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) creates what to an econometrician approaching the data with a different model would appear to be time varying risk premia. The model thus provides a re-interpretation of statistical evidence 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. We further argue that the "hidden" factor that help predict excess returns as documented Duffee (2008) is an intrinsic feature of models with imperfectly informed traders. The model is estimated using monthly data on US short to medium term Treasuries from 1964 to 2007. The model provides a good fit of the data and demonstrates that it is feasible to estimate dynamic models with privately informed agents.

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 this price differs from what an individual trader would be willing to pay for the bond if he had to hold the bond until it matures, "speculative behavior" in the sense of Harrison and Kreps (1978) arises. When traders do not share the same information sets it can be rational for individual traders to believe that other traders currently entertain incorrect expectations regarding future short rates. Traders can then take advantage of what they perceive to be current market misperceptions about future short rates to predict and benefit from excess returns, that is, expected returns above what holding a sequence of short term bonds would yield.

In this paper we present a model populated with traders that engage in the type of speculative behavior described above. The model is used to show the following. If traders' information sets are non-nested, (i) individual traders can systematically predict excess returns even in a model with a constant price of risk. (ii) An econometrician estimating a popular, but if information sets are non-nested, fundamentally misspecified affine three factor no-arbitrage model will find overwhelming but misleading statistical evidence in favor of time varying price of risk.

Date: March 25, 2009. The author thanks David Backus, Francisco Barillas, Thomas Sargent and seminar participants at New York University 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 .

A necessary but not sufficient 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. Three factor no-arbitrage models have proven to be very good statistical models of the term structure of interest rates (e.g. Duffie and Kan 1996 and Duffee 2002). Implicitly, these models assume that the factors are observable by traders, which is often justified by making the factors an invertible function of yields. In a few recent and 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 (1983) 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] Apart from 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 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

[1]What 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 will be used later in this paper.

percentile of individual forecasts of the 3-month Treasury Bill rate 4 quarters ahead fluctuates between 80 and 220 basis points.

The empirical model presented below will display 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 explain the evidence in Duffee (2008) by computing Duffee's impulse responses estimated on artificial data 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 and a constant price of risk 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 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

[2] See for instance Townsend (1983), Sargent (1991) and Makarov and Rytchov (2009).

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 a 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 paper 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, and specifically the absence of a state dependent price of risk, will help highlighting 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

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

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 $\widetilde{b}_t^1$ so that

$$\widetilde{b}_t^1 = b_t^n - c^n. \tag{2.2}$$

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

$$\widetilde{b}_t^1 = -r_t \tag{2.3}$$

and apply (5.9) recursively. The log price of a two period bond then is

$$\widetilde{b}_t^2 = -r_t - \int E\left[r_{t+1} \mid \Omega_t(j)\right] dj + \eta_t^2 \tag{2.4}$$

The price of a three period bond according to (5.9) 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 (5.9) with $n = 3$ gives

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

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 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 $t+2$.

[3]Alternatively, and without any substantial implications for the model, η could also be viewed as measurement errors in model without random bond supply (see Duffee (2008)).

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

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

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

$$y_t^n = -n^{-1}\widetilde{b}_t^n \tag{2.7}$$

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 risk premia are constant over time the expectations hypothesis hold. 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

$$y_t^n \equiv c^n + n^{-1}\sum_{s=0}^{n-1} E\left[r_{t+s} \mid \Omega_t\right] + \gamma_t^n \tag{3.1}$$

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 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

[4]See Backus and Wright (2007).

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*

$$EX^2 < \infty \tag{3.2}$$

and with inner-product

$$\langle X, Y \rangle \equiv E\left(XY\right) : X, Y \in L^2 \tag{3.3}$$

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*

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

for any $\omega \in \Omega$.

Orthogonal projections have the following properties that will be useful to us:

(1) The projection $\mathcal{P}_\Omega X$ coincides with the expectation $E\left[X \mid \Omega\right]$ 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.
$$X = \mathcal{P}_\Omega X + \left(I - \mathcal{P}_\Omega\right) X \tag{3.5}$$
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 $\mathcal{P}_{\Omega 1} X = \mathcal{P}_{\Omega 1} \mathcal{P}_{\Omega 2} X$ 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 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*

$$f_t^n \equiv \widetilde{b}_t^n - \widetilde{b}_t^{n+1} \tag{3.6}$$

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

$$f_t^n = \mathcal{P}_{\Omega t} r_{t+n} \quad \forall t, n \tag{3.7}$$

This formulation turns out to be more convenient than, though equivalent to, stating that the residuals γ_t^n in (3.1) are orthogonal to Ω_t, or that excess returns are not predictable. 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

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

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 $\Omega_0 \subseteq \Omega_1 \subseteq \Omega_2 ... \subseteq \Omega_t$. 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

$$\widetilde{b}_t^3 = -r_t - \mathcal{P}_{\Omega t} r_{t+1} - \mathcal{P}_{\Omega t} \mathcal{P}_{\Omega t+1} r_{t+2}. \tag{3.9}$$

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 $\{E\left[r_{t+n} \mid \Omega_{t+s}\right]\}_{s=0}^{n}$ is a martingale. Applying this result to the bond pricing equation (3.8) gives

$$\widetilde{b}_t^3 = -r_t - \mathcal{P}_{\Omega t} r_{t+1} - \mathcal{P}_{\Omega t} r_{t+2} \tag{3.10}$$

or equivalently

$$y_t^3 = \frac{1}{3} \sum_{s=0}^{2} E\left[r_{t+s} \mid \Omega_t\right] \tag{3.11}$$

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

$$\begin{aligned} f_t^2 &= \widetilde{b}_t^2 - \widetilde{b}_t^3 && (3.12)\\ &= r_t + \mathcal{P}_{\Omega t} r_{t+1} && (3.13)\\ &\quad -r_t - \mathcal{P}_{\Omega t} r_{t+1} - \mathcal{P}_{\Omega t}\mathcal{P}_{\Omega t+1} r_{t+2} && (3.14)\\ &= \mathcal{P}_{\Omega t} r_{t+2} && (3.15)\end{aligned}$$

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 excess returns

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 the following statements formally: Individual traders can predict and take advantage of deviations from the expectations hypothesis conditional on his information set. In spite of individual traders being able to systematically make better predictions than implied forward rates, a conditioning down type argument similar to but weaker than that made by Hansen and Sargent (1991) still holds: An econometrician using only public information would find only trivial deviations from the expectations hypothesis due to exogenous supply shocks.

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 if*

$$\mathcal{P}_{t,j} r_{t+n} = \mathcal{P}_{t+s,i} r_{t+n} \tag{4.1}$$

with probability zero for all j, i, t and $s < n$.

Defining non-nested information sets by the implications for projections is more general than basing a definition on the properties of signals. It could be that signals differ across traders, but that a common subset of the signals are sufficient to optimally predict interest rates. The definition above gets around this issue by defining information sets as non-nested only if they imply that expectations about future short rates will differ across agents.[5]

With non-nested information sets, we thus cannot simply replace the average expectation in period t of the average expectation in period $t+1$ of the short rate in period $t+2$ with a projection onto any individual trader's information set. All traders will entertain

[5]Sargent (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.

different expectations, and they cannot all coincide with the forward rate. The latter is straightforward to show more formally.

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*

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

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

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

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. □

That forward rates differ from individual traders' projections of future short rates imply that traders can systematically predict excess returns. A trader j who believes that the short rate at period $t+n$ will be higher than the current forward rate f_t^n, that is $\mathcal{P}_{t,j} r_{t+n} > f_t^n$, can make a profit (in expectations) by shorting bonds. By optimality of projections, $\mathcal{P}_{t,j} r_{t+n}$ will on average be a more accurate prediction of r_{t+n} than f_t^n so on average other traders will revise their expectations of r_{t+n} towards $\mathcal{P}_{t,j} r_{t+n}$ in period $t+1$. For the example with $\mathcal{P}_{t,j} r_{t+n} > f_t^n$, this will cause bond prices to fall and make the shorting strategy profitable. (A symmetric argument explains why going long in bonds is profitable when $\mathcal{P}_{t,j} r_{t+n} < f_t^n$.)

It is also straightforward to show that with non-nested information sets individual traders can predict average predictions errors.

Proposition 2. *Average period t+s projection errors of the short rate in period t+n*

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

will be orthogonal to $\Omega_t(j)$ so that

$$\mathcal{P}_{t,j} \left(r_{t+n} - \int \mathcal{P}_{t+s,j'} r_{t+n} dj' \right) = 0 \tag{4.4}$$

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

Proof. First note that the expression (4.4) can be rearranged to

$$\mathcal{P}_{t,j} r_{t+n} = \mathcal{P}_{t,j} \int \mathcal{P}_{t+s,j'} r_{t+n} dj'. \tag{4.5}$$

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

$$\mathcal{P}_{t,j}\int \mathcal{P}_{t+s,j'}r_{t+n}dj' = \mathcal{P}_{t,j}\mathcal{P}_{t+s,i}r_{t+n} \text{ for all } i,j\in(0,1): i\neq j. \tag{4.6}$$

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

$$\mathcal{P}_{t,j}r_{t+n} = \mathcal{P}_{t,j}\mathcal{P}_{t+s,i}r_{t+n} \tag{4.7}$$

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

Interestingly, and in spite of bond prices being determined by the actions of individual traders who systematically benefit from predictable excess returns, we can now prove the following: An econometrician using only public information would only be able to reject the expectations hypothesis trivially due to the exogenous supply shocks. That is, the yield dynamics that are due to non-nested information sets are orthogonal to public information.

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 3. *The forward rate f_t^n can be decomposed into the projection of r_{t+n} onto Ω_t^p, terms that are orthogonal to Ω_t^p and exogenojus supply shocks η_t^n and η_t^{n+1}.*

Proof. For convenience, first define the notation

$$\prod_{s=0}^{n-1}\int \mathcal{P}_{t+s,j}r_{t+n} \equiv \int \mathcal{P}_{t,j}\int \mathcal{P}_{t+1,j'}...\int \mathcal{P}_{t+n-1,j''}r_{t+n}dj''...dj'dj. \tag{4.8}$$

and rewrite the definition of the n period forward rate as

$$f_t^n = \prod_{s=0}^{n-1}\int \mathcal{P}_{t+s,j}r_{t+n} + \left(\eta_t^n - \eta_t^{n+1}\right). \tag{4.9}$$

Add and subtract $\prod_{s=0}^{n-2}\int \mathcal{P}_{t+s,j}r_{t+n}$ from the r.h.s. of (4.9) to get

$$\begin{aligned} f_t^n \;=\; & \prod_{s=0}^{n-2}\int \mathcal{P}_{t+s,j}r_{t+n} \\ & -\prod_{s=0}^{n-2}\int \mathcal{P}_{t+s,j}\left(r_{t+n} - \int \mathcal{P}_{t+n-1,j'''}r_{t+n}dj'''\right) \\ & +\left(\eta_t^n - \eta_t^{n+1}\right) \end{aligned} \tag{4.10}$$

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 second line of (4.10) as

$$\begin{aligned} & \prod_{s=0}^{n-2}\int \mathcal{P}_{t+s,j}\left(r_{t+n}-\int \mathcal{P}_{t+n-1,j'''}r_{t+n}dj'''\right) \\ = & \prod_{s=0}^{n-2}\int \mathcal{P}_{t+s,j}^{\perp p}\left(r_{t+n}-\int \mathcal{P}_{t+n-1,j'''}r_{t+n}dj'''\right) \\ & +\prod_{s=0}^{n-2}\int \mathcal{P}_{t+s}^{p}\left(r_{t+n}-\int \mathcal{P}_{t+n-1,j'''}r_{t+n}dj'''\right) \end{aligned} \tag{4.11}$$

where the second term on the r.h.s. is zero since the prediction errors in brackets are orthogonal to Ω_t^p since $\Omega_t^p \subseteq \Omega_t(j)$ for all j. Substituting this back into (4.10) gives

$$\begin{aligned} f_t^n = & \prod_{s=0}^{n-2}\int \mathcal{P}_{t+s,j}r_{t+n} \\ & -\prod_{s=0}^{n-2}\int \mathcal{P}_{t+s,j}^{\perp p}\left(r_{t+n}-\int \mathcal{P}_{t+n-1,j'''}r_{t+n}dj'''\right) \\ & +\left(\eta_t^n-\eta_t^{n+1}\right) \end{aligned} \tag{4.12}$$

Repeat this procedure by adding and subtracting $\prod_{s=0}^{n-m}\int \mathcal{P}_{t+s,j}r_{t+n}$ for $m = 3, 4, ...n$ until the forward rate is expressed as a sum of the average predicted short rate in period $t+n$ and a sum of average projections of higher order prediction errors onto the orthogonal complement to the public information in trader's information sets (and supply shocks)

$$\begin{aligned} f_t^n = & \int \mathcal{P}_{t,j}r_{t+n} \\ & +\sum_{m=1}^{n}\prod_{s=0}^{n-m}\int \mathcal{P}_{t+s,j}^{\perp p}\left(r_{t+n}-\int \mathcal{P}_{t+m-1,j}r_{t+n}\right) \\ & +\left(\eta_t^n-\eta_t^{n+1}\right) \end{aligned} \tag{4.13}$$

Finally, decompose $\int \mathcal{P}_{t,j}r_{t+n}$ into $\mathcal{P}_t^p r_{t+n} + \int \mathcal{P}_{t,j}^{\perp p}r_{t+n}$

$$\begin{aligned} f_t^n = & \mathcal{P}_t^p r_{t+n}+\int \mathcal{P}_{t,j}^{\perp p}r_{t+n} \\ & +\sum_{m=1}^{n}\prod_{s=0}^{n-m}\int \mathcal{P}_{t+s,j}^{\perp p}\left(r_{t+n}-\int \mathcal{P}_{t+m-1,j}r_{t+n}\right) \\ & +\left(\eta_t^n-\eta_t^{n+1}\right) \end{aligned} \tag{4.14}$$

so that all terms on the right hand side are orthogonal to Ω_t^p except $\mathcal{P}_t^p r_{t+n}$, which concludes the proof. □

The proof shows that even though equilibrium bond prices are affected by traders exploiting what they perceive to be deviations from the expectations hypothesis, these deviations are orthogonal to public information. 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 is thus orthogonal to public information. The additional dynamics caused by speculative behavior in our model must therefore also 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 alterative 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, a conditioning down argument similar to, but weaker than, 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

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

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

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

$$\varepsilon_t \sim N(0, I) \tag{5.3}$$

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 next period wealth W_{t+1} according to the CARA utility function

$$U_t(j) = E_t\left[e^{-\gamma(W_{t+1}(j))} \mid \Omega_t(j)\right] \tag{5.4}$$

and subject to the law of motion for wealth given by

$$W_{t+1}(j) = \mathbf{b}_{t+1}^{(-1)\prime}\mathbf{d}_t(j) + \left[W_t(j) - \mathbf{b}_t'\mathbf{d}_t(j)\right](1 + r_t) \tag{5.5}$$

where $\mathbf{b}_t$ is a vector of prices of bonds of different maturities and $\mathbf{b}_{t+1}^{(-1)}$ is a vector of prices of the same bonds in the next period (when they have one period less to maturity). The vector $\mathbf{d}_t(j)$ is trader j's holdings of bonds of corresponding maturities. Maximizing the utility function (5.4) subject to the law of motion of wealth (5.5) gives an expression for trader j's demand for bonds $\mathbf{d}_t(j)$ that we can approximate log-linearly as

$$\widetilde{\mathbf{d}}_t(j) \approx E\left[\widetilde{\mathbf{b}}_{t+1}^{(-1)} \mid \Omega_t(j)\right] - r_t - \widetilde{\mathbf{b}}_t \tag{5.6}$$

(The expression for trader j's demand (5.6) is derived in the Appendix.) The vector of bond supply $\mathbf{s}_t$ is an exogenous process that in logs is Gaussian white noise

$$\mathbf{s}_t = e^{\eta_t} \tag{5.7}$$
$$\eta_t \sim N\left(\mathbf{0}, \mathbf{\Sigma}_\eta\right) \tag{5.8}$$

where Σ_η is assumed to be diagonal. The supply shocks η_t are uncorrelated with all other shocks in the model.

Equating aggregate log demand and supply gives the log price $\widetilde{b}_t^n$ for an n periods to maturity zero coupon bond

$$\widetilde{b}_t^n = -r_t + \int E\left[\widetilde{b}_{t+1}^{n-1} \mid \Omega_t(j)\right] dj - \eta_t^n \tag{5.9}$$

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 (5.9) 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

$$y_t^n = -n^{-1}\widetilde{b}_t^n \tag{5.10}$$

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 the empirical exercise below where I use monthly data, the selected yields are the 3, 12, 24, 36, 48 and 60 month interest rates.) 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 $\zeta_t(j)$

$$s_t(j) = x_t^1 + \zeta(j) \tag{5.11}$$
$$\zeta_t(j) \sim N\left(0, \sigma_\zeta^2\right) \tag{5.12}$$

The vector $S_t(j)$ defined as

$$S_t(j) = \left[\begin{array}{ccc} s_t(j) & r_t & \mathbf{y}_t' \end{array}\right]' \tag{5.13}$$

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

$$X_t = MX_{t-1} + Nu_t \tag{5.14}$$
$$\mathbf{y}_t = D_1X_t + D_2r_{t-1} + D_3\eta_t \tag{5.15}$$

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 $\mathbf{x}_t$ $\left(\equiv \begin{bmatrix} x_t^1 & x_t^2 \end{bmatrix}'\right)$

$$X_t \equiv \begin{bmatrix} \mathbf{x}_{t|t}^{(0)\prime} & \mathbf{x}_{t|t}^{(1)\prime} & \cdots & \mathbf{x}_{t|t}^{(\overline{k})\prime} \end{bmatrix}' \tag{5.16}$$

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 $\overline{k}$ is the maximum order of expectation considered and can be chosen to achieve an arbitrarily close approximation to the limit as $k \to \infty$. 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 1.

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}	$5.4\times10^{-5}- 4.5\times10^{-3}$
	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}	$9.1\times10^{-5}- 2.1\times10^{-2}$
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 Random Walk Metropolis-Hastings algorithm.

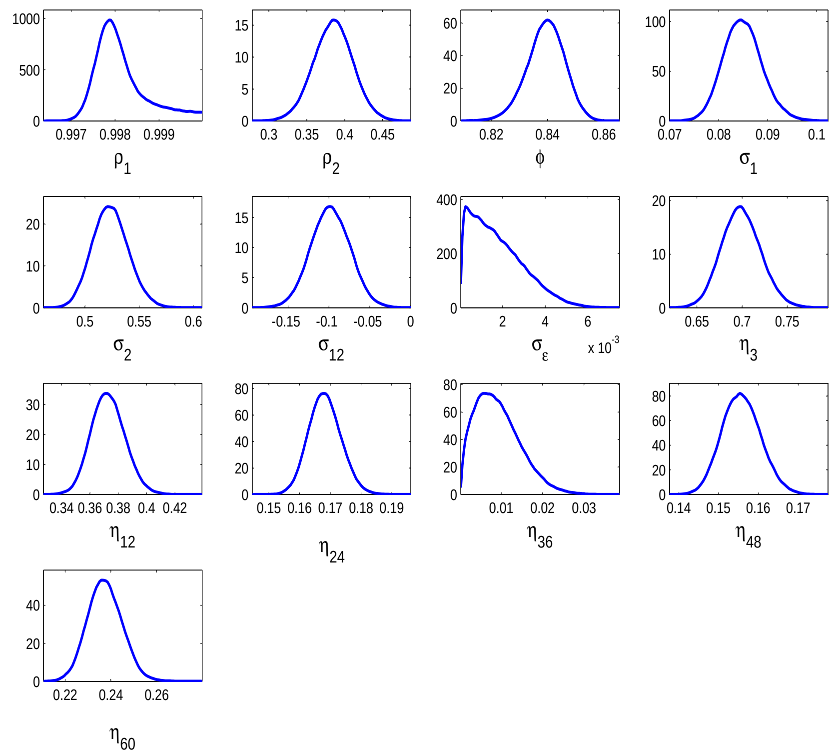

FIGURE 1. Estimated posterior densities of structural parameters generated using 1 000 000 draws from an Adaptive Random Walk Metropolis-Hastings algorithm.

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.

[TBC]

5.5. **The estimated dispersion of expectations. [TBC]**

[6]As a convergence check, Figure 5 in the Appendix shows recursive plots of the diagonal of the covariance matrix of the MCMC of the posterior.

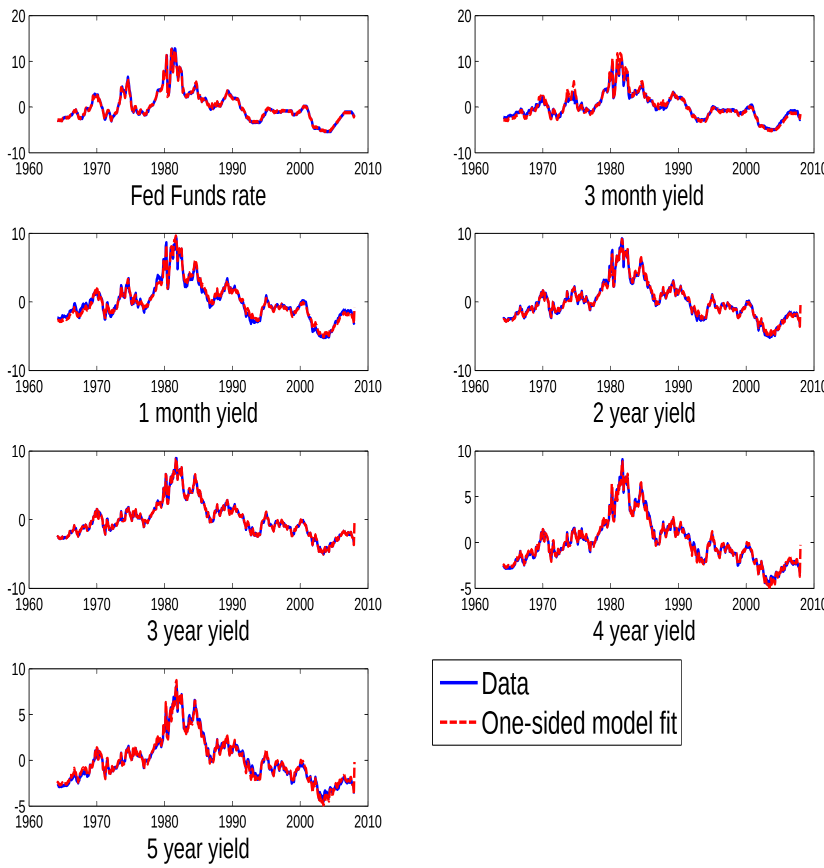

Figure 2. Data (blue solid) and one-sided model fit (red dashed).

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.

6.1. **A three factor no-arbitrage model.** Affine 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 mispecified.[7] It is

[7]Perhaps 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.

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. $\overline{k} = 6$). 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

$$f_t = \phi f_{t-1} + \Sigma e_t \tag{6.1}$$

where ϕ is diagonal and Σ 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

$$r_t = \delta' f_{t-1} \tag{6.2}$$

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

$$\widetilde{b}_t^n = B_n' f_t \tag{6.3}$$

where

$$B_n' = B_{n-1}' (\phi - \Sigma\lambda) - \delta' \tag{6.4}$$

As before, yields can be computed as

$$y_t^n = -n^{-1} \widetilde{b}_t^n \tag{6.5}$$

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 is repeated 260 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

[8] The procedure is very time consuming and the 260 repetitions took about 5 days to compute on a not unusually slow computer.

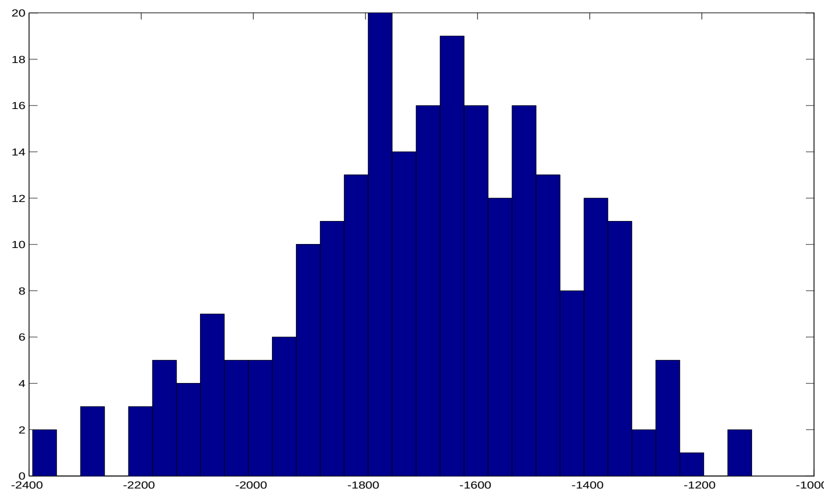

FIGURE 3. Distribution of log likelihood ratios of restricted and unrestricted three factor yield model.

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.

6.2. **The Duffee Hidden Factor.** 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

$$x_t^\dagger = D^\dagger x_{t-1}^\dagger + \Sigma^\dagger \epsilon_t \tag{6.6}$$
$$y_t = A + B^\dagger x_t^\dagger + \eta_t^\dagger \tag{6.7}$$

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.

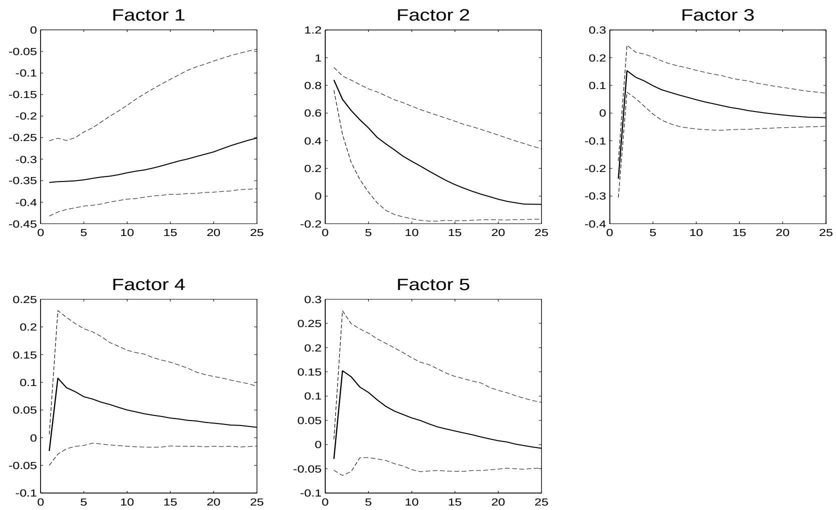

FIGURE 4. Impulse responses to orthogonal factor innovations, dashed lines are 5th and 95th percentile.

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.

[TBC]

6.3. **The Cochrane and Piazzesi Factor.** [TBC]

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 wether 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

[9]The 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.

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] 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.

[2] 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.

[3] Backus, D. Foresi and C. Telmer, (1998), Discrete Time Bond Pricing, unpublished manuscript, New York University.

[4] 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.

[5] Backus, D. and J. Wright, (2007), "Cracking the conundrum", NBER working paper13419.

[6] 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.

[7] Brockwell, P.J. and R.A. Davis, (2006), *Time Series: Theory and Methods*, Springer-Verlag.

[8] Cochrane, J. and M. Piazzesi, (2005), "Bond Risk Premia", *American Economic Review.*

[9] Cochrane, J. and M. Piazzesi, (2008), "Decomposing the Yield Curve", mimeo University of Chicago.

[10] Duffee, Gregory (2002), "Term Premia and Interest Rate Forecasts in Affine Models," *Journal of Finance*, 57, pp.405-443.

[11] Duffie, Darrell and Rui Kan (1996), "Yield Factor Models of Interest Rates," *Mathematical Finance*, 64, pp.379-406.

[12] Duffee, Greg, (2008), Infomration in (and not in) the term structure of interest rates, mimeo, Johns Hopkins University.

[13] Evans, G. W. and S. Honkapohja, (2005). "An Interview With Thomas J. Sargent," *Macroeconomic Dynamics*, vol. 9(04), pp 561-583.

[14] Haario, H, E. Saksman and J. Tamminen, 2001, "An Adaptive Metropolis Algorithm", *Bernoulli*, 7(2), 223-242.

[15] Hansen, L.P. and T.J. Sargent, 1991, Rational Expectations Econometrics, Westview Press.

[16] Harrison, J.M. and D. Kreps, (1978), "Speculative Behaviour in a Stock Market with Heterogenous Expectations", *Quarterly Journal of Economics*, pp323-336.

[17] Kasa, Kenneth, 2000, "Forecasting the forecast of others in the frequency domain", *Review of Economic Dynamics,* 3, pp726-756.

[18] Litterman, R., Scheinkman, J., (1991), "Common factors affecting bond returns", Journal of Fixed Income 1, 51–61.

[19] Morris, S. and H.S. Shin, (2002), "The social value of public information", *American Economic Review* 92, pp1521-1534.
[20] Nimark, K., (2007), "Dynamic Higher Order Expectations", working paper, Universitat Pompeu Fabra.
[21] Pearlman, J.G. and T.J. Sargent, 2005, "Knowing the forecasts of others", *Review of Economic Dynamics*, Volume 8, pp480-497.
[22] Piazzesi, M. and M. Schneider, (2006), "Equilibrium Yield Curves", NBER Working Papers 12609.
[23] 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.
[24] Sargent, Thomas J., 1991, "Equilibrium with Signal Extraction from Endogenous Variables", *Journal of Economic Dynamics and Control* 15, pp245-273.
[25] Shiller, R. J.Y. Campbell and K.L. Shoenholtz, (1983), "Forward Rates and Future Policy: Interpeting the Term Structure of Interest Rates", *Brookings Papers on Economic Activity*, pp173-223.
[26] 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.
[27] Xiong, W. and H. Yan, (2008),"Heterogenous Expectations and Bond Markets", mimeo, Princeton University.

Appendix A. The term structure of interest rates

The central bank controls the short interest rate and is taken as exogenous in the model. The price of longer maturity bonds are determined by trader's demand and the supply of bonds which are exogenous.

A.1. **Bond demand.** Traders are indexed by j and maximize next period wealth according to the CARA utility function

$$U_t(j) = E_t \left[e^{-\gamma(W_{t+1}(j))} \mid \Omega_t(j) \right] \tag{A.1}$$

where W_{t+1} is next period wealth given by

$$W_{t+1}(j) = \mathbf{b}'_{t+1}\mathbf{d}_t(j) + \left[W_t(j) - \mathbf{b}'_t\mathbf{d}_t(j)\right](1 + r_t) \tag{A.2}$$

and $\mathbf{b}_t$ is a vector of prices of bonds of different maturities and $\mathbf{q}_t(j)$ is trader j's holdings of bonds of corresponding maturity. Taking logs of () yields

$$\log U_t(j) = -E\left[-\gamma \mathbf{d}'_t \left[\mathbf{b}_{t+1} - (1 + r_t)\,\mathbf{b}_t\right] + (1 + r_t)W_t(j) + \frac{\gamma^2}{2}\mathbf{d}'_t\Sigma\mathbf{d}_t \mid \Omega_t(j)\right] \tag{A.3}$$

where $\mathbf{\Sigma}$ is the conditional covariance matrix of bond prices, that is

$$\Sigma = E\left[\left(E\left[\mathbf{b}_{t+1} \mid \Omega_t(j)\right] - \mathbf{b}_{t+1}\right)\left(E\left[\mathbf{b}_{t+1} \mid \Omega_t(j)\right] - \mathbf{b}_{t+1}\right)'\right] \tag{A.4}$$

The first order conditions w.r.t. bond demand $\mathbf{d}_t$ then is

$$\gamma\left[E\left[\mathbf{b}_{t+1} \mid \Omega_t(j)\right] - (1 + r_t)\,\mathbf{b}_t\right] - \gamma^2\Sigma\mathbf{d}_t = \mathbf{0} \tag{A.5}$$

or

$$\mathbf{d}_t(j) = \frac{1}{\gamma}\Sigma^{-1}\left[E\left[\mathbf{b}_{t+1} \mid \Omega_t(j)\right] - (1 + r_t)\,\mathbf{b}_t\right] \tag{A.6}$$

Taking log approximation

$$\widetilde{\mathbf{d}}_t(j) \approx E\left[\widetilde{\mathbf{b}}_{t+1} \mid \Omega_t(j)\right] - r_t - \widetilde{\mathbf{b}}_t \tag{A.7}$$

A.1.1. *Bond supply.* Bond supply $\mathbf{s}_t$ is (in logs) an exogenous white noise process

$$\mathbf{s}_t = e^{\eta_t} \tag{A.8}$$
$$\eta_t \sim N\left(\mathbf{0}, \mathbf{\Sigma}_\eta\right) \tag{A.9}$$

A.1.2. *Equilibrium prices.* Equating log demand and supply

$$\eta_t = \int E\left[\widetilde{\mathbf{b}}_{t+1} \mid \Omega_t(j)\right] dj - \mathbf{r}_t + \widetilde{\mathbf{b}}_t \tag{A.10}$$

gives log bond prices for an n periods to maturity zero coupon bond $\widetilde{b}_t^n$

$$\widetilde{b}_t^n = -r_t + \int E\left[\widetilde{b}_{t+1}^{n-1} \mid \Omega_t(j)\right] dj - \eta_t^n \tag{A.11}$$

Appendix B. 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 z_t and v_t. Define the exogenous state vector $\mathbf{x}_t$ as

$$\mathbf{x}_t = \left[\begin{array}{cc} x_t^1 & x_t^2 \end{array}\right]' \tag{B.1}$$

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

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

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

$$X_t \equiv \begin{bmatrix} \mathbf{x}_t^{(0)} \\ \mathbf{x}_t^{(1)} \\ \vdots \\ \mathbf{x}_t^{(\overline{k})} \end{bmatrix} \tag{B.3}$$

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

$$X_t = MX_{t-1} + N\mathbf{e}_t \tag{B.4}$$

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

$$Q = \left[\begin{array}{cc} \mathbf{0}_{\overline{k}\times 2} & \begin{array}{c} I_{\overline{k}-2} \\ \mathbf{0}_{2\times\left(\overline{k}-2\right)} \end{array} \end{array}\right] \tag{B.5}$$

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

$$\overline{M} = MQ \tag{B.6}$$

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

$$b_t^n = \begin{bmatrix} \mathbf{1}_{1\times 2} & \mathbf{0} \end{bmatrix} \overline{M}^{n-1} + \phi b_t^{n-1} - \sum_{j=1}^{n} \lambda^j r_{t-1} \tag{B.7}$$

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

$$y_t^n \equiv -n^{-1} b_t^n \tag{B.8}$$

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

$$\begin{aligned} S_t(j) &= \begin{bmatrix} s_t(j) & r_t & \mathbf{y}_t' \end{bmatrix}' & \text{(B.9)} \\ &= L_1 X_t + L_2 \mathbf{u}_t & \text{(B.10)} \end{aligned}$$

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

$$\mathbf{u}_t = \begin{bmatrix} \eta_t & \varepsilon_t & \eta_t \end{bmatrix}' \tag{B.11}$$

Agent j's updating equation of his state estimate will then follow

$$X_{t|t} = MX_{t-1} + K\left(S_t(j) - L_1 M X_{t-1}\right) \tag{B.12}$$

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 $\mathbf{x}_{t|t}^{(0:\overline{k})}$

$$\begin{aligned} M &= \begin{bmatrix} \rho & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + \begin{bmatrix} \mathbf{0}_{2\times 2} & \mathbf{0} \\ \mathbf{0} & [M - KLM]_{11} \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ [KLM]_{11} \end{bmatrix} & \text{(B.13)} \\ N &= \begin{bmatrix} \mathbf{1}_{1\times 2} & 0 \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ [KLN]_{11} \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & K_3 \end{bmatrix} & \text{(B.14)} \end{aligned}$$

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

$$\begin{aligned} K &= \left(PL_1' + G\right)\left(LPL' + L_2 \Sigma_{uu} L_2'\right)^{-1} & \text{(B.15)} \\ P &= W\left(P - \left(PL_1' + G\right)\left(L_1 P L_1' + L_2 \Sigma_{uu} L_2'\right)^{-1}\left(PL_1' + G\right)'\right)W' + \Sigma_{ee} & \text{(B.16)} \end{aligned}$$

where

$$G = E\left(\mathbf{e}_t \mathbf{u}_t' L_2'\right) \tag{B.17}$$

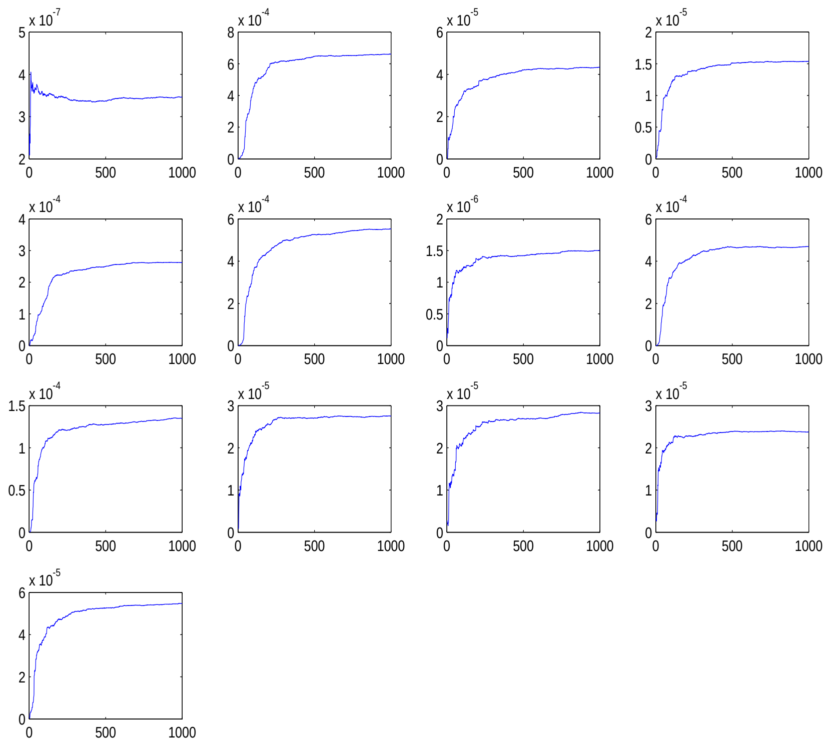

FIGURE 5. Convergence of MCMC: Recursive plots of diagonal from covariance matrix of MCMC