THE OUT-OF-SAMPLE SUCCESS OF TERM STRUCTURE MODELS AS EXCHANGE RATE PREDICTORS: A STEP BEYOND

Richard Clarida, Lucio Sarno, Mark P Taylor and Giorgio Valente

INTERNATIONAL MACROECONOMICS
THE OUT-OF-SAMPLE SUCCESS OF TERM STRUCTURE MODELS AS EXCHANGE RATE PREDICTORS: A STEP BEYOND

Richard Clarida, Columbia University
Lucio Sarno, Warwick Business School and CEPR
Mark P Taylor, Warwick University and CEPR
Giorgio Valente, Warwick Business School

Discussion Paper No. 3281
March 2002

Centre for Economic Policy Research
90–98 Goswell Rd, London EC1V 7RR, UK
Tel: (44 20) 7878 2900, Fax: (44 20) 7878 2999
Email: cepr@cepr.org, Website: www.cepr.org

This Discussion Paper is issued under the auspices of the Centre’s research programme in INTERNATIONAL MACROECONOMICS. Any opinions expressed here are those of the author(s) and not those of the Centre for Economic Policy Research. Research disseminated by CEPR may include views on policy, but the Centre itself takes no institutional policy positions.

The Centre for Economic Policy Research was established in 1983 as a private educational charity, to promote independent analysis and public discussion of open economies and the relations among them. It is pluralist and non-partisan, bringing economic research to bear on the analysis of medium- and long-run policy questions. Institutional (core) finance for the Centre has been provided through major grants from the Economic and Social Research Council, under which an ESRC Resource Centre operates within CEPR; the Esmée Fairbairn Charitable Trust; and the Bank of England. These organizations do not give prior review to the Centre’s publications, nor do they necessarily endorse the views expressed therein.

These Discussion Papers often represent preliminary or incomplete work, circulated to encourage discussion and comment. Citation and use of such a paper should take account of its provisional character.

Copyright: Richard Clarida, Lucio Sarno, Mark P Taylor and Giorgio Valente
ABSTRACT

The Out-of-Sample Success of Term Structure Models as Exchange Rate Predictors: A Step Beyond*

A large literature suggests that standard exchange rate models cannot outperform a random walk forecast and that the forward rate is not an optimal predictor of the spot rate. There is evidence, however, that the term structure of forward premia contains valuable information for forecasting future spot exchange rates and that exchange rate dynamics display non-linearities. This Paper proposes a term-structure forecasting model of exchange rates based on a regime-switching vector equilibrium correction model which is novel in this context. Our model significantly outperforms both a random walk and, to a lesser extent, a linear term-structure vector equilibrium correction model for four major dollar rates across a range of horizons.

JEL Classification: F31
Keywords: forecasting, foreign exchange, Markov switching, non-linearity and term structure

Richard Clarida  
Head of Economics Department  
International Affairs Building  
Room 1020  
Columbia University  
420 West 118th Street  
New York, NY 10027  
Tel: (1 212) 280 4648  
Fax: (1 212) 854 8059  
Email: rhc2@columbia.edu

Lucio Sarno  
Finance Group  
Warwick Business School  
University of Warwick  
Coventry  
CV4 7AL  
Tel: (44 24) 7652 8219  
Fax: (44 24) 7657 2871  
Email: lucio.sarno@warwick.ac.uk

For further Discussion Papers by this author see:  

For further Discussion Papers by this author see:  
* Forthcoming in the *Journal of International Economics*. This Paper was partly written while Sarno and Taylor were visiting the Federal Reserve Bank of St. Louis and the International Monetary Fund respectively. Sarno and Taylor acknowledge financial support from the Economic and Social Research Council (No. L138251044) and the Leverhulme Trust respectively. The authors are indebted for their constructive comments to the discussants of the Paper, Bruce Hansen and Philip Lane, as well as to Frank Diebold, Charles Engel, Martin Evans, Lutz Kilian, Ken West, Jonathan Wright and other participants to the conference on ‘Empirical Exchange Rate Models’, in addition to participants in seminars at the Federal Reserve Bank of St. Louis, Washington University, the Bank of England, Queen Mary College, the University of Manchester and the University of Warwick. The authors alone are responsible for any errors and for the views expressed in the Paper.

Submitted 5th March 2002
NON-TECHNICAL SUMMARY

The Meese and Rogoff (1983a,b) studies marked a watershed in empirical exchange rate economics. In particular, their robust finding that standard empirical exchange rate models could not outperform a simple no-change (random walk) forecast was at the time seen as devastating. Even with the benefit of twenty years of hindsight, moreover, the random walk remains the standard comparator for exchange rate forecasting and models which consistently and significantly outperform a naive random walk are still elusive (e.g. see Mark, 1995).

A parallel finding in the exchange rate literature, also dating from the early 1980s, was that the forward rate is not an optimal predictor of the future spot exchange rate (see eg. Hansen and Hodrick, 1980; Frankel, 1980; Bilson, 1981), or equivalently that the forward premium is not an optimal predictor of the rate of depreciation, as the efficient markets hypothesis, at least in its risk-neutral formulation, would suggest. Attempts to locate the source of this failure of the risk-neutral efficient markets hypothesis either in the presence of stable, significant and plausible risk premia, or in some sense in the failure of rational expectations when applied to the foreign exchange market as a whole, have also met with mixed and very limited success. Thus, from the early 1980s onward, exchange rate forecasting in general became increasingly seen as a hazardous occupation, and this remains largely the case.

A ray of hope in an otherwise murky environment was, however, provided by Clarida and Taylor (1997), who argued that the failure of the forward rate optimally to predict the future spot rate did not necessarily imply that forward rates did not contain valuable information for forecasting future spot exchange rates. Clarida and Taylor develop what they term an ‘agnostic’ framework for linking spot rate and forward rate movements without assuming anything at all specific about risk premia or expectations formation except that departures from the risk-neutral efficient markets hypothesis (RNEMH) drive at most a stationary wedge between forward and expected future spot rates. This is sufficient to establish the existence of a linear vector equilibrium correction model (VECM) for spot and forward exchange rates. Using this framework, Clarida and Taylor are able to extract sufficient information from the term structure of forward premia to outperform the random walk forecast – and a range of alternative forecasts – for several exchange rates in out-of-sample forecasting. Indeed, at the one-year forecasting horizon, their improvement over the naive random walk is of the order of 40% in terms of root mean square errors.

Alongside the work on exchange rate forecasting, another strand of the literature has developed in which increasingly strong evidence of non-
linearities of one sort or another in exchange rate movements has been reported. One element of this, dating at least to Booth and Glassman (1987), has been the mounting evidence that the conditional distribution of nominal exchange rate changes is well described by a mixture of normal distributions and that, consequently, a Markov switching model may be a logical characterization of exchange rate behaviour. Although Markov-switching models fit nominal exchange rate data very well, in general they do not produce superior forecasts to a random walk or the forward rate on the basis of conventional forecasting criteria. Overall, in fact, the literature on non-linear modelling of exchange rates has produced models that fit satisfactorily and forecast well in sample but that in general fail to beat simple random walk models or linear specifications in out-of-sample forecasting (eg. Engel, 1994).

In the present Paper, we investigate whether allowing for non-linearities in the underlying data-generating process for the term structure yields superior exchange rate forecasts. This is done through estimating a fairly general three-regime Markov-switching vector equilibrium correction model (MS-VECM) for spot rates and the term structure of forward rates which is essentially based on an extension of Markovian regime shifts to a nonstationary framework, for which the underlying econometric theory has recently been developed. Given the evidence of significant regime-switching behavior in exchange rate movements discussed above, this seems a natural way to extend the Clarida-Taylor analysis, even though this involves estimating and forecasting from a sophisticated multivariate non-linear model. Indeed, to the best of the present authors’ knowledge, the research reported in this Paper represents the first application of Markov-switching in a multivariate cointegrated framework to exchange rate modelling and forecasting.

Using weekly data on spot and forward dollar exchange rates for the G5 countries over the period January 1979 through to December 1995, we found strong evidence of the presence of non-linearities in the term structure, which appeared to be modelled well by 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 1998. The forecasting results were impressive. The MS-VECM forecasts were found to be strongly superior to the random walk forecasts at a range of forecasting horizons up to 52 weeks ahead, using standard forecasting accuracy criteria and on the basis of standard tests of significance. Moreover, the non-linear VECM also outperformed, in general, a linear VECM for spot and forward rates in out-of-sample forecasting of the spot rate, although the magnitude of the gain from using a Markov-switching VECM relative to a linear VECM is rather small in magnitude at short horizon.
1 Introduction

The Meese and Rogoff (1983a,b) studies marked a watershed in empirical exchange rate economics. In particular, their robust finding that standard empirical exchange rate models could not outperform a simple random walk forecast was at the time seen as devastating.\footnote{Even with the benefit of twenty years of hindsight, moreover, the random walk remains the standard comparator for exchange rate forecasting and models which consistently and significantly outperform a naive random walk are still elusive (e.g. see Mark, 1995).} A parallel finding in the exchange rate literature, also dating from the early 1980s, was that the forward rate is not an optimal predictor of the future spot exchange rate (see e.g. Hansen and Hodrick, 1980; Frankel, 1980; Bilson, 1981), or equivalently that the forward premium is not an optimal predictor of the rate of depreciation, as the efficient markets hypothesis, at least in its risk-neutral formulation, would suggest (see Frankel and Rose, 1995; Taylor, 1995). Attempts to locate the source of this failure of the risk-neutral efficient markets hypothesis either in the presence of stable, significant and plausible risk premia, or in some sense in the failure of rational expectations when applied to the foreign exchange market as a whole, have also met with mixed and very limited success (see Lewis, 1995; Taylor, 1995). Thus, from the early 1980s onward, exchange rate forecasting in general became increasingly to be seen as a hazardous occupation, and this remains largely the case.

A ray of hope in an otherwise murky environment was, however, provided by Clarida and Taylor (1997), who argued that the failure of the forward rate optimally to predict the future spot rate did not necessarily imply that forward rates did not contain valuable information for forecasting future spot exchange rates. Clarida and Taylor develop what they term an ‘agnostic’ framework for linking spot rate and forward rate movements without assuming anything at all specific about risk premia or expectations formation except that departures from the risk-neutral efficient markets hypothesis (RNEMH) drive at most a stationary wedge between forward and expected future spot rates. This is sufficient to establish the existence of a linear vector equilibrium correction model (VECM) for spot and forward exchange rates. Using this framework, Clarida and Taylor are able to extract sufficient information from the term structure of forward premia to outperform the random walk forecast - and a range of alternative forecasts - for several exchange rates in
out-of-sample forecasting. Indeed, at the one-year forecasting horizon, their improvement over the naive random walk is of the order of 40 percent in terms of root mean square errors.

Alongside the work on exchange rate forecasting, another strand of the literature has developed in which increasingly strong evidence of nonlinearities of one sort or another in exchange rate movements has been reported. One element of this, dating at least to Booth and Glassman (1987), has been the mounting evidence that the conditional distribution of nominal exchange rate changes is well described by a mixture of normal distributions and that, consequently, a Markov switching model may be a logical characterization of exchange rate behavior (e.g. see Engel and Hamilton, 1990; LeBaron, 1992; Engel, 1994; Engel and Hakkio, 1996; Engel and Kim, 1999). However, although Markov-switching models fit nominal exchange rate data very well, in general they do not produce superior forecasts to a random walk or the forward rate on the basis of conventional forecasting criteria (see e.g. Engel, 1994). An exception in this context is the study by Engel and Hamilton (1990), who apply the Markov-switching model developed by Hamilton (1988, 1989) to dollar exchange rate data and show that the model generates better forecasts than a random walk. In the light of the subsequent literature, however, these forecasting results appear to be somewhat fragile. Overall, in fact, the literature on nonlinear modelling of exchange rates has produced models that fit satisfactorily and forecast well in sample but that in general fail to beat simple random walk models or linear specifications in out-of-sample forecasting (e.g. see Diebold and Nason, 1990; Engel, 1994; Meese and Rose, 1990, 1991).

In the present paper, we investigate whether allowing for nonlinearities in the underlying data-generating process for the term structure yields superior exchange rate forecasts. This is done through estimating a fairly general three-regime Markov-switching vector equilibrium correction model (MS-VECM) for spot rates and the term structure of forward rates which is essentially based on an extension of Markovian regime shifts to a nonstationary framework, for which the underlying econometric theory has recently been developed. Given the evidence of significant regime-switching behavior in exchange rate movements discussed above, this seems a natural way to extend the Clarida-Taylor analysis, even though this involves estimating and forecasting from a sophisticated multivariate nonlinear model. Indeed, to the best of the present authors’ knowl-
edge, the research reported in this paper represents the first application of Markov-switching in a multivariate cointegrated framework to exchange rate modelling and forecasting.

Using weekly data since 1979 for four major dollar exchange rates, we are able to replicate the Clarida-Taylor forecasting results in a linear VECM framework. However, we also show that conventional linear VECMs reveal significant residual nonlinearity and are easily rejected when tested against the alternative of an MS-VECM. Finally, we show that allowing for nonlinearities, using an MS-VECM, results in forecasts which are superior to random walk forecasts and, to a lesser extent, to the linear VECM exchange rate forecasts. We thus confirm that the information contained in the term structure of forward rates is indeed valuable for forecasting spot exchange rates but that statistically significant improvements can be made over linear forecasting models by allowing for nonlinearities.

The remainder of the paper is set out as follows. In Section 2 we discuss a general framework for linking spot and forward rates, as a motivation for our multivariate modelling. In Section 3 we briefly set out the econometrics of Markov-switching multivariate models as applied to non-stationary processes and cointegrated systems. In the following section we report our empirical testing, estimation and forecasting results. A final section concludes.

2 The information in the term structure of forward exchange premia

Let $s_t$ and $f_t^{h(k)}$ be, respectively, the spot exchange rate and the $h(k)$-period forward exchange rate, each at time $t$. It is now well documented that nominal exchange rates between the currencies of the major industrialized economies are well described by unit root processes. We can therefore write the spot exchange rate as the sum of two components:

$$s_t = m_t + q_t,$$

where $m_t$ is a unit-root process evolving as a random walk with drift, and $q_t$ is a stationary process having mean zero and a finite variance (Beveridge and Nelson, 1981; Stock and Watson, 1988). If agents are risk-neutral and the market is efficient in the sense that exchange rates fully reflect all
information in a given information set $\Omega_t$ (so that, in effect, the market conforms to the rational expectations hypothesis) then the forward exchange rate $f_t^{h(k)}$ should predict the $h(k)$-period ahead future value of the spot exchange rate optimally given $\Omega_t$. This is the essence of the risk-neutral efficient market hypothesis (RNEMH). There now exists a large literature rejecting the RNEMH, although it is unclear whether rejection is due to a failure of the assumptions of risk neutrality or of rational expectations or of both (e.g. see Taylor, 1995; Sarno and Taylor, 2002, Chapter 2).

Following Clarida and Taylor (1997), we may in general define departures from the RNEMH, due either to the presence of risk premia or to a failure of rational expectations, or both, as follows:

$$\gamma_t \equiv f_t^{h(k)} - E_t (s_{t+h(k)}|\Omega_t),$$  \hspace{1cm} (2)

where $E(\cdot|\Omega_t)$ denotes the mathematical expectation conditional on $\Omega_t$. From (1) and (2) we can obtain:

$$f_t^{h(k)} = \gamma_t + h(k)\theta + E_t (q_{t+h(k)}|\Omega_t) + m_t,$$  \hspace{1cm} (3)

where $\theta$ is the drift of the random walk process $m_t$. Subtracting (1) from (3), we achieve an expression for the forward premium at time $t$:

$$f_t^{h(k)} - s_t = \gamma_t + h(k)\theta + E_t (q_{t+h(k)} - q_t|\Omega_t).$$  \hspace{1cm} (4)

Equation (4) says that if the departure from the RNMEH $\gamma_t$ is stationary, given $q_t \sim I(0)$, the forward premium $\left(f_t^{h(k)} - s_t\right)$ must also be stationary. This implies that forward and spot rates exhibit a common stochastic trend and are cointegrated with cointegrating vector $[1, -1]$. Moreover, since this is true for any $h(k)$, if we consider the vector of forward rates of tenor $h(1)$ to $h(m)$ periods, together with the current spot rate, $[s_t, f_t^{h(1)}, f_t^{h(2)}, f_t^{h(3)}, \ldots, f_t^{h(m)}]'$, then this must be cointegrated with $m$ unique cointegrating vectors, each given by a row of the matrix $[-I, I_m]$, where $I_m$ is an $m$-dimensional identity matrix and $I$ is an $m$-dimensional column vector of ones. Further, by the Granger Representation Theorem (Engle and Granger, 1987) the same
set of forward and spot rates must possess a VECM representation in which the term structure of forward premia plays the part of the equilibrium errors. Clarida and Taylor (1997) exploit this framework and use exactly a linear VECM representation to demonstrate that a large amount of information may be extracted from the term structure in order to forecast the spot exchange rate, even though the forward rate is not an optimal spot rate predictor. Indeed, dynamic out-of-sample forecasts up to one year ahead indicate that the VECM is superior to a range of alternative forecasts, including a random walk and standard spot-forward regressions.

3 Markov-switching equilibrium correction

In this section we outline the econometric procedure employed in order to model regime shifts in the dynamic relationship between spot exchange rates and the term structure of forward rates. The procedure essentially extends Hamilton’s (1988, 1989) Markov-switching regime framework to nonstationary systems, allowing us to apply it to cointegrated vector autoregressive (VAR) and VECM systems (see Krolzig, 1997, 1999).

Consider the following $M$-regime $p$-th order Markov-switching vector autoregression (MS($M$)-VAR($p$)) which allows for regime shifts in the intercept term:\(^4\)

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

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 $\Pi_i$’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 $\Sigma$, $\varepsilon_t \sim NID(0, \Sigma)$. The regime-generating process is assumed to be an ergodic Markov chain with a finite number of states $z_t \in \{1, \ldots, M\}$ governed by the transition probabilities $p_{ij} = \Pr(z_{t+1} = j \mid z_t = i)$, and $\sum_{j=1}^{M} p_{ij} = 1 \forall i, j \in \{1, \ldots, M\}$.

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 indeed there is cointegration, the cointegrated MS-VAR (5) implies a Markov-switching vector equilibrium correction model or MS-VECM of the form:

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

where \( \Gamma_i = - \sum_{j=i+1}^{p} \Pi_j \) are matrices of parameters, and \( \Pi = \sum_{i=1}^{p} \Pi_i - I \) is the long-run impact matrix whose rank \( r \) determines the number of cointegrating vectors (e.g. Johansen, 1995; Krolzig, 1999).

Although, for expositional purposes, we have outlined the MS-VECM framework for the case of regime shifts in the intercept alone, shifts may be allowed for elsewhere. The present application focuses on a multivariate model comprising, for each of the four major dollar exchange rate analyzed, the spot exchange rate and the forward rates at one month (four weeks), three months (thirteen weeks), six months (twenty-six weeks) and twelve months (fifty-two weeks) to maturity (hence \( y_t = [s_t, f_t^{1}, f_t^{2}, f_t^{3}, f_t^{4}]' = [s_t, f_t^{4}, f_t^{13}, f_t^{21}, f_t^{52}]' \), for which, following the reasoning of Section 2, four unique, independent cointegrating relationships should exist. As discussed in Section 4 below, in our empirical work, after considerable experimentation, we selected a specification of the MS-VECM which 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, may be written as follows:

\[
\Delta y_t = \nu(z_t) + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + \Pi y_{t-1} + u_t,
\]

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

An MS-VECM can be estimated using a two-stage maximum likelihood procedure. The first stage of this procedure essentially consists of the implementation of the Johansen (1988, 1991) maximum likelihood cointegration procedure in order to test for the number of cointegrating relationships in the system and to estimate the cointegration matrix. In fact, in the first stage use of the conventional Johansen procedure is legitimate without modelling the Markovian regime shifts.
explicitly (see Saikkonen, 1992; Saikkonen and Luukkonen, 1997). The second stage then consists of the implementation of an expectation-maximization (EM) algorithm for maximum likelihood estimation which yields estimates of the remaining parameters of the model (Dempster, Laird and Rubin, 1977; Hamilton, 1993; Kim and Nelson, 1999; Krolzig, 1999).

We now turn to a brief discussion of our data set and then to our empirical analysis.

4 Empirical results

4.1 Data, unit root tests and cointegration analysis

Our data set comprises weekly observations of spot and 4-, 13-, 26- and 52-week forward US dollar exchange rates among the G5 countries (dollar-franc, dollar-mark, dollar-yen and dollar-sterling) for the period January 7 1979 to December 31 1998, a total of 1,043 observations for each series. In our empirical work, we carried out our estimations over the period January 1979-December 1995, reserving the last three years of data for out-of-sample forecasting tests.

As a preliminary exercise, we tested for unit root behavior of the spot rate and the four forward rate time series examined for each of the four dollar exchange rates under investigation by calculating standard augmented Dickey-Fuller (ADF) test statistics. In each case, the number of lags was chosen such that no residual autocorrelation was evident in the auxiliary regressions. In keeping with the very large number of studies of unit root behavior for these time series, we were in each case unable to reject the unit root null hypothesis at conventional nominal levels of significance. On the other hand, differencing the series did appear to induce stationarity in each case. Hence, the unit root tests clearly indicate that each of the time series examined is a realization from a stochastic process integrated of order one, which suggests that testing for cointegration between \(s_t, f^{14}_t, f^{13}_t, f^{26}_t, f^{52}_t\) is the logical next step.

We then employed the Johansen (1988, 1991) maximum likelihood procedure in a VAR for \(y_t = [s_t, f^{14}_t, f^{13}_t, f^{26}_t, f^{52}_t]\) and an unrestricted constant term. On the basis of the Johansen likelihood ratio test statistics for cointegrating rank (based on the maximal eigenvalue and on the trace of the stochastic matrix), we could strongly reject the hypothesis of three independent cointegrating vectors against the alternative of four, but were not able to reject the hypothesis of
exactly four cointegrating vectors for each exchange rate examined at conventional nominal test sizes.⁹

[TABLE 1 ABOUT HERE]

In order to identify the cointegrating vectors uniquely, we then imposed and tested the over-identifying restrictions on the $\beta'$ matrix of cointegrating coefficients suggested by the framework discussed in Section 2:

$$
\beta' y_t = \begin{bmatrix}
-1 & 1 & 0 & 0 & 0 \\
-1 & 0 & 1 & 0 & 0 \\
-1 & 0 & 0 & 1 & 0 \\
-1 & 0 & 0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
& s_t \\
& f_{1t} \\
& f_{13t} \\
& f_{26t} \\
& f_{52t}
\end{bmatrix},
$$

(8)

As the results reported in Table 1 clearly show, for each currency examined these restrictions were in fact rejected by the data. While these rejections are clearly statistically significant, we proceeded to examine whether the departures from the null hypothesis were large by imposing the following exactly-identifying restrictions:

$$
\beta' y_t = \begin{bmatrix}
-1 & \phi_4 & 0 & 0 & 0 \\
-1 & 0 & \phi_{13} & 0 & 0 \\
-1 & 0 & 0 & \phi_{26} & 0 \\
-1 & 0 & 0 & 0 & \phi_{52}
\end{bmatrix}
\begin{bmatrix}
& s_t \\
& f_{1t} \\
& f_{13t} \\
& f_{26t} \\
& f_{52t}
\end{bmatrix},
$$

(9)

where the $\phi_i$ parameters are unrestricted. This yielded the results reported in Table 2. The results suggest that, consistent with the recent studies by Naka and Whitney (1995) and Luintel and Paudyal (1998), the departure from the overidentifying restrictions, although statistically significant at conventional test sizes, is actually very small in magnitude. Indeed all of the estimated $\phi_i$ coefficients, except for the last two cointegrating relationships obtained on French and German data, are in the range between 0.98 and 1.04 and, therefore, very close indeed to the theoretical value of 1. Thus, rejection of the hypothesis $H_0 : \phi_i = 1 \ \forall i$ may be due to tiny departures from the null hypothesis (due, for example, to tiny data imperfections) which are not economically significant, but which appear as statistically significant given our very large sample
size. The framework we discussed in Section 2 provides strong economic priors in favor of the unity restrictions. Moreover, given that we know that covered interest parity holds strongly among Eurodeposit interest rates (Taylor, 1987, 1989), coefficients different from unity would imply a unit root in international interest rate differentials which seems highly implausible for the countries considered. Given this reasoning, we report below results obtained with the unity restrictions imposed.

[TABLE 2 ABOUT HERE]

### 4.2 Linearity testing and MS-VECM estimation results

We next estimated a standard linear VECM using full-information maximum likelihood (FIML) methods:

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

where $y_t = [s_t, f_{t}^{1}, f_{t}^{13}, f_{t}^{26}, f_{t}^{72}]$, assuming $p = 5$ as suggested by both the Akaike Information Criterion, the Schwartz Information Criterion and the Hannan-Quinn Criterion. Employing the conventional general-to-specific procedure, we obtained fairly parsimonious models for each exchange rate with no significant residual serial correlation. We then applied two fairly general linearity tests to the residuals from the estimated linear VECMs, namely Ramsey’s (1969) RESET test and the Brock, Dechert and Sheinkman (BDS) (1991) test for testing the null hypothesis that the residuals from (10) are independent and identically distributed (iid) against an unspecified alternative. Application of both of these tests provided very strong empirical evidence that the linear VECM fails to capture important nonlinearities in the data generating process, as linearity is rejected with marginal significance levels (p-values) of virtually zero (see Table 3).

[TABLE 3 ABOUT HERE]

We then proceeded to investigate the presence of nonlinearities further through the estimation of a fairly general Markov-switching model of the form:
\[
\Delta y_t - \delta (z_t) = \alpha \left[ \beta' y_{t-1} - \mu (z_t) \right] + \sum_{i=1}^{p-1} \Gamma_i \left[ \Delta y_{t-i} - \delta (z_t) \right] + \omega_t,
\]

where \( y_t = [s_t, f^4_t, f^{13}_t, f^{26}_t, f^{52}_t]' \), \( \delta (z_t) \) is the \((5 \times 1)\) regime-dependent vector of means of the short-run dynamics, and \( \mu (z_t) \) is the \((4 \times 1)\) regime-dependent vector of means of the long-run equilibrium relationships.

Next we applied the conventional ‘bottom-up’ procedure designed to detect Markovian shifts in order to select the most adequate characterization of an \( M \)-regime \( p \)-th order MS-VECM for \( \Delta y_t \). The VARMA representations of the time series suggested in each case that the number of regimes was in the range between two and three. The linearity test indicates in each case the rejection of the linear VECM in favor of a nonlinear alternative model (see Table A1 in the Appendix).

However, for any MS-VECM estimated the implicit assumption that the regime shifts affect only the intercept term of the VECM was found to be inappropriate. In fact, we checked the relevance of conditional homoskedasticity by estimating an MS-VECM where the Gaussian innovation is allowed to be regime-dependent, \( \varepsilon_t \sim NID(0, \Sigma(z_t)) \). We then tested the hypothesis of no regime dependence in the variance-covariance matrix 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 (see Table A1 in the Appendix) indicated very strong rejection of the null of no regime dependence, clearly suggesting that an MS-VECM that allows for shifts in both the intercept and the variance-covariance matrix, namely an MSIH-VECM(\( p \)), is the most appropriate model within its class in the present application. Further in the same spirit of the test for regime-conditional homoskedasticity, we executed a test in order to select the most parsimonious VECM appropriately representing the dynamic relationship between spot and forward exchange rates. In particular, we tested the null of MSIH-VECM(1) against the alternative of MSIH-VECM(\( p \)) and, as may be seen in Table A1 (Appendix), for all currencies examined, we were not able to reject this null hypothesis at standard significance levels.

Finally, in order to discriminate between models allowing for two regimes against models governed by three regimes we also constructed the upper bound LR test of Davies (1977, 1987). The results produced (see Table A1, last column) very large statistics, suggesting that three regimes
may be appropriate in all cases.\textsuperscript{17} Therefore, in spite of parsimony considerations, we allowed for three regimes in our MS-VECM.

As stressed in some recent contributions\textsuperscript{18}, it is instructive to note that model (11), where the regime shifts occur in the drift of the VECM as well as the equilibrium mean of the cointegration relationships, can be equivalently represented by means of an MSI-VECM.\textsuperscript{19} Hence, the final result of the procedure identifies in all countries an MS-VECM model governed by three different regimes that can be written as follows:

\[
\Delta y_t = v(z_t) + \Pi y_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + \omega_t, \tag{12}
\]

where \(\Pi = \alpha\beta\), \(\omega_t \sim N(0, \Sigma(z_t))\) and \(z_t = 1, 2, 3\). We estimated the MSIH-VECM (12) using an EM algorithm for maximum likelihood (Dempster, Laird and Rubin, 1977), for each of France, Germany, Japan and the UK\textsuperscript{20}. With few exceptions, the estimation yields fairly plausible estimates of the coefficients for the VECMs estimated, including the adjustment coefficients in \(\alpha\), which were generally found to be strongly statistically significantly different from zero.\textsuperscript{21,22}

For each country we find that three regimes are appropriate in describing the data, and that in each case the three regimes are driven mainly by the joint variability of spot and forward exchange rates. Shifts from one regime to another appeared to be due largely to shifts in the variance of the term structure equilibrium. On the other hand, shifts in the intercept terms were found to be relatively smaller in magnitude, albeit massively statistically significant. This appears to be in line with the extensive empirical literature investigating the time-varying nature of exchange rates risk premia. One tentative interpretation of our MSIH-VECM is, in fact, in terms of shifts in the mean and variance of foreign exchange returns consistent with deviations from the equilibrium levels implied by conventional macroeconomic fundamentals that may be caused, for example, by ‘peso problems’ or by other kinds of departures from the standard efficient markets hypothesis (see Engel and Hamilton, 1990).
4.3 Forecasting spot exchange rates out-of-sample with the MSIH-VECM

The procedure we have applied so far allowed us to achieve a reliable estimation of the dynamic relationship between spot exchange rates and the whole term structure of forward premia. The exercise conducted by Clarida and Taylor (1997) showed that the term structure of forward rates, embedding a larger information set than a single forward rate, is able to predict the dynamics of the spot exchange rate out-of-sample with a higher precision achieving average gains of 40\% at long horizons with respect to the random walk benchmark model (in terms of root mean square errors).

In order to assess the usefulness of our nonlinear VECM characterization of the term structure, dynamic out-of-sample forecasts of the spot rate were constructed using the MSIH(3)-VECM(1) estimated and described in the previous section. In particular we performed forecasting exercises on the period January 1996-December 1998 with forecast horizons up to 52 weeks ahead.\textsuperscript{23} The out-of-sample forecasts for a given horizon $j = 1, \ldots, 52$ are constructed according to the recursive procedure described in Clarida and Taylor (1993), namely conditional only upon information up to the data of the forecast and with successive re-estimation as the date on which forecasts are conditioned moves through the data set.

It is well known in the literature that forecasting with nonlinear models raises special problems.\textsuperscript{24} We therefore adopt a very general forecasting procedure based on Monte Carlo integration which is capable of producing forecasts virtually identical to the analytical forecasts for a wide range of models. In particular, we forecast the path for $s_{t+j}$ for $j = 1, \ldots, 52$ using Monte Carlo simulations calibrated on the estimated MSIH-VECMs. 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 virtually identical to those which would result from calculating the exact forecast analytically (Gallant, Rossi and Tauchen, 1993; Brown and Mariano, 1984, 1989; Granger and Terasvirta, 1993, Chapter 8; Franses and van Dijk, 2000, Chapters 3-4).

Forecast accuracy is evaluated using absolute and square error criteria, namely the mean abso-
lute error (MAE) and the root mean square error (RMSE). We compared the forecasts produced by the MSIH-VECM to the forecasts generated by a simple random walk benchmark as well as the forecasts generated by the linear term-structure VECM originally proposed by Clarida and Taylor (1997). Further, in order to assess the accuracy of forecasts derived from two different models we employ the Diebold and Mariano (1995) test:

\[ DM = \frac{\overline{d}}{\sqrt{\frac{2\pi \hat{f}(0)}{T}}} \]  \hspace{1cm} (13)

where \( \overline{d} \) is an average (over \( T \) observations) of a general loss differential function and \( \hat{f}(0) \) 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. Consistent with a large literature (see, inter alia, Mark, 1995; Kilian, 1999; Kilian and Taylor, 2000) 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 \( \hat{f}(0) \) 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.25

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 - one of which is 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.26

[TABLE 4 ABOUT HERE]

Table 4 gives detailed results of the accuracy of the forecasts for the dollar-franc, dollar-mark, dollar-sterling and dollar-yen systems respectively, using MAE and RMSE criteria for forecast accuracy. The results provide evidence in favor of the predictive superiority of the MSIH-VECM models against the naive random walk and, to a lesser extent, against linear VECM models. Comparing our results to those obtained using the pure random walk we can see that, across countries, the MSIH-VECM models give very much more accurate forecasts. At the 4-week
horizon we achieve average improvements ranging between 28% and 38% across currencies using
the MAE and between 27% and 31% using the RMSE. At the 52-week horizon we obtain average
improvements ranging between 8% and 70% using the MAE and between 5% and 68% using the
RMSE, with a maximum reduction of 70% in the case of the dollar-yen rate using the MAE. The
statistical significance of these results is confirmed executing the DM test, although the marginal
significance levels reported should be treated with caution in the light of our earlier discussion on
the asymptotic properties of the DM statistic in the present context.

These results extend the findings of Clarida and Taylor (1997) who, using a linear VECM
framework for the term structure of forward foreign exchange premia, were able to provide out-of-
sample forecasts of spot exchange rates which were superior to alternative conventional forecasting
methods. By explicitly incorporating nonlinearity into the modelling framework, we have in the
present analysis been able to improve upon the Clarida-Taylor results as in almost all cases we
are able to improve the forecasting performance of the linear VECM models. In particular, the
gains we obtained relative to the linear VECMs range, on average across currencies, between 1%
and 10% at the 4-week horizon and between 10% and 38% at the 52-week horizon using the MAE.
Using the RMSE the gains range between 1% and 7% at the 4-week horizon and between 10%
and 38% at the 52-week horizon.27 Therefore, the gain from using an MSIH-VECM rather than a
linear VECM is relatively small at short horizon, albeit generally statistically significant; however,
this gain increases with the forecast horizon and becomes substantial at the 52-week horizon.

5 Conclusion

In this paper we have reported what we believe to be the first analysis of spot and forward
exchange rates in a multivariate Markov-switching framework, and in particular we have applied
that framework to exchange rate forecasting. 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 exchange rate modelling, as well as by the
relative forecasting success of the ‘agnostic’ linear VECM model of the term structure of forward
premia.
Using weekly data on spot and forward dollar exchange rates for the G5 countries over the period January 1979 through December 1995, we found strong evidence of the presence of nonlinearities in the term structure, which appeared to be modelled well by 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 1998. The forecasting results were impressive. The MS-VECM forecasts were found to be strongly superior to the random walk forecasts at a range of forecasting horizons up to 52 weeks ahead, using standard forecasting accuracy criteria and on the basis of standard tests of significance. Moreover, the nonlinear VECM also outperformed, in general, a linear VECM for spot and forward rates in out-of-sample forecasting of the spot rate, although the magnitude of the gain from using a Markov-switching VECM relative to a linear VECM is rather small in magnitude at short horizon.

In this research, we have been primarily concerned with providing sound models of exchange rate forecasting and have therefore adopted an ‘agnostic’ approach both with respect to the sources of the underlying departures from the risk-neutral efficient markets hypothesis and in the sources of the underlying nonlinearities. Future research might, therefore, usefully analyze the sources of these nonlinearities further and attempt to improve on the parametric nonlinear formulation proposed in this paper. Possible extensions include the allowance for different equilibrium correction terms (speeds of adjustment towards equilibrium) in different regimes, and the endogeneization of the probability of switching from one regime to another, which might, for example, be made a function of macroeconomic fundamentals.

With regard to the evaluation of forecasting models, although the relevant literature has traditionally focused on accuracy evaluations based on point forecasts, several authors have recently emphasized the importance of evaluating the forecast accuracy of economic models on the basis of density - as opposed to point - forecasting performance (see, inter alia, Diebold, Gunther and Tay, 1998; Granger and Pesaran, 1999; Tay and Wallis, 2000; Timmerman, 2000; Sarno and Valente, 2002). Especially when evaluating nonlinear models, which are capable of producing highly non-normal forecast densities, it would seem appropriate to consider a model’s density forecasting
performance. This is an immediate avenue for future research.

Given the difficulty in beating random walk forecasts using fundamentals-based models - first highlighted by Meese and Rogoff (1983a,b) - as well as the well-known failure of the forward rate optimally to predict the future spot rate, the evidence provided by our results that the term structure of forward rates is powerful in forecasting spot exchange rates is rather striking. In particular, it seems that, notwithstanding the failure of the simple (risk-neutral) efficient markets hypothesis in this context, forward rates may contain more useful information to forecast spot exchange rates than do conventional fundamentals. It seems plausible that important microstructural effects may be responsible for this phenomenon, as argued, for example, by Lyons (2001), Sarno and Taylor (2001) and Evans and Lyons (2002). Understanding the exact nature of this incremental information remains an important challenge in the research agenda.
A Appendix: Markov-Switching-VECM results

Table A1. ‘Bottom-up’ identification procedure

<table>
<thead>
<tr>
<th></th>
<th>LR1</th>
<th>LR2</th>
<th>Davies</th>
</tr>
</thead>
<tbody>
<tr>
<td>France</td>
<td>5007.8</td>
<td>0.4</td>
<td>1305.1</td>
</tr>
<tr>
<td>Germany</td>
<td>3158.6</td>
<td>0.648</td>
<td>440.5</td>
</tr>
<tr>
<td>Japan</td>
<td>3320.9</td>
<td>0.928</td>
<td>487.3</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>1668.5</td>
<td>35.0</td>
<td>477.7</td>
</tr>
</tbody>
</table>

Notes: $LR1$ is a test statistic of the null hypothesis of no regime dependent variance-covariance matrix (i.e. MSI($M$)-VECM($p$) versus MSiH($M$)-VECM($p$)). $LR2$ tests the null hypothesis that the model having autoregressive component of order one is equivalent to another with a higher autoregressive order (i.e. MS($M$)-VECM(1) versus MS($M$)-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 hypothesis due to the existence of 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, and $\{0\}$ indicates $p$-values below $10^{-100}$, which are considered as virtually zero.
References


24


1 See also Meese and Rogoff (1988) and the discussion on the ‘exchange rate disconnect puzzle’ by Obstfeld and Rogoff (2000).

2 The possibility that their forecasting results might be improved upon through exploiting a nonlinear framework was anticipated by Clarida and Taylor, who ended their discussion with the following suggestion: “... in order to establish baseline results, we have restricted ourselves to a linear framework. Further empirical work might usefully extend the present analysis to examine nonlinearities” (Clarida and Taylor, 1997, pp. 361-2). It is the purpose of the present paper to provide exactly this generalization.

3 In our empirical work, we consider forward rates of 4, 13, 26 and 52 weeks maturity, so that in our notation, \( h(1) = 4, h(2) = 13, h(3) = 26, \) and \( h(4) = 52. \)

4 Although, for expositional simplicity, this section focuses on equation (5), clearly a more general formulation of (5) may be considered which allows for all parameters of the model to be conditioned on the state \( z_t. \)

5 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. \( r = K \)).

6 We are grateful to the Bank for International Settlements (BIS) for supplying the data.

7 Results not reported but available from the authors upon request.

8 We allowed for a maximum lag length of five and chose, for each dollar exchange rate, the appropriate lag length on the basis of conventional information criteria.

9 The choice of exactly four independent cointegrating vectors was also suggested by the Hansen-Johansen (1993) recursive procedure for the cointegrating rank. The cointegration test statistics are not reported here in order to conserve space, but are available from the authors on request.

10 Leamer (1978, Chapter 4) points out that classical hypothesis testing will lead to rejection of any null hypothesis with a sufficiently large sample: ‘Classical hypothesis testing at a fixed level of significance increasingly distorts the interpretation of the data against a null hypothesis as the sample size grows. The significance level should consequently be a decreasing function of sample size’ (p. 114).

11 Moreover, standard univariate unit root tests to each of the forward premium series (i.e. with
the unity restrictions imposed) suggested in every case that the forward premium is stationary.

12 We did, however, execute all of the empirical analysis discussed below without imposing the unity restrictions and using instead the estimates of the cointegrating parameters reported in Table 2. The results were quantitatively extremely similar (virtually identical) and qualitatively identical to those reported below.

13 Full details on these estimation results are available from the authors upon request, but are not reported to conserve space and because we mainly focus on nonlinearities and their effect on forecasting in this paper.

14 The BDS test for a series $u_t$ is calculated in the following way. Let $u_{t,v}$ be a set of consecutive terms from $u_t: u_{t,v} \{u_t, u_{t+1}, \ldots, u_{t+v-1}\}$. The pair of vectors $u_{t,v}$ and $u_{s,v}$ are said to be no more than $\varsigma$ apart if $|u_{t+j} - u_{s+j}| \leq \varsigma$ for $j = 0, 1, \ldots, v - 1$. Thus, the correlation integral $C_v(\varsigma)$ is defined as the product of the limit of $T^{-2}$ ($T$ being the number of observations) times the number of $\varsigma$-close pairs $(s,t)$, essentially measuring the probability that the pairs of points $(s,t)$ are within $\varsigma$ of each other. The BDS statistic is then constructed as $S(v, \varsigma) = \hat{C}_v(\varsigma) - [\hat{C}_1(\varsigma)]^v$ for some $v$ and $\varsigma$. Under the null hypothesis that $u_t$ is iid, $\sqrt{T}S(v, \varsigma)$ $\sim N(0, \xi)$, where the variance $\xi$ is a function of $v$ and $\varsigma$. Rejection of the null implies that some form of nonlinearity is present in $u_t$, although the type of nonlinearity cannot be exactly determined under the BDS test. BDS (1991) suggest that the choice of $v$ and, particularly, the choice of $\varsigma$, are crucial for the power of the test, which is reasonably powerful only in large samples. BDS (1991) also suggest values of $\varsigma$ between 0.5 and 1.5 times the standard deviation of $u_t$, whereas the value of $v$ should preferably be such that $(T/v) > 200$.

15 We also compare below the forecasting performance of the linear VECM to that of an MS-VECM.

16 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 checking 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 technical discus-
sion of the bottom-up procedure, see Krolzig (1997); for a more intuitive discussion see Sarno and Valente (2000).

17 It is important to note here that the regularity conditions under which the Davies (1977, 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. Moreover, even if the Davies bound suggested by Krolzig is appropriate, it is possible that it will only be valid if the null model is a linear model with iid errors; in the present case, it is difficult to believe that this condition is met since exchange rate innovations are not homoskedastic, which would induce some distortion. Therefore, the distribution of the LR test is likely to differ from the adjusted \(\chi^2\) distribution proposed by Davies (1977, 1987), and this is why we do not report marginal significance levels for the LR tests. For extensive discussions of the problems related to LR testing in this context, see Hansen (1992, 1996) and Garcia (1998). See Garcia and Perron (1996) for an empirical application. We are thankful to Bruce Hansen for clarifying several econometric issues related to LR testing in the present paper.

18 See, for example, Krolzig (1997) and the references therein.

19 In order to recognize the shifts in the drift of the VECM separately from the ones occurring in the long-run 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 \( \delta(z_t) = \beta_\perp (\alpha_\perp^\prime \beta_\perp)^{-1} \alpha_\perp^\prime v(z_t) \) (\( \perp \) denoting the orthogonal complement) and the equilibrium mean \( \mu(s_t) = - (\beta^\prime \alpha)^{-1} [\beta^\prime v(z_t)] \).

20 These estimation results are not reported to conserve space, but they are available in Clarida, Sarno, Taylor and Valente (2001, Table A2 in the Appendix).

21 Investigation of the estimated smoothed transition probabilities (not reported to conserve space) for the different regimes indicated that, in general, for each country, the three regimes seem to be important in that they characterize a substantial fraction of the joint movements of the spot exchange rates as well as the term structure of forward rates, with each regime being somewhat persistent but with a rather large number of switches over the sample.

22 We also looked at graphs of standardized residuals, the smoothed residuals and the one-step
prediction errors from 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.


24 See Brown and Mariano (1984, 1989) and, for a general discussion, Granger and Terasvirta (1993, Chapter 8) and Franses and van Dijk (2000, Chapters 3–4).

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

26 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. Therefore, 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 (1999) and Kilian and Taylor (2001), although this procedure is computationally very demanding and it is unknown whether it is valid.
in the context of MSIH-VECMs.

27 Note, however, that there is heterogeneity across countries. For example, the gain for the French data is fairly small, whereas the gain for the German data is substantial. The MSIH-VECM for dollar-sterling is arguably the model which performs worse in forecasting and is beaten by a random walk at the 52-week horizon on the basis of the RMSE.
Table 1. Tests of the null hypothesis that independent forward premia comprise a basis for the cointegration space

<table>
<thead>
<tr>
<th></th>
<th>$\chi^2(g)$</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>France</td>
<td>21.6</td>
<td>$5.6 \times 10^{-3}$</td>
</tr>
<tr>
<td>Germany</td>
<td>51.7</td>
<td>$1.8 \times 10^{-8}$</td>
</tr>
<tr>
<td>Japan</td>
<td>41.9</td>
<td>$1.4 \times 10^{-6}$</td>
</tr>
<tr>
<td>UK</td>
<td>53.6</td>
<td>$8.1 \times 10^{-9}$</td>
</tr>
</tbody>
</table>

Notes: The test is a $\chi^2$ version of the test of the overidentifying restrictions on the $\beta$ matrix described in equation (8); $g$ is the number of restrictions imposed.

Table 2. Long-run equilibrium parameters

<table>
<thead>
<tr>
<th>$h(k)$</th>
<th>France</th>
<th>Germany</th>
<th>Japan</th>
<th>UK</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 weeks</td>
<td>0.995 (0.002)</td>
<td>1.006 (0.002)</td>
<td>1.002 (0.001)</td>
<td>0.999 (0.002)</td>
</tr>
<tr>
<td>13 weeks</td>
<td>0.987 (0.007)</td>
<td>1.020 (0.006)</td>
<td>1.005 (0.004)</td>
<td>1.001 (0.007)</td>
</tr>
<tr>
<td>26 weeks</td>
<td>0.976 (0.013)</td>
<td>1.042 (0.011)</td>
<td>1.010 (0.007)</td>
<td>1.004 (0.012)</td>
</tr>
<tr>
<td>52 weeks</td>
<td>0.964 (0.022)</td>
<td>1.084 (0.022)</td>
<td>1.018 (0.013)</td>
<td>1.015 (0.022)</td>
</tr>
</tbody>
</table>

Notes: The table gives the estimated long-run slope parameter for the forward rate at each maturity. Figures in parentheses denote asymptotic standard errors.
Table 3. Linearity tests on the residuals from the linear VECM (10)

*Panel a): RESET tests: p-values*

<table>
<thead>
<tr>
<th></th>
<th>France</th>
<th>Germany</th>
<th>Japan</th>
<th>UK</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta s_t$</td>
<td>$4.1 \times 10^{-60}$</td>
<td>$1.2 \times 10^{-49}$</td>
<td>$3.6 \times 10^{-37}$</td>
<td>$3.1 \times 10^{-18}$</td>
</tr>
<tr>
<td>$\Delta f_t^4$</td>
<td>$3.7 \times 10^{-48}$</td>
<td>$4.1 \times 10^{-58}$</td>
<td>$1.2 \times 10^{-34}$</td>
<td>$7.3 \times 10^{-19}$</td>
</tr>
<tr>
<td>$\Delta f_t^{13}$</td>
<td>$5.2 \times 10^{-56}$</td>
<td>$3.1 \times 10^{-58}$</td>
<td>$2.1 \times 10^{-52}$</td>
<td>$2.3 \times 10^{-4}$</td>
</tr>
<tr>
<td>$\Delta f_t^{26}$</td>
<td>$4.4 \times 10^{-22}$</td>
<td>$2.6 \times 10^{-19}$</td>
<td>$1.4 \times 10^{-16}$</td>
<td>$7.2 \times 10^{-57}$</td>
</tr>
<tr>
<td>$\Delta f_t^{52}$</td>
<td>$1.0 \times 10^{-26}$</td>
<td>$2.6 \times 10^{-27}$</td>
<td>$2.9 \times 10^{-16}$</td>
<td>$1.5 \times 10^{-54}$</td>
</tr>
</tbody>
</table>

(continued ...)

32
Panel b): BDS tests: p-values

<table>
<thead>
<tr>
<th></th>
<th>$v = 2$</th>
<th>$v = 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\zeta = 0.5\sigma_u$</td>
<td>$\zeta = 1.0\sigma_u$</td>
</tr>
<tr>
<td>France</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Delta s_t$</td>
<td>$4.9 \times 10^{-2}$</td>
<td>$3.0 \times 10^{-2}$</td>
</tr>
<tr>
<td>$\Delta f^3_t$</td>
<td>$1.0 \times 10^{-2}$</td>
<td>$1.4 \times 10^{-2}$</td>
</tr>
<tr>
<td>$\Delta f^3_{13}$</td>
<td>$9.3 \times 10^{-3}$</td>
<td>$9.9 \times 10^{-3}$</td>
</tr>
<tr>
<td>$\Delta f^6_t$</td>
<td>$2.1 \times 10^{-3}$</td>
<td>$3.4 \times 10^{-3}$</td>
</tr>
<tr>
<td>$\Delta f^2_{12}$</td>
<td>$2.0 \times 10^{-4}$</td>
<td>$1.2 \times 10^{-4}$</td>
</tr>
<tr>
<td>Germany</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Delta s_t$</td>
<td>$5.7 \times 10^{-2}$</td>
<td>$1.2 \times 10^{-2}$</td>
</tr>
<tr>
<td>$\Delta f^3_t$</td>
<td>$3.9 \times 10^{-2}$</td>
<td>$1.2 \times 10^{-2}$</td>
</tr>
<tr>
<td>$\Delta f^3_{13}$</td>
<td>$7.7 \times 10^{-2}$</td>
<td>$8.8 \times 10^{-3}$</td>
</tr>
<tr>
<td>$\Delta f^6_t$</td>
<td>$3.1 \times 10^{-2}$</td>
<td>$6.8 \times 10^{-3}$</td>
</tr>
<tr>
<td>$\Delta f^2_{12}$</td>
<td>$3.4 \times 10^{-2}$</td>
<td>$5.8 \times 10^{-3}$</td>
</tr>
<tr>
<td>Japan</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Delta s_t$</td>
<td>$9.7 \times 10^{-5}$</td>
<td>$8.4 \times 10^{-5}$</td>
</tr>
<tr>
<td>$\Delta f^3_t$</td>
<td>$4.5 \times 10^{-5}$</td>
<td>$5.9 \times 10^{-5}$</td>
</tr>
<tr>
<td>$\Delta f^3_{13}$</td>
<td>$3.4 \times 10^{-4}$</td>
<td>$1.7 \times 10^{-4}$</td>
</tr>
<tr>
<td>$\Delta f^6_t$</td>
<td>$8.6 \times 10^{-5}$</td>
<td>$9.4 \times 10^{-5}$</td>
</tr>
<tr>
<td>$\Delta f^2_{12}$</td>
<td>$1.1 \times 10^{-10}$</td>
<td>$2.9 \times 10^{-10}$</td>
</tr>
<tr>
<td>UK</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Delta s_t$</td>
<td>$2.1 \times 10^{-2}$</td>
<td>$1.7 \times 10^{-2}$</td>
</tr>
<tr>
<td>$\Delta f^3_t$</td>
<td>$1.0 \times 10^{-2}$</td>
<td>$1.3 \times 10^{-2}$</td>
</tr>
<tr>
<td>$\Delta f^3_{13}$</td>
<td>$1.9 \times 10^{-2}$</td>
<td>$1.0 \times 10^{-2}$</td>
</tr>
<tr>
<td>$\Delta f^6_t$</td>
<td>$5.8 \times 10^{-3}$</td>
<td>$4.9 \times 10^{-3}$</td>
</tr>
<tr>
<td>$\Delta f^2_{12}$</td>
<td>$4.5 \times 10^{-3}$</td>
<td>$1.2 \times 10^{-3}$</td>
</tr>
</tbody>
</table>

Notes: The RESET test statistics were computed considering an alternative model with a quadratic and a cubic term; they are distributed as $F(2, T - m - 3)$ under the null hypothesis of linearity (no misspecification), where $T$ is the number of observations and $m$ is the number of regressors (inclusive of the intercept). The BDS test statistic tests the null hypothesis that a series is iid against the alternative of a realization from an unspecified nonlinear process, as described in the text. The critical values, from the normal distribution, are 1.960 and 2.576 at the five percent and one percent nominal levels of significance respectively. Given that the choices of $v$ and $\zeta$ are crucial for the power of the test, we report the results for different plausible values of $v$ and $\zeta$, as suggested by BDS (1991) and explained in the text; $\sigma_u$ is the standard deviation of the residuals from the linear VECM (10). For both RESET tests and BDS tests, only $p$-values are reported.
Table 4: Forecasting exercises

a) France

<table>
<thead>
<tr>
<th></th>
<th>MSVECM</th>
<th>RW</th>
<th>VECM</th>
<th>DM1(A)</th>
<th>DM2(A)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>0.0037</td>
<td>0.0057</td>
<td>0.0037</td>
<td>1.3×10^{-1}</td>
<td>1.7×10^{-1}</td>
</tr>
<tr>
<td>13</td>
<td>0.0037</td>
<td>0.0066</td>
<td>0.0038</td>
<td>6.9×10^{-3}</td>
<td>3.0×10^{-2}</td>
</tr>
<tr>
<td>26</td>
<td>0.0047</td>
<td>0.0027</td>
<td>0.0049</td>
<td>1.2×10^{-2}</td>
<td>1.7×10^{-6}</td>
</tr>
<tr>
<td>52</td>
<td>0.0066</td>
<td>0.0115</td>
<td>0.0073</td>
<td>3.9×10^{-3}</td>
<td>2.2×10^{-16}</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>MSVECM</th>
<th>RW</th>
<th>VECM</th>
<th>DM1(A)</th>
<th>DM2(A)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>0.0043</td>
<td>0.0062</td>
<td>0.0043</td>
<td>1.0×10^{-1}</td>
<td>3.0×10^{-1}</td>
</tr>
<tr>
<td>13</td>
<td>0.0047</td>
<td>0.0074</td>
<td>0.0048</td>
<td>4.9×10^{-3}</td>
<td>2.2×10^{-2}</td>
</tr>
<tr>
<td>26</td>
<td>0.0055</td>
<td>0.0092</td>
<td>0.0058</td>
<td>1.0×10^{-2}</td>
<td>3.7×10^{-8}</td>
</tr>
<tr>
<td>52</td>
<td>0.0095</td>
<td>0.0147</td>
<td>0.0106</td>
<td>2.1×10^{-3}</td>
<td>1.4×10^{-13}</td>
</tr>
</tbody>
</table>

b) Germany

<table>
<thead>
<tr>
<th></th>
<th>MSVECM</th>
<th>RW</th>
<th>VECM</th>
<th>DM1(A)</th>
<th>DM2(A)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>0.0040</td>
<td>0.0064</td>
<td>0.0042</td>
<td>5.4×10^{-2}</td>
<td>2.7×10^{-1}</td>
</tr>
<tr>
<td>13</td>
<td>0.0043</td>
<td>0.0082</td>
<td>0.0056</td>
<td>1.3×10^{-3}</td>
<td>1.1×10^{-10}</td>
</tr>
<tr>
<td>26</td>
<td>0.0055</td>
<td>0.0121</td>
<td>0.0075</td>
<td>2.1×10^{-7}</td>
<td>7.8×10^{-4}</td>
</tr>
<tr>
<td>52</td>
<td>0.0062</td>
<td>0.0160</td>
<td>0.0100</td>
<td>1.3×10^{-15}</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>MSVECM</th>
<th>RW</th>
<th>VECM</th>
<th>DM1(A)</th>
<th>DM2(A)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>0.0046</td>
<td>0.0066</td>
<td>0.0047</td>
<td>7.9×10^{-2}</td>
<td>3.7×10^{-1}</td>
</tr>
<tr>
<td>13</td>
<td>0.0054</td>
<td>0.0094</td>
<td>0.0068</td>
<td>1.0×10^{-3}</td>
<td>1.8×10^{-17}</td>
</tr>
<tr>
<td>26</td>
<td>0.0066</td>
<td>0.0138</td>
<td>0.0088</td>
<td>7.0×10^{-8}</td>
<td>3.3×10^{-4}</td>
</tr>
<tr>
<td>52</td>
<td>0.0080</td>
<td>0.0190</td>
<td>0.0128</td>
<td>7.0×10^{-12}</td>
<td>0</td>
</tr>
</tbody>
</table>

(continued ...)

34
### c) Japan

<table>
<thead>
<tr>
<th>(k)</th>
<th>MSVECM</th>
<th>RW</th>
<th>VECD</th>
<th>DM1(A)</th>
<th>DM2(A)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>0.0020</td>
<td>0.0032</td>
<td>0.0022</td>
<td>1.2 \times 10^{-2}</td>
<td>5.9 \times 10^{-1}</td>
</tr>
<tr>
<td>13</td>
<td>0.0032</td>
<td>0.0042</td>
<td>0.0032</td>
<td>1.9 \times 10^{-2}</td>
<td>8.9 \times 10^{-1}</td>
</tr>
<tr>
<td>26</td>
<td>0.0040</td>
<td>0.0077</td>
<td>0.0046</td>
<td>6.1 \times 10^{-6}</td>
<td>2.8 \times 10^{-2}</td>
</tr>
<tr>
<td>52</td>
<td>0.0053</td>
<td>0.0179</td>
<td>0.0082</td>
<td>3.9 \times 10^{-14}</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(k)</th>
<th>MSVECM</th>
<th>RW</th>
<th>VECD</th>
<th>DM1(A)</th>
<th>DM2(A)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>0.0026</td>
<td>0.0036</td>
<td>0.0028</td>
<td>4.4 \times 10^{-2}</td>
<td>5.4 \times 10^{-1}</td>
</tr>
<tr>
<td>13</td>
<td>0.0042</td>
<td>0.0052</td>
<td>0.0042</td>
<td>4.6 \times 10^{-2}</td>
<td>9.6 \times 10^{-1}</td>
</tr>
<tr>
<td>26</td>
<td>0.0050</td>
<td>0.0094</td>
<td>0.0058</td>
<td>2.9 \times 10^{-7}</td>
<td>5.6 \times 10^{-4}</td>
</tr>
<tr>
<td>52</td>
<td>0.0074</td>
<td>0.0230</td>
<td>0.0112</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

### d) United Kingdom

<table>
<thead>
<tr>
<th>(k)</th>
<th>MSVECM</th>
<th>RW</th>
<th>VECD</th>
<th>DM1(A)</th>
<th>DM2(A)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>0.0044</td>
<td>0.0061</td>
<td>0.0046</td>
<td>1.8 \times 10^{-2}</td>
<td>3.4 \times 10^{-1}</td>
</tr>
<tr>
<td>13</td>
<td>0.0025</td>
<td>0.0047</td>
<td>0.0028</td>
<td>1.3 \times 10^{-3}</td>
<td>1.4 \times 10^{-1}</td>
</tr>
<tr>
<td>26</td>
<td>0.0042</td>
<td>0.0059</td>
<td>0.0053</td>
<td>5.1 \times 10^{-3}</td>
<td>6.9 \times 10^{-7}</td>
</tr>
<tr>
<td>52</td>
<td>0.0135</td>
<td>0.0147</td>
<td>0.0167</td>
<td>5.1 \times 10^{-1}</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(k)</th>
<th>MSVECM</th>
<th>RW</th>
<th>VECD</th>
<th>DM1(A)</th>
<th>DM2(A)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>0.0049</td>
<td>0.0067</td>
<td>0.0052</td>
<td>2.1 \times 10^{-2}</td>
<td>2.9 \times 10^{-1}</td>
</tr>
<tr>
<td>13</td>
<td>0.0034</td>
<td>0.0054</td>
<td>0.0036</td>
<td>2.1 \times 10^{-3}</td>
<td>3.5 \times 10^{-1}</td>
</tr>
<tr>
<td>26</td>
<td>0.0054</td>
<td>0.0069</td>
<td>0.0069</td>
<td>1.2 \times 10^{-2}</td>
<td>7.5 \times 10^{-13}</td>
</tr>
<tr>
<td>52</td>
<td>0.0186</td>
<td>0.0069</td>
<td>0.0225</td>
<td>4.4 \times 10^{-1}</td>
<td>0</td>
</tr>
</tbody>
</table>

**Notes:** MSVECM, RW and VECD are the average level of the (absolute or square) forecast error obtained by the MSH-VECD, random walk and linear VECD respectively. The forecast errors are obtained by recursive estimation of out-of-sample dynamic forecasts up to \(k = 4, 13, 26, 52\) steps ahead over the period 1996:1-1998:52. DM1(A) is the Diebold-Mariano statistic comparing the forecast errors of the MSH-VECD model with the ones obtained by a driftless random walk. DM2(A) is the Diebold-Mariano statistic comparing the forecast errors of the MSH-VECD model with the ones obtained by the linear VECD. \(A\) is the optimal truncation lag constructed according to the AR(1) Andrews (1991) rule. For the Diebold and Mariano”s statistics only \(p\)—values are reported. 0 indicates \(p\)—values below \(10^{-10}\), which are considered as virtually zero.