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

January 18, 2005

Abstract

In this paper, we analyze the out-of-sample forecasting performance of a number of prominent 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 fairly little evidence that favors either band-threshold or exponential smooth transition autoregressive models over simple linear autoregressive models in terms of out-of-sample forecasting performance. Most of the evidence in support of nonlinear autoregressive specifications comes from a comparison of interval forecasts. Using a long span of annual data, there is more support for an exponential smooth transition autoregressive model over a linear autoregressive model for the U.K.-U.S. real exchange rate. 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

*Corresponding author. This is a significantly revised version of our earlier paper, "Nonlinear Models of Real Exchange Rate Behavior: A Re-examination." We thank Nathan Balke, Mike Bergman, Mike Clements, 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.

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 and 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, several recent studies estimate nonlinear autoregressive (AR) models for U.S. dollar real exchange rates. Obstfeld and Taylor (1997) estimate band-threshold AR (Balke and 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 over the post-Bretton Woods period. 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. Taylor, Peel, and Sarno (2001) consider exponential smooth transition AR (Granger and 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] Bertola and 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. Using monthly data for the U.S. dollar real exchange rate vis-à-vis the U.K.,

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

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

Germany, France, and Japan over the post-Bretton Woods era, Taylor, Peel, and 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, and Sarno (2001) conclude that the real exchange rates they consider are well characterized by nonlinear mean-reversion. Michael, Nobay, and Peel (1997) also estimate ESTAR processes for U.S. dollar real exchange rates. They use monthly interwar data, as well as the Lothian and Taylor (1996) annual data that cover more than two centuries. Michael, Nobay, and Peel (1997) reject linear AR models in favor of ESTAR alternatives and conclude that there is empirical support for theoretical models of real exchange rate behavior incorporating transaction costs.[3]

Obstfeld and Taylor (1997), Taylor, Peel, and Sarno (2001), and Michael, Nobay, and Peel (1997) all 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 and Taylor (1997), Taylor, Peel, and Sarno (2001), and Michael, Nobay, and Peel (1997) to forecast real exchange rate observations that have become available since the models were originally estimated. 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

[3] Sarantis (1999) also estimates ESTAR models for real exchange rates. 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.

evidence in favor of nonlinear model specifications.

We first compare the out-of-sample forecasting performance of nonlinear and linear AR models 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 and 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 and Nason, 1990; De Gooijer and Kumar, 1992; Ramsey, 1996; Stock and 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 and Smith, 1999). In light of this, we also use a weighted version of the Diebold and Mariano (1995) test developed by van Dijk and 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 and Franses (2003) find that forecasts of U.S. output growth generated by the nonlinear "floor and ceiling" model of Pesaran and Potter (1997) are not superior to forecasts generated by a linear AR model according to the conventional Mariano and Diebold (1995) test; however, the floor and ceiling model forecasts are superior to linear AR model forecasts according to the weighted Diebold and Mariano (1995) test.

Our evaluation also includes an analysis of interval and density forecasts generated by linear and nonlinear AR models, and we analyze interval and density forecasts along the lines suggested by Christoffersen (1998) and Diebold, Gunther, and 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 and Smith (2000) with respect to forecasting U.S. GNP growth and unemployment and Siliverstovs and van Dijk (2003) with respect to industrial production growth in the G7 countries. By using the weighted version of the Diebold and

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 and Taylor (1997), Taylor, Peel, and Sarno (2001), and Michael, Nobay, and Peel (1997). Section 3 describes 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 Three Extant Studies

2.1. Obstfeld and Taylor (1997) Band-TAR Model

Motivated by theoretical models of costly arbitrage due to transport costs, Obstfeld and Taylor (1997, OT) estimate Band-TAR models for consumer price index-based U.S. dollar real exchange rates.[4] Their Band-TAR model takes the form,

$$\Delta q_t = \lambda^{out} \cdot (q_{t-1} - c) + \varepsilon_t^{out} \text{ if } q_{t-1} > c;$$

$$\Delta q_t = \varepsilon_t^{in} \text{ if } c \geq q_{t-1} \geq -c; \tag{1}$$

$$\Delta q_t = \lambda^{out} \cdot (q_{t-1} + c) + \varepsilon_t^{out} \text{ if } -c > q_{t-1},$$

where q_t is the log-level of the real exchange rate, Δ is the first-difference operator, $\varepsilon_t^{out} \sim N(0, \sigma^{out^2})$,

[4] 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.

and $\varepsilon_t^{in} \sim N(0, \sigma^{in^2})$.[5] 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 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 $\lambda^{out} < 0$.

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, and Sarno (2001). OT obtain estimates of λ^{out} that are negative and sometimes sizable in absolute value. (They do not report standard errors or *t*-statistics for their point estimates.) 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.[6]

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 through 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 appear below, and they are similar to those in OT.

---

[5] OT measure the log-level of the real exchange rate as $q_t = p_t^1 - p_t^2$, where p_t^2 is the log-level of the U.S. consumer price index and p_t^1 is the sum of the log-level of the foreign consumer price index and the log-level of the U.S. dollar/foreign currency exchange rate. Before entering equation (1), q_t is demeaned or detrended.

[6] OT also employ a likelihood-ratio test of the linear AR null hypothesis against a Band-TAR alternative. The null cannot be rejected for any of the four countries we consider.

[7] Following Obstfeld and Taylor (1997), we first demean q_t before estimating equation (1). Obstfeld and Taylor (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 other studies we consider. Our results are qualitatively unchanged when we use detrended data.

United Kingdom, 1980:02-1994:11:

$$\begin{aligned} \Delta q_t &= -0.084 \cdot (q_{t-1} - 0.163) + \varepsilon_t^{out} \text{ if } q_{t-1} > 0.163,\ \hat{\sigma}^{out} = 0.044; \\ \Delta q_t &= \varepsilon_t^{in} \text{ if } 0.163 \geq q_{t-1} \geq -0.163,\ \hat{\sigma}^{in} = 0.033; \\ \Delta q_t &= -0.084 \cdot (q_{t-1} + 0.163) + \varepsilon_t^{out} \text{ if } -0.163 > q_{t-1},\ \hat{\sigma}^{out} = 0.044 \end{aligned} \tag{2}$$

Germany, 1980:02-1994:12:

$$\begin{aligned} \Delta q_t &= -0.052 \cdot (q_{t-1} - 0.111) + \varepsilon_t^{out} \text{ if } q_{t-1} > 0.111,\ \hat{\sigma}^{out} = 0.038; \\ \Delta q_t &= \varepsilon_t^{in} \text{ if } 0.111 \geq q_{t-1} \geq -0.111,\ \hat{\sigma}^{in} = 0.035; \\ \Delta q_t &= -0.052 \cdot (q_{t-1} + 0.111) + \varepsilon_t^{out} \text{ if } -0.111 > q_{t-1},\ \hat{\sigma}^{out} = 0.038 \end{aligned} \tag{3}$$

France, 1980:02-1994:11:

$$\begin{aligned} \Delta q_t &= -0.048 \cdot (q_{t-1} - 0.107) + \varepsilon_t^{out} \text{ if } q_{t-1} > 0.107,\ \hat{\sigma}^{out} = 0.038; \\ \Delta q_t &= \varepsilon_t^{in} \text{ if } 0.107 \geq q_{t-1} \geq -0.107,\ \hat{\sigma}^{in} = 0.031 \\ \Delta q_t &= -0.048 \cdot (q_{t-1} + 0.107) + \varepsilon_t^{out} \text{ if } -0.107 > q_{t-1},\ \hat{\sigma}^{out} = 0.038 \end{aligned} \tag{4}$$

Japan, 1980:02-1994:11:

$$\begin{aligned} \Delta q_t &= -0.057 \cdot (q_{t-1} - 0.254) + \varepsilon_t^{out} \text{ if } q_{t-1} > 0.254,\ \hat{\sigma}^{out} = 0.030; \\ \Delta q_t &= \varepsilon_t^{in} \text{ if } 0.254 \geq q_{t-1} \geq -0.254,\ \hat{\sigma}^{in} = 0.036; \\ \Delta q_t &= -0.057 \cdot (q_{t-1} + 0.254) + \varepsilon_t^{out} \text{ if } -0.254 > q_{t-1},\ \hat{\sigma}^{out} = 0.030 \end{aligned} \tag{5}$$

2.2. Taylor, Peel, and Sarno (2001) ESTAR Model

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

$$q_t = q_{t-1} - \{1 - \exp[\alpha \cdot (q_{t-1} - \eta)^2]\} \cdot (q_{t-1} - \eta) + \varepsilon_t, \tag{6}$$

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.[8] For the parsimonious ESTAR model, equation (6), 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 $\alpha < 0$).[9] Using monthly data from 1973:01-1996:12, TPS estimate equation (6) using multivariate nonlinear least squares (nonlinear seemingly unrelated regressions) for the real exchange rate for four countries relative to the U.S. dollar: U.K., Germany, France, and Japan.[10] 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 and 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,[11] 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 (6) 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 (6)). The estimated ESTAR models appear below, and the parameter

---

[8] TPS define 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 nominal exchange rate (U.S. dollar price of foreign currency), 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. TPS normalize each real exchange rate series to zero in 1973:01. Note that TPS assume that $\varepsilon_t \sim iid\ N(0,\sigma_\varepsilon^2)$ for their impulse response analysis (described below).

[9] As α increases in absolute value, the nonlinear effect becomes “stronger.”

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

[11] 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, and Tauchen (1993).

estimates are very close to those reported in TPS (see their Table 3).[12]

United Kingdom, 1973:02-1996:12:

$$q_t = q_{t-1} - \{1 - \exp[-0.449 \cdot (q_{t-1} - 0.145)^2]\} \cdot (q_{t-1} - 0.145) + \varepsilon_t, \ \hat{\sigma}_\varepsilon = 0.033 \qquad (7)$$

Germany, 1973:02-1996:12:

$$q_t = q_{t-1} - \{1 - \exp[-0.264 \cdot (q_{t-1} + 0.007)^2]\} \cdot (q_{t-1} + 0.007) + \varepsilon_t, \ \hat{\sigma}_\varepsilon = 0.035 \qquad (8)$$

France, 1973:02-1996:12:

$$q_t = q_{t-1} - \{1 - \exp[-0.289 \cdot (q_{t-1} - 0.049)^2]\} \cdot (q_{t-1} - 0.049) + \varepsilon_t, \ \hat{\sigma}_\varepsilon = 0.033 \qquad (9)$$

Japan, 1973:02-1996:12:

$$q_t = q_{t-1} - \{1 - \exp[-0.165 \cdot (q_{t-1} - 0.515)^2]\} \cdot (q_{t-1} - 0.515) + \varepsilon_t, \ \hat{\sigma}_\varepsilon = 0.033 \qquad (10)$$

2.3. Michael, Nobay, and Peel (1997) ESTAR Model

Michael, Nobay, and Peel (1997, MNP) estimate nonlinear real exchange rate models using monthly interwar data and long spans of annual data. We focus on their results for the U.K.-U.S. real exchange rate and the long span of annual data from Lothian and Taylor (1996) covering 1791-1992. The real exchange rate is based on U.K. and U.S. wholesale price indices.[13] Similar to TPS, MNP begin with a general ESTAR specification and test down to a parsimonious ESTAR model of the form,

$$q_t - \eta = \{\exp[\alpha \cdot (q_{t-1} - \eta)^2]\} \cdot [\beta_1 \cdot (q_{t-1} - \eta) + (1 - \beta_1) \cdot (q_{t-2} - \eta)] + \varepsilon_t \, . \qquad (11)$$

As with the ESTAR model estimated by TPS, the MNP ESTAR model behaves as a random walk in the extreme inner regime, while the speed of mean-reversion increases as the real exchange rate moves away

---

[12] The *t*-statistics corresponding to the estimated α parameters in equations (7)-(10) are very close to those reported in TPS and are significant at conventional levels. Equations (7)-(10) also pass the diagnostic tests used by TPS. The complete estimation results are available upon request from the authors.

[13] MNP define 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 pound price of the U.S. dollar, p_t^* is the log-level of the U.K. wholesale price index, and p_t is the log-level of the U.S. wholesale price index.

from its long-run equilibrium value (assuming $\alpha < 0$). MNP estimate equation (11) via nonlinear least squares and obtain a negative and statistically significant estimate of α, so that the speed of mean-reversion increases with the size of the deviation in the real exchange rate from its long-run equilibrium.

We estimate the MNP ESTAR model for the U.K.-U.S. real exchange rate using the Lothian and Taylor (1996) annual data for the U.K.-U.S. nominal exchange rate and U.K. and U.S. wholesale price indices that span the period 1791-1990.[14] Following MNP, we estimate equation (11) using nonlinear least squares. After allowing for two lags, the usable sample covers 1793-1990. The estimated ESTAR model appears below and is very similar to that in MNP.[15]

United Kingdom, 1793-1990:

$$q_t - 1.567 = \{\exp[-2.437 \cdot (q_{t-1} - 1.567)^2]\} \cdot [1.182 \cdot (q_{t-1} - 1.567) - 0.182 \cdot (q_{t-2} - 1.567)] + \varepsilon_t, \hat{\sigma}_\varepsilon = 0.069 . \quad (12)$$

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 or year 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 monthly data and 2002 for annual data. 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);

---

[14] We downloaded the data from Nelson Mark's web page at http://economics.sbs.ohio-state.edu/Mark/nmark.htm. The data are made available to accompany Mark (2001).

[15] The *t*-statistic corresponding to the estimated α parameter is equation (12) is very close to the *t*-statistic reported in MNP (-4.19 versus -4.13) and is significant at conventional levels. The complete estimation results for equation (12) are available upon request from the authors.

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). For the MNP ESTAR model, this leaves the 1991-2002 out-of-sample period (12 observations).

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 (and q_{t-1} for the MNP ESTAR model, which has a lag order of two) for $t = R, \ldots, R + P - (h-1)$.[16] By using this procedure, we are simulating the situation of a forecaster who uses the fitted nonlinear AR models from the three 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 and MNP 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 AR(1) model, while we compare forecasts from the MNP ESTAR model to those generated by a linear AR(2) model. These are the natural linear AR counterparts to the nonlinear AR models given in equations (1), (6), and (11). 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. Consider the linear AR(1) model,

$$q_{t+1} = \phi_0 + \phi_1 \cdot q_t + u_{t+1}, \tag{13}$$

where $u_{t+1} \sim N(0, \sigma_u^2)$. The point forecast of q_{t+1} given q_t is simply $E(q_{t+1} \mid q_t) = \phi_0 + \phi_1 \cdot q_t$. The

[16] In the taxonomy of West and McCracken (1998), we are using a "fixed" sampling scheme.

density forecast for q_{t+1} is given by $N(\phi_0 + \phi_1 \cdot q_t, \sigma_u^2)$, and the density forecast can be used to calculate an interval forecast of any desired size.[17] Using $q_{t+2} = \phi_0 + \phi_1 \cdot q_{t+1} + u_{t+2}$, the point forecast of q_{t+2} given q_t is $E(q_{t+2} \mid q_t) = \phi_0 + \phi_1 \cdot E(q_{t+1} \mid q_t) = \phi_0 + \phi_1 \cdot (\phi_0 + \phi_1 \cdot q_t)$, and the density forecast is given by $N[\phi_0 + \phi_1 \cdot (\phi_0 + \phi_1 \cdot q_t), (\phi_1^2 + 1) \cdot \sigma_u^2]$. We can proceed in a similar manner to generate point, interval, and density forecasts when $h > 2$.

In general, analytical point, interval, and density forecasts are not available for nonlinear AR models when $h \geq 2$, as $E[f(x)] \neq f[E(x)]$. We use the following simulations-based procedure to generate forecasts for the nonlinear AR models.[18] 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 250th and 750th simulated realizations from the ordered set of 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.

3.2. Evaluating Point Forecasts

Denote the point forecast error at horizon *h* corresponding to the nonlinear AR model as $e_{N,t+h|t}$

---

[17] We treat the parameters of the linear and nonlinear AR models as known in forming forecasts, and so we do not explicitly consider parameters estimation uncertainty in forming interval and density forecasts. This is common in the extant literature. See Hansen (2003) for methods of incorporating parameter estimation uncertainty into interval forecasts for linear models.

[18] Simulations-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 and Smith (1997).

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 $P_h = P-(h-1)$. We use the Diebold and Mariano (1995) procedure to test whether the nonlinear AR model MSFE is significantly less than the linear AR model MSFE. Define the MSFE loss differential as $d_{t+h} \equiv e_{L,t+h|t}^2 - e_{N,t+h|t}^2$, ($t = R,\ldots,R+P-h$), so that $E(d_{t+h}) = 0$ under the null hypothesis of equal predictive ability, while $E(d_{t+h}) > 0$ under the one-sided alternative hypothesis that the nonlinear AR model forecasts have a smaller MSFE than the linear AR model forecasts. Assuming that the loss differential series is covariance stationary, Diebold and Mariano (1995) show that the statistic,

$$DM = \bar{d}\cdot[\hat{V}(\bar{d})]^{-0.5}, \tag{14}$$

can be treated as a standard normal variate asymptotically under the null hypothesis of equal predictive ability, where $\bar{d} = (1/P_h)\sum_{t=R}^{R+P-h} d_{t+h}$ and $\hat{V}(\bar{d})$ is a consistent estimate of the asymptotic variance of $\bar{d}$ that is based on estimates of the autocovariances of d_{t+h} up to order $h-1$.[19] Following Siliverstovs and van Dijk (2003), we adjust the *DM* statistic using the Harvey, Leyboure, and Newbold (1997) correction factor that is designed to account for potential finite-sample size distortions. This yields the modified *DM* statistic,

$$M\text{-}DM = \{[P_h + 1 - 2\cdot h + h\cdot(h-1)/P_h]/P_h\}\cdot DM, \tag{15}$$

where we assess significance using the Student's *t* distribution with $P_h - 1$ degrees of freedom.

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. Following van Dijk and Franses (2003), consider a weighted MSFE loss differential,

[19] This assumes that the *h*-step-ahead forecasts display serial correlation up to order $h-1$. We use the Newey and West (1987) procedure with the Bartlett kernel in estimating the asymptotic variance in order to ensure that $\hat{V}(\bar{d})$ is positive; more specifically, we use $\hat{V}(\bar{d}) = (1/P_h)\cdot\{\hat{\gamma}_0 + 2\cdot\sum_{k=1}^{h-1}[1-(k/h)]\cdot\hat{\gamma}_k\}$, where $\hat{\gamma}_k = (1/P_h)\cdot\sum_{t=R+k}^{R+P-h}(d_{t+h}-\bar{d})\cdot(d_{t+h-k}-\bar{d})$.

$$\bar{d}_w = (1/P_h)\sum_{t=R}^{R+P-h} w(\omega_t)\cdot d_{t+h}\ , \tag{16}$$

for which they define a corresponding weighted Diebold and Mariano (1995) statistic,

$$W\text{-}DM = \bar{d}_w \cdot [\hat{V}(\bar{d}_w)]^{-0.5}\,, \tag{17}$$

where $\hat{V}(\bar{d}_w)$ can be estimated in an analogous way to $\hat{V}(\bar{d})$. Like the *DM* statistic, the *W-DM* statistic can also be treated as a standard normal variate asymptotically. 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 and Franses (2003),

$$w_T(\omega_t) = 1 - \varphi(q_t)/\max[\varphi(q_t)]\,, \tag{18}$$

where $\varphi(\cdot)$ is the density function of q_t. This weight function attaches greater weight to observations in both tails of the distribution of q_t.[20] We again follow Siliverstovs and van Dijk (2003) and use the Harvey, Newbold, and Leyboure (1997) correction factor to obtain the modified *W-DM* statistic,

$$MW\text{-}DM = \{[P_h + 1 - 2\cdot h + h\cdot(h-1)/P_h]/P_h\}\cdot W\text{-}DM, \tag{19}$$

where we again assess significance using the Student's *t* distribution with $P_h - 1$ degrees of freedom.

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 rates, 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 can be analyzed straightforwardly using contingency tables or matrices, where the observed number of outcomes is

---

[20] In our applications, we use a nonparametric kernel density procedure over the in-sample observations of q_t in order to estimate $\varphi(q_t)$ in equation (18).

compared to the expected number under the appropriate null hypothesis.

Consider an interval forecast for q_{t+1} given q_t and denote the lower and upper bounds of the interval forecast as $L_{t+1|t}(\theta)$ and $U_{t+1|t}(\theta)$, respectively, where θ is the given nominal coverage. We compare the actual q_{t+1} value ($t = R,\ldots,R+P-1$) to the interval forecast and obtain the sequence of indicator variables $\{I_{t+1|t}\}_{t=R}^{R+P-1}$, where $I_{t+1|t} = 1$ if $L_{t+1|t}(\theta) \le q_{t+1} \le U_{t+1|t}(\theta)$ and 0 otherwise. The first test evaluates unconditional coverage. Let $n_1 = \sum_{t=R}^{R+P-1} I_{t+1|t}$ (the number of "hits") and $n_0 = P - n_1$ (the number of "misses"). Under the null hypothesis of correct unconditional coverage, the expected number of hits equals $\theta \cdot P$, while the expected number of misses equals $(1-\theta)\cdot P$, so that the Pearson χ^2 statistic is given by

$$\chi^2_{UC} = \frac{[n_0 - (1-\theta)\cdot P]^2}{(1-\theta)\cdot P} + \frac{[n_1 - \theta \cdot P]^2}{\theta \cdot P}. \tag{20}$$

The test for independence is based on the contingency matrix of observed outcomes,

$$\begin{pmatrix} n_{00} & n_{01} \\ n_{10} & n_{11} \end{pmatrix}, \tag{21}$$

where n_{ij} equals the number of transitions of the indicator variable from $I_{t|t-1} = i$ to $I_{t+1|t} = j$ ($i, j = 0,1$). Under the null hypothesis of independence, the expected number of outcomes for n_{ij} equals $e_{ij}^{IND} = P \cdot [(n_{i0} + n_{i1})/P] \cdot [(n_{0j} + n_{1j})/P]$, so that the Pearson χ^2 statistic is given by

$$\chi^2_{IND} = \sum_{i=0}^{1} \sum_{j=0}^{1} (n_{ij} - e_{ij}^{IND})^2 / e_{ij}^{IND}. \tag{22}$$

Both the χ^2_{UC} and χ^2_{IND} statistics follow the $\chi^2(1)$ asymptotic distribution.

The test for conditional coverage combines the unconditional coverage and independence tests. We again use the matrix of observed outcomes given by equation (21). The matrix of expected outcomes under the null hypothesis of correct conditional coverage is given by

$$E^{CC} = \begin{pmatrix} (1-\theta)\cdot(n_{00}+n_{01}) & \theta\cdot(n_{00}+n_{01}) \\ (1-\theta)\cdot(n_{10}+n_{11}) & \theta\cdot(n_{10}+n_{11}) \end{pmatrix}, \tag{23}$$

so that the Pearson χ^2 statistic can be expressed as

$$\chi^2_{CC} = \sum_{i=0}^{1} \sum_{j=0}^{1} (n_{ij} - e^{CC}_{ij})^2 / e^{CC}_{ij} , \tag{24}$$

where e^{CC}_{ij} is the (i, j) element of the matrix E^{CC} given in equation (23). The χ^2_{CC} statistic follows the $\chi^2(2)$ asymptotic distribution.

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 and Patel (1998).[21] 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 and van Dijk (2003), who use the procedure suggested by Diebold, Gunther, and Tay (1998) based on Bonferroni bounds. More specifically, we divide the indicator variable series into *h* sub-groups of independent observations: $(I_{R+h|R}, I_{R+2\cdot h|R+h}, \ldots)$, $(I_{R+h+1|R+1}, I_{R+2\cdot h+1|R+h+1}, \ldots)$, $\ldots$, $(I_{R+h+(h-1)|R+(h-1)}, I_{R+2\cdot h+(h-1)|R+h+(h-1),} \ldots)$. 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. Of course, proceeding in this way can severely 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

---

[21] See Weerahandi (1995, Chapter 5) for a textbook treatment of the calculation of exact *p*-values for Pearson χ^2 statistics corresponding to contingency tables.

indicator variables available in each sub-group as *h* increases also helps to motivate the use of exact *p*-values for inference.

3.4. Evaluating Density Forecasts

Diebold, Gunther, and Tay (1998) provide a framework for evaluating density forecasts. Let $p_{t+1|t}(\cdot)$ be the probability density forecast for q_{t+1} given q_t generated by a given model and let $f_{t+1|t}(\cdot)$ be the actual or true predictive density. Next, define the probability integral transform (PIT) of the observed value of q_{t+1} with respect to the density forecast $p_{t+1|t}(\cdot)$ as

$$z_{t+1|t} = \int_{-\infty}^{q_{t+1}} p_{t+1|t}(u)du = P_{t+1|t}(q_{t+1}), \tag{25}$$

where $P_{t+1|t}(\cdot)$ is the cumulative density function corresponding to the probability density forecast $p_{t+1|t}(\cdot)$. Under the null hypothesis that $p_{t+1|t}(\cdot) = f_{t+1|t}(\cdot)$ (so that $p_{t+1|t}(\cdot)$ is the correct density forecast), Diebold, Gunther, and Tay (1998) show that $\{z_{t+1|t}\}_{t=R}^{R+P-1}$ is distributed *iid* $U(0,1)$. Following Clements and Smith (2000) and Siliverstovs and van Dijk (2003), we test for uniformity using the Kolmogorov-Smirnov (KS) statistic, which tests the significance of the largest absolute deviation of $P_{t+1|t}(q_{t+1})$ from the $U(0,1)$ cumulative distribution function.[22] Berkowitz (2001) recommends transforming the PIT series using the inverse of the standard normal cumulative density function, $z^*_{t+1|t} = \Phi^{-1}(z_{t+1|t})$, where $\Phi(\cdot)$ is the standard normal cumulative density function. Under the null hypothesis that $p_{t+1|t}(\cdot) = f_{t+1|t}(\cdot)$, $\{z^*_{t+1|t}\}_{t=R}^{R+P-1}$ is distributed *iid* $N(0,1)$. We again follow Clements and Smith (2000) and Siliverstovs and van Dijk (2003) and test the null hypothesis that $\{z^*_{t+1|t}\}_{t=R}^{R+P-1}$ is distributed as a standard normal variate

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

using the Doornik and Hansen (1994; DH) statistic.[23] Note that the KS and DH statistics assume independence. In order to explicitly test for independence in the PITs, Diebold, Gunther, and Tay (1998) recommend looking for autocorrelation in the power-transformed PIT series, $\{(z_{t+1|t} - \bar{z}_{t+1|t})^k\}_{t=R}^{R+P-1}$, for $k = 1,\ldots,4$, where $\bar{z}_{t+1|t} = (1/P)\sum_{t=R}^{R+P-1} z_{t+1|t}$. Following Siliverstovs and 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 the h sub-groups defined by $(z_{R+h|R}, z_{R+2\cdot h|R+h}, \ldots)$, $(z_{R+h+1|R+1}, z_{R+2\cdot h+1|R+h+1}, \ldots)$, $\ldots$, $(z_{R+h+(h-1)|R+(h-1)}, z_{R+2\cdot h+(h-1)|R+h+(h-1),} \ldots)$. 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.

4. Empirical Results

4.1. Point Forecasts

Table 1 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.[24] Column 2 reports the ratio of the linear AR model root MSFE (RMSFE) to the RMSFE for a random walk (with drift) model, column 3 reports the ratio of the Band-TAR model RMSFE to the random walk model RMSFE, and column 4 reports the ratio of the Band-TAR model RMSFE to the linear AR model RMSFE. We see from column 4 of Table 1 that 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.

---

[23] The DH statistic follows the $\chi^2(2)$ asymptotic distribution, and we base inference on critical values corresponding to this distribution.

[24] 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. Also note that coefficient estimates for all of the linear AR models are available at http://pages.slu.edu/faculty/rapachde/Research.htm.

At longer horizons, the Band-TAR model RMSFE is actually 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 often sizable. Not surprisingly, we cannot reject the null hypothesis that the Band-TAR model MSFE is 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. For Japan, the relative RMSFE declines as the horizon increases, and the Band-TAR model RMSFE is 12% 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 (see column 5), the Band-TAR model MSFE is significantly less than the linear AR model MSFE for Japan at horizons of 6-24 months. 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 a relative MSFE metric and some support for the Band-TAR model specification for Japan according to an MSFE metric.

There is reason to be cautious about basing inferences on the Student's *t* distribution for the *M-DM* statistic in Table 1. When $h = 1$, McCracken (2004) shows that the *DM* 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 *DM* 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. While, strictly speaking, we are not comparing two nested linear models, as *c* becomes large in the OT Band-TAR model specification, equation (1), and the AR coefficient approaches unity in the linear AR(1) model, we approach the situation where we have two nested linear models. A similar situation holds for the TPS ESTAR model specification, equation (6), as α approaches zero. Given the potential relevance of comparing nearly nested models, we augment our analysis with a parametric bootstrap procedure in order to generate critical values for the *M-DM* statistics. Under the null hypothesis of equal predictive ability, we generate pseudo-data using a linear AR Gaussian process fitted to the original data. Parametric bootstrapped critical values are reported in columns 6-8 of Table 1. 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 become insignificant at every horizon for Japan.

Column 9 of Table 1 reports *MW-DM* statistics, which place greater weight on forecasting real exchange rate values farther out in the tails of the unconditional distribution. Column 9 also gives the *p*-values for the *MW-DM* statistics based on the Student's *t* distribution in brackets, and we report parametric bootstrapped critical values in columns 10-12. 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 critical 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, 2, 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 short horizons. However, none of the *MW-DM* statistics is significant at any horizon according to the bootstrapped critical values. Turning to Japan, the *M-DM* statistic is significant at horizons of 6-24 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 the weighted MSFE criterion. However, as with the U.K., none of the *MW-DM* statistics is significant at any horizon according to the bootstrapped critical values. There does not appear to be reliable evidence that the OT Band-TAR models produce significantly better forecasts in terms of an MSFE criterion weighted toward forecasting extreme real exchange rate values for any of the countries.

Table 2 presents the out-of-sample point forecast evaluation results for the TPS ESTAR models 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 critical values are used). For Germany, the ESTAR model RMSFE is less than the linear AR model RMSFE at all horizons. Furthermore, the *M-DM* statistic is significant at horizons of 18-24 months, while the *MW-DM* statistic is significant at horizons of 15-24 months, when we use the *p*-values based on the Student's *t* distribution. However, neither the *M-DM* nor the *MW-DM* statistic is significant at any horizon when we use the bootstrapped critical values. Turning to France, the ESTAR model RMSFE is less than the linear AR model RMSFE at all horizons, and the *M-DM* (*MW-DM*) statistic is significant at horizons of 1 (1 and 2) and 12-24 (6-24) months. However, similar to the situation for Germany, none of the *M-DM* or *MW-DM* statistics is significant according to the bootstrapped critical 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. While the *MW-DM* statistic is significant horizons of 21 and 24 months according to *p*-values based on the Student's *t* distribution, the statistic is insignificant at both horizons according to the bootstrapped critical values. Overall, there is some evidence in Tables 1 and 2 that the OT Band-TAR and TPS ESTAR models offer forecasting gains at some horizons using monthly data relative to simple linear AR models for some countries. However, this evidence is not particularly robust, as there is no evidence of forecasting gains when we base inferences for the *M-DM* and *MW-DM* statistics on bootstrapped critical values.

In Table 3, we report results for the MNP ESTAR model and a linear AR(2) model. We only consider a horizon of 1 year, as we only have 12 annual out-of-sample observations (1991-2002). Despite the limited number of out-of-sample observations, the *M-DM* statistic is significant at the 10% level according to the bootstrapped critical values (although it is not significant according to the *p*-value based on the Student's *t* distribution). In addition, the *MW-DM* statistic is significant at the 5% level using both the *p*-value based on the Student's *t* distribution and the bootstrapped critical values. According to these tests, the MNP ESTAR model does offer significant forecasting gains for annual real exchange rate observations from 1991-2002, especially with regard to more extreme real exchange rate observations.[25]

[25] We also compared forecasts generated by the MNP ESTAR model and a linear AR(2) for the 1981-2002 out-of-sample period, and the *M-DM* and *MW-DM* statistics (1.34 and 2.28, respectively) were both significant at the 10% level using bootstrapped critical values. In order to compare point forecasts in terms of the predicted direction of change of the real exchange rate, we also employed the Henrikkson and Merton (1981), Cumby and Modest (1987), and Pesaran and 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. There is some significant evidence of market timing ability for the MNP ESTAR model, while there is no significant evidence of market timing ability for the linear AR(2) counterpart.

4.2. Interval Forecasts

Pearson χ^2 statistics used to evaluate interval forecasts for the OT Band-TAR models 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 U.K. linear AR model, correct unconditional converge is rejected for the linear AR model at all three forecast horizons (column 4), independence is rejected at horizons of 2 and 3 months (column 5), and correct conditional coverage is rejected at all three forecast horizons (column 6). With respect to the U.K. Band-TAR model, correct unconditional coverage, independence, and correct conditional coverage are all rejected at the 1-month horizon, but none of these null hypotheses is rejected at horizons of 2 and 3 months. This suggests that interval forecasts generated by the U.K. Band-TAR model are more accurate at horizons beyond 1 month than forecasts generated by the U.K. linear AR model. Focusing on the χ^2_{CC} statistics, a similar situation obtains for Germany at all reported horizons. That is, we reject correct conditional coverage for the linear AR model at horizons of 1-3 months, but we cannot reject correct conditional coverage for the Band-TAR model at the same horizons. For France, correct conditional coverage is rejected at horizons of 1 and 2 months for the linear AR model and at a horizon of 1 month for the Band-TAR model. With respect to Japan, the correct conditional coverage cannot be rejected at any reported horizon for the linear AR model, while it is rejected at all three horizons for the Band-TAR model. Thus, the linear AR model appears to provide more accurate interval forecasts than the Band-TAR model for Japan.

We evaluate inter-quartile interval forecasts for the TPS ESTAR models 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 conditional coverage is rejected at all reported horizons for both models. For Germany, test results for correct conditional coverage are the same for both the ESTAR and linear AR models: correct conditional coverage cannot be rejected at horizons of 1 and 3 months but is rejected at the 2-month horizon. For France, correct conditional coverage cannot be rejected for either model at horizons of 1 and

3 months; at the 2-month horizon, it is rejected for the linear AR model but not the ESTAR model. Both the ESTAR and linear AR 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, there is some evidence supporting the OT Band-TAR specification over a linear AR specification for the U.K., Germany, and France (Table 4) and evidence supporting the TPS ESTAR specification over a linear AR specification for France (Table 5).[26] The results actually support a linear AR specification over the OT Band-TAR specification for Japan (Table 4), while the results do not favor either the TPS ESTAR or linear AR specification for the U.K., Germany, and Japan (Table 5). We also calculated χ^2_{CC} statistics for the MNP ESTAR model and a linear AR(2) counterpart over the 1991-2002 out-of-sample period at the 1-year horizon. We could not reject correct conditional coverage for either model ($\chi^2_{CC} = 0.87$ with a *p*-value of 0.80 for each model).

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 $k = 2$ (column 7) at all reported horizons for both the OT Band-TAR and linear AR models for all four countries. This points to deficiencies in the density forecasts for both the OT Band-TAR and linear AR model specifications. There are also frequent rejections for both models and all countries when $k = 4$ (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 null hypothesis of uniformity is rejected using the KS statistic at horizons of 2 and 3 months for the linear AR model but not the Band-TAR model. In contrast, for Germany and France, the KS statistic rejects uniformity for the Band-TAR model

[26] The better performance of the Band-TAR models for the U.K., Germany, and France in terms of interval forecasts may be primarily explained by the difference in the variance of the disturbance term across regimes in the Band-TAR specification.

at one or more horizons, while uniformity is not rejected at any reported horizon for the linear AR model, supporting the linear AR model over the Band-TAR model. For France, the DH statistic is significant at the 3-month horizon for the Band-TAR model but not the linear AR model. For Japan, the KS and DH statistics reject at two or more horizons for both the linear AR and Band-TAR models.

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. 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 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.

We also calculated the KS, DH, and LB ($k = 1, \ldots, 4$) statistics for the MNP ESTAR model and a linear AR(2) model over the 1991-2002 out-of-sample period at the 1-year horizon. The KS, DH, and LB ($k = 1, \ldots, 4$) statistics were all insignificant for both the MNP ESTAR and linear AR models.

5. Comparing In-Sample Conditional Densities

Looking back to Tables 1 and 2, 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 and 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

---

[27] In a footnote (p. 318), they add, "In other words, *significance* of nonlinearity does not necessarily imply its *economic importance*" [emphasis in original].

relevance of this for the OT Band-TAR and TPS ESTAR models, we follow the suggestion of Pagan (2002) and Breunig, Najarian, and 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 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 1 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 1, 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 appear "very slight." In Figure 2, 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 the comparisons of conditional expectation functions in Figures 1 and 2, the lack of forecasting gains at short horizons provided by fitted OT Band-TAR and TPS ESTAR models appears to result from the absence of "strong" nonlinearities in the conditional means of these nonlinear AR models. In Figure 3, we graph the conditional expectation functions of q_t given q_{t-1} corresponding to the fitted MNP ESTAR model and a linear AR counterpart.[28] Differences between the conditional expectation functions corresponding to the fitted MNP ESTAR and linear AR models are somewhat more apparent in Figure 3 than in Figures 1 and 2, and Table 3 shows that the MNP ESTAR model offers significant forecasting gains relative to a linear AR model.

In order to compare the fitted nonlinear and linear AR models more formally and extensively, we use the recently developed test of Corradi and Swanson (2003). This test allows us to compare the

---

[28] Given that the MNP ESTAR includes two lags of q_t in the conditional expectation function, we have to integrate out q_{t-2} in order to graph the conditional expectation function of q_t given q_{t-1} corresponding to this model. We follow Pagan (2002) and use a nonparametric procedure. More specifically, we simulate 30,000 observations for q_t using the fitted MNP ESTAR model and then apply a nonparametric estimator to the simulated observations in order to evaluate the conditional expectation function at the in-sample data points for q_{t-1}.

conditional densities for q_t given q_{t-1} corresponding to two different fitted models, each of which is possibly misspecified. Let $P^N_{t+1|t}(u|\Omega^*_N)$ denote the conditional cumulative density function for q_t given q_{t-1} corresponding to a given nonlinear AR model and let $P^L_{t+1|t}(u|\Omega^*_L)$ denote the conditional cumulative density function corresponding to a linear AR benchmark model, where Ω^*_N (Ω^*_L) is the probability limit of the quasi maximum likelihood estimator of the vector of parameters of the nonlinear (linear) AR model. The null hypothesis is that the conditional densities corresponding to the nonlinear and linear AR models are equally accurate relative to the true conditional density. Corradi and Swanson (2003) employ a distributional analog of the mean squared error metric, and the null hypothesis can be expressed as

$$H_0: \int_U E\{[P^L_{t+1|t}(u|\Omega^*_L) - F_{t+1|t}(u|\Omega_0)]^2 - [P^N_{t+1|t}(u|\Omega^*_N) - F_{t+1|t}(u|\Omega_0)]^2\}\varphi(u)du = 0, \tag{26}$$

where $F_{t+1|t}(u|\Omega_0)$ is the cumulative density function corresponding to the true conditional probability density function $f_{t+1|t}(\cdot)$, $\int \varphi(u)du = 1$, and $\varphi(u) \geq 0$ for all $u \in U \subset \Re$. The alternative hypothesis is that the conditional density corresponding to the nonlinear AR model is more accurate than the conditional density corresponding to the linear AR benchmark model:

$$H_1: \int_U E\{[P^L_{t+1|t}(u|\Omega^*_L) - F_{t+1|t}(u|\Omega_0)]^2 - [P^N_{t+1|t}(u|\Omega^*_N) - F_{t+1|t}(u|\Omega_0)]^2\}\varphi(u)du > 0. \tag{27}$$

The test statistic takes the form,

$$Z_T = \int_U Z_{T,u}\varphi(u)du, \tag{28}$$

where

$$Z_{T,u} = \frac{1}{\sqrt{R}}\sum_{t=0}^{R-1}\{[1(q_{t+1} \leq u) - P^L_{t+1|t}(u|\hat{\Omega}_L)]^2 - [1(q_{t+1} \leq u) - P^N_{t+1|t}(u|\hat{\Omega}_N)]^2\} \tag{29}$$

and $\hat{\Omega}_N$ ($\hat{\Omega}_L$) is the nonlinear least squares (linear least squares) estimator of the vector of parameters of the nonlinear (linear) AR model. In order to make the test statistic in equation (28) operational, Corradi and Swanson (2003) use the following expression:

$$Z_T = \frac{1}{S}\sum_{i=1}^{S}\left(\frac{1}{\sqrt{R}}\sum_{t=0}^{R-1}\{[1(q_{t+1} \le u_i) - P^{L}_{t+1|t}(u_i \mid \hat{\Omega}_L)]^2 - [1(q_{t+1} \le u_i) - P^{N}_{t+1|t}(u_i \mid \hat{\Omega}_N)]^2\}\right), \tag{30}$$

where S is set equal to a large number. In our applications, we integrate over a grid of u values whose lower and upper limits correspond to the minimum and maximum values of the in-sample q_t observations. We also define a test statistic $R\text{-}Z_T$ that integrates over two grids of u values whose limits correspond to the minimum and maximum value 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.[29]

Corradi and Swanson (2003) test results for the fitted OT Band-TAR, TPS ESTAR, and MNP ESTAR models and linear AR benchmark models are reported in Table 8. Following the recommendation of Corradi and Swanson (2003), we base inferences on block bootstrapped critical values.[30] According to the Z_T statistics in column 3, the null hypothesis of equal conditional density accuracy cannot be rejected for any of the 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 for the OT Band-TAR model for France, TPS ESTAR model for Germany, and MNP ESTAR model for the U.K. The significant $R\text{-}Z_T$ statistic for the MNP ESTAR model in Table 8 is consistent with the significant *MW-DM* statistic in Table 3 and indicates the relevance of the MNP ESTAR model. The significant $R\text{-}Z_T$ statistics in Table 8 for the OT Band-TAR model for France and TPS ESTAR model for Germany suggest that these nonlinear AR models provide better approximations to the true conditional densities than linear AR benchmark models for more extreme q_t

---

[29] In principle, we could apply the Corradi and Swanson (2003) test to our out-of-sample density forecasts. However, Monte Carlo simulations in Corradi and Swanson (2004) suggest that a very large number of forecasts are needed for reliable inference.

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

observations. There is no evidence of statistically significant differences in the conditional densities corresponding to the other six nonlinear AR models and their linear AR counterparts. 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 4.1 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 literature. Overall, we find fairly limited evidence that favors the OT Band-TAR and TPS ESTAR model specifications over simple linear AR benchmark models using monthly data from the post-Bretton Woods period. 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, and we find no robust evidence that the point forecasts generated by these nonlinear models are statistically superior to forecasts generated by linear AR models according to an MSFE criterion (weighted or unweighted). Our analysis of the in-sample conditional expectation functions and conditional densities corresponding to the fitted OT Band-TAR and TPS ESTAR models reveals that these nonlinear models are typically not very different from linear AR benchmark models. There is evidence that interval forecasts generated by the OT Band-TAR and TPS ESTAR models have better coverage than forecast intervals generated by linear AR models for some countries. Evaluation of density forecasts 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 specification. 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. There is more in-sample and out-of-sample support for the MNP ESTAR model specification over a linear AR specification using a long span of annual data.

Finally, we suggest some avenues for future research. Given the support we find for the MNP ESTAR model, which is estimated using over two centuries of data for the U.K.-U.S. real exchange rate, further research using long spans of real exchange rate data may be useful in assessing the relevance of nonlinear real exchange rate behavior, as the post-Bretton Woods period may not be long enough for "strong" evidence of nonlinear real exchange rate behavior to become evident.[31] In addition, Sarno, Taylor, and Chowdhury (2002) and Imbs, et al. (2003) recently report in-sample evidence of nonlinear behavior in real exchange rates based on more disaggregated price indices.[32] 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, it would also be useful to explore model specification issues more extensively by considering, for example, different types of nonlinear behavior, non-normally distributed disturbance terms, and conditional heteroskedasticity in the disturbances.[33] Along these lines, mutlivariate nonlinear model specifications may also prove helpful, as the theoretical models in Goswami, Shrikhande, and 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.

[31] Of course, a drawback to the use of long spans of data is that it potentially combines data from different regimes (Mussa, 1986).
[32] Also see O'Connell and Wei (2002), who find that price discrepancies for individual goods between U.S. cities are nonlinearly mean-reverting to parity.
[33] We do not pursue these extensions in the present paper, as we focus on the nonlinearities in the conditional means of real exchange rates emphasized in the existing studies of OT, TPS, and MNP. Nevertheless, these extensions are potentially important areas for future research.

References

Balke, N.S. and T.B. Fomby, 1997. Threshold Cointegration. *International Economic Review* 38, 627-645.

Benninga, S. and A.A. Protopapadakis, 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. and R.J. Caballero, 1990. Kinked Adjustment Costs and Aggregate Dynamics. In O.J. Blanchard and S. Fischer, eds., *NBER Macroeconomics Annual* (Cambridge, Mass.: MIT Press).

Breunig, R., S. Najarian, and A. Pagan, 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. and M.W. McCracken, 2004. Evaluating Long-Horizon Forecasts. Manuscript, University of Missouri-Columbia.

Clements, M.P. and J. Smith, 1997. The Performance of Alternative Forecasting Methods for SETAR Models. *International Journal of Forecasting* 13, 463-475.

Clements, M.P. and J. Smith, 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. and N.R. Swanson, 2003. A Test for Comparing Multiple Misspecified Conditional Distributions. Manuscript, Rutgers University.

Corradi, V. and N.R. Swanson, 2004. Predictive Density Accuracy Tests. Manuscript, Rutgers University.

Cumby, R.E. and D.M. Modest, 1987. Testing for Market Timing Ability. *Journal of Financial Economics* 19, 169-189.

De Gooijer, J.G. and K. Kumar, 1992. Some Recent Developments in Non-Linear Time Series Modelling, Testing, and Forecasting. *International Journal of Forecasting* 8, 135-156.

Diebold, F.X., T. Gunther, and A. Tay, 1998. Evaluating Density Forecasts With Applications to Financial Risk Management. *International Economic Review* 39, 863-883.

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

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

Doornik, J.A. and H. Hansen, 1994. An Omnibus Test for Univariate and Multivariate Normality. Manuscript, Nuffield College, Oxford.

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

Eitrheim O., and T. Teräsvirta, 1996. Testing the Adequacy of Smooth Transition Autoregressive Models. *Journal of Econometrics* 74, 59-75.

Gallant, A.R., P.E. Ross, and G. Tauchen, 1993. Nonlinear Dynamic Structures. *Econometrica* 61, 871-908.

Goswami, G., M. Shrikhande, and L. Wu, 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. and T. Teräsvirta, 1993. *Modelling Nonlinear Economic Relationships* (Oxford: Oxford University Press).

Hansen, B.E., 2003. Interval Forecasts and Parameter Uncertainty. Manuscript, University of Wisconsin-Madison.

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

Henriksson, R.D. and R.C. Merton , 1981. On Market Timing and Investment Performance. II. Statistical Procedures for Evaluating Forecasting Skills. *Journal of Business* 54, 513-533.

Imbs, J., H. Mumtaz, M.O. Ravn, and Helene Rey, 2003. Non-Linearities and Real Exchange Rate Dynamics. Manuscript, London Business School.

Lothian, J.R. and M.P. Taylor, 1996. Real Exchange Rate Behavior: The Recent Float from the Perspective of the Last Two Centuries. *Journal of Political Economy* 104, 488-510.

Mark, N.C., 2001. *International Macroeconomics and Finance* (Oxford: Blackwell Publishers).

McCracken, M.W., 2004. Asymptotics for Out-of-Sample Tests of Granger Causality. Manuscript, University of Missouri-Columbia.

Mehta, C.R. and N.R. Patel, 1998. Exact Inference for Categorical Data. In P. Armitage and T. Colton, eds., *Encyclopedia of Biostatistics* (Chichester: John Wiley), pp. 1411-1422.

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* 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., V. Zarnowitz, R.S. Tsay, and G.C. Tiao, 1998. Forecasting the U.S. Unemployment Rate. *Journal of the American Statistical Association* 93, 478-493.

Mussa, Michael, 1986. Nominal Exchange Rate Regimes and the Behavior of Real Exchange Rates: Evidence and Implications. *Carnegie Rochester Conference Series on Public Policy* 25, 117-213.

Newey, W. and K.D. West, 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 Bernanke, B. and K. Rogoff, eds., *National Bureau of Economic Research Macroeconomics Annual 2000* (Cambridge, Mass.: NBER and MIT Press), pp. 339-390.

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* 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. and S.-J. Wei, 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. and A.C. Stockman, 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. and S.M. Potter, 1997. A Floor and Ceiling Model of U.S. Output. *Journal of Economic Dynamics and Control* 21, 661-695.

Pesaran, M.H. and A.G. Timmermann, 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, M.P. Taylor, and I. Chowdhury, 2002. Nonlinear Dynamics in Deviations from the Law of One Price. *Journal of International Money and Finance*, forthcoming.

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* 50, 1309-1319.

Siliverstovs, B. and D. van Dijk, 2003. Forecasting Industrial Production with Linear, Nonlinear, and Structural Change Models. Econometric Institute Report EI 2003-16.

Stock, J.H. and M.W. Watson, 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., 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* 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. and P.H. Franses, 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. and M.W. McCracken, 1998. Regression-Based Tests of Predictive Ability. *International Economic Review* 39, 817-840.

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

Weerahandi, S., 1995. *Exact Statistical Methods for Data Analysis* (New York: Springer-Verlag).

Table 1: Out-of-sample point forecast evaluation, linear AR and Obstfeld and Taylor (1997) Band-TAR models

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
					Parametric bootstrapped *M-DM* critical values				Parametric bootstrapped *MW-DM* critical values		
h[a]	AR/RW[b]	Band-TAR/RW[c]	Band-TAR/AR[d]	*M-DM*[e]	10%	5%	1%	*MW-DM*[f]	10%	5%	1%
United Kingdom, 1994:12-2003:07 out-of-sample period											
1	0.99	1.00	1.00	-0.36 [0.64]	1.41	1.82	3.11	**1.61** [0.05]	2.81	3.33	4.26
2	0.99	0.99	1.00	-0.07 [0.53]	1.87	2.51	4.09	**1.52** [0.07]	2.88	3.24	4.19
3	0.99	0.99	1.00	-0.12 [0.54]	1.96	2.50	4.11	**1.58** [0.06]	2.83	3.32	4.40
6	0.97	0.98	1.01	-0.20 [0.58]	1.82	3.33	3.82	**1.58** [0.06]	2.67	3.06	3.90
9	0.94	0.98	1.04	-0.47 [0.68]	1.82	2.20	3.90	1.06 [0.15]	2.42	2.99	3.93
12	0.92	0.98	1.07	-0.63 [0.73]	1.84	2.20	3.86	0.59 [0.28]	2.45	2.86	3.99
15	0.89	0.98	1.09	-0.74 [0.77]	1.87	2.46	3.89	0.26 [0.40]	2.39	2.96	3.95
18	0.88	0.98	1.11	-0.69 [0.75]	1.90	2.67	4.30	0.15 [0.44]	2.39	3.03	4.10
21	0.87	0.98	1.13	-0.72 [0.76]	2.11	2.80	4.44	0.07 [0.47]	2.64	3.29	4.47
24	0.85	0.98	1.16	-0.87 [0.81]	2.29	3.14	4.66	-0.17 [0.57]	2.78	3.37	4.82
Germany, 1995:01-2003:07 out-of-sample period											
1	0.99	1.00	1.00	-0.31 [0.62]	1.31	1.77	3.02	1.03 [0.15]	2.18	3.02	4.04
2	0.99	0.99	1.00	-0.22 [0.59]	1.66	2.43	4.78	0.77 [0.22]	2.55	3.03	4.79
3	0.97	0.99	1.01	-0.45 [0.67]	1.77	2.73	4.60	0.70 [0.24]	2.66	3.26	4.90
6	0.93	0.99	1.06	-1.05 [0.85]	1.85	2.84	4.58	0.17 [0.43]	2.67	3.18	4.81
9	0.89	1.00	1.11	-1.31 [0.90]	1.83	2.62	4.55	-0.28 [0.61]	2.63	3.16	4.74
12	0.86	1.00	1.16	-1.51 [0.93]	1.88	2.95	5.05	-0.61 [0.73]	2.48	3.27	4.62
15	0.83	1.01	1.21	-1.68 [0.95]	1.98	2.87	5.28	-0.86 [0.80]	2.59	3.22	4.84
18	0.82	1.01	1.24	-1.96 [0.97]	2.14	2.99	5.66	-1.13 [0.87]	2.71	3.36	5.20
21	0.80	1.02	1.27	-2.56 [0.99]	2.46	3.38	5.95	-1.84 [0.97]	3.01	3.66	5.94
24	0.79	1.03	1.31	-3.04 [1.00]	2.77	3.80	6.83	-2.60 [0.99]	3.38	4.32	6.51

Notes: *p*-value using the *t* distribution with $P_h - 1$ degrees of freedom is reported in brackets; bold statistic indicates significance at the 10% level according to the *p*-value; bold bootstrapped critical value indicates that the statistic is significant according to the bootstrapped critical value; 0.00 indicates <0.005.

[a]Forecast horizon (months-ahead).

[b]Ratio of the linear AR model RMSFE to the random walk model RMSFE.

[c]Ratio of the Band-TAR model RMSFE to the random walk model RMSFE.

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

[e]Modified Diebold and 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.

[f]Modified weighted Diebold and 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 1 (continued)

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
					Parametric bootstrapped *M-DM* critical values				Parametric bootstrapped *MW-DM* critical values		
h[a]	AR/RW[b]	Band-TAR/RW[c]	Band-TAR/AR[d]	*M-DM*[e]	10%	5%	1%	*MW-DM*[f]	10%	5%	1%
France, 1994:12-2003:06 out-of-sample period											
1	1.00	1.00	1.00	-0.23[0.59]	1.14	1.56	3.21	0.67 [0.25]	2.06	2.73	4.01
2	0.99	0.99	1.00	0.01 [0.50]	1.66	2.26	4.00	0.71 [0.24]	2.35	2.95	4.69
3	0.99	0.99	1.01	-0.18 [0.57]	1.78	2.42	4.85	0.45 [0.33]	2.46	2.99	5.07
6	0.96	1.00	1.03	-0.53 [0.70]	1.77	2.58	4.52	0.24 [0.40]	2.42	2.90	4.70
9	0.94	1.01	1.07	-0.75 [0.77]	1.75	2.56	4.30	-0.12 [0.55]	2.44	3.09	4.22
12	0.92	1.01	1.11	-0.95 [0.83]	1.79	2.65	4.63	-0.38 [0.65]	2.44	3.11	4.32
15	0.90	1.02	1.14	-1.13 [0.87]	1.80	2.82	4.53	-0.55 [0.71]	2.54	3.00	4.48
18	0.89	1.04	1.16	-1.36 [0.91]	1.91	2.98	5.06	-0.75 [0.77]	2.71	3.18	4.65
21	0.88	1.05	1.19	-1.79 [0.96]	2.15	3.13	5.59	-1.17 [0.88]	2.78	3.64	5.58
24	0.87	1.06	1.22	-2.27 [0.99]	2.37	3.73	5.85	-1.68 [0.95]	3.05	3.95	5.97
Japan, 1994:12-2003:06 out-of-sample period											
1	0.99	0.98	0.99	0.76 [0.22]	1.20	1.68	2.67	0.28 [0.39]	2.08	2.53	3.62
2	0.99	0.97	0.98	0.81 [0.21]	2.30	3.26	5.03	0.46 [0.32]	2.91	3.59	5.11
3	0.98	0.96	0.98	0.84 [0.20]	2.44	3.48	5.11	0.62 [0.27]	3.07	3.78	5.37
6	0.97	0.92	0.96	**1.35** [0.09]	2.57	3.29	5.16	**1.69** [0.05]	2.95	3.57	5.33
9	0.94	0.87	0.93	**1.72** [0.04]	2.52	3.31	5.12	**2.33** [0.01]	2.92	3.50	5.16
12	0.93	0.85	0.92	**1.64** [0.05]	2.52	3.44	4.98	**2.20** [0.01]	2.88	3.47	5.67
15	0.92	0.84	0.91	**1.62** [0.05]	2.67	3.65	5.73	**1.99** [0.02]	2.96	3.52	6.34
18	0.91	0.82	0.90	**1.75** [0.04]	2.84	3.81	6.96	**2.23** [0.01]	3.07	3.77	6.97
21	0.91	0.81	0.90	**1.91** [0.03]	3.07	4.29	7.94	**2.49** [0.01]	3.25	4.26	7.75
24	0.90	0.79	0.88	**2.21** [0.01]	3.44	4.68	9.32	**2.65** [0.00]	3.47	4.47	8.87

Table 2: Out-of-sample point forecast evaluation, linear AR and Taylor, Peel, and Sarno (2001) ESTAR models

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
					Parametric bootstrapped *M-DM* critical values				Parametric bootstrapped *MW-DM* critical values		
h[a]	AR/RW[b]	ESTAR/ RW[c]	ESTAR/ AR[d]	*M-DM*[e]	10%	5%	1%	*MW-DM*[f]	10%	5%	1%
United Kingdom, 1997:01-2003:07 out-of-sample period											
1	0.98	0.99	1.01	-1.04 [0.85]	1.35	1.76	2.91	0.89 [0.19]	2.36	2.67	3.39
2	0.98	0.99	1.01	-0.56 [0.71]	1.16	1.52	2.57	1.03 [0.15]	1.97	2.26	2.82
3	0.97	0.99	1.02	-0.57 [0.72]	1.24	1.73	2.71	1.08 [0.14]	2.08	2.37	3.10
6	0.94	0.98	1.04	-0.84 [0.80]	1.45	1.89	2.86	1.09 [0.14]	2.19	2.53	3.57
9	0.90	0.96	1.07	-1.16 [0.87]	1.67	2.07	3.13	0.24 [0.41]	2.26	2.55	3.71
12	0.85	0.92	1.08	-1.42 [0.92]	1.89	2.39	3.70	-0.42 [0.66]	2.26	2.80	3.77
15	0.81	0.88	1.09	-1.47 [0.93]	2.07	2.62	4.09	-0.66 [0.75]	2.49	3.16	4.09
18	0.76	0.85	1.11	-1.99 [0.97]	2.41	3.20	4.60	-1.16 [0.88]	2.71	3.48	4.72
21	0.72	0.79	1.10	-2.02 [0.98]	2.86	3.83	5.57	-1.15 [0.87]	3.05	3.93	5.83
24	0.67	0.75	1.12	-2.70 [1.00]	3.38	4.59	7.55	-1.68 [0.95]	3.53	4.40	6.04
Germany, 1997:01-2003:07 out-of-sample period											
1	1.00	0.99	0.99	0.83 [0.20]	1.67	2.16	3.13	1.05 [0.15]	2.35	2.91	3.51
2	1.00	0.98	0.98	0.57 [0.28]	1.49	1.85	2.62	0.84 [0.20]	2.03	2.28	3.02
3	1.01	0.97	0.97	0.71 [0.24]	1.57	2.03	2.98	0.93 [0.18]	2.16	2.43	3.35
6	1.02	0.96	0.94	0.76 [0.23]	1.82	2.44	3.30	1.04 [0.15]	2.25	2.65	3.89
9	1.03	0.94	0.92	0.80 [0.21]	2.01	2.81	3.70	1.03 [0.15]	2.38	2.84	4.43
12	1.02	0.92	0.90	0.97 [0.17]	2.19	3.20	5.14	1.17 [0.12]	2.52	3.09	4.80
15	1.02	0.90	0.88	1.08 [0.14]	2.69	3.69	6.16	**1.29** [0.10]	2.80	3.42	5.42
18	1.04	0.89	0.86	**1.30** [0.10]	3.10	4.39	7.08	**1.49** [0.07]	2.97	3.91	6.18
21	1.05	0.88	0.84	**1.62** [0.05]	3.52	5.01	7.79	**1.81** [0.04]	3.29	4.28	6.77
24	1.04	0.86	0.82	**2.07** [0.02]	4.12	6.17	10.04	**2.25** [0.01]	3.73	5.25	8.48

Notes: *p*-value using the *t* distribution with $P_h - 1$ degrees of freedom is reported in brackets; bold statistic indicates significance at the 10% level according to the *p*-value; bold bootstrapped critical value indicates that the statistic is significant according to the bootstrapped critical value; 0.00 indicates <0.005.

[a]Forecast horizon (months-ahead).

[b]Ratio of the linear AR model RMSFE to the random walk model RMSFE.

[c]Ratio of the ESTAR model RMSFE to the random walk model RMSFE.

[d]Ratio of the ESTAR model RMSFE to the linear AR model RMSFE.

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

[f]Modified weighted Diebold and 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 ESTAR model weighted MSFE.

Table 2 (continued)

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
					Parametric bootstrapped *M-DM* critical values				Parametric bootstrapped *MW-DM* critical values		
h[a]	AR/RW[b]	ESTAR/ RW[c]	ESTAR/ AR[d]	*M-DM*[e]	10%	5%	1%	*MW-DM*[f]	10%	5%	1%
France, 1997:01-2003:06 out-of-sample period											
1	1.01	0.99	0.98	**1.27** [0.10]	1.56	2.03	3.34	**1.64** [0.05]	2.34	2.56	3.39
2	1.01	0.98	0.97	1.14 [0.13]	1.36	1.70	2.85	**1.45** [0.08]	1.90	2.18	2.92
3	1.01	0.98	0.97	0.91 [0.18]	1.50	1.91	2.85	1.22 [0.11]	2.05	3.39	3.18
6	1.03	0.96	0.93	1.18 [0.12]	1.82	2.53	3.71	**1.44** [0.08]	2.21	2.77	3.60
9	1.05	0.95	0.90	1.16 [0.12]	2.08	2.69	4.40	**1.40** [0.08]	2.33	3.05	3.99
12	1.04	0.92	0.88	**1.52** [0.07]	2.37	3.39	5.24	**1.67** [0.05]	2.47	3.02	4.56
15	1.05	0.91	0.87	**1.65** [0.05]	2.76	4.00	6.08	**1.81** [0.04]	2.73	3.44	5.26
18	1.07	0.91	0.85	**1.84** [0.04]	3.32	4.82	7.48	**2.00** [0.03]	3.29	4.01	6.35
21	1.08	0.91	0.84	**2.32** [0.01]	3.77	5.48	8.83	**2.45** [0.01]	3.52	4.82	6.82
24	1.08	0.89	0.83	**2.83** [0.00]	4.65	6.58	11.72	**2.99** [0.00]	3.98	5.56	9.09
Japan, 1997:01-2003:06 out-of-sample period											
1	0.99	0.99	1.00	-0.68 [0.75]	1.67	2.27	3.34	0.63 [0.26]	2.34	2.85	3.62
2	0.98	0.99	1.01	-0.45 [0.67]	1.51	2.00	3.02	0.61 [0.27]	1.98	2.25	3.41
3	0.97	0.98	1.01	-0.76 [0.77]	1.71	2.35	3.30	0.49 [0.31]	2.13	2.52	3.52
6	0.94	0.96	1.02	-0.86 [0.80]	2.07	2.73	3.73	0.32 [0.37]	2.43	2.86	3.85
9	0.91	0.92	1.01	-0.41 [0.66]	2.24	3.19	4.87	0.59 [0.28]	2.55	3.17	4.44
12	0.88	0.90	1.02	-0.56 [0.71]	2.51	3.30	5.93	0.68 [0.25]	2.72	3.29	5.28
15	0.86	0.88	1.02	-0.61 [0.73]	2.95	3.97	7.22	0.53 [0.30]	2.94	3.62	5.99
18	0.84	0.86	1.02	-0.52 [0.70]	3.42	4.57	8.19	0.96 [0.17]	3.27	4.06	6.63
21	0.82	0.83	1.02	-0.56 [0.71]	4.07	5.87	8.85	**1.60** [0.06]	3.69	5.33	7.64
24	0.79	0.80	1.01	-0.41 [0.66]	4.84	7.21	11.06	**2.41** [0.01]	4.48	6.01	10.49

Table 3: Out-of-sample point forecast evaluation, linear AR and Michael, Nobay, and Peel (1997) ESTAR models

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
					Parametric bootstrapped *M-DM* critical values				Parametric bootstrapped *MW-DM* critical values		
h[a]	AR/RW[b]	ESTAR/ RW[c]	ESTAR/ AR[d]	*M-DM*[e]	10%	5%	1%	*MW-DM*[f]	10%	5%	1%
United Kingdom, 1991-2002 out-of-sample period											
1	0.96	0.93	0.97	1.12 [0.14]	**1.06**	1.41	2.29	**2.21** [0.02]	**1.70**	**2.02**	2.51

Notes: *p*-value using the *t* distribution with $P_h - 1$ degrees of freedom is reported in brackets; bold statistic indicates significance at the 10% level according to the *p*-value; bold bootstrapped critical value indicates that the statistic is significant according to the bootstrapped critical value.

[a]Forecast horizon (years-ahead).

[b]Ratio of the linear AR model RMSFE to the random walk model RMSFE.

[c]Ratio of the ESTAR model RMSFE to the random walk model RMSFE.

[d]Ratio of the ESTAR model RMSFE to the linear AR model RMSFE.

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

[f]Modified weighted Diebold and 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 ESTAR model weighted MSFE.

Table 4: Out-of-sample interval forecast evaluation, linear AR and Obstfeld and 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]
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	**30.15** [0.00]	**3.53** [0.09]	**31.90** [0.00]
Band-TAR	2	0.05	3.77 [0.07], 0.96 [0.40]	0.04 [1.00], 0.02 [1.00]	5.71 [0.07], 5.14 [0.08]
Band-TAR	3	0.033	1.88 [0.23], 0.47 [0.61], 0.12 [0.86]	0.03 [1.00], 0.03 [1.00], 0.03 [1.00]	0.06 [1.00], 0.06 [1.00], 0.06 [1.00]
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	2.81 [0.11]	0.48 [0.54]	3.64 [0.17]
Band-TAR	2	0.05	1.59 [0.26], 1.59 [0.26]	0.70 [0.57], 0.70 [0.57]	0.78 [0.72], 0.78 [0.72]
Band-TAR	3	0.033	2.94 [0.12], 1.88 [0.23], **8.76** [0.00]	0.14 [0.74], 0.14 [0.74], 0.42 [0.72]	0.89 [0.64], 0.89 [0.64], 1.53 [0.51]
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]	1.89 [0.21]	**6.55** [0.04]
Band-TAR	2	0.05	0.96 [0.40], 4.41 [0.05]	0.00 [1.00], 0.00 [1.00]	0.32 [0.88], 0.32 [0.88]
Band-TAR	3	0.033	**9.53** [0.00], 1.06 [0.39], **6.82** [0.01]	0.02 [1.00], 0.02 [1.00], 0.00 [1.00]	0.29 [0.93], 0.29 [0.93], 0.13 [1.00]
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 [0.03], 6.86 [0.03], 6.26 [0.05]
Band-TAR	1	0.10	**4.28** [0.05]	1.04 [0.41]	**4.93** [0.09]
Band-TAR	2	0.05	**5.67** [0.02], 3.31 [0.09]	0.08 [1.00], 0.08 [1.00]	**6.55** [0.04], **6.55** [0.04]
Band-TAR	3	0.033	**7.53** [0.01], **7.53** [0.01], 5.12 [0.04]	0.34 [1.00], 0.34 [1.00], 0.65 [1.00]	**19.09** [0.00], **19.09** [0.00], **18.29** [0.00]

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 (months-ahead).

[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, and 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]
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]
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]
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]
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 (months-ahead).

[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 and 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]
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	**0.15**	1.36	1.08	**15.91**	0.52	2.22
Band-TAR	2	0.05	0.10, 0.10	2.02, 1.09	3.03, 0.06	**17.72**, **11.79**	**4.98**, 0.02	**9.35**, **6.52**
Band-TAR	3	0.033	0.14, 0.11, 0.10	1.68, 2.53, 1.27	0.11, 0.85, 0.04	**8.21**, **10.96**, **10.11**	0.29, 0.97, 0.18	2.90, 2.87, 2.08
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	3.17	1.54	**25.94**	1.21	**6.48**
Band-TAR	2	0.05	0.16, **0.19**	3.41, 4.93	0.72, 0.00	**16.21**, **6.69**	0.12, 0.16	**4.79**, 1.07
Band-TAR	3	0.033	0.19, **0.24**, **0.26**	**17.78**, 2.10, 1.99	0.05, 0.09, 0.18	**7.58**, **11.66**, **13.80**	0.07, 0.34, 0.07	1.41, 2.82, **7.53**
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	0.11	1.99	0.91	**27.36**	0.55	**5.00**
Band-TAR	2	0.05	0.14, **0.19**	1.47, 1.76	0.68, 0.64	**7.75**, **20.35**	0.01, 0.01	1.24, **5.56**
Band-TAR	3	0.033	0.21, 0.18, 0.23	0.93, **19.75**, 2.52	0.03, 0.01, 0.00	**15.59**, 6.58, **13.27**	0.64, 0.00, 0.16	**9.09**, 1.14, **5.23**
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	**3193.77**	1.76	**30.10**	**2.92**	**18.30**
Band-TAR	2	0.05	**0.19**, **0.19**	**1017**, **301**	1.67, 0.55	**24.81**, **9.75**	**5.77**, 1.20	**15.14**, 2.61
Band-TAR	3	0.033	**0.29**, **0.29**, **0.30**	**169**, **107**, **148**	0.03, 0.05, 0.07	**14.33**, **12.20**, **5.80**	0.29, 0.59, 0.00	**5.49**, 4.45, 0.49

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 (months-ahead).

[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, and Sarno (2001) ESTAR models

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Model	h[a]	0.10/h	KS[b]	DH[c]	$LB,\ k=1$[d]	$LB,\ k=2$[d]	$LB,\ k=3$[d]	$LB,\ k=4$[d]
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
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**
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**
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 (months-ahead).

[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
United Kingdom, 1793-1990	ESTAR	0.0074	0.0181	0.0212	0.0273	0.0140	**0.0118**	**0.0135**	0.0151

Notes: bold bootstrapped critical value indicates that the statistic is significant according to the bootstrapped critical value.
[a]Corradi and 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 and 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.

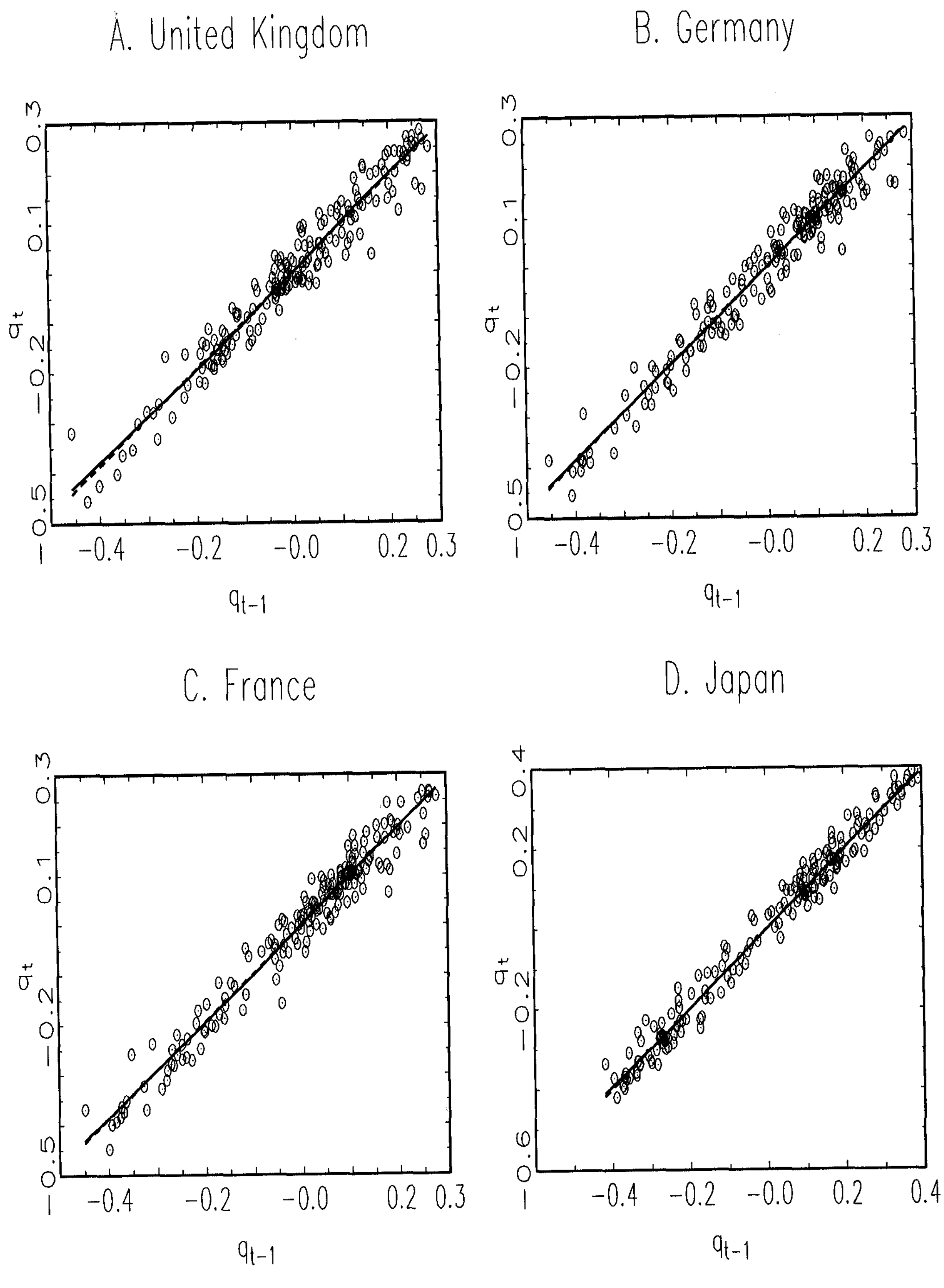

Figure 1: Scatterplot of real exchange rate log-levels (q_t) and lagged real exchange rate log-levels (q_{t-1}), Obstfeld and Taylor (1997) data.

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

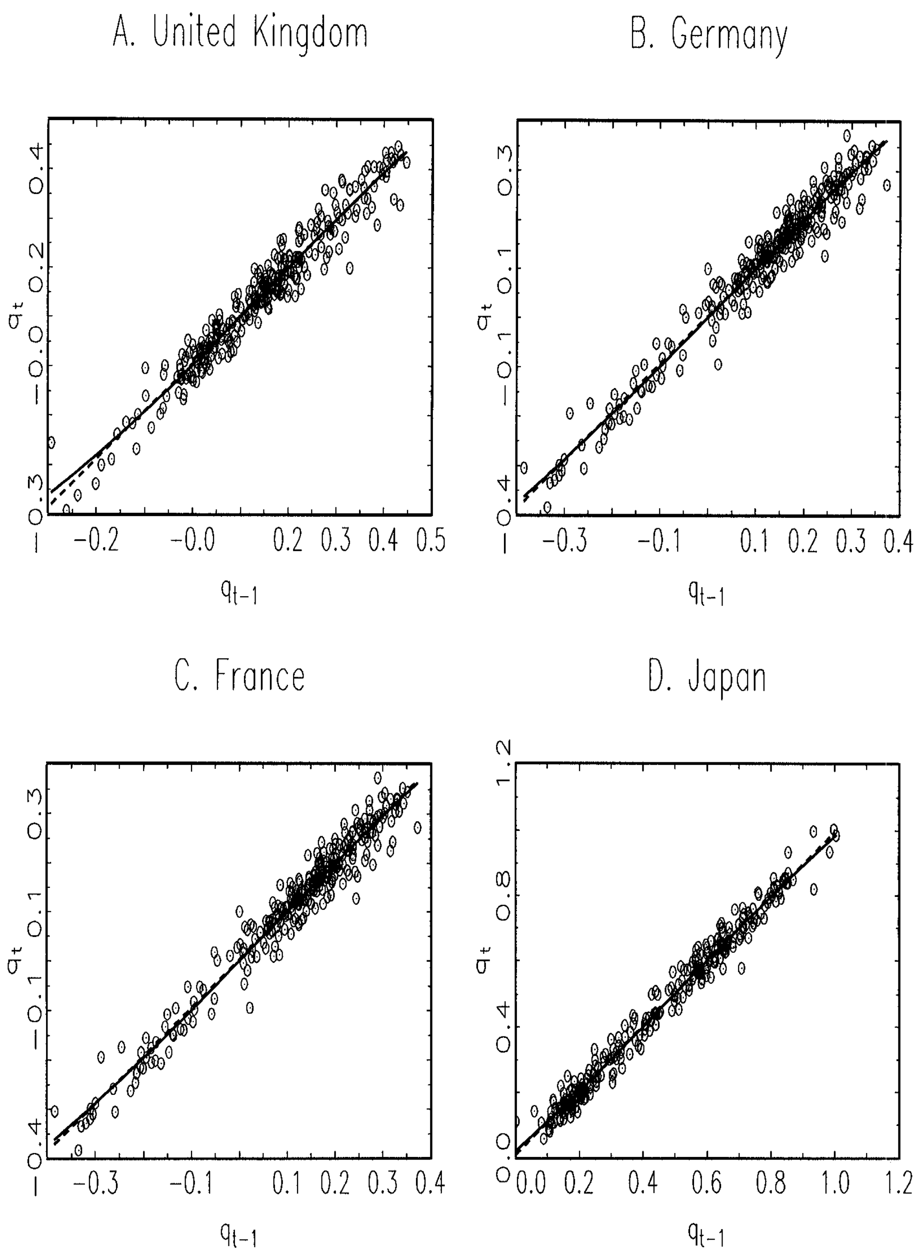

Figure 2: Scatterplot of real exchange rate log-level (q_t) and lagged real exchange rate log-level (q_{t-1}), Taylor, Peel, and Sarno (2001) data.

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

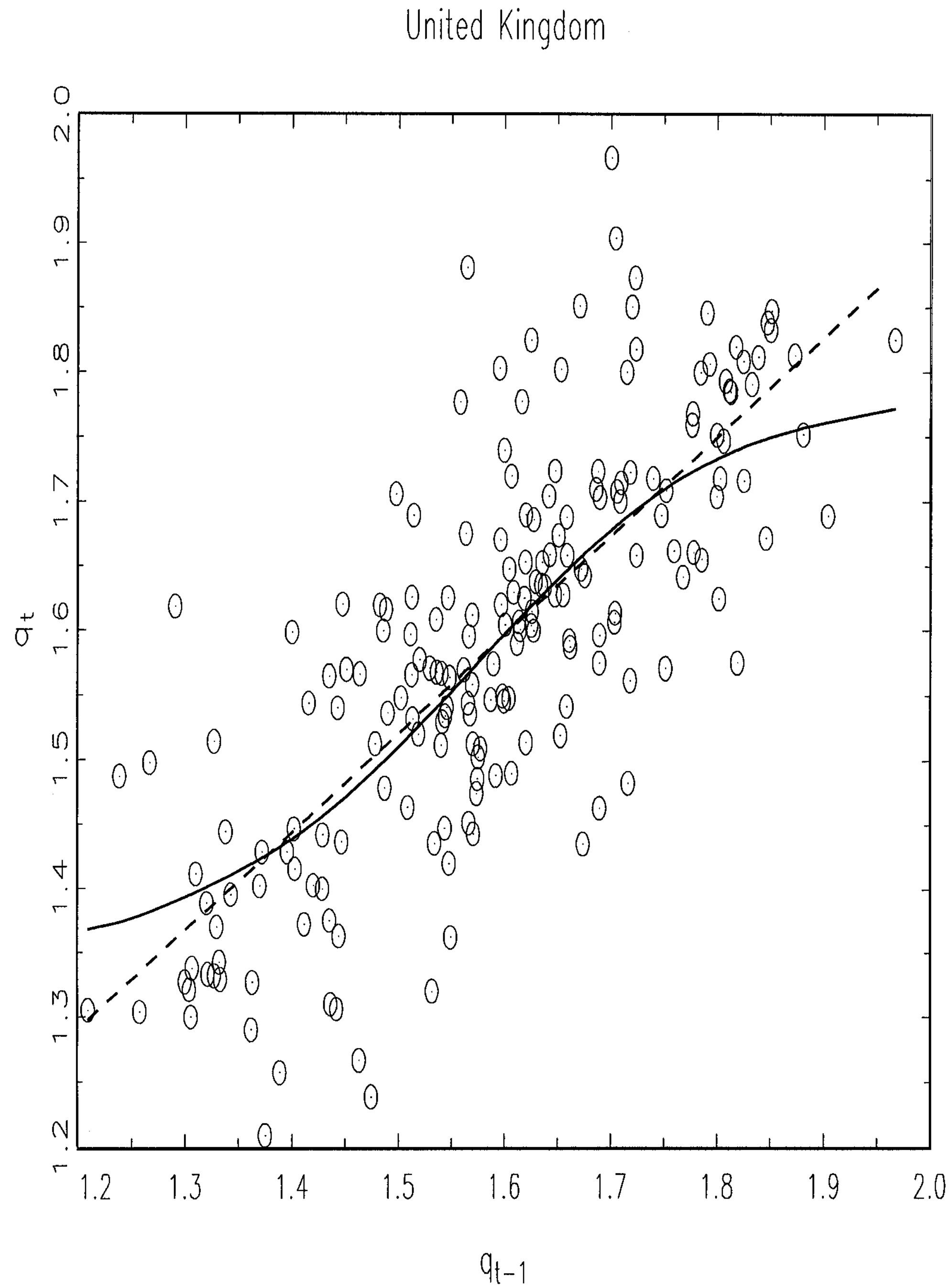

Figure 3: Scatterplot of real exchange rate log-levels (q_t) and lagged real exchange rate log-levels (q_{t-1}), Michael, Nobay, and Peel (1997) data.

Notes: solid line is the conditional expectation function for the fitted Michael, Nobay, and Peel (1997) ESTAR model; dashed line is the conditional expectation function for a fitted linear AR(1) model.