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

KRISTOFFER P. NIMARK

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

1. Introduction

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

In this paper we present a term structure model populated with traders that engage in the type of speculative behavior described above. We use this model to argue that relaxing the assumption that all traders have access to the same information introduces empirically relevant new dynamics to the term structure of interest rates. More specifically, we apply the properties of orthogonal projections to show that if traders' information sets are non-nested,

---

Date: October 2, 2009. The author thanks David Backus, Francisco Barillas, Toni Braun, Jerome Henry, Helen Rey, Thomas Sargent, Chris Telmer, Pontus Rendahl and seminar and conference participants at New York University, Bank of Norway, Sveriges Riksbank, EABCN conference on *Uncertainty Over the Business Cycle*, ESSIM 2009, 3rd European Workshop on General Equilibrium Theory, University of Alicante, SED 2009 and NORMAC 2009 for useful comments and suggestions.

Address: CREI, Universitat Pompeu Fabra, Ramon Trias Fargas 25-27, Barcelona 08005.
e-mail: knimark@crei.cat *webpage*: www.kris-nimark.net .

(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 a 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.

The model is estimated on monthly US bond data. Using the estimated model to generate artificial data, we further show that an econometrician estimating an affine three factor no-arbitrage model would find overwhelming but misleading statistical evidence in favor of a time varying risk premia if the model with non-nested information sets were the true data generating process. The main thrust of this paper is thus to argue that predictable excess returns can be a consequence of traders having non-nested information sets, and not necessarily caused by that traders value a marginal increase in consumption (wealth) differently in different states of the world.

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 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. This is indeed what the evidence in Ludvigsson and Ng (2009) suggests.

If prices do not reveal all information relevant for predicting bond returns, it becomes more probable that traders have non-nested information sets, that is, traders will have access to and use different information when trading.[1] With the exceptions of bond prices, statements

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

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 under non-nested information sets (or one of its synonyms, see Footnote 1). Some examples are Allen, Morris and Shin (2005), Kasa, Walker and Whiteman (2008), Bacchetta and van Wincoop (2005, 2007) and Makarov and Rytchov (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. As far as the author knows, this is the first paper to estimate a model with non-nested information sets and the fit of the model is surprisingly good. The model is also used to quantify the dispersion of expectations implied by the posterior parameter estimates.

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

2. A Bond Pricing Model

In this section we present a simple bond pricing model. The current price of a bond depends on the average expectation of the price of the same bond in the next period, discounted by the one period interest rate. 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 the term structure dynamics of relaxing the assumption that traders all have access to the same information. In what follows, traders are indexed by $j \in (0,1)$ and trader j's information set is denoted $\Omega_t(j)$.

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

$$U_t(j) = \sum_{s=0}^{\infty} \beta^s E_t \left[\log C_{t+s} \mid \Omega_t(j)\right] \tag{2.1}$$

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

$$\alpha_t(j) = \Sigma^{-1}\left(E\left[\mathbf{b}_{t+1}^{(-1)} \mid \Omega_t(j)\right] - \mathbf{b}_t' - r_t\right) + \mathbf{a} \tag{2.2}$$

where $\mathbf{a}$ is a vector of mean portfolio weights and Σ is the conditional covariance of bond prices. The vector of bond supply $\mathbf{s}_t$ is stochastic and proportional to aggregate wealth W_t

$$\mathbf{s}_t = W_t \Sigma^{-1} \left(\mathbf{1} + \eta_t\right) \tag{2.3}$$

$$\eta_t \sim N\left(0, \sigma_\eta^2 \cdot I\right) \tag{2.4}$$

(the term Σ^{-1} in (2.3) simply normalizes the variance of the supply shocks η_t). Equating aggregate demand $W_t \int \alpha_t(j)dj$ and supply gives the log price b_t^n of an n periods to maturity zero coupon bond

$$b_t^n = c^n - r_t + \int E\left[b_{t+1}^{n-1} \mid \Omega_t(j)\right] dj - \eta_t^n. \tag{2.5}$$

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

$$\widetilde{b}_t^1 = b_t^n - c^n. \tag{2.6}$$

2.2. **The term structure of interest rates.** The bond price formula (2.5) can be used to price any maturity bond. The procedure is similar to deriving bond prices under a no arbitrage assumption, though we need to be more careful in specifying the information sets that the expectations that govern prices are conditioned on. As usual, we can start from

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

and apply (2.5) recursively. The log price of a two period bond then is

$$\widetilde{b}_t^2 = -r_t - \int E\left[r_{t+1} \mid \Omega_t(j)\right] dj + \eta_t^2 \tag{2.8}$$

The price of a three period bond according to (2.5) 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 (2.8) by one period and substituting into (2.5) with $n = 3$ gives

$$\begin{aligned}\widetilde{b}_t^3 &= -r_t - \int E\left[r_{t+1} \mid \Omega_t(j)\right] dj \\ &\quad - \int E\left[\int E\left[r_{t+2} \mid \Omega_{t+1}(j')\right] dj' \mid \Omega_t(j)\right] dj \\ &\quad + \eta_t^3\end{aligned} \tag{2.9}$$

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

$$\begin{aligned}\widetilde{b}_t^n &= -r_t - \int E\left[r_{t+1} \mid \Omega_t(j)\right] - \\ &\quad \int E\left[\int E\left[r_{t+2} \mid \Omega_{t+1}(j')\right] dj' \mid \Omega_t(j)\right] dj + ... \\ &\quad ... + \int E\left[\int E\left[... \int E\left[r_{t+n-1} \mid \Omega_{t+n-2}(j'')\right] dj'' ... \mid \Omega_{t+1}(j')\right] dj' \mid \Omega_t(j)\right] dj + \eta_t^n\end{aligned} \tag{2.10}$$

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

$$y_t^n = -n^{-1}\widetilde{b}_t^n \tag{2.11}$$

As has been noted before (e.g. Allen, Morris and Shin (2005)), the fact that (2.10) contains average expectations of average expectations (and so on) prevents us from applying the law of iterated expectations to solve for bond prices. Before analyzing the effects of non-nested information sets, 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 inner product space that the analysis is framed in.

2.3. **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*

$$EX^2 < \infty \tag{2.12}$$

and with inner-product

$$\langle X, Y \rangle \equiv E(XY) : X, Y \in L^2 \tag{2.13}$$

In a linear model with Gaussian shocks, conditional expectations are equivalent to orthogonal projections.

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

$$\langle X - \mathcal{P}_\Omega X, \omega \rangle = 0 \tag{2.14}$$

for any $\omega \in \Omega$.

We can use the equality

$$E(X \mid \Omega) = \mathcal{P}_\Omega X \tag{2.15}$$

to replace the conditional expectations in the bond pricing equation (2.10) 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*

$$f_t^n \equiv \widetilde{b}_t^n - \widetilde{b}_t^{n+1} \tag{2.16}$$

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

$$f_t^n = \mathcal{P}_{\Omega t} r_{t+n} \quad \forall t, n \tag{2.17}$$

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

We can use the definition (2.17) to demonstrate that in the absence of supply shocks, the bond pricing equation (2.10) implies that the expectations hypothesis hold if all agents share the same information set. Consider the 2 period ahead forward rate which by (2.10) and (2.16) is given by

$$\begin{aligned} f_t^2 &= \widetilde{b}_t^3 - \widetilde{b}_t^2 \\ &= \int E\left[\int E\left[r_{t+2} \mid \Omega_{t+1}(j')\right] dj' \mid \Omega_t(j)\right] dj \end{aligned} \tag{2.18}$$

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 taking 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 $\{E[r_{t+n} \mid \Omega_{t+s}]\}_{s=0}^{n}$ is a martingale. Applying this result to the bond pricing equation (2.18) gives

$$f_t^2 = \mathcal{P}_{\Omega t}(\mathcal{P}_{\Omega t+1} r_{t+2}) \quad (2.19)$$

$$= \mathcal{P}_{\Omega t} r_{t+2} \quad (2.20)$$

The expectation hypothesis then holds with respect to the common information set Ω_t.

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. But of course, bond yields can always be backed out from implied forward rates by the identity (2.16).

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 predictions about ones own expectations. 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 in a model with constant risk premia. 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 shown that the speculative dynamics introduced by non-nested information sets are orthogonal to public information which has implications for how the speculative dynamics can be quantified using 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*

$$\mathcal{P}_{t,j} r_{t+n} \neq \mathcal{P}_{t+s,i} r_{t+n} : j \neq i \quad (3.1)$$

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

Defining non-nested information sets through the implications for projections of short rates onto individual trader's information sets is somewhat tailored to the needs of this paper. A more general definition would simply state that information sets are non-nested if projections of *any* random variable onto individual traders' information sets differ. The definition used here is designed to avoid trivial cases where projections only differ about uninteresting quantities. We therefore define information sets as non-nested only if they imply that expectations about future short rates differ across agents.[2]

3.1. **Predictable excess returns and speculative dynamics.** We start by proving that non-nested information sets are sufficient for individual traders to be able to predict excess returns. We do this in two steps. In the first step we prove 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 future short rates.

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

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

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

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

then the ex ante symmetry of traders implies that

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

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

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, $\mathcal{P}_{t,j} r_{t+n} \neq f_t^n$ implies that $\mathcal{P}_{t,j} \left(r_{t+n} - f_t^n \right) \neq 0$, 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 defined as the difference in return between holding an n period bond until it matures and the return on holding a sequence of one period bonds over n periods, i.e. the excess return on an period bond is given by*

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

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

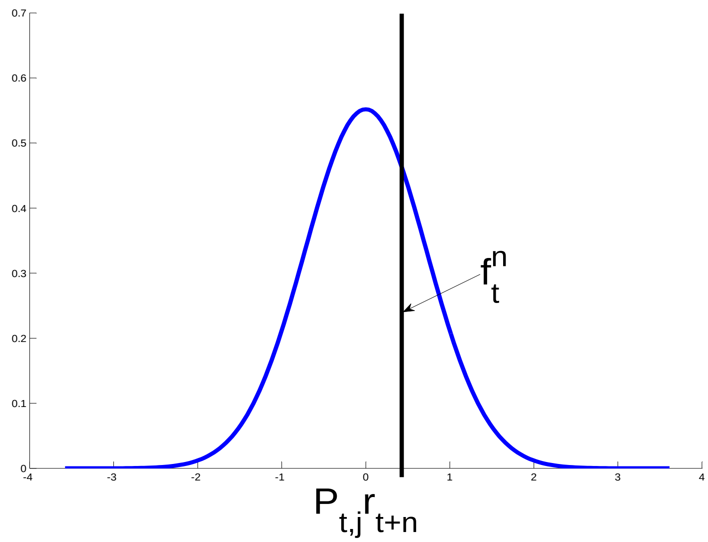

FIGURE 1. Dispersion of expectations of the short rate in period t+n and the implied forward rate.

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

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

$$\mathcal{P}_{t,j}\left(-\widetilde{b}_t^n - (r_t + r_{t+1} + ...r_{t+n-1})\right) \neq 0 \tag{3.5}$$

From the identity

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

we thus have that excess returns on an n period bond are predictable if

$$\sum_{s=1}^{n-1} \mathcal{P}_{t,j}r_{t+s} \neq \sum_{s=1}^{n-1} f_t^s \tag{3.7}$$

By Proposition 1, the inequality (3.7) must hold for at least some n. To see the last point, assume that by for some horizon n (3.7) does not hold, then by Proposition 1 we have that $\mathcal{P}_{t,j}r_{t+n+1} \neq f_t^{n+1}$ which implies that

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

which concludes the proof. □

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

Proposition 3. *Non-nested information sets imply that an individual trader j can systematically predict the average period $t+s$ projection errors of the short rate in period $t+n$, that is*

$$\mathcal{P}_{t,j}\left(r_{t+n} - \int \mathcal{P}_{t+s,j'} r_{t+n} dj'\right) \neq 0 \tag{3.9}$$

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

Proof. The expression (3.9) can be rearranged to

$$\mathcal{P}_{t,j} r_{t+n} = \mathcal{P}_{t,j} \int \mathcal{P}_{t+s,j'} r_{t+n} dj'. \tag{3.10}$$

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

$$\mathcal{P}_{t,j} \int \mathcal{P}_{t+s,j'} r_{t+n} dj' = \mathcal{P}_{t,j} \mathcal{P}_{t+s,i} r_{t+n} \text{ for all } i, j \in (0, 1) : i \neq j. \tag{3.11}$$

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

$$\mathcal{P}_{t,j} r_{t+n} = \mathcal{P}_{t,j} \mathcal{P}_{t+s,i} r_{t+n} \tag{3.12}$$

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. □

Proposition 3 shows that individual traders not only can predict excess returns, they can also systematically predict the average prediction errors made by other traders. In the next proposition, we show that the speculative dynamics introduced by non-nested information sets can be expressed as average predictions about average prediction errors of future short rates. That is, predictions about the difference between future short rates and other traders' predictions about future short rates.

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

Proof. For convenience, first define the notation

$$\prod_{s=0}^{n-1} \int \mathcal{P}_{t+s,j} r_{t+n} \equiv \int \mathcal{P}_{t,j} \int \mathcal{P}_{t+1,j'} ... \int \mathcal{P}_{t+n-1,j''} r_{t+n} dj''...dj'dj. \tag{3.13}$$

and rewrite the definition of the n period forward rate as

$$f_t^n = \prod_{s=0}^{n-1} \int \mathcal{P}_{t+s,j} r_{t+n} + \left(\eta_t^n - \eta_t^{n+1}\right). \tag{3.14}$$

Add and subtract $\int \mathcal{P}_{t,j} r_{t+n}$ from the r.h.s. of (3.14) to get

$$\begin{aligned} f_t^n &= \int \mathcal{P}_{t,j} r_{t+n} \\ &\quad - \int \mathcal{P}_{t,j} \underbrace{\left(r_{t+n} - \prod_{s=0}^{n-1} \int \mathcal{P}_{t+s,j} r_{t+n} \right)}_{n-1 \text{ order prediction error}} \\ &\quad + \left(\eta_t^n - \eta_t^{n+1}\right) \end{aligned} \tag{3.15}$$

□

The term on the second line of (3.15) 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 $t+n$. order prediction error average prediction error across traders. With $n = 2$, this term reduces to the average of the predicted error (3.9).

In a model with perfect or common information, the higher order prediction errors on the second line 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 would simply be 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 frequently. This is also the sense in which the "speculative behavior" in this model conforms to the definition of Harrison and Kreps (1978).

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 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 (**??**) as

$$\begin{aligned} f_t^n &= \mathcal{P}_t^p r_{t+n} + \int \mathcal{P}_{t,j}^{\perp p} r_{t+n} \\ &- \int \mathcal{P}_{t,j}^p \left(r_{t+n} - \prod_{s=0}^{n-1} \int \mathcal{P}_{t+s,j} r_{t+n} \right) \\ &- \int \mathcal{P}_{t,j}^{\perp p} \left(r_{t+n} - \prod_{s=0}^{n-1} \int \mathcal{P}_{t+s,j} r_{t+n} \right) \\ &+ \left(\eta_t^n - \eta_t^{n+1} \right) \end{aligned} \tag{3.16}$$

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

$$\mathcal{P}_{t,j}^p r_{t+n} = \mathcal{P}_{t,j}^p \prod_{s=0}^{n-1} \int \mathcal{P}_{t+s,j} r_{t+n} \tag{3.17}$$

for all all j and $s = 0, 1, ..., m-1$. The term on the second line of (3.16) is thus identically zero and the n period forward rate can thus be expressed as

$$\begin{aligned} f_t^n &= \mathcal{P}_t^p r_{t+n} + \int \mathcal{P}_{t,j}^{\perp p} r_{t+n} \\ &- - \int \mathcal{P}_{t,j}^{\perp p} \left(r_{t+n} - \prod_{s=0}^{n-1} \int \mathcal{P}_{t+s,j} r_{t+n} \right) \\ &+ \left(\eta_t^n - \eta_t^{n+1} \right) \end{aligned} \tag{3.18}$$

which concludes the proof. □

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

This ends the theoretical part of the paper. Before turning to the data, we can summarize our findings so far. Individual traders can identify and take advantage of predictable excess returns, even though risk premia is constant. The model thus provides an alterative explanation of predictable excess returns that is not based on agents valuing a marginal increase in wealth differently in different states of the world. We also demonstrated that a the new dynamics were orthogonal to public information. This has an interesting empirical implication: The 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 introduces new dynamics to the term structure of interest rates and can provide an alternative explanation for predictable excess returns that is not based on a time varying willingness to bear risk. 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 things: A process for the short rate and the information sets of the traders.

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

$$r_t = x_t^1 + x_t^2 + x_t^3 + \phi r_{t-1} \tag{4.1}$$

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

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

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 r_t.

4.2. **Traders' information sets.** All traders observe a vector of public signals containing the current short rate r_t and selected bond yields collected in the vector $\mathbf{y}_t$. 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

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

$$D = \begin{bmatrix} d_1 & 0 \\ 0 & d_2 \end{bmatrix} \tag{4.4}$$

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

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

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

4.3. **The solved model.** When traders have non-nested information sets it becomes optimal to "forecast the forecasts of others", and natural representations of the state in this class of models tend to be infinite.[3] The model is solved using the method proposed by Nimark (2007) which delivers a finite dimensional representation of the form

$$X_t = MX_{t-1} + N\epsilon_t \tag{4.6}$$

$$\mathbf{y}_t = B_1 X_t + B_2 r_{t-1} + \eta_t \tag{4.7}$$

where $\mathbf{y}_t$ is a vector of yields of bonds of different maturities. The matrices M, N, B_1 and B_2 are functions of the structural parameters of the model, i.e. the parameters governing the short rate (4.1), traders' information sets (4.5) and the standard deviation of the bond supply shocks η_t (assumed to be constant across maturities). The vector ϵ_t contains the aggregate shocks in the economy, i.e. the factor innovations ε_t and the bond supply shocks η_t. The state vector X_t contains stacked higher order expectations of the factors

$$X_t \equiv \left[\begin{array}{cccc} \mathbf{x}_t^{(0)\prime} & \mathbf{x}_t^{(1)\prime} & \cdots & \mathbf{x}_t^{(\overline{k})\prime} \end{array} \right]' \tag{4.8}$$

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 to the limit as $k \to \infty$. Common knowledge of the model is used to pin down the law of motion for X_t. That is, first order expectations $\mathbf{x}_t^{(1)}$ are optimal estimates of the actual factors $\mathbf{x}_t$. The knowledge that other traders have model consistent estimates allow traders to form estimates of other traders estimates of the true factors so that second order expectations $\mathbf{x}_t^{(2)}$ are optimal estimates of $\mathbf{x}_t^{(1)}$ given the law of motion for $\mathbf{x}_t^{(1)}$, and so on. The Appendix provides more details on the solution procedure.

4.4. **Posterior Estimates.** The model is in a form that can be estimated directly by likelihood based methods. The vector of parameters to be estimated is denoted $\theta \equiv \{A, C, D, \phi, \sigma_\eta\}$ and consists of a total of thirteen 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. The sample period is from January 1964 to December 2007 (528 monthly observations) and chosen to coincide with the sample period used by Cochrane and Piazzesi (2008) and Duffee (2008). The time series are demeaned. The posterior mode of θ (with uniform truncated priors) is found using a simulated annealing numerical maximizer and the posterior parameter distributions are simulated by 1 000 000 draws from an Adaptive Metropolis algorithm (see Haario, Saksman and Tamminen (2001)). The parameter estimates are reported in Table 1.

[3]See Townsend (1983), Sargent (1991) and Makarov and Rytchov (2009).

Table 1

Parameter Estimates 1964:1-2007:12

θ	Mode	2.5%-97.5%
	Short rate process	
ρ_1	0.994	0.992 - 0.995
ρ_2	0.990	0.987 - 0.994
ρ_3	0.60	0.57 -0.62
ϕ	0	0 - 0.003
c_1	1.44	1.31 - 3.71
c_2	0.65	0.62 - 0.68
c_3	0.43	0.41 - 0.46
c_{21}	-1.43	(-3.72) - (-1.30)
c_{31}	-0.77	(-0.88) - (-0.73)
c_{32}	0.25	0.23 - 0.28
	Noise in private signals	
d_1	0.005	0.003 -0.012
d_2	10	4.39- 10
	Bond supply shocks	
σ_η	0.19	0.18 - 0.20

Computed using 1 000 000 draws from a MCMC generated by an Adaptive Metropolis algorithm.

By themselves, the posterior estimates are not very interesting, but we can note that all parameters appear to be well-identified.[4] The first two factors are very persistent and traders appear to have much more precise private information about the first factor than about the second, that is $d_1 << d_2$. The posterior mode of the standard deviation of the noise in the second private signal is close to the imposed boundary of 10. However, imposing the boundary has little effect since the marginal informativeness of the second private signal is practically zero for much lower levels of noise than the actual boundary. At the other end of the boundary, i.e. zero, the likelihood drops sharply. The parameter is thus well identified in the sense the it is clear from the data that the private signal about the second facto is very imprecise. There seems to be no role for short rate inertia in the model and the parameter ϕ is estimated to be very close to zero so and the lagged short rate in (4.1) could be dispensed of without sacrificing fit. The standard deviation 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 R^2 of a VAR and it is clear that the model can explain most of the observed variance of yields.

4.5. **Historical speculation.** The previous section showed that the speculative term introduced to the term strcure by non-nested information sets coul be expresseda as a higher

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

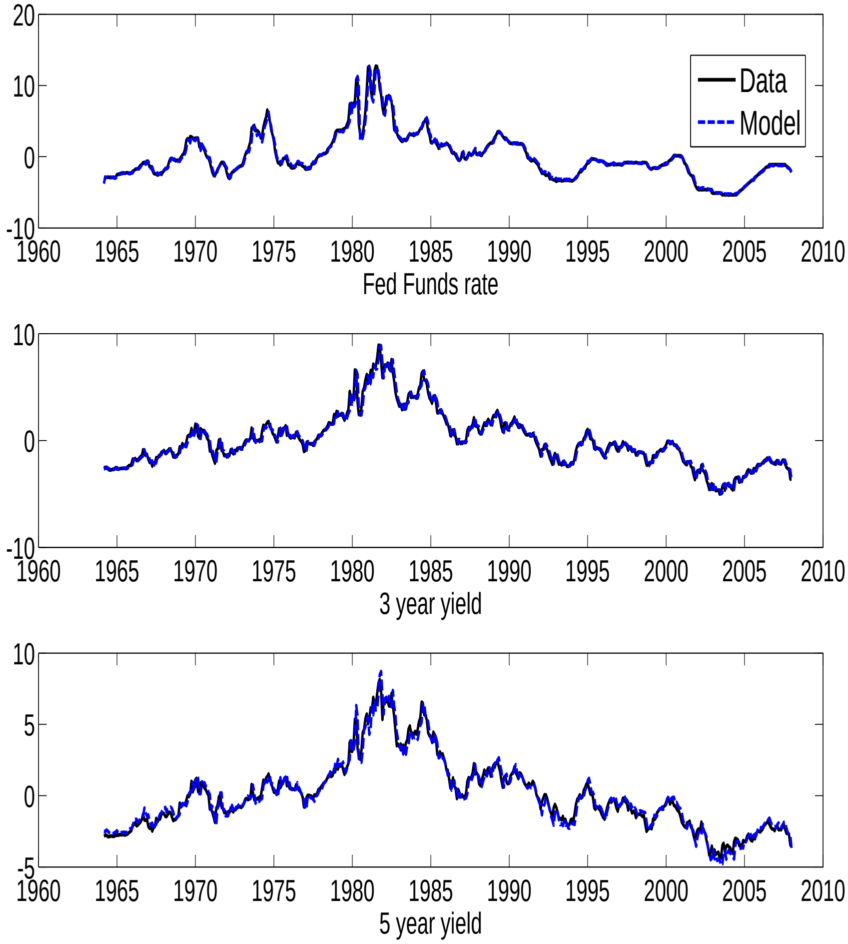

FIGURE 2. Data (black solid) and one-sided model fit (blue dashed).

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 use information for $t+s : s > 0$ to back out information about the higher order prediction error in period t. The Kalman simulation smoother (see for instance Durbin and Koopman (2002)) can be used to draw from $p\left(X^T \mid \mathbf{y}^T\right)$ for a given parameter vector θ. The simulation smoother together with the posterior distribution of θ can be used to construct a posterior distribution of the state X^T. Once we have a posterior distribution of the state, it is straight forward to compute the distribution of the speculative term in the implied n

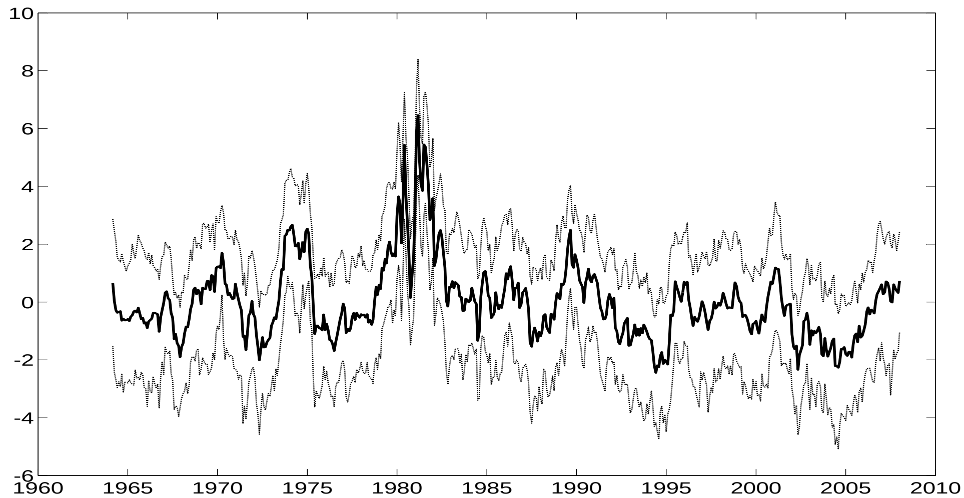

FIGURE 3. Estimated speculative term (percentage points) in 12 month implied forward rate 1964:1 - 2007:12. Median (solid) and 95% probability interval (dotted).

period forward rate. I.e. we can construct a posterior distribution of the time series

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

(More details on how the posterior state distribution can be constructed are given in the Appendix.) Figure 3 displays the median and the 95% probability interval of the speculative term in the implied 12 month forward rate, i.e. for n=12 in (4.9).

It is clear from the figure that the posterior estimates suggest that the speculative term is quantitatively important. The median reaches about 6.5 percentage points in the early 1980s. It also appears to be statistically significant with the lower bound of the 95% probability interval reaching 4 percentage points in the same period. Over the sample, there are about ten episodes when the median reaches either plus or minus two percentage points.

The episode in the early eighties is the most eye-catching in the sample. It coincides with the so-called Volcker disinflation when the then Federal Reserve chairman Paul Volcker raised interest rates sharply to bring inflation under control (e.g. 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.

4.6. **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, though there may be reason to be cautious when interpreting the reported survey expectations to be representative of the expectations of actual bond traders.

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

$$\sqrt{E\left(\mathcal{P}_{t,j}r_{t+n} - \int \mathcal{P}_{t,j'}r_{t+n}\right)^2} \tag{4.10}$$

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 somewhat 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 $0.088 \times 3.9 = 0.34$ or 34 basis points at the median. This is around half of the lower end (80 basis point) of the spread reported by Swanson. At the 97.5th percentile, the model estimate is roughly equal to the lower end of the spread, i.e. $0.226 \times 3.9 = 0.88$. The estimated dispersion of expectations seems reasonable, and should increase our confidence in the model.

Table 2
Estimated dispersion of short rate expectations

	$\sqrt{E\left(\mathcal{P}_{t,j}r_{t+n} - \int \mathcal{P}_{t,j'}r_{t+n}\right)^2}$		
n	**2.5%**	**Median**	**97.5%**
12	0.067	0.088	0.226
36	0.050	0.072	0.185
60	0.037	0.058	0.151

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

Table 2 also shows that the dispersion of expectations are decreasing with the forecast horizon, but remains significant also at the 5 year horizon. Of course, at long enough horizons the dispersion will disappear entirely, 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

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

5.1. **A three factor no-arbitrage model.** Affine three factor no-arbitrage models can provide a very good fit of the term structure of interest rates (e.g. Duffie and Kan 1996). However, if the world is characterized by bond markets where traders with non-nested information sets interact, low dimensional affine no-arbitrage models are fundamentally misspecified. (Perhaps interestingly, in the limit case as the variance of the idiosyncratic noise in traders private signals tend to zero, the model presented here becomes an affine three factor no-arbitrage constant risk premia model.) It is well-known that in dynamic models where agents have non-nested information sets, natural state representations tend to be infinite dimensional (see for instance Townsend (1983), Sargent (1991) and Makarov and Rytchov (2009)). In the estimated model of the previous section, the infinite dimensional representation was approximated with an 24 dimensional (the maximum orders of expectations of thet factors times the number of factors, i.e. $\overline{k} \times 3 = 8 \times 3 = 24$) state vector. This turns out to be sufficient to accurately represent the dynamics of the model with non-nested information, and given the structure imposed by higher order expectations by common knowledge of rationality, the model places stringent restrictions on the observable variables, in spite of th ehigh dimensional state vector. (The number of parameters are significantly fewer than in a standard three factor no-arbitrage factor model.)

In this section, we investigate what an affine three factor no-arbitrage model would find if the estiamted model of the previous section represents the true economy. We are particularly interested in finding out whether a three factor no-arbitrage model will correctly detect that the risk premia is constant in the model that generated the artificial data.

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

$$F_t = \Gamma F_{t-1} + \Psi e_t \tag{5.1}$$

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.) The short rate is a function of the factors

$$r_t = \delta' F_{t-1} \tag{5.2}$$

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

$$\widetilde{b}_t^n = B_n' F_t \tag{5.3}$$

where

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

As before, yields can be computed as

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

Imposing that the risk premia is constant equals setting $\lambda = \mathbf{0}$.

The experiment we conduct is the following. We first draw parameters from the posterior distribution of our model and for each parameter vector θ we generate 528 observations. We then estimate the three factor no-arbitrage model described above with added yield measurement errors of constant variance across maturities, a total of 18 parameters. For each sample of artificial data we also re-estimate the model imposing the restriction that

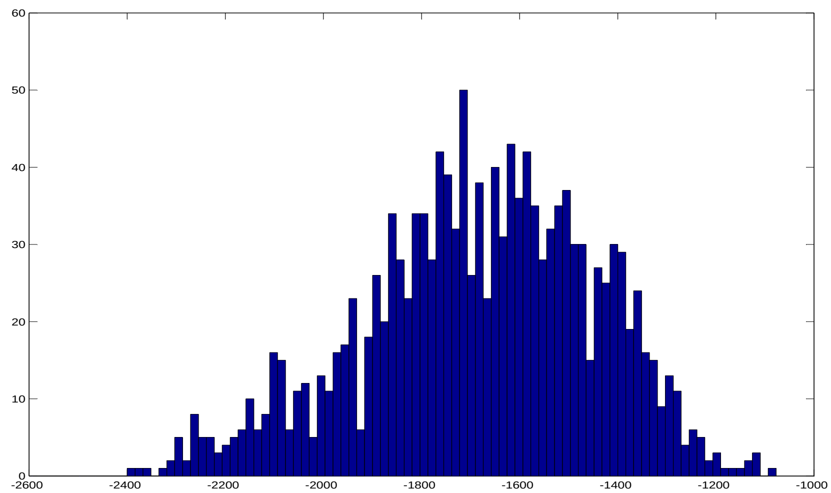

FIGURE 4. Distribution of log likelihood ratios of restricted and unrestricted three factor no-arbitrage model estimated on artificial data.

the risk premia is constant. We then compare the marginal likelihoods of the restricted and unrestricted model. This procedure was repeated 1400 times.[5]

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

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

$$x_t^\dagger = D^\dagger x_{t-1}^\dagger + \Sigma^\dagger \epsilon_t \tag{5.6}$$

$$y_t = A + B^\dagger x_t^\dagger + \eta_t^\dagger \tag{5.7}$$

[5]The procedure is very time consuming and the 1400 repetitions took about 3 1/2 weeks to compute on a not unusually slow computer.

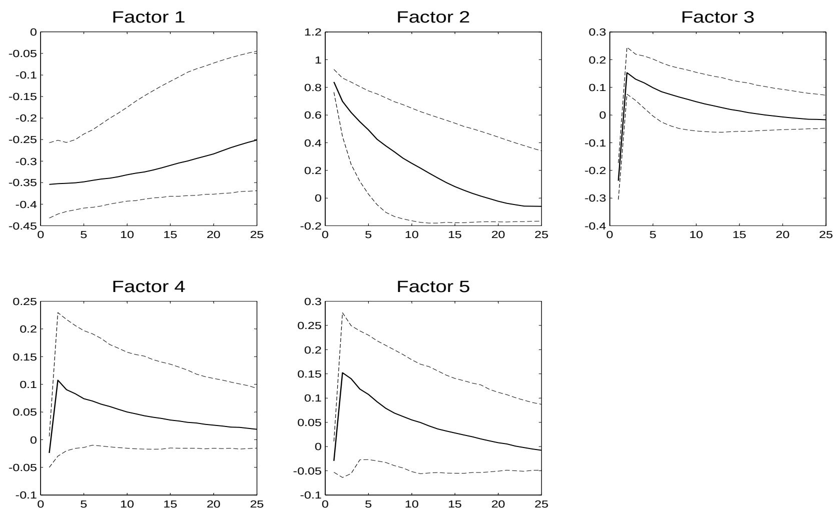

FIGURE 5. Impulse responses to orthogonal factor innovations, dashed lines are 5th and 95th percentile.

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

We investigated whether the hidden factor found by Duffee is consistent with the model presented here by again generating artificial data sets using parameter draws from the estimated posterior of our model with non-nested information sets. For each parameter draw, we simulated 528 months of data and then estimated Duffee's five factor model and performed the rotations to compute the principal components. We then computed impulse responses of short rates to orthogonal innovations to the factors. This procedure was repeated 500 times. Figure 4 shows the median impulse response and the 5th and 95th percentile.[6] As we can see, the fourth and fifth factor have little effect on the short rate in the impact period, but becomes more important at longer time horizons. This is exactly what we should expect from a model where the term structure do not reveal all information about future short rates perfectly. If the state of the model would be revealed perfectly by the cross section of yields, no additional factor beyond the three (level, slope and curvature) that explains the cross

[6]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 Baysian posterior simulation for each draw of artificial data is too time consuming to be feasible.

sections would be useful to predict future yields. However, if the state is not revealed by the cross section, then by definition there must be additional factors that can help predict future yields.

6. Conclusions

In this paper we have showed that introducing private information in a simple model of bond pricing can give rise to speculative behaviour 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, that is, 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 provides a new explanation for predictable excess returns that is not based on the value traders attach to a marginal increase in wealth in different states of the world.

Of course, there exist a vast literature on the term structure of interest rates where predictable excess returns are driven by time varying risk premia and one may ask why a new explanation is needed. For the purposes of this discussion, one can broadly classify existing papers according to whether they quantify or interpret time varying risk premia, with the vast majority of papers falling in the former class. The quantifying class of papers has mainly been concerned with describing the statistical properties of risk premia accurately, either in relation to observed variables, e.g. Litterman and Sheinkman (1991) and Ang and Piazzesi (2003). Some papers are explicit in their purpose of decomposing yields into information about future short rates and information about risk premia, such as the papers by Cochrane and Piazzesi (2005) and Backus and Wright (2007). These papers frame their discussions in the context of statistical models, often with no-arbitrage as the only economic restriction imposed. These models provide a very good fit of the term structure.

Papers attempting to provide an economic explanation of predictable excess returns generally need to impose more structure by populating their models with a representative agent and assuming a specific functional form for the agent's utility function, e.g. Backus, Gregory and Zin (1989), and Wachter (2006). In these models, predictable excess returns are driven by variations in the expected marginal utility of consumption in different states of the world. The cost of more structure is substantially worse fit, and it has proven hard to connect variation in the risk premia with the macro economic aggregates suggested by economic theory. The paper by Wachter is probably the most successful to date, in that it manages to construct a model that fit the broad features of the term structure, e.g. the correlation with the business cycle and the average slope. However, a lot of variation in the term structure is left unexplained.

Here, we have taken a step back and proposed a different theory of what drives predictable excess returns in a model that in some sense has more in common with the purely statistical factor models than with the representative agent macro finance models. In the model, no attempt is made to link predictable excess returns to anything else than the underlying unobserved factors. Still, the model suggest a quite different interpretation of the observed history of bond prices. Instead of being driven by variations in expected interest rates and the willingness to bear risk, we showed here that term structure dynamics can potentially be explained by speculative trade. That is, trade driven not by perceptions about the fundamental value of bonds, but trade driven by the value traders expect other traders to attach to the bond in the future.

The quantitative importance of speculative trade was assessed by estimating the model using likelihood based methods. While theoretical models of private information in asset markets go back to at least Grossman (1976), this paper is to the author's knowledge the first to take such model directly to the data. Given the prevalence in the literature of citations of Keynes' "Beauty Contest" metaphor to describe the stock market (e.g. Allen, Morris and Shin (2005)and Walker (2007)), many find it plausible that trade in assets are at least partly driven by speculative motives. The estimated model suggest that it may also be quantitatively important with the speculative term making up a large fraction of the variance of implied forward rates.

We have also shown that the model presented here can explain some features of statistical factor models. Specifically, our model generates data that when used as a sample for the 5 factor model of Duffee (2008) results in estimates that reproduce the "hidden" factor documented by Duffee. We also demonstrated that a three factor no-arbitrage model may mistakenly attribute the dynamics due to non-nested information to time varying risk premia.

References

[1] Adam, Klaus, (2007), "Optimal Monetary Policy with Imperfect Common Knowledge", *Journal of Monetary Economics*, vol. 54(2), pp 267-301.

[2] Allen, F. S. Morris and H.S., (2006), "Shin, 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, (2007), "Policy with Dispersed Information," NBER Working Papers 13590

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

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

[7] Backus, David & Foresi, Silverio & Mozumdar, Abon & Wu, Liuren, 2001. "Predictable changes in yields and forward rates," *Journal of Financial Economics*, vol. 59(3), pp281-311.

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

[9] Bekaert, G., Hodrick, R. and D.A. Marshall, (1997), "On biases in tests of the expectations hypothesis of the term structure of interest rates" *Journal of Financial Economics*, vol. 44(3), pp309-348.

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

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

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

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

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

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

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

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

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

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

[20] 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.

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

[22] 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, 425-436.

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

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

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

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

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

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

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

[30] Morris, S. and H.S. Shin, (2002), "The social value of public information", *American Economic Review* 92, pp1521-1534.

[31] Nimark, K., (2008), "Dynamic Pricing and Imperfect Common Knowledge", *Journal of Monetary Economics,* 55, pp365-382.

[32] Nimark, K., (2007), "Dynamic Higher Order Expectations", working paper, Universitat Pompeu Fabra.

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

[34] Piazzesi, M. and M. Schneider, (2006), "Equilibrium Yield Curves", NBER Working Papers 12609.

[35] Rudebusch, G.D.and J. C. Williams, (2006), "Revealing the Secrets of the Temple: The Value of Publishing Central Bank Interest Rate Projections" NBER Working Papers 12638.

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

[37] Shiller, R. J.Y. Campbell and K.L. Shoenholtz, (1983), "Forward Rates and Future Policy: Interpeting the Term Structure of Interest Rates", *Brookings Papers on Economic Activity*, pp173-223.

[38] Swanson, E.T., (2006), "Have Increases in Federal Reserve Transparency Improved Private Sector Interest Rate Forecasts?", *Journal of Money, Credit, and Banking* 38(3), April 2006, pp. 791-819.

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

[40] 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.

[41] Xiong, W. and H. Yan, (2008),"Heterogenous Expectations and Bond Markets", mimeo, Princeton University.

Appendix A. Some useful properties of projections

This Appendix reproduces some results and properties of orthogonal projections on inner-product spaces (as in definition x and x). These are used for the analytical results of Section 3 in the main text above. Proofs and more details can be found in for instance Brockwell and Davis (2006).Orthogonal projections

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

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

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 $\omega \in \Omega$.

(1) The projection $\mathcal{P}_\Omega X$ coincides with the conditional expectation $E[X \mid \Omega]$ 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 3 where we decompose bond prices into a component that is the projection of future short rates on public information and into a component that is orthogonal to public information. Property (3) can be used to show that individuals can predict average expectations errors when information sets are non-nested. Finally, Property (4) can be used to show both that in the absence of supply shocks the expectations hypothesis holds in our model with respect to a public information set and that individual traders can predict excess return when information sets are non-nested.

Appendix B. Solving the model

The model is solved by using the method proposed in Nimark (2007). It involves deriving an explicit law of motion for higher order expectation of the exogenous processes x_t^1, x_t^2 and x_t^3. Define the exogenous state vector $\mathbf{x}_t$ as

$$\mathbf{x}_t = \begin{bmatrix} x_t^1 & x_t^2 & x_t^3 \end{bmatrix}' \tag{B.1}$$

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

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

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

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

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

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

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

$$H = \begin{bmatrix} \mathbf{0}_{2\overline{k}\times 2} & \begin{matrix} I_{\overline{k}-2} \\ \mathbf{0}_{2\times(\overline{k}-2)} \end{matrix} \end{bmatrix} \tag{B.5}$$

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

$$\overline{M} = (MH) \tag{B.6}$$

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

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

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

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

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

$$S_t(j) = \begin{bmatrix} \mathbf{z}_t'(j) & r_t & \mathbf{y}_t' \end{bmatrix}' \tag{B.9}$$

$$= L_1 X_t + L_2 r_{t-1} + Q \begin{bmatrix} \zeta_t \\ \eta_t \end{bmatrix} \tag{B.10}$$

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

$$X_{t|t}(j) = MX_{t|t-1}(j) + K\left(S_t(j) - L_1 M X_{t|t-1}(j)\right) \tag{B.11}$$

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

$$M = \begin{bmatrix} \rho & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + \begin{bmatrix} \mathbf{0}_{2\times 2} & \mathbf{0} \\ \mathbf{0} & [M - KL_1M]_{11} \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ [KL_1M]_{11} \end{bmatrix} \tag{B.12}$$

$$N = \begin{bmatrix} \mathbf{1}_{1\times 2} & 0 \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ [KL_1N]_{11} \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & K_3 \end{bmatrix} \tag{B.13}$$

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

$$K = (PL_1' + NQ')(L_1PL_1' + QQ')^{-1} \tag{B.14}$$

$$P = M\left(P - (PL_1' + NQ')(L_1PL_1' + QQ')^{-1}(PL_1' + NQ')'\right)M' + NN' \tag{B.15}$$

The model is solved by finding the fixed point that satisfies (B.13) and (B.14).