IMF Working Paper

Nonlinear Exchange Rate Models: A Selective Overview

Lucio Sarno

INTERNATIONAL MONETARY FUND

IMF Working Paper

IMF Institute

Nonlinear Exchange Rate Models: A Selective Overview

Prepared by Lucio Sarno[1]

Authorized for distribution by Mohsin S. Khan

May 2003

Abstract

The views expressed in this Working Paper are those of the author(s) and do not necessarily represent those of the IMF or IMF policy. Working Papers describe research in progress by the author(s) and are published to elicit comments and to further debate.

This paper provides a selective overview of nonlinear exchange rate models recently proposed in the literature and assesses their contribution to understanding exchange rate behavior. Two key questions are examined. The first question is whether nonlinear autoregressive models of real exchange rates help resolve the "purchasing power parity (PPP) puzzles." The second question is whether recently developed nonlinear, regime-switching vector equilibrium correction models of the nominal exchange rate can beat a random walk model, the standard benchmark in the exchange rate literature, in terms of out-of-sample forecasting performance. Finally, issues related to the adequateness of standard methods of evaluation of (linear and nonlinear) exchange rate models are discussed with reference to different forecast accuracy criteria.

Contents

Page

I. INTRODUCTION

In recent years, a growing body of literature in exchange rate economics has devoted a great deal of attention to the role of nonlinear dynamics in exchange rates. This paper provides a selective overview of nonlinear exchange rate models recently proposed by researchers and assesses their contribution to understanding exchange rate behavior.[2] Two key questions are addressed. The first is concerned with whether nonlinear autoregressive models of real exchange rates help us resolve the "purchasing power parity (PPP) puzzles"—namely the fact that the real exchange rate either displays no tendency to revert to a stable long-run equilibrium level consistent with PPP or shows an implausibly slow speed of mean reversion. The second question is whether recently developed nonlinear, regime-switching vector equilibrium correction models of the nominal exchange rate can beat a random walk model, the standard benchmark in the exchange rate literature, in terms of out-of-sample forecasting performance. Finally, issues related to the adequateness of standard methods of evaluation of (linear and nonlinear) exchange rate models are discussed with reference to both point and density forecast accuracy criteria.

The remainder of the paper is as follows. Section II briefly reviews the literature on testing the validity of long-run PPP and on modeling the real exchange rate, discusses the theoretical rationale for nonlinear mean reversion in the real exchange rate, and then summarizes the empirical evidence on the ability of nonlinear models of real exchange rate behavior to address the PPP puzzles. Section III provides an overview of the literature on out-of-sample exchange rate forecasting using economic fundamentals and the term structure of forward premia, followed by a discussion of the recent work on nonlinear models of the term structure and an assessment of their ability to beat a random walk forecast and to improve upon linear specifications. Section IV reviews some of the recent work on evaluating density forecasts, which is particularly relevant in the context of exchange rates and of nonlinear models that are consistent with nonnormal densities. This section also includes a simple application to density forecast evaluations in the context of linear and nonlinear models of exchange rates. A final section concludes.

II.   REAL EXCHANGE RATE BEHAVIOR AND LONG-RUN PURCHASING POWER PARITY

A. A Brief Overview of the Literature

The purchasing power parity (PPP) hypothesis states that national price levels should be equal when expressed in a common currency. Although very few economists would believe that this simple proposition holds at each point in time, a large literature in international finance has examined empirically the validity of PPP over the long run either by testing whether nominal

---

[2] This paper draws on several papers written by the author jointly with several coauthors, including Ibrahim Chowdhury, Richard Clarida, Mark Taylor, and Giorgio Valente.

exchange rates and relative prices move together in the long run or by testing whether the real exchange rate has a tendency to revert to a stable equilibrium level over time. The latter approach is motivated by the fact that the real exchange rate may be defined as the nominal exchange rate adjusted for relative national price levels. More formally, the real exchange rate, q_t, may be expressed in logarithmic form as

$$q_t \equiv s_t - p_t + p_t^*,$$ (1)

where s_t is the logarithm of the nominal exchange rate (domestic price of foreign currency), and p_t and p_t^* denote the logarithms of the domestic and foreign price levels respectively. The real exchange rate, q_t, may thus be interpreted as a measure of the deviation from PPP and must be stationary for long-run PPP to hold (see the surveys of Froot and Rogoff, 1995; Rogoff, 1996; Sarno and Taylor, 2002).

Although long-run PPP is a very simple proposition about exchange rate behavior, it has attracted the attention of researchers for decades. Indeed, whether long-run PPP holds or whether the real exchange rate is stationary has important economic implications on a number of fronts. In particular, the degree of persistence in the real exchange rate can be used to infer the principal impulses driving exchange rate movements. For example, if the real exchange rate is highly persistent or close to a random walk, then the shocks are likely to be real-side, principally technology shocks, whereas if it is not very persistent, then the shocks must be principally to aggregate demand, such as, for example, innovations to monetary policy (Rogoff, 1996). Further, from a theoretical perspective, if PPP is not a valid long-run international parity condition, this casts doubts on the predictions of much open-economy macroeconomics that is based on the assumption of long-run PPP. Indeed, the implications of open economy dynamic models are very sensitive to the presence or absence of a unit root in the real exchange rate (e.g., Lane, 2001; Sarno, 2001). Finally, estimates of PPP exchange rates are often used for practical purposes such as determining the degree of misalignment of the nominal exchange rate and the appropriate policy response, the setting of exchange rate parities, and the international comparison of national income levels. These practical uses of the PPP concept, and in particular the calculation of PPP exchange rates, would obviously be of very limited use if PPP deviations contain a unit root.

Regardless of the great interest in this area of research, manifested by the large number of papers on PPP published over the last few decades, and regardless of the increasing quality of data sets utilized and of the econometric techniques employed, the validity of long-run PPP and the properties of PPP deviations remain the subject of ongoing controversy. Specifically, earlier cointegration studies generally reported the absence of significant mean reversion of the real exchange rate for the recent floating experience (Taylor, 1988; Mark, 1990), but were supportive of reversion toward PPP for the gold standard period (McCloskey and Zecher, 1984; Diebold, Husted, and Rush, 1991), for the interwar float (Taylor and McMahon, 1988), for the 1950s U.S.-Canadian float (McNown and Wallace, 1989), and for the exchange rates of high-inflation countries (Choudhry, McNown, and Wallace, 1991). More recent applied work on

long-run PPP among the major industrialized economies has, however, been more favorable toward the long-run PPP hypothesis for the recent float (e.g., Corbae and Ouliaris, 1988; Cheung and Lai, 1993a, 1993b, 1994, 1998; Frankel and Rose, 1996).

One well-documented explanation for the inability to find clear-cut evidence of PPP is the low power of conventional statistical tests to reject a false null hypothesis of a unit root in the real exchange rate or no cointegration between the nominal exchange rate and relative prices with a sample span corresponding to the length of the recent float (Frankel, 1986, 1990; Froot and Rogoff, 1995; Lothian and Taylor, 1997). Researchers have sought to overcome the power problem in testing for mean reversion in the real exchange rate either through long span studies (e.g., Kim, 1990; Lothian and Taylor, 1996; Taylor, 2002) or through panel unit root studies (e.g.,Abuaf and Jorion, 1990; Frankel and Rose, 1996; O'Connell, 1998; Papell, 1998; Sarno and Taylor, 1998; Taylor and Sarno, 1998). However, whether or not the long-span or panel-data studies do in fact answer the question whether PPP holds in the long run remains contentious. As far as the long-span studies are concerned, as noted in particular by Frankel and Rose (1996), the long samples required to generate a reasonable level of statistical power with standard univariate unit root tests may be unavailable for many currencies (perhaps thereby generating a `survivorship bias' in tests on the available data) and, in any case, may potentially be inappropriate because of differences in real exchange rate behavior both across different historical periods and across different nominal exchange rate regimes (e.g., Baxter and Stockman, 1989; Hegwood and Papell, 1999; Taylor, 2002). As for panel-data studies, these provide mixed evidence. While, for example, Abuaf and Jorion (1990), Frankel and Rose (1996) and Taylor and Sarno (1998) find results favorable to long-run PPP, O'Connell (1998) rejects it on the basis of their empirical evidence.

In light of the evidence provided by this literature, there remain several unresolved puzzles, among which two are prominent. First, it is still controversial whether long-run PPP is valid during the recent floating exchange rate regime. Second, it is puzzling why the majority of studies find empirical estimates of the persistence of PPP deviations that are too high—the half life of shocks ranges between three and five years—to be explained in light of conventional nominal rigidities and cannot be reconciled with the large short-term volatility of real exchange rates (Rogoff, 1996).

B. Nonlinear Mean Reversion in Real Exchange Rates: Rationale and Modeling Procedures

In the procedures conventionally applied to test for long-run PPP, the null hypothesis is usually that the process generating the real exchange rate series has a unit root, while the alternative hypothesis is that all of the roots of the process lie within the unit circle. Thus, the maintained hypothesis in the conventional framework assumes a linear autoregressive process for the real exchange rate, which means that adjustment is both continuous and of constant speed, regardless of the size of the deviation from PPP. However, the presence of transactions costs may imply a nonlinear process, which has important implications for the conventional unit root tests of long-run PPP. The idea that there may be nonlinearities in real exchange rate adjustment dates at least from Heckscher (1916), who suggested that there may be significant

deviations from the law of one price due to international transactions costs between spatially separated markets. A similar viewpoint can be discerned in the writings of Cassel (e.g., Cassel, 1922) and, to a greater or lesser extent, in other earlier writers (Officer, 1982). More recently, a number of authors have developed theoretical models of nonlinear real exchange rate adjustment arising from transactions costs in international arbitrage (e.g., Benninga and Protopapadakis, 1988; Williams and Wright, 1991; Dumas, 1992; Sercu, Uppal and Van Hulle, 1995). In most of these models, proportional or "iceberg" transport costs ("iceberg" because a fraction of goods are presumed to "melt" when shipped) create a band for the real exchange rate within which the marginal cost of arbitrage exceeds the marginal benefit. Assuming instantaneous goods arbitrage at the edges of the band then typically implies that the thresholds become reflecting barriers.

Drawing on recent work on the theory of investment under uncertainty, some of these studies show that the thresholds should be interpreted more broadly than as simply reflecting shipping costs and trade barriers per se, but also as resulting from the sunk costs of international arbitrage and the resulting tendency for traders to wait for sufficiently large arbitrage opportunities to open up before entering the market (see in particular Dumas, 1992; Obstfeld and Rogoff, 2000).

O'Connell and Wei (2002) extend the iceberg model to allow for fixed as well as proportional costs of arbitrage. This results in a two-threshold model where the real exchange rate is reset by arbitrage to an upper or lower inner threshold whenever it hits the corresponding outer threshold. Intuitively, arbitrage will be heavy once it is profitable enough to outweigh the initial fixed cost, but will stop short of returning the real rate to the PPP level because of the proportional arbitrage costs.

Overall, these models suggest that the exchange rate will become increasingly mean reverting with the size of the deviation from the equilibrium level. In some models the jump to mean-reverting behavior is sudden, whilst in others it is smooth, and Dumas (1994) suggests that even in the former case, time aggregation will tend to smooth the transition between regimes. Moreover, if the real exchange rate is measured using price indices made up of goods prices each with a different size of international arbitrage costs, one would expect adjustment of the overall real exchange rate to be smooth rather than discontinuous.

Some empirical evidence of the effect of transactions costs on tests of PPP is provided by Davutyan and Pippenger (1990). More recently, Obstfeld and Taylor (1997), Taylor (2001), Sarno, Taylor, and Chowdhury (2003), Leon and Najarian (2003) have investigated the nonlinear nature of the adjustment process in terms of a threshold autoregressive (TAR) model (Tong, 1990). The TAR model allows for a transactions costs band within which no adjustment takes place—so that deviations from PPP may exhibit unit root behavior—while outside of the band the process switches abruptly to become stationary autoregressive. While discrete switching of this kind may be appropriate when considering the effects of arbitrage on disaggregated goods prices (Obstfeld and Taylor, 1997), discrete adjustment of the aggregate real exchange rate would clearly be most appropriate only when firms and traded goods are identical. Moreover, several theoretical studies suggest that smooth rather than discrete

adjustment may be more appropriate in the presence of proportional transactions costs and, as suggested by Teräsvirta (1994), Dumas (1994), and Bertola and Caballero (1990), time aggregation and nonsynchronous adjustment by heterogeneous agents is likely to result in smooth aggregate regime switching.

An alternative characterization of nonlinear adjustment, which allows for smooth rather than discrete adjustment, is in terms of a smooth transition autoregressive (STAR) model (Granger and Teräsvirta, 1993). In the STAR model, adjustment takes place in every period but the speed of adjustment varies with the extent of the deviation from parity. A STAR model may be written as follows:

$$[q_t - \mu] = \sum_{j=1}^{p} \beta_j [q_{t-j} - \mu] + \left(\sum_{j=1}^{p} \beta_j^* [q_{t-j} - \mu] \right) \Phi[\theta; q_{t-d} - \mu] + \varepsilon_t, \quad (2)$$

where $\{q_t\}$ is a stationary and ergodic process, $\varepsilon_t \sim iid(0, \sigma^2)$ and $(\theta\mu) \in \{\Re^+ \times \Re\}$, where $\Re$ denotes the real line $(-\infty, \infty)$ and $\Re^+$ the positive real line $(0, \infty)$. The transition function $\Phi[\theta; q_{t-d} - \mu]$ determines the degree of mean reversion and is itself governed by the parameter θ, which effectively determines the speed of mean reversion, and the parameter μ, which is the equilibrium level of $\{q_t\}$; the integer $d > 0$ denotes a delay parameter. A simple transition function suggested by Granger and Teräsvirta (1993) is the exponential function:

$$\Phi[\theta; q_{t-d} - \mu] = 1 - \exp\left\{ -\theta^2 [q_{t-d} - \mu]^2 \right\}, \quad (3)$$

in which case (2) would be termed an exponential STAR or ESTAR model. The exponential transition function is bounded between zero and unity, $\Phi : \Re \longrightarrow [0,1]$, has the properties $\Phi[0] = 0$ and $\lim_{x \longrightarrow \pm\infty} \Phi[x] = 1$, and is symmetrically inverse-bell shaped around zero. These properties of the ESTAR model are attractive in the present modeling context because they allow a smooth transition between regimes and symmetric adjustment of the real exchange rate for deviations above and below the equilibrium level. The transition parameter θ determines the speed of transition between the two extreme regimes, with lower absolute values of θ implying slower transition. The inner regime corresponds to $q_{t-d} = \mu$, when $\Phi = 0$ and (2) becomes a linear AR(p) model:

$$[q_{t-d} - \mu] = \sum_{j=1}^{p} \beta_j [q_{t-j} - \mu] + \varepsilon_t. \quad (4)$$

The outer regime corresponds, for a given θ, to $\lim_{[q_{t-d} - \mu] \longrightarrow \pm\infty} \Phi[\theta; q_{t-d} - \mu]$, where (2) becomes a different AR(p) model:

$$[q_{t-d} - \mu] = \sum_{j=1}^{p} (\beta_j + \beta_j^*)[q_{t-j} - \mu] + \varepsilon_t, \quad (5)$$

with a correspondingly different speed of mean reversion so long as $\beta_j^* \neq 0$ for at least one value of j.[3]

It is also instructive to reparameterize the STAR model (2) as

$$\Delta q_t = \alpha + \rho q_{t-1} + \sum_{j=1}^{p-1} \phi_j \Delta q_{t-j} + \left\{ \alpha^* + \rho^* q_{t-1} + \sum_{j=1}^{p-1} \phi_j^* \Delta q_{t-j} \right\} \Phi[\theta; q_{t-d}] + \varepsilon_t, \quad (6)$$

where $\Delta q_{t-j} \equiv q_{t-j} - q_{t-j-1}$. In this form, the crucial parameters are ρ and ρ^*. Our above discussion of the effect of transactions costs suggests that the larger the deviation from PPP the stronger will be the tendency to move back to equilibrium. This implies that while $\rho \geq 0$ is admissible, we must have $\rho^* < 0$ and $(\rho + \rho^*) < 0$. That is, for small deviations q_t may be characterized by unit root or even explosive behavior, but for large deviations the process is mean reverting. This analysis has implications for the conventional test for a unit root in the real exchange rate process, which is based on a linear AR(p) model, written below as an augmented Dickey-Fuller regression:

$$\Delta q_t = \alpha' + \rho' q_{t-1} + \sum_{j=1}^{p-1} \phi_j' \Delta q_{t-j} + \varepsilon_t. \quad (7)$$

Assuming that the true process for q_t is given by the nonlinear model (6), estimates of the parameter ρ' in (7) will tend to lie between ρ and $(\rho + \rho^*)$, depending upon the distribution of observed deviations from the equilibrium level μ. Hence, the null hypothesis $H_0 : \rho' = 0$ (a single unit root) may not be rejected against the stationary linear alternative hypothesis $H_1 : \rho' < 1$, even though the true nonlinear process is globally stable with $(\rho + \rho^*) < 0$. Thus, failure to reject the unit root hypothesis on the basis of a linear model does not necessarily invalidate long-run PPP.

---

[3] Granger and Teräsvirta (1993) also suggest the logistic function as a plausible transition function for some applications, resulting in a logistic STAR or LSTAR model. Since, however, the LSTAR model implies asymmetric behavior of q according to whether it is above or below the equilibrium level, that model is, a priori, inappropriate for modeling real exchange rate movements. That is to say, it is hard to think of economic reasons why the speed of adjustment of the real exchange rate should vary according to whether the dollar is overvalued or undervalued, especially if one is thinking of goods arbitrage as ultimately driving the impetus towards the long-run equilibrium and one is dealing with major dollar real exchange rates against the currencies of other developed industrialized countries.

C. Empirical Evidence on Nonlinear Mean Reversion in Real Exchange Rates

We now turn to the empirical evidence on nonlinear mean reversion in real exchange rates. Michael, Nobay, and Peel (MNP) (1997) apply the ESTAR model to monthly interwar data for the French franc-U.S. dollar, French franc-pound sterling, and pound sterling-U.S. dollar as well as for the Lothian and Taylor (1996) long span data set. Their results clearly reject the linear framework in favor of an ESTAR process. The systematic pattern in the estimates of the nonlinear models provides strong evidence of mean-reverting behavior for PPP deviations, and helps explain the mixed results of previous studies. However, the periods examined by MNP are ones over which the relevance of long-run PPP is uncontentious (Taylor and McMahon, 1988; Lothian and Taylor, 1996).

Using data for the recent float, however, Taylor, Peel and Sarno (TSP) (2001) report empirical results that provide strong confirmation that four major real bilateral dollar exchange rates are well characterized by nonlinearly mean reverting processes. For example, the estimated model for dollar-sterling over the 1973–1996 sample period is as follows:

$$\hat{q}_t = q_{t-1} - \left[1 - \exp\left\{-0.452\{q_{t-1} + 0.149\}^2\right\}\right]\left[q_{t-1} + 0.149\right],$$

$$(-2.771) \qquad (4.274) \qquad (4.274) \tag{8}$$

$$[0.002]$$

where the hat denotes the fitted value, figures in parentheses are t-ratios; the p-value, calculated by Monte Carlo methods, for the significance of the speed of adjustment parameter in the transition function is 0.002. The recorded R^2 for this simple nonlinear AR(1) is 0.94. This estimated model, which may be seen as representative of the results reported by TPS (2001), implies an equilibrium level of the real exchange rate in the neighborhood of which the behavior of the log-level of the real exchange rate is close to a random walk, becoming increasingly mean reverting with the absolute size of the deviation from equilibrium, consistent with the recent theoretical literature on the nature of real exchange rate dynamics in the presence of international arbitrage costs.

TPS also estimated the impulse response functions corresponding to their estimated nonlinear real exchange rate models by Monte Carlo integration.[4] By taking account of statistically significant nonlinearities, TPS find the speed of real exchange rate adjustment to be typically much faster than the very slow speeds of real exchange rate adjustment hitherto recorded in the literature. For example, the estimated half lives (in months) conditional on average initial history of the real exchange rate are the following:

---

[4] Note that, because of the nonlinearity, the half lives of shocks to the real exchange rates vary both with the size of the shock and with the initial conditions.

Shock (in percent):	40	30	20	10	5	1
U.S. dollar-pound sterling	10	20	22	26	29	32
U.S. dollar-deutsch mark	11	14	18	25	29	33
U.S. dollar-French franc	13	18	24	33	38	40
U.S. dollar-Japanese yen	14	18	24	32	38	42

where in the first row we report the size of the shock (in percentage terms) to the level of the real exchange rate. The estimated half lives of these four major real dollar exchange rates illustrate the nonlinear nature of the response to shocks quite clearly, with larger shocks mean reverting much faster than smaller shocks. The dollar-sterling and dollar-mark models show a marked degree of similarity in terms of the estimated half lives, displaying quite fast mean reversion, ranging from a half life of under one year for the largest shocks of 40 percent to just under three years for very small shocks of 1 percent; for shocks of 5–10 percent, the half lives are just over two years. The dollar-franc displays slightly higher persistence, conditional on average history, with half lives ranging from 13 months for a 40 percent shock to 40 months for a 1 percent shock, while for the dollar-yen the range is 14 to 42 months. For the other sizes of shocks considered, the half lives for dollar-franc and dollar-yen are very close.

These results therefore seem to shed some light on Rogoff's (1996) PPP puzzle. Only for small shocks occurring when the real exchange rate is near its equilibrium do nonlinear models consistently yield half lives in the range of three to five years, which Rogoff (1996) terms "glacial." For dollar-mark and dollar-sterling in particular, even small shocks of one to five percent have a half life under three years, conditional on average history. For larger shocks, the speed of mean reversion is even faster.[5]

---

[5] In a number of Monte Carlo studies calibrated on the estimated nonlinear models, TPS also demonstrate the very low power of standard univariate unit root tests to reject a false null hypothesis of unit root behavior when the true model is nonlinearly mean reverting, thereby suggesting an explanation for the difficulty researchers have encountered in rejecting the linear unit root hypothesis at conventional significance levels for major real exchange rates over the recent floating rate period. Panel unit root tests, however, displayed much higher power in their rejection of the false null hypothesis against an alternative of nonlinear mean reversion, in keeping with the recent literature.

III. OUT-OF-SAMPLE FORECASTS OF THE NOMINAL EXCHANGE RATE

A. The Failure of Conventional Exchange Rate Models

A logical way of examining the empirical ability of exchange rate models is to examine their out-of-sample forecasting performance. In a highly influential paper, Meese and Rogoff (1983) compare the out-of-sample forecasts produced by various exchange rate models with forecasts produced by a random walk model, by the forward exchange rate, by a univariate regression of the spot rate, and by a vector autoregression. They use rolling regressions to generate a succession of out-of-sample forecasts for each model and for various time horizons. The conclusion which emerges from this study is that, on a comparison of root mean square errors (RMSEs), none of the asset-market exchange rate models outperforms the simple random walk, even though actual future values of the right-hand-side variables are allowed in the dynamic forecasts (thereby giving the models a very large informational advantage). In particular, Meese and Rogoff compare random-walk forecasts with those produced by the flexible-price monetary model, Frankel's (1979) real interest rate differential variant of the monetary model, and a synthesis of the monetary and portfolio balance models suggested by Hooper and Morton (1982).

A variant of the Meese-Rogoff approach involves employing a time-varying parameter model. In fact, the poor forecasting performance noted by Meese and Rogoff may be due to the fact the parameters in the estimated equations are unstable. This instability may be rationalized on a number of grounds, in response to policy regime changes as an example of a Lucas critique problem (Lucas, 1976), or because of implicit instability in the money demand or PPP equations, or also because of agents' heterogeneity leading to different responses to macroeconomic developments over time. For example, Schinasi and Swamy (1989) use a Kalman-filter maximum likelihood estimation technique to estimate time-varying parameter models which are found to outperform the random walk model of the exchange rate for certain time periods and currencies.

A general finding in this literature is that researchers have found that one key to improving forecast performance based on economic fundamentals lies in the introduction of equation dynamics. This has been done in various ways: by using dynamic forecasting equations for the forcing variables in the forward-looking, rational expectations version of the flexible-price monetary model, by incorporating dynamic partial adjustment terms into the estimating equation, by using time-varying parameter estimation techniques, and by using dynamic error correction forms (e.g., Koedijk and Schotman, 1990; MacDonald and Taylor, 1993, 1994).

Nevertheless, it remains true that most studies which claim to have beaten the random walk in out-of-sample forecasting turn out to be fragile in the sense that it is generally hard to replicate the superior forecasting performance for alternative periods and alternative currencies.

A related approach, due in the context of foreign exchange market analysis originally to Mark (1995), who considers long-horizon predictability through analysis of equations of the form:

$$\Delta_k s_{t+k} = \alpha + \beta_k \left(z_t - s_t \right) + u_{t+k} ,$$

(9)

where z_t is an exchange rate fundamental, for example that suggested by the monetary class of models, $z_t \equiv \left[\left(m_t - m_t^* \right) - \kappa \left(y_t - y_t^* \right) \right]$, and u_{t+k} is a disturbance term.[6] If the fundamental in question helps forecast the exchange rate, then we should find $\beta_k < 0$ and significantly different from zero. In a series of forecasting tests over long horizons for a number of quarterly dollar exchange rates, Mark finds that equation (9) may be able to predict the nominal exchange rate only at long horizons, such as the four-year horizon. Moreover, both the goodness of in-sample fit and the estimated value of β_k rise as the horizon k rises. Mark interprets this as evidence that, while quarter-to-quarter exchange rate movements may be noisy, systematic movements related to the fundamentals become apparent in long-horizon changes.

In general, long-horizon regressions have been used extensively in the literature, but with mixed success (see Kilian, 1999). One reason may be that previous research has focused on linear models. In fact, in a linear world, it can be argued that there is no rationale for conducting long-horizon forecast tests. The problem is that under linearity k-step ahead forecasts are obtained by linear extrapolation from 1-step ahead forecasts. Thus, by construction there cannot be any gain in power at higher horizons (see Berkowitz and Giorgianni, 1999; Kilian, 1999; Berben and van Dijk, 1998). However, the mounting evidence of nonlinear exchange rate dynamics provides a new rationale for the use of long-horizon regression tests.

B. The Information in the Term Structure of the Forward Premia

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

---

[6] Mark chooses a priori values for κ equal to 1.0 and 0.3.

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

$$s_t = m_t + q_t, \tag{10}$$

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). 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 Ω_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 Ω_t. This is the essence of the 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., Sarno and Taylor, 2002).

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:

$$\gamma_t \equiv f_t^{h(k)} - E\left(s_{t+h(k)}|\Omega_t\right), \tag{11}$$

where $E(./\Omega_t)$ denotes the mathematical expectation conditional on Ω_t. From (10) and (11) we can obtain:

$$f_t^{h(k)} = \gamma_t + h(k)\theta + E\left(q_{t+h(k)}|\Omega_t\right) + m_t, \tag{12}$$

where θ is the drift of the random walk process m_t. Subtracting (10) from (12), we obtain an expression for the forward premium at time t:

$$f_t^{h(k)} - s_t = \gamma_t + h(k)\theta + E\left(q_{t+h(k)} - q_t|\Omega_t\right). \tag{13}$$

Equation (13) says that if the departure from the RNMEH γ_t is stationary, given $q_t \sim I(0)$, the forward premium $\left(f_t^{h(k)} - s_t\right)$ must also be stationary. This implies that forward and spot rates exhibit a common stochastic trend and are cointegrated with cointegrating vector [1,–1]. Moreover, since this is true for any $h(k)$, if we consider the vector of forward rates for $h(1)$ to $h(m)$ periods, together with the current spot rate, $\left[s_t, f_t^{h(1)}, f_t^{h(2)}, f_t^{h(3)}, \ldots f_t^{h(k)}\right]'$, 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 ι is an m-dimensional column vector of ones. Further, by the Granger Representation Theorem (Engle and Granger,

1987) the same set of forward and spot rates must possess a VECM representation in which the term structure of forward premia plays the part of the equilibrium errors.

C. Allowing for Nonlinear Dynamics in a Spot-Forward VECM

Alongside the work on exchange rate forecasting, another strand of the literature has developed in which increasingly strong evidence of nonlinearities of one sort or another in exchange rate movements has been reported. One element of this, dating at least to Booth and Glassman (1987), has been the mounting evidence that the conditional distribution of nominal exchange rate changes is well described by a mixture of normal distributions and that, consequently, a Markov switching model may be a logical characterization of exchange rate behavior (e.g., see Engel and Hamilton, 1990; LeBaron, 1992; Engel, 1994; Engel and Hakkio, 1996; Engel and Kim, 1999). However, although Markov-switching models fit nominal exchange rate data very well, in general they do not produce superior forecasts to a random walk or the forward rate on the basis of conventional forecasting criteria (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 modeling 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).

Clarida, Sarno, Taylor, and Valente (CSTV) (2003) investigate whether allowing for nonlinearities in the underlying data-generating process for the term structure yields superior exchange rate forecasts. This is done through estimating a fairly general three-regime Markov-switching vector equilibrium correction model (MS-VECM) for spot rates and the term structure of forward rates which is essentially based on an extension of Markovian regime shifts to a nonstationary framework, for which the underlying econometric theory has recently been developed (e.g., Krolzig, 1997). The model proposed by CSTV allows regime shifts in both the intercept and the variance-covariance matrix and is governed by three different regimes:

$$\Delta y_t = \nu(z_t) + \Pi y_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + \omega_t, \tag{14}$$

where $\Pi = \alpha\beta'$, $\omega_t \sim NIID\left(0, \sum(z_t)\right)$ and $z_t = 1, 2, 3$. This model—termed Markov-Switching-Intercept-Heteroskedastic-VECM or MSIH-VECM—is estimated by CSTV by maximum likelihood (Dempster, Laird and Rubin, 1977), using dollar rates for each of France, Germany, Japan and the UK, over the sample 1979–1995. With few exceptions, the estimation yields fairly plausible estimates of the coefficients for the VECMs estimated.

For each country we find that three regimes are appropriate in describing the data, and that in each case the three regimes are driven mainly by the joint variability of spot and

forward exchange rates. Shifts from one regime to another appeared to be due largely to shifts in the variance of the term structure equilibrium. On the other hand, shifts in the intercept terms were found to be relatively smaller in magnitude, albeit massively statistically significant. This appears to be in line with the extensive empirical literature investigating the time-varying nature of exchange rates risk premia. One tentative interpretation of our MSIH-VECM is, in fact, in terms of shifts in the mean and variance of foreign exchange returns consistent with deviations from the equilibrium levels implied by conventional macroeconomic fundamentals that may be caused, for example, by "peso problems" or by other kinds of departures from the standard efficient markets hypothesis (see Engel and Hamilton, 1990).

D. Empirical Evidence on the Importance of Nonlinear Dynamics in Out-of-Sample Exchange Rate Forecasting

In order to assess the usefulness of the nonlinear VECM characterization of the term structure, CSTV (2003) construct dynamic out-of-sample forecasts of the spot rate using the MSIH(3)-VECM(1) described in the previous section. In particular the forecasting exercises were performed on the period January 1996–December 1998 with forecast horizons up to 52 weeks ahead. The out-of-sample forecasts for a given horizon $j=1,\ldots, 52$ are constructed according a recursive procedure, 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.[7]

Forecast accuracy is evaluated using absolute and square error criteria, namely the mean absolute error (MAE) and the root mean square error (RMSE). The forecasts produced by the MSIH-VECM are compared 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, the Diebold and Mariano (1995) test is employed:

---

[7] It is well known in the literature that forecasting with nonlinear models raises special problems. CSTV 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, CSTV forecast the path for s_{t+j} for $j=1,\ldots, 52$ using Monte Carlo simulations calibrated on the estimated MSIH-VECMs. The simulation procedure is repeated with identical random numbers 10,000 times and the average of the 10,000 realizations of the sequence of forecasts is taken as the point forecast. Given the large number of simulations, by the Law of Large Numbers this procedure should produce results virtually identical to those which would result from calculating the exact forecast analytically (Gallant, Rossi, and Tauchen, 1993; Brown and Mariano, 1984, 1989; Granger and Teräsvirta, 1993, Chapter 8).

$$DM = \frac{\bar{d}}{\sqrt{\dfrac{2\pi \hat{f}(0)}{T}}},$$ (15)

where $\bar{d}$ is an average (over T observations) of a general loss differential function and $\hat{f}(0)$ is a consistent estimate of the spectral density of the loss differential function at frequency zero. Diebold and Mariano show that the DM statistic is distributed as standard normal under the null hypothesis of equal forecast accuracy. Consistent with a large literature (see, inter alia, Mark, 1995) the loss differential function considered is the difference between the (absolute and square) forecast errors.

The results of the accuracy of the forecasts for the dollar-franc, dollar-mark, dollar-sterling and dollar-yen systems respectively, using MAE and RMSE criteria for forecast accuracy, provide evidence in favor of the predictive superiority of the MSIH-VECM models against the naive random walk and, to a lesser extent, against linear VECM models. The MSIH-VECM models give very much more accurate forecasts than a random walk. At the four-week horizon, the models achieve average improvements ranging between 28–38 percent across currencies using the MAE, and between 27–31 percent using the RMSE. At the 52-week horizon, average improvements range between 8–70 percent using the MAE, and between 5–68 percent using the RMSE, with a maximum reduction of 70 percent in the case of the dollar-yen rate using the MAE. The statistical significance of these results is confirmed executing the DM test.[8]

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 modeling framework, the MS-VECM improves upon the Clarida-Taylor results. In particular, the gains obtained relative to the linear VECMs range, on average across currencies, were between 1–10 percent at the four-week horizon and between 10–38 percent at the 52-week horizon using the MAE. Using the RMSE, the gains range between 1–7 percent at the four-week horizon and between 10–38 percent at the 52-week horizon. Therefore, the gain from using an MSIH-VECM rather than a linear VECM is relatively small at short horizon, albeit generally statistically significant; however, this gain increases with the forecast horizon and becomes substantial at the 52-week horizon.

---

[8] Several problems may arise using DM statistics in small samples as well as taking into account parameter uncertainty (see also West, 1996; West and McCracken, 1998; and McCracken, 2000). In this 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 nonstandard and unknown. Therefore, CSTV emphasize that the marginal significance levels reported below should be interpreted with caution.

IV. FURTHER ISSUES IN EVALUATING NONLINEAR FORECASTING EXCHANGE RATE MODELS

A. Density Forecast Evaluation

A large body of literature in econometrics and applied economics has focused on evaluating the forecast accuracy of economic models (e.g., see the survey of Diebold and Lopez, 1996, and the references therein). Although this literature has traditionally focused on accuracy evaluations based on point forecasts, several authors have recently emphasized the importance of evaluating the forecasting ability of economic models on the basis of density, as opposed to point, forecasting performance (see, inter alia, the survey by Tay and Wallis, 2000, and the references therein).

In a decision-theoretical context, the need to consider the predictive density of a time series—as opposed to considering only its conditional mean and variance—seems fairly accepted in the light of the argument that economic agents may not have loss functions that depend symmetrically on the realizations of future values of potentially non-Gaussian variables. In this case, agents are interested in knowing 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 represents an obvious case (Diebold, Hahn and Tay, 1999; Berkowitz, 2001)—there is an increasingly strong need to provide and evaluate density forecasts. These issues are particularly important in the context of nonlinear models since these models may provide highly non-normal densities.

Several researchers have recently proposed methods for evaluating density forecasts. For example, Diebold, Gunther and Tay (1998) extend previous work on the probability integral transform and show how it is possible to evaluate a model-based predictive density and to test formally the hypothesis that the predictive density implied by a particular model is equal to the true predictive density.[9] Similar ideas have been developed by, inter alia, Anderson, Hall and Titterington (1994), Li (1996), Granger and Pesaran (1999) and Berkowitz (2001). In general, this line of research has produced several methods either to measure the closeness of two density functions or to test the hypothesis that the predictive density generated by a particular model is equal to the true predictive density.

Sarno and Valente (2003) recently proposed a test statistic for the null hypothesis that two competing models have equal density forecast accuracy, in an attempt to provide a more accurate description of the uncertainty surrounding forecasts than traditional methods based on point forecasting. This test is, in the context of density forecasting, the analogue of the test statistic developed by Diebold and Mariano (1995) for testing the null hypothesis that

---

[9] 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 period but only to the forecast period.

two models have equal forecast accuracy in the context of point forecasting. Let $f(y), g_1(y)$ and $g_2(y)$ be three probability density functions with distribution functions F, G_1 and G_2 respectively. Let $f(y)$ be the probability density function of the variable y_t over the period $t = 1,...,T$, whereas $g_1(y)$ and $g_2(y)$ are the probability density functions implied by two competing forecasting models, say M_1 and M_2.

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

$$H_0 : dist[f(y), g_1(y)] = dist[f(y), g_2(y)],$$ (16)

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

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

$$ISD = \int [\phi(x) - \gamma(x)]^2 dx,$$ (17)

where $\phi(.)$ and $\gamma(.)$ denote probability density functions; $ISD \geq 0$, and $ISD=0$ only if $\phi(x) = \gamma(x.)$. Using equation (17) we can rewrite the null hypothesis H_0 as

$$H_0 : \int [f(y) - g_1(y)]^2 dy = \int [f(y) - g_2(y)]^2 dy$$
$$: ISD_1 - ISD_2 = 0,$$ (18)

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

Consider two series of forecasts, $\left\{\hat{y}_{1t}\right\}_{t=1}^{T_1}$ and $\left\{\hat{y}_{2t}\right\}_{t=1}^{T_2}$, obtained from the two competing models M_1 and M_2. Let $g_1(y)$ and $g_2(y)$ be the probability density functions of the two forecast series $\left\{\hat{y}_{1t}\right\}_{t=1}^{T_1}$ and $\left\{\hat{y}_{2t}\right\}_{t=1}^{T_2}$ respectively.[10] With observations $\{y_{1t}\}_{t=1}^{T}$, $\left\{\hat{y}_{1t}\right\}_{t=1}^{T}$ and $\left\{\hat{y}_{2t}\right\}_{t=1}^{T}$ we can estimate the unknown functions $f(y), g_1(y)$ and $g_2(y)$ using kernel estimation, obtaining:

---

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

$$\hat{f}(y) = \frac{1}{Th} \sum_{i=1}^{T} K\left(\frac{y_i - y}{h}\right) \qquad (19)$$

$$\hat{g}_1(y) = \frac{1}{Th} \sum_{i=1}^{T} K\left(\frac{y_{1i} - y}{h}\right) \qquad (20)$$

$$\hat{g}_2(y) = \frac{1}{Th} \sum_{i=1}^{T} K\left(\frac{y_{2i} - y}{h}\right) \qquad (21)$$

where $K(.)$ is the kernel function and h is the smoothing parameter.[11] Using (19)–(21) we can then obtain a consistent estimate of the integrated square differences ISD_1 and ISD_2, say $\hat{ISD}_1$ and $\hat{ISD}_2$. Define $d = \hat{ISD}_1 - \hat{ISD}_2$ as the estimated relative distance of the probability density functions $g_1(y)$ and $g_2(y)$ from $f(y)$. In order to test for the statistical significance of d, the next step is to calculate a confidence interval for it.

In the spirit of the analysis of Hall (1992), define $\{y_i^j\}_{t=1}^{T}$, $\left\{\hat{y}_{1t}^j\right\}_{t=1}^{T}$ and $\left\{\hat{y}_{2t}^j\right\}_{t=1}^{T}$ j-th resample of the original data $\{y_{1t}\}_{t=1}^{T}$, $\left\{\hat{y}_{1t}\right\}_{t=1}^{T}$ and $\left\{\hat{y}_{2t}\right\}_{t=1}^{T}$, drawn randomly with replacement. From these resamples it is possible to obtain consistent bootstrap estimates of the density functions $\hat{f}^j(y)$, $\hat{g}_1^j(y)$, $\hat{g}_2^j(y)$ and, consequently, of $d^j = \hat{ISD}_1^j - \hat{ISD}_2^j$.[12]

Consider a sample path $\{d^j\}_{j=1}^{B}$, where B is the number of bootstrap replications. Under general conditions (See Kendall and Stuart, 1976, Ch. 11), we have

---

[11] In practice, for some models $g_1(y)$ and $g_2(y)$ are known analytically. However, for more complex, nonlinear parametrisations we may not know the probability density functions and therefore need to estimate them. In this paper we consider nonparametric estimation of densities as a general procedure to implement the test described here, but it should be clear that this test is directly applicable also when the probability density functions are known analytically.

[12] Strictly speaking, the data can only be resampled with replacement if they are independently and identically distributed. If there is dependence, the bootstrap procedure needs to be modified to accommodate dependence. However, note that, while resampling blocks generally increase the efficiency of the bootstrap in this case, the available evidence on this issue indicates that the efficiency gains are small (Hall, Horowitz, and Jing, 1995).

$$\sqrt{B}\left(\bar{d}-\mu\right)\xrightarrow{d}N\left(0,\sigma^2\right),$$ (22)

where

$$\bar{d}=\frac{1}{B}\sum_{j=1}^{B}d^j=\frac{1}{B}\sum_{j=1}^{B}\left(\hat{ISD}_1^j-\hat{ISD}_2^j\right)$$ (23)

is the average difference of the estimated relative distances over B bootstrap replications.

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

$$\eta=\frac{\bar{d}}{\sqrt{\dfrac{\hat{\sigma}^2}{B}}}\xrightarrow{d}N(0,1),$$ (24)

where $\hat{\sigma}^2$ is a consistent estimate of σ^2.[13]

This test statistic displays several attractive properties in that it has a known limiting standard normal distribution and—unlike available testing procedures—does not involve testing a joint hypothesis. The test is easy to implement in practice, as illustrated below in an application to exchange rate forecasting. Also, the test is found to have satisfactory empirical size and power properties in a simulation exercise (see Sarno and Valente, 2003). Nevertheless, this test circumvents the problem of testing a joint hypothesis by relying on somewhat stronger assumptions than other methods proposed in the literature that are based on the probability integral transform. Relaxation of these assumptions is an immediate avenue for future research. In particular, the assumption of time-invariance of the densities over the forecast horizon could be relaxed by using recursive kernel estimation, which would allow us to test the null hypothesis of equal density forecast accuracy on time-varying densities period by period (Yamato, 1971; Nobel, Morvai, and Kulkarni, 1998).

B. An Illustrative Application to Exchange Rate Forecasting

We shall illustrate the practical use of the η test statistic with an application to out-of-sample exchange rate forecasting and financial risk management. In recent years, trading

---

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

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 large-scale risk measurement models designed to manage risk. These models generally employ the Value-at-Risk (VaR) methodology. VaR may be defined as the expected maximum loss over a target horizon within a given confidence interval (Jorion, 1997). More formally, VaR is an interval forecast, typically a one-sided 95 or 99 percent interval of the distribution of expected wealth or returns. Users of the VaR methodology generally assume that expected returns are normally or t-distributed. However, this assumption contrasts with the large amount of empirical evidence suggesting that the distribution of exchange rate returns is not normal. Point forecast analysis and testing procedures based upon it do not take into account these features, so that VaR analysis often relies on dubious parametric distributional assumptions. In this example we analyze the out-of-sample forecasting performance of two empirical models of exchange rates and we investigate the implications of these forecasts for a risk manager who has to quantify the risk associated with a simple internationally diversified portfolio over a one-week horizon.

As competing models we consider two multivariate models of the nominal exchange rate. The first model (say Model M_1) is the linear vector error correction model (VECM) based on the spot-forward relationship suggested by Clarida and Taylor (1997). The second model (say Model M_2) is a nonlinear generalization of Model M_1 which allows the intercept and the variance-covariance matrix of the VECM to be regime-shifting, as proposed by CSTV (2003).[14]

Specifically, M_1 can be written as follows:

$$\Delta y_t = \nu + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + \Pi y_{t-1} + u_t, \tag{25}$$

where $y_t = [s_t, f_t]'$; $\Gamma_i = -\sum_{j=i+1}^{p} \Pi_j$ are matrices of parameters; $\Pi = \sum_{i=1}^{p} \Pi_j - I = \alpha\beta'$ is the long-run impact matrix whose rank r determines the number of cointegrating vectors (e.g., Johansen, 1995); and the vector of disturbances $u_t \sim NIID(0, \Sigma)$. On the other hand, M_2 is a Markov-switching VECM which allows the intercept and the variance-covariance matrix to be regime-shifting:

$$\Delta y_t = \nu(z_t) + \Pi y_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + \omega_t, \tag{26}$$

where $y_t = [s_t, f_t]$ $\omega_t \sim NIID(0, \Sigma)$, and the number of regimes $z_t = 1,2$. In this application, we set the number of regimes equal to two for convenience.

---

[14] Specifically, Models M_1 and M_2 are bivariate versions of the VECMs proposed by Clarida and Taylor (1997) and CSTV (2001) respectively. Giorgio Valente is gratefully acknowledged for carrying out the estimation in this section of the paper.

In order to calculate the η test we first estimated the competing models (25) and (26) using weekly bilateral US dollar exchange rate data (domestic price of the foreign currency) vis-à-vis the Japanese yen and the pound sterling. Time series for bilateral dollar exchange rates and one-month forward rates over the sample period from 1979:1 to 2000:52 were obtained from the Bank for International Settlements. We estimated models (25) and (26) using data from 1979:1 to 1991:52, leaving data from 1992:1 to the end of the sample period for calculating out-of-sample dynamic one-week-ahead forecasts.[15] In order to take into account the possibility that the implied one-week-ahead forecast distribution might be time-varying, we split our forecasting period in four non-overlapping equally sized sub-periods: 1992:1–1994:13, 1994:14–1996:27, 1996:28–1998:39, 1998:40–2000:52. One-week-ahead forecasts were constructed recursively, 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.

Inspecting Figures 1 and 2, which show the predictive densities from the two competing models M_1 and M_2 together with the true predictive density for each exchange rate examined and for different sub-sample periods, a clear result arises. Across sub-periods, the predictive densities produced by the linear VECM (25) are more leptokurtic than the ones obtained from the Markov-switching VECM (26). Simple visual inspection of the graphs in Figures 1 and 2 show that the distance between the forecast density of the Markov-switching VECM (26) from the true predictive density is shorter than the distance between the predictive density of the linear VECM (25) and the true predictive density. This visual evidence is, in fact, supported by the results of the η test, reported in Table 1. For both exchange rates examined and across subperiods, the η test, calculated using 100 bootstrap replications, is positive and statistically significant, strongly rejecting the null hypothesis of equidistance of the competing predictive densities from the true predictive density. In turn, these results imply that the Markov-switching model (Model M_2) is superior to the linear model (Model M_1) in terms of out-of-sample forecasting performance, suggesting that nonlinearity plays some role in forecasting exchange rates.

C. Implications for Risk Management and VaR Analysis

We now further investigate the practical implications of the forecasting results reported in the previous sub-section in the context of a simple risk management problem. Given the predictions of the two competing models M_1 and M_2, assume that a risk manager wishes to quantify the one-week-ahead risk associated with an internationally diversified portfolio

---

[15] The lag length, p was set equal to unity for both the linear and regime-switching VECMs. Note that the models do not exploit all of the term structure of the forward premia, which is available up to 52 weeks ahead from the Bank for International Settlements. Nevertheless, we thought this was unnecessary given the merely illustrative nature of the present empirical exercise.

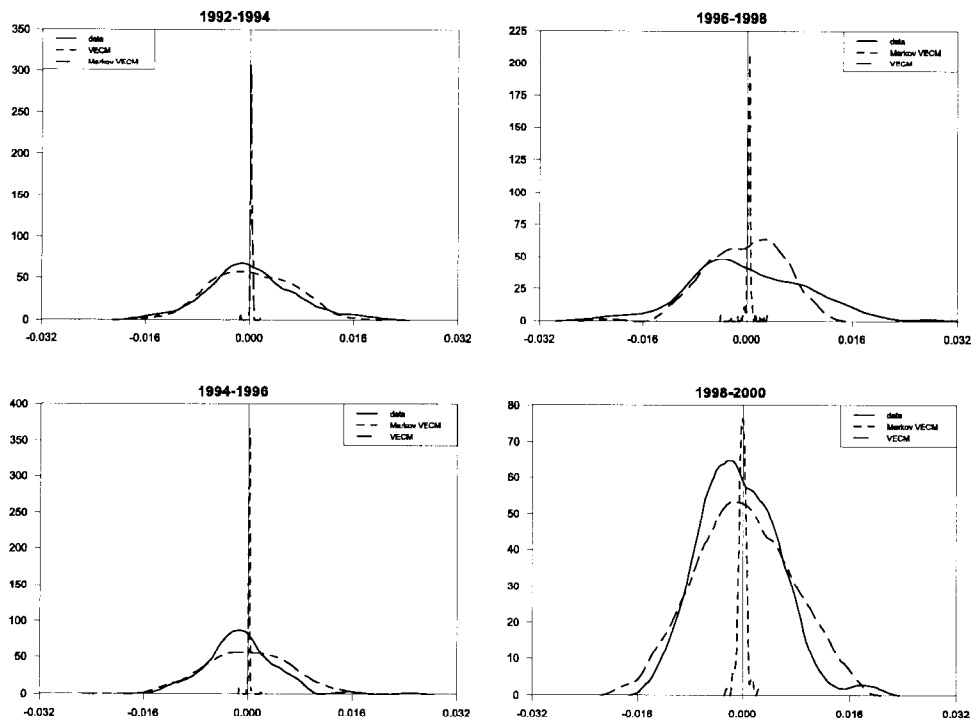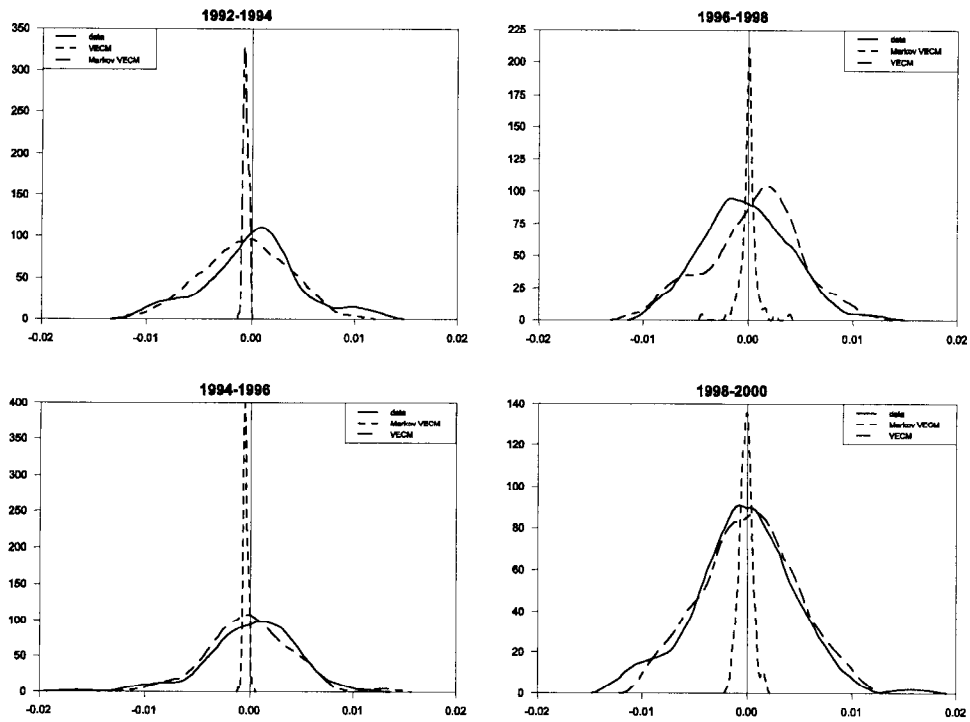# Figure 1. Kernel Density Estimation: Japan

Figure 2. Kernel Density Estimation: United Kingdom

selected at time t. For simplicity, assume that the portfolio comprises only one domestic bond and one foreign bond. These domestic and foreign bonds are identical in all respects except for the currency of denomination and yield the continuously compounded returns r and r^* respectively, expressed in local currency. Given initial wealth $W_t = 1$ and defining ω_t as the predetermined allocation to the foreign bond at time t, the end-of-week wealth is

$$W_{t+1} = (1 - \omega_t)\exp(r) + \omega_t \exp(r^* + \Delta s_{t+1}),\tag{27}$$

where Δs_{t+1} denotes the weekly change in the nominal exchange rate.[16] Given that the two bonds are riskless in local currency, the only source of uncertainty to be taken into consideration by the risk manager is the future nominal exchange rate, Δs_{t+1}. Models M_1 and M_2 provide the one-week-ahead density forecasts of Δs_{t+1}, which, in turn, determine implied densities for the end-of-week wealth. On the basis of these densities the risk manager calculates the VaR of the portfolio as the dollar loss relative to the mean:

$$VaR = E_t(W_{t+1}) - W_{t+1}^*,\tag{28}$$

where $E_t(W_{t+1})$ is the mean of the end-of-week wealth distribution and W_{t+1}^* is the lowest portfolio value at the given confidence level c. In our example the VaR is calculated according to equation (28) as a 99 percent confidence level for losses (i.e. $c = 0.99$), for both models M_1 and M_2.

The results, reported in Table 2, suggest that there is a substantial difference between the estimated values of the VaRs. For both exchange rates and across different subperiods, the linear VECM (25) generates VaRs that are at least ten times smaller than the ones implied by the Markov-switching VECM (26). In turn, this implies that the violations from the VaRs experienced over the sub-sample periods are larger under the linear VECM (25) than under the Markov-switching VECM (26). In fact, the violation rate, calculated as the ratio of the number of violations occurred over a subsample period to the total number of observations in the subsample period, is different for the two competing models, denoting a general tendency for the linear VECM (25) to produce VaRs that underestimate the probability of large losses. On the other hand, the Markov-switching VECM (26) produces VaRs that are generally in line with the theoretical 99 percent confidence level.

---

[16] In our empirical exercise we use one-month Eurorates for Japan, the United Kingdom, and the United States to proxy these riskless rates. r^* refers to the foreign interest rate (either for Japan or for the United Kingdom) and r refers to the domestic interest rate (for the United States). Data are sampled at weekly frequency and were obtained from the Bank for International Settlements.

Table 1. Exchange Rate Forecasting Results

	1992:1–1994:13	1994:14–1996:27	1996:28–1998:39	1998:40–2000:52
	Japan			
η test	7.047	5.373	9.621	7.999
	[0]	$[7.76 \times 10^{-8}]$	[0]	[0]
	United Kingdom			
η test	7.610	8.391	6.016	7.127
	[0]	[0]	[0]	[0]

Notes: Models M_1 and M_2 denote the linear VECM based on the spot-forward relationship (25) and the Markov-switching VECM (26) respectively. η test is the test statistic for the null hypothesis that Models M_1 and M_2 have equal density forecast accuracy, using 100 bootstrap replications. Figures in brackets denote p-values; p-values equal to zero up to the 8th decimal point are recorded as [0].

Table 2. VaR Estimation and Backtesting

Panel A, Japan

	1992:1–1994:13	1994:14–1996:27	1996:28–1998:39	1998:40–2000:52
	Model M_1: linear VECM			
VaR (mean)	−0.032	−0.038	−0.204	−0.063
No. of violations	57	52	50	47
Violation rate	0.487	0.444	0.427	0.401
$LR1$	74.508	73.822	73.307	72.271
	$[6.04 \times 10^{-18}]$	$[8.55 \times 10^{-18}]$	$[1.11 \times 10^{-17}]$	$[1.88 \times 10^{-17}]$
$LR2$	88.654	87.344	83.509	82.558
	$[5.61 \times 10^{-20}]$	$[1.08 \times 10^{-19}]$	$[7.35 \times 10^{-19}]$	$[1.18 \times 10^{-18}]$
$LR3$	14.147	13.522	10.203	10.288
	$[1.69 \times 10^{-4}]$	$[2.36 \times 10^{-4}]$	$[1.40 \times 10^{-3}]$	$[1.34 \times 10^{-3}]$
	Model M_2: Markov-switching VECM			
VaR (mean)	−0.591	−0.734	−0.913	−0.554
No. of violations	2	0	5	0
Violation rate	0.017	0	0.042	0
$LR1$	3.516	NA	13.839	NA
	[0.060]		$[1.99 \times 10^{-4}]$	
$LR2$	3.568	NA	14.103	NA
	[0.167]		$[8.66 \times 10^{-4}]$	
$LR3$	0.052	NA	0.184	NA
	[0.819]		[0.667]	

Table 2 (*continued*). VaR Estimation and Backtesting

Panel B, United Kingdom

	1992:1–1994:13	1994:14–1996:27	1996:28–1998:39	1998:40–2000:52
	Model M_1: linear VECM			
VaR (mean)	–0.036	–0.036	–0.246	–0.125
No. of violations	48	50	27	42
Violation rate	0.410	0.427	0.230	0.358
LR1	72.651	73.307	56.652	69.828
	$[1.55 \times 10^{-17}]$	$[1.11 \times 10^{-17}]$	$[5.20 \times 10^{-14}]$	$[6.47 \times 10^{-17}]$
LR2	83.987	86.209	62.561	80.436
	$[5.79 \times 10^{-19}]$	$[1.91 \times 10^{-19}]$	$[2.49 \times 10^{-14}]$	$[3.42 \times 10^{-18}]$
LR3	11.336	12.902	5.9096	10.607
	$[7.60 \times 10^{-4}]$	$[3.28 \times 10^{-4}]$	$[1.51 \times 10^{-2}]$	$[1.13 \times 10^{-3}]$
	Model M_2: Markov-switching VECM			
VaR (mean)	–0.668	–0.742	–0.830	–0.746
No. of violations	2	4	0	6
Violation rate	0.017	0.034	0	0.051
LR1	3.568	10.882	NA	17.114
	[0.058]	$[9.71 \times 10^{-4}]$		$[3.52 \times 10^{-5}]$
LR2	4.021	12.210	NA	18.112
	[0.133]	$[2.23 \times 10^{-3}]$		$[1.17 \times 10^{-4}]$
LR3	0.453	1.328	NA	0.998
	[0.500]	[0.249]		[0.317]

Notes: This table shows backtests of estimated VaR based upon the one-week-ahead forecasts obtained using the linear VECM (25) and the Markov-switching VECM (26) over four subperiods. VaR (mean) is the Value-at-risk relative to the mean calculated according to equation (28), using the 99 percent confidence level. The number (No.) of violations is calculated as the number of times the realized dollar loss exceeded the estimated VaR over the subsample period. The violation rate is calculated as the ratio of the number of violations to the total number of observations in the subsample period. LR1, LR2 and LR3 are the likelihood ratio tests proposed by Christoffersen (1998): LR1 is a test of unconditional coverage, where the null hypothesis is that the observed violation rate is not different from the target violation rate $c = 1–0.99 = 0.01$; LR2 is a test of conditional coverage where the null hypothesis is that the observed violation rate is not different from the target violation rate and the violations are i.i.d.; LR3 is a test of the null hypothesis that the violations are i.i.d. LR1 and LR2 are distributed as $\chi^2(1)$ and $\chi^2(2)$ under their respective null hypotheses; LR3 is calculated as the difference between LR2 and LR1 and is distributed as $\chi^2(1)$ under the null hypothesis.

This evidence tells us nothing, however, about the statistical significance of these differences. One way to shed light on the statistical significance of the differences is by comparing the targeted violation rate (i.e., 1–0.99 = 0.01) with the observed violation rate and by testing formally for their equality. In general, conventional testing procedures applied to testing for the statistical significance of the coverage rate (i.e. the difference between the target violation rate and the observed violation rate) are known to have low power (Kupiec, 1995; Christoffersen, 1998). Thus, in order to test for the statistical significance of the coverage rate, we calculated the likelihood ratio tests proposed by Christoffersen (1998), who first showed that not only violations should occur one percent of the time, but they should also be i.i.d. over time for them to be nonsystematic. These test statistics ($LR1$, $LR2$ and $LR3$)[17], reported in Table 1, confirm that the departures from VaR of the linear VECM (25) are strongly statistically significant. Indeed, all the likelihood ratio tests $LR1$, $LR2$ and $LR3$ exhibit very low p-values. On the other hand, the Markov-switching VECM (26) produces out-of-sample forecasts that imply VaRs associated with departures that are infrequent (generally close to the theoretical 99 confidence level) and often statistically insignificant at conventional significance levels.

In this simple application we have shown how density forecasts can help us discriminate among competing exchange rate models. In our example, the linear exchange rate model M_1 has produced forecasts that do not capture satisfactorily the higher moments of the predictive distribution of the exchange rate, generating VaRs that underestimate the probability of large losses. However, the Markov-switching model M_2, which does better than the linear model at matching the higher moments of the predictive distribution of exchange rates, produced VaRs that are generally in line with the target violation rate of one percent.

V. Concluding remarks

Exchange rate economics is alive and continues to attract the attention of academics, policymakers, and practitioners. One reason this field receives attention is the large number of puzzles the profession has been unable to resolve. This paper has reviewed a selection of recent papers that have shed some light on various puzzles. The common feature of the literature reviewed here is that it gives special attention to the importance of allowing for nonlinear dynamics in exchange rate behavior. Besides being a fancy way to improve on the

---

[17] $LR1$ is a test of unconditional coverage, where the null hypothesis is that the observed violation rate is not different from the target violation rate c = 1–0.99 = 0.01; $LR2$ is a test of conditional coverage where the null hypothesis is that the observed violation rate is not different from the target violation rate and the violations are i.i.d.; $LR3$ is a test of the null hypothesis that the violations are i.i.d. $LR1$ is distributed as $\chi^2(1)$ under the null hypothesis, whereas $LR2$ is distributed as $\chi^2(2)$ under the null hypothesis; $LR3$ is calculated as the difference between $LR2$ and $LR1$ and is distributed as $\chi^2(1)$ under the null hypothesis.

linear econometric techniques more commonly used by researchers, the nonlinear approach to exchange rate modeling and forecasting has produced results that can be viewed as very promising.

With respect to the behavior of the real exchange rate, the theoretical literature focusing on the importance of international trade costs has led researchers to consider nonlinear models of real exchange rates that give a role of the size of the deviation from equilibrium in determining the speed of mean reversion, a feature that cannot be captured within a linear framework. The empirical evidence provided by these studies suggests that major real bilateral dollar exchange rates are well characterized by nonlinearly mean reverting processes over the floating rate period since 1973. These models imply an equilibrium level of the real exchange rate in the neighborhood of which the behavior of the real exchange rate is close to a random walk, becoming increasingly mean reverting with the absolute size of the deviation from equilibrium. In addition, the half lives of shocks to the real exchange rates implied by these models suggest much faster real exchange rate adjustment than typically recorded in the literature, hence shedding some light on Rogoff's (1996) PPP puzzle.

With respect to the ability of empirical exchange rate models to explain and forecast the nominal exchange rate, the literature is still somewhat in the dark. About 20 years after Meese and Rogoff's (1983) paper, their findings that empirical exchange rate models are not able to beat a random walk have not been convincingly overturned. Economic fundamentals typically suggested by open-economy macro theory do not appear to contain sufficient information as to provide satisfactory out-of-sample forecasts of the exchange rate, especially at short horizons. However, the information contained in the term structure of forward premia appears to be more useful, especially in empirical models that allow for nonlinearities. An analysis of spot and forward exchange rates in a multivariate Markov-switching framework, 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 modeling, produces very satisfactory forecasting results. Indeed, this framework generates forecasts that are 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, this model also outperforms its linear counterpart in out-of-sample forecasting, although the magnitude of the gain from using a nonlinear relative to a linear one is rather small in magnitude at short horizons.

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. Especially when evaluating nonlinear models, which are capable of producing highly nonnormal forecast densities, it would seem appropriate to consider a model's density forecasting performance. In the last part of this paper, we have presented a recently developed testing procedure designed to discriminate between competing forecasting models in terms of density forecast accuracy. An illustration of the potential use of this framework in the

context of exchange rate forecasting highlights the benefits of using a nonlinear model relative to a linear one to be in the forecasts of the higher moments of the exchange rate distribution. In other words, not only does this nonlinear framework produce better point forecasts of the level of the exchange rate, but it also provides a better measure of the uncertainty surrounding these forecasts. This seems particularly important for risk management and also for the policymaker interested in forecasting the probability of large changes in the exchange rate (depreciations or appreciations). The probability of large exchange rate movements risks being seriously underestimated when relying on linear models.

Overall, we end this survey with some degree of optimism. Nonlinear exchange rate models may not be the solution to every puzzle in exchange rate economics, but they have led us to a stage where the faith in PPP professed by many international economists for many decades shows some scientific support, and where we can provide forecasts of exchange rate movements that are better than a model that simply assumes no change.

REFERENCES

Abuaf, N., and P. Jorion, 1990, "Purchasing Power Parity in the Long Run," *Journal of Finance*, Vol. 45, pp. 157–174.

Anderson, N.H., P. Hall, and D.M. Titterington, 1994, "Two-Sample Test Statistics for Measuring Discrepancies Between Two Multivariate Probability Density Functions Using Kernel-Based Density Functions," *Journal of Multivariate Analysis*, Vol. 50, pp. 41–54.

Baxter, M., and A.C. Stockman, 1989, "Business Cycles and the Exchange Rate System," *Journal of Monetary Economics*, Vol. 23, pp. 377–400.

Benninga, S., and A.A. Protopapadakis, 1988, "The Equilibrium Pricing of Exchange Rates and Assets When Trade Takes Time," *Journal of International Economics*, Vol. 7, pp.129–149.

Berkowitz, J., 2001, "Testing Density Forecasts, with Applications to Risk Management," *Journal of Business and Economic Statistics*, Vol. 19, pp. 465–474.

Berkowitz, J., and L. Giorgianni, 2001, "Long-Horizon Exchange Rate Predictability?" *Review of Economics and Statistics*, Vol. 83, pp. 81–91.

Berben, R.-P., and D. van Dijk, 1998, "Does the Absence of Cointegration Explain the Typical Findings in Long-Horizon Regressions?" (unpublished; Rotterdam: Tinbergen Institute, Erasmus University).

Beveridge, S., and C.R. Nelson, 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*, Vol. 7, pp. 151–174.

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

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

————, 1989, "Predictors in Dynamic Nonlinear Models: Large Sample Behaviour," *Econometric Theory*, Vol. 5, pp. 430–452.

Cassel, G., 1922, *Money and Foreign Exchange After 1914* (London: Constable).

Cheung, Y.-W., and K.S. Lai, 1993a, "A Fractional Cointegration Analysis of Purchasing Power Parity," *Journal of Business and Economic Statistics*, Vol. 11, pp. 103–112.

———, 1993b, "Long-Run Purchasing Power Parity During the Recent Float," *Journal of International Economics*, Vol. 34, pp. 181–192.

———, 1994, "Mean Reversion in Real Exchange Rates," *Economics Letters*, Vol. 46, pp. 251–256.

———, 1998, "Parity Reversion in Real Exchange Rates During the Post-Bretton Woods Period," *Journal of International Money and Finance*, Vol. 17, pp. 597–614.

Christoffersen, P.F., 1998, "Evaluating Interval Forecasts," *International Economic Review*, Vol. 39, pp. 841–862.

Clarida, R. H., and M. P. Taylor, 1997, "The Term Structure of Forward Exchange Premiums and the Forecastability of Spot Exchange Rates: Correcting the Errors," *Review of Economics and Statistics*, Vol. 79, pp. 353–361.

Clarida, R. H., L. Sarno, M. P. Taylor, and G. Valente, 2003, "The Out-of-Sample Success of Term Structure Models as Exchange Rate Predictors: A Step Beyond," *Journal of International Economics*, Vol. 60, pp. 61–83.

Corbae, D., and S. Ouliaris, 1988, "Cointegration and Tests of Purchasing Power Parity," *Review of Economics and Statistics*, Vol. 70, pp. 508–511.

Davutyan, N., and J. Pippenger, 1990, "Testing Purchasing Power Parity: Some Evidence of the Effects of Transactions Costs," *Econometric Reviews*, Vol. 9, pp. 211–240.

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

Diebold, F. X., and J. A. Nason, 1990, "Nonparametric Exchange Rate Prediction?" *Journal of International Economics*, Vol. 3-4, pp. 315–332.

Diebold, F. X., S. Husted, and M. Rush, 1991, "Real Exchange Rates Under the Gold Standard," *Journal of Political Economy*, Vol. 99, pp. 1252–1271.

Diebold, F. X., and R. S. Mariano, 1995, "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics*, Vol. 13, pp. 253–263.

Diebold, F. X., and J. A. Lopez, 1996, "Forecast Evaluation and Combination," in *Handbook of Statistics,* ed. by G. S. Maddala and C. A. Rao (Amsterdam, New York and Oxford: Elsevier, North-Holland), Vol.14, pp. 241–268.

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

Diebold, F. X., J. Hahn, and A. S. Tay, 1999, "Multivariate Density Forecast Evaluation and Calibration in Financial Risk Management: High-Frequency Returns on Foreign Exchange," *Review of Economics and Statistics*, Vol. 81, pp. 661–673.

Dumas, B., 1992, "Dynamic Equilibrium and the Real Exchange Rate in Spatially Separated World," *Review of Financial Studies*, Vol. 5, pp. 153–180.

————, 1994, "Partial Equilibrium Versus General Equilibrium Models of the International Capital Market," Chapter 10 in *The Handbook of International Macroeconomics*, ed. by F. Van Der Ploeg (Oxford: Blackwell).

Engel, C., 1994, "Can the Markov Switching Model Forecast Exchange Rates?" *Journal of International Economics*, Vol. 36, pp. 151–165.

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

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

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

Frankel, J. A., 1979, "On the Mark: A Theory of Floating Exchange Rates Based on Real Interest Differentials," *American Economic Review*, Vol. 69, pp. 610–622.

————, 1986, "International Capital Mobility and Crowding Out in the US Economy: Imperfect Integration of Financial Markets or Goods Markets?" in *How Open is the US Economy?*, ed. by R. W. Hafer (Lexington: Lexington Books).

————, 1990, "Zen and the Art of Modern Macroeconomics: The Search for Perfect Nothingness," in *Monetary Policy for a Volatile Global Economy*, ed. by W. Haraf and T.Willett (Washington DC: American Enterprise Institute).

Frankel, J. A. and A. K. Rose, 1996, "A Panel Project on Purchasing Power Parity: Mean Reversion Within and Between Countries," *Journal of International Economics*, Vol. 40, pp. 209–224.

Froot, K. A. and K. Rogoff, 1995, "Perspectives on PPP and Long-Run Real Exchange Rates," in *Handbook of International Economics* , ed. by K. Rogoff and G. Grossman (Amsterdam: North Holland).

Gallant, A. R., P. E. Rossi, and G. Tauchen, 1993, "Nonlinear Dynamics Structures," *Econometrica*, Vol. 61, pp. 871–907.

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

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

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

Hall, P., J. L. Horowitz, and B.-Y. Jing, 1995, "On Blocking Rules for the Bootstrap with Dependent Data," *Biometrika*, Vol. 82, pp. 561–574.

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*, Vol. 12, pp. 385–423.

————, 1989, "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, Vol. 57, pp. 357–384.

Hegwood, N. D., and D. H. Papell, 1998, "Quasi Purchasing Power Parity," *International Journal of Finance and Economics*, Vol. 3, pp. 279–289.

Hooper, P. and J. Morton, 1982, "Fluctuations in the Dollar: A Model of Nominal and Real Exchange Rate Determination," *Journal of International Money and Finance*, Vol. 1, pp. 39–56.

Johansen, S., 1995, *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models* (Oxford: Oxford University Press).

Jorion, P., 1997, *Value at Risk* (New York: Dow Jones Irvin).

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

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

Kim, Y., 1990, "Purchasing Power Parity in the Long Run: A Cointegration Approach," *Journal of Money, Credit, and Banking*, Vol. 22, pp. 491–503.

Koedijk, K. G., and P. Schotman, 1990, "How to Beat the Random Walk: An Empirical Model of Real Exchange Rates," *Journal of International Economics*, Vol. 29, pp. 3–4.

Kupiec, P. H., 1995, "Techniques for Verifying the Accuracy of Risk Measurement Models," *Journal of Derivatives*, Winter, pp. 73–84.

Lane, P. R., 2001, "The New Open Economy Macroeconomics: A Survey," *Journal of International Economics*.

LeBaron, B., 1992, "Forecast Improvements Using a Volatility Index," *Journal of Applied Econometrics*, Vol. 7, Supplement, pp. 137–49.

Leon, H. and S. Najarian, 2003, "Time-Varying Thresholds: An Application to Purchasing Power Parity," (unpublished; Washington D.C.: International Monetary Fund, IMF Institute).

Li, Q., 1996, "Nonparametric Testing of Closeness Between Two Unknown Distribution Functions," *Econometric Reviews*, Vol. 15, pp. 261–274.

Lothian, J. R., and M. P. Taylor, 1996, "Real Exchange Rate Behaviour: The Recent Float from the Perspective of the Past Two Centuries," *Journal of Political Economy*, Vol. 104, pp. 488–510.

————, 1997, "Real Exchange Rate Behaviour: The Problem of Power and Sample Size," *Journal of International Money and Finance*, Vol. 16, pp. 945–954.

Lucas, R. E. Jr., 1976, "Econometric Policy Evaluation: A Critique," *Journal of Monetary Economics 1*, Supplementary Series, pp. 19–46.

MacDonald, R., and M. P. Taylor, 1993, "The Monetary Approach to the Exchange Rate: Rational Expectations, Long-Run Equilibrium, and Forecasting," *Staff Papers*, International Monetary Fund, Vol. 40, No. 1, pp. 89–107.

————, 1994, "The Monetary Model of the Exchange Rate: Long-Run Relationships, Short-Run Dynamics and How to Beat a Random Walk," *Journal of International Money and Finance*, Vol. 13, pp. 276–290.

Mammen, E., 1992, *When Does Bootstrap Work: Asymptotic Results and Simulations*, Lecture Notes in Statistics, Vol. 77 (Berlin: Springer-Verlag).

Mark, N.C., 1990, "Real and Nominal Exchange Rates in the Long Run: An Empirical Investigation," *Journal of International Economics*, Vol. 28, pp. 115–136.

————, 1995, "Exchange Rates and Fundamentals: Evidence on Long-Horizon Predictability," *American Economic Review*, Vol. 85, pp. 201–218.

McCloskey, D. N., and J. R. Zecher, 1984, "The Success of Purchasing Power Parity: Historical Evidence and Its Implications for Macroeconomics," in *A Retrospective on the Classical Gold Standard, 1821-1931*, ed. by M. D. Bordo and A. J. Schwartz, National Bureau of Economic Research Conference Report Series (Chicago: University of Chicago Press), pp. 121–150.

McCracken, M. W., 2000, "Robust Out-of-Sample Inference," *Journal of Econometrics*, Vol. 99, pp. 195–223.

McNown, R., and M. Wallace, 1989, "National Price Levels, Purchasing Power Parity, and Cointegration: A Test of Four High Inflation Economies," *Journal of International Money and Finance*, Vol. 8, pp. 533–545.

Meese, R. A., and K. Rogoff, 1983, "Empirical Exchange Rate Models of the Seventies: Do They Fit Out of Sample?" *Journal of International Economics*, Vol.14, pp. 3–24.

Meese, R. A., and A. K. Rose, 1990, "Nonlinear, Nonparametric, Nonessential Exchange Rate Estimation," *American Economic Review*, Vol. 80, pp. 192–196.

————, 1991, "An Empirical Assessment of Non-linearities in Models of Exchange Rate Determination," *Review of Economic Studies*, Vol. 58, pp. 603–619.

Michael, P., A. R. Nobay, and D. A. Peel, 1997, "Transactions Costs and Nonlinear Adjustment in Real Exchange Rates: An Empirical Investigation," *Journal of Political Economy*, Vol. 105, pp. 862–879.

Nobel, A. B., G. Morvai, and S. R. Kulkarni, 1998, "Density Estimation from an Individual Numerical Sequence," *IEEE Transactions on Information Theory*, Vol. 44, pp. 537–541.

Obstfeld, M., and A. M. Taylor, 1997, "Nonlinear Aspects of Goods-Market Arbitrage and Adjustment: Heckscher's Commodity Points Revisited," *Journal of the Japanese and International Economies*, Vol. 11, pp. 441–479.

Obstfeld, M., K. and Rogoff, 2000, "The Six Major Puzzles in International Macroeconomics: Is There a Common Cause?" in *NBER Macroeconomics Annual 2000*, ed. by B. Bernanke and K. Rogoff (Cambridge, Mass: MIT Press), pp. 339–390.

O'Connell, P. G. J., 1998, "The Overvaluation of Purchasing Power Parity," *Journal of International Economics*, Vol. 44, pp. 1–19.

O'Connell, P. G. J., and S.-J. Wei, 2002, "The Bigger They Are, the Harder They Fall: The Retail Price Differences Across U.S. Cities," *Journal of International Economics*, Vol. 56, pp. 21–54.

Officer, L. H., 1982, *Purchasing Power Parity and Exchange Rates: Theory, Evidence and Relevance* (Greenwich, CT: JAI Press).

Papell, D. H., 1998, "Searching for Stationarity: Purchasing Power Parity Under the Current Float," *Journal of International Economics*, Vol. 43, pp. 313–332.

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

Rogoff, K., 1996, "The Purchasing Power Parity Puzzle," *Journal of Economic Literature*, Vol. 34, pp. 647–668.

Sarno, L., 2001, "Toward a New Paradigm in Open Economy Modeling: Where Do We Stand?" *Federal Reserve Bank of St. Louis Review*, Vol. 83, pp. 21–36.

Sarno, L., and M. P. Taylor, 1998, "Real Exchange Rates Under the Recent Float: Unequivocal Evidence of Mean Reversion," *Economics Letters*, Vol. 60, pp.131–137.

———, 2002, "Purchasing Power Parity and the Real Exchange Rate," *Staff Papers, International Monetary Fund*, Vol. 49, pp. 65–105.

Sarno, L., M. P. Taylor, and I. Chowdhury, 2003, "Nonlinear Dynamics in Deviations from the Law of One Price: A Broad-Based Empirical Study," *Journal of International Money and Finance*, forthcoming.

Sarno, L., and G. Valente, 2003 "Comparing the Accuracy of Density Forecasts from Competing Models," *Journal of Forecasting*, forthcoming.

Schinasi, G. J., and P. A. V. B. Swamy, 1989, "The Out-of-Sample Forecasting Performance of Exchange Rate Models When Coefficients Are Allowed to Change," *Journal of International Money and Finance*, Vol. 8, pp. 375–390.

Sercu, P., R. Uppal, and C. Van Hulle, 1995, "The Exchange Rate in the Presence of Transactions Costs: Implications for Tests of Purchasing Power Parity," *Journal of Finance*, Vol. 50, pp. 1309–1319.

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

Taylor, A. M., 2002, "A Century of Purchasing-Power Parity," *Review of Economics and Statistics*, Vol. 84, pp. 139–150.

Taylor, M. P., 1988, "An Empirical Examination of Long-Run Purchasing Power Parity Using Cointegration Techniques," *Applied Economics*, Vol. 20, pp. 1369–1381.

Taylor, M. P., and P. C. McMahon, 1988, "Long-Run Purchasing Power Parity in the 1920s," *European Economic Review*, Vol. 32, pp. 179–197.

Taylor, M. P., and L. Sarno, 1998, "The Behaviour of Real Exchange Rates During the Post-Bretton Woods Period," *Journal of International Economics*, Vol. 46, pp. 281–312.

Taylor, M. P., D. A. Peel, and L. Sarno, 2001, "Nonlinear Mean-Reversion in Real Exchange Rates: Toward a Solution to the Purchasing Power Parity Puzzles," *International Economic Review*, Vol. 42, pp. 1015–1042.

Teräsvirta, T., 1994, "Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models," *Journal of the American Statistical Association,* Vol. 89, pp. 208–218.

Tong, H., 1990, *Non-Linear Time Series: A Dynamical System Approach* (New York: Oxford University Press).

West, K. D., 1996, "Asymptotic Inference About Predictive Ability," *Econometrica*, Vol. 64, pp. 1067–1084.

West, K. D., and M. W. McCracken, 1998, "Regression-Based Tests of Predictive Ability," *International Economic Review*, Vol. 39, pp. 817–840.

Williams, J. C., and B. D. Wright, 1991, *Storage and Commodity Markets* (Cambridge: Cambridge University Press).

Yamato, H., 1971, "Sequential Estimation of a Continuous Probability Function and Mode," *Bulletin of Mathematical Statistics*, Vol. 14, pp. 1–12.