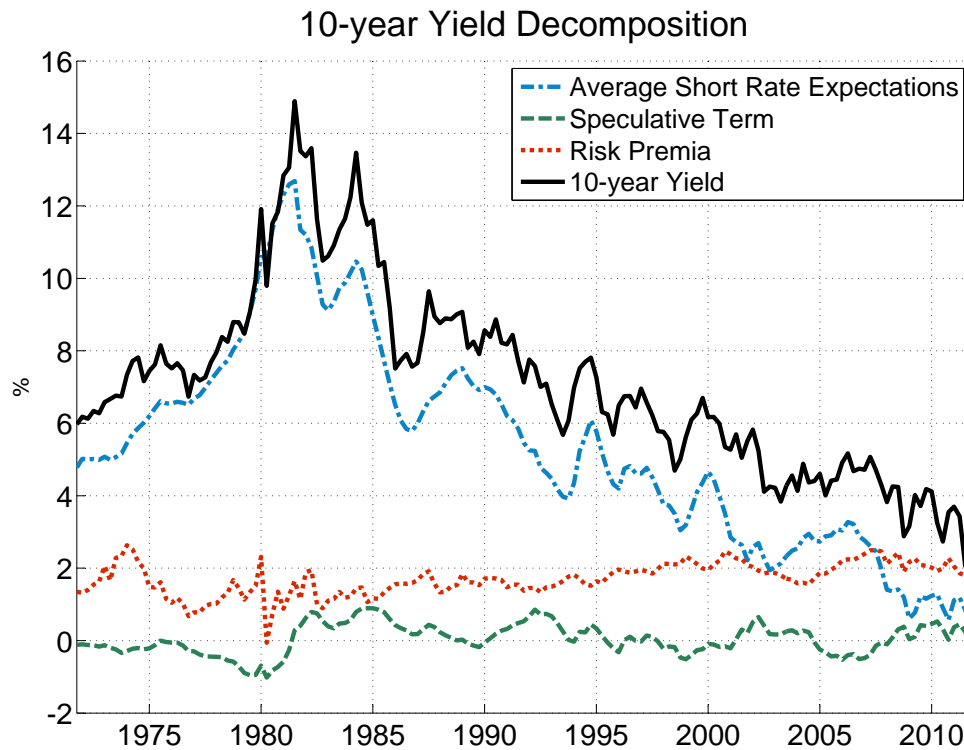


Figure 1 Decomposition of the 10-year yield. This figure plots the 10-year yield along with estimates of average short rate expectations, common risk premia and the speculative component at the posterior mode. The sample is 1971Q2 to 2011Q4.

estimate of the state X_t conditional on the entire history of observables. Since the speculative components of the term structure are linear functions of the state, the smoothed state history $E[X^T | \bar{z}^T]$ can be used to perform the decomposition of historical bond yields as described by (59).

Figure 1 plots the history of the 10 year yield together with a decomposition, splitting the yield into the terms based on average expectations about future short rates, classical risk premia and the speculative term. Most of the variation in yields is driven by variation in average first order expectations about the short rate. The standard deviation of the second most important term, classical risk premia, is 43 basis points, making it somewhat more volatile than the speculative term which has a standard deviation of 34 basis points. In absolute terms, the speculative component is largest around 1980 when it accounts for about a (negative) 1 percentage point of the 10 year yield. The speculative component's contribution as a fraction of the total yield is largest in the low yield environment of the last decade, accounting for up to 60 basis points at a time when the 10 year yield was only 4 percent.

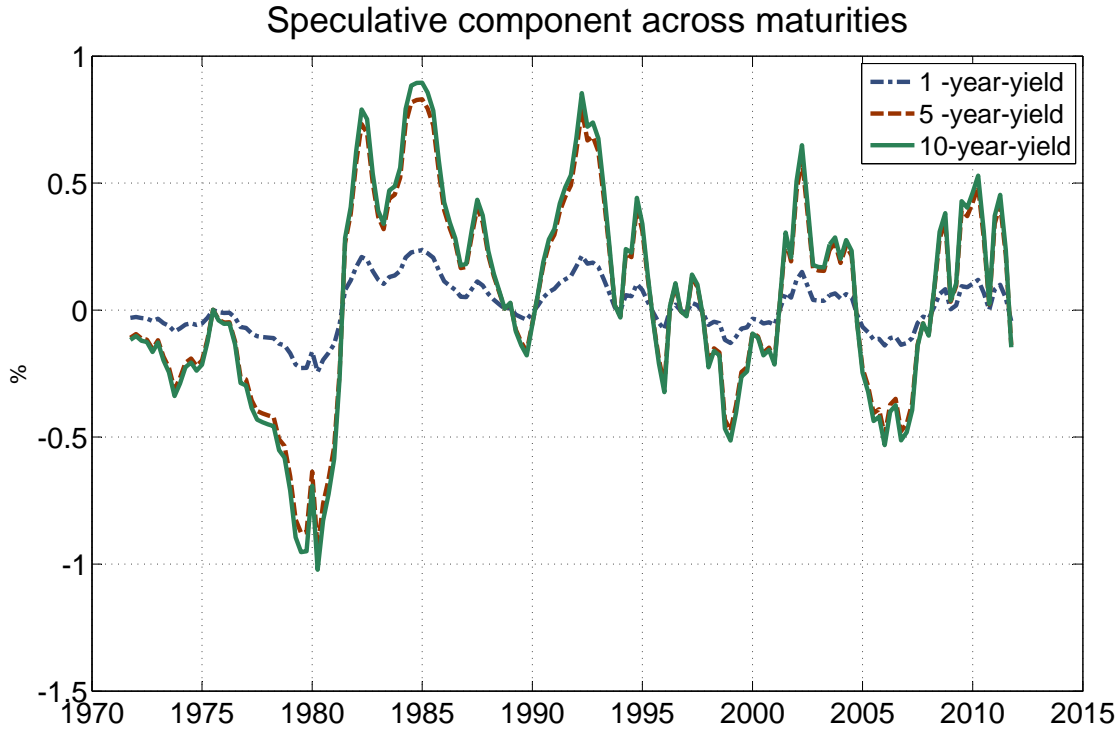


Figure 2 Speculative component across maturities. This figure plots estimates of the speculative component in the 1- 5- and 10-year yield at the posterior mode. The sample is 1971Q2 to 2011Q4.

Speculative dynamics are present at all maturities $n > 2$, but are quantitatively more important in medium- to long-maturity bonds. This is illustrated in Figure 2, where the estimated speculative components in the 1- 5- and 10-year yields are plotted. These speculative components are almost perfectly correlated across maturities and thus appear to have a one-factor structure. The speculative term is most volatile in the 10-year bond yield, but only marginally more so than for the 5-year bond. In comparison, the standard deviation of the speculative term in the 1-year bond is substantially lower at 9 basis points, and it never accounts for more than 25 basis points of the 1 year yield in the sample.

In the equilibrium model of Nimark (2012), agents form higher order expectations only about future risk-free short rates and there are no classical sources of time varying risk-premia. The speculative component extracted from the data using that model have qualitative properties similar to the speculative component extracted here, but is generally more volatile and peaks at around 3.5 percentage points in the early 1980s. The more flexible specification presented here that allows also for classical sources of time varying risk premia thus reduces the importance of the speculative component relative to the more restrictive model in Nimark (2012).

C. Comparison to a full information model

Gaussian affine term structure models have been used, for instance, by Cochrane and Piazzesi (2005) and Joslin, Priebsch and Singleton (2011) to decompose the term structure into risk premia and expected future short rates. Allowing for speculative dynamics may potentially change our estimates of historical risk premia and short rate expectations. In the bottom panel of Figure 3 we have plotted the posterior estimate of the risk premia in the 10-year bond extracted using our model with heterogeneously informed agents together with the risk premia extracted using the full information model of Joslin *et al* (2011), which our model nests as a special case. The full information model is estimated using the same bond yields (i.e. 1-, 5- and 10 year) and the same individual survey responses about the one quarter ahead forecasts for the T-Bill 3 month yield and 10 year bond yield as observables. Since the full information model implies that all agents share the same expectations about future bond yields, we treat the survey data as noisy measures of a common expectation held by all agents as in Kim and Orphanides (2005) and Chernov and Mueller (2012). Figure 3 shows that the speculative term appears to partially “crowd out” the time varying risk premium. The standard deviation of the risk premium is 126 basis points in the full information model but only 43 basis points in the model with heterogeneous information. Allowing for heterogeneous information also changes the cyclical properties of risk premia qualitatively: The correlation between the common risk premia term in the heterogeneous information model and the risk premia extracted using the full information model is -0.62 .

The two models imply different interpretations of recent historical episodes. For instance, the full information model interprets the Volcker disinflation period in the early 1980s as a time when risk premia were unusually high. As can be seen in the top panel of Figure 3, the heterogeneous information model instead attributes more of the high yields of that period to first order expectations about short interest rates and records a much smaller movement in risk premia. The early 1980s is also the period when the speculative term is the largest. In the second quarter of 1980, the speculative term contributed negatively to the 10 year bond yield by about 1 percentage point. Thus, according to the model, this was a period when individual agents believed that other agents underestimated future bond yields. The speculative component switches sign in 1982 and contributes positively to bond yields in the later part of the Volcker disinflation.

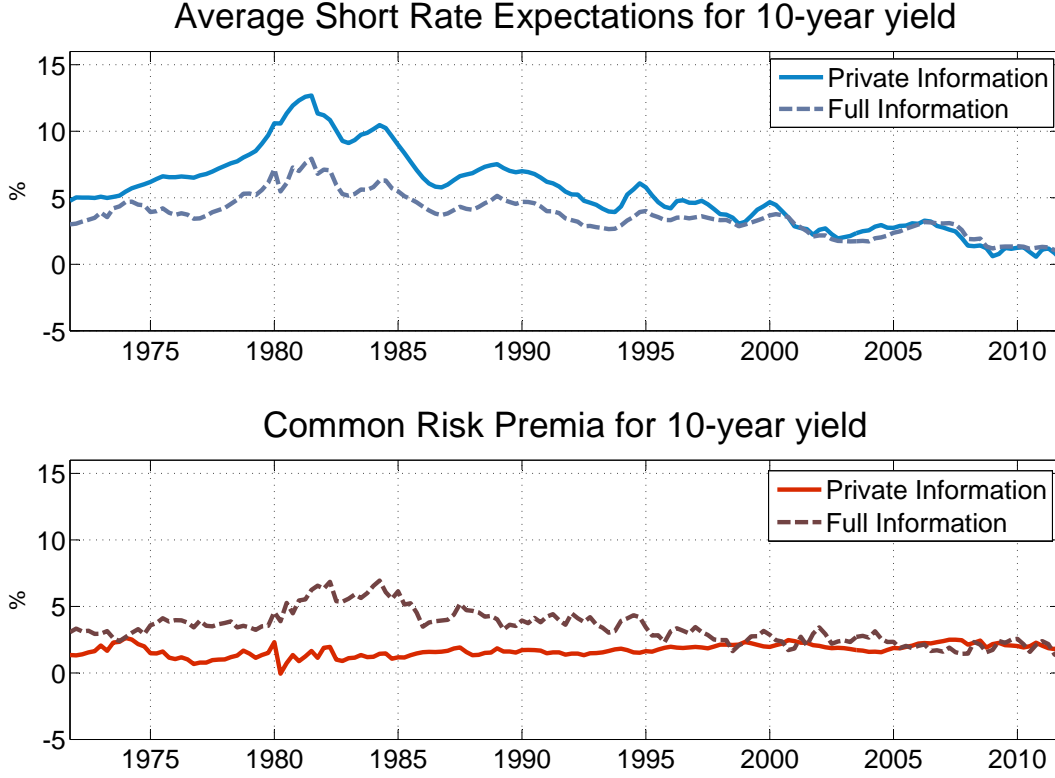


Figure 3 Average short rate expectations and common risk premia for the 10-year yield in the private and full information model This figure plots estimates of average short rate expectations and risk premia for the 10-year yield for the full information model and the private information model. Both estimates are obtained at the posterior mode (MLE estimates). The sample is 1971Q2 to 2011Q4.

D. Short rate and risk premia speculation

The speculative component defined by (59) and derived above is the sum of higher order prediction errors about average future stochastic discount factors. These discount factors are made up both of a time-discount component and risk adjustment component. We can decompose the speculative term in (59) further in order to separate speculation related to future short rates from speculation about future risk premia.

To do so, note that by recursive substitution in (50) the row vector B'_n can be decomposed into a term that captures higher order expectations about the short rate and a term that captures higher order expectations about future risk adjustments. The total speculative term in (59) can then be decomposed as

$$\underbrace{B'_n (I - \bar{H}) X_t}_{\text{total speculative term}} = \underbrace{-\delta_X \sum_{s=0}^{n-1} (\mathcal{F}H)^s (I - \bar{H}) X_t}_{\text{short rate speculation}} - \underbrace{\psi_{n-1-s} \Sigma_a \Lambda_x \sum_{s=0}^{n-2} (\mathcal{F}H)^s (I - \bar{H}) X_t}_{\text{risk premia speculation}} \quad (72)$$

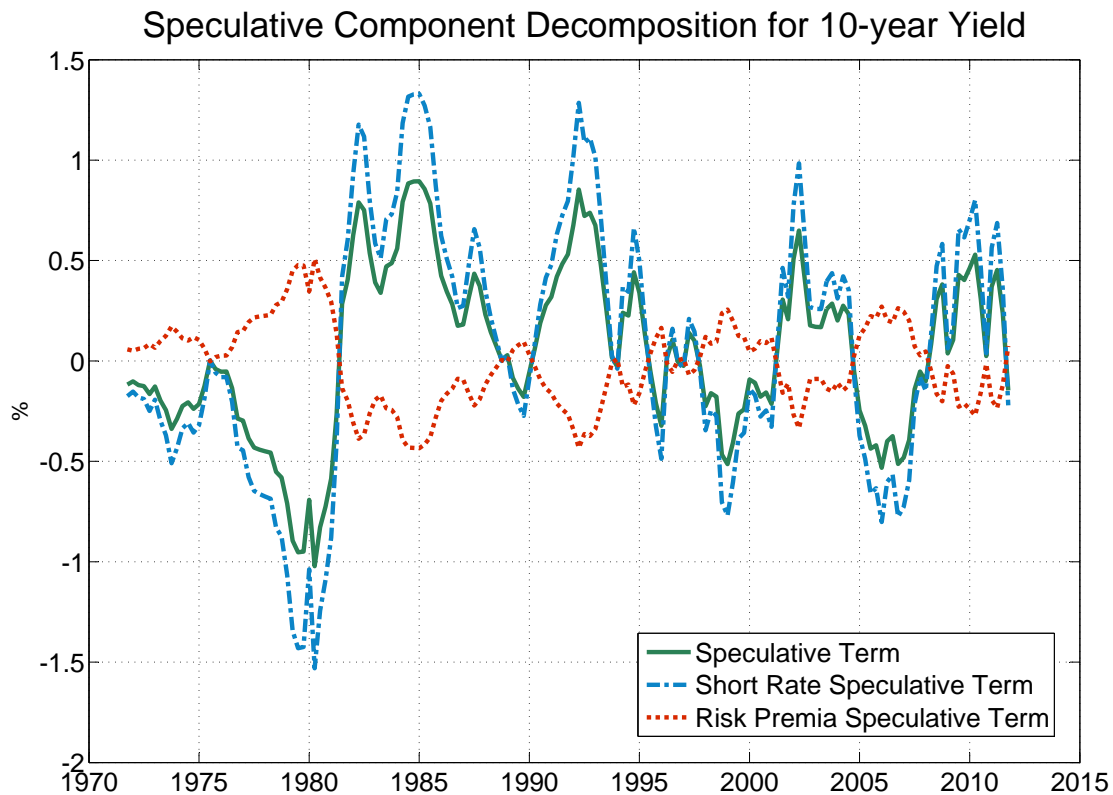


Figure 4 Speculative Component Decomposition for the 10-year yield This figure plots the speculative component decomposition for the 10-year yield into the speculative component attributed to speculation about future short rates and the speculative component related to common risk premia. Estimates are obtained at the posterior mode (MLE estimates). The sample is 1971Q2 to 2011Q4.

Figure 4 displays the posterior estimates of the two terms that make up the speculative component in the 10-year yield. The term capturing speculation about future short rates is substantially more volatile than the total speculative term. At times, speculation about short rates contributed up to as much as 1.5 percentage points to the 10 year yield. However, speculation about short rates is negatively correlated with speculation about future risk premia. When individual agents think that other agents overestimate future risk-free rates, they thus also tend to think that other agents underestimate future risk premia. This negative correlation explains why the total speculative component is less volatile than the component capturing only speculation about future short rates.

E. The total effect of information imperfections

The speculative term quantified above is a function of perceived disagreement between individuals' first and higher order expectations. The speculative term would thus be identically zero if all agents could observe the state perfectly. Yet, the speculative term does not capture the total effect of information imperfections in the model. Independently of the speculative term, agents may have incorrect first order expectations about the latent factors. Like the speculative component, these expectation errors would be identically zero if agents' signals were perfectly precise. In general, agents' first order expectations will not coincide with the true factors and we can use the estimated model to quantify the historical importance of this discrepancy.

To this end, define the counterfactual full information price p_t^{n*} as the price that would prevail if agents first and higher order expectations coincided with the true factors, that is if $x_t^{(k)} = x_t$ for every k . The total effect of information imperfections is then captured by the difference between the actual price and the counterfactual full information price. In Figure 5 we have plotted the difference between the actual 10 year bond yield and the counterfactual full information bond yield, i.e. $n^{-1}(p_t^n - p_t^{n*})$ alongside the speculative component.

The estimated total effect of information imperfections on historical bond yields is quite large, at times accounting for up to 2.5 percentage points of 10 year bond yields and it is almost perfectly correlated with the speculative component (which makes up part of the total effect). The difference between the total effect and the speculative component is due to agents' incorrect first order expectations about the latent factors. Agents' incorrect first order expectations are thus quantitatively at least as important for bond yields as the speculative component.¹¹

F. What drives speculative dynamics?

In order to address the question of what drives speculative dynamics we can decompose the variance of the speculative terms into four orthogonal sources: The three innovations to the exogenous factors in x_t and the maturity specific disturbances v_t . Table 2 displays variance decompositions of the 1-, 5- and 10 year

¹¹These estimates of the importance of incorrect first order expectations about the latent factors are conditional on treating the heterogeneous information model as the data generating process. Estimating an otherwise similar model but with a single imperfectly informed representative agent suggests that information imperfections are quantitatively unimportant. The reason for this difference is that a representative agent model cannot explain the cross-sectional dispersion observed in the survey data. The the cross-sectional variance of the survey data can then also not be used to discipline the parameters of the model that determine the precision of the representative agent's information set.

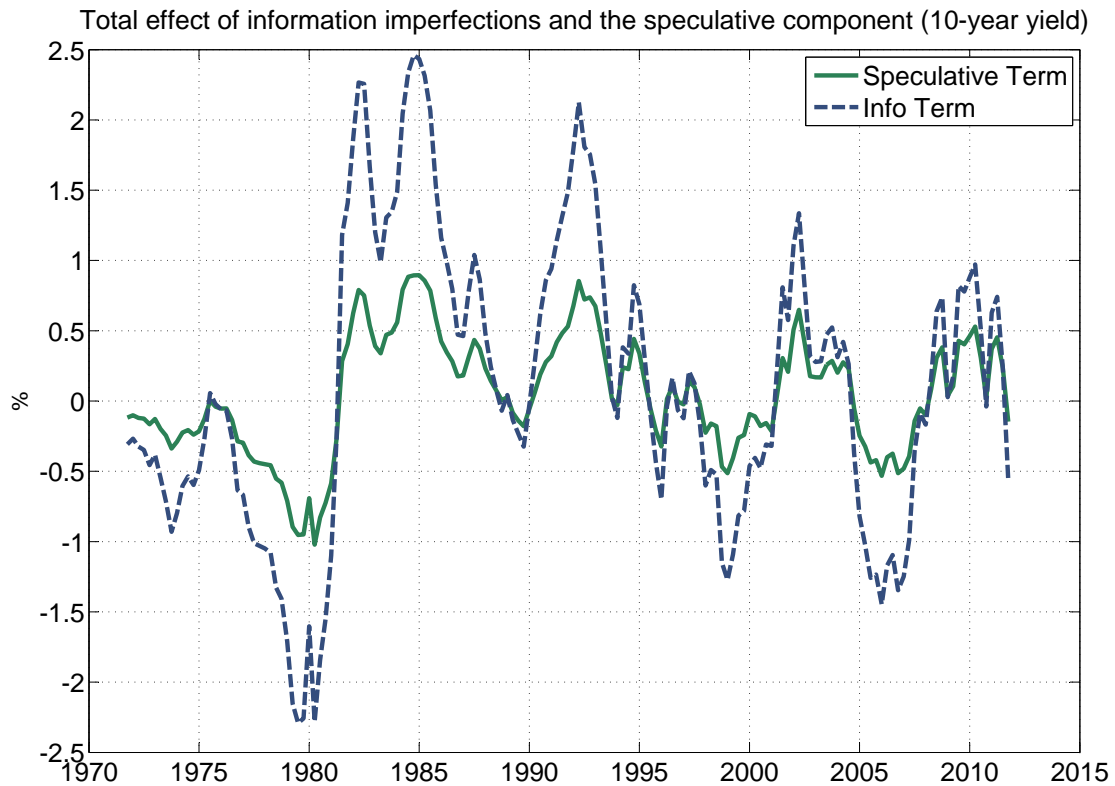


Figure 5 Total effect of information imperfections and the speculative component for the 10-year yield

This figure plots the speculative component for the 10-year yield along with the counterfactual yield (Info Term) that would prevail if all order of expectations coincided with the true factors. Estimates are obtained at the posterior mode (MLE estimates). The sample is 1971Q2 to 2011Q4.

yields, the speculative components in the 1-, 5- and 10 year yields and the first three principal components. Shocks to the first factor explain more than 97 per cent of the variance of all yields and the variance of the first principal component. It also accounts for about 75 per cent of the variance of the speculative component. Shocks to the third and second factor explain little of the variance of yields, but are important drivers of the second and third principal component.

Interestingly, the maturity specific shocks, which explain less than 0.2 per cent of the variance of any bond yield, explain between 14 and 17 per cent of the variance of the speculative term. The maturity specific shocks also explain a substantial fraction of the second and third principal component, but virtually none of the variance of the first.

The maturity specific shocks in the model are not formally equivalent to traditional “pricing errors”. An innovation to v_t^n does not affect only y_t^n but also agents’ estimate of the state since y_t^n is part of agents’

Table II

Variance decomposition of yields, speculative components and principal component of yields

This table reports results of a variance decomposition of yields, speculative components and principal components of yields. These were computed at the posterior mode (MLE estimates) of our model. The sample period is from 1971:Q2 to 2011Q4. In the first three columns we show the percentage of the variance of the 1-, 5- and 10-year yield that is explained by the shocks of the model. The next three columns show related quantities for the 1-, 5- and 10-year speculative component of yields. The last three columns report the results for the first three principal components of yields.

	Yields			Speculative Component			Principal Components		
	$y_t^{(1)}$	$y_t^{(5)}$	$y_t^{(10)}$	$spec_t^{(1)}$	$spec_t^{(5)}$	$spec_t^{(10)}$	pc_t^1	pc_t^2	pc_t^3
ε_t^1	97.3	97.7	97.8	71.9	74.9	75.5	97.7	15.1	27.4
ε_t^2	2.02	1.78	1.76	3.23	2.62	2.57	1.83	35.1	45.4
ε_t^3	0.61	0.31	0.26	7.77	7.23	7.07	0.37	36.9	17.2
v_t	0.08	0.15	0.18	17.1	15.2	14.8	0.04	12.8	9.90

observation vector z_t^j . Through its effect on agents' state estimates, it will also indirectly affect bond yields of maturities other than n . Due to the persistence in agents' estimates of the state, the effect on bond yields of a single maturity specific shock will last for several periods after impact. Our specification is thus not subject to the critique in Hamilton and Wu (2011) who argue that the independent white noise assumption of classical pricing errors is testable and rejected by the data in standard affine term structure models.

It would be interesting to rotate the model into an equivalent representation where the states are the principal components and analyze how each of the principal components affect the speculative term at different maturities. However, except in the full information limit, no such equivalent representation exists. Since the principal components are observable directly from the cross-section of yields, such a representation could not capture the speculative dynamics since these are orthogonal to current bond prices.

G. How useful are the agent-specific signals for predicting excess returns?

We can use the estimated model to quantify how useful the agent-specific signals are in terms of helping agents to forecast future bond prices and excess returns. For each maturity n we first compute the model implied unconditional variance of quarterly excess returns defined as

$$var(rx_t^n) = E (p_{t+1}^{n-1} - p_t^n - r_t)^2 \tag{73}$$

We then compute the variance of expected excess returns conditional on the information set of agent j . Dividing the latter by the former gives the model implied R^2 of excess returns for an n -period bond from the perspective of an agent in the model.

The R^2 of excess returns from the agents perspective varies across maturities. It is largest for very short and very long maturities at around 0.27. The R^2 is smallest for 1 year bonds at about 0.21. These estimates of the model implied R^2 are somewhat smaller than suggested by simple predictability regressions on yields only, but larger than the R^2 of the Sharpe-ratio constrained affine models in Duffee (2010).

Comparing the predictability of excess returns from the agents' perspective with the R^2 of excess returns conditional only on bond yields suggests that the agent-specific signals increase the R^2 of excess returns by about 6 per cent, more or less uniformly across maturities.

VI. Conclusions

In this paper we have presented and estimated an affine no-arbitrage Gaussian model that can be used to analyze the term structure of interest rates when agents have access to heterogeneous information and form rational expectations. We showed that heterogeneous information introduces a speculative component to bond yields that is orthogonal to public information in real time and is quantitatively important, at times accounting for up to a percentage point of yields. The heterogeneous information model gives a different interpretation of historical US bond yields relative to a nested full information affine term structure model. In particular, the heterogeneous information model attributes less of the high bond yields during the Volcker disinflation in the early 1980s to high risk premia and more to high average first order expectations about future risk-free short rates. It may thus be important to control for speculative dynamics when extracting information about interest rates expectations and risk premia from market prices of bonds.

Allowing for heterogeneous information also changes the qualitative properties of risk premia: The risk premia component extracted using the model presented here is negatively correlated with risk premia extracted using the full information model. Relaxing the assumption that all agents have access to the same information may thus potentially change our view on what the economic forces are that drive time variation in risk premia.

The speculative component in bond prices is orthogonal to public information in real time and is thus not an invertible function of current bond yields. This makes the speculative component similar to an un-

spanned factor and the results in Duffee (2011) suggest that the similarities do not stop there. That paper demonstrates that even though there is strong evidence that unspanned factors would help predict future bond prices if they could be observed, (yields-only) models with unspanned factors do not outperform lower dimensional models with only spanned factors in out-of-sample forecasting. The lack of improved out-of-sample forecasting performance when dynamically filtering for the unspanned factors suggests that the unspanned factors are uncorrelated not only with current bond yields, but also with lagged bond yields just like the speculative component in our model. These similarities between unspanned factors and the speculative component are intriguing and, while outside the scope of the present paper, it would be interesting to analyze the relationship between the two types of objects further.

Finally, the stochastic discount factor framework presented here is general and can also be used to price other asset classes. In the present paper, we found that speculative dynamics were quantitatively important in treasury markets even though the value at maturity of a zero-coupon default-free bond is known with certainty, and the only source of uncertainty is future discount rates. The prices of other classes of assets such as stocks and corporate bonds also depend on expectations about future discount rates, but are subject to additional sources of uncertainty due to stochastic cash-flows and the probability of default. It seems plausible that speculative dynamics could be even more important in those asset classes where prices depend on a richer set of variables.

References

- Admati, Anit R., (1985), “A noisy rational expectations equilibrium for multi-asset securities markets”, *Econometrica*, Vol. 53, pp. 629-657.
- Allen, F. S. Morris and H.S. Shin, (2006), “Beauty Contests and Iterated Expectations in Asset Markets”, *Review of Financial Studies*, 19, pp719 – 752.
- Bacchetta, Phillippe and Eric van Wincoop, (2006) “Can Information Heterogeneity Explain the Exchange Rate Determination Puzzle?”, *American Economic Review* vol 96, pp552-576.
- Barillas, F (2013) “Can we Exploit Predictability in Bond Markets?”, Working Paper, Emory University.
- Chernov, M and Mueller, P (2012), “The Term Structure of Inflation Expectations”, forthcoming in *Journal of Financial Economics*.
- Chun, A.L. (2011), Expectations, Bond Yields, and Monetary Policy, *Review of Financial Studies*, vol 24(1), pp208-47.
- Cochrane, J. and M. Piazzesi, (2005), “Bond Risk Premia”, *American Economic Review*.
- Cochrane, J. and M. Piazzesi, (2008), “Decomposing the Yield Curve”, mimeo University of Chicago.
- Dai, Q and K. Singleton, 2000, “Specification Analysis of Affine Term Structure Models,” *Journal of Finance*, vol. 55(5), pages 1943-78.
- Diamond, D and R. Verrecchia, (1981), “Information Aggregation in a Noisy Rational Expectations Economy”, *Journal of Financial Economics* vol. 9, pages 221-235.
- Duffee, Gregory (2002), “Term Premia and Interest Rate Forecasts in Affine Models,” *Journal of Finance*, 57, pp.405-443.
- Duffee, Gregory, (2010), “Sharpe ratios in term structure models”, mimeo, Johns Hopkins University.
- Duffee, Gregory, (2011), “Information in (and not in) the term structure of interest rates”, forthcoming, *Review of Financial Studies*.
- Duffie, Darrell and Rui Kan (1996), “Yield Factor Models of Interest Rates,” *Mathematical Finance*, 64, pp.379-406.

- Durbin, J. and S.J. Koopman, (2002), "A simple and efficient simulation smoother for state space time series analysis", *Biometrika*, 89, pp. 603-615.
- Geweke, John, 2005, *Contemporary Bayesian Econometrics and Statistics*, Wiley-Interscience.
- Grossman, Sanford, 1976, "On the Efficiency of Competitive Stock Markets Where Trades Have Diverse Information", *Journal of Finance*, Vol. 31, No. 2, Papers and Proceedings. pp. 573-585.
- Grossman, S. and J. Stiglitz, (1980), "On the Impossibility of Informationally Efficient Markets", *American Economic Review* vol. 70(3), pages 393-408.
- Gurkaynak, R. S., Sack, B., and Wright, J. H., (2007), "The U.S. treasury yield curve: 1961 to the present", *Journal of Monetary Economics* vol. 54, pages 2291-2304.
- Harrison, J.M. and D. Kreps, (1978), "Speculative Behaviour in a Stock Market with Heterogeneous Expectations", *Quarterly Journal of Economics*, pp323-336.
- Hamilton, J and Wu, C (2011), "Testable Implications of Affine-Term-Structure Models", forthcoming, *Journal of Econometrics*.
- Hellvig, Martin (1980), "On the aggregation of information in competitive markets," *Journal of Economic Theory*, 22, issue 3 pp.477-498.
- Joslin, S., M. Pribsch and K. Singleton, (2010), "Risk Premiums in Dynamic Term Structure Models with Unspanned Macro Risks," working paper, Stanford University.
- Joslin, S., K. Singleton and H. Zhu (2011), "A New Perspective on Gaussian DTSMs", forthcoming in the *Review of Financial Studies*.
- Kim, D. H., Orphanides, A., (2005). "Term structure estimation with survey data on interest rate forecasts" working paper, Board of Governors of the Federal Reserve System.
- Ludvigson, S. and S. Ng, (2009), "Macro Factors in Bond Risk Premia", *Review of Financial Studies* 22(12), 5027-5067.
- Nimark, K., (2011), "Dynamic Higher Order Expectations", working paper, Universitat Pompeu Fabra.

- Nimark, K., (2012), “Speculative Dynamics in the Term Structure of Interest Rates”, working paper, Universitat Pompeu Fabra.
- Piazzesi, M. and M. Schneider, (2011), “Trend and Cycle in Bond Premia”, unpublished manuscript.
- Townsend, Robert M., (1983), Forecasting the Forecasts of Others, *Journal of Political Economy*, vol 91, pp546-588.
- Singleton, Kenneth J., (1987), “Asset prices in a time series model with disparately informed, competitive traders”, in *New Approaches to Monetary Economics*, Eds. W.A. Burnett and K.J. Singleton, Cambridge University Press.
- Xiong, W. and H. Yan, (2010), “Heterogeneous Expectations and Bond Markets”, *Review of Financial Studies*, vol. 23(4), pp. 1433-1466.

Appendix A. Proof of Proposition 2

We want to prove that the speculative term $p_t^n - \bar{p}_t^n$ is orthogonal to public information in real time, i.e.

$$E([p_t^n - \bar{p}_t^n] \omega_t) = 0 : \forall \omega_t \in \Omega_t \quad (\text{A.1})$$

where Ω_t is the public information set defined as in Proposition 1.

Start by taking expectations of $p_t^n - \bar{p}_t^n$ conditional on the public information set Ω_t

$$\begin{aligned} E[p_t^n - \bar{p}_t^n | \Omega_t] &= -E\left(\int E[m_{t+1}^j - m_{t+1}^j | \Omega_t^j] dj | \Omega_t\right) \\ &\quad -E\left(\int E\left[m_{t+2}^j - \int E[m_{t+2}^i | \Omega_{t+1}^i] di | \Omega_t^j\right] dj | \Omega_t\right) - \dots \\ &\quad \dots - E\left(\int E\left[m_{t+n}^j - \int E\left[\dots \int E[m_{t+n}^i | \Omega_{t+n-1}^{i'}] di' \dots\right] di | \Omega_t^j\right] dj | \Omega_t\right). \end{aligned} \quad (\text{A.2})$$

The definition of the public information set (8) implies that $\Omega_t \subseteq \Omega_{t+s}^j$ for all j and $s \geq 0$. Applying the law of iterated expectations to the right hand side of (A.2) then gives

$$E[p_t^n - \bar{p}_t^n | \Omega_t] = E(m_{t+1}^j | \Omega_t) - E(m_{t+1}^j | \Omega_t) \quad (\text{A.3})$$

$$\begin{aligned} &+ E(m_{t+2}^j | \Omega_t) - E(m_{t+2}^j | \Omega_t) + \dots \\ &\dots + E(m_{t+n}^j | \Omega_t) - E(m_{t+n}^j | \Omega_t) \\ &= 0 \end{aligned} \quad (\text{A.4})$$

which completes the proof.

Appendix B. Deriving the bond price equation

This Appendix demonstrates how to find the expressions for A_n and B_n' in the conjectured bond price equation

$$p_t^n = A_n + B_n' X_t + v_t^n. \quad (\text{B.1})$$

The stochastic discount factor of agent j is denoted M_{t+1}^j and in the absence of arbitrage, the relationship

$$p_t^{n+1} = \log E \left[M_{t+1}^j P_{t+1}^n | \Omega_t^j \right] \quad (\text{B.2})$$

must be satisfied for each agent j and maturity n . Following the full information affine asset pricing literature, we specify the logarithm of agent j 's SDF to follow

$$m_{t+1}^j = -r_t - \frac{1}{2} \Lambda_t^{j' \Sigma_a \Lambda_t^j} - \Lambda_t^{j'} a_{t+1}^j : a_{t+1}^j \sim N(0, \Sigma_a). \quad (\text{B.3})$$

The vector a_{t+1}^j spans the conditional one-period-ahead forecast errors of agent j for each maturity n . That is, for each maturity n , agent j 's forecast error can be written as a linear function of the vector a_{t+1}^j

$$p_{t+1}^{n-1} - E \left(p_{t+1}^{n-1} | \Omega_t^j \right) = \psi_{n-1} a_{t+1}^j. \quad (\text{B.4})$$

Given the conjectured bond price equation (B.1), agent j 's forecast in period t of a bond price in period $t + 1$ can be (ex post) incorrect either because his forecast of the state X_{t+1} was incorrect or because of a maturity specific shock v_{t+1}^n , i.e.

$$p_{t+1}^{n-1} - E \left(p_{t+1}^{n-1} | \Omega_t^j \right) = B'_{n-1} \left(X_{t+1} - E(X_{t+1} | \Omega_t^j) \right) + v_{t+1}^{n-1}. \quad (\text{B.5})$$

Both the state forecast error and the maturity specific shocks must therefore be included in the vector a_{t+1}^j

$$a_{t+1}^j \equiv \begin{bmatrix} X_{t+1} - E(X_{t+1} | \Omega_t^j) \\ v_{t+1} \end{bmatrix} \quad (\text{B.6})$$

where

$$v_t \equiv \begin{bmatrix} v_t^2 & \dots & v_t^n \end{bmatrix}'. \quad (\text{B.7})$$

Given these definitions, the row vector ψ_{n-1} that maps the vector a_{t+1}^j into forecast error is given by

$$\psi_{n-1} \equiv \begin{bmatrix} B'_{n-1} & 0 \end{bmatrix} \quad (\text{B.8})$$

if $n = 2$ (since there is no maturity-specific shock in the risk-free one period bond's price) and

$$\psi_{n-1} \equiv \begin{bmatrix} B'_{n-1} & e'_{n-2} \end{bmatrix} \quad (\text{B.9})$$

if $n > 2$ where e_n is a vector of conformable dimensions with a 1 in the n^{th} element and zeros elsewhere so that $e_1 v_t = v_t^2, e_2 v_t = v_t^3, \dots, e_{n-1} v_t = v_t^n$ etc.

Step-by-step derivation of bond price recursions

Start by substituting the stochastic discount factor (B.3) into the no-arbitrage condition (B.2)

$$p_t^n = \log E \left[\exp \left(-r_t - \frac{1}{2} \Lambda_t^{j'} \Sigma_a \Lambda_t^j - \Lambda_t^{j'} a_{t+1}^j + p_{t+1}^{n-1} \right) \middle| \Omega_t^j \right]. \quad (\text{B.10})$$

Use the definition (B.4) of the forecast error to replace p_{t+1}^{n-1}

$$p_t^n = \log E \left[\exp \left(-r_t - \frac{1}{2} \Lambda_t^{j'} \Sigma_a \Lambda_t^j - \Lambda_t^{j'} a_{t+1}^j + E \left(p_{t+1}^{n-1} \middle| \Omega_t^j \right) + \psi_{n-1} a_{t+1}^j \right) \middle| \Omega_t^j \right] \quad (\text{B.11})$$

Take the quantities known by agent j outside the expectation operator

$$\begin{aligned} p_t^n &= -r_t - \frac{1}{2} \Lambda_t^{j'} \Sigma_a \Lambda_t^j + E \left(p_{t+1}^{n-1} \middle| \Omega_t^j \right) \\ &\quad + \log E \left[\exp \left(\left[\psi_{n-1} - \Lambda_t^{j'} \right] a_{t+1}^j \right) \middle| \Omega_t^j \right] \end{aligned} \quad (\text{B.12})$$

and use the fact that the term inside the bracket is log-normally distributed so that

$$\begin{aligned} p_t^n &= -r_t - \frac{1}{2} \Lambda_t^{j'} \Sigma_a \Lambda_t^j + A_{n-1} + B'_{n-1} \mu_X + B'_{n-1} \mathcal{F} H X_t^j \\ &\quad + \frac{1}{2} \left(\psi_{n-1} - \Lambda_t^{j'} \right) \Sigma_a \left(\psi_{n-1} - \Lambda_t^{j'} \right)'. \end{aligned} \quad (\text{B.13})$$

The expression (B.13) can be further simplified to

$$\begin{aligned} p_t^n &= -r_t + A_{n-1} + B'_{n-1} \mu_X + B'_{n-1} \mathcal{F} H X_t^j \\ &\quad + \frac{1}{2} \psi_{n-1} \Sigma_a \psi'_{n-1} - \psi_{n-1} \Sigma_a \Lambda_t^j. \end{aligned} \quad (\text{B.14})$$

Finally, substituting in the expression for Λ_t^j and r_t and expanding we get

$$\begin{aligned}
p_t^n &= -\delta_0 - \delta_X X_t + A_{n-1} + B'_{n-1} \mu_X + B'_{n-1} \mathcal{F} H X_t^j \\
&\quad + \frac{1}{2} \psi_{n-1} \Sigma_a \psi'_{n-1} \\
&\quad - \psi_{n-1} \Sigma_a \Lambda_0 - \psi_{n-1} \Sigma_a \Lambda_x X_t^j - \psi_{n-1} \Sigma_a \Lambda_v v_t.
\end{aligned} \tag{B.15}$$

Matching coefficients in the expression (B.15) for the average agent (defined so that $X_t^j = X_t$) and the conjectured bond price equation (B.1)

$$\begin{aligned}
A_n + B'_n X_t + v_t^n &= -\delta_0 - \delta_X X_t + A_{n-1} + B'_{n-1} \mu_X + B'_{n-1} \mathcal{F} H X_t \\
&\quad + \frac{1}{2} \psi_{n-1} \Sigma_a \psi'_{n-1} \\
&\quad - \psi_{n-1} \Sigma_a \Lambda_0 - \psi_{n-1} \Sigma_a \Lambda_x X_t - \psi_{n-1} \Sigma_a \Lambda_v v_t
\end{aligned} \tag{B.16}$$

gives the expressions for A_n and B'_n

$$\begin{aligned}
A_n &= -\delta_0 + A_{n-1} + B'_{n-1} \mu_X + \frac{1}{2} \psi_{n-1} \Sigma_a \psi'_{n-1} - \psi_{n-1} \Sigma_a \Lambda_0 \\
B_n &= -\delta_X + B'_{n-1} \mathcal{F} H - \psi_{n-1} \Sigma_a \Lambda_x
\end{aligned} \tag{B.17}$$

Restricting Λ_v

The last step that remains is to ensure that the maturity specific shocks v_t^n enter as conjectured in the price function (B.1), i.e. we need to set Λ_v such that

$$v_t^n = -\psi_{n-1} \Sigma_a \Lambda_v v_t \tag{B.18}$$

for each n . Start by stacking the expression for each n on top of each other so that

$$v_t = -\Psi \Sigma_a \Lambda_v v_t \tag{B.19}$$

where

$$\Psi \equiv \begin{bmatrix} \psi_1 \\ \vdots \\ \psi_{\bar{n}-1} \end{bmatrix}. \quad (\text{B.20})$$

Setting

$$\Lambda_v = -(\Psi \Sigma_a)_{right}^{-1} \quad (\text{B.21})$$

where $(\cdot)_{right}^{-1}$ denotes the (right) one-sided inverse of a matrix then ensures that equation (B.18) holds for each n . Letting the SDF depend on the maturity specific shocks thus do not introduce any additional free parameters through Λ_v .

The (right) one-sided inverse of $\Psi \Sigma_a$ exists as long as $rank(\Psi) = \bar{n} - 1$ where \bar{n} is the maximum maturity of any traded bond. This rank condition is likely to be satisfied in the model. To see why, note that Ψ can be written as

$$\Psi = \begin{bmatrix} B & I_{\bar{n}-1} \end{bmatrix} \quad (\text{B.22})$$

where

$$B \equiv \begin{bmatrix} B'_1 \\ \vdots \\ B'_{\bar{n}-1} \end{bmatrix} \quad (\text{B.23})$$

The covariance matrix Σ_a is in turn given by

$$\Sigma_a = \begin{bmatrix} \Sigma_{t+1|t} & CV' \\ VC' & VV' \end{bmatrix} \quad (\text{B.24})$$

where

$$\Sigma_{t+1|t} \equiv E \left(X_{t+1} - E \left[X_{t+1} \mid \Omega_t^j \right] \right) \left(X_{t+1} - E \left[X_{t+1} \mid \Omega_t^j \right] \right)' \quad (\text{B.25})$$

and $V \equiv E [v_t v_t']^{1/2}$. The product $\Psi \Sigma_a$ is then given by

$$\Psi \Sigma_a = \begin{bmatrix} (B \Sigma_{t+1|t} + VC') & (BCV' + VV') \end{bmatrix} \quad (\text{B.26})$$

The right inverse of the matrix $\Psi \Sigma_a$ exists as long as the square matrix $BCV' + VV'$ is of full rank. VV'

is a positive definite (and diagonal) matrix and is thus of full rank. The sum $BCV' + VV'$ will also be of full rank unless adding the BCV' cancels the independence of the rows of VV' . There is nothing in the model's structure that suggest that this should happen and the rank condition on $\Psi\Sigma_a$ can be checked on a case-by-case basis.

Appendix C. Solving the model

Solving the model implies finding a law of motion for the higher order expectations of x_t of the form

$$X_{t+1} = \mu_X + \mathcal{F}X_t + \mathcal{C}u_{t+1} \quad (\text{C.1})$$

where

$$X_t \equiv \begin{bmatrix} x_t^{(0)} \\ x_t^{(1)} \\ \vdots \\ x_t^{(\bar{k})} \end{bmatrix}$$

That is, to solve the model, we need to find the matrices \mathcal{F} and \mathcal{C} as functions of the parameters governing the short rate process, the maturity specific disturbances and the idiosyncratic noise shocks. The integer \bar{k} is the maximum order of expectation considered and can be chosen to achieve an arbitrarily close approximation to the limit as $\bar{k} \rightarrow \infty$. Here, a brief overview of the method is given, but the reader is referred to Nimark (2011) for more details on the solution method.

First, common knowledge of the model can be used to pin down the law of motion for the vector X_t containing the hierarchy of higher order expectations of x_t . Rational, i.e. model consistent, expectations of x_t thus imply a law of motion for average expectations $x_t^{(1)}$ which can then be treated as a new stochastic process. Knowledge that other agents are rational means that second order expectations $x_t^{(2)}$ are determined by the average across agents of the rational expectations of the stochastic process $x_t^{(1)}$. The average third order expectation $x_t^{(3)}$ is then the average of the rational expectations of the process $x_t^{(2)}$, and so on. Imposing this structure on all orders of expectations allows us to find the matrices \mathcal{F} and \mathcal{C} . Section A below describes how this is implemented in practice.

Second, the method exploits that the importance of higher order expectations is decreasing in the order

of expectations. This result has two components:

(i) The variances of higher order expectations of the factors x_t are bounded by the variance of the true process. More generally, the variance of $k + 1$ order expectation cannot be larger than the variance of a k order expectation

$$\text{cov} \left(x_t^{(k+1)} \right) \leq \text{cov} \left(x_t^{(k)} \right) \quad (\text{C.2})$$

To see why, first define the average $k + 1$ order expectation error $\zeta_t^{(k+1)}$

$$x_t^{(k)} \equiv x_t^{(k+1)} + \zeta_t^{(k+1)} \quad (\text{C.3})$$

Since $x_t^{(k+1)}$ is the average of an optimal estimate of $x_t^{(k)}$ the error $\zeta_t^{(k+1)}$ must be orthogonal to $x_t^{(k+1)}$ so that

$$\text{cov} \left(x_t^{(k)} \right) = \text{cov} \left(x_t^{(k+1)} \right) + \text{cov} \left(\zeta_t^{(k+1)} \right). \quad (\text{C.4})$$

Now, since covariances are positive semi-definite we have that

$$\text{cov} \left(\zeta_t^{(k+1)} \right) \geq \mathbf{0} \quad (\text{C.5})$$

and the inequality (C.2) follows immediately. (This is an abbreviated description of a more formal proof available in Nimark 2011.)

That the variances of higher order expectations of the factors are bounded is not sufficient for an accurate finite dimensional solution. We also need (ii) that the impact of the expectations of the factors on bond yields decreases “fast enough” in the order of expectation. The proof of this result is somewhat involved and interested readers are referred to the original reference.

The law of motion of higher order expectations of the factors

To find the law of motion for the hierarchy of expectations X_t we use the following strategy. For given \mathcal{F}, \mathcal{C} in (C.1) and B'_n in (50) we will derive the law of motion for agent j 's expectations of X_t , denoted $X_{t|t}^j \equiv E \left[X_t \mid \Omega_t^j \right]$. First, write the vector of signals z_t^j as a function of the state, the aggregate shocks and

the idiosyncratic shocks

$$z_t^j \equiv \begin{bmatrix} x_t^j \\ r_t \\ y_t \end{bmatrix} = \mu_z + DX_t + R \begin{bmatrix} u_t \\ \eta_t^j \end{bmatrix} \quad (\text{C.6})$$

where μ_z and D are given by

$$\mu_z = \begin{bmatrix} 0 \\ \delta_0 \\ -\frac{1}{2}A_2 \\ \vdots \\ -\frac{1}{\bar{n}}A_{\bar{n}} \end{bmatrix}, \quad D = \begin{bmatrix} I_d & \mathbf{0} \\ B'_1 \\ \vdots \\ \bar{n}^{-1}B'_{\bar{n}} \end{bmatrix} \quad (\text{C.7})$$

and R can be partitioned conformably to the aggregate and the idiosyncratic shocks

$$R = \begin{bmatrix} R_u & R_\eta \end{bmatrix}. \quad (\text{C.8})$$

The matrix R_u picks out the appropriate maturity specific shocks v_t^n from the vector of aggregate shocks u_t so that

$$R_u = \begin{bmatrix} \mathbf{0} \\ V \end{bmatrix} \quad (\text{C.9})$$

and R_η adds the idiosyncratic shocks $Q\eta_t^j$ to the exogenous state x_t to form the agent j specific signal vector x_t^j , i.e.

$$R_\eta = \begin{bmatrix} Q \\ \mathbf{0} \end{bmatrix} \quad (\text{C.10})$$

Agent j 's updating equation of the state $X_{t|t}^j \equiv E[X_{t+1} | \Omega_t^j]$ estimate will then follow

$$X_{t|t}^j = \mu_X + \mathcal{F}X_{t-1|t-1}^j + K \left(z_t^j - D(\mu_X + \mathcal{F}X_{t-1|t-1}^j) \right) \quad (\text{C.11})$$

Rewriting the observables vector z_t^j as a function of the lagged state and current period innovations and taking averages across agents using that $\int \zeta_t(j) dj = 0$ yields

$$X_{t|t} = \mu_X + \mathcal{F}X_{t-1|t-1} \quad (\text{C.12})$$

$$+ K \left(D(\mu_X + \mathcal{F}X_{t-1}) + (DC + R_u)u_t - D(\mu_X + \mathcal{F}X_{t-1|t-1}^j) \right) \quad (\text{C.13})$$

$$= \mu_X + (\mathcal{F} - KDF)X_{t-1|t-1} + KDFX_{t-1} + K(DC + R_u)u_t \quad (\text{C.14})$$

Appending the average updating equation to the exogenous state gives us the conjectured form of the law of motion of $x_t^{(0:\bar{k})}$

$$\begin{bmatrix} x_t \\ X_{t|t} \end{bmatrix} = \mu_X + \mathcal{F} \begin{bmatrix} x_{t-1} \\ X_{t-1|t-1} \end{bmatrix} + \mathcal{C}u_t \quad (\text{C.15})$$

where \mathcal{F} and \mathcal{C} are given by

$$\mathcal{F} = \begin{bmatrix} F^P & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + \begin{bmatrix} \mathbf{0}_{d \times d} & \mathbf{0} \\ \mathbf{0} & [\mathcal{F} - KDF]_- \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ [KDF]_- \end{bmatrix} \quad (\text{C.16})$$

$$\mathcal{C} = \begin{bmatrix} C & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ [K(DC + R_u)]_- \end{bmatrix} \quad (\text{C.17})$$

where $[\cdot]_-$ indicates that the a last row or column has been truncated to make a the matrix $[\cdot]$ conformable, i.e. implementing that $x_t^{(k)} = 0 : k > \bar{k}$. The Kalman gain K in (C.11) is given by

$$K = (\Sigma_{t+1|t}D' + CR_u)(D\Sigma_{t+1|t}D' + RR')^{-1} \quad (\text{C.18})$$

$$\begin{aligned} \Sigma_{t+1|t} &= \mathcal{F} \left(\Sigma_{t+1|t} - (\Sigma_{t+1|t}D' + CR_u)(D\Sigma_{t+1|t}D' + RR')^{-1} (\Sigma_{t+1|t}D' + CR_u)' \right) \mathcal{F}' \\ &+ \mathcal{C}\mathcal{C}' \end{aligned} \quad (\text{C.19})$$

The model is solved by finding a fixed point that satisfies (50), (C.16), (C.17), (C.18) and (C.19).

Appendix D. The matrices in the estimated state space system

The objects in the state space system left undefined in Section IV.G of the main text are given by

$$\mu_t^{\bar{z}} = \begin{bmatrix} -\frac{1}{4}A_4 \\ -\frac{1}{8}A_{20} \\ -\frac{1}{40}A_{40} \\ -\mathbf{1}_{(m \times 1)} \times (A_1 + B'_1 \mu_X) \\ -\frac{1}{40} \mathbf{1}_{(m \times 1)} \times (A_{40} + B'_{40} \mu_X) \end{bmatrix}, \quad \bar{D}_t = \begin{bmatrix} -\frac{1}{4}B'_4 \\ -\frac{1}{8}B'_{20} \\ -\frac{1}{40}B'_{40} \\ -\mathbf{1}_{(m \times 1)} \times B'_1 \mathcal{F}H \\ -\frac{1}{40} \mathbf{1}_{(m \times 1)} \times B'_{40} \mathcal{F}H \end{bmatrix} \quad (\text{D.1})$$

$$\bar{R}_t = \begin{bmatrix} \sigma_v \times \begin{bmatrix} e_{d+4-1} & e_{d+20-1} & e_{d+40-1} \end{bmatrix}' & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & I_m \times B'_1 \mathcal{F} \Sigma_j^{1/2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & I_m \times \frac{1}{40} B'_{40} \mathcal{F} \Sigma_j^{1/2} \end{bmatrix} \quad (\text{D.2})$$

where m is the number of survey responses available in period t and e_i is a vector with a one in the i^{th} position and zeros elsewhere.

Appendix E. Computing the cross-sectional variance Σ_j

The idiosyncratic noise shocks η_t^j are white noise processes that are orthogonal across agents and to the aggregate shocks v_t and ε_t . This implies that the cross-sectional variance of expectations is equal to the part of the unconditional variance of agent j 's expectations that is due to idiosyncratic shocks. This quantity can be computed by finding the variance of the estimates in agent j 's updating equation (C.11), but with the aggregate shocks v_t and ε_t “switched off”. The covariance Σ_j of agent j 's state estimate due to idiosyncratic shocks is defined as

$$\Sigma_j \equiv E \left(E [X_t | \Omega_t^j] - \int E [X_t | \Omega_t^i] dj \right) \left(E [X_t | \Omega_t^j] - \int E [X_t | \Omega_t^i] dj \right)' \quad (\text{E.1})$$

and given by the solution to the Lyapunov equation

$$\Sigma_j = (I - KD) \mathcal{F} \Sigma_j \mathcal{F}' (I - KD)' + KR_\eta R_\eta K'. \quad (\text{E.2})$$

which can be found by simply iterating on (E.2).