A note on long horizon forecasts of nonlinear models of real exchange rates: Comments on Rapach and Wohar (2006)

Buncic, Daniel

School of Economics, University of New South Wales

24 January 2008

A NOTE ON LONG HORIZON FORECASTS OF NONLINEAR MODELS OF REAL EXCHANGE RATES: COMMENTS ON RAPACH AND WOHAR (2006)[‡]

DANIEL BUNCIC[§]

School of Economics
University of New South Wales
Sydney, Australia

First version: December 19, 2007
This version 1.24: January 24, 2008

Abstract

We show that long horizon forecasts from the nonlinear models that are considered in the study by Rapach and Wohar (2006) cannot generate any forecast gains over a simple AR(1) specification. This is contrary to the findings reported in Rapach and Wohar (2006). Moreover, we illustrate graphically that the nonlinearity in the forecasts from the ESTAR model is the strongest when forecasting one step-ahead and that it diminishes as the forecast horizon increases. There exists, therefore, no potential whatsoever for the considered nonlinear models to outperform linear ones when forecasting far ahead. We also illustrate graphically why one step-ahead forecasts from the nonlinear ESTAR model fail to yield superior predictions to a simple AR(1).

Keywords: PPP, regime modelling, nonlinear real exchange rate models, ESTAR, forecast evaluation.
JEL Classification: C22, C52, C53, F31, F47.

[‡]I would like to thank Lance Fisher, Geoff Kingston, Adrian Pagan, Dick van Dijk and Minxian Yang for helpful comments on an earlier version of the paper. Financial support from ARC grant RM 02853 is gratefully acknowledged. All remaining errors are mine.

[§]Contact details: School of Economics, Australian School of Business, University of New South Wales, Sydney 2052, Australia. Tel: +61(2) 938 59732. Fax: +61(2) 938 56337. email: d.buncic@unsw.edu.au. web: www.dbuncic.googlepages.com.

1. Introduction

Rapach and Wohar (2006) examine the out-of-sample forecast performance of two widely used nonlinear statistical models for real exchange rates. The models examined are the Band-TAR model of Obstfeld and Taylor (1997, OT) and the ESTAR model of Taylor, Peel and Sarno (2001, TPS). Their study is important, as it is one of the first to provide a comprehensive assessment of the out-of-sample gains from using nonlinear models to forecast real exchange rates. However, some of the conclusions that are drawn in the paper seem rather surprising and counter intuitive. For example, Rapach and Wohar (2006, pp. $350-352$) conclude that: "*Overall, there is robust evidence in Tables 2 and 3 that the OT Band-TAR and TPS 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.*" Since the nonlinear models utilised in the study are stable and globally covariance stationary, one would expect, *a priori*, long-horizon forecasts from such models to converge to their unconditional means, so that no gains from using nonlinear models, relative to simple linear ones, should be realised when forecasting far ahead. Their conclusion seems even more surprising given that "*[t]here is almost no robust evidence that nonlinear models offer forecasting gains at short horizons for any country.*" (Rapach and Wohar, 2006, p. 352).[1]

Rapach and Wohar also employ graphical methods to visualise the nonlinearity in the conditional means of the fitted models and to understand why they fail to provide superior forecasts at short horizons. Such an approach has recently been advocated by Pagan (2002) and Breunig, Najarian and Pagan (2003). For that purpose, implied conditional means of the competing models are plotted. Surprisingly, though, these plots are drawn in (q_{t-1}, q_t) space (see Figures 2 and 3 in their paper), making it extremely difficult to visually identify any differences between the conditional means of the models. In that respect, the graphical methods are not utilised in the best possible way, leaving the reader to ponder about how much the conditional means of the nonlinear models differ from linear ones. Moreover, there is no motivation provided to support the counter intuitive claim that forecast gains exist at longer horizons, given that no such gains are realised over short horizons, although the competing models that are employed in the study are simple enough to warrant a visual exposition to support this claim.

In the current context, graphical methods, in conjunction with simulation techniques, are particularly useful to show that the nonlinearity in the conditional means of the OT and TPS models vanishes the farther ahead the forecast, so that no potential exists whatsoever to outperform linear models at longer forecast horizons. At the one step ahead horizon, where the nonlinearity in the conditional mean is the strongest and where, thus, the greatest advantage over a linear forecast is realisable, there are two contributing factors that lead to forecast failure. The first one relates to the spread of the out-of-sample observations around the conditional means, yielding a relatively large variance when comput-

[1] Rapach and Wohar also use interval and density forecasts to evaluate one to three step ahead forecasts. However, no conclusive results are reached regarding the superiority of any models forecasts (see the discussion on pp. $352-356$).

ing the Diebold and Mariano (1995, DM) statistic and hence small t-ratios. This is evident in all four empirical real exchange rate series that are used. The second factor relates to the location of the out-of-sample data. For two out of the four series, all out-of-sample observations cluster around an area that is often labelled as the inner regime, where the conditional means of the linear and nonlinear models are extremely difficult to distinguish. Thus, not only is the denominator in the computation of the DM t-ratio large, the numerator is also very small, resulting in a minuscule test statistic. These are important results that can easily be illustrated by plotting the conditional means in $(q_{t-1}, \Delta q_t)$ space and superimposing the out-of-sample data.

Band-TAR and ESTAR type models have become extremely popular in the empirical literature in recent years and are widely used among researchers. For example, the search for 'ESTAR' and 'exchange rate' in the Google Scholar search engine yields around 6220 hits. The OT and TPS models in particular have been cited 265 and 198 times respectively in the Google Scholar Citation Index (GSCI). As a comparison, Hall White's seminal paper on data snooping published in 2000 has received 285 citations in the GSCI. [2] Additionally, from a practitioners perspective it is often of interest to see how well nonlinear models perform out-of-sample, before a decision regarding the implementation of such models is reached. It is, therefore, important to provide a careful assessment of the relative forecast performance of these models.

The purpose of this study is twofold. Firstly, it illustrates graphically why no gains exist when the ESTAR model is used to forecast one step ahead. Secondly, and more importantly, it is shown that there exists no potential for the ESTAR model to generate any gains when forecasting far ahead, as the nonlinearity in the conditional mean dissipates the farther ahead the forecast. The nonlinearity is the strongest at the one step ahead horizon. Throughout the paper, heavy emphasis is placed on a graphical exposition of these findings. Formal statistical tests are also provided to verify the graphical results, however, without any discussion. All computations presented in this study employ the same data set that is used by Rapach and Wohar (2006), which is publicly available from David Rapach's website: http://pages.slu.edu/faculty/rapachde/Research.htm. [3]

The remainder of the paper is structured as follows. Section 2 briefly outlines the competing models that are employed in the forecast evaluation exercise together with a comparison of the conditional means that each model generates. Section 3 illustrates why forecast failure is encountered at the one step ahead forecast horizon, and more importantly, why there exists no potential whatsoever for the nonlinear models that Rapach and Wohar consider to generate any gains when forecasting far ahead. This section illustrates graphically that the nonlinearity in the forecasts from the ESTAR model of TPS is the strongest when forecasting one step ahead and that it diminish as the forecast horizon increases. For horizons greater than 10 steps ahead, no visual signs of nonlinearity in the conditional means are identifiable. The findings of this study are summarised in Section 4.

[2] Citations statistics were accessed on December 19, 2007.

[3] The data can also be obtained in Excel format from http://www.dbuncic.googlepages.com/research.

2. Preliminary Discussion

2.1. Modelling real exchange rates

Rapach and Wohar (2006) study the out-of-sample forecast performance of the empirical real exchange rate series of the United Kingdom (UK), Germany, France and Japan using two nonlinear statistical models. These are the Band-TAR model of Obstfeld and Taylor (1997) and the ESTAR model of Taylor *et al.* (2001) (see equations (1) and (2) in their paper). The Band-TAR model has the following specification:

$$\Delta q_t = \begin{cases} \sigma_{in}\epsilon_t & \text{if } |q_{t-1}| \leq c \quad \text{(inner Regime)} \\ \lambda \, \mathsf{sign}\,(q_{t-1})\,(|q_{t-1}| - c) + \sigma_{out}\epsilon_t & \text{if } |q_{t-1}| > c \quad \text{(outer Regime)} \end{cases} \tag{1}$$

where q_t is the log of the real exchange rate,[4] $\epsilon_t \sim \mathsf{N}\,(0,1)$ and sign is the signum function.[5] Under this specification, the real exchange rate follows a random walk in the regime inner, ie., when $|q_{t-1}| \leq c$, and an equilibrium correcting mechanism in the outer regime (when $|q_{t-1}| > c$), with λ being the speed of adjustment parameter.

In the ESTAR model, the real exchange rate evolves according to:

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

where q_t and ϵ_t are defined as above in (1), η is the long-run equilibrium level of q_t, and $\Phi\,(\alpha, \eta; q_{t-1})$ is the standard exponential regime weighting function which is bounded between zero and unity. Notice here that the ESTAR model was rewritten in Δq_t form to make it consistent with the formulation of the Band-TAR model in (1). As in (1), the real exchange rate has two regimes in the ESTAR model, nevertheless, with the movement between inner and outer regimes being smooth rather than discrete.

The benchmark model in the forecast evaluation of Rapach and Wohar is a simple AR(1) specification, parameterised in the standard way as:

$$\Delta q_t = \delta\,(q_{t-1} - \mu) + \sigma\epsilon_t \tag{3}$$

where q_t and ϵ_t are again defined as above in (1). Note that δ and μ are also, respectively, speed of adjustment parameters and the long-run equilibrium level of q_t, but are not necessarily the same as in the nonlinear formulations given in (1) and (2).

[4] which is defined as, $q_t \equiv s_t + (p_t^* - p_t)$, where s_t is the log of the nominal exchange rate and p_t (p_t^*) are the logs of domestic (foreign) CPIs. See footnote 7 on page 343 in Rapach and Wohar (2006) for more details on the data.

[5] The signum function is defined in the standard way as $\mathsf{sign}\,(x)$ equal to 1 if $x > 0$, 0 if $x = 0$, and -1 if $x < 0$.

2.2. How do the conditional means compare?

It is evident from the specifications given in (1) and (2) that the nonlinearity in the OT and TPS models concerns the conditional mean of the real exchange rate. Rapach and Wohar show plots of the conditional means of q_t, given q_{t-1}, in Figures 2 and 3 in their paper to provide *"a visual feel for how 'close' the fitted linear and nonlinear AR models are in terms of their conditional means"* (Rapach and Wohar, 2006, p. 357). That is, they show plots of $E(q_t|q_{t-1})$, where the expectation is taken with respect to the considered models. Nevertheless, they draw these plots in (q_{t-1}, q_t) space, making it thereby extremely difficult to visually distinguish the models from one another.

A considerably more informative way to present the conditional means of these models is in $(q_{t-1}, \Delta q_t)$ space. In Figure 1 a comparison of the implied conditional means of the OT, TPS and AR(1) models is shown under the two different plot scenarios.[6] Panel (b) of Figure 1 plots the original formulation used in Rapach and Wohar (2006). Notice from this panel how difficult it is to identify any differences between the conditional means of the models. In $(q_{t-1}, \Delta q_t)$ space, as plotted in panel (a) of Figure 1, it is much easier to appreciate how the models differ. One can see that the nonlinear models, as intended, contain a fairly wide non adjustment region, within which q_t follows a random walk process. Outside this non adjustment region the conditional mean of the models changes to allow for a stronger adjustment, relative to the linear specification. The conditional mean of the AR(1), on the other hand, remains linear over all states of the conditioning variable and simply 'slices' through the two nonlinear models.

Why is this illustration useful? Recall that a k step ahead point forecast is formed as $E(\Delta q_{t+k}|\mathscr{F}_t)$, where $\mathscr{F}_t$ is a conditioning set containing all available information to the forecasting agent at time t. Given the simple structure of the models in (1) to (3), once the unknown parameters have been estimated, all that is needed to form the forecast is q_t; that is, one simply evaluates $E(\Delta q_{t+k}|q_t)$. A one step ahead forecast for the ESTAR model, for example, can be computed analytically as $E(\Delta q_{T+1}|q_T) = -(q_T - \eta)\,\Phi\,(\alpha, \eta; q_T)$, where T is the sample size over which the model was estimated, so that q_T is the last in sample observation. Notice then that the one step ahead forecast $E(\Delta q_{T+1}|q_T)$, or alternatively $E(\Delta q_t|q_{t-1})$, is nothing else but the conditional mean, evaluated at q_T. As shown in panel (a) of Figure 1, this quantity can easily and informatively be visualised in $(q_{t-1}, \Delta q_t)$ space. More importantly, though, the out-of-sample data can be superimposed onto a two dimensional plot of $E(\Delta q_t|q_{t-1})$ to graphically assess each model's relative performance.

A k step ahead forecast, denoted by $E(\Delta q_{T+k}|q_T)$, can be formed and utilised in a similar way. Although this quantity may not be available in closed form for the nonlinear models that are considered here, it can still be computed via simulation. The important point is, nevertheless, that we can plot $E(\Delta q_t|q_{t-k})$ in a two dimensional space once the conditional mean has been formed. Any evidence of nonlinearity in $E(\Delta q_t|q_{t-k})$, and its

[6] The model parameters are those of Germany, which are provided in Table 1. An AR(1) was estimated on the German real exchange rate series, yielding estimates of -0.0207 and -0.1448 for δ and μ, respectively.

importance in the forecasts, can readily be examined from the plots. Representing and assessing forecasts in this way is appealing, because a visual representation is likely to provide a much greater insight into why one model dominates over another one than the outcome of a statistical test.

3. Evaluating forecasts

The forecast results that are reported here focus solely on the TPS model. The reasons for this are as follows. Firstly, it is evident from panel (a) of Figure 1 that the OT and TPS models are very similar in terms of their conditional means and hence their forecasts. Presenting the results for both models is thus repetitive. Secondly, the TPS appears to be perceived as the more elegant model in the literature due to the transition between the inner and outer regimes being modelled by a smooth function.

3.1. Why one step ahead forecasts fail

Figure 2 shows plots of the conditional mean $E(\Delta q_t|q_{t-1})$ under the nonlinear ESTAR specification of TPS as well as under an AR(1) model, together with a scatter plot of the out-of-sample data for the four empirical series that are considered. The plots in Figure 2 also show a non-parametric (NP) estimate of $E(\Delta q_t|q_{t-1})$ together with approximate 95% confidence intervals, the in-sample data, as well as dashed vertical lines at the 15^{th} and 85^{th} percentiles of q_{t-1}.[7] The reason why an NP estimate of $E(\Delta q_t|q_{t-1})$ is included in the plots is to provide a purely data based measure of the conditional mean in order to show what the parametric models are trying to fit. The parameter values that were used to plot the conditional means of the TPS model are provided in the lower half of Table 1.

What can be seen from Figure 2? The plots displayed in panels (a) and (d) for the empirical real exchange rate series of the UK and Japan are particularly interesting, because over the entire out-of-sample period that is considered in the study by Rapach and Wohar, not a single observation exists that falls into the region that would be classified as an outer regime. In fact, all out-of-sample observations cluster around an area where the forecasts coming from the nonlinear TPS model and the linear AR(1) are difficult to distinguish. It is also easy to see from these plots why the weighted version of the DM test statistic that Rapach and Wohar (2006) use,[8] which is designed to give a larger weight to observations falling into the tails of the distribution of q_{t-1}, is unlikely to provide any advantage over the unweighted version of the test and why forecast failure is consequently encountered.

For Germany and France, as shown in panels (b) and (c) of Figure 2, the out-of-sample data points show a somewhat wider dispersion, with approximately half of them falling below the 15^{th} percentile of q_{t-1}. Nevertheless, it is evident that not a single observation falls outside the in-sample data range, or close to the extreme ends of the density of q_{t-1},

[7] A local linear regression estimator was used to compute the NP conditional mean (see Pagan and Ullah, 1999, p. 104 for details).

[8] See p. 347 in Rapach and Wohar (2006) for more details on the construction of this test.

where a weight of unity would be received in the computation of the weighted DM statistic. Notice also that the spread of the out-of-sample data points across the conditional means of the two models is substantial. This makes forecast evaluation more difficult, as not only is the difference between the conditional means fairly small over the range of the out-of-sample observations, but the variance of these data points around the conditional means is also considerable, resulting in small t-ratios.

As a final comment, notice also from Figure 2 that over the entire data span — including the in-and out-of-sample observations — the spread around the conditional means is fairly large. It is not clear, therefore, whether a strategy to increase the out-of-sample evaluation period to obtain a better measure of the DM statistic has any potential of being successful. A 'too short' evaluation period is often thought to be one of the contributing reasons why the DM test fails. For completeness and without discussion, the weighted and unweighted DM test results are provided in Table 2.

3.2. Why long horizon forecasts cannot succeed

It is again most informative to resort to a graphical exposition to illustrate why long horizon forecasts from the nonlinear TPS model do not have the potential to generate superior forecasts over a simple linear AR(1). To do so, it is, nevertheless, necessary to use simulation techniques to compute forecasts beyond one step ahead from the TPS model. The approach that is employed by Rapach and Wohar (2006, see p. 346) is fairly standard.[9] They simulate a large number of pseudo realisations of q_{t+k}, $\forall\ k = 1, ..., 24$, given q_t, using the following recursion:

$$\begin{aligned} q^*_{t+1} &= q_t - (q_t - \eta)\, \Phi\left(\alpha, \eta; q_t\right) + \sigma\epsilon^*_{t+1} \\ q^*_{t+2} &= q^*_{t+1} - \left(q^*_{t+1} - \eta\right) \Phi\left(\alpha, \eta; q^*_{t+1}\right) + \sigma\epsilon^*_{t+2} \\ &\vdots \\ q^*_{t+k} &= q^*_{t+k-1} - \left(q^*_{t+k-1} - \eta\right) \Phi\left(\alpha, \eta; q^*_{t+k-1}\right) + \sigma\epsilon^*_{t+k} \end{aligned} \tag{4}$$

where q^*_{t+k} is the k step ahead pseudo realisation of q_{t+k} given q_t and random draws $\left\{\epsilon_{t+j}\right\}_{j=1}^{k}$. The k step ahead conditional forecasts are then constructed by simply taking the arithmetic mean over the simulated pseudo realisations of q^*_{t+k}.

An alternative approach that can be used to obtain the forecasts is to simulate a large number of realisations of q_t from the ESTAR model in (2) and then compute the conditional mean directly using non-parametric methods. The benefit of this approach is that it allows us to evaluate $E(\Delta q_t|q_{t-k})$ over an arbitrary range of values of q_{t-k}. This way one can evaluate forecasts at a sufficient number of points over a given interval, making it possible to draw a line and examine $E(\Delta q_t|q_{t-k})$ visually. Any nonlinearities in the conditional forecasts should then be identifiable from the plots of the NP estimates of

[9] See section 3.5 in Franses and van Dijk (2000) for a textbook style treatment.

$E(\Delta q_t|q_{t-k})$. With the recursive scheme of (4) one generally only evaluates the forecasts at a particular set of conditioning values, which are typically the out-of-sample data points. Although these points could also be used in a scatter plot to visualise the shape of the forecasts, this strategy is rarely employed in the literature.

To illustrate how this approach can be put into practice, 4 million observations of q_t are simulated from the parameter estimates of the UK series in Table 1. A non-parametric estimate of $E(\Delta q_t|q_{t-k})$ over 100 equally spaced points in the interval $[\min(q_t), \max(q_t)]$ is computed and plotted in Figure 3. Short horizons, with $k = [1, 2, 3, 5]$, are displayed in panel (a). Panel (b) shows conditional means corresponding to longer horizons, ie., when $k = [10, 15, 20, 25]$. It is evident from panel (a) of Figure 3 that the nonlinearity in the forecasts is strongest at the one step ahead horizon, that is, when $k = 1$. 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 3 one can see that for forecasts of 10 steps ahead or longer (ie., when $k \geq 10$) no signs of nonlinearity remain. One can also see in panel (b) that as the forecast horizon increases, $E(\Delta q_t|q_{t-k})$ tends towards 0 for all values of the conditioning variable q_{t-k}, which is in line with our prior expectations.

How do long horizon forecasts from the nonlinear model compare to linear AR(1) forecast? Figure 4 shows 24 step ahead forecasts for all four empirical real exchange rate series that are considered by Rapach and Wohar. 24 step ahead forecasts were chosen here as they show the highest level of statistical significance, together with the smallest relative ESTAR/AR(1) root mean squared forecast error (RMSFE) (see the entries for France and Germany in Table 3 in Rapach and Wohar, 2006). Note that the structure of the plots that are shown in Figure 4 have the same format as the ones that were given earlier. However, since the conditional mean of the ESTAR model had to be simulated, we also computed the 24 step ahead forecasts using the recursive scheme of (4). These were evaluated at the out-of-sample data points, using $10,000$ draws over which the arithmetic mean was taken. These are marked by black circles in Figure 4.

Observe initially from Figure 4 how closely the forecasts generated from the recursive scheme of (4) overlap with the NP estimate of $E(\Delta q_t|q_{t-24})$ on the simulated data. This is shown here to provide confidence that the two approaches generate equivalent results. Notice also from these plots that, as expected, there exist no visual signs of nonlinearity in the conditional means, and therefore the forecasts. In fact, the conditional means of the nonlinear ESTAR model overlap very closely with those generated from the linear AR(1). Forecast gains, therefore, cannot be realised. For completeness, and again without discussion, formal statistical test results are provided in Table 2. Only the unweighted version of the DM test is reported due to the lack of nonlinearity in the TPS forecasts.

It should be mentioned here that an attempt was made to understand how the results reported in Rapach and Wohar were arrived at. For that purpose, we obtained the code that is provided on David Rapach's website to compute multi step ahead forecasts with the file `Tps_frap.prg` for the French Franc real exchange rate series. The surprising result that came out of this exercise was that the RMSFE decreased as we increased the horizon of the forecast. For example, setting `hmax` (the forecast horizon) on line 396 to $30, 40, 50, 60$, and 70 months resulted in RMSFE of $0.83, 0.82, 0.80, 0.77$, and 0.67, respectively.[10] For any stable and covariance stationary time series model, this ratio should tend to unity. It seems, therefore, that there is an error in the code that was utilised by Rapach and Wohar (2006).

4. Conclusion

This note has reevaluated the out-of-sample forecast performance of the ESTAR model as recently examined in Rapach and Wohar (2006). Contrary to the findings reported in Rapach and Wohar (2006) that nonlinear models offer forecast gains over a simple linear AR(1) specification when forecasting far ahead, we have shown that no such gains exist. Moreover, we have illustrated with graphical methods that the nonlinearity in the forecasts of the ESTAR model dissipates as the forecast horizon increases. The nonlinearity in the forecasts is strongest when forecasting one step-ahead. Forecasts of 10 steps ahead or longer fail to show any visible signs of nonlinearity so that no potential exists whatsoever for the ESTAR model to outperform a simple AR(1) model when forecasting far ahead.

Another interesting result that was presented in this note relates to the one step-ahead forecasts. Although it was shown that the potential to generate superior forecasts is the largest at the one step-ahead horizon as the nonlinearity is the strongest here, forecast failure still prevails due to the following two reasons. Firstly, for two out of the four empirical real exchange rate series, all out-of-sample observations cluster around an area where the forecasts from the two competing models are extremely hard to discriminate. Secondly, in all four series the are considered the spread of the out-of-sample data around the conditional means is relatively large, resulting in a substantial variance and hence small t-ratios when computing the DM test statistic. Thus, not only is the difference between the forecasts of the competing models over the range of the out-of-sample data minuscule, but the precision available to conduct statistical tests is also low. These are important empirical findings that needed to be highlight.

[10] To gain speed, the section that computes the modified DM statistic from line 473 onwards until the start of procedures was commented out. Also, the following commands were added to print the result to the screen:

```
gnon_24=g_non[.,hmax];glin_24=g_lin[.,hmax];
"RMSE ESTAR/AR: "sqrt(meanc(gnon_24))/sqrt(meanc(glin_24));?;
```

References

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.

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

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

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.

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

White, Halbert (2000): "A Reality Check for Data Snooping", *Econometrica*, **68**(5), 1097–1126.

Tables and Figures

Table 1: Band-TAR and ESTAR Model Parameters.

Band-TAR	United Kingdom	Germany	France	Japan
λ	-0.084469	-0.051648	-0.047584	-0.057013
c	0.163408	0.111319	0.107238	0.254211
σ_{out}	0.044440	0.037765	0.037590	0.029530
σ_{in}	0.032755	0.031245	0.030720	0.035618
ESTAR	**United Kingdom**	**Germany**	**France**	**Japan**
α	-0.449111	-0.263792	-0.289013	-0.164703
η	0.149670	-0.006640	0.048919	0.514759
σ	0.033375	0.034507	0.033061	0.033393

Notes: Band-TAR and ESTAR model parameter estimates as contained in the GAUSS files available on David Rapach's website: http://pages.slu.edu/faculty/rapachde/Research.htm.

Table 2: Unweighted and weighted DM test results for one step-ahead forecast.

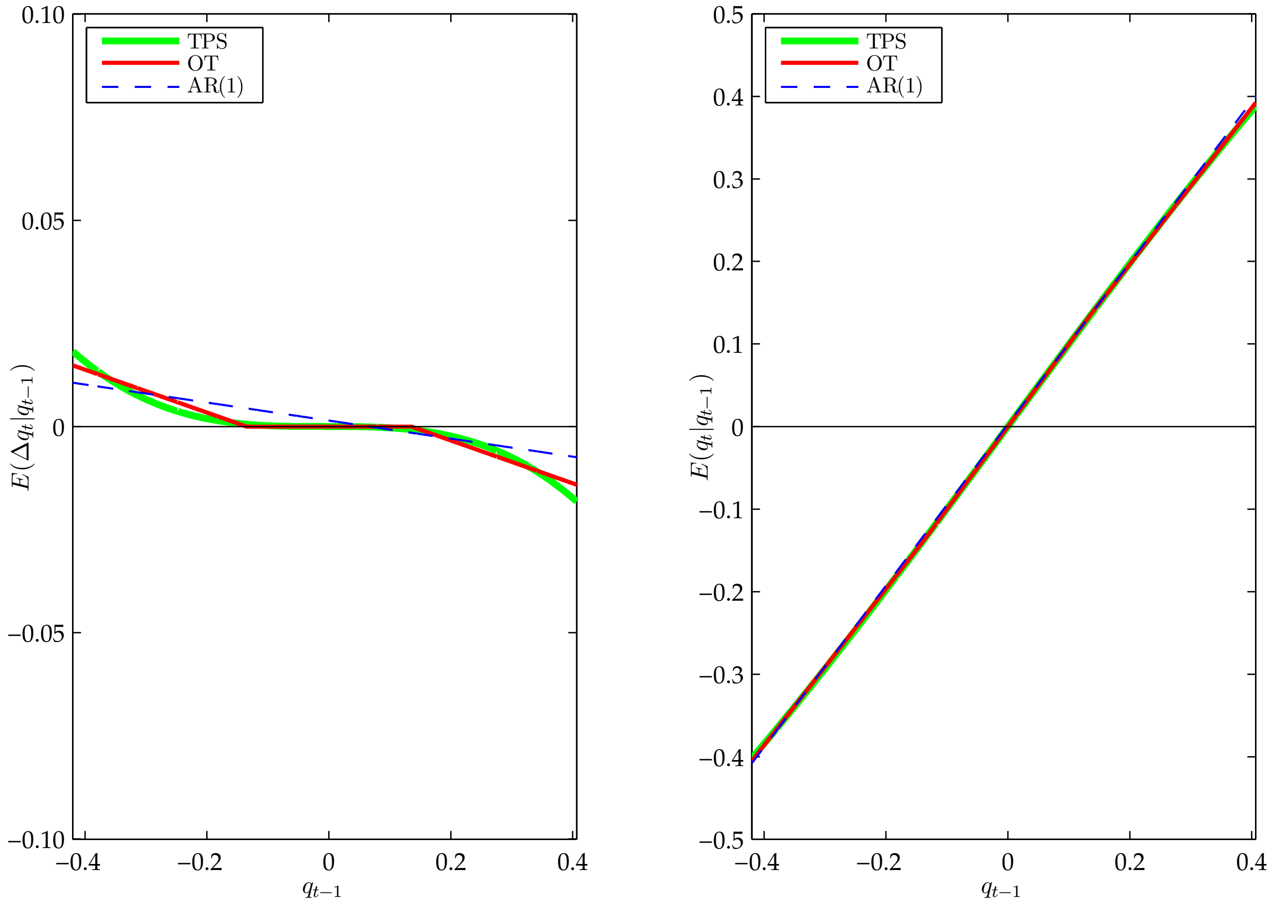
Estimates	United Kingdom	Germany	France	Japan
$\overline{d}$	-6.63×10^{-6}	2.11×10^{-5}	2.57×10^{-5}	-7.81×10^{-6}
$(se.)$	(4.63×10^{-6})	(3.13×10^{-5})	(2.99×10^{-5})	(1.04×10^{-5})
$[t-statistic]$	$[-1.4334]$	[0.6724]	[0.8599]	$[-0.7498]$
$\omega\overline{d}$	-3.68×10^{-6}	8.76×10^{-6}	1.48×10^{-5}	-3.06×10^{-6}
$(se.)$	(1.15×10^{-6})	(1.87×10^{-5})	(2.01×10^{-5})	(2.89×10^{-6})
$[t-statistic]$	$[-3.1954]$	[0.4694]	[0.7382]	$[-1.0578]$

Notes: Unweighted ($\overline{d}$) and weighted ($\omega\overline{d}$) Diebold and Mariano (1995, DM) test statistics. Standard errors ($se.$) are of the Newey and West (1987, NW) type, with a truncation lag of 10. $\overline{d}$ was calculated as the arithmetic mean of $d_t = (\epsilon^{AR}_{t|t-1})^2 - (\epsilon^{TPS}_{t|t-1})^2$ over the out-of-sample period, with $\epsilon^{AR}_{t|t-1}$ and $\epsilon^{TPS}_{t|t-1}$ being the one step ahead forecast errors from the AR(1) and TPS models, respectively. The small sample correction factor of Harvey, Leybourne and Newbold (1997) was used in the construction of both unweighted and weighted test statistics. $\omega\overline{d}$ was computed as the arithmetic mean of $\omega_t d_t$, where $\omega_t = 1 - \hat{f}(q_t)/\max[\hat{f}(q_t)]$ and $\hat{f}(q_t)$ is an estimate of the density function of q_t, evaluated at the out-of-sample data points.

Table 3: Unweighted DM test results for 24 step-ahead forecast.

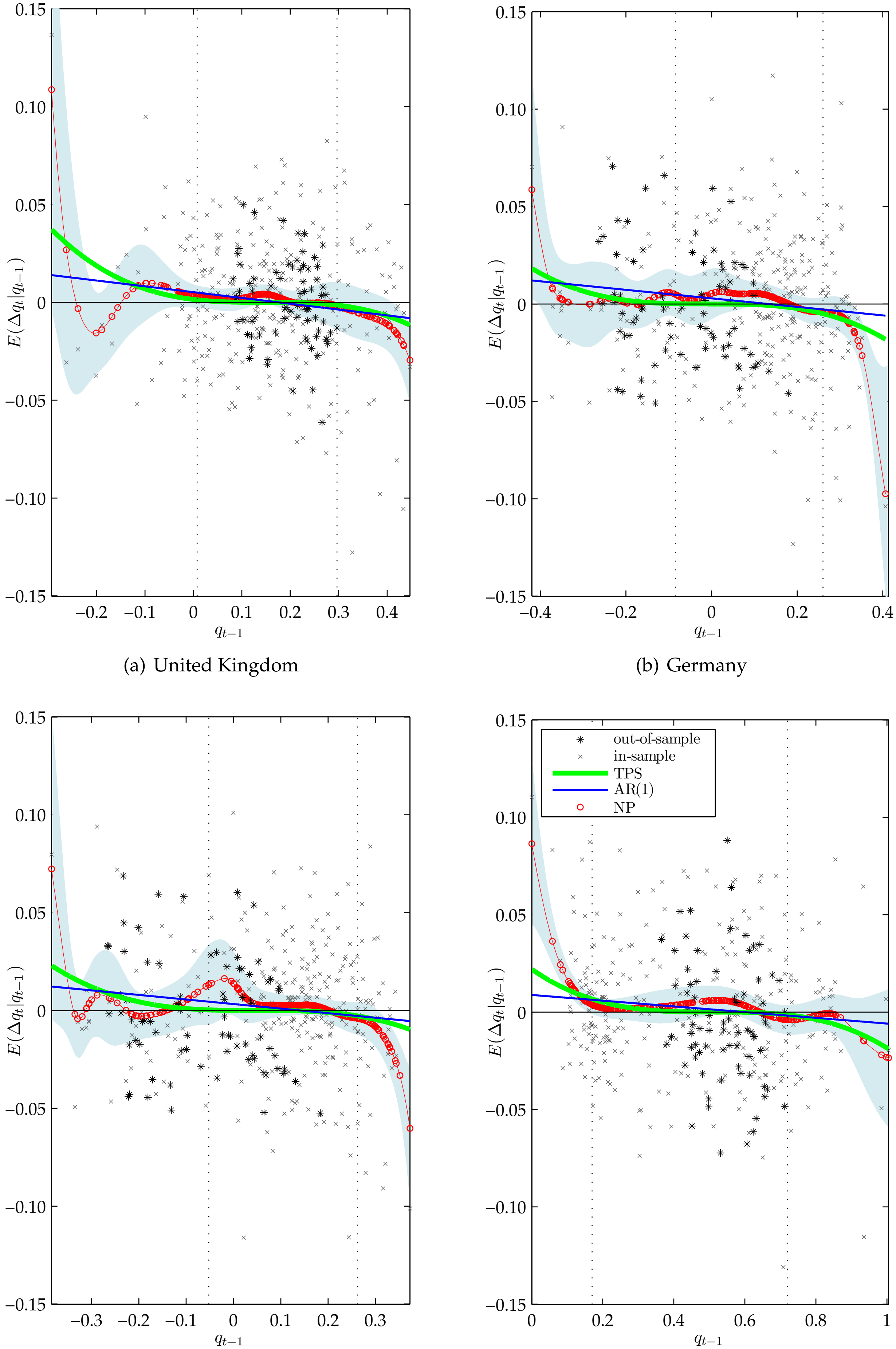
Estimates	United Kingdom	Germany	France	Japan
$\overline{d}$	-2.48×10^{-8}	1.89×10^{-7}	6.50×10^{-8}	2.26×10^{-7}
$(se.)$	(1.54×10^{-7})	(7.48×10^{-7})	(5.94×10^{-7})	(4.76×10^{-6})
$[t-statistic]$	$[-0.1607]$	[0.2531]	[0.1095]	[0.4757]

Notes: The unweighted DM test statistic $\overline{d}$ and its standard error ($se.$) were computed as in Table 2.



(a) $E(\Delta q_t|q_{t-1})$ (b) $E(q_t|q_{t-1})$

Figure 1: Plots of the implied conditional means. The thick green line and the thin red line show respectively the implied conditional means of the Band-TAR model of Obstfeld and Taylor (1997, OT) and the ESTAR model of Taylor *et al.* (2001, TPS). The dashed blue line corresponds to the implied conditional mean of an AR(1).



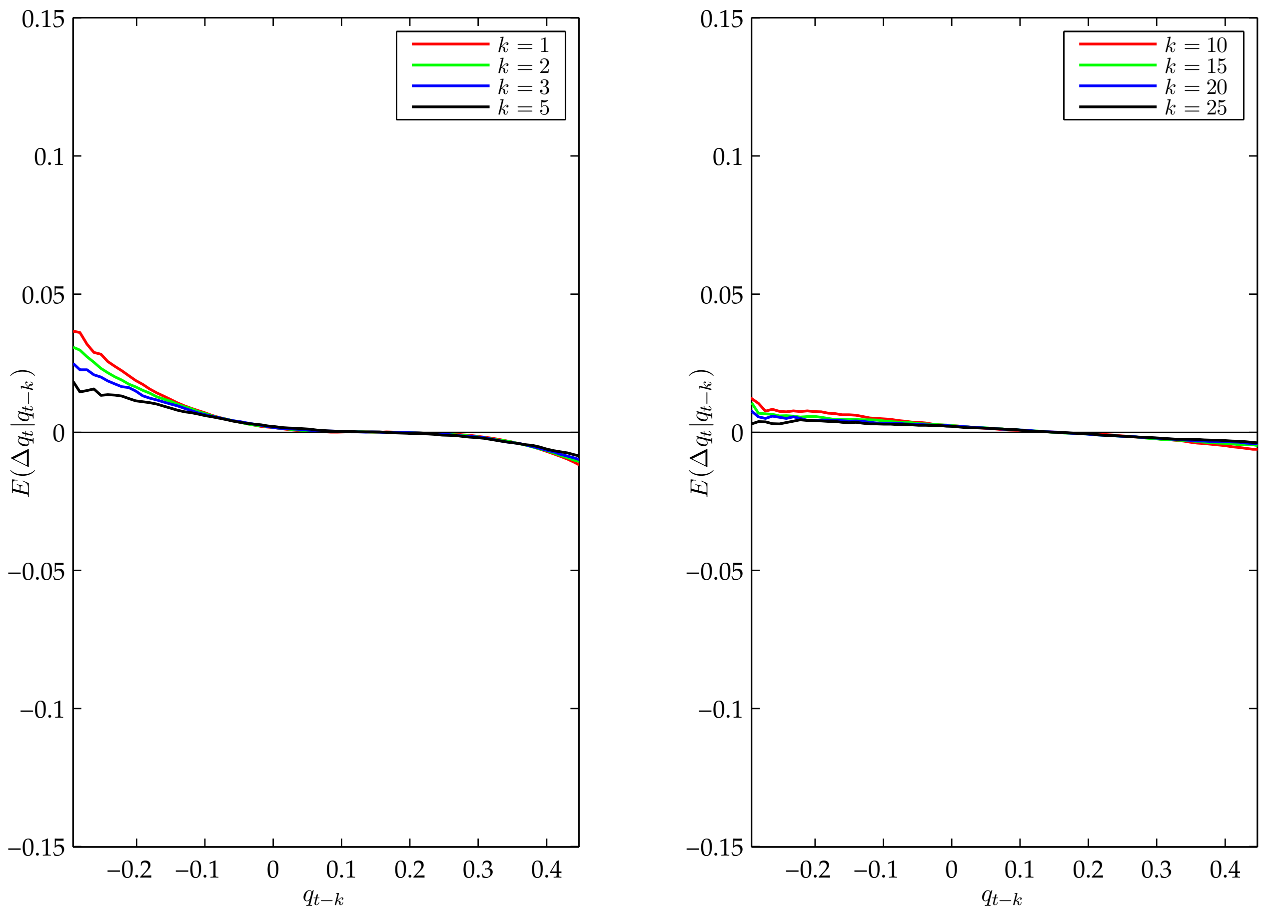
(a) United Kingdom

(b) Germany

(c) France

(d) Japan

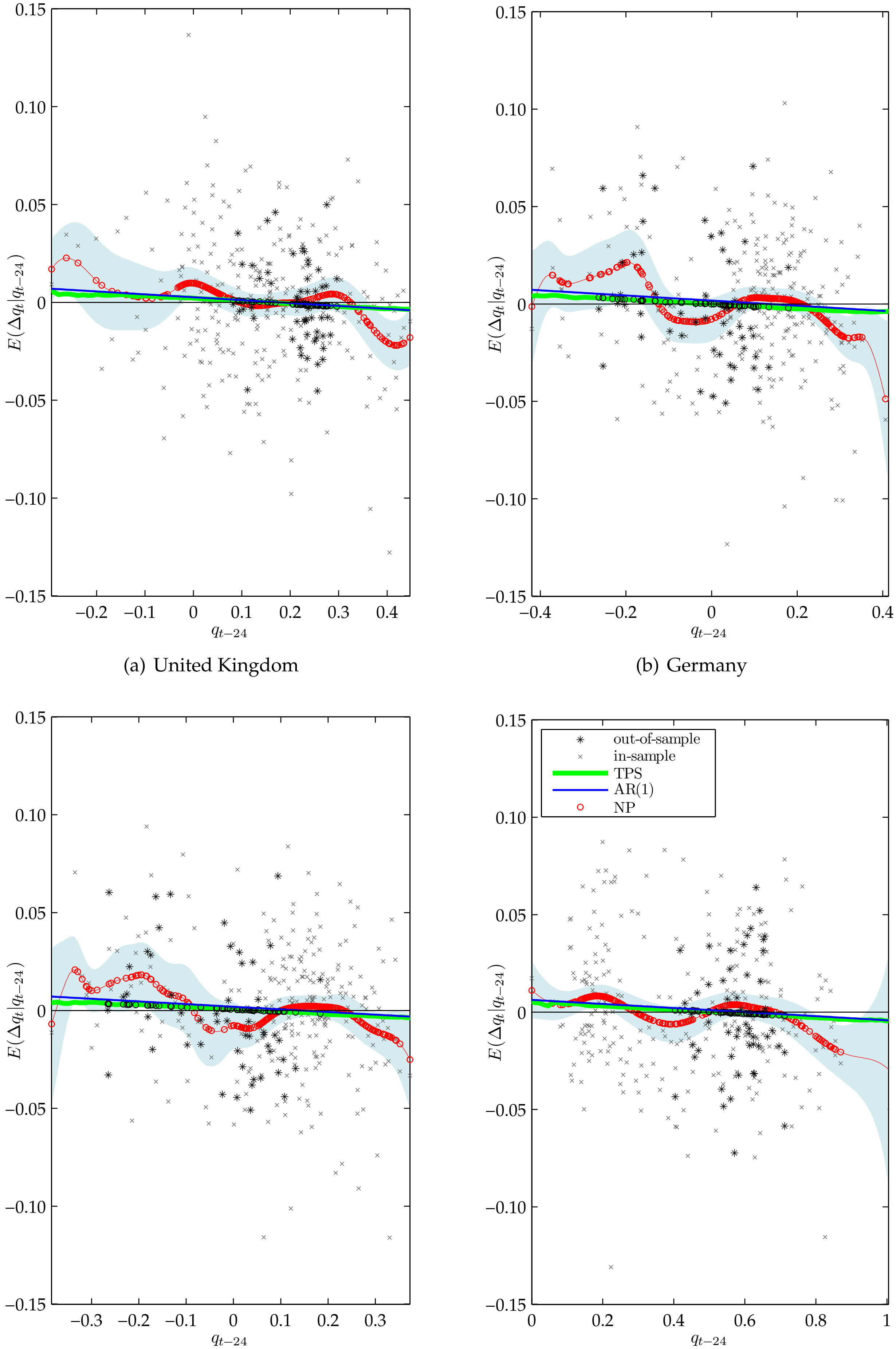
Figure 2: One step-ahead conditional forecasts. The thick green and thin blue lines show the one step-ahead conditional forecasts of the TPS and AR(1) models, respectively. Red circles are the non-parametric conditional means, with 95% confidence intervals drawn as blue shading. Grey crosses mark the 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.



(a) $k = 1, 2, 3, 5$.

(b) $k = 10, 15, 20, 25$.

Figure 3: Conditional means corresponding to k step ahead forecast. These were obtained as non-parametric estimates of the conditional mean $E(\Delta q_t|q_{t-k})$ from 4 million simulated pseudo observations from the ESTAR model of TPS under the parameter setting of the UK series. The conditional mean $E(\Delta q_t|q_{t-k})$ was computed at 100 equally spaced points over the interval $[\min(q_t), \max(q_t)]$.



(a) United Kingdom

(b) Germany

(c) France

(d) Japan

Figure 4: 24 step-ahead conditional forecasts. The content of these plots is the same as in Figure 2. Black circles are superimposed to denote the 24 step-ahead conditional forecast computed from the recursive scheme outlined in (4). The conditional forecasts were averaged over 10,000 draws.