DISCUSSION PAPER SERIES

No. 3024

WHY IS IT SO DIFFICULT TO BEAT THE RANDOM WALK FORECAST OF EXCHANGE RATES?

Lutz Kilian and Mark P Taylor

INTERNATIONAL MACROECONOMICS

Centre for Economic Policy Research

www.cepr.org

ISSN 0265-8003

WHY IS IT SO DIFFICULT TO BEAT THE RANDOM WALK FORECAST OF EXCHANGE RATES?

Lutz Kilian, University of Michigan and CEPR
Mark P Taylor, Warwick Business School and CEPR

Discussion Paper No. 3024
October 2001

Centre for Economic Policy Research
90–98 Goswell Rd, London EC1V 7RR, UK
Tel: (44 20) 7878 2900, Fax: (44 20) 7878 2999
Email: cepr@cepr.org, Website: www.cepr.org

CEPR Discussion Paper No. 3024

October 2001

ABSTRACT

Why is it so Difficult to Beat the Random Walk Forecast of Exchange Rates?

We propose an exchange rate model that can explain both the observed volatility and the persistence of real and nominal exchange rate movements and thus in some measure resolves Rogoff's (1996) purchasing power parity puzzle. Our analysis reconciles the well known difficulties in beating the random walk forecast model with the statistical evidence of non-linear mean reversion in deviations from fundamentals. Our analysis also provides a compelling rationale for the long-horizon predictability of exchange rates. We find strong empirical support for long-horizon predictability, and we explain why it is difficult to exploit this predictability in real-time forecasts. Our results not only lend support to economists' beliefs that the exchange rate is inherently predictable, but they also help us to understand the reluctance of applied forecasters to abandon chartist methods in favour of models based on economic fundamentals.

JEL Classification: C53, F31 and F47
Keywords: economic models of exchange rate determination, long-horizon regression tests, purchasing power parity, random walk and real exchange rate

Lutz Kilian
Department of Economics
University of Michigan
Ann Arbor
MI 48109-1220
Tel: (1 734) 764 2320
Fax: (1 734) 764 2769
Email: lkilian@umich.edu

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=132598

Mark P Taylor
Professor of Financial Economics
Warwick Business School
University of Warwick
Coventry
CV4 7AL
Tel: (44 24) 7657 3008
Fax: (44 24) 7652 3032
Email: mark.taylor@wbs.warwick.ac.uk

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=100357

Submitted 21 January 2001

1. Introduction

After nearly two decades of research since Meese and Rogoff's pioneering work on exchange rate predictability (see Meese and Rogoff, 1983a,b), the goal of exploiting economic models of exchange rate determination to beat naïve random walk forecasts remains as elusive as ever (see Taylor, 1995). One possible explanation is simply that standard economic models of exchange rate determination are inadequate. Indeed, this appears to be the response of many professional exchange rate forecasters (see e.g. Cheung and Chinn, 1999), although this interpretation seems to go against deeply held beliefs among many economists.

A more charitable interpretation of the dismal forecast performance of economic exchange rate models is that the theory is fundamentally sound, but its empirical implementation as a linear statistical model is flawed. In that view, economic models of the exchange rate imply long-run equilibrium conditions only, toward which the economy adjusts in a nonlinear fashion. Indeed, there has been recent work documenting various nonlinearities in deviations of the spot exchange rate from economic fundamentals (e.g. Balke and Fomby, 1997; Taylor and Peel, 2000; Taylor, Peel and Sarno, 2001). This literature differs from the earlier literature on nonlinear exchange rate forecasting in that it is based on equilibrium conditions derived from economic theory. The evidence of nonlinear mean reversion in the deviation from equilibrium has raised expectations that, with the appropriate nonlinear structure, economic models of the exchange rate will prove useful for forecasting, at least at longer horizons.

The question of forecast accuracy traditionally has been an important test of the credibility of economic models of exchange rate determination. However, the forecast performance of nonlinear models based on economic theory has been left largely unexplored. Part of the problem relates to technical difficulties in implementing forecast accuracy tests in a nonlinear framework while another part is related to the small samples of data available for empirical work. Moreover, what has been lacking is a convincing economic explanation of the source of the nonlinearities found in empirical work. The most successful nonlinear empirical exchange rate models embody smooth threshold dynamics. One possible explanation of such dynamics are transaction costs, but transaction costs do not provide a compelling explanation of long swings in nominal exchange rates such as the large and persistent overvaluation of the dollar during the mid-1980s, nor do they explain the observed volatility in real and nominal exchange rates.

In Section 2, we discuss a new and complementary explanation of smooth threshold dynamics that arises from the widespread uncertainty about the equilibrium level of the exchange rate. We postulate a distribution of beliefs about the appropriate level of the exchange rate relative to a given mismeasured economic fundamental. Agents assign less probability to levels corresponding to large deviations from the measured fundamental than to values close to the measured fundamental because large deviations are increasingly implausible from a theoretical point of view. Thus, as the spot exchange rate moves away from the latent equilibrium, a consensus will gradually build that the spot exchange rate is misaligned, inducing rational agents to take stronger positions against the prevailing exchange rate and ensuring the ultimate mean reversion of the exchange rate toward the unobserved economic fundamental. We propose an econometric model of the exchange rate that captures this behavior and that endogenously generates nonlinear smooth threshold dynamics in the exchange rate as a function of past departures from equilibrium.

In Section 3, we provide empirical evidence for this type of nonlinear dynamics in the form of estimates of exponential smooth threshold autoregressive (ESTAR) models linking movements in the nominal exchange rate nonlinearly to movements in underlying economic fundamentals. This evidence is important for a number of reasons.

First, it corroborates the evidence presented in Taylor, Peel and Sarno (2001) that there is strong - albeit nonlinear - mean reversion in dollar real exchange rates. Allowing for this nonlinearity goes some way towards solving Rogoff's (1996) 'purchasing power parity puzzle' concerning the apparently very slow speed of adjustment of real exchange rates. Any satisfactory model of exchange rates must be able to explain: (1) the existence of large deviations from macroeconomic fundamentals, (2) the persistence of these deviations over time, and (3) the short-term volatility of deviations from fundamentals. The nonlinear model we propose is consistent with all three of these empirical regularities.

Second, the evidence of smooth threshold dynamics is important because it suggests that the exchange rate should be more predictable at longer forecast horizons, at least for large enough sample sizes. In contrast, in the linear exchange rate models used in the previous literature there is no reason for the exchange rate to be more predictable at longer horizons than at short horizons. For example, Berben and van Dijk (1998) prove that in the linear model the asymptotic power of long-horizon regression tests is constant across forecast horizons. In related work, Kilian (1999) provides simulation evidence for the lack of increased long-horizon power in finite samples.

Third, the evidence of smooth threshold dynamics is important because the standard tests of long-horizon exchange rate predictability used in the literature (see Chinn and Meese 1995; Mark 1995; Mark and Sul 2001) are invalidated by the presence of nonlinear mean reversion. Our analysis suggests that all previous results on long-horizon exchange rate predictability based on linear models must be regarded as invalid. This paper is intended to re-open the debate over long-horizon predictability, taking full account of the nonlinearity of the data generating process.

Although previous research has often implicitly appealed to nonlinearities to justify increased long-horizon predictability, the source and nature of these nonlinearities has remained vague and no attempt has been made to quantify the importance of this source of exchange-rate predictability. In Section 4, we quantify the predictability of the exchange rate in an idealized ESTAR model using response surface techniques. We show that - in the presence of nonlinear mean reversion - the degree of predictability relative to the random walk forecast increases with the forecast horizon. These forecast techniques, however, are difficult to implement in practice.

In section 5, we therefore develop a new empirical methodology for assessing the degree of long-horizon predictability of nominal exchange rates in the presence of smooth-threshold nonlinearities. We propose an easy-to-use statistical test of the relative forecast accuracy of our nonlinear model against the random walk model. The proposed long-horizon regression test not only is highly accurate under the null of no exchange rate predictability, but it has high power against plausible alternatives, even in relatively small samples.

Using this new test, in Section 6 we provide strong empirical evidence for seven OECD countries that the predictability of the spot dollar exchange rate improves dramatically as the forecast horizon is lengthened from one quarter to several years. This evidence is based on "in-sample predictability tests based on fitted models for the entire post-Bretton Woods sample period. For six of seven countries we beat the random walk forecast at forecast horizons of two or three years at conventional significance levels.

If the exchange rate is inherently predictable, why has it been so difficult to beat the random walk forecast in real time? We identify three reasons in this paper. First, we show that near the equilibrium in our model the exchange rate will be well approximated by a random walk. This fact goes a long way toward explaining the success of the random walk forecast for

OECD exchange rates in earlier work (Meese and Rogoff, 1983a, 1983b). Moreover, given our evidence of nonlinearity, forecast accuracy tests based on linear mean reversion toward economic fundamentals will be misspecified, which makes it difficult to judge the success of the random walk forecast using traditional tools.

Second, even if nonlinear mean reversion is modeled correctly, the power of recursive real time ("out-of-sample") forecast tests may be too low to beat the random walk forecast in real time, given the short time span of post-Bretton Woods data. We document the low power of the recursive out-of-sample forecast accuracy tests relative to "in-sample" tests by Monte Carlo simulation. Indeed, in our empirical analysis, we are unable to establish beyond a reasonable doubt that the long-horizon regression forecast is significantly more accurate than the random walk forecast in a recursive real-time forecast setting, although we find strong evidence of increased long-horizon predictability consistent with the in-sample results.

Third, our ESTAR model suggests that the strength of the link between the exchange rate and fundamentals increases nonlinearly with the distance of the exchange rate from the level consistent with economic fundamentals. The closer the exchange rate is to its equilibrium value, the more random and less predictable will be the observed movements in the spot exchange rate. Thus, only unusually large departures from fundamentals in the sample path will reveal the inherent tendency toward mean reversion and such events may be rare along a given sample path, unless the sample size is large. This view is also supported by historical evidence that at least during periods of large departures from economic fundamentals (such as during hyperinflations), the exchange rate does seem to behave as suggested by economic theory (see Frenkel, 1976; Taylor and McMahon, 1988), whereas the evidence is much less clear during normal times (see Taylor, 1995).

2. Sources of Threshold Dynamics in Exchange Rates

Recently, several papers have investigated the evidence of smooth thresholds in the deviation of spot exchange rates from macroeconomic fundamentals (e.g. Balke and Fomby, 1997; Taylor and Peel 2000; Taylor, Peel and Sarno 2001). Such nonlinearities are frequently motivated by the existence of transactions costs (see Dumas, 1992; Taylor, Peel and Sarno, 2001).[1] It might be argued, however, that transactions costs alone could not account for many of the observed very large movements in real exchange rates, either in terms of day-to-day volatility or in terms of periods of substantial and persistent misalignments such as the substantial degree of overvaluation of the U.S. dollar in the 1980s. In this paper, we therefore propose a new and complementary explanation of smooth threshold dynamics.

Our starting point is a model of noise trading and arbitrage along the lines of models proposed by De Long, Shleifer, Summers and Waldmann (1990a,b). Consider a world consisting of two types of foreign exchange traders: (1) noise traders and (2) rational speculators (or arbitrageurs). Noise traders are traders whose demand for foreign exchange is affected by beliefs that are not fully justified by news about fundamentals. These traders follow pseudo-signals about future returns such as the advice of financial analysts or technical analysis. Arbitrageurs are traders who form fully rational expectations about the returns of holding foreign exchange. Arbitrageurs will take advantage of noise traders' mistaken beliefs about future movements in the exchange rate. They will sell foreign exchange when noise traders push prices up and buy when noise traders depress prices, turning a profit in the process.

Why then are noise traders not driven out of the market? As shown by Shleifer and Summers (1990), if risk taking is rewarded in the market, noise traders may earn higher expected returns than rational speculators. With higher expected returns, noise traders as a group may be slow to disappear from the market. Moreover, as new noise traders enter the market, these traders will be subject to the same judgment biases as the current survivors in the market.

More importantly, the unpredictability of noise traders' future opinions creates risk to arbitrageurs that will prevent complete arbitrage. Arbitrage will be limited by two types of risk. The first risk is that future realizations of the fundamental may turn out higher than anticipated. The second risk is that an asset that is overpriced today may be even more overpriced tomorrow because of unpredictable swings in the demand of noise traders. Although this mispricing will be corrected in the long run, arbitrageurs tend to have horizons too short time to take advantage of this arbitrage opportunity. One reason is that most arbitrageurs have to borrow to implement their trades and are subject to per period fees. Another reason is that many traders' performance is evaluated at short intervals, removing incentives for long-term arbitrage against fundamental

[1] See Taylor, Peel and Sarno (2001) for further discussion of this literature and further references.

mispricing.[2]

These arguments alone, however, are likely to understate the limits of arbitrage. They presume that the rational trader actually knows the fundamental value of the asset. This assumption is empirically implausible, given the large degree of uncertainty in exchange rate modeling. In practice, the equilibrium value of the exchange rate cannot be observed directly and arbitrageurs will have as hard a time as econometricians in detecting deviations of the exchange rate from fundamentals. This fact introduces an additional element of risk in arbitrage, as noted by Shleifer and Summers (1990).

Shleifer and Summers implicitly treat this model risk as a constant. Clearly, however, this model risk will diminish, as the exchange rate becomes increasingly overvalued or undervalued. We postulate a distribution of beliefs about the appropriate level of the exchange rate relative to a given mismeasured economic fundamental. Agents assign less probability to levels corresponding to large deviations from the measured fundamental than to values close to the measured fundamental because larger deviations are increasingly implausible from a theoretical point of view. Thus, close to the latent equilibrium, there is no consensus as to whether the exchange rate is overvalued or undervalued. Few rational traders will be inclined to take a strong position because no trader is confident of having the right model. Even if a given arbitrageur has strong views on this question, that arbitrageur has to take account of the heterogeneity of beliefs among other traders. Thus, close to the equilibrium, the exchange rate is driven mainly by noise traders. As the exchange rate moves away from the latent equilibrium, however, a consensus will gradually build that the exchange rate is misaligned, inducing rational agents to take stronger positions against the prevailing exchange rate and ensuring the ultimate mean reversion of the exchange rate toward the unobserved true economic fundamental.

For example, suppose that the latent equilibrium of the Euro-dollar rate is somewhere close to 1. An agent may not be sure whether the Euro-dollar exchange rate is more likely to be 0.95 or 1.05, but the same agent may be fairly certain that a rate of 0.70 is unreasonably low. In

[2] De Long, Shleifer, Summers and Waldmann (1990b) observe that in the presence of positive feedback traders rational arbitrage may even destabilize the exchange rate. Specifically, arbitrageurs may choose to buy more foreign exchange than would be justified by good news about fundamentals alone, in anticipation of positive feedback trading. Such a model helps to explain how traders can rationally expect that the exchange rate will continue to rise, while being fully aware that the exchange rate is already overvalued relative to fundamentals (see Frankel and Froot, 1988). It also suggests that it may be difficult to distinguish genuine noise traders and arbitrageurs in practice.

that case, the agent will be reluctant to take a position either way against a Euro-dollar rate of 0.95 because the risk of being proven wrong is too high. Instead, the agent will be inclined to wait for the Euro to fall to a sufficiently low level before speculating on an appreciation. As a result, the movements of the nominal exchange rate may appear quite random for values close to the equilibrium, but for sufficiently large departures from equilibrium, mean reversion toward the fundamental sets in. This nonlinearity may be described by a smooth threshold model, in which the strength of mean reversion is an increasing function of past deviations from equilibrium. The next section will present empirical evidence that actual exchange rate behavior is broadly consistent with this stylized framework.[3]

3. Econometric Evidence of Nonlinear Mean Reversion in Deviations from Fundamentals

A parsimonious parametric model that captures the nature of nonlinear mean reversion as motivated by our stylized theoretical model is the exponential threshold autoregressive (ESTAR) model of Teräsvirta (1994). We apply this model to quarterly data on bilateral U.S. dollar exchange rates and fundamentals for Canada, France, Germany, Italy, Japan, Switzerland, and the U.K. There are potentially many different type of macroeconomic fundamentals that could drive the spot exchange rate. For example, Chinn and Meese (1995), Mark (1995), Kilian (1999), and Taylor and Peel (2000) focus on so-called monetary fundamentals. These models require numerous auxiliary assumptions that are difficult to verify directly. In this paper, we therefore focus on a much simpler measure of fundamentals, namely relative aggregate prices or ‘purchasing power parity’ (PPP) fundamentals (Taylor, 1995). Taylor, Peel and Sarno (2001) document evidence of nonlinear mean reversion in real (i.e. relative price adjusted) exchange behaviour using monthly data for a number of major dollar exchange rates over the post-Bretton Woods period. We begin by extending the empirical evidence for nonlinear mean reversion in deviations from PPP fundamentals.

We use quarterly data obtained from the International Monetary Fund’s International Financial Statistics data base for the period 1973.I-2000.IV on spot nominal exchange rates (foreign price of dollars), denoted (in logarithms) by e_t, and a PPP fundamental based on relative

---

[3] See also De Grauwe and Dewachter (1993) for a model in which the interaction of traders who base their view on economic fundamentals and of traders who rely on chartist methods induces nonlinear exchange rate dynamics.

consumer price indices, $f_t \equiv p_t - p_t^*$, where p_t is the logarithm of the U.S. CPI and p_t^* is the logarithm of the foreign CPI. Hence, the deviation of the nominal exchange rate from the underlying PPP fundamental, $z_t \equiv e_t - f_t$, is in fact the real exchange rate (in logarithmic form). For uniformity, we demeaned z_t for each country prior to the empirical analysis.

Examination of the partial correlogram for z_t , as proposed in the context of the estimation of nonlinear autoregressive models by Granger and Teräsvirta (1993) and Teräsvirta (1994), revealed second-order serial correlation in the data, suggesting a nonlinear AR(2) model[4]. Specifically, we postulated a smooth transition autoregressive model or STAR model of the form

$$z_t - \mu_z = \Phi\left[\left\langle z_{t-d} \right\rangle_{d=1}^{\bar{d}}; \mu; \Gamma\right]\left(\phi_1(z_{t-1} - \mu_z) + \phi_2(z_{t-2} - \mu_z)\right) + u_t, \quad u_t \sim iid(0, \sigma^2).$$

The transition function $\Phi\left[\left\langle z_{t-d} \right\rangle_{d=1}^{\bar{d}}; \mu; \Gamma\right]$ determines the degree of nonlinearity in the model and is a function of lagged movements in the real exchange rate, $\left\langle z_{t-d} \right\rangle_{d=1}^{\bar{d}}$, of the equilibrium level of the real exchange rate, $\mu \in (-\infty, \infty)$, and of the vector of transition parameters $\Gamma \in (-\infty, 0]^{\bar{d}}$. Previous work (Taylor and Peel, 2000; Taylor, Peel and Sarno, 2001) suggests that an exponential form of the transition function is particularly applicable to real exchange rate movements. This functional form also makes good economic sense in this application because it implies symmetric adjustment of the exchange rate above and below equilibrium. Granger and Teräsvirta (1993) term STAR models employing exponential transition functions exponential STAR or ESTAR models.

In any empirical application, it is of course necessary to determine the delay $\bar{d}$ (the dimension of Γ and the number of lagged values of the real exchange rate influencing the transition function) and whether any of its elements are zero. In general, applied practice with ESTAR models has favored restricting Γ to be a singleton (see e.g. Teräsvirta, 1994; Taylor,

[4] There was, however, no evidence of higher-order serial correlation in the *nominal* exchange rate, suggesting that the standard random walk comparator is still applicable.

Peel and Sarno, 2001), and Granger and Teräsvirta (1993) and Teräsvirta (1994) suggest a series of nested tests for determining the appropriate delay in this case. In the present application to quarterly real exchange rate data, however, we found that the model which worked best for each country - in terms of goodness of fit, statistical significance of parameters, and adequate diagnostics – set the dimension of Γ to five, with each element equal to the same negative value γ. This parameterization seems reasonably intuitive since it allows the effects of deviations from equilibrium to affect the nonlinear dynamics with a single lag (rather than suddenly kicking in at a higher lag) and also allows the effects of persistent deviations to be cumulative.

In addition, we subsequently found that the restriction $\phi_1 + \phi_2 = 1$ could not be rejected at standard significance levels for any of the countries. Hence, the model which we estimated for each country was of the form:

$$z_t - \mu_z = \left(\exp\left\{\gamma \sum_{d=1}^{5} (z_{t-d} - \mu_z)^2\right\}\right)\left(\phi_1 (z_{t-1} - \mu_z) + (1-\phi_1)(z_{t-2} - \mu_z)\right) + u_t$$

The exponential transition function $\Phi = \left(\exp\left\{\gamma \sum_{d=1}^{5} (z_{t-d} - \mu_z)^2\right\}\right), \quad \gamma < 0,$ will take the value unity when the last five values of the nominal exchange are equal to the fundamental equilibrium level $\mu_z + f_{t-d}$, or equivalently the real exchange rate z_{t-d} is equal to its equilibrium level μ_z. Thus, at equilibrium, the real exchange rate will follow a unit root process:

$$z_t = \phi_1 z_{t-1} + (1-\phi_1) z_{t-2} + u_t$$

As departures from the fundamental equilibrium increase, however, Φ shrinks towards its limiting value of zero and at any instant the real exchange rate will follow a mean-reverting AR(2) process with mean μ_z and slope coefficients adding up to the instantaneous value of $\Phi < 1$.

The resulting ESTAR models are estimated on our data by nonlinear least squares (see

Gallant, 1987; Gallant and White, 1988).[5] The model estimates are reported in Table 1. The models perform well in terms of providing good fits, statistically significant coefficients and the residual diagnostic statistics are satisfactory in all cases (see Eitrheim and Teräsvirta, 1996).

The estimated standardized transition parameter in each case appears to be strongly significantly different from zero both on the basis of the individual '*t*-ratios'. It should be noted, however, that these '*t*-ratios' must be carefully interpreted since, under the null hypothesis that the transition parameter γ is in fact equal to zero, each of the real exchange rate series would be generated by a unit root process. Hence, the presence of a unit root under the null hypothesis complicates the testing procedure analogously to the way in which the distribution of a Dickey-Fuller statistic cannot be assumed to be approximately Gaussian. We therefore calculated the empirical marginal significance levels of these test statistics by Monte Carlo methods under the null hypothesis that the true data generating process for the logarithm of each of the four real exchange rate series was a unit root AR(2) process, with the parameters of the data generating process calibrated using the actual real exchange rate data over the sample period. The empirical significance levels were based on 5,000 simulations of length 604, from which the first 500 data points were in each case discarded. At each replication, ESTAR equations identical in form to those reported in Table 1 were estimated. The percentage of replications for which a '*t*-ratio' for the estimated transition parameters greater in absolute value than that reported in Table 1 was obtained was then taken as the empirical significance level in each case. From these empirical marginal significance levels (reported in square brackets below the coefficient estimates in Table 1), we see that the estimated transition parameter is significantly different from zero at the one percent significance level in every case. Since these tests may be construed as nonlinear univariate unit root tests, the results indicate strong evidence of nonlinear mean reversion for each of the dollar real exchange rates examined over the post-Bretton Woods period. This is our first significant empirical result and corroborates the evidence of Taylor, Peel and Sarno (2001) based on monthly data. Previous research based on linear models has generally found great difficulty in rejecting the unit root hypothesis at standard test sizes for real exchange rates over the post Bretton Woods period (Taylor, 1995; Rogoff, 1996; Taylor, Peel and Sarno, 2001)[6].

[5] Regularity conditions for the consistency and asymptotic normality of this estimator are discussed by Gallant (1987), Gallant and White (1988), Klimko and Nelson (1978) and, in the present context, Tjøstheim (1986).

We also calculated the half lives of various sizes of shock to the real exchange rate using our estimated models. The results are shown in Table 2.[7] The half-life estimates demonstrate the nonlinear nature of the estimated real exchange rate models, with larger shocks mean reverting much faster than smaller shocks. Indeed, very large shocks of twenty percent have a half life of only four or five quarters, while – at the other extreme – small shocks of one percent have a half life of three to four years. These results therefore again confirm the findings of Taylor, Peel and Sarno (2001) and shed light on Rogoff's (1996) 'purchasing power parity puzzle' concerning the very slow speed of adjustment of real exchange rate shocks, and constitute our second significant empirical finding.[8] Only for small shocks occurring when the real exchange rate is near its equilibrium do our nonlinear models consistently yield 'glacial' speeds of adjustment with half lives in the three to five years range.

An additional test of the model is its ability to generate endogenously data that exhibit both high short-term volatility and large and persistent deviations from fundamentals of the same magnitude as in actual data. Figure 1 shows a representative realization from the fitted ESTAR model for Germany, plotted against the actual real exchange rate. Not only is there clear evidence of long swings in the real exchange rate data, but the simulated data also display short-term volatility. Thus, our model is consistent with the three empirical regularities any satisfactory model of exchange rates must be able to explain: (1) the existence of large deviations from macroeconomic fundamentals, (2) the persistence of these deviations over time, and (3) the short-term volatility of deviations from fundamentals.

So far our empirical evidence has shed light on three important exchange rate puzzles. First, by allowing for nonlinearities we have provided evidence of mean reversion in the real

---

[6] As is now well known, the difficulty in rejecting the unit root hypothesis for real exchange rates may also be largely due to a lack of statistical power in unit root tests with sample spans corresponding to the post Bretton Woods period (Lothian and Taylor, 1997). However, attempts to overcome this problem by the use of very long data sets (e.g. Lothian and Taylor, 1996) or by using panel unit root tests (e.g. Abuaf and Jorion, 1990) may introduce additional problems (Taylor, 1995; Taylor, Peel and Sarno, 2001) concerning possible regime shifts in the case of long-span studies (Hegwood and Papell, 1998) and problems in the interpretation of test procedures in the case of panel unit root tests (Taylor and Sarno, 1998; Sarno and Taylor, 2001).

[7] The half lives were calculated by Monte Carlo integration conditional on average initial history as described in Taylor, Peel and Sarno (2001), except that a non-parametric bootstrap of the estimated residuals was used rather than draws from the normal distribution.

[8] Rogoff (1996) argues that the very long half lives of three to five years typically reported for real exchange rates are puzzling because real exchange rates must be driven largely by monetary and financial factors (because of their volatility), which one would expect to adjust much faster.

exchange rate that has hitherto proved notoriously elusive. Second, we have shown that the puzzlingly slow speeds of adjustment or real exchange rates previously found may also be largely due to a failure to allow for nonlinear adjustment. Third, we have shown that the model generates data that are consistent with the empirical regularities found in actual data. A fourth important exchange rate puzzle is the difficulty of beating a simple random walk forecast with a model based on exchange rate fundamentals, and was first brought to the profession's attention by the work of Meese and Rogoff (1983a, 1983b). In the remainder of the paper we demonstrate how light may also be shed on this fourth puzzle by allowing for nonlinearities in exchange rate movements.

4. Long-Horizon Predictability Due to Nonlinear Mean Reversion: A Monte Carlo Study

The evidence in Table 2 of long-run purchasing power parity suggests that the spot exchange rate should be predictable at least at long horizons for sample sizes large enough to allow accurate estimation of the ESTAR model. In this section we construct a response surface to demonstrate this point. We generate repeated trials from the stylized ESTAR population model

$$z_t = \left(\exp\left\{\gamma \sum_{d=1}^{5} (z_{t-d})^2\right\}\right)\left(\phi z_{t-1} + (1-\phi) z_{t-2}\right) + u_t, \quad u_t \sim N(0,\sigma^2),$$

where $z_t \equiv e_t - f_t$. For expositional purposes we assume that $f_t = 0 \;\forall\; t$. Thus, the data generated by this model corresponds to the spot exchange rate. This assumption greatly simplifies the simulation design without affecting the main insights. For the benchmark model we use $\gamma = -0.7$, $\phi_1 = 1.2$ and $\sigma = 0.05$. These parameter values are roughly consistent with the range of estimates in Table 1.

We generate repeated draws of exchange rate data from this process and compare the root prediction mean squared error (RPMSE) of the ESTAR model to that of the no-change forecast. The forecast gains will be expressed as percent reductions of the RPMSE of the random walk forecast for each horizon. For expository purposes, we assume that the parameters of the ESTAR model are known. This allows us to abstract from small-sample distortions. Qualitatively similar results are obtained when the parameters are estimated for realistic sample sizes. We first study the extent and pattern of predictability across forecast horizons for the benchmark model. Figure 2 shows that on average the ESTAR forecast is unambiguously more accurate than the random walk forecast. The forecast gains tend to increase monotonically with the horizon from 6% for k = 1 to 22% for k = 16. This simple experiment convincingly demonstrates that in principle the random walk model can be beaten in a world characterized by threshold dynamics.

The degree to which the ESTAR model outperforms the random walk model of course depends on the design of the process. We therefore analyze the sensitivity of the simulation results to the choice of key parameters. Figure 3 shows that the qualitative results are robust to parameter changes. The differences are only a matter of degree. We find that predictability generally increases with σ^2, ϕ and γ. Low values of σ^2 are associated with low predictability because it takes large deviations from the fundamental to give the ESTAR forecast model an advantage. In the ESTAR model the strength of the link between the exchange rate and fundamentals increases nonlinearly with the distance of the exchange rate from the level consistent with economic fundamentals. Conversely, as the exchange rate approaches the fundamental equilibrium, its behavior becomes less and less mean reverting and in fact approaches that of a unit root process. Hence, if the innovation variance is small, we would expect the exchange rate to remain close to the fundamental equilibrium, resulting in low predictability relative to the random walk.

More generally, this result suggests that predictability will be low relative to the random walk forecast if the exchange rate remains close to its fundamental value for a given sample path. In essence, the ESTAR model and random walk will be almost observationally equivalent in that case. This phenomenon may help to explain the difficulties in beating the random walk forecast based on very short samples of data for OECD countries under the recent float (see Meese and Rogoff, 1983a,b). We would expect, however, that these difficulties could be overcome if the sample period is extended far enough for the sample path to become representative for the underlying ESTAR process and if the nonlinear mean reversion is modeled correctly.

Although the simulation results in Figures 2 and 3 are highly suggestive, they are based on simplifying assumptions, most importantly that $f_t = 0 \; \forall \; t$. In practice, it is not enough to model z_t, if we are interested in forecasting the spot exchange rate, rather we need to model the *joint* time series process for fundamentals and spot exchange rates. The latter task is considerably more demanding than estimating a univariate model and may involve estimation of a large number of parameters. The next section will propose some easy-to-use econometric tests that avoid these difficulties.

5. A New Approach to Generating Bootstrap Critical Values for Long-Horizon Regression Tests

The response surface in Figure 2 suggests that it might be possible in practice to beat the constant change forecast for sufficiently large sample sizes and forecast horizons. In this section, we will propose an easy-to-use econometric test of that proposition. Note that for small sample sizes the estimation of the full bivariate nonlinear model for $x_t = (e_t, f_t)'$ is an extremely difficult exercise. However, we can greatly reduce the number of parameters to be estimated by utilizing the well-known technique of long-horizon regression tests as a diagnostic tool. Long-horizon regressions take the form

$$e_{t+k} - e_t = a_k + b_k z_t + \varepsilon_{t+k}, \qquad k = 1, 4, 8, 12, 16$$

where the error term in general will be serially correlated. Mean reversion in exchange rates may be detected by a *t*-test of $H_0: b_k = 0$ versus $H_1: b_k < 0$ for a given forecast horizon *k*, or jointly for all forecast horizons as $H_0: b_k = 0 \;\forall\, k$ versus $H_1: b_k < 0$ for some *k*. It is well known that asymptotic critical values for the *t*-test statistics are severely biased in small samples. In order to mitigate these size distortions, critical values are usually calculated based on the bootstrap approximation of the finite sample distribution of the test statistic under the null hypothesis of no exchange rate predictability.

Alternatively, the out-of-sample prediction mean-squared error of the two models may be evaluated using the *DM* test of Diebold and Mariano (1995). A formal test may be based on a sequence of rolling or recursive forecasts and involves comparing the null of equal forecast accuracy against the one-sided alternative that forecasts from the long-horizon regression are more accurate than random walk forecasts. The distribution of the *DM* test statistic in long-horizon regression problems is not known in general (see McCracken, 1999). In practice, it is common to rely on the bootstrap approach to construct critical values for the *DM* test.

Long-horizon regression tests have been used extensively in the past (e.g., Mark 1995;

Chinn and Meese 1995), but without much success (see Kilian 1999). The reason is that previous research focused on linear models. In a world of linear mean reversion there is no rationale for conducting long-horizon regression tests. The problem is that under linearity *k*-step ahead forecasts are obtained by linear extrapolation from 1-step ahead forecasts. Thus, by construction there can be no gain in power at longer horizons (see Berben and van Dijk 1998; Kilian 1999; Berkowitz and Giorgianni 2001). Our assumption of nonlinear mean reversion, in contrast, provides a new and compelling rationale for the use of long-horizon regression tests.

It is often believed that conventional long-horizon regression tests have power against nonlinear processes of unknown form (such as peso problems or fads). What is not always understood is that the conventional approach of using bootstrap critical values for long-horizon regression tests does not allow for that possibility. The reason is that these critical values are obtained under the explicit assumption of a linear data generating process of the form

$$\begin{aligned}\Delta e_t - \mu_e &= u_{1t} \\ z_t - \mu_z &= \sum_{j=1}^{J} b_j \ (z_{t-j} - \mu_z) \ + u_{2t}\end{aligned}$$

As noted by Kilian (1999) if the true process is nonlinear, these critical values are invalid under the null hypothesis and the resulting bootstrap *p*-values cannot be given meaningful interpretations.

We therefore propose a modification of the bootstrap methodology for long-horizon regression tests. We postulate that under the null hypothesis that the nominal exchange rate follows a random walk (possibly with drift), the data generating process may be approximated by the model:

$$\begin{aligned}\Delta e_t - \mu_e &= u_{1t} \\ z_t - \mu_z &= \left(\exp\left\{\gamma \sum_{d=1}^{5} (z_{t-d} - \mu_z)^2\right\}\right)\left(\phi_1 (z_{t-1} - \mu_z) + (1-\phi_1)(z_{t-2} - \mu_z)\right) + u_{2t}\end{aligned}$$

where the innovations $u_t = (u_{1t}, u_{2t})'$ are assumed to be independently and identically distributed. The reason for postulating this particular statistical model is that it is consistent with our stylized economic model and that it embodies the nonlinear dynamics that we showed to be a prominent feature of the actual data.[9]

This system of equations may be estimated by nonlinear least squares. We treat the estimate of this process as the bootstrap data generating process. Bootstrap *p*-values for the long-horizon regression test statistics under the null may be obtained by generating repeated trials from this bootstrap data generating process, re-estimating the long-horizon regression test statistic for each set of bootstrap data, and evaluating the empirical distribution of the resulting long-horizon regression test statistics. A detailed description of the bootstrap algorithm can be found in the appendix.

The next section will apply this long-horizon regression test to exchange rate data for the UK, Germany, Japan, France, Switzerland, Canada and Italy. An obvious concern is the accuracy of the proposed bootstrap test under the null hypothesis that the exchange rate is indeed a random walk (possibly with a drift). While a comprehensive Monte Carlo study of the size of the test is beyond the scope of this paper, we will illustrate the accuracy of the proposed method for a representative data generating process of the form:

$$\begin{aligned}
\Delta e_t + 0.005 &= u_{1t} \\
z_t - 0.096 &= \left(\exp\left\{-0.7941\sum\nolimits_{d=1}^{5}(z_{t-d} - 0.096)^2\right\}\right)\left(1.2333(z_{t-1} - 0.096) + (1 - 1.2333)(z_{t-2} - 0.096)\right) + u_{2t}
\end{aligned}$$

where the parameter values correspond to the estimates obtained for the U.S.-German exchange rate data and the innovation vector u_t is obtained by random sampling with replacement from the actual regression residuals.

Figure 4 shows the effective size of the long-horizon regression test at the nominal 10% significance level for the actual sample size of 104. Approximate two-standard error bands for the rejection rates under the null of a 10% significance level are indicated by two horizontal lines. The statistics *t*(20) and *t*(A) refer to in-sample *t*-tests for the slope coefficient of the long-horizon regression. The two tests differ only in the computation of the standard error of the slope coefficient. The former test uses a Newey-West standard error based on a fixed truncation lag of 20; the latter uses a truncation lag based on Andrews's (1991) procedure. *DM*(20) and *DM*(A) refer to the corresponding Diebold-Mariano tests of out-of-sample forecast accuracy.

[9] Note that it is essential for our proposal to have a fully specified econometric model of the DGP motivated by economic theory. Our approach would not be valid in the presence of nonlinearities of unknown form.

The out-of-sample test are implemented based on a sequence of recursive forecasts, starting with a sample size of 32 quarters. The *joint* tests refer to tests of the random walk null against predictability at some horizon. They are based on the distribution of the maximum value of a given test statistic across all horizons. A detailed description of these tests can be found in Mark (1995).

The first panel of Figure 4 shows that even for sample sizes as small as 104 observations the bootstrap test is remarkably accurate. The effective size of all four tests is reasonably close to the nominal significance level of 10% and remains fairly constant across forecast horizons. This result means that any evidence of increased long-horizon predictability is unlikely to be caused by size distortions. Thus, we may have confidence in any evidence of increased long-horizon predictability in empirical work. The test is even more accurate, if we double the sample size, as shown in the second panel of Figure 4.

Next we will analyze the finite-sample power of the long-horizon regression test. The power of the test will in general depend on the alternative model. We will consider three examples of processes that may be considered empirically plausible under the joint alternative hypothesis of exchange rate predictability and nonlinear mean reversion in z_t. Modeling the power of the test requires an estimate of the joint DGP of $\{e_t, f_t\}$ or equivalently of $\{e_t, z_t\}$ or $\{f_t, z_t\}$. We clearly have little hope of correctly identifying the underlying complicated nonlinear dynamics of the nominal exchange rate from actual data. Instead, we focus on the easier task of finding a reasonable approximation to the time series process of the fundamental, f_t. For expository purposes we postulate that the DGP for z_t is the same as in the size study. Given the DGP for z_t, selecting a DGP for f_t will pin down the implied DGP for e_t by the identity $e_t \equiv z_t + f_t$. Our starting point is once again the U.S.-German data set. Preliminary tests did not reject the assumption that the German fundamentals follow a linear time series process. We selected the following three models as our DGPs:

DGP 1:

$$\Delta f_t = -0.0052 + u_{1t}$$

$$z_t - 0.096 = \left(\exp\left\{-0.7941\sum_{d=1}^{5}(z_{t-d} - 0.096)^2\right\}\right)\left(1.2333(z_{t-1} - 0.096) + (1 - 1.2333)(z_{t-2} - 0.096)\right) + u_{2t}$$

DGP 2:

$$\Delta f_t = -0.0030 - 0.1733\Delta f_{t-1} + 0.1419\Delta f_{t-2} + 0.1860\Delta f_{t-3} + 0.2417\Delta f_{t-4} + u_{1t}$$

$$z_t - 0.096 = \left(\exp\left\{-0.7941\sum\nolimits_{d=1}^{5}(z_{t-d} - 0.096)^2\right\}\right)\left(1.2333(z_{t-1} - 0.096) + (1 - 1.2333)(z_{t-2} - 0.096)\right) + u_{2t}$$

DGP 3:

$$\Delta f_t = -0.0039 + 0.2022\Delta f_{t-2} + 0.0643\Delta e_{t-2} + u_{1t}$$

$$z_t - 0.096 = \left(\exp\left\{-0.7941\sum\nolimits_{d=1}^{5}(z_{t-d} - 0.096)^2\right\}\right)\left(1.2333(z_{t-1} - 0.096) + (1 - 1.2333)(z_{t-2} - 0.096)\right) + u_{2t}$$

DGP 1 and DGP 2 were selected by the Schwarz information criterion and the Hannan-Quinn criterion, respectively, among the class of linear regressions of Δf_t on an intercept and up to eight autoregressive lags. DGP 3 was selected among all possible linear regressions involving up to 4 lags of Δf_t and Δe_t each and an intercept. The innovation vector u_t is again obtained by random sampling with replacement from the actual regression residuals.

The power of the long-horizon regression test against each of these alternatives is shown in Figure 5. All power results are based on the nominal 10% bootstrap test. As the actual test size is very close to the nominal size, there is no need for size corrections. In the first panel, the sample size is T = 104 as in the actual data. Figure 5 suggests several important conclusions. First, the proposed long-horizon regression test not only is highly accurate under the null of no exchange rate predictability, but has high power against empirically plausible alternatives, even in small samples. Second, whether the test is conducted in-sample or out-of-sample, Figure 5 suggests that our ability to predict the exchange rate will improve at intermediate horizons. The latter point is important because it provides the rationale for conducting long-horizon regression tests in practice. For example, for the three DGPs considered, the power of the long-horizon regression test tends to be lowest at the one-quarter horizon. As the forecast horizon is lengthened, power tends to improve initially, but ultimately falls again, resulting in a hump-shaped pattern with a peak at horizons of about one or two years. Third, power is considerably lower for recursive out-of-sample tests than for tests based on the full sample. The power of the in-sample tests is typically close to 90%, whereas the power of the out-of-sample tests is closer to 50% or 60%.

Why are we not able to beat the random walk model more often in real time when the null is false by construction? Part of the problem with our real-time exercise is the loss of power resulting from the small number of recursive forecast errors in the sample. Moreover, the small estimation sample underlying the out-of-sample exercise makes it unlikely that we obtain reliable estimates of the mean reversion parameter b_k. For example, at the beginning of the out-of-sample forecast exercise we use only 8 years worth of observations to construct the long-horizon forecast. Clearly, that may not be enough to capture nonlinear mean reversion. Increasing the initial sample size would seem to be the obvious solution, except that this increase in turn would further reduce the number of recursive forecast errors and thus would further lower the power of the out-of-sample test. Hence, short of obtaining a much larger sample, there is no obvious solution to the low power of the out-of-sample tests. As both types of tests are equally reliable under the null hypothesis, this evidence suggests that in empirical work the in-sample test of the random walk hypothesis will be preferable.

To confirm our interpretation that the much lower power of the out-of-sample test for T = 104 is an artifact of the sample size, we also experimented with a sample size of T = 208. The improvement in power is striking. The second panel of Figure 5 shows that the in-sample tests for T = 208 have power of virtually 100% for all horizons. For the out-of-sample tests, power is typically in the range of 96% to 99% with a peak at the 8 quarter horizon.

Although our power analysis is limited to three representative DGPs, we conclude that several qualitative implications of our model of nonlinear mean reversion are likely to be robust and can be tested empirically. First, if our model is supported by the data, the degree of predictability should be highest at intermediate horizons. Second, we expect to find less decisive empirical results for the out-of-sample tests than for the in-sample tests in our empirical work. This weaker evidence, however, need not indicate a failure of the model. It is fully expected given the lower power of out-of-sample tests in small samples. A third testable implication that emerges from the power analysis is that, to the extent that the random walk null hypothesis is false, the pattern of predictability for the in-sample and out-of-sample tests ought to be similar, even if the level of significance is much lower out-of-sample than in-sample.

6. Empirical Evidence of Long-Horizon Predictability Relative to the Random Walk Model

Figure 6 shows the bootstrap p-values for our four long-horizon regression tests of the random walk null. Separate results are shown for horizons of k = 1, 4, 8, 12 and 16 quarters. As the exchange rate becomes more predictable at longer horizons, these p-values should fall. The horizontal bar indicates the nominal significance level of 10%. Any p-value below 0.10 implies a rejection of the random walk null hypothesis at the 10% significance level. The results in Figure 6 are generally consistent with all three testable implications developed in section 5. Predictability generally is highest at intermediate horizons. The in-sample evidence is much stronger than the out-of-sample evidence, and the pattern of predictability across forecast horizons is broadly similar for in-sample and out-of-sample tests.

We will first focus on the results for the in-sample *t*-tests in columns 1 and 2. If our model of exchange rate determination is correct, we would expect to see a clear pattern of increased long-horizon predictability in the form of p-values that fall as the horizon grows. This is indeed what we find. There is little difference between the *t*(20) and *t*(A) test results, suggesting that the results are not sensitive to the choice of truncation lag. In virtually all cases, p-values fall as we increase k from 1 to 4 and 8. With the exception of France, we also find that p-values rise again for very long horizons, resulting in an U-pattern. This result is not surprising given the smaller effective sample size as the forecast horizon is lengthened. It is consistent with a loss of power at longer horizons, as suggested by Figure 5. In addition to the pattern of predictability, in many cases we find that the long-horizon regression is significantly more accurate than the random walk at longer horizons. For example, for k =12, we are able to reject the random walk model at the 10% significance level for six (five) of the seven countries using the *t*(A) (*t*(20)) test. In four (two) cases even the joint test statistic is significant at the 10% level. This number rises to six (four) out of seven if we focus on the 15% significance level. This evidence allows us for the first time to reject conclusively the random walk forecast model.

Does this result mean that we can also beat the random walk forecast in real time? The power study in Figure 5 suggests that beating the random walk model in real time will be much more difficult, given the smaller effective sample size. This is indeed what the empirical results suggest. Columns 3 and 4 show the corresponding p-values for the *DM* test of out-of-sample accuracy. These test results are based on recursive (or real-time) estimation of the forecast model starting with a sample size of 32 quarters.[10] Using a conventional significance level of 10 percent, with the exception of the U.K. and of Switzerland at the 3-year horizon, there is no

[10] Qualitative similar results are obtained with an initial sample size of 48 quarters. Note that the larger the sample size, the smaller is the number of recursive forecasts and the less reliable is the *DM* test. This tradeoff suggests that our choice of 32 quarters is a reasonable compromise.

evidence that the long-horizon regression beats the random walk. Moreover, none of the joint tests are significant at even the 15% level. This result is consistent with the evidence of a drastic loss of power in Figure 5 for the out-of-sample tests relative to the in-sample tests. There is, however, clear evidence for all seven countries that predictability improves as the forecast horizon is increased from one quarter to 1, 2 and 3 years, before deteriorating at the 4-year horizon. This pattern is generally similar to the pattern of the in-sample *t*-test p-values. The existence of a U-pattern in p-values is consistent with the hump-shaped power pattern we documented in Figure 5, although the locations and depth of the troughs suggest a somewhat different DGP than those that we considered in the power study.

We conclude that despite clear evidence of nonlinear mean reversion consistent with economic models, the goal of forecasting nominal exchange rates in real time is likely to remain elusive for the foreseeable future. Our analysis suggests that the difficulty of beating the random walk model in real time does not reflect a problem with the forecast model based on economic fundamentals; rather it is a natural consequence of the small time span of data available for empirical work. Our empirical results not only help us to understand the reluctance of applied forecasters to abandon chartists methods in favor of models based on economic fundamentals, especially at shorter horizons (Taylor and Allen, 1992), but they also lend support to economists' beliefs that the exchange rate is inherently predictable.

7. Concluding Remarks

The land-mark work of Meese and Rogoff (1983a, 1983b), published nearly two decades ago, launched the profession on a crusade to find the holy grail of beating the random walk model of exchange rates. Like the true Holy Grail, the goal of exploiting economic models of exchange rate determination to beat naïve *constant change* forecasts has remained elusive.

Alongside this difficulty in forecasting the nominal exchange rate or – almost equivalently – in distinguishing the nominal exchange rate from a random walk, researchers have also found it extremely difficult to reject a unit root in the real exchange rate. Moreover, even where researchers have been able to reject the unit root hypothesis for real exchange rates, using panel unit root tests or long spans of data, the apparent extreme persistence of the real exchange has remained puzzling (see Rogoff, 1996).[11] Recently, empirical evidence has been forthcoming that the relationship between the nominal exchange rate and the underlying fundamentals may be inherently nonlinear and that this finding may resolve these puzzles concerning the real exchange rate (Taylor and Peel, 2000; Taylor, Peel and Sarno, 2001). In the present paper, we have explored the question of whether evidence of nonlinearity in the relationship between the nominal exchange rate and macroeconomic fundamentals may also help to explain the difficulties in forecasting the nominal exchange rate.

We provided empirical support for threshold dynamics in the form of estimates of exponential smooth threshold autoregressive (ESTAR) models fitted to quarterly data on dollar exchange rates and PPP fundamentals for seven countries over the entire post Bretton Woods period. Our analysis sheds light on three central questions in exchange rate forecasting: 1) Do we understand why earlier exchange rate forecast models failed to beat the random walk? 2) Does this failure mean that we have to abandon standard economic models of exchange rate determination? 3) Why do professional exchange rate forecasters ignore economic models?

We showed that linear representations of the process that generates exchange rate data are fundamentally misspecified. As a result, the failure of previous research to provide conclusive evidence in favor of economic models of exchange rate determination is not surprising. In fact, our model suggests that close to the equilibrium the exchange rate will be well approximated by a random walk. It is only following large departures from equilibrium, that the mean reversion of the process becomes apparent. Thus, our model is also consistent with the tendency of spot exchange rate to respond to economic fundamentals during periods of hyperinflation, but much less so during normal times (see Frenkel 1976; Taylor 1995).

Allowing for smooth threshold nonlinearities in the econometric analysis goes a long way toward showing that economic models of the exchange rate are fundamentally sound. We proposed a new long-horizon regression test designed to detect nonlinear long-horizon predictability and provided strong empirical evidence against the random walk model. For example, based on in-sample tests, at the 3-year horizon we were able to reject the random walk model at the 10% level for five or six of the seven countries, depending on the choice of test

[11] As noted above, however, evidence based on panel unit root tests – is now seen as far less convincing than hitherto because of problems in the interpretation of these tests (Taylor and Sarno, 1998; Sarno and Taylor, 2001).

statistic. This evidence supports economists who have continued to use economic models to explain exchange rate fluctuations.

At the same time, our results also rationalize the reluctance of foreign exchange traders to rely on economic models of exchange rate determination, in particular for shorter horizons (see Allen and Taylor 1990, 1992; Taylor and Allen 1992; Cheung and Chinn 1999). We showed that it is difficult to exploit nonlinear threshold dynamics for real-time prediction. In practice, only unusually large departures from fundamentals will reveal the exchange rate's inherent tendency toward mean reversion and such events tend to be rare along a given sample path, resulting in low power, unless the sample size is large. Moreover, we found that nonlinear economic forecast models perform best at horizons of two to three years. In contrast, Frankel and Froot (1990) note that most foreign exchange traders appear to be interested in horizons of less than six months. There is no evidence that economic models improve forecast accuracy at such short horizons. We conclude that the goal of beating the random walk model in real time is likely to remain elusive for the foreseeable future.

References

Abuaf, N. and P. Jorion (1990), "Purchasing Power Parity in the Long Run," *Journal of Finance*, 45, 157-174.

Allen, H. and M.P. Taylor (1990), "Charts, Noise and Fundamentals in the Foreign Exchange Market," *Economic Journal*, 100, 49-59.

Allen, H. and M.P. Taylor (1992), "Chartist Analysis," in M. Milgate, P. Newman and J. Eatwell (eds.), *The New Palgrave Dictionary of Money and Finance*, London: MacMillan.

Andrews, D.W.K. (1991), "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817-858.

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

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

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

Cheung, Y.-W., and M.D. Chinn (1999), "Macroeconomic Implications of the Beliefs and Behavior of Foreign Exchange Traders," NBER Working Paper No. 7417.

Chinn, M.D., and R.A. Meese (1995), "Banking on Currency Forecasts: How Predictable is Change in Money?" *Journal of International Economics*, 38, 161-178.

De Grauwe, P., and H. Dewachter (1993), "A Chaotic Model of the Exchange Rate: The Role of Fundamentalists and Chartists," *Open Economies Review*, 4, 351-379.

De Long, J.B., A. Shleifer, L.H. Summers, and R.J. Waldmann (1990a), "Noise Trader Risk in Financial Markets," *Journal of Political Economy*, 98, 703-738.

De Long, J.B., A. Shleifer, L.H. Summers, and R.J. Waldmann (1990b), "Positive Feedback Investment Strategies and Destabilizing Rational Speculation," *Journal of Finance*, 45, 379-395.

Diebold, F.X., and R. Mariano (1995), "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics*, 13, 253-262.

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

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

Frankel, J.A., and K.A. Froot (1990), "Chartists, Fundamentalists and the Demand for Dollars," in A.S. Courakis and M.P. Taylor (eds.), *Private Behaviour and Government Policies in Interdependent Economies*, Oxford: Oxford University Press, 73-126.

Frenkel, J. (1976), "A Monetary Approach to the Exchange Rate: Doctrinal Aspects and Empirical Evidence," *Scandinavian Journal of Economics*, 78, 200-224.

Gallant, A.R. (1987), *Nonlinear Statistical Models*, New York: Wiley.

Gallant, A.R. and H. White (1988), *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*, Oxford: Oxford University Press.

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

Hegwood, N.D. and D.H. Papell (1998), "Quasi Purchasing Power Parity," *International Journal of Finance and Economics*, 3, 279-289.

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

Klimko, L.A., and P.I. Nelson (1978), "On Conditional Least Squares Estimation for Stochastic Processes," *Annals of Statistics*, 6, 629-42.

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

Lothian, J.R. and M.P. Taylor (1997), "Real Exchange Rate Behavior: The Problem of Power and Sample Size," *Journal of International Money and Finance*, 16, 945-954.

Mark, N.C. (1995), "Exchange Rates and Fundamentals: Evidence on Long-Horizon Predictability," *American Economic Review*, 85, 201-218.

Mark, N.C., and D. Sul (2001), "Nominal Exchange Rates and Monetary Fundamentals: Evidence from a Small Bretton Woods Panel," *Journal of International Economics*, 53, 29-52.

McCracken, M.W. (1999), "Asymptotics for Out-of-Sample Tests of Causality," manuscript, Department of Economics, Lousiana State University.

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

Meese, R. and K. Rogoff (1983b), “The Out-of-.Sample Failure of Empirical Exchange Rate Models: Sampling Error or Misspecification?” in J.A. Frenkel (ed.) *Exchange Rates and International Macroeconomics*, Chicago: Chicago University Press and National Bureau of Economic Research.

Rogoff, K. “The Purchasing Power Parity Puzzle,” *Journal of Economic Literature*, 34, 647-668.

Sarno, L. and M.P. Taylor (2001), “Deviations of Exchange Rates from Purchasing Power Parity: A Cautionary Note,” forthcoming: *International Monetary Fund Staff Papers*.

Shleifer, A., and L. Summer (1990), “The Noise Trader Approach to Finance,” *Journal of Economic Perspectives*, 4, 19-33.

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

Taylor, M.P., and H. Allen (1992), “The Use of Technical Analysis in the Foreign Exchange Market,” *Journal of International Money and Finance*, 11, 304-314.

Taylor, M.P. and P.C. McMahon (1998), “Long-Run Purchasing Parity in the 1920s,” *European Economic Review*, 32, 179-197.

Taylor, M.P., and D.A. Peel (2000), “Nonlinear Adjustment, Long Run Equilibrium and Exchange Rate Fundamentals,” *Journal of International Money and Finance,* 19, 33-53.

Taylor, M.P., D.A. Peel and L. Sarno (2001), “Nonlinear Adjustment in Real Exchange Rates: Towards a Solution to the Purchasing Power Parity Puzzles,” forthcoming: *International Economic Review.*

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.

Tjøstheim, D. (1986), “Estimation in Nonlinear Time Series Models,” *Stochastic Processes and Their Applications*, 21, 251-73.

Table 1: ESTAR Estimates by Country
PPP Fundamental

Canada

$$z_t = \left(\exp\left\{-\underset{\substack{(16.3074)\\[0.0000]}}{0.7060}\sum_{d=1}^{5} z_{t-d}{}^2\right\}\right)\left(\underset{(12.7201)}{1.1811} z_{t-1} + (1 - \underset{(12.7201)}{1.1811}) z_{t-2}\right) + \hat{u}_t$$

R^2=0.96 s=0.0192 DW=2.03 AR(1)=[0.09] AR(1-4)=[0.11]

France

$$(z_t - \underset{(2.8544)}{0.0954}) = \left(\exp\left\{-\underset{\substack{(5.75312)\\[0.0015]}}{0.8638}\sum_{d=1}^{5}(z_{t-d} - \underset{(2.8544)}{0.0954})^2\right\}\right)\left(\underset{(13.8396)}{1.3219}(z_{t-1} - \underset{(2.8544)}{0.0954}) + (1 - \underset{(13.8396)}{1.3219})(z_{t-2} - \underset{(2.8544)}{0.0954})\right) + \hat{u}_t$$

R^2=0.91 s= 0.0473 DW=1.96 AR(1)=[0.65] AR(1-4)=[0.19]

Germany

$$(z_t - \underset{(2.4076)}{0.0960}) = \left(\exp\left\{-\underset{\substack{(6.3540)\\[0.0012]}}{0.7941}\sum_{d=1}^{5}(z_{t-d} - \underset{(2.4076)}{0.0960})^2\right\}\right)\left(\underset{(11.6493)}{1.2333}(z_{t-1} - \underset{(2.4076)}{0.0960}) + (1 - \underset{(11.6493)}{1.2333})(z_{t-2} - \underset{(2.4076)}{0.0960})\right) + \hat{u}_t$$

R^2=0.90 s= 0.0530 DW=1.91 AR(1)=[0.91] AR(1-4)=[0.31]

Italy

$$z_t = \left(\exp\left\{-\underset{\substack{(3.6754)\\[0.0028]}}{0.9092}\sum_{d=1}^{5} z_{t-d}{}^2\right\}\right)\left(\underset{(10.2805)}{1.1540} z_{t-1} + (1 - \underset{(10.2805)}{1.1540}) z_{t-2}\right) + \hat{u}_t$$

R^2=0.87 s= 0.0540 DW=1.91 AR(1)=[0.82] AR(1-4)=[0.21]

(continued….)

(… Table 1 continued)

Japan

$$z_t = \left(\exp\left\{-\underset{\substack{(7.7301)\\[0.0010]}}{0.7256}\sum_{d=1}^{5} z_{t-d}^{\;2}\right\}\right)\left(\underset{(13.5378)}{1.3500}z_{t-1} + (1-\underset{(13.5378)}{1.3500})z_{t-2}\right) + \hat{u}_t$$

R^2=0.94 s= 0.0571 DW=1.79 AR(1)=[0.15] AR(1-4)=[0.19]

Switzerland

$$z_t = \left(\exp\left\{-\underset{\substack{(5.2450)\\[0.0019]}}{0.7242}\sum_{d=1}^{5} z_{t-d}^{\;2}\right\}\right)\left(\underset{(13.0796)}{1.2922}z_{t-1} + (1-\underset{(13.0796)}{1.2922})z_{t-2}\right) + \hat{u}_t$$

R^2=0.88 s= 0.0599 DW=1.93 AR(1)=[0.45] AR(1-4)=[0.65]

United Kingdom

$$z_t = \left(\exp\left\{-\underset{\substack{(3.3603)\\[0.0031]}}{1.0696}\sum_{d=1}^{5} z_{t-d}^{\;2}\right\}\right)\left(\underset{(11.1052)}{1.1448}z_{t-1} + (1-\underset{(11.1052)}{1.1448})z_{t-2}\right) + \hat{u}_t$$

R^2=0.86 s= 0.0520 DW=1.96 AR(1)=[0.45] AR(1-4)=[0.21]

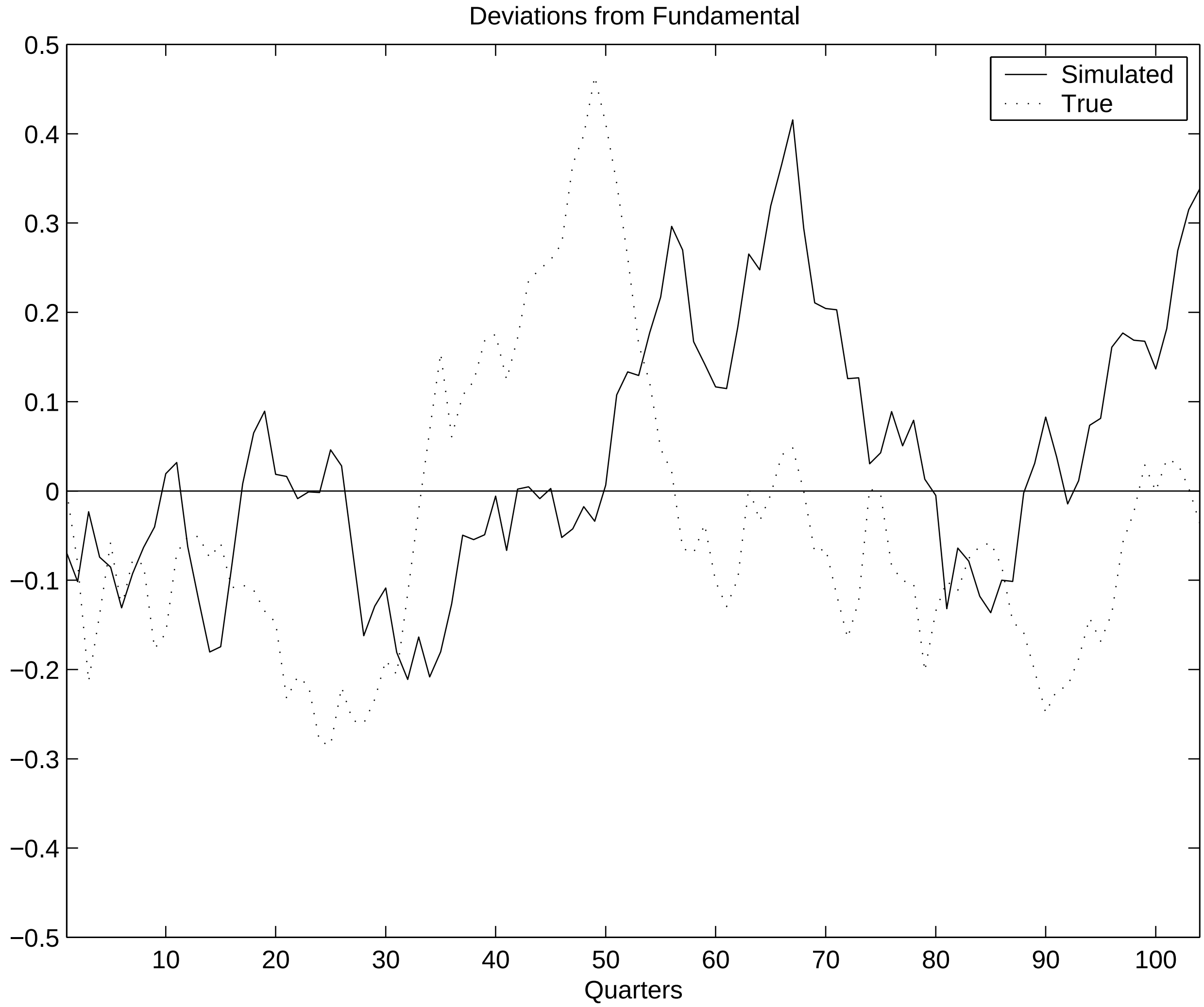
Notes: R^2 denotes the coefficient of determination, s is the standard error of the regression. AR(1) and AR(1-4) are Lagrange multiplier test statistics for first-order and up to fourth-order serial correlation in the residuals respectively, constructed as in Eitrheim and Teräsvirta (1996). Figures in parentheses below coefficient estimates denote the ratio of the estimated coefficient to the estimated standard error of the coefficient estimate. Figures given in square brackets denote marginal significance levels. The marginal significance levels for the estimated transition parameters were calculated by a non-parametric bootstrap under the null hypothesis of a unit root AR(2) process.

Table 2: Estimated Half Lives in Quarters

Shock:	20%	10%	5%	1%
Canada	5	10	14	16
France	4	9	12	13
Germany	4	9	12	14
Italy	4	9	12	14
Japan	5	10	13	15
Switzerland	5	10	14	16
United Kingdom	4	9	12	13

Notes: The half lives are the number of complete quarters taken for a shock of size $\log_e(1+x/100)$ to the logarithm of the real exchange rate (equivalent to a shock of x% to the level of the real exchange rate) to decline to 0.5 $\log_e(1+x/100)$. They were estimated by Monte Carlo integration, using a non-parametric bootstrap approach, conditional on the average initial history - as described in Taylor, Peel and Sarno (2001) - and on the estimated models in Table 1.

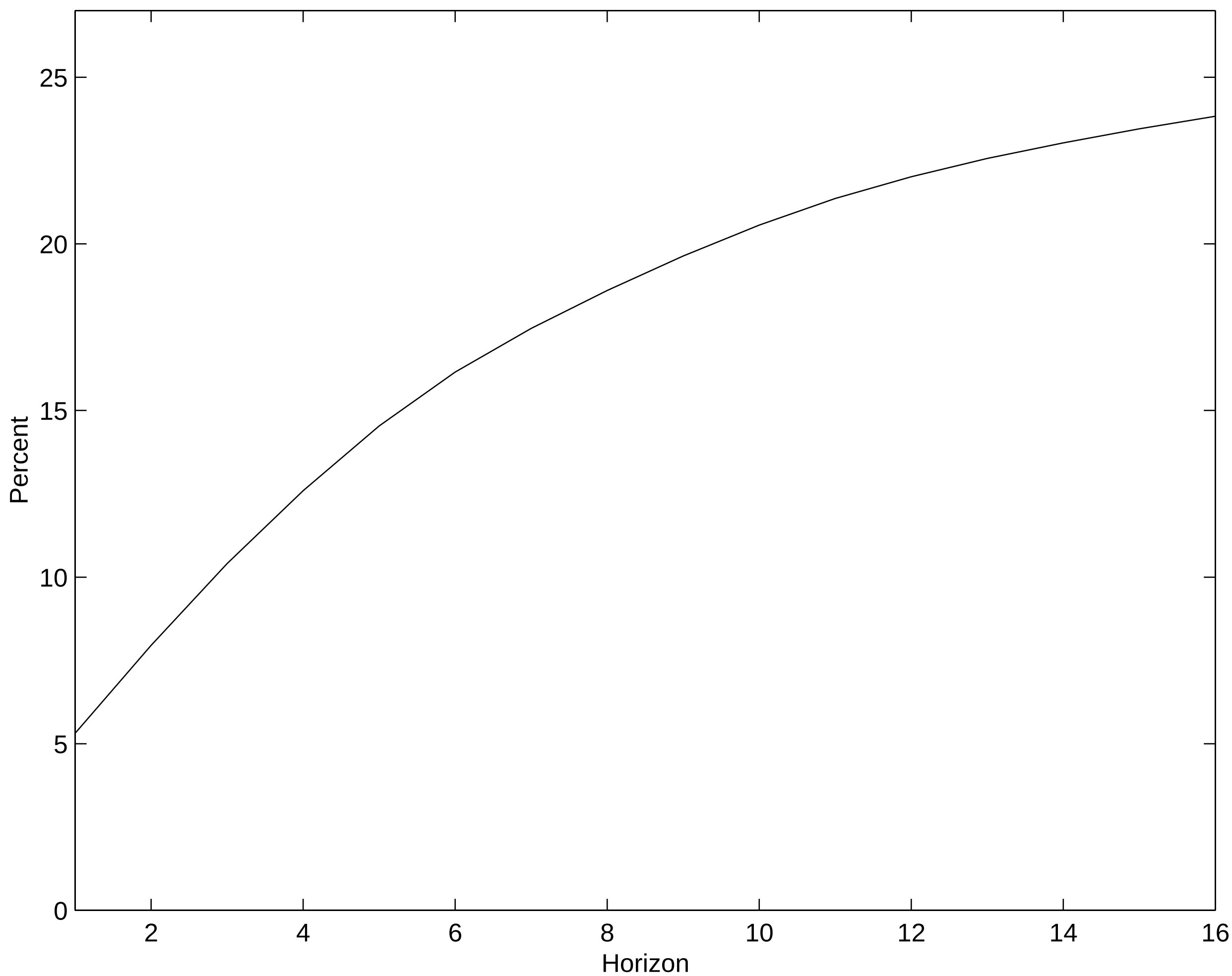
Figure 1: Actual and Simulated Real Exchange Rates
Based on a Representative Random Draw
From the Fitted ESTAR Model for the Deutsche Mark-Dollar Rate



Notes: The simulated data were generated by a nonparametric bootstrap approach based on the fitted model for Germany in Table 1.

Figure 2: Response Surface for Forecast Accuracy Gains of ESTAR Model Relative to Random Walk Model

Benchmark Model with Known Parameters



Notes: The benchmark model is described in the text. The RPMSE of the ESTAR model was estimated by Monte Carlo integration.

Figure 3: Response Surface for Forecast Accuracy Gains of ESTAR Model Relative to Random Walk Model

Sensitivity Analysis of Benchmark Model with Known Parameters

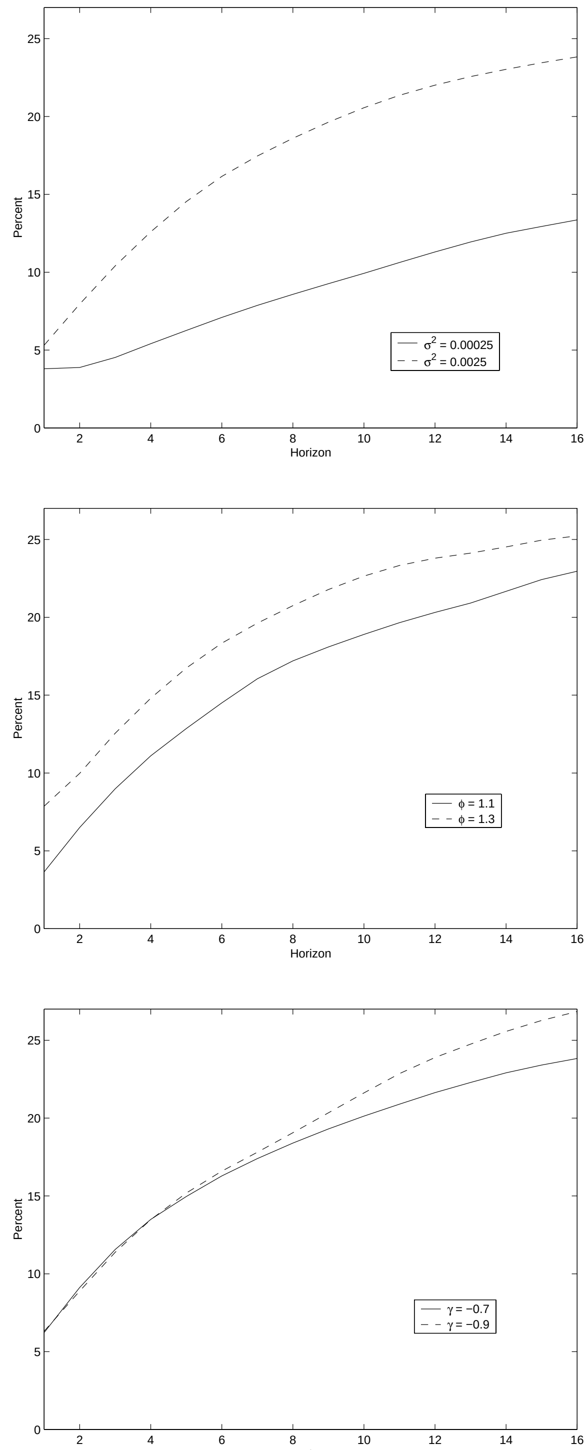
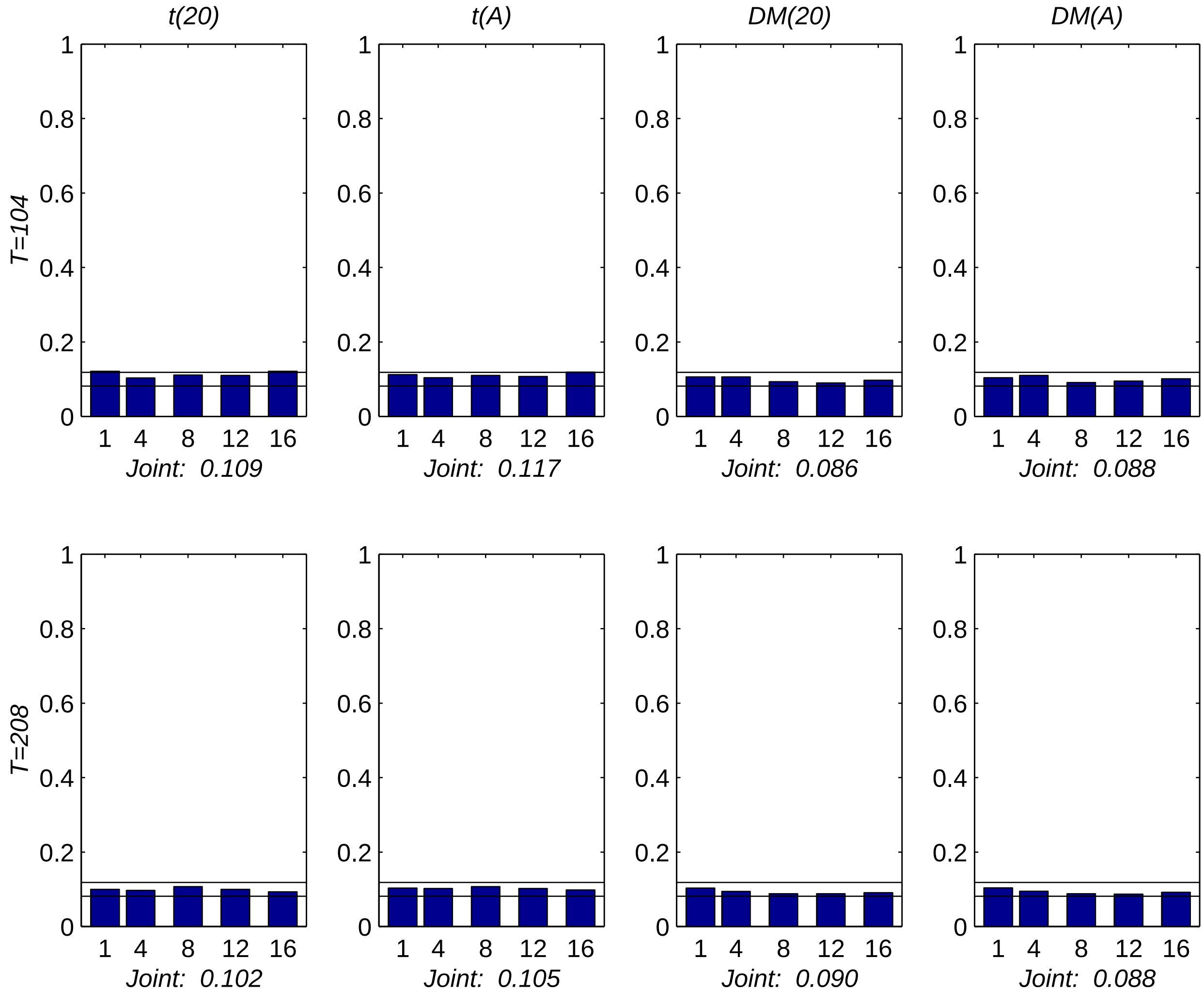


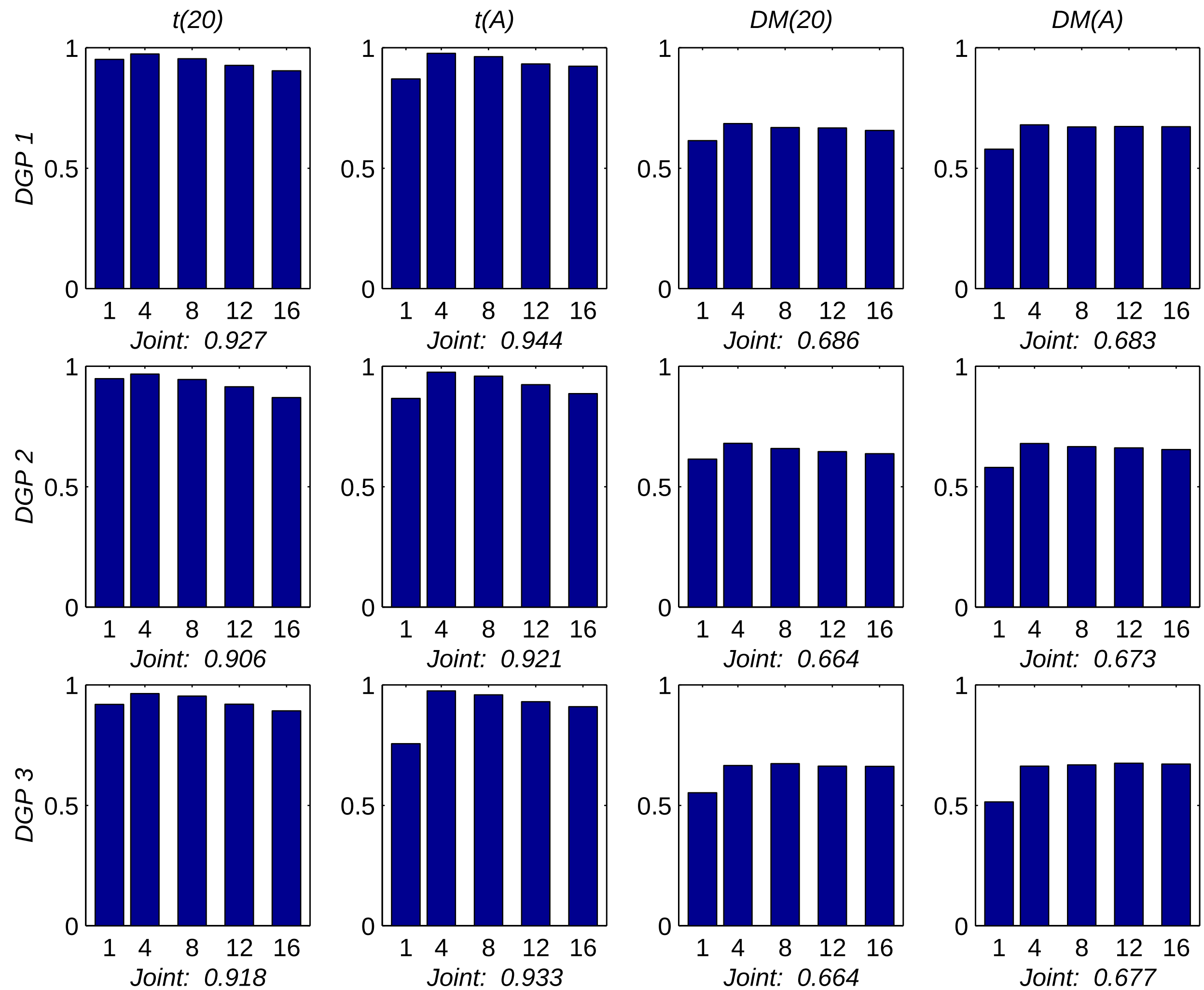
Figure 4: Effective Size of Bootstrap Test Procedure under ESTAR Null



Notes: Based on 1000 Monte Carlo trials with 2,000 bootstrap replications each. The DGP is described in the text.

Figure 5: Power of Bootstrap Test Procedure at the 10% Significance Level

(a) T = 104



Notes: Based on 1000 Monte Carlo trials with 2,000 bootstrap replications each. The DGP for z_t is the same as for the size analysis, but the DGP for f_t differs:

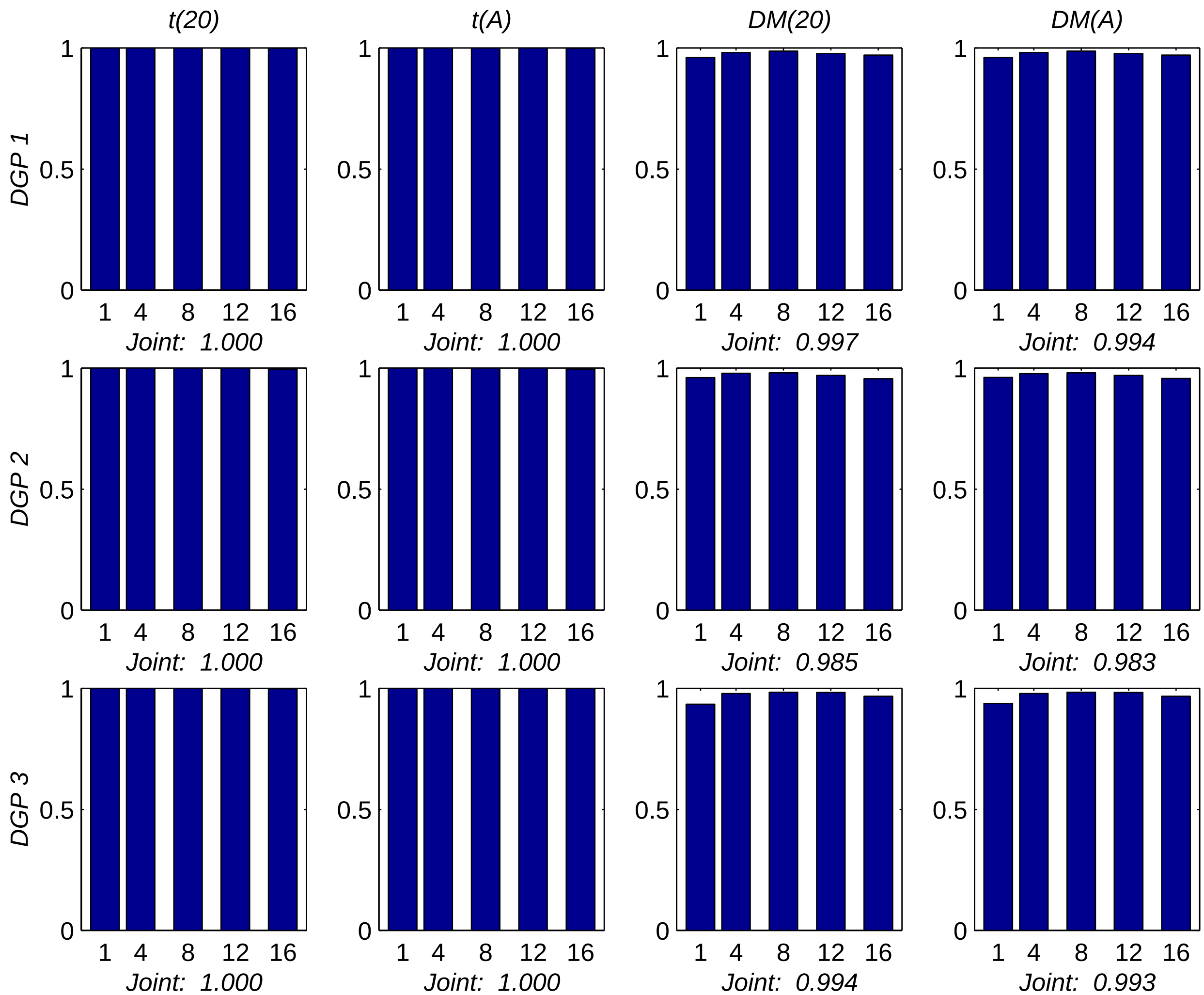
DGP1: Δf_t regressed on an intercept

DGP2: Δf_t regressed on 4 lags of itself and an intercept.

DGP3: Δf_t regressed on an intercept, Δf_{t-2} and Δe_{t-2}.

Figure 5 (contd.)

(b) T = 208



Notes: Based on 1000 Monte Carlo trials with 2,000 bootstrap replications each. The DGP for z_t is the same as for the size analysis, but the DGP for f_t differs:

DGP1: Δf_t regressed on an intercept

DGP2: Δf_t regressed on 4 lags of itself and an intercept.

DGP3: Δf_t regressed on an intercept, Δf_{t-2} and Δe_{t-2}.

Figure 6: Bootstrap p-Values under ESTAR Null
PPP Fundamental

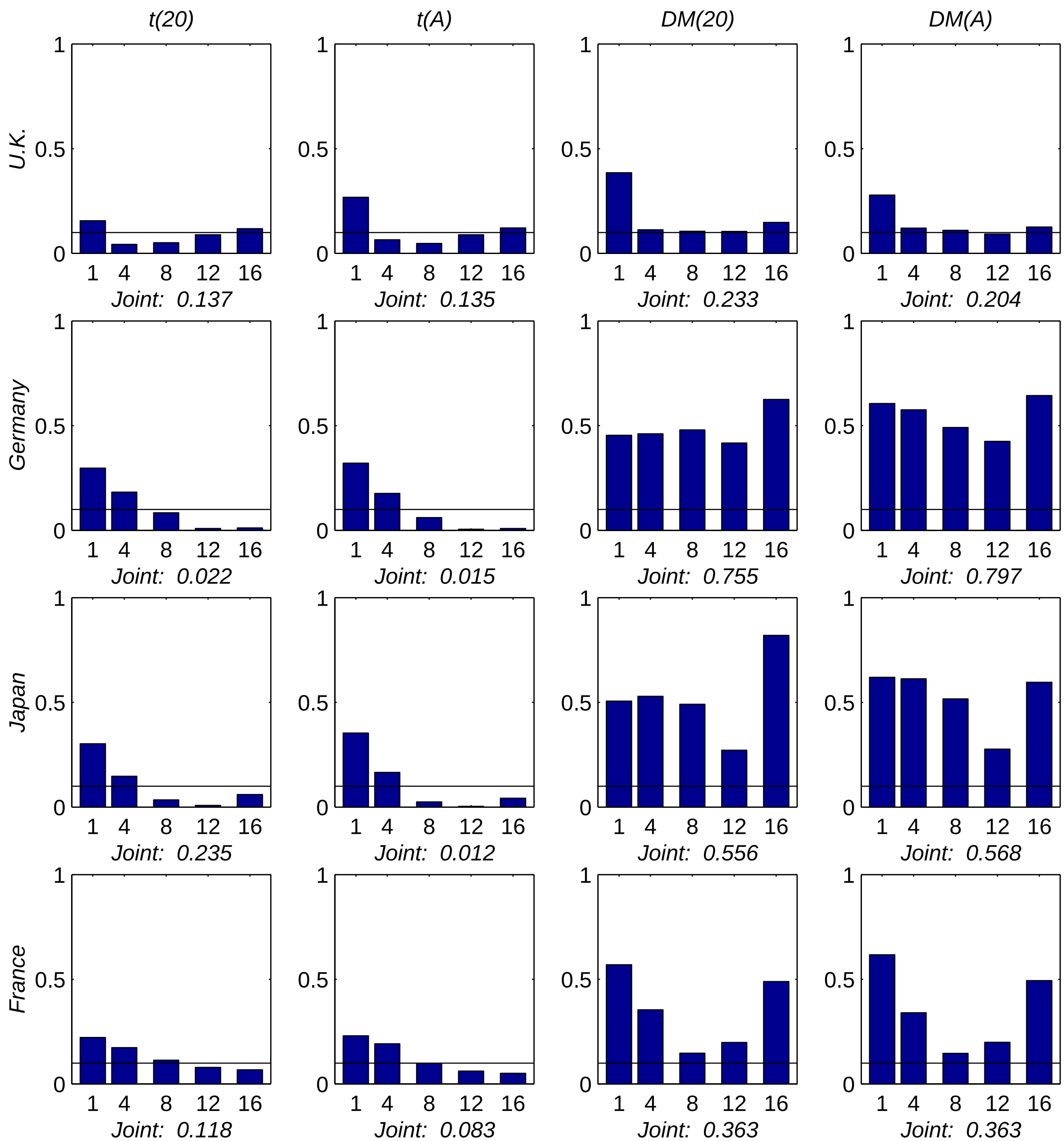
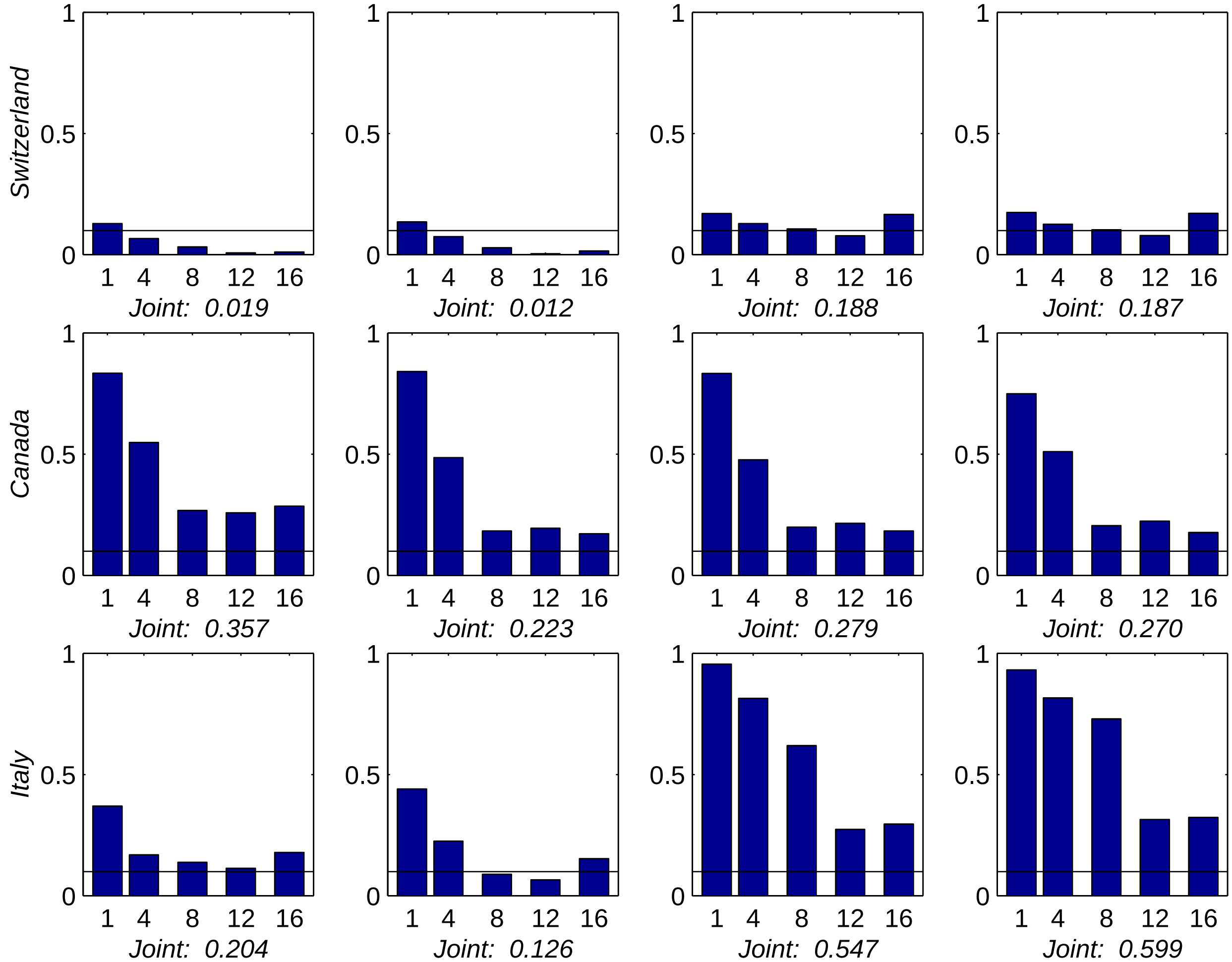


Figure 6 (contd.)



Notes: Based on 2,000 bootstrap replications. Quarterly IFS data for 1973.I-2000.IV.

APPENDIX: BOOTSTRAP ALGORITHM FOR LONG-HORIZON REGRESSION TEST

1. Given the sequence of observations $\{x_t\}$ where $x_t = (e_t, f_t)$ρ, define $z_t \equiv e_t - f_t$, estimate the long-horizon regression

$$e_{t+k} - e_t = a_k + b_k z_t + \varepsilon_{t+k}, \qquad k = 1, 4, 8, 12, 16,$$

and for each k construct the test statistic $\hat{\theta}_k$.

2. Postulate a nonlinear data generating process of the form

$$\Delta e_t - \mu_e = u_{1t}$$
$$z_t - \mu_z = \left(\exp\left\{\gamma \sum_{d=1}^{5} (z_{t-d} - \mu_z)^2\right\}\right)\left(\phi_1 (z_{t-1} - \mu_z) + (1 - \phi_1)(z_{t-2} - \mu_z)\right) + u_{2t}$$

where the restriction that the exchange rate follows a random walk under H_0 has been imposed and the innovations $u_t = (u_{1t}, u_{2t})$ρ are assumed to be zero mean, independent and identically distributed. Estimate this process by nonlinear least-squares.

3. Based on the fitted model generate a sequence of pseudo observations $\{x_t^*\}$ of the same length as the original data series $\{x_t\}$, where $x_t^* = (e_t^*, f_t^*)$ρ is obtained from the realizations of the bootstrap data generating process:

$$\Delta e_t^* - \hat{\mu}_e = u_{1t}^*$$
$$z_t^* - \hat{\mu}_z = \left(\exp\left\{\hat{\gamma} \sum_{d=1}^{5} (z_{t-d}^* - \hat{\mu}_z)^2\right\}\right)\left(\hat{\phi}_1 (z_{t-1}^* - \hat{\mu}_z) + (1 - \hat{\phi}_1)(z_{t-2}^* - \hat{\mu}_z)\right) + u_{2t}^*$$

The pseudo innovation term $u_t^* = (u_{1t}^*, u_{2t}^*)$ρ is random and drawn with replacement from the set of recentered residuals $\tilde{u}_t = \hat{u}_t - T^{-1} \sum_{t=p+1}^{T} \hat{u}_t$. The process may be initialized with $z_{t-j}^* = 0$ and $\Delta e_{t-j}^* = 0$ for $j = p, \ldots, 1$. We discard the first 500 transients.

3. Repeat the preceding step 2,000 times. For each of the 2,000 bootstrap replications $\{x_t^*\}$

estimate the long-horizon regression

$$e^*_{t+k} - e^*_t = a_k + b_k z^*_t + \varepsilon^*_{t+k}, \qquad k = 1, 4, 8, 12, 16$$

and construct the test statistics of interest, $\hat{\theta}^*_k$.

4. Use the empirical distribution of the 2,000 replications of the bootstrap test statistic $\hat{\theta}^*_k$ to determine the *p*-value of the test statistic $\hat{\theta}_k$, where $k = 1, 4, 8, 12, 16$.