Nonlinear Models of Real Exchange Rate Behavior: A Re-examination

David E. Rapach*
Albers School of Business and Economics
Seattle University
900 Broadway
Seattle, WA 98122-4340
Phone: 206-296-5705
Fax: 206-296-2486
E-mail: rapachd@seattleu.edu

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

July 17, 2003

Abstract

In this paper, we re-examine a number of nonlinear models of U.S. dollar real exchange rate behavior from the extant empirical literature. Our re-examination employs econometric tests that go beyond the current "standard" analysis of nonlinear models, including parametric encompassing tests of sample moments, graphical analysis of conditional expectation functions, model selection based on out-of-sample forecasting performance, and "flexible" parametric nonlinearity tests. Using these econometric tests and data from the post-Bretton Woods float, we find little statistical evidence to recommend either band-threshold autoregressive models or exponential smooth transition autoregressive models over simple linear autoregressive models. Using data spanning more than two centuries, we find more statistical support for an exponential smooth transition autoregressive model for the U.K.-U.S. real exchange rate. Finally, there is some statistical backing for some Markov-switching autoregressive models estimated using post-Bretton Woods data. Overall, our re-examination provides rather limited statistical support for nonlinear models of real exchange rate behavior from the extant literature, especially for nonlinear models estimated using data from the post-Bretton Woods period.

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

Key words: Real exchange rate; Transaction costs; Threshold autoregressive model; Smooth transition autoregressive model; Markov-switching autoregressive model

*Corresponding author. We thank Mike Bergman, Christian Dahl, Adrian Pagan, Hailong Qian, and Jack Strauss for very helpful comments on earlier drafts. The usual disclaimer applies. The results reported in this paper were generated using GAUSS 3.6.

1. Introduction

Purchasing power parity (PPP) states that national price levels should be equal when expressed in a common currency. While few (any?) economists believe that PPP holds at each point in time, "most instinctively believe in some variant of purchasing power parity as an anchor for long-run exchange rates" (Rogoff, 1996, p. 647). Testing for long-run PPP is tantamount to testing for a stationary real exchange rate,[1] and numerous papers have investigated the stationarity of real exchange rates. Clearly problematic from the standpoint of PPP, many studies are unable to reject the unit root null hypothesis for real exchange rates, especially over the post-Bretton Woods period.[2] However, several recent papers are able to reject the unit root null hypothesis over sample periods as short as the modern float using panel data methods (Flood and Taylor, 1996; Frankel and Rose, 1996; Oh, 1996; Wu, 1996; Papell, 1997; Taylor and Sarno, 1998; Wu and Wu, 2001). In addition, other recent studies find significant evidence of mean-reversion in real exchange rates using data covering very long periods, typically a century or more (Diebold, Husted, and Rush, 1991; Glen, 1992; Cheung and Lai, 1994; Lothian and Taylor, 1996; A. Taylor, 2002). Recent studies using panel data and long spans of data thus appear to have "rescued" PPP as a long-run proposition.[3]

While studies utilizing panel procedures and long spans of data have generally been successful in rejecting the unit root hypothesis for real exchange rates, these studies also find that deviations from PPP are very persistent, with half-lives for PPP deviations ranging from approximately 3-5 years.[4] This gives rise to what Rogoff (1996) calls the "purchasing power parity puzzle": real exchange rates are highly volatile in the short term, and this volatility is usually attributed to financial factors such as monetary

[1] The interest is typically in relative PPP, so that the real exchange rate is not required to be unity. Relative PPP recognizes that consumer price indices are typically used to construct real exchange rates.

[2] There are numerous surveys of the PPP literature; see, for example, Breuer (1994), M. Taylor (1995), Froot and Rogoff (1995), Rogoff (1996), and Sarno and Taylor (2002).

[3] However, these recent findings have not gone unchallenged; see, for example, O'Connell (1998a), Engel (2000), and Breuer, McNown, and Wallace (2001).

[4] Indeed, it is primarily because deviations from PPP are persistent that panel procedures and long spans of data are typically required to reject the unit root null hypothesis: unit root tests applied to a single real exchange rate series over a period as short as the modern float can have low power in distinguishing a persistent, but stationary, alternative process from a unit root null process.

shocks; although such shocks can have important real effects in the presence of sticky nominal prices, it seems implausible that nominal prices are so rigid as to be consistent with half-lives for PPP deviations of 3-5 years. Although it is quite plausible that real shocks (such as shocks to tastes and technologies) could have such persistent real effects, models where real shocks predominate are generally inconsistent with the observed short-term volatility in real exchange rates.

While most of the extant empirical literature, including the studies discussed above, assumes that real exchange rates follow linear autoregressive (AR) processes, a recent strand of the empirical literature considers nonlinear AR models. Nonlinear models of real exchange rate behavior are well supported by economic intuition, and the notion that real exchange rates could follow nonlinear processes dates back at least to Heckscher (1916), who suggested that deviations from the law of one price might be due to international transaction costs between spatially separated markets (Obstfeld and Taylor, 1997). A number of elegant theoretical models formalize the notion of nonlinear real exchange rate behavior due to transaction costs (Benninga and Protopapadakis, 1988; Williams and Wright, 1991; Dumas, 1992; Coleman, 1995; Sercu, Uppal, and Van Hulle, 1995; Ohanian and Stockman, 1997; O'Connell, 1998b; Obstfeld and Rogoff, 2000). Transaction costs can be broadly defined to include transportation costs, tariffs and nontariff barriers, and any other costs that agents incur in international trade (Obstfeld and Rogoff, 2000). In these models, arbitrage is not profitable if price discrepancies do not exceed transaction costs, giving rise to a "band of inactivity," where deviations from PPP are not corrected. However, "large" price discrepancies that exceed transaction costs are quickly corrected, so that the real exchange rate is driven back to the band of inactivity where arbitrage is no longer profitable.

As demonstrated by Obstfeld and Taylor (1997) and A. Taylor (2001), nonlinear adjustment in real exchange rates provides a potential explanation for the PPP puzzle. Via Monte Carlo simulations, Obstfeld and Taylor (1997) and A. Taylor (2001) show that estimates of the half-life for PPP deviations based on a linear AR model can seriously overstate the actual half-life (relevant outside the band of inactivity) when the true data-generating process is a band-threshold AR (Balke and Fomby, 1997; Band-TAR) model. For the Band-TAR models in Obstfeld and Taylor (1997) and A. Taylor (2001), there is an inner regime (corresponding to the band of inactivity) where the real exchange rate behaves as a unit root. Outside of the band, the Band-TAR model is characterized by reversion to the edge of the band. If a researcher wrongly assumes a linear AR structure, this incorrectly combines data from two separate regimes—an inner regime with no mean-reversion and an outer regime with possibly rapid reversion to the band's edge—and leads a researcher to overstate the half-life.

Motivated by theoretical models incorporating transaction costs and the potential specification bias associated with linear models, several recent studies have estimated nonlinear models for U.S. dollar real exchange rates. Obstfeld and Taylor (1997) estimate Band-TAR models for a large number of U.S. dollar real exchange rates based on both broad and disaggregated consumer price indices. They find evidence of reversion to the edge of the band of inactivity in the outer regime. They also find that half-lives corresponding to reversion to the edge of the band in the outer regime for their estimated Band-TAR models are often much shorter than the half-lives implied by estimated linear AR models.

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.[5] Bertola and Caballero (1990), Dumas (1994), and Teräsvirta (1994) suggest that time aggregation and nonsynchronous 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., Germany, France, and Japan over the post-Bretton Woods era, Taylor, Peel, and Sarno (2001) estimate parsimonious ESTAR models 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. Their

[5] As with the Band-TAR models in Obstfeld and Taylor (1997) and A. Taylor (2001), the ESTAR model is characterized by symmetric adjustment.

estimated ESTAR models pass a battery of diagnostic tests, and 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) 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.[6]

Another recent study that estimates nonlinear models for U.S. dollar real exchange rates is Bergman and Hansson (2000). They estimate two-state Markov-switching AR (Hamilton, 1989; MS-AR) models for six U.S. dollar real exchange rates over the post-Bretton Woods period. Their preferred specification allows for the intercept alone in an AR(1) process to follow a two-state Markov process. While this specification is not necessarily implied by theoretical models incorporating transaction costs, it is in the spirit of the "long swings" model of Engel and Hamilton (1990). Bergman and Hansson (2000) strongly reject a linear AR model in favor of their MS-AR model for most real exchange rates. In addition, they find that out-of-sample forecasts from their MS-AR model typically outperform forecasts generated by a random walk model over the 1990s. Overall, Bergman and Hansson (2000) claim considerable statistical support for an MS-AR model of real exchange rate behavior.

Obstfeld and Taylor (1997), Taylor, Peel, and Sarno (2001), Michael, Nobay, and Peel (1997), and Bergman and Hansson (2000) all report evidence of nonlinear behavior in U.S. dollar real exchange rates. In the present paper, we re-examine a number of the nonlinear models from these papers in an effort to better assess the overall strength of the statistical evidence for nonlinear behavior in U.S. dollar real exchange rates. Given the growing popularity of nonlinear models of real exchange rate behavior in the empirical literature, as well as the promise that they offer in providing a solution to the PPP puzzle, the

[6] 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.

degree of statistical support for nonlinear models of real exchange rate behavior is an important research topic. Our paper can also be viewed as a contribution to the broader literature that assesses the empirical importance of nonlinear behavior in economic variables; see, for example, Hess and Iwata (1997) and Potter and Koop (2000, 2001) with respect to the behavior of U.S. real GDP.

Our re-examination employs a set of econometric tests recently developed by Pagan (2002), Hamilton (2001), and Dahl and González-Rivera (2002), in addition to a comparison of out-of-sample forecasts from fitted linear and nonlinear AR models. This allows for an analysis of empirical nonlinear models that goes beyond the current "standard" approach, which typically contains two components. The first component involves testing the null hypothesis of linearity against the alternative hypothesis of a particular type of nonlinear process. For example, Michael, Nobay, and Peel (1997) use the Lagrange-multiplier test of Teräsvirta (1994) to test the null hypothesis of a linear AR model against the alternative of an ESTAR model, and Bergman and Hansson (2000) use the Hansen (1992, 1996) procedure to test the null hypothesis of a linear AR model against the alternative of a two-state MS-AR model. The second component of the standard approach subjects the residuals from an estimated nonlinear model to tests for any remaining serial correlation and/or nonlinearity. For example, Taylor, Peel, and Sarno (2001) test for any remaining serial correlation and any remaining ESTAR nonlinearity in the residuals of their fitted ESTAR models using tests from Eitrheim and Teräsvirta (1996), and Bergman and Hansson (2000) employ tests developed by Hamilton (1996) to test for autocorrelation and ARCH in the residuals of their fitted MS-AR models.

While these standard tests certainly play an important role in model evaluation, Pagan (2002) argues that they should be supplemented with other diagnostic tests. He suggests using a Hausman (1978)-type test (what Breunig, Najarian, and Pagan (2002) call a "parametric encompassing test") in order to compare the moments implied by the estimated nonlinear model for the endogenous variable to the endogenous variable's sample moments. If a given moment implied by an estimated nonlinear model is significantly different from the sample moment, this can indicate a serious specification problem with the fitted nonlinear model. Pagan (2002, p. 2) also argues that "statistical tests are often much less informative than a few simple graphical comparisons," and he advocates supplementing formal methods of evaluating a model with informal graphical analysis. More specifically, Pagan (2002) suggests graphing the conditional expectation functions for the estimated nonlinear model and a linear counterpart, as well as a scatterplot of the data. The conditional expectation functions for the nonlinear and linear models can be compared, and the "degree of nonlinearity" can be visually assessed.

Out-of-sample forecasting is widely regarded as a way to prevent model overfitting. In our re-examination, we compare out-of-sample forecasts from fitted linear and nonlinear AR models. We use the Diebold and Mariano (1995) test for equal predictive ability, as well as the Harvey, Leybourne, and Newbold (1998) test for forecast encompassing, to test whether forecasts generated by the nonlinear AR model are superior to those generated by the linear AR model. If the forecasts from the nonlinear model are statistically superior to those from the linear model, we select the nonlinear AR model over the linear AR model. This type of model selection fits well with the empirical exchange rate literature, as it is in the spirit of Meese and Rogoff (1983). Note that Bergman and Hansson (2000) compare out-of-sample forecasts from their fitted MS-AR model to those from a random walk model. However, in contrast to Bergman and Hansson (2000) and Meese and Rogoff (1983), we compare out-of-sample forecasts from a fitted nonlinear AR model of interest to those from an unrestricted linear AR model (and not a random walk), as we are interested in comparing globally stationary nonlinear and linear AR model specifications. Preliminary Monte Carlo evidence in Liu and Enders (2002) suggests that out-of-sample forecast comparisons perform well in selecting between nonlinear and linear AR models.

We also test for nonlinear behavior in real exchange rates using the recently developed Hamilton (2001) "flexible" parametric approach to nonlinear inference, which is based on the concept of random fields. Hamilton (2001) strikes a middle ground between a parametric approach that assumes a particular nonlinear structure and a flexible nonparametric approach that is not well suited for hypothesis testing and model specification. Among the tests developed by Hamilton (2001) is a Largrange-multiplier test of the null hypothesis of linearity against an alternative hypothesis that allows for a broad class of deterministic nonlinear functions. Dahl and González-Rivera (2002) develop additional Lagrange-multiplier tests based on random fields. In extensive Monte Carlo simulations, they find their tests to be quite powerful for detecting a number of different types of nonlinearities, including TAR and ESTAR processes. We employ the Hamilton (2001) and Dahl and González-Rivera (2002) tests in our re-examination in order to assess the robustness of the evidence for nonlinear behavior reported in Obstfeld and Taylor (1997), Michael, Nobay, and Peel (1997), and Bergman and Hansson (2000), who rely on nonlinearity tests that assume a particular type of nonlinear alternative.

Previewing our empirical results, we find rather limited evidence of nonlinear behavior in U.S. dollar real exchange rates. When we re-examine the fitted Band-TAR models in Obstfeld and Taylor (1997) and ESTAR models in Taylor, Peel, and Sarno (2001) using monthly data from the post-Bretton Woods float, we find little difference in the conditional expectation functions for the fitted nonlinear models and linear counterparts. Furthermore, out-of-sample forecasts generated by the nonlinear models do not outperform forecasts generated by linear models over the 1990s, and Hamilton (2001) and Dahl and González-Rivera (2002) tests fail to provide significant evidence of nonlinear behavior. We find more support for the ESTAR model of Michael, Nobay, and Peel (1997), which is estimated using a long span of annual data for the U.S. dollar-U.K. sterling real exchange rate. There are noticeable differences between the conditional expectation functions for the fitted ESTAR model and a linear AR counterpart, and the fitted ESTAR model generates forecasts that are significantly superior to those from a linear AR model over the 1961-1990 out-of-sample period. Using quarterly data from the post-Bretton Woods period, support for the estimated MS-AR models in Bergman and Hansson (2000) is mixed. For the U.K. and Switzerland, the fitted MS-AR models do not encompass the sample variance, pointing to possible specification problems with the MS-AR models for these countries. The estimated MS-AR models for some other countries do encompass the sample variance and first-order autocorrelation, and some of the models generate out-of-sample forecasts that are significantly superior to those from a linear AR model over the 1990s. Overall, while nonlinear models are well motivated by economic theory and provide a potential explanation for the PPP puzzle, we find rather limited statistical support for extant nonlinear models of U.S. dollar real exchange rate behavior in our re-examination, especially nonlinear models estimated using data from the post-Bretton Woods period.

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), Michael, Nobay, and Peel (1997), and Bergman and Hansson (2000). Section 3 describes the econometric tests that we use in our re-examination. Section 4 presents our empirical results. Section 5 concludes.

2. Review of Four Extant Studies

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

Motivated by models of costly arbitrage due to transport costs, Obstfeld and Taylor (1997) estimate Band-TAR models for consumer price index-based U.S. dollar real exchange rates.[7] Their Band-TAR model takes the form,

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

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

[7] Obstfeld and Taylor (1997) 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.

introduction, using Monte Carlo simulations, Obstfeld and Taylor (1997) show that incorrectly assuming a linear AR structure can lead a researcher to substantially understate the speed of adjustment in real exchange rate deviations when the data are actually generated by a Band-TAR process such as equation (1).

Obstfeld and Taylor (1997) estimate equation (1) via maximum likelihood using monthly data covering 1980-1994 for a large number of countries. We concentrate on their results for Canada, France, Germany, Italy, U.K., and Japan, as this constitutes a set of countries similar to those used in the other studies that we re-examine. Obstfeld and Taylor (1997) 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. Obstfeld and Taylor (1997) 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.) They also find that the half-life calculated from the estimated Band-TAR model's outer regime is often substantially lower than the half-life implied by a fitted linear AR(1) model. 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.[8]

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

Taylor, Peel, and Sarno (2001, TPS) consider the following smooth transition AR model,

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

[8] Obstfeld and Taylor (1997) 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 six countries that we focus on.

where q_t is stationary and ergodic, $\varepsilon_t \sim iid(0,\sigma^2)$, and η is the long-run equilibrium level for q_t. The transition function, $\Phi(\alpha; q_{t-d} - \eta)$, allows for nonlinear mean-reversion in the real exchange rate process. TPS use the exponential transition function (Granger and Teräsvirta, 1993),

$$\Phi(\alpha; q_{t-d} - \eta) = 1 - \exp[\alpha \cdot (q_{t-d} - \eta)^2], \tag{3}$$

where $\alpha < 0$, so that equation (2) constitutes an ESTAR model. In contrast to the Band-TAR model, the ESTAR model allows for smooth transition between two extreme regimes. TPS view the ESTAR model specification as being consistent with theoretical models based on transaction costs in international arbitrage.

Using monthly data from 1973:01-1996:12, TPS estimate equation (2) for the real exchange rate for four countries relative to the U.S. dollar (U.K., Germany, France, Japan). They define the log-level of the real exchange rate as $q_t \equiv s_t - p_t + p_t^*$, 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 to zero in 1973:01. Based on inspection of partial autocorrelation functions (Teräsvirta, 1994), they set $p = 1$ in equation (2) for each country, and they set the delay parameter, *d*, equal to one for each country according to specification searches. TPS estimate equation (2) using multivariate nonlinear least squares (nonlinear seemingly unrelated regressions),[9] and they cannot reject the restrictions that $\beta_1 = -\beta_1^* = 1$, so they report estimation results for the following parsimonious ESTAR model,

$$q_t = q_{t-1} - \{1 - \exp[\alpha \cdot (q_{t-1} - \eta)^2]\} \cdot (q_{t-1} - \eta) + \varepsilon_t . \tag{4}$$

For the parsimonious ESTAR model, equation (4), 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 $\alpha < 0$). TPS obtain negative and statistically

---

[9] This is equivalent to maximum likelihood estimation under the assumption that the disturbance terms are distributed normally.

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, and no remaining logistic nonlinearity.

TPS use their estimated ESTAR models for two additional purposes. First, they calculate impulse responses,[10] 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). Furthermore, via Monte Carlo simulations based on their estimated ESTAR models, they show that standard unit root tests have low power to detect nonlinear mean-reversion in real exchange rates over the post-Bretton Woods period, while panel unit root tests are much more powerful in detecting nonlinear mean-reversion. Overall, TPS argue that their estimated ESTAR processes for real exchange rates can explain the slow speeds of mean-reversion found in the empirical literature, as well as the inability to reject the unit root null hypothesis on a country-by-country basis over the post-Bretton Woods era (and the ability to reject the unit root null using panel methods).

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, and they 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.

[10] 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).

Like TPS, MNP consider an ESTAR specification. Based on inspection of the partial autocorrelation function, MNP select $p = 2$ in equation (2).[11] Considering delay parameter values from 1-3, MNP test the null hypothesis of a linear model against an ESTAR alternative specification using the *F*-statistic version of the Lagrange-multiplier test in Teräsvirta (1994). This test is based on the artificial regression,

$$q_t = \beta_{00} + \sum_{j=1}^{2} (\beta_{0j} \cdot q_{t-j} + \beta_{1j} \cdot q_{t-j} \cdot q_{t-d} + \beta_{2j} \cdot q_{t-j} \cdot q_{t-d}^2) + e_t \, , \tag{5}$$

where the null hypothesis of linearity is represented by $\beta_{1j} = \beta_{2j} = 0$ for $j = 1,2$. MNP are able to reject the null hypothesis of linearity for $d = 1$. They interpret this as evidence of nonlinear adjustment in the real exchange rate, and they proceed to estimate the ESTAR model via nonlinear least squares.[12] They are unable to reject the restrictions that $\beta_1 + \beta_2 = 1$, $\beta_1 = -\beta_1^*$, and $\beta_2 = -\beta_2^*$, so they report estimation results for the restricted ESTAR model,[13]

$$q_t - \eta = \{\exp[\alpha \cdot (q_{t-1} - \eta)^2]\} \cdot [\beta_1 \cdot (q_{t-1} - \eta) + (1 - \beta_1) \cdot (q_{t-2} - \eta)] + \varepsilon_t \, . \tag{6}$$

As with the parsimonious ESTAR model estimated by TPS, the restricted ESTAR model in MNP behaves as a random walk in the extreme inner regime. MNP obtain a negative and statistically significant estimate of α, so that the speed of mean-reversion increases with the deviation in the real exchange rate from its long-run equilibrium. Like TPS, MNP interpret their results as evidence of nonlinear behavior in the U.S.-U.K. real exchange rate that is consistent with theoretical models of international arbitrage in the presence of transaction costs.

---

[11] MNP actually specify q_t in deviations from the mean, so they omit η in equation (2). In order to facilitate comparisons across studies, we leave η in equation (2) in our empirical tests in Section 4 below. This has no important effects on our results.

[12] As is the case for estimation of the ESTAR models in TPS, this is equivalent to maximum likelihood estimation under the assumption that the disturbance term is distributed normally.

[13] MNP actually work with an augmented Dickey-Fuller form of equation (6), but their restrictions on the augmented Dickey-Fuller form of equation (6) are equivalent to those given in the text. We use the specification in equation (6) instead of the augmented Dickey-Fuller form of equation (6) in order to facilitate comparison with the ESTAR model specification in TPS.

2.4. Bergman and Hansson (2000) MS-AR(1) Model

Instead of using a Band-TAR or ESTAR specification, Bergman and Hansson (2000) model post-Bretton Woods real exchange rates using an MS-AR specification (Hamilton, 1989). They include a single lag, resulting in an MS-AR(1) model that takes the form,

$$q_t = \zeta_{s_t} + \phi_{s_t} \cdot q_{t-1} + \varepsilon_t \,, \tag{7}$$

where $\varepsilon_t \mid s_t \sim N(0, \sigma^2_{s_t})$ and the unobserved random state variable, s_t, can take on a value of 1 or 2 and follows the two-state Markov process,

$$\Pr(s_t = j \mid s_{t-1} = i) = p_{ij} \text{, for } i, j = 1,2 \,, \tag{8}$$

and $\sum_{j=1}^{2} p_{ij} = 1$ for $i = 1,2$. Bergman and Hansson (2000) use quarterly data covering 1973:1-1997:4 for six U.S. dollar real exchange rates (U.K., France, Germany, Switzerland, Canada, and Japan) based on consumer price indices, where the real exchange rate is defined as the nominal exchange rate (U.S. dollar price of foreign currency) times the foreign consumer price index divided by the U.S. consumer price index. They normalize the real exchange rate to unity in 1973:2, and they use 100 times the log-level of the real exchange rate. Bergman and Hansson (2000) estimate the MS-AR(1) model via maximum likelihood using data through 1990:4, reserving the 1991:1-1997:4 period for evaluation of out-of-sample forecasts. Using Wald and likelihood-ratio tests, they generally cannot reject the restrictions that ϕ_{s_t} and σ_{s_t} are state-independent, so they report estimation results for the following restricted MS-AR(1) model for each country,

$$q_t = \zeta_{s_t} + \phi \cdot q_{t-1} + \varepsilon_t \,, \tag{9}$$

where $\varepsilon_t \sim N(0, \sigma^2)$. They find that the intercept term can differ markedly across the two state regimes, and their estimates of ϕ range from 0.871-0.958, indicating a globally stationary process for each real exchange rate. In addition, they test for a one-state AR(1) model against their two-state restricted MS-

AR(1) alternative model. Using the grid procedure in Hansen (1992, 1996), they reject the one-state model in favor of their MS-AR(1) model for every country except the U.K.

In the spirit of Meese and Rogoff (1983), Bergman and Hansson (2000) proceed to compare out-of-sample forecasts from a random walk model to their MS-AR(1) model. Considering forecast horizons of 1-4 quarters for each country over the 1991:1-1997:4 out-of-sample period, forecasts generated by the MS-AR(1) model have a lower mean squared error (MSE) than forecasts generated by the random walk with drift model for each country at each horizon. Moreover, using the Harvey, Leybourne, and Newbold (1998) forecast encompassing test, they generally reject the null hypothesis that the forecasts from the random walk with drift model encompass the forecasts from the MS-AR(1) model, indicating that the forecasts from the MS-AR(1) model contain information useful for forecasting beyond that already contained in the random walk with drift model. Finally, using Monte Carlo simulations, Bergman and Hansson (2000) find that conventional unit root tests typically have low power to reject the unit root null hypothesis when the true data-generating process is a stationary MS-AR(1) model.

3. Alternative Econometric Tests for Nonlinear Models

3.1. Parametric Encompassing Tests of Sample Moments

The first diagnostic tool that we use to re-examine nonlinear models from the four studies cited above is the Pagan (2002) parametric encompassing test. While “standard” tests of nonlinear model adequacy involve testing the fitted nonlinear model residuals for serial correlation and any remaining nonlinearities, Pagan (2002) recommends comparing sample moments to those implied by the fitted nonlinear model as an additional check on model adequacy. If the fitted nonlinear model cannot encompass (or “match”) particular sample moments, this can indicate a serious problem with fitted nonlinear model specification. As emphasized by Pagan (2002), the particular moments to consider will vary across applications. When evaluating nonlinear real exchange rate models, the ability of a fitted nonlinear model to encompass the sample variance and first-order autocorrelation is of special interest, as the volatility and persistence of real exchange rates are the key stylized facts that constitute the PPP puzzle. In our applications below, we thus focus on the sample variance, $\hat{\sigma}_q^2 = (1/T)\sum_{t=1}^{T}(q_t - \overline{q})^2$, where $\overline{q} = (1/T)\sum_{t=1}^{T} q_t$, and sample first-order autocorrelation, $\hat{\rho} = [1/(T-1)]\sum_{t=2}^{T}(q_t - \overline{q})(q_{t-1} - \overline{q}) / \hat{\sigma}_q^2$.

In order to form a statistic for testing the null hypothesis that the sample moment equals the moment implied by the fitted nonlinear model, Pagan (2002) considers the differential, $\hat{\mu} - \mu(\hat{\theta})$, where $\hat{\mu}$ is the sample moment and $\mu(\hat{\theta})$ is the value of the moment implied by the fitted nonlinear model ($\hat{\theta}$ denotes the estimated vector of the parameters for the nonlinear model). Pagan (2002) recommends calculating $\mu(\hat{\theta})$ via simulation, and, as in Pagan (2002), we use the following procedure. With the fitted nonlinear model serving as the data-generating process, we use a Monte Carlo simulation to generate 30,000 observations for the log-level of the real exchange rate and calculate the variance and first-order autocorrelation for the 30,000 observations. By simulating a very large number of observations, we essentially limit the sampling uncertainty associated with $\mu(\hat{\theta})$ to the estimated parameters of the nonlinear model. As pointed out by Pagan (2002), $\hat{\mu} - \mu(\hat{\theta})$ is a version of the Hausman (1978) test, as $\hat{\mu}$ is a consistent estimate and $\mu(\hat{\theta})$ is an efficient estimate under the null hypothesis that the fitted nonlinear model is valid. We could thus use the *t*-statistic, $\{\mathrm{var}[\hat{\mu} - \mu(\hat{\theta})]\}^{-0.5}[\hat{\mu} - \mu(\hat{\theta})]$, to test the null hypothesis that $\hat{\mu} - \mu(\hat{\theta}) = 0$, and the *t*-statistic will have a standard normal asymptotic distribution. When θ is estimated efficiently, Hausman (1978) shows that the variance of $\hat{\mu} - \mu(\hat{\theta})$ is given by $\mathrm{var}[\hat{\mu} - \mu(\hat{\theta})] = \mathrm{var}(\hat{\mu}) - \mathrm{var}[\mu(\hat{\theta})]$. Given that it can be difficult to estimate $\mathrm{var}[\mu(\hat{\theta})]$, Pagan (2002) suggests using the modified *t*-statistic, $[\mathrm{var}(\hat{\mu})]^{-0.5}[\hat{\mu} - \mu(\hat{\theta})]$. This will be a conservative test, as $\mathrm{var}(\hat{\mu}) \geq \mathrm{var}(\hat{\mu}) - \mathrm{var}[\mu(\hat{\theta})]$. The modified *t*-statistic is relatively easy to compute, as we can calculate a heteroskedastic and autocorrelation consistent (HAC) standard error for the sample moment using the

Newey and West (1987) procedure.[14] To compute the standard error for the sample variance and first-order autocorrelation, we regress $(q_t - \overline{q})^2$ and $(q_t - \overline{q})(q_{t-1} - \overline{q}) / \hat{\sigma}_q^2$, respectively, on a constant only and calculate the HAC standard error corresponding to the constant. We also calculate a conservative χ^2-statistic in order to test the null hypothesis that the sample variance and first-order autocorrelation jointly encompass the sample variance and first-order autocorrelation.[15]

3.2. Graphical Analysis of Conditional Expectation Functions

As a more informal way of evaluating a fitted nonlinear model using graphical analysis, Pagan (2002) proposes comparing the conditional expectation function for the fitted nonlinear model to a scatterplot of the data, as well as to the conditional expectation function for a fitted linear model. By comparing the conditional expectation function for the fitted nonlinear model to a scatterplot of the data, we can develop a visual sense of how well the estimated nonlinear model fits the data and examine whether a few outliers are primarily responsible for any nonlinearities. By comparing the conditional expectation function for the fitted nonlinear model to that of a fitted linear model, we can get a visual sense of the "degree of nonlinearity" in the estimated nonlinear model.

In our applications in Section 4 below, we consider the expectation of q_t conditional on q_{t-1}, as all of the studies that we re-examine use an AR framework. It is straightforward to calculate the conditional expectation function for the fitted Band-TAR and ESTAR models in Obstfeld and Taylor (1997) and TPS. It is more complicated to calculate the conditional expectation function for the ESTAR and MS-AR(1) models in MNP and Bergman and Hansson (2000), as this involves integrating out some variables. For these models, we follow Pagan (2002) and use a nonparametric procedure. More specifically, we generate 30,000 observations for q_t using the fitted nonlinear model and then apply a

[14] In our applications in Section 4 below, we use the Bartlett kernel and a lag truncation of 12 (8, 4) in the Newey and West (1987) procedure for monthly (quarterly, annual) data.

[15] The χ^2-statistic is distributed as a χ^2 random variable with two degrees of freedom under the null hypothesis.

nonparametric estimator to the generated observations in order to evaluate the conditional expectation function at the sample data points for q_{t-1}.[16] For all of the nonlinear models that we re-examine, we compare the conditional expectation function for the fitted nonlinear model to that of a fitted linear AR(1) model.

3.3. Model Selection Based on Out-of-Sample Forecasting

We compare one-step-ahead out-of-sample forecasts from nonlinear and linear AR models using the following procedure. We first divide the total sample of *T* observations into in-sample and out-of-sample portions, where the in-sample portion spans the first *R* observations and the out-of-sample portion the last *P* observations for q_t. We then estimate the nonlinear and linear AR models using the in-sample portion of the total sample, and we use the estimated models to generate two series of *P* one-step-ahead out-of-sample forecasts, one corresponding to the fitted nonlinear AR model and the other to the fitted linear AR model. Denote the out-of-sample forecast errors for the fitted nonlinear AR model by $\{\hat{u}_{NL,t+1}\}_{t=R}^{T-1}$ and those for the fitted linear AR model by $\{\hat{u}_{L,t+1}\}_{t=R}^{T-1}$. We formally test whether the nonlinear AR model forecasts are superior to the linear AR model forecasts using the Diebold and Mariano (1995) and Harvey, Leybourne, and Newbold (1998) statistics.

We use the Diebold and Mariano (1995) statistic, what we label the *MSE-T* statistic (following Clark and McCracken, 2001), to test the null hypothesis that the nonlinear AR model forecast MSE is equal to the linear AR model forecast MSE against the one-sided (upper-tail) alternative hypothesis that the nonlinear AR model forecast MSE is less than the linear AR model forecast MSE. The *MSE-T* statistic is

[16] Following Pagan (2002), we use the Nadaraya-Watson estimator with a Gaussian kernel and window width of the form in Silverman (1986).

based on the loss differential, $\hat{d}_{t+1} = (\hat{u}_{L,t+1})^2 - (\hat{u}_{NL,t+1})^2$. Let $\bar{d} = P^{-1}\sum_{t=R}^{T-1}\hat{d}_{t+1} = MSE_L - MSE_{NL}$, where $MSE_i = P^{-1}\sum_{t=R}^{T-1}(\hat{u}_{i,t+1})^2$, $i = L, NL$, and let $V_d = P^{-1}\sum_{t=R}^{T-1}(\hat{d}_{t+1} - \bar{d})^2$. The *MSE-T* statistic is given by

$$MSE\text{-}T = P^{0.5} \cdot \bar{d} \cdot V_d^{-0.5}. \tag{10}$$

A significant *MSE-T* statistic indicates that the nonlinear AR model forecasts are statistically superior to those of the linear AR model. West (1996) shows that the *MSE-T* statistic has a standard normal limiting distribution.[17]

Our other out-of-sample statistic, the Harvey, Leybourne, and Newbold (1998) statistic, or what we label the *ENC-T* statistic (again following Clark and McCracken, 2001), relates to the concept of forecast encompassing. Forecast encompassing is based on optimally constructed composite forecasts.[18] Intuitively, if the forecasts from the linear AR model encompass the forecasts from the nonlinear AR model, the nonlinear AR model provides no useful information for predicting real exchange rates beyond that already contained in the linear AR model; if the linear AR model forecasts do not encompass the nonlinear AR model forecasts, then the nonlinear AR model does contain information useful for predicting real exchange rates beyond that already contained in the linear AR model. Tests for forecast encompassing are tantamount to testing whether the weight attached to the nonlinear AR model forecast is zero in an optimal composite forecast that takes the form of a convex combination of the linear and nonlinear AR model forecasts. The *ENC-T* statistic is given by

$$ENC\text{-}T = P^{0.5} \cdot \bar{c} \cdot V_c^{-0.5}, \tag{11}$$

where $\hat{c}_{t+1} = \hat{u}_{L,t+1}(\hat{u}_{L,t+1} - \hat{u}_{NL,t+1})$, $\bar{c} = P^{-1}\sum_{t=R}^{T-1}\hat{c}_{t+1}$, and $V_c = P^{-1}\sum_{t=R}^{T-1}(\hat{c}_{t+1} - \bar{c})^2$. Under the null hypothesis, the weight attached to the nonlinear AR model forecast in the optimal composite forecast is

[17] McCracken (2000) shows that the limiting distribution of the *MSE-F* statistic is non-standard when comparing forecasts from two nested linear models.

[18] See Clements and Hendry (1998) for a textbook discussion of forecast encompassing.

zero, and the linear AR model forecasts encompass the nonlinear AR model forecasts. Under the one-sided (upper-tail) alternative hypothesis, the weight attached to the nonlinear AR model forecast in the optimal composite forecast is greater than zero, so that the linear AR model forecasts do not encompass the nonlinear AR model forecasts. Similar to the *MSE-T* statistic, the theory in West (1996) can be used to show that the *ENC-T* statistic has a standard normal limiting distribution.[19]

3.4. Hamilton (2001) and Dahl and González-Rivera (2002) Nonlinearity Tests

Hamilton (2001) develops what he terms a "parametric approach to flexible nonlinear inference." Consider a regression model of the form, $y_t = \xi(x_t) + \varepsilon_t$, where x_t is a $k \times 1$ vector (which may include lagged values of the regressand, as it does for our applications), $\varepsilon_t \sim N(0, \sigma^2)$, and the functional form, $\xi(\cdot)$, is unknown. Hamilton (2001) proposes treating $\xi(\cdot)$ itself as the outcome of a random process, and he considers the following specification,

$$y_t = \beta_0 + x_t' \cdot \beta_1 + \lambda \cdot m(g \circ x_t) + \varepsilon_t, \tag{12}$$

where $m(z)$, for any *z*, represents a realization of a stationary and homogeneous Gaussian random field. The linear component of the conditional mean of y_t is given by $\beta_0 + x_t' \cdot \beta_1$, while the nonlinear component is given by $\lambda \cdot m(g \circ x_t)$, where λ represents the contribution of the nonlinear component to the conditional mean, $g \geq 0$ is a vector of parameters that governs the curvature of the conditional mean (it is related to the covariance function associated with the random field), and $\circ$ denotes element-by-element multiplication. Hamilton (2001) develops a Lagrange-multiplier test of the null hypothesis of linearity ($\lambda^2 = 0$) against the alternative hypothesis of nonlinearity, and, following Dahl and González-Rivera (2002), we label this the λ_H^E statistic. Under the null hypothesis, *g* is unidentified, and Hamilton (2001) solves this problem by fixing *g* to the mean of its prior distribution, which is tantamount to

[19] Similar to the case for the *MSE-T* statistic, Clark and McCracken (2001) show that the *ENC-T* statistic has a non-standard limiting distribution when comparing forecasts from two nested linear models.

assuming that the random field's covariance function is known. The Hamilton (2001) approach is attractive in that it creates a framework for nonlinearity testing that does not "force" a particular functional form on the regression model under the alternative hypothesis, and in this sense is a "flexible" parametric approach.[20]

Dahl and González-Rivera (2002) develop three new statistics to test the null of linearity against the alternative of nonlinearity—what they label the λ_{OP}^{E}, λ_{OP}^{A}, and g_{OP} statistics—that are closely related to the Hamilton (2001) statistic. Like the Hamilton (2001) λ_{H}^{E} statistic, the Dahl and González-Rivera (2002) λ_{OP}^{E} and λ_{OP}^{A} statistics test the null hypothesis that $\lambda^2 = 0$ in equation (12). Like the Hamilton (2001) λ_{H}^{E} test, the Dahl and González-Rivera (2002) λ_{OP}^{E} test assumes that the random field's covariance function is known. In contrast, the Dahl and González-Rivera (2002) λ_{OP}^{A} test does not assume a known covariance function, as it is based on a higher-order Taylor series approximation of the covariance function. The Dahl and González-Rivera (2002) g_{OP} statistic is used to test for nonlinearity under the null hypothesis that $g = 0$ in equation (12), and it is free of the nuisance parameter, λ, under the null hypothesis. In extensive Monte Carlo simulations involving a host of nonlinear models, Dahl and González-Rivera (2002) find that their statistics are often more powerful than the Hamilton (2001) statistic in detecting nonlinearity. For the purposes of the present paper, it is worth noting that Dahl and González-Rivera (2002) find the λ_{OP}^{A} statistic to be particularly powerful against TAR and ESTAR alternatives.

[20] In extensive Monte Carlo simulations, Dahl (2002) finds that the Hamilton (2001) test is more powerful than tests based on the multivariate spline smoother (which also do not require the specification of a particular functional form under the alternative hypothesis).

4. Empirical Results

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

We first re-examine the Band-TAR models in Obstfeld and Taylor (1997). We use monthly data from Obstfeld and Taylor (1997) for the U.S. dollar real exchange rate with respect to Canada, France, Germany, Italy, U.K., and Japan.[21] The real exchange rate data begin in 1980:01 for each country and end in 1994:10 (1994:11; 1994:12) for Italy (Canada, France, U.K., Germany; Japan). Following Obstfeld and Taylor (1997), we first demean the log-level of the real exchange rate before estimating the Band-TAR model, equation (1). In their tables, Obstfeld and Taylor (1997) report detailed results for detrended real exchange rates, but note that the results are qualitatively similar when they use demeaned data. We report the results for demeaned data, as this is consistent with the treatment of the real exchange rate in the other studies we re-examine.[22] Again following Obstfeld and Taylor (1997), we estimate equation (1) via maximum likelihood. We implement maximum likelihood estimation through a grid search over possible values of c, and we require both the inner and outer regimes to contain at least 30 observations for Δq_t. After allowing for differencing, the usable sample begins in 1980:02 for each country.

Table 1 reports an estimated linear AR(1) model (estimated via least squares) and the estimated Band-TAR model for each country. The estimated linear AR(1) models indicate that the real exchange rates are fairly persistent. However, the estimated Band-TAR models indicate that the speed of mean-reversion is faster—sometimes considerably so—when we allow for a band of inaction, in line with the results in Obstfeld and Taylor (1997). The estimated values of c, which define the band of inaction, appear plausible, ranging from 0.07-0.31, and are similar to those reported in Obstfeld and Taylor (1997).

Next, we use the Pagan (2002) test to examine whether the variance and first-order autocorrelation implied by the fitted Band-TAR models encompass their sample counterparts. The results are reported in Table 1. Note that the variances and first-order autocorrelations are calculated for the

[21] We thank Alan Taylor for generously providing the data used in Obstfeld and Taylor (1997).
[22] Our results are qualitatively unchanged when we use detrended data.

levels of q_t, again to be consistent with the other studies that we re-examine. From Table 1, it is evident that we cannot reject (at conventional significant levels) the null hypothesis that the variance and first-order autocorrelation implied by the fitted Band-TAR model equal their sample counterparts (individually and jointly) for each country. According to this metric, the fitted Band-TAR models are not out of line with the data.

In Figure 1, we graph the conditional expectation functions for q_t given q_{t-1} for the fitted linear AR(1) and Band-TAR models over the range of q_{t-1} values in the data, as well as a scatterplot of the data. Visual inspection of the conditional expectation functions indicates little apparent difference between the fitted linear and nonlinear AR models over the range of q_{t-1} values in the data, and the scatterplots do not evince obvious nonlinear patterns. The "closeness" of the fitted linear and nonlinear models is confirmed by the in-sample forecast comparisons reported in Table 1. The in-sample root mean squared errors (RMSEs) for the linear and nonlinear model forecasts are very close to one another for each country, differing by less than half of 1% for each country. When we compare out-of-sample forecasts, reserving the period beginning in 1991:01 and running through the end of the available sample for the formation of out-of-sample forecasts, again there is very little difference between the RMSEs for the fitted linear and nonlinear model forecasts, as they differ by less than 1% for each country. Not surprisingly, neither the *MSE-T* nor *ENC-T* statistic is significant at conventional levels for any country, so that the out-of-sample forecasts generated by the Band-TAR model are not significantly superior to those generated by a linear AR(1) model.

The Hamilton (2001) and Dahl and González-Rivera (2002) test results, also reported in Table 1, show that the null hypothesis of linearity cannot be rejected against the alternative of nonlinearity at conventional significance levels for any statistic for any country, although the g_{OP} statistic is nearly significant at the 10% level for the U.K.[23] Recall from Section 2.1 that Obstfeld and Taylor (1997) are unable to reject the null hypothesis of linearity against a TAR alternative using the Tsay (1989) test for Canada, Italy, U.K, and Japan, while they can reject the null hypothesis for France and Germany. Our results for the flexible parametric tests of Hamilton (2001) and Dahl and González-Rivera (2002) reported in Table 1 provide further evidence against nonlinear real exchange rate behavior for Canada, Italy, U.K., and Japan, and they indicate that the evidence of nonlinear behavior for France and Germany reported in Obstfeld and Taylor (1997) is not particularly robust.[24]

Whether comparing conditional expectation functions or out-of-sample forecasts for Band-TAR and linear AR models or testing for nonlinearity using the Hamilton (2001) and Dahl and González-Rivera (2002) tests, we find little reason in Table 1 to favor a nonlinear specification over a linear AR specification using the data in Obstfeld and Taylor (1997).[25]

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

Table 2 reports results for our re-examination of the ESTAR models in TPS. Following TPS, our monthly data for U.S. dollar real exchange rates with respect to the U.K., Germany, France, and Japan begin in 1973:01 and end in 1996:12. We compute real exchange rates using consumer price indices and nominal exchange rates from the International Monetary Fund's *International Financial Statistics* CD-ROM (September 2001), and, following TPS, we normalize the log-level of the real exchange rate to be equal to zero in 1973:01. Again following TPS, we estimate the ESTAR model, equation (4), for each country using multivariate nonlinear least squares (nonlinear seemingly unrelated regressions). After allowing for

[23] The *p*-values reported in Table 1 were generated using the bootstrap procedure in Dahl and González-Rivera (2002). We thank Christian Dahl for providing us with a GAUSS program used to calculate the Hamilton (2001) and Dahl and González-Rivera (2002) statistics and *p*-values. Note that we found relatively little evidence of ARCH in the data, so that the Dahl and González-Rivera (2002) bootstrap procedure, which does not allow for ARCH in the residuals, should not be adversely affected.

[24] It is worth noting that the λ_{OP}^{A} statistic is insignificant for every country in Table 1, since, as noted in Section 3.4 above, Dahl and González-Rivera (2002) find this statistic to be particularly powerful against TAR alternatives.

[25] It should be noted that Obstfeld and Taylor (1997) believe that the Band-TAR model specification is more appropriate for real exchange rates based on disaggregated price indices. We discuss this further in the conclusion.

the single lag, our usable sample begins in 1973:02. Table 2 reports an estimated linear AR(1) model, in addition to the estimated ESTAR model, for each country. The estimated parameters for the ESTAR models reported in Table 2 are very close to those reported in TPS (see their Table 3).

Examining the parametric encompassing test results in Table 2, we see that the *t*-statistics and χ^2-statistic are insignificant for the U.K., Germany, and France. For Japan, the variance implied by the fitted ESTAR model is significantly less than the sample variance according to the *t*-statistic (and we reject the null hypothesis that the estimated nonlinear model jointly encompasses the sample variance and first-order autocorrelation at the 10% level). This calls into question the adequacy of the estimated ESTAR model for Japan.[26]

Figure 2 plots the conditional expectation function for q_t given q_{t-1} for the fitted linear AR(1) model and the fitted ESTAR model over the range of q_{t-1} values in the sample, as well as a scatterplot of the data, for each country. As was the case for the fitted Band-TAR models in Figure 1, there is little visual evidence of important differences between the fitted linear AR model and the fitted ESTAR model over the range of q_{t-1} values in the sample for each country. According to the in-sample forecast comparisons in Table 2, there is again little difference between the in-sample RMSEs for the fitted linear and nonlinear AR models, as all are less than 0.20%.

When we compare out-of-sample forecasts over the 1992:01-1996:12 period, there is also little difference between the forecasts generated by the linear AR model and those generated by the ESTAR model for the U.K., Germany, and France in terms of RMSE, and neither the *MSE-T* nor *ENC-T* statistic indicates that the out-of-sample forecasts generated by the ESTAR model are superior to those generated by the linear AR model for these countries. For Japan, the out-of-sample forecast RMSE for the ESTAR model

[26] It is interesting to note that, in TPS, the residuals from the fitted ESTAR model for Japan pass a battery of standard diagnostic tests for remaining serial correlation and nonlinearity. We also used a Monte Carlo simulation to generate finite-sample critical values for the *t*-statistic corresponding to the variance for Japan, and the *t*-statistic remains significant (at the 10% level). Note that it is unnecessary to simulate critical values for the other *t*-statistics in Table 2, as these *t*-statistics are all insignificant, and, if anything, the finite-sample critical values will increase (in absolute value).

is almost 13% *higher* than the out-of-sample forecast RMSE for the linear AR model, and, not surprisingly, the *MSE-T* and *ENC-T* statistics do not indicate that the ESTAR model out-of-sample forecasts are superior to the linear AR model forecasts.

We do not find support for nonlinear behavior in real exchange rates when we use the Hamilton (2001) and Dahl and González-Rivera (2002) tests, as none of the statistics for any of the four countries can reject the null hypothesis of a linear AR model. It is noteworthy that the null hypothesis cannot be rejected using the λ_{OP}^{A} statistic, which Dahl and González-Rivera (2002) find to be particularly powerful against ESTAR alternatives. Overall, there appears to be little to recommend an ESTAR specification over a simple linear AR specification in Table 2 for the TPS data.

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

We next re-examine the MNP ESTAR model for the U.K.-U.S. real exchange rate estimated using a long span of data. As in MNP, we use 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 1793-1990 period.[27] Following MNP, we estimate equation (6) using nonlinear least squares. After allowing for two lags, the usable sample covers 1793-1990. We report the estimated ESTAR model, along with an estimated linear AR(2) model, in Table 3. The estimated ESTAR model reported in Table 3 is very similar to that in MNP.

Looking first at the parametric encompassing test results in Table 3, we see that the fitted ESTAR model is able to encompass the sample variance and first-order autocorrelation (individually and jointly). In Figure 3, there are apparent differences between the conditional expectation functions for the fitted ESTAR model and a fitted linear AR model, especially near the tails of the sample, where the fitted ESTAR model appears to track the data better. Turning back to Table 3, we see that the in-sample fit of the estimated

[27] 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).

ESTAR model represents an improvement of almost 2% in terms of RMSE over the fitted linear AR(2) model.

Over the 1961-1990 out-of-sample period, the forecast RMSE for the ESTAR model is 3.38% lower than the forecast RMSE for the linear AR(2) model. The *MSE-T* statistic is significant at the 10% level (recall that this is a one-tailed test), indicating that the ESTAR model forecast MSE is significantly lower than the linear AR(2) model forecast MSE. In addition, the *ENC-T* statistic is significant at the 5% level, so that the ESTAR model contains information useful in predicting real exchange rates beyond that already contained in the linear AR(2) model.

Finally, we can reject the null hypothesis of linearity according to the Dahl and González-Rivera (2002) λ_{OP}^{A} statistic at the 10% significance level, although we cannot reject it using the other statistics (but the g_{OP} statistic is nearly significant at the 10% level). Recall from Section 3.4 that Dahl and González-Rivera (2002) find the λ_{OP}^{A} statistic to be especially powerful in detecting ESTAR alternatives in Monte Carlo simulations. The significant λ_{OP}^{A} statistic reported in Table 3 shows that MNP's rejection of the linear AR specification against an ESTAR alternative using the Teräsvirta (1994) test is reasonably robust.

In general, the results reported in Table 3 lend support to the ESTAR specification for the U.K.-U.S. real exchange rate proposed by MNP that is estimated using over two centuries of data.

4.4. Bergman and Hansson (2000) MS-AR(1) Model

Our final re-examination is for the MS-AR(1) models in Bergman and Hansson (2000). Following Bergman and Hansson (2000), we use quarterly data from 1973:1-1997:4 for U.S. dollar real exchange rates with respect to the U.K., France, Germany, Switzerland, Canada, and Japan based on consumer price indices.[28] Again following Bergman and Hansson (2000), we normalize the real exchange rate to unity in 1973:2 (using 100 times the log-level of the real exchange rate) and estimate the MS-AR(1) model for

---

[28] We thank Michael Bergman and Jesper Hansson for generously providing the data from Bergman and Hansson (2000).

each country via maximum likelihood using data through 1990:4. After allowing for a single lag, the usable sample starts in 1973:2. (The 1991:1-1997:4 period is reserved for out-of-sample forecasting, as in Bergman and Hansson, 2000.) Table 4 reports the estimated MS-AR(1) model for each country, and the estimated models in Table 4 exactly match those in Bergman and Hansson (2000).[29]

Looking at the parametric encompassing test results in Table 4, we see that the variance and first-order autocorrelation implied by the fitted nonlinear model encompass the sample variance and first-order autocorrelation (individually and jointly) for France, Germany, Canada, and Japan. However, the variance implied by the fitted MS-AR(1) model does not encompass the sample variance for the U.K. and Switzerland, as the variance implied by the fitted nonlinear model is significantly greater than the sample variance for both countries. This indicates a possible specification problem with the fitted MS-AR(1) model for the U.K. and Switzerland.

The conditional expectation functions for q_t given q_{t-1} are plotted for a fitted linear AR(1) model and the fitted MS-AR(1) model for each country in Figure 4. While important differences between the conditional expectation functions for the linear and nonlinear models are not necessarily apparent in Figure 4, this diagnostic tool is probably less useful for the Bergman and Hansson (2000) restricted MS-AR(1) model specification, which has a state-independent AR coefficient (recall that the intercept alone follows a two-state Markov process in their preferred specification). Looking at the in-sample RMSE for the fitted linear and nonlinear models in Table 4, we see that the fitted MS-AR(1) model has a substantially lower in-sample RMSE than the fitted linear AR(1) model for each country. For the U.K., France, Germany, and Switzerland, the in-sample RMSE for the fitted MS-AR(1) model is over 35% lower than the in-sample RMSE for the fitted linear AR(1) model. However, recall that the fitted MS-AR(1) models for the U.K. and Switzerland fail to encompass the sample variance.

[29] Our estimated models exactly match those in Bergman and Hansson (2000), as we use their data and the same GAUSS program to implement maximum likelihood estimation. The GAUSS program is authored by Anders Warne and is available at http://www.warne.texlips.org/.

Comparing out-of-sample forecasts over 1991:1-1997:4, the MS-AR(1) model forecasts are not significantly superior to those of the linear AR(1) model for the U.K. and Switzerland according the *MSE-T* and *ENC-T* statistics. For France, Germany, Canada, and Japan, the *ENC-T* statistic is significant, indicating that the fitted MS-AR(1) model contains predictive information beyond that already contained in the fitted linear AR(1) model. In addition, the *MSE-T* statistic is significant for Canada. It is interesting to note that the two countries for which we fail to find significant out-of-sample predictive ability, the U.K. and Switzerland, are also the two countries for which the fitted MS-AR(1) model fails to encompass the sample variance, so that the parametric encompassing tests and out-of-sample forecast comparisons both point to potential specification problems for the U.K. and Switzerland.

Finally, Table 4 also reports Hamilton (2001) and Dahl and González-Rivera (2002) nonlinearity test results.[30] All four statistics are insignificant for the U.K., France, Germany, Canada, and Japan, so that the evidence of nonlinear real exchange rate behavior in Bergmann and Hansonn (2000), who use the Hansen (1992, 1996) test to test a linear AR specification against their MS-AR(1) alternative, does not appear to be particularly robust for these countries.[31] The λ_{OP}^{E} and g_{OP} statistics are significant for Switzerland, providing evidence of nonlinear real exchange rate behavior for Switzerland. However, as discussed above, the parametric encompassing and out-of-sample forecasting tests do not support the MS-AR(1) specification for Switzerland, so an alternative nonlinear specification is apparently required for Switzerland.

Overall, the out-of-sample forecasting tests provide support for the MS-AR(1) model over a linear AR(1) model for France, Germany, Canada, and Japan, while the evidence strongly suggests that the MS-AR(1) model is misspecified for the U.K. and Switzerland. As we noted in the introduction, the Bergmann and Hansonn (2000) MS-AR(1) specification is not readily suggested by theoretical models

[30] The tests are applied over the 1973:2-1990:4 in-sample period, as Bergman and Hansson (2000) use this sample when applying the Hansen (1992, 1996) test.

[31] While Dahl and González-Rivera (2002) consider a large number of nonlinear model specifications in their Monte Carlo experiments, they do not consider MS-AR specifications. Thus, it may be the case that the Hamilton (2001) and Dahl and González-Rivera (2002) nonlinearity tests are not particularly powerful against MS-AR alternatives.

incorporating transaction costs. It would seem to be more compatible with models that allow for infrequent real shocks or changes in investor sentiment that alter the long-run equilibrium real exchange rate.

5. Conclusion

In this paper, we re-examine nonlinear models of U.S. dollar real exchange rate behavior from the extant empirical literature. Our re-examination employs parametric encompassing tests of sample moments and graphical analysis of conditional expectation functions (Pagan, 2002), model selection based on out-of-sample forecasting, and flexible parametric nonlinearity tests (Hamilton, 2001; Dahl and González-Rivera, 2002). These econometric tests allow us to move beyond the standard analysis of nonlinear models used in extant studies, so that we can better assess the extent of the statistical support for nonlinear models of real exchange rate behavior. When we employ these tests, we find relatively little support for a number of fitted nonlinear real exchange rate models appearing in the extant literature. According to these tests, the Obstfeld and Taylor (1997) Band-TAR models and the Taylor, Peel, and Sarno (2001) ESTAR models differ little from simple linear AR specifications when estimated using data from the post-Bretton Woods period. The Michael, Nobay, and Peel (1997) ESTAR model for the U.K.-U.S. real exchange rate, which is estimated using over two centuries of data, receives considerably more support from the econometric tests we employ. Bergman and Hansson (2000) MS-AR(1) models estimated for the U.K. and Switzerland over the modern float have difficulty matching basic characteristics of the data, pointing to likely model misspecification, while estimated MS-AR(1) models for France, Germany, Canada, and Japan are favored over liner AR models according to out-of-sample forecast. Overall, we find rather limited support for nonlinear models of U.S. dollar real exchange rate behavior from the extant literature, especially for nonlinear models estimated using data from the post-Bretton Woods period.

While we fail to find strong statistical support for a number of nonlinear models from the extant literature, we think it is premature to dismiss nonlinear models of U.S. dollar real exchange rate behavior for two reasons. First, theoretical models based on transaction costs make an almost compelling case for nonlinear Band-TAR and ESTAR specifications of real exchange rate behavior.[32] Second, simulations and impulse response analysis in Obstfeld and Taylor (1997), Taylor, Peel, and Sarno (2001), and Bergman and Hansson (2000) show that nonlinear real exchange rate models provide a potential explanation for the PPP puzzle. Nevertheless, the fact that many fitted real exchange rate models from the extant literature do not receive strong statistical support over simple linear models indicates that further research is needed in order to establish the empirical relevance of nonlinear real exchange rate behavior.

Given that we find more support for the ESTAR model in Michael, Nobay, and Peel (1997), which is estimated using over two centuries of data for the U.K.-U.S. real exchange rate, the post-Bretton Woods era may simply be too brief for strong evidence of nonlinear real exchange rate behavior to manifest itself. Further research using long spans of real exchange rate data may therefore be useful in assessing the empirical relevance of nonlinear real exchange rate behavior.[33] In addition, there may be stronger statistical evidence of nonlinear behavior in real exchange rates based on more disaggregated price indices. For example, Sarno, Taylor, and Chowdhury (2002) estimate Band-TAR models using data for five major bilateral U.S. dollar exchange rates and price indices for nine goods sectors during the modern float. They find significant evidence of nonlinear mean-reversion in deviations from the law of one price.[34] It would be interesting to see if this finding is supported by the econometric tests used in the present paper. Finally, theoretical models developed in Goswami, Shrikhande, and Wu (2002) suggest

[32] It is thus somewhat ironic that the Bergmann and Hansonn (2000) MS-AR(1) specification, which is not necessarily implied by theoretical models based on transaction costs, receives more support than Band-TAR and ESTAR specifications for most countries over the post-Bretton Woods period.
[33] Of course, a drawback to the use of long spans of data is that it potentially combines data from different regimes (Mussa, 1986).
[34] 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.

that univariate nonlinear models of real exchange rate behavior are inadequate, as the dynamics of the real exchange rate depend critically on a state variable such as the capital stock or trade balance. Mutlivariate nonlinear models may thus be better able to detect nonlinear real exchange rate behavior than univariate nonlinear AR models. Further investigation in these areas is warranted, as the jury is still out on the empirical relevance of nonlinear models of U.S. dollar real exchange rate behavior.

References

Balke, N.S., Fomby, T.B., 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.

Bergman, U.M. and J. Hansson, 2000. Real Exchange Rates and Switching Regimes. *Journal of International Money and Finance*, forthcoming.

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).

Breuer, J., 1994. An Assessment of the Evidence on Purchasing Power Parity. In J. Williamson, ed., *Estimating Equilibrium Exchange Rates* (Washington, D.C.: Institute for International Economics).

Breuer, J., R. McNown, and M. Wallace, 2001. Misleading Inference from Panel Unit Root Tests with an Illustration from Purchasing Power Parity. *Review of International Economics* 9, 482-493.

Breunig, R., S. Najarian, and A. Pagan, 2002. Specification Testing of Markov Switching Models. Manuscript, Australian National University.

Cheung, Y.-W. and K.S. Lai, 1994. Mean Reversion in Real Exchange Rates. *Economic Letters* 46, 251-256.

Clark, T.E. and M.W. McCracken, 2001. Tests of Equal Forecast Accuracy and Encompassing for Nested Models. *Journal of Econometrics* 105, 85-110.

Clements, M.P. and D.F. Hendry, 1998. *Forecasting Economic Time Series*. Cambridge University Press, Cambridge, U.K.

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.

Dahl, C.M., 2002. An Investigation of Tests For Linearity and the Accuracy of Likelihood Based Inference Using Random Fields. *Econometrics Journal* 1, 1-25.

Dahl, C.M. and G. González-Rivera. 2002. Testing For Neglected Nonlinearity in Regression Models Based on the Theory of Random Fields. *Journal of Econometrics*, forthcoming.

Diebold, F.X., S. Husted, and M. Rush, 1991. Real Exchange Rates under the Gold Standard. *Journal of Political Economy* 99, 1252-1271.

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

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

Dumas, B., 1994. Partial Equilibrium Versus General Equilibrium Models of the International Capital Market. In E Van Der Ploeg, ed., *The Handbook of International Macroeconomics* (Oxford: Blackwell).

Engle, C. 2000. Long-Run PPP May Not Hold After All. *Journal of International Economics* 51, 243-273.

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

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

Flood, R.P. and M.P. Taylor, 1996. Exchange Rate Economics: What's Wrong with the Conventional Macro Approach? In J. A. Frankel, G. Galli, and A. Giovannini, eds., *The Microstructure of Foreign Exchange Markets* (Chicago: Chicago University Press).

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

Froot, K.A., and K. Rogoff, 1995. Perspectives on PPP and Long-run Real Exchange Rates. In Rogoff, K. and G. Grossman, eds., *Handbook of International Economics* (Amsterdam: North-Holland).

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

Glen, J.D., 1992. Real Exchange Rates in the Short, Medium, and Long Run. *Journal of International Economics* 33, 147-166.

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

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

Hamilton, J.D., 1996. Specification Testing in Markov-Switching Time-Series Models. *Journal of Econometrics* 70, 127-157.

Hamilton, J.D., 2001. A Parametric Approach To Flexible Nonlinear Inference. *Econometrica*, 69, 537-573.

Hansen, B.E., 1992. The Likelihood Ratio Test Under Non-Standard Conditions: Testing the Markov Switching Model of GNP, *Journal of Applied Econometrics* 7, S61-S82.

Hansen, B.E., 1996. Erratum: The Likelihood Ratio Test Under Non-Standard Conditions: Testing the Markov Switching Model of GNP, *Journal of Applied Econometrics* 11, 195-198.

Harvey, D.I., S.J. Leybourne, and P. Newbold, 1998. Tests for Forecast Encompassing. *Journal of Business and Economic Statistics* 16, 254-259.

Hausman, J.A., 1978. Specification Tests in Econometrics. *Econometrica* 46, 1251-1271.

Heckscher, E.F., 1916. Vaxelkursens Grundval vid Pappersmyntfot. Ekono-misk Tidskrift. 18, 309-312.

Hess, G.D. and S. Iwata, 1997. Measuring and Comparing Business-Cycle Features, *Journal of Business and Economics Statistics* 15, 432-444.

Liu, Y., and W. Enders, 2002. Out-of-Sample Forecasts and Non-linear Model Selection With an Example of the Term-Structure of Interest Rates. *Southern Economics Journal*, forthcoming.

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., 2000. Asymptotics for Out-of-Sample Tests of Causality. Manuscript, University of Missouri at Columbia.

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

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.

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* 10, 129-162.

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).

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., 1998a. The Overvaluation of Purchasing Power Parity. *Journal of International Economics* 44, 1-19.

O'Connell, P.G.J., 1998b. 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.

Oh, K.-Y., 1996. Purchasing Power Parity and Unit Root Tests Using Panel Data. *Journal of International Money and Finance* 15, 405-418.

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.

Papell, D.H., 1997. Searching for Stationarity: Purchasing Power Parity Under the Current Float. *Journal of International Economics* 43, 313-332.

Potter, S. and G. Koop, 2000. Nonlinearity, Structural Breaks or Outliers in Economic Time Series? In William Barnett et al., eds., *Nonlinearity and Econometrics* (Cambridge: Cambridge University Press).

Potter, S. and G. Koop. 2001. Are Apparent Findings of Nonlinearity Due to Structural Instability in Economic Time Series? *Econometrics Journal* 4, 37-55.

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. and M.P. Taylor, 2002. Purchasing Power Parity and the Real Exchange Rate. *International Monetary Fund Staff Papers* 49, 65-105.

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.

Silvermann, B., 1986. *Density Estimation For Statistics and Data Analysis* (London: Chapman and Hall).

Taylor, A.M., 2001. Potential Pitfalls for the Purchasing-Power Parity Puzzle: Sampling and Specification Biases in Mean-Reversion Tests of the Law of One Price. *Econometrica* 69, 473-498.

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

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

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.

Taylor, M.P. and L. Sarno, 1998. The Behavior of Real Exchange Rates During the post-Bretton Woods Period. *Journal of International Economics* 46, 281-312.

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.

West, K.D., 1996. Asymptotic inference about predictive ability. *Econometrica* 64, 1067-1084.

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

Wu, Y, 1996. Are Real Exchange Rates Non-Stationary?: Evidence from a Panel-Data Test. *Journal of Money, Credit and Banking* 28, 54-63.

Wu, J.-L. and S. Wu, 2001. Is Purchasing Power Parity Overvalued? *Journal of Money, Credit, and Banking* 33, 804-812.

Table 1: Test results, Obstfeld and Taylor (1997) Band-TAR model

A. Canada

Estimated linear AR(1) model:

$\Delta\hat{q}_t = -0.0065 q_{t-1}$
$\hat{\sigma} = 0.0109$

Estimated Band-TAR model:

$\Delta q_t = -0.0460(q_{t-1} - 0.0715)$ if $q_{t-1} > 0.0715$
$\Delta q_t = -0.0460(q_{t-1} + 0.0715)$ if $-0.0715 > q_{t-1}$
$\hat{\sigma}^{out} = 0.0098$; $\hat{\sigma}^{in} = 0.0116$

Parametric encompassing test, Band-TAR model:

Sample var. = 0.005; Band-TAR var. = 0.0056: *t*-stat. = –0.636
Sample autocor.= 0.980; Band-TAR autocor. = 0.989; *t*-stat. = –0.032
Joint χ^2-stat. = 0.407

In-sample forecast comparison, 1980:02-1994:11:

Linear AR(1) model RMSE = 0.0109
Band-TAR model RMSE = 0.0109
% improvement = 0.2445

Out-of-sample forecast comparison, 1991:01-1994:11:

Linear AR(1) model RMSE = 0.0120
Band-TAR model RMSE = 0.0120
% improvement = 0.0152
MSE-T = 0.061
ENC-T = 0.152

Hamilton (2001), Dahl and González-Rivera (2002) nonlinearity tests:

$\lambda_H^E = 0.459$ [*p*-value = 0.429]
$\lambda_{OP}^E = 2.706$ [*p*-value = 0.242]
$\lambda_{OP}^A = 2.464$ [*p*-value = 0.433]
$g_{OP} = 1.317$ [*p*-value = 0.252]

B. France

Estimated linear AR(1) model:

$\Delta\hat{q}_t = -0.0214 q_{t-1}$
$\hat{\sigma} = 0.0293$

Estimated Band-TAR model:

$\Delta q_t = -0.0510(q_{t-1} - 0.1910)$ if $q_{t-1} > 0.1910$
$\Delta q_t = -0.0510(q_{t-1} + 0.1910)$ if $-0.1910 > q_{t-1}$
$\hat{\sigma}^{out} = 0.0332$; $\hat{\sigma}^{in} = 0.0278$

Parametric encompassing test, Band-TAR model:

Sample var. = 0.029; Band-TAR var. = 0.032; *t*-stat. = –0.291
Sample autocor. =0.975; Band-TAR autocor. =0.986; *t*-stat. = –0.041
Joint χ^2-stat. = 0.098

In-sample forecast comparison, 1980:02-1994:11:

Linear AR(1) model RMSE = 0.0293
Band-TAR model RMSE = 0.0294
% improvement = –0.3027

Out-of-sample forecast comparison, 1991:01-1994:11:

Linear AR(1) model RMSE = 0.0277
Band-TAR model RMSE = 0.0277
% improvement = 0.1342
MSE-T = 1.187
ENC-T = 1.222

Hamilton (2001), Dahl and González-Rivera (2002) nonlinearity tests:

$\lambda_H^E = 0.001$ [*p*-value = 0.978]
$\lambda_{OP}^E = 1.207$ [*p*-value = 0.524]
$\lambda_{OP}^A = 1.574$ [*p*-value = 0.896]
$g_{OP} = 1.245$ [*p*-value = 0.648]

Table 1 (continued)

C. Germany

Estimated linear AR(1) model:

$$\Delta\hat{q}_t = -0.0205 q_{t-1}$$
$$\hat{\sigma} = 0.0295$$

Estimated Band-TAR model:

$$\Delta q_t = -0.0476(q_{t-1} - 0.1736) \text{ if } q_{t-1} > 0.1736$$
$$\Delta q_t = -0.0476(q_{t-1} + 0.1736) \text{ if } -0.1736 > q_{t-1}$$
$$\hat{\sigma}^{out} = 0.0340 \text{ ; } \hat{\sigma}^{in} = 0.0275$$

Parametric encompassing test, Band-TAR model:

Sample var. = 0.030; Band-TAR var.= 0.029 : *t*-stat. = 0.043
Sample autocor.= 0.974; Band-TAR autocor. =0.984; *t*-stat. = –0.039
Joint χ^2-stat. = 0.004

In-sample forecast comparison, 1980:02-1994:12:

Linear AR(1) model RMSE = 0.0295
Band-TAR model RMSE = 0.0296
% improvement = –0.1853

Out-of-sample forecast comparison, 1991:01-1994:12:

Linear AR(1) model RMSE = 0.0279
Band-TAR model RMSE = 0.0279
% improvement = –0.1402
MSE-T = –0.110
ENC-T = 0.180

Hamilton (2001), Dahl and González-Rivera (2002) nonlinearity tests:

λ_H^E = 0.003 [*p*-value = 0.959]
λ_{OP}^E = 0.993 [*p*-value = 0.584]
λ_{OP}^A = 1.561 [*p*-value = 0.947]
g_{OP} = 1.350 [*p*-value = 0.571]

D. Italy

Estimated linear AR(1) model:

$$\Delta\hat{q}_t = -0.0140 q_{t-1}$$
$$\hat{\sigma} = 0.0292$$

Estimated Band-TAR model:

$$\Delta q_t = -0.0947(q_{t-1} - 0.2268) \text{ if } q_{t-1} > 0.2268$$
$$\Delta q_t = -0.0947(q_{t-1} + 0.2268) \text{ if } -0.2268 > q_{t-1}$$
$$\hat{\sigma}^{out} = 0.0360 \text{ ; } \hat{\sigma}^{in} = 0.0264$$

Parametric encompassing test, Band-TAR model:

Sample var. = 0.035; Band-TAR var. = 0.031; *t*-stat. = 0.440
Sample autocor. = 0.986; Band-TAR autocor. = 0.987; *t*-stat. = –0.004
Joint χ^2-stat. = 0.194

In-sample forecast comparison, 1980:02-1994:10:

Linear AR(1) model RMSE = 0.0292
Band-TAR model RMSE = 0.0291
% improvement = 0.4425

Out-of-sample forecast comparison, 1991:01-1994:10:

Linear AR(1) model RMSE = 0.0343
Band-TAR model RMSE = 0.0340
% improvement = 0.8354
MSE-T = 1.110
ENC-T = 1.164

Hamilton (2001), Dahl and González-Rivera (2002) nonlinearity tests:

λ_H^E = 0.756 [*p*-value = 0.301]
λ_{OP}^E = 2.943 [*p*-value = 0.198]
λ_{OP}^A = 2.598 [*p*-value = 0.327]
g_{OP} = 1.387 [*p*-value = 0.181]

Table 1 (continued)

E. U.K.

Estimated linear AR(1) model:

$\Delta\hat{q}_t = -0.0227 q_{t-1}$
$\hat{\sigma} = 0.0303$

Estimated Band-TAR model:

$\Delta q_t = -0.0741(q_{t-1} - 0.1991)$ if $q_{t-1} > 0.1991$
$\Delta q_t = -0.0741(q_{t-1} + 0.1991)$ if $-0.1991 > q_{t-1}$
$\hat{\sigma}^{out} = 0.0388$; $\hat{\sigma}^{in} = 0.0273$

Parametric encompassing test, Band-TAR model:

Sample var. = 0.025; Band-TAR var. = 0.030; *t*-stat. = –0.697
Sample autocor. = 0.977; Band-TAR autocor. = 0.984; *t*-stat. = –0.027
Joint χ^2-stat. = 0.486

In-sample forecast comparison, 1980:02-1994:11:

Linear AR(1) model RMSE = 0.0303
Band-TAR model RMSE = 0.0303
% improvement = –0.1336

Out-of-sample forecast comparison, 1991:01-1994:11:

Linear AR(1) model RMSE = 0.0321
Band-TAR model RMSE = 0.0323
% improvement = –0.7277
MSE-T = –0.944
ENC-T = –0.838

Hamilton (2001), Dahl and González-Rivera (2002) nonlinearity tests:

$\lambda_H^E = 0.620$ [*p*-value = 0.372]
$\lambda_{OP}^E = 2.845$ [*p*-value = 0.174]
$\lambda_{OP}^A = 2.436$ [*p*-value = 0.338]
$g_{OP} = 1.429$ [*p*-value = 0.107]

F. Japan

Estimated linear AR(1) model:

$\Delta\hat{q}_t = -0.0026 q_{t-1}$
$\hat{\sigma} = 0.0302$

Estimated Band-TAR model:

$\Delta q_t = -0.1331(q_{t-1} - 0.3130)$ if $q_{t-1} > 0.3130$
$\Delta q_t = -0.1331(q_{t-1} + 0.3130)$ if $-0.3130 > q_{t-1}$
$\hat{\sigma}^{out} = 0.0229$; $\hat{\sigma}^{in} = 0.0315$

Parametric encompassing test, Band-TAR model:

Sample var. = 0.049; Band-TAR var. = 0.054; *t*-stat. = –0.473
Sample autocor. = 0.981; Band-TAR autocor. = 0.992; *t*-stat. = –0.040
Joint χ^2-stat. = 0.224

In-sample forecast comparison, 1980:02-1994:11:

Linear AR(1) model RMSE = 0.0302
Band-TAR model RMSE = 0.0301
% improvement = 0.2698

Out-of-sample forecast comparison, 1991:01-1994:11:

Linear AR(1) model RMSE = 0.0225
Band-TAR model RMSE = 0.0225
% improvement = 0.0570
MSE-T = 0.038
ENC-T =0.527

Hamilton (2001), Dahl and González-Rivera (2002) nonlinearity tests:

$\lambda_H^E = 0.243$ [*p*-value = 0.596]
$\lambda_{OP}^E = 2.274$ [*p*-value = 0.376]
$\lambda_{OP}^A = 2.630$ [*p*-value = 0.396]
$g_{OP} = 1.372$ [*p*-value = 0.251]

Table 2: Test results, Taylor, Peel, and Sarno (2001) ESTAR model

A. U.K.

Estimated linear AR(1) model:

$\hat{q}_t = 0.005 + 0.970 q_{t-1}$
$\hat{\sigma} = 0.033$

Estimated ESTAR model:

$\hat{q}_t = q_{t-1} - \{1 - \exp[-0.448(q_{t-1} - 0.150)^2]\}(q_{t-1} - 0.150)$
$\hat{\sigma} = 0.033$

Parametric encompassing test, ESTAR model:

Sample var. = 0.019; ESTAR var. = 0.023; t-stat. = –0.864
Sample autocor. = 0.968; ESTAR autocor. = 0.976; t-stat. = –0.041
Joint χ^2-stat. = 0.749

In-sample forecast comparison, 1973:02-1996:12:

Linear AR(1) model RMSE = 0.0333
ESTAR model RMSE = 0.0333
% improvement = –0.145

Out-of-sample forecast comparison, 1992:01-1996:12:

Linear AR(1) model RMSE = 0.031
ESTAR model RMSE = 0.031
% improvement = –0.463
MSE-T = –0.900
ENC-T = –0.792

Hamilton (2001), Dahl and González-Rivera (2002) nonlinearity tests:

λ_H^E = 0.237 [p-value = 0.637]
λ_{OP}^E = 1.717 [p-value = 0.392]
λ_{OP}^A = 1.654 [p-value = 0.721]
g_{OP} = 1.221 [p-value = 0.546]

B. Germany

Estimated linear AR(1) model:

$\hat{q}_t = 0.003 + 0.978 q_{t-1}$
$\hat{\sigma}$ =0.035

Estimated ESTAR model:

$\hat{q}_t = q_{t-1} - \{1 - \exp[-0.263(q_{t-1} + 0.007)^2]\}(q_{t-1} + 0.007)$
$\hat{\sigma} = 0.034$

Parametric encompassing test, ESTAR model:

Sample var. = 0.028; ESTAR mean = 0.031; t-stat. = –0.406
Sample autocor. = 0.977; ESTAR autocor. = 0.981; t-stat. = –0.016
Joint χ^2-stat. = 0.165

In-sample forecast comparison, 1973:02-1996:12:

Linear AR(1) model RMSE = 0.0345
ESTAR model RMSE = 0.0345
% improvement = 0.035

Out-of-sample forecast comparison, 1992:01-1996:12:

Linear AR(1) model RMSE = 0.029
ESTAR model RMSE = 0.029
% improvement = 0.416
MSE-T = 0.277
ENC-T = 0.626

Hamilton (2001), Dahl and González-Rivera (2002) nonlinearity tests:

λ_H^E = 0.265 [p-value = 0.579]
λ_{OP}^E = 0.479 [p-value = 0.719]
λ_{OP}^A = 2.680 [p-value = 0.190]
g_{OP} = 1.199 [p-value = 0.846]

Table 2 (continued)

C. France

Estimated linear AR(1) model:

$\hat{q}_t = 0.003 + 0.977 q_{t-1}$
$\hat{\sigma} = 0.033$

Estimated ESTAR model:

$\hat{q}_{t-1} = q_{t-1} - \{1 - \exp[-0.287(q_{t-1} - 0.049)^2]\}(q_{t-1} - 0.049)$
$\hat{\sigma} = 0.033$

Parametric encompassing test, ESTAR model:

Sample var. = 0.024; ESTAR var. = 0.029; *t*-stat. = –0.660
Sample autocor. = 0.976 ; ESTAR autocor. = 0.981; *t*-stat. = –0.026
Joint χ^2-stat. = 0.436

In-sample forecast comparison, 1973:02-1996:12:

Linear AR(1) model RMSE = 0.033
ESTAR model RMSE = 0.033
% improvement = –0.186

Out-of-sample forecast comparison, 1992:01-1996:12:

Linear AR(1) model RMSE = 0.029
ESTAR model RMSE = 0.029
% improvement = 0.398
MSE-T = 0.796
ENC-T = 0.847

Hamilton (2001), Dahl and González-Rivera (2002) nonlinearity tests:

λ_H^E = 0.229 [*p*-value = 0.636]
λ_{OP}^E = 0.606 [*p*-value = 0.712]
λ_{OP}^A = 2.753 [*p*-value = 0.169]
g_{OP} = 1.629 [*p*-value = 0.367]

D. Japan

Estimated linear(1) AR model:

$\hat{q}_t = 0.009 + 0.985 q_{t-1}$
$\hat{\sigma} = 0.033$

Estimated ESTAR model:

$\hat{q}_{t-1} = q_{t-1} - \{1 - \exp[-0.164(q_{t-1} - 0.513)^2]\}(q_{t-1} - 0.513)$
$\hat{\sigma} = 0.033$

Parametric encompassing test, ESTAR model:

Sample var. = 0.058; ESTAR var. = 0.037; *t*-stat. = 2.271
Sample autocor. = 0.983; ESTAR autocor.= 0.985; *t*-stat. = –0.006
Joint χ^2-stat. = 5.158

In-sample forecast comparison, 1973:02-1996:12:

Linear AR(1) model RMSE = 0.0335
ESTAR model RMSE = 0.0334
% improvement = 0.195

Out-of-sample forecast comparison, 1992:01-1996:12:

Linear AR(1) model RMSE = 0.032
ESTAR model RMSE = 0.035
% improvement = –12.70
MSE-T = –1.472
ENC-T = –0.006

Hamilton (2001), Dahl and González-Rivera (2002) nonlinearity tests:

λ_H^E = 0.336 [*p*-value = 0.515]
λ_{OP}^E = 0.484 [*p*-value = 0.818]
λ_{OP}^A = 1.755 [*p*-value = 0.905]
g_{OP} = 1.181 [*p*-value = 0.728]

Table 3: Test results, Michael, Nobay, and Peel (1997) ESTAR model

U.K.

Estimated linear AR(2) model:

$\hat{q}_t = 0.193 + 0.997 q_{t-1} - 0.119 q_{t-2}$
$\hat{\sigma}$ = 0.070

Estimated ESTAR model:

$\hat{q}_t - 1.567 = \{\exp[-2.437(q_{t-1} - 1.567)^2]\}[1.182(q_{t-1} - 1.567) - 0.182(q_{t-2} - 1.567)]$
$\hat{\sigma}$ = 0.069

Parametric encompassing test, ESTAR model:

Sample var. = 0.024; ESTAR var. = 0.024; *t*-stat. = 0.091
Sample autocor. = 0.882; ESTAR autocor. = 0.889; *t*-stat. = –0.049
Joint χ^2-stat. = 0.011

In-sample forecast comparison, 1793-1990:

Linear AR(2) model RMSE = 0.0700
ESTAR model RMSE = 0.0687
% improvement = 1.8979

Out-of-sample forecast comparison, 1961-1990:

Linear AR(2) model RMSE = 0.0783
ESTAR model RMSE = 0.0756
% improvement = 3.3784
MSE-T = 1.381
ENC-T = 1.750

Hamilton (2001), Dahl and González-Rivera (2002) nonlinearity tests:

λ_H^E = 0.011 [*p*-value = 0.920]
λ_{OP}^E = 0.694 [*p*-value = 0.621]
λ_{OP}^A = 9.228 [*p*-value = 0.097]
g_{OP} = 2.477 [*p*-value = 0.107]

Table 4: Test results, Bergman and Hansson (2000) MS-AR(1) model

A. U.K.

Estimated linear AR(1) model:

$\hat{q}_t = -0.778 + 0.952 q_{t-1}$
$\hat{\sigma}$ = 5.357

Estimated MS-AR(1) model:

$\hat{q}_t = 3.554(|\, s_t = 1) - 5.096(|\, s_t = 2) + 0.928 q_{t-1}$; p_{11} = 0.67; p_{22} = 0.69
$\hat{\sigma}$ = 3.181

Parametric encompassing test, MS-AR(1) model:

Sample var. = 209.588; MS-AR(1) var. = 333.059; *t*-stat. = –2.053
Sample autocor.= 0.901; MS-AR(1) autocor. = 0.956; *t*-stat. = –0.180
Joint χ^2-stat. = 4.269

In-sample forecast comparison, 1973:2-1990:4:

Linear AR(1) model RMSE = 5.3568
MS-AR(1) model RMSE = 3.1809
% improvement = 40.6208

Out-of-sample forecast comparison, 1991:1-1997:4:

Linear AR(1) model RMSE = 4.7862
MS-AR(1) model RMSE = 4.8751
% improvement = –1.8573
MSE-T = –0.473
ENC-T = 0.125

Hamilton (2001), Dahl and González-Rivera (2002) nonlinearity tests:

λ_H^E = 0.554 [*p*-value = 0.385]
λ_{OP}^E = 2.782 [*p*-value = 0.170]
λ_{OP}^A = 2.214 [*p*-value = 0.389]
g_{OP} = 2.004 [*p*-value = 0.360]

B. France

Estimated linear AR(1) model:

$\hat{q}_t = -0.092 + 0.963 q_{t-1}$
$\hat{\sigma}$ = 4.985

Estimated MS-AR(1) model:

$\hat{q}_t = 6.131(|\, s_t = 1) - 2.845(|\, s_t = 2) + 0.904 q_{t-1}$; p_{11} = 0.68 ; p_{22} = 0.83
$\hat{\sigma}$ = 2.776

Parametric encompassing test, MS-AR(1) model:

Sample var. = 281.224; MS-AR(1) var. = 316.658; *t*-stat. = –0.328
Sample autocor. = 0.947; MS-AR(1) autocor. = 0.960; *t*-stat. = –0.039
Joint χ^2-stat. = 0.109

In-sample forecast comparison, 1973:2-1990:4:

Linear AR(1) model RMSE = 4.9854
MS-AR(1) model RMSE = 2.7760
% improvement = 44.3183

Out-of-sample forecast comparison, 1991:1-1997:4:

Linear AR(1) model RMSE = 4.4642
MS-AR(1) model RMSE = 4.4292
% improvement = 0.7847
MSE-T = 0.156
ENC-T = 1.429

Hamilton (2001), Dahl and González-Rivera (2002) nonlinearity tests:

λ_H^E = 0.300 [*p*-value = 0.566]
λ_{OP}^E = 1.669 [*p*-value = 0.433]
λ_{OP}^A = 1.845 [*p*-value = 0.595]
g_{OP} = 1.801 [*p*-value = 0.453]

Table 4 (continued)

C. Germany

Estimated linear AR(1) model:

$\hat{q}_t = 0.036 + 0.960 q_{t-1}$
$\hat{\sigma} = 5.337$

Estimated MS-AR(1) model:

$\hat{q}_t = 6.569(s_t = 1) - 2.676(| s_t = 2) + 0.888 q_{t-1}$; $p_{11} = 0.68$; $p_{22} = 0.83$
$\hat{\sigma} = 3.274$

Parametric encompassing test, MS-AR(1) model:

Sample var. = 318.316; MS-AR(1) var. = 287.573; *t*-stat. = 0.263
Sample autocor. = 0.950; MS-AR(1) autocor. = 0.950; *t*-stat. = 0.002
Joint χ^2-stat. = 0.069

In-sample forecast comparison, 1973:2-1990:4:

Linear AR(1) model RMSE = 5.3368
MS-AR(1) model RMSE = 3.2740
% improvement = 38.6519

Out-of-sample forecast comparison, 1991:1-1997:4:

Linear AR(1) model RMSE = 4.6274
MS-AR(1) model RMSE = 4.5614
% improvement = 1.4258
MSE-T = 0.406
ENC-T = 1.616

Hamilton (2001), Dahl and González-Rivera (2002) nonlinearity tests:

$\lambda_H^E = 0.000$ [*p*-value = 0.991]
$\lambda_{OP}^E = 0.916$ [*p*-value = 0.647]
$\lambda_{OP}^A = 1.615$ [*p*-value = 0.870]
$g_{OP} = 0.519$ [*p*-value = 0.599]

D. Switzerland

Estimated linear AR(1) model:

$\hat{q}_t = -1.665 + 0.940 q_{t-1}$
$\hat{\sigma} = 5.741$

Estimated MS-AR(1) model:

$\hat{q}_t = 2.390(| s_t = 1) - 6.556(| s_t = 2) + 0.958 q_{t-1}$; $p_{11} = 0.79$; $p_{22} = 0.72$
$\hat{\sigma} = 3.676$

Parametric encompassing test, MS-AR(1) model:

Sample var. = 269.047; MS-AR(1) var. = 808.442; *t*-stat. = –7.484
Sample autocor. = 0.902 ; MS-AR(1) autocor. = 0.980 ; *t*-stat. = –0.253
Joint χ^2-stat. = 56.138

In-sample forecast comparison, 1973:2-1990:4:

Linear AR(1) model RMSE = 5.7412
MS-AR(1) model RMSE = 3.6760
% improvement = 35.9712

Out-of-sample forecast comparison, 1991:1-1997:4:

Linear AR(1) model RMSE = 5.1595
MS-AR(1) model RMSE = 5.3296
% improvement = –3.2968
MSE-T = –0.582
ENC-T = 0.461

Hamilton (2001), Dahl and González-Rivera (2002) nonlinearity tests:

$\lambda_H^E = 1.374$ [*p*-value = 0.132]
$\lambda_{OP}^E = 5.014$ [*p*-value = 0.023]
$\lambda_{OP}^A = 3.922$ [*p*-value = 0.113]
$g_{OP} = 1.737$ [*p*-value = 0.025]

Table 4 (continued)

E. Canada

Estimated linear AR(1) model:

$\hat{q}_t = 0.274 + 0.974 q_{t-1}$
$\hat{\sigma}$ = 1.572

Estimated MS-AR(1) model:

$q_t = 1.693(| s_t = 1) - 0.306(| s_t = 2) + 0.922 q_{t-1}$; p_{11} = 0.95; p_{22} = 0.94
$\hat{\sigma}$ = 1.282

Parametric encompassing test, MS-AR(1) model:

Sample var. = 61.627; MS-AR(1) var. = 83.685; *t*-stat. = –1.506
Sample autocor. = 0.967; MS-AR(1) autocor. = 0.986; *t*-stat. = –0.057
Joint χ^2-stat. = 2.270

In-sample forecast comparison, 1973:2-1990:4:

Linear AR(1) model RMSE = 1.5720
MS-AR(1) model RMSE = 1.2820
% improvement = 18.4497

Out-of-sample forecast comparison, 1991:1-1997:4:

Linear AR(1) model RMSE = 2.0799
MS-AR(1) model RMSE = 1.8857
% improvement = 9.3332
MSE-T = 2.191
ENC-T = 2.517

Hamilton (2001), Dahl and González-Rivera (2002) nonlinearity tests:

λ_H^E = 0.790 [*p*-value = 0.273]
λ_{OP}^E = 3.437 [*p*-value = 0.129]
λ_{OP}^A = 3.192 [*p*-value = 0.208]
g_{OP} = 1.398 [*p*-value = 0.184]

F. Japan

Estimated linear AR(1) model:

$\hat{q}_t = -1.573 + 0.957 q_{t-1}$
$\hat{\sigma}$ = 5.420

Estimated MS-AR(1) model:

$\hat{q}_t = -0.370(| s_t = 1) - 8.982(| s_t = 2) + 0.871 q_{t-1}$; p_{11} = 0.91; p_{22} = 0.82
$\hat{\sigma}$ = 3.985

Parametric encompassing test, MS-AR(1) model:

Sample var. = 361.308; MS-AR(1) var. = 364.008; *t*-stat. = –0.032
Sample autocor. = 0.929; MS-AR(1) autocor. = 0.959; *t*-stat. = –0.096
Joint χ^2-stat. = 0.010

In-sample forecast comparison, 1973:2-1990:4:

Linear AR(1) model RMSE = 5.4201
MS-AR(1) model RMSE = 3.9848
% improvement = 26.4808

Out-of-sample forecast comparison, 1991:1-1997:4:

Linear AR(1) model RMSE = 5.0758
MS-AR(1) model RMSE = 4.9048
% improvement = 3.3679
MSE-T = 0.547
ENC-T = 1.946

Hamilton (2001), Dahl and González-Rivera (2002) nonlinearity tests:

λ_H^E = 0.134 [*p*-value = 0.700]
λ_{OP}^E = 0.631 [*p*-value = 0.793]
λ_{OP}^A = 1.808 [*p*-value = 0.853]
g_{OP} = 0.997 [*p*-value = 0.814]

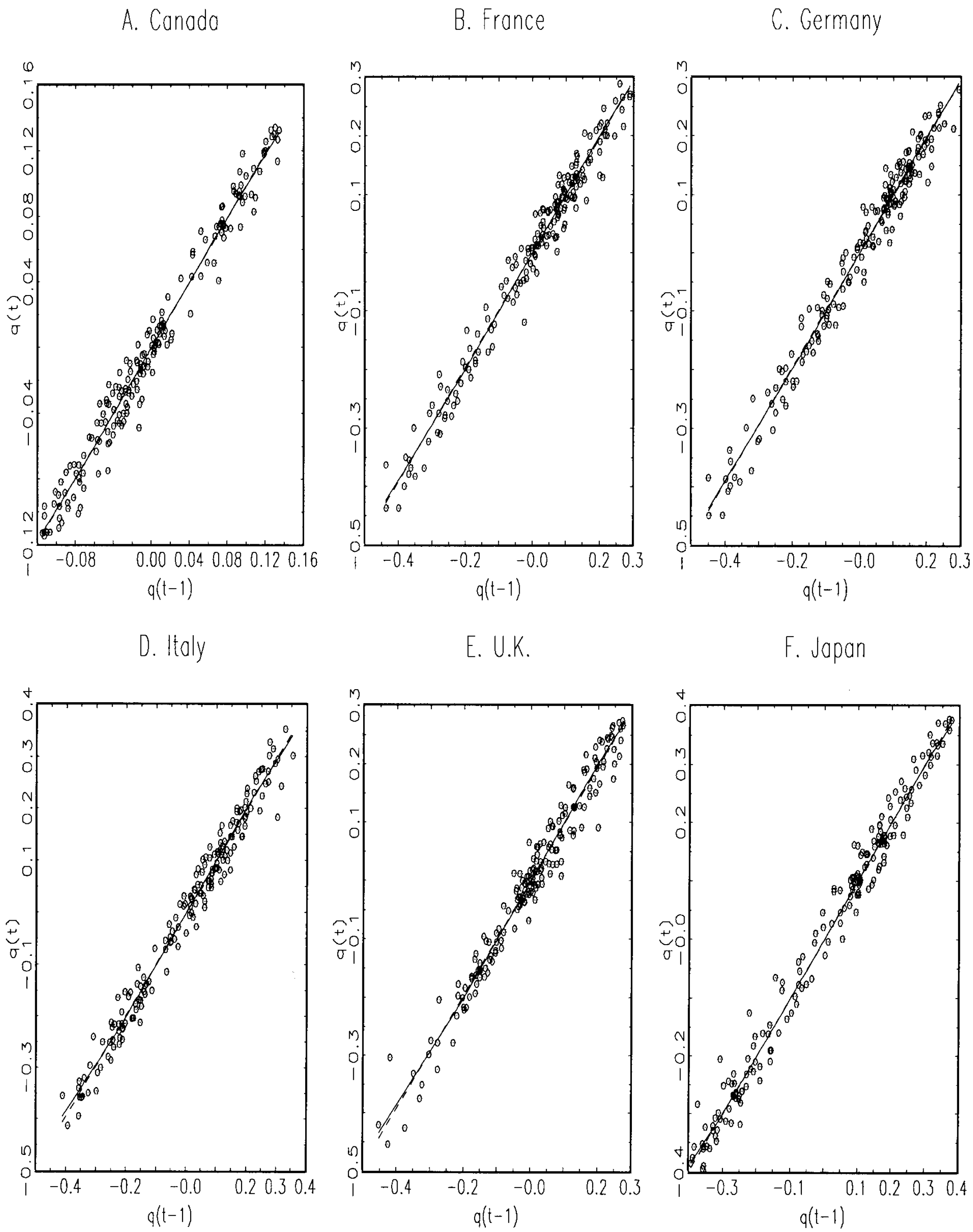

Figure 1: Scatterplot of real exchange rate log-levels ($q(t)$) and lagged real exchange rate log-levels ($q(t-1)$), Obstfeld and Taylor (1997) data.

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.

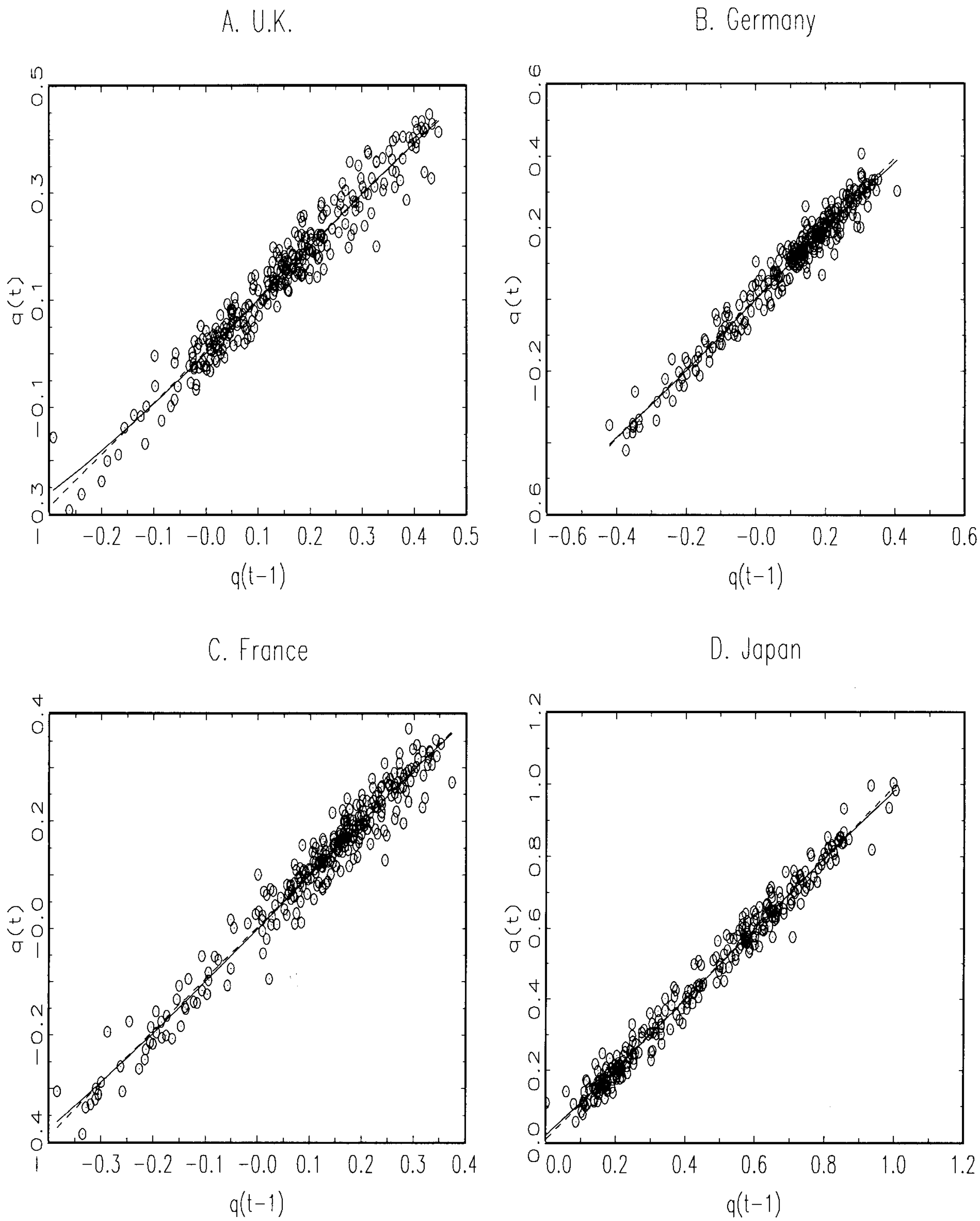

Figure 2: Scatterplot of real exchange rate log-level ($q(t)$) and lagged real exchange rate log-level ($q(t-1)$), Taylor, Peel, and Sarno (2001) data.

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.

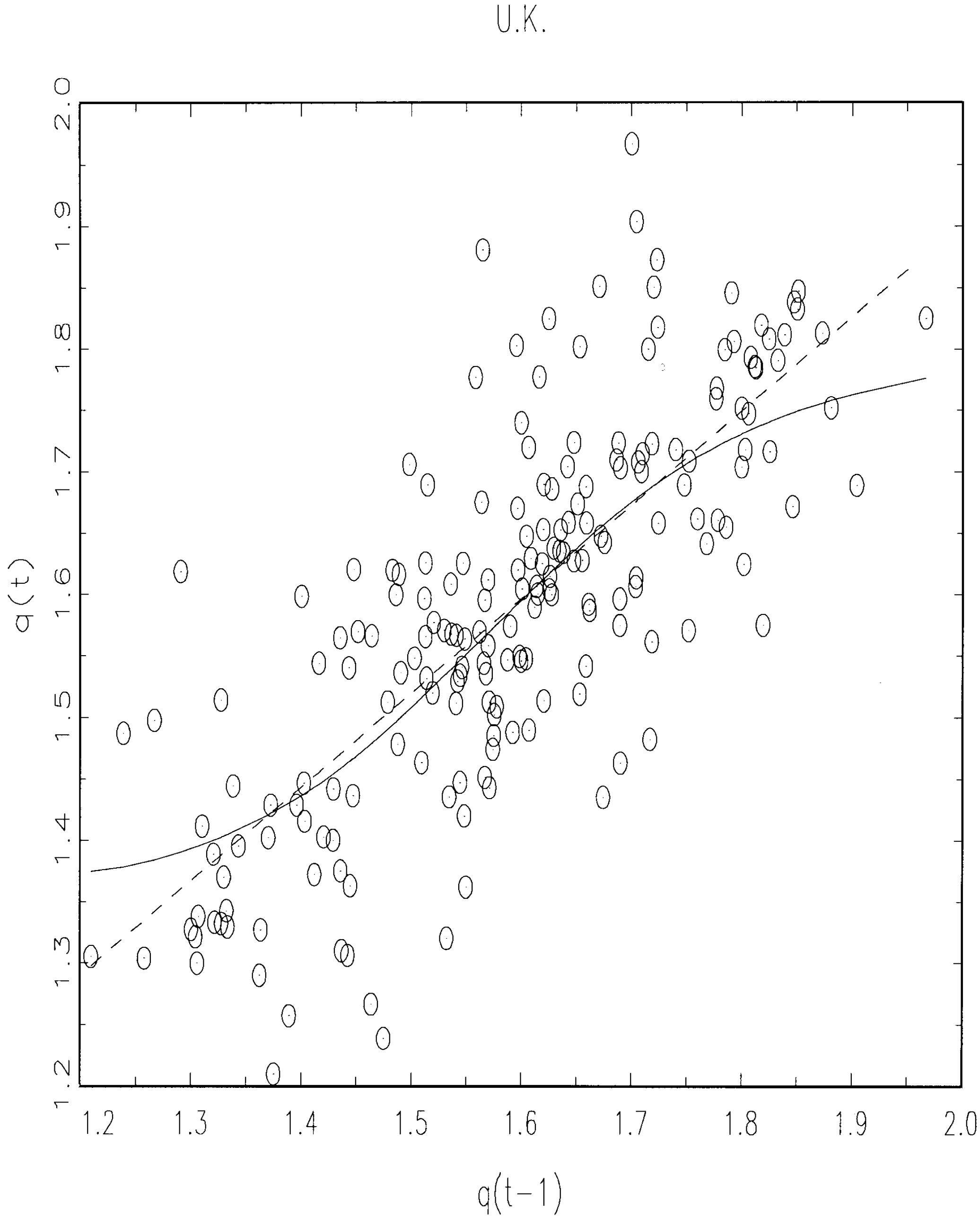

Figure 3: Scatterplot of real exchange rate log-levels ($q(t)$) and lagged real exchange rate log-levels ($q(t-1)$), Michael, Nobay, and Peel (1997) data.

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.

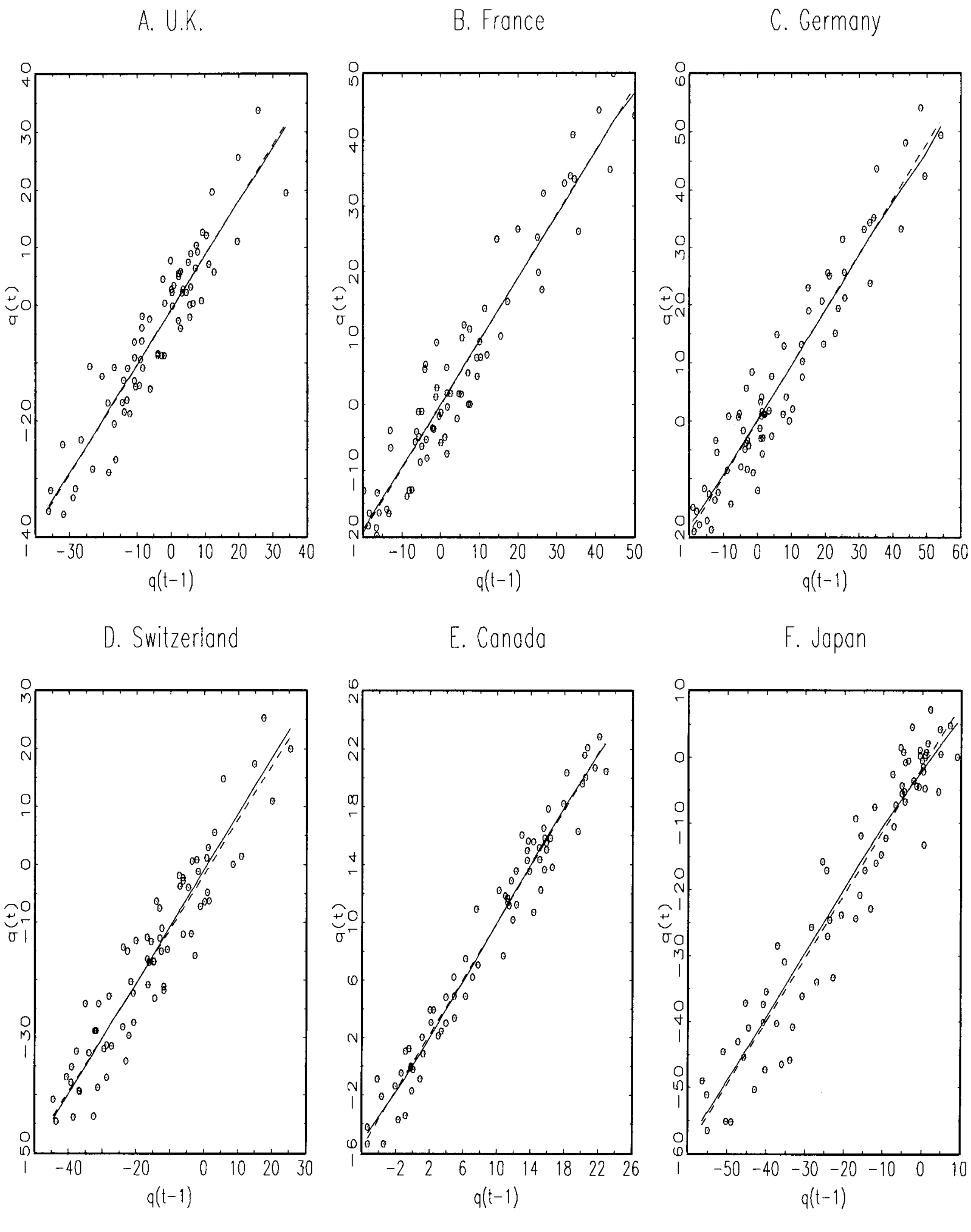

Figure 4: Scatterplot of real exchange rate log-levels ($q(t)$) and lagged real exchange rate log-levels ($q(t-1)$), Bergman and Hansson (2000) data.

Notes: solid line is the conditional expectation function for the fitted Bergman and Hansson (2000) MS-AR(1) model; dashed line is the conditional expectation function for a fitted linear AR(1) model.