SPECULATIVE DYNAMICS IN THE TERM STRUCTURE OF INTEREST RATES

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. It is demonstrated that (i) dispersion of expectations about future short rates is sufficient for individual traders to systematically predict excess returns even in a model with constant risk premia and (ii) the new term structure dynamics driven by speculative trade is orthogonal to public information in real time. The model is estimated using monthly data on US short to medium term Treasuries from 1964 to 2007 and it provides a very good fit of the data, comparing favorably to a standard affine three factor no-arbitrage model. Speculative dynamics are also found to be quantitatively important, accounting for a substantial fraction of the variation of bond yields and is more important at long maturities.

Keywords: Term structure of interest rates; Speculative dynamics; Excess returns; Non-nested information; Private information

1. INTRODUCTION

When 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 rational 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 empirically relevant new dynamics to the term structure of interest rates. More specifically, we show that (i) individual traders can systematically predict excess returns (defined as the difference in return on holding an n period bond until it matures and holding a sequence of short bonds for n periods) even in a model with constant risk premia. (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.

Despite the fact that the speculative dynamics are orthogonal to public information, we can both quantify their importance as well as back out an estimated historical time series of their evolution in the past. This is possible since we as econometricians have access to the full sample of data and the speculative term is orthogonal only to public information available to traders in real time. That is, we can use public information available in period $t+s : s > 0$ to back out an estimate of the speculative term in period t. The estimated model suggests that speculative dynamics are quantitatively important and explains a substantial fraction of the variation in observed US bond yields.

A necessary condition for traders to have any relevant 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 optimally 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 linear combinations of current bond yields are not sufficient to predict future bond yields optimally.

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.

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. 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 which provides us with another way of gauging the plausibility of this assumption. 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.

There is a growing literature analyzing asset pricing models under non-nested information sets (or one of its synonyms, see Footnote 1). Singleton (1987) is an early reference of a dynamic asset pricing model with an information structure similar to the one presented here. More recent examples include papers by Allen, Morris and Shin (2005), Kasa, Walker and Whiteman (2008), Bacchetta and van Wincoop (2005, 2007), Cespa and Vives (2009) and Makarov and Rytchkov (2009). 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 uniform priors truncated only to ensure that the model is stationary and that variances are non-negative. To the best of my knowledge, this is the first paper to estimate a model with non-nested information sets. The fit of the model is surprisingly good and compares favorably to the fit of a standard affine three factor no-arbitrage model. The model is also used to quantify the dispersion of expectations implied by the posterior parameter estimates. A small implied cross sectional dispersion of short rate expectations is found to be sufficient to generate quantitatively important speculative dynamics.

Using the estimated model to generate artificial data, we further show that an econometrician estimating an affine three factor no-arbitrage model would find overwhelming but misleading statistical evidence in favor of time varying risk premia if the model with non-nested information sets was the true data generating process. The estimated model also displays similar dynamics to those documented by Duffee (2008) and Cochrane and Piazzesi (2005, 2008). Factors that play practically no 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

---

[1]What I in this paper call *non-nested information sets* is also known as *disparately informed traders* (Singleton 1987), *private information* (Sargent 1991), *heterogenous information* (Bacchetta and van Wincoop 2006), *dispersed information* (Angeletos and Pavan 2009) and *imperfect common knowledge* (Woodford 2002, Adam 2006 and Nimark 2008). The term *non-nested information* connects naturally to the language of orthogonal projections used in this paper.

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.

The next section presents a simple bond pricing model that allows for traders to have non-nested information sets. Section 3 uses the properties of orthogonal projections to derive the main theoretical results and Section 4 contains the estimated model and discusses the quantitative importance of speculative dynamics. Section 5 relates the model to a popular class of alternative models. Section 6 concludes and the Appendix contains details of how the model was solved and lists some of the properties of orthogonal projections used in Section 3.

2. A BOND PRICING MODEL

In this section we present a simple bond pricing model. The 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, and a stochastic supply shock. In the absence of private information, the model implies that the expectations hypothesis holds and 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.

2.1. **Demand for long maturity bonds.** As in Allen, Morris and Shin (2006) there are overlapping generations of agents who each live for two periods. Time is discrete and indexed by t. Each generation consist of a continuum of households with unit mass. Each household are endowed with one unit of wealth that they invest when young. When old, households unwind their asset positions and consume. Unlike in the model of Allen *et al*, the owners of wealth, i.e. the households, do not trade assets themselves. Instead, a continuum of traders, indexed by $j \in (0,1)$, trade on behalf of the households, with households diversifying their available funds across the continuum of traders. While not modeled explicitly here, this set up can be motivated as a perfectly competitive limit case of the monopolistically competitive mutual funds model of Garcia and Vanden (2009) that allow uninformed households to benefit from mutual funds private information, while diversifying away idiosyncratic risk associated with individual funds.

Trader j invests one unit of wealth in period t on behalf of all households born in period t. In period $t+1$ trader j unwinds the position of the now old generation of households who then use the proceeds to consume. Traders are infinitely lived and perform the same service for the next generation of households. The set up with short lived investors but infinitely lived traders implies that private information is long-lived while abstracting from wealth dynamics.[2]

There are two types of assets. A risk free one period bond with (log) return r_t and risky zero-coupon bonds of various maturities. Trader j chooses a vector of portfolio weights $\mathbf{a}_t(j)$ in order to maximize the discounted expected log of wealth under management $W_{t+1}(j)$ in

---

[2]Readers interested in the interaction between wealth dynamics and heterogenous beliefs can consult Xiong and Yan (2009) who analyze this issue in a model with boundedly rational traders.

period $t + 1$. That is, trader j solves the problem

$$\max_{\mathbf{a}_t(j)} E\left[\log W_{t+1}(j) \mid \Omega_t(j)\right] \tag{2.1}$$

subject to

$$W_{t+1}(j) = 1 + r_t^p(j)$$

where $\Omega_t(j)$ denotes trader j's information set and $r_t^p(j)$ is the log return of the portfolio chosen by trader j in period t. In the model presented below, equilibrium log returns of individual bonds are normally distributed. However, log returns of a portfolio of assets with individually log normal returns are not normally distributed. Following Campbell and Viceira (2002a, 2002b) we therefore use a second order Taylor expansion to approximate the log excess portfolio return as

$$r_t^p(j) - r_t = \mathbf{a}_t'(j)\left(\mathbf{b}_{t+1}^- - \mathbf{b}_t - \mathbf{r}_t\right) + \frac{1}{2}\mathbf{a}_t'(j) diag\left[\Sigma_{b,t}(j)\right] - \frac{1}{2}\mathbf{a}_t'(j)\Sigma_{b,t}(j)\mathbf{a}_t(j) \tag{2.2}$$

where $\mathbf{b}_t$ is a vector of (log) bond prices for bonds of different maturities. The vector $\mathbf{b}_{t+1}^-$ contains the prices of the same bonds in the next period when they have one period less to maturity and $\mathbf{r}_t$ is a vector of the risk free rate r_t. The difference $\mathbf{b}_{t+1}^- - \mathbf{b}_t - \mathbf{r}_t$ is thus a vector of excess returns on bonds of different maturities. The matrix $\Sigma_{b,t}(j)$ is the covariance of log bond returns

$$\Sigma_{b,t}(j) \equiv E\left[\left(\mathbf{b}_{t+1}^- - \mathbf{b}_t - \mathbf{r}_t\right)\left(\mathbf{b}_{t+1}^- - \mathbf{b}_t - \mathbf{r}_t\right)' \mid \Omega_t(j)\right] \tag{2.3}$$

conditional on trader j's information set. In the next section, we give general conditions for traders' information sets that are sufficient to generate speculative behavior. In the empirical part of the paper we will be explicit about the exact nature of traders's information sets. At this stage, it is sufficient to state that we will restrict our attention to specifications such that in equilibrium, conditional returns are normally distributed, time invariant and with a common variance across all traders

$$E\left[\left(\mathbf{b}_{t+1}^- - \mathbf{b}_t - \mathbf{r}_t\right) \mid \Omega_t(j)\right] \sim N\left(\mu_{\mathbf{b}}, \Sigma_b\right) \forall j, t \tag{2.4}$$

We will also maintain the assumption throughout the paper that current bond prices $\mathbf{b}_t$ and the current short rate r_t are observed perfectly by all traders so that

$$\{\mathbf{b}_t, r_t\} \in \Omega_t(j) \forall j \tag{2.5}$$

This assumption is introduced already here, in order to simplify notation in the derivation of the term structure in Section 2.4. Under these assumptions, maximizing the (second order Taylor approximation) of expected return (2.2) on wealth under management by trader j with respect to $\mathbf{a}_t(j)$ gives the optimal portfolio weights

$$\mathbf{a}_t(j) = \Sigma_b^{-1}\left(E\left[\mathbf{b}_{t+1}^- \mid \Omega_t(j)\right] - \mathbf{b}_t - \mathbf{r}_t\right) + \frac{1}{2}\Sigma_b^{-1} diag\left[\Sigma_b\right] \tag{2.6}$$

Since each trader j has one unit of wealth to invest, taking average across traders of the portfolio weights (2.6) yields the aggregate demand for bonds.

2.2. **Bond supply.** The vector of bond supply $\mathbf{s}_t$ is stochastic

$$\mathbf{s}_t = \Sigma_b^{-1}\mathbf{v}_t : \mathbf{v}_t \sim N\left(0, \Sigma_{\mathbf{v}}\right) \tag{2.7}$$

where to simplify notation, the vector of supply shocks $\mathbf{v}_t$ are normalized by the inverse of the conditional variance of bond prices Σ_b^{-1}. The supply shocks $\mathbf{v}_t$ play a similar role here as the noise traders in Admati (1985). That is, they will prevent equilibrium prices from fully revealing the information held by other traders.

2.3. **Equilibrium bond prices.** Equating aggregate demand $\int \mathbf{a}_t(j)dj$ and supply $\mathbf{s}_t$

$$\Sigma_b^{-1}\mathbf{v}_t = \Sigma_b^{-1}\left(\int E\left[\mathbf{b}_{t+1}^- \mid \Omega_t(j)\right]dj - \mathbf{b}_t - \mathbf{r}_t\right) + \frac{1}{2}\Sigma_b^{-1}diag\left[\Sigma_b\right] \tag{2.8}$$

and solving for the vector of current log bond prices $\mathbf{b}_t$ gives

$$\mathbf{b}_t = \frac{1}{2}diag\left[\Sigma_b\right] - \mathbf{r}_t + \int E\left[\mathbf{b}_{t+1}^- \mid \Omega_t(j)\right]dj - \mathbf{v}_t \tag{2.9}$$

A generic element of $\mathbf{b}_t$ is thus the log price b_t^n of an n periods to maturity zero coupon bond given by

$$b_t^n = \alpha^n - r_t + \int E\left[b_{t+1}^{n-1} \mid \Omega_t(j)\right]dj - v_t^n \tag{2.10}$$

where α^n and v_t^n are the relevant elements of $\frac{1}{2}diag\left[\Sigma_b\right]$ and $\mathbf{v}_t$ respectively. The price of an n periods to maturity bond in period t thus depends on 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.

2.4. **The term structure of interest rates.** The bond price formula (2.10) 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 timing of the information sets that the expectations that govern prices are conditioned on. As usual, we can start from

$$b_t^1 = -r_t \tag{2.11}$$

and apply (2.10) recursively. The log price of a two period bond then is

$$b_t^2 = \alpha^2 - r_t - \int E\left[r_{t+1} \mid \Omega_t(j)\right]dj + v_t^2 \tag{2.12}$$

The price of a three period bond according to (2.10) 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 (2.12) by one period and substituting into (2.10) with $n = 3$ gives

$$\begin{aligned} b_t^3 = {} & \alpha^2 + \alpha^3 - r_t - \int E\left[r_{t+1} \mid \Omega_t(j)\right]dj \\ & - \int E\left[\int E\left[r_{t+2} \mid \Omega_{t+1}(j')\right]dj' \mid \Omega_t(j)\right]dj \\ & + v_t^3 \end{aligned} \tag{2.13}$$

Applying the same procedure recursively to derive the price of an n periods to maturity bond gives

$$b_t^n = \sum_{i=2}^{n} \alpha^n - r_t - \int E\left[r_{t+1} \mid \Omega_t(j)\right] - \tag{2.14}$$

$$\int E\left[\int E\left[r_{t+2} \mid \Omega_{t+1}(j')\right] dj' \mid \Omega_t(j)\right] dj + \dots$$

$$\dots + \int E\left[\int E\left[\dots \int E\left[r_{t+n-1} \mid \Omega_{t+n-2}(j'')\right] dj'' \dots \mid \Omega_{t+1}(j')\right] dj' \mid \Omega_t(j)\right] dj + v_t^n$$

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

$$y_t^n = -n^{-1} b_t^n \tag{2.15}$$

As has been noted before (e.g. Allen, Morris and Shin (2006)), the fact that (2.14) contains average expectations of average expectations (and so on) prevents us from applying the law of iterated expectations to solve for bond prices if traders have non-nested information sets. Before analyzing the consequences of this, we first establish that in the absence of supply shocks and with only common information, the model does indeed imply that the expectation hypothesis hold. Below, the properties of orthogonal projections will be used extensively and we therefore first define orthogonal projections and the relevant inner product space.

2.5. **Deviations from means.** Since we are primarily interested in the effects of information on the dynamics of bond yields we will conduct most of the analysis in terms of deviations of the log price from its mean. We therefore define the log deviation $\widetilde{b}_t^n$ of the bond price as

$$\widetilde{b}_t^n \equiv b_t^n - \mu_b^n \tag{2.16}$$

where $\mu_b^n \equiv E\left[b_t^n\right]$. Similarly, define the deviation of the short interest rate r_t from its mean μ_r

$$\widetilde{r}_t \equiv r_t - \mu_r \tag{2.17}$$

where $\mu_r \equiv E\left[r_t\right]$. Subtracting μ_b^n and μ_r from (2.14) and substituting into the expression (2.15) of bond yields give the deviation of an n period bond yield from its mean

$$\widetilde{y}_t^n = -n^{-1} \widetilde{b}_t^n. \tag{2.18}$$

2.6. **Projections and a common information benchmark.** In the next section, the properties of orthogonal projections will be used to analyze the implications of non-nested information sets. However, we first apply some of these tools to the familiar case where all traders share the same information set. This section can thus be thought of as a benchmark, confirming that in the absence of private information the model implies that the expectation hypothesis hold. (For readers unfamiliar with orthogonal projections, the Appendix lists some of the properties that will be particularly useful below. For more details see Brockwell and Davis (2006).)

In the model presented below, all bond yields, the factors that drive them and the signals that traders observe will be elements of the inner product space L^2, which we now define.

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*

$$EX^2 < \infty \tag{2.19}$$

and with inner-product

$$\langle X, Y \rangle \equiv E\left(XY\right) : X, Y \in L^2 \tag{2.20}$$

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

$$\langle X - \mathcal{P}_\Omega X, \omega \rangle = 0 \tag{2.21}$$

for any $\omega \in \Omega$.

In a linear model with Gaussian shocks, conditional expectations are equivalent to orthogonal projections. We can thus use the equality

$$E\left(X \mid \Omega\right) = \mathcal{P}_\Omega X \tag{2.22}$$

to replace the conditional expectations in the bond pricing equation (2.14) with projections and rephrase the expectation hypothesis in the following way.

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*

$$f_t^n \equiv \widetilde{b}_t^n - \widetilde{b}_t^{n+1} \tag{2.23}$$

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

$$f_t^n = \mathcal{P}_{\Omega t}\widetilde{r}_{t+n} \quad \forall t, n \tag{2.24}$$

This is a standard definition of the expectation hypothesis (e.g. Backus, Foresi, Mozumdar and Wu, 2001), apart from the explicit reference to the information set expectations are conditioned on.

We can use the definition (2.24) to demonstrate that in the absence of supply shocks, the bond pricing equation (2.14) implies that the expectations hypothesis hold if all agents share the same information set. Consider the 2 period ahead forward rate which by (2.14) and (2.23) is given by

$$\begin{aligned} f_t^2 &= \widetilde{b}_t^3 - \widetilde{b}_t^2 \\ &= \int E\left[\int E\left[\widetilde{r}_{t+2} \mid \Omega_{t+1}(j')\right] dj' \mid \Omega_t(j)\right] dj \end{aligned} \tag{2.25}$$

For now, let $\Omega_t(j) = \Omega_t$ for all j, that is, let the information set Ω_t be common across all traders. If traders do not forget, the sequence of information sets $\{\Omega_t\}_{t=0}^\infty$ is a filtration so that $\Omega_0 \subseteq \Omega_1 \subseteq \Omega_2... \subseteq \Omega_t$. It is a property of projections that repeated projections onto nested information sets reduces to the projection onto the smallest information set, i.e. $\mathcal{P}_{\Omega t}\mathcal{P}_{\Omega t+s}X = \mathcal{P}_{\Omega t}X$ for $s \geq 0$. This is simply the law of iterated expectations and implies that traders cannot predict in period t how they will revise their expectation in period

$t + 1$ of the period $t + 2$ short rate. Stated differently, the sequence $\{E\left[\widetilde{r}_{t+n} \mid \Omega_{t+s}\right]\}_{s=0}^{n}$ is a martingale. Applying this result to the bond pricing equation (2.25) gives

$$
\begin{aligned}
f_t^2 &= \mathcal{P}_{\Omega t}\left(\mathcal{P}_{\Omega t+1}\widetilde{r}_{t+2}\right) & (2.26) \\
&= \mathcal{P}_{\Omega t}\widetilde{r}_{t+2} & (2.27)
\end{aligned}
$$

The expectation hypothesis then holds for a three period bond with respect to the common information set Ω_t. A similar argument can be generalized to an n-period bond.

In the next section we analyze how the dynamics of the term structure changes when traders have non-nested information sets. The direct link between forward rates and expectations about future short rates make it more convenient to frame the analysis in terms of forward rates rather than bond yields. Of course, bond yields can always be backed out from implied forward rates by the identity (2.23).

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

In the previous section, the fact that rational traders cannot predict the direction that they will revise their own expectations in the future allowed us to solve for bond prices as a function of the common period t expectation about future short rates. With non-nested information sets, predictions about the expectations of others are distinct from one's own expectations. If a trader's current prediction about future short rates differ from his expectation about other traders' predictions, it is rational for him to believe that other traders in the future will revise their predictions in the direction towards what he considers the best prediction, as more information becomes available. This section draws out the consequences of this fact for the term structure and contains the main theoretical contributions of the paper.

First, it is demonstrated that dispersed expectation about future short rates is sufficient for traders to be able to predict excess returns, even though risk premia are constant in the model. Secondly, we show that with non-nested information sets, individual traders can predict the average prediction error made by others which introduces speculative behavior in the sense of Harrison and Kreps (1978). Third, it is shown that the speculative dynamics introduced by non-nested information sets are orthogonal to public information which has interesting implications for how the speculative dynamics can (and cannot) be quantified using public bond price data. We start by defining what it means for information sets to be non-nested.

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*

$$
\mathcal{P}_{t,j}\widetilde{r}_{t+n} \neq \mathcal{P}_{t+s,i}\widetilde{r}_{t+n} : j \neq i \tag{3.1}
$$

for $s = 0, 1, 2, ..n - 1$ and for at least some $t = 0, 1, 2...$

Information sets are thus considered non-nested if two traders in at least some period t disagree about the best prediction of the short rate in period $t + n$ and expects to disagree up until period $t + n$ when the short rate $\widetilde{r}_{t+n}$ is observed. Defining non-nested information

sets through the implications for projections of short rates onto individual trader's informa-
tion 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 (see Brockwell and Davis 2006). The definition
used here is designed to avoid trivial cases where projections only differ about uninterest-
ing quantities. We therefore define information sets as non-nested only if they imply that
expectations about future short rates differ across agents.

3.1. **Predictable excess returns.** 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. First, we show that if individual traders' projections of future short rates are dispersed,
then the n period forward rate f_t^n cannot coincide generally with traders' expectations about
the corresponding 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 $\widetilde{r}_{t+n}$ onto trader
j's information set $\Omega_t(j)$ so that*

$$\mathcal{P}_{t,j}\widetilde{r}_{t+n} = f_t^n \tag{3.2}$$

*holds. The equality (3.2) can only hold generally, i.e. for all traders at all times, when
traders' information sets coincide.*

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

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

then the ex ante symmetry of traders implies that

$$\mathcal{P}_{t,j}\widetilde{r}_{t+n} = \mathcal{P}_{t,i}\widetilde{r}_{t+n} \forall j, i, t, n \tag{3.3}$$

or that the forward rate f_t^n is the best predictor of $\widetilde{r}_{t+n}$ for trader j at all times t and
at all horizons n only when it is also the best prediction for all others traders, i.e. when
information sets are nested. □

In words, Proposition 1 simply states that if the distribution across traders of expected
future short rates is non-degenerate, all points on the support of the distribution cannot
coincide with the forward rate, which is a single number. This is illustrated in Figure 1 and
may seem like an obvious statement, but the uniqueness of orthogonal projections makes it
nevertheless interesting. To see why, note that by the uniqueness of orthogonal projections,
$\mathcal{P}_{t,j}\widetilde{r}_{t+n} \neq f_t^n$ implies that $\mathcal{P}_{t,j}(\widetilde{r}_{t+n} - f_t^n) \neq 0$, i.e. trader j can systematically predict the
forecast error a person would make who used the forward rate as a forecast of the short rate.
The next proposition shows that this in turn implies that traders can systematically predict
excess returns.

Definition 6. *(Excess return) The excess return on an n period bond is the difference in
return between holding an n period bond until maturity and the return on holding a sequence
of one period bonds over n periods, i.e.*

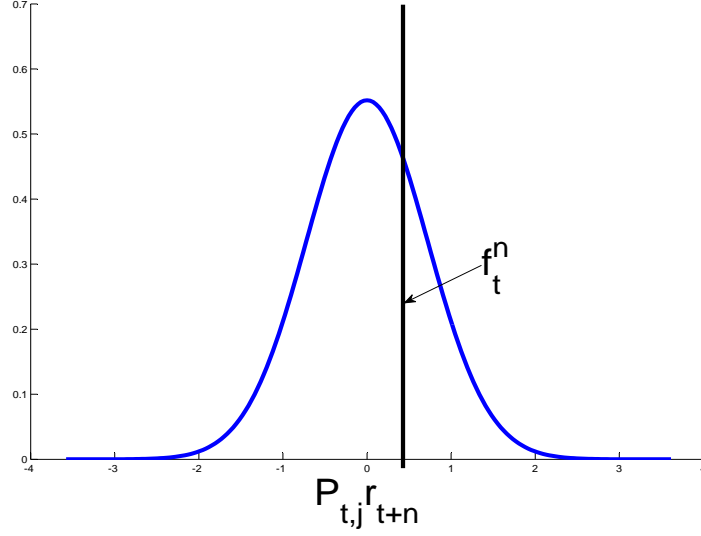$$-\widetilde{b}_t^n - (\widetilde{r}_t + \widetilde{r}_{t+1} + ...\widetilde{r}_{t+n-1}) \tag{3.4}$$

FIGURE 1. Dispersion of expectations of the short rate in period t+n and the implied forward rate.

An alternative (and more common) definition of excess returns would be to define it as the difference in return of holding an n period bond for one period minus the one period yield

$$\widetilde{b}_{t+1}^{n-1} - \widetilde{b}_t^n - \widetilde{r}_t \tag{3.5}$$

This definition is equivalent to (3.4) in the sense that if the expectations hypothesis hold, expected excess returns will be zero according to both definitions (see Singleton 2006). The definition (3.4) is somewhat more convenient to work with in the present setting, but either definition could be used without changing the substance of the results presented below. We now use the result from Proposition 1 to show that non-nested information sets imply that individual traders can predict excess returns.

Proposition 2. *If traders' information sets are non-nested, excess returns are predictable by trader j for at least some maturity n^*.*

Proof. Excess return on an n^*-period bond are predictable with respect to trader j's information set if

$$\mathcal{P}_{t,j}\left(-\widetilde{b}_t^{n^*} - (\widetilde{r}_t + \widetilde{r}_{t+1} + ...\widetilde{r}_{t+n^*-1})\right) \neq 0 \tag{3.6}$$

for some t. From the identity

$$-\widetilde{b}_t^{n^*} \equiv \widetilde{r}_t + f_t^1 + ... + f_t^{n^*-1} \tag{3.7}$$

excess returns on an n^* period bond are thus predictable if

$$\sum_{s=1}^{n^*-1} \mathcal{P}_{t,j}\widetilde{r}_{t+s} \neq \sum_{s=1}^{n^*-1} f_t^s \tag{3.8}$$

We know from Proposition 1 that $\mathcal{P}_{t,j}\widetilde{r}_{t+n} \neq f_t^n$ holds for at least some, t, n and j if information sets are non-nested. To prove the proposition we need to show that this in turn implies that the inequality (3.8) must be true for at least some maturity n^*. The proposition holds trivially for $n = 2$ since the inequality (3.8) then is true by assumption. For $n > 2$, consider the case when (3.8) is an equality at maturity $n + 1$ so that

$$\sum_{s=1}^{n} \mathcal{P}_{t,j}\widetilde{r}_{t+s} = \sum_{s=1}^{n} f_t^s \tag{3.9}$$

which can equivalently be written as

$$\sum_{s=1}^{n-1} \mathcal{P}_{t,j}\widetilde{r}_{t+s} + \mathcal{P}_{t,j}\widetilde{r}_{t+n} = \sum_{s=1}^{n-1} f_t^s + f_t^n \tag{3.10}$$

Since by assumption, $\mathcal{P}_{t,j}\widetilde{r}_{t+n} \neq f_t^n$ we must have that

$$\sum_{s=1}^{n-1} \mathcal{P}_{t,j}\widetilde{r}_{t+s} \neq \sum_{s=1}^{n-1} f_t^s \tag{3.11}$$

which implies that excess returns on an $n^* = n$ period bond is predictable. To complete the proof, it is sufficient to note that the proposition is trivially true for $n^* = n + 1$ if (3.9) fails to hold. That is, if $\mathcal{P}_{t,j}\widetilde{r}_{t+n} \neq f_t^n$ excess returns cannot simultaneously be unpredictable for bonds of both maturity n and $n + 1$.   □

3.2. **Non-nested information sets and speculative trade.** Proposition 1 and 2 combined demonstrated that a non-degenerate distribution of short rate expectations is sufficient for excess returns to be predictable by individual traders. The next 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. *Non-nested information sets imply that an individual trader j can systematically predict the average period $t + s$ projection error of the short rate in period $t + n$, that is*

$$\mathcal{P}_{t,j}\left(\widetilde{r}_{t+n} - \int \mathcal{P}_{t+s,j'}\widetilde{r}_{t+n}dj'\right) \neq 0 \tag{3.12}$$

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

Proof. The expression (3.12) can be rearranged to

$$\mathcal{P}_{t,j}\widetilde{r}_{t+n} = \mathcal{P}_{t,j}\int \mathcal{P}_{t+s,j'}\widetilde{r}_{t+n}dj'. \tag{3.13}$$

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

$$\mathcal{P}_{t,j}\int \mathcal{P}_{t+s,j'}\widetilde{r}_{t+n}dj' = \mathcal{P}_{t,j}\mathcal{P}_{t+s,i}\widetilde{r}_{t+n} : i \neq j. \tag{3.14}$$

That is, an individual trader j's expectation about average expectations coincide with his expectation of trader i's expectation for any $i \neq j$. By property (4) of projections we know that

$$\mathcal{P}_{t,j}\widetilde{r}_{t+n} = \mathcal{P}_{t,j}\mathcal{P}_{t+s,i}\widetilde{r}_{t+n} \tag{3.15}$$

if and only if $\Omega_t(j) \subseteq \Omega_{t+s}(i)$ which contradicts the definition of non-nested information sets and completes the proof. $\qquad\square$

Proposition 3 showed that individual traders can systematically predict the average prediction errors made by other traders. To see how this induces speculative behavior, consider an example with $n = 2$ and $s = 0$ in (3.12) and

$$\mathcal{P}_{t,j}\left(\widetilde{r}_{t+2} - \int \mathcal{P}_{t,j'}\widetilde{r}_{t+2}dj'\right) = \Delta \tag{3.16}$$

where $\Delta > 0$ so that trader j's expectation about $\widetilde{r}_{t+2}$ is higher than what he thinks the current average expectation of $\widetilde{r}_{t+2}$ is. Since trader j knows that other traders will receive new information in period $t + 1$, he thinks that other traders on average will revise their expectations in period $t + 1$ towards what trader j considers the best prediction of $\widetilde{r}_{t+2}$ so that

$$\mathcal{P}_{t,j}\left(\widetilde{r}_{t+2} - \int \mathcal{P}_{t+1,j'}\widetilde{r}_{t+2}dj'\right) < \Delta \tag{3.17}$$

A positive Δ thus implies that trader j believes that other traders will revise their expectations of $\widetilde{r}_{t+2}$ upward in $t + 1$

$$\mathcal{P}_{t,j}\int \mathcal{P}_{t+1,j'}\widetilde{r}_{t+2}dj' > \mathcal{P}_{t,j}\int \mathcal{P}_{t,j'}\widetilde{r}_{t+2}dj' \tag{3.18}$$

and trader j's expectation of the one period return on a 3 period bond $(\widetilde{b}_{t+1}^2 - \widetilde{b}_t^3)$ is lower than his expectation about other traders expectations about the one period return, since by (2.12) the price of a 2 period bond in $t + 1$ depends negatively on the average expectation in $t + 1$ of $\widetilde{r}_{t+2}$.[3] Ceteris paribus, for a positive Δ trader j expects to hold less than the average trader of the 3 period bond, as he thinks others on average expects it to have a higher return. A symmetric argument holds if Δ is negative.

3.3. **Forward rates and average higher order prediction errors.** So far, we have discussed the implications of non-nested information sets from the perspective of the individual trader. We now turn to the implications of non-nested information sets for aggregate bond prices, or more specifically, for implied forward rates. In the next proposition, we show that the speculative dynamics introduced by non-nested information sets can be expressed as average predictions about higher order prediction errors of future short rates. That is, predictions about the difference between future short rates and other traders' predictions about future short rates.

---

[3]The argument will hold as long as there is not a perfect negative correlation between revisions to $\int \mathcal{P}_{t+s,j'}r_{t+1}$ and $\int \mathcal{P}_{t+s,j'}r_{t+2}$ for $s = 0, 1$.

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

Proof. For convenience, first define the notation

$$\prod_{s=0}^{n-1} \int \mathcal{P}_{t+s,j}\widetilde{r}_{t+n} \equiv \int \mathcal{P}_{t,j} \int \mathcal{P}_{t+1,j'}... \int \mathcal{P}_{t+n-1,j''}\widetilde{r}_{t+n}dj''...dj'dj. \tag{3.19}$$

and rewrite the definition of the n period forward rate as

$$f_t^n = \prod_{s=0}^{n-1} \int \mathcal{P}_{t+s,j}\widetilde{r}_{t+n} + \left(v_t^n - v_t^{n+1}\right). \tag{3.20}$$

Add and subtract $\int \mathcal{P}_{t,j}\widetilde{r}_{t+n}$ from the r.h.s. of (3.20) to get

$$\begin{aligned} f_t^n &= \int \mathcal{P}_{t,j}\widetilde{r}_{t+n} \\ &\quad - \int \mathcal{P}_{t,j}\underbrace{\left(\widetilde{r}_{t+n} - \prod_{s=0}^{n-1} \int \mathcal{P}_{t+s,j}\widetilde{r}_{t+n}\right)}_{n-1 \text{ order prediction error}} \\ &\quad + \left(v_t^n - v_t^{n+1}\right) \end{aligned} \tag{3.21}$$

□

The term on the second line of (3.21) is the average prediction of the $n-1$ order prediction error, i.e. the average prediction of the difference between the actual short rate in period $t+n$ and the $n-1$ order expectation of the short rate in period $t+n$.[4] The expression (3.21) thus demonstrates that even in the absence of time varying risk premia, forward rates do not necessarily reflect average expectations about future short rates.

In a model with perfect or common information, the higher order prediction errors on the second line of (3.21) would of course be zero and the n period forward rate would be a function of only the period t average expectation of the the short rate in period $t+n$ (and the exogenous supply shocks). This would also be true in a model where bonds are only traded when they are issued and then held until maturity. In such a setting, the expectation of other traders' expectations would not matter for the equilibrium price, since the price of a zero coupon bond at maturity is simply its face value, which is known to all traders. The price of the bond at the date of issue would then simply be such that the return on an n period bond equals the expected return on the alternative investment. (By imposing this condition for all maturities n, it can be shown that this alternative return is the average expectation of the cumulative return of holding a series of one period bonds for n periods.) The new dynamics introduced to the term structure by non-nested information sets contained in the higher order prediction error term is thus dependent on the fact that long bonds are traded

---

[4]In a different context, Bacchetta and Wincoop (2006) shows that a similar term (which they label the "higher order wedge") can be expressed as an average expectation error of the innovations to the fundamental process in their model.

frequently. This is also the sense in which the "speculative behavior" in this model conforms to the definition of Harrison and Kreps (1978).

There are also some differences between the model presented here and the set up of Harrison and Kreps (1978) that are worth noting. The most important of these is perhaps that Harrison and Kreps rule out short sales, with the implication that the price of the asset in their model is bounded below by what any single trader would be willing to pay for it, where he to hold on to the asset forever. In our model, there are no short sales constraints and the price of a bond can be either above or below what the equilibrium price would be if the bond could not be traded before maturity.

3.4. Speculative dynamics and public information.

It is straightforward to show that the speculative dynamics due to higher order prediction errors are orthogonal to public information. Before proving this statement, we first define two relevant information sets.

Definition 7. *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 8. *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 $\widetilde{r}_{t+n}$ 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 (3.21) as

$$
\begin{aligned}
f_t^n \;=\;& \mathcal{P}_t^p \widetilde{r}_{t+n} + \int \mathcal{P}_{t,j}^{\perp p} \widetilde{r}_{t+n} \\[2mm]
& - \int \mathcal{P}_{t,j}^p \left(\widetilde{r}_{t+n} - \prod_{s=0}^{n-1} \int \mathcal{P}_{t+s,j} \widetilde{r}_{t+n} \right) \\[2mm]
& - \int \mathcal{P}_{t,j}^{\perp p} \left(\widetilde{r}_{t+n} - \prod_{s=0}^{n-1} \int \mathcal{P}_{t+s,j} \widetilde{r}_{t+n} \right) \\[2mm]
& + \left(v_t^n - v_t^{n+1} \right)
\end{aligned}
\tag{3.22}
$$

Since $\Omega_t^p \subseteq \Omega_{t+s}(j)$ for all j and $s = 0, 1, ..., m-1$ and by Property 4 of orthogonal projections (see Appendix A) we have that

$$
\mathcal{P}_{t,j}^p \widetilde{r}_{t+n} = \mathcal{P}_{t,j}^p \prod_{s=0}^{n-1} \int \mathcal{P}_{t+s,j} \widetilde{r}_{t+n}
\tag{3.23}
$$

for all all j and $s = 0, 1, ..., m - 1$. The term on the second line of (3.22) is thus identically zero and the n period forward rate can thus be expressed as

$$
\begin{aligned}
f_t^n &= \mathcal{P}_t^p \widetilde{r}_{t+n} + \int \mathcal{P}_{t,j}^{\perp p} \widetilde{r}_{t+n} \\
&\quad - \int \mathcal{P}_{t,j}^{\perp p} \left(\widetilde{r}_{t+n} - \prod_{s=0}^{n-1} \int \mathcal{P}_{t+s,j} \widetilde{r}_{t+n} \right) \\
&\quad + \left(v_t^n - v_t^{n+1} \right)
\end{aligned}
\tag{3.24}
$$

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 which by definition is common knowledge. This is intuitive, since all traders know that all traders know, and so on, that all traders know that all traders observe a public signal and that the public signal therefore cannot be used to predict the errors that other traders will make. 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 can summarize our findings so far. Individual traders can identify and take advantage of predictable excess returns, even though risk premia are 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. We also demonstrated that the new dynamics introduced by speculative behavior is orthogonal to public information. This has an interesting empirical implication: Speculative dynamics cannot be detected using public data in real time. However, as econometricians we can use public information from periods $t + s : s > 0$ to extract an estimate of the term due to the speculative dynamics in period t. To do so, we need to specify a process for the short rate and traders' information sets.

4. The Estimated Model

In the previous section it was demonstrated that non-nested information sets introduce new dynamics to the term structure of interest rates. Here, we address the question whether these dynamics are quantitatively important. Above, bond prices were derived as functions of higher order expectations of future short rates. In order to have an operational model that can be estimated, we need to specify two more objects: A process for the short rate and the information sets of the traders.

4.1. **The short rate.** The short interest rate $\widetilde{r}_t$ is the sum of three exogenous factors

$$
\widetilde{r}_t = x_t^1 + x_t^2 + x_t^3
\tag{4.1}
$$

where the factors $\mathbf{x}_t \equiv \begin{bmatrix} x_t^1 & x_t^2 & x_t^3 \end{bmatrix}'$ follow the vector autoregressive process

$$
\mathbf{x}_t = A\mathbf{x}_{t-1} + C\mathbf{u}_t : \mathbf{u}_t \sim N(0, I_3)
\tag{4.2}
$$

The diagonal structure of A and the lower triangular structure of C

$$A = \begin{bmatrix} \rho_1 & 0 & 0 \\ 0 & \rho_2 & 0 \\ 0 & 0 & \rho_3 \end{bmatrix}, C = \begin{bmatrix} c_1 & 0 & 0 \\ c_{21} & c_2 & 0 \\ c_{31} & c_{32} & c_3 \end{bmatrix}$$

are normalizations that do not restrict the dynamics of $\widetilde{r}_t$. The three factor structure is motivated by two considerations. It gives a sufficiently high dimensional latent state to make the filtering problem of traders interesting, while keeping the model computationally tractable. In addition, using three factors implies that if traders were perfectly informed, the model would simply be a three factor affine no-arbitrage model with a constant price of risk.

4.2. **Traders' information sets.** All traders observe a vector of public signals containing the current short rate $\widetilde{r}_t$ and selected bond yields collected in the vector $\mathbf{y}_t$. Non-nested information sets are introduced through individual signals about the first two factors x_t^1 and x_t^2. Each signal is the sum of a the true factor and an idiosyncratic noise component and the noise is uncorrelated across signals and time. The vector of private signals $\mathbf{z}_t(j)$ observed by trader j is thus given by

$$\mathbf{z}_t(j) = \begin{bmatrix} x_t^1 \\ x_t^2 \end{bmatrix} + Q\zeta_t(j) : \zeta_t \sim N(0, I_2) \tag{4.3}$$

$$Q = \begin{bmatrix} q_1 & 0 \\ 0 & q_2 \end{bmatrix} \tag{4.4}$$

Since the short rate is observed directly and is the sum of the three factors, it is without loss of generality that traders observe private signals only about the first two factors. The vector

$$S_t(j) = \begin{bmatrix} \mathbf{z}_t'(j) & \widetilde{r}_t & \mathbf{y}_t' \end{bmatrix}' \tag{4.5}$$

contains all the signals that trader j observes in period t. Trader j's information set in period is thus given by

$$\Omega_t(j) = \{S_t(j), \Omega_{t-1}(j)\} \tag{4.6}$$

implying that traders condition their expectations on the entire history of observed signals.

4.3. **The solved model.** When traders have non-nested information sets it becomes optimal to form expectation about other traders' expectations, and natural representations of the state in this class of models tend to be infinite.[5] The model is solved using the method proposed in Nimark (2010) which delivers a law of motion for the (finite dimensional) state X_t of the form

$$X_t = MX_{t-1} + N\mathbf{e}_t \tag{4.7}$$

The state vector X_t contains stacked higher order expectations of the factors

$$X_t \equiv \begin{bmatrix} \mathbf{x}_t^{(0)\prime} & \mathbf{x}_t^{(1)\prime} & \cdots & \mathbf{x}_t^{(\overline{k})\prime} \end{bmatrix}' \tag{4.8}$$

---

[5]See Townsend (1983), Sargent (1991) and Makarov and Rytchkov (2009).

where the higher order expectations of the vector of factors $\mathbf{x}_t$ are defined recursively as

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

starting from $\mathbf{x}_t^{(0)} = \mathbf{x}_t$. The integer $\overline{k}$ is the maximum order of expectation considered and can be chosen to achieve an arbitrarily close approximation in the limit as $\overline{k} \to \infty$.

Common knowledge of the model is used to pin down the law of motion for X_t, that is, to find M and N in (4.7). As usual in rational expectations models, first order expectations $\mathbf{x}_t^{(1)}$ are optimal, i.e. model consistent estimates of the actual factors $\mathbf{x}_t$. The knowledge that other traders have model consistent estimates allow traders to treat average first order expectations as a stochastic process with known properties when they form second order expectations. Common knowledge of the model thus implies that second order expectations $\mathbf{x}_t^{(2)}$ are optimal estimates of $\mathbf{x}_t^{(1)}$ given the law of motion for $\mathbf{x}_t^{(1)}$. Imposing this structure on all orders of expectations allows us to find the law of motion for the complete hierarchy of expectations as functions of the structural parameters of the model, i.e. the parameters governing the short rate (4.1), traders' information sets (4.6) and the parameter σ_v governing the standard deviation of the bond supply shocks v_t. The vector e_t contains the aggregate shocks in the economy, i.e. the factor innovations $\mathbf{u}_t$ and the bond supply shocks $\mathbf{v}_t$.

For a given law of motion (4.7), bond prices can be derived using the average expectation operator $H : \mathbb{R}^{\overline{k}} \to \mathbb{R}^{\overline{k}}$ that annihilates the lowest order expectation of a hierarchy so that

$$\begin{bmatrix} \mathbf{x}_t^{(1)} \\ \mathbf{x}_t^{(2)} \\ \vdots \\ \mathbf{x}_t^{(\overline{k}+1)} \end{bmatrix} = H \begin{bmatrix} \mathbf{x}_t^{(0)} \\ \mathbf{x}_t^{(1)} \\ \vdots \\ \mathbf{x}_t^{(\overline{k})} \end{bmatrix} \tag{4.9}$$

and where $\mathbf{x}_t^{(k)} = 0 : k > \overline{k}$. Combing the operator H that moves expectations one step up in orders of expectations and the matrix M from the law of motion (4.7) that moves expectations one step forward in time allows us to compute the higher order expectation in the bond pricing equation (2.14) as

$$\prod_{s=0}^{n-1} \int \mathcal{P}_{t+s,j} \widetilde{r}_{t+n} = \begin{bmatrix} \mathbf{1}_{1\times 3} & \mathbf{0} \end{bmatrix} (MH)^{s-1} X_t \tag{4.10}$$

so that bond prices are given by

$$\widetilde{b}_t^n = -\sum_{s=0}^{n-1} \begin{bmatrix} \mathbf{1}_{1\times 3} & \mathbf{0} \end{bmatrix} (MH)^{s-1} X_t + v_t^n \tag{4.11}$$

The matrix M governs the actual dynamics of $\widetilde{r}_t$ while bonds are priced as if X_t was observed by all agents and $\widetilde{r}_t$ followed a process governed by MH. The matrices M and MH are thus analogous to the "physical" and "risk neutral" dynamics in a standard no-arbitrage framework, though the interpretation is different.

Since bond prices are a linear function of the state X_t plus the supply shocks $\mathbf{v}_t$ we can use the definition of bond yields $y_t^n = -n^{-1}\widetilde{b}_t^n$ to also write the vector of bond yields $\mathbf{y}_t$ as

linear function of the state and the supply shocks

$$\mathbf{y}_t = BX_t + \mathbf{v}_t \tag{4.12}$$

The state equation (4.7) and the yield equation (4.12) constitutes a state space system that can be used to let the observed history of bond yields inform us about the most likely values of the parameters of the model. More details on how the model was solved can be found in the Appendix.

4.4. **Posterior Estimates.** The solved model, i.e. the state space system (4.7) and (4.12), is in a form that can be estimated directly by likelihood based methods. The standard deviations of an n period bond supply shock v_t^n is specified as $n\sigma_v$, and supply shocks are assumed to be independent across maturities. The presence of the supply shocks $\mathbf{v}_t$ thus add σ_v as a single new parameter to estimate along with the the parameters of the short rate process (4.2) and the private signals of traders (4.3). The vector of parameters to be estimated is denoted $\theta \equiv \{A, C, Q, \sigma_v\}$ and consists of a total of 12 parameters. 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. These are the same maturities that the traders inside the model are assumed to observe. 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 parameter distributions was simulated by 500 000 draws from an Adaptive Metropolis algorithm (see Haario, Saksman and Tamminen (2001)), initialized from a parameter vector found by maximizing the posterior using the simulated annealing maximizer of Goffe (1996). There are some large discrete mean adjustments in the first 200 000 draws of the Markov chain, but visual inspection suggests that both first and second moments have converged after 500 000 draws.[6] The results based on the last 250 000 draws are reported in Table 1. The posterior mode $\widehat{\theta}$ is the parameter vector of the simulated posterior Markov chain that achieves the highest posterior likelihood.

---

[6]More details, including the simulated posterior Markov chain and recursive plots of parameter standard deviations, are available to download from the author's web page.

Table 1

Posterior Parameter Estimates 1964:1-2007:12

θ	Mode $\widehat{\theta}$	Prior dist.	Posterior 2.5%-97.5%
Short rate process			
ρ_1	0.990	$U\left(0,0.99\right)$	0.989 - 0.990
ρ_2	0.989	$U\left(0,0.99\right)$	0.987 - 0.990
ρ_3	0.696	$U\left(0,0.99\right)$	0.655 -0.750
c_1	11.2	$U\left(0,50\right)$	7.42 - 26.8
c_2	0.62	$U\left(0,50\right)$	0.59 - 0.65
c_3	0.48	$U\left(0,50\right)$	0.45 - 0.51
c_{21}	-11.9	$U\left(-50,50\right)$	(-27.4) - (-8.11)
c_{31}	-0.49	$U\left(-50,50\right)$	(-0.54) - (-0.45)
c_{32}	0.38	$U\left(-50,50\right)$	0.34 - 0.43
Noise in private signals			
q_1	0.35	$U\left(0,50\right)$	0.24 -1.14
q_2	1.01	$U\left(0,50\right)$	0.61- 1.34
Bond supply shocks			
σ_v	0.21	$U\left(0,50\right)$	0.20 - 0.21

Log likelihood at $\widehat{\theta}$: -248.18

By themselves, the posterior estimates are not very interesting, but we can note that all parameters appear to be well-identified. The first two factors are very persistent and traders appear to have more precise private information about the first factor than about the second, that is $q_1 < q_2$. The standard deviation σ_v of the supply shocks are similar to the estimated standard deviation of measurement errors in latent factor models (e.g. Duffee 2008).

Figure 2 displays one period ahead fitted values together with the actual (demeaned) data series for selected yields. The interpretation is similar to the fit of a VAR and it is clear that the model can explain most of the observed variation in yields.

4.5. **Historical speculation.** The previous section showed that the speculative term introduced to the term structure by non-nested information sets could be expressed as a higher order prediction error that is orthogonal to public information. Nevertheless, we can quantify this term using public price data since the period t higher order prediction error is only orthogonal to public information known up to time t. As econometricians, we can use the full sample and exploit information for $t+s : s > 0$ to back out information about the higher order prediction error in period t. The Kalman simulation smoother can be used to draw from $p\left(X^T \mid \mathbf{y}^T, \theta\right)$ for a given parameter vector θ (e.g. Durbin and Koopman 2002). The simulation smoother together with the posterior distribution of θ can thus be used to construct the posterior distribution of the state $p\left(X^T \mid \mathbf{y}^T\right)$. Once we have a posterior distribution of the state, it is straightforward to compute the distribution of the speculative
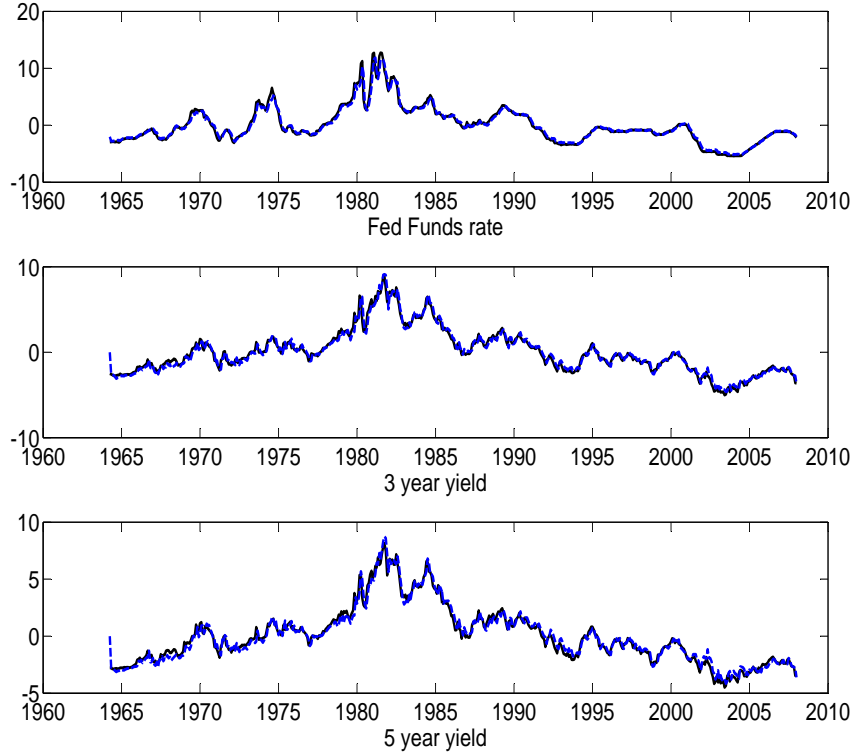
FIGURE 2. Data (black solid) and one-sided model fit (blue dashed).

term (3.21) in the implied n period forward rate by using that

$$-\int \mathcal{P}_{t,j}\left(\widetilde{r}_{t+n} - \prod_{s=1}^{n-1}\int \mathcal{P}_{t+s,j}\widetilde{r}_{t+n}\right) \tag{4.13}$$

$$= \begin{bmatrix} \mathbf{1}_{1\times 3} & \mathbf{0} \end{bmatrix}\left[(MH)^n - M^n H\right]X_t$$

where the equality follows from (4.10) and the fact that

$$\int \mathcal{P}_{t,j}\widetilde{r}_{t+n} \;=\; \begin{bmatrix} \mathbf{1}_{1\times 3} & \mathbf{0} \end{bmatrix} M^n H X_t \tag{4.14}$$

$$=\; \mathbf{1}_{1\times 3} \times A^n \int \mathcal{P}_{t,j}\mathbf{x}_t^{(0)} dj \tag{4.15}$$

Figure 3 displays the median and the 95% (point-wise) posterior probability intervals for the speculative term in the one $(n = 12)$ and five $(n = 60)$ year forward rates.

The speculative term in the two maturities are strongly correlated and potentially quantitatively important at both short and long maturities. At their peaks, around 1981, the

median estimate is 3.5% and 2.5% for respectively, the 1 and 5 year forward rates. The 95% probability interval reaches even higher, and is above 5% for the the speculative term in the 1 year forward rate. One way to interpret these numbers is the following. If a "naive" observer in 1981 took the implied forward rate as an indicator of market expectations about the short rate in 1982, the magnitude of the mistake he would make by not controlling for the speculative term would be 3.5%.

Singleton (2006) points out that violations of the expectation hypothesis in US data are most pronounced when the period 1979-1983 is included in sample, and the episode in the early 1980's is also the most eye-catching (and volatile) in Figure 3. Since the speculative term (4.13) is (up to the supply shocks $\mathbf{v}_t$) the difference between the implied forward rate and the average expectations of the future short rate, the speculative term has some features that other models may attribute to time varying risk premia. However, the interpretation our model suggest is quite different. The period 1979 - 1983 coincides with the so-called Volcker disinflation when the then Federal Reserve chairman Paul Volcker raised interest rates sharply to bring inflation under control (see for instance the account in Goodfriend and King 2005). Once inflation credibility had been established, short interest rates began to fall, though long rates stayed long for some time. The estimated model suggests that this was an episode when first order expectations of future short rates where significantly lower than higher order expectations. That is, individual traders may have found it credible that Volcker would be able to keep future short rates low before they believed that other traders had been convinced as well. This contrasts with the alternative explanation that this was a period when traders demanded either more compensation to hold a given amount of risk, or when the amount of risk was perceived to be higher than usual.

4.6. **Quantifying the importance of speculative dynamics.** We can also use the estimated model to quantify the relative importance of speculation at different maturities. The left panel of Figure 4 displays the median and the 95 per cent probability interval of the relative standard deviation of the speculative term and forward rates. At the one year horizon, the relative standard deviation of the speculative term is around 15% at the median and rises to approximately 30% at the peak at around a maturity of three years ($n = 36$). It is also evident from the figure that the speculative term is not only quantitatively important, but it is also statistically significant, with the 95% probability interval clearly bounded away from zero. (The statistical significance of the speculative term is somewhat obscured by the point-wise nature of the probability intervals in Figure 3.)

The right panel of Figure 4 displays the relative standard deviation of the speculative term in yields, computed as the recursive average of the speculative term in forward rates. We can see that the speculative term is quantitatively important also for yields, and more so at long horizons, peaking at the four year maturity where it reaches 22.5 % at the median. The fact that speculative dynamics appear to be more important for longer maturities may also help explain the evidence in Gürkaynak, Sack and Swanson (2005) who argue that current macro models of the term structure have trouble explaining the "excess" variability of long bond yields. Embedding a non-nested information structure in a macro model may improve these models' ability to match the variance of long term yields.
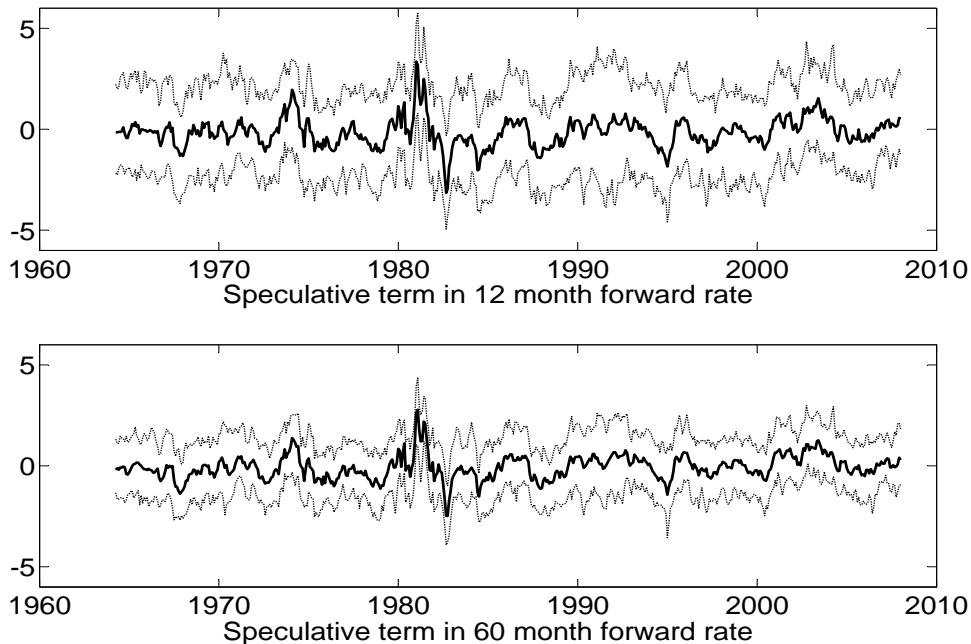
FIGURE 3. Estimated speculative term (percentage points) in 12 and 60 month implied forward rate 1964:1 - 2007:12. Median (solid) and 95% probability interval (dotted).

4.7. **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 that 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. Since no information about expectation dispersion is used in the estimation process, this can be thought of as an informal test of over-identifying restrictions.

The standard deviation of individual traders expectations of the short rate n months ahead is given by

$$\sqrt{E\left(\mathcal{P}_{t,j}\widetilde{r}_{t+n} - \int \mathcal{P}_{t,j'}\widetilde{r}_{t+n}\right)^2} \tag{4.16}$$

The 95 per cent probability intervals for $n = 12$, 36 and 60 of the posterior distribution of dispersion of first order expectations are reported in Table 2. The first row ($n = 12$) is directly comparable to the survey evidence reported by Swanson (2006). The dispersion implied by the estimated model is significantly smaller than that of the Blue Chip survey. The spread between the 10th and the 90th percentile of a Gaussian distribution is approximately 3.9 standard deviations. The estimated spread is thus approximately $0.12 \times 3.9 = 0.47$ or
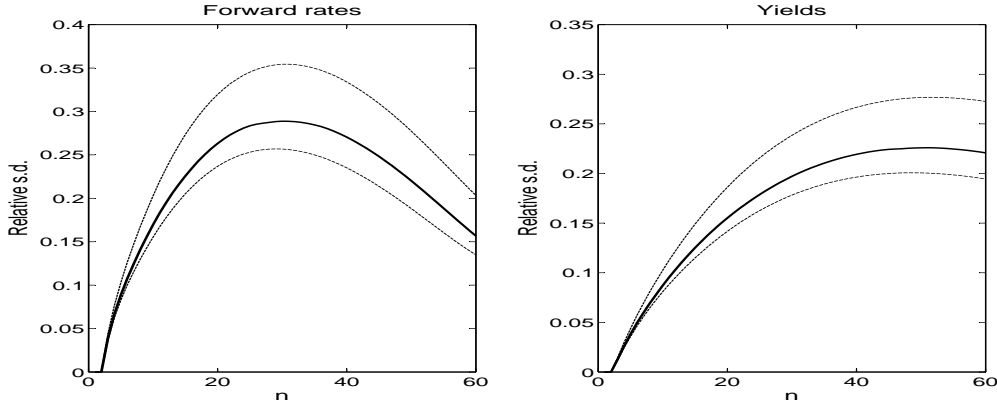
FIGURE 4. Estimated relative standard deviation of speculative term in forward rates (left panel) and yields (right panel). Median (solid) and 95% probability interval (dotted).

around 50 basis points at the median and around 60 basis points at the 97.5 percentile. The estimated dispersion is thus smaller than even the lower end (80 basis point) of the spread reported by Swanson. The model can thus not be considered to rely on an implausibly large dispersion of expectations to fit the data. This should increase our confidence in the model.

Table 2

Estimated dispersion of short rate expectations

n	2.5%	$\sqrt{E\left(\mathcal{P}_{t,j}\widetilde{r}_{t+n} - \int \mathcal{P}_{t,j'}\widetilde{r}_{t+n}\right)^2}$ Median	97.5%
12	0.08	0.12	0.15
36	0.06	0.09	0.11
60	0.05	0.07	0.08

All numbers in percentage points (1 basis point = 0.01).

Table 2 also reports the dispersion of expectations at the three and five year forecast horizon. The spread is slightly decreasing at longer horizons. Of course, at long enough horizons the dispersion will disappear as expectations of all agents will converge to the unconditional mean of short rates.

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

In this section, we relate the estimated model above to some of the statistical evidence from no-arbitrage factor models. We do three things. First, we show that the model with non-nested information sets compares favorably with an affine three factor no-arbitrage model in terms of fitting the dynamics of the term structure, in spite of having fewer parameters. Secondly, we demonstrate that if the world is characterized by the model presented here, an affine three factor no-arbitrage model would find overwhelming but misleading evidence

in favor of time varying risk premia. Thirdly, we show that the model can account for the "hidden" factors found by Duffee (2008).

5.1. **An estimated affine three factor no-arbitrage model.** A three factor no arbitrage model can be described by the following equations (see Ang and Piazzesi 2003). The three factors in the vector X_t^F follow

$$X_t^F = \Gamma X_{t-1}^F + \Psi e_t \tag{5.1}$$

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, see Joslin, Singleton and Zhu 2010.) The short rate is a linear function of the factors

$$\widetilde{r}_t = \delta' X_t^F \tag{5.2}$$

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

$$\widetilde{b}_t^n = B_n^F X_t^F \tag{5.3}$$

where

$$B_n^F = B_{n-1}^F \left(\Gamma - \Psi \lambda \right) - \delta' \tag{5.4}$$

As before, yields can be computed as

$$y_t^n = -n^{-1} \widetilde{b}_t^n + v_t^F \tag{5.5}$$

The affine model is estimated using likelihood based methods. The parameter vector θ^F to be estimated consists of $\Gamma, \Psi, \delta, \lambda$ and σ_{vF}^2 where σ_{vF}^2 is the variance of the yield measurement errors v_t^F (assumed constant across maturities). First, a preliminary mode is found by maximizing the likelihood over the free parameters using a simulated annealing algorithm. This mode is then taken as the initial value for the parameter vector in an Adaptive Metroplis algorithm used to construct a posterior estimate of the parameters. The final mode $\widehat{\theta}^F$ is the vector in the posterior distribution that achieves the highest log likelihood $\mathcal{L}\left(\mathbf{y}^T \mid \theta^F, \mathcal{M}_F\right)$. To formally compare the fit of the no-arbitrage model and the non-nested information model, one should ideally compute the posterior odds ratio. This involves computing the marginal likelihoods of the two models which is computationally demanding and the results are often inaccurate. One approximation of the posterior odds ratio that has the benefit of being easy to compute (see Canova 2007) is the Schwarz approximation

$$\mathcal{L}\left(\mathbf{y}^T \mid \widehat{\theta}^F, \mathcal{M}_F\right) - \mathcal{L}\left(\mathbf{y}^T \mid \widehat{\theta}, \mathcal{M}\right) - \frac{1}{2}\left(\dim \theta^F - \dim \theta\right) \ln T \tag{5.6}$$

where $\mathcal{M}_F$ and $\mathcal{M}$ denotes respectively, the no-arbitrage and the non-nested information model. The Schwarz approximation is thus similar to a (log) likelihood ratio but with a correction for the number of parameters.

Estimating the no-arbitrage model on the same sample used in Section 4 yields a log likelihood at the mode of -893.99, which is significantly lower than the log likelihood (-260.77) of the non-nested information model.[7] This is in spite of the fact that the number of parameters in the non-nested information model is significantly fewer than in

---

[7]More details on the estimation of the affine three factor no-arbitrage model is available from the author's web page.
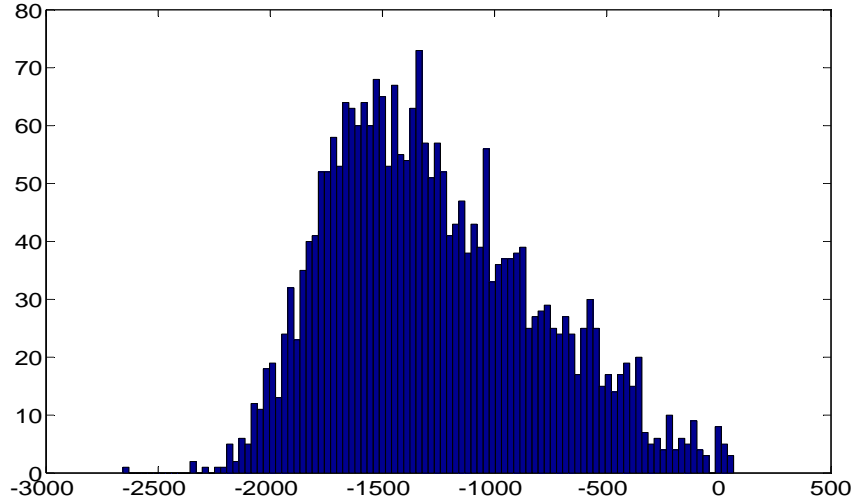
FIGURE 5. Histogram of Schwarz approximation to the posterior odds ratio of the restricted and unrestricted three factor no-arbitrage model estimated on artificial data.

the three factor no-arbitrage model (12 vs 19). The Schwarz approximation will thus further punish the no-arbitrage model and gives an approximate log posterior odds ratio of $-893.99 - (-260.77) - \frac{1}{2}(19 - 12)\ln 528 = -655.16$. The Schwarz approximation thus suggests that the no-arbitrage model is $e^{-655.16} \approx 0$ times as likely as the non-nested information model.

To be fair, affine no-arbitrage models are able to also match the average slope of the yield curve, while here we estimate both models on demeaned data. It is nevertheless interesting to see that the more parsimonious model matches the deviations from the sample mean better. In principle, we could also use 7 additional parameters to adjust the average supply of bonds so that the non-nested information model also matches average yields. The number of parameters of each model would then be the same.

5.2. **Detecting constant risk premia with a no-arbitrage model.** Affine three factor no-arbitrage models are known to provide a 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 misspecified. The reason is 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 Rytchkov (2009)). In the estimated model of the previous section, the true model is approximated with a 75 dimensional state vector. Given the structure imposed by higher order expectations by common knowledge of rationality, the model places stringent restrictions on the dynamics of the observable variables, in spite of the high dimensional state vector.

In this section, we investigate what an affine three factor no-arbitrage model would find if the estimated model of the previous section represents the true data generating process. We are particularly interested in finding out whether a three factor no-arbitrage model will correctly detect that risk premia are constant in the model that generated the artificial data. The experiment we conduct is the following. We first draw parameters from the posterior distribution of the non-nested information model and for each draw we generate 528 observations. The no-arbitrage model is then estimated by maximum likelihood and for each sample of artificial data we also re-estimate the model imposing the restriction that risk premia is constant, which implies setting λ in (5.4) equal to zero. We then compute the Schwarz approximation to the posterior odds ratio of the restricted and unrestricted model. This procedure was repeated 2500 times.[8]

Figure 5 displays the histogram of the posterior odds ratios of the restricted model with $\lambda = \mathbf{0}$ and the unrestricted model. If the restriction of no time-varying risk premia was supported by the data, the (log of the) posterior odds ratio should on average equal zero. Instead, the average log posterior odds ratio is around -1400. The Schwarz approximation to the posterior odds ratio then suggest that the restricted model is only $e^{-1400} \approx 0$ as likely as the unrestricted model. In other words, the restricted model has approximately 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* risk premia. Of course, this does not prove that the actual data generating process is a model with non-nested information sets and constant risk premia. However, it does demonstrate that the evidence from this popular class of statistical models is not sufficient reason to conclude that risk premia are time varying.

5.3. **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

$$x_t^\dagger = D^\dagger x_{t-1}^\dagger + \Sigma^\dagger \epsilon_t \tag{5.7}$$
$$y_t = A + B^\dagger x_t^\dagger + v_t^\dagger \tag{5.8}$$

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 manifest itself as 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 by maximum

---

[8]The procedure is very time consuming and the 2500 repetitions took about 5 weeks to compute.
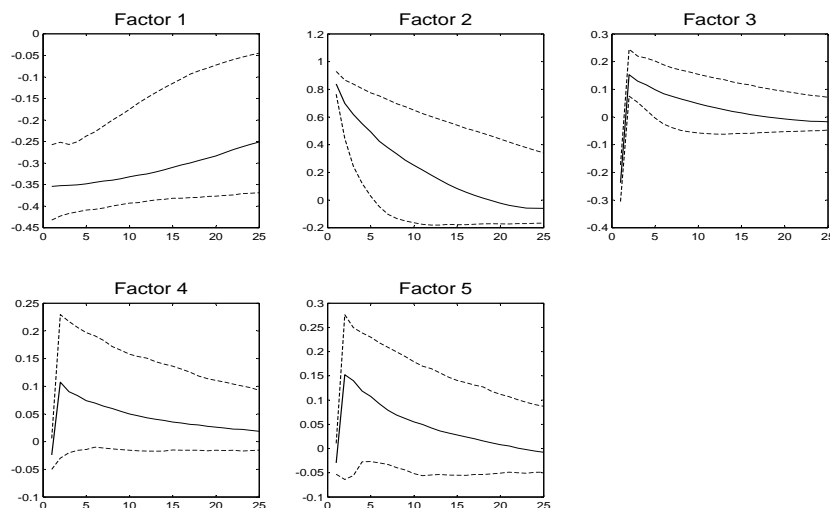
FIGURE 6. Impulse responses to orthogonal factor innovations, dashed lines
are 5th and 95th percentile.

likelihood and performed the rotations to find the principal components. We then computed
impulse responses of short rates to orthogonal innovations to the factors. This procedure
was repeated 450 times. Figure 7 shows the median impulse response and the 5th and 95th
percentile.[9] As we can see, the fourth and fifth principal components 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 does 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 factors 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 construction there
must be additional factors that can help predict future yields.

6. CONCLUSIONS

Introducing private information in a model of bond pricing can give rise to speculative
behavior in the sense of Harrison and Kreps (1978). For traders in the model to engage in
speculative behavior it is sufficient that information sets are non-nested, or equivalently that
there is dispersion across traders' forecasts of future short rates. Dispersed expectations are
also sufficient for traders to be able to predict excess returns, and the model thus provides a
new explanation for predictable excess returns that is not based on the value traders attach
to a marginal increase of wealth in different states of the world.

---

[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. A full Bayesian posterior simulation for each
draw of artificial data is too time consuming to be feasible.

Theoretical models of private information in asset markets dates back at least to Grossman (1976). More recently, Allen, Morris and Shin (2006) presented a single risky asset, finite horizon model with an information structure similar to the one presented here. They show that a concern among market participants about other market participants opinions as described in Keynes' (1936) "beauty contest" metaphor of the stock market, can be present among fully rational traders if traders have access to private information. Here we have demonstrated that even if the estimated amount of private information is small in the sense that the dispersion of (first order) expectations across traders is low, speculative trade driven by attempts to exploit perceived market mispricing of bonds can be quantitatively important. We showed this by formulating an empirically plausible model of the term structure that was estimated using likelihood based methods. The model has fewer parameters than a popular class of affine term structure models, but nevertheless fit the dynamics of bond yields better. We also demonstrated how a historical time series of the effect of speculative dynamics on implied forward rates can be estimated from public price data, in spite of the fact that speculation according to the model is orthogonal to real time public information.

Excess returns in the model presented here are predictable with respect to individual traders' information sets. While the speculative behavior this induces is shown to be potentially quantitatively important, in the model it is orthogonal to public price data in real time. One may then ask how these results relate to the large body of evidence documenting that excess returns are predictable using public information, e.g. Litterman and Sheinkman (1991), Ang and Piazzesi (2003) and Cochrane and Piazzesi (2005). First of all, there is nothing in the mechanism proposed here that excludes speculative dynamics from co-existing with time varying risk premia.

Having said this, we also demonstrated that a class of popular statistical models would find statistically significant, but nevertheless misleading, evidence in favor of time varying risk premia if the model with non-nested information sets is the true data generating process. Since neither model nests the other, it is difficult to say anything with confidence about the reasons for this, but one can speculate that three factor no-arbitrage models may be over-fitting the data. This view is also supported by several recent papers. Bauer (2009) and Barillas and Collin-Dufresne (2009) both show that affine factor models of the term structure may be biased toward overstating the predictability of excess returns. There is also survey evidence suggesting that excess returns may be less predictable than previously thought. Piazzesi and Schneider (2008) show in a no-arbitrage model with learning that if one imposes that the dynamics of expectations about future short rates should be consistent with survey evidence, then the estimated variability of risk premia decreases. Bacchetta, Mertens and van Wincoop (2008) show that predictable forecast errors from survey data are closely related to predictable excess returns in both bond and stock markets.

Recent studies have found that "unspanned" macro factors help predict excess returns, e.g. Ludvigson and Ng (2009) and Joslin, Priebsch and Singleton (2010) while the model here abstracts from public information not spanned by bond prices. Potentially, this could have implications for the estimated importance speculative dynamics. One possibility is that there are publicly observed unspanned factors that have exactly offsetting effects on short rate expectations and risk premia (as in the model of Joslin *et al* 2010). These factors would then appear not to be priced, i.e. these factors would not be spanned by current

bond prices, but could still help predict future excess returns. The public information set of the model presented here would then be too small and the importance of speculative dynamics could be overstated. However, there is another interpretation of the results of Ludvigsson and Ng (2009) and Joslin *et al* (2010) that is more favorable to our model. In both studies cited above, the macro information is dated according to the calender time of the quantity that it measures. As an example, the Consumer Price Index (CPI) for say January is dated as January in the bond pricing model and not dated as February, when the data is actually made public. In a perfect information setting like Joslin *et al*'s this distinction is unimportant as traders are assumed to observe the true state of the world at all times. However, in imperfect information models this distinction becomes important. To see why, note that a natural explanation why the predictive content of January CPI may appear not to be spanned by bond prices in January is that it is not observed until February. If interest rates moves at the time of the data release (i.e. in February), this will create the impression that January CPI in the bond price model helps predict excess returns. The fact that bond prices respond at the time of the many macro economic data releases (see for instance Faust, Rogers, Wang and Wright 2007) lends support to the latter interpretation. Whether there is relevant public information that is unspanned by bond prices but captured by macro economic data that *is available in real time* is thus arguably still an open question.

Ultimately, one would like to have a model that encompasses both time varying risk premia and speculative trade. In such a model, the identifying distinction that excess returns due to time varying price of risk should be predictable from public information in real time could potentially form the basis for a three-way statistical decomposition of the yield curve, adding a speculative term to the type of decomposition made by for instance Cochrane and Piazzesi (2005). The relative importance of time varying risk premia and speculative dynamics could then be addressed more directly.

Finally, the zero-coupon bonds traded in the present model have a known value at maturity. The uncertainty about future bond prices arise solely from the uncertainty about the discount rates that apply between the current period and the period when a bond matures. Arguably, these discount rates should matter also for the pricing of other assets that will be traded and pay dividends in the future. To the extent that there is additional uncertainty about dividend payments and returns on other classes of assets, speculative dynamics may be even more important in other markets than the bond market.

References

[1] Admati, Anit R., (1985), "A noisy rational expectations equilibrium for multi-asset securities markets" *Econometrica*, Vol. 53, pp. 629-657

[2] Allen, F. S. Morris and H.S. Shin, (2006), "Beauty Contests and Iterated Expectations in Asset Markets", *Review of Financial Studies*, 19, pp719 − 752.

[3] 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.

[4] Angeletos, G.M. and A. Pavan, (2009), "Policy with Dispersed Information," Journal of the European Economic Association 7:1.

[5] Bacchetta, Phillippe and Eric van Wincoop, (2006) "Can Information Heterogeneity Explain the Exchange Rate Determination Puzzle?", *American Economic Review* vol 96, pp552-576.

[6] Bacchetta, P. E. Mertens and E. van Wincoop, (2008) "Predictability in in Financial Markets: What do Survey Expectations Tell Us?", forthcoming, *Journal of International Finance*.

[7] Backus, D., A. Gregory and S. Zin, (1989), "Risk premiums in the term structure: Evidence from artificial economies", *Journal of Monetary Economics* 24, pp371-399.

[8] Backus, D, S. Foresi, a. Mozumdar and L. Wu, (2001), "Predictable changes in yields and forward rates," *Journal of Financial Economics*, vol. 59(3), pp281-311.

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

[10] Barillas, F. and P. Collin-Dufresne, (2009), "Small Sample Biases in Bond Predictive Regressions", mimeo, New York University.

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

[12] Cambell, J.Y. and L.M. Viceira, (2002), Strategic Asset Allocation, Cambridge University Press, Cambridge, UK.

[13] Canova, F., (2007), *Methods for Applied Macroeconomic Research*, Princeton University Press.

[14] Cespa, G. and X. Vives, (2009), "Dynamic Trading and Asset Prices: Keynes vs. Hayek", mimeo, IESE Business School.

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

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

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

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

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

[20] Durbin, J. and S.J. Koopman, (2002), "A simple and efficient simulation smoother for state space time series analysis", *Biometrica*, 89, pp. 603-615.

[21] Faust, J., J. Rogers, S. Wang, and J. Wright, (2007), "The high-frequency response of exchange rates and interest rates to macroeconomic announcements", *Journal of Monetary Economics*, vol. 54(4), pp1051-1068.

[22] Garcia, D. and J. Vanden, (2009), "Information acquisition and mutual funds", *Journal of Economic Theory*, 144(5), 1965-1995.

[23] Goffe, Bill, (1996), "SIMANN: A Global Optimization Algorithm using Simulated Annealing,", *Studies in Nonlinear Dynamics and Econometrics*, vol. 1, no. 3, pp169–176.

[24] Goodfriend, M. and R.King, (2005),"The Incredible Volcker Disinflation", *Journal of Monetary Economics,* 52, pp981-1015.

[25] Grossman, Sanford, 1976, "On the Efficiency of Competitive Stock Markets Where Trades Have Diverse Information", *Journal of Finance*, Vol. 31, No. 2, Papers and Proceedings. pp. 573-585.

[26] Grossman, S. and J. Stiglitz, (1980), "On the Impossibility of Informationally Efficient Markets", *American Economic Review* vol. 70(3), pages 393-408.

[27] Gürkaynak, R.S., B. Sack and E. Swanson, (2005). The Sensitivity of Long-Term Interest Rates to Economic News: Evidence and Implications for Macroeconomic Models, *American Economic Review* 95, pp425-436.

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

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

[30] Joslin, S., M. Priebsch and K. Singleton, (2010), "Risk Premiums in Dynamic Term Structure Models with Unspanned Macro Risks," working paper, Stanford University.

[31] Joslin, S., K. Singleton and H. Zhu (2010), "A New Perspective on Gaussian DTSMs", working paper, Stanford University.

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

[33] Kasa, K., T. Walker, and C. Whiteman, (2007), "Asset Prices in a Time Series Model with Perpetually Disparately Informed, Competitive Traders", mimeo, Indiana University.

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

[35] Ludvigsson, S. and S. Ng, (2009), "Macro Factors in Bond Risk Premia", forthcoming in the *Review of Financial Studies*.

[36] Makarov, I. and O. Rytchkov, (2009), "Forecasting the Forecasts of Others: Implications for Asset Pricing", mimeo, London Business School.

[37] Nimark, K., (2010), "Dynamic Higher Order Expectations", working paper, Universitat Pompeu Fabra.

[38] Nimark, K., (2009), "A Low Dimensional Kalman Filter for Systems with Lagged Observables ", working paper, Universitat Pompeu Fabra.

[39] Pearlman, J.G. and T.J. Sargent, (2005), "Knowing the forecasts of others", *Review of Economic Dynamics*, Volume 8, pp480-497.

[40] Piazzesi, M. and M. Schneider, (2009), "Trend and Cycle in Bond Premia", unpublished manuscript.

[41] Townsend, Robert M., (1983), Forecasting the Forecasts of Others, *Journal of Political Economy*, vol 91, pp546-588.

[42] Sargent, Thomas J., (1991), "Equilibrium with Signal Extraction from Endogenous Variables", *Journal of Economic Dynamics and Control* 15, pp245-273.

[43] Simon, Dan, (2006), *Optimal Filtering: Kalman, H_∞ and Non-Linear Approaches*, Wiley-Interscience.

[44] Singleton, Kenneth J., (1987), "Asset prices in a time series model with disparately informed, competitive traders", in *New Approaches to Monetary Economics,* Eds. W.A. Burnett and K.J. Singleton, Cambridge University Press.

[45] Singleton, Kenneth J., (2006), *Empirical Dynamic Asset Pricing*, Princeton University Press.

[46] Swanson, Eric.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.

[47] Walker, T., (2007), "How Equilibrium Prices Reveal Information in Time Series Models with Disparately Informed, Competitive Traders", *Journal of Economic Theory*.

[48] 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.

[49] Xiong, W. and H. Yan, (2009),"Heterogenous Expectations and Bond Markets", Review of Finacial Studies, forthcoming.

Appendix A. Some useful properties of projections

This Appendix reproduces some results and properties of orthogonal projections on inner-product spaces that are used in Section 3 of the main text above. 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 C of all random variables X with finite variance*

$$EX^2 < \infty \tag{A.1}$$

and with inner-product

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

Definition 2. *Let Ω be a subspace of L^2. An orthogonal projection of X on Ω, denoted $P_\Omega X$, is the unique element in L^2 satisfying*

$$\langle X - \mathcal{P}_\Omega X, \omega \rangle = 0 \tag{A.3}$$

for any $\omega \in \Omega$.

Orthogonal projections have the following useful properties:

(1) The projection $\mathcal{P}_\Omega X$ coincides with the conditional expectation $E\left[X \mid \Omega\right]$ in linear models with Gaussian shocks.
(2) Let Ω' be a subspace of Ω and $\Omega'^\perp$ its orthogonal complement in Ω. Then each $\omega \in \Omega$ has a representation as a sum of an element in Ω' and an element of $\Omega'^\perp$, i.e.

$$\omega = \mathcal{P}_{\Omega'}\omega + \mathcal{P}_{\Omega'^\perp}\omega \tag{A.4}$$

(3) $X \in \Omega^\perp$ if and only if $\mathcal{P}_\Omega X = 0$, where $\Omega^\perp$ is the orthogonal complement to Ω.
(4) $\Omega_1 \subseteq \Omega_2$ 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) - (4) to analyze traders's expectations in a model with linear constraints and Gaussian shocks. Property (2) was used in the proof of Proposition 5 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) is used to show that individuals can predict average expectations errors when information sets are non-nested. Property (4) is 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 other traders' prediction errors as well as excess returns when information sets are non-nested.

Appendix B. Solving the model

Solving the model implies substituting out the higher order expectations from the bond pricing equation (2.14). We are looking for a solution of the form

$$X_t = MX_{t-1} + N\mathbf{e}_t \tag{B.1}$$
$$\mathbf{y}_t = BX_t + \mathbf{v}_t \tag{B.2}$$

where

$$X_t \equiv \begin{bmatrix} \mathbf{x}_t^{(0)} \\ \mathbf{x}_t^{(1)} \\ \vdots \\ \mathbf{x}_t^{(\bar{k})} \end{bmatrix}, \mathbf{e}_t = \begin{bmatrix} \mathbf{u}_t \\ \mathbf{v}_t \end{bmatrix}$$

That is, to solve the model, we need to find the matrices M, N and B as functions of the parameters governing the short rate process, the stochastic supply shocks and the idiosyncratic noise shocks. The integer $\bar{k}$ is the maximum order of expectation considered and can be chosen to achieve an arbitrarily close approximation to the limit as $\bar{k} \to \infty$. Here, a brief overview of the method is given, but the reader is referred to Nimark (2010) for more details on the solution method.

First, common knowledge of the model can be used to pin down the law of motion for the vector X_t containing the hierarchy of higher order expectations of $\mathbf{x}_t$. Rational, i.e. model consistent, expectations of $\mathbf{x}_t$ thus implies a law of motion for average expectations $\mathbf{x}_t^{(1)}$ which can then be treated as a new stochastic process. Knowledge that other traders are rational, means that second order expectations $\mathbf{x}_t^{(2)}$ are determined by the average across traders of the rational expectations of the stochastic process $\mathbf{x}_t^{(1)}$. Third order expectations

$\mathbf{x}_t^{(3)}$ are then the average of the rational expectation of the process $\mathbf{x}_t^{(2)}$, and so on. Imposing this structure on all orders of expectations allows us to find the matrices M and N. Section $B.2$ below describes how this is implemented in practice.

Second, the method exploits that the importance of higher order expectations are decreasing with the order of expectation. This has two components:

(i) The variance of higher order expectations of the factors $\mathbf{x}_t$ are bounded by the variance of the true process, or more generally, the variance of $k + 1$ order expectation cannot be larger than the variance of a k order expectation

$$E\left[\mathbf{x}_t^{(k+1)}\mathbf{x}_t'^{(k+1)}\right] \leq E\left[\mathbf{x}_t^{(k)}\mathbf{x}_t'^{(k)}\right] \tag{B.3}$$

To see why, note that by the identity

$$\mathbf{x}_t^{(k)} \equiv \mathbf{x}_t^{(k+1)} + \varepsilon_t^{(k+1)} \tag{B.4}$$

and the fact that since $\mathbf{x}_t^{(k+1)}$ is the average of an optimal estimate of $\mathbf{x}_t^{(k)}$ the $k = 1$ order error $\varepsilon_t^{(k+1)}$ must be orthogonal to $\mathbf{x}_t^{(k+1)}$ we have that

$$E\left[\mathbf{x}_t^{(k)}\mathbf{x}_t'^{(k)}\right] = E\left[\mathbf{x}_t^{(k+1)}\mathbf{x}_t'^{(k+1)}\right] + E\left[\varepsilon_t^{(k+1)}\varepsilon_t'^{(k+1)}\right]. \tag{B.5}$$

Since $E\left[\varepsilon_t^{(k+1)}\varepsilon_t'^{(k+1)}\right]$ is a covariance

$$E\left[\varepsilon_t^{(k+1)}\varepsilon_t'^{(k+1)}\right] \geq \mathbf{0} \tag{B.6}$$

and the inequality (B.3) then follows immediately. (This is an abbreviated version of a more formal proof available in Nimark (2010).)

That the variances of higher order expectations of the factors are bounded is not sufficient for an accurate finite dimensional solution. We also need (ii) that the impact of the expectations of the factors on bond yields are decreasing "fast enough" with the order of expectation. To understand why this is the case, it helps to first note that the full information rational expectation equilibrium is a special case of the solution (B.1) - (B.2) where

$$\mathbf{x}_t^{(k)} = \mathbf{x}_t^{(k+1)} \forall t, k \tag{B.7}$$

That is, under full information, all orders of expectations of the factors coincide with the true factors at all times. Define the full information solution, i.e. bond yields as a function of the true state $\mathbf{x}_t^{(0)}$ as

$$\mathbf{y}_t = \widetilde{B}\mathbf{x}_t^{(0)} + \mathbf{v}_t \tag{B.8}$$

where the rows of the matrix $\widetilde{B}$ can be found by stacking the matrices $\frac{1}{n}\sum_{s=0}^{n-1}\left[\begin{array}{cc}\mathbf{1}_{1\times 3} & \mathbf{0}\end{array}\right]A^{s-1}$ for the appropriate choices of n. By chance, for some t, all orders of expectations can coincide also in the non-nested information case. Yields in the non-nested information model should then be the same as the yields in the full information model: If everybody believes that everybody agrees about what the current "true" state $\mathbf{x}_t^{(0)}$ is, they must also agree about their expectations about future short rates. For this to then result in the same yields as a

function of the state as the full information solution, the sub-matrices of B must sum to $\widetilde{B}$ since

$$\widetilde{B}\mathbf{x}_t^{(0)} = BX_t$$

$$= \begin{bmatrix} B_0 & B_1 & \cdots & B_{\overline{k}} \end{bmatrix} \begin{bmatrix} \mathbf{x}_t^{(0)} \\ \mathbf{x}_t^{(0)} \\ \vdots \\ \mathbf{x}_t^{(0)} \end{bmatrix} \tag{B.9}$$

$$= \sum_0^\infty B_j \mathbf{x}_t^{(0)} \tag{B.10}$$

if $\mathbf{x}_t^{(0)} = \mathbf{x}_t^{(k)} \forall k$ and where the B_js are the sub matrices of B in the non-nested information solution (B.2). Since $\widetilde{B}$ is finite, the sequence $B_0, B_1, ..., B_k, ...$ must be a convergent series, which implies that $\lim_{k \to \infty} B_k = 0$.

Summing up, we can use common knowledge of the model to derive a law of motion for the hierarchy of expectations of the factors $\mathbf{x}_t$. We can then find an approximate solution for bond yields of the form

$$\mathbf{y}_t = \begin{bmatrix} B_0 & B_1 & \cdots & B_{\overline{k}} \end{bmatrix} \begin{bmatrix} \mathbf{x}_t^{(0)} \\ \mathbf{x}_t^{(1)} \\ \vdots \\ \mathbf{x}_t^{(\overline{k})} \end{bmatrix} + Q\mathbf{e}_t \tag{B.11}$$

where the coefficients B_k that tend to zero as $k \to \infty$, are multiplied with a k order expectation, where the variance of the expectations are bounded from above by the variance of the true state $\mathbf{x}_t^{(0)}$. The accuracy of the solution for the $\overline{k}$ chosen in the estimated model is discussed below in section $B.4$.

B.1. **Bond yields as a function of the state.** For a given law of motion (4.7), bond prices can be derived using the average expectation operator $H : \mathbb{R}^{\overline{k}} \to \mathbb{R}^{\overline{k}}$ that annihilates the lowest order expectation of a hierarchy so that

$$\begin{bmatrix} \mathbf{x}_t^{(1)} \\ \mathbf{x}_t^{(2)} \\ \vdots \\ \mathbf{x}_t^{(\overline{k}+1)} \end{bmatrix} = H \begin{bmatrix} \mathbf{x}_t^{(0)} \\ \mathbf{x}_t^{(1)} \\ \vdots \\ \mathbf{x}_t^{(\overline{k})} \end{bmatrix} \tag{B.12}$$

and where $\mathbf{x}_t^{(k)} = 0 : k > \overline{k}$. The matrix H is then given by

$$H = \begin{bmatrix} \mathbf{0}_{3\overline{k} \times 3} & I_{\overline{k}-3} \\ & \mathbf{0}_{3 \times (\overline{k}-3)} \end{bmatrix} \tag{B.13}$$

Combing the operator H that moves expectations one step up in orders of expectations and the matrix M from the law of motion (4.7) that moves expectations one step forward in time

allows us to compute the higher order expectation in the bond pricing equation (2.14) as

$$\widetilde{b}_t^n = -\sum_{s=0}^{n-1} \begin{bmatrix} \mathbf{1}_{1\times 3} & \mathbf{0} \end{bmatrix} (MH)^{s-1} X_t + \widetilde{v}_t^n \tag{B.14}$$

where $\widetilde{v}_t^n \equiv n v_t^n$. The yield on a bond with n periods to maturity is then given by

$$y_t^n = \frac{1}{n} \sum_{s=0}^{n-1} \begin{bmatrix} \mathbf{1}_{1\times 3} & \mathbf{0} \end{bmatrix} (MH)^{s-1} X_t + v_t^n \tag{B.15}$$

By stacking the yield formula (B.1) for appropriate maturities n gives the matrix B in (B.2).

B.2. **The law of motion of higher order expectations of the factors.** To find the law of motion for the hierarchy of expectations X_t we use the following strategy. For a given M, N and B in (B.1) - (B.2) we will derive the law of motion for trader j's expectations of X_t, denoted $X_{t|t}(j) \equiv E[X_t \mid \Omega_t(j)]$. First, write the vector of signals $S_t(j)$ as a function of the state, the aggregate shocks and the idiosyncratic shocks

$$S_t(j) = \begin{bmatrix} \mathbf{z}_t'(j) & \widetilde{r}_t & \mathbf{y}_t' \end{bmatrix}' \tag{B.16}$$

$$= DX_t + R \begin{bmatrix} \zeta_t(j) \\ \mathbf{u}_t \\ \mathbf{v}_t \end{bmatrix} \tag{B.17}$$

where the matrix D is given by the definitions (4.1), (4.3) and (4.12)

$$D = \begin{bmatrix} I_2 & \mathbf{0} \\ \mathbf{1}_{1\times 3} & \mathbf{0} \\ & B \end{bmatrix} \tag{B.18}$$

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

$$R = \begin{bmatrix} R_j & R_A \end{bmatrix}.$$

Agent j's updating equation of the state $X_{t|t}(j)$ estimate will then follow

$$X_{t|t}(j) = MX_{t|t-1}(j) + K\left(S_t(j) - DMX_{t|t-1}(j)\right) \tag{B.19}$$

Rewriting the observables vector $S_t(j)$ as a function of the lagged state and taking averages across traders using that $\int \zeta_t(j)dj = 0$ yields

$$X_{t|t} = MX_{t|t-1} + K\left(DMX_{t-1} + (DN + R_A)\mathbf{e}_t + -DMX_{t|t-1}(j)\right) \tag{B.20}$$

$$= (M - KDM)X_{t|t-1} + KDMX_{t-1} + K(DN + R_A)\mathbf{e}_t \tag{B.21}$$

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

$$\begin{bmatrix} \mathbf{x}_t \\ X_{t|t} \end{bmatrix} = M \begin{bmatrix} \mathbf{x}_{t-1} \\ X_{t-1|t-1} \end{bmatrix} + N\mathbf{e}_t$$

where M and N are given by

$$M = \begin{bmatrix} A & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + \begin{bmatrix} \mathbf{0}_{3\times3} & \mathbf{0} \\ \mathbf{0} & [M - KDM]_- \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ [KDM]_- \end{bmatrix} \tag{B.22}$$

$$N = \begin{bmatrix} C & 0 \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ [K(DN + R_A)]_- \end{bmatrix} \tag{B.23}$$

where $[\cdot]_-$ indicates that the a last row or column has been canceled to make a the matrix $[\cdot]$ conformable, i.e. implementing that $\mathbf{x}_t^{(k)} = 0 : k > \overline{k}$. The Kalman gain K in (B.19) is given by

$$K = (PD' + NR')(DPD' + RR')^{-1} \tag{B.24}$$

$$P = M\left(P - (PD' + NR')(DPD' + RR')^{-1}(PD' + NR')'\right)M' + NN' \tag{B.25}$$

The model is solved by finding a fixed point that satisfies (B.2), (B.22), (B.23), (B.24) and (B.25).

B.3. **Numerical issues.** It is well-known that due to finite machine precision, the Kalman filter may be unstable in applications where some state variables (or some linear combination of state variables) are estimated very precisely compared to others, see Simon (2006). This manifests itself as a Kalman filter covariance matrix P that is numerically not symmetric and/or positive semi-definite. In our model, there are three latent factors, but a linear combination of them (i.e. the short rate $\widetilde{r}_t$) is observed perfectly and solving the model as above may (and for some parameterizations, do) give rise to numerical inaccuracies. To avoid this problem, we can rewrite the short rate process as

$$\widetilde{r}_t = x_t^1 + x_t^2 + x_t^3 \tag{B.26}$$

$$= x_t^1 + x_t^2 + \rho_3\left(\widetilde{r}_{t-1} - x_{t-1}^1 - x_{t-1}^2\right) + \begin{bmatrix} c_{31} & c_{32} & c_3 \end{bmatrix} \mathbf{u}_t \tag{B.27}$$

We can still solve the model as above if we redefine the state as

$$\widehat{X}_t = \begin{bmatrix} \widehat{\mathbf{x}}_t^{(0)} \\ \widehat{\mathbf{x}}_t^{(1)} \\ \vdots \\ \widehat{\mathbf{x}}_t^{(\overline{k})} \end{bmatrix}, \quad \widehat{\mathbf{x}}_t^{(0)} \equiv \begin{bmatrix} x_t^1 \\ x_t^2 \end{bmatrix}$$

to get a system of similar form as in Section $B.2$ above

$$\widehat{X}_t = \widehat{M}\widehat{X}_{t-1} + \overline{N}\mathbf{e}_t \tag{B.28}$$

$$\mathbf{y}_t = \widehat{B}\widehat{X}_t + R\mathbf{e}_t + B_r\widetilde{r}_t \tag{B.29}$$

but where yields now also depend on the current short rate directly. We also need to adjust the measurement equation for trader j to be

$$S_t(j) = D_1\widehat{X}_t + D_2\widehat{X}_{t-1} + D_r\widetilde{r}_{t-1} + \widehat{R}\begin{bmatrix} \zeta_t(j) \\ \mathbf{u}_t \\ \mathbf{v}_t \end{bmatrix} \tag{B.30}$$

where

$$\widehat{R} = \left[\begin{array}{cc} R_j & R_A \end{array} \right] + \left[\begin{array}{c} \mathbf{0} \\ \left[\begin{array}{ccccc} \mathbf{0} & c_{31} & c_{32} & c_3 & \mathbf{0} \end{array} \right] \\ \mathbf{0} \end{array} \right] \tag{B.31}$$

The updating equation of trader $j's$ state estimate is now given by

$$\widehat{X}_{t|t}(j) = \widehat{M}\widehat{X}_{t|t-1}(j) + \widehat{K} \left[S_t(j) - D_1\widehat{M}\widehat{X}_{t-1|t-1}(j) - D_2\widehat{X}_{t-1|t-1}(j) - D_r\widetilde{r}_{t-1} \right] \tag{B.32}$$

Since the additional term $D_2\widehat{X}_{t-1}$ in trader j's measurement equation is a function of the lagged state $\widehat{X}_{t-1}$, the innovation representation is non-standard. The Kalman gain $\widehat{K}$ in (B.32) is the steady state $(t \to \infty)$ Kalman gain of the modified Kalman filter

$$\widehat{K} = P_{+1}D_1' \tag{B.33}$$
$$\times \left[\left(D_1\widehat{M} + D_2 \right) P \left(D_1\widehat{M} + D_2 \right)' + \left(D_1\widehat{C} + \widehat{R} \right) \left(D_1\widehat{C} + \widehat{R} \right)' \right]^{-1}$$

$$P = P_{+1} \tag{B.34}$$
$$- \widehat{K} \left[\left(D_1\widehat{M} + D_2 \right) P \left(D_1\widehat{M} + D_2 \right)' + \left(D_1\widehat{C} + \widehat{R} \right) \left(D_1\widehat{C} + \widehat{R} \right)' \right] \widehat{K}'$$

$$P_{+1} = \widehat{M}P\widehat{M} + \widehat{C}\widehat{C}' \tag{B.35}$$

For a derivation of the modified filter, see Nimark (2009).

$\widehat{M}$ and $\widehat{N}$ are found in the same way as in (B.22), (B.23), i.e. by taking averages of the update equation (B.32) and amending it to the "new" true factor process

$$\left[\begin{array}{c} x_t^1 \\ x_t^2 \end{array} \right] = \widehat{A} \left[\begin{array}{c} x_{t-1}^1 \\ x_{t-1}^2 \end{array} \right] + \widehat{C}\mathbf{u}_t \tag{B.36}$$

$$= \left[\begin{array}{cc} \rho_1 & 0 \\ 0 & \rho_2 \end{array} \right] \left[\begin{array}{c} x_{t-1}^1 \\ x_{t-1}^2 \end{array} \right] + \left[\begin{array}{ccc} c_1 & 0 & 0 \\ c_{21} & c_2 & 0 \end{array} \right] \mathbf{u}_t \tag{B.37}$$

The resulting system is algebraically equivalent to the system in $B.2$, but does not suffer from numerical instability.

B.4. **The accuracy of the solution.** While it can be shown that the impact coefficients of higher order expectations on yields are a convergent sum, exactly how many orders of expectations that are needed for an accurate solution in a given application depends on the parameters of the model. In the model estimated in Section 4 $\bar{k} = 25$. Figure 8 displays the posterior distribution of the loadings onto the different orders of expectations of the factors for the 12, 36 and 60 month yields, that is, the posterior distributions of the odd (left column) and even (right column) numbered elements of the rows of the relevant rows of B in (4.12). As we can see, the loadings approaches zero quite quickly and the loadings onto the factors for the 12 month yield are approximately zero for $k > 12$. For longer maturities (36 and 60 months shown), the loadings are practically zero for $k > 15$. Choosing $\bar{k} = 25$ thus seems more than sufficient.
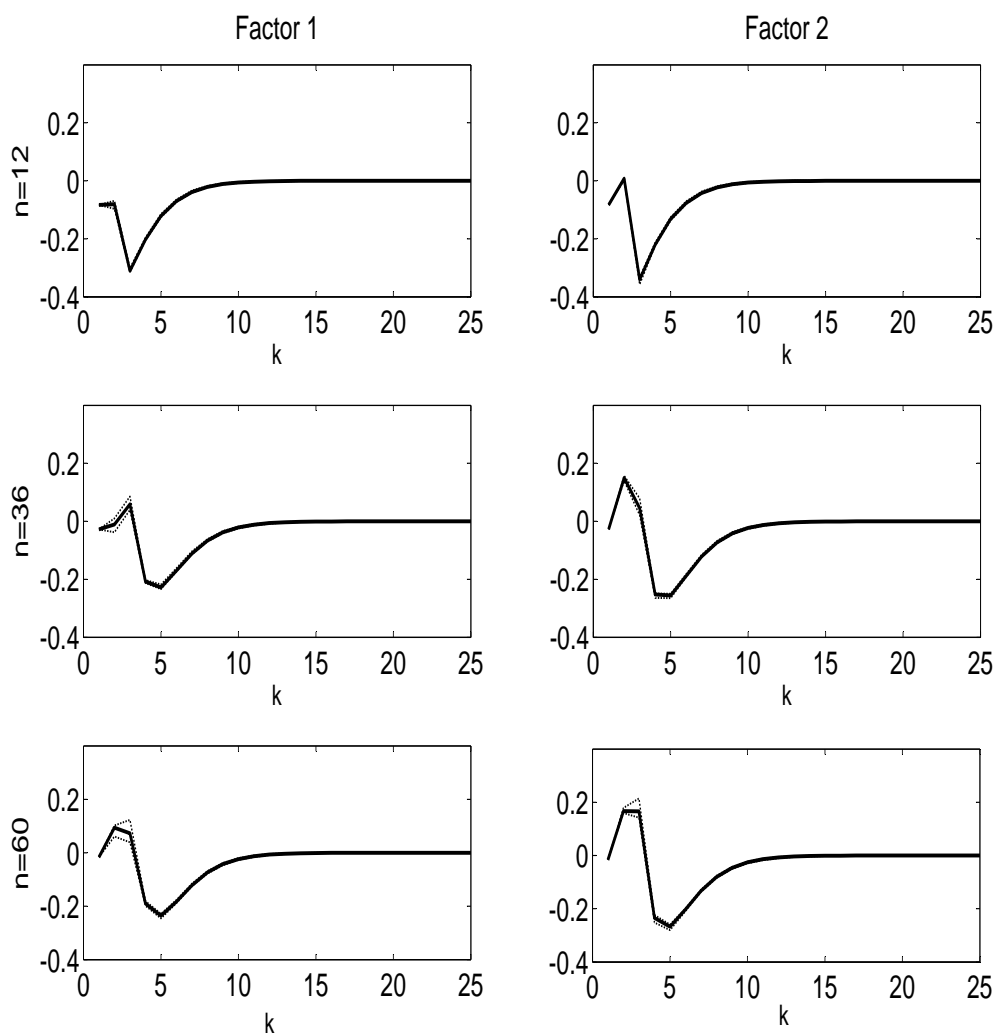
FIGURE 7. Estimated posterior (higher order) factor loadings, median (solid) and 95% probability interval (dotted). First column contains loadings onto the (higher order expectations of the) first factor x_t^1, with order of expectation k on the x-axis for yields of 1, 3 and 5 years to maturity. Second column contains the same information regarding the second factor x_t^1

Code and data used to estimate the model and generate figures are available at the author's web page.