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 (CEPR)
c: Federal Reserve Bank of St. Louis

First version: August 2002 - Second revised version: May 2004

Abstract

We examine the forecasting performance of a range of time-series models of the daily US effective federal funds (FF) rate recently proposed in the literature. We find that: (i) most of the models and predictor variables considered produce satisfactory one-day-ahead forecasts of the FF rate; (ii) the best forecasting model is a simple univariate model where the future FF rate is forecast using the current difference between the FF rate and its target; (iii) combining the forecasts from various models generally yields modest improvements on the best performing model. These results have a natural interpretation and clear policy implications.

JEL classification: E43; E47.
Keywords: federal funds rate; forecasting; term structure; nonlinearity.

1. Introduction

The importance of the effective federal funds (FF) rate in US financial markets is unquestionable. The Federal Reserve (Fed) implements monetary policy by targeting the effective FF rate. The ability of market participants to predict the FF rate is important to modern analyses of monetary policy in that other interest rates are believed to be linked to the FF rate by the market expectation of monetary policy actions that directly affect the FF rate. It is therefore not surprising that a vast body of research has studied the behavior of the FF rate and proposed empirical models designed to explain it. One strand of this literature focuses on the FF rate using data at monthly or quarterly frequency to establish the extent to which arguments of interest to the Fed - such as inflation and the output gap - are sufficient to explain the variation in the FF rate (e.g. Taylor, 1993, 1999; Clarida, Gali and Gertler, 1998, 2000, and the references therein). A related literature investigates the impact of monetary policy shocks on key macroeconomic aggregates, again using low frequency data, identifying shocks to monetary policy using the FF rate in structural vector autoregressions (e.g. Christiano, Eichenbaum and Evans, 1999, and the references therein). Other studies focus on the high-frequency behavior of the FF rate, using data at the daily frequency. This frequency is appealing because each day the Trading Desk of the Federal Reserve Bank of New York (hereafter Desk) conducts open market operations designed to move the FF rate in the desired direction (e.g. Hamilton, 1996; Roberds, Runkle and Whiteman, 1996; Balduzzi, Bertola and Foresi, 1997; Taylor, 2001).[1]

The literature has suggested that several variables have predictive power for explaining FF rate movements: FF futures rates (Krueger and Kuttner, 1996), the FF rate target (Rudebusch, 1995; Taylor, 2001), and other interest rates linked to the FF rate via no-arbitrage conditions or the term structure of interest rates (e.g. Enders and Granger, 1998; Hansen and Seo, 2002; Sarno and Thornton, 2003; Clarida, Sarno, Taylor and Valente, 2004). A number of models, both univariate and multivariate, linear and nonlinear, have been proposed to capture the unknown process that drives FF rate movements. To date, however, there appears to be no consensus on what variables and models best characterize the behavior of the FF rate at the daily frequency. This paper attempts to fill this gap in the literature by examining a variety of univariate and multivariate, linear and nonlinear empirical models of the FF rate, largely taken directly from or inspired by previous research in this context. We estimate these models using daily data for the period from January 1 1990 through December 31 1996 and generate forecasts over the remaining four years of data. We also examine the potential to improve on the individual or 'primitive' models by using combinations of forecasts (see, *inter alia*, Diebold, 2001; Stock and Watson, 1999b, 2003; Swanson and Zeng, 2001).[2]

To anticipate our results, we find that, in general, most of the models and predictor variables considered produce satisfactory one-day-ahead forecasts of the FF rate. However, the best forecasting model is a very parsimonious univariate model where the one-day-ahead funds rate is forecast using the current difference between the funds rate and its target. This model can be thought of as a simple variant of the Desk's reaction function proposed by Taylor (2001). Combining the forecasts from various models provides generally modest improvements on this Desk's reaction function. We argue that these results have a natural interpretation and that they are in line with the growing empirical evidence suggesting that the Fed's policy is well described as a forward-looking interest rate rule.[3]

The remainder of the paper is set out as follows. In Section 2 we describe the empirical models of the daily FF rate considered in the paper. We also briefly discuss some econometric issues relevant to estimation of these models. In Section 3 we describe the data. Some preliminary data analysis and the in-sample estimation results are given in Section 4, while in Section 5 we present our forecasting results using both the primitive models and the combinations of forecasts, including evidence on point forecast accuracy and market timing ability. A final section concludes.

2. Empirical models of the federal funds rate

This section describes the empirical models of the daily US effective FF rate considered in this paper. Our aim is to model the daily behavior of the FF rate. However, we cannot rely on standard macroeconomic variables (such as inflation and output) that one may expect to drive monetary policy decisions and interest rate movements since these variables are not available at the daily frequency. Thus, the information set considered includes the FF rate target, other US interest rates to which the FF rate is likely to be linked via no-arbitrage conditions, and FF futures rates, in addition to lagged values of the FF rate.[4]

2.1 Univariate models

The first specification considered is a simple driftless random walk (RW) model:

$$\Delta i_t^{FF} = u_t, \tag{1}$$

where Δi_t^{FF} is the daily change in the effective FF rate. Although several researchers have concluded that the interest rate fails to follow a random walk (e.g. Shiller, Campbell and Schoenholtz, 1983; Campbell, 1987; Barrett, Slovin and Sushka, 1988; Lasser, 1992; Hamilton, 1996; Roberds, Runkle and Whiteman, 1996; Balduzzi, Bertola and Foresi, 1997; Lanne, 1999, 2000), most studies cannot reject the unit root hypothesis for interest rates (e.g. Stock and Watson, 1988, 1999a). Given the large empirical work suggesting that very persistent series with a root very close (if not equal) to unity are better approximated by $I(1)$ processes than by stationary ones (e.g. see Stock, 1997), it seems reasonable to consider a RW model for the FF rate as one of our specifications.[5]

A more general model of FF rate movements is a linear ARMA (p,q) model:

$$\Delta i_t^{FF} = \sum_{j=1}^{p} \gamma_j \Delta i_{t-j}^{FF} + \varepsilon_t + \sum_{s=1}^{q} \phi_s \varepsilon_{t-s}. \tag{2}$$

This model generalizes the RW model to account for higher-order autoregressive dependence and for moving-average serial correlation in the residual error.

The third specification considered is a variant of the model recently proposed by Taylor (2001):

$$\Delta i_t^{FF} = \xi \left(i_{t-1}^{FF} - i_{t-1}^{T} \right) + error\ term \tag{3}$$

where $\left(i_{t-1}^{FF} - i_{t-1}^{T} \right)$ is the lagged difference between the FF rate, i_{t-1}^{FF} and the FF rate target, i_{t-1}^{T}. Under this model, future daily changes in the FF rate are driven by current deviations of the FF rate from its target. Taylor (2001) suggests that the Desk strives to keep the FF rate close to its target level, which, since 1994, is publicly announced by the Federal Open Market Committee (FOMC). According to the literature on forward-looking interest rate rules (Clarida, Gali and Gertler, 1998, 2000), the FF rate target is set on the basis of considerations that may be parsimoniously summarized in expected inflation and output gap. Hence, equation (3) may be seen as the Desk's reaction function designed to minimize deviations of the FF rate from the target, which is set by the FOMC at a level believed to deliver the desired inflation and output objectives. Taylor argues that the adjustment to departures of the FF rate from its target is partial at daily frequency, implying that $-1 < \xi < 0$.[6] [7]

Another univariate model of the daily FF rate, first examined by Hamilton (1996), is the exponential generalized autoregressive conditional heteroskedasticity or EGARCH model. An EGARCH(p, q) for the daily change in the FF rate may be written as follows:

$$\Delta i_t^{FF} = \eta\left(i_{t-1}^{FF} - i_{t-3}^{FF}\right) + \sigma_t v_t$$

$$\log\left(\sigma_t^2\right) = \omega + \sum_{j=1}^{p} \vartheta_j \sigma_{t-j}^2 + \sum_{s=1}^{q}\left[\rho_s\left(\left|\frac{v_{t-s}}{\sigma_{t-s}}\right| - E\left|\frac{v_{t-s}}{\sigma_{t-s}}\right|\right) + \varsigma_s \frac{v_{t-s}}{\sigma_{t-s}}\right] \tag{4}$$

where $\left(i_{t-1}^{FF} - i_{t-3}^{FF}\right)$ is the cumulative change of the FF rate over the preceding two days, and v_t denotes independently and identically distributed (i.i.d.) innovations with zero mean and unit variance. The conditional variance σ_t^2 is modeled in the spirit of the analysis of Nelson (1990), where in order to take into account the asymmetric effect between future conditional variances and current FF rate changes, both sign and magnitude of the innovations are taken into account.

Another univariate nonlinear model we consider is a Markov-switching pth-order autoregressive model with M regimes, MS-AR(M, p) (Hamilton, 1988, 1989; Gray, 1996):

$$\Delta i_t^{FF} = \sum_{j=1}^{p} \varphi_j\left(z_t\right)\Delta i_{t-j}^{FF} + \sigma\left(z_t\right)\omega_t \qquad z_t = 1, 2, \ldots, M. \tag{5}$$

Model (5) allows for the autoregressive structure, $\sum_{j=1}^{p} \varphi_j\left(z_t\right)\Delta i_{t-j}^{FF}$, and the variance of the error term, $\sigma^2\left(z_t\right)$, to be shifting over time across regimes. In this model, the variance of the innovations ω_t is time-varying but, unlike for the EGARCH, the dynamics is governed by an unobservable variable z_t which is assumed to follow a first-order Markov chain. Markov-switching models have often been employed to model interest rates with some degree of success. Such models are a plausible alternative to models of the ARCH family designed to capture fat-tailed disturbances (e.g. Hamilton, 1988; Hamilton and Susmel, 1994; Ait-Sahalia, 1996; Gray, 1996; Bansal and Zhou, 2002; Dai, Singleton and Yang, 2003; Clarida, Sarno, Taylor and Valente, 2004).

2.2 Multivariate models

Among the multivariate models we consider, we include the Momentum-Threshold Autoregressive (M-TAR) model proposed by Enders and Granger (1998):

$$\Delta y_t = \sum_{j=1}^{p-1} \Lambda_j \Delta y_{t-j} + I_t \left[\alpha_1 \beta' y_{t-1} \right] + \left(1 - I_t\right) \left[\alpha_2 \beta' y_{t-1} \right] + error \tag{6}$$

$$I_t = \begin{matrix} 1 \text{ if } \beta' y_{t-1} \geq 0 \\ 0 \text{ if } \beta' y_{t-1} < 0 \end{matrix}$$

where $y_t = \left[i_t^{FF}, i_t^{TB} \right]$; i_t^{TB} is the 3-month Treasury Bill (TB) rate. This model explicitly takes into account the existence of asymmetries that may occur in the adjustment process along the short-end of the yield curve. Essentially, the adjustment parameters, α_1, α_2, differ depending on whether the slope of the short-end of the yield curve is positive or negative.

Another multivariate threshold model, applied to the FF rate by Hansen and Seo (2002), is the bivariate TAR or BTAR model:

$$\Delta y_t = I_t \left[\sum_{j=1}^{p-1} \Lambda_{1j} \Delta y_{t-j} + \alpha_1 \beta' y_{t-1} \right] + \left(1 - I_t\right) \left[\sum_{j=1}^{p-1} \Lambda_{2j} \Delta y_{t-j} + \alpha_2 \beta' y_{t-1} \right] + error \tag{7}$$

$$I_t = \begin{matrix} 1 \text{ if } \beta' y_{t-1} \geq \overline{k} \\ 0 \text{ if } \beta' y_{t-1} < \overline{k} \end{matrix}$$

where $y_t = \left[i_t^{FF}, i_t^{TB} \right]$. Although both the M-TAR model (6) and the BTAR model (7) belong to the family of threshold autoregressive models, they differ in several respects. First, in the BTAR model (7) the Heaviside indicator function I_t is equal to zero or unity according to whether the value of the cointegrating residual is smaller (or larger) than a threshold $\overline{k}$, which must be estimated. In contrast, in the M-TAR model (6) the threshold is assumed to be equal to zero. Second, in the BTAR model (7) the whole set of parameters, Λ_i, α_i, is shifting over time, while in the M-TAR model (6) only the speed of adjustment parameter, namely α_i (for $i = 1, 2$), is shifting over time.

The final model considered is a Markov-Switching Vector Error Correction Model or MS-VECM of the term structure of FF futures rates. This model is partly inspired by the recent work of Krueger and Kuttner (1996) and Kuttner (2001), where it is shown that FF futures rates are useful in predicting future changes in the FF rate. If the FF rate and FF futures rates are $I(1)$ variables, it is straightforward to demonstrate that the FF rate and the FF futures rates should cointegrate with a cointegrating vector $[1, -1]$.[8] In turn, via the Granger representation theorem (Granger, 1986), the joint dynamics of the FF rate and the FF futures rates can be described by a VECM:

$$\Delta y_t = \sum_{j=1}^{p-1} \Gamma_i (z_t) \Delta y_{t-j} + \Pi (z_t) y_{t-1} + \Sigma^{1/2} (z_t) \varepsilon_t \tag{8}$$

where $y_t = \left[i_t^{FF}, f_t^1, f_t^2 \right]$, and the long-run impact matrix $\Pi(z_t) = \alpha(z_t) \beta'$. The VECM in equation (8) has been generalized to a Markov-switching framework where the parameters can shift to take into account the evidence that interest rate changes are heteroskedastic and that their distribution is well approximated by a mixture of normal distributions (e.g. see Hamilton, 1988, 1996; Gray 1996; Bansal and Zhou, 2002; Dai, Singleton and Yang, 2003; Clarida, Sarno, Taylor and Valente, 2004).[9]

3. Data issues

The data set consists of daily observations on the effective FF rate $i_t^{FF,nc}$, the 3-month T-bill i_t^{TB}, the FF rate target i_t^T, and the FF futures rate f_t^m for $m = 1, 2$. The FF rate 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. Because reserve requirements are binding at the end of the reserve maintenance period, called settlement Wednesday, the FF rate tends to be more volatile on settlement Wednesdays.[10] The federal funds rate time series was adjusted in order to eliminate the effect of the increased volatility on settlement days, as done, for example, in Sarno and Thornton (2003).[11] This yielded the time series for the corrected FF rate, i_t^{FF}, which is the time series we employ in our empirical work.

f_t^m is the rate on a FF futures contract with maturity m, traded on the Chicago Board of Trade (CBT). Futures contracts are designed to hedge against or speculate on the FF rate. The CBT offers FF futures contracts at several maturities; however, the most active contracts are for the 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 i_t^{FF} and the 3-month T-Bill rate i_t^{TB} were obtained from the Federal Reserve Bank of St. Louis database, *Federal Reserve Economic Data* (FRED). The FF rate target, i_t^T was taken from Thornton and Wheelock (2000). The FF futures rates f_t^m were obtained from the CBT.

The sample period spans from January 1 1990 through December 31 2000, yielding a total of $2,869$ observations. This sample period was chosen for two reasons. First, while the Fed has never explicitly stated when it began targeting the FF rate in implementing monetary policy, an emerging consensus view is that the Fed has been explicitly targeting the FF rate since at least the late 1980s (e.g. see Meulendyke, 1998; Hamilton and Jordá, 2001; Poole, Rasche and Thornton, 2002). Second, while the FF futures market has existed since October 1988, trading activity in this market was initially small (Krueger and Kuttner, 1996, p. 867). 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.

Since we are interested in the predictive power of alternative time series models, we initially estimate each model over the period January 1 1990 through December 31 1996. Forecasts over the remaining four years of data are generated using a recursive forecasting procedure, described in Section 5.1.

4. Empirical results

4.1 Preliminary data analysis and unit root tests

Table 1 presents summary statistics for the series of interest, both in levels (Panel a) and first difference (Panel b). These summary statistics show that all rates - the (non-corrected and corrected)

FF rate, the TB rate, the FF futures rates, and the FF rate target - display similar values for the mean, variance, skewness and kurtosis. It is, however, clear that the correction for settlement Wednesdays discussed in the previous section make the mean of the FF rate closer to the mean of the FF rate target; indeed, after this correction the mean difference between the FF rate and the FF rate target, $\left(i_t^{FF} - i_t^T\right)$, is only 0.025 and statistically insignificantly different from zero when one takes into consideration its standard deviation. Also, 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 may be non-normal. This is confirmed by the rejections of the Jarque-Bera test for normality reported in the last row of Panels a-b in Table 1.[12]

4.2 Estimation results and in-sample performance

We estimate each model described in Section 2 over the sample period January 1 1990 and December 31 1996. Although the core of our empirical work relates to the out-of-sample performance of these models, this section provides details on the ability of each model to explain the FF rate in sample. The full estimation results for each model are given in Tables A1 to A3 in the Appendix.

Table A1 reports the estimation results for the univariate models of the FF rate. For model (2) we found an ARMA(2,1) specification to be satisfactory in that no serial correlation was left in the residuals (Table A1, column 2). Estimation of model (3), namely the Desk's reaction function, suggests very fast, albeit partial, adjustment of the FF rate toward the FF rate target (Table A1, column 3). The speed of reversion parameter ξ is consistent with over 70 percent of the daily difference between the FF rate and the target being dissipated the following day, which is in line with the evidence presented in Taylor (2001). The EGARCH model (4) was estimated following the specification procedure adopted in Hamilton (1996); see Table A1, column 4.[13]

The univariate Markov-switching model (5) was specified and estimated by employing the 'bottom-up' procedure suggested by Krolzig (1997, Ch. 6). This procedure is designed to detect Markovian shifts in order to select the most adequate characterization of an *M*-regime *p*-th order MS-AR for Δi_t^{FF}. The bottom-up procedure suggested in each case that two regimes were sufficient to characterize the dynamics of FF rate changes, confirming previous results reported in the literature on modeling short-term interest rates (e.g. Gray, 1996; Ang and Bekaert, 2002). Further, a specification which allows the autoregressive structure and the variance to shift over time (Markov-Switching-Autoregressive-Heteroskedastic-AR, or MSAH-AR) was selected since a regime-shifting intercept was found to be insignificant at conventional statistical levels (Table A1, column 5).

Turning to the multivariate models, the M-TAR (6) and the BTAR (7) were estimated following the procedures described in Enders and Granger (1998) and Hansen and Seo (2002) respectively (see Table A2 in the Appendix). In particular, both models (6) and (7) were estimated by imposing the cointegrating vector $\beta' = [1, -1]$. With respect to the BTAR model (7), the threshold parameter was estimated, as in Hansen and Seo (2002), using 300 gridpoints, and the trimming parameter was set to 0.05. Finally, the MS-VECM (8) was estimated according to the conventional procedure suggested by Krolzig (1997) and employed, for example, by Clarida, Sarno, Taylor and Valente (2003, 2004), designed to jointly select the appropriate lag length and the number of regimes characterizing the dynamics of the FF rate and the FF futures rates. The VARMA representations of the series suggested in each case that there are between two and three regimes. We adopted a specification which allows the whole set of parameters - i.e. intercept, autoregressive structure, cointegrating matrix and variance-covariance matrix - to shift over time (Markov-Switching-Intercept-Autoregressive-Heteroskedastic-VECM or MSIAH-VECM) with the number of regimes $M = 3$, which was found to adequately characterize the joint dynamics of the FF rate and FF futures rates - see Table A3 in the Appendix.

The estimation yields fairly plausible estimates of the coefficients for all the specifications considered. Further, for any of the regime-shifting models (5)-(8) the null hypothesis of linearity is rejected in all cases with very low *p*-values, suggesting that nonlinearities and asymmetries of the kind modeled here may be important ingredients for characterizing in sample the dynamics of the effective FF rate. Statistics measuring the in-sample performance of the estimated models (2)-(8) are given in Table 2, where the $\overline{R}^2$ and conventional information criteria (namely, AIC, BIC and HQ) are reported. The goodness of fit of the models is satisfactory and all the $\overline{R}^2$ are larger than 0.11, which is a satisfactory $\overline{R}^2$ if one considers that we are modeling a daily interest rate time series in first difference. Five models out of eight exhibit an $\overline{R}^2 > 0.20$ and two of them, namely Taylor's (2001) Desk reaction function and the MSIAH-VECM, display an $\overline{R}^2 > 0.30$. Inspection of the information criteria tells us that, although the MSIAH-VECM has the highest $\overline{R}^2$ recorded, this model may be overparameterized. In fact, within the group of multivariate models, the MSIAH-VECM is outperformed by the two competing TAR models and it is also outperformed by the Taylor's (2001) Desk reaction function, which displays the second highest $\overline{R}^2$ and the lowest information criteria within the set of competing models.

5. Out-of-sample forecasting results

5.1 Methodological issues

In order to evaluate the forecasting performance of the empirical models of the FF rate considered, dynamic out-of-sample forecasts of the FF rate were constructed using each of the models estimated in the previous section. In particular, we calculated one-step-ahead (one-day-ahead) forecasts over the period January 1 1997 and December 31 2000. The out-of-sample forecasts are constructed according to a recursive procedure that is conditional only upon information available up to the date of the forecasts and with successive re-estimation as the date on which forecasts are conditioned moves through the data set. [14]

We assess the forecasting performance of each of the eight individual models examined and then consider combinations of forecasts (models) using the forecast pooling approach proposed by Stock and Watson (1999b, 2003). For each time series we choose two separate periods: (i) a start-up period over which forecasts are produced using the eight individual competing models but not the pooling procedures; (ii) the simulated real-time forecast period over which recursive forecasts are produced using all individual models as well as the pooling procedures. Let T_0 be the date of the first observation used in this study (namely January 1 1990) and T_1 be the first observation for the forecast period (namely January 1 1997). Then the start-up period ends at $T_2 = T_1 + 261$ (until the end of

1997) and the forecast period goes from T_2 to T_3, where T_3 is the date of the final observation in our data set (December 31 2000). All the forecasting results reported in the following sub-sections refer to the simulated real-time forecast period T_2 to T_3 (inclusive).[15]

5.2 Forecasting results: 'primitive' models

In Table 3 we report the forecasting results obtained using the eight models estimated in Section 4, which we term 'primitive' models. In Panel a) of Table 3 we report the mean absolute error (MAE) and the root mean square error (RMSE) for each of the estimated models. Using the random walk (RW) model as a benchmark in assessing the relative forecasting performance of the primitive models, we then report the *p*-values of tests for the null hypothesis of equal point forecast accuracy based on both the MAE and the RMSE.[16]

We use the RW model as a benchmark since it is the benchmark used in the analysis of much research on the properties of the FF rate, cited in Section 2.1, and in particular the theoretical benchmark considered by Hamilton (1996). Moreover, the RW model is often used as a benchmark in forecasting studies on financial variables, such as, for example, exchange rates (Meese and Rogoff, 1983). An alternative benchmark we might have chosen is the ARMA model in equation (2) (e.g. Nelson, 1972). However, the exercise carried out in this paper is the first comprehensive out-of-sample forecast comparison of FF rate models. Out-of-sample forecast comparisons are popular in applied economic and finance largely because some landmark papers (e.g. Nelson, 1972; Meese and Rogoff, 1983) found that simple benchmarks do as well as theory-based models.

Our calculations appear to suggest that the Taylor (2001) Desk reaction function (3) exhibits the best out-of-sample performance: the MAE and the RMSE obtained for the Desk reaction function are the lowest obtained across all models. However, the *p*-values for the null of equal predictive accuracy, calculated by bootstrap[17], indicate that the null hypothesis is not rejected in each case. Hence, the differences in terms of MAEs and RMSEs reported in Panel a) of Table 3 are not statistically significant and do not enable us to discriminate among the models examined. Nevertheless, this result should be taken with caution as the non-rejection of the null of equal point forecast accuracy may be due to the low power of the relevant test statistic (e.g. see Kilian and Taylor, 2003).[18] Clark and West (2004) recently investigated out-of-sample mean squared prediction errors (MSPEs) to evaluate the null that a given series follows a zero-mean martingale difference against the alternative that it is linearly predictable. Despite the fact that under the null of no predictability the population MSPE of the null model is equal to the MSPE of the linear alternative, Clark and West show that the alternative model's sample MSPE is greater than the null's, which is what we find in our results in Panel a) of Table 3. Clark and West's simulations suggest that the test power is about 50 percent in their setup. This led us to consider additional tests.

Formal comparisons of the predicted and actual FF rate changes can be obtained in a variety of ways. We consider a set of tests for market timing ability of the competing models, including the 'hit' ratio (HR), calculated as the proportion of times the sign of the future FF rate change is correctly predicted over the whole forecast period, as well as the tests proposed by Henriksson and Merton (1981) and McCracken and West (1998) - hereafter HM and WM tests. The idea behind the HM test is that there is evidence of market timing if the sum of the estimated conditional probabilities of correct forecasts (that is the probability of correct forecast sign either when the FF rate is rising or falling) exceeds unity. The HM test statistic is given by:

$$HM = \frac{n_{11} - \frac{n_{01}n_{10}}{n}}{\sqrt{\frac{n_{01}n_{10}n_{20}n_{02}}{n^2(n-1)}}} \sim N(0,1) \tag{9}$$

where n_{11} is the number of correct forecasts when the FF rate is rising; n_{01}, n_{10} are the number of positive FF rate changes and forecasts of positive FF rate changes respectively, while n_{02} and n_{20} denote the number of negative FF rate changes and forecasts of negative FF rate changes respectively. The total number of evaluation periods is denoted by n. The HM test is asymptotically equivalent to a one-tailed test on the significance of the slope coefficient in the following regression:

$$I_{\left\{\Delta i_{t+1}^{FF} > 0\right\}} = \varrho_0^{HM} + \varrho_1^{HM} I_{\left\{\widetilde{\Delta i_{t+1}^{FF}} > 0\right\}} + error\ term \tag{10}$$

where $\Delta i_{t+1}^{FF}, \widetilde{\Delta i_{t+1}^{FF}}$ denote the realized and forecast first difference of the FF rate respectively, and I is the indicator function equal to unity when its argument is true and zero otherwise.

The other test employed is the 'efficiency test' introduced by Mincer and Zarnowitz (1969) and studied by McCracken and West (1998). This test extends the HM test to take into account not only the sign of the realized returns, but also their magnitude. This involves estimating the auxiliary regression:

$$\Delta i_{t+1} - \widetilde{\Delta i_{t+1}} = \phi_0^{MW} + \phi_1^{MW} \widetilde{\Delta i_{t+1}} + error\ term \qquad (11)$$

or alternatively

$$\Delta i_{t+1} = \varrho_0^{MW} + \varrho_1^{MW} \widetilde{\Delta i_{t+1}} + error\ term \qquad (12)$$

where $\varrho_0^{MW} = \phi_0^{MW}$ and $\varrho_1^{MW} = \left(1 + \phi_1^{MW}\right)$. As for the HM test, the null hypothesis of no market timing ability is that the slope coefficient ϱ_1^{HM} is equal to zero against the one-sided alternative that it is strictly greater than zero. Differently, for the WM test the null hypothesis of market timing ability is that the slope coefficient ϱ_1^{MW} is equal to unity (or alternatively the slope coefficient ϕ_1^{MW} is equal to zero). The results from calculating the hit ratio and executing the HM and WM tests are reported in Panel b) of Table 3. A fairly clear-cut result emerges from these tests. The analysis of the hit ratio statistics shows that most of the models (with the exception of the BTAR) exhibit evidence of market timing - i.e. the hit ratio is above 50 percent. This finding is corroborated by the results of the regression-based tests of market timing, which provide general evidence that the primitive models have market timing ability. However, the evidence of market timing ability is clearly stronger for the more parsimonious univariate models than for the more sophisticated nonlinear multivariate models, as evidenced by the fact that the *p*-values for the null of no market timing are drastically lower for the univariate models. This result is in line with the general finding that nonlinear models do not substantively outperform linear models in out-of-sample forecasting (e.g. see Clements and Krolzig, 1998; Stock and Watson, 1999b). It seems reasonable to conclude that, while all models examined display evidence of market timing ability, univariate models perform better than multivariate models. Moreover, the Taylor (2001) Desk reaction function exhibits the best performance in terms of hit ratios and market timing, displaying *p*-values much lower than the ones recorded by the alternative models.

Several caveats are in order. We have selected the best performing model on the basis of comparisons of the hit ratios and the *p*-values from carrying out market timing tests, essentially selecting the best performing model as the one for which the rejection of the null hypothesis of no market timing is strongest. We did not directly test a model against another in terms of market timing, however, since a test for equal market timing ability between competing models is not available to date.[19]

Also, one may be concerned about the importance of the settlement days in our forecasting exercise even if we corrected the FF rate time series to eliminate the effect of settlement Wednesdays prior to beginning the empirical work. Hence, we checked the robustness of a fraction of the results in Table 3 by comparing the forecasting performance of the simplest models - the RW model, the ARMA model, and Taylor Desk reaction function - when the final day of the settlement period is taken out of the calculations in constructing the forecasts errors used in the forecasting comparison of the models. The results, reported in Table C1 in Appendix C, suggest that excluding settlement days does not affect our results reported in Table 3 in that there is no qualitative difference and only small quantitative differences between the results in Table 3 and in Appendix C.

5.3 Combinations of forecasts

In this section we investigate whether there may be gains from combining forecasts from the primitive models. Following Stock and Watson (1999b, 2003), we employ five combinations of forecasts: simple combination forecasts; regression-based combination forecasts; median combination forecasts; discounted MSFE forecasts; and shrinkage forecasts. These methods differ in the way they use historical information to compute the combination forecast and in the extent to which the weight given to a primitive model's forecast is allowed to change over time.[20]

Let $\widetilde{\Delta i_{t+1,j}^{FF}}$ be the one-step-ahead forecast of the FF rate change at time t implied by the primitive model $j = 1,...,N$. Most of the combination forecasts are weighted averages of the primitive models' forecasts, i.e. $\widetilde{\Delta i_{t+1,c}^{FF}} = \sum_{j=1}^{N} w_{t,j} \widetilde{\Delta i_{t+1,j}^{FF}}$ where $w_{t,j}$ is the weight associated at time t with model j. In general, the weights $\{w_{t,j}\}$ depend on the historical performance of the individual forecast from model j. As discussed in Section 5.1, in order to obtain the first estimates of the weights $w_{t,j}$, we 'train' the individual models during the start-up period (i.e. 1997) and then we apply the following combination schemes.

The simple combination forecasts scheme computes the weights based on the relative forecasting performance, measured by the *MSFE*, of the primitive models. This relative performance is controlled by a parameter ω, which is set to zero in the simplest scheme that would place equal weight on all the forecasts. As ω increases, more importance is given to the model that has been performing better, in terms of MSFEs, in the past. In our empirical exercise we use values of $\omega = 0,1,5$.[21]

The regression-based combination forecasts scheme computes the weights applied to the combination forecast as the result of estimating a regression of Δi_{t+1}^{FF} on the one-step-ahead forecast of the FF rate change at time t implied by each primitive model and an intercept term (Granger and Ramanathan, 1984; Diebold, 2001).

If forecast errors are non-normal, then linear combinations are no longer optimal. The median combination forecasts scheme takes this into account and computes the combination forecasts as the median from a group of models. This scheme avoids placing too big a weight on forecasts that are strongly biased upwards or downwards for reasons such as parameter breaks or parameters which have been estimated by achieving local (rather than global) optima.

The discounted MSFE forecasts scheme (Diebold and Pauly, 1987; Diebold, 2001) computes the combination forecast as a weighted average of the primitive forecasts, where the weights depend inversely on the historical performance of each individual forecast according to a discount factor, d which we set equal to 0.95 and 0.90 in our calculations.

The shrinkage forecasts scheme computes the weights as an average of the recursive Granger-Ramanathan regression-based estimates of the weights and equal weighting (see Diebold and Pauly, 1990; Giacomini, 2002).

The results of the combination forecast exercises are reported in Table 4. Panel a) of Table 4 shows the mean absolute error (MAE) and the root mean square error (RMSE) for each combination forecast and the best performing primitive model - i.e. the Taylor (2001) Desk reaction function. The results suggest that the performance of the best primitive model is difficult to match even by using sophisticated pooling forecast techniques. The results of the test for equal point forecast accuracy indicate that we are not able to reject the null of equal predictive accuracy in each case. Hence, the differences in MAEs and RMSEs reported in Table 4 are not statistically significant. This conclusion is of course subject to the caveat that the test statistic for the null of equal point forecast accuracy may have low power in this context.

Tests for market timing ability are reported in Panel b) of Table 4. Using this metric, we find clear evidence of market timing for all of the forecasts examined. This finding is corroborated by the values of the hit ratios and the results of the tests of market timing (HM and WM tests). On the basis of the size of the *p*-values (smaller *p*-values indicate stronger market timing ability), the best performing primitive model displays stronger market timing ability than most of the combination forecasts. The exceptions, which appear to have stronger market timing ability than Taylor's (2001) Desk reaction function, are the discounted MSFE forecast scheme and, very marginally, the shrinkage forecasts scheme.

5.4 Interpreting the forecasting results

The forecasting results reported in this section provide several insights. First, we confirm that, in general, most of the models and predictor variables considered produce satisfactory one-day-ahead forecasts of the FF rate. Second, the best forecasting model is a simple and very parsimonious univariate model where the future FF rate is forecast using the current difference between the funds rate and its target. Third, combining the forecasts from various models may improve on the best performing model, but the improvements are generally modest.

These results may be seen as consistent with the growing empirical evidence uncovering that the Fed's policy is well described as targeting the FF rate according to a forward-looking version of Taylor's (1993) interest rate rule. For example, if the Fed implements monetary policy on the basis of expectations of future inflation and output gap, then the FF rate target set by the Fed will presumably contain information about future inflation and output gaps, which *a priori* one would expect to be important in predicting future interest rates (Clarida, Gali and Gertler, 1998, 2000).

Further, our results are also consistent with the Fed's description of its monetary policy operating procedure and its understanding by the economics profession. Indeed, there seems to be general agreement that the Fed has explicitly targeted the funds rate at least since the late 1980s and, therefore, throughout the sample period under investigation in this paper (see Meulendyke, 1998; Hamilton and Jordá, 2001). Also, since 1994 - hence throughout the forecast period examined - the Fed has announced target changes immediately upon making them. Before 1994, target changes were not announced: 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; Thornton, 2004). If one believes that the FF rate does in fact display reversion toward the FF rate target, then clearly this procedure would make it easier for the market to forecast the next-day FF rate by publicly announcing what the Fed's desired FF rate is.

Our results support the view that reversion to the target is a prominent feature of FF rate behavior during the sample examined and it is a crucial feature in forecasting out of sample the FF rate at the daily frequency. Although the mechanism linking the FF rate to the FF rate target is one of partial, not full, adjustment at the daily frequency, we found it very hard to improve on the simple Desk reaction function linking the FF rate and the target by using much more sophisticated multivariate or nonlinear models and alternative predictor variables. At the very least, our results suggest that the simple univariate Desk reaction function model suggested by Taylor (2001) is a very good first approximation to the FF rate behavior and represents a difficult benchmark to beat in one-day-ahead forecasting of the FF rate.

6. Conclusion

In this paper we reported what we believe to be the first broad-based analysis of a variety of empirical models of the daily FF rate and examined their performance in forecasting out-of-sample the one-day-ahead FF rate. Our research was inspired by encouraging results previously reported in the literature on the predictability of the FF rate using linear and nonlinear models and on the explanatory power of variables such as the FF futures rates, the FF rate target and other US interest rates to which the FF rate is likely to be linked via no-arbitrage conditions.

Using daily data over the period from January 1 1990 through December 31 1996, we confirmed that the predictor variables suggested by the literature have substantial explanatory power on the FF rate in sample and that accounting for nonlinearity in the unknown true data generating process governing the FF rate may yield satisfactory characterizations of the time-series properties of the FF rate. We then used a wide range of univariate and multivariate, linear and nonlinear models to forecast out of sample over the period January 1 1997 through to December 31 2000, using both conventional measures of point forecast accuracy based on mean absolute errors and root mean squared errors as well as hit ratios and market timing tests designed to evaluate the ability of the models to forecast both the direction and the magnitude of future FF rate changes. The forecasting results were interesting. Using conventional measures of point forecast accuracy we found that the reaction function proposed by Taylor (2001), where the gap between the FF rate and its target is used to predict the next-day FF rate, produces the lowest mean absolute errors and root mean square errors. However, general tests for equal point forecast accuracy did not enable us to distinguish among the competing models, possibly because of the low power of these tests in this context.

Using hit ratios and market timing tests, we found that the simple univariate reaction function emerges as the best performing model, forecasting correctly over 66 percent of the times the direction of the next-day FF rate and showing satisfactory market timing ability. Combining the forecasts from various models may improve on the best performing model, but the improvements are generally modest in size.

In turn, these results have a natural interpretation and may be seen as consistent with the growing empirical evidence suggesting that the Federal Reserve's policy may be characterized as a forward-looking interest rate rule. Our results support the view that reversion to the target is a key ingredient in models designed for characterizing in sample and forecasting out of sample the FF rate at the daily frequency. The simple univariate Desk reaction function suggested by Taylor (2001) is a very good first approximation to the FF rate behavior and represents a difficult benchmark to beat in one-day-ahead forecasting of the FF rate.

Table 1. Summary statistics

Panel a): Levels

	$i_t^{FF,nc}$	i_t^{FF}	i_t^{T}	$\left(i_t^{FF}-i_t^{T}\right)$	i_t^{TB}	f_t^1	f_t^2
Mean	5.265	5.249	5.224	0.025	5.056	5.277	5.304
Variance	1.895	1.893	1.832	0.059	1.578	1.771	1.741
Skewness	0.241	0.239	0.179	2.593	0.354	0.139	0.097
Kurtosis	3.080	3.065	3.010	31.642	3.197	3.001	2.976
JB	0	0	0	0	0	0	$13x10^{-4}$

Panel b): First differences

	$\Delta i_t^{FF,nc}$	Δi_t^{FF}	Δi_t^{T}	$\Delta\left(i_t^{FF}-i_t^{T}\right)$	Δi_t^{TB}	Δf_t^1	Δf_t^2
Mean	-0.0010	-0.0010	-0.0006	-0.0005	-0.0006	-0.0006	-0.0007
Variance	0.084	0.080	0.001	0.081	0.002	0.001	0.001
Skewness	0.825	0.755	2.294	0.821	0.109	2.275	0.091
Kurtosis	25.025	23.829	142.691	23.256	19.901	49.385	23.996
JB	0	0	0	0	0	0	0

Notes: $i_t^{FF,nc}, i_t^{FF}$, i_t^{T}, i_t^{TB}, f_t^1, f_t^2, denote the non-corrected effective federal funds rate, the corrected effective federal funds rate (adjusted to account for settlement days, as discussed in Section 3), the federal funds rate target, the 3-month T-Bill rate, the one-month and two-month federal funds futures rates respectively. The sample period spans from January 1 1990 through December 31 2000. JB denotes the Jarque-Bera test for the null hypothesis of normality, for which only *p*-values are reported; 0 denotes *p*-values lower than 10^{-4}. Δ is the first difference operator.

Table 2. In-sample performance

	$\overline{R^2}$	*AIC*	*SIC*	*HQ*
RW	–	–	–	–
ARMA(2,1)	0.282	0.077	0.079	0.078
Taylor (2001)	0.355	0.069	0.069	0.069
EGARCH(1,1)	0.115	0.099	0.111	0.103
MSAH-AR(2,1)	0.117	0.100	0.102	0.100
M-TAR	0.256	0.081	0.085	0.083
BTAR	0.244	0.083	0.088	0.085
MSIAH-VECM(3,1)	0.402	0.094	0.119	0.103

Notes: The definitions of the models are given in Section 2. All models are estimated over the sample period from 1 January 1990 to 31 December 1996. AIC, BIC and HQ denote the Akaike, Schwartz and Hannan-Quinn information criteria respectively.

Table 3. Out-of-sample performance: primitive models

a) Point forecast evaluation

	MAE	$p-value$	$RMSE$	$p-value$
RW	0.124	–	0.196	–
ARMA(2,1)	0.114	0.773	0.176	0.796
Taylor (2001)	0.105	0.497	0.167	0.672
M-TAR	0.121	0.931	0.189	0.943
EGARCH(1,1)	0.119	0.887	0.186	0.918
MSAH-AR(2,1)	0.126	0.967	0.194	0.986
BTAR	0.139	0.616	0.202	0.923
MSIH-VECM(3,1)	0.138	0.656	0.216	0.791

b) Regression-based test for market timing

	HR	HM	WM
RW	–	–	–
ARMA(2,1)	0.634	0	0.914 (0.09)
Taylor (2001)	0.664	0	0.980 (0.07)
M-TAR	0.577	0	0.823 (0.11)
EGARCH(1,1)	0.615	0	1.197 (0.13)
MSAH-AR(2,1)	0.508	0.122	0.671 (0.27)
BTAR	0.474	0.055	0.621 (0.22)
MSIH-VECM(3,1)	0.519	0.104	0.505 (0.29)

Notes: The definitions of the models are given in Section 2. The forecast period goes from January 1 1998 to December 31 2000. *Panel a)*: MAE and $RMSE$ denote the mean absolute error and the root mean square error respectively. $p-value$ is the *p*-value from executing test statistics for the null hypothesis that the model considered has equal point forecast accuracy as the random walk (RW), calculated by bootstrap using the procedure described in Appendix B. *Panel b)*: HR is the hit ratio calculated as the proportion of correctly predicted signs. HM and WM are the Henriksson and Merton (1981) and the 'efficiency test' calculated as in West and McCracken (1998), using the auxiliary regressions (10) and (12) respectively (see Section 5.2). HM and WM is calculated using the Newey-West (1987) autocorrelation- and heteroskedasticity-consistent covariance matrix. For the HM test statistics only *p*-values are reported; 0 denotes *p*-values lower than 10^{-4}. Values in parentheses are estimated standard errors.

Table 4. Out-of-sample performance: combinations of forecasts

a) Point forecast evaluation

	MAE	$p-value$	$RMSE$	$p-value$
Taylor (2001)	0.105	–	0.167	–
LCF, $\omega=0$	0.114	0.752	0.179	0.848
LCF, $\omega=1$	0.113	0.791	0.177	0.869
LCF, $\omega=5$	0.108	0.936	0.170	0.952
RCF	0.110	0.891	0.174	0.949
Median	0.115	0.700	0.179	0.828
DCF, $d=0.95$	0.102	0.906	0.159	0.915
DCF, $d=0.90$	0.102	0.906	0.159	0.915
SCF, $\kappa=0.25$	0.117	0.754	0.189	0.845
SCF, $\kappa=0.50$	0.116	0.768	0.188	0.851
SCF, $\kappa=1.00$	0.115	0.791	0.187	0.862

b) Regression-based test for market timing

	HR	HM	WM
Taylor (2001)	0.664	0	1.006 (0.06)
LCF, $\omega=0$	0.621	0	1.072 (0.09)
LCF, $\omega=1$	0.634	0	1.077 (0.09)
LCF, $\omega=5$	0.677	0	1.089 (0.09)
RCF	0.655	0	1.052 (0.07)
Median	0.606	0	1.093 (0.10)
DCF, $d=0.95$	0.782	0	1.005 (0.06)
DCF, $d=0.90$	0.783	0	1.004 (0.05)
SCF, $\kappa=0.25$	0.666	0	1.006 (0.06)
SCF, $\kappa=0.50$	0.669	0	1.005 (0.06)
SCF, $\kappa=1.00$	0.669	0	1.004 (0.06)

Notes: See Notes to Table 3. The full forecast period goes January from 1 1997 to December 31 2000, while the start-up period spans from January 1 1997 and December 31 1997. LCF denotes the combination of forecasts where the weights assigned to each primitive model's forecasts are calculated by using the simple combination forecasts scheme described in Section 5.3; and $\omega=0,1,5$ is a parameter controlling the relative weight of the best performing model within the panel. RCF denotes the combination of forecasts where the weights assigned to each primitive model's forecasts are calculated using the regression-based method of Granger and Ramanathan (1984). Median is the combination of forecasts calculated as the median of the forecasts from the primitive models. DCF denotes the combination of forecasts calculated according to the discounted forecasts scheme, and the discount factor $d=0.95,0.90$. SCF is the combination of forecasts where the weights assigned to each primitive model's forecasts are calculated as an average of the recursive ordinary-least-squares estimator and equal weighting, with a degree of shrinkage κ=0.25, 0.5, 1.0.

Appendix A. Further estimation results

Table A1. Univariate models

	ARMA(2,1)	Taylor (2001)	EGARCH(1,1)	MSAH-AR(2,1)
γ_1	0.298 (0.03)	–	–	–
γ_2	-0.100 (0.03)	–	–	–
ϕ_1	-0.865 (0.02)	–	–	–
ξ	–	-0.714 (0.02)	–	–
η	–	–	-0.176 (0.01)	–
ω	–	–	-2.027 (0.05)	–
ϑ_1	–	–	0.441 (0.02)	–
ρ_1	–	–	0.854 (0.02)	–
ς_1	–	–	-0.204 (0.02)	–
$\varphi_1(z=1)$	–	–	–	0.112 (0.03)
$\varphi_1(z=2)$	–	–	–	0.356 (0.05)
σ	0.277	0.263	0.308	–
$\sigma(z_t=1)$	–	–	–	0.101
$\sigma(z_t=2)$	–	–	–	0.611
ARCH	0	0	0.899	0.241
LR linearity	–	–	–	0.307

Notes: The definitions of the models are given in Section 2. This table reports estimates of the parameters of the models defined by equations (2)-(5) in Section 2. The sample period used for the estimation goes from 1 January 1990 to 31 December 1996. σ denotes the residual standard deviation. Values in parenthesis are standard errors calculated using the Newey-West autocorrelation and heteroskedasticity consistent covariance matrix. ARCH is the Lagrange multiplier (LM) test for autoregressive conditional heteroskedasticity (ARCH) in the residuals (Engle 1982). The LR linearity test is a Davies (1977, 1987) test for the null hypothesis that the true model is a linear AR(p) against the alternative of a MS-AR(M, p). Its *p*-value is calculated as in Davies (1977, 1987). For each of ARCH and LR linearity tests, we only report *p*-values; 0 denotes *p*-values lower than 10^{-4}.

Table A2. TAR models

Panel a) M-TAR

$$\Lambda_1 = \begin{bmatrix} -0.130 & -0.241 \\ (0.02) & (0.14) \\ -0.005 & 0.100 \\ (0.004) & (0.02) \end{bmatrix}; \Lambda_2 = \begin{bmatrix} -0.132 & -0.04 \\ (0.02) & (0.14) \\ -0.006 & -0.092 \\ (0.003) & (0.02) \end{bmatrix}; \quad \alpha_1 = \begin{bmatrix} -0.474 \\ (0.02) \\ 0.002 \\ (0.005) \end{bmatrix}; \alpha_2 = \begin{bmatrix} -0.221 \\ (0.03) \\ -0.003 \\ (0.004) \end{bmatrix}$$

LR1 : $\{0\}$

Panel b) BTAR

$$\Lambda_{11} = \begin{bmatrix} -0.06 & -0.777 \\ (0.01) & (0.25) \\ -0.004 & 0.155 \\ (0.005) & (0.03) \end{bmatrix}; \alpha_1 = \begin{bmatrix} -0.498 \\ (0.02) \\ 0.006 \\ (0.004) \end{bmatrix}; \quad \Lambda_{21} = \begin{bmatrix} -0.056 & -0.171 \\ (0.03) & (0.17) \\ -0.003 & 0.072 \\ (0.004) & (0.024) \end{bmatrix}; \alpha_2 = \begin{bmatrix} -0.244 \\ (0.04) \\ 0.014 \\ (0.005) \end{bmatrix};$$

$\bar{k} = 0.456$, $supLM^0$: $\{0.024\}$

Notes: The definitions of the models are given in Section 2. This table reports estimates of the parameters of the models defined by equations (6) and (7) in Section 2. The sample period used for the estimation is from 1 January 1990 to 31 December 1996. *Panel a)*: the M-TAR model (6) is estimated imposing $\beta' = [1, -1]$, as in Enders and Granger (1998, p. 310), and selecting the lag length according to the results of AIC, BIC and HQ criteria. $LR1$ is the likelihood ratio test for the null hypothesis that $\alpha_1 = \alpha_2$, for which we report the *p*-value; 0 denotes a *p*-value lower than 10^{-4}. *Panel b)*: the BTAR model is estimated by imposing $\beta' = [1, -1]$. The threshold parameter is estimated as in Hansen and Seo (2002) and computed with 300 gridpoints, whereas the trimming parameter is set to 0.05 (see, Andrews, 1993). Figures in parentheses are asymptotic standard errors. $supLM^0$ is the *p*-value from executing a test for threshold cointegration, computed by parametric bootstrap (Hansen and Seo, p. 299).

Table A3. Markov-switching VECM

MSIAH-VECM with three regimes and one lag

$$\tilde{\Gamma}_1(z=1)=\begin{bmatrix} 0.034 & 0.554 & -0.333 \\ (0.017) & (0.205) & (0.178) \\ -0.011 & -0.042 & 0.160 \\ (0.004) & (0.05) & (0.064) \\ -0.012 & -0.087 & 0.197 \\ (0.006) & (0.064) & (0.066) \end{bmatrix};\tilde{\Gamma}_1(z=2)=\begin{bmatrix} -0.05 & -0.978 & 0.528 \\ (0.019) & (0.215) & (0.182) \\ -0.001 & -0.055 & 0.037 \\ (0.001) & (0.023) & (0.016) \\ -0.0001 & 0.017 & -0.003 \\ (0.002) & (0.028) & (0.021) \end{bmatrix};$$

$$\tilde{\Gamma}_1(z=3)=\begin{bmatrix} 0.174 & -3.737 & 1.878 \\ (0.116) & (2.189) & (2.153) \\ 0.015 & 0.600 & -0.389 \\ (0.015) & (0.445) & (0.350) \\ 0.015 & 0.714 & -0.456 \\ (0.017) & (0.516) & (0.395) \end{bmatrix};$$

$$\tilde{\alpha}(z=1)=\begin{bmatrix} -0.106 & 0.042 \\ (0.054) & (0.041) \\ 0.018 & 0.001 \\ (0.015) & (0.011) \\ 0.010 & 0.005 \\ (0.019) & (0.015) \end{bmatrix};\tilde{\alpha}(z=2)=\begin{bmatrix} -0.134 & 0.0622 \\ (0.007) & (0.006) \\ 0.005 & 0.0008 \\ (0.005) & (0.005) \\ -0.0029 & 0.004 \\ (0.008) & (0.006) \end{bmatrix};$$

$$\tilde{\alpha}(z=3)=\begin{bmatrix} -0.160 & 0.061 \\ (0.036) & (0.030) \\ 0.350 & -0.345 \\ (0.056) & (0.055) \\ 0.209 & -0.218 \\ (0.055) & (0.050) \end{bmatrix};$$

(continued ...)

(...Table A3 continued)

$$\widetilde{\sum}(z=1)=\begin{bmatrix}0.0871 & & \\ 0.0015 & 0.0012 & \\ 0.0013 & 0.0013 & 0.0017\end{bmatrix}; \qquad \widetilde{\sum}(z=2)=\begin{bmatrix}0.0063 & & \\ 0.0001 & 0.0001 & \\ 0.0007 & 0.0001 & 0.0003\end{bmatrix};$$

$$\widetilde{\sum}(z=3)=\begin{bmatrix}0.5979 & & \\ 0.0091 & 0.0209 & \\ 0.0089 & 0.0165 & 0.0198\end{bmatrix};$$

$$\widetilde{\mathbf{P}}=\begin{bmatrix}0.601 & 0.173 & 0.254\\ 0.256 & 0.738 & 0.458\\ 0.143 & 0.089 & 0.288\end{bmatrix}$$

$\rho(A)=0.352;$ LR linearity test: 0

Notes: The model estimated is given by equation (8) in Section 2. Tildes denote estimated values obtained using the expectation maximization (EM) algorithm for maximum likelihood (Dempster, Laird and Rubin, 1977). Figures in parentheses are asymptotic standard errors. Symbols are defined as in equation (8). P denotes the $M \times M$ transition matrix. $\rho(A)$ 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, and for stationarity $|\rho(A)| < 0$ is required. The LR linearity test is a Davies-type test for the null hypothesis that the true model is a linear $\text{VECM}(p-1)$ against the alternative of a $\text{MSIAH}(M, p-1)$-VECM (Davies, 1977, 1987). For the LR linearity test, we only the report *p*-value; 0 indicates a $p-$value below 10^{-4}.

Appendix B. Bootstrap procedure for the *p*-value of the test of equal point forecast accuracy

The bootstrap algorithm used to determine the *p*-values of the test statistic of equal point forecast accuracy consists of the following steps:

1. Given the sequence of observations $\{x_t\}$ where $x_t = \left(\Delta i_t^{FF}, z_t\right)'$ and z_t denotes the explanatory variables, estimate each of the models (2)-(8) in Section 2 and construct the test statistic of interest, $\hat{\theta}$ (i.e. test statistic for the null hypothesis, H_0, of equal point forecast accuracy).

2. Postulate a data generating process (DGP) for each of the models (2)-(8) given in Section 2, where the FF rate is assumed to follow a driftless random walk under H_0 and the innovations are assumed to be i.i.d.

3. Based on the model specified in step 2), generate a sequence of pseudo observations $\{x_t^*\}$ of the same length as the original data series $\{x_t\}$ and discard the first 1,000 transient. The pseudo innovation terms are random and drawn with replacement from the set of observed residuals. Repeat this step 5,000 times.

4. For each of the 5,000 bootstrap replications $\{x_t^*\}$, estimate each of the models (2)-(8) and construct the test statistic of interest, $\hat{\theta}^*$.

5. Use the empirical distribution of the 5,000 replications of the bootstrap test statistic, $\hat{\theta}^*$ to determine the *p*-value of the test statistic $\hat{\theta}$.

Appendix C. Robustness results

Table C1. Out-of-sample performance of three primitive models without using the forecasts of settlement days

a) Point forecast evaluation

	MAE	$p-value$	$RMSE$	$p-value$
RW	0.123	–	0.197	–
ARMA(2,1)	0.112	0.751	0.174	0.771
Taylor (2001)	0.103	0.503	0.165	0.648

b) Regression-based test for market timing

	HR	HM	WM
RW	–	–	–
ARMA(2,1)	0.596	0	1.023 (0.09)
Taylor (2001)	0.618	0	1.073 (0.08)

Notes: The definitions of the models are given in Section 2. The forecast period goes from January 1 1998 to December 31 2000. *Panel a)*: MAE and $RMSE$ denote the mean absolute error and the root mean square error respectively. $p-value$ are *p*-values from executing test statistics for the null hypothesis that the model considered has equal point forecast accuracy as the random walk (RW), calculated by bootstrap using the procedure described in Appendix B. *Panel b)*: HR is the hit ratio calculated as the proportion of correctly predicted signs. HM and WM are the Henriksson and Merton (1981) and the 'efficiency test' as in West and McCracken (1998) calculated using the auxiliary regressions (10) and (12) respectively (see Section 5.2). HM and WM is calculated using the Newey-West (1987) autocorrelation and heteroskedasticity consistent covariance matrix. For the HM test statistics only *p*-values are reported; 0 denotes *p*-values lower than 10^{-4}. Values in parentheses are estimated standard errors.

References

Ait-Sahalia, Yacine (1996). "Testing Continuous-Time Models of the Spot Interest Rate," Review of Financial Studies, 9, 385-426.

Andrews, Donald W.K. (1993). "Tests for Parameter Instability and Structural Change with Unknown Change Point," Econometrica, 61, 821-856.

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

Balduzzi, Pierluigi, Giuseppe Bertola and Silverio Foresi (1997). "A Model of Target Changes and the Term Structure of Interest Rates," Journal of Monetary Economics, 39, 223-249.

Bansal, Ravi and Hao Zhou (2002). "Term Structure of Interest Rates with Regime Shifts," Journal of Finance, 57, 1997-2043.

Bates, J.M. and Clive W.J. Granger (1969). "The Combination of Forecasts," Operations Research Quarterly, 20, 451–468.

Berret, W. Brian, Myron B. Slovin, and Marie E. Sushka, (1988). "Reserve Regulation and Recourse as a Source of Risk Premia in the Federal Fund Market," Journal of Banking and Finance, 12, 575-584.

Campbell, John Y. (1987). "Money Announcements, the Demand for Bank Reserves, and the Behavior of the Federal Funds Rate within the Statement Week," Journal of Money, Credit and Banking, 19, 56-67.

Clarida, Richard H., Jordi Gali, and Mark Gertler (1998). "Monetary Policy Rules in Practice: Some International Evidence," European Economic Review, 42, 1033-1067.

Clarida, Richard H., Jordi Gali, and Mark Gertler (2000). "Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory," Quarterly Journal of Economics, 115, 147-180.

Clarida, Richard H., Lucio Sarno, Mark P. Taylor, and Giorgio Valente (2003). "The Out-of-Sample Success of Term Structure Models as Exchange Rate Predictors: A Step Beyond," Journal of International Economics, 60, 61-83.

Clarida, Richard H., Lucio Sarno, Mark P. Taylor, and Giorgio Valente (2004). "The Role of Asymmetries and Regime Shifts in the Term Structure of Interest Rates," Journal of Business, forthcoming.

Clements, Michael P. and Hans-Martin Krolzig (1998). "A Comparison of the Forecast Performance of Markov-Switching and Threshold Autoregressive Models of US GNP," Econometrics Journal, 1, C47-75.

Christiano, Lawrence J., Martin Eichembaum, and Charles L. Evans (1999). "Monetary Policy Shocks. What Have We Learned at the End?". In Handbook of Macroeconomics, vol. 1A, edited by Taylor, John B. and Michael Woodford. North-Holland Elsevier Science, 65-148.

Cook, Timothy and Thomas Hahn (1989). "The Effect of Changes in the Federal Funds Rate Target on Market Interest Rates in the 1970s," Journal of Monetary Economics, 24, 331-351.

Cumby, Robert E. and David M. Modest (1987). "Testing for Market Timing Ability, A Framework for Forecast Evaluation," Journal of Financial Economics, 19, 169-189.

Dai, Qiang, Kenneth J. Singleton, and Wei Yang (2003). "Are Regime Shifts Priced in the U.S. Treasury Market?", Stanford University and New York University, mimeo.

Davies, Robert B.(1977). "Hypothesis Testing when a Nuisance Parameter is Present Only Under the Alternative," Biometrika, 64, 247-254.

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

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

Deutsch, M., Clive W.J. Granger, and Timo Terasvirta (1994). “The Combination of Forecasts Using Changing Weights,” International Journal of Forecasting, 10, 47-57.

Diebold, Francis X. (2001). Elements of Forecasting, South-Western College Publishing.

Diebold, Francis X. and Lutz Kilian (2000). “Unit Root Tests Are Useful for Selecting Forecasting Models,” Journal of Business and Economics Statistics, 18, 265-273.

Diebold, F.X. and Roberto S. Mariano (1995). “Comparing Predictive Accuracy,” Journal of Business and Economics Statistics, 13, 253-263.

Diebold, F.X. and P. Pauly (1987). “Structural Change and the Combination of Forecasts,” Journal of Forecasting, 6, 21-40.

Diebold, F.X. and P. Pauly (1990). “The Use of Prior Information in Forecast Combination,” International Journal of Forecasting, 6, 503-508.

Elliott, Graham and Allan Timmermann (2002a). “Optimal Forecasts Combinations Under General Loss Functions and Forecast Error Distributions,” University of California San Diego, mimeo.

Elliott, Graham and Allan Timmermann (2002b). “Optimal Forecasts Combination Under Regime Switching,” University of California San Diego, mimeo.

Enders, Walter and Clive W.J. Granger (1998). “Unit-Root Tests and Asymmetric Adjustment with an Example Using the Term Structure of Interest Rates,” Journal of Business and Economic Statistics, 16, 304-311.

Engle, Robert E. (1982). “Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of U.K. Inflation,” Econometrica, 50, 987-1008.

Engle, Robert E. and Clive W.J. Granger (1987). “Co-integration and equilibrium correction Representation, Estimation and Testing,” Econometrica, 55, 251-276.

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

Fuller, Wayne A. (1976). Introduction to Statistical Time Series, 2nd ed., New York: Wiley.

Giacomini, Raffaella (2002). “Tests of Conditional Predictive Ability,” University of California San Diego, mimeo.

Gilbert, Scott (2001). “Sampling Schemes and Tests of Regression Models,” Southern Illinois University at Carbondale, mimeo.

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

Granger, Clive W.J. and R. Ramanathan (1984) “Improved Methods of Combining Forecasts,” Journal of Forecasting, 3, 197-204.

Granger, Clive W.J. and Timo Terasvirta (1993). Modeling Nonlinear Economic Relationships, Oxford, UK: Oxford University Press.

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

Hall, Anthony D., Heather M. Anderson and Clive W.J. Granger (1992). “A Cointegration Analysis of Treasury Bill Yields,” Review of Economics and Statistics, 74, 116-126.

Hall, Anthony D., Vance Martin, and Adrian Pagan (1996). "Modelling the Term Structure." In Handbook of Statistics, 14, edited by G.S. Maddala and Rao, C.R.. North Holland Elsevier Science. 91-118.

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

Hamilton, James D. and Raul Susmel (1994). "Autoregressive Conditional Heteroskedasticity and Changes in Regime," Journal of Econometrics, 64, 307-333.

Hamilton, James D. (1996). "The Daily Market for Federal Funds," Journal of Political Economy, 104, 26-56.

Hamilton, James and Oscar Jordá (2001). "A Model for the Federal Funds Rate Target," Journal of Political Economy, 5, 1135-1167.

Hansen, Bruce E. (2003). "Structural Changes in the Cointegrated Vector Autoregressive Model," Journal of Econometrics, 114, 261-295.

Hansen, Bruce E. and Byeongseon Seo (2002). "Testing for Two-Regime Threshold Cointegration in Vector Error-Correction Models," Journal of Econometrics, 110, 293-318.

Henriksson, Roy D. and Robert C. Merton (1981). "On Market Timing and Investment Performance II. Statistical Procedures for Evaluating Forecasting Skills," Journal of Business, 54, 513-533.

Inoue, Atsushi and Lutz Kilian (2003a). "On the Selection of Forecasting Models," University of Michigan, mimeo.

Inoue, Atsushi and Lutz Kilian (2003b). "In-Sample or Out-of-Sample Tests of Predictability: Which One Should We Use?", University of Michigan, mimeo.

Karlsen, Hans A. (1990). A Class of Non-Linear Time Series Models, University of Bergen, unpublished PhD dissertation.

Kendall, Maurice G. and Alan Stuart (1976). The Advanced Theory of Statistics, Vol. 1-2, 4th edition, London: Charles Griffin and Co.

Kilian, Lutz (1999). "Exchange Rates and Monetary Fundamentals: What Do We Learn from Long-Horizon Regressions?", Journal of Applied Econometrics, 14, 491-510.

Kilian, Lutz and Mark P. Taylor (2003). "Why Is It so Difficult to Beat the Random Walk Forecast of Exchange Rates?", Journal of International Economics, 60, 85-107.

Krolzig, Hans-Martin (1997). Markov-Switching Vector Autoregressions, New York, NY: Springer.

Krueger, Joel T. and Kenneth N. Kuttner (1996). "The Fed Funds Futures Rate as a Predictor of Federal Reserve Policy," Journal of Futures Markets, 16, 865-879.

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

Lanne, Markku J. (1999). "Near Unit Roots and the Predictive Power of Yield Spreads for Changes in Long-Term Interest Rates," Review of Economics and Statistics, 81, 393-398.

Lanne, Markku J. (2000). "Near Unit Roots, Cointegration, and the Term Structure of Interests Rates," Journal of Applied Econometrics, 15, 513-529.

Lasser, Dennis J. (1992). "The Effect of Contemporaneous Reserve Accounting on the Market for Federal Funds," Journal of Banking and Finance, 16,1047-1056.

MacKinnon, James G. (1991). "Critical Values for Cointegration Tests." In Readings in Cointegration, edited by Engle, Robert E. and Clive W.J. Granger. Oxford: Oxford University Press, Ch. 13.

Mark, Nelson C. (1995). "Exchange Rates and Fundamentals: Evidence on Long-Horizon Predictability," American Economic Review, 85, 201-218.

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

Meese, Richard A. and Kenneth Rogoff (1983). "Empirical Exchange Rate Model of the Seventies," Journal of International Economics, 14, 3-24.

Mincer, J. and V. Zarnowitz (1969). "The Evaluation of Economic Forecast." In Economic Forecasts and Expectations, edited by Mincer, J. National Bureau for Economic Research: New York.

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

Nelson, Charles R. (1972). "The Prediction Performance of the FRB-MIT-PENN Model of the U.S. Economy," American Economic Review, 62, 902-917.

Nelson, Daniel B. (1990). "Conditional Heteroskedasticity in Asset Returns: A New Approach," Econometrica, 59, 347-370.

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

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

Rahbek, Anders and Neil Shephard (2002). "Inference and Ergodicitiy in the Autoregressive Conditional Root Model," Nuffield College, University of Oxford, mimeo.

Roberds, William, David Runkle and Charles H. Whitman (1996). "A Daily View of Yield Spreads and Short-Term Interest Rate Movements," Journal of Money, Credit, and Banking, 28, 34-53.

Rudebusch, Glenn D.(1995). "Federal Reserve Interest Rate Targeting, Rational Expectations, and the Term Structure," Journal of Monetary Economics, 35, 245-274.

Sarno, Lucio and Daniel L. Thornton (2003). "The Dynamic Relationship Between the Federal Funds Rate and the Treasury Bill Rate: An Empirical Investigation," Journal of Banking and Finance, 27, 1079-1110.

Shiller, Robert J., John Y. Campbell and Kermit L. Schoenholtz (1983). "Forward Rates and Future Policy: Interpreting the Term Structure of Interest Rates," Brookings Papers on Economic Activity, 0, 173-217.

Stock, James H. (1997). "Cointegration, Long-run Movements, and Long-Horizon Forecasting." In Advances in Economics and Econometrics: Theory and Applications, Seventh World Congress, Vol. III, edited by Kreps, David M. and Kenneth F. Wallis, Cambridge, UK: Cambridge University Press.

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

Stock, James H. and Mark W. Watson (1999a). "Business Cycles Fluctuations in US Macroeconomic Time Series." In Handbook of Macroeconomics, Vol. 1A, edited by Taylor, John B. and Michael Woodford. North Holland Elsevier Science. 1-64.

Stock, James H. and Mark W. Watson (1999b). "A Comparison of Linear and Nonlinear Univariate Models for Forecasting Macroeconomic Time Series." In Cointegration, Causality and Forecasting: A Festschrift for Clive W.J. Granger, edited by Engle, Robert and Halbert L. White. Oxford: Oxford University Press. 1-44.

Stock, James H. and Mark W. Watson (2003). "Combination Forecasts of Output Growth in a Seven-Country Data Set," Harvard University and Princeton University, mimeo.

Swanson, Norman R. and Tian Zeng (2001). "Choosing Among Competing Econometric Forecasts: Regression-Based Forecast Combination Using Model Selection," Journal of Forecasting, 20, 425-440.

Taylor, John B. (1993). "Discretion Versus Policy Rules in Practice," Carnegie Rochester Conference Series on Public Policy, 39, 195-214.

Taylor, John B. (1999). Monetary Policy Rules, National Bureau of Economic Research Conference Report series. Chicago and London: University of Chicago Press.

Taylor, John B. (2001). "Expectations, Open Market Operations, and Changes in the Federal Funds Rate," Federal Reserve Bank of St. Louis Review, 83, 33-47.

Thornton, Daniel L. (2004). "The Fed and Short-Term Rates: Is It open Market Operations, Open Mouth Operations or Interest Rate Smoothing?", Journal of Banking and Finance, 28, 475-498.

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

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

West, Kenneth D. and Dongchul Cho (1995). "The Predictive Ability of Several Models of Exchange Rate Volatility," Journal of Econometrics, 69, 367-391.

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

White, Halbert L. (2000)."A Reality Check for Data Snooping," Econometrica, 68, 1097-1126.

---

[*] *Acknowledgments:* This paper was partly written while Lucio Sarno was a Visiting Scholar at the Federal Reserve Bank of St. Louis and the International Monetary Fund, whose hospitality is gratefully acknowledged. We are grateful to Ken West (editor) and three anonymous referees for constructive comments and suggestions which led to a substantially improved paper. We are also thankful to Mike Dueker, Oscar Jordá, Lutz Kilian, Chris Neely and participants to the 2003 Royal Economic Society Annual Conference at the University of Warwick for useful conversations or comments on previous drafts. The views expressed do not necessarily reflect the views of the Federal Reserve Bank of St. Louis, the Board of Governors of the Federal Reserve System or the International Monetary Fund.

[1] See also Shiller, Campbell and Schoenholtz (1983), Campbell (1987), Barrett, Slovin and Sushka (1988), Lasser (1992), Krueger and Kuttner (1996).

[2] Although throughout the paper we follow much previous research in referring to our forecasting exercise as an 'out-of-sample' (OOS) exercise, a more precise definition would be 'pseudo out-of-sample' (POOS) or 'simulated out-of-sample' (SOOS) exercise. This procedure is of course different from a true OOS. The POOS is based on fitting each candidate model on the first part of the sample period available and constructing out-of-sample forecasts according to a recursive procedure that is conditional only upon information available up to the date of the forecasts and with successive re-estimation as the date on which forecasts are conditioned moves through the data set. See Inoue and Kilian (2003a,b) and Stock and Watson (2003).

[3] We stick to one-step-ahead forecasting in this paper. In a previous version of the paper we also analyzed forecasts up to eight weeks ahead, although an exercise of this kind would ideally require a

model of how the FF rate target changes. Given our focus on one-day-ahead forecasts, changes in the target affect very few of the observations and hence the models considered here are not intended to explore issues related to explaining or forecasting target changes.

[4] In each of the models of FF rate changes described below, we do not allow for a constant term since, for all models, in preliminary estimations the constant was always found to be very small in size and statistically insignificantly different from zero at conventional significance levels.

[5] Further useful references on this issue include Lanne (1999) and Diebold and Kilian (2000).

[6] It should be noted that the original formalization in Taylor (2001) considers as the dependent variable the change in the Federal supply balances which, in turn, are adjusted by the Fed in order to induce the desired change in the FF rate. However, we use the FF rate change as the left-hand-side variable in equation (3). Note also that this model is similar to the model used by Rudebusch (1995), where the FF rate is modeled as the target plus an error term which is allowed to follow an autoregressive process of order one. For a model of the FF rate target, see Hamilton and Jordá (2001), who propose a autoregressive conditional hazard model which is specifically designed for discrete-valued time series, such as the target.

[7] We also investigated whether the adjustment coefficient ξ has changed over the sample by estimating equation (3) recursively. In principle, it seems possible that this coefficient has varied over time if there are many unmodeled factors that are important in determining this coefficient. However, recursive estimation revealed little, if any, time variation in the adjustment coefficient ξ, suggesting that this parameter has been fairly stable over the sample period examined. An alternative specification we did not consider in this paper involves modeling the spread between the FF rate and FF rate target.

[8] Empirically, the [1,-1] restriction is often rejected in studies of the Expectations Hypothesis of the term structure of interest rates using US data (e.g., Hall, Anderson and Granger, 1992; Hall, Martin and Pagan, 1996). On exception is Hansen (2003), who argues that the rejection may be due to non-constant parameters.

[9] An alternative model to consider in future work is that of Rahbek and Shephard (2002); in that model, the error correction is either active or inactive, where the probability of the error correction being active increases with the deviation from equilibrium. Also, in addition to the nonlinear multivariate models described in this section, we considered in each case their linear multivariate counterparts. However, these linear multivariate models performed worse in forecasting than their nonlinear generalizations and, in fact, they performed worse than most of the univariate models considered in this paper. Therefore, to conserve space, we do not report the empirical results from multivariate linear models in this version of the paper, although the relevant empirical results are available upon request.

[10] Since February 1984 the reserve maintenance period has been two weeks for all institutions. Before 1984 it was one week for most large institutions. For a more detailed discussion of the Federal Reserve's reserve requirements and the microstructure of the federal funds market, see, for example, Taylor (2001).

[11] Precisely, the adjusted time series for the FF rate is the ordinary least squares residual from the regression of the FF rate on a dummy variable that equals zero on non-settlement days and unity on settlement days.

[12] Before proceeding to estimating the models described in Section 2, we carried out preliminary unit root tests for each of i_t^{FF}, i_t^{TB}, f_t^1, f_t^2 and i_t^T (available upon request). Using standard augmented Dickey-Fuller (ADF) tests we are unable to reject the unit root null hypothesis for any of the rates. However, differencing the interest rate time series appears to induce stationarity, consistent with much previous research (e.g., Stock and Watson, 1988, 1999a). Although it is unlikely that interests rates are strictly I(1), it seems useful to treat them as such for forecasting purposes, also because the least squares estimator of a root near unity is downwards biased and, therefore, it may be better to impose unity rather than estimating it.

[13] However, differently from Hamilton (1996), the distribution of the innovations is assumed to be Gaussian for simplicity. Under this specification $E\left|\frac{v_{t-s}}{\sigma_{t-s}}\right| = \sqrt{\frac{2}{\pi}} \quad \forall s$ in equation (4).

[14] While forecasting with linear models is straightforward, it is well known in the literature that forecasting with nonlinear models is more difficult because of the need to condition on the distribution of future exogenous shocks whose conditional expectation may be zero in a linear framework but not in a nonlinear framework. However, given that we compute one-step-ahead forecasts, the procedure often suggested in the literature that involves implementing numerical integration using Monte Carlo methods is not required as the one-step-ahead forecasts can be calculated analytically for our models (see, *inter alia*, Granger and Teräsvirta, 1993; Franses and van Dijk, 2000).

[15] As a robustness check, we investigated shorter sample periods of one year from 1998 to 1999, and from 1999 to 2000; the results (not reported in the paper to conserve space but available upon request) were qualitatively identical, suggesting that our results for the full forecast horizon are unlikely to have been generated by a particular subperiod.

[16] This would be the test proposed by Diebold and Mariano (1995) under certain circumstances, and specifically if we had models with no estimated parameters.

[17] The finite-sample distribution of the test statistic for equal point forecast accuracy may deviate from normality; this problem is particularly severe in the presence of estimated parameters (see West 1996; West and McCracken 1998; McCracken 2000; Gilbert, 2001). The *p*-values reported in this paper were calculated by bootstrap (see Mark, 1995; Kilian, 1999), and a description of the procedure employed is given in Appendix B. Note that in this bootstrap procedure we do not allow for conditional heteroskedasticity in the data generating process or for serial correlation in the denominator of the test statistics. West (1996) proves that allowance for serial correlation may not be necessary in RMSE comparisons, but it may be necessary for MAE comparisons. Hence, we judge the RMSE results as more informative, although the results reported below for RMSE and MAE comparisons are qualitatively identical.

[18] Although Kilian and Taylor (2003) document the low power of this class of test statistics in finite sample, admittedly they are working with a smaller sample than the one used in this paper, and they point out that if their sample size were to double the test would have good power.

[19] White (2000) considers a similar comparison problem of forecasting models, examining also directional forecasts. However, his setup does not account for the non-differential nature of the loss function and for forecasts that are based on estimated parameters.

[20] A further extension of our work which we leave for future research might involve considering nonlinear combination schemes (Granger and Terasvirta, 1993; Deutsch, Granger and Terasvirta, 1994; Elliott and Timmermann, 2002a,b).

[21] Note that $\omega = 1$ corresponds to Bates and Granger's (1969) suggestion for the choice of optimal linear weighting scheme when individual forecasts are uncorrelated.