Journal of International Money and Finance 28 (2009) 1045–1063

Contents lists available at ScienceDirect

Journal of International Money and Finance

journal homepage: www.elsevier.com/locate/jimf

New evidence on nominal exchange rate predictability

Jyh-Lin Wu [a,b,*], Yu-Hau Hu [c]

[a] Institute of Economics, National Sun Yat-Sen University, Taiwan
[b] Department of Economics, National Chung-Cheng University, Chia-Yi, Taiwan
[c] Department of International Trade, Cheng-Shiu Technological University, Kaohsiung, Taiwan

JEL classification:
F31
F47
C53

Keywords:
Purchasing power parity
Nominal exchange rates
Real exchange rates
Random walks
Long-horizon regression tests

A B S T R A C T

The Meese–Rogoff puzzle, one of the well-known puzzles in international economics, concerns the weak relationship between nominal exchange rates and market fundamentals. The purpose of this paper is to show that market fundamentals do in fact matter in forecasting nominal exchange rates. In particular, we emphasize the importance of the Harrod–Balassa–Samuelson effect in modeling deviations from purchasing power parity. Based on the post-Bretton Woods period, we provide solid out-of-sample evidence that rejects the random walk forecast model at medium-term and long-term forecast horizons. We also find mild evidence for out-of-sample predictability of nominal exchange rates over the short term.

1. Introduction

Since the publication of the seminal paper by Meese and Rogoff (1983), the predictability of exchange rates has been the subject of an ongoing scholarly debate in empirical international finance and has inspired a large volume of papers over the past two decades. Meese and Rogoff (1983) compare the predictive ability of a variety of exchange rate models and conclude that no existing structural exchange rate models can reliably beat random walks at short- or medium-term forecast horizons in out-of-sample forecast contests. Their finding is robust to assumptions of different fundamentals such as purchasing power parity (PPP) and uncovered interest rate parity, as well as to the use of the realized value of fundamentals in the forecast period and is now known as the Meese–Rogoff Puzzle. This puzzle is also sometimes referred to as the exchange rate disconnect puzzle (Obstfeld and Rogoff,

* Corresponding author. Institute of Economics, National Sun Yat-Sen University, 70 Lien-hai Road, 804 Kaohsiung, Taiwan.
E-mail addresses: jlwu2@mail.nsysu.edu.tw, ecdjlw@ccu.edu.tw (J.-L. Wu), hincent@csu.edu.tw (Y.-H. Hu).

0261-5606/$ – see front matter
doi:10.1016/j.jimonfin.2008.12.008

2000). Meese and Rogoff's findings are quite striking since a random walk model is not embedded within any economic wisdom or theory. The goals of research on exchange rate predictability over the past two decades have been to uncover the reasons which explain the Meese–Rogoff puzzle and to provide evidence which rejects the random walk forecast model.

A number of authors have found evidence which beats random walk forecasts in out-of-sample contests (e.g. Chinn and Meese, 1995; Mark, 1995; MacDonald and Marsh, 1997). The empirical procedure and the robustness of results in the previously mentioned literature have been challenged by a number of authors (e.g. Cheung et al., 2005; Kilian, 1999; Berkowitz and Giorgianni, 2001; Berben and van Dijik, 1998; Rossi, 2005). Recently several authors have applied the panel approach to show evidence of co-integration between exchange rates and fundamentals and then have provided evidence which beats random walk forecasts (e.g. Mark and Sul, 2001; Groen, 2000). However, these articles fail to examine the significance of deviations between two forecast errors.

One possible explanation for the Meese–Rogoff puzzle is that the linear forecasting model fails to capture important non-linearities in the data. However, a number of authors have found that allowing for regime switching in exchange rate models does not improve the out-of-sample predictability of the models (e.g. Engel and Hamilton, 1990; Engel, 1994). Other forms of non-linearity have also been found to be largely unimportant for exchange rates (e.g. Diebold and Nason, 1990; Meese and Rose, 1991). Recent empirical work has supported non-linear, mean-reverting adjustment of real exchange rates and has shown that the exponential smooth transition autoregressive (ESTAR) model provides a parsimonious fit to the data (e.g. Michael et al., 1997; Taylor et al., 2001; Taylor and Peel, 2000). Given the fact that real exchange rates follow an ESTAR process, Kilian and Taylor (2003) provide in-sample evidence to beat the random walk forecast at horizons of 2–3 years, but their out-of-sample evidence is fairly weak. Based on simulation results, Kilian and Taylor (2003) argue that the reason for the poor out-of-sample predictability of their model may be due to the short sample period available for empirical analysis.

Most existing literature which applies the ESTAR model to describe real exchange rate dynamics assumes that the long-run equilibrium of real exchange rates is time invariant and hence ignores the effects of real factors on the equilibrium of real exchange rates. Several articles argue that productivity differentials between countries affect real exchange rates (e.g. Harrod, 1939; Balassa, 1964; Samuelson, 1964). The Harrod–Balassa–Samuelson (HBS) effect suggests that, under some assumptions, fast growing economies will experience a rising relative price of non-tradables and hence a real appreciation over time. In this case, deviations from PPP will revert to an equilibrium trend instead of a constant mean implying that the real exchange rate would exhibit a trend behavior if one takes the Harrod–Balassa–Samuelson (HBS) effects into account.[1] Based on the idea of differential productivity growth in tradables and non-tradables, Obstfeld (1993) develops a simple stochastic model in which real exchange rates contain a pronounced deterministic trend. Kilian and Taylor (2003) argue that the HBS effect is significant when long historical data are used, but may not be significant in the shorter post-Bretton Woods period. Their arguments are also underscored by recent empirical findings with long historical data (e.g. Lothian, 1990; Cuddington and Liang, 2000; Lothian and Taylor, 2000, 2008; Taylor, 2002; Peel and Venetis, 2003; Taylor and Taylor, 2004). However, Bergin et al. (2006) point out that "the HBS effect has not always been a fact of economic life, and appears to be a phenomenon of only the postwar period." Their empirical evidence reveals that the effect virtually vanishes from the data if one looks back fifty years or more. Several recent studies also point out the significance of the HBS effect on equilibrium real exchange rates over the post-Bretton Woods period, which indicates the significance of the HBS effect even in relatively short spans of data (Paya et al., 2003; Paya and Peel, 2003; Sollis, 2005).[2]

[1] There are a number of papers pointing out the limitations of the theory pronounced by Harrod–Balassa–Samuelson. Notwithstanding these limitations, the theory is useful in explaining long-run trend deviations from PPP. Asea and Corden (1994) provide an interesting overview on the Harrod–Balassa–Samuelson theory.

[2] Sollis (2005) finds that real exchange rates for most industrialized countries are stationary around a gradually changing deterministic trend function. Paya et al. (2003) adopt an ESTAR model with a non-constant equilibrium to describe the dynamics of real exchange rates within the European Monetary System. Both deterministic trends and relative price of non-tradables are applied to proxy for the HBS effect. Paya and Peel (2003) find that the estimated half-life of the dollar–yen real exchange rate is less than 2 years based on a non-linear model which incorporates a deterministic trend to approximate the equilibrium level.

If the HBS effect is significant and is neglected in an empirical exchange rate model, then the estimation would suffer from omitted-variable bias, which could in turn result in the failure of the model in its out-of-sample predictability. The purpose of this paper is two-fold. First, we combine the non-linear adjustment of real exchange rates and the HBS effect in a model and then apply it to address the issue of nominal exchange rate predictability. Second, we investigate the significance of the HBS effect in enhancing exchange rate forecasts by examining whether the failure to reject the random walk forecast as observed in most existing literature can be explained solely by the restriction of short sample period.

Using the data over the post-Bretton Woods period for several industrialized countries, we obtained the following significant findings. First, our model fits the data reasonably well and shows that real effects (such as the HBS effect) on the equilibrium real exchange rate are important even in relatively short spans of data. Second, we provide strong out-of-sample evidence which rejects the random walk forecast of nominal exchange rates at horizons of 2–4 years. We also find mild evidence for out-of-sample predictability of nominal exchange rates when forecast horizons are less than one year. These results are robust to several newly developed statistical tests, to different sample periods and to different initial windows of estimation. Our findings are illuminating since they indicate that taking the HBS effect into account in an ESTAR model can help to strengthen the recursive out-of-sample predictability of the long-horizon regression equation.[3] Using a less parsimonious model we obtain evidence of out-of-sample predictability. To the best of our knowledge, our paper is the first paper that provides solid out-of-sample evidence to reject random walk forecasts at short-term, medium-term and long forecast horizons.[4]

Third, simulation results point out that our bootstrap tests have correct size and good power given the modified ESTAR dynamics of real exchange rates. There is no indication that the power of bootstrap tests increases with forecast horizons. The contribution of the paper is to complement the existing literature by providing new out-of-sample evidence on the predictability of nominal exchange rates, which enriches our understanding of the Messe–Rogoff puzzle.

The article proceeds as follows. Section 2 provides a brief discussion of the HBS effect and several issues in its empirical application. Section 3 describes the estimation results of our model. Section 4 offers the bootstrap tests provided by Kilian (1999) and Kilian and Taylor (2003) given the modified ESTAR dynamics of real exchange rates. The size and power analysis of the bootstrap tests are given in Section 5. Section 6 concludes our discussion.

2. The Harrod–Balassa–Samuelson effect

The classical model of the HBS effect implies that the relative price of non-tradable goods in terms of tradable goods (or real exchange rates) is determined entirely by the production technology. The HBS effect relies on the following four assumptions. First, there are two symmetric countries in the world and each country has two, traded and non-traded, sectors. Both factor and final goods markets are perfectly competitive. Second, production takes place under constant returns to scale. Third, capital is perfectly mobile internationally. Finally, labor is internationally immobile but mobile between the tradable and non-tradable sectors.

Under these assumptions, one can derive the following equation for real exchange rates:

$$z = (1-\theta)\left[\pi\left(a_{\mathrm{T}}^{*}-a_{\mathrm{T}}\right)-\left(a_{\mathrm{N}}^{*}-a_{\mathrm{N}}\right)+b\right],$$

where z, a_{N} (a_{T}), θ are the real exchange rate, the productivity in the non-tradable (tradable) sector, and

3 In the case of recursive forecasts, we use all the data to determine the model specification and then divide the data sample to generate forecasts.

4 There are several papers providing out-of-sample evidence to reject random walk forecasts, which either fail to examine the statistical significance of the difference between two forecast errors (Mark and Sul, 2001) or fail to construct the finite sample distribution of the DM statistic (Chinn and Meese, 1995; MacDonald and Marsh, 1997; Clarida et al., 2003) or fail to beat random walk forecasts when forecast horizons are short or medium term (Mark, 1995).

the weight of the traded good price in the price index, respectively.[5] b is a constant, π is a parameter, and the superscripted '*' indicates a foreign variable. Based on the above equation, it can be shown that the relative price of non-traded goods depends only on the technology of traded and non-traded sectors, and a positive productivity shock to the domestic traded sector leads to a real appreciation of domestic currency. This explains why rich countries (or countries with a relatively high level of productivity) will tend to have a higher exchange rate-adjusted price level on average.

The empirical evidence of the HBS effect provides mixed results (see, e.g., the recent survey studies by Froot and Rogoff, 1995 and Sarno and Taylor, 2002). Bergin et al. (2006) point out that the HBS effect virtually disappears if we look back fifty years or more. However, several articles, based on OECD countries over the period after 1950, provide only weak evidence of the HBS effect (Froot and Rogoff, 1991; Asea and Mendoza, 1994; Fitzgerald, 2003). The conventional explanation of the previous findings is that since the HBS effect relies upon relative productivity differentials, it would apply better to price differentials between developed and developing countries rather than to those between developed countries. Recently, Bergin et al. (2006) utilize a model with a continuum of goods differentiated by productivity, monopolistic competition, transaction costs and endogenous tradability to examine the stylized fact of the HBS effect.[6] Their model implies that the HBS effect should be more noticeable even if countries have only a small gap in their GDPs. In addition, some articles also support the significance of the HBS effect in OECD countries during the period after 1973 (Paya and Peel, 2003; Paya et al., 2003; Sollis, 2005). This evidence justifies our empirical analysis based on major industrialized countries during the post-Bretton Woods period.

Empirically, a common, straightforward strategy to model the impact of the HBS effect is to allow for a linear, deterministic trend in a real exchange rate process. However, the linear trend strategy is not theoretically supported by Bergin et al. (2006). They point out that the innovation of endogenous tradability plus the heterogeneous productivity growth allows us to alter the impact of the single shock process resulting in a non-linear effect on real exchange rates. Moreover, Lothian and Taylor (2000, 2008) point out that a cubic trend specification is appropriate for the real sterling–dollar rates if the time trend proxies for the HBS effect. This leads us to speculate upon the significance of non-linear trends in capturing the HBS effect. In this paper the impact of the HBS effect on real exchange rates is modeled with a non-linear deterministic trend.

3. The estimation of the ESTAR model

To combine the HBS effect and the non-linear ESTAR dynamics of real exchange rates into a model, we specify an ESTAR model with a non-constant equilibrium for the real exchange rate. Let s_t, p_t, and p_t^* be the logarithm of the spot nominal exchange rates (US dollar per foreign currency), domestic (US) and foreign consumer price indices, respectively, and let $f_t \equiv p_t - p_t^*$ denote the PPP fundamental. The real exchange rate is defined as the deviation of the nominal exchange rate from relative prices (the PPP fundamental). Hence $z_t = s_t - f_t$ is the real exchange rate. Following Kilian and Taylor (2003), we apply an ESTAR(2) model to describe its non-linear dynamics:

$$z_t = g_t + [\alpha_1(z_{t-1} - g_{t-1}) + (1-\alpha_1)(z_{t-2} - g_{t-2})]F[z_{t-d}, \gamma, g_{t-d}] + u_t, \tag{1}$$

$$F(z_{t-d}, \gamma, g_{t-d}) = \exp\left[\gamma(z_{t-d} - g_{t-d})^2\right], \quad u_t \sim \text{i.i.d.}\left(0, \sigma^2\right),$$

where, g_t denotes the equilibrium of the real exchange rate. If one neglects the HBS effect, then g_t is time invariant and hence assumed to be a constant (Kilian and Taylor, 2003); otherwise, g_t is time varying and assumed to be a polynomial up to a cubic trend: $g_t = a_0 + a_1 t + a_2 t^2 + a_3 t^3$. The inclusion

[5] A detailed derivation and discussion of the HBS effect can be found in Balassa (1964), Samuelson (1964), Mark (2001) and Lothian and Taylor (2008).

[6] Let's assume that the pre-condition of the HBS effect is that technological shocks hit traded sectors. Suppose shocks hit non-traded sectors initially, then those which receive positive technology shocks and pay for transaction costs become traded sectors, which meets the pre-condition of the HBS effect and the HBS effect arises endogenously.

of a non-linear deterministic trend is crucial since it not only agrees with recent arguments by Bergin et al. (2006) and Lothian and Taylor (2000, 2008), but also allows empirical results from our model to have a straightforward comparison with those found in Kilian and Taylor (2003).

The term $z_t - g_t$ can be viewed as short-run fluctuations of the real exchange rate around its long-run equilibrium path, g_t. The transition function, $F[z_{t-d}, \gamma, g_{t-d}]$, is a function of the lagged real exchange rate and its lagged long-run equilibrium, which captures the non-linearity of the model. The parameter γ (<0) denotes the speed of transition between regimes and d is the delay lag.[7]

Given the exponential transition function, F is bounded by zero and one, and the value of F is close to one when the deviation between z_{t-d} and g_{t-d} is small. In this case, the real exchange rate is highly persistent. As departures from equilibrium increase, F moves toward zero and hence real exchange rates follow a stationary AR(2) process.

Another popular model for capturing the non-linearity of real exchange rates is the regime switching model that allows for occasional discrete changes in exchange rate behavior. Spurred by Hamilton (1989), the Markov switching model is widely applied to model exchange rates (Bergman and Hansson, 2005; Siddique and Sweeney, 1998; Engel, 1994; Engel and Hamilton, 1990).[8] In these models, the number of parameters to be estimated grows rapidly with the number of real exchange rate regimes. As a result, usually only two or three regimes are allowed, which is due to the fact that estimating many parameters implies a lower power.[9] Cheung and Erlandsson (2005) demonstrate that data frequency and sample length are crucial for determining the number of regimes.[10] The estimation of Markov switching models may be spurious without a formal test for the existence of multiple regimes. The ESTAR model allows a smooth transition between regimes so that there can be a continuum of states between regimes. This specification can be thought of as a model in the spirit of Markov switching where the probability of switching between regimes is a function of an observed variable instead of an unobserved state variable. The ESTAR model provides a flexible approach to allow for regime switching and has the advantage of simplicity in estimation, and hence is appealing in dealing with the dynamics of financial markets. Recently, there have been a number of articles which lend support to the appropriateness of the ESTAR model in modeling real exchange rate dynamics (Taylor et al., 2001; Michael et al., 1997).

The quarterly data of the consumer price index and nominal exchange rate for Germany (GER), France (FRA), Italy (ITA), Japan (JAP), Canada (CAN), Switzerland (SWI), and the United Kingdom (UK) over the period of the first quarter of 1973 to the last quarter of 1998 were obtained from IMF's International Financial Statistics. The data frequency and sample period in our empirical analysis are the same as those in Kilian and Taylor (2003).

The specification of the trend for each country is determined by sequentially testing the significance of the high-order trend term in the model. In other words, we first include t, t^2, and t^3 in the model and then estimate the model with the method of non-linear least squares. If the coefficient of t^3 is significant, then we report our estimation results; otherwise, we exclude the t^3 term from the trend function and then re-estimate the model with t and t^2 in the trend function. If there is no evidence for including a trend term in the model, we then assume a time invariant mean in our ESTAR model. The lag order d in the model is based on the model that works well in terms of goodness of fit, statistical significance of parameters, and diagnostic tests of residuals. Based on the above sequential strategy, we adopt a linear trend function for JAP, a quadratic trend function for GER and FRA, a cubic trend function

[7] If the transition parameter equals zero, then the model degenerates to a linear unit-root process. Hence, testing the significance of the transition parameter is equivalent to testing the hypothesis that the real exchange rate follows a linear unit-root process against the alternative of a mean-reverting ESTAR process.

[8] On the other hand, several authors point out that Markov switching models for exchange rates are unstable over time or the forecast performance of the models is sensitive to misclassification of regimes (Marsh, 2000; Dacco and Satchell, 1999) and implies that Markov switching models are not suitable for forecasting exchange rates.

[9] One reason for the absence of a formal test for the number of regimes results from the fact that the commonly used test statistics do not have their usual asymptotic distributions (Cheung and Erlandsson, 2005).

[10] Cheung and Erlandsson (2005) present a systematic and extensive empirical study on the presence of Markov switching dynamics in three dollar-based exchange rates and find that monthly data instead of quarterly data offer unambiguous evidence of the presence of Markov switching dynamics.

for CAN and ITA, and a constant mean for UK and SWI. The lag order, d, is set to be five for JAP, GER, UK, SWI, and FRA, and four for ITA, and CAN.

Paya and Peel (2003) point out that the asymptotic distribution of trend coefficients in an ESTAR model is non-standard; we, therefore, construct a finite sample distribution of t-statistic through bootstrap based on their method.[11] Estimation results from the best fitting model are reported in Table 1, indicating that the estimated coefficients are all significant, and that the model fits the data reasonably well for all countries. Rows 3–5 of Table 1 report the t-statistics of trend coefficients and their p-values constructed through bootstrap, which reveals the significance of the trend coefficient at the 10% level in different countries except for UK and SWI.

Conventional wisdom indicates that the economic model at the heart of the HBS effect would apply better to developing economies instead of developed economies. It is interesting to find that the coefficients in the trend function are significant at conventional levels in five out of seven industrialized countries indicating that the HBS effect is significant in explaining the movement of real exchange rates. In addition, results from Table 1 indicate significant non-linear trend movements in real exchange rates between the dollar and the currencies of major industrial countries (except for JAP, UK and SWI). These findings are interesting since they support the contention that the HBS effect could be significant in major industrialized countries and imply that the relative productivity differentials between the US and other major industrialized countries vary over time.

After comparing the corresponding trend coefficients for those countries that have a similar trend pattern, we find that their absolute values are similar to each other. This indicates that the impact of the HBS effect on real exchange rates (captured by trend terms) is similar for the countries in our sample. These countries are all developed countries and have relatively small GDP gaps between each other. It is therefore reasonable to find a similar impact of the HBS effect on real exchange rates. In addition, the intercept of the trend function differs across countries indicating initial heterogeneous technology among them.

It is worth noting that the asymptotic distribution for the t-statistic of the transition parameter ($\gamma = 0$) is not conventional as pointed out by Taylor and Peel (2000). We therefore construct the empirical marginal significance levels for the transition parameter through a non-parametric bootstrap method under the null hypothesis of a unit-root AR(2) process (AR(3) process for UK) with either a linear, a non-linear or no trend.[12] Based on the bootstrapped critical values, we reject the hypothesis that the transitional parameter is zero for all countries which indicates the existence of a non-linear adjustment of real exchange rates. As for residual diagnostics, there is no serial correlation and autoregressive conditional heteroscedasticity (ARCH) in residuals at the 5% level as indicated by Q and LM statistics, and Q^2 and ARCH statistics, respectively.[13] There is also no evidence of a misspecification of the model in estimated residuals at the conventional level of significance for all countries as indicated by the RESET statistic.[14] Overall, our findings support the appropriateness of an ESTAR model with the HBS effect in describing the dynamics of real exchange rates for all countries except UK and SWI.[15]

[11] The data-generating process under the null hypothesis is the ESTAR model without trends, where the parameters of the process are taken from Table 1 with the trend coefficient being set at zero. The empirical significance levels were based on 5000 simulations of length 204, from which the first 100 were in each case discarded (leaving 104 data points, corresponding to the size of our data set). At each replication an ESTAR equation with trends was estimated for each artificial data set.

[12] Based upon the findings in Table 1, we adopt a linear trend function for JAP, a quadratic trend function for GER and FRA, a cubic trend function for CAN and ITA and a constant mean for UK and SWI.

[13] The Q and LM statistics are the Ljung–Box autocorrelation test and the Breusch–Goldfrey test for serial correlation of residuals. The ARCH statistic is the Lagrange multiplier test proposed by Engle (1982) for the autoregressive conditional heteroscedasticity of residuals. The Q^2 statistic is the Q statistic of squared residuals that is applied to examine the ARCH effect in residuals.

[14] The RESET statistic is the Ramsey's regression specification error test.

[15] We also plot the actual and simulated real exchange rates for GER, and the figure indicates that simulated real exchange rates from the fitted ESTAR model reveal not only long swings in real exchange rates, but also short-run volatility. The figure is not reported here but is available upon request from authors.

Table 1
The estimation of ESTAR models with the HBS effect: 1973Q1–1998Q4.

$$z_t = g_t + [\alpha_1(z_{t-1} - g_{t-1}) + \alpha_2(z_{t-2} - g_{t-2}) + (1 - \alpha_1 - \alpha_2)(z_{t-3} - g_{t-3})]F[z_{t-d}, \gamma, g_{t-d}] + u_t,$$
$$F(z_{t-d}, \gamma, g_{t-d}) = \exp\left[\gamma(z_{t-d} - g_{t-d})^2\right], \quad u_t \sim \text{i.i.d.}(0, \sigma^2),$$
$$g_t = a_0 + a_1 t + a_2 t^2 + a_3 t^3$$

	JAP	UK	GER	ITA	CAN	SWI	FRA
a_0	−5.20	0.379	−0.223	−7.184	0.109	−0.409	−1.404
	(−106.70)	(18.16)	(−3.14)	(−99.10)	(3.73)	(−14.037)	(−26.64)
a_1	0.006	–	−0.015	−0.023	−0.021	–	−0.015
	[0.001]		**[0.021]**	**[0.067]**	**[0.011]**		**[0.003]**
a_2	–	–	1.4×10^{-4}	5×10^{-4}	4.4×10^{-4}	–	1.3×10^{-4}
			[0.018]	**[0.066]**	**[0.013]**		**[0.006]**
a_3	–	–	–	-2.5×10^{-6}	-2.8×10^{-6}	–	–
				[0.088]	**[0.011]**		
α_1	1.120	1.156	1.069	1.153	1.001	1.015	1.126
	(10.27)	(10.89)	(9.68)	(10.49)	(9.53)	(9.796)	(10.25)
α_2	–	−0.334	–	–	–	–	–
		(−2.13)					
γ	−4.36	−4.49	−5.01	−5.35	−22.89	−2.365	−4.65
	[0.04]	**[0.01]**	**[0.01]**	**[0.05]**	**[0.06]**	**[0.012]**	**[0.01]**
Q(12)	0.91	0.37	0.49	0.43	0.20	0.72	0.78
Q^2(12)	0.37	0.82	0.84	0.61	0.83	0.86	0.94
LM(1)	0.40	0.09	0.05	0.71	0.66	1.00	0.81
LM(2)	0.64	0.16	0.13	0.24	0.72	0.53	0.94
ARCH(1)	0.43	0.90	0.25	0.55	0.93	0.67	0.69
ARCH(2)	0.71	0.19	0.42	0.62	0.99	0.78	0.90
RESET(1)	0.49	0.15	0.10	0.06	0.67	0.08	0.41
RESET(2)	0.76	0.15	0.21	0.15	0.49	0.15	0.11

Note: The number in a parenthesis (square bracket) under an estimate is its t-statistic (p-value). LM(p) and Q(p) are respectively the Lagrange multiplier and the Ljung–Box autocorrelation tests for up to a pth-order autocorrelation. They are χ^2 distributions with p degrees of freedom. ARCH(p) is a test statistics for up to a pth-order autoregressive conditional heteroscedasticity. It has χ^2 distribution with p degrees of freedom. RESET is the Ramsey's regression specification error test that has an F distribution. The p-values for the estimated transition parameter, γ, are constructed based on a non-parametric bootstrap. The dash, "–", indicates that a statistic is not constructed. The number in boldface indicates significance at the 10% level.

Based upon previous findings, we examine the possibility of whether or not a model based on PPP fundamentals rejects the random walk forecast model of nominal exchange rates. Following Kilian and Taylor (2003), we specify the model for the nominal exchange rate change (Δs_t) under the hypothesis that it is unpredictable. The dynamic process of real exchange rates (z_t) is described based on the fitted ESTAR model in Table 1.

4. The bootstrap tests of long-horizon predictability

Given the fact that our model fits the data well, we evaluate the predictive accuracy of the long-horizon regression equation relative to a random walk model. The long-horizon regression equation is described as follows:

$$s_{t+k} - s_t = \alpha_k + \beta_k\left(z_t - \widehat{g}_t\right) + \varepsilon_{t+k}, \quad k = 1, 4, 8, 12, 16, \tag{2}$$

where $\widehat{g}_t$ is the estimate of g_t. The specification in Eq. (2) is different from the conventional long-horizon regression equation in which z_t instead of $z_t - \widehat{g}_t$ is used. The reason for using a trend adjusted series ($z_t - \widehat{g}_t$) is that the long-run equilibrium of real exchange rates is affected by the HBS effect which is approximated by a non-linear trend. The non-predictability of nominal exchange rates can be examined by testing the hypothesis of H_0: $\beta_k = 0$ versus H_1: $\beta_k < 0$ for a given forecast horizon k. The estimated residuals from Eq. (2) are serially correlated if k is greater than one. We therefore apply the method provided by Newey and West (1987) to construct the autocorrelation consistent estimate of the covariance matrix.

Following Mark (1995) and Kilian and Taylor (2003), two different truncation lags are used to construct the Newey–West covariance matrix. One is to arbitrarily set the truncation lag at 20 and the other is to select the lag order based on Andrews' (1991) procedure. Corresponding to these two lag truncations in computing the standard error of the slope coefficient in the long-horizon regression equation, the t-statistics of the slope coefficient are denoted by $t(20)$ and $t(\mathrm{A})$, respectively. We also apply a joint test to examine whether the smallest t-ratio among the five horizons, $t(j)_{\min} = \min\{t_k(j): j = 20, \mathrm{A};\ k = 1, 4, 8, 12, 16\}$, is significant.[16]

To evaluate the out-of-sample predictability of a model based on the PPP fundamental relative to the random walk with drift forecast model, we obtain a sequence of recursive forecasts from the long-horizon regression equation and the random walk model, respectively.[17, 18] Following Kilian and Taylor (2003), the first 32 quarters are reserved for estimation, and hence the out-of-sample forecast period starts from the first quarter in 1981. We then apply the DM statistic provided by Diebold and Mariano (1995) to examine the null hypothesis of no difference in the accuracy of two competing forecasts.[19] Here, we define the loss differential under a given forecast horizon k: $d_t^k = (u_{\mathrm{R}t,k}^2 - u_{\mathrm{L}t,k}^2)$, where $u_{\mathrm{L}t,k}$ and $u_{Rt,k}$ are date-t forecast errors from the long-horizon regression equation and the random walk forecast model, respectively. The t-type DM statistic under a given forecast horizon is given as follows:

$$\mathrm{DM}_k = \frac{\bar{d}^k}{\sqrt{2\pi\hat{f}^k(0)/N_f^k}}, \tag{3}$$

where $\bar{d}^k$ is the average of the loss differentials under a given forecast horizon k, N_f^k is the number of k-step ahead forecasts, and $\hat{f}^k(0)$ is the Newey–West estimate of the spectral density of the loss differential function, d_t^k at frequency 0. Corresponding to the two lag truncations used in Newey–West estimate, the DM statistics are denoted by DM(20) and DM(A). Again, a joint test is applied to examine whether the largest DM statistic among the five horizons, $\mathrm{DM}(j)_{\max} = \max\{\mathrm{DM}_k(j): j = 20, \mathrm{A};\ k = 1, 4, 8, 12, 16\}$, is significant.

The t-type DM test as found in Eq. (3) is designed for forecasts that do not rely on regression estimates and hence are not subject to parameter estimation error. To account for parameter estimation uncertainty, Clark and McCracken (2001, 2004) provide the following F-type test to examine the hypothesis of equal forecast accuracy under a given forecast horizon k:

$$\mathrm{GM}_k = N_f^k \frac{\bar{d}^k}{\mathrm{MSE}_{\mathrm{L}}^k}, \tag{4}$$

where $\mathrm{MSE}_{\mathrm{L}}^k = (N_f^k)^{-1}\sum_{j=1}^{N_f^k}\hat{u}_{\mathrm{L},j}^2$ is the mean squared error from the long-horizon regression equation under a given forecast horizon k. Gilbert (2001) considers the same statistic in the split-sample context.

[16] A detailed description of the joint test can be found in Mark (1995).

[17] Kilian (1999) highlights the importance of being careful about the distinction between pure random walk and random walk with drift hypotheses. Since the null bootstrap model given in Eq. (5) includes a drift in the nominal exchange rate, we therefore allow a drift in the random walk forecast model.

[18] There are two conventional schemes in constructing out-of-sample forecasts. The rolling scheme fixes the estimation window size and drops distant observations as recent ones are added. The recursive scheme uses all available data. We follow Mark (1995) and Kilian and Taylor (2003) to construct recursive forecasts based on the long-horizon prediction equation and random walk with drift, respectively. In other words, we re-estimate Eq. (2) in each rolling sample with expanding length and then use the updated parameters to predict nominal exchange rates under different forecast horizons. Different out-of-sample statistics are constructed based on these forecasts.

[19] Rossi (2005) points out that the long-horizon regression forecasts are biased by the estimation error when nominal exchange rates and fundamentals are highly persistent, but not exactly co-integrated and therefore provides a test for equal accuracy under this case. In our paper the PPP fundamental is adopted and the deviation from the fundamental is the real exchange rate. We find that the size of the bootstrap test adopted in our paper is accurate even when the real exchange rate is highly persistent as shown in our simulation results presented in Section 5 of the paper. We therefore do not apply Rossi's statistic to investigate the out-of-sample predictability of nominal exchange rates.

Inoue and Kilian (2004) refer to Eq. (4) as the Gilbert–McCracken (GM) test statistic. Although both the GM and DM tests are based on the average loss of differentials, the key difference between these two tests is that the GM test uses a different normalization designed to account for parameter estimation uncertainty in the forecast models. Moreover, Clark and McCracken (2004) point out that, based on a linear structure, the power of the GM test is superior to that of the DM test when two models are nested. A joint test is also applied to examine whether the largest GM statistic among the five horizons, $GM(j)_{max} = \max\{GM_k(j): j = 20, A; k = 1, 4, 8, 12, 16\}$, is significant.

Recently, Clark and West (2006) point out that under the hypothesis of no predictive ability of a variable in the alternative model, the mean squared prediction error (MSPE) of the null model (σ_1^2) should be smaller than that of the alternative model (σ_2^2). Therefore they suggest using the MSPE adjusted series to test the hypothesis of no predictability in nested models. The test examines whether the adjusted mean squared error difference is zero, which is defined as follows:

$$\mathrm{MSPEA}^k = \sigma_{1,k}^2 - \left(\sigma_{2,k}^2 - \mathrm{adj}_k\right),$$

where $\sigma_{1,k}^2$, $\sigma_{2,k}^2$ and adj_k are the sample average of $u_{\mathrm{R}t,k}^2$, $u_{\mathrm{L}t,k}^2$ and $(u_{\mathrm{L}t,k} - u_{\mathrm{R}t,k})^2$, respectively. We then test whether MSPEA is significantly different from zero. Let's define:

$$w_{t+k} = u_{\mathrm{R}t,k}^2 - \left[u_{\mathrm{L}t,k}^2 - \left(u_{\mathrm{L}t,k} - u_{\mathrm{R}t,k}\right)^2\right].$$

We regress w_{t+k} on a constant and use the resulting t-statistic (CW_k) for a zero coefficient. The Newey–West method is applied to construct an autocorrelation consistent standard error. A joint test is also applied to examine whether the largest CW statistic among the five horizons, $CW(j)_{max} = \max\{CW_k(j): j = 20, A; k = 1, 4, 8, 12, 16\}$, is significant.

The least square estimate of β_k is biased (Stambaugh, 1986), and the limiting distributions of the DM, GM and CW statistics are non-standard when models are nested (Clark and McCracken, 2001, 2004; Clark and West, 2006, 2007).[20] We therefore apply the bootstrap strategy provided by Kilian and Taylor (2003) to simulate the finite sample distribution of the DM, GM and CW statistics.[21] Following Kilian and Taylor (2003), we first generate bootstrapped data using the following restricted data-generating process (DGP):

$$\Delta s_t - c = \varepsilon_{1t}, \tag{5}$$

$$z_t = g_t + [\alpha_1(z_{t-1} - g_{t-1}) + (1 - \alpha_1)(z_{t-2} - g_{t-2})]\exp\left[\gamma(z_{t-d} - g_{t-d})^2\right] + \varepsilon_{2t},$$

where g_t is a trend function that is determined based on the empirical results in Table 1.

The hypothesis of the exchange rate following the process of random walk with drift is imposed in the DGP. The innovations of the system $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t})'$ are assumed to be independently and identically distributed. The coefficients of α_1,γ and the estimated coefficients in the trend function are obtained from Table 1. Second, after generating the bootstrapped data, we estimate Eq. (2) and then construct t_k, GM_k, DM_k, CW_k, t_{min}, GM_{max}, DM_{max} and CW_{max} statistics, respectively. Finally, repeating the previous procedures 2000 times, we obtain the empirical distribution for all statistics, respectively.

Fig. 1 reports the marginal significance level (p-values) of all statistics for each country at different forecast horizons of 1, 4, 8, 12, and 16 quarters under the hypothesis that nominal exchange rates follow

[20] The small sample bias of estimates arises from the fact that regressors are highly persistent and their innovations are highly correlated with return innovations.

[21] Clark and West (2007) point out that the limiting distribution of the CW statistic under the null hypothesis is standard normal when the estimation is performed using a rolling regression. However, the asymptotic distribution needs to be simulated when the estimation is performed recursively. They suggest that, for recursive specifications, one-tail standard normal critical values can be applied to test the null hypothesis of equal forecasting power. This is because their simulation results show that normal standard critical values result in little size distortion. Moreover, their simulation results also show that the critical values from bootstrap distribution lead to even smaller size distortion. We therefore construct the finite sample distribution of the CW statistic based on the bootstrap method.

the process of a random walk with drift. The horizontal line in the figure indicates the nominal significance level of 10% and any p-value below it indicates the rejection of the random walk hypothesis at the 10% level of significance. A number on the bar indicates a p-value of a test. The in-sample and out-of-sample results are reported in the first two and last five columns, respectively. Several interesting findings can be observed in Fig. 1. First, evidence from in-sample test indicates the superior predictability of the long-horizon regression relative to the random walk model for JAP, GER, CAN and SWI. Evidence from the joint test also indicates similar findings. Among these four countries, we also observed that there is a clear pattern showing that the p-value decreases as the forecast horizon increases. These findings indicate increased in-sample predictability as the forecast horizon increases, which is consistent with other work (Mark, 1995 and Engel and West, 2005). Engel and West (2005) point out that the near random walk behavior of the exchange rate is manifest if the discount factor is close to one and fundamentals follow a unit-root process. Their findings suggest the difficulty of predicting exchange rates with short horizons.

Second, the p-values from the last five columns of Fig. 1 indicate that the out-of-sample predictability of nominal exchange rates is significant for the countries JAP, UK, GER, CAN, and SWI when forecast horizons are long (sixteen quarters) or medium term (eight quarters). The out-of-sample predictability, based on DM and GM statistics, is observed for CAN at all forecast horizons, for JAP, UK and GER when forecast horizons are not shorter than eight quarters, and for ITA (SWI) when the forecast horizon is one (eight and twelve) quarter. Based on the CW statistic, we find support for the out-of-sample predictability for CAN at all forecast horizons, for JAP (SWI) when forecast horizons are not shorter than eight (twelve) quarters and for GER when the forecast horizon is eight and twelve quarters. The joint test provides evidence of out-of-sample predictability for JAP, UK, GER and CAN based on DM and GM statistics, and for JAP, CAN, SWI when the CW statistic is applied. Frankel and Rose (1995) and Taylor (1995) point out that there is no fundamental-based exchange rate model available that is capable of beating random walk forecasting model at short term. Our findings of out-of-sample predictability over short term for CAN are therefore important, which also indicate the significance of combining both the HBS effect and non-linearity in modeling the dynamics of real exchange rates. Third, there is no significant difference between in- and out-of-sample evidence of the predictability of nominal exchange rates, and the pattern of superiority in prediction across forecast horizons is similar for in-sample and out-of-sample tests.

We are able to reject the random walk forecast for three or four out of seven countries with different joint tests, but there is no rejection in Kilian and Taylor (2003) with the joint DM test.[22] Our findings are interesting since they point out that the evidence of beating random walk forecasts will be strengthened, given the short spans of data, if the HBS effect is allowed for in modeling real exchange rate dynamics.

Are the previous findings sensitive to the length of the initial estimation window? To address this issue, we repeat the previous out-of-sample bootstrap tests with four different initial windows: 32, 40, 50 and 60 quarters, and report the results of a joint test in Table 2.[23] It is worth noting that the number of out-of-sample forecasts decreases as the length of initial window increases, which in turn reduces the power of out-of-sample bootstrap tests. Findings from Table 2 indicate that the random walk forecasts are rejected by the GM test for JAP at different initial windows, for CAN when the length of the initial window is 32, 40 and 50 quarters, for SWI when the initial window is 40, 50 and 60, and for GER with the initial window of 32 and 40 quarters. Therefore, results from the GM test are robust to different initial windows in general. DM tests reject random walk forecasts for JAP and CAN when the initial window is 32 and 40 quarters. As for CW tests, the superiority of random walk forecasts was

[22] Kilian and Taylor (2003) apply only the DM statistic in their out-of-sample test and find no evidence to reject the random walk forecasts. We also re-investigate the out-of-sample predictability of nominal exchange rates with DM, GM and CW statistics based on the Kilian–Taylor model over the same period as in Fig. 1. We find that the hypothesis of non-predictability of nominal exchange rates is not rejected based on the joint DM test. The hypothesis is rejected only for UK (CAN) with the p-value of 0.063 (0.063) when the joint GM (CW) test is applied. Empirical results are not reported here but are available upon request from authors.

[23] We report joint statistics in Table 2 to save space. Similar findings are observed based on results from different statistics at different forecast horizons, which are not reported here but are available upon request from authors.

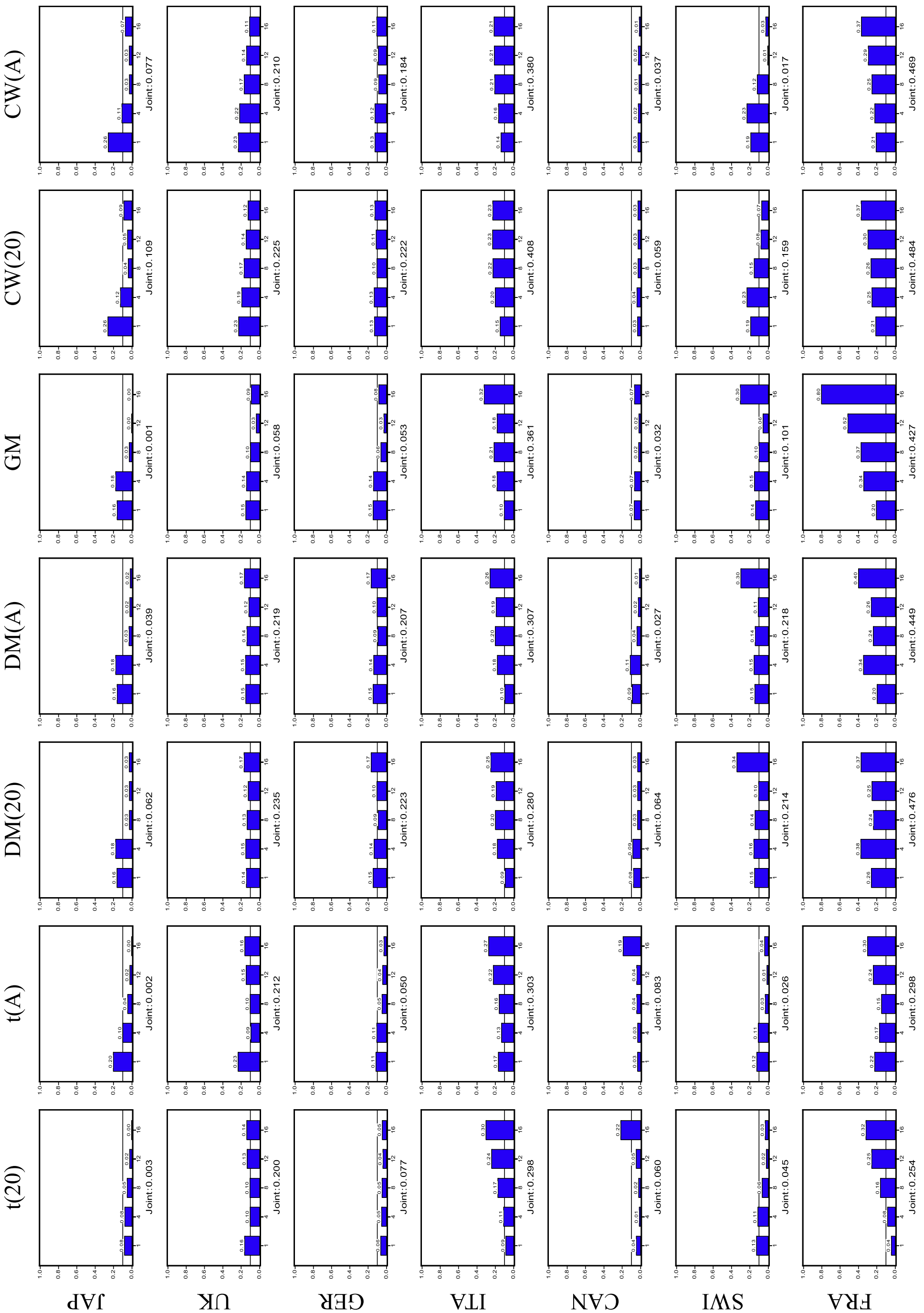

Fig. 1. Bootstrapped *P*-values for different statistics: 1973Q1–1998Q4.

Table 2
Bootstrapped *p*-value for joint statistics with different initial windows: 1973Q1–1998Q4.

	IW	$t(20)_{min}$	$t(A)_{min}$	$DM(20)_{max}$	$DM(A)_{max}$	GM_{max}	$CW(20)_{max}$	$CW(A)_{max}$
JAP	32Q	**0.003**	**0.002**	**0.062**	**0.039**	**0.001**	**0.109**	**0.077**
	40Q	**0.003**	**0.002**	**0.063**	**0.048**	**0.001**	**0.098**	**0.098**
	50Q	**0.003**	**0.002**	0.114	0.146	**0.006**	0.155	0.179
	60Q	**0.003**	**0.002**	0.251	0.257	**0.009**	0.298	0.353
UK	32Q	0.200	0.212	0.235	0.219	**0.058**	0.225	0.210
	40Q	0.200	0.212	0.248	0.282	0.106	0.254	0.265
	50Q	0.200	0.212	0.321	0.406	0.359	0.248	0.286
	60Q	0.200	0.212	0.409	0.479	0.455	0.459	0.460
GER	32Q	**0.077**	**0.050**	0.223	0.207	**0.053**	0.222	0.184
	40Q	**0.077**	**0.050**	0.246	0.238	**0.024**	0.174	0.139
	50Q	**0.077**	**0.050**	0.287	0.369	0.342	0.147	0.134
	60Q	**0.077**	**0.050**	0.269	0.396	0.299	0.155	0.146
ITA	32Q	0.298	0.303	0.280	0.307	0.361	0.408	0.380
	40Q	0.298	0.303	0.402	0.415	0.362	0.443	0.402
	50Q	0.298	0.303	0.511	0.512	0.523	0.608	0.613
	60Q	0.298	0.303	0.340	0.406	0.467	0.385	0.396
CAN	32Q	**0.060**	**0.083**	**0.064**	**0.027**	**0.032**	**0.059**	**0.037**
	40Q	**0.060**	**0.083**	**0.088**	**0.058**	**0.037**	**0.037**	**0.026**
	50Q	**0.060**	**0.083**	0.174	0.142	**0.060**	**0.068**	**0.051**
	60Q	**0.060**	**0.083**	0.443	0.409	0.139	0.309	0.270
SWI	32Q	**0.045**	**0.026**	0.214	0.218	0.101	0.159	**0.017**
	40Q	**0.045**	**0.026**	0.161	0.151	**0.003**	0.133	**0.079**
	50Q	**0.045**	**0.026**	0.132	0.195	**0.036**	**0.048**	**0.037**
	60Q	**0.045**	**0.026**	0.212	0.307	**0.071**	0.153	0.100
FRA	32Q	0.254	0.298	0.476	0.449	0.427	0.484	0.469
	40Q	0.254	0.298	0.513	0.519	0.609	0.511	0.474
	50Q	0.254	0.298	0.776	0.669	0.691	0.743	0.743
	60Q	0.254	0.298	0.481	0.546	0.547	0.479	0.510

Note: IW indicates the length of initial estimation window. 32Q, 40Q, 50Q and 60Q are 32, 40, 50 and 60 quarters, respectively. $t(20)$ and $t(A)$ are *t*-statistics with the truncation lag set at 20 and with that determined based on Andrew's (1991) procedure. DM, GM and CW indicate Diebold–Mariano, Gilbert–McCracken and Clark–West statistics, respectively. The number in boldface indicates significance at the 10% level.

rejected for CAN and SWI (for JAP) when the initial window is 32, 40 and 50 (32 and 40) quarters, respectively. Among these different out-of-sample tests, the evidence of rejection is stronger based on the GM test. A reasonable explanation for this finding is that the power of the GM test is higher than that of DM and CW tests, and hence it is less likely to reject random walk forecasts with DM or CW tests when the length of initial window increases. The previous conjecture is confirmed as one can see from the empirical investigation in Section 5.

Bergin et al. (2006) point out that if the trade pattern truly is endogenously determined then it takes time for the HBS effect to emerge. This indicates that the HBS effect appears to be more significant if we have longer post war data. To justify their argument, we re-investigated the previous out-of-sample forecast contest by extending the sample period from the last quarter of 1998 to the second quarter of 2006 for JAP, CAN, SWI, and UK.[24] We re-estimated the ESTAR model with the extended sample period and observed a quadratic trend function for JAP and UK, a cubic trend function for CAN, and a cubic trend for SWI. These results reveal significant HBS effects in all of the four countries with the extended sample period supporting the theoretical implication of Bergin et al. (2006).

[24] We excluded the Deutschemark–dollar, French franc–dollar and lira–dollar rates from our sample since the German Mark, French franc and Italian lira went out of circulation due to the introduction of the Euro in January 1999.

Results of forecast contests based on the above models are reported in Fig. 2.[25] Two out of four countries reject random walk forecasts based on in-sample statistics. The random walk forecasts are rejected at the 10% level of significance for all four countries when the joint GM test is applied and for two out of four countries if the joint CW or DM test is applied. In addition, we also find evidence of rejecting random walk forecasts with different out-of-sample statistics when forecast horizons are medium term or long. It is interesting to find that the out-of-sample predictability of nominal exchange rates is observed, based on the GM test, for UK and CAN regardless of forecast horizons when the sample period is long. As for SWI, we find evidence of rejecting random walk forecasts with different out-of-sample statistics when forecast horizons are medium term or long. It is interesting to find that the evidence of out-of-sample predictability of nominal exchange rates under different forecast horizons is weak for the UK when the sample period is short, but it is strengthened when the sample period is long. In short, the out-of-sample evidence of beating the random walk forecast model is strengthened with the extended sample period, which echoes the theoretical implication of Bergin et al. (2006).

Given the fact that our model beats the random walk model for most countries as indicated by Figs. 1 and 2, it is interesting to ask: Are our findings due to the problem of the size distortion or because of the high power of bootstrap tests? To address this issue we apply the bootstrap strategy provided by Kilian and Taylor (2003) to examine the size and power of bootstrap tests.

5. Size and power of bootstrap tests

Since our bootstrap tests reject the random walk forecast model across forecast horizons, especially for an extended sample, we are interested in whether these are reliable tests. A test is said to be unreliable if its effective size exceeds its nominal size. Following the strategy provided by Kilian and Taylor (2003), we impose the hypothesis that nominal exchange rate changes are unpredictable and then postulate the following data-generating process for the representative country, GER, as follows:

$$\begin{aligned}
\Delta s_t + 0.005 &= \hat{\varepsilon}_{1t} \\
z_t &= \hat{g}_t + \left\{ \exp\left[-5.011\left(z_{t-5} - \hat{g}_{t-5}\right)^2 \right] \right\} \left[1.069\left(z_{t-1} - \hat{g}_{t-1}\right) - 0.069\left(z_{t-2} - \hat{g}_{t-2}\right) \right] + \hat{\varepsilon}_{2t}, \\
\hat{g}_t &= -0.223 - 0.015 \times t + 1.4 \times 10^{-4} \times t^2,
\end{aligned}$$

where the DGP of z_t for GER is set to be the same as that in Table 1. Residuals, $\hat{\varepsilon}_t = (\hat{\varepsilon}_{1t}, \hat{\varepsilon}_{2t})$, are obtained by a random draw with replacement from actual regression residuals. A three-step procedure is applied to construct the size of in-sample and out-of-sample tests, based on a nominal 10% level of significance. In step 1, we generate artificial data based on the DGP given above by Monte Carlo simulations. In step 2, based on the simulated data, we construct different test statistics and bootstrap their finite sample distribution with 2000 replications. In step 3, repeating the previous procedures 1000 times, we obtain the size of bootstrap tests. A detailed description of the simulation procedure can be found in Kilian (1999) and Kilian and Taylor (2003). Fig. 3 reports the effective size of the bootstrap test at the nominal 10% level of significance. The effective size of all statistics is close to 0.1 at different forecast horizons, implying no significant size distortion for those countries.

To evaluate the power of the bootstrap tests, we follow the strategy provided by Kilian and Taylor (2003) in simulating data for the real exchange rates (z_t) and the PPP fundamental (f_t), which allows us to construct the nominal exchange rate (s_t), since $s_t = f_t + z_t$. The DGP of z_t for GER is again set to be the same as that in Table 1, and the three different DGPs for f_t are taken from Kilian and Taylor (2003).[26]

[25] Estimation results from the extended sample period are not reported here but are available upon request from authors.

[26] Although our sample period is the same as that of Kilian and Taylor (2003), our data version (IFS CD-ROM) is different from theirs. This is the reason why the estimated coefficients in Δf_t equations are not the same as those in Kilian and Taylor (2003).

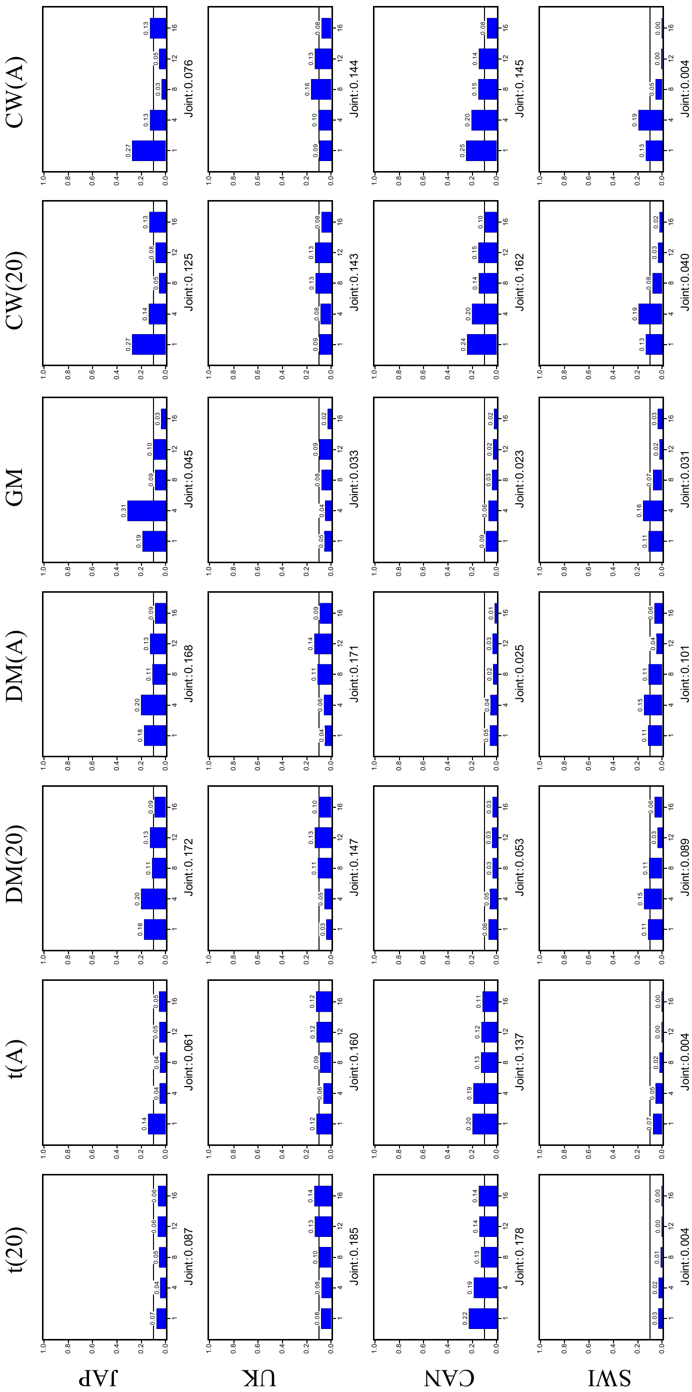

Fig. 2. Bootstrapped *P*-values for different statistics: 1973Q1–2006Q2.

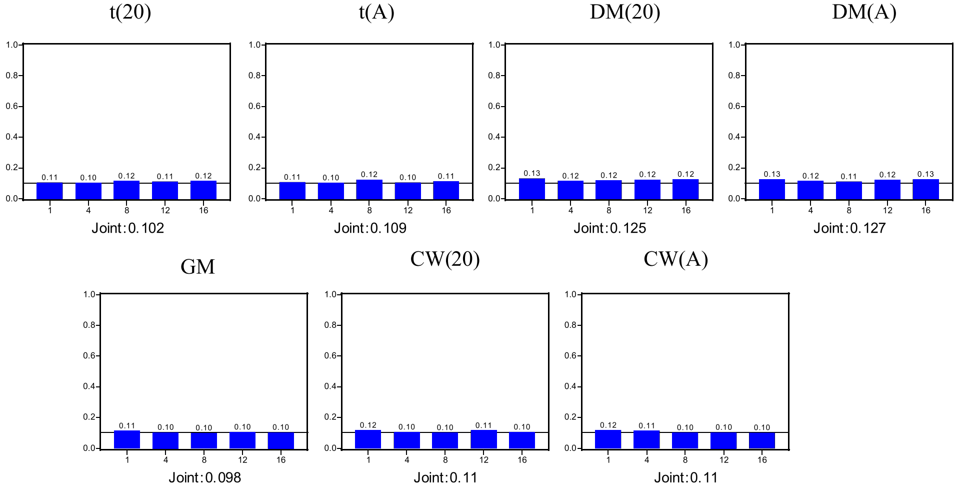

Fig. 3. Effective sizes of bootstrap tests at the 10% level of significance: Deutschemark to dollar rate, 1973Q1–1998Q4. Note: *p*-values are obtained based on 1000 Monte Carlo simulations with 2000 bootstrap replications each.

DGP1:

$$\begin{aligned}
\Delta f_t &= 0.0050 + \hat{\varepsilon}_{1t} \\
z_t &= \hat{g}_t + \left\{ \exp\left[-5.011\left(z_{t-5} - \hat{g}_{t-5}\right)^2 \right] \right\} \left[1.069\left(z_{t-1} - \hat{g}_{t-1}\right) - 0.069\left(z_{t-2} - \hat{g}_{t-2}\right) \right] + \hat{\varepsilon}_{2t}, \\
\hat{g}_t &= -0.223 - 0.015 \times t + 1.4 \times 10^{-4} \times t^2
\end{aligned}$$

DGP2:

$$\begin{aligned}
\Delta f_t &= 0.0010 + 0.3554\Delta f_{t-1} - 0.1162\Delta f_{t-2} + 0.1938\Delta f_{t-3} + 0.3395\Delta f_{t-4} + \hat{\varepsilon}_{1t} \\
z_t &= \hat{g}_t + \left\{ \exp\left[-5.011\left(z_{t-5} - \hat{g}_{t-5}\right)^2 \right] \right\} \left[1.069\left(z_{t-1} - \hat{g}_{t-1}\right) - 0.069\left(z_{t-2} - \hat{g}_{t-2}\right) \right] + \hat{\varepsilon}_{2t}, \\
\hat{g}_t &= -0.223 - 0.015 \times t + 1.4 \times 10^{-4} \times t^2
\end{aligned}$$

DGP3:

$$\begin{aligned}
\Delta f_t &= 0.0037 + 0.2158\Delta f_{t-2} + 0.0191\Delta e_{t-2} + \hat{\varepsilon}_{1t} \\
z_t &= \hat{g}_t + \left\{ \exp\left[-5.011\left(z_{t-5} - \hat{g}_{t-5}\right)^2 \right] \right\} \left[1.069\left(z_{t-1} - \hat{g}_{t-1}\right) - 0.069\left(z_{t-2} - \hat{g}_{t-2}\right) \right] + \hat{\varepsilon}_{2t}, \\
\hat{g}_t &= -0.223 - 0.015 \times t + 1.4 \times 10^{-4} \times t^2
\end{aligned}$$

The residuals for each equation are bootstrapped from actual regression residuals.

Findings from Fig. 4 indicate that the power of the in-sample bootstrap test varies from 0.81 to 1.0 for different data-generating processes. As for the power of out-of-sample bootstrap tests, it varies from 0.62 to 0.95, 0.79 to 0.99 and 0.53 to 0.99 for the DM, GM and CW tests, respectively.[27] Based on the joint test, we find that the power of GM test is higher than that of DM and CW tests.[28] Our finding of a high power for the GM test (relative to DM test) is consistent with that in Clark and McCracken (2004). There is no observation that the power of the tests increases with forecast horizons which is

[27] It is meaningless to compare the power in Fig. 4 with those of Kilian and Taylor (2003) since our DGPs are different from theirs.

[28] We also allow the sample size to increase from $T = 104$ to $T = 208$ in simulations and find that the power of bootstrap tests increases with the sample size (results are not reported here, but are available upon request from authors).

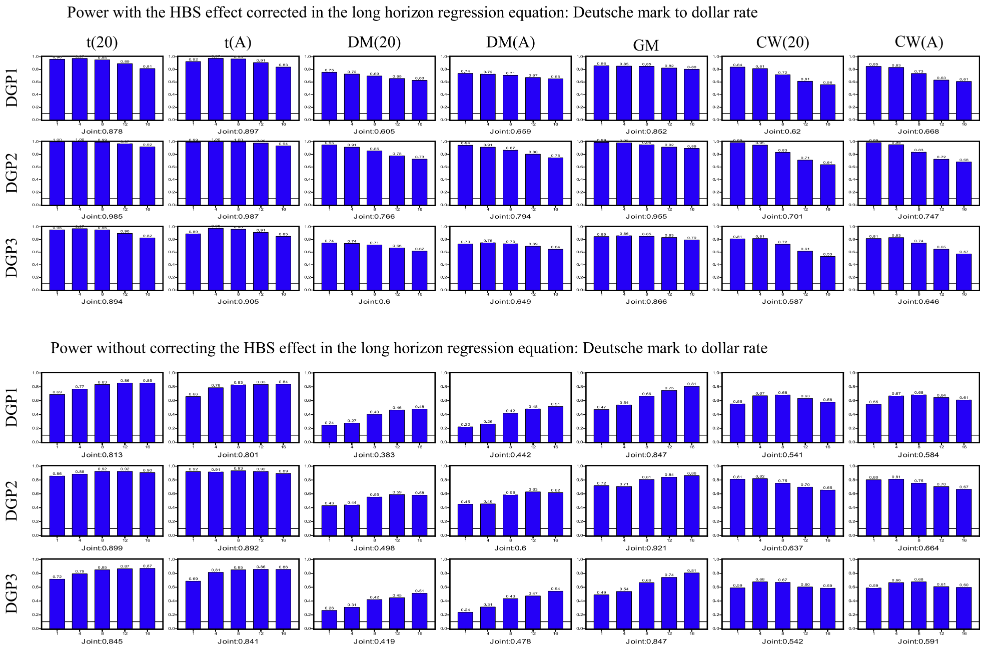

Fig. 4. Power of bootstrap tests at the 10% level of significance: 1973Q1–1998Q4.

consistent with the finding in Berkowitz and Giorgianni (2001). In addition, the power of in-sample tests is higher than that of out-of-sample tests, given the same test size. This finding is due to the fact that an out-of-sample analysis based on sample-splitting results in a loss of information, which in turn leads to lower power in a small sample.

We are also interested in the case where the HBS effect does exist in data but is neglected in long-horizon regression equation and in bootstrapping the finite sample distribution of in- and out-of-sample statistics. We re-construct the power of bootstrap tests in this case. Results from the lower panel of Fig. 4 indicate that the power of in-sample bootstrap tests varies from 0.69 to 0.93 and the power of out-of-sample bootstrap tests varies from 0.24 to 0.51 (0.47 to 0.81) for DGP1 and DGP3, and from 0.43 to 0.63 (0.71 to 0.86) for DGP2 based on the DM (GM) test. As for the power of CW test, it varies from 0.55 to 0.82, and the power does not vary significantly across different DGPs. The power of the GM test is higher than that of the DM and CW tests based on joint tests. However, we observe that the power of tests increases with forecast horizons in general. The power comparison between the previous two cases is meaningless since their DGPs are different. However, it is reasonable to state that, given the small time span of data, the out-of-sample tests may not achieve reasonably high power if the HBS effect exists in data but is not accounted for in regression.

6. Conclusion

Although our knowledge on exchange rate behavior has improved during the past two decades, economists are still puzzled by the failure to beat random walk models in out-of-sample forecasting contests. Recently, a number of authors have suggested that the adjustment in the real exchange rate is in fact non-linear. Kilian and Taylor (2003) focus on PPP fundamentals and argue that an ESTAR model with a constant equilibrium is appropriate in describing real exchange rate dynamics over the period of post-Bretton Woods for seven major OECD countries. Based on in-sample evidence, Kilian and Taylor (2003) conclusively beat the random walk forecast model, but their out-of-sample evidence is less than satisfactory.

In this paper we incorporate both a non-linear adjustment and HBS effect in a model to re-examine the predictability of nominal exchange rates wherein we arrive at several conclusions. First, empirical results have shown that an ESTAR model embodying the HBS effect provides a parsimonious representation of real exchange rate data during the recent float. Second, empirical results from in-sample and out-of-sample tests allow us to beat the random walk forecast model, given the short spans of data, at short-term, medium-term and long forecast horizons. Moreover, the previous findings are robust to different statistical tests, sample periods and initial estimation windows. Third, results from simulations indicate that our bootstrap tests have correct size and good power, and there is no observation that the power of tests increases with forecast horizons.

Are nominal exchange rates predictable? The recursive out-of-sample evidence from Kilian and Taylor (2003) does not reject the random walk forecast model. They conclude that "this stylized empirical fact appears to be a natural consequence of the small time span of data available for empirical work." Based on a careful empirical investigation, we provide solid evidence to beat the random walk forecast model. We argue that, given the short time span of data, combining the HBS effect with the non-linear adjustments of real exchange rates are useful in providing evidence of nominal exchange rate predictability. Our empirical findings therefore shed new light on understanding the Meese–Rogoff puzzle.

Acknowledgements

For helpful comments we appreciate James Lothian (the editor), an anonymous referee, Alan Taylor, Paul Bergin as well as participants in seminars at UC Davis and National Sun Yat-sen University. Part of the paper was finished while Wu was visiting at UC Davis where the Institute of Governmental Affairs and Department of Economics graciously provided him with superior research facilities. Wu is grateful for financial support from National Science Council in Taiwan.

References

Andrews, D., 1991. Heteroscedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 59 (3), 817–858.
Asea, P.K., Mendoza, E., 1994. A Balassa–Samuelson model: a general equilibrium appraisal. Review of International Economics 2 (3), 244–267.
Asea, P.K., Corden, W.M., 1994. The Balassa–Samuelson model: an overview. Review of International Economics 2 (3), 191–200.
Balassa, B., 1964. The purchasing-power parity doctrine: a reappraisal. Journal of Political Economy 72 (6), 584–596.
Berben, R.B., van Dijik, D.J., 1998. Does the absence of cointegration explain the typical findings in long horizon regressions. Report 9814. Econometrics Institute, Erasmus University of Rotterdam.
Bergin, P., Glick, R., Taylor, A., 2006. Productivity, tradability, and the long-run price puzzle. Journal of Monetary Economics 53 (8), 2041–2066.
Bergman, U.M., Hansson, J., 2005. Real exchange rates and switching regimes. Journal of International Money and Finance 24 (1), 121–138.
Berkowitz, J., Giorgianni, L., 2001. Long-horizon exchange rate predictability? Reviews of Economics and Statistics 83 (1), 81–91.
Cheung, Y.W., Erlandsson, U.G., 2005. Exchange rates and Markov switching dynamics. Journal of Business and Economic Statistics 23 (3), 314–320.
Cheung, Y.-W., Chinn, M.D., Pascual, A.G., 2005. Empirical exchange rate models of the nineties: are they fit to survive? Journal of International Money and Finance 24 (7), 1150–1175.
Chinn, M., Meese, R., 1995. Banking on currency forecasts: how predictable is change in money? Journal of International Economics 38 (1–2), 161–178.
Clarida, R.H., Sarno, L., Taylor, M.P., Valente, G., 2003. The out-of-sample success of term structure models as exchange rate predictors: a step beyond. Journal of International Economics 60 (1), 61–83.
Clark, T.E., McCracken, M.W., 2001. Tests of equal forecast accuracy and encompassing for nested models. Journal of Econometrics 105 (1), 85–110.
Clark, T.E., McCracken, M.W., 2004. Evaluating long-horizon forecasts. Working Papers: 0302. Department of Economics, University of Missouri.
Clark, T.E., West, K.D., 2006. Using out-of-sample mean squared prediction errors to test the martingale difference hypothesis. Journal of Econometrics 135 (1–2), 155–186.
Clark, T.E., West, K.D., 2007. Approximately normal tests for equal predictive accuracy in nested models. Journal of Econometrics 138 (1), 291–311.
Cuddington, J.T., Liang, H., 2000. Purchasing power parity over two centuries. Journal of International Money and Finance 19 (5), 753–757.
Dacco, R., Satchell, S., 1999. Why do regime-switching models forecast so badly? Journal of Forecasting 18 (1), 1–16.
Diebold, F.X., Nason, J.A., 1990. Nonparametric exchange rate prediction. Journal of International Economics 28 (3–4), 315–332.
Diebold, F.X., Mariano, R.S., 1995. Comparing predictive accuracy. Journal of Business and Economic Statistics 13 (3), 253–262.

Engel, C., 1994. Can the Markov switching model forecast exchange rates? Journal of International Economics 36 (1–2), 151–165.
Engel, C., Hamilton, J.D., 1990. Long swings in the dollar: are they in the data and do markets know it? American Economic Review 80 (4), 689–713.
Engel, C., West, K.D., 2005. Exchange rates and fundamentals. Journal of Political Economy 113 (3), 485–517.
Engle, R.F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica 50 (4), 987–1007.
Fitzgerald, D., 2003. Terms-of-trade effects, interdependence and cross-country differences in price levels. Working Paper. UC Santa Cruz.
Frankel, J.A., Rose, A., 1995. Empirical research on nominal exchange rates. In: Grossmann, G., Rogoff, K. (Eds.), Handbook of International Economics, vol. III. North Holland, Amsterdam, pp. 1689–1729.
Froot, K., Rogoff, K., 1995. Perspectives on PPP and long-run real exchange rates. In: Grossman, G., Rogoff, K. (Eds.), Handbook of International Economics, vol. III. North Holland, Amsterdam, pp. 1647–1688.
Froot, K., Rogoff, K., 1991. The EMS, the EMU, and the transition to a common currency. In: Fischer, S., Balanchard, O. (Eds.), NBER Macroeconomics Annual. MIT Press, Cambridge, MA.
Gilbert, S., 2001. Sampling Schemes and Tests of Regression Models. Manuscript. Department of Economics, Southern Illinois University at Carbondale.
Groen, J.J., 2000. The monetary exchange rate model as a long-run phenomenon. Journal of International Economics 52 (2), 299–319.
Hamilton, J., 1989. A new approach to the economic analysis of nonstationary time series and the business cycle. Econometrica 57 (2), 357–384.
Harrod, R., 1939. International Economics. Cambridge University Press, Cambridge.
Inoue, A., Kilian, L., 2004. In-sample or out-of-sample tests of predictability: which one should we use? Econometric Review 23 (4), 371–402.
Kilian, L., 1999. Exchange rates and monetary fundamentals: what do we learn from long-horizon regressions? Journal of Applied Econometrics 14 (5), 491–510.
Kilian, L., Taylor, M.P., 2003. Why is it so difficult to beat the random walk forecast of exchange rates? Journal of International Economics 60 (1), 85–107.
Lothian, J.R., Taylor, M.P., 2008. Real exchange rates over the past two centuries: how important is the Harrod–Balassa–Samuelson effect? Economic Journal 118 (532), 1742–1763.
Lothian, J.R., Taylor, M.P., 2000. Purchasing power parity over two centuries: strengthening the case for real exchange rate stability: a reply to Cuddington and Liang. Journal of International Money and Finance 19 (5), 759–764.
Lothian, J.R., 1990. A century plus of yen exchange rate behavior. Japan and the World Economy 2 (1), 47–70.
MacDonald, R., Marsh, I.W., 1997. On fundamentals and exchange rates: a Casselian perspective. Review of Economics and Statistics 79 (4), 655–664.
Mark, N.C., 2001. International Macroeconomics and Finance: Theory and Econometric Methods. Blackwell Publishers, Massachusetts, USA.
Mark, N.C., 1995. Exchange rates and fundamentals: evidence on long-horizon predictability. American Economic Review 85 (1), 201–218.
Mark, N.C., Sul, D., 2001. Nominal exchange rates and monetary fundamentals: evidence from a small post-Bretton Woods panel. Journal of International Economics 53 (1), 29–52.
Marsh, I.W., 2000. High-frequency Markov-switching models in the foreign exchange market. Journal of Forecasting 19 (2), 123–134.
Meese, R.A., Rose, A.K., 1991. An empirical assessment of non-linearities in models of exchange rate determination. Review of Economic Studies 58 (3), 603–619.
Meese, R.A., Rogoff, K., 1983. Empirical exchange rate models of seventies: do they fit out-of-sample? Journal of International Economics 14 (1–2), 3–24.
Michael, P., Nobay, R.A., Peel, D.A., 1997. Transaction costs and nonlinear adjustment in real exchange rates: an empirical investigation. Journal of Political Economy 105 (4), 862–879.
Newey, W.K., West, K.D., 1987. A simple positive semi-definite heteroscedasticity and autocorrelation consistent covariance matrix. Econometrica 55 (3), 703–708.
Obstfeld, M., 1993. Model trending real exchange rates. Center for International and Development Economic Research, Working Paper No. C93–011.
Obstfeld, M., Rogoff, K., 2000. The six major puzzles in international macroeconomics: is there a common cause?. In: NBER Macroeconomics Annual, vol. 15.
Paya, I., Venetis, I.A., Peel, D.A., 2003. Further evidence on PPP adjustment speeds: the case of effective real exchange rates and the EMS. Oxford Bulletin of Economics and Statistics 65 (4), 421–437.
Paya, I., Peel, D.A., 2003. Purchasing power parity adjustment speeds in high frequency data when the equilibrium real exchange rate is proxied by a deterministic trend. Manchester School (Suppl.), 39–53.
Peel, D.A., Venetis, I.A., 2003. Purchasing power parity over two centuries: trends and non-linearity. Applied Economics 35 (5), 609–617.
Rossi, B., 2005. Testing long-horizon predictive ability with high persistence, and the Meese–Rogoff puzzle. International Economic Review 46 (1), 61–92.
Samuelson, P., 1964. Theoretical notes on trade problem. Reviews of Economics and Statistics 46 (2), 145–154.
Sarno, L., Taylor, M.P., 2002. The Economics of Exchange Rates. Cambridge University Press, Cambridge.
Siddique, A., Sweeney, R.J., 1998. Forecasting real exchange rates. Journal of International Money and Finance 17, 63–70.
Sollis, R., 2005. Evidence on purchasing power parity from univariate models: the case of smooth transition trend stationarity. Journal of Applied Economics 20 (1), 79–98.
Stambaugh, R.F., 1986. Bias in Regression with Lagged Stochastic Regressors. Mimeo. University of Chicago.
Taylor, A.M., 2002. A century of purchasing power parity. Review of Economics and Statistics 84 (1), 139–150.

Taylor, A.M., Taylor, M.P., 2004. The purchasing power parity debate. Journal of Economic Perspective 18 (4), 135–158.
Taylor, M.P., 1995. The economics of exchange rates. Journal of Economic Literature 33 (1), 13–47.
Taylor, M.P., Peel, D.A., Sarno, L., 2001. Nonlinear mean-reversion adjustment in real exchange rates: towards a solution to the purchasing power parity puzzles. International Economic Reviews 42 (4), 1015–1042.
Taylor, M.P., Peel, D.A., 2000. Nonlinear adjustment, long-run equilibrium and exchange rate fundamentals. Journal of International Money and Finance 19 (1), 33–53.