

# High Dimensional Yield Curves: Models and Forecasting

Clive G. Bowsher\* and Roland Meeks  
Nuffield College, University of Oxford, Oxford, OX1 1NF, U.K.  
`clive.bowsher@nuffield.ox.ac.uk`  
`roland.meeks@nuffield.ox.ac.uk`

October 2, 2006

## Abstract

Functional Signal plus Noise (FSN) models are proposed for analysing the dynamics of a large cross-section of yields or asset prices in which contemporaneous observations are functionally related. The FSN models are used to forecast high dimensional yield curves for US Treasury bonds at the one month ahead horizon. The models achieve large reductions in mean square forecast errors relative to a random walk for yields and readily dominate both the Diebold and Li (2006) and random walk forecasts across all maturities studied. We show that the Expectations Theory (ET) of the term structure completely determines the conditional mean of any zero-coupon yield curve. This enables a novel evaluation of the ET in which its 1-step ahead forecasts are compared with those of rival methods such as the FSN models, with the results strongly supporting the growing body of empirical evidence against the ET. Yield spreads do provide important information for forecasting the yield curve, especially in the case of shorter maturities, but not in the manner prescribed by the Expectations Theory.

*Keywords:* Yield curve, term structure, expectations theory, FSN models, functional time series, forecasting, state space form, cubic spline.

*JEL classification:* C33, C51, C53, E47, G12.

## 1 Introduction

In this paper we develop a novel econometric framework for modelling and forecasting high dimensional yield curves consisting of the yields on a large number of discount bonds with different maturities. Forecasting one month ahead, we are able to improve upon the forecasting performance of all other existing methods, including a new method introduced here in which the forecasts implement the conditional mean implied by the expectations theory of the term structure. Two obvious lacunae exist in the term structure literature: the vast majority of empirical dynamic studies concentrate on a small subset of maturities rather than entire term structures; and there is a relative paucity of work concerned primarily with forecasting the yield

---

\*Corresponding author: *tel:* +44 1865 278969, *fax:* +44 1865 278621. Comments are welcome and should be directed to the above e-mail address. Functional Signal plus Noise (FSN) models were first introduced in the earlier working paper, Bowsher (2004).

curve. The first stems from the econometric difficulties involved in modelling the dynamics of a high dimensional cross-section of interest rates that are functionally related to one another. The second exists despite the considerable importance of forecasting the term structure for bond portfolio management, derivatives pricing and monetary policy. The methods introduced here address both areas and can also readily be applied to time series of other economically important functions such as the supply and demand curves of the limit order book of a financial exchange (see Bowsher 2004).

The main contributions of the paper may be summarised as follows. First, *Functional Signal plus Noise (FSN)* time series models are introduced for analysing the dynamics of a large cross-section of yields or asset prices in which contemporaneous observations are functionally related. The FSN models specify the evolution over time of stochastic functions, a problem that has received relatively little attention in the econometrics and statistics literature, and the models may conveniently be written in linear state space form. Second, we show that the Expectations Theory (ET) of the term structure completely determines the conditional mean of any zero-coupon yield curve, given an information set that includes the currently observed, complete yield curve of adequate dimension. We are thus able to derive and implement the minimal mean square error point forecasts implied by the theory. Third, FSN models are used to forecast high-dimensional yield curves for US Treasury bonds at the 1 month ahead horizon and their performance compared in an out-of-sample experiment to the ET forecasts and the Diebold and Li (2006) dynamic Nelson-Siegel model. The preferred FSN models achieve large reductions in mean square forecast error (MSFE) relative to a random walk for the yield curve, especially at the shorter maturity end, and readily dominate the ET, Diebold and Li (2006), and random walk forecasts in terms of MSFE across all maturities studied. A novel and particularly direct evaluation of the ET is thus also provided in which its 1-step ahead forecasts are compared with those of rival models. Our results, obtained using two different datasets, strongly support the growing body of empirical evidence against the ET.

In the Functional Signal plus Noise models developed here, the information about the functional, cross-sectional relationship between contemporaneously observed yields is captured by modelling the observed curves as the sum of a smooth ‘signal function’ or latent yield curve,  $\mathcal{S}(\tau)$ , and noise. The signal function used is a cubic spline uniquely determined by the yields

that correspond to the knots of the spline, the *knot-yields*. The state equation of the FSN model then determines the stochastic evolution of the spline function by specifying that the knot-yields follow a vector autoregression which may be written as an equilibrium correction model (ECM) in which the spreads between the knot-yields appear as regressors. We find that yield spreads provide important information for forecasting the yield curve, especially in the case of shorter maturities, but not in the manner prescribed by the ET.

The work presented is also a contribution to applied *functional time series* analysis. Each yield curve may be regarded as a finite dimensional vector, albeit of very high dimension. However, the standard approaches of multivariate time series or panel data econometrics are of little help in this setting, owing to the high dimensionality of the curves and the close, functional relationship between the yields. Indeed, the analysis of time series of stochastic functions is in a state of relative infancy.<sup>1</sup> Previous work in the statistics literature includes the use of Functional AutoRegressive (FAR) models for forecasting entire smooth, continuous functions by Besse and Cardot (1996) and Besse, Cardot, and Stephenson (2000). The FSN models may be interpreted as a special type of dynamic factor model (see Stock and Watson 2006, pp. 524) in which the knot-yields are the factors and the factor loadings are determined by the requirement that the latent yield curve (i.e. the vector of ‘common components’) be a natural cubic spline function, rather than the factor loadings being parameters for estimation. The semiparametric FSN approach thus allows quasi-maximum likelihood estimation using the state space form and the Kalman filter even when the cross-sectional dimension of the data is very large.

Two previous studies examine out-of-sample forecasting of high dimensional yield curve functions. Diebold and Li (2006) introduce a dynamic version of the Nelson and Siegel (1987) yield curve in which the three parameters or factors describing the curve each follow an AR(1) process. They report only comparable forecasting performance to that of a random walk for the yield curve (denoted  $RW^{YC}$ ) in terms of MSFEs at the 1 month ahead forecast horizon, but substantial improvement in MSFEs over the  $RW^{YC}$  and other benchmark forecasting models at a horizon of 12 months. Kargin and Onatski (2004) forecast approximate forward rate curves (derived from Eurodollar futures rates) 1 year ahead using a functional autoregression and their new predictive factor decomposition. They uniformly outperform the MSFEs of the Diebold and

---

<sup>1</sup>*Functional Data Analysis* by Ramsay and Silverman (1997) is a landmark in this area but does not consider the time series case in which the functions are dependent.

Li (2006) and  $RW^{YC}$  procedures across all maturities, but report small reductions in MSFEs relative to the forecast based on the average yield curve observed to date.

The two studies reflect a common theme in the term structure literature, namely the difficulty of outperforming naïve forecasting devices, particularly the  $RW^{YC}$  or ‘no change’ forecast. Duffee (2002) documents that forecasts made using the standard class of (‘completely’) affine term structure models typically perform worse than the  $RW^{YC}$  at horizons of 3, 6 and 12 months ahead. His ‘essentially affine’ models produce forecasts somewhat better than the  $RW^{YC}$  for the three maturities reported and at all 3 horizons. Ang and Piazzesi (2003) present a VAR model in the affine class for yields of 5 different maturities which imposes no-arbitrage restrictions and performs slightly better than the  $RW^{YC}$  at the 1 month ahead forecast horizon for 4 of the 5 maturities used. Incorporating macroeconomic factors in the model further improves the forecast performance for those 4 maturities. Finally, Swanson and White (1995) show that the premium of the forward rate over the 1 month short rate, and lags thereof, can be used to improve the 1-step ahead MSFE of the 1 month rate relative to a random walk with drift.

For the first time in the literature, we present forecasting models whose 1 month ahead forecasts strongly outperform a random walk for the yield curve. Furthermore, the models are for high dimensional yield curves rather than for a small subset of maturities. The structure of the paper is as follows. Section 2 develops two new methods for modelling and forecasting high dimensional yield curves – FSN models and forecasts based on the Expectations Theory of the term structure. Theorem 1 contains our result on the conditional expectation of yield curves under the ET. A procedure for evaluating the ET based on Theorem 1 is also proposed and previous empirical work in this area reviewed. Section 3 applies both methods to the task of forecasting high dimensional US Treasury zero-coupon yield curves, whilst Section 4 focuses on forecasting a complete set of US Treasury bill yields. Section 5 then concludes. The Appendix provides the necessary mathematical details on cubic spline functions. These are denoted here, as a function of maturity  $\tau$ , by  $\mathcal{S}(\tau)$ .

## 2 Models for High Dimensional Yield Curves

A zero-coupon or discount bond with face value \$1 and *maturity*  $\tau$  is a security that makes a single payment of \$1  $\tau$  periods from today. Its *yield to maturity*,  $y_t(\tau)$ , is defined as the per

period, continuously compounded return obtained by holding the bond from time  $t$  to  $t + \tau$ , so that

$$y_t(\tau) = -\tau^{-1}p_t(\tau), \quad (1)$$

where  $p_t(\tau)$  is the log price of the bond at  $t$ . The (zero-coupon) *yield curve* consists of the yields on discount bonds of different maturities and is denoted generically here by the vector  $y_t(\boldsymbol{\tau}) := (y_t(\tau_1), y_t(\tau_2), \dots, y_t(\tau_N))'$ . The purpose of the current paper is to develop forecasting methods for the empirically relevant case where the cross-sectional dimension  $N$  of the observed yield curve is large. In Section 3 the task will be to forecast one month ahead a  $36 \times 1$  yield curve with  $\boldsymbol{\tau} = (1.5, 2, 3, \dots, 11, 12, 15, 18, \dots, 81, 84)$  and where maturities are measured in months. A 3-dimensional plot of the dataset used there may usefully be previewed at this stage by examining the first panel of Figure 2. The plot strongly suggests the suitability of treating each observed yield curve as a smooth function perturbed by noise. Taking the contemporaneous pairwise correlations of yields for adjacent maturities using the dataset shown there gives correlations that all lie between 0.9964 and 0.9997.

In this section we present two new approaches to such a forecasting problem. First we describe the flexible, semiparametric family of Functional Signal plus Noise models. Second, we ask what the Expectations Theory of the term structure implies about forecasting yields when a history of complete yield curves is available upon which to base the forecasts. An  $n$ -dimensional yield curve is said to be *complete* here when yields are observed for all maturities  $\tau \in \{1, 2, \dots, n\}$ . In the dataset of Section 3, complete yield curves are available with  $n = 85$ . It turns out that the ET completely determines the conditional mean of any yield curve  $y_t(\boldsymbol{\tau})$ , no matter how large its dimension, given an information set that includes the current, complete  $(\tau_N + 1)$ -dimensional yield curve. In a comparison of the forecast performance of the FSN model and ET forecasts in Section 3, we find that the FSN models prove very effective in forecasting high dimensional yield curves, whilst the conditional mean implied by the ET does not give the MSFE-minimising point forecasts of the yields.

## 2.1 Functional Signal plus Noise Models

In the semiparametric Functional Signal plus Noise models proposed here, the information about the functional, cross-sectional relationship between contemporaneously observed yields is cap-

tured by modelling the observed curves as the sum of a smooth latent ‘signal function’,  $\mathcal{S}(\boldsymbol{\tau})$ , and noise. The exposition below is given in terms of yield curves. However, the FSN models can also readily be applied to time series of other economically important functions such as the supply and demand curves of an electronic limit order book (see Bowsher 2004).

The signal function used is a cubic spline uniquely determined by the yields that correspond to the knots of the spline, the knot-yields. The state equation of the FSN model then determines the stochastic evolution of the spline function by specifying that the knot-yields follow a vector autoregression which may be written as an equilibrium correction model (ECM) in which the spreads between the knot-yields appear as regressors. It is shown how to write the FSN-ECM model in linear state space form, thus allowing the use of the Kalman filter to compute both the Gaussian quasi-likelihood function and one-step ahead point predictions.<sup>2</sup> Harvey and Koopman (1993) were the first to describe a linear state space model with a state equation determining the stochastic evolution of a cubic spline function. There, as in Koopman and Ooms (2001), the stochastic spline is used to model the latent, time-varying periodic pattern of a scalar time series. This contrasts the present work in which the stochastic splines are used as a tool in functional time series analysis and assume the role of smooth approximations to the observed functional data.

### 2.1.1 FSN-ECM models

The FSN-ECM model consists of a dynamically evolving, natural cubic spline signal function denoted by  $S_{\gamma_t}(\boldsymbol{\tau})$ , plus a noise process. A cubic spline is essentially a piecewise cubic function with pieces that join together to form a smooth function overall.<sup>3</sup> The spline signal function or latent yield curve,  $S_{\gamma_t}(\boldsymbol{\tau}) := (S_{\gamma_t}(\tau_1), \dots, S_{\gamma_t}(\tau_N))'$ , has  $m$  knots positioned at the maturities  $k = (1, k_2, \dots, k_m)$ , which are deterministic and fixed over time. The notation  $S_{\gamma_t}(\boldsymbol{\tau})$  is used to imply that the spline interpolates to the latent yields  $\gamma_t = (\gamma_{1t}, \dots, \gamma_{mt})'$  – i.e.  $S_{\gamma_t}(k_j) = \gamma_{jt}$  for  $j = 1, \dots, m$ . We refer to the vector  $\gamma_t$  as the *knot yields* of the spline. An illustrative spline signal function is shown in Figure 1. Another terminology is to refer to  $S_{\gamma_t}(\boldsymbol{\tau})$  as a natural cubic spline on  $(k; \gamma_t)$ , since the spline passes through the points  $(k_j, \gamma_{jt})_{j=1}^m$ , which together

---

<sup>2</sup>The term *Kalman filter* should always be taken here to refer to the recursions as they are conveniently stated in Koopman, Shephard, and Doornik (1999, Section 4.3, pp. 122-123). For a textbook exposition of the Kalman filtering procedure, see Harvey (1989, Ch. 3).

<sup>3</sup>The essential aspects of cubic spline theory and related definitions needed here are described in the Appendix.

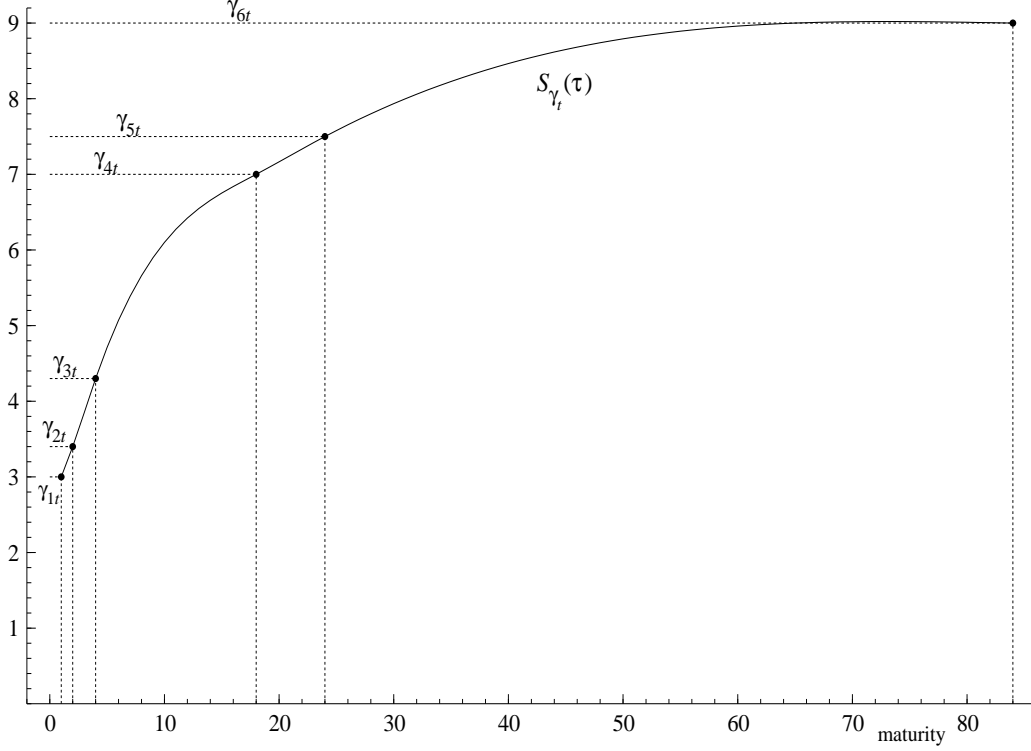


Figure 1: An illustrative spline signal function or latent yield curve,  $S_{\gamma_t}(\tau)$ . The yields at the knots,  $\gamma_{jt}$ , are labelled and plotted using filled circles; the knots are given by  $k = (1, 2, 4, 18, 24, 84)$ .

determine the remainder of the spline function uniquely.

A formal definition of the FSN-ECM model now follows.

**Definition 1 FSN-ECM Model.** *The model for the time series of  $N$ -dimensional observed yield curves,  $\{y_t(\tau)\}$ , is given by*

$$\begin{aligned} y_t(\tau) &= S_{\gamma_t}(\tau) + \epsilon_t \\ &= W(k, \tau)\gamma_t + \epsilon_t, \end{aligned} \quad (2)$$

$$\Delta\gamma_{t+1} = \alpha(\beta'\gamma_t - \mu_s) + \Psi\Delta\gamma_t + \nu_t, \quad (3)$$

for  $t = 1, \dots, T$ . Here  $S_{\gamma_t}(\tau)$  is a natural cubic spline on  $(k; \gamma_t)$ , the  $N \times m$  deterministic matrix  $W(k, \tau)$  is given by Theorem 2 of the Appendix, the  $m \times (m - 1)$  matrix  $\alpha$  has full rank, and the matrix  $\beta$  is defined uniquely by  $\beta'\gamma_t = (\gamma_{j+1,t} - \gamma_{jt})_{j=1}^{m-1}$ . The initial state  $(\gamma'_1, \gamma'_0)'$  has finite first and second moments given by  $\mu^*$  and  $\Omega^*$  respectively. The series  $\{u_t := (\epsilon'_t, \nu'_t)'\}$  has a finite second moment for all  $t$  and satisfies, for all  $t$ ,  $E[u_t] = 0$ ,  $E[\epsilon_t \epsilon'_t] = \Omega_\epsilon$ ,  $E[\nu_t \nu'_t] = \Omega_\nu$ ,  $E[\epsilon_t \nu'_t] = 0$ ,  $E[u_t u'_s] = 0 \forall s \neq t$ , and  $E[u_t(\gamma'_1, \gamma'_0)] = 0$ . Note that  $\{u_t\}$  is a vector white noise process.

The **Gaussian FSN-ECM Model** is the FSN-ECM model with the additional condition imposed that both  $u_t$  and  $(\gamma'_1, \gamma'_0)'$  have multivariate Normal distributions.

The state equation (3) is motivated by the cointegration-based yield curve literature discussed in Section 2.2.1 and is consistent with the case where the knot yields  $\gamma_t$  are I(1) and the  $(m - 1)$  spreads between them are cointegrating relations.<sup>4</sup> In that case,  $E[\Delta\gamma_{t+1}] = 0$  which excludes deterministic trends, and  $\mu_s = E[\beta'\gamma_t]$ , the stationary mean of the spreads. Furthermore the  $N$ -dimensional latent yield curve,  $S_{\gamma_t}(\boldsymbol{\tau})$ , is in that case I(1) and itself has  $(N - 1)$  linearly independent, stationary yield spreads which are cointegrating relations.

Under the conditions of Definition 1, the FSN-ECM model can be written in linear state space form, as defined by Harvey (1989, pp. 100-104). It is important to note that the deterministic matrix  $W(k, \boldsymbol{\tau})$  depends only on the vector of maturities  $\boldsymbol{\tau}$  and the knot positions  $k$  thus allowing the spline signal function  $S_{\gamma_t}(\boldsymbol{\tau})$  to be written as the linear function,  $W(k, \boldsymbol{\tau})\gamma_t$ , of the knot yields. The state vector at  $t$  can be taken to be  $(\gamma'_t, \gamma'_{t-1})'$  or, in an isomorphic representation, to be  $(\gamma_{1t}, (\beta'\gamma_t)', \gamma_{1,t-1}, (\beta'\gamma_{t-1})')'$  – see also equation (10). We use the latter in our computational work. The choice of a cubic spline as the signal function in FSN models has two advantageous features. First, the stochastic evolution over time of the signal function is determined completely by the time series properties of the  $m$ -dimensional vector  $\gamma_t$ , where  $m$  is relatively small (e.g.  $m = 5$ ), thus allowing the construction of a parsimonious model when the dimension of the yield curve is very much larger. Second, since the model has the linear state space form, the Kalman filter may be used to perform both quasi-maximum likelihood estimation (QMLE) and 1-step ahead, linear point prediction. These two features, together with the flexibility of cubic splines as approximating functions, make the cubic spline framework adopted here a particularly attractive one.

In Sections 3 and 4, the parameters of the various FSN-ECM models are estimated by maximising the likelihood of the corresponding Gaussian FSN-ECM model, which may be computed using the Kalman filter and widely available software for state space time series models. This procedure gives the QMLEs for the parameters. The FSN-ECM forecasts for the models used in Sections 3 and 4 are the 1-step ahead point predictions given by the Kalman filter,  $[\hat{y}_t(\boldsymbol{\tau})|$

---

<sup>4</sup>However, we prefer a broader definition of the FSN-ECM model and thus do not impose additional conditions on the roots of the characteristic polynomial of the VAR in (3) and on  $\det[\alpha'_\perp(I - \Psi)\beta_\perp]$ .



$y_{t-1}(\boldsymbol{\tau}), \dots, y_1(\boldsymbol{\tau}); \theta]_{KF}$ , with the parameter vector of the model set equal to some estimated value,  $\theta$ . Note that  $[\hat{y}_t(\boldsymbol{\tau}) | y_{t-1}(\boldsymbol{\tau}), \dots, y_1(\boldsymbol{\tau}); \theta]_{KF}$  is a linear function of the past observations  $(y_{t-1}(\boldsymbol{\tau}), \dots, y_1(\boldsymbol{\tau}))$  and has minimum MSFE amongst the class of such linear predictors when  $\theta$  is equal to the true parameter vector.<sup>5</sup>

## 2.2 Expectations Theory Forecasts

Complete  $n$ -dimensional yield curves are denoted here by  $y_t(1 : n) := (y_t(1), y_t(2), \dots, y_t(n))'$ . The vector of spreads between the yields and the short rate is written as  $s_{nt} := (s_t(2, 1), \dots, s_t(n, 1))'$ , where  $s_t(\tau_j, \tau_i)$  is defined as the spread  $y_t(\tau_j) - y_t(\tau_i)$ . The Expectations Theory states that a longer-term  $\tau$ -period yield differs only by a time-invariant constant from the conditionally expected per period log return obtained by successively rolling over 1-period discount bonds for  $\tau$  periods.<sup>6</sup> A formal definition is as follows.

**Definition 2** *The **Expectations Theory (ET)** is the statement that*

$$y_t(\tau) = \left\{ \tau^{-1} \sum_{i=0}^{\tau-1} \mathbb{E}[y_{t+i}(1) | \mathcal{F}_t] \right\} + \rho(\tau), \quad \tau = 1, 2, \dots, \quad (4)$$

where the constants  $\rho(\tau) \in \mathbb{R}$  are known as term premia,  $\rho(1) = 0$ , and  $\{\mathcal{F}_t\}$  denotes the filtration of publicly available information which includes the natural filtration of the yield curve.

Note that in Definition 2 the physical units of time used to measure bond maturity and the time interval between observations are necessarily the same.

Theorem 1 below states that the ET fully determines the conditional expectation of the  $(n - 1)$ -dimensional yield curve,  $y_{t+1}(1 : n - 1)$ , given any information set that includes the current  $n$ -dimensional yield curve,  $y_t(1 : n)$ . Furthermore, that conditional expectation is an affine function of the current spread vector,  $s_{nt}$ . The key result is that

$$\mathbb{E}[\Delta y_{t+1}(\tau) | \mathcal{F}_t] = \frac{\tau + 1}{\tau} \{s_t(\tau + 1, 1) - \rho(\tau + 1)\} - \{s_t(\tau, 1) - \rho(\tau)\}, \quad \tau = 1, 2, \dots \quad (5)$$

which in matrix form yields the following theorem.

<sup>5</sup>Recall that when  $\theta$  is equal to the truth,  $[\hat{y}_t(\boldsymbol{\tau}) | y_{t-1}(\boldsymbol{\tau}), \dots, y_1(\boldsymbol{\tau}); \theta]_{KF}$  is only guaranteed to equal  $\mathbb{E}_\theta[y_t(\boldsymbol{\tau}) | y_{t-1}(\boldsymbol{\tau}), \dots, y_1(\boldsymbol{\tau})]$  for the Gaussian FSN-ECM model.

<sup>6</sup>Campbell, Lo, and MacKinlay (1997, Ch.10) provide a useful summary of different forms of the ET. We follow the majority of the empirical literature by examining the logarithmic form, rather than working with bond prices in dollars and the associated gross yields to maturity.

**Theorem 1 (Conditional Expectation of Yield Curve)** Let  $n \geq 2$  and suppose that the Expectations Theory (Definition 2) is satisfied. Then,

$$y_{t+1}(1 : n - 1) = y_t(1 : n - 1) + \bar{\alpha}_{n-1}^{ET}(s_{nt} - \rho_n) + \nu_{t+1}, \quad (6)$$

where  $E[\nu_{t+1}|\mathcal{F}_t] = 0$  and  $\rho_n = (\rho(2), \dots, \rho(n))'$ . The  $(n - 1) \times (n - 1)$  matrix  $\bar{\alpha}_{n-1}^{ET}$  is, for  $n > 2$ , given by

$$\bar{\alpha}_{n-1}^{ET} = \begin{pmatrix} 2 & 0 & 0 & 0 & \dots & 0 & 0 \\ -1 & 3/2 & 0 & 0 & \dots & 0 & 0 \\ 0 & -1 & 4/3 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & -1 & \frac{n}{n-1} \end{pmatrix}, \quad (7)$$

and  $\bar{\alpha}_1^{ET} = 2$ . It follows from the definition of excess returns on a  $\tau$  period bond, denoted here by  $rx_{t+1}(\tau)$ , that under the Expectations Theory

$$E[rx_{t+1}(\tau)|\mathcal{F}_t] = \tau\rho(\tau) - (\tau - 1)\rho(\tau - 1), \quad \tau = 2, 3, \dots \quad (8)$$

We prove Theorem 1 elsewhere (see the proof of Theorem 1 of Bowsher and Meeks 2006). The important point to note here is that when one works with a complete term structure of maturities,  $\boldsymbol{\tau} = (1, 2, \dots, n)$ , the matrix  $\bar{\alpha}_{n-1}^{ET}$  and hence the conditional mean of the  $(n - 1)$ -dimensional yield curve are both *entirely determined* by the ET. This has not previously been recognised. Put another way, under the null of the ET *alone* (i.e. Definition 2), the MSFE-minimising, 1-step ahead forecasts are a known linear function of the difference between the spread and term premia vectors. Note from equation (8) that conditionally expected excess returns from holding a  $\tau$  period bond for 1 period (that is, in excess of the 1 period yield,  $y_t(1)$ , and realised at time  $t + 1$ ), are deterministic and constant over time under the ET.

### 2.2.1 Evaluating the ET

Theorem 1 gives an alternative, theory-based approach to forecasting high dimensional yield curves. Note that these yield curves need not be complete. Since  $E[\Delta y_{t+1}(\tau)|\mathcal{F}_t]$  involves the spread  $s_t(\tau + 1, 1)$ , the method is feasible provided that the data used to form the forecasts contains the yields for the required maturities (see equation 5). However, this will not usually create any difficulty. In Sections 3.4 and 4 below, the FSN-ECM model forecasts are compared to these ET forecasts, with the results strongly favouring the former. This not only subjects

the proposed FSN-ECM models to comparison with a broader set of models, but also provides a new method of evaluating the ET.

Under the null of the ET the forecasts implied by (6) have smaller MSFE for each maturity  $\tau$  than any other forecasting method since they are based on the conditional mean. Models that outperform the forecasting equation (6) in terms of MSFE thus constitute evidence against the ET. The proposed evaluation method in which the forecasts implied by (6) are compared with those of rival models has several advantages. First, the method is applicable *to any dimension of yield curve*,  $N$ , no matter how large. Second, if the ET is taken to hold no matter what the physical units of time  $t$  are in Definition 2, then forecasting equation (6) holds under the null of the ET for any time series frequency (e.g. weekly, monthly or yearly time series). Third, unlike tests of the stationarity of yield spreads which assume that the yield curve is  $I(1)$  and hence that the spreads are cointegrating relations under the null of the ET, the method is free of auxiliary ‘nuisance’ assumptions concerning the order of integration of the yield curve.

In contrast to our approach here, the cointegration-based literature seeks to evaluate the ET by analysing the monthly time series of a low dimensional  $N$ -vector of zero-coupon yields, where  $N$  is typically less than 10. Attention usually focuses on inference concerning the cointegration rank and cointegrating relations, with different studies reaching varying conclusions that are sensitive to a largely unknown degree on the subset of maturities chosen. Hall, Anderson, and Granger (1992) find that three linearly independent spreads are cointegrating relations when T-bills with maturities of 1, 2, 3, and 4 months are modelled. Pagan, Hall, and Martin (1996) work with a 5-dimensional yield curve and report rejection of the standard hypothesis test that 4 linearly independent spreads are cointegrating relations. They note however that the point estimates are quite close to the situation where such cointegrating relations hold. These authors also highlight the major impact on the critical values of the test of a levels effect of the short rate in the disturbance of the VAR, which may result in the test rejecting erroneously. Shea (1992) reaches a variety of conclusions depending on which maturities are modelled, finding both cases with 2 common trends and a single common trend. He reports that the cointegrating relations can often be written as linear combinations of the yield spreads.

It is possible to construct DGPs *inconsistent* with the ET (for example using FSN-ECM processes with  $\Omega_\epsilon = 0$ ) in which the  $N$ -dimensional yield curve is  $I(1)$  and there are  $(N -$

1) linearly independent spreads that are stationary cointegrating relations, but in which the conditional expectations of the yields do not satisfy Theorem 1. The evaluation method proposed here thus utilises the implications of the ET more fully than cointegration-based approaches which test only a weak implication of the theory, namely that  $(N - 1)$  linearly independent spreads are cointegrating relations when the yield curve is I(1).

The important regression studies of Fama and Bliss (1987), Campbell and Shiller (1991) and Cochrane and Piazzesi (2005) all find evidence that, contrary to the ET, conditional expected excess returns  $E[rx_{t+1}(\tau)|\mathcal{F}_t]$  are stochastic and time varying (cf equation 8). Campbell and Shiller (1991) can be interpreted as finding a strong positive effect on  $E[rx_{t+1}(\tau)|\mathcal{F}_t]$  of the spread  $s_t(\tau, 1)$  for  $\tau$  ranging from a few months to 10 years, where the time series frequency is monthly. Fama and Bliss (1987) and Cochrane and Piazzesi (2005) work instead with yearly time series. The important finding is that conditional expected excess returns depend on (one period) forward rates at  $t$ , which may be expressed as a linear combination of spreads plus the one period interest rate. Indeed, Cochrane and Piazzesi (2005) find that a *single*, ‘tent-shaped’ linear combination of forward rates predicts excess returns on 2 to 5 year maturity bonds realised in 1 year’s time with an  $R^2$  of up to 0.44. In order to see the connection of these regression studies with our approach, note that the definition of excess returns implies the identity

$$E[\Delta y_{t+1}(\tau)|\mathcal{F}_t] = \frac{\tau + 1}{\tau} s_t(\tau + 1, 1) - s_t(\tau, 1) - \frac{1}{\tau} E[rx_{t+1}(\tau + 1)|\mathcal{F}_t], \quad \tau = 1, 2, \dots \quad (9)$$

Thus, if for example spreads or lagged changes in yields enter  $E[rx_{t+1}(\tau + 1)|\mathcal{F}_t]$ , the conditional expected change in the yield  $E[\Delta y_{t+1}(\tau)|\mathcal{F}_t]$  then deviates from the one implied by the ET in (5). Indeed, the above regression studies lead us to expect that spreads will be useful in forecasting the yield curve but not in the way prescribed by the ET.

We show later in Sections 3.4 and 4 that the 1-month ahead, out-of-sample forecasts of certain FSN-ECM models have much smaller MSFEs than the ET forecasts based on equation (6), *for all maturities* forecast in the two datasets examined there (36 different maturities ranging from 1 month to 7 years). The ET forecast errors are also positively autocorrelated at a lag of 1 month, with an average sample autocorrelation of approximately 0.3. We conclude that the conditional mean of Theorem 1 implied by the ET is far from being the optimal MSFE predictor and that the ET is very wide of the mark. Our findings can be interpreted as evidence that conditional

expected excess returns for the one month ahead horizon,  $E[rx_{t+1}(\tau)|\mathcal{F}_t]$ , are stochastic and time varying for maturities  $\tau$  ranging widely from 1 month to 7 years. The findings are thus complementary to the results reported for the one year ahead horizon in Cochrane and Piazzesi (2005).

### 3 Forecasting High Dimensional Yield Curves

This section presents the results of an out-of-sample forecasting exercise in which the new FSN-ECM models and ET forecasts are used to forecast high dimensional yield curves and their performance compared to that of the main competing models. The task set is a difficult one, namely to forecast one month ahead a  $36 \times 1$  dimensional zero-coupon yield curve. The data are first described in detail before moving on to discussion of the specification and selection of the FSN-ECM models used, and the forecasting results obtained.

#### 3.1 Data

We use the same dataset of Unsmoothed Fama Bliss (UFB) forward rates as Diebold and Li (2006), running from November 1984 to December 2000 inclusive. The dataset is available from and has been constructed by Robert Bliss using data from the CRSP government bond files.<sup>7</sup> Zero-coupon UFB yields are then obtained by averaging the appropriate UFB forward rates. As is discussed below, the set of maturities for which yields are observed is not the same for every  $t$ . Although our FSN-ECM models can readily accommodate this feature using a time-varying but deterministic matrix  $W(k, \boldsymbol{\tau}_t)$  in the observation equation (3), we work here with the fixed vector of maturities  $\boldsymbol{\tau} = (1.5, 2, 3, \dots, 11, 12, 15, 18, \dots, 81, 84)$ , where maturities are in months and 1 month is taken to equal 30.4375 days. This approach enables the disaggregation by maturity of forecast performance over time and facilitates comparison with earlier work. Where a yield  $y_t(\tau_i)$  is not directly observed, a linear interpolation between the two nearest maturity observations is performed, as in Diebold and Li (2006). Note that we include a greater number of maturities between 1.5 months and 7 years than these authors (36 maturities compared to 14).

A 3-dimensional plot of the final dataset is shown in Figure 2, together with a plot in

---

<sup>7</sup>For further details see Bliss (1997) and the notes accompanying the “Bliss Term Structure Generating Programs.” The latter may be obtained together with our dataset, `ufb2full.dat`, by request to Robert Bliss.

the lower panel of the maturities directly observed at each date. The latter highlights the clear time-variation in the set of directly observed maturities. The minimum and maximum maturities of 1.5 and 84 months respectively were chosen in order mostly to avoid interpolations using observations separated by a relatively large maturity span. Note in particular that it is difficult to construct a reliable 1 month yield using this dataset since there is frequently no observed maturity less than or equal to 30.4375 days.<sup>8</sup> Diebold and Li (2006, Table 1) provides descriptive statistics for a subset of our maturities.

Amongst US government bonds, only Treasury Bills are pure discount bonds whilst others are coupon bearing. Thus the zero-coupon yield curve must first be constructed from the observed bond prices. Bliss (1997) discusses and compares the leading term structure estimation methods and finds that the Unsmoothed Fama Bliss (UFB) method used here performs best overall. The UFB method (see Fama and Bliss 1987, p.690) essentially constructs a piecewise constant forward rate curve, constant over the intervals between the maturities of the included bonds, that exactly prices each bond (under the assumption that coupon bonds are priced as bundles of synthetic discount bonds). All existing studies of yield curve dynamics employ such a cross-sectional estimation of the yield curve prior to and separate from modelling its dynamics.

This is somewhat unsatisfactory from the econometric viewpoint of wishing to model and forecast *observable* data within a single inferential framework. The FSN framework is ideally suited to this task since it consists of a latent yield curve and data observed with measurement error. One possibility would be to retain the ECM state equation (3) for the latent knot-yields, transform these using the relation in (1) to obtain latent knot-prices, and then form a latent discount function as the natural cubic spline interpolating to those knot-prices. The observation equation is then non-linear, and expresses the observed prices of coupon bonds as a function of the future coupon payments for each included bond, the latent discount function and the measurement error,  $\epsilon_t$ . Fitting such a non-linear FSN-ECM model results in estimation of the discount function using a cubic spline, as in the widely used McCulloch (1971) procedure, but does so by making coherent use of both the time series and cross-sectional information in the data. This extension of our approach would allow forecasting of future coupon bond prices and

---

<sup>8</sup>In the absence of such an observation, there is no valid point to use as the lower maturity point in a linear interpolation. An alternative, not adopted here, would be to derive a 1 month yield by assuming yields are constant for all maturities up to and including the first observed maturity.

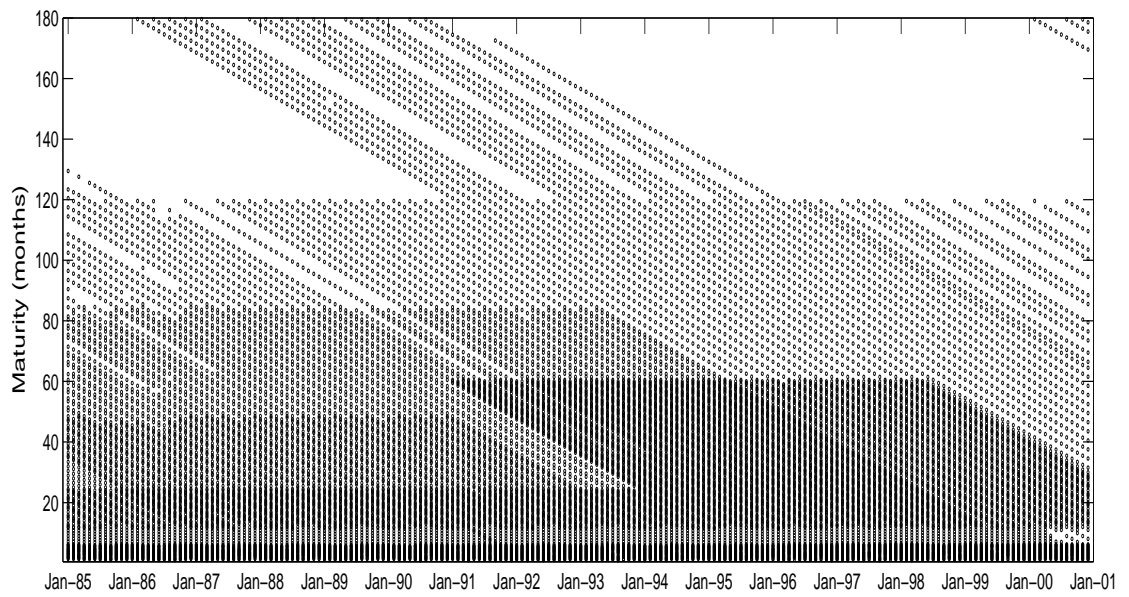
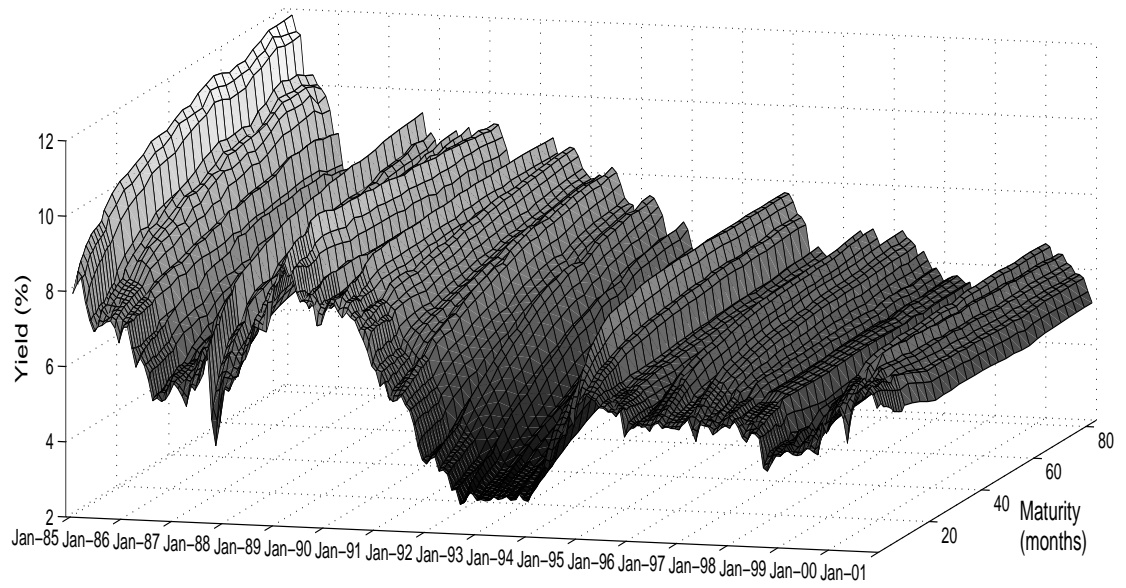


Figure 2: Zero-coupon, Unsmoothed Fama Bliss yields on US Treasury bonds. The upper panel shows a 3-dimensional plot of the dataset used in Section 3; yields are measured in percentage points per annum. The lower panel plots as circles the maturities observed at each date.

is left to future research.

### 3.2 FSN-ECM forecasting models

The model nomenclature FSN( $m$ )-ECM( $p$ ) is used to denote a model in which the spline,  $\mathcal{S}_{\gamma_t}(\boldsymbol{\tau})$ , has knots positioned at  $m$  different maturities (i.e. the knot vector  $k$  has dimension  $m$ ) and the maximum lag entering the ECM state equation (3) is the  $p$ th lag of  $\gamma_{t+1}$ . We consider models here with  $m \in \{5, 6\}$  and  $p \in \{1, 2\}$ . The following non-singular transformation of the state equation is useful in what follows. The transformed vector,  $\varphi_t := Q\gamma_t$ , consisting of the (latent) short rate and inter-knot (latent) yield spreads is given by

$$\varphi_t := \left( \gamma_{1t}, \gamma_{2t} - \gamma_{1t}, \dots, \gamma_{mt} - \gamma_{m-1,t} \right)' = \begin{pmatrix} 1 & 0_{1 \times (m-1)} \\ \beta' & \end{pmatrix} \gamma_t = Q\gamma_t, \quad (10)$$

where  $\beta'$  is defined as in (3). The state equation may then be written equivalently as the VAR

$$\Delta\varphi_{t+1} = Q\alpha(\beta'Q^{-1}\varphi_t - \mu_s) + Q\Psi Q^{-1}\Delta\varphi_t + \eta_t, \quad (11)$$

where  $\eta_t = Q\nu_t$ , and we define  $\Omega_\eta = \text{Var}[\eta_t] = Q\Omega_\nu Q'$ .

For all of the FSN( $m$ )-ECM( $p$ ) forecasting models used in Sections 3 and 4, the covariance matrix  $\Omega_\eta$  is diagonal and  $\Omega_\epsilon = \sigma_\epsilon^2 I_N$  has the one free parameter,  $\sigma_\epsilon^2$ . In all cases the Kalman filter is initialised using  $(\gamma'_1, \gamma'_0)' \sim (\mu^*, \Omega^*)$ , where  $\Omega^* = 0$  and  $\mu^*$  is set equal to the yields,  $(y_0(k)', y_{-1}(k)')$ , that correspond to the knot maturities and are *observed* in the data for the two periods prior to our estimation period (i.e. 1984:11 and 1984:12).<sup>9</sup> FSN( $m$ )-ECM(1) models are obtained by setting  $\Psi = 0$ , whilst we impose that  $Q\Psi Q^{-1}$  is diagonal in all FSN( $m$ )-ECM(2) models. The latter restriction means that only *own* lagged changes of the short rate and spreads enter each equation in (11). The matrix  $\alpha$  in (3) determines the loadings on the spread regressors,  $\beta'\gamma_t$ . The FSN(5)-ECM( $p$ ) models considered always employ an unrestricted  $\alpha$ , whilst the FSN(6)-ECM( $p$ ) models use a restricted  $\alpha$  in which the first 5 rows form an upper triangular matrix and the last row consists entirely of zeros. We dub this form, which draws on the empirical findings of Hall, Anderson, and Granger (1992), ‘triangular  $\alpha$ ’ and adopt the nomenclature T-FSN( $m$ )-ECM( $p$ ) for models with the restriction imposed. In such triangular

<sup>9</sup>The initialisation procedure used legitimately conditions on pre-sample information and avoids augmenting the parameter vector of the model with the initial state vector. Diffuse initialisation was found not to perform well in forecasting in this context. The motivation is the use of the approximation that the observed knot-yields follow a random walk for the 2 periods in question. Note that in Section 3.4 it is necessary to use  $\gamma_{1t} = y_{t-1}(1.5)$  for  $t = 0, 1$  as 1 month yields are not observed in the dataset used there.



models the change in a given knot-yield,  $\Delta\gamma_{j,t+1}$ , depends only on time  $t$  spreads involving knot-yields of the same and longer maturities, i.e. on  $(\gamma_{j+1,t} - \gamma_{j,t}, \dots, \gamma_{m,t} - \gamma_{m-1,t})'$ . Thus, when  $\Psi = 0$ ,  $\Delta\gamma_{m,t+1}$  follows a random walk.

### 3.3 Model selection procedure

The data from 1985:1 to 1993:12 inclusive was used as ‘training’ data for the purpose of an in-sample model selection stage in which the number of knots ( $m$ ) and their positions ( $k$ ) were determined. One-step ahead forecasts of the data for each month from 1994:1 to 2000:12 inclusive were then made using a small subset of models carried forward from the in-sample stage, and the forecasts compared across those models. It was felt that there was insufficient data to hold back some time periods for additional evaluation of a single forecasting model or procedure selected after the second stage.

The in-sample stage for knot selection is based on assessing the cross-sectional fit for a large number of different knot vectors,  $k$ . Specifically, we fit using OLS a natural cubic spline with knot vector  $k$  to each observed yield curve,  $y_t(\tau)$ , and then compute the mean across time of the residual sum of squares (RSS) from each of the cross-sectional regressions. For  $m = 5$  and  $m = 6$ , the top twenty knot vectors in terms of minimisation of the mean RSS amongst all possible  $k$  with end knots at 1 and 84 months and internal knots lying in the set  $(2, 3, \dots, 11, 12, 15, 18, \dots, 78, 81)$ . This is the same set of maturities used in the dataset described above, but excluding the shortest and longest maturities.<sup>10</sup> Other values of  $m$  were not considered. The procedure has the advantage that it is computationally feasible to search over a very large model space in the manner described (46,376 knot vectors for  $m = 6$ , and 5984 for  $m = 5$ ). The criterion is cross-sectional fit rather than dynamic forecasting, but the ability to mimic the shape of observed yield curves is a preliminary desideratum for the FSN-ECM model to perform well in forecasting. Three knot vectors for  $m = 6$  and one for  $m = 5$  were then carried forward to the second stage, avoiding  $k$ ’s in which neighbouring knots occupied adjacent positions in the ordered set  $(2, 3, \dots, 11, 12, 15, 18, \dots, 81, 84)$ . It was found that in-sample estimation using such knots and the training data alone resulted in poorly behaved estimates of  $\Omega_\eta$  that involved zero variances.

---

<sup>10</sup>We present models with the first knot positioned at 1 month, i.e.  $k_1 = 1$ , rather than at the first maturity present in the dataset, i.e.  $k_1 = 1.5$ , for two reasons:  $k_1 = 1$  is a more natural, generally applicable choice and its use here allows the estimation of a 1 month base rate using the smoothed estimates,  $\hat{\gamma}_{1t}$ , if desired.

### 3.4 Forecast Evaluation

The forecast performance of the FSN-ECM models selected from the first stage is compared here to three rival models: a RW for the yield curve (i.e. the ‘no change’ forecast,  $RW^{YC}$ ), the Diebold and Li (2006) dynamic Nelson-Siegel model (henceforth DNS) and the ET forecast derived in Theorem 1. Two different estimation schemes are used in forecasting: either the parameters are updated recursively (R) by adding an observation to the data used for estimation each time a new forecast is made, or parameters are held constant (C) at the in-sample estimates obtained using the 1985:1 to 1993:12 training data. Implementation of the second and third rival models is described before proceeding to a discussion of the forecasting results.

We implement the version of the DNS model preferred by Diebold and Li (2006) in which each of the three latent factors follows an AR(1) process. The three factors parametrise the Nelson and Siegel (1987) latent yield curve at each time  $t$  and may be interpreted as the ‘level, slope and curvature’ of the latent yield curve. Rather than use the 2-stage OLS estimation procedure of Diebold and Li (2006), we use the state space form of the model and the Kalman filter to perform QML estimation (as with the FSN-ECM models). The DNS model specification is almost identical to the ‘yields-only’ model in Diebold, Rudebusch, and Aruoba (2006), with unrestricted and diagonal covariance matrices for the disturbances of the state and observation equations respectively. The only difference is that for the observation disturbance, the elements of the diagonal of the covariance matrix are restricted to be equal within 8 different maturity groupings, owing to the higher dimension of the yield curve in this setting.<sup>11</sup> The state equation was initialised using the unconditional mean and variance of the state vector, as in Diebold, Rudebusch, and Aruoba (2006).

Since our dataset does not include a 1 month yield, we produce ET forecasts of  $y_t(3 : 84)$  based on equation (6) with the first two rows excluded, where  $y_t(3 : 84) := (y_t(3), y_t(4), \dots, y_t(84))'$ . Twenty five additional yields, namely  $(y_t(13), y_t(16), \dots, y_t(85))'$ , are thus included in the information set on which the ET forecasts are based, compared both to the FSN-ECM and DNS forecasts.<sup>12</sup> The vector of term premia,  $\rho(3 : 85) := (\rho(3), \dots, \rho(85))'$ , is estimated by OLS

---

<sup>11</sup>Each maturity in (1.5, 2, 3, 4) has its own parameter; a single parameter then corresponds to each of the maturity groupings (5, 6, ..., 10), (11, 12, 15, ..., 24), (27, ..., 79), and (79, ..., 84).

<sup>12</sup>Clearly the FSN-ECM and DNS forecasts can be interpreted, for purposes of assessing whether the ET forecasts of  $y_t(\tau)$  are MSFE minimising forecasts, as being conditional on this broader information set even though use is not made of the additional yields.

using the following regression derived from equation (6)

$$\Delta y_{t+1}(3 : 84) = \bar{\alpha}_{3:84}^{ET} [s_t(3 : 85) - \rho(3 : 85)] + \nu_{t+1}, \quad (12)$$

where  $s_t(3 : 85) := (s_t(3, 1), \dots, s_t(85, 1))'$ ,  $\bar{\alpha}_{3:84}^{ET}$  denotes the 3rd to 84th rows inclusive of  $\bar{\alpha}_{84}^{ET}$  in equation (6), and the term premia are assumed to lie on a cubic spline with knot vector  $(1, 3, 4, 27, 85)$  and  $\rho(1) = 0$ . The parameters estimated by OLS are thus the term premia for the knot maturities  $(3, 4, 27, 85)$  – see Theorem 2 of the Appendix and Poirier (1973).

Figure 3 plots by maturity the percentage increase in MSFE relative to the  $RW^{YC}$  (negative values thus representing superior performance compared to the  $RW^{YC}$ ) for the following models: the T-FSN(6)-ECM(2) model with triangular alpha and  $k = (1, 2, 4, 18, 24, 84)$ ; the DNS model; and the ET forecasting equation in (12). In all cases estimation is performed recursively, except for the additional line plotted for the ET case with parameters held constant (C) throughout the forecast evaluation period. Not only does the triangular FSN(6)-ECM(2) model outperform all of the rival models (including the  $RW^{YC}$ ) *at all maturities*, but the percentage reductions obtained in MSFE compared to the  $RW^{YC}$  are also substantial, particularly at the shorter maturity end of the yield curve. Considered across the entire span of maturities, the gains over the DNS model are large, with DNS performing particularly poorly and worse than the  $RW^{YC}$  for maturities between 12 and 32 months. The Diebold and Li (2006) DNS method is the only previously published method for forecasting high dimensional yield curves. The authors report better performance for the method at forecast horizons of 6 and 12 months ahead than at the 1 month ahead horizon.

Strikingly, the ET forecasts perform worse than the  $RW^{YC}$  for the majority of the 82 maturities forecast. ET forecasts produced holding the term premia parameters constant (C) have higher MSFE for all maturities than those produced using recursive estimation (R). Presumably recursive estimation improves the forecasts by enabling a degree of variation over time in the term premia,  $\rho(3 : 85)$ , which are of course assumed to be time-invariant constants under the ET. The average MSFEs across the 82 maturities are 113% and 121% of that for the  $RW^{YC}$  in the recursive and constant parameter cases respectively. It is clear from Figure 3 that the conditional mean of Theorem 1 implied by the ET is far from being the optimal MSFE predictor since the MSFEs of the T-FSN(6)-ECM(2) model are much smaller than those for either the

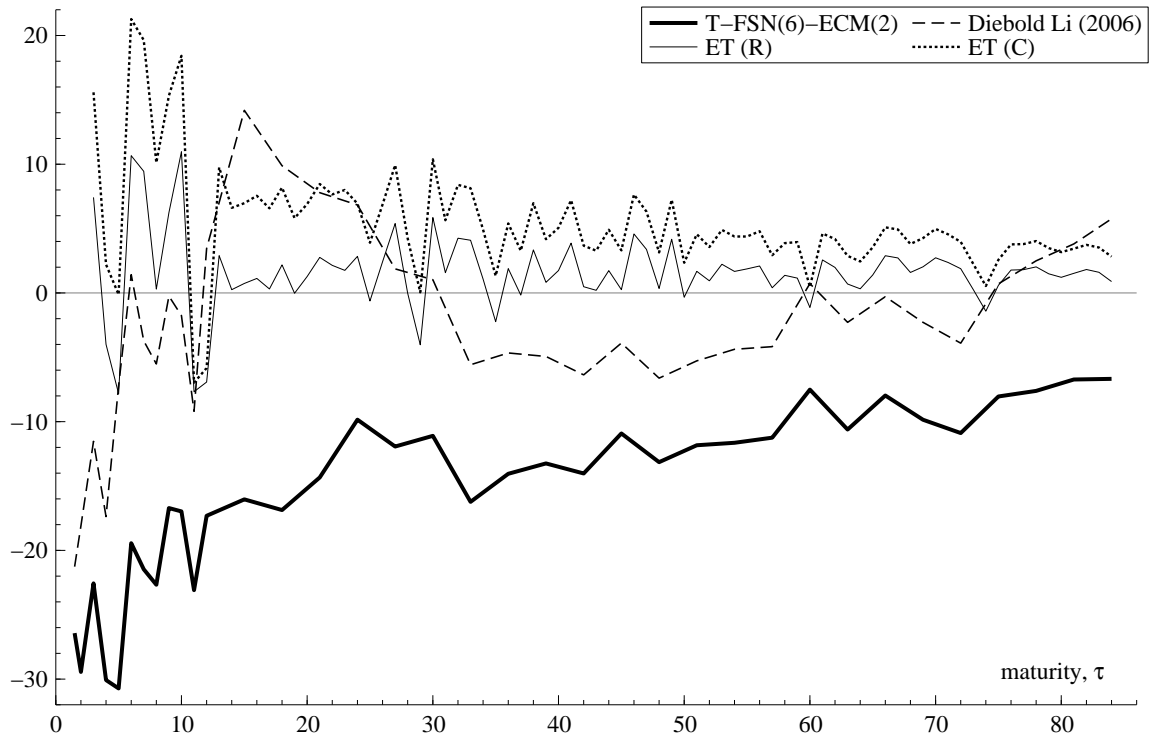


Figure 3: Percentage reduction in model MSFEs by maturity relative to those of the  $RW^{YC}$ . Shown are the results for the triangular (T-) FSN(6)-ECM(2) model with  $k = (1, 2, 4, 18, 24, 84)$ , the Diebold and Li (2006) DNS model and the Expectations Theory (ET) forecasting equation (12). (R) stands for recursive estimation and (C) for forecasts produced using constant parameters. The horizontal axis is maturity measured in months.

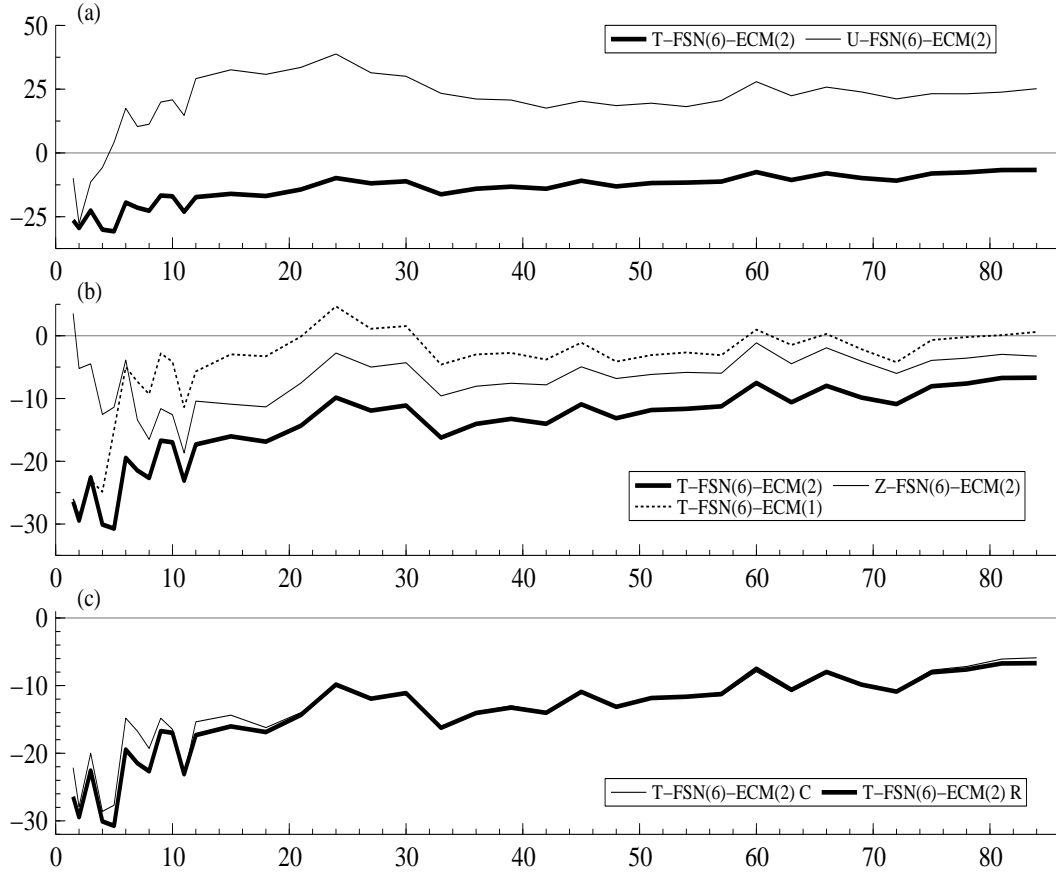


Figure 4: Percentage reduction in FSN(6)-ECM( $p$ ) model MSFEs by maturity relative to those of the  $RW^{YC}$ . The knot vector used in all cases is  $k = (1, 2, 4, 18, 24, 84)$ . (a) FSN(6)-ECM(2) models with triangular (T-) and unrestricted (U-)  $\alpha$  matrices. (b) Z- stands for a model with  $\alpha = 0$ . (c) R stands for recursive estimation and C for forecasts produced using constant parameters. The horizontal axes are maturity measured in months.

ET (R) or ET (C) cases *for all maturities included*. In line with this conclusion, the average across maturities of the absolute value of the first order autocorrelations of the forecast errors are 0.295 (R) and 0.288 (C), compared to a value of zero under the null that the conditional mean of Theorem 1 is the true one. The new method for evaluating the ET proposed in Section 2.2.1 thus finds that the ET is very wide of the mark. We reach the same conclusion using a differently constructed dataset consisting of the yields on a complete set of T-Bills in Section 4.

Figure 4 presents analogous plots for FSN(6)-ECM( $p$ ) models, all of which have  $k = (1, 2, 4, 18, 24, 84)$  and are estimated recursively unless indicated otherwise. In each panel the T-FSN(6)-ECM(2) model of Figure 3 is shown as a thick solid line for ease of comparison. Panel (a) compares that triangular model with an otherwise identical specification in which  $\alpha$  is unre-

stricted (U). The enormous benefit of the triangular restriction on  $\alpha$  is clearly evident and is thought to stem from imprecise estimation of the large number of parameters in an unrestricted  $\alpha$  with  $m = 6$  and time series of this length. Panel (b) highlights the impact of the restrictions  $\Psi = 0$  and  $\alpha = 0$ . Imposing  $\Psi = 0$  on the T-FSN(6)-ECM(2) model to obtain the T-FSN(6)-ECM(1) model is very costly in terms of MSFE for all but the shortest maturities. Imposing  $\alpha = 0$  (to obtain the Z- or Zero model) is somewhat less costly for maturities greater than or equal to 7 months, but is drastically costly for the shortest maturities. Thus inclusion of the spreads  $\beta' \gamma_t$  as regressors in the state equation (3) is crucially important for forecasting the short end of the yield curve, but also continues to play a role at the long end (witness that the line for the T-FSN(6)-ECM(2) model is still below that for the Z-FSN(6)-ECM(2) model for the longer maturities in panel (b)). It is important for forecasting (in terms of MSFEs by maturity) to retain both spreads and lagged changes in knot-yields as regressors in the ECM state equation. Finally in panel (c) and for the triangular FSN(6)-ECM(2) model, recursive estimation (R) produces similar MSFEs to holding the parameters constant (C), the largest differences being for the shorter maturities and in favour there of recursive estimation. This observation suggests that parameter non-constancy is not a significant problem when forecasting the data used here with FSN-ECM models.

Table 1 reports measures of forecast performance both for the models considered in Figures 3 and 4, and for a broader range of specifications and estimation schemes. The focus is on summary MSFE-based measures, although the average across maturity of the absolute value of some sample autocorrelations of the forecast errors and of the mean forecast errors are also reported. Although the average across maturity of the MSFEs, or equivalently the trace of the MSFE matrix (denoted  $mMSFE$ ), is an intuitively reasonable measure it is not invariant to non-singular linear transformations of the data even when linear predictors are used. For example, the model ordering implied by the  $tr(mMSFE)$  can in principle change when the data to be forecast is expressed as a vector consisting of the shortest rate and spreads relative to that rate, rather than as a yield curve. Also reported therefore is the determinant of the MSFE matrix, which has the desired invariance property (see Clements and Hendry 1993).<sup>13</sup>

---

<sup>13</sup>To see this, let the transformed data  $y_i^*(\tau) = Py_i(\tau)$  for some non-singular  $P$ , and denote the corresponding MSFE matrices of the forecast errors by  $mMSFE^*$  and  $mMSFE$ . Provided that the forecast  $\hat{y}_i^*(\tau) = P\hat{y}_i(\tau)$ , then  $\det(mMSFE^*) = \det(P)\det(mMSFE)\det(P')$ . It then follows immediately that model rankings do not depend on the choice of  $P$ . Furthermore, restricting attention to cases where  $\det(P) = 1$ , as is the case for

Model type	R/C	$\alpha$	$tr/N$	$MSFE^{(1.5)}$	$det^{1/N}$	$mean$	$ \rho(1) $	$ \rho(6) $	$ \rho(12) $
<b>FSN(<math>m</math>)-ECM(<math>p</math>)</b>									
$k = (1, 2, 4, 18, 24, 84)$									
FSN(6)-ECM(2)	R	T	85.9	73.9	86.7	0.018	0.106	0.060	0.065
FSN(6)-ECM(2)	C	T	86.6	77.8	87.4	0.027	0.096	0.060	0.061
FSN(6)-ECM(2)	R	U	120.3	89.9	87.3	0.028	0.209	0.058	0.047
FSN(6)-ECM(2)	R	Z	93.5	103.7	88.2	0.014	0.204	0.057	0.126
FSN(6)-ECM(1)	R	T	95.7	73.9	85.2	0.021	0.300	0.050	0.118
$k = (1, 2, 4, 12, 30, 84)$									
FSN(6)-ECM(2)	R	T	86.8	75.2	87.6	0.018	0.090	0.062	0.061
$k = (1, 2, 4, 15, 24, 84)$									
FSN(6)-ECM(2)	R	T	85.9	74.6	87.1	0.018	0.099	0.061	0.066
$k = (1, 2, 4, 27, 84)$									
FSN(5)-ECM(2)	R	U	91.6	78.4	88.0	0.017	0.147	0.091	0.050
FSN(5)-ECM(2)	R	T	88.9	73.3	88.3	0.021	0.093	0.060	0.051
FSN(5)-ECM(1)	R	U	101.3	78.4	85.7	0.030	0.331	0.062	0.106
FSN(5)-ECM(1)	R	T	96.6	73.9	86.6	0.022	0.295	0.046	0.120
<b>Rival Models</b>									
RW <sup>YC</sup>	-	-	$64.9 \times 10^{-3}$	$73.4 \times 10^{-3}$	$1.02 \times 10^{-3}$	0.013	0.305	0.061	0.132
Diebold Li (2006)	R	-	97.9	79.1	86.6	0.026	0.318	0.056	0.125
Diebold Li (2006)	C	-	97.2	73.9	86.6	0.034	0.309	0.054	0.120

Table 1: Forecasting high dimensional yield curves: summary measures of forecast performance. The FSN( $m$ )-ECM( $p$ ) models are grouped according to the knot vector  $k$ ; the parameter  $\alpha$  is either unrestricted (U), triangular (T) or equal to zero (Z). The estimation procedure used is either recursive (R) or holds parameters constant at their in-sample estimates (C). The MSFE-based measures are denoted  $tr/N$  for the average MSFE,  $MSFE^{(1.5)}$  for the MSFE of the 1.5 month yield, and  $det^{1/N}$  for  $[det(mMSFE)]^{1/N}$ . With the exception of the RW<sup>YC</sup> for which the measures are reported directly, all 3 MSFE-based measures are expressed as a percentage of the corresponding measure for the RW<sup>YC</sup>. The average across maturity of the  $mean$  forecast errors and of the absolute values of the sample autocorrelations of the forecast errors at lag  $h$ ,  $|\rho(h)|$ , are also reported.

The reported average MSFEs are reflective of the comments made above concerning Figures 3 and 4. Note that the average MSFE of the triangular FSN(6)-ECM(2) models is 86 to 87 per cent of that for the  $RW^{YC}$  forecast, compared to 89 per cent for the triangular 5 knot FSN(5)-ECM(2) model and 97 to 98 per cent for the DNS model. For triangular  $m = 6$  and  $m = 5$  models, the average MSFE is higher for the ECM(1) models than for the ECM(2) ones. Unlike the  $m = 6$  case, the FSN(5)-ECM(2) model with an unrestricted  $\alpha$  matrix performs quite well, presumably due to the reduction in the number of estimated model parameters.

Large reductions compared to the  $RW^{YC}$  of about 25 per cent in the MSFE of the shortest (1.5 month) yield are achieved by all of the triangular FSN( $m$ )-ECM( $p$ ) models with  $m \in \{5, 6\}$  and  $p \in \{1, 2\}$ . Thus, very substantial gains over the random walk forecast of a short rate can be realised using the FSN models and the information contained in the yield curve alone. Interestingly, all of the models included in Table 1 perform similarly and better than the  $RW^{YC}$  in terms of the  $\det(mMSFE)$  measure, the best performer being the triangular FSN(6)-ECM(1) model. The triangular FSN( $m$ )-ECM(2) models have substantially lower average absolute autocorrelations at lags of 1 and 12 months than the  $RW^{YC}$  and DNS models.

The triangular FSN(6)-ECM(2) models (e.g. the one with  $k = (1, 2, 4, 18, 24, 84)$  featured in Figures 3 and 4) are strong performers across the entire range of summary measures considered in Table 1. These T-FSN(6)-ECM(2) models dominate the rival forecasts of the  $RW^{YC}$  and Diebold and Li (2006) DNS methods in terms of MSFE across all maturities (recall Figure 3); have a much lower average MSFE than the DNS forecasts; achieve very large reductions in the MSFE of the short rate, conditioning on the information in past yield curves alone; and perform better than the  $RW^{YC}$  and very similarly to the DNS methods in terms of the  $\det(mMSFE)$  measure. Having considered a broad range of measures of forecast performance, we conclude that such T-FSN(6)-ECM(2) models outperform the existing methods for forecasting high dimensional yield curves.

### 3.5 Parameter estimates for the T-FSN(6)-ECM(2) model

We report here the QMLEs used to produce the constant parameter (C) forecasts of the triangular FSN(6)-ECM(2) model with  $k = (1, 2, 4, 18, 24, 84)$ , whose MSFEs are reported in Figure 

---

 transformation to ‘short rate and spreads’, the numerical value of the loss for a given model is also invariant to  $P$ .



$\mu_s$	$Q\alpha$					$diag(Q\Psi Q^{-1})$	$diag(\Omega_\eta)$
0.2188	1.223	0.0568	-0.2440	0.9495	-0.0347	0.3405	0.0979
0.1757	-1.223	0.6665	0.1461	-0.4561	0.0003	0.4623	0.0172
0.7146	0	-0.7233	0.2535	-0.5864	0.0340	0.4364	0.0073
0.2093	0	0	-0.1556	0.2685	-0.0028	0.1533	0.0267
1.048	0	0	0	-0.1757	0.0286	0.2309	0.0009
	0	0	0	0	-0.0255	0.1090	0.0196
$\sigma_\epsilon^2$							
0.0032							

Table 2: QMLEs of the triangular FSN(6)-ECM(2) model with  $k = (1, 2, 4, 18, 24, 84)$  obtained using the training data from 1985:1 to 2000:12 inclusive as the estimation data. The operation  $diag(X)$  gives the diagonal of the matrix  $X$  as a column vector.

4(c). This choice stems from the fact that this model specification performs well across the range of measures of forecast performance considered above, and the MSFEs for constant and recursive estimates in Figure 4(c) are very similar.

With the parameters of the state equation (3) set equal to their QMLEs, the knot-yields  $\{\gamma_t\}$  follow an I(1) process and the  $(m - 1)$  spreads between them,  $\beta'\gamma_t$ , are stationary cointegrating relations. This follows since the ranks of  $\hat{\alpha}$  and  $\hat{\alpha}\beta'$  are both equal to 5, the roots  $z$  of the characteristic polynomial of the VAR in (3) then satisfy either  $z = 1$  or  $|z| > 1$ , and the determinant of  $[\alpha'_\perp(I - \Psi)\beta_\perp]$  is non-zero (see Johansen 1996, Theorem 4.2). The characteristic polynomial of the VAR in (3) thus has exactly one unit root.

Table 2 reports the QMLEs of the transformed state equation (11), together with  $\hat{\sigma}_\epsilon^2$ . Recall that the state vector in (11) is  $\varphi_t = [\gamma_{1t}, (\beta'\gamma_t)']$ , the vector consisting of the latent short rate and inter-knot yield spreads. The estimates of the stationary means of the spreads  $\beta'\gamma_t$  are all positive, implying that the latent yield curve is ‘upward sloping’ on average. It follows from (11) that

$$\Delta\gamma_{1,t+1} = Q\alpha_{[1]}(\beta'\gamma_t - \mu_s) + \Delta\gamma_{1,t} + \eta_{1t}, \quad (13)$$

where  $Q\alpha_{[1]}$  denotes the first row of the matrix  $Q\alpha$ . Bearing in mind that  $k = (1, 2, 4, 18, 24, 120)$ , the estimates of the row  $Q\alpha_{[1]}$  in Table 2 are largest in absolute value for the spreads between the first and second, third and fourth, and fourth and fifth knot-yields. Notice that all elements of the last column of  $Q\alpha$  are close to zero, indicating that the spread between the fifth and sixth knot-yields is unimportant as a regressor in (11). It also follows from (11) that the vector of

spreads follows the VAR

$$\Delta(\beta' \gamma_{t+1}) = Q\alpha_{[2:6]}(\beta' \gamma_t - \mu_s) + (Q\Psi Q^{-1})_{[2:6]} \Delta(\beta' \gamma_t) + \eta_{[2:6],t}, \quad (14)$$

where  $X_{[2:6]}$  denotes the 2nd to 6th rows of some matrix,  $X$ . With parameters set equal to the QMLEs in Table 2, this VAR is stationary. Furthermore, a given spread at  $(t + 1)$  then depends *only* on time  $t$  spreads involving the same and longer maturities (since  $Q\alpha_{[2:6]}$  is upper triangular and  $(Q\Psi Q^{-1})_{[2:6]}$  is diagonal) and the same spread at time  $(t - 1)$ . Note that all elements of the estimate of  $Q\Psi Q^{-1}$  are positive and of the estimate of the diagonal of  $Q\alpha_{[2:6]}$  are negative.

## 4 Forecasting Treasury Bill Yields

The focus of this section is on forecasting a complete set of Treasury (T-) bills with maturities ranging from 1 to 11 months inclusive. The dataset used is taken from the widely available and studied ‘Fama Treasury Bill term structure files’ contained in the CRSP monthly US Treasury database. Since zero-coupon yields on T-bills are directly observable, the analysis in this section is free of any initial method used to estimate the zero-coupon yield curve prior to analysing its dynamics. A further advantage is that the use of a yield curve dataset constructed differently to the one in Section 3 allows a check on the robustness of the findings there concerning the Expectations Theory. The results presented demonstrate the flexibility of the FSN-ECM models in forecasting different portions of the yield curve. The information set on which the forecasts are based is smaller since current and past yields with maturities greater than 11 months are excluded. Nevertheless, the use of less heavily parametrised FSN-ECM models with a smaller number of knots,  $m$ , provides an effective forecasting method in this setting. Less computationally burdensome models and datasets such as these are helpful when, for example, using particle filtering to form predictive densities.

The dataset is constructed by selecting for each month the T-bill closest in maturity to 12 months and then following that bill to maturity. The assumption is that the yield curve is constant over the interval between the maturity that is actually observed and the ‘target’ maturity, so the method chooses not to use linear interpolation. For a given month our dataset consists of the 11 monthly yields to maturity of T-bills,  $y_t(1 : 11)$ , and covers the period 1984:11 to 2000:3 inclusive. This is the same period as covered by the dataset used in Section 3, except

that the last 9 months of the year 2000 are omitted.<sup>14</sup> A month is taken to equal 30.4 days when computing the yields.

The five best performing knot vectors in terms of cross-sectional fit, with  $m = 4$  knots and first and last knots at 1 and 11 months respectively, were selected using the same in-sample procedure as that described in Section 3.3. The two internal knots of each knot vector were chosen from the set  $(2, 3, \dots, 10)$ . Only FSN(4)-ECM( $p$ ) models are considered here, again with  $p \in \{1, 2\}$ . The 2 knot vectors  $(1, 2, 3, 11)$  and  $(1, 2, 5, 11)$  were then carried forward to the forecast evaluation stage. The restrictions on  $\Omega_\eta$ ,  $\Omega_\epsilon$ ,  $\alpha$ , and  $Q\Psi Q^{-1}$  are exactly as described in Section 3.2. As there, the Kalman filter is initialised using  $(\gamma'_1, \gamma'_0)' \sim (\mu^*, \Omega^*)$ , where  $\Omega^* = 0$  and  $\mu^*$  is set equal to the yields,  $(y_0(k)', y_{-1}(k)')'$ , that correspond to the knot maturities and are *observed* in the data for the two periods prior to our estimation period (i.e. 1984:11 and 1984:12). The estimation or ‘training data’ period is again 1985:1 to 1993:12 inclusive, and forecasts are made for each period from 1994:1 to 2000:3 inclusive.

Figure 5 plots by maturity the percentage increase in MSFE relative to the  $RW^{YC}$  (negative values thus representing superior performance compared to the  $RW^{YC}$ ) for the FSN(4)-ECM(2) model with unrestricted (U-) alpha and  $k = (1, 2, 3, 11)$ , and for the ET forecasts of  $y_t(1 : 10)$  from equation (6) for both recursively estimated (R) and constant (C) parameters.<sup>15</sup> The vector of term premia,  $\rho_{11}$ , is as before estimated by OLS using the regression derived from (6), with the term premia assumed to lie on a cubic spline with knot vector  $(1, 2, 3, 11)$  and  $\rho(1) = 0$ .

As in Section 3, the FSN-ECM model outperforms the rival  $RW^{YC}$  and ET forecasting methods at all maturities in terms of MSFEs. The average MSFE of the U-FSN(4)-ECM(2) model is 77.6 per cent of that of the  $RW^{YC}$ . The ET forecasts perform worse than the  $RW^{YC}$  for all but the 1 month yield, and ET forecasts produced holding the term premia parameters constant (C) have higher (or equal) MSFE for all maturities than those produced using recursive estimation (R). The average MSFEs across the 10 maturities are 102.3% and 109.6% of that for the  $RW^{YC}$  in the recursive and constant parameter cases respectively. It is clear from Figure 5 that the conditional mean of Theorem 1 implied by the ET is not the optimal MSFE predictor

<sup>14</sup>The 12 month yield and the last 9 months of the year 2000 are omitted from the dataset analysed in this section as the inclusion of either results in missing observations. Further details concerning the Fama T-Bill term structure files may be found in the CRSP monthly US Treasury database guide (pp. 29-31).

<sup>15</sup>Comparison is not made with the Diebold and Li (2006) DNS method in this section. Since one of the dynamic factors (the ‘level factor’) of the model is equal to the limit of the latent yield curve as the maturity tends to infinity, it seems that the model is not well suited to modelling the short end of the yield curve alone.

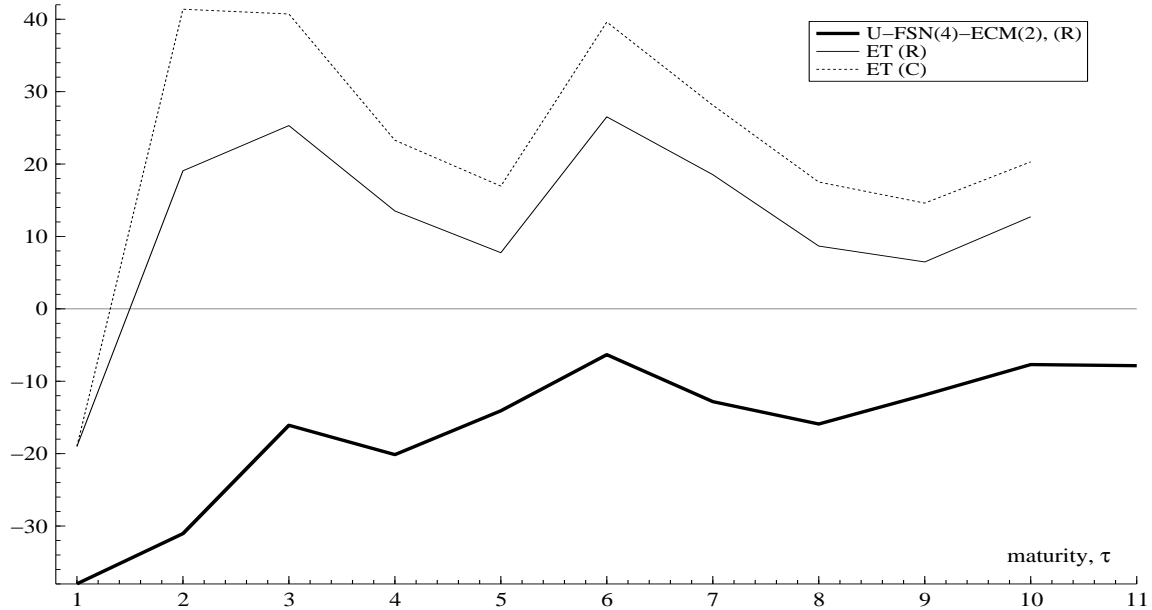


Figure 5: Percentage reduction in model MSFEs by maturity relative to those of the  $RW^{YC}$ . Shown are the results for the FSN(4)-ECM(2) model with unrestricted (U-)  $\alpha$  matrix and  $k = (1, 2, 3, 11)$ , and for the Expectations Theory (ET) forecasting equation (6). (R) stands for recursive estimation and (C) for forecasts produced using constant parameters. The horizontal axis is maturity measured in months.

since the MSFEs of the U-FSN(4)-ECM(2) model are much smaller than those for either the ET (R) or ET (C) cases for all maturities included. In line with this conclusion, the average across maturities of the positive, first order autocorrelations of the forecast errors are 0.278 (R) and 0.281 (C), compared to a value of zero under the null that the conditional mean of Theorem 1 is the true one. As in Section 3, we find that the ET is very wide of the mark.

Figure 6 presents analogous plots for various FSN(4)-ECM( $p$ ) models. The models are estimated recursively unless indicated otherwise. In each panel the U-FSN(4)-ECM(2) model of Figure 5 is shown as a thick solid line for ease of comparison. Panel (a) compares that model with an otherwise identical specification in which  $\alpha$  is triangular (T) instead of unrestricted. In contrast to Figure 4(a), the triangular restriction results in higher MSFEs for most of the T-bill maturities. Of course the inter-knot spreads,  $\beta' \gamma_t$ , involved are between different maturities in the two cases. Panel (b) shows the impact of imposing the restrictions  $\Psi = 0$  or  $\alpha = 0$  on the U-FSN(4)-ECM(2) model. As in Figure 4(b), the  $\alpha = 0$  restriction is extremely costly for short maturities such as those of T-bills, supporting the earlier conclusion that inclusion of the spreads,  $\beta' \gamma_t$ , as regressors in the state equation (3) is very important for forecasting the short

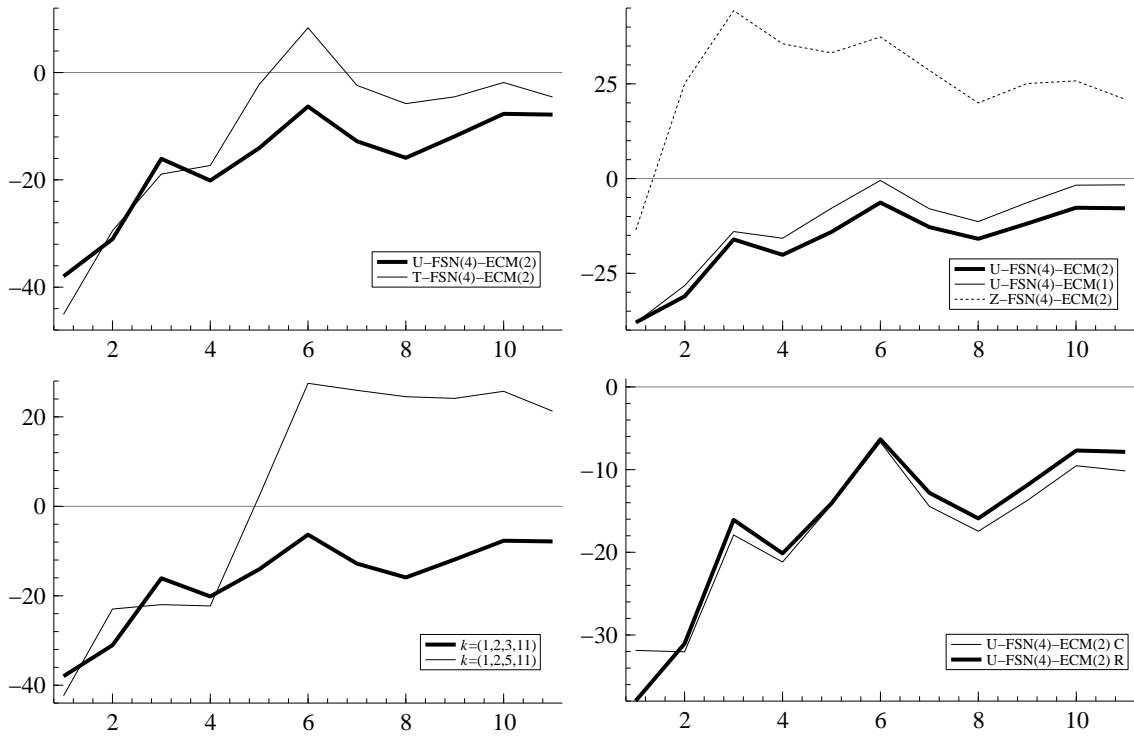


Figure 6: Percentage reduction in FSN(4)-ECM( $p$ ) model MSFEs by maturity relative to those of the  $RW^{YC}$ . (a) FSN(4)-ECM(2) models with triangular (T-) and unrestricted (U-)  $\alpha$  matrices, and  $k = (1, 2, 3, 11)$ . (b) All models shown are for  $k = (1, 2, 3, 11)$ ; Z- stands for a model with  $\alpha = 0$ . (c) U-FSN(4)-ECM(2) models with the 2 different knot vectors indicated. (d) R stands for recursive estimation and C for forecasts produced using constant parameters. The horizontal axes are maturity measured in months.

Model type	R/C	$\alpha$	$tr/N$	$MSFE^{(1)}$	$det^{1/N}$	$mean$	$ \rho(1) $	$ \rho(6) $	$ \rho(12) $
$k = (1, 2, 3, 11)$									
FSN(4)-ECM(2)	R	U	77.6	62.0	59.8	0.051	0.155	0.089	0.111
FSN(4)-ECM(2)	C	U	78.8	68.3	60.3	0.056	0.170	0.101	0.115
FSN(4)-ECM(2)	R	T	79.4	55.1	65.2	0.053	0.204	0.108	0.140
FSN(4)-ECM(2)	R	Z	114.1	86.7	68.9	0.047	0.461	0.137	0.161
FSN(4)-ECM(1)	R	U	80.9	62.0	59.6	0.058	0.230	0.088	0.105
FSN(4)-ECM(1)	R	T	81.5	54.5	65.2	0.055	0.288	0.119	0.126
$k = (1, 2, 5, 11)$									
FSN(4)-ECM(2)	R	U	92.1	57.7	66.6	0.062	0.187	0.087	0.130
FSN(4)-ECM(2)	R	T	124.4	80.5	88.1	0.057	0.121	0.075	0.220
FSN(4)-ECM(2)	R	Z	111.5	83.6	70.2	0.045	0.465	0.145	0.168
FSN(4)-ECM(1)	R	U	84.8	72.2	60.5	0.059	0.247	0.098	0.111
FSN(4)-ECM(1)	R	T	83.1	57.1	65.5	0.049	0.318	0.135	0.112
RW <sup>YC</sup>	-	-	$57.9 \times 10^{-3}$	0.217	$5.55 \times 10^{-3}$	0.037	0.352	0.152	0.132

Table 3: Forecasting T-Bills: summary measures of forecast performance. The FSN(4)-ECM( $p$ ) models are grouped according to the knot vector  $k$ ; the parameter  $\alpha$  is either unrestricted (U), triangular (T) or equal to zero (Z). The estimation procedure used is either recursive (R) or holds parameters constant at their in-sample estimates (C). The MSFE-based measures are denoted  $tr/N$  for the average MSFE,  $MSFE^{(1)}$  for the MSFE of the 1 month yield, and  $det^{1/N}$  for  $[det(mMSFE)]^{1/N}$ . With the exception of the RW<sup>YC</sup> for which the measures are reported directly, all 3 MSFE-based measures are expressed as a percentage of the corresponding measure for the RW<sup>YC</sup>. The average across maturity of the  $mean$  forecast errors and of the absolute values of the sample autocorrelations of the forecast errors at lag  $h$ ,  $|\rho(h)|$ , are also reported.

end of the yield curve. Panel (c) compares two U-FSN(4)-ECM(2) models with  $k = (1, 2, 3, 11)$  and  $k = (1, 2, 5, 11)$ , with the poor performance of the latter knot vector clearly evident for maturities greater than 4 months. The knot selection procedure described in Section 3.3, whilst a very useful tool for searching over a large model space, does not ensure good forecasting performance.<sup>16</sup> Finally, panel (d) shows that for the U-FSN(4)-ECM(2) model of Figure 5 with  $k = (1, 2, 3, 11)$ , recursive estimation produces similar MSFEs to holding the parameters constant.

Table 3 reports the summary measures of forecast performance used earlier for the two selected knot vectors and a range of model specifications, including those in Figure 6. Concentrating on the  $k = (1, 2, 3, 11)$  case, the U-FSN(4)-ECM( $p$ ) models achieve a very large reduction of 38 per cent in the MSFE of the 1 month yield relative to that of RW<sup>YC</sup>, whilst the T-FSN(4)-

<sup>16</sup>One possibility not explored here would be to compare the in-sample, 1-step ahead prediction errors for selected knot vectors using the training data alone, prior to out-of-sample forecasting.

$\mu'_s$	0.2361	0.1043	0.3902	$diag(Q\Psi Q^{-1})$	0.0994	0.0935	0.1531	0.1592
$Q\alpha$	2.4258	-4.0320	0.8578	$diag(\Omega_\eta)$	0.1002	0.0484	0.0044	0.0215
	-2.1054	3.9191	-0.6196					
	-0.2904	0.0584	0.0012	$\sigma_\epsilon^2$	0.0023			
	-0.1431	0.0435	-0.1482					

Table 4: QMLEs of the U-FSN(4)-ECM(2) model with  $k = (1, 2, 3, 11)$  obtained using the training data from 1985:1 to 2000:3 inclusive as the estimation data. The operation  $diag(X)$  gives the diagonal of the matrix  $X$  as a row vector.

ECM( $p$ ) models show an even larger reduction of about 45 per cent. As in Section 3, very substantial gains over the random walk forecast of a short rate can be realised using the FSN-ECM models and the information contained in the yield curve alone. The U-FSN(4)-ECM( $p$ ) models are the best performers in terms of the  $\det(mMSFE)$  measure, with large reductions of 40 per cent compared to the  $RW^{YC}$ . The U-FSN(4)-ECM(2) model with  $k = (1, 2, 3, 11)$  featured in Figure 5 is thus a strong performer across the range of summary measures considered in Table 3. For completeness, Table 4 reports for this model the QMLEs used to produce the constant parameter (C) forecasts whose MSFEs are reported in Figure 6(d). As in Section 3.5, with the parameters of the state equation (3) set equal to these QMLEs, the knot-yields  $\{\gamma_t\}$  follow an I(1) process and the  $(m - 1)$  spreads between them,  $\beta'\gamma_t$ , are stationary cointegrating relations.

## 5 Conclusion

This paper has developed Functional Signal plus Noise (FSN) time series models for analysing the dynamics of a large cross-section of yields in which contemporaneous observations are functionally related. The FSN models specify the evolution over time of stochastic functions, a problem that has received relatively little attention in the econometrics and statistics literature, and the models may conveniently be written in linear state space form. The functional, cross-sectional relationship between contemporaneously observed yields is captured by modelling the observed curves as the sum of a smooth, cubic spline ‘signal function’ and noise. The state equation of the FSN-ECM model determines the stochastic evolution of the spline function by specifying that the yields corresponding to the knots of the spline follow a vector autoregression which may be written as an Equilibrium Correction Model in which the spreads between the

knot-yields appear as regressors.

We prove a new theorem establishing that the Expectations Theory of the term structure completely determines the conditional mean of any zero-coupon yield curve, given an information set that includes the current, complete yield curve of adequate dimension. We are thus able to derive and implement the minimal mean square error point forecasts implied by the theory. This provides a novel means of evaluating the Expectations Theory in which its 1-step ahead forecasts are compared with those of rival models. Our results, obtained using two different datasets, strongly support the growing body of empirical evidence against the theory.

The FSN-ECM models are used to forecast high-dimensional, zero-coupon yield curves of US Treasury bonds at the one month ahead horizon, and their performance compared in an out-of-sample experiment to several important rival models. The triangular FSN(6)-ECM(2) models achieve large reductions in mean square forecast error relative to a random walk for yields, especially at the shorter maturity end of the yield curve, and readily dominate the Diebold and Li (2006), Expectations Theory and random walk forecasts across all maturities. These models achieve very large reductions in the MSFE of the short rate (i.e. the 1 or 1.5 month yields), conditioning on the information in current and past yield curves alone, and also perform well in terms of the forecast evaluation measure which is the determinant of the mean square forecast error matrix. It turns out that (latent) yield spreads provide important information for forecasting the yield curve, especially in the case of shorter maturities, but not in the manner prescribed by the Expectations Theory. It is also important for forecasting to retain lagged changes in knot-yields as regressors in the ECM state equation. The constant parameter FSN-ECM forecasting models examined performed well and use an ECM state equation in which the knot-yields are  $I(1)$  and the spreads between them are stationary, cointegrating relations.

Overall, we find that the proposed FSN-ECM models readily outperform all other existing methods of which we are aware for forecasting high dimensional yield curves. For the first time in the literature, we present forecasting models whose one month ahead forecasts strongly outperform a random walk for the yield curve. Furthermore, the models are for high dimensional yield curves rather than for a small subset of maturities. Two topics in particular merit investigation in future research. First, whilst we have concentrated on demonstrating how much can be achieved in forecasting the yield curve using the information in current and past yield curves



alone, macroeconomic factors can readily be included as regressors in the state equation of the FSN-ECM models. Second, the use of non-linear FSN-ECM models in which the observation equation expresses coupon bond prices as the sum of a cubic spline discount function and noise would remove the need to estimate zero-coupon yield curves in an initial stage separate from modelling the dynamics of the term structure.

## Acknowledgments

We would like to thank Peter Andrews, Jennifer Castle, David Hendry, Siem Jan Koopman, Anthony Murphy, Bent Nielsen, Mark Salmon, Neil Shephard, Stefan de Wachter, and participants at the Warwick Frontiers in Finance 2006 conference for their comments. Any remaining errors are of course our own. Computational work was performed using Version 3.2 of Ox (Doornik 2001) and SsfPack 2.2 (Koopman, Shephard and Doornik 1999). Financial support from the ESRC under awards R42200034061 and R00023839 (Bowsher) and from the British Academy postdoctoral fellowship scheme (Bowsher and Meeks) is gratefully acknowledged.

## APPENDIX

### A A primer on cubic spline theory

A cubic spline is essentially a piecewise cubic function in which the pieces join together to form a smooth function overall.

**Definition 3** *Cubic spline on  $(k; \gamma)$ .* Consider an interval of the real line  $[a, b]$ , subdivided by a vector,  $k$ , of points

$$k = (k_j)_{j=1}^m,$$

where  $k_1 = a, k_m = b$ , and  $k_{j+1} > k_j$  for  $j = 1, \dots, m - 1$ . Each point  $k_j$  is referred to as a **knot**, and for  $m \geq 3$ , the  $(k_j)_{j=2}^{m-1}$  are called the **internal knots**. Denote by  $\gamma$  a vector of real-valued ordinates,  $(\gamma_j)_{j=1}^m$ . A function  $S(x)$  with domain  $[a, b]$  is a **cubic spline interpolating to  $\gamma$  with knots  $k$** , or more concisely a **cubic spline on  $(k; \gamma)$** , if and only if:

$$(i) S(k_j) = \gamma_j \quad (j = 1, \dots, m);$$

(ii)  $S(x)$  coincides with a polynomial of degree at most three on the sub-intervals  $[k_j, k_{j+1}]$  ( $j = 1, \dots, m - 1$ ); and

(iii)  $S(x)$  is twice continuously differentiable on  $[a, b]$ .

If, in addition, the second derivatives at the end knots,  $S''(k_1)$  and  $S''(k_m)$ , are both zero,  $S(x)$  is said to be a **natural cubic spline on  $(k; \gamma)$** .

In this context since the knots are positioned at deterministic maturities that are fixed throughout the analysis, whereas the states  $\gamma$  to which the spline interpolates are stochastic, a cubic spline  $S(x)$  on  $(k; \gamma)$  is denoted by  $S_\gamma(x)$ . The object of interest here is usually the restriction of  $S_\gamma(x)$  to a finite vector of points  $\boldsymbol{\tau} = (\tau_1, \dots, \tau_N)$ . The cubic spline is then written as the finite dimensional vector  $S_\gamma(\boldsymbol{\tau}) := (S_\gamma(\tau_1), \dots, S_\gamma(\tau_N))'$ .

A well known result that arises by combining conditions on  $S''(k_1)$  and  $S''(k_m)$  with the conditions (i), (ii) and (iii) of Definition 3 is that  $S_\gamma(\boldsymbol{\tau})$  is a linear function of the ordinate vector  $\gamma$ . This result, stated for the case of a natural cubic spline in the theorem below, allows the FSN-ECM models to be written in linear state space form.

**Theorem 2** *Let  $S_\gamma(x)$  be a natural cubic spline on  $(k; \gamma)$ , with  $k$  and  $\gamma$  vectors of dimension  $m$ . Also let  $\boldsymbol{\tau} = (\tau_1, \dots, \tau_N)$  be a finite vector of points in  $[a, b]$ , and  $S_\gamma(\boldsymbol{\tau}) := (S_\gamma(\tau_1), \dots, S_\gamma(\tau_N))'$ . Then*

$$S_\gamma(\boldsymbol{\tau}) = W(k, \boldsymbol{\tau})\gamma,$$

where the  $N \times m$  interpolation matrix,  $W(k, \boldsymbol{\tau})$ , depends only on  $\boldsymbol{\tau}$  and the knot positions  $k$ .<sup>17</sup> Details of how to compute  $W(k, \boldsymbol{\tau})$  may be found in equations (2.5), (2.6), (2.11), (2.12) and (2.14) of Poirier (1973, pp. 517-518), where  $\pi_0$  and  $\pi_k$  are set to zero in the case of a natural cubic spline.

**Proof.** See, for example, Poirier (1973, pp. 517-518). ■

## References

- 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, 745–787.

---

<sup>17</sup>In particular,  $W(k, \boldsymbol{\tau})$  does not depend on  $\gamma$ .

- Besse, P. C. and H. Cardot (1996). Approximation spline de la prévision d'un processus fonctionnel autorégressif d'ordre 1. *Canadian Journal of Statistics* 24, 467–487.
- Besse, P. C., H. Cardot, and D. B. Stephenson (2000). Autoregressive forecasting of some functional climatic variations. *Scandinavian Journal of Statistics* 27, 673–687.
- Bliss, R. R. (1997). *Testing Term Structure Estimation Methods*, Volume 9 of *Advances in Futures and Options Research*, pp. 197–231. Greenwich, Connecticut: JAI Press.
- Bowsher, C. G. (2004). Modelling the dynamics of cross-sectional price functions: an econometric analysis of the bid and ask curves of an automated exchange. Nuffield College Economics Discussion Paper, 2004-W21.
- Bowsher, C. G. and R. Meeks (2006). The (near) impossibility of stationary yield spreads and  $I(1)$  yields under the expectations theory of the term structure. Nuffield College Economics Discussion Paper, 2006-W05.
- Campbell, J. Y., A. W. Lo, and A. C. MacKinlay (1997). *The Econometrics of Financial Markets*. Princeton: Princeton University Press.
- Campbell, J. Y. and R. J. Shiller (1991). Yield spreads and interest rate movements: A bird's eye view. *Review of Economic Studies* 58, 495–514.
- Clements, M. P. and D. F. Hendry (1993). On the limitations of comparing mean square forecast errors. *Journal of Forecasting* 12, 617–637.
- Cochrane, J. H. and M. Piazzesi (2005). Bond risk premia. *American Economic Review* 95, 138–160.
- Diebold, F. X. and C. Li (2006). Forecasting the term structure of government bond yields. *Journal of Econometrics* 130, 337–364.
- Diebold, F. X., G. D. Rudebusch, and S. B. Aruoba (2006). The macroeconomy and the yield curve. *Journal of Econometrics* 131, 309–338.
- Duffee, G. R. (2002). Term premia and interest rate forecasts in affine models. *Journal of Finance* 57, 405–443.
- Fama, E. F. and R. R. Bliss (1987). The information in long-maturity forward rates. *American Economic Review* 77, 680–692.

- Hall, A. D., H. M. Anderson, and C. W. Granger (1992). A cointegration analysis of treasury bill yields. *Review of Economics and Statistics* 74, 116–126.
- Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.
- Harvey, A. C. and S. J. Koopman (1993). Forecasting hourly electricity demand using time-varying splines. *Journal of the American Statistical Association* 88, 1228–1236.
- Johansen, S. (1996). *Likelihood-based inference in cointegrated vector autoregressive models*. Oxford: Oxford University Press.
- Kargin, V. and A. Onatski (2004). Curve forecasting by functional autoregression. Mimeo, Economics Department, Columbia University.
- Koopman, S. J. and M. Ooms (2001). Time series modelling of daily tax revenues. Tinbergen Institute Discussion Paper, TI2001-032/4.
- Koopman, S. J., N. Shephard, and J. A. Doornik (1999). Statistical algorithms for models in state space using SsfPack 2.2. *Econometrics Journal* 2, 107–160.
- McCulloch, J. H. (1971). Measuring the term structure of interest rates. *The Journal of Business* 44, 19–31.
- Nelson, C. and A. Siegel (1987). Parsimonious modeling of yield curves. *Journal of Business* 60, 473–489.
- Pagan, A. R., A. Hall, and V. Martin (1996). *Modeling the Term Structure*, Volume 14 of *Handbook of Statistics*, pp. 91–118. Amsterdam: Elsevier, North-Holland.
- Poirier, D. J. (1973). Piecewise regression using cubic spline. *Journal of the American Statistical Association* 68, 515–524.
- Ramsay, J. O. and B. W. Silverman (1997). *Functional Data Analysis*. New York: Springer.
- Shea, G. (1992). Benchmarking the expectations hypothesis of the interest-rate term structure: An analysis of cointegration vectors. *Journal of Business and Economic Statistics* 10, 347–366.
- Stock, J. H. and M. W. Watson (2006). *Forecasting with Many Predictors*, Volume 1 of *Handbook of Economic Forecasting*, pp. 515–554. Amsterdam: Elsevier, North-Holland.

Swanson, N. R. and H. White (1995). A model-selection approach to assessing the information in the term structure using linear models and artificial neural networks. *Journal of Business and Economic Statistics* 13, 265–275.