Federal Funds Rate Prediction

Lucio Sarno[a,b,c], Daniel L. Thornton[c] and Giorgio Valente[a]

a: University of Warwick

b: Centre for Economic Policy Research

c: Federal Reserve Bank of St. Louis

January 2003

Abstract

Recent research has reported that both the federal funds rate futures market and the federal funds target contain valuable information for explaining the behavior of the US effective federal funds rate. A parallel literature on interest rate modelling has recorded evidence that the dynamics of interest rates displays significant regime-switching behavior. In this paper we produce out of sample forecasts of the federal funds rate at horizons up to 8 weeks ahead using linear and nonlinear, regime-switching equilibrium correction models of the funds rate and employing both point and density measures of forecast accuracy. We cannot discriminate among the models considered in terms of point forecast accuracy. However, in terms of density forecast accuracy, we find that the term structure model of the federal funds futures rate is significantly better than the other models considered, and that regime-switching models provide a substantial forecasting improvement relative to their linear counterparts and relative to individual series of the futures rate.

JEL classification: E43; E47.

Keywords: federal funds rate; term structure of interest rates; forecasting; nonlinearity.

1 Introduction

The importance of the effective federal funds rate in US financial markets is unquestionable. The Federal Reserve (Fed) implements monetary policy by targeting the effective federal funds rate. The ability of market participants to predict the federal funds rate is important to modern analyses of monetary policy in that other interest rates are believed to be linked to the federal funds rate by the market expectation of monetary policy actions that directly affect the funds rate. This paper investigates the ability of three models to generate out-of-sample forecasts of the daily federal funds rate over forecast horizons ranging from one to eight weeks.

We consider three alternative models of the federal funds rate. The first model is inspired by the analysis of Taylor (2001), who demonstrates that the federal funds rate target contains valuable information for explaining the behavior of the federal funds rate. Specifically, Taylor shows that the funds rate responds to deviations of it from the Fed's target. If the funds rate responds relatively quickly to deviations from the funds rate target, this model might be expected to forecast well at short horizons but less well at longer horizons.

Our second model incorporates information about the future federal funds rate that is reflected in the term structure of the federal funds futures rate. There is a growing literature that suggests that federal funds futures rates contain substantial information about future monetary policy actions and therefore the future federal funds rate (e.g., Krueger and Kuttner, 1996; Kuttner, 2001; Poole, Rasche and Thornton, 2002).

Our intuition is that the first model may perform relatively well at short horizons given that deviations of the funds rate from its target are relatively short lived, while the model based on the term structure of the futures rates might perform better at longer horizons–up to the maturity of the longer futures contract considered in the model. Hence, we consider a third model that incorporates the information in both the federal funds rate target and the term structure of the federal funds futures rates.

For all three models we motivate and estimate a vector equilibrium correction (VECM) to capture the equilibrium and dynamics of the relationship between the federal funds rate and the predictor variables. Our approach is distinctive in that we consider both linear and nonlinear, regime-switching VECMs based on the three specifications described above in order to generate out-of-sample forecasts of the federal funds rate. Our consideration of regime-switching models is motivated by a recent strand of the empirical literature (e.g., Garcia and Perron, 1996; Gray, 1996; Ang and Bekaert, 2002) that provides convincing evidence that explicit modelling of regime switches in interest rates may be key to produce satisfactory statistical fit of interest rates data. Whether allowing for nonlinearities in the underlying data-generating process for the federal funds rate yields superior federal funds rate forecasts is investigated using a fairly general three-regime Markov-switching vector equilibrium correction model (MS-VECM). To the best of our knowledge, this paper represents the first application of Markov-switching in a multivariate cointegrated framework to interest rate modelling and forecasting.

Using daily data since 1990, we first provide confirmatory evidence that each of the federal funds rate target and the term structure of the federal funds futures rates contain valuable information to explain a substantial fraction of movements in the federal funds rate in a linear VECM framework. However, we show that a conventional linear VECM displays significant residual nonlinearity and is easily rejected when tested against the alternative of an MS-VECM.

We then compare the performance of the models considered in an out-of-sample forecasting exercise, where each linear and nonlinear model is tested against each other as well as against the simple futures rate time series, for forecast horizons up to 8 weeks.[1] The evaluation of the relative performance of the competing models is based on conventional statistical criteria for point forecasting performance as well as on the ability of the models to forecast the true predictive density of federal funds rate out of sample.[2] We argue that density forecasting accuracy is more appropriate for evaluating the competing models because the federal funds rate is not normally distributed and because we are considering nonlinear models consistent with non-normal densities (see, *inter alia*, Diebold, Gunther and Tay, 1998; Granger and Pesaran, 1999; Timmerman, 2000).

To anticipate our forecasting results, we find that all models appear to produce equally good point forecasts of the federal funds rate in that none of the models can be rejected against any of the others and against the simple time series of the futures rate. Nevertheless, our density forecasting

[1] Under the market efficiency hypothesis, the federal funds futures rate is the optimal predictor of the future federal funds rate.

[2] By true predictive density of the data we mean the density of the data estimated over the chosen forecast period. Therefore, no forecast is in fact carried out in this case, and the term 'predictive' simply refers to the fact that the density in question is not estimated over the full sample but only over the forecast period.

results suggest that models based on the term structure of the federal funds futures rate are the best forecasting models of the funds rate and that nonlinear VECMs provide sizable forecasting improvements relative to their linear counterparts. Furthermore, both linear and nonlinear term structure models outperform the simple futures rate time series in terms of density forecasting performance.

The remainder of the paper is set out as follows. In Section 2 we motivate our empirical framework for modelling and forecasting the federal funds rate using the information contained in the federal funds rate target and in the term structure of the federal funds futures rates. In Section 3 we briefly set out the econometrics of Markov-switching multivariate models as applied to nonstationary processes and cointegrated systems. In the following section we describe the data and discuss some of the key features of the federal funds rate and federal funds futures markets. In Section 5, we report our modelling and testing results, while in the following section we report and discuss our forecasting results. A final section concludes.

2 The information in the federal funds rate target and in the term structure of the federal funds futures rate

In this section we consider alternative ways of modelling the US effective federal funds (FF) rate. The first specification is based on the relationship between the FF rate and the official FF rate target. Let s_t and s_t^T be, respectively, the FF rate and the FF rate target on date t. If the Fed uses open market operations to keep the FF rate close to the FF rate target, a logical way to model daily movements in the FF rate is to assume that

$$\Delta s_t = constant + \theta d_{t-1} + error\ term, \tag{1}$$

where Δ is the first-difference operator; $d_t = (s_t - s_t^T)$; and $\theta < 0$ governs the speed at which the FF rate responds to deviations from the FF rate target. Assuming that both s_t and s_t^T are better characterized as unit root or I(1) processes, this then implies that s_t and s_t^T cointegrate with a cointegrating vector $[1, -1]$. In turn, by the Granger Representation Theorem (Engle and Granger, 1987), the dynamic relationship between the FF rate and the FF rate target must be characterized by an equilibrium correction model where the FF rate responds to lagged deviations of the FF rate from the target, which plays the role of the equilibrium correction term. This equilibrium correction model may therefore take the form of equation (1) or a more general variant of it which also includes lags of the change in the FF rate and the FF rate target as right-hand-side variables.[3]

Taylor (2001) suggests that the Trading Desk of the Federal Reserve Bank of New York (hereafter Desk) strives to keep the FF rate close to its target level. Hence, equation (1) can be thought of as the Desk's reaction function. Taylor argues that the adjustment to departures of the federal funds rate from its target is not full at daily frequency, which implies $-1 < \theta \leq 0$.[4] However, the usefulness of the funds rate target for forecasting the funds rate in a VECM depends not only on the equilibrium correction term but also on the dynamic relationship between the funds rate and the funds rate target.

The second model we consider relies on information contained in federal funds futures rates. Define f_t^h the FF futures rate at time t for a contract maturing in month h. The pre-tax profits for an investor long in federal funds contracts purchased on date t for delivery in month h is given by:

$$\pi_t^h = f_t^h - \frac{1}{M} \sum_{i=1}^{M} s_{t+Mh+(M-t)-i+1}, \qquad (2)$$

where π_t^h denotes the pre-tax profits at the maturity of the futures contracts; M is the number of calendar days in month h; and $\frac{1}{M} \sum_{i=1}^{M} s_{t+Mh+(M-t)-i+1}$ is the settlement price based on month h's average effective FF rate.[5]

[3] As discussed in our empirical analysis below, using standard unit root test statistics, we found clear evidence that both the FF rate and the FF rate target are first-difference stationary, supporting the stylized fact that interest rates are $I(1)$ processes (e.g., Stock and Watson, 1988, 1999). There is, however, an apparent conflict between conventional economic and finance theory, which often assumes that interest rates are stationary processes (e.g., see the vast finance literature assuming a Vasicek (1977) model of interest rates, which is simply a mean-reverting process representable as an Ornstein-Uhlenbeck equation) and the mainstream empirical literature on interest rates, which (at least since Engle and Granger, 1987) either assumes or finds that interest rates are nonstationary processes. In our discussion in this section and in our empirical work below, we follow the empirical practice because very persistent series with a root very close to unity are better approximated by $I(1)$ processes than by stationary ones (see, for example, Stock, 1997, and the references therein).

[4] Thornton (2001) offers a different motivation for the Desk's behavior. Specifically, he suggests that the Desk frequently uses the funds rate as an indicator of reserve demand, which is difficult to estimate. While Thornton's motivation for the Desk's behavior differs from Taylor's, the implication is the same: deviations of the funds rate from the funds rate target induce changes in reserve supply that tend to push the funds rate towards the funds rate target. Thornton provides evidence that is consistent with either interpretation.

[5] On the microstructure of the FF futures market, see, *inter alia*, Carlson, McIntire and Thomson (1995), Krueger

The rational expectations efficient market hypothesis implies the absence of arbitrage opportunities, so that $E_t \pi_t^h = 0$. Imposing this no-arbitrage condition yields:[6]

$$
\begin{aligned}
f_t^h &= E_t \left[\frac{1}{M} \sum_{i=1}^{M} s_{t+Mh+(M-t)-i+1} \right] \\
&= \frac{1}{M} \sum_{i=1}^{M} E_t s_{t+Mh+(M-t)-i+1}. \qquad (3)
\end{aligned}
$$

Assuming that each of the effective FF rate and FF futures rate is $I(1)$[7], the possibility that these two rates cointegrate is suggested by rearranging equation (2) as follows:

$$
\begin{aligned}
f_t^h - s_t &= \frac{1}{M} \left[\sum_{i=1}^{M} E_t s_{t+Mh+(M-t)-i+1} - M s_t \right] \\
&= \frac{1}{M} \sum_{i=1}^{M} E_t \left(s_{t+Mh+(M-t)-i+1} - s_t \right). \qquad (4)
\end{aligned}
$$

Since $\left(s_{t+Mh+(M-t)-i+1} - s_t \right)$ is stationary for all $i = 1, \ldots, M$ if s_t is $I(1)$, the right-hand-side of equation (4) is stationary. Thus, it follows that the left-hand-side of equation (4) must also be stationary, which implies that f_t^h and s_t cointegrate with a cointegrating vector $\begin{bmatrix} 1 & -1 \end{bmatrix}'$. Because this is true for any h, we consider the vector $[s_t, f_t^1, f_t^2, f_t^3, ...f_t^m]'$. Hence, there must be m unique cointegrating vectors, each given by a row of the matrix $[-\iota, I_m]$, where I_m is an m-dimensional identity matrix and ι is an m-dimensional column vector of ones. Further, by the Granger Representation Theorem (Engle and Granger, 1987) the same set of FF and FF futures rates must possess a vector equilibrium correction representation in which the term structure of the deviations of the futures FF rates from the FF rate (say the futures premia) play the part of the equilibrium errors. We exploit this framework by estimating a VECM to extract the information in the term structure of FF futures rates for the purpose of forecasting the FF rate.

Note that, under the market efficiency hypothesis, the FF futures rate should predict optimally the FF rate (see Kuttner, 1996, 2001). In some sense, therefore, the existence of a VECM for the FF rate and the FF futures rate implies that there is a wedge between the FF futures rate

and Kuttner (1996), Furfine (1999) and Kuttner (2001); see also Stigum (1990).

[6] Differently from Kuttner (2001), for clarity of exposition we do not explicitly consider the premium accruing to investors long in the spot-month futures contract in this section of the paper.

[7] See footnote 3.

and the expected FF rate that may be due, for example, to constant or time-varying risk premia or to departures from the rational expectations hypothesis that underlies the market efficiency hypothesis.

A third model (Model III) we consider incorporates *both* the information in the FF rate target and the information in the term structure of FF futures rates. If the FF rate target and the term structure of futures rates each contain independent information valuable for forecasting the FF rate, the third model should produce better modelling and forecasting results relative to either Model I or Model II. Finally, we consider both linear and nonlinear (regime-switching) variants of Models I to III.

3 Markov-switching equilibrium correction

In this section, the econometric procedure employed in order to model regime shifts in the dynamic relationships represented in all three models is outlined. The procedure essentially extends Hamilton's (1988, 1989) Markov-switching framework to nonstationary systems, allowing us to apply it to cointegrated vector autoregressive (VAR) and VECM systems (see Krolzig, 1997, 1999).

Consider the following Q-regime p-th order Markov-switching vector autoregression (MS(Q)-VAR(p)) which allows for regime shifts in the intercept term:

$$y_t = \nu(z_t) + \sum_{i=1}^{p} \Pi_i y_{t-i} + \varepsilon_t, \tag{5}$$

where y_t is a K-dimensional observed time series vector, $y_t = [y_{1t}, y_{2t}, \ldots, y_{Kt}]'$; $\nu(z_t)$ is a K-dimensional column vector of regime-dependent intercept terms, $\nu(z_t) = [\nu_1(z_t), \nu_2(z_t), \ldots, \nu_K(z_t)]'$; the Π_i's are $K \times K$ matrices of parameters; $\varepsilon_t = [\varepsilon_{1t}, \varepsilon_{2t}, \ldots, \varepsilon_{Kt}]'$ is a K-dimensional vector of Gaussian white noise processes with covariance matrix Σ, $\varepsilon_t \sim NID(\mathbf{0}, \Sigma)$. The regime-generating process is assumed to be an ergodic Markov chain with a finite number of states $z_t \in \{1, \ldots, Q\}$ governed by the transition probabilities $p_{ij} = \Pr(z_{t+1} = j \mid z_t = i)$, and

$\sum_{j=1}^{Q} p_{ij} = 1 \; \forall i, j \in \{1, \ldots, Q\}$.[8]

A standard case in economics and finance is that y_t is nonstationary but first-difference stationary, i.e. $y_t \sim I(1)$. Then, given $y_t \sim I(1)$, there may be up to $K-1$ linearly independent cointegrating relationships, which represent the long-run equilibrium of the system, and the equilibrium error (the deviation from the long-run equilibrium) is measured by the stationary stochastic process $u_t = \alpha' y_t - \beta$ (Granger, 1986; Engle and Granger, 1987). If there is cointegration, the cointegrated MS-VAR (5) implies a Markov-switching vector equilibrium correction model (MS-VECM) of the form:

$$\Delta y_t = \nu(z_t) + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + \Pi y_{t-1} + \varepsilon_t, \tag{6}$$

where $\Gamma_i = -\sum_{j=i+1}^{p} \Pi_j$ are matrices of parameters, and $\Pi = \sum_{i=1}^{p} \Pi_i - \mathbf{I}$ is the long-run impact matrix whose rank r determines the number of cointegrating vectors (e.g. Johansen, 1995).[9] For expositional purposes we have only presented the MS-VECM framework for regime shifts in the intercept alone; however, the procedure can easily be extended to regime shifts elsewhere.

For Model I $y_t = [s_t, s_t^T]'$, one unique cointegrating relationship should exist. Federal funds futures contracts from 1 through 5 months were initially considered for Model II. However, FF futures rates for contracts longer than 2 months did not appear to have any incremental information over the information contained in 1- and 2-month contracts. Hence, we decided in favor of a parsimonious model where $y_t = [s_t, f_t^1, f_t^2]'$. Model III combines Model I and Model II, so that $y_t = [s_t, s_t^T, f_t^1, f_t^2]'$.

We consider both linear and regime-switching versions of these VECMs for each model. As discussed in Section 5.3 below, after considerable experimentation, we selected a specification of the MS-VECM that allows for regime shifts in the intercept as well as in the variance-covariance matrix. This model, the Markov-Switching-Intercept-Heteroskedastic-VECM or MSIH-VECM,

[8] To be precise, z_t is assumed to follow an ergodic Q-state Markov process with transition matrix

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1Q} \\ p_{21} & p_{22} & \cdots & p_{2Q} \\ \vdots & \vdots & \ddots & \vdots \\ p_{Q1} & p_{Q2} & \cdots & p_{QQ} \end{bmatrix},$$

where $p_{iQ} = 1 - p_{i1} - \ldots - p_{iQ-1}$ for $i \in \{1, \ldots, Q\}$.

[9] In this section it is assumed that $0 < r < K$, implying that y_t is neither purely difference-stationary and non-cointegrated (i.e. $r = 0$) nor is a stationary vector (i.e. $r = K$).

may be written as follows:

$$\Delta y_t = v(z_t) + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + \Pi y_{t-1} + u_t, \tag{7}$$

where $\Pi = \alpha\beta'$, $u_t \sim NIID(\mathbf{0}, \Sigma(z_t))$ and $z_t \in \{1, \ldots, Q\}$.

An MS-VECM can be estimated using a two-stage maximum likelihood procedure. In the first stage, Johansen's (1988, 1991) maximum likelihood cointegration procedure is employed in order to determine the number of cointegrating relationships in the system and to estimate the cointegration matrix. Use of the conventional Johansen procedure is legitimate without modelling the Markovian regime shifts explicitly (see Saikkonen, 1992; Saikkonen and Luukkonen, 1997). In the second stage, an expectation-maximization (EM) algorithm for maximum likelihood estimation is implemented to obtain estimates of the remaining parameters of the model (Dempster, Laird and Rubin, 1977; Hamilton, 1990; Krolzig, 1999).

4 The federal funds rate, the target, and the futures market: discussion and data issues

The data set consists of daily observations on s_t, s_t^T, and f_t^j for $j = 1, 2$. The FF rate, s_t is a weighted average of the rates on federal funds transactions of a group of federal funds brokers who report their transactions daily to the Federal Reserve Bank of New York. Federal funds are deposit balances at Federal Reserve banks that institutions (primarily depositories, e.g. banks and thrifts) lend overnight to each other. These deposit balances are used to satisfy reserve requirements of the Federal Reserve System.[10] f_t^j is the rate on a federal funds futures contract with maturity j, traded on the Chicago Board of Trade (CBT). Futures contracts are designed to hedge against or speculate on the effective FF rate. The CBT offers contracts on the FF rate at a variety of maturities; however, the most active contracts are current month and a few months into the future. The contracts are marked to market on each trading day, and final cash settlement occurs on the first business day following the last day of the contract month. The FF rate was obtained from the Federal Reserve Bank of St. Louis database, *Federal Reserve Economic Data* (FRED). The FF rate target was taken from Thornton and Wheelock (2000, Table B1). The FF futures rates

[10] For a discussion of the Federal Reserve's reserve requirements, see, for example, Taylor (2001).

were obtained from the CBT.

The sample period spans from January 2 1990 through December 29 2000, a total of $2,869$ observations. This period was chosen mainly for two reasons. First, while the Fed has never explicitly stated when it began targeting the federal funds rate in implementing monetary policy, an emerging consensus view is that the Fed has been explicitly targeting the funds rate since at least the late 1980s.[11] [12] Second, while the FF futures market has existed since October 1988, trading activity in this market was initially small. To insure against the possibility that the empirical analysis would be affected by the thinness of the FF futures market during the early years of its operation, we decided to begin the sample in January 1990.

It is important to note that the Fed made two procedural changes in 1994 that may have affected the market's ability to predict the FF rate target. First, in February 1994 the Fed began the practice of announcing FF rate target changes immediately upon making them. Prior to that, target changes were not announced. Consequently, the market had to infer the Fed's actions by observing open market operations and the FF rate (e.g., Cook and Hahn, 1989; Rudebusch, 1995; Taylor, 2001; Poole, Rasche, and Thornton, 2002). Second, in 1994 the FOMC began the practice of changing the funds rate target primarily at regularly scheduled FOMC meetings.[13] Prior to that, most target changes were made during the inter-meeting period and at the discretion of the Chairman. We allow for these procedural changes by including a dummy variable for the 1994 procedural break.

Since in this paper we are mainly interested in the predictive power of alternative time series models, we estimate each model considered over the period January 2 1990 through December 29 1995 and generate forecasts over the remaining five years of data.

[11] See, for example, Meulendyke (1998), Hamilton and Jordá (2001), Poole, Rasche and Thornton (2002).

[12] Also, since October 1989, the Fed has followed the practice of changing the funds rate target by either 25 or 50 basis points, whereas the previous practice involved making target changes of various amounts. (There was one exception: on October 15 1994 the Fed raised the funds rate target by 75 basis points.)

[13] In fact, during our sample period, with two exceptions, the FF rate target was changed at regularly scheduled FOMC meetings. The exceptions occurred on April 18 1994 and October 15 1998.

5 Empirical results[14]

5.1 Preliminary data statistics, unit root and cointegration tests

Table 1 presents summary statistics and the results of unit root tests for the series of interest in this paper. The summary statistics reported in Panel a) of Table 1 show that all four rates display very similar values for the mean, variance, skewness and kurtosis. Indeed, an examination of the third and fourth moments indicates the existence of both excess skewness and kurtosis, suggesting that the underlying distribution of each of these time series appears to be non-normal. This is clearly confirmed by the strong rejections of the Jarque-Bera test for normality reported in the last row of Panel a).[15]

Before presenting the estimates of the VECMs, the results of preliminary unit root and cointegration tests are presented. Tests for a unit root in each of s_t, s_t^T, f_t^1 and f_t^2 are reported in Panel b) of Table 1. The standard augmented Dickey-Fuller (ADF) test does not enable us to reject the unit root null hypothesis for any of the four rates.[16] Moreover, differencing the series appears to induce stationarity, confirming that each of the time series examined is an I(1) process.

The results of the Johansen (1988, 1991) maximum likelihood procedure for testing for cointegration for each of the three models are summarized in Table 2. Consistent with our priors, discussed in Section 2, the Johansen likelihood ratio test statistics (based on the maximal eigenvalue and on the trace of the stochastic matrix) indicate that there is one cointegrating vector in Model I, two cointegrating vectors in Model II, and three cointegrating vectors in Model III.[17]

Tests of the over-identifying restrictions on the β' matrix of cointegrating coefficients are reported in Panel b) of Table 2. The results indicate that the unity restrictions implied by the framework described in Section 2 could not be rejected at conventional levels of significance for

[14] In all statistical tests executed in this and subsequent sections, we use a five percent nominal significance level, unless otherwise specified.

[15] Indeed, nonparametric estimation of the density of the FF rate clearly shows that there are three modes, in addition to excess skewness and kurtosis. These features of the higher moments are also present in funds rate changes.

[16] The lag length was chosen to be the number of lags such that no residual autocorrelation was evident in the auxiliary regressions. However, using non-augmented Dickey Fuller tests or augmented Dickey-Fuller tests with any number of lags in the range from 1 to 10 yielded qualitatively identical results. See also footnote 3.

[17] We allowed for a maximum lag length of five and chose, for each model, the appropriate lag length on the basis of conventional information criteria. However, the cointegration results were found to be robust when using any number of lags in the range between one and five. In each VAR, we allowed for a unrestricted constant term.

any of the cointegrating relationships.

5.2 Linear VECM estimation results and linearity testing

Next, we estimate a standard linear VECM of the form

$$\Delta y_t = \nu + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + \Pi y_{t-1} + u_t \tag{8}$$

using full-information maximum likelihood (FIML) methods for each of Models I-III over the sample period January 1990-December 1995. We set $p = 1$, as suggested by the Akaike Information Criterion, the Schwartz Information Criterion and the Hannan-Quinn Criterion. A dummy variable is included in the equilibrium correction term to account for the 1994 procedural change discussed in Section 4. The dummy variable equals zero until 1994 and unity thereafter. The results, reported in Table A1 in Appendix A, suggest that the dummy is not statistically significant in Model II, so it was dropped from that model in the final estimation. However, the dummy is significant in Models I and III.[18] The incremental value of the adjustment parameters due to the dummy is negative, indicating that the FF rate adjusts more rapidly towards the equilibrium level determined by the FF rate target after the Fed began the practice of announcing publicly changes in the FF rate target upon making them in 1994.

An analysis of the residuals, presented in Table A1 in Appendix A, shows that none of the models exhibit significant residual serial correlation. All three models perform reasonably well in sample. The largest adjusted coefficient of determination ($\overline{R}^2$) in the estimated VECMs is always for the FF rate equation, with values of 0.35, 0.28, and 0.35 for Models I, II and III respectively. While these estimates are rather similar, it appears that the models that include the FF rate target (namely Models I and III) explain slightly more of the variation in the FF rate than models based on the term structure of the FF futures rates. In light of the finding that there is little difference in the estimates of the $\overline{R}^2$ for Models I and III and that the estimate of the $\overline{R}^2$ for Model II is smaller than that of the other two models, one might be tempted to conclude that futures rates may not have incremental information over the FF rate target. There are two reasons to be cautious, however. First, in-sample fit is not necessarily a good indicator of out-of-sample predictive power.

[18] In particular, in Model III the dummy was significant only for the equilibrium correction term involving the target rate s_t^T. In fact, in the final estimation this is the only dummy considered.

Second, because they do not take into account moments higher than the second in evaluating the ability of a model to explain a particular variable, standard goodness-of-fit measures such as the $\overline{R}^2$ are less reliable in the presence of non-normally distributed variables.

Ramsey's (1969) RESET test was also applied to the residuals from the three estimated linear VECMs. The results, reported in Table A1 in Appendix A, reveal evidence that each of the linear VECMs fails to capture statistically significant nonlinearities in the unknown data generating process driving the FF rate. Furthermore, the Jarque-Bera tests also indicate that the residuals from the linear VECMs are highly non-normal, suggesting that these linear VECMs fail to capture the intrinsic non-normality of the FF rate.[19]

5.3 MS-VECM estimation results

We account for nonlinearities in the data by estimating a Markov-switching model of the form:

$$\Delta y_t - \delta(z_t) = \alpha\left[\beta' y_{t-1} - \mu(z_t)\right] + \sum_{i=1}^{p-1} \Gamma_i \left[\Delta y_{t-i} - \delta(z_t)\right] + \omega_t, \tag{9}$$

where Δy_t is defined as before for each of Models I-III; $\delta(z_t)$ is the regime-dependent vector of means of the short-run dynamics; and $\mu(z_t)$ is the regime-dependent vector of means of the long-run equilibrium relationships.

The conventional 'bottom-up' procedure was applied. This procedure is designed to detect Markovian shifts in order to select the most adequate characterization of a Q-regime p-th order MS-VECM for Δy_t.[20] The VARMA representations of the series (see Poskitt and Chung, 1996; and Krolzig, 1997) suggested in each case that there are between two and three regimes. The linearity test, reported in the first column of Table A2 in Appendix A, indicates the rejection of the linear VECM in favor of a nonlinear alternative model. An analysis of the in-sample

[19] We also tested, in Model II, whether the futures rates' implicit funds rate forecasts are unbiased, which would imply market efficiency. Market efficiency also implies that the changes in the futures rates are unpredictable, so that the coefficients on lagged differences and the equilibrium correction term in the futures rate equation should be zero under market efficiency. However, we strongly rejected market efficiency at conventional significance levels (results not reported to conserve space but available upon request).

[20] Essentially, the bottom-up procedure consists of starting with a simple but statistically reliable Markov-switching model by restricting the effects of regime shifts on a limited number of parameters and testing the model against alternatives. In such a procedure, most of the structure contained in the data is not attributed to regime shifts, but explained by observable variables, consistent with the general-to-specific approach to econometric modelling. For a comprehensive discussion of the bottom-up procedure, see Krolzig (1997).

and out-of-sample performance of the nonlinear counterparts of the linear VECMs I, II and III (discussed in detail below) revealed that Model II was the best in terms of both in-sample and out-of-sample forecasting performance. Consequently, to conserve space only the results for Model II are reported in Table A3 (Appendix A).[21]

For all of models considered, the assumption that the regime shifts affect only the intercept term of the VECM was found to be inappropriate. A check for conditional homoskedasticity by estimating an MS-VECM where the innovation is allowed to be regime-dependent, $\varepsilon_t \sim NID(\mathbf{0}, \Sigma(z_t))$, yielded evidence of regime dependence. The hypothesis of no regime dependence in the variance-covariance matrix was tested using a likelihood ratio (LR) test of the type suggested by Krolzig (1997, p. 135-6), in addition to constructing an LR test for the null hypothesis of no regime dependent intercept. The results indicated strong rejection of the null of no regime dependence, suggesting that an MS-VECM that allows for shifts in both the intercept *and* the variance-covariance matrix, MSIH-VECM(p), is the most appropriate model within its class for this application.

To determine the order of the MSIH-VECM(p) we tested the null of an MSIH-VECM(1) against the alternative of an MSIH-VECM(p). The results, reported in the second column of Table A2 (Appendix A), show that we are unable to reject the null hypothesis. Consequently, it appears that a first-order specification is sufficient to characterize the dynamics between the FF rate and the FF futures rates.

Finally, in order to discriminate between models allowing for two regimes against models governed by three regimes the upper bound LR test of Davies (1987) was constructed. The results, reported in the last column of Table A2 (Appendix A), suggest that three regimes may be appropriate.[22] Therefore, we allowed for three regimes.

It is instructive to note that model (9), where the regime shifts occur in the drift of the VECM as well as the equilibrium mean of the cointegrating relationships, can be equivalently represented by means of an MSI-VECM.[23] Hence, the MS-VECM, governed by three different regimes, can

[21] The results of the other MS-VECMs are available upon request.

[22] It is important to note here that the regularity conditions under which the Davies (1987) test is valid are violated, since the Markov model has both a problem of nuisance parameters and a problem of 'zero score' under the null hypothesis. Therefore, the distribution of the LR test is likely to differ from the adjusted χ^2 distribution proposed by Davies (1987). For extensive discussions of the problems related to LR testing in this context, see Hansen (1992, 1996) and Garcia (1998). We are thankful to Bruce Hansen for clarifying several econometric issues related to LR testing in the present context.

[23] In order to recognize the shifts in the drift of the VECM separately from the ones occurring in the long-

be written as follows:

$$\Delta y_t = v(z_t) + \Pi y_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + \omega_t, \tag{10}$$

where $\Pi = \alpha\beta'$, $\omega_t \sim NIID(\mathbf{0}, \Sigma(z_t))$ and $z_t = 1, 2, 3$. In Table A3 in Appendix A, we report the final MSIH-VECM estimation results for Model II. The estimation yields fairly plausible estimates of the coefficients, including the adjustment coefficients in α, which were generally found to be statistically significantly different from zero.[24] [25] Also, the RESET test for the null of no misspecification and the Jarque-Bera test for the null of normality of the residuals suggest non-rejection in each case. Hence, the MSIH-VECM appears to have captured some important features of the unknown data generating process driving the FF rate.

It is difficult to give a precise interpretation to the three regimes identified by the MSIH-VECM. Shifts from one regime to another are due largely to shifts in the variance of the term structure equilibrium–shifts in the intercepts were found to be smaller in magnitude, albeit statistically significant. One possibility is that regimes are capturing a time-varying risk premium that reflects a departure from the no-arbitrage condition defined in Section 2.

We now turn to our forecasting results.

---

term equilibrium mean, consistent with the standard theoretical literature on multiple cointegrated time series, it is possible to decompose the shifts in the intercept term $v(z_t)$ into changes in the drift of the system $\delta(z_t) = \beta_{\perp}(\alpha'_{\perp}\beta_{\perp})^{-1}\alpha'_{\perp}v(z_t)$ ($\perp$ denoting the orthogonal complement) and the equilibrium mean $\mu(s_t) = -(\beta'\alpha)^{-1}[\beta' v(z_t)]$.

[24] Investigation of the estimated smoothed transition probabilities (not reported to conserve space) for the different regimes indicated that, in general, all of the three regimes seem to be important in that they characterize a substantial fraction of the joint movements of the FF rate and the term structure of futures rates.

[25] We also examined graphs of standardized residuals, the smoothed residuals and the one-step prediction errors for each estimated MSIH-VECM. The difference is concerned with the weighting of the residuals. Loosely speaking, the smoothed residuals are the closest to the sample residuals from a linear regression model; however, they overestimate the explanatory power of the Markov-switching model due to the use of the full-sample information covered in the smoothed regime vector. The standardized residuals are conditional residuals. The one-step prediction errors are based on the predicted regime probabilities. Unfortunately, many conventional diagnostic tests, such as standard residual serial correlation tests, may not have their conventional asymptotic distribution when the residuals come from Markov-switching models and are therefore not reported here. However, the graphs of standardized residuals, the smoothed residuals and the one-step prediction errors provided no visual evidence of residual serial correlation in any of the residuals series plotted. See, for example, Krolzig (1997) for a discussion of residual-based model checking in this context.

6 Out-of-sample forecasting results

6.1 Linear VECM forecasting results

The predictive ability of our linear and nonlinear models is investigated by calculating up to forty steps-ahead (one to eight weeks ahead) forecasts using each of the linear models over the period January 1996-December 2000. The out-of-sample forecasts for a given horizon are constructed recursively. That is, the forecasts are conditional only upon information available up to the date of the forecasts because the model is re-estimated as the date on which forecasts are conditioned moves.

The forecasting results for the linear VECMs are summarized in Table 3. The ability of the linear VECMs to produce accurate point forecasts of the FF rate is evaluated using several criteria. Panels a) and b) of Table 3 show the difference between the mean absolute errors (MAEs) and mean square errors (MSEs) for each of the estimated models compared to the remaining two. Further, in order to assess the accuracy of forecasts derived from two different models we employ the Diebold and Mariano (1995) test:

$$DM = \frac{\overline{d}}{\sqrt{\frac{2\pi\widehat{f}(0)}{T}}} \tag{11}$$

where $\overline{d}$ is an average (over T observations) of a general loss differential function and $\widehat{f}(0)$ is a consistent estimate of the spectral density of the loss differential function at frequency zero. Diebold and Mariano show that the DM statistic is distributed as standard normal under the null hypothesis of equal forecast accuracy. The loss differential function we consider is the difference between the (absolute and square) forecast errors. A consistent estimate of the spectral density at frequency zero $\widehat{f}(0)$ is obtained using the method of Newey and West (1987), where the optimal truncation lag has been selected using the Andrews (1991) AR(1) rule.[26]

On the basis of MAEs and MSEs, Model II seems to underperform (produces higher MAEs and MSEs) relative to the other two alternative models across the forecast horizons considered (5, 10,

[26] The rule is implemented as follows: we estimated an AR(1) model to the quantity d_t obtaining the autocorrelation coefficient $\widehat{\rho}$ and the innovation variance from the AR(1) process $\widehat{\sigma}^2$. Then the optimal truncation lag A for the Parzen window in the Newey-West estimator is given by the Andrews rule $A = 2.6614\left[\widehat{\zeta}(2)\,T\right]^{1/5}$ where $\widehat{\zeta}(2)$ is a function of $\widehat{\rho}$ and $\widehat{\sigma}^2$. The Parzen window minimizes the mean square error of the estimator (Gallant, 1987, p. 534).

20 and 40 steps ahead, that is 1, 2, 4 and 8 weeks ahead). However, the results of the Diebold-Mariano test of the null hypothesis of equal forecast accuracy indicate that the null hypothesis cannot be rejected for any pair of models. Hence, the differences in terms of MAEs and MSEs reported in Table 3 are not statistically significant.[27]

The results suggest that the FF rate target and the term structure of FF futures rates have equal predictive power in forecasting the FF rate. This suggests that a more stringent criterion may be necessary in order to differentiate between the forecasting ability of competing models. Given that the FF rate appears to be non-normal it seems reasonable to evaluate the ability of the competing models to forecast the FF rate out of sample in terms of predictive density. Hence, we investigate which model provides the predictive density that is closest to the true predictive density of the data over the sample period used to perform the out-of-sample forecasting exercise.

A large body of literature in financial econometrics has recently focused on evaluating the forecast accuracy of models on the basis of density forecasting performance (see, *inter alia*, Diebold, Gunther and Tay, 1998; Diebold, Hahn and Tay, 1999; Granger and Pesaran, 1999; Timmerman, 2000). In general, this line of research has produced several methods either to measure the closeness of two density functions or to test the hypothesis that the predictive density generated by a particular model corresponds to the true predictive density. However, these tests do not allow us to test directly the null hypothesis of equal density forecast accuracy between competing models.

In order to test formally the null hypothesis of equal density forecast accuracy between the competing linear models, we employed the η test recently proposed by Sarno and Valente (2002). The η test is similar in spirit to the test suggested by Diebold and Mariano (1995) but it involves

[27]Several problems may arise using DM statistics in small sample as well as taking into account parameter uncertainty (see also West, 1996; West and McCracken, 1998; and McCracken, 2000). In our case, where we are dealing with nested competing forecasting models - some of which in the next subsection are nonlinear - and with multi-step-ahead forecasts, the asymptotic distribution of the DM statistic is non-standard and unknown. Therefore, the marginal significance levels reported below should be interpreted with caution. Clark and McCracken (2001) derive the asymptotic distributions of two standard tests in this context for one-step-ahead forecasts from nested linear models. Their results are, unfortunately, not directly applicable to our case because we are dealing with multi-step-ahead forecasts from nested models, and because one of the competing models is a Markov-switching model. Our case is one for which the asymptotic theory of the DM statistic is unknown at the present time. A possible solution involves calculating the marginal significance levels by bootstrap methods using a variant of the bootstrap procedure proposed by Kilian and Taylor (2001), although this procedure is computationally very demanding and it is unknown whether it is valid in the context of MS-VECMs.

the analysis of the whole predictive density instead of point forecasts.[28] This test statistic is constructed as follows:

$$\eta = \frac{\overline{d}}{\sqrt{\frac{\widehat{\sigma^2}}{B}}}, \tag{12}$$

where $\overline{d} = \frac{1}{B}\sum_{j=1}^{B} d^j = \frac{1}{B}\sum_{j=1}^{B}\left(\widehat{ISD}_1^j - \widehat{ISD}_2^j\right)$ is an average (over $j = 1, ..., B$ bootstrap replications) of the difference between two estimated integrated square differences, $\widehat{ISD}_1^j$ and $\widehat{ISD}_2^j$; $\widehat{ISD}_1^j$ is the integrated square difference between a generic model M_1 and the true predictive density, and $\widehat{ISD}_2^j$ is the integrated square difference between a generic model M_2 and the true predictive density; $\widehat{\sigma^2}$ is a consistent bootstrap estimate of the variance of the difference d.[29] Under general conditions, the η test is asymptotically distributed as standard normal under the null hypothesis that the two competing models M_1 and M_2 have equal density forecast accuracy.

The results from applying the η test to our linear VECMs, reported in the last section of Table 3, are in sharp contrast to those obtained using standard out-of-sample statistics that only consider point forecast accuracy. For every forecast horizon, Model II produces the best density forecasts, with the null hypothesis of equal density forecast accuracy being rejected with p-values of virtually zero. Hence, while it is not possible to discriminate between the linear forecasting models in terms of their point forecasting performance, Model II emerges clearly as the best forecasting model of the FF rate in terms of density forecasting. Before providing potential explanations of this finding, we examine the forecasting performance of the MS-VECMs.

6.2 MS-VECM forecasting results

In order to evaluate the usefulness of allowing for nonlinearities and measure the gain from using a nonlinear model, we used MS-VECMs to produce dynamic out-of-sample forecasts of the FF rates up to forty steps ahead (eight weeks ahead) over the period January 1996-December 2000. We first attempt to discriminate between the three MS-VECMs. We then compare the best MS-VECM to each of the linear VECMs estimated in the previous section in order to identify the overall best

[28] See Appendix B for details on the calculation and the properties of the η-test.

[29] The estimated integrated square difference $\widehat{ISD} = \int \left[\widehat{\phi}(x) - \widehat{\gamma}(x)\right]^2 dx$ is obtained by estimating the density functions ϕ and γ by means of the Gaussian kernel estimator.

forecasting model of the FF rate. Finally, we compare the best linear and nonlinear models to the futures rate time series, which would be the optimal predictor under market efficiency.

It is well known in the literature that forecasting with nonlinear models raises special problems. We therefore adopt a very general forecasting procedure based on Monte Carlo integration which is capable of producing forecasts virtually identical to analytical forecasts for a wide range of models. In particular, we forecast the out-of-sample path for the FF rate using Monte Carlo simulations calibrated on our estimated MSIH-VECMs. The vector of Gaussian innovations is set consistent with the estimated covariance matrices. The simulation procedure is repeated with identical random numbers 10,000 times and the average of the 10,000 realizations of the sequence of forecasts is taken as the point forecast. Since we use a large number of simulations, by the Law of Large Numbers this procedure should produce results that are virtually identical to those which would result from calculating the exact forecast analytically (see, *inter alia*, Brown and Mariano, 1984, 1989; Granger and Teräsvirta, 1993, chapter 8; Franses and van Dijk, 2000, chapters 3-4). Again, the forecasts are evaluated in terms of both point and density forecasting performance.

The first two panels of Table 4 show the difference between MAEs and MSEs from each estimated MS-VECM compared to the other two MS-VECMs. On the basis of these criteria, Model II seems to outperform the other two nonlinear competing models across the forecasting horizons considered. However, the results of the Diebold and Mariano (1995) test indicate that the null hypothesis is rejected only when comparing Model II with Model III and using the MAE measure. This suggests that a nonlinear model exploiting the information in the term structure of FF rates performs better in out-of-sample point forecasting than a nonlinear model combining the term structure of FF futures rates and the FF rate target.[30] As was the case with the linear models, in general, the differences in terms of MAEs and MSEs reported in Table 4 are not statistically significant.

The last section of Table 4, however, suggests that the MSIH-VECM II significantly outperforms the other two Markov-switching models in terms of density forecasting performance (using the η-test), with p-values close to zero. This result is consistent with the outcome of the forecasting

[30] In turn, this suggests that although there may be some independent information in the FF rate target and in the term structure of FF futures rates, the incremental information in the FF rate target is not sufficiently large to offset the loss of parsimony that occurs when the target is entered in the VECM.

exercise using linear models. While we are not able to choose among competing models on the basis of point forecasting criteria, we could do so on the basis of their density forecasting performance. Once again, the model that uses only the information in the term structure of the FF futures rates outperforms the alternative models.

Next, in order to quantify the gain from using a nonlinear model, we compare the out-of-sample forecasts from the linear VECMs discussed in the previous section to the forecasts provided by the best nonlinear model (i.e. MS-VECM II). Table 5 reports the difference between the MAEs and MSEs for each of the estimated linear models compared to the MSIH-VECM II. On the basis of these criteria, the MSIH-VECM II produces higher MAEs and MSEs than all the alternative models across the various forecast horizons, but in no instance does the Diebold-Mariano test reject the null of equal predictive accuracy.

However, the test of predictive density forecast accuracy, reported in the last section of Table 5, confirms the importance of the nonlinearities that are explicitly considered in the MSIH-VECM. The null hypothesis is easily rejected in each case. More precisely, for each alternative model and each forecast horizon, the MSIH-VECM II produces the best density forecasts, with the null hypothesis of equal density forecast accuracy being rejected with p-values of virtually zero.

As a final exercise, we investigated whether the simple time series of the FF futures rate can outperform the predictions embedded in the term structure of the FF futures rates at different maturities. To this end, we compared the out-of-sample forecasting performance of the best linear and nonlinear models, VECM II and MSIH-VECM II respectively, to the simple time series of the FF futures rate at 1 and 2 months maturity. Again, the predictive performance was evaluated using absolute and square differences between the predictions of the futures rates and the two competing models at forecast horizons corresponding to the maturity of the FF futures rate (namely 4 and 8 weeks or 1 and 2 months respectively).

As the results in Table 6 show, both linear and nonlinear VECMs outperform the simple FF futures rates across the forecast horizons examined in terms of mean errors. As before, the differences were not statistically significant on the basis of the DM test. If one is concerned about forecasting the whole predictive density, however, the term structure models–linear or nonlinear–significantly outperform the simple FF futures rates at these forecast horizons.

The results of our density forecasting exercises are summarized in Figures 1-2, which report (for each forecast horizon considered) the true predictive density of the FF rate (in first difference) over the forecast period together with the predictive densities implied by the three linear VECMs reported in Table A1 and our preferred MS-VECM, based on the term structure of the FF futures rates. For each of Figures 1 and 2 (which include 1, 2-, 4- and 8-week ahead forecasts), the true predictive density of the FF rate data and the predictive densities of the forecasts from the three linear VECMs are reported in the left panel. The right panel of Figures 1-2 show the true predictive density of the FF rate data and the predictive densities from the best linear model (Model II) and the best MS-VECM (MSIH-VECM (10)).

The differences in predictive density forecasting, largely due to excess kurtosis of the predictive densities of the model forecasts, are apparent. The best performing model in terms of density forecasting performance (closeness of a model's predictive density to the true predictive density of the data) is the regime-switching version of Model II.[31] The second best model is the linear model of the term structure. Hence, it appears that our ability to match reasonably well the predictive density of the FF rate is due, in part, to the information in the term structure of FF futures rates. The fact that the nonlinear model is statistically superior to the linear model, however, suggests that it is also important to allow for regime shifts in the term structure VECM.

6.3 Summing up the forecasting results

These forecasting exercises suggest the following results. First, it is very difficult, if at all possible, to discriminate between models of the FF rate target and models of the term structure of the FF futures rates in terms of their predictive power using conventional point forecasting accuracy criteria, such as mean (absolute or square) errors. The forecasts from several alternative (linear and nonlinear) models could not be distinguished statistically from one another or from the simple futures rate forecast on the basis of point forecasts (Table 6).

Second, it appears that one can discriminate statistically between these models on the basis of their density forecasting performance. The evidence suggests that, by this criterion, the linear (nonlinear) model based on the term structure of FF futures rates is the best forecasting model

[31] In fact, the η-test results discussed earlier essentially confirm that the visual evidence provided by the right panels of Figures 1-2 is significant in a statistical sense.

among the linear or nonlinear models considered.

Third, we find that the general MSIH-VECM (10) performs significantly better than the other linear and nonlinear models considered (as well as the simple time series of the FF futures rates) in terms of predicting the out-of-sample density of the FF rate. Indeed, the performance of the linear term structure model was significantly improved by explicitly allowing for regime switches in both the intercept and the variance-covariance structure of the VECM. This result suggests that not only is the term structure of the FF futures rates important for capturing the behavior of the funds rate, but also that the parametric formulation of the model is important in enhancing the predictive power of the FF futures rates. The difference between the predictive density of the term structure linear VECM and the term structure MS-VECM, graphed in Figures 1-2, clearly quantifies the gain from estimating and forecasting from a nonlinear VECM rather than a linear VECM.

7 Conclusion

In this paper we reported what we believe to be the first analysis of the federal funds rate, the federal funds futures rate and the federal funds rate target in a multivariate Markov-switching framework, and in particular we have applied that framework to forecast the future federal funds rate. Our research was inspired by encouraging results previously reported in the literature on the presence of nonlinearities (and particularly by the success of Markov-switching models) in the context of interest rate modelling in general as well as by the relative success of the term structure of federal funds futures rates and the federal funds rate target in explaining the behavior of the federal funds rate.

Using daily data on the federal funds rate, the funds rate target and the 1- and 2-months futures rates over the period January 1990 through December 1995, we found strong evidence of the explanatory power of the term structure of futures rates and the funds rate target and of the presence of nonlinearities in each of three models of the federal funds rate, which appeared to be modelled well using a multivariate three-regime Markov-switching VECM that allows for shifts in both the intercept and in the covariance structure. We then used this model to forecast dynamically out of sample over the period January 1996 through to December 2000 and compared these forecasts to the forecasts obtained from using a variety of linear and nonlinear models. The forecasting results were interesting. We have found that both the term structure of futures rates and nonlinearity are very important in forecasting the federal funds rate. However, their importance is not obvious when the forecasting ability of our proposed nonlinear VECM is evaluated on the basis of conventional point forecasting criteria. Using general tests for equal point forecast accuracy (e.g. Diebold and Mariano statistics) we could not distinguish among different alternative models using as predictor variable the federal funds futures, the federal funds rate target or both.

In order to measure the forecasting ability of our nonlinear model more accurately we employed a test of the null hypothesis of equal density forecast accuracy, which revealed three clear findings. First, models exploiting the information embedded in the term structure of the federal funds futures rate are the best forecasting models (among the models considered) as they outperform models that use the information in the federal funds rate target. Second, nonlinear VECMs provide sizable improvements relative to their linear counterparts, capturing adequately the non-normality of the federal funds rate and producing more accurate out-of-sample density forecasts. Third, the linear and nonlinear term structure models outperform the simple futures rates time series in terms of density forecasting performance.

Overall, these findings suggest that it is possible to build time series models that can forecast fairly accurately the federal funds rate at horizons up to 2 months or so, using available information. However, obtaining accurate forecasts requires estimation of and forecasting from a nonlinear VECM which allows for intercept corrections and shifts in the variance-covariance matrix. In terms of point forecasting the performance of the linear and nonlinear models considered in this paper is essentially identical, and we were only able to establish the forecasting superiority of the nonlinear VECM of the term structure of futures rates using a test that evaluates the relative density forecasting performance.

While these results aid the profession's understanding of the behavior of the federal funds rate, we view our nonlinear model as a tentatively adequate characterization of the data, which nevertheless is capable of improvement. In particular, while the model used here is fairly general and flexible, the evidence we document suggests that the federal funds rate, the federal funds rate target and the term structure of federal funds futures rates are linked by very complex dynamic interactions. Much more work needs to be carried out to shed light on these relationships. More importantly, while this paper adds to the literature that proposes the use of density forecasting measures in addition to point forecasts, it suffers from a common criticism in this literature. We refer to the fact that, while density forecasts contain a lot more information than point forecasts, it remains unclear how practitioners and policy makers could use this information and act upon it. These questions remain on the agenda for future research.

Table 1. Preliminary data statistics and unit root tests

Panel a): Summary statistics

Levels

	s_t	f_t^1	f_t^2	s_t^T
Maximum	10.390	8.530	8.660	8.250
Minimum	2.580	2.760	2.820	3.000
Mean	5.265	5.277	5.304	5.224
Variance	1.895	1.771	1.741	1.832
Skewness	0.241	0.139	0.097	0.179
Kurtosis	3.080	3.001	2.976	3.010
JB	1.00×10^{-6}	9.31×10^{-4}	1.01×10^{-3}	4.39×10^{-5}

First differences

	Δs_t	Δf_t^1	Δf_t^2	Δs_t^T
Maximum	2.830	0.550	0.380	0.750
Minimum	-2.700	-0.320	-0.330	-0.500
Mean	-2.43×10^{-3}	-1.68×10^{-3}	-1.73×10^{-3}	-1.75×10^{-3}
Variance	1.01×10^{-1}	1.79×10^{-3}	1.97×10^{-3}	2.11×10^{-3}
Skewness	0.791	2.116	0.210	1.449
Kurtosis	21.804	42.447	20.632	110.149
JB	0	0	0	0

Panel b): Unit root tests

Levels

	s_t	f_t^1	f_t^2	s_t^T
ADF test	-2.073	-2.055	-2.051	-2.103

First Differences

	Δs_t	Δf_t^1	Δf_t^2	Δs_t^T
ADF test	-26.768	-24.129	-20.947	-15.419

Notes: s_t, f_t^1, f_t^2, s_t^T denote the effective federal funds rate, the one-month federal funds futures rate, the two-month federal funds futures rate, and the federal funds rate target respectively. Δ is the first difference operator. In Panel a), JB denotes the Jarque-Bera test for normality, for which only p-values are reported. 0 indicates $p-$values below 10^{-500}, which are considered as virtually zero. In Panel b), statistics are augmented Dickey-Fuller test statistics for the null hypothesis of a unit root process, calculated allowing for a constant term in the auxiliary regression. The asymptotic critical value at the 5 (10) significance level is -2.86 (-2.57) (see Fuller, 1976; MacKinnon, 1991).

Table 2. Johansen maximum likelihood cointegration procedure

Panel a): Cointegration tests

Model I : s_t and s_t^T

$\lambda_{\max}$				λ_{trace}			
H_0	H_1	LR	5%	H_0	H_1	LR	5%
$r = 0$	$r = 1$	553.70	15.7	$r = 0$	$r \geq 1$	560.50	20.0
$r \leq 1$	$r = 2$	6.79	9.2	$r \leq 1$	$r = 2$	6.79	9.2

Model II: f_t^1, $f_t^2$and s_t

$\lambda_{\max}$				λ_{trace}			
H_0	H_1	LR	5%	H_0	H_1	LR	5%
$r = 0$	$r = 1$	429.80	22.0	$r = 0$	$r = 1$	479.80	34.9
$r \leq 1$	$r = 2$	47.95	15.7	$r \leq 1$	$r = 2$	51.62	20.0
$r \leq 2$	$r = 3$	3.86	9.2	$r \leq 2$	$r = 3$	3.86	9.2

Model III: s_t *and* s_t^T, f_t^1, f_t^2

$\lambda_{\max}$				λ_{trace}			
H_0	H_1	LR	5%	H_0	H_1	LR	5%
$r = 0$	$r = 1$	567.60	28.1	$r = 0$	$r = 1$	710.80	53.1
$r \leq 1$	$r = 2$	84.46	22.0	$r \leq 1$	$r = 2$	143.20	34.9
$r \leq 2$	$r = 3$	55.01	15.7	$r \leq 2$	$r = 3$	58.78	20.0
$r \leq 3$	$r = 4$	3.76	9.2	$r \leq 3$	$r = 4$	3.76	9.2

(continued ...)

(... Table 2 continued)

Panel b): Test for the restrictions on the cointegration space

$$\textit{Model I}: \beta' y_t = \begin{bmatrix} -1 & 1 \end{bmatrix} \begin{bmatrix} s_t \\ s_t^T \end{bmatrix}$$

$$\textit{Model II}: \beta' y_t = \begin{bmatrix} -1 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} s_t \\ f_t^1 \\ f_t^2 \end{bmatrix}$$

$$\textit{Model III}: \beta' y_t = \begin{bmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} s_t \\ s_t^T \\ f_t^1 \\ f_t^2 \end{bmatrix}$$

	$\chi^2(g)$	
Model I	0.1185	{0.7307}
Model II	0.0613	{0.9698}
Model III	0.0746	{0.9947}

Notes: *Panel a):* H_0 and H_1 denote the null hypothesis and the alternative hypothesis respectively; r denotes the number of cointegrating vectors; $\lambda_{\max}$ and λ_{trace} denote the likelihood ratio cointegration test based on the maximum eigenvalue of the stochastic matrix and the likelihood ratio test based on the trace of the stochastic matrix respectively (Johansen, 1995); the 5% critical values reported in the last column are taken from Osterwald-Lenum (1992). *Panel b):* The test is a χ^2 version of the test of the overidentifying restrictions on the β' matrix; g is the number of restrictions imposed. Figures in braces are p-values.

Table 3. Forecasting exercises - linear models

Diebold and Mariano (1995) test: mean absolute errors

k	Model I - Model II	Model I - Model III	Model II - Model III
1	-0.1858	0.0033	0.1858
	{0.852}	{0.997}	{0.852}
2	-0.1858	0.0026	0.1849
	{0.852}	{0.997}	{0.853}
4	-0.1841	0.0032	0.1852
	{0.853}	{0.997}	{0.853}
8	-0.1735	-0.0047	0.1696
	{0.862}	{0.996}	{0.865}

Diebold and Mariano (1995) test: mean square errors

k	Model I - Model II	Model I - Model III	Model II - Model III
1	-0.0609	-0.0037	0.0559
	{0.951}	{0.997}	{0.955}
2	-0.0618	-0.0038	0.0567
	{0.951}	{0.997}	{0.955}
4	-0.0600	-0.0037	0.0550
	{0.952}	{0.997}	{0.956}
8	-0.0547	-0.0063	0.0473
	{0.956}	{0.995}	{0.962}

Density forecasting test: η test

k	Model I - Model II	Model I - Model III	Model II - Model III
1	5.2545	0.0333	-5.2311
	$\{1.4 \times 10^{-7}\}$	$\{9.7 \times 10^{-1}\}$	$\{1.7 \times 10^{-7}\}$
2	5.6407	0.0778	-5.5310
	$\{1.7 \times 10^{-8}\}$	$\{9.4 \times 10^{-1}\}$	$\{3.2 \times 10^{-8}\}$
4	4.5284	0.0837	-4.7780
	$\{5.9 \times 10^{-6}\}$	$\{9.3 \times 10^{-1}\}$	$\{1.8 \times 10^{-6}\}$
8	5.6182	0.2481	-4.7518
	$\{1.9 \times 10^{-8}\}$	$\{8.0 \times 10^{-1}\}$	$\{2.0 \times 10^{-6}\}$

Notes: Model i - Model j ($i = I, II; j = II, III$) is the Diebold and Mariano test statistics obtained using the difference between the forecast errors of out-of-sample dynamic forecast of the competing linear models up to $k = 1, 2, 4, 8$ weeks ahead over the period 1996-2000. The asymptotic variance of the Diebold and Mariano statistics has been calculated choosing the optimal truncation lag according to the AR(1) Andrews's (1991) rule. The η statistics (Sarno and Valente, 2002) have been calculated using a Gaussian kernel and setting the number of bootstrap replications equal to 100 (see Appendix B).

Table 4. Forecasting exercises - Markov-switching models

Diebold and Mariano (1995) test: mean absolute errors

k	MS I - MS II	MS I - MS III	MS II - MS III
1	1.3648	-0.6011	-2.0641
	{0.172}	{0.548}	{0.039}
2	1.2966	-0.6963	-1.9871
	{0.195}	{0.488}	{0.047}
4	1.2897	-0.6799	-1.9808
	{0.197}	{0.497}	{0.048}
8	1.2878	-0.7038	-2.0364
	{0.198}	{0.482}	{0.042}

Diebold and Mariano (1995) test: mean square errors

k	MS I - MS II	MS I - MS III	MS II - MS III
1	0.6726	-0.2081	-0.8985
	{0.501}	{0.835}	{0.369}
2	0.6809	-0.2203	-0.9356
	{0.496}	{0.826}	{0.349}
4	0.6821	-0.2144	-0.9463
	{0.495}	{0.830}	{0.344}
8	0.6840	-0.2347	-0.9654
	{0.494}	{0.814}	{0.334}

Density forecasting test: η test

k	MS I - MS II	MS I - MS III	MS II - MS III
1	5.9802	-0.1005	-6.6720
	$\{2.2 \times 10^{-9}\}$	$\{9.2 \times 10^{-1}\}$	$\{2.5 \times 10^{-11}\}$
2	5.1880	-0.0715	-5.3790
	$\{2.1 \times 10^{-7}\}$	$\{9.4 \times 10^{-1}\}$	$\{7.5 \times 10^{-8}\}$
4	6.3792	-0.6149	-6.9900
	$\{1.8 \times 10^{-10}\}$	$\{5.4 \times 10^{-1}\}$	$\{2.8 \times 10^{-12}\}$
8	5.6560	-0.1799	-3.5590
	$\{1.6 \times 10^{-8}\}$	$\{8.6 \times 10^{-1}\}$	$\{3.7 \times 10^{-4}\}$

Notes: MS i - MS j ($i = I, II; j = II, III$) is the Diebold and Mariano test statistics obtained using the difference between the forecast errors of out-of-sample dynamic forecast of the competing Markov-switching models up to $k = 1, 2, 4, 8$ weeks ahead over the period 1996-2000. The asymptotic variance of the Diebold and Mariano statistics has been calculated choosing the optimal truncation lag according to the AR(1) Andrews's (1991) rule. The η statistics (Sarno and Valente, 2002) have been calculated using a Gaussian kernel and setting the number of bootstrap replications equal to 100 (see Appendix B).

Table 5. Forecasting exercises - comparison

Diebold and Mariano (1995) test: mean absolute errors

k	Model I - MS II	Model II - MS II	Model III - MS II
1	-0.4105	-0.2891	-0.4092
	{0.6814}	{0.7725}	{0.6823}
2	-0.4301	-0.3199	-0.4274
	{0.6671}	{0.7490}	{0.6691}
4	-0.4228	-0.3121	-0.4225
	{0.6724}	{0.7549}	{0.6581}
8	-0.4045	-0.3047	-0.4006
	{0.6858}	{0.7605}	{0.6887}

Diebold and Mariano (1995) test: mean square errors

k	Model I - MS II	Model II - MS II	Model III - MS II
1	-0.1586	-0.1163	-0.1543
	{0.8739}	{0.9074}	{0.8773}
2	-0.1610	-0.1182	-0.1565
	{0.8720}	{0.9059}	{0.8756}
4	-0.1591	-0.1196	-0.1536
	{0.8735}	{0.9048}	{0.8779}
8	-0.1467	-0.1107	-0.1396
	{0.8833}	{0.9118}	{0.8889}

Density forecasting test: η test

k	Model I - MS II	Model II - MS II	Model III - MS II
1	6.6762	2.4747	6.3221
	$\{2.46 \times 10^{-11}\}$	$\{1.33 \times 10^{-2}\}$	$\{2.59 \times 10^{-10}\}$
2	6.5371	3.1195	6.6586
	$\{6.30 \times 10^{-11}\}$	$\{1.81 \times 10^{-3}\}$	$\{2.78 \times 10^{-11}\}$
4	5.6688	2.9299	6.0354
	$\{1.44 \times 10^{-8}\}$	$\{3.39 \times 10^{-3}\}$	$\{1.59 \times 10^{-9}\}$
8	6.4174	3.3841	6.3761
	$\{1.39 \times 10^{-10}\}$	$\{7.14 \times 10^{-4}\}$	$\{1.82 \times 10^{-10}\}$

Notes: Model i - MS ($i = I, II, III$) is the Diebold and Mariano test statistics obtained using the difference between the forecast errors of out-of-sample dynamic forecast of the competing models up to $k = 1, 2, 4, 8$ weeks ahead over the period 1996-2000. The asymptotic variance of the Diebold and Mariano statistics has been calculated choosing the optimal truncation lag according to the AR(1) Andrews's (1991) rule. The η statistics (Sarno and Valente, 2002) have been calculated using a Gaussian kernel and setting the number of bootstrap replications equal to 100 (see Appendix B).

Table 6. Predictive ability of the federal funds futures rate

Diebold and Mariano (1995) test: mean absolute errors

k	Δf_t^1 - Model II	Δf_t^1 - MS II
4	0.6340	0.2130
	{0.5260}	{0.8313}
	Δf_t^2- Model II	Δf_t^2- MS II
8	0.6570	0.2144
	{0.5111}	{0.8302}

Diebold and Mariano (1995) test: mean square errors

k	Δf_t^1 - Model II	Δf_t^1 - MS II
4	0.5685	0.4273
	{0.5696}	{0.6691}
	Δf_t^2- Model II	Δf_t^2- MS II
8	0.5647	0.4234
	{0.5722}	{0.6720}

Density forecasting test: η test

k	Δf_t^1 - Model II	Δf_t^2 - MS II
4	22.265	18.871
	{0}	{0}
	Δf_t^2- Model II	Δf_t^2- MS II
8	21.384	19.163
	{0}	{0}

Notes: f_t^1 and f_t^2 denote the the one- and two-month federal funds futures rates respectively; Δ is the first-difference operator. Model II and MS II denote the linear and nonlinear VECMs for the term structure of futures rates respectively. Figures denote the Diebold and Mariano test statistics obtained using the difference between the forecast errors from using the futures rates and the out-of-sample dynamic forecast of the competing models up to $k = 1, 2, 4, 8$ weeks ahead over the period 1996-2000. The asymptotic variance of the Diebold and Mariano statistics has been calculated choosing the optimal truncation lag according to the AR(1) Andrews's (1991) rule. The η statistics (Sarno and Valente, 2002) have been calculated using a Gaussian kernel and setting the number of bootstrap replications equal to 100 (see Appendix B). {0} indicates $p-$values below 10^{-500}, which are considered as virtually zero.

A Linear VECM and MS-VECM estimation results

Table A1. Linear VECM estimation results

$$Model\ I: y = \left[\begin{array}{cc} s_t & s_t^T \end{array}\right]'$$

$$\Delta y_t = \nu + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + \alpha \left[\begin{array}{cc} \beta' y_{t-1} & \beta' y_{t-1} d_{94} \end{array}\right] + u_t$$

$$\widetilde{\Gamma}_1 = \left[\begin{array}{cc} 0.0996 & -0.0585 \\ (0.025) & (0.004) \\ -0.0014 & -0.0026 \\ (0.143) & (0.025) \end{array}\right]; \quad \widetilde{v} = \left[\begin{array}{c} -0.00207 \\ (0.0065) \\ -0.00187 \\ (0.0011) \end{array}\right]; \quad \widetilde{\alpha} = \left[\begin{array}{cc} -0.7511 & -0.1628 \\ (0.0307) & (0.0055) \\ 0.0074 & -0.0216 \\ (0.0843) & (0.0151) \end{array}\right];$$

$$\widetilde{\Sigma} = \left[\begin{array}{cc} 0.0658 & 0.0010 \\ 0.0010 & 0.0021 \end{array}\right]; \quad \text{AIC = -3.1954;} \quad \text{BIC = -3.1509}$$

	$\overline{R}^2$	LM1	ARCH	WHITE	JB	RESET
s_t equation	0.3539	0.7737	2.66×10^{-7}	2.79×10^{-18}	1.83×10^{-164}	4.65×10^{-4}
s_t^T equation	0.0021	0.3111	0.7207	0.0148	0	7.02×10^{-4}

$$Model\ II: \ y = \left[\begin{array}{ccc} s_t & f_t^1 & f_t^2 \end{array}\right]'$$

$$\Delta y_t = \nu + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + \alpha\beta' y_{t-1} + u_t$$

$$\widetilde{\Gamma}_1 = \left[\begin{array}{ccc} 0.0182 & -0.5188 & 0.3141 \\ (0.0257) & (0.2887) & (0.27756) \\ -0.0046 & 0.0408 & 0.0294 \\ (0.00394) & (0.044) & (0.0425) \\ -0.0061 & 0.0601 & 0.0304 \\ (0.0041) & (0.047) & (0.045) \end{array}\right]; \quad \widetilde{v} = \left[\begin{array}{c} -0.0002 \\ (0.0069) \\ -0.0004 \\ (0.0010) \\ -0.0009 \\ (0.0011) \end{array}\right]; \quad \widetilde{\alpha} = \left[\begin{array}{cc} -1.0911 & 0.5024 \\ (0.0803) & (0.0644) \\ 0.0906 & -0.0728 \\ (0.0123) & (0.0098) \\ 0.0525 & -0.0423 \\ (0.0130) & (0.0104) \end{array}\right];$$

$$\widetilde{\Sigma} = \left[\begin{array}{ccc} 0.0728 & 0.0010 & 0.0009 \\ 0.0010 & 0.0017 & 0.0015 \\ 0.0009 & 0.0015 & 0.0019 \end{array}\right]; \quad \text{AIC = -7.8534;} \quad \text{BIC = -7.7506}$$

	$\overline{R}^2$	LM1	ARCH	WHITE	JB	RESET
s_t equation	0.2846	0.346	2.12×10^{-10}	4.93×10^{-14}	1.41×10^{-162}	3.31×10^{-2}
f_t^1equation	0.0438	0.877	0.727	5.10×10^{-11}	0	8.16×10^{-2}
f_t^2equation	0.0195	0.854	0.975	2.05×10^{-6}	0	2.71×10^{-2}

(continued ...)

(... Table A1 continued)

$Model\ III : y = \begin{bmatrix} s_t & s_t^T & f_t^1 & f_t^2 \end{bmatrix}'$

$\Delta y_t = \nu + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + \alpha \begin{bmatrix} \beta' y_{t-1} & \beta_1' y_{t-1} d_{94} \end{bmatrix} + u_t$

$$\widetilde{\Gamma}_1 = \begin{bmatrix} 0.1044 & -0.1837 & -0.0913 & -0.0638 \\ (0.0254) & (0.2833) & (0.2663) & (0.1558) \\ -0.0023 & 0.0707 & 0.0130 & -0.0234 \\ (0.0040) & (0.045) & (0.042) & (0.0249) \\ -0.0051 & 0.0751 & 0.0240 & -0.0196 \\ (0.0043) & (0.048) & (0.046) & (0.0266) \\ -0.0012 & 0.0630 & -0.0272 & -0.0174 \\ (0.0044) & (0.049) & (0.046) & (0.0272) \end{bmatrix}; \quad \widetilde{v} = \begin{bmatrix} -0.0011 \\ (0.0066) \\ -0.0007 \\ (0.0010) \\ -0.0011 \\ (0.0011) \\ -0.0003 \\ (0.0011) \end{bmatrix};$$

$$\widetilde{\alpha} = \begin{bmatrix} -0.1232 & 0.1023 & -0.7371 & -0.1693 \\ (0.1083) & (0.0691) & (0.0601) & (0.084) \\ 0.1068 & -0.0796 & -0.0077 & -0.0646 \\ (0.0173) & (0.0110) & (0.0096) & (0.0013) \\ 0.0563 & -0.0441 & -0.0002 & -0.0374 \\ (0.0185) & (0.0118) & (0.0103) & (0.0145) \\ -0.0639 & 0.0028 & 0.0694 & -0.0303 \\ (0.0189) & (0.0121) & (0.0105) & (0.0148) \end{bmatrix}; \quad \widetilde{\Sigma} = \begin{bmatrix} 0.0655 & 0.0008 & 0.0009 & 0.0011 \\ 0.0008 & 0.0016 & 0.0014 & 0.0006 \\ 0.0009 & 0.0014 & 0.0019 & 0.0006 \\ 0.0011 & 0.0006 & 0.0006 & 0.0020 \end{bmatrix};$$

AIC = -11.4644
BIC = -11.2793

	$\overline{R}^2$	LM1	ARCH	WHITE	JB	RESET
s_t equation	0.3553	0.8259	3.61×10^{-7}	6.29×10^{-15}	9.48×10^{-164}	6.31×10^{-2}
f_t^1equation	0.0591	0.2124	0.8884	5.92×10^{-20}	0	8.48×10^{-5}
f_t^2equation	0.0239	0.4027	0.9587	3.37×10^{-5}	0	8.98×10^{-2}
s_t^T equation	0.0489	0.0835	0.7729	3.70×10^{-7}	0	1.24×10^{-5}

Notes: Tildes denote estimated values obtained using FIML estimation; figures in parentheses are estimated standard errors. d_{94} denotes the dummy variable for the 1994 procedural change. AIC and BIC are the Akaike Information Criterion and Bayesian Information Criterion respectively. $\overline{R}^2$ is the adjusted coefficient of determination, LM1 is an LM-type test statistic for residual serial correlation (Godfrey, 1988); ARCH is a test statistic for autoregressive conditional heteroskedasticity (Engle, 1982); WHITE is the White (1980) test for heteroskedasticity calculated without the cross products; JB is the Jarque-Bera test for normality of residuals; RESET is a RESET test calculated using a third order polynomial (Ramsey, 1969). For each of LM1, ARCH, WHITE, JB and RESET we only report p-values.

Table A2. 'Bottom-up' identification procedure

	$LR1$	$LR2$	Davies
	3411.49	18.847	236.511
p-value	0	4.43×10^{-3}	7.79×10^{-93}

Notes: $LR1$ is a test statistic of the null hypothesis of no regime dependent variance-covariance matrix (i.e. MSI(Q)-VECM(p) versus MSIH(Q)-VECM(p)). $LR2$ is a test statistic of the null hypothesis of no regime dependent intercept (i.e. MSH(Q)-VECM(p) versus MSIH(Q)-VECM(p)). Both tests are constructed as $2(\ln L^* - \ln L)$, where L^* and L represent the unconstrained and the constrained maximum likelihood respectively. Those tests are distributed as $\chi^2(g)$ where g is the number of restrictions imposed. *Davies* is the upper bound of the likelihood ratio test where the model is not identified under the null due to the nuisance parameters. In this case it tests the null hypothesis that the model with two regimes is equivalent to the model with three regimes. Figures in braces denote $p-$values, calculated as in Davies (1987), and $\{0\}$ indicates $p-$values below 10^{-500}, which are considered as virtually zero.

Table A3. MSIH(3)-VECM(1): estimation of Model II, $y = \begin{bmatrix} s_t & f_t^1 & f_t^2 \end{bmatrix}'$

Model: $\Delta y_t = \nu(z_t) + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + \alpha\beta' y_{t-1} + u_t$ with $u_t \sim N[0, \Sigma(z_t)]$

$$\widetilde{\Gamma}_1 = \begin{bmatrix} -0.0630 & -0.3787 & 0.0961 \\ (0.0170) & (0.1535) & (0.1335) \\ -0.0040 & 0.0013 & 0.0365 \\ (0.0022) & (0.0244) & (0.0203) \\ -0.0032 & 0.0472 & 0.0120 \\ (0.0028) & (0.0307) & (0.0263) \end{bmatrix}; \quad \widetilde{\alpha} = \begin{bmatrix} -0.9814 & 0.0168 & 0.0052 \\ (0.0636) & (0.0073) & (0.0092) \\ 0.4680 & -0.0045 & 0.0022 \\ (0.0448) & (0.0055) & (0.0071) \end{bmatrix};$$

$$\widetilde{v}(z_1) = \begin{bmatrix} -0.0073 \\ (0.0140) \\ -0.0039 \\ (0.0016) \\ -0.0029 \\ (0.0019) \end{bmatrix}; \quad \widetilde{v}(z_2) = \begin{bmatrix} -0.0221 \\ (0.0033) \\ -0.0003 \\ (0.0005) \\ -0.0007 \\ (0.0006) \end{bmatrix}; \quad \widetilde{v}(z_3) = \begin{bmatrix} 0.2403 \\ (0.0861) \\ -0.0034 \\ (0.0155) \\ -0.0061 \\ (0.0151) \end{bmatrix};$$

$$\widetilde{\sum}(z_1) = \begin{bmatrix} 0.0871 & & \\ 0.0015 & 0.0012 & \\ 0.0013 & 0.0013 & 0.0017 \end{bmatrix}; \quad \widetilde{\sum}(z_2) = \begin{bmatrix} 0.0063 & & \\ 0.0001 & 0.0001 & \\ 0.0007 & 0.0001 & 0.0003 \end{bmatrix};$$

$$\widetilde{\sum}(z_3) = \begin{bmatrix} 0.5979 & & \\ 0.0091 & 0.0209 & \\ 0.0089 & 0.0165 & 0.0198 \end{bmatrix};$$

$$\widetilde{\mathbf{P}} = \begin{bmatrix} 0.583 & 0.205 & 0.483 \\ 0.323 & 0.770 & 0.329 \\ 0.094 & 0.025 & 0.188 \end{bmatrix}; \quad \widetilde{\xi} = \begin{bmatrix} 0.355 \\ 0.585 \\ 0.058 \end{bmatrix};$$

$\rho(A) = 0.0556$; LR linearity test: 0; JB: 0.541; RESET: 0.492

Notes: Tildes denote estimated values obtained using the EM algorithm for maximum likelihood (Dempster, Laird and Rubin, 1977). Figures in parentheses are asymptotic standard errors. Symbols are defined as in equation (10). P and ξ denote the $Q \times Q$ transition matrix and the Q-dimensional ergodic probabilities vector respectively. $\rho(A)$ is the spectral radius of the matrix A calculated as in Karlsen (1990). It can be thought as a measure of stationarity of the MS-VECM. The LR linearity test is a Davies (1987)-type test checking the hypothesis that the true model is a linear VECM against the alternative of a MSIH-VECM. Its $p-$value is calculated as in Davies (1987). JB is the Jarque-Bera test for normality of the standardized residuals; RESET is a RESET test calculated using a third-order polynomial (Ramsey, 1969). For each of LR, JB and RESET we only report p-values; 0 indicates $p-$values below 10^{-500}, which are considered as virtually zero.

B The η-test for equal density forecast accuracy

This appendix briefly outlines the derivation of the η-test statistic for the null hypothesis of equal density forecast accuracy. Let $f(y)$, $g_1(y)$ and $g_2(y)$ be three probability density functions with distribution functions F, G_1 and G_2 respectively, and F, G_1 and G_2 are absolutely continuous with respect to the Lesbegue measure in $\Re^p$. Let $f(y)$ be the probability density function of the variable y_t over the period $t = 1, ..., T$, whereas $g_1(y)$ and $g_2(y)$ are the probability density functions implied by two competing forecasting models, say M_1 and M_2.

We are interested in testing the null hypothesis of equidistance of the probability densities $g_1(y)$ and $g_2(y)$ from $f(y)$, that is

$$H_0 : dist\left[f(y), g_1(y)\right] = dist\left[f(y), g_2(y)\right], \tag{B1}$$

where the operator *dist* denotes a generic measure of distance.

A conventional measure of global closeness between two functions is the integrated square difference (ISD) (e.g. see Pagan and Ullah, 1999):

$$ISD = \int \left[\phi(x) - \gamma(x)\right]^2 dx, \tag{B2}$$

where $\phi(\cdot)$ and $\gamma(\cdot)$ denote probability density functions; $ISD \geq 0$, and $ISD = 0$ only if $\phi(x) = \gamma(x)$. Using (B2) we can rewrite the null hypothesis H_0 in (B1) as follows:

$$\begin{aligned} H_0 \quad &: \quad \int \left[f(y) - g_1(y)\right]^2 dy = \int \left[f(y) - g_2(y)\right]^2 dy \\ &: \quad ISD_1 - ISD_2 = 0. \end{aligned} \tag{B3}$$

In (B3) the null hypothesis of equal density forecast accuracy of models M_1 and M_2 is written as the null hypothesis of equality of two integrated square differences or, equivalently, as the null hypothesis that the difference between two integrated square differences is zero.

Consider three series of realizations from $f(y)$, $g_1(y)$ and $g_2(y)$, say $\{y_t\}_{t=1}^{T}$, $\{\widehat{y}_{1t}\}_{t=1}^{T_1}$ and $\{\widehat{y}_{2t}\}_{t=1}^{T_2}$ respectively.[32] With observations $\{y_t\}_{t=1}^{T}$, $\{\widehat{y}_{1t}\}_{t=1}^{T}$ and $\{\widehat{y}_{2t}\}_{t=1}^{T}$ we can consistently estimate the unknown functions $f(y)$, $g_1(y)$ and $g_2(y)$ using kernel estimation, obtaining:

$$\widehat{f}(y) = \frac{1}{Th}\sum_{i=1}^{T} K\left(\frac{y_i - y}{h}\right) \tag{B4}$$

$$\widehat{g_1}(y) = \frac{1}{Th}\sum_{i=1}^{T} K\left(\frac{y_{1i} - y}{h}\right) \tag{B5}$$

$$\widehat{g_2}(y) = \frac{1}{Th}\sum_{i=1}^{T} K\left(\frac{y_{2i} - y}{h}\right) \tag{B6}$$

[32] For simplicity and for clarity of exposition, throughout this section, we consider the case where $T_1 = T_2 = T$, although the results derived below can be easily extended to the more general case where $T_1 \neq T_2$.

where $K(\cdot)$ is the kernel function and h is the smoothing parameter.[33] Using (B4)-(B6) we can then obtain a consistent estimate of the integrated square differences ISD_1 and ISD_2, $\widehat{ISD}_1$ and $\widehat{ISD}_2$. Define $d = \widehat{ISD}_1 - \widehat{ISD}_2$ as the estimated relative distance between the probability density functions. In order to test for the statistical significance of d, the next step is to calculate a confidence interval for d.

In the spirit of the analysis of Hall (1992), define $\left\{y_i^j\right\}_{i=1}^T$, $\left\{\widehat{y}_{1i}^j\right\}_{i=1}^T$, $\left\{\widehat{y}_{2i}^j\right\}_{i=1}^T$ as the $j-$th resample of the original data $\{y_t\}_{t=1}^T$, $\{\widehat{y}_{1t}\}_{t=1}^T$, $\{\widehat{y}_{2t}\}_{t=1}^T$, drawn randomly with replacement. From these resamples it is possible to obtain consistent bootstrap estimates of the density functions $\widehat{f}^j(y)$, $\widehat{g_1}^j(y)$, $\widehat{g_2}^j(y)$ and, consequently, of $d^j = \widehat{ISD}_1^j - \widehat{ISD}_2^j$.[34]

Consider a sample path $\left\{d^j\right\}_{j=1}^B$, where B is the number of bootstrap replications. Under general conditions[35], we have:

$$\sqrt{B}\left(\overline{d} - \mu\right) \xrightarrow{d} N\left(0, \sigma^2\right), \tag{B7}$$

where

$$\overline{d} = \frac{1}{B}\sum_{j=1}^{B} d^j = \frac{1}{B}\sum_{j=1}^{B}\left(\widehat{ISD}_1^j - \widehat{ISD}_2^j\right) \tag{B8}$$

is the average difference of the estimated relative distances over B bootstrap replications. Because in large samples the average difference $\overline{d}$ is approximately normally distributed with mean μ and variance σ^2/B, the large-sample statistic for testing the null hypothesis that models M_1 and M_2 have equal density forecast accuracy is:

$$\eta = \frac{\overline{d}}{\sqrt{\frac{\widehat{\sigma}^2}{B}}} \xrightarrow{d} N(0,1), \tag{B9}$$

where $\widehat{\sigma}^2$ is a consistent estimate of σ^2.[36] [37]

[33] In practice, for several econometric models $\widehat{g_1}(y)$ and $\widehat{g_2}(y)$ are known analytically. However, for more complex, nonlinear models we may not know the probability density function and therefore need to estimate it. In this paper we consider nonparametric estimation of density as a general procedure to implement the test statistic discussed below, but it should be clear that the test is directly applicable also when the probability density function is known analytically.

[34] Note that the data $\left\{y_i^j\right\}_{i=1}^T$ can only be resampled with replacement if it is independently and identically distributed. If there is dependence, the bootstrap procedure needs to be modified to accommodate dependence.

[35] See Kendall and Stuart (1976, Ch. 11).

[36] On the consistency of the bootstrap estimates of σ^2 in this context see Hall (1992) and Mammen (1992).

[37] Using Monte Carlo methods designed to investigate the size and power properties of this test statistic, Sarno and Valente argue that the η-test has satisfactory empirical size and power properties in finite sample in a variety of circumstances with a number of boostrap replications equal to 100 or so.

References

Andrews, D.W.K. (1991), "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817-858.

Ang, A. and Bekaert, G. (2002), "Regime Switches in Interest Rates," *Journal of Business and Economic Statistics*, 20, 163-182.

Brown, B.Y. and Mariano, R.S. (1984), "Residual Based Stochastic Predictors and Estimation in Nonlinear Models," *Econometrica*, 52, 321-343.

Brown, B.Y. and Mariano, R.S. (1989), "Predictors in Dynamic Nonlinear Models: Large Sample Behaviour," *Econometric Theory*, 5, 430-452.

Carlson, J.B., McIntire J.M., and Thomson, T.J. (1995), "Federal Funds Futures as an Indicator of Future Monetary Policy: A Primer," Federal Reserve Bank of Cleveland *Economic Review,* 31, 1, 20-30.

Clark, T.E. and McCracken, M.W. (2001), "Tests of Equal Forecast Accuracy and Encompassing for Nested Models," *Journal of Econometrics*, 105, 85-110.

Davies, R.B. (1987), "Hypothesis Testing when a Nuisance Parameter is Present Only Under the Alternative," *Biometrika*, 74, 33-43.

Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977), "Maximum Likelihood Estimation from Incomplete Data Via the EM Algorithm," *Journal of the Royal Statistical Society*, 39, Series B, 1-38.

Diebold, F.X. and Mariano, R.S. (1995), "Comparing Predictive Accuracy," *Journal of Business and Economics Statistics*, 13, 253-263.

Diebold, F.X., Gunther, T.A. and Tay, A.S. (1998), "Evaluating Density Forecasts with Applications to Financial Risk Management," *International Economic Review*, 39, pp. 863-883.

Diebold, F.X., Hahn, J., and Tay, A.S. (1999), "Multivariate Density Forecast Evaluation and Calibration in Financial Risk Management: High-Frequency Returns on Foreign Exchange," *Review of Economics and Statistics*, 81, 661-73.

Engle, R.E. and Granger, C.W.J. (1987), "Co-integration and equilibrium correction Representation, Estimation and Testing," *Econometrica*, 55, 251-276.

Franses, P.H. and van Dijk, D. (2000), *Non-linear Time Series Models in Empirical Finance*, Cambridge and New York: Cambridge University Press.

Furfine, C. (1999), "The Micro Structure of the Federal Funds Market," *Financial Markets, Institutions, and Instruments*, 8, pp. 24-44.

Gallant A.R (1987), *Nonlinear Statistical Models*, New York, NY: John Wiley.

Garcia, R. (1998), "Asymptotic Null Distribution of the Likelihood Ratio Test in Markov Switching Models," *International Economic Review*, 39, 763-788.

Garcia, R. and Perron, P. (1996), "An Analysis of the Real Interest Rate Under Regime Shift," *Review of Economics and Statistics*, 78, 111-125.

Granger, C.W.J. (1986), "Developments in the Study of Cointegrated Variables," *Oxford Bulletin of Economics and Statistics*, 48, 213-228.

Granger, C.W.I and Pesaran, M.H. (1999), "A Decision Theoretic Approach to Forecast Evaluation," in Chan, W.S., Lin, W.K. and Tong, H. (eds.), *Statistics and Finance: An Interface*, London: Imperial College Press.

Granger, C.W.J. and Terasvirta, T. (1993), *Modelling Nonlinear Economic Relationships*, Oxford, UK: Oxford University Press.

Gray, S.F. (1996), "Modeling the Conditional Distribution of Interest Rates as a Regime-Switching Process," *Journal of Financial Economics*, 42, 27-62

Hall, P. (1992), "Effect of Bias Estimation on Coverage Accuracy of Bootstrap Confidence Intervals for a Probability Density," *Annals of Statistics*, 20, pp. 675-694.

Hamilton, J.D. (1988), "Rational Expectations Econometric Analysis of Changes in Regime. An Investigation of the Term Structure of Interest Rates," *Journal of Economics Dynamics and Control*, 12, 385-423.

Hamilton, J.D. (1989), "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, 57, 357-384.

Hamilton, J.D. (1993), "Estimation, Inference and Forecasting of Time Series Subject to Changes in Regime," in Maddala, G.S., Rao, C.R. and Vinod, H.D. (eds.), *Handbook of Econometrics*, Vol. 4, Elsevier, Amsterdam.

Hamilton, J. and Jordá, O. (2001), "A Model for the Federal Funds Rate Target," University of California, San Diego, mimeo.

Hansen, B.E. (1992), "The Likelihood Ratio Test under Nonstandard Conditions: Testing the Markov Switching Model of GNP," *Journal of Applied Econometrics*, 7, S61-S82.

Hansen, B.E. (1996), "Erratum: The Likelihood Ratio Test under Nonstandard Conditions: Testing the Markov Switching Model of GNP," *Journal of Applied Econometrics*, 11, 195-198.

Johansen, S. (1998), "Statistical Analysis of Cointegrating Vectors," *Journal of Economic Dynamics and Control*, 12, 231-254.

Johansen, S. (1991), "Estimation and Hypothesis Testing of Cointegrating Vectors in Gaussian Vector Autoregressive Models," *Econometrica*, 59, 1551-1580.

Kendall, M.G. and Stuart, A. (1976), *The Advanced Theory of Statistics*, Vol. 1, 4th edition, London: Charles Griffin and Co.

Kilian, L. and Taylor, M.P. (2001), "Why Is It so Difficult to Beat the Random Walk Forecast of Exchange Rates?," *Journal of International Economics*, forthcoming.

Krolzig, H.-M. (1997), *Markov-Switching Vector Autoregressions*, New York, NY: Springer.

Krolzig, H.-M. (1999), "Statistical Analysis of Cointegrated VAR Processes with Markovian Regime Shifts," Nuffield College and Department of Economics, University of Oxford, mimeo.

Krueger, J.T., and Kuttner, K.N. (1996), "The Fed Funds Futures Rate as a Predictor of Federal Reserve Policy," *Journal of Futures Markets*, 16, 865-79

Kuttner, K.N. (2001), "Monetary Policy Surprises and Interest Rates: Evidence from the Fed Funds Futures Market," *Journal of Monetary Economics*, 47, pp. 523-544.

McCracken, M.W. (2000), "Robust Out-of-Sample Inference," *Journal of Econometrics*, 99, 195-223.

Muelendyke, A.-M. (1998), *US Monetary Policy and Financial Markets*, New York, NY: Federal Reserve Bank of New York.

Newey, W.K. and West, K.D. (1987), "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 3, 703-708.

Pagan, A. and Ullah, A. (1999), *Nonparametric Econometrics*, Cambridge, New York and Melbourne: Cambridge University Press.

Poole, W., Rasche, R.H. and Thornton, D.L. (2002), "Market Aniticpations of Monetary Policy Actions," *Federal Reserve Bank of St. Louis Review*, 4, 65-94.

Poskitt, D.S., and Chung, S.H. (1996), "Markov chain models, time series analysis and extreme value theory," *Advances in Applied Probability*, 28, 405-425

Ramsey, J.B. (1969), "Tests for Specification errors in Classical Linear Least Squares Regression Analysis," *Journal of the Royal Statistical Society B*, 31, 350-371.

Saikkonen, P. (1992), "Estimation and Testing of Cointegrated Systems by an Autoregressive Approximation," *Econometric Theory*, 8, 1-27.

Saikkonen, P. and Luukkonen, R. (1997), "Testing Cointegration in Infinite Order Vector Autoregressive Processes," *Journal of Econometrics*, 81, 93-126.

Sarno, L. and Valente, G. (2002), "Comparing the Accuracy of Density Forecasts from Competing Models," University of Warwick and Centre for Economic Policy Research, London, mimeo.

Stigum, M. (1990), *The Money Market*, revised edition, Homewood, IL: Irwin Publisher.

Stock, J.H. (1997), "Cointegration, Long-run Movements, and Long-Horizon Forecasting," in Kreps, D.M. and Wallis, K.F. (eds.), *Advances in Economics and Econometrics: Theory and Applications, Seventh World Congress*, Vol. III, Cambridge, UK: Cambridge University Press.

Stock, J.H. and Watson, M.W. (1988), "Testing for Common Trends," *Journal of the American Statistical Association*, 83, pp. 1097-1107.

Stock, J.H. and Watson, M.W. (1999), "Business Cycles Fluctuations in US Macroeconomic Time Series," in Taylor, J.B. and Woodford, M. (eds.), *Handbook of Macroeconomics*, Vol. 1A, pp. 1-64.

Taylor, J.B. (2001) "Expectations, Open Market Operations, and Changes in the Federal Funds Rate," Federal Reserve Bank of St. Louis *Review*, 83, pp. 33-48.

Thornton, D.L. (2001), "Identifying the Liquidity Effect at the Daily Frequency," Federal Reserve Bank of St. Louis *Review*, 83 pp. 59-78.

Thornton, D.L. and Wheelock, D.C. (2000), "A History of the Asymmetric Policy Directive," Federal Reserve Bank of St. Louis *Review*, 82, pp. 1-16.

Timmermann, A. (2000), "Density Forecasting in Economics and Finance: Editorial," *Journal of Forecasting*, 19, 231-234.

Vasicek, O. (1977), "An Equilibrium Characterisation of the Term Structure," *Journal of Financial Economics*, 5, pp. 177-188.

West, K.D. (1996), "Asymptotic Inference about Predictive Ability," *Econometrica*, 64, 1067-1084.

West, K.D. and McCracken, M. W. (1998), "Regression-based Tests of Predictive Ability," *International Economic Review*, 39, 817-840.

Figure 1: Kernel predictive density estimation

Federal Funds Rate Prediction - 1 and 2 weeks ahead

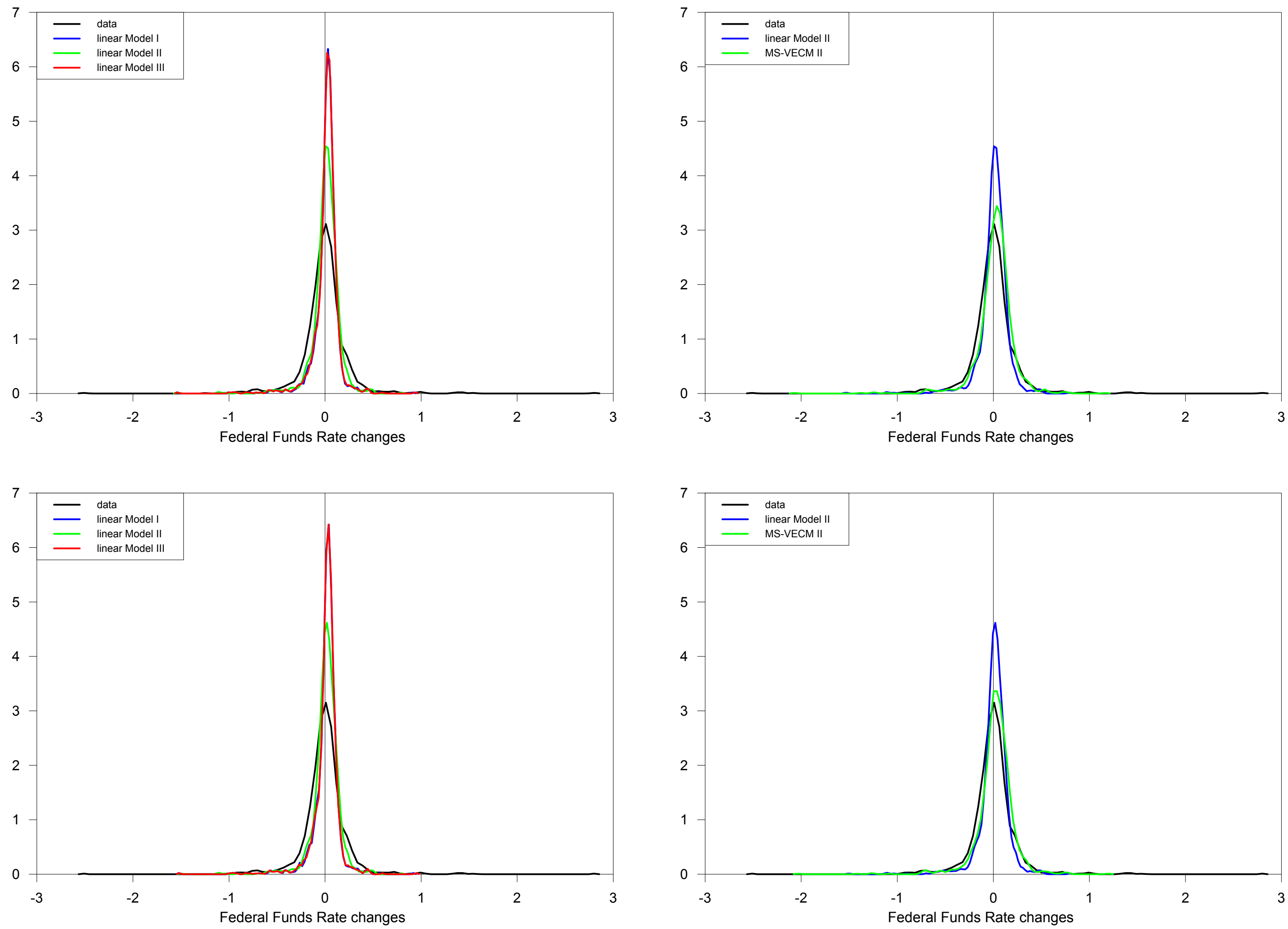

Figure 2: Kernel predictive density estimation

Federal Funds Rate Prediction - 4 and 8 weeks ahead