The Out-of-Sample Forecasting Performance of Nonlinear Models of Real Exchange Rate Behavior

David E. Rapach*
Department of Economics
Saint Louis University
3674 Lindell Boulevard
Saint Louis, MO 63108-3397
Phone: 314-977-3601
Fax: 314-977-1478
E-mail: rapachde@slu.edu

Mark E. Wohar
Department of Economics
University of Nebraska at Omaha
RH 512K
Omaha, NE 68182-0286
Phone: 402-554-3712
Fax: 402-554-2853
E-mail: mwohar@mail.unomaha.edu

September 7, 2005 (Revised)

*Corresponding author. This is a significantly revised version of our earlier paper, "Nonlinear Models of Real Exchange Rate Behavior: A Re-examination." We thank two anonymous referees, an Associate Editor, and an Editor (Michael Clements), as well as Nathan Balke, Mike Bergman, Christian Dahl, Tom Fomby, David Harvey, Steve Leybourne, Mike McCracken, Adrian Pagan, Hailong Qian, Mark Salmon, Lucio Sarno, Aaron Smallwood, Jack Strauss, and Dick van Dijk for very helpful comments. We are also grateful to seminar participants at the University of Alabama, University of California at San Diego, University of Nottingham, University of Oklahoma, University of Warwick, Saint Louis University, Southern Methodist University, and the 2003 Missouri Economics Conference. The usual disclaimer applies. The results reported in this paper were generated using GAUSS 5.0. The GAUSS programs are available at http://pages.slu.edu/faculty/rapachde/Research.htm.

The Out-of-Sample Forecasting Performance of Nonlinear Models of Real Exchange Rate Behavior

Abstract

We analyze the out-of-sample forecasting performance of nonlinear models of U.S. dollar real exchange rate behavior from the extant empirical literature. Our analysis entails a comparison of point, interval, and density forecasts generated by nonlinear and linear autoregressive models. Using monthly data from the post-Bretton Woods period, there is little evidence to recommend either band-threshold or exponential smooth transition autoregressive models over simple linear autoregressive models in terms of out-of-sample forecasting performance at short horizons. Nonlinear models appear to offer more accurate point forecasts at long horizons for some countries. Overall, our results suggest that any nonlinearities in monthly real exchange rate data from the post-Bretton Woods period are quite "subtle" for band-threshold and exponential smooth transition autoregressive model specifications. Further evidence of this is provided by in-sample comparisons of the conditional densities implied by nonlinear and linear autoregressive models.

JEL classifications: C22, C52, C53, F31, F47

Key words: Real exchange rate; Transaction costs; Band-threshold autoregressive model; Exponential smooth transition autoregressive model; Point forecast; Interval forecast; Density forecast

Biographies: David E. Rapach is an Assistant Professor of Economics at Saint Louis University. His research interests include time series econometrics, macroeconomics, international finance, and financial economics. He has published in numerous journals, including *Economic Inquiry*, *International Journal of Forecasting*, *Journal of Applied Econometrics*, *Journal of Forecasting*, *Journal of International Economics*, *Journal of International Money and Finance*, *Journal of Macroeconomics*, and *Journal of Money, Credit, and Banking*.

Mark E. Wohar is the Distinguished Enron Professor of Economics and Finance at the University of Nebraska at Omaha. His research interests include time series econometrics, macroeconomics, international finance, financial institutions, and financial economics. He has published more than 65 articles in economic and financial journals, including the *American Economic Review*, *International Journal of Forecasting*, *Journal of Applied Econometrics*, *Journal of Finance*, *Journal of International Economics*, *Journal of International Money and Finance*, *Journal of Money, Credit, and Banking*, and *Review of Economics and Statistics*.

1. Introduction

There is growing interest in nonlinear models of real exchange rate behavior in the empirical international finance literature. This is not surprising, as nonlinear model specifications of real exchange rate behavior are well motivated by theoretical models incorporating transaction costs.[1] Transaction costs can be broadly defined to include transportation costs, tariffs and nontariff barriers, as well as any other costs that agents incur in international trade (Obstfeld & Rogoff, 2000). Intuitively, transaction costs give rise to a band of inactivity where arbitrage is not profitable, so that nominal exchange rate deviations from purchasing power parity (real exchange rate fluctuations) are not corrected inside of the band. However, if the real exchange rate moves outside of the band, arbitrage works to bring the real exchange rate back to the edge of the band.

Motivated by theoretical models incorporating transaction costs, two recent studies estimate nonlinear autoregressive (AR) models for U.S. dollar real exchange rates over the post-Bretton Woods period. Obstfeld & Taylor (1997) estimate band-threshold AR (Balke & Fomby, 1997; Band-TAR) models for a large number of U.S. dollar real exchange rates based on both broad and disaggregated consumer price indices. In line with the theoretical models cited above, the Band-TAR model is characterized by unit-root behavior in an inner regime and reversion to the edge of the unit-root band in an outer regime. A second study, Taylor, Peel, & Sarno (2001), considers exponential smooth transition AR (Granger & Teräsvirta, 1993; ESTAR) models of U.S. dollar real exchange rate behavior. In contrast to the discrete regime switching that characterizes the Band-TAR model, the ESTAR model allows for smooth transition between regimes.[2] As pointed out by Taylor, Peel, & Sarno (2001), Bertola & Caballero (1990), Dumas (1994), and Teräsvirta (1994) suggest that time aggregation and non-synchronous adjustment by heterogeneous agents is likely to lead to smooth regime switching, rather than discrete switching, and this is especially likely to be the case for real exchange rates based on broad price indices.

---

[1] Theoretical models incorporating transaction costs include Benninga & Protopapadakis (1988), Williams & Wright (1991), Dumas (1992), Coleman (1995), Sercu, Uppal, & Van Hulle (1995), Ohanian & Stockman (1997), O'Connell (1998), and Obstfeld & Rogoff (2000).

[2] Like the Band-TAR model, the ESTAR model is characterized by symmetric adjustment.

Using monthly data for the U.S. dollar real exchange rate vis-à-vis the U.K., Germany, France, and Japan over the post-Bretton Woods era, Taylor, Peel, & Sarno (2001) estimate a parsimonious ESTAR model for each country. For their ESTAR models, the real exchange rate follows a random walk in the extreme inner regime near the long-run equilibrium, while the speed of reversion to the long-run equilibrium increases the farther the real exchange rate deviates from the long-run equilibrium. Taylor, Peel, & Sarno (2001) conclude that the real exchange rates they consider are well characterized by nonlinear mean-reversion.[3]

Obstfeld & Taylor (1997) and Taylor, Peel, & Sarno (2001) report evidence of nonlinear behavior in U.S. dollar real exchange rates. All of the evidence reported in these well-known and oft-cited papers is based on in-sample tests. In the present paper, we add to the existing empirical literature on nonlinear real exchange rate behavior by undertaking an extensive evaluation of the out-of-sample forecasting performance of the nonlinear AR models from these papers. Tests of out-of-sample forecasting performance are widely viewed as an important component of model evaluation and a way of guarding against model overfitting. In our evaluation, we simulate the situation of a forecaster who uses the fitted nonlinear AR models from Obstfeld & Taylor (1997) and Taylor, Peel, & Sarno (2001) to forecast real exchange rate observations that have become available since the models were originally estimated.[4] We compare the out-of-sample real exchange rate forecasts generated by these fitted nonlinear AR models to out-of-sample forecasts generated by fitted linear AR models. If the forecasts generated by nonlinear AR models are superior to those generated by simple linear AR models, this can be construed as strong empirical evidence in favor of nonlinear model specifications.

We first compare the out-of-sample forecasting performance of nonlinear and linear AR models

---

[3] Sarantis (1999) also estimates ESTAR models for real exchange rates using monthly data from the post-Bretton Woods period. However, he estimates ESTAR models under the assumption that real exchange rate levels are nonstationary, while theoretical models suggest estimating nonlinear models under the assumption that real exchange rate levels are globally stationary.

[4] Michael, Nobay, & Peel (1997) also estimate ESTAR models for U.S. dollar real exchange rates using the Lothian & Taylor (1996) annual data that cover more than two centuries. We do not consider this study in detail in the present paper, as there are only a relatively small number of annual observations that have become available since the models were originally estimated.

in terms of mean squared forecast error (MSFE), and we test whether the nonlinear AR model forecasts are significantly superior to the linear AR model forecasts using the popular Diebold & Mariano (1995) test. A number of studies have pointed out that, despite in-sample evidence of nonlinear behavior, nonlinear models typically offer small forecasting gains relative to linear models in terms of MSFE (Diebold & Nason, 1990; De Gooijer & Kumar, 1992; Ramsey, 1996; Stock & Watson, 1999), and it may be the case that forecasting gains associated with nonlinear models will only be evident in certain regimes (Montgomery *et al*., 1998; Clements & Smith, 1999). In light of this, we also use a weighted version of the Diebold & Mariano (1995) test developed by van Dijk & Franses (2003) in order to focus on forecasting real exchange rate observations in the tails of the unconditional distribution. This approach is especially well suited to the applications in the present paper, as theoretical models of nonlinear real exchange rate behavior predict that the adjustment process will be faster for real exchange rate realizations that are far from equilibrium. In an application, van Dijk & Franses (2003) find that forecasts of U.S. output growth generated by the nonlinear "floor and ceiling" model of Pesaran & Potter (1997) are not superior to forecasts generated by a linear AR model according to the conventional Diebold & Mariano (1995) test; however, the floor and ceiling model forecasts are superior to linear AR model forecasts according to the weighted Diebold & Mariano (1995) test.

Our evaluation also includes an analysis of interval and density forecasts generated by nonlinear and linear AR models, and we analyze interval and density forecasts along the lines suggested by Christoffersen (1998) and Diebold, Gunther, & Tay (1998). A number of recent studies have found that—despite failing to produce superior point forecasts relative to linear models—nonlinear models produce superior interval and density forecasts; see, for example, Clements & Smith (2000) with respect to forecasting U.S. GNP growth and unemployment and Siliverstovs & van Dijk (2003) with respect to industrial production growth in the G7 countries. By using the weighted version of the Diebold & Mariano (1995) test and evaluating interval and density forecasts in addition to point forecasts, we hope to maximize the opportunity for any out-of-sample forecasting gains associated with the nonlinear real exchange rate models to become evident (if they exist).

The rest of the paper is organized as follows. Section 2 reviews in more detail the studies of Obstfeld & Taylor (1997) and Taylor, Peel, & Sarno (2001). Section 3 outlines the econometric tests we use to evaluate point, interval, and density forecasts. Section 4 presents the empirical results. Section 5 compares in-sample conditional densities corresponding to nonlinear and linear AR models in an effort to better understand the out-of-sample forecasting results. Section 6 concludes.

2. Review of two extant studies

2.1. Obstfeld & Taylor (1997) Band-TAR model

Motivated by theoretical models of costly arbitrage due to transport costs, Obstfeld & Taylor (1997, OT) estimate Band-TAR models for consumer price index-based U.S. dollar real exchange rates.[5] Their Band-TAR model takes the form,

$$\begin{aligned} \Delta q_t &= \lambda_{out} \cdot (q_{t-1} - c) + \varepsilon_t^{out} \text{ if } q_{t-1} > c\,; \\ \Delta q_t &= \varepsilon_t^{in} \text{ if } c \geq q_{t-1} \geq -c\,; \qquad (1) \\ \Delta q_t &= \lambda_{out} \cdot (q_{t-1} + c) + \varepsilon_t^{out} \text{ if } -c > q_{t-1}\,, \end{aligned}$$

where q_t is the log-level of the real exchange rate, Δ is the first-difference operator, $\varepsilon_t^{out} \sim N(0, \sigma_{out}^2)$, and $\varepsilon_t^{in} \sim N(0, \sigma_{in}^2)$.[6] From equation (1), it is evident that the real exchange rate follows a random walk inside the "band of inaction" defined by $[-c, c]$, as, for example, transaction costs prevent arbitrage from correcting real exchange rate disturbances inside of the band; outside of the band, arbitrage forces correct deviations so that the real exchange rate moves back to the edge of the band when $\lambda_{out} < 0$.

[5] OT also estimate Band-TAR models for U.S. dollar real exchange rates based on various categories of consumer price indices in order to test the law of one price. We concentrate on their overall consumer price index-based real exchange rates, as such real exchange rates are the focus of the present paper. We discuss nonlinear real exchange rate models based on disaggregated consumer price indices in Section 6.

[6] OT measure the log-level of the real exchange rate as $q_t = s_t + p_t^* - p_t$, where s_t is the log-level of the U.S. dollar/foreign currency exchange rate, p_t is the log-level of the U.S. consumer price index, and p_t^* is the log-level of the consumer price index for the relevant country. Before entering equation (1), q_t is demeaned or detrended.

OT estimate equation (1) using maximum likelihood and monthly data covering 1980-1994 for a large number of countries. We concentrate on their results for the U.K., Germany, France, and Japan, as this is the same set of countries considered by Taylor, Peel, & Sarno (2001). OT obtain estimates of λ_{out} that are negative and sometimes fairly sizable in absolute value. Using the Tsay (1989) test for a TAR alternative against a linear AR null hypothesis, they reject the linear null hypothesis for France and Germany.

Using data from the International Monetary Fund's *International Financial Statistics* database, we estimate the OT Band-TAR model, equation (1), for the U.K., Germany, France, and Japan. We use the same sample period as OT for each country: 1980:01-1994:11 (1980:01-1994:12) for the U.K., France, and Japan (Germany). One observation is lost when we allow for the lag in equation (1).[7] Following OT, we estimate equation (1) via maximum likelihood. We implement maximum likelihood estimation via a grid search over possible values of *c*, and we require the outer regime to contain at least 15% of the observations for q_{t-1}. The estimated Band-TAR models for each country are reported in Panel A of Table 1, as well as the number of observations in the outer and inner regimes (n_{out} and n_{in}, respectively). The parameter estimates are similar to those reported in OT.[8]

2.2. Taylor, Peel, & Sarno (2001) ESTAR model

After testing down from a more general ESTAR specification, Taylor, Peel, & Sarno (2001, TPS) consider the following parsimonious ESTAR model,

$$q_t = q_{t-1} - \{1 - \exp[\alpha \cdot (q_{t-1} - \eta)^2]\} \cdot (q_{t-1} - \eta) + \varepsilon_t , \tag{2}$$

[7] Following OT, we first demean q_t before estimating equation (1). OT (1997) report detailed results for detrended q_t, but note that the results are qualitatively similar when they use demeaned data. We report results for demeaned data, as this is consistent with the treatment of the real exchange rate in the nonlinear models in Taylor, Peel, & Sarno (2001). Our results are qualitatively unchanged when we use detrended data.

[8] OT do not report a standard error or *t*-statistic for their estimate of c. We generate a standard error for our estimate of c using a parametric bootstrap procedure.

where q_t is stationary and ergodic, $\varepsilon_t \sim iid(0,\sigma_\varepsilon^2)$, and η is the long-run equilibrium level for q_t.[9] For the parsimonious ESTAR model, equation (2), the real exchange rate behaves as a random walk in the extreme inner regime ($q_{t-1} = \eta$), and the speed of mean-reversion increases as the real exchange rate moves away from its long-run equilibrium value (assuming $\alpha < 0$).[10] Using monthly data from 1973:01-1996:12, TPS estimate equation (2) using multivariate nonlinear least squares (iterative nonlinear seemingly unrelated regressions) for the real exchange rate for four countries relative to the U.S. dollar: U.K., Germany, France, and Japan.[11] TPS obtain negative and statistically significant estimates of α for each of the four real exchange rates. For each country, the fitted parsimonious ESTAR model passes Eitrheim & Teräsvirta (1996) tests for no remaining serial correlation in the residuals, no remaining ESTAR nonlinearity with delay from 2-12 months, and no remaining logistic nonlinearity. TPS also use their estimated ESTAR models to calculate impulse responses,[12] and they find that "large" shocks to the real exchange rate have half-lives that are considerably shorter than the "glacial" half-lives cited by Rogoff (1996).

We estimate equation (2) via multivariate nonlinear least squares using data from the *International Financial Statistics* database and the same sample period as TPS (1973:02-1996:12 after allowing for the lag in equation (2)). The estimated ESTAR models are reported in Panel B of Table 1, and the parameter estimates are very close to those reported in TPS (see their Table 3).

[9] TPS also define the log-level of the real exchange rate as $q_t = s_t - p_t + p_t^*$. TPS normalize each real exchange rate series to zero in 1973:01. Note that TPS assume $\varepsilon_t \sim iid\ N(0,\sigma_\varepsilon^2)$ for their impulse response analysis (described below).

[10] As α increases in absolute value, the nonlinear effect becomes "stronger."

[11] Multivariate nonlinear least squares estimation is equivalent to maximum likelihood estimation under the assumption that the disturbance terms are Gaussian.

[12] Impulse response analysis is considerably more complex for nonlinear, in contrast to linear, AR models. For nonlinear models, the impulse response function is not invariant to the size of the shock, past shocks, and future shocks. TPS calculate impulse response functions using the Monte Carlo integration method in Gallant, Ross, & Tauchen (1993).

3. Constructing and evaluating point, interval, and density forecasts

3.1. Constructing point, interval, and density forecasts

Define an in-sample period that contains the first R observations for q_t and an out-of-sample period that spans P additional observations for q_t. In our applications, the in-sample period corresponds to the sample used to estimate the nonlinear AR model in the original studies. Using the *International Financial Statistics* database, we compile real exchange rate observations for an out-of-sample period. Each out-of-sample period begins in the month immediately following the end of the in-sample period and ends in the most recent period where data are available, either 2003:06 or 2003:07. For the OT Band-TAR models, this leaves the following out-of-sample periods: U.K., 1994:12-2003:07 (104 observations); Germany, 1995:01-2003:07 (103 observations); France and Japan, 1994:12-2003:06 (103 observations). For the TPS ESTAR models, this leaves the following out-of-sample periods: U.K. and Germany, 1997:01-2003:07 (79 observations); France and Japan, 1997:01-2003:06 (78 observations). Graphs of the real exchange rate for the United Kingdom, Germany, France, and Japan covering both the in-sample and out-of-sample periods are presented in Figure 1.

We use the fitted nonlinear AR models reported in Section 2 above to calculate out-of-sample point, interval, and density forecasts of q_{t+h} conditional on q_t for $t = R,\ldots,R+P-h$.[13] By using this procedure, we are simulating the situation of a forecaster who uses the fitted nonlinear AR models from the two published studies to forecast real exchange rate observations that have become available since the models were originally estimated. In this sense, we can see how the original nonlinear AR models "hold up" when forecasting out-of-sample observations.

We are interested in whether the out-of-sample point, interval, and density forecasts generated by the OT Band-TAR and TPS ESTAR models are superior to those generated by simple linear AR models. We compare forecasts from the OT Band-TAR and TPS ESTAR models to those generated by a linear

[13] In the taxonomy of West & McCracken (1998), we are using a "fixed" sampling scheme to generate out-of-sample forecasts.

AR(1) model, as this is the natural linear AR counterpart to the nonlinear AR models given in equations (1) and (2). In line with OT and TPS, we assume that the disturbance terms in the nonlinear AR models are Gaussian. For purposes of comparison, we assume that the disturbance terms in the linear AR models are also Gaussian.

It is straightforward to analytically generate point, interval, and density forecasts for a linear AR model under the assumption that the disturbance term is Gaussian.[14] However, analytical point, interval, and density forecasts are generally not available for nonlinear AR models when $h \geq 2$ even when the disturbance term is Gaussian, as $E[f(x)] \neq f[E(x)]$. We use the following simulation-based procedure to generate forecasts for the nonlinear AR models.[15] Consider, for example, the TPS ESTAR model, which we write compactly as $q_{t+1} = f(q_t) + \varepsilon_{t+1}$, and $h = 2$. For a given q_t, we simulate a realization for q_{t+1} as $q^*_{t+1} = f(q_t) + \sigma_\varepsilon \cdot e^*_{t+1}$, where e^*_{t+1} is a random draw from the standard normal distribution. We then simulate a realization for q_{t+2} as $q^*_{t+2} = f(q^*_{t+1}) + \sigma_\varepsilon \cdot e^*_{t+2}$. We repeat this process 1,000 times, giving us 1,000 simulated realizations for q^*_{t+2} given q_t. The point forecast of q_{t+2} given q_t is the mean of the 1,000 simulated realizations. In order to form, say, an inter-quartile forecast for q_{t+2} given q_t, we use the 250$^{\text{th}}$ and 750$^{\text{th}}$ simulated realizations from the sorted set of 1,000 simulated realizations. It is also straightforward to form an empirical density forecast for q_{t+2} given q_t using the set of simulated realizations. A similar procedure can be used to generate point, interval, and density forecasts for any q_{t+h} given q_t.

[14] We treat the parameters of the linear and nonlinear AR models as known in forming forecasts, and so we do not explicitly consider parameter estimation uncertainty in forming interval and density forecasts. This is common in the extant literature. See Hansen (2004) for methods of incorporating parameter estimation uncertainty into interval forecasts for linear models.

[15] Simulation-based procedures are more computationally intensive, but appear to work better, than other methods for generating forecasts for nonlinear AR models; see, for example, Clements & Smith (1997).

3.2. Evaluating point forecasts

Denote the point forecast error at horizon *h* corresponding to the nonlinear AR model as $e_{N,t+h|t}$ and that corresponding to the linear AR model as $e_{L,t+h|t}$ ($t = R,\ldots,R+P-h$). We focus on the popular MSFE criterion, $MSFE_i = (1/P_h)\sum_{t=R}^{R+P-h} e_{i,t+h|t}^2$ ($i = N, L$), where $P_h = P - (h-1)$, and use the popular Diebold & Mariano (1995) procedure to test the null hypothesis of equal predictive ability against the one-sided alternative hypothesis that the nonlinear AR model has a smaller MSFE than the linear AR model. Following Siliverstovs & van Dijk (2003), we use the modified Diebold & Mariano statistic (*M-DM*) of Harvey, Leyboure, & Newbold (1997) based on a correction factor designed to account for potential finite-sample size distortions, and we assess significance using the Student's *t* distribution with $P_h - 1$ degrees of freedom.[16] As discussed in the introduction, it may be more appropriate to focus on certain observations when evaluating the forecasting performance of a Band-TAR or ESTAR model. We use a weighted Diebold & Mariano (1995) statistic recently developed by van Dijk & Franses (2003), where different weights are attached to observations in different regions. Given that the Band-TAR and ESTAR models assume symmetric adjustment to the long-run equilibrium, we use the first weight function suggested by van Dijk & Franses (2003), $w_T(\omega_t) = 1 - \varphi(q_t)/\max[\varphi(q_t)]$, where $\varphi(q_t)$ is the density function of q_t. This weight function attaches greater weight to observations in both tails of the distribution of q_t.[17] We again follow Siliverstovs & van Dijk (2003) and adjust the weighted statistic using the Harvey, Newbold, & Leyboure (1997) correction factor to obtain the modified *W-DM* statistic, *MW-DM*, and we again assess significance using the Student's *t* distribution with $P_h - 1$ degrees of freedom.

[16] We use the Newey & West (1987) procedure with the Bartlett kernel and a truncation lag of $h-1$ in estimating the asymptotic variance used to compute the *M-DM* statistic.

[17] In our applications, we use a nonparametric kernel density procedure over the in-sample observations of q_t in order to estimate $\varphi(q_t)$.

3.3. Evaluating interval forecasts

In order to evaluate interval forecasts, we follow Wallis (2003), who builds on the likelihood ratio tests developed by Christoffersen (1998). According to Christoffersen (1998), good interval forecasts should have good coverage, and observations that fall inside or outside of the forecast intervals should be independently distributed over time, so that they do not “cluster.” Christoffersen (1998) develops likelihood ratio tests of unconditional coverage, independence, and conditional coverage. We use the Pearson χ^2 versions of these tests advocated by Wallis (2003), which we denote by χ^2_{UC}, χ^2_{IND}, and χ^2_{CC}, respectively. These statistics are based on indicator variables that take on a value of unity if the actual observation is contained in the interval forecast and zero otherwise, and they can be analyzed straightforwardly using contingency tables or matrices, where the observed number of outcomes is compared to the expected number under the appropriate null hypothesis. Instead of basing inferences on the asymptotic distributions of the χ^2_{UC}, χ^2_{IND}, and χ^2_{CC} statistics, we follow the recommendation of Wallis (2003) and calculate exact *p*-values based on the observed and expected outcomes using the theory described in Mehta & Patel (1998).[18] This allows for sharper inference, especially when the available number of out-of-sample forecasts is not very large. When $h \geq 2$, we have to modify the above procedure in order to account for the fact that optimal forecasts at horizon *h* are characterized by autocorrelation of order $h-1$, as the indicator variables used to construct the Pearson χ^2 statistics will also exhibit autocorrelation of order $h-1$ when the forecasts are optimal. We follow Siliverstovs & van Dijk (2003), who use the procedure based on Bonferroni bounds suggested by Diebold, Gunther, & Tay (1998). This procedure divides the indicator variable series into *h* sub-groups that are independent under the null hypothesis. We then apply the χ^2_{UC}, χ^2_{IND}, and χ^2_{CC} tests to each of the *h* sub-groups and reject the relevant null hypothesis for a given test at an overall significance level of α if we reject the null hypothesis for any of the sub-groups at the α / h significance level. Proceeding in this way can severely

[18] See Wallis (2003) for details on the construction of the χ^2 statistics and computation of exact *p*-values.

restrict the number of indicator variables in each sub-group as h increases, so that practical limits are placed on the maximum h we can consider. The declining number of indicator variables available in each sub-group as h increases also helps to motivate our use of exact p-values for inference.

3.4. Evaluating Density Forecasts

Diebold, Gunther, & Tay (1998) provide a framework for evaluating density forecasts based on the probability integral transform (PIT). Under the null hypothesis that the density forecast generated by a given forecasting model is correct, Diebold, Gunther, & Tay (1998) show that the PIT series is distributed $iid\ U(0,1)$. Following Clements & Smith (2000) and Siliverstovs & van Dijk (2003), we test for uniformity using the Kolmogorov-Smirnov (KS) statistic.[19] Berkowitz (2001) suggests transforming the PIT series using the inverse of the standard normal cumulative density function, and under the null hypothesis that the density forecast is correct, the transformed PIT series is distributed $iid\ N(0,1)$. We again follow Clements & Smith (2000) and Siliverstovs & van Dijk (2003) and test for standard normality using the Doornik & Hansen (1994, DH) statistic.[20] Note that the KS and DH statistics assume independence. In order to explicitly test for independence in the PITs, Diebold, Gunther, & Tay (1998) recommend looking for autocorrelation in the power-transformed PIT series. Following Siliverstovs & van Dijk (2003), we use the Ljung-Box statistic to test for first-order autocorrelation in the power-transformed PIT series. Finally, when $h \geq 2$, we proceed analogously as described above in Section 3.3 and divide the PITs into h sub-groups that are independent under the null hypothesis. We then apply the KS, DH, and Ljung-Box tests to each of the h sub-groups and reject the null hypothesis at an overall significance level of α if we reject the null hypothesis for any sub-group at the α / h significance level.

[19] We generate critical values for the KS statistic using the formulae given in Miller (1956).

[20] Berkowitz (2001) recommends using the transformed PIT series in order to have tests with greater power. The DH statistic follows the $\chi^2(2)$ asymptotic distribution, and we base inference on critical values corresponding to this distribution.

4. Empirical results

4.1. Point forecasts

Table 2 presents out-of-sample point forecast evaluation results for the OT Band-TAR models and linear AR(1) counterparts for the U.K., Germany, France, and Japan.[21] Columns 2 and 6 report the root MSFE (RMSFE) for the linear AR model, and columns 3 and 7 report the ratio of the Band-TAR model RMSFE to the linear AR model RMSFE (relative RMSFE). The relative RMSFE is close to unity at short horizons for each country, indicating that the point forecasting performance of the linear AR and Band-TAR models is very similar at short horizons. At longer horizons, the Band-TAR model RMSFE is greater than the linear AR model RMSFE for the U.K., Germany, and France, and the gap between the RMSFE for the two models is usually about 3%-8%. Not surprisingly, we cannot reject the null hypothesis that the Band-TAR model MSFE is not less than the linear AR model MSFE using the *M-DM* test and *p*-values based on the Student's *t* distribution at any horizon for the U.K., Germany, and France (see columns 4 and 8). For Japan, the relative RMSFE declines as the horizon increases, and the Band-TAR model RMSFE is actually 29% less than the linear AR model RMSFE at a horizon of 24 months. According to the *M-DM* statistics and *p*-values based on the Student's *t* distribution, the Band-TAR model MSFE is significantly less than the linear AR model MSFE for Japan at horizons of 6-24 months (see column 8). At this point, there is no support for the Band-TAR model specification over the linear AR model specification for the U.K., Germany, and France according to an MSFE metric. There is significant support for the Band-TAR model specification for Japan at longer forecast horizons according to an MSFE metric.[22]

There is reason to be cautious about basing inferences on the Student's *t* distribution for the *M*-

[21] Recall that q_t is demeaned before estimating the Band-TAR model. In order to form forecasts of q_t, we add the estimated mean over the in-sample period to the forecasts generated by the Band-TAR model. The coefficient estimates for all of the linear AR models are available upon request from the authors.

[22] We also used the Harvey, Leybourne, and Newbold (1998) procedure to test whether the linear AR model forecasts encompass the OT Band-TAR model forecasts. The results are very similar to those reported in Table 2 for the *M-DM* statistic: we can only reject the null hypothesis that the linear AR model forecasts encompass the OT Band-TAR model forecasts at longer horizons for Japan. The complete results are available upon request from the authors.

DM statistic in Table 2. When $h = 1$, McCracken (2004) shows that the Diebold & Mariano (1995) statistic has a non-standard limiting distribution when comparing forecasts from two nested linear models. When $h \geq 2$ and comparing forecasts from nested linear models, Clark and McCracken (2004) show that the Diebold & Mariano (1995) statistic has a non-standard limiting distribution that is not free of nuisance parameters, so that critical values cannot be tabulated, and they recommend using a bootstrap procedure to calculate critical values. The OT Band-TAR and linear AR models are nested, since as c approaches zero in the OT Band-TAR model specification, equation (1), this model approaches a linear AR(1) model. Given the potential relevance of comparing nested models, we augment our analysis with a parametric bootstrap procedure in order to generate critical values for the *M-DM* statistics. Assuming that the data are generated by a linear AR(1) process under the null hypothesis, we simulate a large number of pseudo-samples and compute the *M-DM* statistic for each pseudo-sample. The bootstrapped *p*-value is the proportion of *M-DM* statistics corresponding to the pseudo-samples greater than the *M-DM* statistic corresponding to the original sample. Bootstrapped *p*-values for the *M-DM* statistics are reported in curly brackets in columns 4 and 8 of Table 2. When we base inference on the parametric bootstrapped critical values, the *M-DM* statistics remain insignificant at every horizon for the U.K., Germany, and France, and they are only significant (at the 10% level) at horizons of 6 and 9 months for Japan.

Columns 5 and 9 of Table 2 report *MW-DM* statistics, which place greater weight on forecasting real exchange rate values farther out in the tails of the unconditional distribution. Columns 5 and 9 also report *p*-values for the *MW-DM* statistics based on the Student's *t* distribution in brackets and bootstrapped *p*-values in curly brackets. The results for Germany and France indicate that the *MW-DM* statistics are insignificant at all horizons according to either the *p*-values based on the Student's *t* distribution or the bootstrapped *p*-values. For the U.K., the *MW-DM* statistic is significant according to the *p*-value based on the Student's *t* distribution at horizons of 1, 3, and 6 months. This suggests that when we focus on forecasting more extreme real exchange rate observations, the Band-TAR model has superior forecasting performance over the linear AR model at shorter horizons. However, none of the

MW-DM statistics are significant at any horizon for the U.K. according to the bootstrapped *p*-values. Turning to Japan, the *MW-DM* statistic is significant at horizons of 9-24 months according to the *p*-values based on the Student's *t* distribution, so that the Band-TAR appears to produce superior forecasts in terms of a weighted MSFE criterion at horizons of 9-24 months. The *MW-DM* statistic remains significant at horizons of 9, 12, 15, 21, and 24 months according to the bootstrapped *p*-values. The only reliable evidence in Table 2 that the OT Band-TAR model produces significantly better point forecasts than a linear AR model is for Japan at long horizons using a weighted MSFE criterion.

Table 3 presents the out-of-sample point forecast evaluation results for the TPS ESTAR and linear AR(1) models. For the U.K., the ESTAR model RMSFE is greater than the linear AR model RMSFE at all reported horizons, and neither the *M-DM* nor *MW-DM* statistic is significant at any horizon (regardless of what *p*-values are used). For Germany, the ESTAR model RMSFE is less than the linear AR model RMSFE at all horizons, and the ESTAR model RMSFE is 18% lower than the linear AR model RMSFE at the 24-month horizon. Furthermore, the *M-DM* and *MW-DM* statistics are significant at horizons of 18-24 months when we use the *p*-values based on the Student's *t* distribution, and they remain significant at the 24-month horizon (at the 10% level) when we use the bootstrapped *p*-values. Turning to France, the ESTAR model RMSFE is less than the linear AR model RMSFE at all reported horizons, and the reduction in RMSFE reaches 17% at the 24-month horizon. The *M-DM* (*MW-DM*) statistic is significant at horizons of 1 (1 and 2) and 12-24 (6-24) months for France. When we use the bootstrapped critical values, The *M-DM* (*MW-DM*) statistic is significant at horizons of 2 and 18-24 (18-24) months according to the bootstrapped *p*-values. For Japan, the ESTAR model RMSFE is always greater than the linear AR model RMSFE, and none of the *M-DM* statistics is significant at any horizon according to either *p*-value. The *MW-DM* statistic is significant horizons of 21 and 24 months according to *p*-values based on the Student's *t* distribution and is significant at the 24-month horizon according to the bootstrapped *p*-values.[23]

Overall, there is robust evidence in Tables 2 and 3 that the OT Band-TAR and TPS ESTAR models offer forecasting gains at long horizons relative to simple linear AR models for some countries, especially when we use a weighted MSFE criterion. There is almost no robust evidence that nonlinear models offer forecasting gains at short horizons for any country.[24]

4.2. Interval forecasts

Pearson χ^2 statistics used to evaluate interval forecasts for the OT Band-TAR and linear AR(1) models are provided in Table 4 for $h = 1,2,3$. Following Wallis (2003), we focus on inter-quartile interval forecasts (corresponding to the 0.25 and 0.75 quantiles). For the both the linear AR and Band-TAR models for the U.K., correct unconditional coverage and correct conditional coverage are rejected at all three reported horizons (see columns 4 and 6), while independence is rejected at horizons of 2 and 3 months (1 and 2 months) for the linear AR (Band-TAR) model. For Germany, correct unconditional coverage is rejected at the 1-month horizon and correct conditional coverage is rejected at all three horizons for the linear AR model, while the only rejection for the Band-TAR model is correct conditional coverage at the 3-month horizon. With respect to France, correct unconditional and conditional coverage are rejected at the 1-month horizon for both the linear AR and Band-TAR models. The only other rejection for the linear AR model is conditional coverage at the 2-month horizon, while independence and correct conditional coverage are both rejected at the 3-month horizon for the Band-TAR model. Turning to Japan, there are no rejections for the linear AR model and one rejection (for correct unconditional coverage at the 1-month horizon) for the Band-TAR model.

---

[23] We also tested whether the linear AR model forecasts encompass the TPS ESTAR model forecasts. We can only reject forecast encompassing at some horizons (especially long horizons) for Germany and France. The complete results are available upon request from the authors.

[24] In order to compare point forecasts in terms of the predicted direction of change of the real exchange rate, we employed the Henrikkson & Merton (1981), Cumby & Modest (1987), and Pesaran & Timmermann (1992) tests of market timing ability. There is no evidence that out-of-sample forecasts generated by the OT Band-TAR and TPS ESTAR models are superior to forecasts generated by linear AR(1) models in terms of market timing ability for any country. The complete results are available upon request from the authors.

We evaluate inter-quartile interval forecasts for the TPS ESTAR and linear AR(1) models in Table 5. Both the ESTAR and linear AR models are deficient in terms of coverage for the U.K., as correct unconditional and conditional coverage are rejected at all reported horizons for both models. Independence is also rejected for both models at the 1-month horizon. For Germany, correct conditional coverage is rejected for both the ESTAR and linear AR models at the 2-month horizon, while independence is also rejected at the 2-month horizon for the linear AR model. With respect to France, independence and correct conditional coverage are rejected for the linear AR model at the 2-month horizon, while none of the statistics are significant at any reported horizon for the ESTAR model. Both the linear AR and ESTAR models seem to be adequate in terms of coverage for Japan, as no statistic for either model is significant at any reported horizon.

Summarizing the results in Tables 4 and 5 concerning interval forecast evaluation, we fail to find strong evidence to support the nonlinear model specifications over their linear counterparts. There is not much evidence in Table 4 to recommend the OT Band-TAR model over a linear AR counterpart for the U.K., France, and Japan, while there is some evidence that the OT Band-TAR model provides better coverage than the linear AR model for Germany. There is not much evidence that differentiates the TPS ESTAR from the linear AR specification for the U.K., Germany, and Japan in Table 5, and there is some evidence supporting the TPS ESTAR specification over a linear AR specification for France.[25]

4.3. Density forecasts

Density forecast evaluation results for the OT Band-TAR and linear AR(1) models are reported in Table 6. The aspect of the table that stands out is the rejection of independent PITs for $k = 2$ (see column 7) at all reported horizons for both the OT Band-TAR and linear AR models for all four countries. These rejections point to deficiencies in the density forecasts for both the OT Band-TAR and linear AR model

[25] As described in the text, the interval forecasts are inter-quartile interval forecasts computed using the 250th and 750th realizations of the sorted 1,000 simulated realizations. We also computed highest-density interval forecasts for the nonlinear models; see Hyndman (1995). The results are similar to those reported in Tables 4 and 5. The complete results are available upon request from the authors.

specifications. There are also frequent rejections for both models and all countries when $k = 4$ (see column 9). Looking at the KS and DH statistics in columns 4 and 5 of Table 6, there is some support for the Band-TAR model over the linear AR model for the U.K., as the KS statistic is significant at all reported horizons for the linear AR model, but it is only significant at the 1-month horizon for the Band-TAR model. For Germany and France, the KS statistic is not significant for either model at any reported horizon, while the DH statistic is significant at the 2-month horizon for the linear AR model for Germany. For Japan, the DH statistic is significant at all reported horizons for both the linear AR and Band-TAR models. Overall, the results in Table 6 provide little support for the Band-TAR models over their linear AR counterparts.

Table 7 reports density forecast evaluation results for the TPS ESTAR and linear AR(1) models. As in Table 6, we reject the null hypothesis of independence using the LB statistic when $k = 2$ for both the ESTAR and linear AR models for every country at all reported horizons (see column 7). The null of independence is also rejected when $k = 4$ at one or more reported horizons for both the ESTAR and linear AR models for all countries (see column 9). Again, this indicates deficiencies in both the ESTAR and linear AR model specifications. The KS statistic rejects uniformity at the 1-month horizon for both the ESTAR and linear AR models for the U.K., and the DH statistic is significant at horizons of 1-3 (1) months for both the ESTAR and linear AR models for Japan (Germany). For France, the KS statistic is significant at horizons of 1 and 2 months for the linear AR model but not the ESTAR model. Overall, the results in Table 7 point to problems with both the ESTAR and linear AR model density forecasts and fail to provide much support for the ESTAR models over the linear AR models.

4.4. Robustness Checks

We also checked whether the results reported in Tables 2-7 are significantly changed if we estimate the parameters of the nonlinear and linear AR models using a recursive (or expanding) rather than a fixed window when forming the out-of-sample forecasts. The results are similar, as the parameter estimates for all of the models change relatively little as we use observations beyond the in-sample period corresponding to the fixed estimation window. Furthermore, we checked whether the results in Tables 2-7 are considerably affected when we relax the assumption that the disturbance terms in the linear and nonlinear AR models are normally distributed and instead bootstrap the in-sample residuals to generate forecasts for the linear and nonlinear AR models. The results are similar to those reported in Tables 2-7.[26]

5. Comparing in-sample conditional densities

Looking back to Tables 2 and 3, we see that point forecasts generated by the fitted Band-TAR and ESTAR models are very similar in terms of MSFE to point forecasts generated by their linear AR counterparts at short horizons. Diebold & Nason (1990) offer a number of reasons why nonlinear models may fail to offer sizable forecasting gains relative to linear models. One of their reasons is that "very slight conditional-mean nonlinearities might be truly present and be detectable with large datasets, while nevertheless yielding negligible ex ante forecast improvement" (p. 318).[27] In order to examine the relevance of this for the OT Band-TAR and TPS ESTAR models, we follow the suggestion of Pagan (2002) and Breunig, Najarian, & Pagan (2003) and graphically compare the conditional expectation functions for q_t given q_{t-1} corresponding to the fitted OT Band-TAR and TPS ESTAR models and their linear AR(1) counterparts. This gives us a visual feel for how "close" the fitted linear and nonlinear AR models are in terms of their conditional means.

Figure 2 presents the conditional expectation functions for q_t given q_{t-1} corresponding to the fitted OT Band-TAR models and their linear AR counterparts, as well as a scatterplot of the in-sample data. From Figure 2, we see that the conditional expectation functions corresponding to the fitted Band-TAR and linear AR models are very near each other, so that any nonlinearities in the conditional means

[26] We do not report the complete results for the robustness checks to conserve space. They are available upon request from the authors.
[27] In a footnote (p. 318), they add, "In other words, *significance* of nonlinearity does not necessarily imply its *economic importance*" [emphasis in original].

appear "very slight." In Figure 3, we graph the conditional expectation functions corresponding to the fitted TPS ESTAR models and linear AR counterparts. Again, the conditional expectation functions appear very close to one another. Based on comparisons of the conditional expectation functions in Figures 2 and 3, the lack of sizable forecasting gains at short horizons provided by fitted OT Band-TAR and TPS ESTAR models relative to linear AR counterparts appears to result from the absence of "strong" nonlinearities in the conditional means of these nonlinear AR models.

In order to compare the fitted nonlinear and linear AR models more formally and extensively, we use the recently developed test of Corradi & Swanson (2003). Their test is based on a distributional analog of the mean squared error metric, and it allows us to compare the conditional densities for q_t given q_{t-1} corresponding to two different fitted models, each of which is possibly misspecified. More specifically, we use the Corradi & Swanson (2003) Z_T statistic to test the null hypothesis that the conditional densities corresponding to the fitted linear and nonlinear AR models are equally accurate relative to the true conditional density against the alternative hypothesis that the conditional density corresponding to the nonlinear AR model is more accurate than the conditional density corresponding to the linear AR benchmark model. In order to make the test operational, we compute the Z_T statistic by integrating over a fine grid covering the minimum and maximum values of the in-sample q_t observations. We also define a test statistic R-Z_T that integrates over two grids of values whose limits correspond to the minimum and maximum values of the first and fourth quartiles of the in-sample q_t observations. This allows us to focus our comparison of the conditional distributions corresponding to the fitted linear and nonlinear AR models on the tails of the distributions of in-sample q_t observations.

Corradi & Swanson (2003) test results for the fitted OT Band-TAR and TPS ESTAR models and linear AR counterparts are reported in Table 8. Following the recommendation of Corradi & Swanson (2003), we base inferences on block bootstrapped critical values.[28] According to the Z_T statistics in column 3, the null hypothesis of equal conditional density accuracy cannot be rejected for any of the

[28] Details of the block bootstrap procedure are provided in Corradi & Swanson (2003).

nonlinear AR models relative to the AR benchmark models. This indicates that the conditional densities for q_t given q_{t-1} corresponding to the Band-TAR and TPS ESTAR models are not significantly different from the conditional densities corresponding to linear AR benchmark models in terms of their accuracies. When we focus on the first and fourth quartiles of the in-sample q_t observations, the $R\text{-}Z_T$ statistic is significant only for the OT Band-TAR model for France and TPS ESTAR model for Germany. Overall, the results in Table 8 indicate that, in most cases, fitted nonlinear AR models from the extant literature are quite "close" to fitted linear AR models. This helps to explain the typical inability of point and density forecasts generated by the OT Band-TAR and TPS ESTAR models to improve upon forecasts generated by linear AR models at short horizons in Section 4 above.

6. Conclusion

In this paper, we undertake an extensive evaluation of the out-of-sample forecasting performance of a number of nonlinear models of real exchange rate behavior from the extant empirical literature. Using monthly data from the post-Bretton Woods period, we show that point forecasts generated by the OT Band-TAR and TPS ESTAR models are very similar to forecasts generated by linear AR models at short horizons. For some countries, we find that the nonlinear models offer forecasting gains at long horizons of around two years, especially using a weighted MSFE criterion. Analysis of interval forecasts generated by the OT Band-TAR and TPS ESTAR models at horizons of 1-3 months provides little evidence to support the Band-TAR and ESTAR specifications over linear AR counterparts. Evaluation of density forecasts at horizons of 1-3 months points to deficiencies in the OT Band-TAR, TPS ESTAR, and linear AR models and does not lead one to favor either the nonlinear or linear AR model specifications. Our analysis of in-sample one-month-ahead conditional expectation functions and conditional densities reveals that the fitted OT Band-TAR and ESTAR models are typically not very different from fitted linear AR models, helping to explain why the nonlinear models typically fail to offer sizable forecasting gains at short horizons. Overall, our analysis indicates that any nonlinearities in monthly real exchange rate data from the post-Bretton Woods period are “slight” or “subtle” for the OT Band-TAR and TPS ESTAR model specifications.

Finally, we suggest some avenues for future research. Sarno, Taylor, & Chowdhury (2004) and Imbs, *et al*. (2003) recently report in-sample evidence of nonlinear behavior in real exchange rates based on more disaggregated price indices.[29] It would be interesting to apply the evaluation techniques used in the present paper to these real exchange rates, as nonlinearities may be more evident for real exchange rates based on disaggregated price indices (Granger, 2001). Given that the density forecasts point to deficiencies in the OT Band-TAR, TPS ESTAR, and linear AR model specifications over the post-Bretton Woods period, it would also be useful to explore model specification issues more extensively by considering, for example, different types of nonlinear behavior and conditional heteroskedasticity in the disturbances. Along these lines, mutlivariate nonlinear model specifications may also prove helpful, as the theoretical models in Goswami, Shrikhande, & Wu (2002) suggest that the dynamics of the real exchange rate depend critically on a state variable such as the capital stock or trade balance.

[29] Also see O’Connell & Wei (2002), who find that price discrepancies for individual goods between U.S. cities are nonlinearly mean-reverting to parity.

References

Balke, N.S. & Fomby, T.B. (1997). Threshold cointegration, *International Economic Review*, 38, 627-645.

Benninga, S. & Protopapadakis, A.A. (1988). The equilibrium pricing of exchange rates and assets when trade takes time, *Journal of International Economics*, 7, 129-149.

Berkowitz, J. (2001). Testing density forecasts, with applications to risk management, *Journal of Business and Economic Statistics*, 19, 465-474.

Bertola, G. & Caballero, R.J. (1990). Kinked adjustment costs and aggregate dynamics. In: O.J. Blanchard and S. Fischer (Eds.), *NBER Macroeconomics Annual*. Cambridge, Mass.: MIT Press, pp. 237-287.

Breunig, R., Najarian, S., & Pagan, A. (2003). Specification testing of Markov switching models, *Oxford Bulletin of Economics and Statistics*, 65, 703-725.

Christoffersen, P. (1998). Evaluating interval forecasts, *International Economic Review*, 39, 841-862.

Clark, T.E. & McCracken, M.W. (2004). Evaluating long-horizon forecasts, manuscript, University of Missouri at Columbia.

Clements, M.P. & Smith, J. (1997). The performance of alternative forecasting methods for SETAR models, *International Journal of Forecasting*, 13, 463-475.

Clements, M.P. & Smith, J. (1999). A Monte Carlo study of the forecasting performance of empirical SETAR Models, *Journal of Applied Econometrics*, 14, 123-141.

Clements, M.P. and J. Smith, 2000. Evaluating the forecast densities of linear and non-linear models: Applications to output growth and unemployment, *Journal of Forecasting*, 19, 255-276.

Coleman, A.M.G. (1995). Arbitrage, storage and the law of one price: New theory for the time series analysis of an old problem, manuscript, Princeton University.

Corradi, V. & Swanson, N.R. (2003). A test for comparing multiple misspecified conditional distributions, manuscript, Rutgers University.

Cumby, R.E. & Modest, D.M. (1987). Testing for market timing ability, *Journal of Financial Economics*, 19, 169-189.

De Gooijer, J.G. & Kumar, K. (1992). Some recent developments in non-linear time series modelling, testing, and forecasting, *International Journal of Forecasting*, 8, 135-156.

Diebold, F.X., Gunther, T., & Tay, A. (1998). Evaluating density forecasts with applications to financial risk management, *International Economic Review*, 39, 863-883.

Diebold, F.X. & Mariano, R.S. (1995). Comparing predictive accuracy, *Journal of Business and Economics Statistics*, 13, 253-263.

Diebold, F.X. & Nason, J.A. (1990). Nonparametric exchange rate prediction? *Journal of International Economics*, 28, 315-332.

Doornik, J.A. & Hansen, H. (1994). An omnibus test for univariate and multivariate normality, manuscript, Nuffield College, Oxford.

Dumas, B. (1992). Dynamic equilibrium and the real exchange rate in spatially separated worlds, *Review of Financial Studies*, 5, 153-180.

Dumas, B. (1994). Partial equilibrium versus general equilibrium models of the international capital market. In: F. Van Der Ploeg (Ed.), *The Handbook of International Economics*. Oxford: Blackwell, Chapter 10.

Eitrheim, O. & Teräsvirta, T. (1996). Testing the adequacy of smooth transition autoregressive models, *Journal of Econometrics*, 74, 59-75.

Gallant, A.R., Ross, P.E., & Tauchen, G. (1993). Nonlinear dynamic structures, *Econometrica*, 61, 871-908.

Goswami, G., Shrikhande, M, & Wu, L. (2002). A dynamic equilibrium model of real exchange rates with general transaction costs, manuscript, Fordham University.

Granger, C.W.J. (2001). An overview of nonlinear macroeconometric empirical models, *Macroeconomic Dynamics*, 5, 466-481.

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

Hansen, B.E. (2004). Interval forecasts and parameter uncertainty, *Journal of Econometrics*, forthcoming.

Harvey, D., Leybourne, S., & Newbold, P. (1997). Testing the equality of prediction mean squared errors, *International Journal of Forecasting*, 13, 281-291.

Harvey, D., Leybourne, S., & Newbold, P. (1998). Tests for forecast encompassing, *Journal of Business and Economic Statistics*, 16, 254-259.

Henriksson, R.D. & Merton, R.C. (1981). On market timing and investment performance. II. Statistical procedures for evaluating forecasting skills, *Journal of Business*, 54, 513-533.

Hyndman, R.J. (1995). Highest-density forecast regions for non-linear and non-normal time series models, *Journal of Forecasting*, 14, 431-432.

Imbs, J., Mumtaz, H., Ravn, M.O., & Rey, H. (2003). Non-linearities and real exchange rate dynamics, manuscript, London Business School.

Lothian, J.R. & Taylor, M.P. (1996). Real exchange rate behavior: The recent float from the perspective of the last two centuries, *Journal of Political Economy*, 104, 488-510.

McCracken, M.W. (2004). Asymptotics for out-of-sample tests of Granger causality, manuscript, University of Missouri at Columbia.

Mehta, C.R. & Patel, N.R. (1998). Exact inference for categorical data. In: P. Armitage and T. Colton (Eds.), *Encyclopedia of Biostatistics*. Chichester: John Wiley, pp. 1411-1422.

Michael, P., Nobay, A.R., & Peel, D.A. (1997). Transactions costs and nonlinear adjustment in real exchange rates: An empirical investigation, *Journal of Political Economy*, 105, 862-879.

Miller, L.H. (1956). Table of percentage points of Kolmogorov statistics, *Journal of the American Statistical Association*, 51, 111-121.

Montgomery, A.L., Zarnowitz, V., Tsay, R.S., & Tiao, G.C. (1998). Forecasting the U.S. unemployment rate, *Journal of the American Statistical Association*, 93, 478-493.

Newey, W. & West, K.D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica*, 55, 703-708.

Obstfeld, M. & Rogoff, K. (2000). The six major puzzles in international macroeconomics: Is there a common cause? In: B. Bernanke & K. Rogoff (Eds.), *NBER Macroeconomics Annual 2000*. Cambridge, Mass.: MIT Press, pp. 339-390.

Obstfeld, M. & Taylor, A.M. (1997). Nonlinear aspects of goods-market arbitrage and adjustment: Heckscher’s commodity points revisited, *Journal of the Japanese and International Economies*, 11, 441-479.

O’Connell, P.G.J. (1998). Market frictions and real exchange rates, *Journal of International Money and Finance*, 17, 71-95.

O’Connell, P.G.J. & Wei, S.-J. (2002). The bigger they are, the harder they fall: Retail price differences across U.S. cities, *Journal of International Economics*, 56, 21-53.

Ohanian, L.E. & Stockman, A.C. (1997). Arbitrage costs and exchange rates, manuscript, University of Rochester.

Pagan, A. (2002). Learning about models and their fit to data, *International Economic Journal*, 16, 1-18.

Pesaran, M.H. & Potter, S.M. (1997). A floor and ceiling model of U.S. output, *Journal of Economic Dynamics and Control*, 21, 661-695.

Pesaran, M.H. & Timmermann, A.G. (1992). A simple nonparametric test of predictive performance, *Journal of Business and Economic Statistics*, 10, 461-465.

Ramsey, J.B. (1996). If nonlinear models cannot forecast, what use are they? *Studies in Nonlinear Dynamics and Econometrics*, 1, 65-86.

Rogoff, K. (1996). The purchasing power parity puzzle, *Journal of Economic Literature*, 34, 647-668.

Sarantis, N. (1999). Modeling non-linearities in real effective exchange rates, *Journal of International*

Money and Finance, 18, 27-45.

Sarno, L., Taylor, M.P., & Chowdhury, I. (2004). Nonlinear dynamics in deviations from the law of one price, *Journal of International Money and Finance*, 23, 1-25.

Sercu, P., Uppal, R., & Van Hulle, C. (1995). The exchange rate in the presence of transactions costs: Implications for tests of purchasing power parity, *Journal of Finance*, 50, 1309-1319.

Siliverstovs, B. & van Dijk, D. (2003). Forecasting industrial production with linear, nonlinear, and structural change models, Econometric Institute Report EI 2003-16.

Stock, J.H. & Watson, M.W. (1999). A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series. In: R.F. Engle and H. White (Eds.), *Cointegration, Causality, and Forecsting: A Festschrift in Honour of Clive W.J. Granger*. Oxford: Oxford University Press, pp. 1-44.

Taylor, M.P., Peel, D.A., & Sarno, L. (2001). Nonlinear mean reversion in real exchange rates: Toward a solution to the purchasing power parity puzzles, *International Economic Review*, 42, 1015-1042.

Teräsvirta, T. (1994). Specification, estimation and evaluation of smooth transition autoregressive models, *Journal of the American Statistical Association*, 89, 208-218.

Tsay, R.S. (1989). Testing and modelling threshold autoregressive processes, *Journal of the American Statistical Association*, 84, 231-240.

van Dijk, D. & Franses, P.H. (2003). Selecting a nonlinear time series model using weighted tests of equal forecast accuracy, *Oxford Bulletin of Economics and Statistics*, 65, 727-744.

Wallis, K. (2003). Chi-squared tests of interval and density forecasts, and the Bank of England's fan charts, *International Journal of Forecasting*, 19, 165-175.

West, K.D. & McCracken, M.W. (1998). Regression-based tests of predictive ability, *International Economic Review*, 39, 817-840.

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

Table 1: In-sample parameter estimates for the Obstfeld & Taylor (1997) Band-TAR and Taylor, Peel, & Sarno (2001) ESTAR models

(1)	(2)	(3)	(4)	(5)
	United Kingdom	Germany	France	Japan
A. Obstfeld & Taylor (1997) Band-TAR model				
Sample	1980:02-1994:11	1980:02-1994:12	1980:02-1994:11	1980:02-1994:11
λ_{out}	-0.084 (0.058)	-0.052 (0.031)	-0.048 (0.029)	-0.057 (0.048)
c	0.163 (0.066)	0.111 (0.067)	0.107 (0.067)	0.254 (0.078)
σ_{out}	0.044	0.038	0.038	0.030
σ_{in}	0.033	0.031	0.031	0.036
n_{out}	54	98	92	63
n_{in}	124	81	86	115
B. Taylor, Peel, & Sarno (2001) ESTAR model				
Sample	1973:02-1996:12	1973:02-1996:12	1973:02-1996:12	1973:02-1996:12
α	-0.449 (0.164)	-0.264 (0.104)	-0.289 (0.109)	-0.165 (0.059)
η	0.150 (0.037)	-0.007 (0.041)	0.049 (0.038)	0.515 (0.045)
σ_{ε}	0.033	0.035	0.033	0.033

Notes: standard errors are reported in parentheses; the Obstfeld & Taylor (1997) Band-TAR model is given in equation (1) in the text; the Taylor, Peel, & Sarno (2001) ESTAR model is given in equation (2) in the text.

Table 2: Out-of-sample point forecast evaluation, linear AR and Obstfeld & Taylor (1997) Band-TAR models

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
h[a]	AR[b]	Band-TAR/AR[c]	*M-DM*[d]	*MW-DM*[e]	AR[b]	Band-TAR/AR[c]	*M-DM*[d]	*MW-DM*[e]
	A. United Kingdom, 1994:12-2003:07 out-of-sample period				C. France, 1994:12-2003:06 out-of-sample period			
1	0.02	1.00	-0.36 [0.64] {0.45}	1.61 [**0.05**] {0.24}	0.03	1.00	-0.23 [0.59] {0.43}	0.67 [0.25] {0.35}
2	0.03	1.01	-0.33 [0.63] {0.44}	1.28 [0.10] {0.21}	0.04	1.01	-0.31 [0.62] {0.47}	0.39 [0.35] {0.37}
3	0.03	1.00	0.08 [0.47] {0.33}	1.54 [**0.06**] {0.17}	0.05	1.01	-0.68 [0.75] {0.56}	0.04 [0.48] {0.45}
6	0.04	1.01	-0.14 [0.56] {0.43}	1.64 [**0.05**] {0.16}	0.07	1.02	-0.59 [0.72] {0.52}	0.24 [0.40] {0.42}
9	0.05	1.03	-0.41 [0.66] {0.49}	1.10 [0.14] {0.26}	0.08	1.02	-0.77 [0.78] {0.56}	-0.20 [0.58] {0.48}
12	0.06	1.04	-0.46 [0.68] {0.51}	0.65 [0.26] {0.36}	0.10	1.02	-0.72 [0.76] {0.54}	-0.26 [0.60] {0.49}
15	0.07	1.05	-0.47 [0.68] {0.52}	0.36 [0.36] {0.40}	0.12	1.04	-1.09 [0.86] {0.61}	-1.10 [0.86] {0.59}
18	0.07	1.03	-0.24 [0.60] {0.48}	0.43 [0.33] {0.40}	0.13	1.06	-1.84 [0.97] {0.73}	-1.86 [0.97] {0.74}
21	0.08	1.02	-0.14 [0.56] {0.45}	0.48 [0.32] {0.40}	0.14	1.06	-2.39 [0.99] {0.79}	-2.15 [0.98] {0.76}
24	0.08	1.03	-0.21 [0.58] {0.46}	0.37 [0.36] {0.40)	0.14	1.08	-3.29 [1.00] {0.84}	-2.72 [1.00] {0.82}
	B. Germany, 1995:01-2003:07 out-of-sample period				D. Japan, 1994:12-2003:06 out-of-sample period			
1	0.03	1.00	-0.31 [0.62] {0.42}	1.03 [0.15] {0.25}	0.04	0.99	0.76 [0.22] {0.16}	0.28 [0.39] {0.30}
2	0.04	1.00	-0.39 [0.65] {0.44}	0.48 [0.32] {0.31}	0.05	0.97	0.69 [0.25] {0.16}	0.30 [0.38] {0.29}
3	0.05	1.01	-0.50 [0.69] {0.51}	0.51 [0.31] {0.33}	0.07	0.96	0.82 [0.21] {0.16}	0.39 [0.35] {0.31}
6	0.07	1.03	-1.12 [0.87] {0.64}	-0.28 [0.61] {0.48}	0.10	0.91	1.33 [**0.09**] {0.12}	0.97 [0.17] {0.24}
9	0.09	1.03	-1.20 [0.88] {0.65}	-0.36 [0.64] {0.48}	0.12	0.84	2.16 [**0.02**] {**0.08**}	2.66 [**0.00**] {**0.05**}
12	0.11	1.02	-0.76 [0.78] {0.62}	-0.06 [0.52] {0.41}	0.14	0.81	2.43 [**0.01**] {**0.08**}	3.03 [**0.00**] {**0.04**}
15	0.13	1.03	-1.09 [0.86] {0.59}	-0.66 [0.74] {0.49}	0.16	0.80	2.43 [**0.01**] {0.11}	2.54 [**0.01**] {**0.08**}
18	0.15	1.04	-1.51 [0.93] {0.67}	-1.15 [0.87] {0.55}	0.18	0.76	2.59 [**0.01**] {0.12}	2.56 [**0.01**] {0.10}
21	0.16	1.05	-1.98 [0.97] {0.72}	-1.52 [0.93] {0.60}	0.20	0.74	2.59 [**0.01**] {0.15}	2.85 [**0.00**] {**0.09**}
24	0.17	1.07	-2.45 [0.99] {0.74}	-1.83 [0.96] {0.63}	0.22	0.71	2.96 [**0.00**] {0.14}	3.19 [**0.00**] {**0.09**}

Notes: *p*-value using the Student's *t* distribution with $P_h - 1$ degrees of freedom is reported in brackets; bootstrapped *p*-value is reported in curly brackets; bold *p*-value indicates significance at the 10% level according to the *p*-value; 0.00 indicates <0.005; 1.00 indicates ≥ 0.995.

[a]Forecast horizon (in months).

[b]Linear AR model RMSFE.

[c]Ratio of the Band-TAR model RMSFE to the linear AR model RMSFE.

[d]Modified Diebold & Mariano (1995) test statistic for the null hypothesis that the linear AR model MSFE equals the Band-TAR model MSFE against the alternative hypothesis that the linear AR model MSFE is greater than the Band-TAR model MSFE.

[e]Modified weighted Diebold & Mariano (1995) test statistic for the null hypothesis that the linear AR model weighted MSFE equals the Band-TAR model weighted MSFE against the alternative hypothesis that the linear AR model weighted MSFE is greater than the Band-TAR model weighted MSFE.

Table 3: Out-of-sample point forecast evaluation, linear AR and Taylor, Peel, & Sarno (2001) ESTAR models

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
h[a]	AR[b]	Band-TAR/AR[c]	*M-DM*[d]	*MW-DM*[e]	AR[b]	Band-TAR/AR[c]	*M-DM*[d]	*MW-DM*[e]
	A. United Kingdom, 1997:01-2003:07 out-of-sample period				C. France, 1997:01-2003:06 out-of-sample period			
1	0.02	1.01	-1.04 [0.85] {0.75}	0.89 [0.19] {0.43}	0.03	0.98	1.27 [**0.10**] {0.14}	1.64 [**0.05**] {0.24}
2	0.03	1.01	-0.56 [0.71] {0.62}	1.03 [0.15] {0.27}	0.04	0.97	1.14 [0.13] {**0.09**}	1.45 [**0.08**] {0.21}
3	0.03	1.02	-0.57 [0.72] {0.60}	1.08 [0.14] {0.30}	0.05	0.97	0.91 [0.18] {0.21}	1.22 [0.11] {0.32}
6	0.04	1.04	-0.84 [0.80] {0.67}	1.09 [0.14] {0.32}	0.08	0.93	1.18 [0.12] {0.15}	1.44 [**0.08**] {0.25}
9	0.05	1.07	-1.16 [0.87] {0.75}	0.24 [0.41] {0.54}	0.10	0.90	1.16 [0.12] {0.18}	1.40 [**0.08**] {0.25}
12	0.06	1.08	-1.42 [0.92] {0.82}	-0.42 [0.66] {0.62}	0.12	0.88	1.52 [**0.07**] {0.12}	1.67 [**0.05**] {0.19}
15	0.07	1.09	-1.47 [0.93] {0.81}	-0.66 [0.75] {0.65}	0.14	0.87	1.65 [**0.05**] {0.12}	1.81 [**0.04**] {0.14}
18	0.07	1.11	-1.99 [0.97] {0.90}	-1.16 [0.88] {0.78}	0.16	0.85	1.84 [**0.04**] {**0.10**}	2.00 [**0.03**] {**0.09**}
21	0.07	1.10	-2.02 [0.98] {0.88}	-1.15 [0.87] {0.80}	0.18	0.84	2.32 [**0.01**] {**0.08**}	2.45 [**0.01**] {**0.06**}
24	0.07	1.12	-2.70 [1.00] {0.92}	-1.68 [0.95] {0.91}	0.20	0.83	2.83 [**0.00**] {**0.04**}	2.99 [**0.00**] {**0.04**}
	B. Germany, 1997:01-2003:07 out-of-sample period				D. Japan, 1997:01-2003:06 out-of-sample period			
1	0.03	0.99	0.83 [0.20] {0.19}	1.05 [0.15] {0.39}	0.04	1.00	-0.68 [0.75] {0.65}	0.63 [0.26] {0.41}
2	0.04	0.98	0.57 [0.28] {0.26}	0.84 [0.20] {0.40}	0.05	1.01	-0.45 [0.67] {0.59}	0.61 [0.27] {0.35}
3	0.05	0.97	0.71 [0.24] {0.21}	0.93 [0.18] {0.39}	0.06	1.01	-0.76 [0.77] {0.69}	0.49 [0.31] {0.44}
6	0.07	0.94	0.76 [0.23] {0.24}	1.04 [0.15] {0.36}	0.09	1.02	-0.86 [0.80] {0.68}	0.32 [0.37] {0.51}
9	0.09	0.92	0.80 [0.21] {0.26}	1.03 [0.15] {0.34}	0.09	1.01	-0.41 [0.66] {0.56}	0.59 [0.28] {0.42}
12	0.12	0.90	0.97 [0.17] {0.25}	1.17 [0.12] {0.31}	0.11	1.02	-0.56 [0.71] {0.58}	0.68 [0.25] {0.40}
15	0.14	0.88	1.08 [0.14] {0.22}	1.29 [0.10] {0.27}	0.13	1.02	-0.61 [0.73] {0.59}	0.53 [0.30] {0.44}
18	0.15	0.86	1.30 [**0.10**] {0.17}	1.49 [**0.07**] {0.22}	0.13	1.02	-0.52 [0.70] {0.56}	0.96 [0.17] {0.30}
21	0.17	0.84	1.62 [**0.05**] {0.14}	1.81 [**0.04**] {0.13}	0.14	1.02	-0.56 [0.71] {0.57}	1.60 [**0.06**] {0.13}
24	0.19	0.82	2.07 [**0.02**] {**0.10**}	2.25 [**0.01**] {**0.07**}	0.14	1.01	-0.41 [0.66] {0.54}	2.41 [**0.01**] {**0.06**}

Notes: *p*-value using the Student's *t* distribution with $P_h - 1$ degrees of freedom is reported in brackets; bootstrapped *p*-value is reported in curly brackets; bold *p*-value indicates significance at the 10% level according to the *p*-value; 0.00 indicates <0.005; 1.00 indicates ≥ 0.995.

[a]Forecast horizon (in months).

[b]Linear AR model RMSFE.

[c]Ratio of the Band-TAR model RMSFE to the linear AR model RMSFE.

[d]Modified Diebold & Mariano (1995) test statistic for the null hypothesis that the linear AR model MSFE equals the Band-TAR model MSFE against the alternative hypothesis that the linear AR model MSFE is greater than the ESTAR model MSFE.

[e]Modified weighted Diebold & Mariano (1995) test statistic for the null hypothesis that the linear AR model weighted MSFE equals the ESTAR model weighted MSFE against the alternative hypothesis that the linear AR model weighted MSFE is greater than the Band-TAR model weighted MSFE.

Table 4: Out-of-sample interval forecast evaluation, linear AR and Obstfeld & Taylor (1997) Band-TAR models

(1) Model	(2) h[a]	(3) $0.10/h$	(4) χ^2_{UC}[b]	(5) χ^2_{IND}[c]	(6) χ^2_{CC}[d]
A. United Kingdom, 1994:12-2003:07 out-of-sample period					
Linear AR	1	0.10	**41.88** [0.00]	0.97 [0.36]	**41.60** [0.00]
Linear AR	2	0.05	**19.69** [0.00], **16. 49** [0.00]	**7.29** [0.02], **7.03** [0.02]	**23.44** [0.00], **22.50** [0.00]
Linear AR	3	0.033	**16.94** [0.00], **16.94** [0.00], **16.94** [0.00]	**9.22** [0.03], **9.22** [0.03], **9.22** [0.03]	**22.87** [0.00], **22.87** [0.00], **22.87** [0.00]
Band-TAR	1	0.10	**24.04** [0.00]	**6.17** [0.02]	**28.08** [0.00]
Band-TAR	2	0.05	**11.08** [0.00], **14.29** [0.00]	**7.29** [0.02], **7.03** [0.02]	**23.44** [0.00], **22.50** [0.00]
Band-TAR	3	0.033	**16.94** [0.00], **9.53** [0.00], **16.94** [0.00]	2.83 [0.15], 2.83 [0.15], 2.83 [0.15]	**17.49** [0.00], **17.49** [0.00], **17.49** [0.00]
B. Germany, 1995:01-2003:07 out-of-sample period					
Linear AR	1	0.10	**5.14** [0.03]	0.02 [1.00]	**5.66** [0.06]
Linear AR	2	0.05	0.49 [0.49], 2.37 [0.16]	0.59 [0.53], 0.59 [0.53]	**5.65** [0.05], **5.65** [0.05]
Linear AR	3	0.033	0.12 [0.86], 0.00 [1.00], 0.27 [0.61]	5.30 [0.05], 5.30 [0.05], 4.89 [0.05]	**8.38** [0.02], **8.38** [0.02], **7.53** [0.02]
Band-TAR	1	0.10	1.64 [0.24]	0.94 [0.42]	2.85 [0.25]
Band-TAR	2	0.05	0.18 [0.78], 0.49 [0.49]	0.47 [0.57], 0.47 [0.57]	1.74 [0.44], 1.74 [0.44]
Band-TAR	3	0.033	0.47 [0.61], 0.12 [0.86], 0.03 [1.00]	5.30 [0.05], 5.30 [0.05], 4.89 [0.05]	**8.38** [0.02], **8.38** [0.02], **7.53** [0.02]
C. France, 1994:12-2003:06 out-of-sample period					
Linear AR	1	0.10	**5.14** [0.03]	0.92 [0.41]	**5.62** [0.07]
Linear AR	2	0.05	0.49 [0.49], 1.59 [0.26]	1.49 [0.33], 1.49 [0.33]	**7.78** [0.02], **7.78** [0.02]
Linear AR	3	0.033	0.12 [0.86], 0.47 [0.61], 0.27 [0.61]	2.83 [0.14], 2.83 [0.14], 3.56 [0.13]	5.08 [0.09], 5.08 [0.09], 5.33 [0.07]
Band-TAR	1	0.10	**5.14** [0.03]	0.92 [0.41]	**5.62** [0.07]
Band-TAR	2	0.05	1.59 [0.26], 1.59 [0.26]	1.00 [0.35], 1.00 [0.35]	6.01 [0.05], 6.01 [0.05]
Band-TAR	3	0.033	0.47 [0.61], 0.12 [0.86], 0.03 [1.00]	**9.43** [0.01], **9.43** [0.01], **10.93** [0.00]	**12.05** [0.00], **12.05** [0.00], **12.05** [0.00]
D. Japan, 1994:12-2003:06 out-of-sample period					
Linear AR	1	0.10	2.81 [0.11]	0.75 [0.42]	3.24 [0.20]
Linear AR	2	0.05	2.37 [0.13], 0.02 [0.89]	0.26 [0.76], 0.26 [0.76]	4.16 [0.14], 4.16 [0.14]
Linear AR	3	0.033	1.06 [0.39], 0.47 [0.61], 0.76 [0.49]	0.05 [1.00], 0.05 [1.00], 0.17 [1.00]	6.86 [0.033], 6.86 [0.033], 6.26 [0.05]
Band-TAR	1	0.10	**4.28** [0.05]	0.37 [0.68]	4.28 [0.12]
Band-TAR	2	0.05	1.59 [0.26], 0.96 [0.40]	0.00 [1.00], 0.00 [1.00]	0.72 [0.74], 0.72 [0.74]
Band-TAR	3	0.033	0.12 [0.86], 0.47 [0.61], 0.76 [0.39]	0.59 [0.70], 0.59 [0.70], 0.91 [0.44]	4.19 [0.13], 4.19 [0.13], 3.95 [0.16]

Notes: statistics are reported for each of the h subgroups; exact p-value is reported in brackets; bold statistic indicates significance at the $0.10/h$ level according to the exact p-value; 0.00 indicates <0.005.

[a]Forecast horizon (in months).

[b]Pearson χ^2 test statistic for the null hypothesis that the prediction intervals have correct unconditional coverage.

[c]Pearson χ^2 test statistic for the null hypothesis that the "hits" relating to the prediction intervals are independent.

[d]Pearson χ^2 test statistic for the null hypothesis that the prediction intervals have correct conditional coverage.

Table 5: Out-of-sample interval forecast evaluation, linear AR and Taylor, Peel, & Sarno (2001) ESTAR models

(1)	(2)	(3)	(4)	(5)	(6)
Model	h[a]	0.10/h	χ^2_{UC} [b]	χ^2_{IND} [c]	χ^2_{CC} [d]
A. United Kingdom, 1997:01-2003:07 out-of-sample period					
Linear AR	1	0.10	**21.28** [0.00]	**4.05** [0.06]	**25.49** [0.00]
Linear AR	2	0.05	**9.2**6 [0.00], **9.26** [0.00]	0.08 [1.00], 0.08 [1.00]	**20.67** [0.00], **20.67** [0.00]
Linear AR	3	0.033	**9.85** [0.00], **12.46** [0.00], **11.56** [0.00]	0.19 [1.00], 0.19 [1.00], 0.20 [1.00]	**17.70** [0.00], **17.70** [0.00], **16.73** [0.00]
ESTAR	1	0.10	**21.28** [0.00]	**4.05** [0.06]	**25.49** [0.00]
ESTAR	2	0.05	**9.2**6 [0.00], **9.26** [0.00]	0.08 [1.00], 0.08 [1.00]	**20.67** [0.00], **20.67** [0.00]
ESTAR	3	0.033	3.85 [0.08], **15.38** [0.00], **11.56** [0.00]	0.30 [1.00], 0.30 [1.00], 0.20 [1.00]	**17.73** [0.00], **17.73** [0.00], **16.73** [0.00]
B. Germany, 1997:01-2003:07 out-of-sample period					
Linear AR	1	0.10	1.03 [0.37]	0.01 [1.00]	1.30 [0.56]
Linear AR	2	0.05	0.25 [0.75], 0.03 [1.00]	**7.67** [0.01], **7.67** [0.01]	**9.01** [0.01], **9.01** [0.01]
Linear AR	3	0.033	0.15 [0.85], 1.38 [0.33], 1.96 [0.23]	0.04 [1.00], 0.04 [1.00], 0.00 [1.00]	0.08 [1.00], 0.08 [1.00], 0.17 [1.00]
ESTAR	1	0.10	2.14 [0.18]	0.05 [1.00]	2.56 [0.31]
ESTAR	2	0.05	1.26 [0.34], 0.64 [0.52]	2.75 [0.14], 2.75 [0.14]	**7.54** [0.02] **7.54** [0.02]
ESTAR	3	0.033	0.15 [0.85], 0.62 [0.44], 0.36 [0.69]	0.89 [0.43], 0.89 [0.43], 0.49 [0.68]	1.23 [0.55], 1.23 [0.55], 1.14 [0.62]
C. France, 1997:01-2003:06 out-of-sample period					
Linear AR	1	0.10	0.00 [1.00]	1.58 [0.26]	1.59 [0.49]
Linear AR	2	0.05	0.64 [0.52], 0.00 [1.00]	**8.53** [0.01], **7.80** [0.01]	**8.53** [0.02], **7.82** [0.02]
Linear AR	3	0.033	1.38 [0.33], 1.96 [0.23], 1.00 [0.42]	0.04 [1.00], 0.00 [1.00], 0.00 [1.00]	0.08 [1.00], 0.00 [1.00], 0.00 [1.00]
ESTAR	1	0.10	0.05 [0.91]	2.97 [0.11]	3.09 [0.23]
ESTAR	2	0.05	0.23 [0.75], 0.11 [0.87]	1.65 [0.30], 1.07 [0.47]	4.17 [0.13], 4.24 [0.13]
ESTAR	3	0.033	1.38 [0.33], 1.00 [0.42], 1.96 [0.23]	0.04 [1.00], 0.00 [1.00], 0.00 [1.00]	0.08 [1.00], 0.00 [1.00], 0.00 [1.00]
D. Japan, 1997:01-2003:06 out-of-sample period					
Linear AR	1	0.10	2.51 [0.14]	0.79 [0.48]	3.68 [0.17]
Linear AR	2	0.05	0.03 [0.87], 1.68 [0.26]	0.02 [1.00], 0.03 [1.00]	2.65 [0.28], 3.29 [0.22]
Linear AR	3	0.033	0.00 [1.00], 1.00 [0.42], 0.36 [0.69]	0.26 [0.69], 0.11 [1.00], 0.11 [1.00]	1.25 [0.55], 1.60 [0.53], 1.60 [0.53]
ESTAR	1	0.10	1.85 [0.21]	0.64 [0.49]	2.82 [0.26]
ESTAR	2	0.05	0.23 [0.75], 0.95 [0.42]	0.21 [0.74], 0.05 [1.00]	1.15 [0.58], 1.37 [0.54]
ESTAR	3	0.033	0.62 [0.44], 1.00 [0.42], 0.04 [1.00]	0.62 [0.66], 0.38 [0.67], 0.38 [0.67]	2.53 [0.31], 3.00 [0.22], 3.00 [0.22]

Notes: statistics are reported for each of the h subgroups; exact p-value is reported in brackets; bold statistic indicates significance at the 0.10/h level according to the exact p-value; 0.00 indicates <0.005.

[a]Forecast horizon (in months).

[b]Pearson χ^2 test statistic for the null hypothesis that the prediction intervals have correct unconditional coverage.

[c]Pearson χ^2 test statistic for the null hypothesis that the "hits" relating to the prediction intervals are independent.

[d]Pearson χ^2 test statistic for the null hypothesis that the prediction intervals have correct conditional coverage.

Table 6: Out-of-sample density forecast evaluation, linear AR and Obstfeld & Taylor (1997) Band-TAR models

(1) Model	(2) h[a]	(3) $0.10/h$	(4) KS[b]	(5) DH[c]	(6) $LB,\ k=1$[d]	(7) $LB,\ k=2$[d]	(8) $LB,\ k=3$[d]	(9) $LB,\ k=4$[d]
A. United Kingdom, 1994:12-2003:07 out-of-sample period								
Linear AR	1	0.10	**0.19**	0.91	1.29	**13.54**	0.46	1.33
Linear AR	2	0.05	**0.20**, **0.20**	2.00, 1.05	**2.85**, 0.00	**12.81**, **10.50**	**4.98**, 0.10	**6.24**, **4.82**
Linear AR	3	0.033	**0.27**, **0.25**, **0.24**	1.45, 1.58, 0.63	0.37, 0.58, 0.08	**4.62**, **6.28**, 4.43	0.29, 0.45, 0.35	0.64, 0.60, 0.52
Band-TAR	1	0.10	**0.14**	1.18	1.54	**15.62**	0.46	1.68
Band-TAR	2	0.05	0.17, 0.17	1.74, 0.83	3.32, 0.01	**14.99**, **11.50**	**4.49**, 0.00	**6.59**, **6.03**
Band-TAR	3	0.033	0.21, 0.18, 0.23	1.66, 2.17, 1.24	0.18, 1.09, 0.01	**5.46**, **5.47**, **6.29**	0.09, 0.58, 0.20	1.05, 0.49, 0.45
B. Germany, 1995:01-2003:07 out-of-sample period								
Linear AR	1	0.10	0.09	2.75	2.14	**25.20**	1.73	**6.08**
Linear AR	2	0.05	0.12, 0.16	1.68, **5.98**	0.38, 0.00	**14.47**, **4.71**	0.01, 0.06	**5.23**, 0.33
Linear AR	3	0.033	0.15, 0.14, 0.13	1.94, 0.46, 1.41	0.00, 0.05, 0.14	4.14, **5.97**, **12.14**	0.00, 0.39, 0.13	0.18, 1.56, **7.71**
Band-TAR	1	0.10	0.10	3.91	2.08	**24.41**	1.55	**5.79**
Band-TAR	2	0.05	0.14, 0.18	1.63, 5.13	0.57, 0.00	**13.28**, **4.66**	0.02, 0.11	**3.88**, 0.33
Band-TAR	3	0.033	0.17, 0.14, 0.16	1.05, 1.41, 1.42	0.12, 0.00, 0.05	3.68, **6.73**, **11.22**	0.03, 0.23, 1.14	0.14, 1.33, **7.68**
C. France, 1994:12-2003:06 out-of-sample period								
Linear AR	1	0.10	0.11	1.65	1.23	**26.97**	0.81	**4.62**
Linear AR	2	0.05	0.14, 0.12	3.26, 0.62	0.85, 0.61	**5.68**, **17.25**	0.08, 0.01	0.62, **5.04**
Linear AR	3	0.033	0.12, 0.12, 0.12	1.07, 1.04, 1.00	0.31, 0.02, 0.01	**12.42**, 4.16, **7.60**	1.36, 0.00, 0.15	**7.14**, 0.22, 2.81
Band-TAR	1	0.10	0.10	1.90	0.81	**26.38**	0.64	**4.36**
Band-TAR	2	0.05	0.14, 0.15	3.05, 0.70	0.86, 0.62	**5.25**, **15.43**	0.00, 0.00	0.38, **3.66**
Band-TAR	3	0.033	0.12, 0.15, 0.14	0.55, 0.52, 1.92	0.26, 0.06, 0.00	**10.92**, 3.67, **9.07**	2.12, 0.01, 0.33	**5.91**, 0.14, 2.74
D. Japan, 1994:12-2003:06 out-of-sample period								
Linear AR	1	0.10	**0.16**	**23.07**	2.05	**28.16**	**3.74**	**15.75**
Linear AR	2	0.05	**0.21**, **0.24**	4.19, **12.24**	2.07, 0.38	**21.23**, **4.68**	**6.18**, 0.39	**10.47**, 0.49
Linear AR	3	0.033	0.23, **0.25**, **0.28**	3.87, 2.85, **8.72**	0.03, 0.01, 0.05	**8.72**, **6.14**, 2.54	0.05, 0.05, 0.07	1.54, 0.58, 0.10
Band-TAR	1	0.10	0.12	**3215.26**	2.13	**30.35**	**3.19**	**19.01**
Band-TAR	2	0.05	0.16, 0.16	4.97, **105.76**	2.38, 0.53	**19.35**, **6.92**	**6.17**, 0.71	**10.66**, 0.88
Band-TAR	3	0.033	0.16, 0.21, 0.22	**344**, 5.00, **55.85**	0.00, 0.05, 0.00	**11.08**, **8.66**, 2.75	0.32, 0.26, 0.04	3.24, 2.94, 0.17

Notes: statistics are reported for each of the h subgroups; bold statistic indicates significance at the $0.10/h$ level; 0.00 indicates <0.005.

[a]Forecast horizon (in months).

[b]Kolmogorov-Smirnov test statistic for the null hypothesis that $z_t \sim U(0,1)$.

[c]Doornik and Hansen (1994) test statistic for the null hypothesis that $z_t^* \sim N(0,1)$.

[d]Ljung-Box test statistic for the null hypothesis of no first-order autocorrelation in $(z_t - \bar{z})^k$, $k = 1,\ldots,4$.

Table 7: Out-of-sample density forecast evaluation, linear AR and Taylor, Peel, & Sarno (2001) ESTAR models

(1) Model	(2) h[a]	(3) $0.10/h$	(4) KS[b]	(5) DH[c]	(6) $LB,\ k=1$[d]	(7) $LB,\ k=2$[d]	(8) $LB,\ k=3$[d]	(9) $LB,\ k=4$[d]
A. United Kingdom, 1997:01-2003:07 out-of-sample period								
Linear AR	1	0.10	**0.16**	0.91	0.70	**9.39**	0.58	1.13
Linear AR	2	0.05	0.17, 0.17	0.72, 1.64	0.01, 2.59	**5.13**, **9.57**	0.06, 2.77	1.18, **4.18**
Linear AR	3	0.033	0.19, 0.19, 0.22	1.37, 0.76, 0.05	0.30, 0.00, 1.36	**4.63**, 3.94, 4.04	0.11, 0.21, 1.67	0.27, 0.43, 1.37
ESTAR	1	0.10	**0.15**	1.37	0.56	**9.46**	0.44	1.06
ESTAR	2	0.05	0.18, 0.19	0.47, 1.94	0.08, 3.00	**4.86**, **12.72**	0.03, 2.75	1.21, **6.54**
ESTAR	3	0.033	0.20, 0.24, 0.22	1.73, 1.55, 0.47	0.52, 0.12, 1.04	3.72, **5.84**, 4.04	0.12, 0.07, 1.36	0.10, 0.51, 1.10
B. Germany, 1997:01-2003:07 out-of-sample period								
Linear AR	1	0.10	**0.16**	**4.89**	2.31	**17.60**	2.04	2.80
Linear AR	2	0.05	0.21, **0.23**	2.18, 4.31	0.03, 0.07	**7.01**, 3.34	0.24, 0.03	0.71, 0.22
Linear AR	3	0.033	0.22, 0.23, 0.20	1.33, 0.43, 1.10	0.09, 0.55, 0.01	2.94, **4.94**, **11.65**	0.03, 0.70, 0.70	0.07, 1.00, **7.78**
ESTAR	1	0.10	0.12	**5.61**	2.32	**19.34**	2.37	**4.30**
ESTAR	2	0.05	0.17, 0.18	2.73, 3.83	0.11, 0.11	**7.80**, **4.22**	0.25, 0.01	1.24, 0.52
ESTAR	3	0.033	0.15, 0.16, 0.15	1.04, 0.58, 2.34	0.04, 0.59, 0.01	3.58, **6.07**, **11.46**	0.04, 0.84, 0.04	0.25, 1.73, **8.43**
C. France, 1997:01-2003:06 out-of-sample period								
Linear AR	1	0.10	**0.15**	3.48	2.48	**20.55**	2.33	**3.79**
Linear AR	2	0.05	**0.25**, 0.20	2.19, 2.96	0.04, 0.91	**8.46**, 2.83	0.10, 0.00	0.85, 0.15
Linear AR	3	0.033	0.22, 0.22, 0.24	1.53, 1.16, 1.62	0.04, 0.37, 0.10	2.77, **4.82**, **10.75**	0.01, 0.48, 1.49	0.06, 0.76, **6.10**
ESTAR	1	0.10	0.12	3.12	2.52	**19.50**	2.24	**4.03**
ESTAR	2	0.05	0.21, 0.14	2.12, 3.23	0.09, 1.21	**10.30**, 3.67	0.11, 0.00	1.87, 0.33
ESTAR	3	0.033	0.17, 0.20, 0.20	1.76, 1.92, 1.89	0.02, 0.14, 0.05	3.91, **6.25**, **11.83**	0.00, 0.38, 0.80	0.26, 1.33, **7.31**
D. Japan, 1997:01-2003:06 out-of-sample period								
Linear AR	1	0.10	0.13	**16.62**	0.83	**20.05**	0.35	**10.81**
Linear AR	2	0.05	0.18, 0.16	**11.71**, 3.62	0.07, 0.33	**5.70**, **17.62**	0.36, 3.15	0.89, **7.23**
Linear AR	3	0.033	0.20, 0.20, 0.16	5.04, **9.27**, 1.84	0.32, 0.27, 0.00	4.53, 3.36, **8.07**	0.00, 0.29, 0.05	0.37, 0.58, 2.29
ESTAR	1	0.10	0.11	**60.55**	0.80	**20.08**	0.33	**11.40**
ESTAR	2	0.05	0.17, 0.15	**62.06**, **4.68**	0.06, 0.27	**5.72**, **18.00**	0.28, 2.71	0.91, **8.60**
ESTAR	3	0.033	0.18, 0.18, 0.14	4.60, **57.17**, 1.82	0.28, 0.42, 0.00	**5.36**, 3.72, **8.38**	0.02, 0.49, 0.00	0.60, 0.79, 2.66

Notes: statistics are reported for each of the h subgroups; bold statistic indicates significance at the $0.10/h$ level; 0.00 indicates <0.005.

[a]Forecast horizon (in months).

[b]Kolmogorov-Smirnov test statistic for the null hypothesis that $z_t \sim U(0,1)$.

[c]Doornik and Hansen (1994) test statistic for the null hypothesis that $z_t^* \sim N(0,1)$.

[d]Ljung-Box test statistic for the null hypothesis of no first-order autocorrelation in $(z_t - \bar{z})^k$, $k = 1,\ldots,4$.

Table 8: In-sample comparison of conditional densities corresponding to fitted nonlinear and linear AR models

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
			Block bootstrapped Z_T critical values				Block bootstrapped R-Z_T critical values		
Country, in-sample period	Nonlinear model	Z_T[a]	10%	5%	1%	R-Z_T[b]	10%	5%	1%
United Kingdom, 1980:02-1994:11	Band-TAR	0.0005	0.0057	0.0072	0.0111	0.0008	0.0036	0.0048	0.0137
Germany, 1980:02-1994:12	Band-TAR	0.0006	0.0022	0.0033	0.0084	0.0007	0.0039	0.0044	0.0054
France, 1980:02-1994:11	Band-TAR	0.0010	0.0027	0.0034	0.0071	0.0024	**0.0013**	**0.0023**	0.0029
Japan, 1980:02-1994:11	Band-TAR	0.0012	0.0035	0.0038	0.0047	0.0019	0.0049	0.0061	0.0031
United Kingdom, 1973:02-1996:12	ESTAR	-0.0006	0.0033	0.0051	0.0088	-0.0018	0.0030	0.0044	0.0072
Germany, 1973:02-1996:12	ESTAR	0.0010	0.0029	0.0039	0.0043	0.0030	**0.0025**	**0.0030**	0.0034
France, 1973:02-1996:12	ESTAR	0.0004	0.0026	0.0034	0.0050	0.0026	0.0032	0.0037	0.0051
Japan, 1973:02-1996:12	ESTAR	0.0009	0.0027	0.0039	0.0043	0.0005	0.0024	0.0032	0.0047

Notes: bold bootstrapped critical value indicates that the statistic is significant according to the bootstrapped critical value.

[a]Corradi & Swanson (2003) test statistic for the null hypothesis that the conditional densities corresponding to the nonlinear and linear AR models are equally accurate relative to the true conditional density against the alternative hypothesis that the conditional density corresponding to the nonlinear AR model is more accurate than the conditional density corresponding to the linear AR model.

[b]Corradi & Swanson (2003) test statistic for the null hypothesis that the conditional densities corresponding to the nonlinear and linear AR models are equally accurate relative to the true conditional density against the alternative hypothesis that the conditional density corresponding to the nonlinear AR model is more accurate than the conditional density corresponding to the linear AR model for values of q_t in the upper and lower quartiles of the in-sample observations.

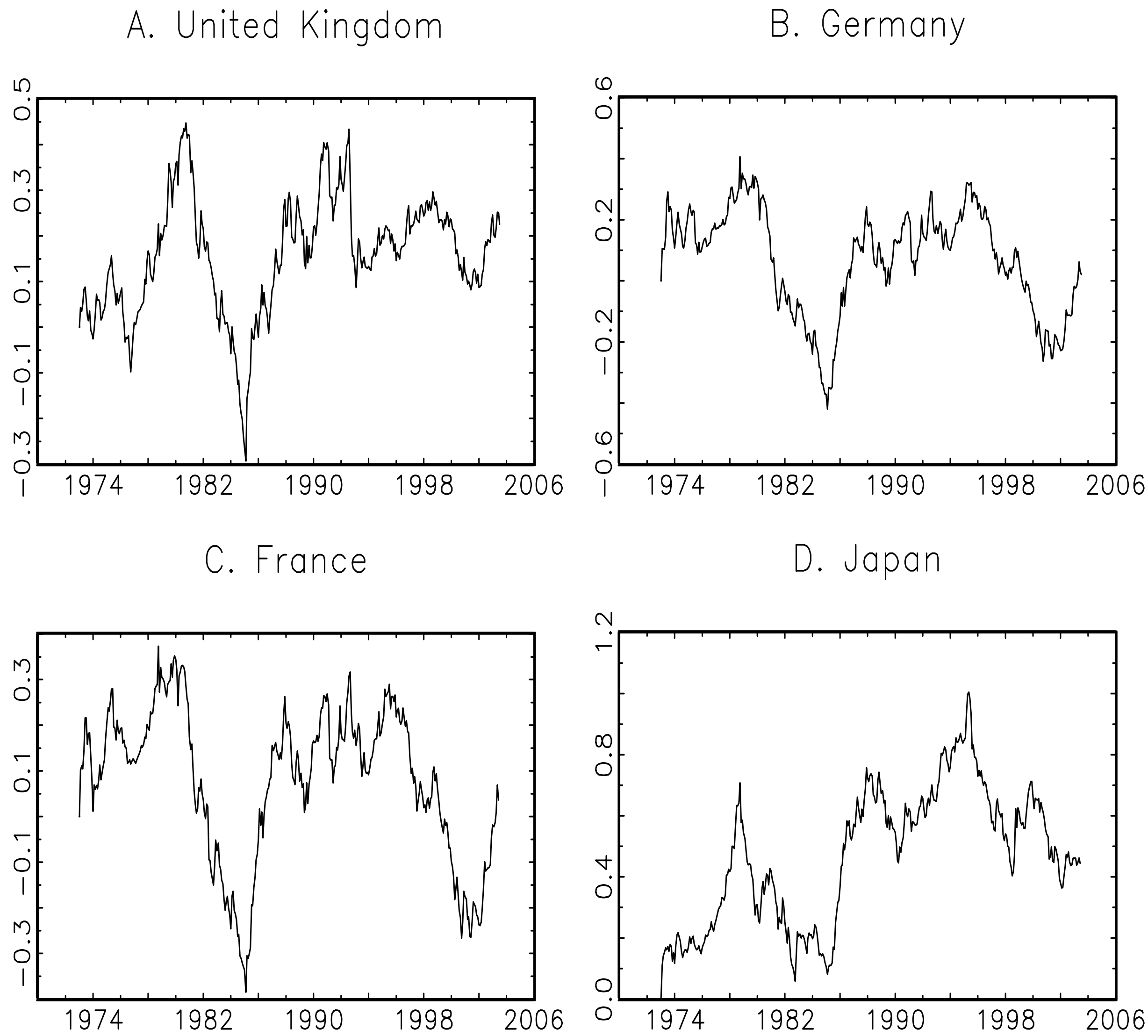

Figure 1: Real exchange rate log-levels, 1973:01-2003:07 (United Kingdom, Germany) and 1973:01-2003:06 (France, Japan)

Note: the real exchange rate log-levels are normalized to zero in 1973:01

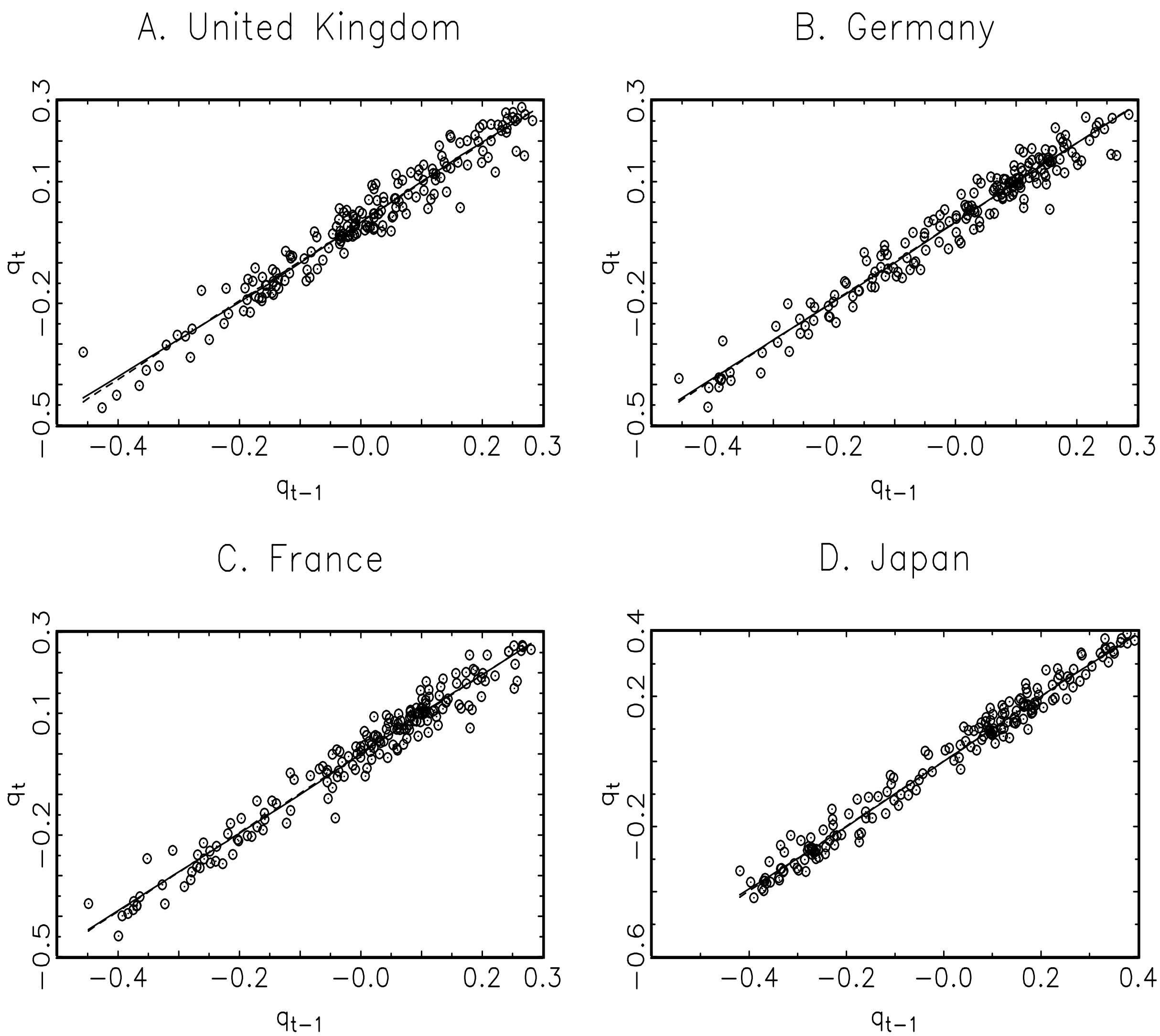

Figure 2: Scatterplot of real exchange rate log-levels (q_t) and lagged real exchange rate log-levels (q_{t-1}), Obstfeld & Taylor (1997) data.

Notes: solid line is the conditional expectation function for the fitted Obstfeld & Taylor (1997) Band-TAR model; dashed line is the conditional expectation function for a fitted linear AR(1) model.

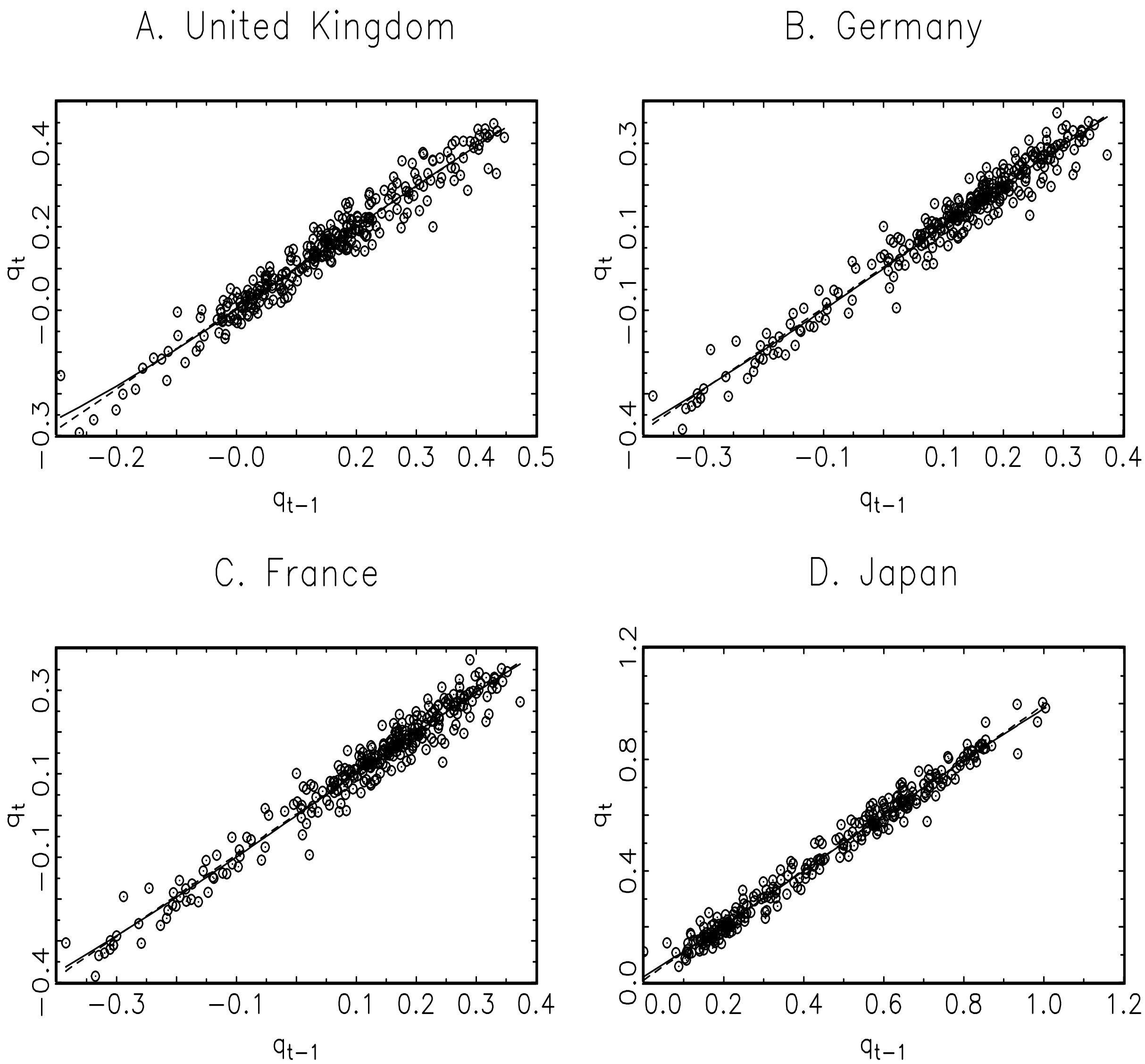

Figure 3: Scatterplot of real exchange rate log-level (q_t) and lagged real exchange rate log-level (q_{t-1}), Taylor, Peel, & Sarno (2001) data.

Notes: solid line is the conditional expectation function for the fitted Taylor, Peel, & Sarno (2001) ESTAR model; dashed line is the conditional expectation function for a fitted linear AR(1) model.