

# The Dynamics of Economic Functions: Modelling and Forecasting the Yield Curve

Clive G. Bowsher\*  
Nuffield College, Oxford University

Roland Meeks†  
Federal Reserve Bank of Dallas

March 28, 2008

## Abstract

The class of Functional Signal plus Noise (FSN) models is introduced that provides a new, general method for modelling and forecasting time series of economic functions. The underlying, continuous economic function (or ‘signal’) is a natural cubic spline whose dynamic evolution is driven by a cointegrated vector autoregression for the ordinates (or ‘ $y$ -values’) at the knots of the spline. The natural cubic spline provides flexible cross-sectional fit and results in a linear, state space model. This FSN model achieves dimension reduction, provides a coherent description of the observed yield curve and its dynamics as the cross-sectional dimension  $N$  becomes large, and can feasibly be estimated and used for forecasting when  $N$  is large. The integration and cointegration properties of the model are derived. The FSN models are then applied to forecasting 36-dimensional yield curves for US Treasury bonds at the one month ahead horizon. The method consistently outperforms the Diebold and Li (2006) and random walk forecasts on the basis of both mean square forecast error criteria and economically relevant loss functions derived from the realised profits of pairs trading algorithms. The analysis also highlights in a concrete setting the dangers of attempts to infer the relative economic value of model forecasts on the basis of their associated mean square forecast errors.

*Keywords:* FSN-ECM models, functional time series, term structure, forecasting interest rates, natural cubic spline, state space form.

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

## 1 INTRODUCTION

This paper develops a novel econometric framework for modelling and forecasting time series of economic functions. In financial economics, market prices at a given point in time are a function of the characteristics of the asset traded such as its maturity date, transaction volume or strike price. It is often appropriate to consider the underlying price as a continuous function of these characteristics. Prominent examples include the zero-coupon yield curve, the ask and bid (or inverse demand and supply) curves of the limit order book of a financial exchange, and the implied volatility surface of options prices. Despite the importance of such functions there has been little development of general methods to study their dynamics that are applicable and feasible when the cross-sectional dimension of the data is moderate or large. To address this need we introduce Functional Signal plus Noise models and demonstrate their usefulness in the concrete setting of modelling and forecasting zero-coupon yield curves. The vast majority of empirical dynamic studies of the yield curve concentrate on a small subset of bond maturities as a result of the econometric difficulties

---

\*Corresponding author: Nuffield College, Oxford OX1 1NF, UK. [clive.bowsher@nuffield.ox.ac.uk](mailto:clive.bowsher@nuffield.ox.ac.uk), tel: +44 (0)1865 278969, fax: +44 (0)1865 278621. Functional Signal plus Noise (FSN) models first appeared in the earlier working paper, Bowsher (2004).

†Research Department, Federal Reserve Bank of Dallas, 2200 N. Pearl St., Dallas TX 75201, USA. [roland.meeks@dal.frb.org](mailto:roland.meeks@dal.frb.org), tel: +1 214 922 6804. The views expressed in this paper are not necessarily those of the Federal Reserve Bank of Dallas or the Federal Reserve System.

involved in modelling dynamic panel data in which cross-sections consist of functionally related observations. By contrast, our approach is based on a continuous, smooth underlying yield curve (or ‘signal’ function) that is observed with measurement error (or ‘noise’). The Functional Signal plus Noise models appear to be the first systematically to outperform a random walk for the yield curve when forecasting one month ahead, a horizon that has particularly challenged previous forecasting methods.

The main contributions of the paper may be summarised as follows. First, the class of Functional Signal plus Noise (FSN) models is introduced that provides a new, general method for modelling and forecasting time series of economic functions. The underlying continuous, smooth economic function (or ‘signal’) is a natural cubic spline whose dynamic evolution is driven by a cointegrated VAR for the ordinates (‘ $y$ -values’) at the knots of the spline. The natural cubic spline provides flexible cross-sectional fit and results in a linear state space model. Since the model’s state equation is the cointegrated VAR written and parametrised in Error Correction form (see Johansen 1996), we call it the FSN-ECM model. This model achieves dimension reduction, provides a coherent description of the observed price or yield curve and its dynamics as the cross-sectional dimension  $N$  becomes large, and can feasibly be estimated and used for forecasting when  $N$  is large. Second, under the assumption that the  $m$  knot ordinates (or yields) follow a cointegrated I(1) process with cointegrating rank  $r$ , a theorem is derived showing that the observed and latent yield curves of the FSN-ECM process with dimension  $N$  are I(1) processes with cointegrating rank  $[N - (m - r)]$ . Third, the FSN-ECM models are applied to forecasting 36-dimensional yield curves for US Treasury bonds at the one month ahead horizon. The method consistently outperforms both the widely used Diebold and Li (2006) and random walk forecasts on the basis of both mean square forecast error (MSFE) criteria and economically relevant loss functions derived from the realised profits of pairs trading algorithms that each period construct an arbitrage portfolio of discount bonds. Yield spreads are found to provide important information for forecasting the yield curve, but not in the manner prescribed by the Expectations Theory. The analysis also highlights in a concrete setting the dangers of attempts to infer the relative economic value of model forecasts on the basis of their associated MSFEs.

The work presented is a contribution to *functional time series* analysis. Each yield curve, for example, may be regarded as a finite dimensional vector, albeit of moderately or very high dimension. However, the standard approaches of multivariate time series or panel data analysis are of little help in this setting owing to the cross-sectional dimension (preventing e.g. the standard use of cointegrated VARs) and the close, functional relationship between the yields. Indeed, the analysis of time series of stochastic, continuous functions is in a state of relative infancy. 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). Kargin and Onatski (2007) develop a new technique, the predictive factor decomposition, for estimation of the autoregressive operator of FARs. The method is applied to the point prediction at a three month horizon of cubic spline interpolations of the term structure of Eurodollar futures rates, but is found to have uniformly larger MSFEs across a wide range of maturities than the Diebold and Li (2006) forecasting method that is implemented.

Our FSN-ECM models may be interpreted as a special type of dynamic factor model (see Stock and Watson 2006, p. 524) in which the ordinates (yields) at the knots of the spline are the factors and the factor loadings are determined by the requirement that the signal function (i.e. the ‘common component’) be a natural cubic spline, rather than the factor loadings being parameters for estimation. The semiparametric FSN approach thus allows quasi-maximum likelihood estimation and prediction using a linear state space form and the filtering methods of Jungbacker and Koopman (2007), even when the cross-sectional dimension ( $N$ ) of the data is very large. Our method may be contrasted with non-parametric approaches to estimation of factor models using either static or dynamic principal components in which asymptotic consistency is established for both large cross-sectional dimension  $N$  and large number of time series observations  $T$  (see Boivin and Ng 2005 for discussion of these methods from a forecasting perspective). Of course, parametric and non-parametric approaches are complementary and have well-established relative advantages and disadvantages. In the context of modelling economic functions, the FSN-ECM approach exploits the *a priori* information that the cross-section consists of (noisy) observations of a continuous, smooth underlying function. Furthermore, the FSN-ECM approach is applicable to panels with only moderate cross-sectional dimension.

A recurring theme in the term structure literature is the considerable difficulty of outperforming naïve forecasting devices, particularly the random walk for the yield curve or ‘no change’ forecast (denoted  $RW^{YC}$ ). Since this difficulty has been found to be particularly pronounced for one month ahead forecasts and term structure forecasting models are almost always specified at a monthly frequency, our study focuses on forecasting the yield curve at the one month horizon. Diebold and Li (2006) introduced 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. The method is now widely used. However, Diebold and Li (2006) report only comparable forecasting performance to  $RW^{YC}$  in terms of mean square forecast errors (MSFEs) at the one month ahead horizon. De Pooter, Ravazzolo, and van Dijk (2007) tried several variants of this forecasting method, but were unable to improve its average root MSFE performance. Almost none of the currently popular term structure forecasting methods that they implemented resulted in lower average root MSFE

than the  $RW^{YC}$  at the one month horizon, and the inclusion of macroeconomic information resulted in only marginal reductions in average root MSFE at best. Several studies have assessed the forecasting performance of affine term structure models using only a small number of yields. Duffee (2002) finds that forecasts made using the standard class of ('completely') affine models typically perform worse than the  $RW^{YC}$  even 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 these three horizons. Ang and Piazzesi (2003) present a VAR model in the affine class for yields of five different maturities which imposes no-arbitrage restrictions and performs only slightly better than the  $RW^{YC}$  at the one month ahead horizon for four of the five maturities used. Incorporating macroeconomic factors in the model improves the forecast performance for those four maturities.

For the first time in the literature, we present forecasting models that consistently outperform a random walk for the yield curve at the one month horizon, both under MSFE-based criteria and ones based on realised trading profits. Furthermore, a broad range of 36 different maturities is used in the forecast evaluation exercise. The structure of the paper is as follows. Section 2 introduces the new class of FSN-ECM models and states a theorem on their integration and cointegration properties in Section 2.2. The choice of a natural cubic spline for the continuous, economic signal function is motivated in Section 2.3 in terms of the good interpolating and approximation properties of such splines. The use of cubic splines in term structure estimation is reviewed and it is argued that potential criticisms in terms of their extrapolatory behaviour misunderstand their role here as piecewise, local approximations. Section 3 presents the specification and selection of FSN-ECM forecasting models for the zero-coupon yield curve and compares their out-of-sample performance to rival models using MSFE-based criteria. Section 4 considers forecast evaluation under economically relevant loss functions derived from the realised profits of pairs trading algorithms, whilst Section 5 concludes. The Appendix provides the necessary mathematical details on natural cubic spline functions so that the paper is self-contained. These are denoted here as a function of  $\tau$  by  $S(\tau)$ .

## 2 FUNCTIONAL SIGNAL PLUS NOISE MODELS

Functional Signal plus Noise (FSN) models provide a general method for modelling and forecasting time series of economic functions. Such a method must have the ability to fit flexibly the shape of the cross-sectional data whilst providing an accurate description of the dynamics of the time series. The FSN models employ a natural cubic spline to model the underlying, continuous economic function, the dynamic evolution of which is driven by a cointegrated Vector AutoRegression (VAR) or Error Correction Model (ECM) of relatively small

dimension. The natural cubic spline is expected to provide the property of good cross-sectional fit in a wide range of settings as a result of its desirable properties as an approximating and interpolating function. The use of the cointegrated VAR ensures that the dynamic properties of both the state (or ‘factor’) and observation vectors are well understood. Furthermore, since cointegrated VARs have been very successful as models of financial variables that may be written in terms of log prices, their incorporation is likely to result in good empirical performance. The resultant FSN-ECM model combines the virtues of parsimony and parametric interpretability. The method is applicable when the underlying economic function is believed to be smooth and, in heuristic terms, varies such that its general shape can be described by a vector of ordinates (‘ $y$ -values’) of relatively small dimension. The method then exploits the dimension-reduction property inherent in the functional data. For concreteness, the exposition below is given in terms of yield curves but is general in scope. For example, Bowsher (2004) provides an application of FSN models to the bid and ask curves of the electronic limit order book of a financial exchange.

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 here by the vector  $\mathbf{y}_t(\boldsymbol{\tau}) := (y_t(\tau_1), y_t(\tau_2), \dots, y_t(\tau_N))'$ . We develop methods for the empirically important case where the cross-sectional dimension  $N$  of the observed yield curve is too large to allow the use of standard, multivariate time series methods. In Section 3 the task will be to forecast one month ahead a 36-dimensional yield curve with  $\boldsymbol{\tau} = (1.5, 2, 3, \dots, 11, 12, 15, 18, \dots, 81, 84)$  and maturities 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). The FSN models proposed below have exactly this form and capture the functional, cross-sectional relationship between contemporaneous observations (yields) by modelling the observed curves as the sum of a smooth, latent ‘signal function’  $\mathcal{S}(\boldsymbol{\tau})$  and noise.

## 2.1 FSN-ECM models

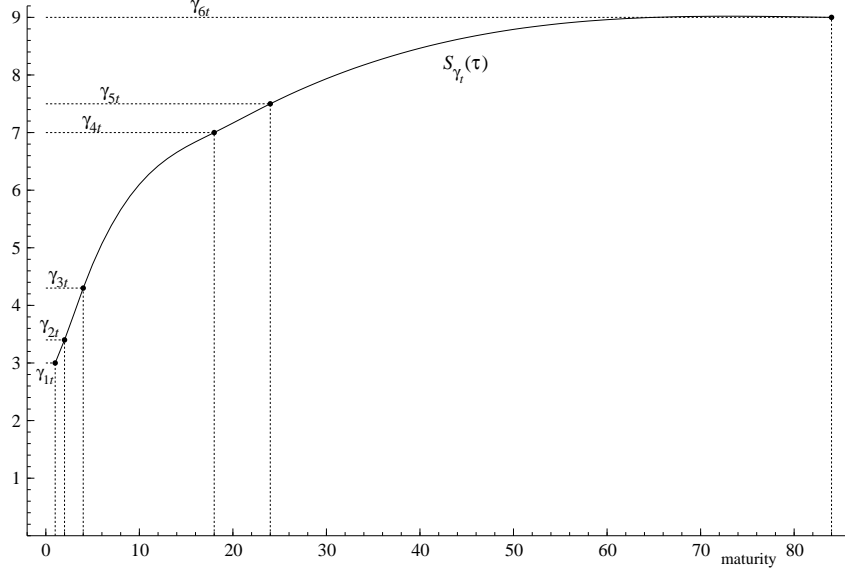
The signal function used is a natural cubic spline uniquely determined by the ordinates (yields) that correspond to the knots of the spline,  $\gamma_t$ . The state equation of the FSN model then determines the stochastic evolution of the spline function by specifying that  $\gamma_t$  follows a cointegrated VAR (or ECM). The resultant FSN-ECM model may be written in linear state space form, thus allowing the use of the Kalman filter to compute both the Gaussian quasi-likelihood function and 1-step ahead point predictions. 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 (2003), 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. The term *Kalman filter* is 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).

A natural cubic spline (NCS) is essentially a piecewise cubic function with pieces that join together to form a twice continuously differentiable function overall (see Appendix B). The NCS signal function or latent yield curve evaluated at the maturities  $\tau$  is written as  $\mathcal{S}_{\gamma_t}(\tau) := (\mathcal{S}_{\gamma_t}(\tau_1), \dots, \mathcal{S}_{\gamma_t}(\tau_N))'$ . The NCS has  $m$  knots positioned at the maturities  $\mathbf{k} = (k_1, k_2, \dots, k_m)$  which are deterministic and fixed over time. The notation  $\mathcal{S}_{\gamma_t}(\tau)$  is used to imply that the spline interpolates to the latent yields  $\gamma_t = (\gamma_{1t}, \dots, \gamma_{mt})'$  – i.e.  $\mathcal{S}_{\gamma_t}(k_j) = \gamma_{jt}$  for  $j = 1, \dots, m$ . We refer to the vector  $\gamma_t$  as the *knot-yields* of the spline. An illustrative NCS signal function is shown in Figure 1. Another terminology is to refer to  $\mathcal{S}_{\gamma_t}(\tau)$  as a NCS on  $(\mathbf{k}; \gamma_t)$ , since the spline passes through the points  $(k_j, \gamma_{jt})_{j=1}^m$ , which together determine the remainder of the spline function uniquely. The vectors  $\mathbf{k}$  and  $\tau$  need not have any elements in common and, in the context of FSN-ECM models, the knot vector typically has much smaller dimension  $m$  than the number of maturities  $N$ .

A formal definition of the FSN( $m$ )-ECM( $p$ ) model now follows, where  $m$  is the number of knots and the  $p$ th lag of  $\gamma_{t+1}$  is the maximum lag that enters the ECM state equation. Without loss of generality, we focus on the case  $p \leq 2$  for expositional simplicity. The definition is stated for the case where the maturities  $\tau$  for which yields are observed in the data are constant over time. Extension to allow for a deterministic, time-varying maturity vector  $\tau_t$  is straightforward both conceptually and computationally (even when its dimension varies over time), but this is not needed in the sequel.

**Definition 1 FSN( $m$ )-ECM(2) Model.** *Let the vector of observed maturities be  $\tau = (\tau_1, \dots, \tau_N)$ , with  $\tau_1 \geq k_1 = 1$  and  $\tau_N \leq k_m$ . The model for the time series of  $N$ -dimensional observation vectors  $\mathbf{y}_t(\tau) := (y_t(\tau_1), y_t(\tau_2), \dots, y_t(\tau_N))'$  is*

Figure 1: An illustrative NCS signal function or latent yield curve,  $S_{\gamma_t}(\tau)$ .



Note: The yields at the knots,  $\gamma_{jt}$ , are labelled and plotted using filled circles; the knots are given by  $\mathbf{k} = (1, 2, 4, 18, 24, 84)$ .

given by

$$\begin{aligned} \mathbf{y}_t(\boldsymbol{\tau}) &= \mathbf{S}_{\gamma_t}(\boldsymbol{\tau}) + \boldsymbol{\varepsilon}_t(\boldsymbol{\tau}) \\ &= \mathbf{W}(\mathbf{k}, \boldsymbol{\tau}) \boldsymbol{\gamma}_t + \boldsymbol{\varepsilon}_t(\boldsymbol{\tau}), \end{aligned} \quad (2)$$

$$\Delta \boldsymbol{\gamma}_{t+1} = \boldsymbol{\alpha}(\boldsymbol{\beta}' \boldsymbol{\gamma}_t - \boldsymbol{\mu}) + \boldsymbol{\Psi} \Delta \boldsymbol{\gamma}_t + \mathbf{v}_t, \quad (3)$$

for  $t = 1, 2, \dots$ . Here  $\mathbf{S}_{\gamma_t}(\boldsymbol{\tau})$  is a natural cubic spline on  $(\mathbf{k}; \boldsymbol{\gamma}_t)$ , the  $N \times m$  deterministic matrix  $\mathbf{W}(\mathbf{k}, \boldsymbol{\tau})$  is given by Lemma 7 of the Appendix, and both  $\boldsymbol{\alpha}$  and  $\boldsymbol{\beta}$  are  $m \times r$  matrices with rank equal to  $r < m$ . Letting  $\mathbf{A}_\gamma(z)$  denote the characteristic polynomial for the process  $\{\boldsymbol{\gamma}_t\}$ , it is imposed that  $|\mathbf{A}_\gamma(z)| = 0$  implies that  $|z| > 1$  or  $z = 1$ .

The initial state  $(\boldsymbol{\gamma}'_1, \boldsymbol{\gamma}'_0)'$  has finite first and second moments given by  $\boldsymbol{\gamma}^*$  and  $\boldsymbol{\Omega}^*$  respectively. The series  $\{\mathbf{u}_t := (\boldsymbol{\varepsilon}_t(\boldsymbol{\tau})', \mathbf{v}_t')'\}$  has a finite second moment for all  $t$  and satisfies, for all  $t$ ,  $\mathbf{E}[\mathbf{u}_t] = \mathbf{0}$ ,  $\mathbf{E}[\boldsymbol{\varepsilon}_t(\boldsymbol{\tau}) \boldsymbol{\varepsilon}_t(\boldsymbol{\tau})'] = \boldsymbol{\Omega}_\varepsilon$ ,  $\mathbf{E}[\mathbf{v}_t \mathbf{v}_t'] = \boldsymbol{\Omega}_v$ ,  $\mathbf{E}[\boldsymbol{\varepsilon}_t(\boldsymbol{\tau}) \mathbf{v}_t'] = \mathbf{0}$ ,  $\mathbf{E}[\mathbf{u}_t \mathbf{u}_s'] = \mathbf{0} \forall s \neq t$ , and  $\mathbf{E}[\mathbf{u}_t (\boldsymbol{\gamma}'_1, \boldsymbol{\gamma}'_0)'] = \mathbf{0}$ . Note that  $\{\mathbf{u}_t\}$  is a vector white noise process. The parameters of the model are thus  $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\Psi}, \boldsymbol{\Omega}_v, \boldsymbol{\Omega}_\varepsilon)$ .

The **Gaussian FSN(m)-ECM(2) Model** has the additional condition imposed that both  $\mathbf{u}_t$  and  $(\boldsymbol{\gamma}'_1, \boldsymbol{\gamma}'_0)'$  have multivariate Normal distributions.

The choice of a NCS as the signal function or latent yield curve  $\mathbf{S}(\boldsymbol{\tau})$  is discussed in detail in Section 2.3 below, and stems from the desirable properties that the NCS has as a smooth approximating and interpolating function. Note that the latent signal function is expected to be smoother than the observed 'curve'  $\mathbf{y}_t(\boldsymbol{\tau})$  and captures the underlying economic function of interest. The ECM state equation (3) is motivated by the numerous studies that successfully model a relatively small vector of yields using a cointegrated VAR in

which the process for the yields, here  $\{\gamma_t\}$ , is  $I(1)$  and  $\{\beta' \gamma_t - \mu\}$  is a stationary, mean zero vector of cointegrating relations. Given the reduced rank  $r < m$  of both  $\alpha$  and  $\beta$  imposed by Definition 1, a necessary and sufficient condition for  $\gamma_t$  to have these properties is that  $\det[\alpha'_\perp(\mathbf{I} - \Psi)\beta_\perp] \neq 0$ . Note that then  $\mu = E[\beta' \gamma_t]$  is the stationary mean of the cointegrating relations and  $E[\Delta\gamma_{t+1}] = 0$ , thus excluding deterministic trends.

The FSN-ECM model combines the virtues of parsimony and parametric interpretability. Dimension reduction is achieved provided that  $m \ll N$  and the approach remains feasible as the cross-sectional dimension  $N$  of the observed yield curve becomes large. Consider varying the vector of observed maturities  $\tau$  in Definition 1 by increasing  $N$  but holding the maximum observed maturity  $\tau_N$  fixed so that the yield curve is ‘filled in’. Since the ECM state equation remains unchanged when the model is well-specified, and provided that the latent signal function captures all or most of the cross-sectional dependence (e.g. the number of parameters in  $\Omega_\epsilon$  does not vary with  $N$ ), the FSN-ECM model continues to provide a parsimonious forecasting model. Jungbacker and Koopman (2007) provide new, computationally efficient methods for implementing the Kalman filter for dynamic factor models that are applicable to the FSN-ECM model and are ideally suited to the large  $N \gg m$  case just described. QML estimation and prediction for the FSN-ECM model thus remain computationally feasible in this case. Using the FSN-ECM forecasting models described in Section 3.2, we would expect the method to be computationally feasible for, say,  $N \approx 5000$ .

The Expectations Theory (ET) of the term structure implies for an  $I(1)$ ,  $m$ -dimensional vector of observed yields that the cointegrating rank  $r$  is  $(m - 1)$  and that  $(m - 1)$  linearly independent yield spreads are cointegrating relations. These hypotheses have received considerable attention in the cointegration-based yield curve literature (see *inter alia* Hall, Anderson, and Granger 1992, Shea 1992 and Pagan, Hall, and Martin 1996) with the findings supporting  $r = (m - 1)$  or  $r = (m - 2)$  and cointegrating relations that can be written as linear combinations of yield spreads. There is a tendency to reject the implications of the ET using hypothesis testing techniques based on asymptotic critical values. However, it is important to pay attention to the validity of the critical values used. For example, Pagan, Hall, and Martin (1996) report rejection of the standard hypothesis test but note that the point estimates obtained are quite close to the situation where the cointegrating relations implied by the ET hold. They highlight the major impact on the critical values of a levels effect of the short rate in the disturbance of the VAR, which may result in the test rejecting erroneously. More generally, the conclusions of in-sample hypothesis testing do not necessarily carry over to the context of out-of-sample forecasting using potentially mis-specified models.

The ECM state equation has the advantage that the dynamic properties of both the knot-yields  $\gamma_t$  and the yield curves are well understood. The following section derives the integration and cointegration properties



of the latent and observed yield curves of an FSN-ECM process.

## 2.2 FSN-ECM model dynamics

We begin from the empirically well-motivated assumption that the process for the  $m$  knot-yields  $\{\gamma_t\}$  is a cointegrated  $I(1)$  process and prove that the latent and observed yield curves are also then cointegrated  $I(1)$  processes with cointegrating rank  $[N - (m - r)]$ , where  $r = \text{rank}(\alpha) = \text{rank}(\beta)$ . The  $(m - r)$  common trends of the knot-yields, and of the latent and observed yield curves are identical and given by  $\alpha'_\perp \sum_{i=1}^t v_i$ . (When  $\delta$  is an  $n \times r$  matrix of full rank, we use the notation  $\delta_\perp$  for some  $n \times (n - r)$  matrix of full rank such that  $\delta' \delta_\perp = \mathbf{0}$ ).

**Theorem 2 (Integration and Cointegration Properties of FSN-ECM Model)** *Let  $\mathbf{y}_t(\tau)$  follow an FSN( $m$ )-ECM(2) process with  $\det[\alpha'_\perp(\mathbf{I} - \Psi)\beta_\perp] \neq \mathbf{0}$ . Then  $\{\gamma_t\}$  is  $I(1)$  and  $\{\beta'\gamma_t - \mu\}$  is a stationary, mean zero vector of cointegrating relations. It follows from Definition 1 that the processes for the observed and latent yield curves,  $\{\mathbf{y}_t(\tau)\}$  and  $\{\mathcal{S}_{\gamma_t}(\tau)\}_{t=1,2,\dots}$  resp., are  $I(1)$  cointegrated processes with the matrix of cointegrating vectors given by  $\phi = [\mathbf{W}(\mathbf{k}, \tau)\beta_\perp]_\perp$  in both cases. The cointegrating rank is thus equal to  $\text{rank}(\phi) = N - (m - r)$  in both cases.*

**Proof.** The necessity and sufficiency of the condition  $\det[\alpha'_\perp(\mathbf{I} - \Psi)\beta_\perp] \neq \mathbf{0}$  follows directly from Theorem 4.2 of Johansen (1996), as does the moving average (MA) representation

$$\gamma_t = \mathbf{C} \sum_{i=1}^t v_i + \mathbf{B}(L)(v_t - \alpha\mu) + \mathbf{D}, \quad (4)$$

where  $\mathbf{D}$  depends on initial values such that  $\beta'\mathbf{D} = \mathbf{0}$ ,  $\mathbf{C} = \beta_\perp[\alpha'_\perp(\mathbf{I} - \Psi)\beta_\perp]^{-1}\alpha'_\perp$ , and the power series for  $\mathbf{B}(z)$  is convergent for  $|z| < 1 + \delta$  for some  $\delta > 0$ .

We thus have the following MA representation for the latent yield curve

$$\mathcal{S}_{\gamma_t}(\tau) = \mathbf{W}(\mathbf{k}, \tau)\mathbf{C} \sum_{i=1}^t v_i + \mathbf{W}(\mathbf{k}, \tau)\mathbf{B}(L)(v_t - \alpha\mu) + \mathbf{W}(\mathbf{k}, \tau)\mathbf{D}. \quad (5)$$

Notice that  $\text{rank}[\mathbf{W}(\mathbf{k}, \tau)] = m$  and  $\text{rank}[\mathbf{W}(\mathbf{k}, \tau)\mathbf{C}] = m - r > 0$ . In particular,  $\mathbf{W}(\mathbf{k}, \tau)\mathbf{C} \neq \mathbf{0}$ . It follows immediately that both  $\mathcal{S}_{\gamma_t}(\tau)$  and  $\mathbf{y}_t(\tau) = \mathcal{S}_{\gamma_t}(\tau) + \varepsilon_t$  are  $I(1)$  processes (since the sum of an  $I(1)$  process and an  $I(0)$  process is itself  $I(1)$ ).

Note also that  $\mathbf{W}(\mathbf{k}, \tau)\beta_\perp$  has full rank equal to  $(m - r)$  and hence  $\text{rank}(\phi) = N - (m - r)$ . The process  $\{\phi'\mathcal{S}_{\gamma_t}(\tau)\}$  is  $I(0)$  since  $\phi'\mathbf{W}(\mathbf{k}, \tau)\mathbf{C} = \phi'\mathbf{W}(\mathbf{k}, \tau)\mathbf{D} = \mathbf{0}$  and hence

$$\phi'\mathcal{S}_{\gamma_t}(\tau) = \phi'\mathbf{W}(\mathbf{k}, \tau)\mathbf{B}(L)(v_t - \alpha\mu). \quad (6)$$

The process  $\{\phi'\mathbf{y}_t(\tau)\}$  is also  $I(0)$  since  $\mathbf{E}[\varepsilon_t v_s'] = \mathbf{0} \forall s, t$ . ■

For a particular choice of  $\beta$  the matrix of cointegrating vectors for the yield curve may thus be computed as  $\phi = [\mathbf{W}(\mathbf{k}, \tau)\beta_\perp]_\perp$ . The simple case where  $\beta'\gamma_t$  consists of the  $(m - 1)$  spreads between the knot-yields is given in the following example.

**Example 3 (FSN-ECM Model with Stationary Yield Spreads)** Consider the case of the FSN( $m$ )-ECM(2) process in Theorem 2 with the  $m \times (m-1)$  matrix  $\beta$  set to  $\beta_s$  where  $\beta_s' \gamma_t = (\gamma_{j+1,t} - \gamma_{jt})_{j=1}^{m-1}$  is the stationary vector of spreads between the knot-yields. Then  $\mathbf{W}(\mathbf{k}, \tau) \beta_{\perp}$  is an  $N$ -vector with all elements identical and we can take  $\phi' \mathbf{y}_t(\tau) = [y_t(\tau_i) - y_t(\tau_1)]_{i=2}^N$ . Thus, e.g. when  $\tau_1 = 1$ , the observed spreads  $[y_t(\tau_i) - y_t(1)]_{i=2}^N$  are a stationary vector for any  $N$ .

## 2.3 Cubic Spline Signal Functions

We motivate here the choice of a natural cubic spline for the economic signal function or latent yield curve  $S(\tau)$ , and discuss the use of cubic splines in term structure estimation. A spline may usefully be viewed as a set of polynomial pieces each of which is a *local* approximation to the function of interest, with the polynomial pieces joined together to form a smooth function overall. Spline functions are a centrepiece of the modern theory that deals with the numerical approximation of functions (see Powell 1996). Cubic splines are used frequently in practice since they provide a balance between accurate approximation and smoothness.

Suppose that a latent yield curve  $y_*(\tau)$  is in  $C^2[1, k_m]$ . Then it is known (see Powell 1996, Theorem 20.3) that the least maximum or ‘minimax’ error achievable by a cubic spline approximation to  $y_*(\tau)$  on  $[1, k_m]$ , with arbitrary number and positioning of knots, has the upper bound  $3h^2 \sup_{\tau \in [1, k_m]} \{y_*''(\tau)\}$  where  $h$  is the maximum interval between adjacent knots. Thus the family of cubic splines is able, using a twice continuously differentiable spline function, to approximate  $y_*(\tau)$  to any required accuracy by using a sufficiently large number of knots (perhaps spaced uniformly, although this is usually sub-optimal). This approximation ability of cubic splines motivates their use as the signal function in FSN models. A question of importance in the forecasting context is then whether sufficiently accurate cross-sectional fit can be achieved using a number of knots that also allows formulation of a parsimonious FSN-ECM model, a question that we are able to answer strongly in the affirmative in what follows when forecasting the yield curve.

A slightly different way to motivate the use of a natural cubic spline as the latent, smooth economic function in the FSN-ECM model is as follows. Suppose that the ‘true’ latent yields at  $t$  corresponding to the maturities  $\mathbf{k} = (1, k_2, \dots, k_m)$  are known to be  $\gamma_t = (\gamma_{1t}, \dots, \gamma_{mt})'$  and one seeks an interpolating function  $S(\tau) \in C^2[1, k_m]$  that passes through the points  $(k_j, \gamma_{jt})_{j=1}^m$ . Then it is known that, of *all* functions in  $C^2[1, k_m]$ , the one that minimises the roughness penalty  $\int_1^{k_m} [S''(\tau)]^2 d\tau$  is the NCS  $S_{\gamma_t}(\tau)$  (see Powell 1996, Theorem 23.2). In this sense, the NCS is the *least* rough or ‘oscillatory’ choice. If the latent economic function is believed to be smooth and its ‘general shape’ can be described by a vector of ordinates  $\gamma_t$  of relatively small dimension, then a NCS interpolating between the  $\gamma_{jt}$  is a good way to describe the function. Problems would arise if the latent economic functions vary rapidly over certain maturity ranges and those ranges also change significantly over time, thus necessitating a large number of knots  $m$  to avoid oversmoothing. Such a large  $m$  would then result

in too many parameters associated with the ECM state equation for the FSN-ECM method to be useful for forecasting, unless very long time series were available.

Under the conditions of Definition 1, the FSN-ECM model can be written in *linear* state space form (see Harvey 1989, pp. 100-104) since the deterministic matrix  $\mathbf{W}(\mathbf{k}, \tau)$  depends only on the vector of maturities  $\tau$  and the knot positions  $\mathbf{k}$ , thus allowing the NCS signal function  $S_{\gamma_i}(\tau)$  to be written as the linear function  $\mathbf{W}(\mathbf{k}, \tau) \gamma_i$  of the knot-yields. This is a further advantage of the use of a NCS in the FSN-ECM model since the linear state space form enables use of the Kalman filter to perform both quasi-maximum likelihood estimation (QMLE) and 1-step ahead, linear point prediction. The state vector at  $t$  can be taken to be  $(\gamma'_{t'}, \gamma'_{t-1})'$  or, in an isomorphic representation, to be  $(\gamma_{1t}, (\boldsymbol{\beta}' \gamma_t)', \gamma_{1,t-1}, (\boldsymbol{\beta}' \gamma_{t-1})')'$  – see also equation (7). We use the latter in our computational work.

We consider now the use of cubic splines to estimate the zero-coupon term structure at some point in time. The flexible McCulloch (1975) procedure fits the *discount* function using a cubic spline regression, is widely used and continues to be regarded as a leading method amongst the existing parsimonious methods (see Jeffrey, Linton, and Nguyen 2006 and Bliss 1997). Both Fisher, Nychka, and Zervos (1995) and Waggoner (1997) use smoothing splines to penalise large variations in the estimated forward rate curve that can result from over-fitting. However, Waggoner (1997, p.14) concludes that the results produced by the McCulloch (1975) procedure were very similar both in terms of fit and smoothness to those obtained using the variable roughness penalty (VRP), smoothing spline method. Jeffrey, Linton, and Nguyen (2006) show that a recently developed, non-parametric kernel smoothing method tends to perform better than the McCulloch (1975) regression spline. However, it is not at all clear how to extend this method to incorporate time series dynamics. To the best of our knowledge, there is no previous work that directly fits the *yield curve* using a cubic spline as we do here. Such an approach automatically ensures that the implied discount function is everywhere positive and is equal to unity for  $\tau = 0$ .

Section 3.5 below will demonstrate that natural cubic splines provide a better cross-sectional fit to a widely used dataset of Unsmoothed Fama Bliss yields than the popular Nelson and Siegel (1987) functional form – as shown in Figure 5, the average over time of the squared OLS errors for the fitted yields (the darkly shaded ‘static component’ there) is smaller for all maturities in the case of the NCS. The number and position of the knots used for the NCS in Figure 5(a) were determined using a different, non-overlapping dataset (see Section 3.3) and also result in a parsimonious, FSN-ECM out-of-sample forecasting model. As noted by Waggoner (1997, p.14), any tendency towards excessive variation or ‘oscillation’ of a regression cubic spline can be controlled through the number and spacing of the knots. In our procedure, that number is automatically kept

low by the requirement that the FSN-ECM state equation not involve too many parameters to be effective in out-of-sample forecasting. The cross-sectional fit obtained from the regressions is still good, suggesting that the direct use of cubic splines to fit *yield curves* using coupon bond price data deserves attention in future research.

Cubic spline methods are sometimes criticised in terms of the divergent behaviour of the *extrapolated* term structure as the maturity tends to infinity. A number of points may be made in connection with such criticisms. First, cubic splines should not be used for such extrapolation since they are intended as piecewise, local approximations. The spline approximation is designed to hold over a bounded interval rather than globally. In our method, the availability of additional data on very long term yields would naturally prompt the addition of further knots in order to allow local approximation over the new maturity range, rather than extrapolation beyond what was previously an 'end knot'. Second, our method is designed to forecast yields for maturities that lie within (or close to) the previously observed range on the basis of the past data. Extrapolation is not the usual aim either of term structure estimation or forecasting methods and, if required, would employ alternative, tailored methods. Third, Campbell, Lo, and MacKinlay (1997, p.413) warn with good reason that, "[...] yield curves should be treated with caution if they are extrapolated beyond the maturity of the longest traded [and observed] bond." To criticise the use of cubic splines for term structure forecasting on the basis of their extrapolation behaviour would be to put them to a use for which they are not intended or designed.

### 3 FORECASTING YIELD CURVES

The new FSN-ECM models are now applied to forecasting the zero-coupon yield curve of US Treasury bonds. The task set is a difficult one, namely to forecast one month ahead a 36-dimensional yield curve. The out-of-sample performance of the FSN-ECM models is compared below to the main competing models. In addition to the MSFE-based criteria considered in this section, forecast evaluation using carefully constructed, economically relevant loss functions based on realised trading profits is undertaken in Section 4. The data used 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 Zero-coupon yield curve data

We use the same dataset of Unsmoothed Fama Bliss (UFB) forward rates as Diebold and Li (2006), which runs 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 (see Bliss 1997). Zero-coupon

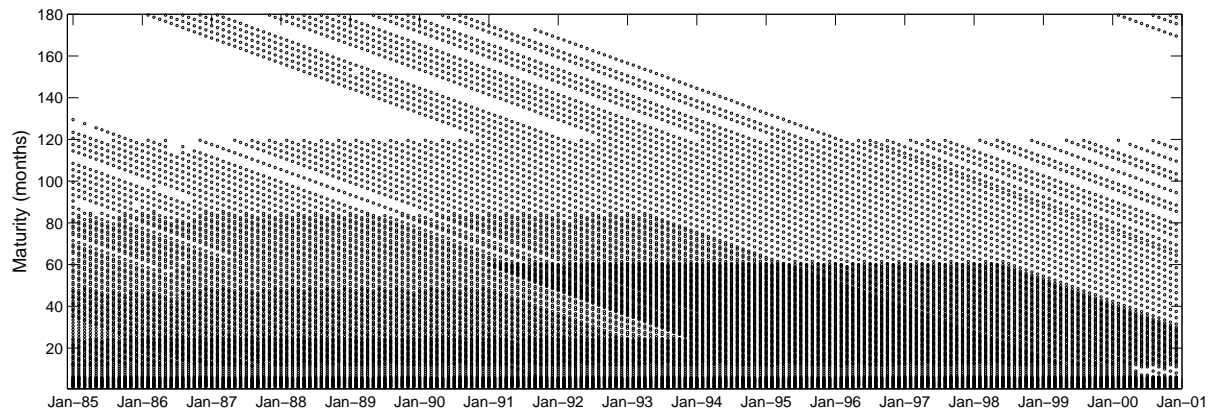
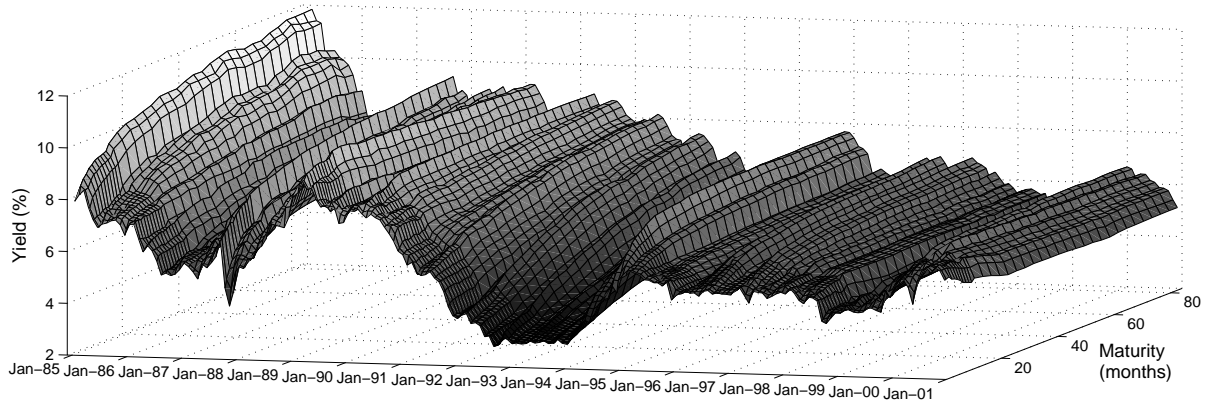
UFB yields are then obtained by averaging the appropriate UFB forward rates. As is discussed below, the set and number of maturities for which yields are observed is not the same for every  $t$ . Our FSN-ECM models can readily accommodate this feature using a time-varying but deterministic matrix  $\mathbf{W}(\mathbf{k}, \boldsymbol{\tau}_t)$  in the observation equation (3) and the particularly straightforward means available for dealing with missing observations when using the Kalman filter. However, we work instead 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 one month is taken to equal 30.4375 days. Importantly, 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 and 84 months 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 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 one month yield using this dataset since there is frequently no observed maturity less than or equal to 30.4375 days. Diebold and Li (2006, Table 1) provides descriptive statistics for a subset of our maturities.

Amongst US government bonds, Treasury Bills are pure discount bonds whilst others are coupon bearing. Thus the zero-coupon yield curve must usually 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 well. 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 a cross-sectional estimation of the zero-coupon 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 and employ a non-linear observation equation that expresses the observed coupon bond prices as the sum of the coupon payments priced using the latent yield curve and measurement error. Fitting such a non-linear

Figure 2: Zero-coupon, Unsmoothed Fama Bliss yields on US Treasury bonds



*Note:* The upper panel shows a 3-dimensional plot of the dataset used in Sections 3 and 4; yields are measured in percentage points per annum. The lower panel plots as circles the maturities directly observed at each date. For further details on the data set (ufb2full.dat), see the notes accompanying the “Bliss Term Structure Generating Programs.”

FSN-ECM model makes coherent use of both the time series and cross-sectional information in the data and would allow forecasting of coupon bond prices. This extension is left to future research.

### 3.2 FSN-ECM forecasting models

The FSN( $m$ )-ECM( $p$ ) forecasting models used have  $m \in \{5, 6\}$  and  $p \in \{1, 2\}$ . A number of *a priori* parameter restrictions are imposed on Definition 1 to obtain the parsimonious forecasting models used in Sections 3 and 4. The matrix  $\beta$  is set to  $\beta_s$ , where  $\beta_s' \gamma_t = (\gamma_{j+1,t} - \gamma_{jt})_{j=1}^{m-1}$  is the vector of spreads between the knot-yields. This choice is motivated by the well-known predictive ability of yield spreads in forecasting the yield curve and the interpretability of the resultant FSN-ECM model. It is important to note that setting  $\beta$  equal to  $\beta_s$  does *not* imply that the yield curve of the FSN-ECM model satisfies the Expectations Theory (ET). As is clearly shown in Section 3.4, the FSN-ECM model forecasts obtained with  $\beta$  equal to  $\beta_s$  differ greatly from those obtained under the ET (see Figure 3 of that section in particular). Section 3.4 and Appendix A discuss computation of the ET forecasts, and the latter also presents a simple example of an FSN-ECM process with  $\beta = \beta_s$  that does not satisfy the ET.

Additional parameter restrictions are expressed in terms of the non-singular matrix  $\mathbf{Q}$ , where  $\varphi_t := \mathbf{Q}\gamma_t$  is the transformed state vector consisting of the (latent) short rate and inter-knot (latent) yield spreads

$$\varphi_t := \left( \gamma_{1t}, \gamma_{2t} - \gamma_{1t}, \dots, \gamma_{mt} - \gamma_{m-1,t} \right)' = \begin{pmatrix} 1 & \mathbf{0}_{1 \times (m-1)} \\ \beta_s' \end{pmatrix} \gamma_t = \mathbf{Q}\gamma_t. \quad (7)$$

The ECM state equation may then be written equivalently as the VAR

$$\Delta\varphi_{t+1} = \mathbf{Q}\alpha(\beta_s'\mathbf{Q}^{-1}\varphi_t - \mu) + \mathbf{Q}\Psi\mathbf{Q}^{-1}\Delta\varphi_t + \eta_t, \quad (8)$$

where  $\eta_t = \mathbf{Q}\mathbf{v}_t$ , and we define  $\mathbf{\Omega}_\eta = \text{Var}[\eta_t] = \mathbf{Q}\mathbf{\Omega}_v\mathbf{Q}'$ . Diagonality of the covariance matrix  $\mathbf{\Omega}_\eta$  is imposed rather than, for example, diagonality of  $\mathbf{\Omega}_v$ , which is less plausible. The covariance matrix of the measurement error  $\mathbf{\Omega}_\epsilon = \sigma_\epsilon^2 \mathbf{I}_N$  and thus has one free parameter,  $\sigma_\epsilon^2$ . Working as in Eq. (8) with a transformed state vector  $\varphi_t$  that includes the cointegrating relations  $\beta_s' \gamma_t$  has general utility for developing parameter restrictions in other areas of application of FSN-ECM models.

FSN( $m$ )-ECM(1) models are obtained by setting  $\Psi = \mathbf{0}$ , whilst we impose that  $\mathbf{Q}\Psi\mathbf{Q}^{-1}$  is diagonal in all FSN( $m$ )-ECM(2) models. The latter restriction means that only *own* lagged changes of the short rate or spreads enter each equation in (8). The matrix  $\alpha$  in (3) determines the loadings on the spread regressors,  $\beta_s' \gamma_t$ . 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 five 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 models the change in the  $j$ th 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  $\Psi = \mathbf{0}$ ,  $\Delta\gamma_{m,t+1}$  follows a random walk.

The approach taken was thus in line with the ‘‘Keep it sophisticatedly simple’’ or KISS principle of Zellner (1992). Given the amount of data available it was necessary to impose some *a priori* structure on the FSN-ECM models in order to obtain reasonably precise parameter estimates. It is evident from the empirical results reported in Sections 3.4 and 4 that the parameter restrictions discussed above work well in forecasting and result in FSN-ECM models that compare very favourably to existing competitor models. Different choices, for example of  $\beta$ , may perhaps in future turn out to perform better but are clearly not necessary to establish the utility of the FSN-ECM models for yield curve forecasting.

The parameters of the restricted 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 are the 1-step ahead point predictions given by the Kalman filter,  $[\hat{y}_{t+1}(\tau)|y_t(\tau), \dots, y_1(\tau); \hat{\theta}_t]_{KF}$ , with the parameter vector of the model set equal to the QMLE,  $\hat{\theta}_t$ , based on data up to and including time  $t$ . Such forecasts are often denoted by  $\hat{y}_{t+1|t}(\tau)$  in what follows. In all cases the Kalman filter is initialised using  $(\gamma'_1, \gamma'_0)' \sim (\gamma^*, \Omega^*)$ , where  $\Omega^* = \mathbf{0}$  and  $\gamma^*$  is set equal to the yields,  $(y_0(\mathbf{k})', y_{-1}(\mathbf{k})')'$ , 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). This initialisation procedure 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 procedure used is motivated by the approximation that the observed knot-yields follow a random walk for the two 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 one month yields are not observed in the dataset used there).

### 3.3 Knot 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 ( $\mathbf{k}$ ) 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,  $\mathbf{k}$ . Specifically, we fit using OLS a natural cubic spline with knot vector  $\mathbf{k}$  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 were determined in terms of minimisation of the mean RSS amongst all possible  $\mathbf{k}$  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 present in our dataset, but excluding the shortest and longest maturities. Values of  $m$  smaller than five were found to result in much larger mean RSS and were not considered further.

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  $\mathbf{k}$ '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.

### 3.4 MSFE-based forecast evaluation

The forecast performance of the FSN-ECM models with the knot vectors selected above was then compared to the forecasts from three rival models: a RW for the yield curve (the 'no change' forecast,  $RW^{YC}$ ), the Diebold and Li (2006) dynamic Nelson-Siegel model (henceforth DNS) and a forecast that embodies the full implications of the Expectations Theory (ET). 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 forecasts 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 (for comparability with the FSN-ECM forecasts). 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 eight different maturity groupings, owing to the higher dimension of the yield curve in this setting. (Each maturity in  $\{1.5, 2, 3, 4\}$  has its own parameter; a single parameter then corresponds to each of the maturity groupings  $\{5, 6, \dots, 10\}$ ,  $\{11, 12, 15, \dots, 24\}$ ,  $\{27, \dots, 75\}$ , and  $\{78, \dots, 84\}$ ). The state equation was initialised using the unconditional mean and variance of the state vector, as in Diebold, Rudebusch, and Aruoba (2006).

Appendix A shows that forecasting Eq. (16) of Lemma 4 there is equivalent to the Expectations Theory (ET) holding. Implementation of these forecasts is thus an ideal way to evaluate and make comparisons with the ET in this context. Since our dataset does not include a one month yield, we produce ET forecasts of  $\mathbf{y}_t(3 : 84)$  based on Eq. (16) with the first two rows excluded, where  $\mathbf{y}_t(3 : 84) := (y_t(3), y_t(4), \dots, y_t(84))'$ . 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. The vector of term premia,  $\rho(3 : 85) := (\rho(3), \dots, \rho(85))'$ , is estimated by OLS using the following regression derived from equation (16)

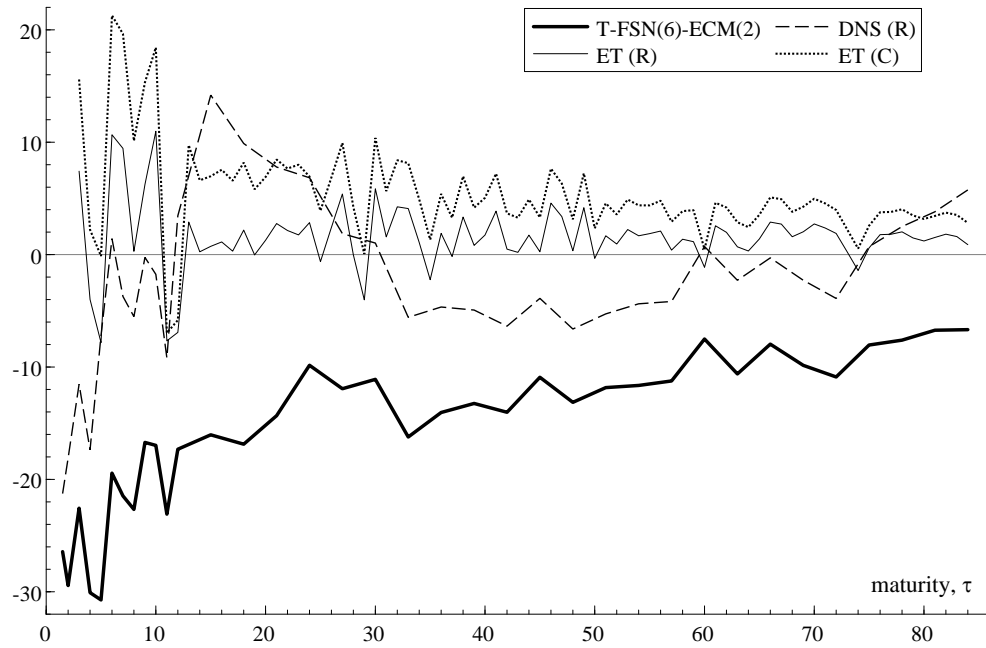
$$\Delta \mathbf{y}_{t+1}(3 : 84) = \alpha_{3:84}^{ET} [\mathbf{s}_t(3 : 85) - \rho(3 : 85)] + \mathbf{v}_{t+1}, \quad (9)$$

where the vector of yield spreads  $\mathbf{s}_t(3 : 85) := (s_t(3, 1), \dots, s_t(85, 1))'$ ,  $\alpha_{3:84}^{ET}$  denotes the 3rd to 84th rows inclusive of  $\alpha_{84}^{ET}$  in equation (16), 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 Lemma 7 of Appendix B 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  $\mathbf{k} = (1, 2, 4, 18, 24, 84)$ ; the DNS model; and the ET forecasting equation in (9). 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. The percentage increase in MSFE relative to the  $RW^{YC}$  is used as the evaluation criterion rather than the MSFE itself because this measure is invariant whether one considers the MSFE for forecasts of yields  $y_{t+1}(\tau)$ , log holding period returns  $r_{t+1}(\tau)$ , or excess log returns  $r_{t+1}(\tau) - y_t(1)$  (a fact which follows from Eq. 13). Such invariance is clearly a desirable property of evaluation criteria in this context. (The values of the MSFEs themselves for the T-FSN(6)-ECM(2) model with  $\mathbf{k} = (1, 2, 4, 18, 24, 84)$  and the DNS model are plotted as the uppermost lines in Figures 5(a) and 5(b) respectively).

Inspection of Figure 3 reveals that the triangular FSN(6)-ECM(2) model outperforms all of the rival models,

Figure 3: Percentage increase in model MSFEs by maturity relative to those of the  $RW^{YC}$

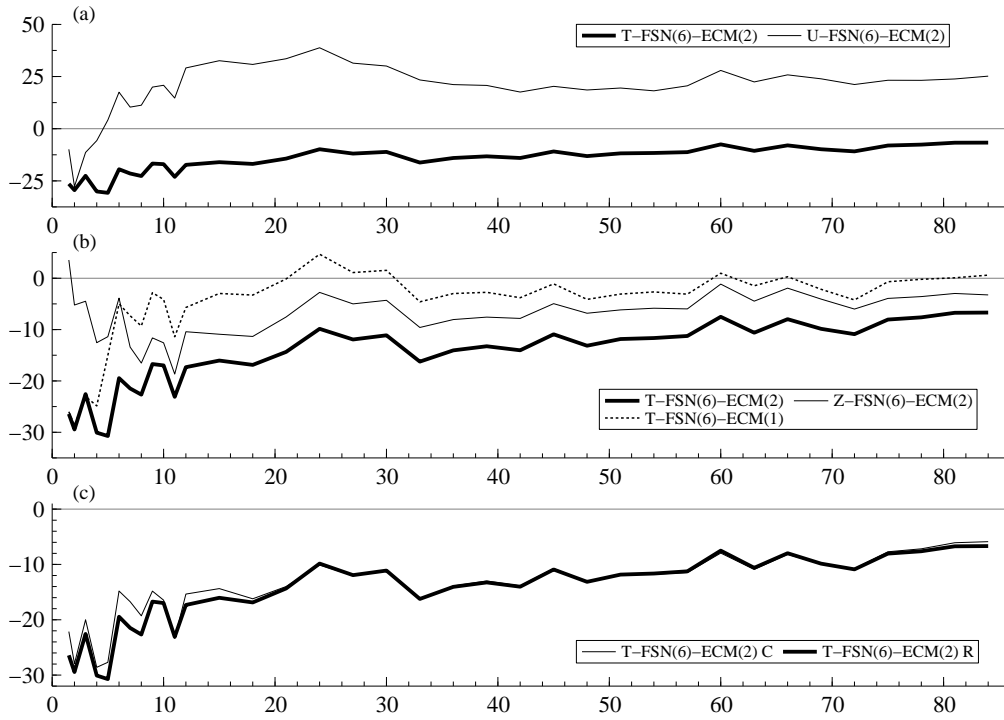


Note: Shown are the results for the triangular (T-) FSN(6)-ECM(2) model with  $k = (1, 2, 4, 18, 24, 84)$ , the Diebold and Li (2006) DNS model and the Expectations Theory (ET) forecasting equation (9). (R) stands for recursive estimation and (C) for forecasts produced using constant parameters. The horizontal axis is maturity measured in months.

including the  $RW^{YC}$ , at all maturities. Furthermore, the percentage reductions obtained in MSFE compared to the  $RW^{YC}$  are 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 most prominent for forecasting yield curves of moderate to high dimension and is widely used. The authors report better performance for the method at forecast horizons of 6 and 12 months ahead than at the one month ahead horizon.

Interestingly, 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 implied by the ET (Lemma 4) is far from being the optimal MSFE predictor. This novel method for evaluating the ET thus finds that the theory is very wide of the mark for US Treasury data and the maturities studied.

Figure 4: Percentage increase in FSN(6)-ECM( $p$ ) model MSFEs by maturity relative to those of the RW<sup>YC</sup>



Note: The knot vector used in all cases is  $\mathbf{k} = (1, 2, 4, 18, 24, 84)$ . (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.

Figure 4 presents analogous plots for FSN(6)-ECM( $p$ ) models, all of which have  $\mathbf{k} = (1, 2, 4, 18, 24, 84)$  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 unrestricted (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  $\Psi = \mathbf{0}$  and  $\alpha = \mathbf{0}$ . Imposing  $\Psi = \mathbf{0}$  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  $\alpha = \mathbf{0}$  (to obtain the Z- or Zero model) is somewhat less costly for maturities greater than or equal to seven months, but is drastically costly for the shortest maturities. Thus inclusion of the spreads  $\beta'_s \gamma_t$  as regressors in the ECM 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  $\mathbf{MSFE}$ ), 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  $tr(\mathbf{MSFE})$  can in principle change when the data to be forecast is expressed as a vector consisting of the shortest yield and spreads relative to that yield, 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).

The reported average MSFEs reflect 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, five 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  $\det(\mathbf{MSFE})$  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  $\mathbf{k} = (1, 2, 4, 18, 24, 84)$  shown in Figure 3) 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 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

Table 1: Summary measures of forecast performance.

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>									
$\mathbf{k} = (1, 2, 4, 18, 24, 84)$									
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
$\mathbf{k} = (1, 2, 4, 12, 30, 84)$									
FSN(6)-ECM(2)	R	T	86.8	75.2	87.6	0.018	0.090	0.062	0.061
$\mathbf{k} = (1, 2, 4, 15, 24, 84)$									
FSN(6)-ECM(2)	R	T	85.9	74.6	87.1	0.018	0.099	0.061	0.066
$\mathbf{k} = (1, 2, 4, 27, 84)$									
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>									
$RW^{YC}$	-	-	$64.9 \times 10^{-3}$	$73.4 \times 10^{-3}$	$1.02 \times 10^{-3}$	0.013	0.305	0.061	0.132
DNS	R	-	97.9	79.1	86.6	0.026	0.318	0.056	0.125
DNS	C	-	97.2	73.9	86.6	0.034	0.309	0.054	0.120

Note: The FSN( $m$ )-ECM( $p$ ) models are grouped according to the knot vector  $\mathbf{k}$ ; 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  $[det(\mathbf{MSFE})]^{1/N}$ . With the exception of the  $RW^{YC}$  for which the measures are reported directly, all 3 MSFE-based measures are expressed as a percentage of the corresponding measure for the  $RW^{YC}$ . 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.

perform better than the  $RW^{YC}$  and very similarly to the DNS methods in terms of the  $\det(\mathbf{MSFE})$  measure. Section 4 will demonstrate that such a T-FSN(6)-ECM(2) model also strongly outperforms the DNS and  $RW^{YC}$  models in terms of the real-time profitability of various trading algorithms implemented using the same dataset. The following section performs a direct comparison between the T-FSN(6)-ECM(2) and DNS models of the components of their MSFEs for different maturities. This decomposition provides insight into the reasons underlying the comparative performance of the models.

### 3.5 MSFE Decomposition

The dominance of the DNS model by the T-FSN(6)-ECM(2) model in terms of MSFEs across maturity could be due primarily to an improved cross-sectional fitting of the yield curve by use of the natural cubic spline instead of the Nelson-Siegel curve, or due primarily to superior forecasting of the (model-specific) latent factors, or a result of some combination of the two. In order to address this question the following decomposition of the MSFE for maturity  $\tau$  is performed.

$$MSFE(\tau) = R^{-1} \sum_{r=1}^R [y_r^{OLS}(\tau) - y_r(\tau)]^2 + R^{-1} \sum_{r=1}^R [\hat{y}_{r|r-1}(\tau) - y_r^{OLS}(\tau)]^2 + 2R^{-1} \sum_{r=1}^R [\hat{y}_{r|r-1}(\tau) - y_r^{OLS}(\tau)][y_r^{OLS}(\tau) - y_r(\tau)], \quad (10)$$

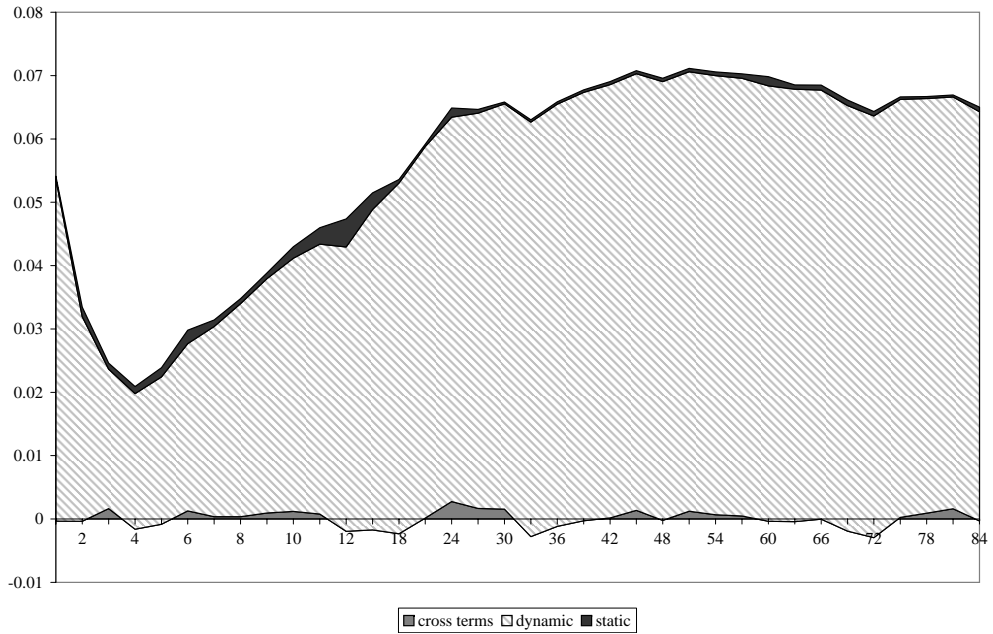
where  $r$  indexes the observations of the forecast evaluation period and  $y_r^{OLS}(\tau)$  is the fitted value of  $y_r(\tau)$  resulting from the OLS fitting of either a natural cubic spline with  $\mathbf{k} = (1, 2, 4, 18, 24, 84)$  (FSN-ECM) or Nelson-Siegel curve (DNS) to the time  $r$  yield curve *alone*. The first or ‘static’ term in Eq. (10) is thus the time average of squared OLS residuals for maturity  $\tau$  – it measures the mean square error made in ‘forecasting’ the observed  $y_r(\tau)$  when the latent (model specific) yield curve at time  $r$  is essentially known. The second or ‘dynamic’ term measures the mean square error in forecasting that latent yield curve on the basis of time  $(r - 1)$  information. The third or ‘cross’ term is negative when the OLS residual and the (latent) forecast error tend to have opposite signs.

In order to fit a Nelson-Siegel curve by OLS we fix the exponential decay parameter  $\lambda$  to a predetermined value as in Diebold and Li (2006, Section 3.2). The value chosen is the QMLE of  $\lambda$  obtained using the training data alone, i.e. the one used for the constant parameter (C) DNS forecasts in Table 1, namely  $\lambda = 0.0766$ . Forecasts  $\hat{y}_{r|r-1}(\tau)$  and their associated  $MSFE(\tau)$ s for the DNS model were then also computed using recursively updated estimates for all parameters of the Nelson-Siegel curve except  $\lambda$ . The  $MSFE(\tau)$ s thus obtained were identical to those where the estimate of  $\lambda$  was also updated recursively, i.e. the DNS(R) method of Figure 3.

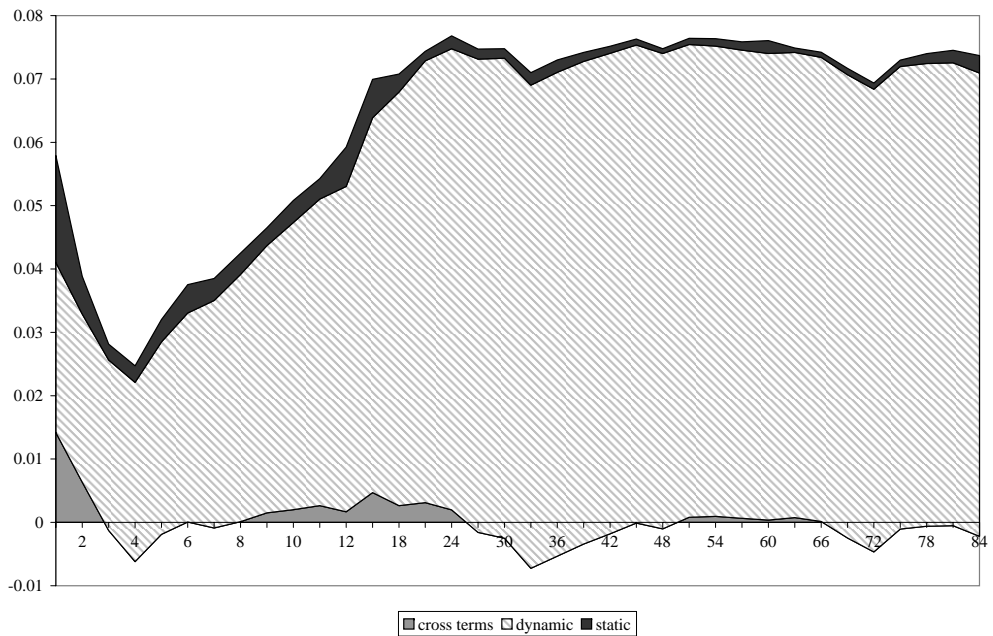
Figure 5 shows the contribution of the static, dynamic and cross terms of Eq. (10) to the total  $MSFE(\tau)$  for both the T-FSN(6)-ECM(2) model with  $\mathbf{k} = (1, 2, 4, 18, 24, 84)$  and the DNS model (upper and lower panels

Figure 5: MSFE decomposition

(a) T-FSN(6)-ECM(2)



(b) DNS



Note: The height at each  $\tau$  of a given shaded area is equal to the magnitude of that particular component of  $MSFE(\tau)$ . The static component is given by  $R^{-1} \sum_{r=1}^R [y_r^{OLS}(\tau) - y_r(\tau)]^2$ , the dynamic component by  $R^{-1} \sum_{r=1}^R [\hat{y}_{r|\tau-1}(\tau) - y_r^{OLS}(\tau)]^2$ , and the cross term component by  $2R^{-1} \sum_{r=1}^R [\hat{y}_{r|\tau-1}(\tau) - y_r^{OLS}(\tau)][y_r^{OLS}(\tau) - y_r(\tau)]$  (see Eq. 10).



respectively). The height at each  $\tau$  of a given shaded area is equal to the magnitude of that particular component (static, dynamic, or cross term). The dynamic component emerges for both models as the largest positive contributor to the total  $MSFE(\tau)$  for all maturities  $\tau$ . For all maturities that satisfy  $3 \leq \tau \leq 84$ , both the static and dynamic components are positive for DNS and FSN-ECM, and larger for DNS than FSN-ECM. Furthermore, the differences between these components (defined as ‘DNS minus FSN-ECM’) are always larger for the dynamic than the static component, and where the difference in cross terms is negative it is more than offset by the positive difference in the dynamic components. In the case of  $\tau = 1.5$  and  $\tau = 2$ , the static component is much larger for DNS than FSN but this difference is more than offset by the dynamic component, the larger  $MSFE(\tau)$  for DNS then being the result of its large, positive cross term for these maturities.

We conclude that for most maturities, both static and dynamic components contribute to the DNS model having the larger  $MSFE(\tau)$ , but that inferior forecasting of the model-specific, latent factors is more important in this regard. This illustrates a strength of the FSN-ECM framework highlighted previously – namely that a very considerable body of literature exists on how empirically to model economic variables such as log yields using cointegrated VARs. By contrast, the DNS slope and curvature factors are highly specialised to the yield curve context and much less is known about how to specify time series models of factors such as these.

### 3.6 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, again with  $\mathbf{k} = (1, 2, 4, 18, 24, 84)$ . Recall that the MSFEs for constant and recursive estimates in Figure 4(c) are very similar. With the parameters of the ECM state equation (3) set equal to the QMLEs used for the constant parameter forecasts, the knot-yields  $\gamma_t$  follow an I(1) process and the  $(m - 1)$  spreads between them  $\beta'_s \gamma_t$  are cointegrating relations. This follows since the ranks of  $\hat{\alpha}$  and  $\hat{\alpha} \beta'_s$  are both equal to five, the roots  $z$  of the characteristic polynomial of the VAR in (3) then satisfy either  $z = 1$  or  $|z| > 1$ , and the determinant of  $[\hat{\alpha}'_{\perp}(\mathbf{I} - \Psi)\beta'_{s\perp}]$  is non-zero. The characteristic polynomial of the VAR in (3) thus has exactly one unit root.

Table 2 reports the QMLEs of the transformed state equation (8), together with  $\hat{\sigma}_{\varepsilon}^2$ . Recall that the state vector in (8) is  $\varphi_t = [\gamma_{1t}, (\beta'_s \gamma_t)']$ , the vector consisting of the latent short rate and inter-knot yield spreads. The estimates of the stationary means of the spreads  $\beta'_s \gamma_t$  are all positive, implying that the latent yield curve is ‘upward sloping’ on average. It follows from (8) that

$$\Delta\gamma_{1,t+1} = (\mathbf{Q}\alpha)_{[1]}(\beta'_s \gamma_t - \mu) + \Delta\gamma_{1,t} + \eta_{1t}, \quad (11)$$

Table 2: QMLEs of the triangular FSN(6)-ECM(2) model with  $\mathbf{k} = (1, 2, 4, 18, 24, 84)$  obtained using the training data from 1985:1 to 2000:12 inclusive as the estimation data

$\boldsymbol{\mu}$	$\mathbf{Q}\boldsymbol{\alpha}$					$\text{diag}(\mathbf{Q}\boldsymbol{\Psi}\mathbf{Q}^{-1})$	$\text{diag}(\boldsymbol{\Omega}_\eta)$
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
$\sigma_\varepsilon^2$							
0.0032							

Note: The operation  $\text{diag}(\mathbf{X})$  gives the diagonal of the matrix  $\mathbf{X}$  as a column vector.

where  $(\mathbf{Q}\boldsymbol{\alpha})_{[1]}$  denotes the first row of the matrix  $\mathbf{Q}\boldsymbol{\alpha}$ . Bearing in mind that  $\mathbf{k} = (1, 2, 4, 18, 24, 84)$ , the estimates of the row  $(\mathbf{Q}\boldsymbol{\alpha})_{[1]}$  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  $\mathbf{Q}\boldsymbol{\alpha}$  are close to zero, indicating that the spread between the fifth and six knot-yields is unimportant as a regressor in Eq. (8). It also follows from (8) that the vector of spreads follows the VAR

$$\Delta(\boldsymbol{\beta}'_s \boldsymbol{\gamma}_{t+1}) = (\mathbf{Q}\boldsymbol{\alpha})_{[2:6]}(\boldsymbol{\beta}'_s \boldsymbol{\gamma}_t - \boldsymbol{\mu}) + (\mathbf{Q}\boldsymbol{\Psi}\mathbf{Q}^{-1})_{[2:6]} \Delta(\boldsymbol{\beta}'_s \boldsymbol{\gamma}_t) + \boldsymbol{\eta}_{[2:6],t} \quad (12)$$

where  $\mathbf{X}_{[2:6]}$  denotes the 2nd to 6th rows of some matrix,  $\mathbf{X}$ . 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  $(\mathbf{Q}\boldsymbol{\alpha})_{[2:6]}$  is upper triangular and  $(\mathbf{Q}\boldsymbol{\Psi}\mathbf{Q}^{-1})_{[2:6]}$  is diagonal) and the same spread at time  $(t - 1)$ . Note that all elements of the estimate of  $\mathbf{Q}\boldsymbol{\Psi}\mathbf{Q}^{-1}$  are positive and of the estimate of the diagonal of  $(\mathbf{Q}\boldsymbol{\alpha})_{[2:6]}$  are negative.

## 4 PROFIT-BASED FORECAST EVALUATION

It is now widely recognised that when comparing forecasting models, each of which may be to some extent mis-specified, no close relationship is guaranteed between model evaluations based on conventional error-based measures such as MSFE and those based on the *ex post* realised profit (or utility) from using each model's forecasts to solve a given economic decision or trading problem. Leitch and Tanner (1993) make just this point in the context of interest rate forecasting. More recently, Granger and Pesaran (2000) have argued in favour of a closer link between the forecast evaluation and decision problems. It is perhaps uncomfortable but nonetheless unavoidable that orderings of forecasting models can vary depending upon the use for which they are intended.

This section therefore examines to what extent the superior, MSFE-based performance of the triangular

FSN(6)-ECM(2) model reported above carries over to a forecast evaluation method based on realised trading profits. Such a method requires the selection of a particular trading problem. We focus on the month-by-month construction of arbitrage portfolios of zero-coupon bonds. In a simple case (Algorithm 1 below), monthly profit then equates to the negative of a particular loss function for evaluating the forecast of the sign of the excess return on the long bond (see Eq. 14). The loss function punishes incorrect forecasts proportionally to the absolute magnitude of the realised return. It is demonstrated in this case that despite the relatively small percentage reduction in the MSFE of the relevant excess return compared to the  $RW^{YC}$ , the FSN-ECM forecasting model results in a much higher realised profit (smaller loss) than both the DNS and  $RW^{YC}$  models.

We retain for this analysis the constant maturity, unsmoothed Fama Bliss zero-coupon yield dataset described in Section 3.1, thus allowing direct comparisons to be drawn with the forecasting results presented in Section 3.4. Since the zero-coupon bonds of this dataset are synthetic rather than corresponding to traded securities, we neither claim that the profit numbers reported below were necessarily available to market participants in real time nor attempt to quantify likely transaction costs. Rather, the results allow the relative performance of the forecasting models to be assessed under loss functions whose form is well motivated economically. The type of transactions we envisage could be implemented in practice by trading in Treasury STRIPS (Separate Trading of Registered Interest and Principal of Securities), which allow investors easily and at negligible cost to create and trade zero-coupon bonds. We expect that the natural cubic spline of the FSN-ECM model would be sufficiently flexible to avoid problems that the Nelson-Siegel curve has in fitting some features of STRIPS data (see Sack 2000), but leave investigation of this aspect to future research.

We denote by  $R_{t+1}(\tau) := [P_{t+1}(\tau - 1)/P_t(\tau)] - 1$  the 1 month net holding period return (HPR) on a bond with  $\tau$  months to maturity (realised at time  $t + 1$ ), and define the corresponding log HPR  $r_{t+1}(\tau) := \log[1 + R_{t+1}(\tau)]$ . Table 3 depicts the cashflows resulting from what we term a pairwise arbitrage portfolio, formed at time  $t$  by trading in a pair of zero-coupon bonds with maturities  $\tau_2$  and  $\tau_1$  ( $\tau_2 > \tau_1$ ). The dollar cashflow at  $t$  resulting from the position taken in the  $\tau_1$ -month bond is denoted  $d_t$ , with  $d_t > 0$  representing a (possibly short) sale and  $d_t < 0$  a purchase. The portfolio is designed as self-financing, i.e. the time  $t$  cashflow relating to the  $\tau_2$ -month bond is  $-d_t$ , this being achieved by matching the purchase (sale) of  $|d_t|/P_t(\tau_1)$  units of the  $\tau_1$ -month bond by the sale (purchase) of  $|d_t|/P_t(\tau_2)$  units of the  $\tau_2$ -month bond. The absolute value  $|d_t|$  is referred to as the size of the position and is fixed at  $\$1,000,000 = \$1M$  for convenience in what follows. We stipulate that the positions are always unwound at date  $t + 1$ , whereupon net cashflow is seen to equal  $d_t[R_{t+1}(\tau_2) - R_{t+1}(\tau_1)] \simeq d_t[r_{t+1}(\tau_2) - r_{t+1}(\tau_1)]$ .

Table 3: Pairwise arbitrage portfolio cashflows

Maturity	Quantity	Cashflow (\$)	
		$t$	$t + 1$
$\tau_1$	$ d_t /P_t(\tau_1)$	$d_t$	$-[d_t/P_t(\tau_1)]P_{t+1}(\tau_1 - 1)$
$\tau_2$	$ d_t /P_t(\tau_2)$	$-d_t$	$[d_t/P_t(\tau_2)]P_{t+1}(\tau_2 - 1)$
Net Cashflow (\$)		0	$d_t[R_{t+1}(\tau_2) - R_{t+1}(\tau_1)]$

*Note:* Component and net dollar cashflows at times  $t$  and  $t + 1$  from a pairwise arbitrage portfolio of discount bonds with maturities  $\tau_1$  and  $\tau_2$ .  $R_{t+1}(\tau)$  denotes the 1 month net holding period return (HPR) realised at  $t + 1$  for a bond with maturity  $\tau$  at time  $t$ . The dollar cashflow at  $t$  resulting from the position taken in the  $\tau_1$ -month bond is denoted  $d_t$ , with  $d_t > 0$  representing a (possibly short) sale and  $d_t < 0$  a purchase.

A positive net cashflow may thus be generated by correctly forecasting the *sign* of the realised excess return on the  $\tau_2$ -month bond over the  $\tau_1$ -month bond, and choosing  $d_t$  to have that same sign. Since the realised excess return is very well approximated by  $r_{t+1}(\tau_2) - r_{t+1}(\tau_1)$ , we generate the forecasts

$$\hat{r}_{t+1|t}(\tau) = \tau y_t(\tau) - (\tau - 1)\hat{y}_{t+1|t}(\tau - 1), \quad \tau = \tau_1, \tau_2, \quad (13)$$

where the yield forecast  $\hat{y}_{t+1|t}(\tau - 1)$  of Section 3.4 has been substituted into the identity for log HPRs, and then forecast the sign of the relevant excess return using the forecast  $\tilde{d}_t = \text{sgn}[\hat{r}_{t+1|t}(\tau_2) - \hat{r}_{t+1|t}(\tau_1)]$ . We then set  $d_t = \$1M \cdot \tilde{d}_t$ .

We consider three algorithms, each of which involves forming various arbitrage portfolios at each time  $t$ . The first two algorithms arose from the observation that over our training period 1985:1 to 1993:12, the average value of absolute, realised excess returns  $|R_{t+1}(\tau_2) - R_{t+1}(\tau_1)|$  is increasing in the maturity of the longer bond  $\tau_2$  and decreasing in the shorter maturity  $\tau_1$ . The largest such average value over the training period was 1.9% per month for the 82- and 3-month bonds, which are respectively the longest and shortest maturities for which we are able to compute HPRs in our dataset. Algorithms 1 and 2 thus emphasise pairwise arbitrage portfolios of long and short ( $\tau_1 = 3$ ) maturity bonds, whilst Algorithm 3 involves choosing  $\tau_2$  by optimising the forecast absolute excess return  $|\hat{r}_{t+1|t}(\tau_2) - \hat{r}_{t+1|t}(3)|$  at each time  $t$ . All three algorithms thus arose from and are motivated by *a priori* considerations not involving data from the evaluation period. The first and simplest algorithm is as follows:

**Algorithm 1 (Wide term spread)** *At all times  $t$  form a pairwise arbitrage portfolio with  $\tau_1 = 3$ ,  $\tau_2 = 82$  and  $d_t$  chosen as follows. If*

$$\begin{aligned} \hat{r}_{t+1|t}(82) - \hat{r}_{t+1|t}(3) &> 0 \quad \text{then} \quad d_t = \$1M \\ \hat{r}_{t+1|t}(82) - \hat{r}_{t+1|t}(3) &< 0 \quad \text{then} \quad d_t = -\$1M; \end{aligned}$$

*else do nothing ( $d_t = \$0$ ). At date  $t + 1$ , close the position. The dollar profit realised at  $t + 1$  is given by  $\Pi_{t+1} =$*

$$d_t[R_{t+1}(82) - R_{t+1}(3)].$$

A natural generalisation of Algorithm 1 involves taking positions on bonds across a broader range of maturities for which a return prediction is made. We again restrict attention to  $\tau_1 = 3$  and rather than taking equally sized positions, weight in favour of pairwise arbitrage portfolios with high, in-sample average absolute, realised excess returns.

**Algorithm 2 (Weighted pairs trades)** *At all times  $t$  and for  $j = 1, \dots, 34$ , form the  $j$ th pairwise arbitrage portfolio with  $\tau_1 = 3$ ,  $\tau_{2j}$  given by the  $j$ th element of the vector  $(3, 4, \dots, 13, 16, \dots, 79, 82)'$  and  $d_{jt}$  chosen as follows. If*

$$\begin{aligned} \hat{r}_{t+1|t}(\tau_{2j}) - \hat{r}_{t+1|t}(3) &> 0 \quad \text{then} \quad d_{jt} = \$1M \times w_j \\ \hat{r}_{t+1|t}(\tau_{2j}) - \hat{r}_{t+1|t}(3) &< 0 \quad \text{then} \quad d_{jt} = -\$1M \times w_j; \end{aligned}$$

*else do nothing ( $d_{jt} = \$0$ ). Letting  $\overline{RX^a}(\tau_{2j}, 3)$  denote the in-sample average of the absolute, realised excess returns  $|R(\tau_{2j}) - R(3)|$  over the training period, the weights  $w_j$  are given by  $w_j = \overline{RX^a}(\tau_{2j}, 3) / \sum_j \overline{RX^a}(\tau_{2j}, 3)$ . At date  $t + 1$ , close all positions. The dollar profit realised at  $t + 1$  is given by  $\Pi_{t+1} = \sum_j d_{jt}[R_{t+1}(\tau_{2j}) - R_{t+1}(3)]$ .*

Note that in Algorithm 2,  $\sum_j |d_{jt}| = \$1M$ , thus ensuring that the total size of the positions taken at  $t$  is the same in all three algorithms. In practice, the weights  $w_j$  increase monotonically from 0.001 for  $\tau_{2j} = 4$  to 0.065 for  $\tau_{2j} = 82$ . The final algorithm we consider selects  $\tau_2$  in order to maximise the forecast absolute excess return.

**Algorithm 3 (Optimised pairs trade)** *At all times  $t$  form a pairwise arbitrage portfolio with  $\tau_1 = 3$ ,  $\tau_{2t} = \arg \max_{\tau \neq 3} |\hat{r}_{t+1|t}(\tau) - \hat{r}_{t+1|t}(3)|$  and  $d_t$  chosen as follows. If*

$$\begin{aligned} \hat{r}_{t+1|t}(\tau_{2t}) - \hat{r}_{t+1|t}(3) &> 0 \quad \text{then} \quad d_t = \$1M \\ \hat{r}_{t+1|t}(\tau_{2t}) - \hat{r}_{t+1|t}(3) &< 0 \quad \text{then} \quad d_t = -\$1M; \end{aligned}$$

*else do nothing ( $d_t = \$0$ ). At date  $t + 1$ , close the positions. The dollar profit realised at  $t + 1$  is given by  $\Pi_{t+1} = d_t[R_{t+1}(\tau_{2t}) - R_{t+1}(3)]$ .*

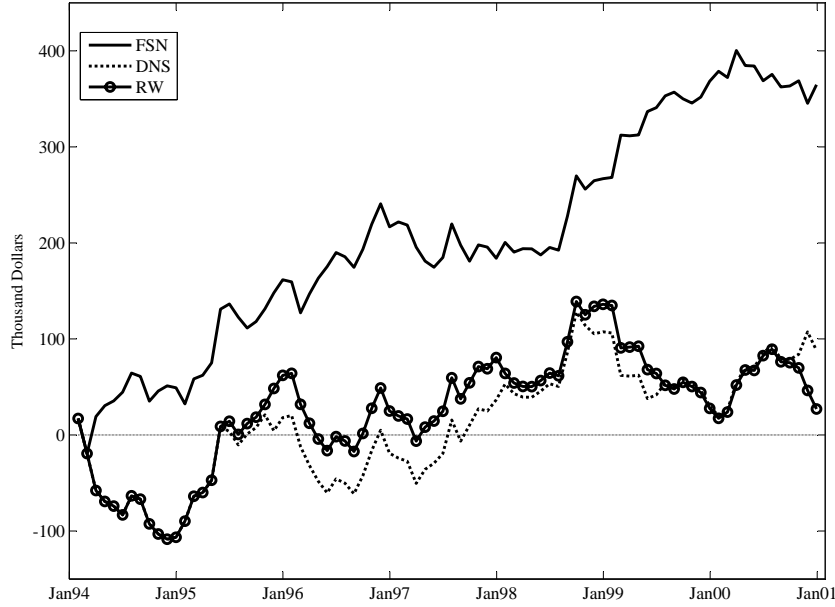
The performance of the three trading algorithms was analysed during the evaluation period 1994:1 to 2000:12 using the triangular FSN(6)-ECM(2) model to forecast returns, and compared to the performance obtained using the DNS and random walk ( $RW^{YC}$ ) forecast models. Recursively updated, out-of-sample forecasts for yields are produced as in Section 3.4 for the DNS and T-FSN(6)-ECM(2) models, with  $\mathbf{k} = (1, 2, 4, 18, 24, 84)$  used for the latter (as in Figure 3). Time series of cumulative profits under Algorithm 1 are plotted in Figure 6. The FSN-ECM model strongly outperforms the DNS and  $RW^{YC}$  models resulting in a final cumulative profit equal to \$364,700, that is 4.1 and 13.5 times that for DNS and  $RW^{YC}$  respectively. The cumulative profit series generated by the naïve, 'no-change in yield'  $RW^{YC}$  forecasts tracks that for the DNS model quite closely for much of the evaluation period. The percentage gains in average monthly profit over the  $RW^{YC}$  from using FSN-ECM and DNS forecasts are 1246% and 225% respectively. These may be

Table 4: Arbitrage portfolio trading algorithm profits, January 1994-December 2000

	Cumulative profit (\$)	Median monthly profit (\$)	Percentiles (\$)	Directional accuracy +	Directional accuracy -
			10th 90th		
Algorithm 1 (Term spread):					
FSN	364,700	4,456	-16,700 26,010	29/44 (65.9%)	21/40 (52.5%)
DNS	88,030	143	-22,150 21,430	37/44 (84.1%)	5/40 (12.5%)
RW <sup>YC</sup>	27,090	-711	-23,070 19,980	36/44 (81.8%)	4/40 (10%)
Algorithm 2 (Weighted pairs):					
FSN	223,400	2,033	-9,192 15,010	1159/1501 (77.2%)	563/1271 (44.3%)
DNS	50,710	-123	-8,111 12,460	1174/1501 (78.2%)	302/1271 (23.8%)
RW <sup>YC</sup>	-17,820	-597	-11,930 11,450	1145/1501 (76.3%)	266/1271 (20.9%)
Algorithm 3 (Optimised pair):					
FSN	316,100	4,488	-14,880 21,100	31/46 (67.4%)	21/38 (55.3%)
DNS	-22,870	-769	-20,490 18,180	28/42 (66.7%)	12/42 (28.6%)
RW <sup>YC</sup>	-32,000	678	-17,400 13,810	45/48 (93.8%)	2/36 (5.56%)

Note: Algorithms 1-3 are described in the text. The column headed *Directional accuracy +* reports, in the case of positive realised excess returns, the proportion of the forecasts of the sign of the excess return(s) that were correct (whilst *Directional accuracy -* reports the same for negative realised excess returns). For Algorithm 2, directional accuracy is computed using totals across all maturity pairs. For Algorithm 3 the maturity  $\tau_2$  of the long bond varies both across time and models. FSN denotes the Triangular FSN(6)-ECM(2) model with  $\mathbf{k} = (1, 2, 4, 18, 24, 84)$ , DNS the Dynamic Nelson-Siegel model and RW<sup>YC</sup> the random walk for yields. All excess return forecasts are made recursively out-of-sample using the yield curve forecasts of Section 3.4.

Figure 6: Cumulative trading profits for Algorithm 1, January 1994-December 2000



Note: Shows the cumulated dollar profits earned over time under trading Algorithm 1 using pairwise arbitrage portfolios constructed from the short (3 month) and long (82 month) bonds. FSN denotes the Triangular FSN(6)-ECM(2) model with  $k = (1, 2, 4, 18, 24, 84)$ , DNS the Dynamic Nelson-Siegel model and RW the random walk for yields. All forecasts are made recursively out-of-sample using the yield curve forecasts of Section 3.4.

compared with the percentage *improvement* in the MSFE of the excess return  $[R_{t+1}(82) - R_{t+1}(3)]$  obtained using the same underlying forecast  $[\hat{r}_{t+1|t}(82) - \hat{r}_{t+1|t}(3)]$ , which are equal to 6.6% and -3.8% for FSN-ECM and DNS respectively. Such a comparison demonstrates in a concrete, applied setting how misleading it can be to attempt to infer the relative economic value of model forecasts from their associated MSFEs.

Table 4 reports summary statistics for the empirical distributions of monthly profits under the three algorithms and for the directional accuracy of the forecasts  $\tilde{d}_t$  of the sign of the excess returns. Under Algorithm 1, the median monthly profit for FSN-ECM is 31 times that for DNS, whilst that for the  $RW^{YC}$  is negative, and graphical inspection of the empirical distributions of monthly profits (not shown) reveals that FSN-ECM first order stochastically dominates both alternatives. Note that although Algorithm 1 requires forecasts of only two yields as input, the FSN-ECM and DNS models both utilise the history of the complete range of maturities in the information set to form their predictions. Under Algorithms 2 and 3, cumulative profit for FSN-ECM has the same order of magnitude as under Algorithm 1 and continues to strongly outperform DNS, which in turn again outperforms  $RW^{YC}$ . Both algorithms result in lower cumulative profit for all three models. Algorithm 3 is perhaps the most demanding in terms of forecast model specification. It results in a decrease in cumulative profit relative to Algorithm 1 of \$110,900 (126%) for DNS compared to just \$48,600 (13.3%) for FSN-ECM. The mean and median monthly profit for DNS both become negative. For FSN-ECM, this optimising algorithm results in a marginally higher median monthly profit, narrower 90%

interquartile range and increased directional accuracy of excess return forecasts.

In order better to understand the relative performance of the three models, note that for a single pairwise arbitrage portfolio the realised cumulative profit may be written as

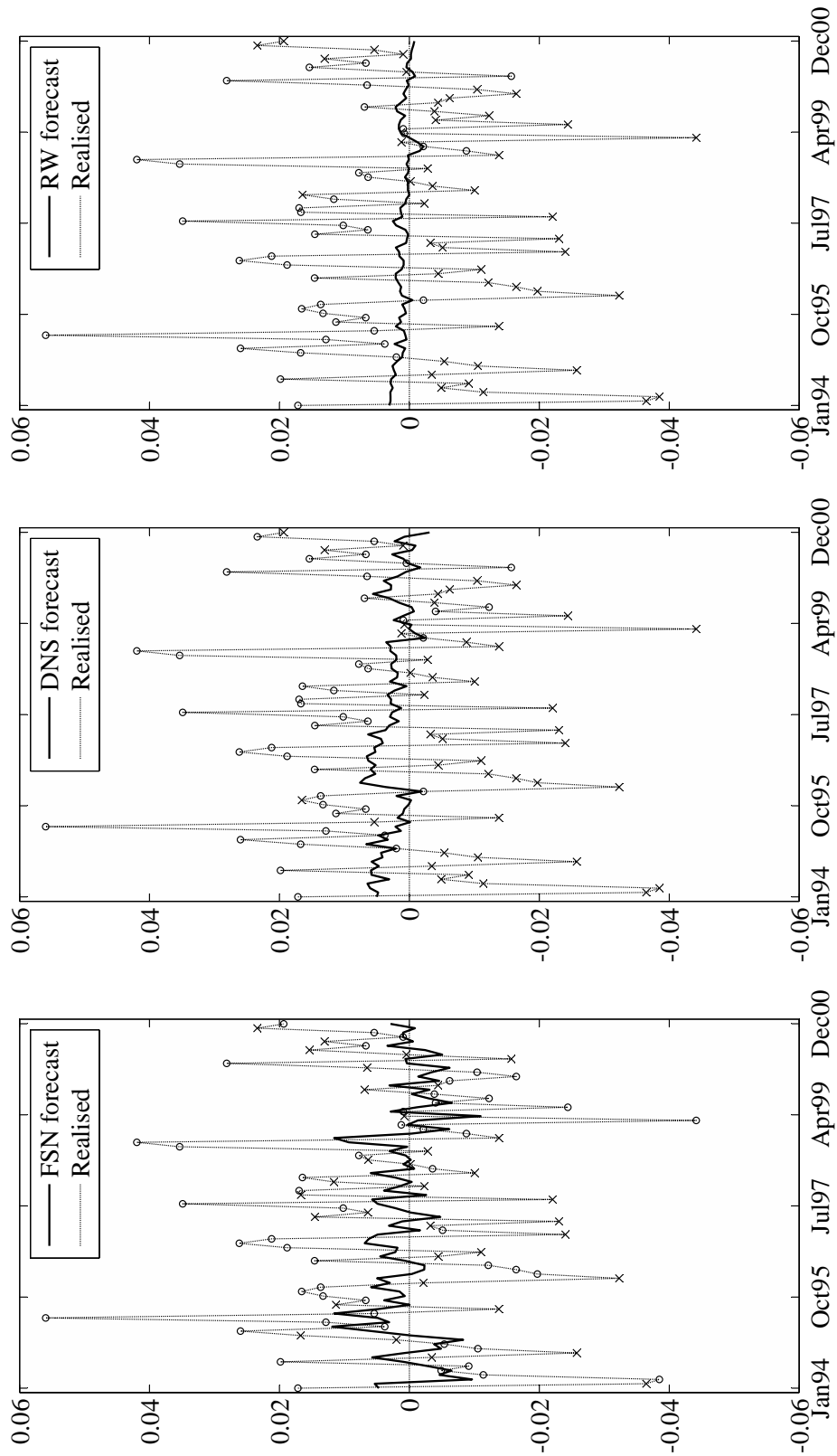
$$\sum_t \Pi_{t+1} = 1M \sum_t -|R_{t+1}(\tau_2) - R_{t+1}(\tau_1)| \cdot l(d_t, R_{t+1}(\tau_2), R_{t+1}(\tau_1)), \quad (14)$$

where the *loss* function  $l_{t+1} := l(d_t, R_{t+1}(\tau_2), R_{t+1}(\tau_1)) = -\{(d_t/1M) \cdot \text{sgn}[R_{t+1}(\tau_2) - R_{t+1}(\tau_1)]\}$  equals  $-1$  when  $\tilde{d}_t$  correctly forecasts the sign of the excess return and equals  $+1$  otherwise. Thus profit performance depends on the ‘directional accuracy’ of the excess return forecasts but with (in)correctly forecast signs rewarded (punished) proportionally to the absolute magnitude of the realised return. Focusing on Algorithm 1, Figure 7 plots for each of the three models the realised excess return  $[R_{t+1}(82) - R_{t+1}(3)]$  and its associated forecast  $[\hat{r}_{t+1|t}(\tau_2) - \hat{r}_{t+1|t}(\tau_1)]$  over the out-of-sample evaluation period. Excess returns whose signs are correctly forecast by  $\tilde{d}_t = \text{sgn}[\hat{r}_{t+1|t}(\tau_2) - \hat{r}_{t+1|t}(\tau_1)]$  are indicated by open circles and by crosses otherwise, corresponding to  $l_{t+1} = -1$  and  $l_{t+1} = +1$  respectively. Both DNS and  $RW^{YC}$  models produce a rather smooth time series of predominantly positive excess return predictions, whereas the FSN-ECM model tracks the pattern of excess returns more closely. This results in many more correctly forecast signs in the case of negative excess returns for the FSN-ECM model compared to both DNS and  $RW^{YC}$  – 53% compared to 13% and 10% (see Table 4) – which is the dominant factor in explaining its superior profitability. Whilst the proportion of correctly forecast signs in the case of positive excess returns is somewhat lower for FSN-ECM – 66% compared to 84% and 82% resp. – large, positive returns still usually have  $l_{t+1} = -1$  (open circles). Overall the comparative gain in profits from correctly forecasting the signs of negative returns outweighs the reduction attaching to positive returns.

Section 3.4 concluded that the T-FSN(6)-ECM(2) model outperformed the DNS and other competitor models in terms of out-of-sample, MSFE-based forecast evaluation measures. The above analysis finds the same conclusion for realised trading profits from trading algorithms that construct arbitrage portfolios of discount bonds using the model forecasts. The dominance of the T-FSN(6)-ECM(2) model is thus found to be robust to the use of very different loss functions. Whereas the largest reductions in MSFE relative to the random walk were obtained for the short end of the yield curve (see Figure 3), the ability of the FSN-ECM model to forecast the signs of the excess returns of the longest bonds over the three-month return underlies its superior trading profitability. The analysis highlights in an applied setting the dangers of attempts to infer the relative economic value of model forecasts on the basis of their associated MSFEs.



Figure 7: Forecast and realised excess returns for Algorithm 1, January 1994-December 2000



Note: Shows plots of the forecast excess return  $\hat{r}_{t+1|\mu}(82) - \hat{r}_{t+1|\mu}(3)$  against the realised excess return  $R_{t+1}(82) - R_{t+1}(3)$ . Realised excess returns whose signs are correctly forecast by  $\hat{d}_t = \text{sgn}[\hat{r}_{t+1|\mu}(\tau_2) - \hat{r}_{t+1|\mu}(\tau_1)]$  are indicated by open circles and by crosses otherwise. FSN denotes the Triangular FSN(6)-ECM(2) model with  $\mathbf{k} = (1, 2, 4, 18, 24, 84)$ , DNS the Dynamic Nelson-Siegel model and RW the random walk for yields. All forecasts are made recursively out-of-sample using the yield curve forecasts of Section 3.4.

## 5 CONCLUSION

Functional Signal plus Noise (FSN) models are proposed that provide a new, general method for modelling and forecasting time series of economic functions. In financial economics, market prices at a given point in time are a function of the characteristics of the asset traded and it is often appropriate to consider the underlying price as a continuous function of these characteristics. Despite their importance, general methods for the study of the dynamics of economic functions that are applicable to observation vectors of moderate or large cross-sectional dimension have received little attention in the literature. The FSN-ECM models specify the evolution over time of stochastic functions and combine the virtues of parsimony and parametric interpretability. A natural cubic spline is used to model the underlying, smooth economic function (or in this context, term structure), the dynamic evolution of which is driven by a cointegrated Vector Autoregression for the ordinates (yields) at the knots of the spline. The natural cubic spline provides flexible cross-sectional fit and results in a linear state space model, thus enabling use of the Kalman filter. This FSN-ECM model achieves dimension reduction, provides a coherent description of the observed yield curve and its dynamics as the cross-sectional dimension  $N$  becomes large, and can feasibly be estimated and used for forecasting when  $N$  is large. Under the assumption that the  $m$  knot-yields follow a cointegrated I(1) process with cointegrating rank  $r$ , a theorem is derived showing that the observed and latent yield curves of the FSN-ECM process with dimension  $N$  are I(1) processes with cointegrating rank  $[N - (m - r)]$ , and giving an expression for the associated matrix of cointegrating vectors. There are  $(m - r)$  identical common trends that drive both knot-yields and yield curves.

The FSN-ECM models are used to forecast 36-dimensional zero-coupon yield curves for US Treasury bonds at the one month ahead horizon. Their out-of-sample performance is compared to important rival models using both MSFE-based criteria and economically relevant loss functions derived from the realised profits earned by implementing pairs trading algorithms. The algorithms construct an arbitrage portfolio of discount bonds each period on the basis of the different model forecasts. The analysis highlights in a concrete, applied setting the dangers of attempts to infer the relative economic value of model forecasts on the basis of their associated MSFEs.

The triangular (T-) 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) Dynamic Nelson-Siegel, Expectations Theory and random walk forecasts across all maturities. A decomposition of MSFE by maturity reveals that both superior cross-sectional fit of the natural cubic spline and better forecasting of the model-specific, latent factors contribute to the T-FSN(6)-

ECM(2) model having smaller MSFEs than the Dynamic Nelson-Siegel model, but that better forecasting of the factors is more important in this regard. It turns out that yield spreads provide important information for forecasting the yield curve, 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.

In all three trading algorithms investigated, the T-FSN(6)-ECM(2) forecasts result in cumulative, realised trading profits that are an order of magnitude larger than those obtained using both the random walk for yields and Dynamic Nelson-Siegel forecasts. The dominance of the T-FSN(6)-ECM(2) model is thus found to be robust to the use of very different loss functions. For the simplest and most profitable algorithm, the empirical distribution of monthly profits for the T-FSN(6)-ECM(2) forecasts first order stochastically dominates both alternatives. Whereas the largest reductions in MSFE relative to the random walk were obtained for the short end of the yield curve, the ability of the T-FSN(6)-ECM(2) model to forecast the sign of excess returns of the longest bonds over the 3-month return underlies its superior trading profitability.

We thus find that the proposed FSN-ECM models consistently outperform both the prominent Diebold and Li (2006) Dynamic Nelson-Siegel model and the random walk at the one month ahead horizon using a broad range of forecast evaluation criteria. Furthermore, the FSN-ECM approach is scalable in the cross-sectional dimension of the yield curve and is based on a continuous, smooth underlying term structure. 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 FSN-ECM models with a latent natural cubic spline yield curve but a non-linear observation equation in order to model and forecast coupon bond prices 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, Andrew Harvey, David Hendry, Siem Jan Koopman, Anthony Murphy, Bent Nielsen, Mark Salmon, Neil Shephard, Stefan de Wachter, participants at the EC<sup>2</sup> conference 2006, the Warwick Frontiers in Finance conference 2006, the Federal Reserve Bank of Dallas, the Federal Reserve Bank of San Francisco, the Bank of England, Southampton University, the Statistical Laboratory University of Cambridge and the anonymous referees for their comments. 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 FORECASTING UNDER THE EXPECTATIONS THEORY

This section considers what the Expectations Theory (ET) 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  $\mathbf{y}_t(\boldsymbol{\tau})$  is said to be *complete* here when  $\boldsymbol{\tau} = (1, 2, \dots, n)$ , and is denoted by  $\mathbf{y}_t(\boldsymbol{\tau}_1^n)$ . The associated vector of spreads between the yields and the short rate is written as  $\mathbf{s}_{nt} := (s_t(2, 1), \dots, s_t(n, 1))'$ , where  $s_t(\tau_j, \tau_i) := y_t(\tau_j) - y_t(\tau_i)$ . Lemma 4 below shows that the ET fully determines the conditional expectation of the  $(n - 1)$ -dimensional yield curve  $\mathbf{y}_{t+1}(\boldsymbol{\tau}_1^{n-1})$  given the public information set  $\mathcal{F}_t$  and that this is an affine function of the current spread vector  $\mathbf{s}_{nt}$ . Furthermore, the ET holds *if and only if* the conditional expectation is given by this affine function for all  $n \geq 2$  (Eq. 16).

**Lemma 4 (Conditional Expectation of Yield Curve under the ET)** *Let  $n \in \{2, 3, \dots\}$  and suppose that*

$$y_t(\tau) = \left\{ \tau^{-1} \sum_{i=0}^{\tau-1} \mathbb{E}[y_{t+i}(1)|\mathcal{F}_t] \right\} + \rho(\tau), \quad \forall t, \forall \tau, \quad (15)$$

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. When Eq. (15) is satisfied we say that the Expectations Theory (ET) holds. Then it is possible to show that the ET holds *if and only if*

$$\mathbb{E}[\Delta \mathbf{y}_{t+1}(\boldsymbol{\tau}_1^{n-1})|\mathcal{F}_t] = \boldsymbol{\alpha}_{n-1}^{ET} (\mathbf{s}_{nt} - \boldsymbol{\rho}_n) \quad \forall t, \forall n \in \{2, 3, \dots\}, \quad (16)$$

where  $\boldsymbol{\rho}_n = (\rho(2), \dots, \rho(n))'$ . The  $(n - 1) \times (n - 1)$  matrix  $\boldsymbol{\alpha}_{n-1}^{ET}$  is given, for  $n > 2$ , by

$$\boldsymbol{\alpha}_{n-1}^{ET} = \begin{pmatrix} 2 & 0 & 0 & 0 & \dots & 0 & 0 \\ -1 & 3/2 & 0 & 0 & \dots & 0 & 0 \\ 0 & -1 & 4/3 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & -1 & \frac{n}{n-1} \end{pmatrix}, \quad (17)$$

and  $\alpha_1^{ET} = 2$ .

**Proof.** See the proof of Theorem 1 of Bowsher and Meeks (2006). ■

The following example establishes that setting  $\boldsymbol{\beta} = \boldsymbol{\beta}_s$  in the FSN-ECM model does not imply that the ET holds.

**Example 5 (FSN-ECM Model with Stationary Yield Spreads)** *Consider the case of the FSN( $m$ )-ECM(2) process in Theorem 2 with  $\boldsymbol{\beta} = \boldsymbol{\beta}_s$  where  $\boldsymbol{\beta}'_s \boldsymbol{\gamma}_t = (\gamma_{j+1,t} - \gamma_{jt})_{j=1}^{m-1}$ ,  $\boldsymbol{\Omega}_\varepsilon = \mathbf{0}$  and  $\boldsymbol{\tau} = (1, 2, \dots, n)$  with  $n \gg m$  and  $n$  fixed. Then by Theorem 2 the vector of yield spreads  $\boldsymbol{\phi}' \mathbf{y}_t(\boldsymbol{\tau}) = [y_t(i) - y_t(1)]_{i=2}^n$  is a vector of stationary cointegrating relations. However, it is straightforward to show that  $\mathbb{E}[\Delta \mathbf{y}_{t+1}(\boldsymbol{\tau}_1^{n-1})|\mathcal{F}_t]$  does not satisfy Eq. (16) and hence that the FSN( $m$ )-ECM(2) process does not satisfy the ET.*

## B BACKGROUND 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 6** *Natural Cubic spline (NCS) on  $(\mathbf{k}; \boldsymbol{\gamma})$ .* Consider an interval of the real line  $[a, b]$ , subdivided by a vector,  $\mathbf{k}$ , of points

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

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  $(k_1, k_m)$  are called the end knots. Denote by  $\boldsymbol{\gamma}$  a vector of real-valued ordinates,  $(\gamma_j)_{j=1}^m$ . A function  $\mathcal{S}(x)$  with domain  $[a, b]$  is a **cubic spline interpolating to  $\boldsymbol{\gamma}$  with knots  $\mathbf{k}$** , or more concisely a **cubic spline on  $(\mathbf{k}; \boldsymbol{\gamma})$** , if and only if:

(i)  $\mathcal{S}(k_j) = \gamma_j$  ( $j = 1, \dots, m$ );

(ii)  $\mathcal{S}(x)$  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)  $\mathcal{S}(x)$  is twice continuously differentiable on  $[a, b]$ .

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

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

A well known result that arises by combining conditions on  $\mathcal{S}''(k_1)$  and  $\mathcal{S}''(k_m)$  with the conditions (i), (ii) and (iii) of Definition 6 is that  $\mathcal{S}_\boldsymbol{\gamma}(\boldsymbol{\tau})$  is a linear function of the ordinate vector  $\boldsymbol{\gamma}$ . This result, stated for the case of a NCS in the theorem below, allows the FSN-ECM models to be written in linear state space form.

**Lemma 7** *Let  $\mathcal{S}_\boldsymbol{\gamma}(x)$  be a NCS on  $(\mathbf{k}; \boldsymbol{\gamma})$ , with  $\mathbf{k}$  and  $\boldsymbol{\gamma}$  vectors of dimension  $m$ . Also let  $\boldsymbol{\tau} = (\tau_1, \dots, \tau_N)$  be a finite vector of points in  $[a, b]$ , and  $\mathcal{S}_\boldsymbol{\gamma}(\boldsymbol{\tau}) := (\mathcal{S}_\boldsymbol{\gamma}(\tau_1), \dots, \mathcal{S}_\boldsymbol{\gamma}(\tau_N))'$ . Then*

$$\mathcal{S}_\boldsymbol{\gamma}(\boldsymbol{\tau}) = \mathbf{W}(\mathbf{k}, \boldsymbol{\tau}) \boldsymbol{\gamma},$$

where the  $N \times m$  interpolation matrix  $\mathbf{W}(\mathbf{k}, \boldsymbol{\tau})$  depends only on  $\boldsymbol{\tau}$  and the knot positions  $\mathbf{k}$ . Details of how to compute  $\mathbf{W}(\mathbf{k}, \boldsymbol{\tau})$  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 NCS.

**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.
- 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.
- Boivin, J. and S. Ng (2005). Understanding and comparing factor-based forecasts. *International Journal of Central Banking* 1, 117–151.
- 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.
- Clements, M. P. and D. F. Hendry (1993). On the limitations of comparing mean square forecast errors. *Journal of Forecasting* 12, 617–637.
- De Pooter, M., F. Ravazzolo, and D. van Dijk (2007). Predicting the term structure of interest rates: Incorporating parameter uncertainty, model uncertainty and macroeconomic information. Tinbergen Institute Discussion Paper 2007-028/4.
- 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.
- Fisher, M., D. Nychka, and D. Zervos (1995). Fitting the term structure of interest rates with smoothing splines. Working Paper 95-1, Finance and Economics Discussion Series, Federal Reserve Board.
- Granger, C. and H. Pesaran (2000). Economic and statistical measures of forecast accuracy. *Journal of Forecasting* 19, 537–560.
- 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.
- Jeffrey, A., O. Linton, and T. Nguyen (2006). Flexible term structure estimation: which method is preferred? *Metrika* 63, 99–122.
- Johansen, S. (1996). *Likelihood-based inference in cointegrated vector autoregressive models*. Oxford: Oxford University Press.
- Jungbacker, B. and S. J. Koopman (2007). Likelihood-based analysis for dynamic factor models. Mimeo, VU University Amsterdam.
- Kargin, V. and A. Onatski (2007). Curve forecasting by functional autoregression. Mimeo, Economics Department, Columbia University.
- Koopman, S. J. and M. Ooms (2003). Time series modelling of daily tax revenues. *Statistica Neerlandica* 57, 439–469.
- 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.
- Leitch, G. and J. E. Tanner (1993). Economic forecast evaluation: profits versus the conventional error measures. *American Economic Review* 81(3), 580–590.
- McCulloch, J. H. (1975). The tax adjusted yield curve. *Journal of Finance* 30, 811–829.
- 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.
- Powell, M. J. D. (1996). *Approximation Theory and Methods*. Cambridge: Cambridge University Press.
- Sack, B. (2000). Using Treasury STRIPS to measure the yield curve. Federal Reserve Board FEDS 2000-42.
- 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.
- Waggoner, D. (1997). Spline methods for extracting interest rate curves from coupon bond prices. Working Paper 97-10, Federal Reserve Bank of Atlanta.
- Zellner, A. (1992). Statistics, science and public policy. *Journal of the American Statistical Association* 87, 1–6.