NBER WORKING PAPER SERIES

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

Richard H. Clarida
Lucio Sarno
Mark P. Taylor
Giorgio Valente

The Out-of-Sample Success of Term Structure Models
as Exchange Rate Predictors: A Step Beyond
Richard H. Clarida, Lucio Sarno, Mark P. Taylor and Giorgio Valente
NBER Working Paper No. 8601
November 2001
JEL No. F31, F37

## ABSTRACT

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. However, there is evidence that the term structure of forward premia contains valuable information for forecasting future spot exchange rates and that exchange rate dynamics display nonlinearities. 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 a linear term-structure vector equilibrium correction model for four major dollar rates across a range of horizons.

Richard Clarida
NBER and Columbia University
Room 1014
420 W. 118th Street
New York, NY 10027
rhc2@columbia.edu

Lucio Sarno
CEPR and University of Warwick
Coventry CV4 7AL
UK

Mark Taylor
CEPR and University of Warwick
Coventry CV4 7AL
UK

Giorgio Valente
CEPR and University of Warwick
Coventry CV4 7AL
UK

# 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.[1] 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 a number of 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

---

[1] See also Meese and Rogoff (1988).

2

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.[2] 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' 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 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 significantly superior even 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 significant improvements can be made over linear forecasting models by allowing for nonlinearities.

---

[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.

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 nonstationary 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 $t$.[3] 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

$$s_t = m_t + q_t, \tag{1}$$

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 $_t$ (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 $_t$. 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).

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:

$$\gamma_t \equiv f_t^{h(k)} - E\left(s_{t+h(k)}|\ _t\right), \tag{2}$$

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

$$f_t^{h(k)} = \gamma_t + h(k)\theta + E_t\left(q_{t+h(k)}|\ _t\right) + m_t, \tag{3}$$

---

[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 $h(4) = 52$.

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 $t$:

$$f_t^{h(k)} - s_t = \gamma_t + h(k)\theta + E_t\left(q_{t+h(k)} - q_t\mid \ \ _t\right). \tag{4}$$

Equation (4) says that if the departure from the RNMEH $\gamma_t$ is stationary, given $q_t \sim I(1)$, 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 $[-\iota, I_m]$, where $I_m$ is an $m$-dimensional identity matrix and $\iota$ 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 vector equilibrium correction (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 strikingly 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]

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

where $y_t$ is a $K$-dimensional observed time series vector, $y_t = [y_{1t}, y_{2t}, \ldots, y_{Kt}]'$; $\nu(z_t) = [\nu_1(z_t), \nu_2(z_t), \ldots, \nu_K(z_t)]'$ is a $K$-dimensional column vector of regime-dependent intercept terms; the $\Pi_i$'s are $K \times K$ matrices of parameters; $\varepsilon_t = $

---

[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 $z_t$.

$[\varepsilon_{1t}, \varepsilon_{2t}, \ldots, \varepsilon_{Kt}]'$ is a $K$-dimensional vector of Gaussian white noise processes with covariance matrix $\Sigma$, $\varepsilon_t \sim NID(\mathbf{0}, \Sigma)$. The regime-generating process is assumed to be an ergodic Markov chain with a finite number of states $z_t \in \{1, \ldots, M\}$ governed by the transition probabilities $p_{ij} = \Pr(z_{t+1} = j \mid z_t = i)$, and $\sum_{j=1}^{M} p_{ij} = 1 \ \forall i, j \in \{1, \ldots, M\}$.[5]

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 (MS-VECM) of the form:

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

where $\Gamma_i = -\sum_{j=i+1}^{p} \Pi_j$ are matrices of parameters, and $\Pi = \sum_{i=1}^{p} \Pi_i - \mathbf{I}$ is the long-run impact matrix whose rank $r$ determines the number of cointegrating vectors (e.g. Johansen, 1995; Krolzig, 1999)[6].

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 $y_t = [s_t, f_t^{h(1)}, f_t^{h(2)}, f_t^{h(3)}, f_t^{h(4)}]' = [s_t, f_t^4, f_t^{13}, f_t^{26}, f_t^{52}]'$), 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:

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

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

---

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

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1M} \\ p_{21} & p_{22} & \cdots & p_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ p_{M1} & p_{M2} & \cdots & p_{MM} \end{bmatrix},$$

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

[6] 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. $r = K$).

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, 1990; 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.[7] 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.[8] 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[9]. 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 $s_t$, $f_t^4$, $f_t^{13}$, $f_t^{26}$, and $f_t^{52}$, is the logical next step.

We then employed the Johansen (1988, 1991) maximum likelihood procedure in a VAR for $y_t = [s_t, f_t^4, f_t^{13}, f_t^{26}, f_t^{52}]'$ and an unrestricted constant term[10]. On

---

[7] We are grateful to the Bank for International Settlements (BIS) for supplying the data.

[8] Moreover, using non-augmented Dickey Fuller tests or augmented Dickey-Fuller tests with any number of lags in the range from 1 to 12 yielded qualitatively identical results. In addition to the ADF tests, we also executed Johansen likelihood ratio tests (Johansen, 1988, 1991) with a 'univariate VAR'. The results were consistent with the ADF test results.

[9] Results not reported but available from the authors upon request.

[10] 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.

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.[11]

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:

$$\beta' y_t = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} s_t \\ f_t^4 \\ f_t^{13} \\ f_t^{26} \\ f_t^{52} \end{bmatrix}. \tag{8}$$

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:

$$\beta' y_t = \begin{bmatrix} -1 & \phi_4 & 0 & 0 & 0 \\ -1 & 0 & \phi_{13} & 0 & 0 \\ -1 & 0 & 0 & \phi_{26} & 0 \\ -1 & 0 & 0 & 0 & \phi_{52} \end{bmatrix} \begin{bmatrix} s_t \\ f_t^4 \\ f_t^{13} \\ f_t^{26} \\ f_t^{52} \end{bmatrix}, \tag{9}$$

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), that 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.[12] The framework we discussed in Section 2 provides strong economic priors in favor

---

[11] The choice of exactly four independent cointegrating vectors was also strongly 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.

[12] 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

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.[13] Given this reasoning, we report below results obtained with the unity restrictions imposed.[14]

## 4.2 Linearity testing and MS-VECM estimation results

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

$$\Delta y_t = \nu + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + \Pi y_{t-1} + u_t \tag{10}$$

where $y_t = [s_t, f_t^4, f_t^{13}, f_t^{26}, f_t^{52}]'$, 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.[15] We then applied two fairly general linearity tests to the residuals from the estimated linear VECMs, namely Ramsey's (1969) RE-SET 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.[16] Application of both of

---

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). See also Berkson (1938).

[13] 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.

[14] 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, which is not suprising given that the estimated parameters are very close to unity.

[15] 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.

[16] The BDS test for a series $u_t$ is calculated in the following way. Let $\mathbf{u}_{t,v}$ be a set of consecutive terms from $u_t : \mathbf{u}_{t,v} \{u_t, u_{t+1}, \ldots, u_{t+v-1}\}$. The pair of vectors $\mathbf{u}_{t,v}$ and $\mathbf{u}_{s,v}$ 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 $S(v, \varsigma) = \widehat{C}_v(\varsigma) - [\widehat{C}_1(\varsigma)]^v$ 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 $(T/v) > 200$.

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).[17]

We then proceeded to investigate the presence of nonlinearities further through the estimation of a fairly general Markov-switching model of the form:

$$\Delta y_t - \delta\left(z_t\right) = \alpha\left[\beta' y_{t-1} - \mu\left(z_t\right)\right] + \sum_{i=1}^{p-1}\Gamma_i\left[\Delta y_{t-i} - \delta\left(z_t\right)\right] + \omega_t, \qquad (11)$$

where $y_t = [s_t, f_t^4, f_t^{13}, f_t^{26}, f_t^{52}]'$, $\delta\left(z_t\right)$ is the $(5 \times 1)$ regime-dependent vector of means of the short-run dynamics, and $\mu\left(z_t\right)$ 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 $\Delta y_t$[18]. 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, $\varepsilon_t \sim NID(\mathbf{0}, \Sigma(z_t))$. 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 of 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

---

[17] We also discuss below how the forecasting performance of the linear VECM is inferior to that of an MS-VECM.

[18] 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 comprehensive discussion of the bottom-up procedure, see Krolzig (1997).

(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.[19] Therefore, in spite of parsimony considerations, we allowed for three regimes in our MS-VECM.

As stressed in some recent contributions[20], 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[21]. 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:

$$\Delta y_t = v\left(z_t\right) + \Pi y_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + \omega_t, \qquad (12)$$

where $\Pi = \alpha\beta'$, $\omega_t \sim NIID(\mathbf{0}, \Sigma(z_t))$ and $z_t = 1, 2, 3$. In Table A2 of the Appendix, we report the final MSIH-VECM estimation results for each of France, Germany, Japan and the UK. 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.[22], [23]

---

[19] 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.

[20] See, for example, Krolzig (1997) and the references therein.

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

[22] 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. Also, regime identification seems to be mainly driven by the frequent shifts in the conditional variance rather than the ones involving the intercept of the MSIH-VECM.

[23] We also looked at graphs of standardized residuals, the smoothed residuals and the one-

### 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.[24] 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.[25] 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 reported in Table A2 in the Appendix. The vector of Gaussian innovations is set consistent with the covariance matrices given in Panel B) of Table A2. 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

---

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. See, for example, Krolzig (1997) for a discussion of residual-based model checking in this context.

[24] For a description of the procedure of out-of-sample forecasting in a Markov-switching framework, see Hamilton (1993, 1994). See also Franses and van Dijk (2000).

[25] 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).

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 absolute error or MAE, the mean square error or MSE, and the root mean square error or 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:

$$DM = \frac{\overline{d}}{\sqrt{\frac{2\pi \widehat{f}(0)}{T}}} \qquad (13)$$

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

Several problems may arise using DM statistics in small sample as well as taking into account parameter uncertainty (see also West, 1996; West and Mc-Cracken, 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.[27]

---

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

[27] 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.

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, MSE 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* against linear VECM models. Comparing our results to those obtained using the pure random walk we can see that, across criteria and across countries, the MSIH-VECM models give very much more accurate forecasts. At the 4-week horizon we achieve average improvements ranging between 24% and 26% across currencies and accuracy criteria, while at the 52-week horizon we obtain average improvements ranging between 43% and 76% with a maximum reduction of 90% in the case of dollar-yen. 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.[28]

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 even further and in all cases and in almost all the time horizons we are able to improve the forecasting performance of the linear VECM models. In particular, the gains that we obtained relative to the linear VECMs range on average, across criteria and across currencies, between 1-11% (4 weeks) and 13%-46% (52 weeks).

## 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

---

[28]Note that the MSIH-VECM for dollar-sterling is, in general, the model which performs worse in forecasting.

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 strongly outperformed a linear VECM for spot and forward rates in out-of-sample forecasting of the spot rate for all exchange rates and horizons. In forecasts of the dollar-yen, for example, the MS-VECM forecasts were approximately sixty percent better than the linear VECM forecasts, and ninety percent better than the random walk forecasts.

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 underlying departures from the risk-neutral efficient markets hypothesis and in the sources of the underlying nonlinearities. Further 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; Pesaran and Skouras, 2001). 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 Evans and Lyons (1999)[29], Lyons

---

[29]Evans and Lyons (1999) provide a model which sheds light on the role of order flow in determining exchange rates. In their model, order flow is a proximate determinant of price. Using data on signed order flow from the Reuters dealing system, Evans and Lyons find that order flow is a significant determinant of some major bilateral exchange rates, obtaining coefficients of determination significantly larger than the ones usually obtained using standard macroeconomic models of nominal exchange rates. Evans and Lyons also show that their

(1999), and Sarno and Taylor (2001). Understanding the exact nature of this incremental information remains an important challenge in the research agenda.

---

model of daily exchange rate changes produces good out-of-sample forecasts at short horizons, beating the alternative of a random walk model in standard measures of forecast accuracy.

**Table 1.  Tests of the null hypothesis that independent forward premia comprise a basis for the cointegration space**

|  | $\chi^2\,(g)$ {$p$-value} |
|---|---|
| France | 21.643 {$5.62\times10^{-3}$} |
| Germany | 51.746 {$1.88\times10^{-8}$} |
| Japan | 41.915 {$1.40\times10^{-6}$} |
| UK | 53.633 {$8.13\times10^{-9}$} |

**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**

| $h(k)$ | France | Germany | Japan | UK |
|---|---|---|---|---|
| 4 weeks | 0.99549 (0.002) | 1.0068 (0.002) | 1.0021 (0.001) | 0.99997 (0.002) |
| 13 weeks | 0.98768 (0.007) | 1.0208 (0.006) | 1.0056 (0.004) | 1.0015 (0.007) |
| 26 weeks | 0.97687 (0.013) | 1.0420 (0.011) | 1.0108 (0.007) | 1.0043 (0.012) |
| 52 weeks | 0.96401 (0.022) | 1.0842 (0.022) | 1.0180 (0.013) | 1.0150 (0.022) |

**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*

|  | France | Germany | Japan | UK |
|---|---|---|---|---|
| $\Delta s_t$ | $4.14 \times 10^{-60}$ | $1.28 \times 10^{-49}$ | $3.68 \times 10^{-37}$ | $3.10 \times 10^{-18}$ |
| $\Delta f_t^4$ | $3.77 \times 10^{-48}$ | $4.17 \times 10^{-58}$ | $1.26 \times 10^{-34}$ | $7.36 \times 10^{-19}$ |
| $\Delta f_t^{13}$ | $5.29 \times 10^{-56}$ | $3.12 \times 10^{-58}$ | $2.15 \times 10^{-52}$ | $2.38 \times 10^{-4}$ |
| $\Delta f_t^{26}$ | $4.43 \times 10^{-22}$ | $2.61 \times 10^{-19}$ | $1.45 \times 10^{-46}$ | $7.22 \times 10^{-57}$ |
| $\Delta f_t^{52}$ | $1.04 \times 10^{-26}$ | $2.67 \times 10^{-27}$ | $2.96 \times 10^{-16}$ | $1.53 \times 10^{-54}$ |

(continued)

18

*Panel b):  BDS tests: p-values*

| | | $v = 2$ | | | $v = 3$ | |
|---|---|---|---|---|---|---|
| | $\varsigma = 0.5\sigma_u$ | $\varsigma = 1.0\sigma_u$ | $\varsigma = 1.5\sigma_u$ | $\varsigma = 0.5\sigma_u$ | $\varsigma = 1.0\sigma_u$ | $\varsigma = 1.5\sigma_u$ |
| France | | | | | | |
| $\Delta s_t$ | $4.97{\times}10^{-2}$ | $3.00{\times}10^{-2}$ | $7.25{\times}10^{-3}$ | $5.48{\times}10^{-5}$ | $2.06{\times}10^{-4}$ | $3.44{\times}10^{-5}$ |
| $\Delta f_t^4$ | $1.01{\times}10^{-2}$ | $1.48{\times}10^{-2}$ | $4.20{\times}10^{-3}$ | $5.36{\times}10^{-5}$ | $1.18{\times}10^{-4}$ | $1.85{\times}10^{-5}$ |
| $\Delta f_t^{13}$ | $9.30{\times}10^{-3}$ | $9.91{\times}10^{-3}$ | $4.22{\times}10^{-3}$ | $7.04{\times}10^{-5}$ | $1.69{\times}10^{-4}$ | $4.09{\times}10^{-5}$ |
| $\Delta f_t^{26}$ | $2.19{\times}10^{-3}$ | $3.49{\times}10^{-3}$ | $6.08{\times}10^{-4}$ | $3.54{\times}10^{-5}$ | $1.18{\times}10^{-4}$ | $1.04{\times}10^{-5}$ |
| $\Delta f_t^{52}$ | $2.04{\times}10^{-4}$ | $1.26{\times}10^{-4}$ | $2.17{\times}10^{-5}$ | $1.70{\times}10^{-5}$ | $1.75{\times}10^{-5}$ | $2.43{\times}10^{-6}$ |
| Germany | | | | | | |
| $\Delta s_t$ | $5.70{\times}10^{-2}$ | $1.29{\times}10^{-2}$ | $4.41{\times}10^{-3}$ | $2.09{\times}10^{-5}$ | $2.55{\times}10^{-5}$ | $5.78{\times}10^{-6}$ |
| $\Delta f_t^4$ | $3.92{\times}10^{-2}$ | $1.23{\times}10^{-2}$ | $2.56{\times}10^{-3}$ | $8.46{\times}10^{-6}$ | $1.98{\times}10^{-5}$ | $2.29{\times}10^{-6}$ |
| $\Delta f_t^{13}$ | $7.71{\times}10^{-2}$ | $8.83{\times}10^{-3}$ | $1.96{\times}10^{-3}$ | $3.90{\times}10^{-5}$ | $8.29{\times}10^{-6}$ | $2.39{\times}10^{-6}$ |
| $\Delta f_t^{26}$ | $3.14{\times}10^{-2}$ | $6.83{\times}10^{-3}$ | $1.71{\times}10^{-3}$ | $5.23{\times}10^{-6}$ | $4.86{\times}10^{-6}$ | $2.21{\times}10^{-6}$ |
| $\Delta f_t^{52}$ | $3.41{\times}10^{-2}$ | $5.80{\times}10^{-3}$ | $1.56{\times}10^{-3}$ | $2.07{\times}10^{-5}$ | $1.57{\times}10^{-6}$ | $3.12{\times}10^{-6}$ |
| Japan | | | | | | |
| $\Delta s_t$ | $9.72{\times}10^{-5}$ | $8.47{\times}10^{-5}$ | $1.06{\times}10^{-4}$ | $3.41{\times}10^{-6}$ | $5.42{\times}10^{-6}$ | $2.00{\times}10^{-5}$ |
| $\Delta f_t^4$ | $4.56{\times}10^{-5}$ | $5.90{\times}10^{-5}$ | $7.57{\times}10^{-5}$ | $7.26{\times}10^{-7}$ | $2.76{\times}10^{-6}$ | $1.30{\times}10^{-5}$ |
| $\Delta f_t^{13}$ | $3.42{\times}10^{-4}$ | $1.71{\times}10^{-4}$ | $1.12{\times}10^{-4}$ | $6.35{\times}10^{-7}$ | $5.38{\times}10^{-6}$ | $1.80{\times}10^{-5}$ |
| $\Delta f_t^{26}$ | $8.68{\times}10^{-5}$ | $9.40{\times}10^{-5}$ | $2.96{\times}10^{-4}$ | $2.78{\times}10^{-8}$ | $1.14{\times}10^{-6}$ | $5.28{\times}10^{-5}$ |
| $\Delta f_t^{52}$ | $1.10{\times}10^{-10}$ | $2.93{\times}10^{-10}$ | $6.76{\times}10^{-8}$ | $2.00{\times}10^{-15}$ | $3.21{\times}10^{-13}$ | $3.29{\times}10^{-9}$ |
| UK | | | | | | |
| $\Delta s_t$ | $2.13{\times}10^{-2}$ | $1.71{\times}10^{-2}$ | $7.78{\times}10^{-8}$ | $3.60{\times}10^{-3}$ | $2.17{\times}10^{-4}$ | $1.68{\times}10^{-4}$ |
| $\Delta f_t^4$ | $1.07{\times}10^{-2}$ | $1.31{\times}10^{-2}$ | $6.04{\times}10^{-3}$ | $1.42{\times}10^{-4}$ | $1.53{\times}10^{-4}$ | $1.13{\times}10^{-4}$ |
| $\Delta f_t^{13}$ | $1.97{\times}10^{-2}$ | $1.06{\times}10^{-2}$ | $7.14{\times}10^{-3}$ | $2.94{\times}10^{-4}$ | $1.26{\times}10^{-4}$ | $1.23{\times}10^{-4}$ |
| $\Delta f_t^{26}$ | $5.86{\times}10^{-3}$ | $4.90{\times}10^{-3}$ | $5.15{\times}10^{-3}$ | $2.47{\times}10^{-5}$ | $3.45{\times}10^{-5}$ | $6.85{\times}10^{-5}$ |
| $\Delta f_t^{52}$ | $4.55{\times}10^{-3}$ | $1.28{\times}10^{-3}$ | $2.23{\times}10^{-3}$ | $1.10{\times}10^{-5}$ | $2.01{\times}10^{-6}$ | $2.25{\times}10^{-5}$ |

**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 $\varsigma$ are crucial for the power of the test, we report the results for different plausible values of $v$ and $\varsigma$, 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 colspan="6">Mean Absolute Errors</th></tr>
<tr><th>$k$</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>0.00376</td><td>0.00577</td><td>0.00379</td><td>$1.38\times10^{-1}$</td><td>$1.77\times10^{-1}$</td></tr>
<tr><td>13</td><td>0.00373</td><td>0.00662</td><td>0.00384</td><td>$6.96\times10^{-3}$</td><td>$3.09\times10^{-2}$</td></tr>
<tr><td>26</td><td>0.00471</td><td>0.08272</td><td>0.00494</td><td>$1.22\times10^{-2}$</td><td>$1.74\times10^{-6}$</td></tr>
<tr><td>52</td><td>0.00664</td><td>0.01159</td><td>0.00734</td><td>$3.93\times10^{-3}$</td><td>$2.22\times10^{-16}$</td></tr>
</tbody>
</table>

<table>
<thead>
<tr><th colspan="6">Mean Square Errors</th></tr>
<tr><th>$k$</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>$2.16\times10^{-5}$</td><td>$4.97\times10^{-5}$</td><td>$2.19\times10^{-5}$</td><td>$1.96\times10^{-1}$</td><td>$1.99\times10^{-1}$</td></tr>
<tr><td>13</td><td>$2.28\times10^{-5}$</td><td>$6.26\times10^{-5}$</td><td>$2.46\times10^{-5}$</td><td>$7.38\times10^{-3}$</td><td>$5.09\times10^{-2}$</td></tr>
<tr><td>26</td><td>$3.09\times10^{-5}$</td><td>$1.00\times10^{-4}$</td><td>$3.41\times10^{-5}$</td><td>$1.53\times10^{-2}$</td><td>$1.74\times10^{-6}$</td></tr>
<tr><td>52</td><td>$9.88\times10^{-5}$</td><td>$2.35\times10^{-4}$</td><td>$1.21\times10^{-4}$</td><td>$1.90\times10^{-3}$</td><td>$2.22\times10^{-14}$</td></tr>
</tbody>
</table>

<table>
<thead>
<tr><th colspan="6">Root Mean Square Errors</th></tr>
<tr><th>$k$</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>0.00433</td><td>0.00620</td><td>0.00436</td><td>$1.07\times10^{-1}$</td><td>$3.00\times10^{-1}$</td></tr>
<tr><td>13</td><td>0.00471</td><td>0.00747</td><td>0.00488</td><td>$4.98\times10^{-3}$</td><td>$2.27\times10^{-2}$</td></tr>
<tr><td>26</td><td>0.00556</td><td>0.00927</td><td>0.00583</td><td>$1.02\times10^{-2}$</td><td>$3.71\times10^{-8}$</td></tr>
<tr><td>52</td><td>0.00953</td><td>0.01477</td><td>0.01060</td><td>$2.14\times10^{-3}$</td><td>$1.47\times10^{-13}$</td></tr>
</tbody>
</table>

(continued ...)

(... Table 4 continued)

*b) Germany*

| | | Mean Absolute Errors | | | |
|---|---|---|---|---|---|
| $k$ | MSVECM | RW | VECM | DM1(A) | DM2(A) |
| 4 | 0.00402 | 0.00640 | 0.00425 | $5.42\times10^{-2}$ | $2.77\times10^{-1}$ |
| 13 | 0.00437 | 0.00826 | 0.00567 | $1.30\times10^{-3}$ | $1.17\times10^{-10}$ |
| 26 | 0.00558 | 0.01216 | 0.00757 | $2.19\times10^{-7}$ | $7.84\times10^{-4}$ |
| 52 | 0.00620 | 0.01607 | 0.01009 | $1.31\times10^{-15}$ | 0 |

| | | Mean Square Errors | | | |
|---|---|---|---|---|---|
| $k$ | MSVECM | RW | VECM | DM1(A) | DM2(A) |
| 4 | $2.39\times10^{-5}$ | $5.47\times10^{-5}$ | $2.63\times10^{-5}$ | $9.77\times10^{-2}$ | $1.51\times10^{-1}$ |
| 13 | $3.06\times10^{-5}$ | $1.00\times10^{-4}$ | $4.86\times10^{-5}$ | $3.77\times10^{-3}$ | $7.27\times10^{-14}$ |
| 26 | $4.48\times10^{-5}$ | $2.06\times10^{-4}$ | $8.01\times10^{-5}$ | $8.25\times10^{-8}$ | $1.50\times10^{-3}$ |
| 52 | $6.92\times10^{-5}$ | $3.81\times10^{-4}$ | $1.69\times10^{-4}$ | $1.03\times10^{-14}$ | $2.44\times10^{-14}$ |

| | | Root Mean Square Errors | | | |
|---|---|---|---|---|---|
| $k$ | MSVECM | RW | VECM | DM1(A) | DM2(A) |
| 4 | 0.00462 | 0.00667 | 0.004781 | $7.93\times10^{-2}$ | $3.75\times10^{-1}$ |
| 13 | 0.00546 | 0.00946 | 0.006863 | $1.08\times10^{-3}$ | $1.81\times10^{-17}$ |
| 26 | 0.00668 | 0.01384 | 0.008882 | $7.06\times10^{-8}$ | $3.37\times10^{-4}$ |
| 52 | 0.00809 | 0.01909 | 0.012876 | $7.08\times10^{-12}$ | 0 |

(continued ...)

(... Table 4 continued)

*c) Japan*

| | | | Mean Absolute Errors | | |
|---|---|---|---|---|---|
| $k$ | MSVECM | RW | VECM | DM1(A) | DM2(A) |
| 4 | 0.00208 | 0.00320 | 0.00220 | $1.29 \times 10^{-2}$ | $5.95 \times 10^{-1}$ |
| 13 | 0.00324 | 0.00423 | 0.00326 | $1.95 \times 10^{-2}$ | $8.92 \times 10^{-1}$ |
| 26 | 0.00405 | 0.00770 | 0.00468 | $6.10 \times 10^{-6}$ | $2.85 \times 10^{-2}$ |
| 52 | 0.00532 | 0.01795 | 0.00822 | $3.97 \times 10^{-14}$ | 0 |
| | | | Mean Square Errors | | |
| $k$ | MSVECM | RW | VECM | DM1(A) | DM2(A) |
| 4 | $8.01 \times 10^{-6}$ | $1.43 \times 10^{-5}$ | $8.36 \times 10^{-6}$ | $5.91 \times 10^{-2}$ | $8.31 \times 10^{-1}$ |
| 13 | $1.84 \times 10^{-5}$ | $2.87 \times 10^{-5}$ | $1.80 \times 10^{-5}$ | $4.01 \times 10^{-2}$ | $7.92 \times 10^{-1}$ |
| 26 | $2.51 \times 10^{-5}$ | $9.08 \times 10^{-5}$ | $3.46 \times 10^{-5}$ | $5.85 \times 10^{-5}$ | $3.14 \times 10^{-4}$ |
| 52 | $6.03 \times 10^{-5}$ | $5.48 \times 10^{-4}$ | $1.33 \times 10^{-4}$ | $2.74 \times 10^{-15}$ | $1.09 \times 10^{-5}$ |
| | | | Root Mean Square Errors | | |
| $k$ | MSVECM | RW | VECM | DM1(A) | DM2(A) |
| 4 | 0.00264 | 0.00368 | 0.00281 | $4.40 \times 10^{-2}$ | $5.45 \times 10^{-1}$ |
| 13 | 0.00421 | 0.00529 | 0.00420 | $4.60 \times 10^{-2}$ | $9.64 \times 10^{-1}$ |
| 26 | 0.00500 | 0.00942 | 0.00588 | $2.92 \times 10^{-7}$ | $5.63 \times 10^{-4}$ |
| 52 | 0.00745 | 0.02301 | 0.01125 | 0 | 0 |

(continued ...)

(... Table 4 continued)

*d) United Kingdom*

| | | Mean Absolute Errors | | | |
|---|---|---|---|---|---|
| $k$ | MSVECM | RW | VECM | DM1(A) | DM2(A) |
| 4 | 0.00445 | 0.00617 | 0.00469 | $1.80{\times}10^{-2}$ | $3.42{\times}10^{-1}$ |
| 13 | 0.00254 | 0.00477 | 0.00287 | $1.38{\times}10^{-3}$ | $1.48{\times}10^{-1}$ |
| 26 | 0.00429 | 0.00593 | 0.00534 | $5.15{\times}10^{-3}$ | $6.92{\times}10^{-7}$ |
| 52 | 0.01352 | 0.01475 | 0.01673 | $5.13{\times}10^{-1}$ | 0 |
| | | Mean Square Errors | | | |
| $k$ | MSVECM | RW | VECM | DM1(A) | DM2(A) |
| 4 | $2.59{\times}10^{-5}$ | $4.97{\times}10^{-5}$ | $2.84{\times}10^{-5}$ | $2.72{\times}10^{-2}$ | $3.93{\times}10^{-1}$ |
| 13 | $1.21{\times}10^{-5}$ | $3.08{\times}10^{-5}$ | $1.33{\times}10^{-5}$ | $5.16{\times}10^{-3}$ | $2.89{\times}10^{-1}$ |
| 26 | $3.02{\times}10^{-5}$ | $4.97{\times}10^{-5}$ | $4.96{\times}10^{-5}$ | $2.00{\times}10^{-2}$ | $3.21{\times}10^{-7}$ |
| 52 | $3.50{\times}10^{-4}$ | $4.30{\times}10^{-4}$ | $5.08{\times}10^{-4}$ | $3.28{\times}10^{-1}$ | 0 |
| | | Root Mean Square Errors | | | |
| $k$ | MSVECM | RW | VECM | DM1(A) | DM2(A) |
| 4 | 0.00497 | 0.00677 | 0.00522 | $2.11{\times}10^{-2}$ | $2.99{\times}10^{-1}$ |
| 13 | 0.00346 | 0.00543 | 0.00362 | $2.12{\times}10^{-3}$ | $3.56{\times}10^{-1}$ |
| 26 | 0.00546 | 0.00694 | 0.00697 | $1.22{\times}10^{-2}$ | $7.51{\times}10^{-13}$ |
| 52 | 0.01867 | 0.00694 | 0.02251 | $4.40{\times}10^{-1}$ | 0 |

**Notes:** MSVECM, RW and VECM are the average level of the (absolute or square) forecast error obtained by the MSIH-VECM, Random Walk and linear VECM respectively. The forecast errors are obtained by recursive estimation of out-of-sample dynamic forecast 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 MSIH-VECM model with the ones obtained by a driftless Random Walk. DM2(A) is the Diebold-Mariano statistic comparing the forecast errors of the MSIH-VECM model with the ones obtained by the linear VECM. 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 $10^{-100}$, which are considered as virtually zero.

## References

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

Berkson, J. (1938), "Some Difficulties of Interpretation Encountered in the Application of the Chi-Square Test", *Journal of the American Statistical Association*, 53, 28-38.

Beveridge, S. and Nelson, C.R. (1981), "A New Approach to Decomposition of Economic Time Series into permanent and Transitory Components with Particular Attention to Measurement of the 'Business Cycle", *Journal of Monetary Economics*, 7, 151-74.

Bilson, J.F.O. (1981), "The "Speculative Efficiency" Hypothesis", *Journal of Business*, 54, 435-51.

Boothe, P.M. and Glassman, D.A. (1987), "The Statistical Distribution of Exchange Rates: Empirical Evidence and Economic Implications", *Journal of International Economics*, 22, 297-319.

Brock, W.A., Dechert, W.D. and Scheinkman, J. (1991), *Nonlinear Dynamics, Chaos, and Instability*, Cambridge, Mass.: MIT Press.

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

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

Cheung, Y.W. and Lai, K.S. (1993), "Finite-Sample Sizes of Johansen's Likelihood Ratio Tests for Cointegration", *Oxford Bulletin of Economics and Statistics*, 55, 313-328.

Clarida, R.H. and Taylor, M.P. (1993), "The Term Structure of Forward Exchange Premiums and The Forecastability of Spot Exchange Rates: Correcting the Errors", National Bureau of Economic Research Working Paper No. 4442.

Clarida, R.H. and Taylor, M.P. (1997), "The Term Structure of Forward Exchange Premiums and The Forecastability of Spot Exchange Rates: Correcting the Errors", *Review of Economics and Statistics*, 89, 353-361.

Clark, T.E. and McCracken, M.W. (2000), "Tests of Equal Forecast Accuracy and Encompassing for Nested Models," *Journal of Econometrics*, 105, 85-110.

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

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

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

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

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

Diebold, F.X. and Nason, J.A. (1990), "Nonparametric Exchange Rate Prediction?", *Journal of International Economics*, 28, pp. 315-332.

Engel, C. (1994), "Can the Markov Switching Model Forecast Exchange Rates?", *Journal of International Economics*, 36, 151-165.

Engel, C. and Hakkio, C.S. (1996), "The Distribution of Exchange Rates in the EMS", *International Journal of Finance and Economics*, 1, 55-67.

Engel, C. and Hamilton, J.D. (1990), "Long Swings in the Dollar: Are They in the Data and Do Markets Know It?", *American Economic Review*, 80, 689-713.

Engel, C. and Kim, C.J. (1999), "The Long-Run U.S./U.K. Real Exchange Rate", *Journal of Money, Credit, and Banking*, 31, 335-356.

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

Evans, M.D.D. and Lyons, R.K. (1999), "Order Flow and Exchange Rate Dynamics," National Bureau of Economic Research Working Paper No. 7317, *Journal of Political Economy*, forthcoming.

Frankel, J.A. (1980), "Tests of Rational Expectations in the Forward Exchange Market", *Southern Economic Journal*, 46, 1083-1101.

Frankel, J.A. and Rose, A. (1995), "Empirical Research on Nominal Exchange Rates", in Grossman, G.M. and Rogoff, K., (eds.), *Handbook of International Economics*, Vol. 3, 1689-1729.

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

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

Garcia, R. (1988), "Asymptotic Null Distribution of the Likelihood Ratio Test in Markov Switching Models", *International Economic Review*, 39, 763-788.

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

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

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

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

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

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

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

Hamilton, J.D. (1994), *Time Series Analysis*, Princeton, NJ: Princeton University Press

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

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

Hansen, L.P., Hodrick, R.J. (1980), "Forward Exchange Rates as Optimal Predictors of Future Spot Rates: An Econometric Analysis", *Journal of Political Economy*, 88, 5, 829-853.

Hansen, L.P. and Johansen, S. (1993), "Recursive Estimation in Cointegrated VAR Models", 1993, University of Copenhagen, Institute of Mathematical Statistics Working Paper 93-1.

Harvey, D., Leibourne, S. and Newbold, P. (1997), "Testing the Equality of Prediction Mean Squared Errors", *International Journal of Forecasting*, 13, 281-291.

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

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

Johansen, S. (1995), "Identifying Restrictions of Linear Equations with Applications to Simultaneous Equations and Cointegration", *Journal of Econometrics*, 69, 111-132.

Karlsen, H.A. (1990), *A Class of Non-Linear Time Series Models*, University of Bergen, unpublished Ph.D. dissertation

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

Kilian, L. and Taylor, M.P. (2000), "Why Is It so Difficult to Beat the Random Walk Forecast of Exchange Rates?", University of Michigan and University of Warwick, mimeo.

Kim, C.-J. and Nelson, C.R. (1999), "State-Space Models with Regime Switching," Cambridge, Mass. and London, UK: MIT Press.

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

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

Leamer, E.E. (1978), *Specification Searches: Ad Hoc Inference with Nonexperimental Data*. New York: John Wiley and Sons.

LeBaron, B. (1992), "Forecast Improvements Using a Volatility Index", *Journal of Applied Econometrics*, 7, S137-49.

Lewis, K.K. (1995), "Puzzles in International Financial Markets", in Grossman, G.M. and Rogoff, K., (eds.), *Handbook of International Economics*, Vol. 3, 1913-1971.

Luintel, K.B. and Paudyal, K. (1998), "Common Stochastic Trends between Forward and Spot Exchange Rates", *Journal of International Money and Finance*, 17, 279-297.

Lyons, R.K. (1999), *The Microstructure Approach to Exchange Rates*, MIT Press, forthcoming.

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

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

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

Meese, R.A. and Rogoff, K. (1983b), "The Out-of-sample Failure of Empirical Exchange Rate Models: Sampling Error or Misspecification?", in Frenkel, J.A. (ed.), *Exchange Rate and International Economics*, Chicago: University of Chicago Press.

Meese, R.A. and Rogoff, K. (1988), "Was It Real? The Exchange Rate-Interest Differential Relation over the Modern Floating-Rate Period", *Journal of Finance*, 43, 933-948.

Meese, R.A. and Rose, A.K (1990), "Nonlinear, Nonparametric, Nonessential Exchange Rate Estimation", *American Economic Review*, 80, 192-196.

Meese, R.A. and Rose, A.K (1991), "An Empirical Assessment of Non-linearities in Models of Exchange Rate Determination", *Review of Economic Studies*, 58, 603-619.

Naka, A. and Whitney, G. (1995), "The Unbiased Forward Rate Hypothesis Re-examined", *Journal of International Money and Finance*, 14, 857-867.

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

Osterwald-Lenum, M. (1992), "A Note with Quantiles of the Asymptotic Distribution of the Maximum Likelihood Cointegration Rank Test Statistics", *Oxford Bulletin of Economics and Statistics*, 54, 461-472.

Pesaran, M.H. and Skouras, S. (2001), "Decision-based Methods for Forecast Evaluation," in Clements, M.P. and Hendry, D.F. (eds.), *A Companion to Economic Forecasting*, Oxford: Blackwell, forthcoming.

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

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

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

Sarno, L. and Taylor, M.P. (2001), *The Microstructure of the Foreign Exchange Market*, Princeton Studies in International Economics, 89, International Economics Section, Princeton University.

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

Tay, A.S. and Wallis, K.F. (2000), "Density Forecasting: A Survey," *Journal of Forecasting*, 19, pp. 235-254.

Taylor, M.P. (1987), "Covered Interest Parity: A High-Frequency, High-Quality Data Study," *Economica*, 54, 429-438.

Taylor, M.P. (1989), "Covered Interest Arbitrage and Market Turbulence," *Economic Journal*, 99, 376-391.

Taylor, M.P. (1995), "The Economics of Exchange Rates", *Journal of Economic Literature*, 33, 13-47.

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

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

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

# A  Appendix:  Markov-Switching-VECM estimation results

**Table A1. 'Bottom-up' identification procedure**

|  | $LR1$ | $LR2$ | Davies |
|---|---|---|---|
| France | 5007.83 {0} | 0.4322  {0.999} | 1305.10 |
| Germany | 3158.56 {0} | 21.7836 {0.648} | 440.47 |
| Japan | 3320.89 {0} | 15.5266 {0.928} | 487.27 |
| United Kingdom | 1668.53 {0} | 35.0404 {0.128} | 477.66 |

**Notes:** $LR1$ is a test statistic of the null hypothesis of no regime dependent variance-covariance matrix (ie. 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 existance 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.

**Table A2.   MSIH(3)-VECM(1) estimation results**

a) *France*

| *Panel A:* |
|---|

$$\widetilde{\Gamma}_1 = \begin{bmatrix} 0.089 & -0.844 & 1.094 & -0.291 & -0.048 \\ (0.23) & (0.39) & (0.19) & (0.42) & (0.26) \\ 0.090 & -0.865 & 1.148 & -0.322 & -0.049 \\ (0.23) & (0.38) & (0.17) & (0.41) & (0.26) \\ 0.072 & -0.867 & 1.178 & -0.319 & -0.063 \\ (0.23) & (0.38) & (0.15) & (0.40) & (0.26) \\ 0.039 & -0.875 & 1.263 & -0.344 & -0.080 \\ (0.23) & (0.37) & (0.17) & (0.40) & (0.26) \\ 0.088 & -0.915 & 0.941 & 0.150 & -0.265 \\ (0.22) & (0.38) & (0.26) & (0.41) & (0.26) \end{bmatrix} ;$$

$$\widetilde{v}(z_1) = \begin{bmatrix} -1.15 \times 10^{-3} \\ (1.12 \times 10^{-3}) \\ -1.44 \times 10^{-3} \\ (1.14 \times 10^{-3}) \\ -1.46 \times 10^{-3} \\ (1.17 \times 10^{-3}) \\ -1.41 \times 10^{-3} \\ (1.29 \times 10^{-3}) \\ -1.59 \times 10^{-3} \\ (1.50 \times 10^{-3}) \end{bmatrix} ; \widetilde{v}(z_2) = \begin{bmatrix} 1.12 \times 10^{-4} \\ (1.49 \times 10^{-5}) \\ 1.25 \times 10^{-4} \\ (8.81 \times 10^{-6}) \\ 1.25 \times 10^{-4} \\ (2.28 \times 10^{-5}) \\ 1.20 \times 10^{-4} \\ (4.48 \times 10^{-5}) \\ 1.25 \times 10^{-4} \\ (7.91 \times 10^{-5}) \end{bmatrix} ; \widetilde{v}(z_3) = \begin{bmatrix} -6.49 \times 10^{-5} \\ (5.92 \times 10^{-4}) \\ -1.06 \times 10^{-5} \\ (5.87 \times 10^{-4}) \\ -1.34 \times 10^{-5} \\ (5.64 \times 10^{-4}) \\ -2.07 \times 10^{-5} \\ (5.35 \times 10^{-4}) \\ 1.27 \times 10^{-5} \\ (5.09 \times 10^{-4}) \end{bmatrix} ;$$

$$\widetilde{\alpha} = \begin{bmatrix} -0.708 & 0.989 & -0.422 & 0.013 \\ (0.50) & (0.20) & (0.20) & (0.11) \\ -0.487 & 0.931 & -0.440 & 0.018 \\ (0.50) & (0.19) & (0.20) & (0.11) \\ -0.918 & 1.245 & -0.500 & 0.004 \\ (0.49) & (0.16) & (0.19) & (0.11) \\ -0.850 & 0.906 & -0.191 & 0.075 \\ (0.50) & (0.16) & (0.20) & (0.11) \\ -0.796 & 0.951 & -0.263 & 0.050 \\ (0.53) & (0.22) & (0.23) & (0.12) \end{bmatrix} ;$$

(continued)

$$\widetilde{\sum}(z_1) = \begin{bmatrix} 8.05 \times 10^{-5} \\ 7.60 \times 10^{-5} & 8.59 \times 10^{-5} \\ 7.24 \times 10^{-5} & 8.48 \times 10^{-5} & 9.14 \times 10^{-5} \\ 7.19 \times 10^{-5} & 8.66 \times 10^{-5} & 9.95 \times 10^{-5} & 1.15 \times 10^{-4} \\ 7.11 \times 10^{-5} & 8.91 \times 10^{-5} & 1.08 \times 10^{-4} & 1.32 \times 10^{-4} & 1.63 \times 10^{-4} \end{bmatrix};$$

$$\widetilde{\sum}(z_2) = \begin{bmatrix} 4.03 \times 10^{-5} \\ 4.02 \times 10^{-5} & 4.01 \times 10^{-5} \\ 4.00 \times 10^{-5} & 3.99 \times 10^{-5} & 3.97 \times 10^{-5} \\ 3.96 \times 10^{-5} & 3.96 \times 10^{-5} & 3.94 \times 10^{-5} & 3.92 \times 10^{-5} \\ 3.89 \times 10^{-5} & 3.89 \times 10^{-5} & 3.88 \times 10^{-5} & 3.86 \times 10^{-5} & 3.82 \times 10^{-5} \end{bmatrix};$$

$$\widetilde{\sum}(z_3) = \begin{bmatrix} 4.55 \times 10^{-5} \\ 4.49 \times 10^{-5} & 4.44 \times 10^{-5} \\ 4.34 \times 10^{-5} & 4.31 \times 10^{-5} & 4.22 \times 10^{-5} \\ 4.15 \times 10^{-5} & 4.13 \times 10^{-5} & 4.07 \times 10^{-5} & 3.97 \times 10^{-5} \\ 3.96 \times 10^{-5} & 3.96 \times 10^{-5} & 3.93 \times 10^{-5} & 3.88 \times 10^{-5} & 3.88 \times 10^{-5} \end{bmatrix};$$

$$\widetilde{\mathbf{P}} = \begin{bmatrix} 0.68 & 0.01 & 0.06 \\ 0.07 & 0.93 & 0.09 \\ 0.25 & 0.06 & 0.85 \end{bmatrix}; \widetilde{\xi} = \begin{bmatrix} 0.09 \\ 0.55 \\ 0.36 \end{bmatrix};$$

$\rho(A) = 0.1515$
LR linearity test: 2430.01

**Notes:** Tildes denote estimated values obtained using the EM algorithm for maximum likelihood (Dempster, Laird and Rubin, 1977). Figures in parentheses are asymptotic standard errors. Symbols are defined as in equation (12). $P$ and $\xi$ denote the $M \times M$ transition matrix and the $M$-dimensional ergodic probabilities vector respectively. $\rho(A)$ is the spectral radius of the matrix $A$ calculated as in Karlsen (1990). It can be tought as a measure of stationarity of the MS-VECM. The LR linearity test is a Davies (1987)-type test testing the hypothesis that the true model is a linear VECM against the alternative of an MSIH-VECM.

*b) Germany*

---

*Panel A:*

---

$$
\widetilde{\Gamma}_1 =
\begin{bmatrix}
-2.972 & 3.440 & 1.112 & -1.452 & -0.145 \\
(0.14) & (0.16) & (0.12) & (0.09) & (0.11) \\
-2.939 & 3.362 & 1.159 & -1.438 & -0.160 \\
(0.12) & (0.13) & (0.09) & (0.08) & (0.11) \\
-2.922 & 3.341 & 1.142 & -1.401 & -0.175 \\
(0.10) & (0.08) & (0.05) & (0.05) & (0.11) \\
-2.911 & 3.193 & 1.358 & -1.466 & -0.187 \\
(0.10) & (0.10) & (0.06) & (0.06) & (0.11) \\
-2.549 & 2.724 & 1.245 & -1.170 & -0.266 \\
(0.20) & (0.27) & (0.20) & (0.15) & (0.13)
\end{bmatrix} ;
$$

$$
\widetilde{v}(z_1) =
\begin{bmatrix}
-2.95 \times 10^{-3} \\
\left(1.00 \times 10^{-3}\right) \\
-2.90 \times 10^{-3} \\
\left(9.98 \times 10^{-4}\right) \\
-2.90 \times 10^{-3} \\
\left(9.86 \times 10^{-4}\right) \\
-2.89 \times 10^{-3} \\
\left(1.01 \times 10^{-3}\right) \\
-3.00 \times 10^{-3} \\
\left(1.15 \times 10^{-3}\right)
\end{bmatrix} ;
\widetilde{v}(z_2) =
\begin{bmatrix}
-1.01 \times 10^{-3} \\
\left(1.66 \times 10^{-4}\right) \\
-9.79 \times 10^{-4} \\
\left(1.70 \times 10^{-4}\right) \\
-9.96 \times 10^{-4} \\
\left(1.61 \times 10^{-4}\right) \\
-9.43 \times 10^{-4} \\
\left(1.54 \times 10^{-4}\right) \\
-9.27 \times 10^{-4} \\
\left(1.58 \times 10^{-4}\right)
\end{bmatrix} ;
\widetilde{v}(z_3) =
\begin{bmatrix}
-5.47 \times 10^{-4} \\
\left(1.34 \times 10^{-5}\right) \\
-5.27 \times 10^{-4} \\
\left(6.47 \times 10^{-6}\right) \\
-5.12 \times 10^{-4} \\
\left(9.70 \times 10^{-6}\right) \\
-5.09 \times 10^{-4} \\
\left(1.91 \times 10^{-5}\right) \\
-5.11 \times 10^{-4} \\
\left(4.21 \times 10^{-5}\right)
\end{bmatrix} ;
$$

$$
\widetilde{\alpha} =
\begin{bmatrix}
7.641 & -4.107 & -0.235 & 0.570 \\
(0.15) & (0.06) & (0.06) & (0.02) \\
7.973 & -4.248 & -0.223 & 0.571 \\
(0.11) & (0.05) & (0.06) & (0.02) \\
7.445 & -3.874 & -0.292 & 0.554 \\
(0.05) & (0.06) & (0.06) & (0.02) \\
7.525 & -4.100 & -0.092 & 0.502 \\
(0.09) & (0.14) & (0.09) & (0.02) \\
7.584 & -3.903 & -0.276 & 0.541 \\
(0.27) & (0.27) & (0.16) & (0.04)
\end{bmatrix} ;
$$

---

(continued)

$$\widetilde{\sum}(z_1) = \begin{bmatrix} 9.20 \times 10^{-5} \\ 9.06 \times 10^{-5} & 8.98 \times 10^{-5} \\ 8.86 \times 10^{-5} & 8.89 \times 10^{-5} & 8.98 \times 10^{-5} \\ 8.58 \times 10^{-5} & 8.73 \times 10^{-5} & 9.05 \times 10^{-5} & 9.45 \times 10^{-5} \\ 8.29 \times 10^{-5} & 8.63 \times 10^{-5} & 9.29 \times 10^{-5} & 1.01 \times 10^{-4} & 1.17 \times 10^{-4} \end{bmatrix};$$

$$\widetilde{\sum}(z_2) = \begin{bmatrix} 3.85 \times 10^{-5} \\ 3.85 \times 10^{-5} & 3.84 \times 10^{-5} \\ 3.81 \times 10^{-5} & 3.81 \times 10^{-5} & 3.78 \times 10^{-5} \\ 3.75 \times 10^{-5} & 3.75 \times 10^{-5} & 3.73 \times 10^{-5} & 3.69 \times 10^{-5} \\ 3.73 \times 10^{-5} & 3.73 \times 10^{-5} & 3.72 \times 10^{-5} & 3.70 \times 10^{-5} & 3.74 \times 10^{-5} \end{bmatrix};$$

$$\widetilde{\sum}(z_3) = \begin{bmatrix} 4.36 \times 10^{-5} \\ 4.36 \times 10^{-5} & 4.36 \times 10^{-5} \\ 4.36 \times 10^{-5} & 4.37 \times 10^{-5} & 4.37 \times 10^{-5} \\ 4.36 \times 10^{-5} & 4.36 \times 10^{-5} & 4.37 \times 10^{-5} & 4.37 \times 10^{-5} \\ 4.35 \times 10^{-5} & 4.36 \times 10^{-5} & 4.36 \times 10^{-5} & 4.37 \times 10^{-5} & 4.39 \times 10^{-5} \end{bmatrix};$$

$$\widetilde{\mathbf{P}} = \begin{bmatrix} 0.57 & 0.11 & 0.01 \\ 0.36 & 0.79 & 0.04 \\ 0.07 & 0.10 & 0.95 \end{bmatrix}; \widetilde{\xi} = \begin{bmatrix} 0.08 \\ 0.28 \\ 0.64 \end{bmatrix};$$

$\rho(A) = 0.1243$
LR linearity test: 3586.70

**Notes:** Tildes denote estimated values obtained using the EM algorithm for maximum likelihood (Dempster, Laird and Rubin, 1977). Figures in parentheses are asymptotic standard errors. Symbols are defined as in equation (12). $P$ and $\xi$ denote the $M \times M$ transition matrix and the $M$-dimensional ergodic probabilities vector respectively. $\rho(A)$ is the spectral radius of the matrix $A$ calculated as in Karlsen (1990). It can be tought as a measure of stationarity of the MS-VECM. The LR linearity test is a Davies (1987)-type test testing the hypothesis that the true model is a linear VECM against the alternative of an MSIH-VECM.

*c) Japan*

| Panel A: |
|---|

$$\widetilde{\Gamma}_1 = \begin{bmatrix} -1.627 & 1.947 & 0.499 & -0.788 & 0.021 \\ (0.11) & (0.13) & (0.10) & (0.09) & (0.05) \\ -1.574 & 1.939 & 0.447 & -0.796 & 0.037 \\ (0.09) & (0.11) & (0.08) & (0.07) & (0.05) \\ -1.652 & 2.143 & 0.303 & -0.791 & 0.050 \\ (0.06) & (0.06) & (0.05) & (0.05) & (0.05) \\ -1.617 & 2.110 & 0.351 & -0.865 & 0.072 \\ (0.06) & (0.07) & (0.05) & (0.04) & (0.05) \\ -1.551 & 2.154 & -0.021 & -0.504 & -0.028 \\ (0.16) & (0.23) & (0.18) & (0.17) & (0.07) \end{bmatrix} ;$$

$$\widetilde{v}(z_1) = \begin{bmatrix} -1.25 \times 10^{-4} \\ (1.93 \times 10^{-4}) \\ -1.46 \times 10^{-4} \\ (1.60 \times 10^{-4}) \\ -2.28 \times 10^{-4} \\ (1.69 \times 10^{-4}) \\ -3.35 \times 10^{-4} \\ (3.23 \times 10^{-4}) \\ -5.58 \times 10^{-4} \\ (6.87 \times 10^{-4}) \end{bmatrix} ; \widetilde{v}(z_2) = \begin{bmatrix} 3.08 \times 10^{-4} \\ (1.62 \times 10^{-5}) \\ 3.09 \times 10^{-4} \\ (1.45 \times 10^{-5}) \\ 3.01 \times 10^{-4} \\ (1.74 \times 10^{-5}) \\ 3.52 \times 10^{-4} \\ (4.40 \times 10^{-5}) \\ 3.95 \times 10^{-4} \\ (8.67 \times 10^{-4}) \end{bmatrix} ; \widetilde{v}(z_3) = \begin{bmatrix} 2.61 \times 10^{-4} \\ (4.49 \times 10^{-4}) \\ 2.75 \times 10^{-4} \\ (4.47 \times 10^{-4}) \\ 2.28 \times 10^{-4} \\ (4.36 \times 10^{-4}) \\ 2.93 \times 10^{-4} \\ (4.36 \times 10^{-4}) \\ 2.58 \times 10^{-4} \\ (4.68 \times 10^{-4}) \end{bmatrix} ;$$

$$\widetilde{\alpha} = \begin{bmatrix} 1.309 & -1.133 & 0.476 & 0.015 \\ (0.12) & (0.12) & (0.13) & (0.10) \\ 1.848 & -1.360 & 0.462 & 0.034 \\ (0.09) & (0.11) & (0.11) & (0.10) \\ 1.579 & -1.087 & 0.340 & 0.050 \\ (0.07) & (0.11) & (0.10) & (0.09) \\ 1.913 & -1.546 & 0.580 & 0.017 \\ (0.09) & (0.12) & (0.10) & (0.09) \\ 2.470 & -1.681 & 0.384 & 0.109 \\ (0.18) & (0.16) & (0.16) & (0.10) \end{bmatrix} ;$$

(continued)

$$
\widetilde{\textstyle\sum}(z_1) = \begin{bmatrix}
8.02 \times 10^{-5} & & & & \\
7.99 \times 10^{-5} & 8.06 \times 10^{-5} & & & \\
7.93 \times 10^{-5} & 8.21 \times 10^{-5} & 8.77 \times 10^{-5} & & \\
7.75 \times 10^{-5} & 8.34 \times 10^{-5} & 9.52 \times 10^{-5} & 1.12 \times 10^{-4} & \\
7.44 \times 10^{-5} & 8.63 \times 10^{-5} & 1.10 \times 10^{-4} & 1.45 \times 10^{-4} & 2.17 \times 10^{-4}
\end{bmatrix};
$$

$$
\widetilde{\textstyle\sum}(z_2) = \begin{bmatrix}
3.50 \times 10^{-5} & & & & \\
3.50 \times 10^{-5} & 3.49 \times 10^{-5} & & & \\
3.49 \times 10^{-5} & 3.49 \times 10^{-5} & 3.48 \times 10^{-5} & & \\
3.50 \times 10^{-5} & 3.50 \times 10^{-5} & 3.49 \times 10^{-5} & 3.49 \times 10^{-5} & \\
3.49 \times 10^{-5} & 3.50 \times 10^{-5} & 3.50 \times 10^{-5} & 3.50 \times 10^{-5} & 3.54 \times 10^{-5}
\end{bmatrix};
$$

$$
\widetilde{\textstyle\sum}(z_3) = \begin{bmatrix}
3.55 \times 10^{-5} & & & & \\
3.47 \times 10^{-5} & 3.43 \times 10^{-5} & & & \\
3.33 \times 10^{-5} & 3.33 \times 10^{-5} & 3.30 \times 10^{-5} & & \\
3.19 \times 10^{-5} & 3.23 \times 10^{-5} & 3.27 \times 10^{-5} & 3.34 \times 10^{-5} & \\
3.13 \times 10^{-5} & 3.23 \times 10^{-5} & 3.34 \times 10^{-5} & 3.50 \times 10^{-5} & 3.86 \times 10^{-5}
\end{bmatrix};
$$

$$
\widetilde{\mathbf{P}} = \begin{bmatrix}
0.47 & 0.08 & 0.05 \\
0.49 & 0.90 & 0.04 \\
0.04 & 0.02 & 0.91
\end{bmatrix}; \widetilde{\xi} = \begin{bmatrix}
0.12 \\
0.68 \\
0.20
\end{bmatrix};
$$

$\rho(A) = 0.1594$
LR linearity test: 4110.68

**Notes:** Tildes denote estimated values obtained using the EM algorithm for maximum likelihood (Dempster, Laird and Rubin, 1977). Figures in parentheses are asymptotic standard errors. Symbols are defined as in equation (12). $P$ and $\xi$ denote the $M \times M$ transition matrix and the $M$-dimensional ergodic probabilities vector respectively. $\rho(A)$ is the spectral radius of the matrix $A$ calculated as in Karlsen (1990). It can be tought as a measure of stationarity of the MS-VECM. The LR linearity test is a Davies (1987)-type test testing the hypothesis that the true model is a linear VECM against the alternative of an MSIH-VECM.

*d) United Kingdom*

---

*Panel A:*

---

$$
\widetilde{\Gamma}_1 = \begin{bmatrix}
-0.472 & 0.839 & 0.031 & 0.042 & -0.438 \\
(0.34) & (0.17) & (0.31) & (0.30) & (0.28) \\
-0.422 & 0.704 & 0.165 & -0.002 & -0.443 \\
(0.34) & (0.14) & (0.30) & (0.30) & (0.28) \\
-0.471 & 0.722 & 0.245 & -0.061 & -0.434 \\
(0.33) & (0.09) & (0.28) & (0.29) & (0.28) \\
-0.498 & 0.562 & 0.650 & -0.323 & -0.389 \\
(0.33) & (0.11) & (0.28) & (0.29) & (0.28) \\
-0.435 & 0.350 & 0.763 & -0.286 & -0.392 \\
(0.36) & (0.28) & (0.35) & (0.33) & (0.28)
\end{bmatrix} ;
$$

$$
\widetilde{v}(z_1) = \begin{bmatrix}
-6.09 \times 10^{-4} \\
(1.31 \times 10^{-3}) \\
-6.18 \times 10^{-4} \\
(1.31 \times 10^{-3}) \\
-6.11 \times 10^{-4} \\
(1.27 \times 10^{-3}) \\
-5.30 \times 10^{-4} \\
(1.20 \times 10^{-3}) \\
-4.18 \times 10^{-4} \\
(1.12 \times 10^{-3})
\end{bmatrix} ;
\widetilde{v}(z_2) = \begin{bmatrix}
2.30 \times 10^{-4} \\
(1.10 \times 10^{-5}) \\
2.38 \times 10^{-4} \\
(8.37 \times 10^{-6}) \\
2.55 \times 10^{-4} \\
(8.57 \times 10^{-6}) \\
2.58 \times 10^{-4} \\
(1.58 \times 10^{-5}) \\
2.45 \times 10^{-4} \\
(2.52 \times 10^{-5})
\end{bmatrix} ;
\widetilde{v}(z_3) = \begin{bmatrix}
-3.71 \times 10^{-5} \\
(2.85 \times 10^{-5}) \\
-3.83 \times 10^{-5} \\
(3.01 \times 10^{-5}) \\
-4.80 \times 10^{-5} \\
(2.83 \times 10^{-5}) \\
-7.45 \times 10^{-5} \\
(4.37 \times 10^{-5}) \\
-9.39 \times 10^{-5} \\
(7.30 \times 10^{-5})
\end{bmatrix} ;
$$

$$
\widetilde{\alpha} = \begin{bmatrix}
-1.379 & 1.713 & -0.861 & 0.178 \\
(0.40) & (0.31) & (0.32) & (0.06) \\
-1.060 & 1.623 & -0.889 & 0.188 \\
(0.39) & (0.30) & (0.31) & (0.05) \\
-1.593 & 1.980 & -0.974 & 0.185 \\
(0.37) & (0.29) & (0.31) & (0.05) \\
-1.612 & 1.793 & -0.789 & 0.138 \\
(0.37) & (0.29) & (0.31) & (0.05) \\
-1.556 & 1.962 & -1.024 & 0.215 \\
(0.43) & (0.35) & (0.33) & (0.06)
\end{bmatrix} ;
$$

---

(continued)

$$\widetilde{\sum}(z_1) = \begin{bmatrix} 5.62 \times 10^{-5} & & & & \\ 5.52 \times 10^{-5} & 5.44 \times 10^{-5} & & & \\ 5.30 \times 10^{-5} & 5.25 \times 10^{-5} & 5.12 \times 10^{-5} & & \\ 4.99 \times 10^{-5} & 4.97 \times 10^{-5} & 4.87 \times 10^{-5} & 4.71 \times 10^{-5} & \\ 4.60 \times 10^{-5} & 4.60 \times 10^{-5} & 4.55 \times 10^{-5} & 4.45 \times 10^{-5} & 4.29 \times 10^{-5} \end{bmatrix};$$

$$\widetilde{\sum}(z_2) = \begin{bmatrix} 2.53 \times 10^{-5} & & & & \\ 2.53 \times 10^{-5} & 2.53 \times 10^{-5} & & & \\ 2.52 \times 10^{-5} & 2.52 \times 10^{-5} & 2.50 \times 10^{-5} & & \\ 2.49 \times 10^{-5} & 2.49 \times 10^{-5} & 2.48 \times 10^{-5} & 2.46 \times 10^{-5} & \\ 2.44 \times 10^{-5} & 2.44 \times 10^{-5} & 2.43 \times 10^{-5} & 2.42 \times 10^{-5} & 2.38 \times 10^{-5} \end{bmatrix};$$

$$\widetilde{\sum}(z_3) = \begin{bmatrix} 5.72 \times 10^{-5} & & & & \\ 5.73 \times 10^{-5} & 5.74 \times 10^{-5} & & & \\ 5.77 \times 10^{-5} & 5.78 \times 10^{-5} & 5.83 \times 10^{-5} & & \\ 5.84 \times 10^{-5} & 5.86 \times 10^{-5} & 5.92 \times 10^{-5} & 6.02 \times 10^{-5} & \\ 5.93 \times 10^{-5} & 5.95 \times 10^{-5} & 6.03 \times 10^{-5} & 6.14 \times 10^{-5} & 6.31 \times 10^{-5} \end{bmatrix};$$

$$\widetilde{\mathbf{P}} = \begin{bmatrix} 0.81 & 0.01 & 0.05 \\ 0.02 & 0.88 & 0.07 \\ 0.17 & 0.10 & 0.88 \end{bmatrix}; \widetilde{\xi} = \begin{bmatrix} 0.15 \\ 0.33 \\ 0.51 \end{bmatrix};$$

$\rho(A) = 0.1876$
LR linearity test: 2090.66

**Notes:** Tildes denote estimated values obtained using the EM algorithm for maximum likelihood (Dempster, Laird and Rubin, 1977). Figures in parentheses are asymptotic standard errors. Symbols are defined as in equation (12). $P$ and $\xi$ denote the $M \times M$ transition matrix and the $M$-dimensional ergodic probabilities vector respectively. $\rho(A)$ is the spectral radius of the matrix $A$ calculated as in Karlsen (1990). It can be tought as a measure of stationarity of the MS-VECM. The LR linearity test is a Davies (1987)-type test testing the hypothesis that the true model is a linear VECM against the alternative of an MSIH-VECM.