The forecasting performance of SETAR models: an empirical application

Gianna Boero* Federico Lampis†

Abstract

The aim of this paper is to evaluate the forecasting performance of SETAR models with an application to the industrial production index of four major European countries over a period which includes the last Great Recession. Both point and interval forecasts are considered at different horizons against those obtained from two linear models. We follow the approach suggested by Teräsvirta, van Dijk, and Medeiros [2005] according to which a dynamic specification may improve the forecast performance of the nonlinear models with respect to the linear models. We re-specify the models every 12 months and we find that the advantages of this procedure are particularly evident in theforecast rounds immediately following the re-specification.

Keywords: SETAR models, point forecasts, interval forecasts, forecasting accuracy, industrial production index. **JEL Classification**: C22, C52, C53.

1 Introduction

Due to their ability to represent asymmetrical movements, the nonlinear time series models have been applied to macroeconomic variables to study the business cycle. The most common nonlinear models include Smooth-Transition Autoregressive (STAR) models, Self-Exciting Threshold Autoregressive (SETAR) models and Markov-Switching models. During the last two

*University of Warwick, Coventry, UK. E-mail: gianna.boero@warwick.ac.uk
†University of Birmingham, Birmingham, UK. E-mail: f.lampis@bham.ac.uk

decades many researchers compared the forecasting performance of nonlinear models to their linear counterparts. A common finding is that, even if the nonlinear models can provide a better in-sample fit than the linear models, they cannot always predict better (see Clements and Smith [1999]and Stock and Watson [1999], amongst others). A possible explanation for this poor forecast performance lies in the fact that the nonlinearities could be highly significant in-sample but not in the out-of-sample period, as suggested by Diebold and Nason [1990] in their application to exchange rate series.

Another line of research has examined under which conditions the nonlinear time series forecasts may outperform linear models. Boero and Marrocu [2002, 2004], for example, evaluate the point and interval forecasts of SETAR models conditional on regimes and find significant improvements in the quality of the SETAR forecasts in correspondence of specific regimes. Macroeconomic data are typically non-stationary in mean and may exhibit strong seasonal patterns, then in most cases a transformation of the raw data is necessary. However, several studies warn that the transformation applied to the data may introduce some bias which could affect the nonlinear characteristics of the original data (see Ghysels et al. [1996], de Bruin and Franses [1998] and Franses and de Bruin [2000]). Moreover, such transformations can have an impact on the forecasting performance of the models. In this respect, Franses and van Dijk [2005] examine the forecasting performance of nonlinear models relative to that of linear models, for quarterly series of industrial production from 17 OECD countries. According to their results, linear autoregressive models with a simple description of seasonality outperform nonlinear models at short forecast horizons, whereas nonlinear models with more elaborate seasonal patterns across regimes dominate at longer horizons. More recently, a number of studies have re-examined the forecasting performance of nonlinear time series models. Teräsvirta, van Dijk, and Medeiros [2005] conduct a study using 47 monthly macroeconomic variables of the G7 economies and find that STAR models have a superior forecast performance than linear models in a large number of cases. These authors emphasise that nonlinear features in time series data could be more or less pronounced in different periods of time, and these could be better captured by frequent re-specification of the models.

Ferrara, Marcellino, and Mogliani [2012] analyse the forecasting performance of nonlinear models (STAR, TAR, time varying specifications and Markov Switching models) for 18 OECD countries and 23 variables. The models are estimated with data from 1970 to 2003 and point forecasts are evaluated over the period 2004q1 to 2009q4, using both fixed moving windows and expanding windows. The models are re-estimated each time another observation is added to the information set, but the specification is assumed to remain unchanged over the forecast period. The study concludes that, on average, the nonlinear models do not outperform the linear models even during the Great Recession period.

In this paper we study the forecasting accuracy of SETAR models taking into account the limitations and recommendations of the studies mentioned above. As in Ferrara et al. [2012], our sample period includes the Great Recession of 2008-2009, which provides a good platform to compare and evaluate the relative out-of-sample forecasting performance of alternative models in periods of recessions and expansions. In several European countries the last recession has been very intense and the economy has recovered very slowly after that. The variable used in this analysis is the seasonally unadjusted monthly Industrial Production Index (IPI) for France, Italy, Spain and the United Kingdom. The IPI is one of the key indicators of the business cycle fluctuations for these countries, given the dimension of their industrial sector. Our analysis is intended to evaluate the models on their ability to produce both point and interval forecasts. The models are estimated on an expanding window of data starting with 1975.1-2005.12 and using a recursive scheme. The forecasting sample ranges from 2006.1 to 2011.12, covering the years before, during and after the Great Recession of 2008-2009. Following the recommendation of Teräsvirta et al. [2005], the models are fully re-specified every twelve months, while they are re-estimated for each new monthly observation included in the sample, and a new set of 1, 3, 6 and 12-step-ahead forecasts are computed. This procedure enables us to replicate a genuine "real time" forecasting environment. As benchmark models we use a standard linear autoregressive model and a seasonal ARMA model that are re-specified every year like the SETAR models.

The most important conclusions can be summarised as follows. The results of point forecast evaluation suggest that there are some gains in the forecast performance of the SETAR models associated with a frequent re-specification. These gains are particularly evident for the 1-step-ahead forecasts, moreover, and in line with the findings by Ferrara et al. [2012], these advantages are stronger outside the recession period. The rest of the paper is organised as follows. In Section 2 we describe the SETAR model and the methodological issues associated with their specification, estimation and their use for forecasting. In Section 3 we present the data and we report the results of the estimation of the models. In Section 4 we describe the forecasting exercise and discuss the results of the evaluation of point and interval forecasts. Section 5 contains concluding remarks.

2 Model description

In general a Self-Exciting Threshold Autoregressive (SETAR) model can be considered a linear AR model where the autoregressive parameters depend on the regime or state. A SETAR model with two regimes is defined as:

$$Y_t = \begin{cases} \alpha_0 + \alpha_1 Y_{t-1} + \cdots + \alpha_{p_1} Y_{t-p_1} + \epsilon_{1t}, & \text{if } \mathrm{Y_{t-d}} \leq \gamma \\ \beta_0 + \beta_1 Y_{t-1} + \cdots + \beta_{p_2} Y_{t-p_2} + \epsilon_{2t}, & \text{if } \mathrm{Y_{t-d}} > \gamma \end{cases} \tag{1}$$

where $t = 1, ..., T$, $d \geq 0$ is an integer called the delay parameter, Y_{t-d} is the threshold variable that defines which regime is operating at the time t, γ is the threshold parameter and $\epsilon_{r,t} \overset{iid}{\sim} N(0, \sigma_r^2)$, $r = 1, 2$. The autoregressive orders in the two regimes, p_1 and p_2, need not to be identical but they must be greater than one. The parameters α_j $(j = 0, \cdots, p_1)$ are the coefficients of the lower regime when $(Y_{t-d} \leq \gamma)$, and β_j $(j = 0, \cdots, p_2)$ are the coefficients of the upper regime when $(Y_{t-d} > \gamma)$.

The models are estimated following the three-stage procedure suggested by Tong [1990]. In the first stage, for given values of γ and d, depending on whether or not $Y_{t-d} \leq \gamma$, the data are assigned to a lower and an upper regime with n_1 and n_2 observations, respectively, and separated autoregressive models are estimated by Maximum Likelihood. The order of each autoregression is chosen according to the usual Akaike Information Criterion (AIC). In the second stage, γ is searched over a set of possible values, while d remains fixed. The re-estimation of the separate autoregressive models allows the determination of the γ parameter, as the one for which the overall AIC (equal to the sum of the AIC in each regime) attains its minimum value. Note that the search of the threshold value γ is usually restricted to be between two predetermined percentiles of Y_t, for example, in our analysis below, we conduct the search between the 15^{th} and the 85^{th} percentiles. In stage three, d is searched over values between 1 and p, where p is set to a maximum value (in our case we set $1 \leq d \leq 6$). The search over d is carried out by repeating both stage 1 and stage 2, and the selected value of d is, again, the value that minimises the AIC.

The use of the SETAR model for forecasting purposes leads to some typical problems of nonlinear models. Specifically, the computation of multi-step-ahead forecasts from nonlinear models involves the solution of complex analytical calculations and the use of numerical integration techniques, or alternatively, the use of simulation methods. In this study, the forecasts are obtained by applying the Monte Carlo simulation method, so that each point forecast is obtained as the average of 1000 replications. For example, the 2-step-ahead Monte Carlo forecast is computed by

$$\hat{Y}_{t+2|t} = \frac{1}{k}\sum_{i=1}^{k} F(\hat{Y}_{t+1|t} + e_i; \theta) \tag{2}$$

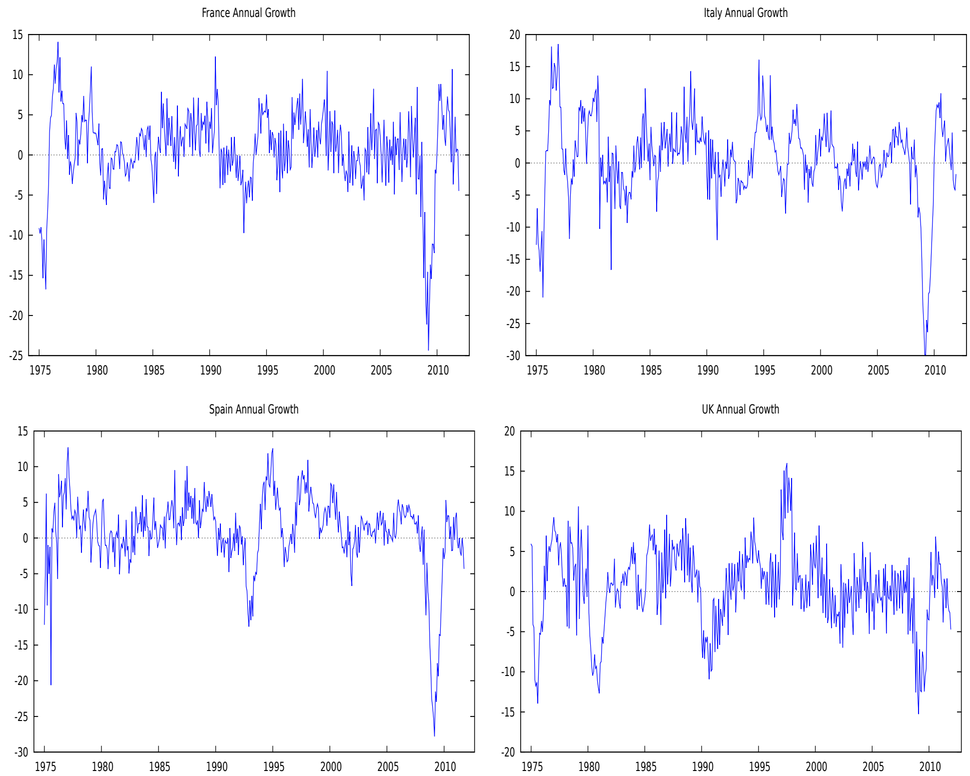
where F(.) is the nonlinear function that represents the SETAR model of equation (1), k is the number of iterations of the Monte Carlo ($k = 1000$), $\hat{Y}_{t+1|t}$ is the 1-step-ahead forecast, θ is the vector of the parameters α_j, β_j defined above, and e_i is the realisation of the error process drawn from the distribution of $\epsilon_{r,t+1}$, $r = 1, 2$ (see Franses and van Dijk [2000] and Cryer and Chan [2008]). Notice that the drawing in period $t + 2$ is made from a distribution with a variance appropriate for the regime the process is in, which is determined by $\hat{Y}_{t+1|t}$.

3 Empirical analysis

We use data on monthly unadjusted series on Industrial Production for four EU countries: France, Italy, Spain and the United Kingdom. The data are taken from the OECD Main Economic Indicators. The sample runs from January 1975 to December 2011, the base year for the indices is 2005, the series are analysed as annual (twelve-month) growth: $y_t = 100 \ln(IPI_t/IPI_{t-12})$. The use of seasonally unadjusted series is advocated in several studies in order to avoid the undesirable effects of filters such as the X-11, which may obscure the features of the data and the distinction between regimes. Similar transformation of this variable has been used by Franses and van Dijk [2005], Teräsvirta et al. [1994], Granger and Teräsvirta [1993] and Teräsvirta and Anderson [1992]. The variables are plotted in Figure 1. As we can observe, the Great Recession of 2008-2009 has been very deep in France, Italy and Spain, with the largest fall reaching 30% in Italy. In the case of the United

Kingdom, the shrink of the IPI during the crisis of 2008-2009 has been less pronounced. In order to detect the presence of nonlinearities in the series, we perform two commonly used tests: the Tsay [1986] test and the Likelihood Ratio (LR) test proposed by Tong [1990].

Figure 1: Annual rate of IPI growth



As in Teräsvirta et al. [2005], the tests have been performed sequentially, once every 12 months, starting with the period 1975.01-2005.12 and ending with the period 1975.01-2011.12. For all the countries there is strong evidence of nonlinearities over all of the samples considered, with the exception of Italy for which linearity is rejected only when the last three years (2009, 2010, 2011) are added to the sample (see linearity tests in the Appendix). In what follows, we estimate three types of models: a SETAR, a simple AR, and a seasonal ARMA. All the models are estimated recursively, that is, the first estimation is performed using the sample 1975.1-2005.12, and a first set of 1, 3, 6 and 12-months ahead forecasts are calculated.

Then, each time the models are re-estimated by expanding the sample with one observation, a new set of forecasts are computed. This process is repeated until the last available data point, that is, 1, 3, 6, 12 months before December 2011, depending on the forecast horizon. These forecasts can be considered genuine forecasts, as in the specification and estimation stage we completely ignore the information embodied in the forecasting period. Additionally, the models are re-specified once a year, such that the first specification is based on data up to December 2005 and the last specification on data up to December 2010. For all the models the optimal lag length is selected on the basis of the Akaike information criterion.

In the case of France, Spain and the United Kingdom, the autoregressive order of the AR model remains unchanged during most of the estimation periods considered ($p = 15$, $p = 14$ and $p = 16$, respectively), whilst, in the case of Italy, the order p changes depending on the sample employed (from $p = 12$ to $p = 15$). Also the structure of the seasonal ARMA model remains unchanged over the entire sample considered: for France, Italy and the United Kingdom the model consists of three autoregressive terms in the regular part and a moving-average term in the seasonal part, while, for Spain, the seasonal component is captured by an autoregressive term. The identification process for the SETAR models follows the description in Section 2. As it can be observed from Table 1, the specification of the SETAR models changes over time in terms of the number of lags entering each regime and the value of the delay parameter. We also estimated various STAR models, but in most cases they collapsed to a SETAR model. The behaviour of the variables analysed is characterised by the presence of many sharp and abrupt changes, so the switching mechanism of the SETAR model is more suitable than that of the STAR model. All the estimations and forecasts for the SETAR models have been carried out with the library TSA of R, for more information on the working of this library see Cryer and Chan [2008].

Table 1: SETAR model specifications

France								Italy							
Sample	p_1	p_2	d	γ	n_1	n_2	AIC	Sample	p_1	p_2	d	γ	n_1	n_2	AIC
1975.01-2005.12	4	14	1	-1.907	65	292	1656	1975.01-2005.12	7	12	6	-3.28	54	306	1890
1975.01-2006.12	4	14	1	-1.907	66	303	1720	1975.01-2006.12	7	12	6	-2.85	64	308	1959
1975.01-2007.12	4	14	1	-1.907	68	313	1777	1975.01-2007.12	7	12	6	-2.85	64	320	2017
1975.01-2008.12	4	14	1	-1.852	76	317	1891	1975.01-2008.12	7	12	6	0.91	184	212	2089
1975.01-2009.12	7	14	3	-1.335	101	305	1956	1975.01-2009.12	7	13	6	-2.37	91	316	2159
1975.01-2010.12	7	14	3	-1.335	104	314	2021	1975.01-2010.12	7	15	6	-2.37	95	322	2209

Spain								United Kingdom							
Sample	p_1	p_2	d	γ	n_1	n_2	AIC	Sample	p_1	p_2	d	γ	n_1	n_2	AIC
1975.01-2005.12	7	12	6	0.246	116	244	1656	1975.01-2005.12	4	16	5	1.072	172	184	1772
1975.01-2006.12	7	12	6	0.246	118	254	1692	1975.01-2006.12	5	16	4	2.372	220	148	1835
1975.01-2007.12	7	12	6	0.246	118	266	1745	1975.01-2007.12	5	16	4	2.372	229	151	1884
1975.01-2008.12	7	12	5	-1.149	71	325	1828	1975.01-2008.12	5	16	4	2.372	238	154	1961
1975.01-2009.12	2	12	5	-0.884	93	315	1907	1975.01-2009.12	5	16	4	2.372	250	154	2030
1975.01-2010.12	4	13	5	-1.149	78	341	1886	1975.01-2010.12	4	16	5	1.072	215	201	2087

p_1: autoregressive order lower regime; p_2: autoregressive order upper regime; γ: threshold parameter; AIC: Akaike Information Criterion; n_1: number of observations of the lower regime; n_1: number of observations of the upper regime.

4 Forecasting exercise

The forecast accuracy of the models is assessed over the period 2006.01-2011.12, which includes a period of expansion (2006-2007), a deep recession (2008-2009) and a slow recovery (2010-2011). For both point and interval forecasts we use the recursive scheme described in the previous section.

4.1 Point forecasts

The accuracy of the point forecasts is measured by the Root Mean Squared Forecast Error (RMSFE). We then calculate the ratio between the RMSFE of each of the linear models and that of the SETAR model: $R_{AR} = RMSFE_{AR}/RMSFE_{SETAR}$ and $R_{SARMA} = RMSFE_{SARMA}/RMSFE_{SETAR}$. A ratio greater than one means that the SETAR model outperforms the linear model. For each country in Table 2 we report the results for forecasts with horizon h equal to 1, 3, 6 and 12 months. In order to assess the sensitivity of the results to specific sub-samples, the RMSFEs are calculated over the 12 forecasts obtained in each year, with the exception of the 2006 ratios which are based on a different number of forecasts, depending on the forecasting horizon. Precisely, the ratios for 2006 are based on 12 forecasts for $h = 1$ month, 10 forecasts for $h = 3$ months, 7 forecasts for $h = 6$ months and 1 forecast for $h = 12$ months. Additionally, for each horizon, the last row of Table 2 reports the percentage of times the ratios are greater than one ($ratios\%$), that is, the percentage of times the SETAR model outperforms the linear benchmarks over the whole forecast period. The last column reports a synthetic measure of the performance of the models: the overall mean across countries and years. The results are mixed, as they vary across countries, years and forecast horizon. From Table 2 we can see that the highest gains of the SETAR models are shown in the case of Spain with RMSFE ratios greater than one up to 73% times.

Table 2: RMSFE Ratios by forecast sample period; h=1, 3, 6, 12.

h = 1 month

Forecast	France		Italy		Spain		United Kingdom		overall
period	R_{AR}	R_{SARMA}	R_{AR}	R_{SARMA}	R_{AR}	R_{SARMA}	R_{AR}	R_{SARMA}	mean
2006:1-12	**1.06**	0.93	**1.02**	**1.02**	**1.26**	**1.27**	0.99	0.85	**1.05**
2007:1-12	**1.02**	**1.10**	**1.02**	**1.01**	0.98	**1.05**	**1.25**	**1.41**	**1.11**
2008:1-12	0.89	0.88	0.96	0.93	**1.07**	**1.09**	0.94	0.90	0.96
2009:1-12	**1.14**	**1.21**	**1.11**	**1.03**	**1.21**	**1.13**	0.76	0.78	**1.05**
2010:1-12	**1.02**	0.89	0.90	0.90	0.86	0.69	0.78	0.80	0.85
2011:1-12	**1.07**	**1.11**	0.94	**1.12**	0.79	0.82	0.82	0.77	0.93
ratios%	60%	51%	49%	58%	56%	65%	54%	38%	54%

h = 3 months

Forecast	France		Italy		Spain		United Kingdom		overall
period	R_{AR}	R_{SARMA}	R_{AR}	R_{SARMA}	R_{AR}	R_{SARMA}	R_{AR}	R_{SARMA}	mean
2006:3-12	**1.05**	0.95	**1.02**	**1.17**	**1.20**	**1.34**	0.97	0.78	**1.06**
2007:1-12	0.93	**1.01**	**1.04**	**1.10**	**1.03**	**1.09**	0.99	**1.20**	**1.05**
2008:1-12	0.94	0.94	0.98	1.00	**1.01**	**1.08**	0.92	0.86	0.97
2009:1-12	0.76	0.78	0.89	0.89	0.96	0.94	0.84	0.87	0.87
2010:1-12	0.94	0.79	0.83	0.89	0.71	0.51	0.75	0.70	0.76
2011:1-12	**1.04**	0.93	0.90	**1.05**	**1.07**	1.08	0.73	0.63	0.93
ratios%	54%	34%	46%	53%	67%	73%	24%	17%	46%

h = 6 months

Forecast	France		Italy		Spain		United Kingdom		overall
period	R_{AR}	R_{SARMA}	R_{AR}	R_{SARMA}	R_{AR}	R_{SARMA}	R_{AR}	R_{SARMA}	mean
2006:6-12	**1.03**	0.86	0.92	**1.29**	**1.27**	**1.61**	0.90	0.94	**1.10**
2007:1-12	**1.02**	**1.02**	0.97	**1.14**	1.00	**1.14**	0.89	**1.03**	**1.03**
2008:1-12	0.96	0.99	0.99	**1.05**	**1.06**	**1.16**	**1.05**	**1.04**	**1.04**
2009:1-12	0.65	0.68	0.84	0.85	0.92	0.99	0.96	0.97	0.86
2010:1-12	0.41	0.48	0.93	**1.09**	0.78	0.62	0.64	0.69	0.71
2011:1-12	**1.41**	**1.07**	**1.03**	0.91	0.90	0.96	0.57	0.59	0.93
ratios%	46%	33%	37%	60%	39%	63%	21%	16%	39%

h = 12 months

Forecast	France		Italy		Spain		United Kingdom		overall
period	R_{AR}	R_{SARMA}	R_{AR}	R_{SARMA}	R_{AR}	R_{SARMA}	R_{AR}	R_{SARMA}	mean
2006:12-12	**1.10**	**1.08**	0.89	**1.10**	**2.15**	**2.83**	0.72	0.84	**1.34**
2007:1-12	0.98	1.02	0.96	**1.10**	**1.05**	**1.30**	0.78	0.82	1.00
2008:1-12	1.00	**1.03**	0.96	**1.08**	0.99	**1.03**	0.98	**1.01**	**1.01**
2009:1-12	0.99	0.99	1.00	1.00	**1.05**	**1.07**	**1.01**	**1.04**	**1.02**
2010:1-12	0.11	0.22	0.37	0.66	0.64	0.62	0.57	0.73	0.49
2011:1-12	**1.37**	**1.23**	0.76	0.49	0.62	0.49	0.61	0.68	0.78
ratios%	43%	51%	31%	46%	54%	61%	15%	25%	41%

R_{AR} is the ratio $RMSFE_{AR}/\ RMSFE_{SETAR}$; R_{SARMA} is the ratio $RMSFE_{SARMA}/RMSFE$-$_{SETAR}$. The last row of each panel reports the percentage of times the SETAR model outperforms the linear counterparts, indicated as *ratios*%. The overall mean is the mean value across each row.

There are cases of clear advantages for the SETAR model with ratios around 1.61 (Spain, $h = 6$, forecast period 2006:6-2006:12), 1.41 (UK, $h = 1$, forecast period 2007:1-2007:12), 1.37 (France, $h = 12$, forecast period 2011:1-2011:12). However these outstanding results do not correspond to an homogeneous forecast superiority of the SETAR model across all countries, sample periods and horizons, as reflected by our synthetic measure of forecast accuracy. In the 1-month-ahead forecasts the SETAR model has on average a better performance 54% of the times across all countries and forecast sample periods, whereas in the 3, 6 and 12-months-ahead forecasts the SETAR model outperforms the linear models 46%, 39% and 41% of the times, respectively.

Surprisingly, but in line with the findings in Ferrara et al. [2012], there is no clear superior performance of the SETAR models during the years of recession. This reflects the inability to capture the turning points of the business cycle even if a dynamic specification approach is adopted. Perhaps a more frequent re-specification during the years of high instability could play in favour of the SETAR model, although this would have a high computational cost. Indeed the SETAR model in various occasions is able to produce forecasts with an error up to 21 times smaller than that of the linear model. For example the actual value of the IPI annual growth in 2007.01 for the UK is 2.60, the SETAR 1-month-ahead forecast is 2.45, the AR forecast is -0.66 and the seasonal ARMA forecast is -0.45, producing values for R_{AR} and R_{SARMA} of 21 and 20 respectively. These remarkably accurate forecasts typically occur in the months immediately after the models are re-specified. The models are fully specified once a year, with the first specification based on data up to 2005.12 and the last re-specification on data up to 2010.12. In order to explore further the potential benefits of a full re-specification of the models and the persistence of these benefits, for each forecast horizon, we perform a sequential calculation of the RMSFE ratios, starting with the first round of forecasts obtained after each of the six model re-specification, over the entire forecast period 2006.1-2011.12. We then add in the calculation of the RMSFEs, one at a time, and for each forecast horizon, the forecast errors from the next 11 rounds of forecasts based on the same model specification as defined in December, though the models are recursively estimated for each additional monthly observation included in the sample. The results of this exercise, for each country and forecast horizon h=1, 3, 6, and 12, are reported in Table 3. The first row of each panel reports the ratios of the RMSFEs for the first round of forecasts, that is the January forecasts for h=1, the March forecasts for h=3, the June forecasts for h=6 and the December forecasts for h=12. The second row of each panel reports the results for the first and second rounds of forecasts, and so on, up to the last row, which reports the results for all of the 72 forecasts for h=1, 70 forecasts for h=3, 67 forecasts for h=6 and 61 forecasts for h=12.

Table 3 also reports, in the last column, the overall mean of the RMSFE ratios across countries. As we can see, there are notable overall gains across countries for the SETAR models, relative to the linear benchmarks, for the first round of forecasts, with gains in the order of 36% for h=1, 10% for h=3, and 33% for h=6. Looking at the results by country, the re-specification yields large immediate benefits to the SETAR models, with gains up to 69% (France, h=1), 45% (Spain, h=1), 46% (UK, h=1), but it is also evident from Table 3 that the gains to the SETAR models are consequently reduced in the successive forecast rounds. There are no gains, overall, to the SETAR models for h=12, although we can observe some benefits immediately after the re-specification, with gains of 13% for Italy and 7% for Spain. Finally, it is of interest to note that the benefits of the re-specifications of the SETAR models persist for several months for h= 1, while the results for h=3 and h=6 show a clear tendency for the gains to the SETAR models to disappear faster. These results lend support to the dynamic specification approach suggested by Teräsvirta et al. [2005].

Table 3: Impact of model specification: RMSFE Ratios by forecast rounds; h=1, 3, 6, 12. Forecast period: 2006.1-2011.12

h = 1 month

Forecast	France		Italy		Spain		United Kingdom			overall
rounds	R_{AR}	R_{SARMA}	R_{AR}	R_{SARMA}	R_{AR}	R_{SARMA}	R_{AR}	R_{SARMA}	N.F.	mean
January	**1.69**	**1.56**	**1.10**	**1.13**	**1.45**	**1.20**	**1.46**	**1.31**	6	**1.36**
Jan-Feb	**1.47**	**1.34**	**1.10**	**1.13**	**1.66**	**1.28**	**1.01**	0.88	12	**1.23**
Jan-Mar	**1.28**	**1.25**	**1.27**	**1.23**	**1.20**	**1.13**	**1.02**	0.95	18	**1.17**
Jan-Apr	**1.27**	**1.25**	**1.27**	**1.23**	**1.05**	**1.01**	0.98	0.88	24	**1.12**
Jan-May	**1.16**	**1.16**	**1.13**	**1.07**	**1.11**	**1.06**	0.99	0.92	30	**1.08**
Jan-June	**1.14**	**1.15**	**1.12**	**1.07**	**1.08**	**1.03**	1.00	0.91	36	**1.06**
Jan-July	**1.07**	**1.07**	**1.08**	**1.02**	**1.08**	0.99	**1.01**	0.91	42	**1.03**
Jan-Aug	**1.05**	**1.04**	**1.07**	**1.02**	**1.07**	0.98	0.98	0.93	48	**1.02**
Jan-Sept	**1.06**	**1.04**	**1.05**	1.00	1.00	0.91	0.95	0.90	54	0.99
Jan-Oct	**1.05**	**1.01**	**1.03**	0.99	0.96	0.90	0.96	0.91	60	0.98
Jan-Nov	**1.01**	0.99	**1.02**	0.99	**1.03**	0.97	0.91	0.87	66	0.98
Jan-Dec	**1.01**	1.00	**1.01**	0.99	**1.02**	0.99	0.88	0.86	72	0.97

h = 3 months

Forecast	France		Italy		Spain		United Kingdom			overall
rounds	R_{AR}	R_{SARMA}	R_{AR}	R_{SARMA}	R_{AR}	R_{SARMA}	R_{AR}	R_{SARMA}	N.F.	mean
March	**1.24**	**1.31**	**1.02**	**1.07**	**1.04**	**1.10**	**1.01**	0.98	6	**1.10**
Mar -Apr	**1.36**	**1.39**	**1.10**	**1.13**	0.90	0.98	0.91	0.81	12	**1.07**
Mar-May	**1.13**	**1.21**	**1.11**	**1.15**	0.86	0.98	0.90	0.83	18	**1.02**
Mar-Jun	**1.03**	**1.10**	**1.04**	**1.07**	0.82	0.85	0.91	0.83	24	0.96
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Mar -Feb	0.88	0.85	0.90	0.93	0.94	0.94	0.86	0.82	70	0.89

h = 6 months

Forecast	France		Italy		Spain		United Kingdom			overall
rounds	R_{AR}	R_{SARMA}	R_{AR}	R_{SARMA}	R_{AR}	R_{SARMA}	R_{AR}	R_{SARMA}	N.F.	mean
June	**1.98**	**2.05**	**1.07**	**1.18**	**1.10**	**1.20**	**1.02**	**1.02**	6	**1.33**
Jun-Jul	**1.21**	**1.20**	**1.16**	**1.23**	**1.07**	**1.02**	0.92	0.92	12	**1.09**
Jun-Aug	0.60	0.59	**1.13**	**1.20**	0.92	0.83	0.87	0.97	18	0.89
Jun-Sept	0.55	0.50	0.86	0.86	0.88	0.71	0.89	0.96	24	0.78
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Jun-May	0.71	0.72	0.87	0.90	0.93	0.98	0.87	0.89	67	0.86

h = 12 months

Forecast	France		Italy		Spain		United Kingdom			overall
rounds	R_{AR}	R_{SARMA}	R_{AR}	R_{SARMA}	R_{AR}	R_{SARMA}	R_{AR}	R_{SARMA}	N.F.	mean
December	0.74	0.87	**1.02**	**1.13**	**1.07**	1.00	0.97	1.01	6	0.98
Dec-Jan	0.58	0.61	0.98	0.97	0.92	0.90	0.98	1.02	11	0.87
Dec-Feb	0.32	0.33	0.90	0.88	0.89	0.89	0.93	0.94	16	0.76
Dec-Apr	0.29	0.32	0.74	0.74	0.85	0.85	0.92	0.96	21	0.71
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Dec-Nov	0.40	0.43	0.79	0.83	0.91	0.92	0.90	0.95	61	0.77

N.F.: Number of forecasts used to calculate the RMSFEs.

4.2 Interval forecasts

In this section we broaden our forecast comparison to the ability of the model to produce correct interval forecasts. An interval forecast for a variable is the probability that the future outcome will fall within a stated interval. The lower and upper limits of the interval forecast are given as the corresponding percentiles. We use central intervals, so that for example, the 90% forecast interval is formed by the 5^{th} and 95^{th} percentiles. Evaluation of interval forecasts is conducted by means of the likelihood ratio test of correct conditional coverage (LR_{CC}) as proposed by Christoffersen [1998]. Christoffersen [1998] shows that a correctly conditionally calibrated interval forecast will provide a hit sequence I_t (for $t = 1, 2, \cdots, T$), with value 1 if the realisation is contained in the forecast interval, and 0 otherwise, that is distributed *i.i.d.* Bernoulli, with the desired success probability c.

As stressed by Christoffersen [1998], a simple test for correct unconditional coverage (LR_{UC}) is insufficient in the presence of dynamics in higher-order moments (conditional heteroscedasticity, for example) because it does not have power against the alternative that the zeros and ones are clustered in time-dependent fashion. in this sections see Christoffersen [1998].

In order to overcome this limitation, Christoffersen [1998] proposes a test for independence (LR_{IND}) which assumes a binary first-order Markov chain for the indicator function I_t. Under the null hypothesis of independence, the test follows a χ^2 distribution with one degree of freedom. The joint test of correct conditional coverage, LR_{CC}, is obtained as the sum of LR_{UC} and LR_{IND}, and is asymptotically χ^2 distributed with two degrees of freedom. For a detailed description of the tests we refer the reader to Christoffersen [1998].

The interval forecasts at 1, 3, 6 and 12-months-ahead have been computed using the recursive scheme on expanding windows as described previously.

The intervals of the SETAR models are calculated using the corresponding percentiles of the 1000 replications performed during the forecasting process. The results of the LR tests reported in Table 4 refer only to the 1-month-ahead forecasts, while the results for the 3, 6 and 12-months-ahead are available upon request. The table reports different nominal coverages c in the range (0.95-0.5), the empirical coverage π and the *p-values* of the three LR tests. Unlike the point forecast exercise, the tests on the interval forecasts have been performed over the entire forecasting period 2006.01-2011.12.

The results show that for Spain and Italy, for almost all levels of coverage, the SETAR model is the only model to pass all the three tests at the 10% significance level, while both linear models show some evidence of violation of the independence assumption, therefore failing the independence test. In the case of France, none of the models performs satisfactorily in terms of the correct conditional coverage test, due to failure of correct unconditional coverage at all intervals. Specifically, all the models appear to generate interval forecasts with actual coverage (π) smaller than the nominal coverage (c). That is, the interval forecasts corresponding to 95, 90, 75 and 50% are too small, as less than 95, 90, 75 and 50% observations actually fall into those intervals. These results may be attributed to an under-estimate of the standard errors used in the calculation of the forecast intervals. Both linear and non-linear models appear to perform equally well in the case of the UK. For longer forecast horizons, all models showed a deterioration of the accuracy of the interval forecasts in terms of both unconditional coverage and independence.

Table 4: Interval Forecasts evaluation: LR tests for unconditional coverage, independence and conditional coverage. 1-step-ahead forecasts; forecasting period: 2006.1-2011.12.

France

	AR				SARMA				SETAR			
c	π	LR_{UC}	LR_{IND}	LR_{CC}	π	LR_{UC}	LR_{IND}	LR_{CC}	π	LR_{UC}	LR_{IND}	LR_{CC}
0.95	0.85	**0.030**	0.213	**0.044**	0.83	**0.016**	0.31	**0.032**	0.79	**0.002**	0.208	**0.003**
0.90	0.73	**0.01**	0.257	**0.016**	0.73	**0.008**	**0.058**	**0.005**	0.66	**0.000**	0.317	**0.001**
0.75	0.59	**0.054**	0.486	0.123	0.59	**0.054**	0.178	**0.063**	0.49	**0.002**	0.389	**0.006**
0.50	0.31	**0.032**	0.203	**0.045**	0.41	0.308	0.137	0.197	0.31	**0.032**	0.305	**0.060**

Italy

	AR				SARMA				SETAR			
c	π	LR_{UC}	LR_{IND}	LR_{CC}	π	LR_{UC}	LR_{IND}	LR_{CC}	π	LR_{UC}	LR_{IND}	LR_{CC}
0.95	0.92	0.421	**0.055**	0.114	0.94	0.874	**0.009**	**0.031**	0.90	0.271	0.100	0.141
0.90	0.92	0.769	**0.055**	0.151	0.90	0.979	0.100	0.258	0.89	0.818	0.163	0.368
0.75	0.77	0.749	**0.063**	0.168	0.77	0.749	**0.023**	**0.072**	0.73	0.823	0.130	0.310
0.50	0.55	0.584	0.257	0.453	0.55	0.584	0.200	0.379	0.48	0.814	0.393	0.676

Spain

	AR				SARMA				SETAR			
c	π	LR_{UC}	LR_{IND}	LR_{CC}	π	LR_{UC}	LR_{IND}	LR_{CC}	π	LR_{UC}	LR_{IND}	LR_{CC}
0.95	0.87	**0.098**	**0.005**	**0.005**	0.89	0.167	**0.047**	**0.053**	0.90	0.271	0.414	0.391
0.90	0.83	0.240	**0.014**	**0.024**	0.87	0.633	**0.084**	0.201	0.79	**0.068**	0.409	0.135
0.75	0.73	0.823	0.163	0.369	0.76	0.892	**0.044**	0.130	0.68	0.357	0.503	0.522
0.50	0.52	0.814	**0.021**	**0.069**	0.56	0.481	0.745	0.740	0.41	0.308	0.307	0.353

United Kingdom

	AR				SARMA				SETAR			
c	π	LR_{UC}	LR_{IND}	LR_{CC}	π	LR_{UC}	LR_{IND}	LR_{CC}	π	LR_{UC}	LR_{IND}	LR_{CC}
0.95	0.94	0.874	0.649	0.890	0.94	0.874	0.649	0.890	0.90	0.271	0.796	0.528
0.90	0.92	0.769	0.488	0.753	0.93	0.565	0.566	0.719	0.86	0.473	0.348	0.498
0.75	0.80	0.487	0.295	0.454	0.82	0.375	0.752	0.641	0.79	0.613	0.409	0.625
0.50	0.51	0.938	0.937	0.994	0.54	0.696	0.951	0.925	0.59	0.308	0.246	0.304

In bold are reported those *p-values* for which the H_0 of the LR tests is rejected at 10%. c indicates the nominal coverage and π indicates the actual unconditional coverage.

5 Conclusions

In this paper we have studied the forecast accuracy of SETAR models in an application to the monthly Industrial Production Index of four major European countries. This assessment has been done on the point and interval forecasts and using as benchmark the forecasts of two standard linear models. Using data covering the last Great Recession up to December 2011, we followed the recommendations of Teräsvirta et al. [2005] who suggest that frequent model re-specification increases the forecast accuracy of nonlinear models. So, in our forecasting exercise, the models were re-estimated each time a new observation was added to the sample, using a recursive scheme on expanding windows and fully re-specified at the beginning of each forecasting window.

The forecast evaluation was conducted at the 1, 3, 6 and 12-months-ahead for both point and interval forecasts. SETAR models produced superior point forecasts in case of 1-month-ahead forecasts 56% of the times, and only 46%, 39%, 41% for the 3, 6 and 12-months-ahead forecasts. As in Teräsvirta et al. [2005] we found that dynamic re-specification of the SETAR models resulted in a better forecasting performance, and this finding, in our case, was particularly evident for the 1-month-ahead forecasts. Interestingly, and in line with previous findings by Ferrara et al. [2012], the advantages of the SETAR models are less pronounced during the recession period. The interval forecasts performance also varied with the forecast horizon. For the 1-month-ahead forecasts there were cases of clear superior performance of the SETAR models, while for longer horizons, both linear and nonlinear models showed a tendency to deteriorate by failing either the unconditional coverage or the independence test, or both.

References

G. Boero and E. Marrocu. The performance of non-linear exchange rate models: A forecasting comparison. *Journal of Forecasting*, 20(2):305–320, 2002.

G. Boero and E. Marrocu. The performance of setar models: a regime conditional evaluation of point, interval and density forecasts. *International Journal of Forecasting*, 20(2):305–320, 2004.

P. Christoffersen. Evaluating interval forecasts. *International Economic Review*, 53, 1998.

M. P. Clements and J. Smith. A monte carlo study of the forecasting performance of empirical SETAR models. *Journal of Applied Econometrics*, 14: 123–141, 1999.

J. Cryer and K. Chan. *Time series analysis with applications in R*. Springer, 2008.

P. de Bruin and P. Franses. On data transformation and evidence of nonlinearity. *Erasmus University Rotterdam, Econometric Institute,*, 1998.

F. Diebold and J. Nason. Nonparametric exchange rate prediction? *Journal of International Economics*, 28:315–332, 1990.

L. Ferrara, M. Marcellino, and M. Mogliani. Macroeconomic forecasting during the Great Recession: The return of non-linearity? Working papers 383, Banque de France, May 2012.

P. Franses and D. van Dijk. *Nonlinear time series models in empirical finance.* Cambridge University Press, Cambridge, 2000.

P.H. Franses and P. de Bruin. Seasonal adjustment and the business cycle in unemployment. *Studies in Nonlinear Dynamics and Econometrics*, 4: 73–84, 2000.

P.H. Franses and D. van Dijk. The forecasting performance of various models for seasonality and nonlinearity for quarterly industrial production. *International Journal of Forecasting*, 21:87102, 2005.

E. Ghysels, C.W.J. Granger, and P. Siklos. Is seasonal adjustment a linear or nonlinear data filtering process? *Journal of Business and Economics Statistics*, 14:374–386, 1996.

C. Granger and T. Teräsvirta. *Modelling Nonlinear Economic relationships.* Oxford University Press, Oxford, 1993.

J. Stock and M Watson. A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series. *Cointegration, causality and forecasting. A festschrift in honour of Clive W.J. Granger*, pages 1–44, 1999.

T. Teräsvirta and H. Anderson. Modelling nonlinearities in business cycle using smooth transition autoregressive models. *Journal of Applied Econometrics*, 7:119–136, 1992.

T. Teräsvirta, D. Tjostheim, and C. Granger. Aspect of modelling nonlinear time series. In *Handbook of Econometrics*, chapter 48, pages 2917–2957. Elsevier Science, Amsterdam, 1994.

T. Teräsvirta, D. van Dijk, and M. Medeiros. Linear models, smooth transition autoregressions, and neural networks for forecasting macroeconomic time series: A re-examination. *International Journal of Forecasting*, 21: 755–774, 2005.

H. Tong. *Nonlinear time series, a dynamical system approach.* Oxford University Press, London, 1990.

R. Tsay. Nonlinearity test for time series. *Biometrika*, 73:461–466, 1986.