Empir Econ (2012) 43:399–426
DOI 10.1007/s00181-011-0460-5

Understanding forecast failure of ESTAR models of real exchange rates

Daniel Buncic

Received: 23 July 2009 / Accepted: 26 January 2011 / Published online: 25 March 2011

Abstract The forecast performance of the empirical ESTAR model of Taylor et al. (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 nonlinearity in the conditional means (or 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.

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

D. Buncic (✉)
University of St. Gallen, Institute of Mathematics and Statistics,
Bodanstrasse 6, 9000, St. Gallen, Switzerland
e-mail: daniel.buncic@unisg.ch

data. 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.[1]

The broad question of whether non-linear models do provide any significant forecast gains, relative to simple linear models when evaluated out-of-sample, is still a contentious topic in the academic literature. One of the earliest studies investigating the forecast performance of non-linear time series models is that of De Gooijer and Kumar (1992). Their sobering conclusion was that no clear evidence in terms of out-of-sample forecasting in favour of non-linear models could be found. Since then, the interest in forecasting economic variables with non-linear time series models and assessing their performance relative to simple linear models has grown enormously (see, for example, the studies by Ramsey (1996), Zhang et al. (1998), Lundbergh and Teräsvirta (2002), Teräsvirta et al. (2005), Teräsvirta (2006), and many others).

Nevertheless, despite the increasing interest in non-linear models for economic and financial time series in general, there does not appear to be much emphasis in the existing literature on portraying non-linear models in a way that the forecast improvements over linear models can be easily understood. This seems to be partially due to the fact that for many non-linear models closed form forecasts are only available at the one-step ahead horizon, with forecasts beyond one period ahead requiring the use of numerical techniques. It thus often appears to be the case that, when an applied forecaster compares the forecasts from linear and non-linear models with statistical tests, an intuitive feel about the importance of the non-linearity in the model is missing and it may not be obvious why a particular statistical outcome is arrived at.

The objective of this study is to learn about the forecasts from the popular ESTAR model for real exchange rates. To this end, I take the well-known empirical ESTAR model of Taylor et al. (2001) and evaluate its forecast performance relative to a simple AR(1) specification over an 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 standard tests for point as well as density forecasts, considering forecast horizons of up to 22 steps ahead. In addition to standard forecast evaluation tests, the study makes use of simulation and non-parametric techniques to visualise and to provide an intuitive picture of how the forecasts from the two competing models differ. The intention here is to use graphical techniques as much as possible to learn about the models and their forecasts. 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. In addition, this study is well-known and widely cited in the international finance literature, receiving well over 400 citations in the Google Scholar citations index.[2]

[1] One notable exception is the study by Rapach and Wohar (2006), who assess the out-of-sample performance of the Band-TAR model of Obstfeld and Taylor (1997) as well as the ESTAR model of Taylor et al. (2001) for four bilateral real exchange rates. The overall conclusion that Rapach and Wohar (2006) arrive at is that non-linear models do not offer forecast gains at short horizons, but that more accurate point forecasts at long horizons are possible for some countries.

[2] Citation statistics were accessed on July 14th, 2010.

Since it is often the case that a visual inspection of the forecasts from two competing models is far more informative to the applied forecaster than the outcome of a statistical test, the aim 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 display of the conditional means should thus be helpful to applied forecasters who want to learn about the forecasts from the models and their out-of-sample fit to the data.

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 outperforms a simple AR(1) specification at any of the 1–22 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 and regardless of whether a fixed or rolling forecasting scheme is used.

Secondly, the graphical analysis that is carried out shows that the variation of the empirical data around the one step ahead point forecasts (or conditional means) of the two competing models is substantial, making it difficult for a statistical procedure to discriminate 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 forecast gains are realised at the one step ahead horizon, where the non-linearity in the conditional mean is the strongest, there exists no potential whatsoever for the fitted ESTAR models to outperform a simple AR(1) at any of the longer forecast horizon that are considered.

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 result 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.

The remainder of the paper is organised in the following sections. Sect. 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 Sect. 3, point and density forecasts are formed, visualised, statistically tested and discussed. Sect. 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 stochastic process:

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

where the error term ϵ_t is assumed to be independently and identically distributed, with zero mean and unit variance.[3] 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.

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.[4]

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 this analysis, I 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. I also include the Swiss real exchange rate series in this analysis.[5] 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 Fig. 1, since approximately the beginning

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

[4] The data can be downloaded from http://www.mathstat.unisg.ch/buncic/data/rer_data.xls.

[5] 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.

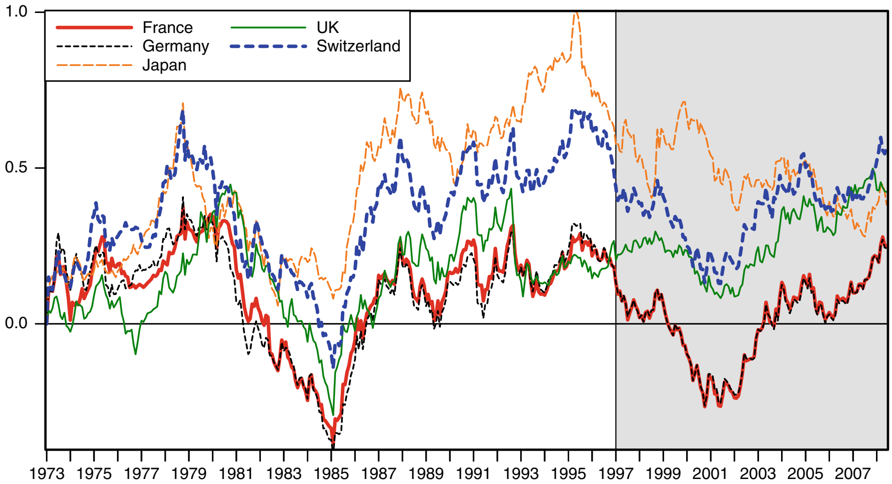


Fig. 1 Time series plot of the normalised real exchange rates over the period from January 1973 to June 2008. The non-shaded and shaded areas denote the in-sample (January 1973 – December 1996) and out-of-sample (January 1997 – June 2008) periods, respectively

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, I do not report the forecast evaluation results for the German real exchange rate series.[6]

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 robust standard errors (SE), the maximum of the log-likelihood function ($L(\gamma, \eta)$ 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 1,029 in Taylor et al.

6 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.

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

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

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 $L(\cdot)$. $\mathsf{LM}_{\mathsf{AR(1)}}$ and $\mathsf{LM}_{\mathsf{AR(1-4)}}$ 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). $\mathsf{LM}_{\mathsf{NL3}}$ is the F-statistics for a test for remaining ESTAR non-linearity (see Eitrheim and Teräsvirta 1996, page 65)

(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 that I do not provide any discussion relating to model misspecification 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 I 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), I preferred to fit the non-linear models to the data set and use these in the forecast evaluation. The IMF's IFS database is highly reliable so that the in-sample data should correlate strongly with, if not exactly match, the data set used in Taylor et al. (2001).[7]

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

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

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

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 that the results that are presented below follow the methodological approach of a "genuine" out-of-sample forecast evaluation. In the terminology of McCracken and West (2002) this is referred to as a "fixed" forecasting scheme. That is, I 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. Nevertheless, I do also provide statistical test results based on a recursive forecast scheme, using a rolling window of $T = 287$ observations. These additional results are reported and documented in more detail in the Appendix of the paper.

The reasons for why I implement a fixed forecasting scheme are as follows. Firstly, the objective of the paper is to use graphical techniques as much as possible to learn about the models, their forecasts and the data. The intention is to provide the applied forecaster with some intuition about the non-linearity in the data and about the fit of the two models to the data. It often appears to be the case that forecast comparisons involving non-linear models give a sense of a black-box mechanism, leaving the forecaster with little appreciation of why one model performs better than the other. In this context it is not possible to present the conditional means in an intuitive and easily understood way if one updates the parameters recursively at each out-of-sample data point, as well over 10 years of out-of-sample data are available, resulting in over 100 conditional means to be displayed for each model. The intuition that one gains from a visual comparison of the out-of-sample data points to the conditional means implied by the models would thus be lost.

Secondly, and perhaps more importantly, the unit-root restriction on the inner regime of the ESTAR model of Taylor et al. (2001) makes the functional form rather rigid in the sense that many observations at the extremes of q_{t-1} are needed for any noticeable changes in the shape of the ESTAR conditional mean to occur. Nevertheless, contrary to what is needed and hence to the disadvantage of the ESTAR model, the out-of-sample observations that are available are rather homogenous and cluster largely around the centre of the density of q_{t-1}. The combination of these two effects results in only marginal variation in the recursively updated parameter estimates and thus also only small changes in the implied conditional means.

In order to illustrate this point further, consider the scatter plots of all in and out-of-sample data points and the (one step ahead) conditional means shown in Fig. 2 (these plots are discussed in more detail in Sect. 3.1.1). 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 smaller (greater) than η. Notice from Panels (c) and (d) of Fig. 2, nonetheless, that for Japan and Switzerland nearly all out-of-sample data fall in the inner unit-root regime of the ESTAR model. The effect of recursively updating the parameter estimates on the shape of the conditional mean is thus most likely negligible. For the UK and France the spread of the out-of-sample data is noticeably larger than for the Japanese and Swiss real exchange rate series. But there does not seem to be any indication of an obvious ESTAR model consistent response of Δq_t to q_{t-1}, portraying instead random variation across the 0 axis. It seems, therefore, again unlikely that recursively updating the parameter estimates will lead to any important differences in the conditional means.

To gain some intuition for the above presented argumentation, I show plots of the recursively estimated γ parameter under an expanding and rolling fixed $T = 287$ window scenario in Panels (a) and (b) of Fig. 3 over the 138 out-of-sample observations from January 1997 to June 2008. Notice from the plots in Fig. 3 how small the changes in the γ parameter estimates are. Under the expanding window shown in Panel (a), all real exchange rates series except the UK one show very little variation in the γ estimate. Note that the estimate for the UK series drops somewhat towards the end of the out-of-sample period, resulting in even weaker non-linearity in the ESTAR conditional mean. Under the rolling window scheme plotted in Panel (b), a little more variation is visible for the French and UK series in the second half of the out-of-sample period, 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 do not impact on the shape of the conditional mean in any important way.

Thirdly, because I consider a test of equal mean squared errors (MSE) of two parametric models, where the first order optimality conditions are essentially moment conditions that provide consistent estimates of the model parameters, no adjustments to the (asymptotic) standard errors in the computation of the Diebold and Mariano (1995) (DM) test need to be made that would normally arise because of 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 more detailed treatment of this result).[8] Also, as 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 recursively updated parameter estimates under an

8 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).

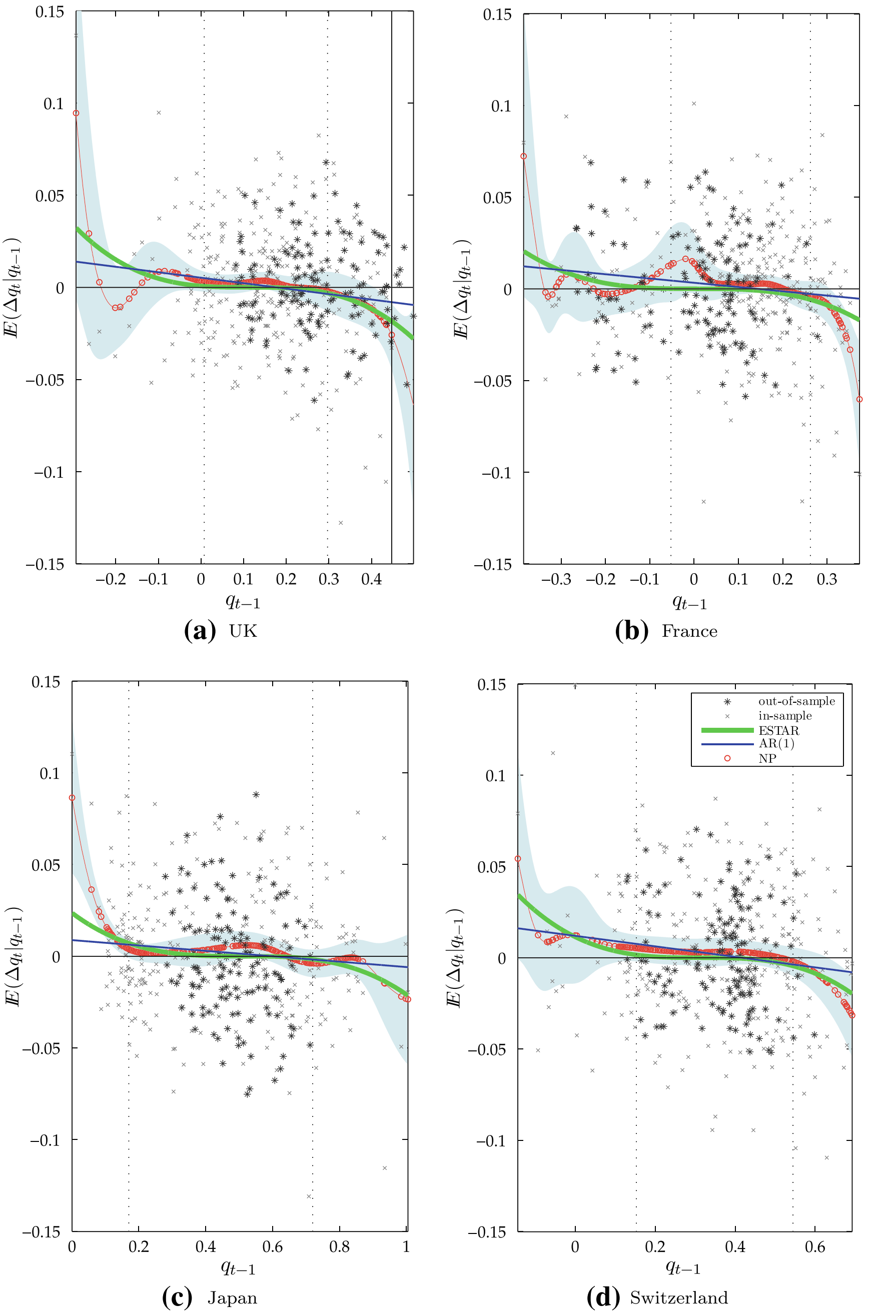


Fig. 2 One step ahead point forecasts. The *thick green* and *thin blue lines* show the one step ahead conditional forecasts of the ESTAR and AR(1) models, respectively. *Red circles* are the non-parametric conditional means, with 95% confidence intervals drawn as *blue shading*. *Grey crosses mark* the in-sample data. *Vertical dotted lines* are drawn at the 15^{th} and 85^{th} percentiles of q_{t-1}. *Black asterisks* denote the out-of-sample data

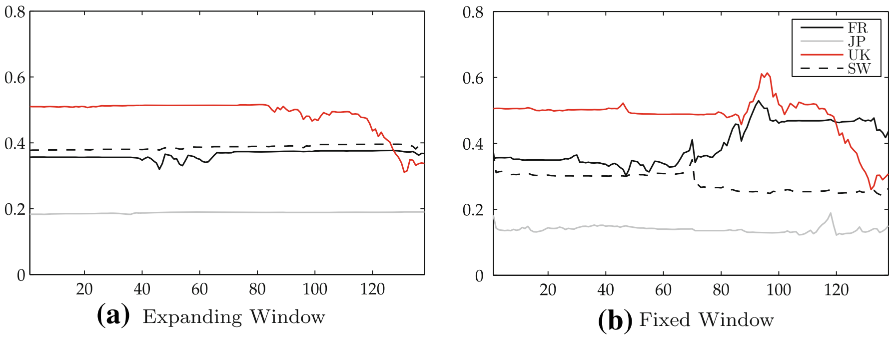


Fig. 3 Recursive estimates of the γ parameter. Panel **a** shows the estimates under an expanding window, where one extra observation from the out-of-sample data is included in each updating step. Panel **b** shows the estimates under a fixed $T = 287$ sample size when rolling through the out-of-sample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

expanding or rolling forecasting scheme can be considerable and in many situations not practical for an applied forecaster.[9]

3.1 Point forecasts

Note that under an MSE loss function, the optimal point forecast of the change in the real exchange rate series, h periods ahead, is $\mathbb{E}\left(\Delta q_{T+h}|\Omega_T\right)$, 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 $\mathbb{E}\left(\Delta q_{T+h}|\Omega_T\right)$ 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 I 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 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 this 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.

9 For example, the recursive fixed window results that are reported in the Appendix of the paper require just under 38 h computation time on a 3 Ghz quad-core processor with a parallel implementation of the forecast path simulation for each of the four series.

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. I have also superimposed the in-sample as well as the out-of-sample data by means of a scatter plot in Fig. 2, and additionally graph a non-parametric (NP) estimate of $\mathbb{E}\,(\Delta q_t | q_{t-1})$ (with 95% confidence bands) to provide a purely data driven measure of $\mathbb{E}\,(\Delta q_t | q_{t-1})$.[10] The dashed vertical lines in Fig. 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 Fig. 2 marks the bound on the in-sample data.

What can we see from Fig. 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 Fig. 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 it was 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 Fig. 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, i.e. 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).

Examining the plots of the UK and French real exchange rate series shown in Panels (a) and (b) of Fig. 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

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 outer regimes (see p. 84 in Franses and van Dijk 2000).

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 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

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

and

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

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

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

In order 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

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

Two standard statistical tests are used 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 stemming from non-linearity in the tails of q_{t-1}. I should stress 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).

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

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

DM statistic	UK	France	Switzerland	Japan
$\overline{d}$	-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]
$\omega\overline{d}$	-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]

Standard ($\overline{d}$) and weighted ($\omega\overline{d}$) Diebold and Mariano (1995, DM) test statistics. Standard errors (SE) are of the Newey and West (1987, NW) type. $\overline{d}$ was calculated as the arithmetic mean of $d_{T+1} \equiv (\varepsilon^{AR}_{T+1|T})^2 - (\varepsilon^{ESTAR}_{T+1|T})^2$ over the out-of-sample data, with $\varepsilon^{AR}_{T+1|T}$ and $\varepsilon^{ESTAR}_{T+1|T}$ 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. $\omega\overline{d}$ 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 and a 'plug in' bandwidth were used to compute the density estimate (see Silverman 1986)

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 Fig. 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 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, it can thus be concluded 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, Sect. 3.5). To implement this, I simulate a large number of pseudo realisations of q_{T+h}, $\forall h > 1$, conditional on q_T, using the following recursion

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

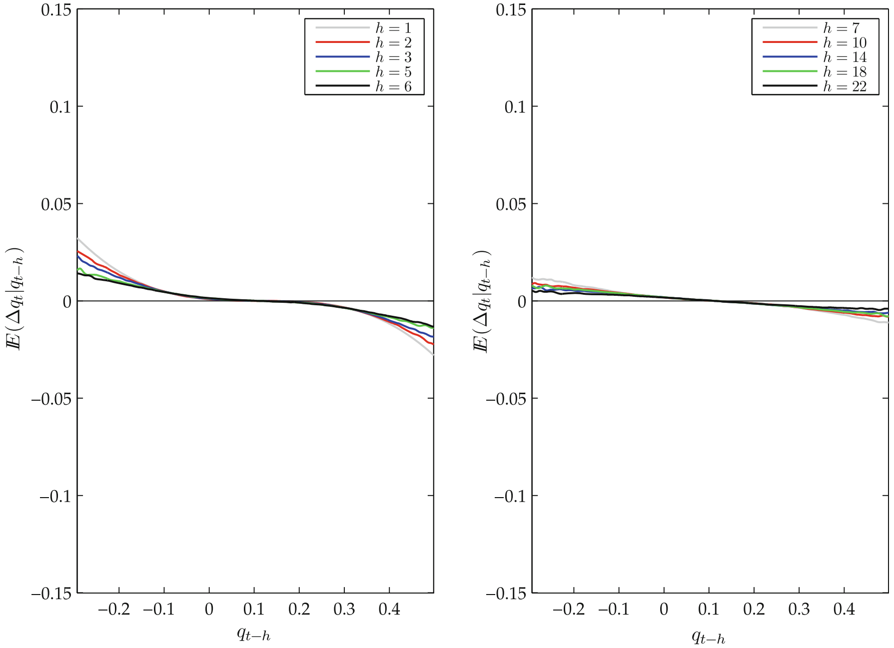
The realisation $\tilde{q}^{j}_{T+h|T}$ is thus the $j^{\text{th}} h$ step ahead pseudo value of q_{T+h}, given q_T and shock sequence $\{\tilde{\epsilon}^{j}_{T+i}\}_{i=1}^{h}$. The h step ahead point forecasts can then be approximated by computing the arithmetic mean over the J simulated $\tilde{q}^{j}_{T+h|T}$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

$$
\mathbb{E}_J(\tilde{q}_{T+h|T}) = J^{-1} \sum_{j=1}^{J} \tilde{q}^{j}_{T+h|T} \tag{7}
$$

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

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 $\mathbb{E}(\Delta q_{T+h}|q_T)$ 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 $\mathbb{E}(\Delta q_t|q_{t-h})$ 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 $\mathbb{E}(\Delta q_t|q_{t-h})$ graphically. As with the visualisation at the one step ahead horizons discussed in Sect. 3.1.1, any non-linearities in the conditional forecasts should then be identifiable from the plots of the non-parametric estimates of $\mathbb{E}(\Delta q_t|q_{t-h})$. In order 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 1,000 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 287 in-sample



(a) $h = 1, 2, 3, 5, 6.$ **(b)** $h = 7, 10, 14, 18, 22.$

Fig. 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 1,000 equally spaced grid points in the interval $[\min(q_t), \max(q_t)]$

in the interval $[\min(q_t), \max(q_t)]$ was used to compute and plot the non-parametric estimate of $\mathbb{E}(\Delta q_t|q_{t-h})$.[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, I plot $\mathbb{E}(\Delta q_t|q_{t-h})$ 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 Fig. 4, respectively. Notice from Panel (a) of Fig. 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 Fig. 4, it is evident that for forecasts of 10 steps ahead or longer (i.e., when $h \geq 10$) no visual signs of non-linearity remain to be seen.

data points and a fairly large number of draws are needed, it was preferred to generate the ϵ_t sequence parametrically from a standard normal density.

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

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 Fig. 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 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.

I 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 (6), where J was set to 10,000 and the $\tilde{\epsilon}^j_{T+h}$ were drawn from a standard normal distribution. It is evident form the results reported in Table 3 that the statistical tests 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 were considered. 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, I show plots of the 10 step ahead point forecasts for all four series in Fig. 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 fall.

[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.

[16] The contents of the plot are the same as in Fig. 2. The ESTAR conditional mean (solid green line) was computed non-parametrically from 1 million simulated draws. Fig. 5 also shows a scatter of the 10 step ahead conditional forecast constructed with the recursive scheme outlined in (6). These are superimposed onto the solid green line with black circles to show how they compare to one another.

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

DM statistic	h	UK	France	Switzerland	Japan
$\omega\bar{d}$	2	-1.74×10^{-5}	6.80×10^{-6}	-1.07×10^{-6}	-1.17×10^{-6}
(SE)		(1.25×10^{-5})	(1.35×10^{-5})	(5.35×10^{-6})	(1.37×10^{-6})
$[t - statistic]$		[−1.3991]	[0.5021]	[−0.2003]	[−0.8544]
$\omega\bar{d}$	3	-1.60×10^{-5}	3.34×10^{-6}	-1.99×10^{-6}	-1.44×10^{-6}
(SE)		(8.70×10^{-6})	(1.20×10^{-5})	(5.24×10^{-6})	(1.55×10^{-6})
$[t - statistic]$		[−1.8420]	[0.2779]	[−0.3789]	[−0.9292]
$\omega\bar{d}$	5	-7.84×10^{-6}	2.32×10^{-6}	1.18×10^{-6}	-8.85×10^{-7}
(SE)		(3.90×10^{-6})	(1.02×10^{-5})	(4.52×10^{-6})	(9.86×10^{-7})
$[t - statistic]$		[−2.0083]	[0.2288]	[0.2613]	[−0.8979]
$\omega\bar{d}$	6	-4.39×10^{-6}	9.68×10^{-7}	1.80×10^{-6}	5.89×10^{-7}
(SE)		(2.89×10^{-6})	(9.20×10^{-6})	(3.76×10^{-6})	(1.00×10^{-6})
$[t - statistic]$		[−1.5161]	[0.1051]	[0.4792]	[0.5858]
$\omega\bar{d}$	7	-3.20×10^{-6}	4.38×10^{-6}	1.39×10^{-6}	4.55×10^{-7}
(SE)		(2.30×10^{-6})	(7.78×10^{-6})	(3.73×10^{-6})	(9.53×10^{-7})
$[t - statistic]$		[−1.3896]	[0.5629]	[0.3725]	[0.4774]
$\omega\bar{d}$	10	-2.65×10^{-6}	2.59×10^{-7}	-1.86×10^{-6}	2.86×10^{-7}
(SE)		(9.40×10^{-7})	(6.63×10^{-6})	(3.04×10^{-6})	(8.02×10^{-7})
$[t - statistic]$		[−2.8219]	[0.0391]	[−0.6113]	[0.3573]
$\omega\bar{d}$	14	-1.48×10^{-6}	-6.98×10^{-7}	-2.11×10^{-6}	4.80×10^{-7}
(SE)		(9.78×10^{-7})	(4.46×10^{-6})	(2.29×10^{-6})	(7.13×10^{-7})
$[t - statistic]$		[−1.5170]	[−0.1565]	[−0.9221]	[0.6732]
$\omega\bar{d}$	18	-1.39×10^{-6}	-3.08×10^{-6}	-1.84×10^{-6}	2.81×10^{-7}
(SE)		(7.89×10^{-7})	(3.93×10^{-6})	(1.98×10^{-6})	(3.34×10^{-7})
$[t - statistic]$		[−1.7564]	[−0.7843]	[−0.9306]	[0.8413]
$\omega\bar{d}$	22	-6.10×10^{-7}	-1.82×10^{-6}	-1.16×10^{-6}	-1.09×10^{-7}
(SE)		(5.88×10^{-7})	(3.40×10^{-6})	(1.33×10^{-6})	(3.05×10^{-7})
$[t - statistic]$		[−1.0378]	[−0.5347]	[−0.8718]	[−0.3581]

Weighted version of the DM test statistic $\omega\bar{d}$ and its standard error (SE) for multiple step ahead point forecasts. The statistics were computed as documented in Table 2

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, I 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.

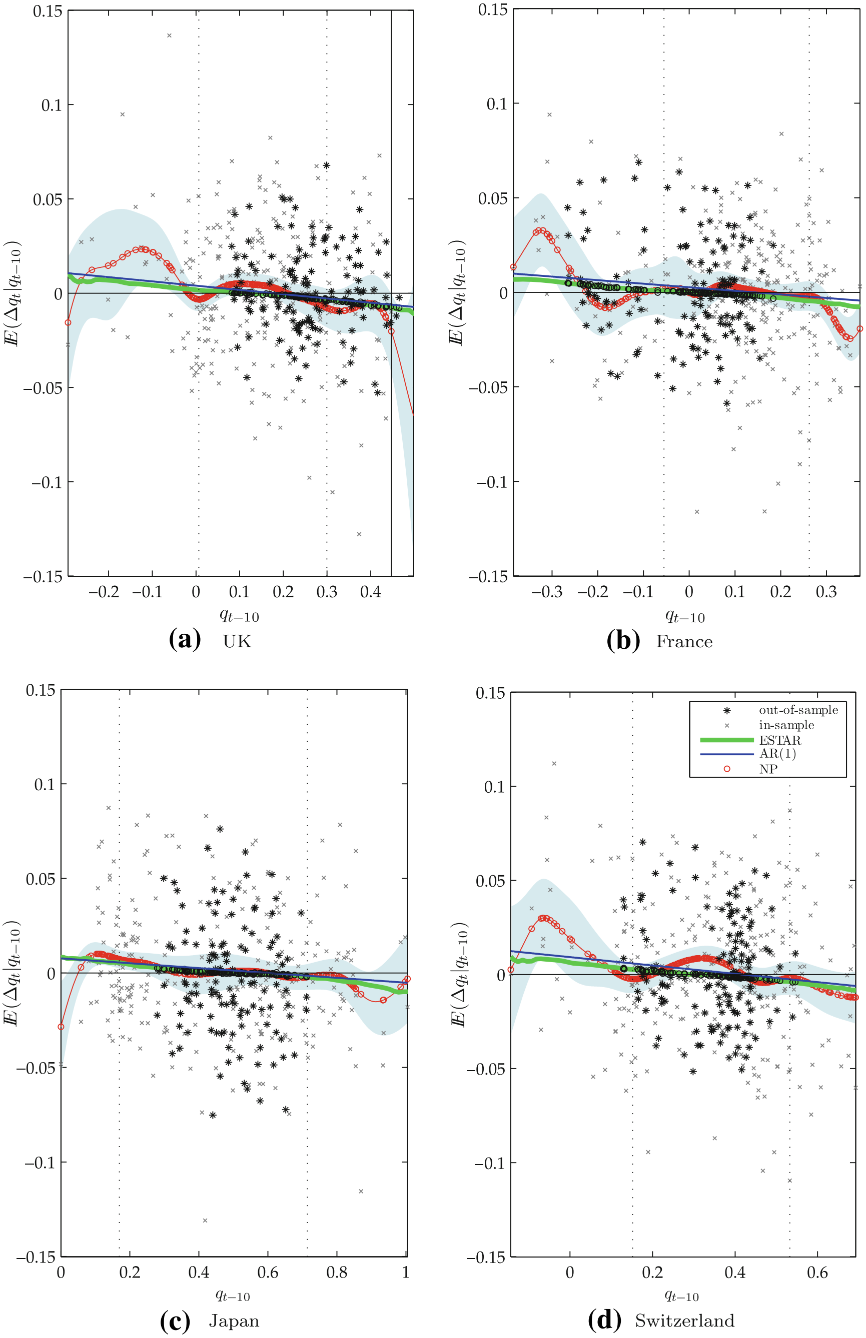


Fig. 5 10 step ahead point forecasts. The contents are the same as in Fig. 2. *Black circles* are superimposed onto the NP conditional mean (*solid green line*) to mark the 10 step-ahead conditional forecast computed from the recursive scheme outlined in (3) to facilitate the comparison to the NP conditional mean computed directly from 1 million simulated ESTAR realisations

3.2.1 Assessing one step ahead density forecasts

In the given context, i.e., 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 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 Sect. 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 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 Sect. 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; 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

[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).

defining the (log) score difference

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

and evaluating the null hypothesis of equal average scores by means of a DM type test as in Sect. 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, I 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

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

[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).

Table 4 DM test statistic for multiple step ahead density forecasts

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

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)

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 to resort to the recursive simulation scheme of (6) 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 (6) 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 Fig. 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 (6) 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 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}.

In order to illustrate how the conditional density estimate $\hat{f}^{NP}(\Delta q_t | q_{t-h})$ can be visualised, I 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 5th, 25th, 50th, 75th 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 (see Scott 1992). 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 Fig. 6. What is particularly interesting to notice from Fig. 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 Fig. 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 Fig. 6 how closely the densities overlap at the different conditioning values of q_{t-h} and how the spread in

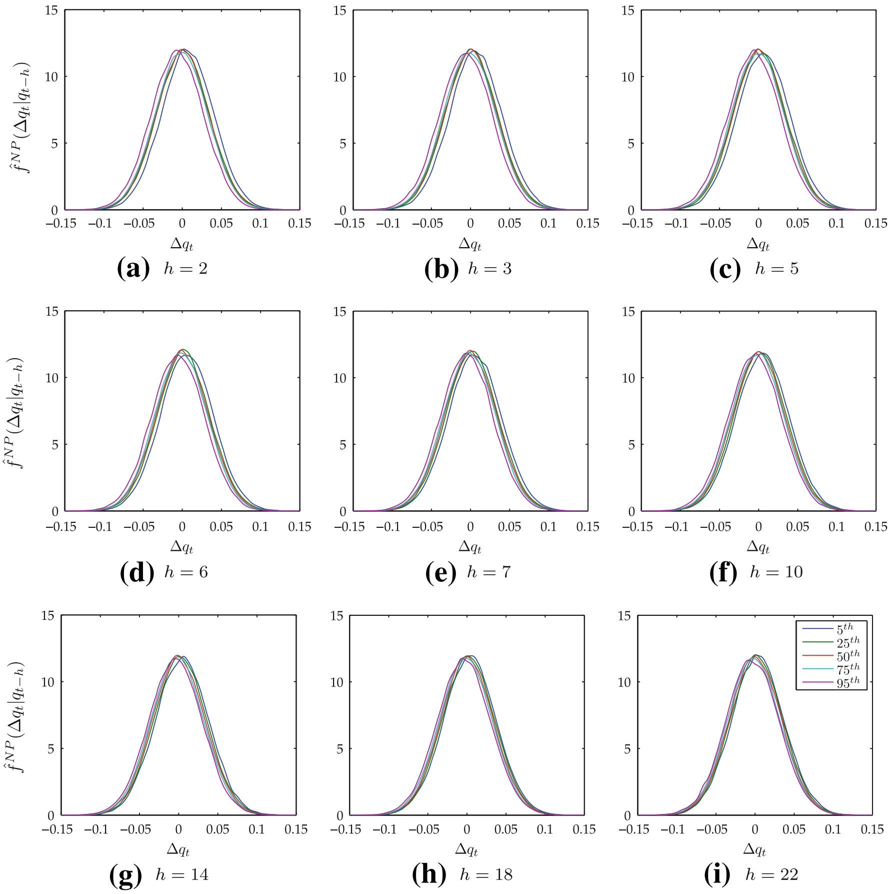


Fig. 6 Multiple step ahead density forecasts of the ESTAR model. These were constructed non-parametrically from 1 million simulated realisations of the ESTAR model in (1) at the parameter values of the UK series. Gaussian univariate and bivariate kernels were used, together with plug in bandwidths that are proportional to the covariance matrix of the data (see Scott 1992)

the densities remains observationally constant. This is indicative of a relatively weak non-linear relationship between Δq_t and q_{t-1}.

Before I proceed to provide some formal statistical evidence to support any of the 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 Fig. 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 Fig. 7 were computed from the 10,000

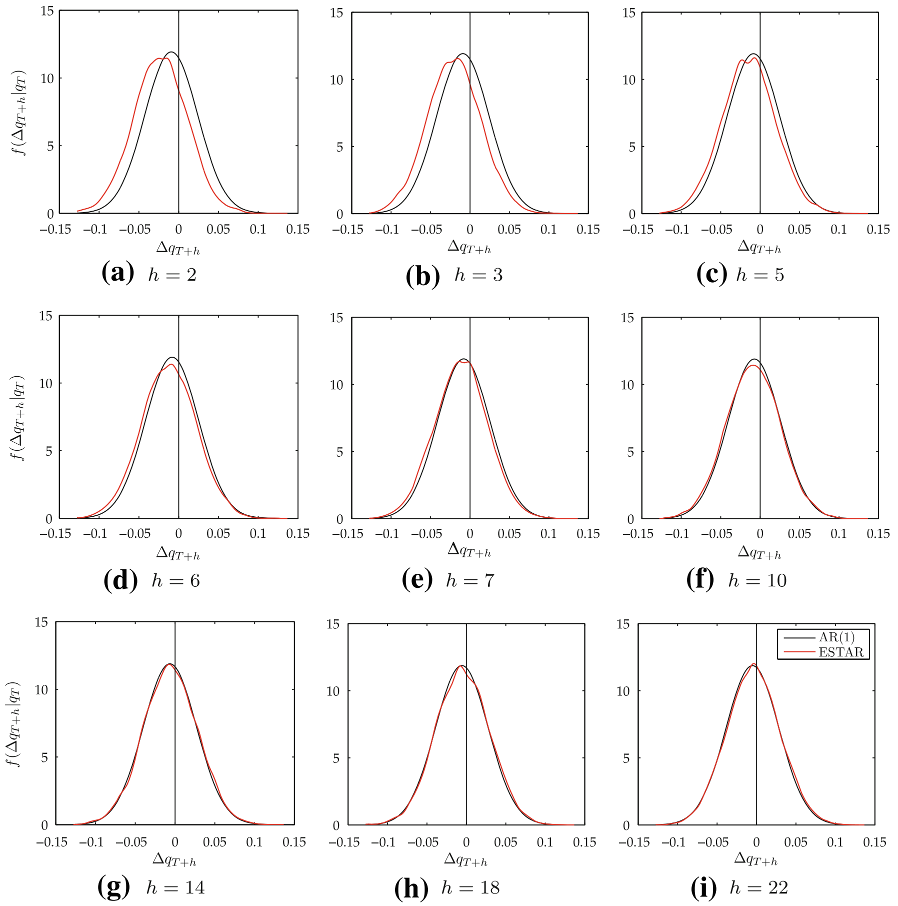


Fig. 7 Comparison of the multiple step ahead density forecasts of the AR(1) and ESTAR models for the UK real exchange rate series. The AR(1) densities were calculated from (2). The ESTAR densities were computed non-parametrically, using the 10,000 recursively constructed pseudo draws from (3). The conditioning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

pseudo observations $\{\tilde{q}^{j}_{T+h|T}\}^{J}_{j=1}$ outlined in (6), using again standard kernel density estimation methods.

The comparison of the multiple step ahead densities plotted in Fig. 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 Fig. 7. Firstly, notice that at the 2–7 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 $h \rightarrow \infty$), it is clear that the overall increase in the spread is very small, so that any differences are hard to identify visually from Fig. 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 four series of interest — at all forecast horizons that were considered — 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 1–22 steps ahead forecast horizon that were considered. This holds true for conditional mean (or point) forecasts, as well as for density forecasts, employing either a fixed or rolling window forecasting scheme. 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.

Acknowledgements I would like to thank, without any implications, Adrian Pagan, Valentyn Panchenko, Lance Fisher and Geoff Kingston for helpful comments on an earlier version of the paper. Financial support from ARC grant RM 02853 is gratefully acknowledged.

Appendix

A Forecast evaluation based on a rolling window scheme

The results presented below are based on a rolling forecasting scheme, where the sample size is fixed at $T = 287$ and the parameter estimates are updated as new data become available. That is, we get the first set of parameter estimates based on a sample from 1 to T, and proceed to forecast 1–22 steps ahead, then we get the second set of parameter estimates based on a sample running from 2 to $T + 1$, and again forecast 1–22 steps ahead and so on.

The results of the DM test that are reported below in Table A.5 make use of the asymptotics presented in Giacomini and White (2006). These tests are based on the weighted and unweighted versions of the DM test that are described in the main part of the paper in Sect. 3. Note that this is still an unconditional test of predictive ability.

The results of the rolling window forecasting scheme with recursively updated parameters presented in Table A.5 are qualitatively in line with the results of the DM tests provided in Table 2. For all four real exchange rate series, the null hypothesis of equal forecast performance cannot be rejected at any conventional significance levels, again regardless of whether a weighted or unweighted version of the DM test is used. It is again evident that the rejections are fairly strong in the sense that the $t-$statistics remain well below 1 in absolute value and are once again negative and sizable for the UK series.

Table A.5 Unweighted and weighted DM test results for one step ahead point forecasts based on a rolling fixed $T = 287$ window

DM statistic	UK	France	Switzerland	Japan
$\overline{d}$	-2.00×10^{-5}	8.01×10^{-6}	3.26×10^{-6}	3.93×10^{-7}
(SE)	(1.40×10^{-5})	(8.98×10^{-6})	(6.94×10^{-6})	(8.75×10^{-6})
$[t - statistic]$	[−1.43427]	[0.8921]	[0.4694]	[0.0449]
$\omega\overline{d}$	-1.21×10^{-5}	5.42×10^{-6}	9.75×10^{-7}	-7.81×10^{-7}
(SE)	(8.21×10^{-6})	(6.53×10^{-6})	(2.77×10^{-6})	(1.68×10^{-6})
$[t - statistic]$	[−1.4742]	[0.8324]	[0.3519]	[−0.4652]

Standard ($\overline{d}$) and weighted ($\omega\overline{d}$) Diebold and Mariano (1995, DM) test statistics. Standard errors (SE) are of the (Newey and West, 1987, NW) type. $\overline{d}$ was calculated as the arithmetic mean of $d_{T+1} \equiv (\varepsilon^{AR}_{T+1|T})^2 - (\varepsilon^{ESTAR}_{T+1|T})^2$ over the out-of-sample data, with $\varepsilon^{AR}_{T+1|T}$ and $\varepsilon^{ESTAR}_{T+1|T}$ 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. $\omega\overline{d}$ 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 and a 'plug in' bandwidth were used to compute the density estimate (see Silverman 1986)

Table A.6 Weighted DM test results for multiple step ahead point forecasts based on a rolling fixed $T = 287$ window

DM statistic	h	UK	France	Switzerland	Japan
$\omega\overline{d}$	2	-1.69×10^{-5}	5.14×10^{-6}	-9.28×10^{-7}	-1.46×10^{-6}
(SE)		(2.41×10^{-5})	(2.24×10^{-5})	(9.30×10^{-6})	(2.56×10^{-6})
$[t - statistic]$		[−0.7028]	[0.2298]	[−0.0998]	[−0.5696]
$\omega\overline{d}$	3	-1.70×10^{-5}	3.89×10^{-6}	-2.30×10^{-6}	-1.30×10^{-6}
(SE)		(1.44×10^{-5})	(1.85×10^{-5})	(8.63×10^{-6})	(2.90×10^{-6})
$[t - statistic]$		[−1.1795]	[0.2107]	[−0.2670]	[−0.4467]
$\omega\overline{d}$	5	-8.92×10^{-6}	3.22×10^{-6}	1.58×10^{-6}	-7.85×10^{-7}
(SE)		(5.92×10^{-6})	(1.69×10^{-5})	(7.25×10^{-6})	(1.69×10^{-6})
$[t - statistic]$		[−1.5078]	[0.1908]	[0.2174]	[−0.4638]
$\omega\overline{d}$	6	-4.56×10^{-6}	2.04×10^{-6}	2.40×10^{-6}	6.28×10^{-7}
(SE)		(4.25×10^{-6})	(1.63×10^{-5})	(6.59×10^{-6})	(1.52×10^{-6})
$[t - statistic]$		[−1.0718]	[0.1252]	[0.3648]	[0.4141]
$\omega\overline{d}$	7	-3.62×10^{-6}	4.23×10^{-6}	2.32×10^{-6}	3.75×10^{-7}
(SE)		(3.59×10^{-6})	(1.42×10^{-5})	(7.01×10^{-6})	(1.62×10^{-6})
$[t - statistic]$		[−1.0065]	[0.2980]	[0.3312]	[0.2309]
$\omega\overline{d}$	10	-3.09×10^{-6}	2.03×10^{-6}	-2.22×10^{-6}	3.62×10^{-7}
(SE)		(1.54×10^{-6})	(1.07×10^{-5})	(5.12×10^{-6})	(1.42×10^{-6})
$[t - statistic]$		[−2.0010]	[0.1896]	[−0.4333]	[0.2554]
$\omega\overline{d}$	14	-1.21×10^{-6}	-2.43×10^{-7}	-2.18×10^{-6}	4.22×10^{-7}
(SE)		(1.60×10^{-6})	(7.12×10^{-6})	(4.16×10^{-6})	(1.28×10^{-6})
$[t - statistic]$		[−0.7512]	[−0.0341]	[−0.5236]	[0.3297]
$\omega\overline{d}$	18	-1.33×10^{-6}	-3.00×10^{-6}	-1.52×10^{-6}	2.45×10^{-7}
(SE)		(1.45×10^{-6})	(7.11×10^{-6})	(2.84×10^{-6})	(5.83×10^{-7})
$[t - statistic]$		[−0.9213]	[−0.4210]	[−0.5337]	[0.4196]
$\omega\overline{d}$	22	-8.20×10^{-7}	-1.85×10^{-6}	-9.65×10^{-7}	-1.10×10^{-7}
(SE)		(1.04×10^{-6})	(6.67×10^{-6})	(2.32×10^{-6})	(4.54×10^{-7})
$[t - statistic]$		[−0.7883]	[−0.2780]	[−0.4157]	[−0.2431]

Weighted version of the 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 A.5

In order to avoid repetition, I again present results for multiple steps ahead forecasts from the rolling window forecasting scheme with recursively updated parameters only for the weighted version of the DM test. These test results are shown in Table A.6 below.

Looking over these results it becomes evident that as with the rolling window one step ahead out-of-sample forecast comparison, there are qualitatively no differences in the results when compared to the fixed forecasting scheme. The overall hypothesis of no forecast gains cannot be rejected in favour of the non-linear ESTAR model.

References

Amisano G, Giacomini R (2007) Comparing density forecasts via weighted likelihood ratio tests. J Bus Econ Stat 25(2):177–190

Bao Y, Lee TH, Saltoğlu B (2007) Comparing density forecast models. J Forecast 26(3):203–225

Breunig RV, Najarian S, Pagan AR (2003) Specification testing of Markov switching models. Oxford Bullet Econ Stat 65(S1):703–725
De Gooijer JG, Kumar K (1992) Some recent developments in non-linear time series modelling, testing, and forecasting* 1. Int J Forecast 8(2):135–156
Diebold FX, Mariano RS (1995) Comparing predictive accuracy. J Bus Econ Stat 13(1):253–263
Eitrheim Ø, Teräsvirta T (1996) Testing the adequacy of smooth transition autoregressive models. J Econ 74(1):59–75
Franses PH, van Dijk D (2000) Nonlinear time series models in empirical finance. Cambridge University Press, Cambridge
Gallant AR (1987) Nonlinear statistical models. John Wiley, New York
Giacomini R, White H (2006) Tests of conditional predictive ability. Econometrica 74(6):1545–1578
Granger CWJ, Teräsvirta T (1993) Modelling nonlinear economic relationships. Oxford University Press, Oxford
Harvey DI, Leybourne SJ, Newbold P (1997) Testing the equality of prediction mean squared errors. Int J Forecast 13(2):281–291
Lundbergh S, Teräsvirta T (2002) Forecasting with smooth transition autoregressive models. In: Clements MP, Hendry DF (eds) A Companion to economic forecasting. Blackwell Publishers, Oxford pp 485–509
McCracken MW, West KD (2002) Inference about predictive ability. In: Clements MP, Hendry DF (eds) A companion to economic forecasting. Blackwell Publishers, Oxford pp 299–321
Mitchell J, Hall SG (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 Bull Econ Stat 67(S1):995–1033
Newey WK, West KD (1987) A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55(3):703–708
Obstfeld M, Taylor AM (1997) Nonlinear aspects of goods-market arbitrage and adjustment: Hecksher's commodity points revisited. J Japanese Int Econ 11(4):441–479
Pagan AR (2002) Learning about models and their fit to data. Int Econ J 16(2):1–18
Pagan AR, Ullah A (1999) Nonparametric Econometrics. Cambridge University Press, New York
Ramsey JB (1996) If nonlinear models cannot forecast, what use are they. Studies Nonlin Dynamic Econ 1(2):65–86
Rapach DE, Wohar ME (2006) The out-of-sample forecasting performance of nonlinear models of real exchange rate behavior. Int J Forecast 22(2):341–361
Scott DW (1992) Multivariate density estimation: theory, practice, and visualization. John Wiley and Sons, New York, Chichester
Silverman BW (1986) Density estimation for statistics and data analysis. Chapman and Hall, London
Taylor MP, Peel DA, Sarno L (2001) Nonlinear mean-reversion in real exchange rates: Towards a solution to the purchasing power parity puzzles. Int Econ Rev 42(4):1015–1042
Teräsvirta T (1994) Specification, estimation, and evaluation of smooth transition autoregressive models. J Am Stat Assoc 89(425):208–218
Teräsvirta T (2006) Forecasting economic variables with nonlinear models. In: Elliott G, Granger C, Timmermann A (eds) Handbook of economic forecasting, vol 1. Elsevier, Amsterdam, pp 413–457
Teräsvirta T, van Dijk D, Medeiros MC (2005) Smooth transition autoregressions, neural networks, and linear models in forecasting macroeconomic time series: A re-examination. Int J Forecast 21(4): 755–774
van Dijk D, Franses PH (2003) Selecting a nonlinear time series model using weighted tests of equal forecast accuracy. Oxford Bull Econ Stat 18(S1):727–744
Zhang G, Patuwo BE, Hu MY (1998) Forecasting with artificial neural networks: the state of the art. Int J Forecast 14(1):35–62