

NON-NESTED INFORMATION SETS AND THE TERM STRUCTURE OF INTEREST RATES *PRELIMINARY AND INCOMPLETE*

KRISTOFFER P. NIMARK

ABSTRACT. If long maturity bonds are traded frequently and traders have non-nested information sets, speculative behavior in the sense of Harrison and Kreps (1978) arises. Using a term structure model displaying such speculative behavior, this paper proposes an empirically plausible re-interpretation of predictable excess returns that is not based on the value traders attach to a marginal increase of wealth in different states of the world. Applying the properties of orthogonal projections, it is demonstrated that individual traders can systematically predict excess returns even in a model with a constant price of risk if information sets are non-nested. The model is estimated using monthly data on US short to medium term Treasuries from 1964 to 2007. It provides a good fit of the data and demonstrates that it is feasible to estimate dynamic models with privately informed agents. We show that a three factor no-arbitrage factor model would find overwhelming but misleading evidence in favor of a time varying price of risk if the world is characterized by the model with non-nested information sets presented here. We further argue that the “hidden” factor that help predict excess returns as documented by Duffee (2008) is an intrinsic feature of models with imperfectly informed traders.

1. INTRODUCTION

If long bonds are traded before they mature, the price an individual trader will be willing to pay for a bond depends on how much he thinks other traders will be willing to pay for it in the future. If traders have access to different information, this price may differ from what an individual trader would be willing to pay for the bond if he had to hold it until it matures and “speculative behavior” in the sense of Harrison and Kreps (1978) arises. That is, the possibility of reselling a bond changes its equilibrium price as traders exploit what they perceive to be market misperceptions about future short rates.

In this paper we present a term structure model populated with traders that engage in the type of speculative behavior described above. We use this model to argue that relaxing the assumption that all traders have access to the same information introduces interesting and empirically relevant new dynamics to the term structure of interest rates. More specifically, we apply the properties of orthogonal projections to show that if traders’ information sets are non-nested, (i) individual traders can systematically predict excess returns even in a

Date: April 20, 2009. The author thanks David Backus, Francisco Barillas, Jerome Henry, Thomas Sargent and seminar participants at New York University and Sveriges Riksbank for useful comments and suggestions.

Address: CREI, Universitat Pompeu Fabra, Ramon Trias Fargas 25-27, Barcelona 08005.

e-mail: knimark@crei.cat *webpage:* www.kris-nimark.net .

model with a constant price of risk. (ii) individual traders can predict and take advantage of other traders prediction errors even though no trader on average is better informed than other traders and (iii) that the speculative dynamics introduced by non-nested information sets are orthogonal to public information.

The model is estimated on monthly US bond data. Using the estimated model to generate artificial data, we further show that an econometrician estimating an affine three factor no-arbitrage model will find overwhelming but misleading statistical evidence in favor of a time varying price of risk. The main thrust of this paper is thus to argue that predictable excess returns can be a consequence of traders having non-nested information sets, rather than that traders value a marginal increase in consumption (wealth) differently in different states of the world.

A necessary condition for traders to have private information about future bond yields is that bond prices cannot perfectly reveal the state of the economy. Recent statistical evidence appear to support this view. In a few closely related papers, Cochrane and Piazzesi (2005, 2008) and Duffee (2008) present evidence suggesting that the factors that can be found by inverting yields are not sufficient to predict future bond returns. They find that while the usual level, slope and curvature factors explain virtually all of the cross sectional variation, additional factors are needed to forecast excess returns. Ludvigson and Ng (2009) provide more evidence that current bond yields are not sufficient to optimally forecast bond returns. They show that drawing on a very large panel of macroeconomic data helps predict deviations from the expectations hypothesis, or equivalently, future excess returns, compared to using only yield data. Stated another way, these statistical models all suggest that bond yields are not Markov.

In addition to the empirical evidence cited above, we also have a priori reasons to believe that bond prices should not reveal all information relevant to predicting future bond returns. Grossman and Stiglitz (1980) argued that if it is costly to gather information and prices are observed costlessly, prices cannot fully reveal all information relevant for predicting future returns. For the bond market, the most important variable to forecast is the short interest rate. In most developed countries, the short interest rate is set by a central bank that responds to macroeconomic developments. If it is costly to gather information about the macro economy, Grossman and Stiglitz's argument implies that bond prices cannot reveal all information relevant to predict future bond returns. This is indeed what the evidence in Ludvigsson and Ng (2009) suggests.

If prices do not reveal all information relevant for predicting bond returns, it becomes more probable that traders have non-nested information sets, that is, traders will have access to and use different information when trading.¹ With the exceptions of bond prices, statements by central bank officials and some well publicized macroeconomic data releases, it is hard to think of sources of information that are public in the strong common knowledge sense of the word. In this paper we allow for traders to have private information that they can exploit when trading. Apart from it's intuitive appeal, this also seems to accord well with casual

¹What I in this paper call *non-nested information sets* is also known as *private information* (Sargent 1989), *heterogenous information* (Bacchetta and van Wincoop 2006), *dispersed information* (Angeletos 2008) and *imperfect common knowledge* (Woodford 2002, Adam 2006 and Nimark 2008). I like the term non-nested since it naturally connects to the language of orthogonal projections that is used in this paper.

observation that at least one motive for trade in assets is possession of information that is not, or at least is not believed to be, already reflected in prices.

One implication of non-nested information sets is that expectations across individual traders will differ. While bond traders' expectations are unobservable, Swanson (2006) presents evidence that professional forecasters' expectations of future interest rates are surprisingly widely dispersed. Citing numbers from the Blue Chip Survey of professional forecasters from 1992-2004, Swanson reports that the spread between the 10th and the 90th percentile of individual forecasts of the 3-month Treasury Bill rate 4 quarters ahead fluctuates between 80 and 220 basis points.

The estimated model presented below displays similar dynamics to those documented by Duffee (2008) and Cochrane and Piazzesi (2005, 2008). Factors that play a very limited role in explaining the cross section of bond yields have predictive power for future yields. This is arguably an intrinsic feature of models with imperfectly informed traders. If the true state of the economy could be summarized by three factors that are an exact linear function of yields, no other factor could possibly add predictive power. We demonstrate that the model presented here can account for the evidence in Duffee (2008) by computing Duffee's impulse responses estimated on artificial data generated from our model.

In addition, when traders have non-nested information sets it becomes optimal to "forecast the forecasts of others", and natural representations of the state in this class of models tend to be infinite.² Here, we will solve a model with non-nested information sets using a method proposed by Nimark (2007). This method delivers a finite (but still relatively high) dimensional state representation. If our model is a good characterization of the world, a low dimensional factor model is fundamentally misspecified, though it would nest our model in the special case when traders are perfectly informed. Through a Monte Carlo exercise using simulated data generated from our theoretical model, we demonstrate that a three factor no-arbitrage model will not be able to fit simulated data from this model without using the additional flexibility allowed for by a time varying price of risk, even though the price of risk in our model is constant. The time varying risk premia measured by a three factor no arbitrage model could thus be picking up dynamics due to misspecification, rather than a true time varying willingness to bear risk.

The present paper is not the first to suggest that dispersed expectations can explain what appears to be time varying risk premia, though to the author's knowledge, this paper is the first to do so in a model with rational traders. Xiong and Yan (2008) analyze a model where two groups of traders have different (and boundedly rational) beliefs about the relationship between the variables that they observe and the underlying fundamental determining future short rates. This leads the two groups of traders to take on speculative positions against each other, and the relative wealth dynamics of the two groups create dynamics that mimics some of those found in the data.

There exists a vast literature on the term structure of interest rates, and for the purposes that we are interested in here, one can broadly classify papers according to whether they quantify or interpret risk premia, with the vast majority of papers falling in the former class. The quantifying class of papers has mainly been concerned with describing the statistical

²See Townsend (1983), Sargent (1991) and Makarov and Rytchov (2009).

properties of risk premia accurately, either in relation to observed variables (e.g. Ang and Piazzesi 2003) or using unobservable factors (e.g. Litterman and Sheinkman 1991). Some papers are explicit in their purpose of decomposing yields into information about future short rates and information about risk premia, e.g. Cochrane and Piazzesi (2005) and Backus and Wright (2007). These papers frame their discussions in the context of statistical models with few economic assumptions (often no-arbitrage affine models) that provide a very good fit of bond yields. Papers attempting to provide an economic explanation of time varying risk premia generally need to impose more structure by populating their models with a representative agent and assuming a specific functional form for the agent's utility function. The price of more structure is worse fit, but the gain is potentially a better understanding of what economic motives drive term structure dynamics.

Backus, Gregory and Zin (1989) is an early reference belonging to the interpretative class of term structure papers. In that paper, the authors set up an artificial economy populated by a representative agent with power utility and find that the model cannot account for neither the sign of the average risk premia nor the magnitude of its variation over time. A more recent paper in the interpretative class that, like Backus et al, relies on variation in expected marginal utility to explain the failure of the expectations hypothesis is Wachter (2006). She sets up a model where agents have habit preferences so that current consumption is evaluated relative to a 10 year weighted average of past consumption. Wachter's model can match some attributes of time varying risk premia, as well as the upward slope of the average yield curve. Piazzesi and Schneider (2006) present a model with recursive preferences that also imply that the expectations hypothesis will fail (though the focus of that paper is to analyze the effect of inflation surprises on bond yields).

There is also a large and growing literature analyzing asset pricing under non-nested information sets (or one of its synonyms, see Footnote 1). Some examples are Allen, Morris and Shin (2005), Kasa, Walker and Whiteman (2008), Bacchetta and van Wincoop (2005, 2007) and Makarov and Rytchov (2008). These papers either present purely theoretical models or models calibrated to explain some feature of the data. In this paper, we estimate the model directly using Bayesian methods with uninformative priors. As far as the author knows, this is the first paper to do so and the fit of the model is surprisingly good.

The next section presents a bond pricing equation that prices bonds as a function of higher order expectations of future short rates. Section 3 defines the expectations hypothesis and discusses some properties of orthogonal projections that we will use for the formal analysis in Section 4 that presents the main analytical results of the paper. Section 5 presents and estimates an empirical model that conforms to the bond pricing equation analyzed in Section 2 - 4. Section 6 shows that a popular three factor no-arbitrage model will mistakenly attribute the dynamics of our model to time varying price of risk. In Section 6 it is also demonstrated that the model can account for some of the findings of Duffee (2008). Section 7 concludes and the Appendix contains details of how the model was solved.

2. EXPECTATIONS AND THE TERM STRUCTURE

In this section we present a simple bond pricing equation, by which the current price of a bond depends on the average expectation of the price of the same bond in the next period,

discounted by the one period interest rate. For now, we will take this bond pricing equation as given, though in the empirical section below a model specification resulting in just such an relationship is presented. The pricing equation is simple, and does not allow for a time varying price of risk. The simplicity of the model helps to highlight the consequences for term structure dynamics of relaxing the assumption that traders all have access to the same information. In what follows, traders are indexed by $j \in (0, 1)$ and trader j 's information set is denoted $\Omega_t(j)$. The log price of a zero coupon bond with n periods to maturity in period t is given by

$$b_t^n = c^n - r_t + \int E [b_{t+1}^{n-1} | \Omega_t(j)] dj - \eta_t^n \quad (2.1)$$

where c^n is a maturity specific constant, r_t is the short interest rate and η_t^n is a maturity specific supply shock.³ The price of an n periods to maturity bond in period t thus depends the average expectation in period t of the price of a $n - 1$ period bond in period $t + 1$. The more a trader expects to be able to sell a bond for in the future, the more is he willing to pay for it today. However, risk aversion prevents the most optimistic trader from demanding all of the available supply. Since we are primarily interested in the effects of information we will dispense with constant c^n for most of the analysis below, and we will denote the deviation of the log price of a bond from the constant $\tilde{b}_t^1$ so that

$$\tilde{b}_t^1 = b_t^n - c^n. \quad (2.2)$$

The bond price formula (2.1) can be used to price any maturity bond. The procedure is similar to deriving bond prices under a no arbitrage assumption, though we need to be more careful in specifying the information sets that the expectations that govern prices are conditioned on. As usual, we can start from

$$\tilde{b}_t^1 = -r_t \quad (2.3)$$

and apply (5.8) recursively. The log price of a two period bond then is

$$\tilde{b}_t^2 = -r_t - \int E [r_{t+1} | \Omega_t(j)] dj + \eta_t^2 \quad (2.4)$$

The price of a three period bond according to (5.8) is given by the average expectation of the price of a two period bond in $t + 1$, discounted by the short rate r_t . Leading (2.4) by one period and substituting into (5.8) with $n = 3$ gives

$$\begin{aligned} \tilde{b}_t^3 = & -r_t - \int E [r_{t+1} | \Omega_t(j)] dj \\ & - \int E \left[\int E [r_{t+2} | \Omega_{t+1}(j')] dj' | \Omega_t(j) \right] dj \\ & + \eta_t^3 \end{aligned} \quad (2.5)$$

It is important to note that the term on the second line of (2.5) contains average expectations of average expectations where the time period that the average expectations are conditioned on are not the same. That is, the price of the three period bond does not depend on the

³Alternatively, the η s could also be viewed as measurement errors in a model without random bond supply (see Duffee (2008)).

average expectation in period t of the short rate in period $t+2$, but on the average expectation in period t of the average expectation in period $t+1$ of the short rate in period $t+2$.

We can apply the same procedure that we used to derive (2.5) to price an n periods to maturity bond

$$\begin{aligned} \tilde{b}_t^n = & -r_t - \int E[r_{t+1} | \Omega_t(j)] - \\ & \int E \left[\int E[r_{t+2} | \Omega_{t+1}(j')] dj' | \Omega_t(j) \right] dj + \dots \\ & \dots + \int E \left[\int E \left[\dots \int E[r_{t+n-1} | \Omega_{t+n-2}(j'')] dj'' \dots | \Omega_{t+1}(j') \right] dj' | \Omega_t(j) \right] dj + \eta_t^n \end{aligned} \quad (2.6)$$

The yield of a bond with n periods to maturity is (as usual) given by dividing the log bond price by n

$$y_t^n = -n^{-1} \tilde{b}_t^n \quad (2.7)$$

If all traders share the same information set, we can compute the expectations terms in the bond price equation by applying the law of iterated expectations and the expression (2.6) collapses to the expectations hypothesis. This is demonstrated in the next section, where we also define the expectations hypothesis and its equivalence with non-predictable excess returns.

3. LINEAR PROJECTIONS AND THE EXPECTATIONS HYPOTHESIS

In this section we will define three closely related concepts; the expectations hypothesis, excess returns and time varying price of risk. Roughly speaking, the following relationships hold. The expectations hypothesis implies no predictable excess returns, and if the expectations hypothesis holds, risk premia are constant over time. We will need to be more careful than usual when using this language. The reason is that in our model, excess returns may or may not be predictable depending on what information set one conditions on. Because of this, we will always specify the information set that a statement about the expectations hypothesis, or predictable excess returns, are conditioned on.

We start by defining what we mean by the expectations hypothesis using an accounting framework, or an “arithmetic identity”.⁴ For any model and an associated information set Ω_t we can use the decomposition

$$y_t^n \equiv c^n + n^{-1} \sum_{s=0}^{n-1} E[r_{t+s} | \Omega_t] + \gamma_t^n \quad (3.1)$$

as an accounting identity where c^n is a maturity specific constant and γ_t^n a maturity specific time varying “risk premium”. Here we define the expectations hypothesis as meaning that γ_t^n is orthogonal to the information set Ω_t . Predictable excess returns are just the inverse of this property: The failure of the expectations hypothesis implies that excess returns, defined as the return above what holding a series of short rates would yield, are predictable. The quotation marks around “risk-premium” are there because time varying compensation for

⁴See Backus and Wright (2007).

risk is not the only possible explanation for this term. (Indeed, the purpose of this paper is to propose an alternative explanation of the term γ_t^n that is not based on a time varying price of risk.) In practise, γ_t often functions as a residual, absorbing all time variation in yields that are not caused by variations in expected future short rates as predicted by the model in question. In this paper, we will use the term *predictable excess returns*, rather than “risk-premia”, though the reader should be aware that in most models, time varying price of risk is the main source of predictable excess returns.

3.1. Some useful properties of projections. In preparation for the next section, which contains the main analytical results of the paper, we here state some definitions and properties of orthogonal projections on inner-product spaces. Proofs and more details can be found in for instance Brockwell and Davis (2006).

Definition 1. (*The inner-product space L^2 .*) The inner product space L^2 is the collection $\mathcal{C}$ of all random variables X with finite variance

$$EX^2 < \infty \quad (3.2)$$

and with inner-product

$$\langle X, Y \rangle \equiv E(XY) : X, Y \in L^2 \quad (3.3)$$

In the model presented below, all bond yields, the factors that drive them and the signals that traders observe will be elements in L^2 .

Definition 2. Let Ω be a subspace of L^2 . An orthogonal projection of X on Ω , denoted $\mathcal{P}_\Omega X$, is the unique element in L^2 satisfying

$$\langle X - \mathcal{P}_\Omega X, \omega \rangle = 0 \quad (3.4)$$

for any $\omega \in \Omega$.

Orthogonal projections have the following useful properties:

- (1) The projection $\mathcal{P}_\Omega X$ coincides with the expectation $E[X | \Omega]$ in linear models with Gaussian shocks.
- (2) Each $X \in L^2$ has a unique representation as a sum of an element in Ω and an element of $\Omega^\perp$, i.e.

$$X = \mathcal{P}_\Omega X + (I - \mathcal{P}_\Omega) X \quad (3.5)$$

where $\Omega^\perp$ is the orthogonal complement of Ω in L^2

- (3) $X \in \Omega$ if and only if $\mathcal{P}_\Omega X = X$.
- (4) $X \in \Omega^\perp$ if and only if $\mathcal{P}_\Omega X = 0$.
- (5) $\Omega_1 \subseteq \Omega_2$ if and only if $\mathcal{P}_{\Omega_1} X = \mathcal{P}_{\Omega_2} X$ for all $X \in L^2$.

Property (1) is obviously useful as it allows us to use property (2) - (5) to analyze traders's expectations in a model with linear constraints and Gaussian shocks. Property (2) will be used in the proof of Proposition 3 where we decompose bond prices into a component that is the projection of future short rates on public information and into a component that is orthogonal to public information. Property (3) and (4) are useful to show that individuals can predict average expectations errors when information sets are non-nested. Finally, Property (5) can be used to show both that in the absence of supply shocks the expectations hypothesis

holds in our model with respect to a public information set and that individual traders can predict excess return when information sets are non-nested.

3.2. The Expectations Hypothesis and A Common Information Benchmark. We are now in a position to more concisely state what we mean by the expectations hypothesis.

Definition 3. (*The Expectations Hypothesis.*) *The expectations hypothesis of the term structure of interest rates is said to hold with respect to Ω_t if the implied forward rate*

$$f_t^n \equiv \tilde{b}_t^n - \tilde{b}_t^{n+1} \quad (3.6)$$

equals the projection of the short rate in period $t + n$ on to Ω_t

$$f_t^n = \mathcal{P}_{\Omega_t} r_{t+n} \quad \forall t, n \quad (3.7)$$

This formulation turns out to be more convenient to work with than an equivalent formulation based on the orthogonality of the residuals γ_t^n in (3.1). The important aspect of this definition for our purposes is that it makes it clear that whether the expectations hypothesis hold or not depends on what information set one conditions on.

To help intuition for how the term structure of interest rates are affected by non-nested information sets, we first confirm that the expectations hypothesis hold for the bond price equation above when all traders share a common information set. Consider the 3 month bond price equation

$$\tilde{b}_t^3 = -r_t - \int E[r_{t+1} | \Omega_t(j)] dj - \int E \left[\int E[r_{t+2} | \Omega_{t+1}(j')] dj' | \Omega_t(j) \right] dj \quad (3.8)$$

and start from the assumption that the period t information set Ω_t is common across all traders. If traders do not forget, the sequence of information sets $\{\Omega_t\}_{t=0}^\infty$ is a filtration so that $\Omega_0 \subseteq \Omega_1 \subseteq \Omega_2 \dots \subseteq \Omega_t$. Using this property one can derive the expectations hypothesis in the following way.

In a linear model with Gaussian shocks we can replace the expectations with orthogonal projections and write (3.8) as

$$\tilde{b}_t^3 = -r_t - \mathcal{P}_{\Omega_t} r_{t+1} - \mathcal{P}_{\Omega_t} \mathcal{P}_{\Omega_{t+1}} r_{t+2}. \quad (3.9)$$

By property (5) of projections above, we have that $\mathcal{P}_{\Omega_t} \mathcal{P}_{\Omega_{t+1}} r_{t+2} = \mathcal{P}_{\Omega_t} r_{t+2}$ since $\Omega_t \subseteq \Omega_{t+1}$. The interpretation of this for our model is that traders in period t cannot predict how they will revise their expectation in period $t + 1$ of the period $t + 2$ short rate, i.e. the sequence $\{E[r_{t+n} | \Omega_{t+s}]\}_{s=0}^n$ is a martingale. Applying this result to the bond pricing equation (3.8) gives

$$\tilde{b}_t^3 = -r_t - \mathcal{P}_{\Omega_t} r_{t+1} - \mathcal{P}_{\Omega_t} r_{t+2} \quad (3.10)$$

or equivalently

$$y_t^3 = \frac{1}{3} \sum_{s=0}^2 E[r_{t+s} | \Omega_t] \quad (3.11)$$

The expectation hypothesis then holds with respect to Ω_t .

We can also verify the assertion that the formulation of the expectations hypothesis stated as in Definition 3 above using implied forward rates, is equivalent to the statement that γ_t^n

in (3.1) is orthogonal to Ω_t . To see why, write out the expression for the two period forward rate

$$f_t^2 = \tilde{b}_t^2 - \tilde{b}_t^3 \quad (3.12)$$

$$= -r_t - \mathcal{P}_{\Omega_t} r_{t+1} \quad (3.13)$$

$$+ r_t + \mathcal{P}_{\Omega_t} r_{t+1} + \mathcal{P}_{\Omega_t} \mathcal{P}_{\Omega_{t+1}} r_{t+2} \quad (3.14)$$

$$= \mathcal{P}_{\Omega_t} r_{t+2} \quad (3.15)$$

The implication that the n period forward rate should be an optimal predictor of spot rate in period $t+n$ (or some equivalent formulation) forms the basis for most empirical tests of the expectations hypothesis, see for instance Backus, Foresi, Mozumdar and Wu (2001). In the next subsection we demonstrate that with privately informed traders, the implied forward rates will generally not be equal to traders' expectations about future short rates.

4. NON-NESTED INFORMATION SETS AND THE TERM STRUCTURE OF INTEREST RATES

We now turn to the implications for the term structure of interest rates of relaxing the assumption that traders share the same information. This section contains the main theoretical results of the paper and we prove that if information sets are non-nested, individual traders can predict excess returns even if the price of risk is constant. Individual traders can also predict average prediction errors made by other agents, which turns out to be an integral part of the new dynamics introduced by non-nested information sets. However, we are also able to show that these new dynamics are orthogonal to public information.

We start by defining what we mean by non-nested information sets.

Definition 4. *The subspace $\Omega_t(j)$ is the space spanned by the history of variables observed by trader j at period t . Projections onto $\Omega_t(j)$ are denoted $\mathcal{P}_{t,j}$.*

Definition 5. *Information sets of traders indexed by $j, i \in (0, 1)$ are said to be non-nested so that $\Omega_t(j) \not\subseteq \Omega_{t+s}(i)$ and $\Omega_t(i) \not\subseteq \Omega_{t+s}(j)$ if and only if*

$$\mathcal{P}_{t,j} r_{t+n} \neq \mathcal{P}_{t+s,i} r_{t+n} \quad (4.1)$$

for all $t = 0, 1, 2, \dots : s = 0, 1, 2, \dots, n$ and $j \neq i$.

Defining non-nested information sets through the implications for projections of short rates onto individual trader's information sets is somewhat tailored to the needs of this paper. A more general definition would simply state that information sets are non-nested if projections of *any* random variable onto individual traders' information sets differ. The definition used here is designed to avoid trivial cases where projections only differ about uninteresting quantities. We therefore define information sets as non-nested only if they imply that expectations about future short rates differ across agents.⁵

We start by proving that non-nested information sets are sufficient for individual traders to be able to predict excess returns. We do this in two steps. In the first (perhaps trivial step), we prove that if individual traders' projections of future short rates are dispersed, then the

⁵Sargent (1991), Kasa (2000) and Pearlman and Sargent (2005) all show (using different methods) that in the model of Townsend (1983), agents expectations coincide even though agents observe private signals. The reason is that in that model equilibrium prices reveal the information of other agents perfectly.

n period forward rate f_t^n cannot coincide generally with traders' expectations about future short rates.

Proposition 1. *The forward rate f_t^n is agent j 's optimal prediction of the short rate n periods ahead, if and only if it coincides with the orthogonal projection of r_{t+n} onto trader j 's information set $\Omega_t(j)$ so that*

$$\mathcal{P}_{t,j}r_{t+n} = f_t^n \quad (4.2)$$

holds. The equality (4.2) will only hold generally when traders' information sets coincide.

Proof. The first half of the proposition holds by the uniqueness and optimality of orthogonal projections. The second half states that $\mathcal{P}_{t,j}r_{t+n} = f_t^n$ hold generally only when information sets are nested. To see why this is true, note that if it was true generally that

$$\mathcal{P}_{t,j}r_{t+n} = f_t^n \forall j, t, n$$

then the ex ante symmetry of traders implies that

$$\mathcal{P}_{t,j}r_{t+n} = \mathcal{P}_{t,i}r_{t+n} \forall j, i, t, n \quad (4.3)$$

or that the forward rate is the best prediction of trader j only when it is also the best prediction for all others traders, i.e. when information sets are nested. $\square$

Intuitively, if the distribution across traders of expected future short rates are non-degenerate, all points on the support of the distribution cannot coincide with the forward rate, which is a single number. This may seem like an obvious result, but the uniqueness of orthogonal projections makes it nevertheless interesting. The reason is that the uniqueness of orthogonal expectations implies that the forecast errors one would make if forward rates are taken at face value as predictions of future short rates are systematically predictable by individual traders since $\mathcal{P}_{t,j}r_{t+n} \neq f_t^n$ implies that $\mathcal{P}_{t,j}(r_{t+n} - f_t^n) \neq 0$ since $\mathcal{P}_{t,j}f_t^n = f_t^n$ if $f_t^n \in \Omega_t(j)$. The next proposition shows that this in turn implies that traders can systematically predict excess returns.

Proposition 2. *If the projection of the period $t+n$ short rate onto trader j 's information set differ from the n period forward rate, so that $\mathcal{P}_{t,j}r_{t+n} = f_t^n$ does not hold for all n and t , excess returns are predictable by trader j .*

Proof. Excess returns on an n -period bond are predictable if

$$-\tilde{b}_t^n \neq r_t + \mathcal{P}_{t,j}r_{t+1} + \dots \mathcal{P}_{t,j}r_{t+n-1} \quad (4.4)$$

that is if the return $-\tilde{b}_t^n$ on holding an n period bond to maturity differ from the trader j 's expected return of holding a sequence of one period bonds for n number of periods. By the identity

$$-\tilde{b}_t^n \equiv r_t + f_t^1 + \dots + f_t^{n-1} \quad (4.5)$$

we thus have that excess returns are predictable if

$$\sum_{s=1}^n \mathcal{P}_{t,j}r_{t+s} = \sum_{s=1}^n f_t^s \quad (4.6)$$

which by Proposition 1 cannot hold for all $n = 1, 2, \dots$. To see the last point, assume that for some horizon n the equality (4.6) holds, then by Proposition 1 we have that

$$\sum_{s=1}^n \mathcal{P}_{t,j} r_{t+s} + \mathcal{P}_{t,j} r_{t+n+1} \neq \sum_{s=1}^n f_t^n + f_t^{n+1} \quad (4.7)$$

This concludes the proof. $\square$

Proposition 1 and 2 demonstrated that a non-degenerate distribution of expectations is sufficient for excess returns to be predictable by individual traders. The next two propositions helps us understand more about the dynamics introduced to the term structure by non-nested information sets and how these dynamics relate to public and private information. First, we demonstrate that non-nested information sets imply that individual traders can predict the average prediction errors made by other traders.

Proposition 3. *Average period $t + s$ projection errors of the short rate in period $t + n$*

$$r_{t+n} - \int \mathcal{P}_{t+s,j'} r_{t+n} dj'$$

are orthogonal to $\Omega_t(j)$

$$\mathcal{P}_{t,j} \left(r_{t+n} - \int \mathcal{P}_{t+s,j'} r_{t+n} dj' \right) = 0 \quad (4.8)$$

if and only if $\Omega_t(j) \subseteq \Omega_{t+s}(i)$ for all $s = 0, 1, 2, \dots$ and all $j \neq i \in (0, 1)$.

Proof. The expression (4.8) can be rearranged to

$$\mathcal{P}_{t,j} r_{t+n} = \mathcal{P}_{t,j} \int \mathcal{P}_{t+s,j'} r_{t+n} dj'. \quad (4.9)$$

Since traders do not receive signals that are informative about the idiosyncratic noise in other traders' signals, we have that

$$\mathcal{P}_{t,j} \int \mathcal{P}_{t+s,j'} r_{t+n} dj' = \mathcal{P}_{t,j} \mathcal{P}_{t+s,i} r_{t+n} \text{ for all } i, j \in (0, 1) : i \neq j. \quad (4.10)$$

That is, an individual trader j 's expectation about average expectations coincide with his expectation of trader i 's expectation for any $j \neq i$. By property (5) of projections we know that

$$\mathcal{P}_{t,j} r_{t+n} = \mathcal{P}_{t,j} \mathcal{P}_{t+s,i} r_{t+n} \quad (4.11)$$

if and only if $\Omega_t(j) \subseteq \Omega_{t+s}(i)$ which completes the proof. $\square$

Proposition 3 shows that individual traders not only can predict excess returns, they can also systematically predict the average prediction errors made by other traders. In the next proposition, we show that we can express forward rates (and in extension, the term structure) as a function of first order average projections of future short rates and higher order prediction errors, that is, predictions about the difference between future short rates and other traders' predictions about future short rates.

Proposition 4. *The forward rate f_t^n can be decomposed into the the average first order projection of r_{t+n} , a sum of higher order projection errors and the exogenous supply shocks η_t^n and η_t^{n+1} .*

Proof. For convenience, first define the notation

$$\prod_{s=0}^{n-1} \int \mathcal{P}_{t+s,j} r_{t+n} \equiv \int \mathcal{P}_{t,j} \int \mathcal{P}_{t+1,j'} \dots \int \mathcal{P}_{t+n-1,j''} r_{t+n} dj'' \dots dj' dj. \quad (4.12)$$

and rewrite the definition of the n period forward rate as

$$f_t^n = \prod_{s=0}^{n-1} \int \mathcal{P}_{t+s,j} r_{t+n} + (\eta_t^n - \eta_t^{n+1}). \quad (4.13)$$

Add and subtract $\prod_{s=0}^{n-2} \int \mathcal{P}_{t+s,j} r_{t+n}$ from the r.h.s. of (4.13) to get

$$\begin{aligned} f_t^n &= \prod_{s=0}^{n-2} \int \mathcal{P}_{t+s,j} r_{t+n} \\ &\quad - \prod_{s=0}^{n-2} \int \mathcal{P}_{t+s,j} \left(r_{t+n} - \int \mathcal{P}_{t+n-1,j'''} r_{t+n} dj''' \right) \\ &\quad + (\eta_t^n - \eta_t^{n+1}) \end{aligned} \quad (4.14)$$

The term in brackets on the second line of (4.14) is the average prediction error across traders. However, the first line still contains $n - 2$ order expectations. We can repeatedly add and subtract $\prod_{s=0}^{n-m} \int \mathcal{P}_{t+s,j} r_{t+n}$ until we get

$$\begin{aligned} f_t^n &= \int \mathcal{P}_{t,j} r_{t+n} \\ &\quad - \sum_{m=1}^n \prod_{s=0}^{n-m} \int \mathcal{P}_{t+s,j} \left(r_{t+n} - \int \mathcal{P}_{t+m-1,j} r_{t+n} \right) \\ &\quad + (\eta_t^n - \eta_t^{n+1}) \end{aligned} \quad (4.15)$$

that is, an expression for the forward rate containing the average first order expectations (the first line), a sum of higher order prediction errors (second line) and the supply shocks η_t^n and η_t^{n+1} (third line). $\square$

In a model with perfect or common information, the higher order prediction errors on the second line would of course be zero. The new dynamics introduced to the term structure by non-nested information sets are thus completely contained in the higher order prediction error term. By Proposition 3 above, we know that individual traders can predict average prediction errors. However, it is straightforward to show that these new dynamics here expressed as a sum of higher order prediction errors are orthogonal to public information. Before proving this statement, we first define two relevant information sets.

Definition 6. The subspace Ω_t^p is the space spanned by the history of publicly observable variables in period t so that $\Omega_t^p \subseteq \Omega_t(j)$ for all j . Projections onto Ω_t^p are denoted $\mathcal{P}_t^p$.

Definition 7. The subspace $\Omega_t^{\perp p}(j)$ is the orthogonal complement of Ω_t^p in $\Omega_t(j)$. Projections onto $\Omega_t^{\perp p}(j)$ are denoted $\mathcal{P}_{t,j}^{\perp p}$.

Proposition 5. The forward rate f_t^n can be decomposed into the projection of r_{t+n} onto the public information set Ω_t^p , the supply shocks and terms that are orthogonal to public information.

Proof. Use that any projection onto $\Omega_t(j)$ can be decomposed into a sum of the projection onto Ω_t^p and a projection onto the orthogonal complement $\Omega_t^{\perp p}(j)$ to rewrite the expression of the forward rate (4.15) as

$$\begin{aligned} f_t^n &= \mathcal{P}_t^p r_{t+n} + \int \mathcal{P}_{t,j}^{\perp p} r_{t+n} \\ &\quad - \sum_{m=1}^n \prod_{s=0}^{n-m} \int \mathcal{P}_{t+s,j}^p \left(r_{t+n} - \int \mathcal{P}_{t+m-1,j} r_{t+n} \right) \end{aligned} \quad (4.16)$$

$$\begin{aligned} &\quad - \sum_{m=1}^n \prod_{s=0}^{n-m} \int \mathcal{P}_{t+s,j}^{\perp p} \left(r_{t+n} - \int \mathcal{P}_{t+m-1,j} r_{t+n} \right) \\ &\quad + (\eta_t^n - \eta_t^{n+1}) \end{aligned} \quad (4.17)$$

Since $\Omega_t^p \subseteq \Omega_{t+s}(j)$ for all j and $s = 0, 1, \dots, m-1$ we have that

$$\mathcal{P}_{t+s,j}^p r_{t+n} = \mathcal{P}_{t+s,j}^p \int \mathcal{P}_{t+m-1,j} r_{t+n} \quad (4.18)$$

for all j and $s = 0, 1, \dots, m-1$ by Property 5 of orthogonal projections. The term on the second line of is thus identically zero and the forward rate can thus be expressed as

$$\begin{aligned} f_t^n &= \mathcal{P}_t^p r_{t+n} + \int \mathcal{P}_{t,j}^{\perp p} r_{t+n} \\ &\quad - \sum_{m=1}^n \prod_{s=0}^{n-m} \int \mathcal{P}_{t+s,j}^{\perp p} \left(r_{t+n} - \int \mathcal{P}_{t+m-1,j} r_{t+n} \right) \\ &\quad + (\eta_t^n - \eta_t^{n+1}) \end{aligned} \quad (4.19)$$

which concludes the proof. $\square$

Proposition 5 demonstrates that the new term structure dynamics introduced by non-nested information sets are orthogonal to public information. This is natural, since by definition, all traders know that all traders know, and so on, that all traders know that all traders observe the public signal. The component of other traders' projection errors that are predictable by an individual trader j must therefore be orthogonal to public information.

This ends the theoretical part of the paper. Before turning to the data, we summarize our findings so far. Individual traders can identify and take advantage of predictable excess returns, even though the price of risk is constant. The model thus provides an alternative explanation of predictable excess returns that is not based on agents valuing a marginal

increase in wealth differently in different states of the world. If one conditions on non-public information, excess returns *should* be predictable and the expectations hypothesis *should* not hold, even in a model with constant price of risk. However, we also demonstrated that a conditioning down argument similar to that made by Hansen and Sargent (1991) (in a different context) still holds: Implied forward rates coincide with optimal predictions of future short rates made by an econometrician using only public signals up to an exogenous supply shock process. That is, the yield dynamics due to traders attempting to exploit average prediction errors are orthogonal to public information.

5. AN EMPIRICAL MODEL

In this section, an explicit model of the term structure is presented in which traders have private information that is relevant for predicting future short rates. Apart from the information structure, the model is kept as simple as possible.

5.1. The short rate. The short interest rate r_t is an inertial exogenous process given by

$$r_t = x_t^1 + x_t^2 + \phi r_{t-1} \quad (5.1)$$

where the factors x_t^1 and x_t^2 follow the vector auto regressive process

$$\begin{bmatrix} x_t^1 \\ x_t^2 \end{bmatrix} = \begin{bmatrix} \rho_1 & 0 \\ 0 & \rho_2 \end{bmatrix} \mathbf{x}_{t-1} + \Sigma \varepsilon_t \quad (5.2)$$

$$\varepsilon_t \sim N(0, I) \quad (5.3)$$

and Σ is a lower triangular matrix.

5.2. Demand and supply of long maturity bonds. Traders are indexed by $j \in (0, 1)$ and maximize the discounted expected utility of future log consumption

$$U_t(j) = \sum_{s=0}^{\infty} \beta^s E_t [\log C_{t+s} \mid \Omega_t(j)] \quad (5.4)$$

where consumption is financed solely from wealth. Campbell and Viceira (2002) show that the optimal portfolio weights for trader j 's holdings of bond of different maturities can be approximated by the vector $\alpha_t(j)$

$$\alpha_t(j) = \Sigma^{-1} \left(E \left[\mathbf{b}_{t+1}^{(-1)} \mid \Omega_t(j) \right] - \mathbf{b}'_t - r_t \right) + \mathbf{c}^d \quad (5.5)$$

where $\mathbf{c}^d$ is a vector of mean portfolio weights. The vector of bond supply $\mathbf{s}_t$ is stochastic and proportional to aggregate wealth W_t

$$\mathbf{s}_t = W_t \Sigma^{-1} (1 + \eta_t) \quad (5.6)$$

$$\eta_t \sim N(0, \Sigma_\eta) \quad (5.7)$$

where the term Σ^{-1} normalizes the variance of the supply shocks η_t . Equating aggregate log demand $W_t \int \alpha_t(j) dj$ and supply gives the deviation of the log price $\tilde{b}_t^n$ of an n periods to maturity zero coupon bond

$$\tilde{b}_t^n = -r_t + \int E \left[\tilde{b}_{t+1}^{n-1} \mid \Omega_t(j) \right] dj - \eta_t^n. \quad (5.8)$$

The price of an n periods to maturity bond in period t thus depends the average expectation in period t of the price of a $n - 1$ period bond in period $t + 1$. The more a trader expects to be able to sell a bond for in the future, the more is he willing to pay for it today. However, risk aversion prevents the most optimistic trader from demanding all of the available supply. The bond price formula (5.8) can be used to price any maturity bond. The yield of a bond with n periods to maturity is (as usual) given by dividing the log bond price by

$$y_t^n = -n^{-1}\tilde{b}_t^n \quad (5.9)$$

5.3. Trader's information sets. All traders observe a vector of public signals containing the current short rate r_t and selected bond yields collected in the vector $\mathbf{y}_t$. In addition, each trader also observes a private signal $s_t(j)$ which is a sum of the first factor x_t^1 and an idiosyncratic noise component $\zeta_t(j)$

$$s_t(j) = x_t^1 + \zeta_t(j) \quad (5.10)$$

$$\zeta_t(j) \sim N(0, \sigma_\zeta^2) \quad (5.11)$$

The vector $S_t(j)$ defined as

$$S_t(j) = [s_t(j) \quad r_t \quad \mathbf{y}_t']' \quad (5.12)$$

then contains the signals that trader j observes in period t .

The model can be solved using the method proposed in Nimark (2007) and then put in the form

$$X_t = MX_{t-1} + Nu_t \quad (5.13)$$

$$\mathbf{y}_t = D_1X_t + D_2r_{t-1} + D_3\eta_t \quad (5.14)$$

where $\mathbf{y}_t$ is a vector of yields of bonds of different maturities, and X_t is a vector with stacked higher order expectations of the factors contained in $\mathbf{x}_t \left(\equiv [x_t^1 \quad x_t^2]' \right)$

$$X_t \equiv \left[\mathbf{x}_{t|t}^{(0)'} \quad \mathbf{x}_{t|t}^{(1)'} \quad \cdots \quad \mathbf{x}_{t|t}^{(\bar{k})'} \right]' \quad (5.15)$$

The higher order expectations of $\mathbf{x}_t$ are defined recursively as

$$\mathbf{x}_{t|t}^{(k)} \equiv \int E \left[\mathbf{x}_{t|t}^{(k-1)} \mid \Omega_t(j) \right] dj$$

starting from $\mathbf{x}_{t|t}^{(0)} = \mathbf{x}_t$. 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 $k \rightarrow \infty$. The Appendix provide details on the solution procedure.

5.4. The Estimated Model. The model is in a form that can be estimated directly by likelihood based methods. I use monthly data of the Federal Funds rate and the 3, 12, 24, 36, 48 and 60 month annualized interest rates on Treasuries taken from the CRSP data base. The sample period is from January 1964 to December 2007 (528 monthly observations) and chosen to coincide with the sample period used by Cochrane and Piazzesi (2008) and Duffee (2008). The time series are demeaned. The posterior mode (based on uniform truncated priors) is found using a simulated annealing numerical maximizer and the posterior parameter distributions are simulated by 1 000 000 draws from an Adaptive Random Walk Metropolis

Hastings algorithm (see Haario, Saksman and Tamminen (2001)). The results are reported in Table 1 and the same information is given graphically in Figure 6 in the Appendix.

Table 1
Parameter Estimates 1964:1-2007:12

	Mode	2.5%-97.5%
Short rate process		
ρ_1	0.998	0.997 - 0.999
ρ_2	0.38	0.33 - 0.43
ϕ	0.84	0.82 - 0.85
σ_1	0.08	0.07 - 0.09
σ_2	0.52	0.49 - 0.56
σ_{12}	-0.10	(-0.14) - (-0.05)
Noise in private signal		
σ_ζ	1.2×10^{-5}	$5.4 \times 10^{-5} - 4.5 \times 10^{-3}$
Bond supply shocks		
η_3	0.70	0.65 - 0.74
η_{12}	0.37	0.35 - 0.40
η_{24}	0.16	0.16 - 0.18
η_{36}	1.6×10^{-5}	$9.1 \times 10^{-5} - 2.1 \times 10^{-2}$
η_{48}	0.15	0.14 - 0.16
η_{50}	0.24	0.22 - 0.25
Probability interval computed using 1 000 000 draws from a MCMC generated by an Adaptive Metropolis algorithm.		

By themselves, the posterior estimates are not very interesting, but we can note that all parameters appear to be well-identified.⁶ We can also note that the private signal appears to be very precise as evidenced by the low estimated standard deviation of the idiosyncratic noise component $\zeta_t(j)$. Also, supply shocks (or measurement errors) in the 3 year bond yields are of several orders of magnitude less than that of the supply shocks of bonds of the other maturities.

Figure 2 displays one period ahead fitted values together with the actual (demeaned) data series. The fit looks pretty good, though admittedly, these are time series that are relatively easy to fit given their very high persistence.

5.5. The estimated dispersion of expectations. The dynamics of the model depend importantly on that traders information sets are non-nested. As noted in the introduction, one implication of non-nested information sets is the expectations will be dispersed. This fact can be used as an independent check to gauge whether the estimated model requires a reasonable degree of expectations dispersion to fit the data. In Figure 3, the estimated posterior histograms of the standard deviation of individual traders expectations of the short rate 12, 36 and 60 months ahead are plotted. That is, Figure 3 shows the histograms of the

⁶As a convergence check, Figure 7 in the Appendix shows recursive plots of the diagonal of the covariance matrix of the MCMC of the posterior.

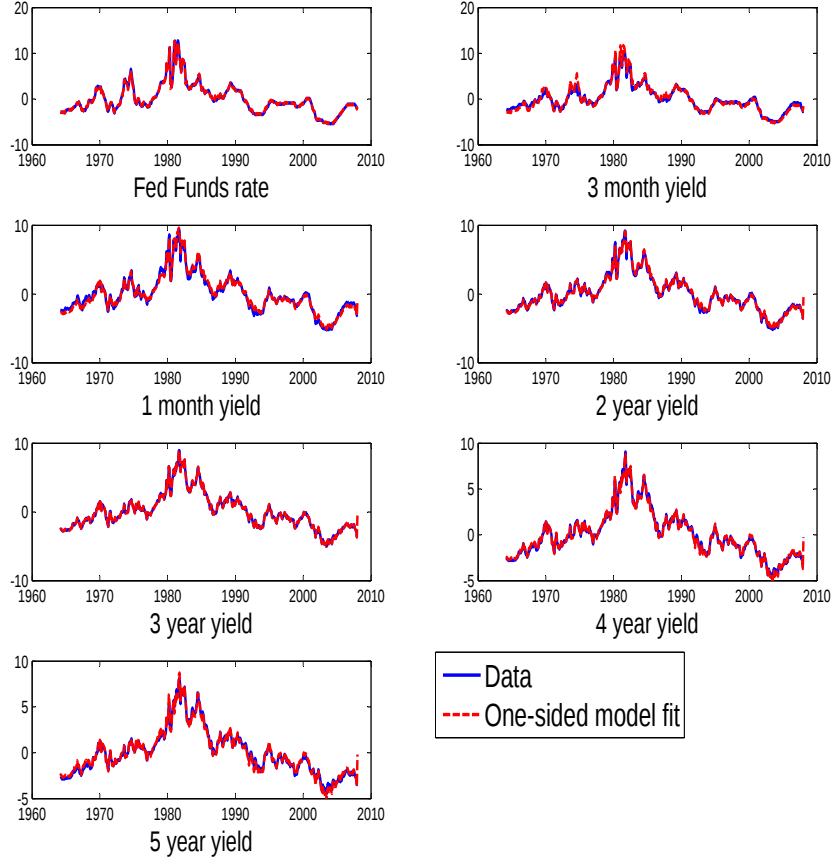


FIGURE 1. Data (blue solid) and one-sided model fit (red dashed).

quantity

$$\sqrt{E \left(\mathcal{P}_{t,j} r_{t+n} - \int \mathcal{P}_{t,j'} r_{t+n} \right)^2} \quad (5.16)$$

for $n = 12, 36$ and 60 computed for 10 000 draws from the estimated posterior parameter distribution. Two things stand out. First, the standard deviations of the dispersion of individual traders' expectations is quite small. Most of the posterior probability weight (95%) is below 1.1×10^{-6} , 7.7×10^{-4} and 4.6×10^{-3} (percentage points) respectively for the 12, 36 and 60 month horizon. The dispersion is thus significantly less than that of the forecast surveys analyzed by Swanson (2006). Secondly, the dispersion is increasing with horizon. The dispersion of expectations is fundamentally driven by the dispersion of traders' estimates of the factors x_t^1 and x_t^2 that drive short rates. These differences accumulate over time because

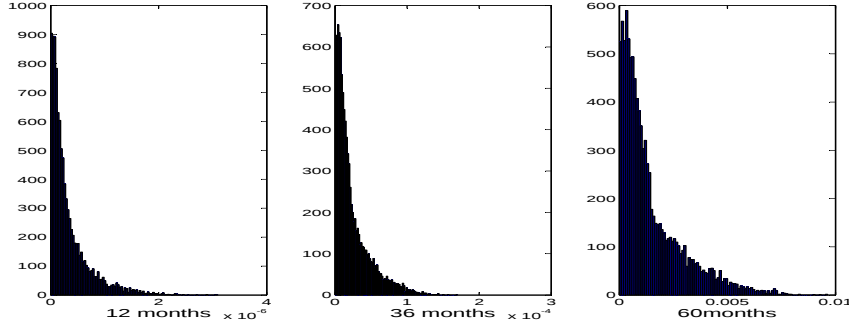


FIGURE 2. Histograms of standard deviations of dispersion of individual trader's expectations of short rates 12, 36 and 60 months ahead.

of the inertial character of short interest rates which explains why expectations about the short rate 5 years ahead are more dispersed than expectations about the short rate 1 year ahead.

While the dispersion of individual expectations is an intuitive measure of disagreement, it does not matter directly for the dynamics of the term structure as only average (higher order) expectations enter the bond pricing formula. We therefore also plot the posterior histograms of the standard deviations of the difference between the average expectation of the short rate in $t+n$ and the n -period ahead higher order expectations term in the forward rate. That is, Figure 4 displays the histogram of the quantity

$$\sqrt{E \left(\int \mathcal{P}_{t,j} r_{t+n} - \prod_{s=0}^{n-1} \int \mathcal{P}_{t+s,j} r_{t+n} \right)^2} \quad (5.17)$$

for $n = 12, 36$ and 60 computed for 10 000 draws from the estimated posterior parameter distribution. This provide a measure of the dispersion of different orders of expectations. The dispersion between average first and higher order expectations are larger than the dispersion of expectations across traders, but especially for short horizons, still quite small. The minimum length 95% probability intervals range from practically zero (for all horizons) to 3.4×10^{-5} , 0.01 and 1.2 percentage points for the 12, 36 and 60 month horizon. Again, and for the same reasons as before, the dispersion increases with horizon. The importance of higher order expectations for the dynamics of the term structure thus increases with maturity.

5.6. Quantifying the gains from private information. [TBC]

6. THE MODEL AND THE EVIDENCE FROM STATISTICAL TERM STRUCTURE MODELS

This section demonstrates that the estimated model can account for some of the findings of statistical models of the term structure.

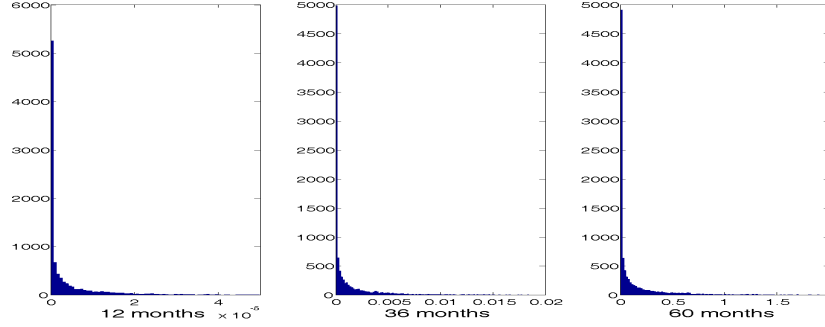


FIGURE 3. Histograms of estimated posterior variance of difference between average expectations of short rate and higher order term in forward rate 12, 36 and 60 months ahead.

6.1. A three factor no-arbitrage model. Affine three factor no-arbitrage models can provide a very good fit of the term structure of interest rates (e.g. Duffie and Kan 1996). However, if the world is characterized by bond markets where traders with non-nested information sets interact, low dimensional affine no-arbitrage models are fundamentally misspecified.⁷ It is well-known that in dynamic models where agents have non-nested information sets, natural state representations tend to be infinite dimensional (see for instance Townsend (1983), Sargent (1991) and Makarov and Rytchov (2009)). In the estimated model of the previous section, the infinite dimensional representation was approximated with an 12 dimensional state vector (i.e. $\bar{k} = 6$). This turns out to be sufficient to accurately represent the dynamics of the model with non-nested information. We now investigate what an affine three factor no-arbitrage model would find if the model of the previous section represents the true economy. We are particularly interested in finding out whether a three factor no-arbitrage model will correctly detect that the price of risk is constant in the model that generated the artificial data.

A three factor no arbitrage model can be described by the following equations (see Ang and Piazzesi (2003) for more details.) The three factors in the vector F_t follow

$$F_t = \Gamma F_{t-1} + \Psi e_t \quad (6.1)$$

where Γ is diagonal and Ψ is lower triangular with ones on the diagonal. (These are normalizations that do not affect the estimated yield dynamics.) The short rate is a function of the factors

$$r_t = \delta' F_{t-1} \quad (6.2)$$

and the deviation of the log price of an n period bond from its mean is then given by

$$\tilde{b}_t^n = B_n' F_t \quad (6.3)$$

where

$$B_n' = B_{n-1}' (\Gamma - \Psi \lambda) - \delta' \quad (6.4)$$

⁷Perhaps interestingly, in the limit case as the variance of the idiosyncratic noise in traders private signals tend to zero, the model presented here becomes an affine two factor no-arbitrage constant price of risk model.

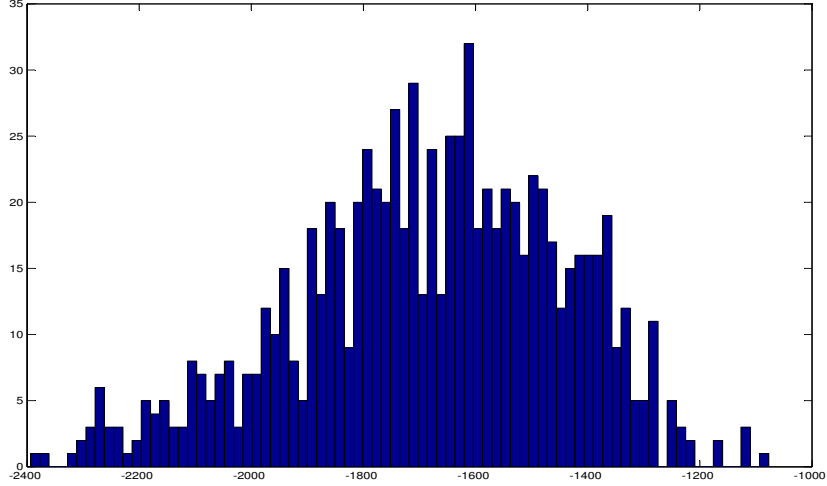


FIGURE 4. Distribution of log likelihood ratios of restricted and unrestricted three factor no-arbitrage model estimated on artificial data.

As before, yields can be computed as

$$y_t^n = -n^{-1} \tilde{b}_t^n \quad (6.5)$$

Imposing that the price of risk is constant equals setting $\lambda = \mathbf{0}$ (which also implies imposing that the expectations hypothesis hold).

The experiment we conduct is the following. We first draw parameters from the posterior distribution of our model and generate 528 observations. We then estimate the three factor no-arbitrage model described above (with added yield measurement errors) with and without imposing the restriction that the price of risk is constant and compare the marginal likelihoods of the restricted and unrestricted model. This procedure was repeated 800 times.⁸

Figure 3 displays the histogram of the log-likelihoods of the restricted model with $\lambda = \mathbf{0}$ minus the log-likelihood of the unrestricted model. If the restriction of no time-varying risk was supported by the data, this difference should on average equal zero. Instead, the average difference is about 1700. This means that the restricted model is only $e^{-1700} \approx 0$ as likely as the unrestricted model. In other words, the restricted model has zero probability of being the true model compared to the model with time varying risk premia, in spite of the fact that the data is generated by a model with *constant* price of risk. Of course, this does not prove that the actual data generating process is a model with non-nested information sets. However, it does demonstrate that the evidence from statistical models is not sufficient reason to conclude that the price of risk is time varying.

⁸The procedure is very time consuming and the 800 repetitions took about 2 weeks to compute on a not unusually slow computer.

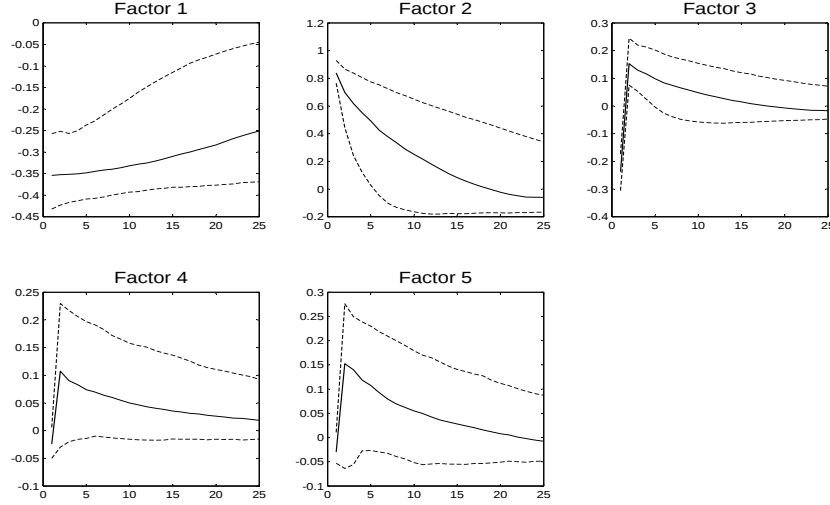


FIGURE 5. Impulse responses to orthogonal factor innovations, dashed lines are 5th and 95th percentile.

6.2. Hidden factors and predictable excess returns. Duffee (2008) provide evidence of a “hidden” factor that is insignificant in explaining the cross-section of yields but important for predicting short rates and in extension, excess returns. Duffee estimates a 5 factor model of the form

$$x_t^\dagger = D^\dagger x_{t-1}^\dagger + \Sigma^\dagger \epsilon_t \quad (6.6)$$

$$y_t = A + B^\dagger x_t^\dagger + \eta_t^\dagger \quad (6.7)$$

on US bond data and the estimated model can be rotated to compute the implied principal components. Duffee finds that while the first three principal components explain almost all of the unconditional variation in yields, the fifth principal component is important for explaining expected future short rates. He illustrates this by impulse response functions of the 5 factors and their effect on the short rate. If a factor is unimportant for the cross section, but important for predicting short rates (and in extension, excess returns) it will be evidenced by an impulse response function of the short rate to the factor in question that originates at zero but then becomes positive (or negative).

We investigated whether the hidden factor found by Duffee is consistent with the model presented here by again generating artificial data sets using parameter draws from the estimated posterior of our model with non-nested information sets. For each parameter draw, we simulated 528 months of data and then estimated Duffee’s five factor model and performed the rotations to compute the principal component factors. We then computed impulse responses of short rates to the orthogonal factors. This procedure was repeated 500 times.

Figure 4 shows the median impulse response and the 5th and 95th percentile.⁹ As we can see, the fourth and fifth factor have little effect on the short rate in the impact period, but becomes more important at longer time horizons. This is exactly what we should expect from a model where the term structure do not reveal all information about future short rates perfectly. If the state of the model would be revealed perfectly by the cross section of yields, no additional factor beyond the three (level, slope and curvature) that explains the cross sections would be useful to predict future yields. However, if the state is not revealed by the cross section, then by definition there must be additional factors that can help predict future yields.

7. CONCLUSIONS

In this paper we have presented a model of the term structure in which we relax the assumption that all traders have access to the same information sets. We used this model to demonstrate that predictable excess return are not necessarily evidence of a time varying price of risk. Instead we offered the alternative explanation that excess return may be more or less predictable depending on whether one conditions on information that is public or not.

Underlying the model is an assumption that the term structure does not reveal all information that is relevant for predicting future bond returns and we have cited statistical evidence that supports this view. While it is possible that traders all use the same information even when the state is not perfectly revealed, it is arguably less likely. If prices do not reveal the state of the world perfectly, traders have an incentive to collect information that can help predict future short rates and in extension excess returns and information will be more profitable if it is not shared with other traders. It appears natural then to assume that traders will seek to obtain private information.

We have also shown that the model presented here can explain some features of statistical factor models. Specifically, our model generates data that when used as a sample for the 5 factor model of Duffee (2008) generate estimates that reproduce the “hidden” factor documented by Duffee. We also demonstrated that a three factor no-arbitrage model mistakenly will attribute the dynamics due to non-nested information to time varying price of risk. To sum up, this paper suggest that the statistical evidence of a time varying price of risk may need to be reconsidered.

[TBC]

REFERENCES

- [1] Adam, Klaus, (2007), “Optimal Monetary Policy with Imperfect Common Knowledge”, *Journal of Monetary Economics*, vol. 54(2), pp 267-301.
- [2] Ang, A. and M. Piazzesi (2003), “A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables”, *Journal of Monetary Economics*, 50, pp.745-787.
- [3] Backus, D., A. Gregory and S. Zin, (1989), “Risk premiums and in the term structure: Evidence from artificial economies”, *Journal of Monetary Economics* 24, pp371-399.
- [4] Backus, D. Foresi and C. Telmer, (1998), Discrete Time Bond Pricing, unpublished manuscript, New York University.

⁹The percentile refer to the percentiles of the point estimates from Duffee’s model estimated on artificial data and can thus not be given a probabilistic interpretation.

- [5] Backus, David & Foresi, Silverio & Mozumdar, Abon & Wu, Liuren, 2001. "Predictable changes in yields and forward rates," *Journal of Financial Economics*, vol. 59(3), pp281-311.
- [6] Backus, D. and J. Wright, (2007), "Cracking the conundrum", NBER working paper13419.
- [7] Bekaert, G., Hodrick, R. and D.A. Marshall, (1997), "On biases in tests of the expectations hypothesis of the term structure of interest rates" *Journal of Financial Economics*, vol. 44(3), pp309-348.
- [8] Brockwell, P.J. and R.A. Davis, (2006), *Time Series: Theory and Methods*, Springer-Verlag.
- [9] Cambell, J.Y. and L.M. Viceira, 2002, *Strategic Asset Allocation*, Cambridge University Press, Cambridge, UK.
- [10] Cochrane, J. and M. Piazzesi, (2005), "Bond Risk Premia", *American Economic Review*.
- [11] Cochrane, J. and M. Piazzesi, (2008), "Decomposing the Yield Curve", mimeo University of Chicago.
- [12] Duffee, Gregory (2002), "Term Premia and Interest Rate Forecasts in Affine Models," *Journal of Finance*, 57, pp.405-443.
- [13] Duffie, Darrell and Rui Kan (1996), "Yield Factor Models of Interest Rates," *Mathematical Finance*, 64, pp.379-406.
- [14] Duffee, Greg, (2008), Information in (and not in) the term structure of interest rates, mimeo, Johns Hopkins University.
- [15] Evans, G. W. and S. Honkapohja, (2005). "An Interview With Thomas J. Sargent," *Macroeconomic Dynamics*, vol. 9(04), pp 561-583.
- [16] Haario, H, E. Saksman and J. Tamminen, 2001, "An Adaptive Metropolis Algorithm", *Bernoulli*, 7(2), 223-242.
- [17] Hansen, L.P. and T.J. Sargent, 1991, *Rational Expectations Econometrics*, Westview Press.
- [18] Harrison, J.M. and D. Kreps, (1978), "Speculative Behaviour in a Stock Market with Heterogenous Expectations", *Quarterly Journal of Economics*, pp323-336.
- [19] Kasa, Kenneth, 2000, "Forecasting the forecast of others in the frequency domain", *Review of Economic Dynamics*, 3, pp726-756.
- [20] Litterman, R., Scheinkman, J., (1991), "Common factors affecting bond returns", *Journal of Fixed Income* 1, 51-61.
- [21] Morris, S. and H.S. Shin, (2002), "The social value of public information", *American Economic Review* 92, pp1521-1534.
- [22] Nimark, K., (2007), "Dynamic Higher Order Expectations", working paper, Universitat Pompeu Fabra.
- [23] Pearlman, J.G. and T.J. Sargent, 2005, "Knowing the forecasts of others", *Review of Economic Dynamics*, Volume 8, pp480-497.
- [24] Piazzesi, M. and M. Schneider, (2006), "Equilibrium Yield Curves", NBER Working Papers 12609.
- [25] Rudebusch, G.D. and J. C. Williams, (2006), "Revealing the Secrets of the Temple: The Value of Publishing Central Bank Interest Rate Projections" NBER Working Papers 12638.
- [26] Sargent, Thomas J., 1991, "Equilibrium with Signal Extraction from Endogenous Variables", *Journal of Economic Dynamics and Control* 15, pp245-273.
- [27] Shiller, R. J.Y. Campbell and K.L. Shoenholtz, (1983), "Forward Rates and Future Policy: Interpreting the Term Structure of Interest Rates", *Brookings Papers on Economic Activity*, pp173-223.
- [28] Swanson, E.T., (2006), "Have Increases in Federal Reserve Transparency Improved Private Sector Interest Rate Forecasts?", *Journal of Money, Credit, and Banking* 38(3), April 2006, pp. 791-819.
- [29] Woodford, M. 2002, "Imperfect Common Knowledge and the Effects of Monetary Policy," in P. Aghion, R. Frydman, J. Stiglitz, and M. Woodford, eds., *Knowledge, Information, and Expectations in Modern Macroeconomics: In Honour of Edmund S. Phelps*, Princeton: Princeton University Press.
- [30] Xiong, W. and H. Yan, (2008), "Heterogenous Expectations and Bond Markets", mimeo, Princeton University.

APPENDIX A. SOLVING THE MODEL

The model is solved by using the method proposed in Nimark (2007). It involves deriving an explicit law of motion for higher order expectation of the exogenous processes x_t^1 and x_t^2 .

Define the exogenous state vector $\mathbf{x}_t$ as

$$\mathbf{x}_t = \begin{bmatrix} x_t^1 & x_t^2 \end{bmatrix}' \quad (\text{A.1})$$

Define a k th order average expectation of $\mathbf{x}_t$ recursively as

$$\mathbf{x}_{t|t}^{(k)} = \int E \left[\mathbf{x}_{t|t}^{(k-1)} \mid \Omega_t(j) \right] dj \quad (\text{A.2})$$

starting from the convention that $\mathbf{x}_t^{(0)} = \mathbf{x}_t$. Define a hierarchy of expectations from order 0 to k as

$$X_t \equiv \begin{bmatrix} \mathbf{x}_t^{(0)} \\ \mathbf{x}_t^{(1)} \\ \vdots \\ \mathbf{x}_t^{(\bar{k})} \end{bmatrix} \quad (\text{A.3})$$

Nimark (2007) show that a linear model with an exogenous state following a persistent process can be accurately approximated by first conjecturing a law of motion for the hierarchy of expectations in the form

$$X_t = MX_{t-1} + N\mathbf{e}_t \quad (\text{A.4})$$

where $\bar{k}$ is a finite integer. Define the average expectations operator $Q : \mathbb{R}^{\bar{k}} \rightarrow \mathbb{R}^{\bar{k}}$ as

$$Q = \begin{bmatrix} \mathbf{0}_{2\bar{k} \times 2} & I_{\bar{k}-2} \\ \mathbf{0}_{2 \times (\bar{k}-2)} & \end{bmatrix} \quad (\text{A.5})$$

that is Q moves a hierarchy of expectations one step up in order of expectations. Define the average one step ahead expectation operator $\bar{M} : \mathbb{R}^{\bar{k}} \rightarrow \mathbb{R}^{\bar{k}}$

$$\bar{M} = MQ \quad (\text{A.6})$$

so that for a given law of motion (A.4) we can then price bonds recursively using $\bar{M}$ or

$$b_t^n = \begin{bmatrix} \mathbf{1}_{1 \times 2} & \mathbf{0} \end{bmatrix} \bar{M}^{n-1} + \phi b_t^{n-1} - \sum_{j=1}^n \phi^j r_{t-1} \quad (\text{A.7})$$

where the yield on an n periods to maturity bond is given by

$$y_t^n \equiv -n^{-1} b_t^n \quad (\text{A.8})$$

We can write the vector of signals $S_t(j)$ as a function of the state

$$S_t(j) = \begin{bmatrix} s_t(j) & r_t & \mathbf{y}_t' \end{bmatrix}' \quad (\text{A.9})$$

$$= L_1 X_t + L_2 \mathbf{u}_t \quad (\text{A.10})$$

where L is a selector matrix picking out the linear combination of the states that are observables and

$$\mathbf{u}_t = \begin{bmatrix} \eta_t & \varepsilon_t & \eta_t \end{bmatrix}' \quad (\text{A.11})$$

is a shock vector. Agent j 's updating equation of his state estimate will then follow

$$X_{t|t} = MX_{t-1} + K(S_t(j) - L_1 MX_{t-1}) \quad (\text{A.12})$$

Rewriting the observables vector $S_t(j)$ as a function of the lagged state and taking averages across traders and appending it to the exogenous state gives us the conjectured form of the law of motion of $\mathbf{x}_{t|t}^{(0;\bar{k})}$

$$M = \begin{bmatrix} \rho & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + \begin{bmatrix} \mathbf{0}_{2 \times 2} & \mathbf{0} \\ \mathbf{0} & [M - KLM]_{11} \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ [KLM]_{11} \end{bmatrix} \quad (\text{A.13})$$

$$N = \begin{bmatrix} \mathbf{1}_{1 \times 2} & 0 \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ [KLN]_{11} \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & K_3 \end{bmatrix} \quad (\text{A.14})$$

the Kalman gain K in (A.12) is given by

$$K = (PL'_1 + G)(LPL' + L_2\Sigma_{uu}L'_2)^{-1} \quad (\text{A.15})$$

$$P = W(P - (PL'_1 + G)(L_1PL'_1 + L_2\Sigma_{uu}L'_2)^{-1}(PL'_1 + G)')W' + \Sigma_{ee} \quad (\text{A.16})$$

where

$$G = E(\mathbf{e}_t \mathbf{u}'_t L'_2) \quad (\text{A.17})$$

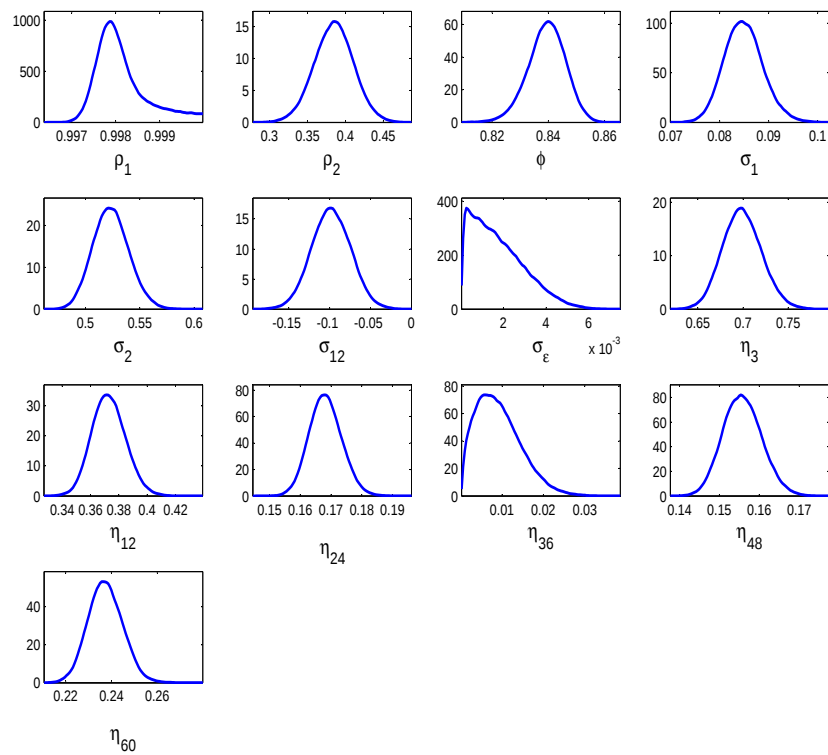


FIGURE 6. Estimated posterior densities of structural parameters generated using 1 000 000 draws from an Adaptive Random Walk Metropolis-Hastings algorithm.

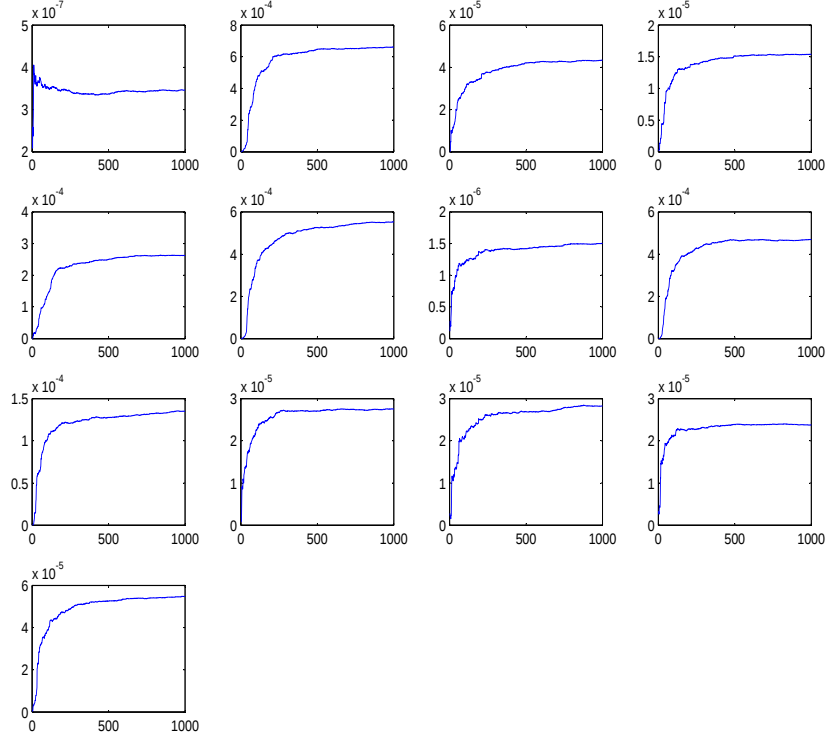


FIGURE 7. Convergence of MCMC: Recursive plots of diagonal from covariance matrix of MCMC