**Understanding forecast failure of
ESTAR models of real exchange rates**

Daniel Buncic

UNDERSTANDING FORECAST FAILURE OF ESTAR MODELS OF REAL EXCHANGE RATES[*]

DANIEL BUNCIC[§]

School of Economics
University of New South Wales
Sydney, Australia

First Version : September 19, 2008
This Version : July 23, 2009

COMMENTS WELCOME

Abstract

The forecast performance of the empirical ESTAR model of Taylor, Peel and Sarno (2001) is examined for 4 bilateral real exchange rate series over an out-of-sample evaluation period of nearly 12 years. Point as well as density forecasts are constructed, considering forecast horizons of 1 to 22 steps head. The study finds that no forecast gains over a simple AR(1) specification exist at any of the forecast horizons that are considered, regardless of whether point or density forecasts are utilised in the evaluation. Non-parametric methods are used in conjunction with simulation techniques to learn about the models and their forecasts. It is shown graphically that the non-linearity in the point forecasts of the ESTAR model decreases as the forecast horizon increases. The non-parametric methods show also that the multiple steps ahead forecast densities are normal looking with no signs of bi-modality, skewness or kurtosis. Overall, there seems little to be gained from using an ESTAR specification over a simple AR(1) model.

Keywords: Purchasing power parity, regime modelling, non-linear real exchange rate models, ESTAR, forecast evaluation, density forecasts, non-parametric methods.
JEL Classification: C22, C52, C53, F31, F47.

1. Introduction

The Exponential Smooth Transition Autoregressive (ESTAR) model introduced by Granger and Teräsvirta (1993) and Teräsvirta (1994) into the economics literature has become *the* workhorse statistical paradigm for the modelling of real exchange rate data. This can be most easily appreciated by performing a search for "*ESTAR*" and "*real exchange rate*" in the Google Scholar search engine. One particular empirical study that has attracted considerable attention is that of Taylor, Peel and Sarno (2001). This study has received over 300 citations in the Google Scholar citations index. As a comparison, Kenneth Rogoff's (1996) seminal paper entitled "*The Purchasing Power Parity Puzzle*" published in the Journal of Economic Literature has received about 1500 citations.[1]

Nonetheless, despite the noticeable popularity in modelling real exchange rates within a non-linear ESTAR framework, little work appears to have been done in the out-of-sample forecast evaluation of these models.[2] This is especially surprising to see given that these models were partially developed with foreign exchange dealers in mind who, based upon the survey evidence provided in Cheung and Chinn (2000) and Cheung *et al.* (2004), are known to employ a non-linear mix of fundamentalist and chartist trading strategies in their foreign exchange dealings.[3] It would therefore seem to be of interest to understand how well such non-linear models can perform out-of-sample, when a decision regarding the implementation of these models needs to be made.

The objective of this study is twofold. Firstly, we take the well known empirical model of Taylor *et al.* (2001) and evaluate its forecast performance relative to a simple AR(1) specification over the out-of-sample period from January 1997 to June 2008 using the bilateral real exchange rates of the UK, France, Japan and Switzerland *vis-à-vis* the US Dollar. The models are assessed using point, as well as, density forecasts, considering forecast horizons of up to 22 steps ahead. Secondly, and more importantly, the objective is to provide an intuitive visual representation of the forecast analysis that is carried out. This is done with the aid of simulation and non-parametric techniques to visualise and understand what the models forecast. The intention here is to use graphical techniques to learn about the models and their fit to real exchange rate data.

The empirical ESTAR model of Taylor *et al.* (2001) is particularly suitable for a graphical analysis, as it is a simple, low-dimensional model, relying only on one conditioning variable to form the forecast. Since it is often the case that a visual inspection of the fore-

---

[1]Citation statistics were accessed on December 23rd, 2008.

[2]One notable exception is the study by Rapach and Wohar (2006), who assess the out-of-sample performance of the Band-TAR as well as the ESTAR model for real exchange rates. However, the conclusions that are drawn in their study appear to be misleading and counter intuitive, as claims of "*forecasting gains at long horizons relative to simple linear AR models*" are raised by Rapach and Wohar (see p. $350 - 352$).

[3]See also the agent based models of Westerhoff and Reitz (2003) and De Grauwe and Grimaldi (2005, 2006) who use fundamentalist and chartist trading strategies in their theoretical models.

casts from two competing models is far more informative to the applied forecaster than the outcome of a statistical test, the objective here is to illustrate how simulation and non-parametric methods can be used to highlight how the forecasts from the two competing models differ, and hence where one model is likely to perform better than the other. The graphical forecast analysis can also be thought of as an informal test of the model, ie., an overall goodness of fit test, where the out-of-sample data represent the evaluation period.

The main findings of this study can be summarised as follows. Firstly, the statistical tests that are conducted provide no evidence to conclude that the ESTAR model outper-forms a simple AR(1) specification at any of the 1 to 22 steps ahead forecast horizons for all four empirical real exchange rate series that are considered. This outcome is reached regardless of whether point or density forecasts are used in the evaluation of the out-of-sample data.

Secondly, the graphical analysis that is carried out shows that the variation of the em-pirical data around the one step ahead point forecasts (or conditional means) of the two competing models is substantial, making it impossible for a statistical procedure to dis-criminate between these two models at the of out-of-sample data points that are available. Furthermore, using simulation and non-parametric techniques it is illustrated graphically that the non-linearity in the h step ahead point forecasts of the ESTAR model decreases monotonically as the forecast horizon increases. These two results imply that, as no fore-cast gains are realised at the one step ahead horizon, where the non-linearity in the con-ditional mean is the strongest, there exists no potential whatsoever for the fitted ESTAR models to outperform a simple AR(1) at any forecast horizon. This result contradicts the finding of forecast gains at long horizons documented in Rapach and Wohar (2006).

The graphical analysis shows also that the forecast densities of the fitted ESTAR models are approximately normal looking, without any indication of skewness and/or kurtosis. This is regardless of the magnitude of the conditioning variable used in the construction of the forecast densities. When testing the forecast densities, the implication of this re-sult is that the statistical comparison boils down to one of equal conditional means and variances. Since the conditional means of the two competing models were found to be indistinguishable from one another, with similar sized variances, it is easy to appreciate why the null of equal forecast densities cannot be rejected.

Although the results presented in this study may seem like a straightforward out-of-sample forecast evaluation exercise of one particular non-linear ESTAR model over one particular out-of-sample data set, the forecast evaluation results have much wider impli-cations for other ESTAR models that have been fitted in the international finance literature. For example, Kilian and Taylor (2003) and Taylor and Peel (2000) use quarterly data with around 100 observations to estimate second order ESTAR models. Michael *et al.* (1997) em-ploy an annual data frequency with up to 200 observations. Similarly, Baum *et al.* (2001) take monthly data with less than 270 observations to fit a number of ESTAR models to a

total of 17 different data series. All of the fitted models in these studies have comparably sized parameter estimates in terms of the threshold parameter and the standard deviation of the shock component, and therefore imply a fit that is similar to what is presented in this study. Despite the fact that the estimated ESTAR models are often deemed to be successful in terms of their in-sample fit, it should be clear from the graphical illustration that is provided in this study that over the combined in and out-of-sample period of over 35 years that we consider, a substantial part of the variation in the sample data is still left unexplained. This makes it difficult to decisively conclude whether an AR(1) or an ESTAR model most closely matches the properties of the empirical real exchange rate data.

The remainder of the paper is organised in the following sections. Section 2 gives a brief description of the ESTAR model, the data that was used and how the model was estimated, with a short discussion of the results. In Section 3, point and density forecasts are formed, visualised, statistically tested and discussed. Section 4 concludes the study with a summary of the findings.

2. Model, data and estimation

The non-linear ESTAR model, the empirical data and the estimation method that is employed in this study are described in this section. Since the model and the data have been widely used in the literature, and as the estimation approach is considered to be rather standard, the description is kept to a minimum.

2.1. The ESTAR model

Taylor *et al.* (2001) specify the real exchange rate q_t to evolve according to the following non-linear process:

$$
\begin{aligned}
\Delta q_t &= -\left(q_{t-1} - \eta\right) \Phi\left(\gamma, \eta; q_{t-1}\right) + \sigma_\eta \epsilon_t \\
\Phi\left(\gamma, \eta; q_{t-1}\right) &= 1 - \exp\left\{-\gamma\left(q_{t-1} - \eta\right)^2\right\}
\end{aligned}
\tag{1}
$$

where the error term ϵ_t is assumed to be independently and identically distributed, with zero mean and unit variance.[4] The exponential weighting function $\Phi\left(\gamma, \eta; q_{t-1}\right)$ determines the regime that governs the evolution of q_t in (1). In the extreme case, that is, when $\Phi\left(\gamma, \eta; q_{t-1}\right)$ is either 0 or 1, q_t evolves either according to a random walk process or an equilibrium correcting mechanism, where η is the long-run equilibrium level of q_t. For all other values of $\Phi\left(\gamma, \eta; q_{t-1}\right)$, q_t evolves as a smooth and continuous non-linear process with a continuum of regimes.

---

[4]One can impose the restriction that ϵ_t is Gaussian, however, this is not needed at the estimation stage.

2.2. Data

As in Taylor *et al.* (2001), end-of-month nominal exchange rate and CPI data were obtained from the IMF's International Financial Statistics database for the US, the UK, Japan, France, Germany, and also for Switzerland over the period from January 1973 to June 2008, yielding 426 observations. The real exchange rates for the UK, Japan, France, Germany and Switzerland — relative to the US — are constructed in the standard way as $q_t \equiv \log(CPI_t^{home}/CPI_t^{US} S_t)$, where S_t is the home currency price of one US Dollar. The series are further normalized to be equal to zero in January 1973. Figure 1 shows a time series plot of these five real exchange rates from January 1973 to June 2008.[5]

Taylor *et al.* (2001) originally estimated the ESTAR models over a sample period from January 1973 to December 1996 for the real exchange rates of the UK, Japan, France and Germany only. This study extends the available data set by nearly 12 years to conduct an out-of-sample evaluation of these models. In our analysis, we use the January 1973 to December 1996 in-sample period to estimate the ESTAR models and then use the remaining data up to June 2008 to evaluate the models out-of-sample. We also include the Swiss real exchange rate series in this analysis.[6] The reason for doing this becomes clear when examining the evolution of the five series over the full sample data. As one can see from Figure 1, since approximately the beginning of 1996 the German and French real exchange rate series start to track one another extremely closely. This is evidently due to the anticipation of the third stage of the European Monetary Union (EMU) commencing in January 1999. As the purpose of this study is to assess how well the fitted non-linear ESTAR models perform over the out-of-sample period from January 1997 to June 2008, it is somewhat uninformative and rather repetitive to include both series in the forecast evaluation. For that reason, we do not report the forecast evaluation results for the German real exchange rate series.[7]

2.3. ESTAR estimation and discussion of results

The ESTAR model in (1) can be consistently estimated by standard non-linear least squares estimation or alternatively, if one is willing to make the assumption that ϵ_t is Gaussian, by maximum likelihood (see Gallant, 1987). The parameter estimates of all five real exchange rate series over the in-sample period from January 1973 to December 1996, together with

---

[5]The data can be downloaded from http://research.economics.unsw.edu.au/dbuncic/data/rer_data.xls.

[6]The Swiss Franc is one of the seven most heavily traded currencies in the world. Although there are other heavily traded currencies that could have been included in the forecast evaluation such as, for example, the Australian, Canadian or the New Zealand Dollars, these are often labelled as commodity currencies, due to their sensitivity to commodity prices. Since the influence of commodity prices can be fairly severe, it becomes difficult to identify adjustment due to Purchasing Power Parity (PPP) deviations or commodity price movements.

[7]The results for the German series are quantitatively very similar to those for the French series and can be obtained from the author upon request.

robust standard errors (se), the maximum of the log-likelihood function ($\mathcal{L}(\gamma, \eta)$ under a Gaussian assumption) and some standard mis-specification tests are reported in the upper part of Table 1.

It is evident from the results that are reported in Table 1 that the parameter estimates of the UK, German, French and Japanese series correspond very closely to the values estimated in previous studies (see Table 3 on page 1029 in Taylor *et al.* (2001) and Table 1 on page 344 in Rapach and Wohar (2006)). Notice also that the estimates for the Swiss series are similar in magnitude to those obtained for the French series and hence fall within the expected range of values found in the literature.

It should be emphasised here that we do not provide any discussion relating to model mis-specification and/or how the particular form of the ESTAR model specified in (1) was arrived at, although some test statistics are reported in Table 1. For details pertaining to these issues we refer the reader to the extensive discussion in Taylor *et al.* (2001). The focus of this study is to evaluate the fitted ESTAR model of Taylor *et al.* (2001) over the out-of-sample period from January 1997 to June 2008. Although it would have been possible to calibrate the parameters of the ESTAR model at the values found in Taylor *et al.* (2001), we preferred to fit the non-linear models to our data set and use these in the forecast evaluation. The IMF's IFS database is highly reliable so that our in sample data should correlate strongly with, if not exactly match, the data set used in Taylor *et al.* (2001).[8]

3. Forecasts and forecast evaluation

The forecast evaluation exercise focuses on point and density forecasts. Point forecasts still appear to be widely used by practitioners as they are easy to implement and interpret. Nonetheless, point forecasts have the drawback of being least informative in the sense that they do not provide any indication of the uncertainty surrounding the forecasts. Probability density forecasts, on the other hand, are the most general and informative forecasts that can be computed, as the whole forecast density is constructed.

The benchmark model that is used in the forecast evaluation exercise is a simple AR(1) specification for the real exchange rate, parameterised in the standard way as

$$\Delta q_t = \delta \left(q_{t-1} - \mu \right) + \sigma_\mu \epsilon_t. \tag{2}$$

The estimates of the AR(1) model parameters are — for reasons of completeness and again without any discussion — reported in the lower part of Table 1.

It should be mentioned here that the methodological approach of our forecast evaluation is a "genuine" out-of-sample forecast evaluation. In the terminology of McCracken

---

[8]Minor differences in the parameter estimates are thus most likely due to different numerical routines or differences in the convergence criteria.

and West (2002) this is referred to as a "fixed" forecasting scheme. That is, we estimate the model parameters over the in-sample period from January 1973 to December 1996 and do not update (or re-estimate) these as new data become available when constructing the out-of-sample forecasts.

There are three reasons why we implement a fixed forecasting scheme. Firstly, because a test of equal mean squared errors (MSE) of two parametric models is considered, where the first order optimality conditions are essentially moment conditions that provide consistent estimates of the model parameters, no adjustment to the (asymptotic) standard errors in the computation of the Diebold and Mariano (1995) (DM) test needs to be made that would normally arise due to the parameters on which the forecasts are based being sample estimates rather than population quantities (see pp. 312-313 in McCracken and West, 2002, for a detailed derivation of this result).[9] Rolling or recursive forecasts would require an adjustment to the standard errors of the DM test.

Secondly, because analytic (closed form) forecasts from the ESTAR model are available only at the one step ahead forecast horizon, where multiple steps ahead forecasts need to be simulated, the computational burden of updating and simulating a new forecast path for each of the updated parameter estimates under a rolling or recursive scheme would be substantial. These would then also need to be adjusted for parameter estimation uncertainty, requiring additional simulations.

And thirdly, because of the relatively rigid functional form of the ESTAR in (1) and a fairly homogenous out-of-sample data structure, the quantitative differences in the forecast errors from the updates to the parameter estimates are likely to remain small. To appreciate this, consider the scatter plots of all in and out-of-sample data and the conditional means shown in Figure 2 (these plots are discussed in more detail in Section 3.1.1). Notice from Panels (c) and (d) how for Japan and Switzerland nearly all out-of-sample data fall in the unit-root inner regime of the ESTAR model. Because of the unit-root restriction in (1), to obtain any noticeable changes in the shape of the conditional mean of the ESTAR model one would require observations away from the centre of q_{t-1} and large positive (negative) responses of Δq_t when q_{t-1} is sufficiently less (greater) than η.

Although the out-of-sample data for the UK and France do spread out more than the Japanese and Swiss real exchange rate series, there seems to be little evidence of an obvious model consistent response of Δq_t to q_{t-1}, portraying instead random variation across 0. Hence, we can expect changes in the parameter estimates due to updating within a recursive or rolling window scheme to be small and thus negligibly impact upon the qualitative forecast results.

To provide some evidence of the lack of substantial variation of the parameter estimates when expanding or updating the sample size over the out-of-sample data, we show

---

[9]In the notation of McCracken and West (2002), the term *F* in equation 14.20 on page 309 is equal to zero (see also Bao *et al.*, 2007, p. 9).

plots of the recursively estimated γ parameter under an expanding and fixed ($T = 287$) window scenario in Panels (a) and (b) of Figure 3 over the 138 out-of-sample observations from January 1997 to June 2008. Notice how small the changes in the γ parameter estimates are. Under the expanding window, all but the UK series show very little variation. The estimate for the UK series drops somewhat towards the end of the out-of-sample period, resulting in an even smaller value of γ. Under the fixed window scheme, a little more variation is visible for the French and UK series in the second half of the out-of-sample data, nevertheless for the French series the γ estimate increases only marginally from below 0.4 to around 0.5, while that of the UK series drops again below 0.5 after a short increase. Such small changes in γ alter the shape of the one step ahead conditional mean only at the margin.

It should be pointed out here that although we only look at one step ahead changes in the parameter estimates, since the non-linearity in the ESTAR conditional mean decreases with the forecast horizon, we can anticipate that such small changes will lead to even smaller differences in the multiple steps ahead forecast errors. It is therefore highly unlikely to see any decisive qualitative changes in the results of the DM test under a recursive forecasting scheme.

3.1. Point forecasts

Recall that under a MSE loss function, the optimal point forecast of the change in the real exchange rate series, h periods ahead, is $I\!E(\Delta q_{T+h}|\Omega_T)$, where $\Omega_T = \{Q^T; \mathcal{M}(\theta)\}$ is the information set available to the forecasting agent at time T when the forecast is made, Q^T is the full history of q_t up to time T and $\mathcal{M}(\theta)$ is the model with parameters θ used to construct the forecast. The $h-$step ahead point forecast $I\!E(\Delta q_{T+h}|\Omega_T)$ is thus nothing more than the conditional mean of Δq_t, given q_{t-h}, evaluated at the out-of-sample data points of the model under consideration.

3.1.1. Assessing one step ahead point forecasts

How do the conditional means of the competing models differ from one another at the one step ahead forecast horizon? Before we proceed to provide any formal statistical evidence to evaluate the out-of-sample forecast performance of the non-linear ESTAR model relative to the simple AR(1) benchmark, it will be informative here to consider an informal graphical approach to visually compare the one step ahead point forecasts of the two models. Such an approach has recently been advocated by Pagan (2002) and Breunig *et al.* (2003) to learn about models and their fit to data. In the current context we can informally assess one step ahead point forecasts by examining plots of the conditional means implied by the competing models over all out-of-sample data points.

Figure 2 shows the implied conditional means of the ESTAR and AR(1) models evaluated at the parameter estimates that are reported in Table 1 for the four real exchange rate series that are considered in the forecast evaluation. We have also superimposed the in-sample as well as the out-of-sample data by means of a scatter plot in Figure 2, and additionally graph a non-parametric (NP) estimate of $I\!\!E(\Delta q_t | q_{t-1})$ (with 95% confidence bands) to provide a purely data driven measure of $I\!\!E(\Delta q_t | q_{t-1})$.[10] The dashed vertical lines in Figure 2 show the 15th and 85th percentiles of the in-sample values of q_{t-1}.[11] The solid vertical line for the UK series in Panel (a) of Figure 2 marks the bound on the in-sample data.

What can we see from Figure 2? Notice initially how the conditional means of the ESTAR model and the AR(1) differ from one another. For the AR(1) model, adjustment towards its long-run equilibrium occurs at a constant rate over all values of q_{t-1}, so that it does not matter how far away one is from PPP when adjusting to any deviations from it. For the ESTAR model, on the other hand, this adjustment is evidently a non-linear function of q_{t-1}. The speed of adjustment towards PPP thus increases — with accelerating speed — the further away q_{t-1} is from η. Nevertheless, despite these important model specific differences between the conditional means of the linear and non-linear models, it is evident from Figure 2 that overall the variation of the empirical data around the conditional means is fairly substantial, so that a significant portion of the movement in Δq_t is not explained by the models.

Notice here also that over the entire out-of-sample period that we consider, covering nearly 12 years of data, only for the UK series are there a handful of observations that fall outside the in-sample data range. There is not a single out-of-sample data point that falls outside the in-sample data range for the French, Japanese and Swiss real exchange rate series. What is particularly interesting to see from Panels (c) and (d) in Figure 2 is that for the Swiss and Japanese series nearly all of the out-of-sample observations cluster around the centre of q_{t-1}, that is, in between the 15th and the 85th percentiles. Recall that in the literature that models real exchange rates with a threshold type model, ie., Obstfeld and Taylor (1997), this region coincides with what is labelled the "*inner regime*", where Δq_t is assumed to be inside the no adjustment threshold band within which q_t follows a random walk process. Given that the conditional means of the ESTAR and AR(1) models overlap fairly closely over this range, one can anticipate that statistical tests will have difficulties in decisively rejecting the (null) hypothesis of no forecast improvement of the ESTAR model over the AR(1).

---

[10]A local linear regression estimator was used to compute the NP conditional means with a Silverman (1986) rule of thumb plug-in bandwidth (see Pagan and Ullah, 1999, p. 104 for details).

[11]Note here that the 15th and 85th percentiles were used as the lower and upper bounds on the η parameter in the initial grid search of the estimation, before a Newton-Raphson type maximisation algorithm was used. In Threshold Autoregressive (TAR) models it is commonly required to have at least 15% of the sample data in each of the two regimes (see p. 84 in Franses and van Dijk, 2000).

Examining the plots of the UK and French real exchange rate series shown in Panels (a) and (b) of Figure 2, one can notice that the out-of-sample data points show a somewhat wider dispersion, with a number of them falling outside the 15th to 85th percentile range. Nonetheless, it is evident also that only very few observations fall close to the extreme tail ends of the density of q_{t-1}, where the non-linearity in the conditional means, and hence the forecasts of the ESTAR model, is most pronounced compared to the linear model. Notice here also that the spread of the out-of-sample data points across the conditional means of the two models is again fairly substantial, so that one can once again anticipate that it will be difficult for a forecast evaluation test to differentiate between these two models.

In order to provide some formal statistical evidence of the conjectured forecast failure of the non-liner ESTAR model at the one step ahead horizon, let the one step ahead forecast errors of the two competing models be defined as

$$\varepsilon^{ESTAR}_{T+1|T} = \Delta q_T + (q_T - \eta)\, \Phi\left(\gamma, \eta; q_T\right) \tag{3}$$

and

$$\varepsilon^{AR}_{T+1|T} = \Delta q_T - \delta\left(q_T - \mu\right), \tag{4}$$

where T is the last observation of the in-sample data set. The loss function at time $T+1$ that we employ to assess the models is a squared error loss function formed as

$$d_{T+1} \equiv (\varepsilon^{AR}_{T+1|T})^2 - (\varepsilon^{ESTAR}_{T+1|T})^2. \tag{5}$$

To evaluate the competing models, it is necessary to investigate how likely it is that the squared error loss d_{T+1} has a population expectation that is different from zero. That is, it is necessary to test the null hypothesis

$$\mathcal{H}_0 : \mathbb{E}(d_{T+1}) = 0$$

against the alternative

$$\mathcal{H}_A : \mathbb{E}(d_{T+1}) > 0.$$

We use two standard statistical tests to assess this. These are the Diebold and Mariano (1995) (DM) test, using the small sample correction factor of Harvey *et al.* (1997) and a weighted version of the DM test, adapted from van Dijk and Franses (2003). The weighted version of the DM test is designed to give more weight to out-of-sample observations that fall towards the extremes of the density of q_{t-1}, where the non-linearity in the ESTAR model is at its strongest.[12] It should thus be more apt in picking up forecast gains stem-

---

[12] See van Dijk and Franses (2003) for the computational details of the weighted version of the test. The weights ω_{T+1} were computed as $1 - \hat{f}(q_{T+1})/\max[\hat{f}(q_{T+1})]$ where $\hat{f}(q_{T+1})$ is a non-parametric estimate of the density function of q_{T+1}, evaluated at the out-of-sample data points. A Gaussian kernel with a plug-in

ming from non-linearity in the tails of q_{t-1}. We should stress here that the use of the DM test is valid as the two competing models are not nested. A restriction of $\gamma = 0$ makes the ESTAR model a random walk process rather than an AR(1).

The DM test results for the one step ahead point forecasts are reported in Table 2 below. These tests confirm the impressions that were formed from the visual inspection of the implied conditional means in Figure 2. All null hypotheses of equal forecast performance cannot be rejected for any of the four empirical series that are considered in the forecast evaluation, at any conventional significance levels and regardless of whether a weighted or an unweighted version of the DM test is used. Notice that the $t-$ratios remain well below unity in absolute value, suggesting that this is a fairly strong failure to reject the null hypothesis. Notice here also that for the UK and Japanese series, the DM test statistic is, in fact, negative, indicating that the ESTAR model generates larger forecast errors than the AR(1) model. Overall, therefore, we can conclude that it is highly unlikely that the ESTAR models that are considered here can outperform a simple AR(1) forecast at the one step ahead horizon.

3.1.2. Assessing multiple steps ahead point forecasts

How likely is it for the non-linear ESTAR model to generate any gains when forming a multiple periods ahead point forecast? We can again informally answer this question by looking at how different the implied conditional means of the ESTAR and AR(1) models are from one another. Moreover, since we saw that the non-linearity in the conditional means of the ESTAR models was quite mild at the one step ahead horizon, given the variation in the empirical data, it will be interesting to observe graphically how the non-linearity in the conditional mean changes as the forecast horizon increases. It should be clear that, because the ESTAR models that were estimated here are stable and stationary, the h step ahead conditional mean should converge to the unconditional mean of Δq_t, as h goes to infinity. As the same holds true for the AR(1) model, one can expect the difference between the forecasts of the two models to disappear as h increases.

Constructing multiple step ahead forecasts for the AR(1) model is straight forward and can be computed recursively in closed form. For the ESTAR model, nevertheless, this is not possible as it is necessary to integrate out non-linear transformations of all future shocks, therefore requiring numerical techniques. The approach that is employed here is Monte Carlo (MC) integration (cf. Franses and van Dijk, 2000, Section 3.5). To implement this, we simulate a large number of pseudo realisations of q_{T+h}, $\forall\, h > 1$, conditional on

---

bandwidth were used to compute $\hat{f}(q_{T+1})$.

q_T, using the following recursion

$$
\begin{aligned}
\tilde{q}^j_{T+1|T} &= q_T - (q_T - \eta)\,\Phi(\gamma, \eta; q_T) + \sigma_\eta \tilde{\epsilon}^j_{T+1} \\
\tilde{q}^j_{T+2|T} &= \tilde{q}^j_{T+1|T} - (\tilde{q}^j_{T+1|T} - \eta)\,\Phi(\gamma, \eta; \tilde{q}^j_{T+1|T}) + \sigma_\eta \tilde{\epsilon}^j_{T+2} \qquad (6)\\
&\vdots \\
\tilde{q}^j_{T+h|T} &= \tilde{q}^j_{T+h-1|T} - (\tilde{q}^j_{T+h-1|T} - \eta)\,\Phi(\gamma, \eta; \tilde{q}^j_{T+h-1|T}) + \sigma_\eta \tilde{\epsilon}^j_{T+h}.
\end{aligned}
$$

The realisation $\tilde{q}^j_{T+h|T}$ is thus the jth h step ahead pseudo value of q_{T+h}, given q_T and shock sequence $\{\tilde{\epsilon}^j_{T+i}\}^h_{i=1}$. The h step ahead point forecasts can then be approximated by computing the arithmetic mean over the J simulated $\tilde{q}^j_{T+h|T}$ entries, that is, one computes

$$
I\!E_J(\tilde{q}_{T+h|T}) = J^{-1} \sum_{j=1}^{J} \tilde{q}^j_{T+h|T} \qquad (7)
$$

which will have the property that $\lim_{J\to\infty} I\!E_J(\tilde{q}_{T+h|T}) = I\!E(q_{T+h}|q_T)$. To get the conditional mean for the changes in the q_t series, one simply constructs $I\!E(\Delta q_{T+h}|q_T)$ as $I\!E_J(\tilde{q}_{T+h|T}) - I\!E_J(\tilde{q}_{T+h-1|T})$.

Although it is appropriate to employ this approach to generate multiple steps ahead forecasts of Δq_t, one drawback when computing the conditional means for visualisation purposes is that the quantity $I\!E(\Delta q_{T+h}|q_T)$ will only be available at the empirical out-of-sample data points. A useful alternative approach that can be employed here to obtain the h step ahead implied conditional mean of q_t is to simulate a large number of realisations of q_t from the ESTAR model in (1) and then use non-parametric methods to compute $I\!E(\Delta q_t|q_{t-h})$ directly on the simulated data. The benefit of this approach lies in its ease of implementation and its ability to cover an arbitrary range of values of q_t. This way one can evaluate forecasts at a sufficient number of points over a given interval so that a line can be drawn to examine $I\!E(\Delta q_t|q_{t-h})$ graphically. As with the visualisation at the one step ahead horizons discussed in Section 3.1.1, any non-linearities in the conditional forecasts should then be identifiable from the plots of the non-parametric estimates of $I\!E(\Delta q_t|q_{t-h})$.

To illustrate how this approach can be implemented to examine the non-linearity of multiple steps ahead forecasts, 1 million observations of q_t were simulated from (1), calibrated at the parameter estimates of the UK series that are provided in Table 1. The ϵ_t were drawn from a standard normal distribution.[13] A grid of 1000 equally spaced points

---

[13]One could also use a non-parametric bootstrap and re-sample the empirical residuals of the UK series if one finds the normality assumption to be too restrictive. However, since there are only 288 in-sample data points and a fairly large number of draws are needed, we preferred to generate the ϵ_t sequence parametrically from a standard normal density.

in the interval $[\min(q_t), \max(q_t)]$ was used to compute and plot the non-parametric estimate of $I\!E(\Delta q_t | q_{t-h})$.[14] Note that the reason why the parameter settings of the UK series was chosen is that it yields the largest estimate of the transition function parameter γ. Recall that, given the range of the transition variable, the strength of the non-linearity in the ESTAR model is governed by the size of the γ parameter, where values close to 0 indicate weaker non-linearity and larger ones stronger non-linearity. To visualise how the non-linearity changes at different forecast horizons, we plot $I\!E(\Delta q_t | q_{t-h})$ for two sets of forecast horizons. These are $h = [1, 2, 3, 5, 6]$ and $h = [7, 10, 14, 18, 22]$ in Panels (a) and (b) of Figure 4, respectively. Notice from Panel (a) of Figure 4 that the non-linearity in the forecasts is strongest at the one step ahead horizon, that is, when $h = 1$. Both, the curvature, as well as the steepness, of the conditional means decreases at the transition points as the forecast horizon increases. For longer horizons shown in Panel (b) of Figure 4, it is evident that for forecasts of 10 steps ahead or longer (ie., when $h \geq 10$) no visual signs of non-linearity remain to be seen.

Why might one find this information useful? If the non-linearity in the conditional mean of the ESTAR model decreases monotonically as the forecast horizon increases, being the strongest at the one step ahead horizon, than it seems highly unlikely that any statistical tests evaluating the performance of the ESTAR model at longer forecast horizons will reject the null hypothesis of equal forecast accuracy. We can remind ourselves here again of the results obtained from the plots of the one step ahead conditional forecasts shown in Figure 2. Recall that not only was the difference between the conditional means of the competing models fairly small, but that the spread of the data around the conditional means was also substantial, so that it was impossible to statistically discriminate between the ESTAR and AR(1) models at the one step ahead out-of-sample data points. Since the non-linearity in the forecasts decreases as h increases, converging to the AR(1) forecast of the unconditional mean of Δq_t, and since the variation of the data around the conditional means remains fairly large, one should be convinced that no possibility exists for the considered ESTAR models to outperform the AR(1) models at any forecast horizon.

We can once again provide some formal statistical evidence in support of this conjecture by computing the weighted DM test for multiple step ahead forecasts considering horizons $h = [2, 3, 5, 6, 7, 10, 14, 18, 22]$. The results of this test are reported in Table 3 below.[15] The multiple steps ahead point forecasts from the ESTAR model — necessary to compute the DM test statistic — were constructed from the recursive scheme that was outlined in (3), where J was set to $10\,000$ and the $\tilde{\epsilon}_{T+h}^{j}$ were drawn from a standard normal distribution. It is evident form the results reported in Table 3 that the statistical tests

---

[14]The $\min(q_t)$ and $\max(q_t)$ values are those of the full sample data.

[15]The reason why only the results of the weighted DM test are reported here is purely to avoid repetition and to allow any potential non-linearity in the tails of q_t to be weighted favourably in the evaluation of the test. There is, nevertheless, qualitatively no difference in the results between the unweighted and weighted versions of the DM test, as both indicate a rather strong non-rejection of the null hypothesis.

confirm the conjectured failure of the ESTAR model. The null hypothesis of equal forecast accuracy cannot be rejected at any conventional significance levels and forecast horizon that we consider. Notice that for the UK series, the test statistic yields negative values which in some cases are large enough to suggest that the AR(1) model provides forecast gains over the non-linear model. Despite these results, however, it should be kept in mind here that the forecasts that the linear and non-linear models generate are very similar at higher forecast horizons. To see how similar they in fact are, regardless of their statistical significance, we show plots of the 10 step ahead point forecasts for all four series in Figure 5.[16] Notice how closely the conditional means of the competing models overlap, especially over intervals where the bulk of the out-of-sample data lies.

In conclusion of this section, it should be mentioned here that our finding of no forecast gains in favour of the ESTAR model is in contrast with the results reported in Rapach and Wohar (2006), who conclude that "...*ESTAR models offer forecasting gains at long horizons relative to simple linear AR models for some countries, especially when we use a weighted MSFE criterion.* "(see Rapach and Wohar, 2006, pp. 350-352).

3.2. Density forecasts

Density forecasts play a fundamental role in the finance literature. In risk management, for example, density forecasts form a building block for risk measures such as Value-at-Risk and Expected Shortfall. As it is often reported in the literature that non-linear models can generate highly skewed and/or bi-modal forecast densities, especially when considering forecasts multiple periods ahead (cf. Lundbergh and Teräsvirta, 2002, p. 505), it is important to analyse how the conditional forecast densities of the fitted ESTAR and AR(1) models differ from one another. Understanding these differences will be of particular interest to a practitioner who relies on forecasts of the conditional distributions to price financial derivatives in risk management scenarios. Throughout this section, we will once again employ informal graphical techniques extensively to provide an intuitive visual assessment of the forecast densities. As in the previous section, formal statistical tests are then used to supplement and validate any conjectures drawn from the visual assessment.

3.2.1. Assessing one step ahead density forecasts

In the given context, ie., under the assumption that the ϵ_t are distributed as a standard normal random variable, it is trivial to compute the one step ahead forecast densities for

---

[16]The contents of the plot are the same as in Figure 2. The ESTAR conditional mean (solid green line) was computed non-parametrically from 1 million simulated draws. Figure 5 also shows a scatter of the 10 step ahead conditional forecast constructed with the recursive scheme outlined in (3). These are superimposed onto the solid green line with black circles to show how they compare to one another.

the AR(1) and ESTAR models. These are, respectively

$$f_{T,1}^{AR}(\Delta q_{T+1}) = \mathsf{N}\left(\delta\left(q_T - \mu\right), \sigma_\mu^2\right) \tag{8}$$

and

$$f_{T,1}^{ESTAR}(\Delta q_{T+1}) = \mathsf{N}\left(-\left(q_T - \eta\right)\Phi(\gamma, \eta; q_T), \sigma_\eta^2\right), \tag{9}$$

where $\mathsf{N}(a, b)$ denotes the Gaussian density function with location and scale parameters a and b respectively.

Notice from (8) and (9) that, because the same functional form for the density of the stochastic process is assumed, a comparison of the one step ahead forecast densities reduces to one of equal conditional means if $\sigma_\eta^2 = \sigma_\mu^2$, and therefore boils down to an evaluation of the point forecasts as in Section 3.1. A statistical test of equal density forecasts should, therefore, lead to the same qualitative conclusion as a test of equal conditional means. Although it is not clear whether the population quantities are such that $\sigma_\eta^2 = \sigma_\mu^2$, it is evident from the estimates of σ_η^2 and σ_μ^2 reported in Table 1 that the difference between the sample quantities is very small. It can therefore be conjectured here that there exists very little evidence to suggest that the forecast densities of the AR(1) and ESTAR models differ from one another at the one step ahead horizon, given that the conditional means were found to be statistically indistinguishable in Section 3.1 and that the differences between sample quantities of σ_η^2 and σ_μ^2 are very small.

This conjecture can be tested formally by comparing the performance of the two density forecasts $f_{T,1}^{AR}(\Delta q_{T+1})$ and $f_{T,1}^{ESTAR}(\Delta q_{T+1})$ relative to the true, but unobserved, density of Δq_{T+1}. The statistical approach implemented here is a logarithmic scoring rule that is based upon the difference of the Kullback-Leibler Information Criterion (KLIC) of the competing density forecasts, which has the interpretation of a goodness of fit test (see Mitchell and Hall, 2005, Bao *et al.*, 2007 and Amisano and Giacomini, 2007). Taking the difference of the KLICs of the competing densities ensures that the term involving the true but unknown density of Δq_{T+1} drops out, so that the comparison based on the KLICs boils down to a comparison of the logarithmic scores.[17]

The idea behind the comparison of the logarithmic scores is to give a higher (lower) score to a density forecast if a given out-of-sample observation falls within a high (low) probability region. The density forecast that yields the highest average score is then preferred. The difference between the average scores can be tested statistically by defining the (log) score difference

$$d_{T+1}^S = \log f_{T,1}^{ESTAR}(\Delta q_{T+1}) - \log f_{T,1}^{AR}(\Delta q_{T+1}) \tag{10}$$

---

[17]The use of the term score here should not be confused with the first order condition in Maximum Likelihood estimation, which is often referred to as the Score (or Fisher Score).

and evaluating the null hypothesis of equal average scores by means of a DM type test as in Section 3.1. Given that both forecast densities follow a Gaussian distribution, (10) reduces to

$$d_{T+1}^S = -\log\left(\frac{\sigma_\eta}{\sigma_\mu}\right) - \frac{1}{2}\left[\left(\frac{\varepsilon_{T+1|T}^{ESTAR}}{\sigma_\eta}\right)^2 - \left(\frac{\varepsilon_{T+1|T}^{AR}}{\sigma_\mu}\right)^2\right] \tag{11}$$

which can then be used to compute the average score over the out-of-sample observations and to construct the corresponding DM test of equal density forecasts.[18]

The results of the DM test of equal density forecasts at the one step ahead horizon are reported in the first row of Table 4. Recall here that the preferred model is the one that yields, on average, the highest log score. Since d_{T+1}^S in (10) is written in such a way that the AR(1) log density is subtracted from the ESTAR log density, we again form the null hypothesis of equal density forecasts as

$$\mathcal{H}_0 : I\!E(d_{T+1}^S) = 0$$

against the alternative

$$\mathcal{H}_A : I\!E(d_{T+1}^S) > 0$$

to test for the superiority of the ESTAR models' density forecasts. A significantly large positive value of the out-of-sample average of d_{T+1}^S would thus suggest that the ESTAR density outperforms the simple AR(1). Nevertheless, notice from the results of this test reported in Table 4 that all $t-$statistics with positive entries remain well below one, while those of the UK and Japanese series even yield negative entries. We can conclude here, therefore, that no statistical evidence exists to suggest that the densities differ from one another at the one step ahead horizon.

3.2.2. Assessing multiple steps ahead density forecasts

For the AR(1) model, multiple steps ahead density forecasts are available in closed form, given the assumption that the ϵ_t are distributed as a standard normal random variable. The h step ahead forecast density takes the form

$$f_{T,h}^{AR}(\Delta q_{T+h}) = \mathsf{N}\left(\delta\rho^{(h-1)}\left(q_T - \mu\right), \left[\sigma_\mu^2 + \frac{\sigma_\mu^2\delta^2}{(1-\rho^2)}\left(1 - \rho^{2(h-1)}\right)\right]\right) \tag{12}$$

where $\rho = \delta + 1$ and $\sigma_\mu^2 + \frac{\sigma_\mu^2\delta^2}{(1-\rho^2)}$ is the unconditional variance of the Δq_t process in (2). For the ESTAR model, nevertheless, no closed form is available so that it is again necessary

---

[18]Notice here, that, as discussed before, when $\sigma_\eta = \sigma_\mu$, then the first term involving the logs disappears, and the second term becomes $(2\sigma_\mu^2)^{-1}[(\varepsilon_{T+1|T}^{AR})^2 - (\varepsilon_{T+1|T}^{ESTAR})^2]$. This is thus a scaled version of the DM test of equal conditional means given previously in (5).

to resort to the recursive simulation scheme of (3) to construct the h step ahead pseudo values $\tilde{q}_{T+h|T}$. These can then be used with non-parametric methods to get an estimate of the forecast density. That is, given the sequence of pseudo realisations $\{\tilde{q}^j_{T+h|T}\}^J_{j=1}$ we can obtain an approximation of the h step ahead forecast density from the ESTAR model by constructing $\Delta\tilde{q}^j_{T+h|T} = \tilde{q}^j_{T+h|T} - \tilde{q}^j_{T+h-1|T}, \forall j = 1, ..., J$ generated from (3) and then compute the density estimate of $f^{ESTAR}_{T,h}(\Delta\tilde{q}_{T+h|T})$ non-parametrically. The kernel density estimate can then be utilised for visualisation purposes and to compute the average of the log score in the DM test.

One drawback with this approach when considering informal graphical methods is that one will again only be able to visualise the h step ahead density at the actual out-of-sample values that are conditioned upon. It will thus not be possible to get a feel for how the forecast density changes as the size of the conditioning variable changes, unless there is substantial variation in the actual out-of-sample observations. To illustrate this, consider the plot of the 10 step ahead conditional point forecasts for the Japanese series shown in Panel (c) of Figure 5. Notice that the out-of-sample values of the conditioning variable denoted by the black asterisks cluster largely around a value of 0.5. If we use the Monte Carlo scheme of (3) to generate 10 000 paths from each of the given q_T to compute the forecast density, we will not know whether the forecast density takes on a different shape when q_T is closer to the extreme tail ends of either 0 or 1.

A more informative approach is to simulate again a large number of draws from the ESTAR models in (1) and then compute an estimate of the conditional density of $\Delta q_t | q_{t-h}$ directly using non-parametric methods. That is, compute

$$\hat{f}^{NP}(\Delta q_t | q_{t-h}) = \frac{\hat{f}^{NP}(\Delta q_t, q_{t-h})}{\hat{f}^{NP}(q_{t-h})}, \tag{13}$$

where $\hat{f}^{NP}(\cdot)$ is a non-parametric estimate of the density. The values of q_{t-h} that are conditioned upon could then be chosen to be some percentiles of interest of q_{t-h}.

To illustrate how the conditional density estimate $\hat{f}^{NP}(\Delta q_t | q_{t-h})$ can be visualised, we simulate 1 million draws from the ESTAR model in (1) under the parameter settings of the UK series and set the conditioning values at the $5^{th}, 25^{th}, 50^{th}, 75^{th}$ and 95th percentiles of q_{t-h}. A Gaussian kernel and a plug in bandwidth that is proportional to the covariance matrix of $(\Delta q_t \ q_{t-h})'$ were used to construct the bivariate density estimates. Plots of the estimates of the conditional densities of $f^{NP}(\Delta q_t | q_{t-h}), \forall h = [2, 3, 5, 6, 7, 10, 14, 18, 22]$ are shown in Figure 6. What is particularly interesting to notice from Figure 6 is that there is no obvious visual indication of skewness or bi-modality in the forecast densities. This is regardless of the forecast horizon considered and the conditioning values from which the forecasts were initiated. Two other features that are interesting to observe from Figure 6 are the lack of a visual widening in the spread of the forecast densities as h increases and

the close overlap of the forecast densities at the different percentiles of q_{t-h}. Both of these are due to the weak correlation between Δq_t and q_{t-1}, or alternatively, the high persistence in q_t.

The easiest way to see why this is the case, consider the AR(1) representation for Δq_t in (2) to be the true process for Δq_t. If $\delta = 0$, then Δq_t and q_{t-1} are uncorrelated and hence independently distributed so that $f(\Delta q_t | q_{t-h}) = f(\Delta q_t) \sim \mathsf{N}(0, \sigma_\mu^2)$. Thus for all conditioning values of q_{t-h} the location of $f(\Delta q_t | q_{t-h})$ is at 0. Similarly, the spread of the density at the different forecast horizons will be fixed at σ_μ^2. Although it is clear here that Δq_t and q_{t-1} are not independent processes as they were simulated from the ESTAR model in (1), it is evident from Figure 6 how closely the densities overlap at the different conditioning values of q_{t-h} and how the spread in the densities remains observationally constant. This is indicative of a relatively weak relationship between Δq_t and q_{t-1}.

Before we proceed to provide some formal statistical evidence to support any of our conjectures, it will be useful here to do a side-by-side comparison of the forecast densities of the two competing models. As we have ruled out that the shape of the ESTAR forecast density changes at different conditioning values of q_T, we can choose a fixed value of q_T and plot the forecast density $f_{T,h}(\Delta q_{T+h})$ for the AR(1) together with the simulated density estimate from the ESTAR model at the forecast horizons of interest to us. Such a comparison is shown in Figure 7, again only for the UK series to avoid unnecessary repetition. The conditioning value used here for q_T is approximately 0.5 (the November 2007 value), which is the (full sample) maximum value of q_T. The ESTAR forecast densities plotted in Figure 7 were computed from the 10 000 pseudo observations $\{\tilde{q}_{T+h|T}^j\}_{j=1}^J$ outlined in (3), using again standard kernel density estimation methods.

The comparison of the multiple step ahead densities plotted in Figure 7 shows a number of notable features. Although these were partially discussed and hence are expected, it is nevertheless informative to outline these once again, however, with a visual reference to Figure 7. Firstly, notice that at the 2 to 7 step ahead forecast horizon the densities are somewhat offset and do not overlap, nevertheless, there is no indication of a markedly different shape or spread of these densities. Evidently, the densities do not overlap as the two models' forecasts of the conditional means differ from one another. For example, at the 2 step ahead horizon, the ESTAR and AR(1) models predict mean changes of about -0.025 and -0.012, respectively. The conditioning value of $q_T \approx 0.5$ for November 2007 was particularly chosen here to amplify this difference in the location of the forecast densities. Secondly, notice how there is no obvious visual increase in the spread of the densities as h increases from 2 to 22. For the AR(1) model, where an analytic expression for the forecast standard error is available, the values range narrowly between 33.4590×10^{-3} and 33.6246×10^{-3} at horizons 2 and 22, respectively. With the unconditional standard error of Δq_t under the AR(1) specification in (2) being 33.6952×10^{-3} (the limit at the h step horizon as $h \to \infty$), it is clear that the overall increase in the spread is very small, so

that any differences are hard to identify visually from Figure 7.

Formal statistical test results of equal h step ahead average density log scores are reported in Table 4. The unweighted version of the DM test was used in the computation of the log score difference in Table 4, again employing the correction factor of Harvey *et al.* (1997). For all 4 series of interest — at all forecast horizons that we consider — the null hypothesis of equal average log scores cannot be rejected at any conventional significance levels. Hence, no evidence seems to exist to indicate that the considered ESTAR model generates any forecast gains over a simple AR(1) specification, regardless of whether point or density forecasts are utilised.

4. Conclusion

This study assessed the forecast performance of the widely cited ESTAR model of Taylor *et al.* (2001) over the out-of-sample period from January 1997 to June 2008. More specifically, point and density forecasts were constructed and evaluated for four empirical real exchange rate series, using a simple AR(1) as the benchmark model. Throughout the study heavy use of graphical methods was made in conjunction with simulation and non-parametric techniques. This was done to supplement the standard formal statistical tests in the analysis and evaluation of the forecasts, and to learn about the models and their fit to the data.

The statistical test results in this study show that there exist no forecast gains from utilising a non-linear ESTAR model over a simple AR(1) specification at any of the 1 to 22 steps ahead forecast horizon that were consider. This holds true for conditional mean (or point) forecasts, as well as for density forecasts. Graphical methods that are utilised throughout the paper show that the non-linearity in the one step ahead point forecasts of the ESTAR model is relatively weak, given the variation in the empirical data, and that it decreases monotonically as the forecast horizon increases. Therefore, as no forecast gains are realised at the one step ahead horizon, there exists no potential whatsoever for any forecast gains to be realised at longer horizons. The graphical analysis shows also that the forecast densities are approximately normal looking without any signs of skewness or kurtosis.

On a broader level, it is interesting to observe from the graphical analysis that over the total of 35 years of real exchange rate data that were utilised, a significant proportion of the variation of the empirical series around the conditional means still remains unexplained. One might feel therefore that a non-linear specification as embodied in the ESTAR model is still short of being satisfactory in explaining the PPP puzzles raised by the international finance literature.

References

Amisano, Gianni and Raffaella Giacomini (2007): "Comparing Density Forecasts via Weighted Likelihood Ratio Tests", *Journal of Business and Economic Statistics*, **25** (2), 177–190.

Bao, Yong, Tae-Hwy Lee and Burak Saltoğlu (2007): "Comparing density forecast models", *Journal of Forecasting*, **26** (3), 203–225.

Baum, Christopher F., John T. Barkoulas and Mustafa Caglayan (2001): "Nonlinear Adjustment to Purchasing Power Parity in the Post-Bretton Woods Era", *Journal of International Money and Finance*, **20** (3), 379–399.

Breunig, Robert V., Serineh Najarian and Adrian R. Pagan (2003): "Specification Testing of Markov Switching Models", *Oxford Bulletin of Economics and Statistics*, **65** (S1), 703–725.

Cheung, Yin-Wong and Menzie D. Chinn (2000): "Currency Traders and Exchange Rate Dynamics: A Survey of the U.S. Market", *Journal of International Money and Finance*, **20** (4), 439–471.

Cheung, Yin-Wong, Menzie D. Chinn and Ian W. Marsh (2004): "How Do UK-Based Foreign Exchange Dealers Think Their Market Operates?" *International Journal of Finance and Economics*, **9** (4), 289–306.

De Grauwe, Paul and Marianna Grimaldi (2005): "Heterogeneity of Agents, Transaction Costs and the Exchange Rate", *Journal of Economic Dynamics and Control*, **29** (4), 691–719.

——————— (2006): "Exchange Rate Puzzles: A Tale of Switching Attractors", *European Economic Review*, **50** (1), 1–33.

Diebold, Francis X. and Roberto S. Mariano (1995): "Comparing Predictive Accuracy", *Journal of Business and Economic Statistics*, **13** (1), 253–263.

Eitrheim, Øyvind and Timo Teräsvirta (1996): "Testing the Adequacy of Smooth Transition Autoregressive Models", *Journal of Econometrics*, **74** (1), 59–75.

Franses, Philip Hans and Dick van Dijk (2000): *Nonlinear Time Series Models in Empirical Finance*, Cambridge: Cambridge University Press.

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

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

Harvey, David I., Stephen J. Leybourne and Paul Newbold (1997): "Testing the Equality of Prediction Mean Squared Errors", *International Journal of Forecasting*, **13** (2), 281–291.

Kilian, Lutz and Mark P. Taylor (2003): "Why Is It So Difficult to Beat the Random Walk Forecast of Exchange Rates?" *Journal of International Economics*, **60** (1), 85–107.

Lundbergh, Stefan and Timo Teräsvirta (2002): "Forecasting with Smooth Transition Autoregressive Models", in *A Companion to Economic Forecasting*, edited by Michael P. Clements and David F. Hendry, Blackwell Publishing, 485–509.

McCracken, Michael W. and Kenneth D. West (2002): "Inference about Predictive Ability", in *A Companion to Economic Forecasting*, edited by Michael P. Clements and David F. Hendry, Oxford: Blackwell Publishers, 299–321.

Michael, Panos, A. Robert Nobay and David A. Peel (1997): "Transaction Costs and Nonlinear Adjustment in Real Exchange Rates: An Empirical Investigation", *Journal of Political Economy*, **105** (4), 862–879.

Mitchell, James and Stephen G. Hall (2005): "Evaluating, Comparing and Combining Density Forecasts Using the KLIC with an Application to the Bank of England and NIESR 'Fan' Charts of Inflation", *Oxford Bulletin of Economics and Statistics*, **67** (S1), 995–1033.

Newey, Whitney K. and Kenneth D. West (1987): "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix", *Econometrica*, **55** (3), 703–708.

Obstfeld, Maurice and Alan M. Taylor (1997): "Nonlinear Aspects of Goods-Market Arbitrage and Adjustment: Hecksher's Commodity Points Revisited", *Journal of the Japanese and International Economies*, **11** (4), 441–479.

Pagan, Adrian R. (2002): "Learning About Models and Their Fit to Data", *International Economic Journal*, **16** (2), 1–18.

Pagan, Adrian R. and Aman Ullah (1999): *Nonparametric Econometrics*, New York: Cambridge University Press.

Rapach, David E. and Mark E. Wohar (2006): "The out-of-Sample Forecasting Performance of Nonlinear Models of Real Exchange Rate Behavior", *International Journal of Forecasting*, **22** (2), 341– 361.

Rogoff, Kenneth S. (1996): "The Purchasing Power Parity Puzzle", *Journal of Economic Literature*, **34** (2), 647–668.

Scott, David W. (1992): *Multivariate Density Estimation: Theory, Practice, and Visualization*, New York, Chichester: John Wiley and Sons.

Silverman, B. W. (1986): *Density Estimation for Statistics and Data Analysis*, London: Chapman and Hall.

Taylor, Mark P. and David A. Peel (2000): "Nonlinear Adjustment, Long-Run Equilibrium and Exchange Rate Fundamentals", *Journal of International Money and Finance*, **19** (1), 33–53.

Taylor, Mark P., David A. Peel and Lucio Sarno (2001): "Nonlinear Mean-Reversion in Real Exchange Rates: Towards a Solution to the Purchasing Power Parity Puzzles", *International Economic Review*, **42** (4), 1015–1042.

Teräsvirta, Timo (1994): "Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models", *Journal of the American Statistical Association*, **89** (425), 208–218.

van Dijk, Dick and Philip Hans Franses (2003): "Selecting a Nonlinear Time Series Model Using Weighted Tests of Equal Forecast Accuracy", *Oxford Bulletin of Economics and Statistics*, **18** (S1), 727–744.

Westerhoff, Frank H. and Stefan Reitz (2003): "Nonlinearities and Cyclical Behavior: The Role of Chartists and Fundamentalists", *Studies in Nonlinear Dynamics and Econometrics*, **7** (4), Article 3.
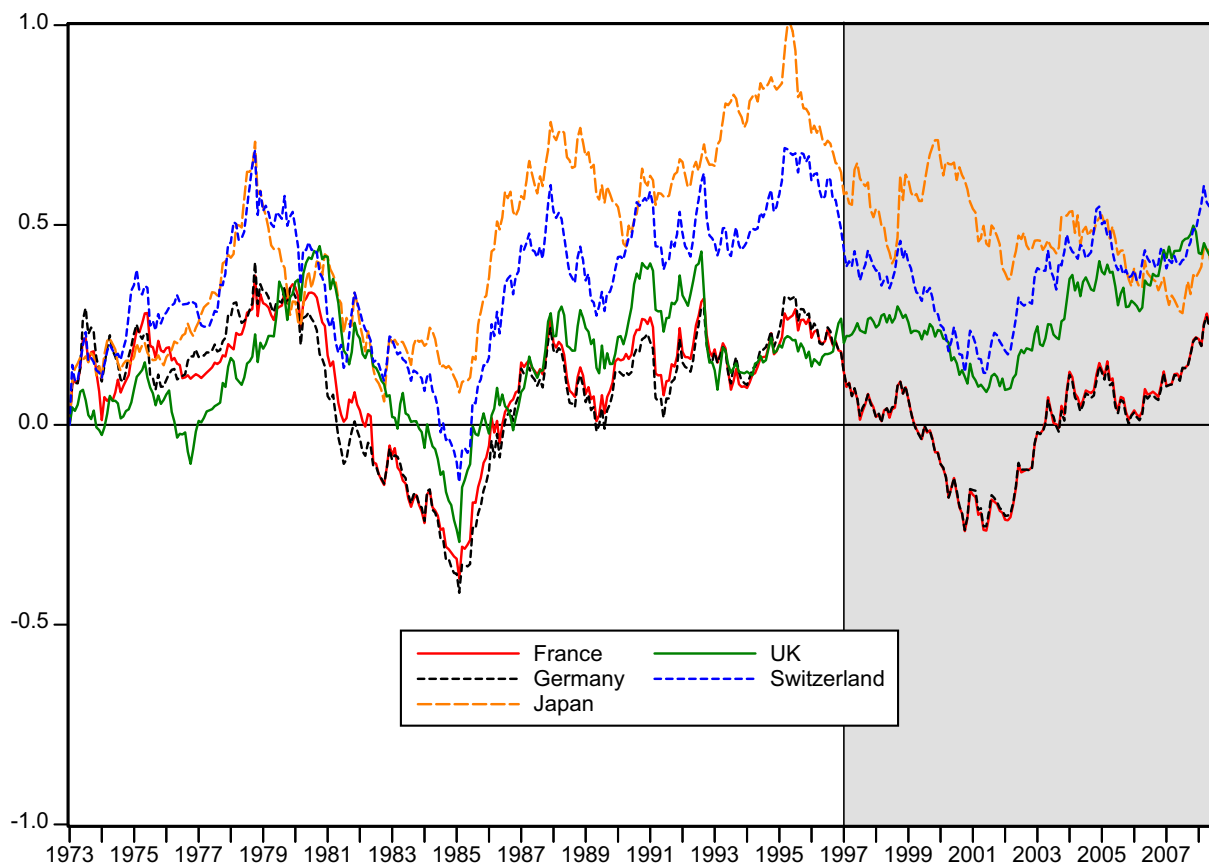
Figures and Tables

Figure 1: Time series plot of the normalised real exchange rates over the period from January 1973 to June 2008. The non-shaded and shaded areas denote the in-sample (January 1973 – December 1996) and out-of-sample (January 1997 – June 2008) periods, respectively.

Table 1: ESTAR and AR(1) in-sample parameter estimates.

ESTAR	UK	Germany	France	Japan	Switzerland
γ (se)	0.5056 (0.0727)	0.2933 (0.2254)	0.3536 (0.2523)	0.1819 (0.1229)	0.3742 (0.2391)
η (se)	0.1125 (0.4103)	-0.0115 (0.0693)	0.0059 (0.0614)	0.5102 (0.0776)	0.3142 (0.0624)
σ_η	0.033324	0.034502	0.033061	0.033390	0.038275
$\mathscr{L}(\gamma, \eta)$	569.99	560.02	572.94	569.42	530.23
$LM_{AR(1)}$ [$p-value$]	0.1691 [0.6812]	0.1478 [0.7009]	0.1561 [0.6930]	0.1656 [0.6843]	0.0818 [0.7751]
$LM_{AR(1-4)}$ [$p-value$]	0.1781 [0.9496]	0.1750 [0.9511]	0.1725 [0.9523]	0.1753 [0.9510]	0.13386 [0.9699]
LM_{NL3} [$p-value$]	1.0697 [0.3623]	1.1747 [0.3197]	1.0856 [0.3554]	0.4142 [0.7429]	0.9334 [0.4248]
AR(1)	**UK**	**Germany**	**France**	**Japan**	**Switzerland**
δ (se)	-0.0297 (0.0199)	-0.0219 (0.0157)	-0.0233 (0.0166)	-0.0147 (0.0096)	-0.0288 (0.0154)
μ (se)	0.1759 (0.0756)	0.1317 (0.0992)	0.1413 (0.0891)	0.5981 (0.1907)	0.4158 (0.0864)
σ_μ	0.033444	0.034640	0.033117	0.033579	0.038385
$\mathscr{L}(\delta, \mu)$	568.96	558.87	571.78	567.81	529.41

Notes: ESTAR and AR(1) parameter estimates over the in-sample period from January 1973 to December 1996. The maximum of the log-likelihood is denoted by $\mathscr{L}(\cdot)$. $LM_{AR(1)}$ and $LM_{AR(1-4)}$ are $F-$statistics of Langrange Multiplier (LM) test for first and first to fourth order serial correlation in the residuals, constructed as in Eitrheim and Teräsvirta (1996). LM_{NL3} is the $F-$statistics for a test for remaining ESTAR non-linearity (see Eitrheim and Teräsvirta, 1996, page 65).
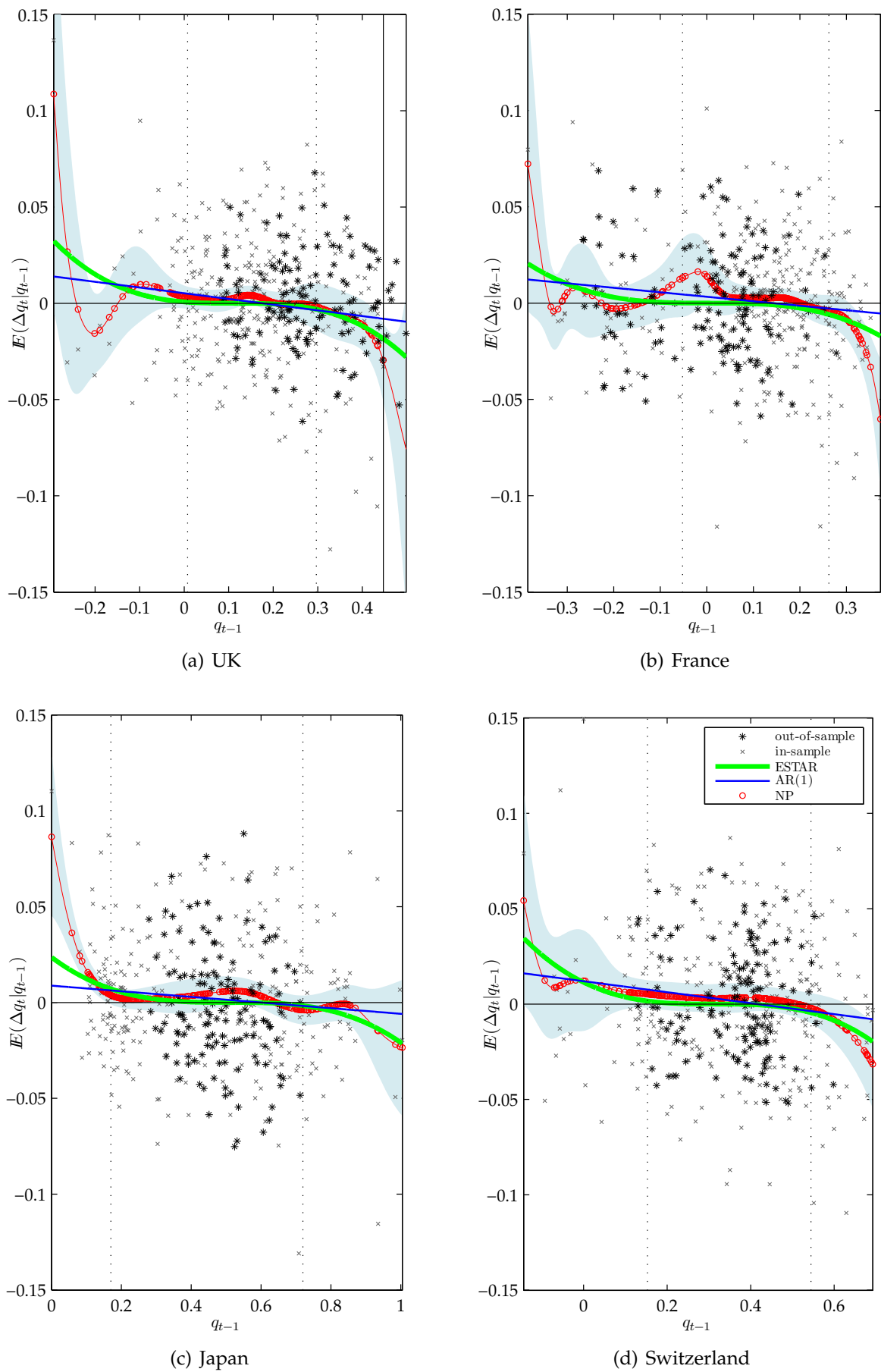
Figure 2: One step ahead point forecasts. The thick green and thin blue lines show the one step ahead conditional forecasts of the ESTAR and AR(1) models, respectively. Red circles are the non-parametric conditional means, with 95% confidence intervals drawn as blue shading. Grey crosses mark the in-sample data. Vertical dotted lines are drawn at the 15th and 85th percentiles of q_{t-1}. Black asterisks denote the out-of-sample data.
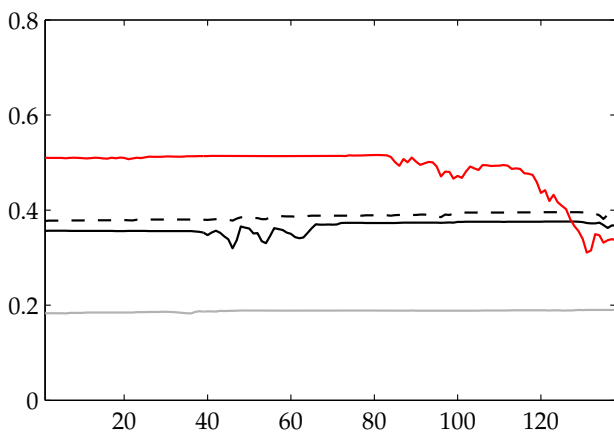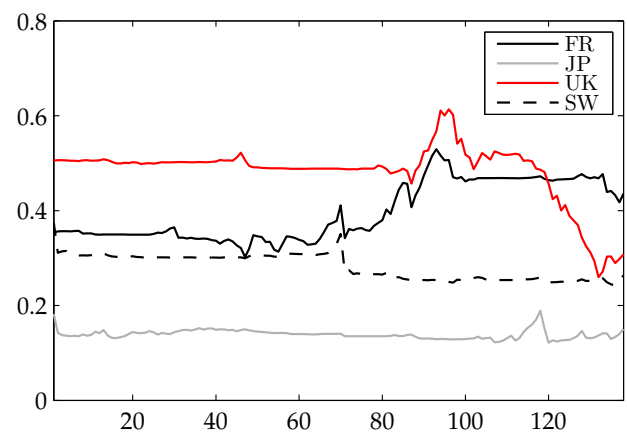
(a) Expanding Window   (b) Fixed Window

Figure 3: Recursive estimates of the γ parameter. Panel (a) shows the estimates under an expanding window, where one extra observation from the out-of-sample data is included in each updating step. Panel (b) shows the estimates under a fixed $T = 287$ sample size when rolling through the out-of-sample data. At each updating step, one observation is dropped at the beginning of the sample as each new one is added at the end.

Table 2: Unweighted and weighted DM test results for one step ahead point forecasts.

DM statistic	UK	France	Switzerland	Japan
$\bar{d}$	-3.37×10^{-5}	9.68×10^{-6}	9.71×10^{-7}	-1.59×10^{-6}
(se)	(2.51×10^{-5})	(2.08×10^{-5})	(1.41×10^{-5})	(8.87×10^{-6})
[t−statistic]	$[-1.3406]$	$[0.4645]$	$[0.0688]$	$[-0.1792]$
$\omega\bar{d}$	-2.32×10^{-5}	6.66×10^{-6}	3.51×10^{-7}	-1.05×10^{-6}
(se)	(1.78×10^{-5})	(1.47×10^{-5})	(6.08×10^{-6})	(1.31×10^{-6})
[t−statistic]	$[-1.3048]$	$[0.4538]$	$[0.0577]$	$[-0.8013]$

Notes: Standard $(\bar{d})$ and weighted $(\omega\bar{d})$ Diebold and Mariano (1995, DM) test statistics. Standard errors (se) are of the Newey and West (1987, NW) type. $\bar{d}$ was calculated as the arithmetic mean of $d_{T+1} \equiv (\varepsilon_{T+1|T}^{AR})^2 - (\varepsilon_{T+1|T}^{ESTAR})^2$ over the out-of-sample data, with $\varepsilon_{T+1|T}^{AR}$ and $\varepsilon_{T+1|T}^{ESTAR}$ being the one step ahead forecast errors from the AR(1) and ESTAR models, respectively. The small sample correction factor of Harvey *et al.* (1997) was used in the construction of both test statistics. $\omega\bar{d}$ was computed as the arithmetic mean of $\omega_{T+1}d_{T+1}$, where $\omega_{T+1} = 1 - \hat{f}(q_{T+1})/\max[\hat{f}(q_{T+1})]$ and $\hat{f}(q_{T+1})$ is an estimate of the density function of q_{T+1}, evaluated at the out-of-sample data points. A Gaussian kernel with a plug in bandwidth were used to compute the density estimate.
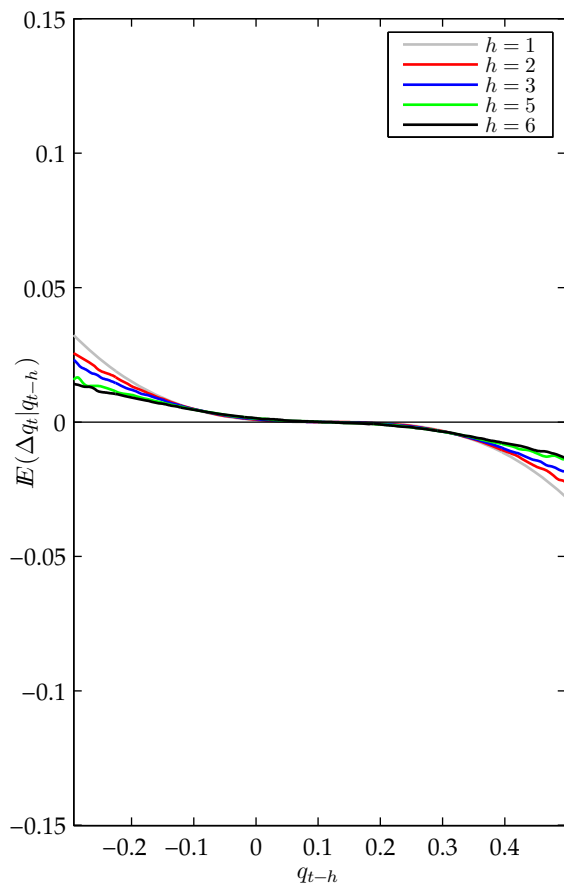
(a) $h = 1, 2, 3, 5, 6$.
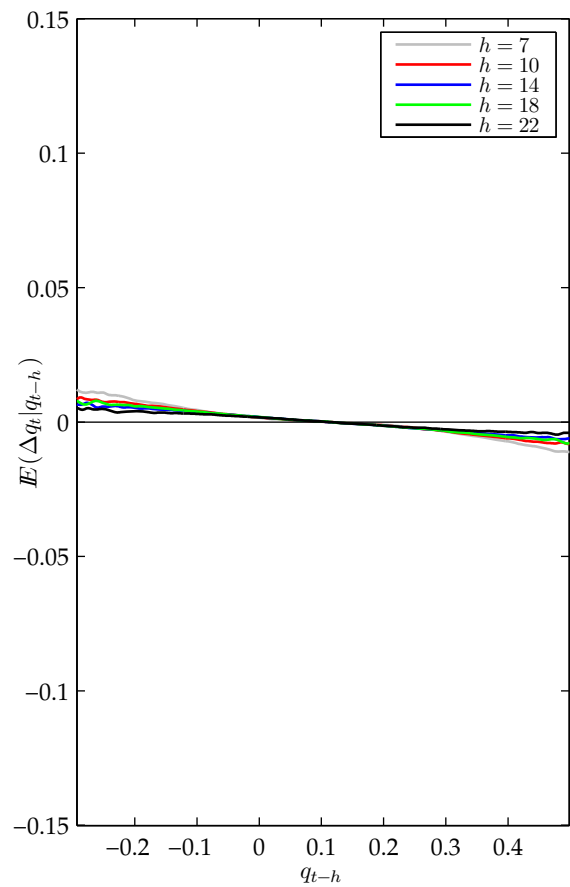
(b) $h = 7, 10, 14, 18, 22$.

Figure 4: Conditional means corresponding to h step ahead forecast. These were obtained as non-parametric estimates of the conditional mean from 1 million simulated pseudo observations from the ESTAR model of Taylor *et al.* (2001) at the parameter values of the UK series. The conditional mean $\mathbb{E}(\Delta q_t | q_{t-k})$ was computed over 1000 equally spaced grid points in the interval $[\min(q_t), \max(q_t)]$.

Table 3: Weighted DM test results for multiple step ahead point forecasts.

| DM statistic	h	UK	France	Switzerland	Japan
$\omega\overline{d}$ (se) [t−statistic]	2	-1.74×10^{-5} (1.25×10^{-5}) $[-1.3991]$	6.80×10^{-6} (1.35×10^{-5}) $[0.5021]$	-1.07×10^{-6} (5.35×10^{-6}) $[-0.2003]$	-1.17×10^{-6} (1.37×10^{-6}) $[-0.8544]$
$\omega\overline{d}$ (se) [t−statistic]	3	-1.60×10^{-5} (8.70×10^{-6}) $[-1.8420]$	3.34×10^{-6} (1.20×10^{-5}) $[0.2779]$	-1.99×10^{-6} (5.24×10^{-6}) $[-0.3789]$	-1.44×10^{-6} (1.55×10^{-6}) $[-0.9292]$
$\omega\overline{d}$ (se) [t−statistic]	5	-7.84×10^{-6} (3.90×10^{-6}) $[-2.0083]$	2.32×10^{-6} (1.02×10^{-5}) $[0.2288]$	1.18×10^{-6} (4.52×10^{-6}) $[0.2613]$	-8.85×10^{-7} (9.86×10^{-7}) $[-0.8979]$
$\omega\overline{d}$ (se) [t−statistic]	6	-4.39×10^{-6} (2.89×10^{-6}) $[-1.5161]$	9.68×10^{-7} (9.20×10^{-6}) $[0.1051]$	1.80×10^{-6} (3.76×10^{-6}) $[0.4792]$	5.89×10^{-7} (1.00×10^{-6}) $[0.5858]$
$\omega\overline{d}$ (se) [t−statistic]	7	-3.20×10^{-6} (2.30×10^{-6}) $[-1.3896]$	4.38×10^{-6} (7.78×10^{-6}) $[0.5629]$	1.39×10^{-6} (3.73×10^{-6}) $[0.3725]$	4.55×10^{-7} (9.53×10^{-7}) $[0.4774]$
$\omega\overline{d}$ (se) [t−statistic]	10	-2.65×10^{-6} (9.40×10^{-7}) $[-2.8219]$	2.59×10^{-7} (6.63×10^{-6}) $[0.0391]$	-1.86×10^{-6} (3.04×10^{-6}) $[-0.6113]$	2.86×10^{-7} (8.02×10^{-7}) $[0.3573]$
$\omega\overline{d}$ (se) [t−statistic]	14	-1.48×10^{-6} (9.78×10^{-7}) $[-1.5170]$	-6.98×10^{-7} (4.46×10^{-6}) $[-0.1565]$	-2.11×10^{-6} (2.29×10^{-6}) $[-0.9221]$	4.80×10^{-7} (7.13×10^{-7}) $[0.6732]$
$\omega\overline{d}$ (se) [t−statistic]	18	-1.39×10^{-6} (7.89×10^{-7}) $[-1.7564]$	-3.08×10^{-6} (3.93×10^{-6}) $[-0.7843]$	-1.84×10^{-6} (1.98×10^{-6}) $[-0.9306]$	4.81×10^{-7} (3.34×10^{-7}) $[1.4395]$
$\omega\overline{d}$ (se) [t−statistic]	22	-6.10×10^{-7} (5.88×10^{-7}) $[-1.0378]$	-1.82×10^{-6} (3.40×10^{-6}) $[-0.5347]$	-1.16×10^{-6} (1.33×10^{-6}) $[-0.8718]$	-1.09×10^{-7} (3.05×10^{-7}) $[-0.3581]$

Notes: The weighted DM test statistic $\omega\overline{d}$ and its standard error (se) for multiple step ahead point forecasts. The statistics were computed as documented in Table 2.
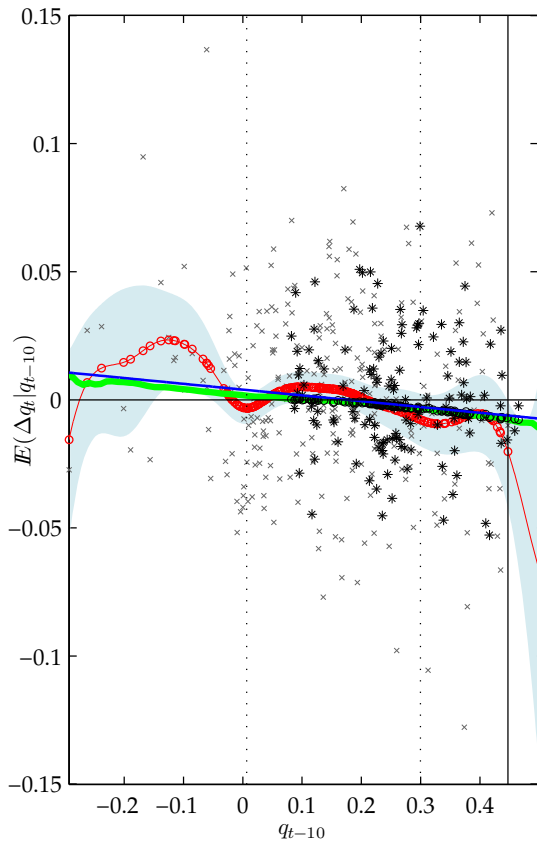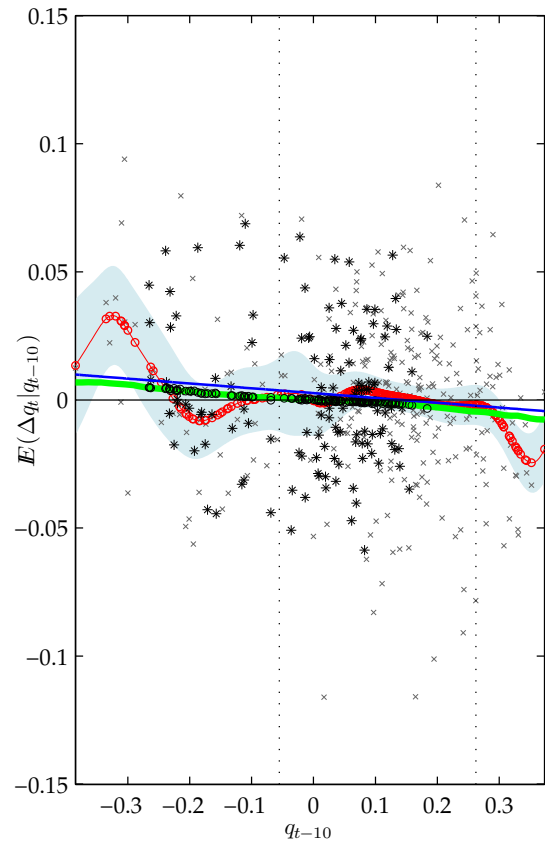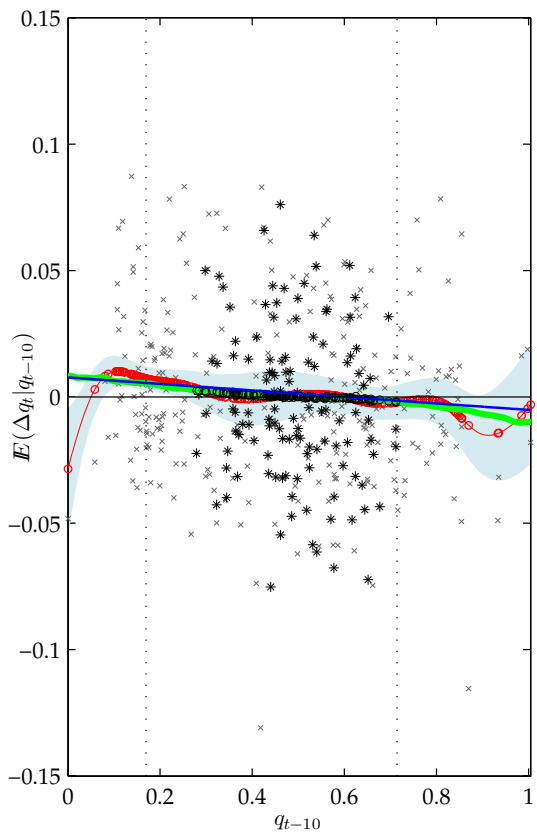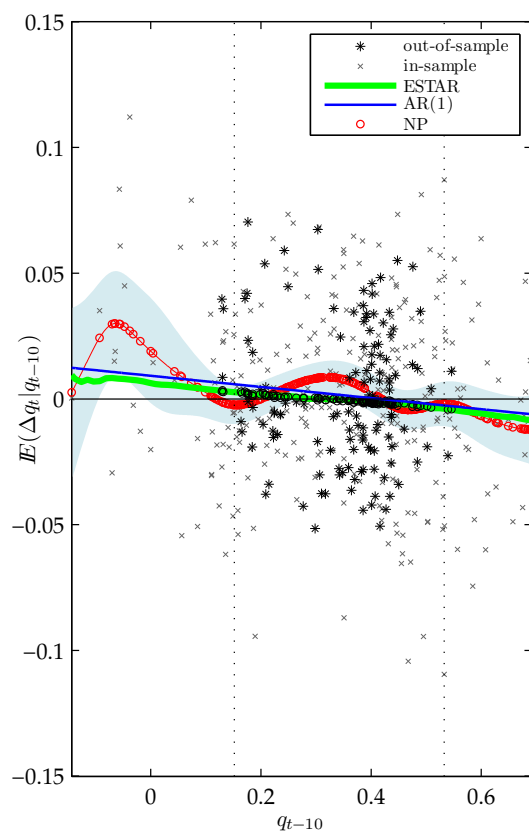
Figure 5: 10 step ahead point forecasts. The contents are the same as in Figure 2. Black circles are superimposed onto the NP conditional mean (solid green line) to mark the 10 step-ahead conditional forecast computed from the recursive scheme outlined in (3) to facilitate the comparison to the NP conditional mean computed directly from 1 million simulated ESTAR realisations.
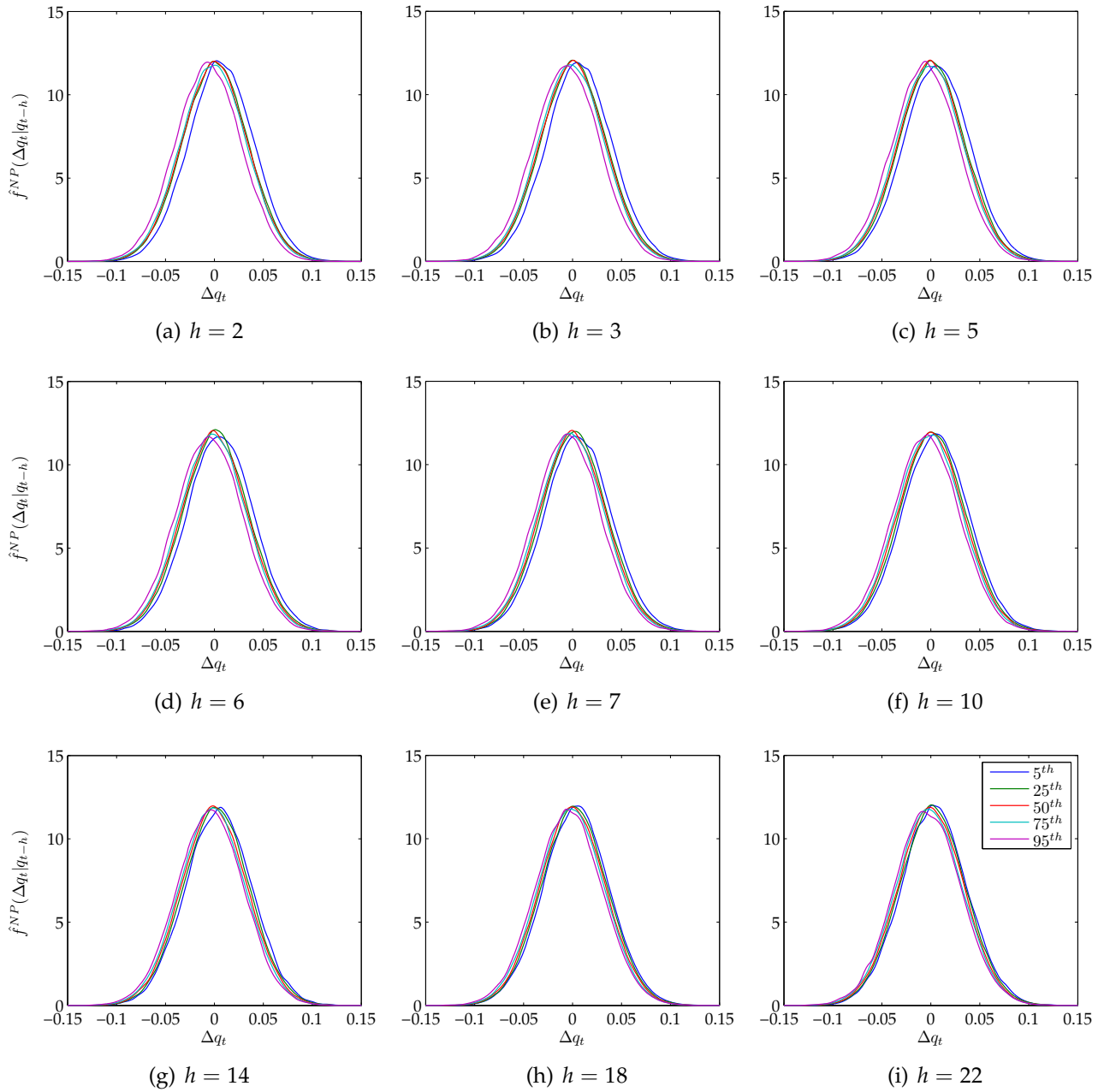
Figure 6: Multiple step ahead density forecasts of the ESTAR model. These were constructed non-parametrically from 1 million simulated realisations of the ESTAR model in (1) at the parameter values of the UK series. Gaussian univariate and bivariate kernels were used, together with plug in bandwidths that are proportional to the covariance matrix of the data (see Scott, 1992).

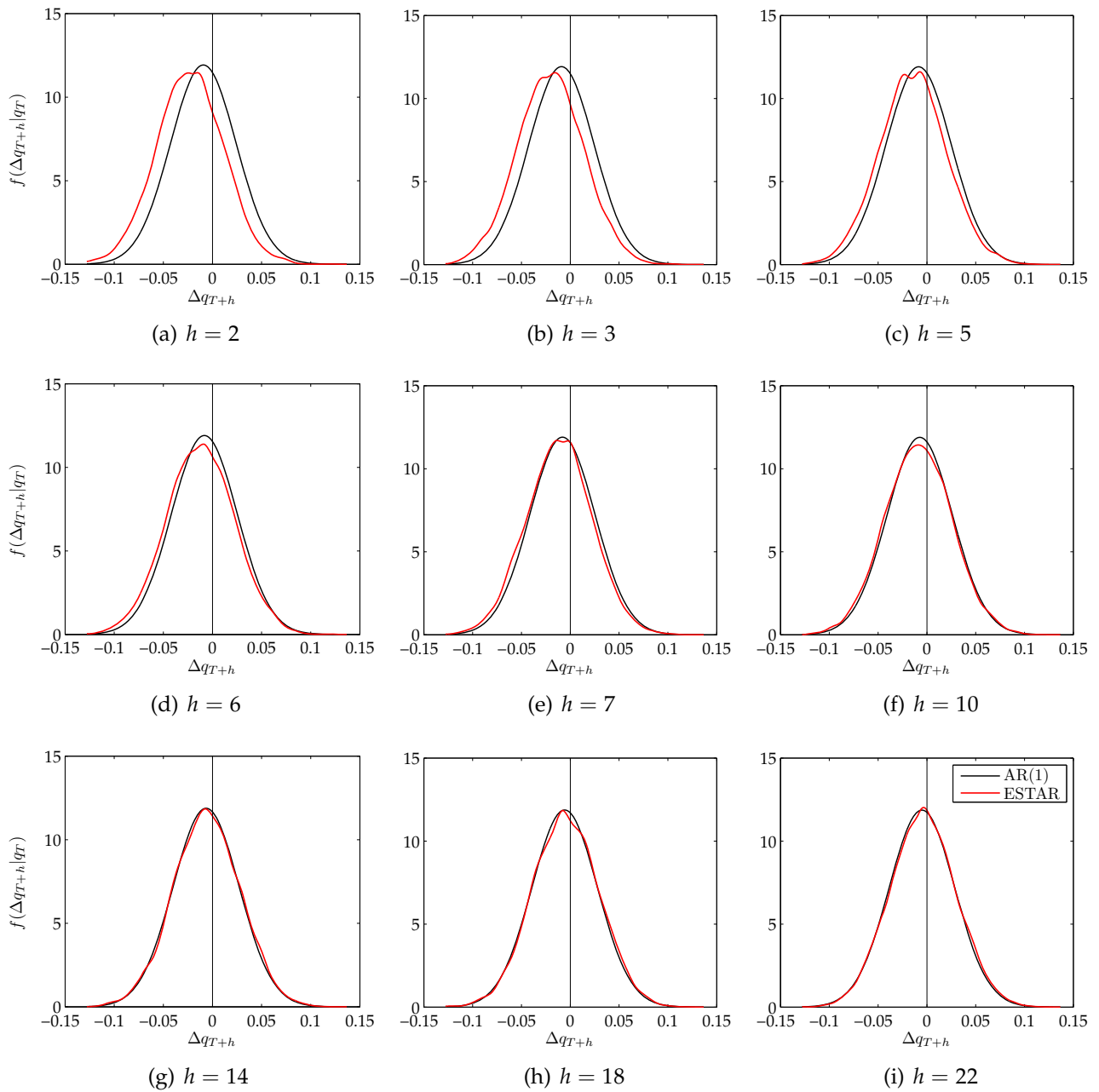**Figure 7:** Comparison of the multiple step ahead density forecasts of the AR(1) and ESTAR models for the UK real exchange rate series. The AR(1) densities were calculated from (2). The ESTAR densities were computed non-parametrically, using the 10 000 recursively constructed pseudo draws from (3). The conditioning value of q_T is approximately 0.5 (November 2007 value) from which the forecasts were initiated.

Table 4: DM test statistic for multiple step ahead density forecasts.

DM statistic	h	UK	France	Switzerland	Japan
$\overline{d^S}$ (se) [t−statistic]	1	-1.33×10^{-2} (1.13×10^{-2}) $[-1.1803]$	5.63×10^{-3} (9.49×10^{-3}) $[0.5934]$	9.94×10^{-4} (4.81×10^{-3}) $[-0.2067]$	-5.89×10^{-4} (4.61×10^{-3}) $[-0.1279]$
$\overline{d^S}$ (se) [t−statistic]	2	-4.19×10^{-2} (3.39×10^{-2}) $[-1.2369]$	4.20×10^{-2} (3.63×10^{-2}) $[1.1556]$	2.22×10^{-2} (2.15×10^{-2}) $[1.0342]$	-2.54×10^{-2} (7.73×10^{-2}) $[-0.3285]$
$\overline{d^S}$ (se) [t−statistic]	3	-2.97×10^{-2} (2.68×10^{-2}) $[-1.1111]$	3.41×10^{-2} (3.51×10^{-2}) $[0.9721]$	1.90×10^{-2} (2.08×10^{-2}) $[0.9141]$	-1.73×10^{-2} (3.19×10^{-2}) $[-0.5436]$
$\overline{d^S}$ (se) [t−statistic]	5	-1.24×10^{-2} (1.93×10^{-2}) $[-0.6444]$	2.69×10^{-2} (3.45×10^{-2}) $[0.7777]$	1.51×10^{-2} (2.05×10^{-2}) $[0.7359]$	-1.18×10^{-1} (1.69×10^{-1}) $[-0.7002]$
$\overline{d^S}$ (se) [t−statistic]	6	-7.64×10^{-3} (1.74×10^{-2}) $[-0.4401]$	2.33×10^{-2} (3.25×10^{-2}) $[0.7162]$	1.30×10^{-2} (2.00×10^{-2}) $[0.6506]$	2.16×10^{-2} (2.95×10^{-2}) $[0.7318]$
$\overline{d^S}$ (se) [t−statistic]	7	-1.10×10^{-3} (1.61×10^{-2}) $[-0.0684]$	2.25×10^{-2} (3.10×10^{-2}) $[0.7234]$	6.77×10^{-3} (1.97×10^{-2}) $[0.3437]$	1.16×10^{-2} (2.91×10^{-2}) $[0.3971]$
$\overline{d^S}$ (se) [t−statistic]	10	1.07×10^{-3} (1.28×10^{-2}) $[0.0833]$	1.30×10^{-2} (2.81×10^{-2}) $[0.4642]$	2.83×10^{-3} (1.77×10^{-2}) $[0.1605]$	-1.38×10^{-1} (1.79×10^{-1}) $[-0.7683]$
$\overline{d^S}$ (se) [t−statistic]	14	5.38×10^{-3} (1.20×10^{-2}) $[0.4500]$	6.92×10^{-3} (2.45×10^{-2}) $[0.2824]$	-1.75×10^{-4} (1.48×10^{-2}) $[-0.0118]$	-1.40×10^{-2} (5.40×10^{-2}) $[-0.2599]$
$\overline{d^S}$ (se) [t−statistic]	18	5.09×10^{-3} (9.65×10^{-3}) $[0.5271]$	5.16×10^{-3} (2.11×10^{-2}) $[0.2442]$	-4.30×10^{-3} (1.32×10^{-2}) $[-0.3267]$	-2.75×10^{-2} (5.86×10^{-2}) $[-0.4692]$
$\overline{d^S}$ (se) [t−statistic]	22	2.00×10^{-3} (8.51×10^{-3}) $[0.2346]$	1.26×10^{-3} (1.77×10^{-2}) $[0.0708]$	-5.14×10^{-3} (1.08×10^{-2}) $[-0.4750]$	-4.60×10^{-2} (7.63×10^{-2}) $[-0.6025]$

Notes: The DM test statistic $\overline{d^S}$ on the log score difference and its standard error (se) for multiple step ahead density forecasts. The DM statistics were computed as documented in Table 2, using the correction factor of Harvey *et al.* (1997).