Empirical exchange rate models and currency risk: some evidence from density forecasts

Lucio Sarno a,b,*, Giorgio Valente a

aFinance Group, Warwick Business School, University of Warwick, Coventry CV4 7AL, UK
bCentre for Economic Policy Research (CEPR), UK

Abstract

A large literature in exchange rate economics has investigated the forecasting performance of empirical exchange rate models using conventional point forecast accuracy criteria. However, in the context of managing exchange rate risk, interest centers on more than just point forecasts. This paper provides a formal evaluation of recent exchange rate models based on the term structure of forward exchange rates, which previous research has shown to be satisfactory in point forecasting, in terms of density forecasting performance. The economic value of the exchange rate density forecasts is investigated in the context of an application to a simple risk management exercise.

© 2004 Elsevier Ltd. All rights reserved.

JEL classification: F31; F37

Keywords: Foreign exchange; Forecasting; Term structure; Density forecast; Nonlinearity

1. Introduction

Exchange rate risk plays a major role in international portfolio diversification and in several aspects of economic policy, including the assessment of the uncertainty
surrounding prices of exports and imports, the value of international reserves and open positions in foreign currency, and the domestic currency value of debt payments and workers’ remittances which, in turn, may affect domestic wages, prices, output and employment. In international financial markets, expectations of future exchange rates affect agents’ decisions in a number of respects, including their investment, hedging, and borrowing and lending decisions. It is not surprising, therefore, that an enormous empirical literature has developed which focuses on modelling and forecasting nominal exchange rates.

The vast majority of the empirical literature on forecasting exchange rates has centered on forecasting the level of nominal exchange rates. This literature is highly influenced by the seminal work of Meese and Rogoff (1983a,b), who first documented that empirical exchange rate models, based on conventional macroeconomic fundamentals suggested by international macroeconomics theory, cannot outperform a simple no-change or random walk forecast of exchange rates in terms of standard measures of point forecast accuracy. After over 20 years of research since the publication of the Meese–Rogoff studies, their findings remain, with a few exceptions, very robust (e.g. Mark, 1995; Neely and Sarno, 2002).  

While macroeconomic fundamentals do not appear to be useful in forecasting exchange rates, however, models which exploit the information in the term structure of forward exchange rates and forward premia have produced satisfactory results. Clarida and Taylor (1997) first argued that, although the forward exchange rate is not an optimal predictor of the future spot exchange rate (e.g. see Lewis, 1995; Engel, 1996; and the references therein), forward rates may still contain valuable information for forecasting future spot exchange rates. Under the assumption of stationary expected foreign exchange excess returns, Clarida and Taylor (1997) derive a linear vector equilibrium correction model (VECM) of spot and forward exchange rates. Using this linear VECM, Clarida and Taylor show that it is possible to extract sufficient information from the term structure of forward premia to outperform the random walk model for several exchange rates in out-of-sample forecasting. Then, following the large literature showing the existence of regime-switching behavior in exchange rate movements (e.g. Engel and Hamilton, 1990; LeBaron, 1992; Engel, 1994; Engel and Hakkio, 1996), Clarida et al. (2003) generalize the linear VECM of spot and forward exchange rates to a nonlinear, three-regime Markov-switching VECM (MS-VECM) which is found to outperform a random walk as well as to improve on the linear VECM in terms of out-of-sample forecasting.

While an extensive body of literature has investigated the performance of exchange rate models in forecasting the level of the exchange rate, surprisingly little attention has been devoted to forecasting the density of exchange rates. In a

---

1 See, for example, the papers published in the special issue of the *Journal of International Economics* (May 2003) on “Empirical Exchange Rate Models.”
decision-theoretical context, the need to consider the density forecast of a variable\textsuperscript{2} – as opposed to considering only its conditional mean and variance – seems fairly accepted on the basis of the argument that economic agents may have loss functions that do not depend symmetrically on the realizations of future values of potentially non-Gaussian variables (Satchell and Timmermann, 1995; Granger, 2003). In this case, agents are interested in forecasting not only the mean and variance of the variables in question, but their full predictive densities. In various contexts in economics and finance – among which the recent boom in financial risk management is an obvious case – there is strong need to evaluate density forecasts.

Several researchers have proposed methods for evaluating density forecasts. These methods allow us to evaluate a model-based predictive density by measuring the closeness of two density functions or by testing the hypothesis that the predictive density generated by a particular model corresponds to the true predictive density (e.g. Diebold et al., 1998; Berkowitz, 2001).\textsuperscript{3} More recently, this line of research has also provided a test statistic to formally evaluate the relative ability of competing models in matching the true predictive density (Corradi and Swanson, 2004, in press). However, a literature focusing on the performance of empirical models to forecast the density of exchange rates has not emerged to date, and the main focus of the relevant literature remains on point forecasting of the nominal exchange rate.

The present paper contributes to the relevant literature in that we re-examine the forecasting performance of term structure models of exchange rates, which were shown to outperform a random walk in out-of-sample point forecasting by Clarida and Taylor (1997) and Clarida et al. (2003). However, we assess the ability of these models to forecast out-of-sample the one-step-ahead density of nominal exchange rates, hence filling, to some extent, the important gap in the literature described above. Our analysis is carried out using the recent techniques on evaluating density forecasts mentioned above as well as on Value-at-Risk (VaR) calculations.

In particular, using weekly data for eight bilateral dollar exchange rates from January 1985 to December 2003, we focus on the ability of both the linear VECM and the MS-VECM to forecast the one-week-ahead exchange rate density. To anticipate our main results, we find that Markov-switching term structure models of exchange rates produce satisfactory density forecasts of exchange rates. In particular, the MS-VECM of the term structure convincingly outperforms a random walk forecast and a linear term structure VECM in our density forecasting exercise, suggesting that the allowance for nonlinearity in these models may be particularly

\textsuperscript{2} A density forecast (or predictive density) of the realization of a random variable at some future time is an estimate of the probability distribution of the possible future values of that variable. It therefore provides a full description of the uncertainty associated with a forecast, in contrast with a point forecast, which contains no description of the associated uncertainty. For a survey of the literature on density forecasting and a discussion of its applications in macroeconomics and finance, see Tay and Wallis (2000). See also Granger and Pesaran (1999), Granger (2003) and Sarno and Valente (2004, in press).

\textsuperscript{3} By ‘true’ predictive density of the data we mean the density of the data over the chosen forecast period. Therefore, no forecast is in fact carried out in this case, and the term ‘predictive’ simply refers to the fact that the density in question does not refer to the full sample but only to the forecast period. Also note that we use the terms ‘predictive density’ and ‘forecast density’ interchangeably below.
important to produce satisfactory out-of-sample density forecasting performance. Finally, we illustrate the practical importance of our results on density forecasting with a simple application to a risk management exercise. In recent years, trading accounts at large financial institutions have shown a dramatic growth and become increasingly more complex. Partly in response to this trend, major trading institutions have developed risk measurement models designed to manage risk. The most common approach employed in this context is based on the VaR methodology, where VaR is defined as the expected maximum loss over a target horizon within a given confidence interval (Jorion, 2001) – more formally, VaR is an interval forecast, typically a one-sided 95% or 99% interval of the distribution of expected wealth or returns. In our simple application we analyze the out-of-sample forecasting performance of term structure models of exchange rates, investigating the implications of these forecasts for a risk manager who has to quantify the risk associated with holding a currency portfolio over a one-week horizon. This application further illustrates how the MS-VECM captures satisfactorily the higher moments of the predictive density of exchange rates, generating VaRs that estimate the probability of large losses better than the other two competing models. Put another way, our findings indicate that better density forecasts of exchange rates, of the type recorded by the regime-switching model considered in this paper, can potentially lead to substantial improvements in risk management and, more precisely, to better estimates of downside risk.

The remainder of the paper is set out as follows. In Section 2, we briefly review the literature employing term structure models of forward premia to forecast exchange rates, both in a linear framework and in a Markov-switching framework. In Section 3, we describe the data and carry out a preliminary unit root and cointegration analysis of the spot and forward exchange rates data. We also report linearity tests and other tests designed to select the most adequate MS-VECM in our context. We then forecast from these models in Section 4, where we report tests designed to assess the performance of the random walk model, the linear VECM and the MS-VECM in terms of density forecasting. We also explore the implications of the density forecasting results in a risk management exercise. A final section briefly summarizes and concludes.

2. Term structure forecasting models of exchange rates: a brief overview

Let $s_t$ and $f_t^k$ be, respectively, the spot exchange rate and the $k$-period forward exchange rate at time $t$. Under the assumptions that (i) each of $s_t$ and $f_t^k$ are well described by unit root processes and that (ii) departures from the risk-neutral efficient markets hypothesis – namely expected foreign exchange excess returns, $f_t^k - E_r(s_{t+k} | \Omega_t)$, defined with respect to a given information set $\Omega_t$ – are stationary, it is straightforward to derive an expression which implies that the forward premium, $f_t^k - s_t$, is stationary (Clarida and Taylor, 1997). In turn, this result implies that forward and spot exchange rates have a common stochastic trend and are cointegrated with cointegrating vector $[1,-1]$. This also implies that, since this is true
for any $k$, if we consider the vector of forward rates of tenor 1 to $m$ periods, together with the current spot rate, $[s_t, f_t^1, f_t^2, f_t^3, \ldots, f_t^m]'$, then this must be cointegrated with $m$ unique cointegrating vectors, each given by a row of the matrix $[-1, I_m]$, where $I_m$ is an $m$-dimensional identity matrix and $\mathbf{1}$ is an $m$-dimensional column vector of ones. Finally, by the Granger Representation Theorem (Granger, 1986; Engle and Granger, 1987) this vector of forward and spot rates must possess a VECM representation in which the term structure of forward premia plays the part of the equilibrium errors.

It is important to note that, while assumption (i) above is uncontroversial, assumption (ii) and its key implication of a stationary forward premium are more controversial. In particular, some authors have suggested that the forward premium may display long memory properties and is well characterized as a nonstationary fractionally integrated process, or $f(d)$ process, where $0.5 < d < 1$. This property of the forward premium would then imply that spot and forward exchange rates are fractionally cointegrated and that the dynamic relationship between spot and forward rates is characterized by a fractional VECM (e.g. see Baillie, 1996; Baillie and Bollerslev, 2000; Maynard and Phillips, 2001). While it is well documented in the literature that the forward premium is persistent and this property is likely to plague statistical inference in the context of testing market efficiency, the specific long-run properties of the forward premium and of the relationship between spot and forward rates are not central to the empirical work carried out in the present study. Our main aim relates to the investigation of the short-run forecasting ability of empirical exchange rate models based on information extracted from the term structure of forward rates. For this purpose, the assumption of a stationary (albeit potentially persistent) forward premium is not crucial in our analysis, and hence we shall stick to the original assumptions of the Clarida–Taylor framework.

The linear VECM used by Clarida and Taylor (1997) may be written as follows:

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

(1)

where $y_t = [s_t, f_t^1, f_t^2, f_t^3, \ldots, f_t^m]'$, with the superscript denoting the number of weeks corresponding to the maturity of the forward contract; $\Pi = \alpha \beta'$ is the long-run impact matrix whose rank determines the number of cointegrating vectors linking spot and forward rates (equal to four in this specific VECM); and $u_t$ is a vector of Gaussian error terms (Johansen, 1988, 1991). Clarida and Taylor (1997) exploit this linear VECM representation to show that sufficient information may be extracted from the term structure in order to forecast the spot dollar exchange rate during the recent floating exchange rate regime. Their dynamic out-of-sample forecasts suggest that the linear VECM is superior to a range of alternative forecasts, including a random walk and standard spot-forward regressions.

Clarida et al. (2003) then generalize the linear VECM in Eq. (1) to a multivariate Markov-switching framework and examine the performance of such a model in out-of-sample exchange rate forecasting. This generalized term structure model 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. Using weekly data on major spot and forward dollar exchange rates over the period 1979 through 1995, Clarida et al. report evidence of the presence of nonlinearities in the term structure, which appeared to be well characterized by a three-regime MS-VECM that allows for shifts in both the intercept and the covariance structure of the error terms. This MS-VECM may be written as follows:

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

where \( y_t \) is as defined in Eq. (1); \( \nu(z_t) \) is a 3-dimensional column vector of regime-dependent intercept terms, \( \nu(z_t) = [\nu_1(z_t), \nu_2(z_t), \nu_3(z_t)]' \); the \( \Pi_i \)'s are \( 3 \times 3 \) matrices of parameters; \( \varepsilon_t = [\varepsilon_{1t}, \varepsilon_{2t}, \varepsilon_{3t}]' \) is a 3-dimensional vector of error terms, \( \varepsilon_t \sim \text{NID}(0, \Sigma(z_t)) \). The regime-generating process is assumed to be an ergodic Markov chain with three states \( z_t \in \{1,2,3\} \), governed by the transition probabilities \( p_{ij} = \Pr(z_{t+1} = j | z_t = i) \), and \( \sum_{j=1}^{3} p_{ij} = 1 \forall i, j \in \{1,2,3\} \). This MS-VECM is termed Markov-Switching-Intercept-Heteroskedastic-VECM or MSIH-VECM. In order to reflect the fact that the model has three regimes and one lag of the dependent variable in each equation, the model is termed MSIH(3)-VECM(1). Estimation of an MS-VECM can be carried out using an expectations maximization (EM) algorithm for maximum likelihood (Dempster et al., 1977; Krolzig, 1997).4

Clarida et al. (2003) use the MSIH(3)-VECM(1) to forecast dynamically out of sample over the period 1996 through 1998. The results suggest that the MS-VECM forecasts are strongly superior to the random walk forecasts at a range of forecasting horizons up to 52 weeks ahead, using standard forecast accuracy criteria. Moreover, the MS-VECM also outperforms a linear VECM for spot and forward rates in out-of-sample forecasting of the spot rate, although the magnitude of the gain, in point forecasting, from using an MS-VECM relative to a linear VECM is rather small at short horizons (about 10% on average at the 4-week forecast horizon). Nevertheless, it is possible that traditional measures of forecast accuracy mask somehow the potential superiority of nonlinear models (Satchell and Timmermann, 1995; Granger, 2003).

Overall, this literature suggests that not only the term structure of forward premia contains valuable information about the future spot exchange rate, but also that the allowance for nonlinear dynamics in the form of regime-switching behavior enhances somewhat this information to produce a satisfactory forecasting exchange rate model. While previous research on term structure models has analyzed forecasting performance focusing primarily 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. Especially when evaluating nonlinear models, which are

---

4 An alternative way to carry out specification, estimation and density forecasting for this class of regime-switching models involves conducting robust statistical inference using Markov Chain Monte Carlo (MCMC) methods (e.g. see Chib, 2001, 2004, and the references therein).
capable of producing non-normal forecast densities, it would seem appropriate to consider a model’s density forecasting performance. This is indeed the primary objective of the empirical work undertaken in this paper, where we carry out density forecasting tests on the linear VECM and the MS-VECM of the term structure of forward premia as well as on a random walk exchange rate model. We then investigate some of the implications of our density forecasting results for exchange rate risk management.

3. Empirical analysis I: modelling

In this section, we describe the data and carry out a preliminary unit root and cointegration analysis of spot and forward exchange rates. We also report linearity tests and other tests designed to select the most adequate MS-VECM in our context, which we then estimate.5

3.1. Data and preliminaries

Our data set comprises weekly observations of eight bilateral spot and 4-, 13-, 26- and 52-week forward US dollar exchange rates (vis-à-vis the UK sterling, Swiss franc, Japanese yen, Canadian dollar, New Zealand dollar, Swedish krona, Norwegian krona, Danish krona). The sample period spans from January 4, 1985 to December 31, 2003.6 Following previous literature (e.g. Hansen and Hodrick, 1980, p. 852), data are Tuesdays of every week, taken from Datastream. From this data set, we constructed the time series of interest, namely the logarithm of the spot exchange rate, $s_t$ and the logarithm of the $k$-week forward exchange rate, $f_t^k$ for $k = 4, 13, 26, 52$. In our empirical work, we carried out our estimations over the period January 1985–December 1995, reserving the last eight years of data for out-of-sample forecasting.

Having first confirmed – by using standard unit root test statistics – the stylized fact that each of the time series examined is a realization from a stochastic process integrated of order one, we then employed the Johansen (1988, 1991) maximum likelihood procedure in a vector autoregression for $y_t = [s_t, f_t^4, f_t^{13}, f_t^{26}, f_t^{52}]'$ and an unrestricted constant term. On the basis of the Johansen likelihood ratio test statistics for cointegrating rank, 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.

5 Throughout our discussion of the empirical results, we employ a nominal significance level of 5% unless explicitly stated otherwise.

6 The start date was chosen since it was the earliest date for which data for all exchange rates examined are available. Full details on the preliminary empirical analysis discussed in this sub-section are available from the authors upon request, but are not reported to conserve space.
When testing the hypothesis that the cointegrating vectors linking spot and forward rates are of the form $[1, -1]$, we rejected this hypothesis but – consistent with Naka and Whitney (1995), Luinlet and Paudyal (1998) and Clarida et al. (2003) – the departure from the overidentifying restrictions, albeit statistically significant at conventional test sizes, was found to be very small in magnitude. Following Clarida et al. (2003), we interpreted the rejection of the unity restrictions on the cointegration space as 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 large sample size. Given the theoretical economic priors in favor of the unity restrictions and the fact that, under covered interest parity, coefficients different from unity would have the implausible implication of a unit root in international interest rate differentials, we carried out our empirical work employing the unity restrictions.

We next estimated a standard linear VECM, as given in Eq. (1), using full-information maximum likelihood (FIML) methods, assuming a maximum lag length of three, as suggested by both the Akaike Information Criterion and the Schwartz Information Criterion. Employing the conventional general-to-specific procedure, we obtained fairly parsimonious models for each exchange rate.

3.2. Linearity tests, identification and estimation of the MS-VECM

We proceeded to investigate the presence of nonlinearities further through the estimation of a general MS-VECM of the form:

$$
\Delta y_t = \nu(z_t) + \Pi(z_t)y_{t-1} + \sum_{d=1}^{p-1} \Gamma_d(z_t)\Delta y_{t-d} + \eta_t,
$$

where all parameters are as defined in Eq. (2), except for the autoregressive parameters $\Gamma_d$'s and the long-run matrix $\Pi$ which are also allowed to be regime shifting, i.e. $\Gamma_d(z_t)$ and $\Pi(z_t) = \alpha(z_t)\beta'$; $\eta_t$ is a vector of error terms, $\eta_t \sim \text{NID}(0, \sum_{n}(z_t))$. The number of regimes, $z_t$ – for which we consider a maximum of three regimes, i.e. $z_t \in \{1, 2, 3\}$ – is identified using a likelihood ratio test specifically designed for this purpose and described below. The MS-VECM in Eq. (3) is indeed slightly more general than the MS-VECM used by Clarida et al. (2003) in that, except for the long-run cointegrating matrix $\beta'$, which is restricted to be consistent with stationary forward premia, every other parameter of the model is allowed to be regime shifting. In essence, this MS-VECM allows for each of the intercepts, the variance–covariance matrix and the autoregressive structure to be regime dependent. To reflect the fact that model (3) allows for each of the parameters on the autoregressive lags of $\Delta y$ to be regime switching, in addition to regime-switching intercept vector and covariance matrix, this VECM is termed Markov-Switching-Intercept-Autoregressive-Heteroskedastic-VECM or MSIAH-VECM.

Next we applied the ‘bottom-up’ procedure designed to detect Markovian shifts in order to select the most adequate characterization of an MS-VECM for $\Delta y_t$. The linearity tests led us, for each exchange rate, to reject the linear VECM against an MS-VECM. These tests provide strong empirical evidence that the linear VECM
fails to capture important nonlinearities in the data generating process, since linearity is rejected with marginal significance levels ($p$-values) of virtually zero – see the last column in Table 1.

For each MS-VECM estimated, we tested the hypothesis of no regime shifts against the alternative hypothesis of regime shifts in each of the vector of intercept terms, the variance–covariance matrix, and the autoregressive structure, respectively. The likelihood ratio (LR) tests LR1, LR2 and LR3, constructed as suggested by Krolzig (1997, pp. 135–136) and reported in Table 1, suggest in each case massive rejections of the null hypotheses tested, clearly indicating that an MS-VECM that allows for shifts in each of the intercept, the variance–covariance matrix and the autoregressive structure, namely an MSIAH-VECM, is the most appropriate model within its class in the present application.

Finally, in order to discriminate between models allowing for two regimes against models governed by three regimes we also calculated an LR test for this null hypothesis. The results produced very low $p$-values (see LR4, Table 1), suggesting that three regimes are appropriate in all cases. Therefore, in spite of parsimony considerations, we allowed for three regimes in the MSIAH-VECM.

Following the results from the identification procedure reported in Table 1, we estimated the MSIAH-VECM(3) for each of the eight exchange rates examined. The estimation yielded fairly plausible estimates of the coefficients, including the adjustment coefficients in $\alpha$, which were generally found to be strongly statistically significantly different from zero.

For each exchange rate we find that three regimes are appropriate in describing the data. Shifts from one regime to another appeared to be due largely to shifts in the variance of the term structure equilibrium. On the other hand, shifts in the intercept

<table>
<thead>
<tr>
<th></th>
<th>LR1</th>
<th>LR2</th>
<th>LR3</th>
<th>LR4</th>
<th>Linearity</th>
</tr>
</thead>
<tbody>
<tr>
<td>UK</td>
<td>0</td>
<td>$1.67 \times 10^{-16}$</td>
<td>$1.59 \times 10^{-14}$</td>
<td>$2.61 \times 10^{-22}$</td>
<td>0</td>
</tr>
<tr>
<td>Switzerland</td>
<td>0</td>
<td>$4.13 \times 10^{-14}$</td>
<td>$7.36 \times 10^{-18}$</td>
<td>$2.23 \times 10^{-24}$</td>
<td>0</td>
</tr>
<tr>
<td>Japan</td>
<td>0</td>
<td>$3.02 \times 10^{-12}$</td>
<td>$2.61 \times 10^{-18}$</td>
<td>$2.83 \times 10^{-37}$</td>
<td>0</td>
</tr>
<tr>
<td>Canada</td>
<td>$3.98 \times 10^{-243}$</td>
<td>$5.93 \times 10^{-3}$</td>
<td>$9.77 \times 10^{-15}$</td>
<td>$5.85 \times 10^{-46}$</td>
<td>$2.24 \times 10^{-275}$</td>
</tr>
<tr>
<td>New Zealand</td>
<td>0</td>
<td>$2.08 \times 10^{-8}$</td>
<td>$2.86 \times 10^{-8}$</td>
<td>$4.01 \times 10^{-82}$</td>
<td>0</td>
</tr>
<tr>
<td>Sweden</td>
<td>0</td>
<td>0</td>
<td>$4.77 \times 10^{-9}$</td>
<td>$1.93 \times 10^{-17}$</td>
<td>0</td>
</tr>
<tr>
<td>Norway</td>
<td>0</td>
<td>$4.23 \times 10^{-27}$</td>
<td>$6.64 \times 10^{-10}$</td>
<td>$2.64 \times 10^{-32}$</td>
<td>0</td>
</tr>
<tr>
<td>Denmark</td>
<td>0</td>
<td>$1.53 \times 10^{-6}$</td>
<td>$2.79 \times 10^{-13}$</td>
<td>$1.09 \times 10^{-50}$</td>
<td>0</td>
</tr>
</tbody>
</table>

Notes: LR1 is a test statistic of the null hypothesis of MSI(3)-VECM(1) vs MSIAH(3)-VECM(1). LR2 is a test statistic of the null hypothesis of MSH(3)-VECM(1) vs MSIAH(3)-VECM(1). LR3 is a test statistic of the null hypothesis of MSIH(3)-VECM(1) vs MSIAH(3)-VECM(1). LR4 is the likelihood ratio test for the null hypothesis that the MSIAH-VECM(1) with two regimes is equivalent to the MSIAH-VECM(1) with three regimes. These LR tests are constructed as $2(\ln L^* - \ln L)$, where $L^*$ and $L$ represent the unconstrained and the constrained maximum likelihood, respectively. These tests are distributed as $\chi^2 (r)$ where $r$ is the number of restrictions. ‘Linearity’ is a linearity test for the null hypothesis that the selected MSIAH-VECM(1) is equivalent to a linear Gaussian VECM(1). $p$-values relative to the LR4 and linearity tests are calculated as in Ang and Bekaert (1998). For all test statistics only $p$-values are reported; $p$-values below $10^{-350}$ are reported as 0.
terms and, to a lesser extent, in the autoregressive structure were found to be relatively smaller in magnitude, albeit strongly statistically significant. This seems consistent with the large empirical literature investigating the time-varying nature of exchange rates risk premia. One tentative interpretation of this MSIAH-VECM is, in fact, in terms of shifts in the mean and variance of foreign exchange returns consistent with deviations from the equilibrium levels implied by conventional macroeconomic fundamentals that may be caused, for example, by ‘peso problems’ or by other kinds of departures from the efficient markets hypothesis (see Engle and Hamilton, 1990; Clarida et al., 2003).

For each of the countries considered, we clearly identified one regime characterized by periods in which the average spot rate and its variability were relatively high compared to the remaining two regimes. We then investigated whether or not the probability of switching between regimes was related to macroeconomic fundamentals, in the spirit of recent research by Bansal and Zhou (2002), Bansal et al. (2004) and Clarida et al. (in press) – in the context of interest rates modelling, these authors find that regime shifts are intimately related to macroeconomic fundamentals. Like in previous research (e.g. Bansal and Zhou, 2002; Clarida et al., in press), because data on the explanatory variables we consider are not available at weekly frequency, we use quarterly data. Hence, from the estimated MSIAH-VECMs, we converted the weekly smoothed probabilities into quarterly probabilities by averaging. Further, in order to obtain a binary variable from the estimated average MSIAH-VECM probabilities we defined a variable which is equal to zero when the average quarterly probability of being in a regime is smaller than 0.5 and equal to unity when this average probability is greater than or equal to 0.5. The fundamentals term we consider in our simple exercise is the deviation of the spot exchange rate from its equilibrium value as predicted by the monetary model of exchange rate determination (see, inter alia, Mark, 1995; Sarno and Taylor, 2003; Sarno et al., 2004).7 These deviations have been squared in order to assign the same weight to negative and positive departures from equilibrium. We then calculated the correlation between the binary indicator variable – capturing the regime of high mean and volatility of spot exchange rate movements – and our proxy for the deviation

---

7 Precisely, we focus on the deviation, say dev, of the nominal exchange rate from its fundamental value: \[ \text{dev}_t = s_t - \text{fund}_t, \] where the fundamentals term, fund, is the long-run equilibrium of the nominal exchange rate determined by the monetary fundamentals. The fundamentals term is given by: \[ \text{fund}_t = (m_t - m^*_t) - (g_t - g^*_t), \] where \( m \) and \( g \) denote the log-levels of money supply and real income respectively; and asterisks denote foreign variables – data for narrow money and real Gross Domestic Product were taken for each country from the *International Financial Statistics* of the International Monetary Fund. fund may be thought of as a typical representation of the long-run equilibrium exchange rate implied by exchange rate determination theory, and forms of this kind are implied, for example, by the monetary approach to exchange rate determination as well as by general equilibrium models with cash-in-advance constraints and several ‘new open economy macroeconomic’ models (see Sarno, 2001; Neely and Sarno, 2002; Sarno and Taylor, 2003, Ch. 4 and 5; Sarno et al. 2004). Hence, the link between monetary fundamentals and the nominal exchange rate is consistent with both traditional models of exchange rate determination based on aggregate functions as well as with more recent microfounded open economy models.
from exchange rate equilibrium. Our results show that on average, across countries, the probability of being in the regime characterized by high average and volatile spot exchange rates is significantly positively correlated with departures from their equilibrium levels. In particular, the average estimated correlation coefficient is about 8%, which is in line with the recent literature on interest rate shifts and macroeconomic fundamentals (i.e. Bansal and Zhou, 2002, p. 2020). This positive relationship suggests that when large deviations (of either sign) from the fundamental exchange rate equilibrium level occur, spot exchange rates are likely to experience periods of high excess returns and volatility.

In general, we note that considerable caution should be exercised in interpreting the regime classification, as there is considerable error in this classification due to parameter estimation error. However, at first glance, our results suggest that the shifts in mean and variance of spot exchange rates depicted by our MSIAH-VECMs may be related, to some extent, to changes in the sort of economic fundamentals one would expect to play a role in driving exchange rates behavior and regimes.

4. Empirical analysis II: forecasting and risk assessment

In this section we compute tests designed to assess the performance of the random walk model, the linear VECM and the MSIAH-VECM in terms of exchange rate density forecasting. We then explore the implications of the density forecasting results in a risk management exercise.

4.1. Out-of-sample density forecasts

To investigate the forecasting ability of our linear and Markov-switching VECMs, we attempt to exploit the whole information provided by the out-of-sample predictions of the models in the context of a density forecasting approach. In particular, it is interesting to examine whether the MSIAH-VECM(3), which was chosen on the basis of the identification procedure which rejected the linear VECM, performs satisfactorily in terms of ‘closeness’ of the predicted moments of the forecast density of the model relative to the true moments of exchange rate movements obtained from the data over the forecast period. This question cannot be addressed fully by using conventional methods of point forecast accuracy evaluation, since these methods consider only the first two moments of the distribution of exchange rates.

A large and growing literature has recently focused on evaluating the forecast accuracy of empirical models on the basis of density, as opposed to point, forecasting performance (see, inter alia, Diebold et al., 1998; Granger and Pesaran, 1999; Tay and Wallis, 2000; Timmermann, 2000; Sarno and Valente, 2004, in press). Several researchers have proposed methods for evaluating density forecasts. For example, Diebold et al. (1998) extend previous work on the probability integral transform and show how it is possible to evaluate a model-based predictive density. Diebold et al. propose the calculation of the probability integral transforms of the actual
realizations of the variables (i.e. exchange rate changes for each country under investigation) over the forecast period, \( \{\Delta s_{t+1}\}_{t=1}^{n} \) with respect to the models’ forecast densities, denoted by \( \{p_t(\Delta s_{t+1})\}_{t=1}^{n} \):

\[
w_t = \int_{-\infty}^{\Delta s_{t+1}} p_t(\varphi) d\varphi \quad t = 1, \ldots, n.
\]

When the model forecast density corresponds to the true predictive density, then the sequence of \( \{w_t\}_{t=1}^{n} \) is iid \([0,1]\). The idea is to evaluate whether the realizations of the data over the forecast period come from the selected forecast density by testing whether the \( \{w_t\} \) series departs from iid uniformity.

Berkowitz (2001) suggests that rather than working with the \( \{w_t\} \) series it may be fruitful to take the inverse normal cumulative distribution function (CDF) transform of the series \( \{w_t\} \), denoted by \( \{x_t\} \). Under the null hypothesis of equality of the model density and the true predictive density, \( \{x_t\} \) is distributed as standard normal, and Berkowitz proposes an LR test for zero mean and unit variance, under the maintained hypothesis of iid normality. We rely on the test of Berkowitz (2001) in our empirical work, since Berkowitz shows that working with the inverse normal transform of the series \( \{w_t\} \) drastically increases the power of the test relative to the version based on uniformity.

While, under general conditions, the linear VECMs forecast densities are easy to calculate analytically (they are in fact multivariate normal distributions with means and variances given by functions of the estimated parameters), the MSIAH-VECM forecast densities can, in general, be obtained analytically only for one-step-ahead forecasts. The MSIAH-VECM forecast densities are mixtures of multivariate normal distributions with weights given by the predicted regime probabilities. In general, the MSIAH-VECM forecast densities are non-normal, asymmetric and heteroskedastic. In this paper we focus on the one-step-ahead forecast density of the MSIAH(3)-VECM(1), which is given by:

\[
p_{t+1}(\Delta y_{t+1}) = \sum_{j=1}^{3} \left\{ \sum_{i=1}^{3} p_{ij} \mathbf{P} \right\} p_{t+1}(\Delta y_{t+1}|z_{t+1} = j, \Omega_t),
\]

where \( p_{ij} = \Pr(z_{t+1} = j|z_t = i) \) are the transition probabilities; \( \mathbf{P} \) is the transition matrix conditional on the information set at time \( t, \Omega_t \); and \( p_{t+1}(\Delta y_{t+1}|z_{t+1} = j, \Omega_t) \) is the regime-conditional forecast density.

We now turn to the evaluation of the probability integral transforms. The null of iid normality is a joint hypothesis and, in the spirit of Diebold et al. (1998), we consider each part of the hypothesis in turn. The iid assumption is tested by calculating the Ljung and Box (1979) test for no serial correlation. In order to take into account the dependence occurring in the higher moments, we consider \((x - \bar{x})^\omega\) for \( \omega \) up to four. In our forecasting exercise, we compare the density forecasting performance of the linear VECM in Eq. (1) and the MSIAH-VECM in Eq. (3) with the standard benchmark in the literature on exchange rate forecasting, namely the
random walk model.\textsuperscript{8} The \( p \)-values from carrying out these tests are reported in Panel a) of Table 2. The results suggest that, for each exchange rate, we are unable to reject the null hypothesis of no serial correlation at conventional significance levels. This finding holds for each of three models examined, namely the random walk model, the linear VECM, and the MSIAH-VECM. We then carry out tests for normality to verify whether the \( \{x_t\} \) series displays any statistically significant skewness or excess kurtosis. The \( p \)-values from these tests, reported in Panel b) of Table 2, suggest that the null hypotheses of no skewness and no excess kurtosis cannot be rejected for any of the exchange rates and any of the three competing models considered. Taken together, the results in Panel a) and Panel b) of Table 2 indicate that the null hypothesis of iid normality of the \( \{x_t\} \) series cannot be rejected.

Given these findings, the maintained hypothesis which is required to carry out the LR test proposed by Berkowitz (2001) is validated by the data. Hence, we calculate the LR tests of zero mean and unit variance proposed by Berkowitz (2001), which we report in Panel a) of Table 3. The results are interesting. On the basis of these LR tests, the only model for which we cannot reject the null hypothesis is the MSIAH-VECM. Indeed, the forecast densities of the random walk model and the linear VECM lead to rejection of the null hypothesis for each exchange rate except for the New Zealand dollar. In other words, except for the New Zealand dollar (where each of three competing models performs satisfactorily), the random walk model and the linear VECM produce density forecasts that are statistically significantly different from the actual density of exchange rates data over the forecast period. On the other hand, the MSIAH-VECM generates, for each exchange rate considered, density forecasts that are statistically identical to the true predictive densities.

One of the limitations of the testing procedure employed above is the fact that, while it allows us to measure how well a model’s predictive density approximates the true predictive density of the data, it does not allow a formal test of which of the competing models considered performs best in terms of density forecasting performance. Heuristically this can be investigated by inspecting the \( p \)-values of the Berkowitz tests for each individual model (i.e. higher \( p \)-values presumably reflect a better density forecast) but it is not possible to use the Berkowitz testing procedure to obtain a \( p \)-value for the null hypothesis that two models perform equally well in forecasting the true predictive distribution. A solution to this problem has recently been proposed by Corradi and Swanson (2004), who derive a test statistic for the null hypothesis that two models have equal density forecast accuracy. In some sense, this test evaluates competing forecasting models in terms of density forecasting in the same spirit of the Diebold and Mariano (1995) test for equal point forecast accuracy of competing models. Using a measure that may be seen as the analogue of the mean square error in the contest of density forecasts, Corradi and Swanson (2004) test a null hypothesis that can be expressed as

\textsuperscript{8} Obviously, the random walk model refers to the level of the exchange rate, implying that in the case of exchange rate changes the model becomes a model of normally distributed exchange rate changes with mean equal to the drift term and variance equal to the residual variance estimated over the sample period available.
Table 2
Out-of-sample performance: tests for iid and normality

Panel a) Tests for iid based upon inverse normal probability transform

<table>
<thead>
<tr>
<th></th>
<th>(x – \bar{x})</th>
<th>(x – \bar{x})^2</th>
<th>(x – \bar{x})^3</th>
<th>(x – \bar{x})^4</th>
</tr>
</thead>
<tbody>
<tr>
<td>RW</td>
<td>VECM</td>
<td>MSIAH-VECM</td>
<td>RW</td>
<td>VECM</td>
</tr>
<tr>
<td>UK</td>
<td>0.6418</td>
<td>0.7148</td>
<td>0.5057</td>
<td>0.5712</td>
</tr>
<tr>
<td>Switzerland</td>
<td>0.2083</td>
<td>0.3228</td>
<td>0.2938</td>
<td>0.4365</td>
</tr>
<tr>
<td>Japan</td>
<td>0.1297</td>
<td>0.3857</td>
<td>0.6756</td>
<td>0.6344</td>
</tr>
<tr>
<td>Canada</td>
<td>0.9946</td>
<td>0.9999</td>
<td>0.9996</td>
<td>0.3179</td>
</tr>
<tr>
<td>New Zealand</td>
<td>0.1137</td>
<td>0.2823</td>
<td>0.2758</td>
<td>0.2287</td>
</tr>
<tr>
<td>Sweden</td>
<td>0.5129</td>
<td>0.7465</td>
<td>0.8697</td>
<td>0.7254</td>
</tr>
<tr>
<td>Norway</td>
<td>0.2161</td>
<td>0.4233</td>
<td>0.4113</td>
<td>0.4269</td>
</tr>
<tr>
<td>Denmark</td>
<td>0.2266</td>
<td>0.3691</td>
<td>0.4011</td>
<td>0.4258</td>
</tr>
</tbody>
</table>

Panel b) Tests for normality based upon inverse normal probability transform

<table>
<thead>
<tr>
<th></th>
<th>Sk</th>
<th>Ku</th>
<th>Sk</th>
<th>Ku</th>
<th>Sk</th>
<th>Ku</th>
<th>Sk</th>
<th>Ku</th>
</tr>
</thead>
<tbody>
<tr>
<td>RW</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VECM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MSIAH-VECM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UK</td>
<td>0.4021</td>
<td>0.2618</td>
<td>0.3431</td>
<td>0.2669</td>
<td>0.5604</td>
<td>0.3330</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Switzerland</td>
<td>0.1823</td>
<td>0.4309</td>
<td>0.1921</td>
<td>0.4089</td>
<td>0.2324</td>
<td>0.3615</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Japan</td>
<td>0.0951</td>
<td>0.4780</td>
<td>0.0817</td>
<td>0.4511</td>
<td>0.2718</td>
<td>0.4734</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Canada</td>
<td>0.2815</td>
<td>0.5165</td>
<td>0.4250</td>
<td>0.4801</td>
<td>0.4418</td>
<td>0.5155</td>
<td></td>
<td></td>
</tr>
<tr>
<td>New Zealand</td>
<td>0.7336</td>
<td>0.4482</td>
<td>0.6906</td>
<td>0.4043</td>
<td>0.6684</td>
<td>0.4273</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sweden</td>
<td>0.4482</td>
<td>0.4069</td>
<td>0.4624</td>
<td>0.4442</td>
<td>0.5021</td>
<td>0.3872</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Norway</td>
<td>0.3649</td>
<td>0.2717</td>
<td>0.4698</td>
<td>0.2537</td>
<td>0.5232</td>
<td>0.3212</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Denmark</td>
<td>0.4204</td>
<td>0.4567</td>
<td>0.4156</td>
<td>0.4464</td>
<td>0.5525</td>
<td>0.2804</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: Panel a) Figures denote p-values for the Ljung and Box (1979) χ^2 test of no first-order serial correlation of the normal inverse probability transform series, x_t, as defined in Section 4.1. Panel b) Figures denote p-values for the tests of the null hypothesis that the skewness (Sk) and excess kurtosis (Ku) of the distribution of the normal inverse probability transform series x_t are equal to 0.
Table 3
Out-of-sample performance: density forecasting tests

Panel a) Berkowitz (2001) LR test

<table>
<thead>
<tr>
<th></th>
<th>RW</th>
<th>VECM</th>
<th>MSIAH-VECM</th>
</tr>
</thead>
<tbody>
<tr>
<td>UK</td>
<td>2.60 × 10^{-2}</td>
<td>3.08 × 10^{-2}</td>
<td>6.92 × 10^{-1}</td>
</tr>
<tr>
<td>Switzerland</td>
<td>5.47 × 10^{-6}</td>
<td>7.03 × 10^{-5}</td>
<td>7.03 × 10^{-1}</td>
</tr>
<tr>
<td>Japan</td>
<td>1.27 × 10^{-10}</td>
<td>4.25 × 10^{-8}</td>
<td>3.19 × 10^{-1}</td>
</tr>
<tr>
<td>Canada</td>
<td>2.91 × 10^{-7}</td>
<td>6.86 × 10^{-5}</td>
<td>3.11 × 10^{-1}</td>
</tr>
<tr>
<td>New Zealand</td>
<td>6.41 × 10^{-1}</td>
<td>6.35 × 10^{-1}</td>
<td>7.04 × 10^{-1}</td>
</tr>
<tr>
<td>Sweden</td>
<td>3.46 × 10^{-9}</td>
<td>3.19 × 10^{-8}</td>
<td>3.44 × 10^{-1}</td>
</tr>
<tr>
<td>Norway</td>
<td>4.88 × 10^{-6}</td>
<td>1.58 × 10^{-5}</td>
<td>8.12 × 10^{-1}</td>
</tr>
<tr>
<td>Denmark</td>
<td>5.70 × 10^{-4}</td>
<td>3.42 × 10^{-4}</td>
<td>3.89 × 10^{-1}</td>
</tr>
</tbody>
</table>

Panel b) Corradi and Swanson (2004) test

<table>
<thead>
<tr>
<th></th>
<th>RW vs VECM</th>
<th>RW vs MSIAH-VECM</th>
<th>VECM vs MSIAH-VECM</th>
</tr>
</thead>
<tbody>
<tr>
<td>UK</td>
<td>−0.0013 [0.0274]</td>
<td>0.0201 [0.0032]</td>
<td>0.0214 [0.0030]</td>
</tr>
<tr>
<td>Switzerland</td>
<td>−0.0071 [0.0221]</td>
<td>0.0077 [0.0237]</td>
<td>0.0148 [0.0212]</td>
</tr>
<tr>
<td>Japan</td>
<td>0.0183 [0.0124]</td>
<td>0.0102 [0.0194]</td>
<td>0.0284 [0.0122]</td>
</tr>
<tr>
<td>Canada</td>
<td>−0.0211 [0.0101]</td>
<td>0.0264 [0.0073]</td>
<td>0.0475 [0.0037]</td>
</tr>
<tr>
<td>New Zealand</td>
<td>0.0074 [0.0220]</td>
<td>0.0285 [0.0092]</td>
<td>0.0211 [0.0129]</td>
</tr>
<tr>
<td>Sweden</td>
<td>0.0010 [0.0299]</td>
<td>0.0117 [0.0280]</td>
<td>0.0107 [0.0307]</td>
</tr>
<tr>
<td>Norway</td>
<td>0.0085 [0.0211]</td>
<td>0.0122 [0.0188]</td>
<td>0.0037 [0.0249]</td>
</tr>
<tr>
<td>Denmark</td>
<td>−0.0107 [0.0203]</td>
<td>0.0047 [0.0013]</td>
<td>0.0154 [0.0009]</td>
</tr>
</tbody>
</table>

Notes: Panel a) Figures denote p-values from computing the LR test statistic of Berkowitz (2001), which is distributed as \(\chi^2(3)\) under the null hypothesis. Panel b) Figures reported are Corradi and Swanson (2004, in press) test statistics for the null hypothesis that the conditional densities from two competing models are equally accurate relative to the true density. The test statistics are constructed by selecting \(u \in [\Delta s_{\min}, \Delta s_{\max}]\) where \(\Delta s\) is the time series of the log-difference of the nominal exchange rate over the entire sample period and 100 equally spaced values of \(u\) across this range were used. Values in brackets are p-values calculated by bootstrap based on subsampling as in Politis et al. (1999), where the number of bootstrap replications is set to 100.

\[
H_0 : \int_U E\{[F_{1,t+1}(\Delta s_{t+1}|u) - F_0(\Delta s_{t+1}|u)]^2 \\
- [F_{2,t+1}(\Delta s_{t+1}|u) - F_0(\Delta s_{t+1}|u)]^2\} \phi(u) du,
\]

where \(u \in U \subset \mathbb{R}, \phi(u) \geq 0\) is a possible unbounded set on the real line; \(\int_U \phi(u) du = 1\); \(F_{q,t+1}(\Delta s_{t+1}|u)\) is the predictive cumulative density function (CDF) implied by model \(q = 1, 2\) at time \(t + 1\) for a given \(u\); and \(F_0(\Delta s_{t+1}|u)\) is the value of the true CDF for a given \(u\). The test statistic for the null hypothesis (6) takes the form of

\[
Z_T = \int_U Z_{T,u} \phi(u) du,
\]

where \(Z_{T,u} = \frac{1}{\sqrt{p}} \sum_{i=0}^{n-1} \{[\tilde{F}_{1,t+1}(\Delta s_{t+1}|u) - F_0(\Delta s_{t+1}|u)]^2 - [\tilde{F}_{2,t+1}(\Delta s_{t+1}|u) - F_0(\Delta s_{t+1}|u)]^2\}\); and \(\tilde{F}_{q,t+1}\) is the estimated counterpart of \(F_{q,t+1}\) for \(q = 1, 2\). The test statistic \(Z_T\) is then calculated by averaging \(Z_{T,u}\) over \(u \in U\). Following Corradi and Swanson (2004), this can be
done by generating a fine grid of \( u \) whose values are equally spaced across the range determined by the minimum and maximum value of \( \Delta s_{t+1} \) over the sample period. Hence, if we assume a grid with \( S \) points, the test statistic (7) becomes:

\[
Z_T = \frac{1}{S} \sum_{s=1}^{S} \left\{ \frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} \left[ \tilde{F}_{s,t+1}(\Delta s_{t+1}|u_s) - F_0(\Delta s_{t+1}|u_s) \right]^2 - \left[ \tilde{F}_{2,s,t+1}(\Delta s_{t+1}|u_s) - F_0(\Delta s_{t+1}|u_s) \right]^2 \right\},
\]

(8)

Differently from the previous testing procedures based on the probability integral transform, the limiting distribution of the test statistic (8) is a functional of a Gaussian process whose covariance kernel is not a martingale difference sequence in the presence of model misspecification. This implies that its limiting distribution is not nuisance-parameter free and therefore cannot be tabulated. Corradi and Swanson (2004) show how to obtain critical values for the distribution of the test statistic (8) by bootstrap.\(^9\),\(^10\)

The results from calculating the Corradi and Swanson (2004) test are reported in Panel b) of Table 3. In general, the null hypothesis of equal density forecast accuracy is rejected for most exchange rates at the 5% significance level. In particular, there is clear evidence that the MSIAH-VECM is able to generate predictive densities which are better approximations of the true predictive density than the ones implied by the random walk model and the linear VECM. Indeed, by inspecting the second and third columns in Panel b) of Table 3, the Corradi–Swanson test statistics are all positive and statistically significant. This means that the distance between the true predictive density implied by the data and the predictive densities generated by the random walk model (column 2) or the linear VECM (column 3) is significantly larger than the distance between the true predictive density and the predictive density generated by the MSIAH-VECM. Further, while it is straightforward to establish, on the basis of these results, that the MSIAH-VECM is the best model in terms of density forecasting performance, it is difficult to discriminate between the linear VECM and the random walk model since there is no clear pattern in the sign of the test statistics.

Summing up, the forecasting results in this section suggest that, in terms of density forecasting performance, the MSIAH-VECM performs better than the linear VECM and the random walk model in terms of explaining the out-of-sample behavior of exchange rate movements. Clearly, this finding, obtained by both testing procedures employed – namely the Berkowitz test and the Corradi–Swanson test

---

\(^9\) In our density forecasting exercise we implement the bootstrap designed to calculate critical values using subsampling, as shown in Politis et al. (1999, Ch. 3) and recently employed by Linton et al. (2003).

\(^10\) Note that another test for equal density forecast accuracy available in the literature is due to Sarno and Valente (2004). While this test has the advantage of being easily applicable since it has a known limiting distribution, we prefer to use the Corradi–Swanson test since this is more general and works under less stringent assumptions and regularity conditions.
given in Table 3 – is due to the allowance for multiple regimes in the MSIAH-VECM, which enhances the information embedded in the forward premia and generates density forecasts that are closer to the true predictive densities, providing a better characterization of the uncertainty surrounding the exchange rate forecasts than the other two competing models.

4.2. The economic value of density forecasts: a simple example of VaR analysis

In this section we further investigate the practical implications of the density forecasting results reported in the previous sub-section in the context of a simple risk management exercise. Given the predictions of the three competing models examined here, assume that an US risk manager wishes to quantify the one-week-ahead risk associated with holding a position in foreign currency. Assume that the position in question is a naive diversified portfolio comprising a domestic asset and $N$ foreign assets which are identical in all respects except for the currency of denomination, e.g. euro-deposit rates. Each asset delivers yields in local currency, and given that diversification of this portfolio is assumed to be naive, the weight on each asset is $1/(N+1)$. Define $i_t$ the domestic (US) one-period interest rate and $i_{ct}$ the foreign one-period interest rate associated with country $c$. Then, given a certain level of wealth at time $t$ invested in this portfolio, say $W_t$, the law of motion of wealth is as follows:

$$W_{t+1} = \left[ \frac{1}{N+1} \exp(i_t) + \sum_{c=1}^{N} \frac{1}{N+1} \exp\left(i_{ct} + \Delta S_{t+1}^c\right) \right] W_t$$

where $\Delta S_{t+1}^c$ is the first difference in the dollar log-exchange rate vis-à-vis the currency of country $c$ (e.g. Mark, 2001; Elton et al., 2003). Given that $i_t$ and $i_{ct}$ are known at time $t$, the only source of risk to be taken in consideration by the risk manager at time $t$ is the future nominal exchange rate $\Delta S_{t+1}^c$ for $c = 1, \ldots, N$. Normalizing $W_t = 1$ for simplicity, the risk manager can use a specific model of exchange rates at time $t$ to produce one-week-ahead density forecasts of $\Delta S_{t+1}^c$, which, in turn, imply one-week-ahead density forecasts for the one-week-ahead wealth $W_{t+1}$. On the basis of these densities the risk manager calculates, for each of the three competing models – the random walk, the linear VECM and the MSIAH-VECM – the VaR as a confidence interval for losses such that

$$\Pr(W_{t+1} < \text{VaR}_{t+1}) = 1 - \phi.$$ (10)

In our example the VaR is a 99% confidence level for losses (i.e. $\phi = 0.99$), for all models; and $N = 8$, corresponding to the eight dollar exchange rates studied in this paper. Eq. (10) simply states that the probability that one-week-ahead wealth, $W_{t+1}$ is less than the VaR is equal to the significance level ($1 - \phi$).$^{11}$

$^{11}$ For this sub-section, we obtained weekly observations of one-week euro rates for each country examined during the sample period January 1996–December 2003 for Datastream.
Summary statistics are reported in Panel a) of Table 4. In order to assess the relative size and relative variability of the VaR estimates across the competing models we use the mean relative bias statistic (MRB) and root mean squared relative bias statistic (RMSRB), suggested by Hendricks (1996). The MRB statistic is calculated as:

\[
\text{MRB}_g = \frac{1}{n} \sum_{h=1}^{n} \frac{\text{VaR}_{g,t+h} - \overline{\text{VaR}}_{t+h}}{\overline{\text{VaR}}_{g,t+h}}
\]

where \( \text{VaR}_{g,t} \) is the estimated VaR from the specific model \( g \) used at time \( t \), \( \overline{\text{VaR}}_{t} \) is the cross-sectional average VaR at time \( t \) over the three competing models, and \( n \) is the number of out-of-sample observations. This statistic gives a measure of size for each estimated VaR relative to the average of all competing models.

The RMSRB statistic is calculated as:

\[
\text{RMSRB}_g = \sqrt{\frac{1}{n} \sum_{h=1}^{n} \left( \frac{\text{VaR}_{g,t+h} - \overline{\text{VaR}}_{t+h}}{\overline{\text{VaR}}_{g,t+h}} \right)^2}.
\]

Table 4
Value-at-Risk calculations

Panel a) Summary statistics

<table>
<thead>
<tr>
<th></th>
<th>RW</th>
<th>VECM</th>
<th>MSIAH-VECM</th>
</tr>
</thead>
<tbody>
<tr>
<td>MRB</td>
<td>-0.0014</td>
<td>-0.0010</td>
<td>0.0024</td>
</tr>
<tr>
<td>RMSRB</td>
<td>2.43 \times 10^{-6}</td>
<td>1.45 \times 10^{-6}</td>
<td>6.65 \times 10^{-6}</td>
</tr>
<tr>
<td>AD</td>
<td>-</td>
<td>0.0132</td>
<td>0.1347</td>
</tr>
<tr>
<td>\text{corr}_{\text{VaR},W}</td>
<td>-0.0121</td>
<td>0.0314</td>
<td>0.1553**</td>
</tr>
</tbody>
</table>

Panel b) VaR backtests

<table>
<thead>
<tr>
<th></th>
<th>RW</th>
<th>VECM</th>
<th>MSIAH-VECM</th>
</tr>
</thead>
<tbody>
<tr>
<td>( V )</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>( \text{VR} )</td>
<td>0.23%*</td>
<td>0.23%*</td>
<td>0.72%</td>
</tr>
<tr>
<td>( \text{CD} )</td>
<td>0.998</td>
<td>0.996</td>
<td>0.994</td>
</tr>
</tbody>
</table>

Notes: Panel a) MRB and RMSRB are the mean relative bias and the root mean square relative bias, calculated as in Hendricks (1996) and given in Eqs. (11) and (12), respectively. AD is the average distance between the realized data and the VaR implied by each of the random walk (RW) model, linear VECM and MSIAH-VECM, standardized by the average distance of the RW model. \( \text{corr}_{\text{VaR},W} \) is the correlation coefficient between the estimated VaR and the realized data, calculated as in Hendricks (1996). ** indicates statistical significance at the 1% significance level. Panel b) \( V \) denotes the number of times that \( W_{t+1} < \text{VaR}_{t+1} \), while VR denotes the estimated violation rate (i.e. \( V \) divided by the number of out-of-sample observations), for each of the three competing models. We test the null hypothesis that the violation rate VR does not exceed the theoretical violation rate of 1% using the Kupiec (1995) test; the asterisk (*) indicates statistical significance at the 5% significance level, while the absence of an asterisk next to VR implies that the Kupiec test statistic does not reject the null hypothesis of equality of the model’s estimated violation rate and 1%. CD is the Christoffersen and Diebold (2000) test for the sample first-order autocorrelation of a binary variable which is equal to unity if a violation occurs and zero otherwise. For the CD tests only \( p \)-values are reported.
This measure provides us with information about the extent to which the estimated VaR tends to vary around the average VaR at time $t$. The results from calculating the MRB and RMSRB, reported in Panel a) of Table 4, suggest that the MSIAH-VECM produces higher VaRs (compared to the average VaR produced by the other two competing models) and it also produces more volatile VaRs (around the average VaR). The higher VaR produced by the MSIAH-VECM is indicative that this model is less conservative, on average, than the other two competing models.

We also report in Panel a) of Table 4 the average distance (AD) between the realized data and the VaR implied by the random walk model, the linear VECM and the MSIAH-VECM, standardized by the average distance of the random walk model; and, finally, we report the correlation coefficient between the estimated VaR from each model and the realized data, calculated as in Hendricks (1996) and termed $\text{corr}_{\text{VaR},H}$. These calculations indicate, on the basis of the AD, that the MSIAH-VECM produces the “closest” VaR to the realized data by some 13% relative to the random walk model and 12% relative to the linear VECM. Also, on the basis of $\text{corr}_{\text{VaR},H}$, the MSIAH-VECM generates the VaR most highly correlated with the realized data across the three models considered.\(^{12}\)

In Panel b) of Table 4, we report some VaR backtests. In particular, we report (as $V$) the number of times that $W_{t+1} < \text{VaR}_{t+1}$, and the implied estimate of the violation rate (i.e., $V$ divided by the number of out-of-sample observations), say VR, for each of the random walk model, the linear VECM and the MSIAH-VECM, respectively. We also test of the null hypothesis that the violation rate VR does not exceed the theoretical 1% violation rate considered in this VaR application, calculated as in Kupiec (1995). Finally, we report the Christoffersen and Diebold (2000) test for the sample first-order autocorrelation of a binary variable which is equal to unity if a violation occurs and zero otherwise. The results in Panel b) indicate that the random walk model and the linear VECM are too conservative in that they both exhibit one violation, whereas the number of violations under the VaR estimated for the MSIAH-VECM is three, which implies an estimated violation rate of 0.72%. Such estimated violation rate is indeed insignificantly different from the theoretical violation rate of 1%, as confirmed by the Kupiec test statistic. On the other hand, the estimated violation rate for the other two competing models is, in each case, equal to about 0.23%, and is statistically significantly different from the theoretical violation rate of 1%. The CD test also suggests that there is no first-order serial correlation in the VaR violations for each of the three models under examination, indicating that these violations are, for each model, non-systematic.\(^{13}\)

Summing up, this simple application further illustrates the satisfactory out-of-sample forecasting performance of the MSIAH-VECM relative to the random walk model.

---

\(^{12}\) The results show the poor performance of the other two models, which both display statistically insignificant correlation coefficients and, for the random walk model, a negatively signed one.

\(^{13}\) In terms of interpretation, a significant autocorrelation coefficient denotes a persistent series of violations, which in turn implies unsatisfactory performance of a model in estimating the VaR. Note that, while our result of no significant serial correlation is obvious for the random walk model and the linear VECM since these models only imply one violation, it is not obvious for the MSIAH-VECM.
benchmark and to the linear VECM. We find that the MSIAH-VECM provides a violation rate which is statistically insignificantly different from the theoretical violation rate of 1%, whereas the other two competing models provide estimated VaRs which are too conservative and, generally, poor estimates of the risk of the portfolio under examination. This may be seen as evidence that the most general model, the MSIAH-VECM does better than the other two competing models at matching the moments of the predictive distribution of exchange rate changes, generating VaRs that are more in line with the desired violation rate, confirming the findings of the previous sub-section on density forecasting performance and illustrating the practical importance of such results.

5. Conclusion

This article has re-examined the performance of some empirical exchange rate models in terms of out-of-sample forecasting of nominal exchange rates. In particular, inspired by the success of recently developed models of the term structure of forward exchange rates in terms of point forecasting, we have carried out a density forecasting analysis, applied to both linear and regime-switching versions of these term structure models. This exercise was aided by the recent developments of sophisticated econometric techniques which allow us to formally evaluate the performance of time series models in terms of density forecasting.

Our main result, using weekly data for eight US dollar exchange rates during the recent floating exchange rate regime, is that a Markov-switching VECM for spot and forward exchange rates that explicitly takes into account the mounting evidence that the conditional distribution of exchange rates is well characterized by a mixture of normal distributions produces very satisfactory one-week-ahead density forecasts. This model was found to outperform its more parsimonious linear counterpart as well as the standard benchmark in the exchange rate forecasting literature, namely the random walk model.

The implication of our findings were further investigated in the context of a simple application of Value-at-Risk methods. In our application we specifically examined the implications of our exchange rate density forecasts for a risk manager who has to quantify the risk associated with holding a currency portfolio over a one-week horizon. This application further illustrated how the Markov-switching VECM captures satisfactorily the moments of the predictive density of exchange rates, generating VaRs that measure the probability of large losses more accurately than the other two competing models.

Overall, the pecking order implied by our density forecasting results – Markov-switching term structure model, linear term structure model, random walk model14 – is the same as recorded in previous work based on conventional point forecasting

14 Specially, this is the result suggested by the Berkowitz test statistic, while using the Corradi–Swanson test we were not able to discriminate between the linear VECM and the random walk model. All tests indicated, however, that the best performing model is the Markov-switching VECM.
criteria. However, when evaluating these models in terms of density forecasting results, the superiority of the Markov-switching VECM relative to the linear VECM becomes much clearer than previously recorded in the literature using point forecast evaluation. Overall, our findings highlight how better density forecasts of exchange rates, of the type recorded in this paper using Markov-switching models of the term structure, can potentially lead to substantial improvements in risk management and, more precisely, to better estimates of downside risk.

Acknowledgments

This paper was partly written while Lucio Sarno was visiting the International Monetary Fund and the Central Bank of Norway. The research reported in this paper was carried out with the aid of financial support from the Economic and Social Research Council (ESRC Grant No. RES-000-22-0404). The authors are grateful for useful conversations and comments to Angelos Kanas, Georgios Kouretas, Jim Lothian, Rich Lyons, Mike Melvin, Athanasios Papadopoulos, Mark Salmon, Nick Sarantis, Mark Taylor, George Tsiotas (discussant) and other participants of the 8th International Conference on Macroeconomic Analysis and International Finance, University of Crete, May 27–29, 2004. The authors alone are responsible for any errors that may remain and for the views expressed in the paper.

References

Corradi, V., Swanson, N.R., 2004. Predictive density accuracy tests. Queen Mary University and Rutgers University, mimeo.


