International Journal of Forecasting 18 (2002) 421–438

www.elsevier.com/locate/ijforecast

Combined forecasts from linear and nonlinear time series models

Nobuhiko Terui[a,*], Herman K. van Dijk[b]

[a]Graduate School of Economics and Management, Tohoku University, 980-8576 Sendai, Japan
[b]Erasmus University Rotterdam, Econometric Institute, P.O. Box 1738, 300 Rotterdam, The Netherlands

Abstract

Combined forecasts from a linear and a nonlinear model are investigated for time series with possibly nonlinear characteristics. The forecasts are combined by a constant coefficient regression method as well as a time varying method. The time varying method allows for a locally (non)linear modeling. The methods are applied to three data sets: Canadian lynx and sunspot series, US annual macro-economic time series — used by Nelson and Plosser (J. Monetary Econ., 10 (1982) 139) — and US monthly unemployment rate and production indices. It is shown that the combined forecasts perform well, especially with time varying coefficients. This result holds for out of sample performance for the sunspot series, the Canadian lynx number series and the monthly series, but it does not uniformly hold for the Nelson and Plosser economic time series. © 2002 International Institute of Forecasters. Published by Elsevier Science B.V. All rights reserved.

Keywords: Combining forecasts; ExpAR model; Locally (non)linear modeling; Threshold model; Time varying coefficient model

1. Introduction

Since the influential work of Bates and Granger (1969) several schemes for combining forecasts of different models have been constructed. Crane and Crotty (1967), Reinmuth and Geurts (1976) and Granger and Ramanathan (1984) propose, for instance, the use of regression methods. The latter authors point out that conventional weighting is equivalent to constrained ordinary least squares where the observations are the dependent variables, the individual forecasts are explanatory variables, and the weights are constrained to sum to one. Furthermore, they show that the unconstrained least squares method can be applied to get a better forecasting performance.

An important motive to combine forecasts from different models is the fundamental assumption that one cannot identify the true process exactly, but different models may play a complementary role in the approximation of the data generating process. We follow this idea and consider the combination of several time series models for analyzing data which show, possibly, nonlinear characteristics. We investigate the properties of combining forecasts of linear and nonlinear models by a constant coefficient

*Corresponding author.
E-mail addresses: terui@ecom.tohoku.ac.jp (N. Terui), hkvdijk@few.eur.nl (H.K. van Dijk).

0169-2070/02/$ – see front matter © 2002 International Institute of Forecasters. Published by Elsevier Science B.V. All rights reserved.
PII: S0169-2070(01)00120-0

regression method as well as time varying regression method.

There are several reasons to consider the proposed methods. Firstly, there exists empirical evidence that nonlinear models perform well for long term forecasting and that a linear model dominates in the short run. In fact, a linear model can be useful as a robust model for analyzing data which exhibit apparently nonlinear characteristics. We note that Tong (1990, pp. 425–429), proposes a simple combination, where a linear and a threshold autoregressive (TAR) model are used alternatively corresponding to upward and downward phases of time series.

Secondly, it is possible for a data generating process to switch its structure over the observation period between a linear and a nonlinear structure. The combined forecast can be based on a locally linear or locally nonlinear model. This is important for economic time series which exhibit structural change. Terui and Kariya (1997a,b) indicate that many economic series show no clear features of nonlinearity. These series appear to be standing on the borderline of linear, Gaussian and nonlinear, non-Gaussian regions.

Thirdly, by using combined forecasts, one can evaluate the contribution of each component for the whole series (constant combination) or at every time point (time varying combination).

As for the class of nonlinear time series models to be combined with a linear model, we use threshold autoregressive (TAR) models and exponential autoregressive (ExpAR) models. One reason for this is that these models have competed with each other in their performances on the Canadian lynx and Wolfe's sunspot data, which are benchmarks for nonlinear models. Studies of their performance are available. The other reason is that these models are suggested for some macroeconomic data.

We include in our framework a linear autoregressive model with time varying coefficients. Then one can investigate whether the nonlinear component or the time variation in the parameters is more important. Note that our approach can be extended to other classes of nonlinear models with possibly time varying parameters. We comment on this in our conclusions.

In the empirical application regarding economic time series, we first test the linearity of each series. Based on the outcome of the test, we choose six annual macro economic series from the well known data set from Nelson and Plosser (1982) and 2-monthly (seasonally adjusted) economic series: the US unemployment rate and the industrial production index.

It is shown that the combined forecasts perform well in most cases, especially, with time varying coefficients. However, the combined forecasts do not necessarily dominate for all series; sometimes a linear model still produces the best forecasts. Our results are in line with those of De Gooijer and Kumar (1992) and Clements and Smith (1999).

2. Nonlinear models and their combinations

In addition to a conventional linear autoregressive (AR) model, we consider two classes of nonlinear time series models; the threshold autoregressive (TAR) models and the exponential autoregressive (ExpAR) models. For stationary time series $\{Y_t\}$, a two regime self-exciting TAR model of order (p_1, p_2), denoted by TAR$(p_1, p_2; d, r)$, is defined as

$$Y_t = \begin{cases} \beta_0^{(1)} + \beta_1^{(1)} Y_{t-1} + \cdots + \beta_{p_1}^{(1)} Y_{t-p_1} + \epsilon_t^{(1)} & \text{if } Y_{t-d} \le r \\ \beta_0^{(2)} + \beta_1^{(2)} Y_{t-1} + \cdots + \beta_{p_2}^{(2)} Y_{t-p_2} + \epsilon_t^{(2)} & \text{if } Y_{t-d} > r \end{cases} \tag{1}$$

where $\{\epsilon_t^{(i)}\}$, $i = 1, 2$, is the innovation process for each regime. Their variances are denoted as $\sigma^2_{(1)}$ and $\sigma^2_{(2)}$. The parameters d and r are called delay and threshold parameters, respectively.

We note that a TAR model can be characterized as a piece-wise linear time series model.

The ExpAR model with order p, denoted by ExpAR(p; d, γ) is defined as

$$\begin{aligned} Y'_t = &\{\phi_1 + \pi_1 \exp(-\gamma {Y'_{t-d}}^2)\}Y'_{t-1} \\ &+ \{\phi_2 + \pi_2 \exp(-\gamma {Y'_{t-d}}^2)\}Y'_{t-2} + \cdots \\ &+ \{\phi_p + \pi_p \exp(-\gamma {Y'_{t-d}}^2)\}Y'_{t-p} + \epsilon_t \end{aligned} \quad (2)$$

where $\{Y'_t\}$ is a mean deleted process and $\{\epsilon_t\}$ is an innovation process with variance σ^2. This model can be interpreted as a smoothly switching model between two extreme regimes, according to the magnitude of the amplitude $|Y'_{t-d}|$.

Estimating the TAR and ExpAR model, that is, determining the lag length as well as estimating the delay and threshold parameters, is conducted by extensive use of the Akaike Information Criterion (AIC).

In order to forecast with a combination of models, we consider two techniques; the constant coefficient and the time varying coefficient method. Each method gives a combined model which is defined as follows.

Constant combination:

$$Y_t = \beta^0 + \beta^l Y^l_t + \beta^t Y^t_t + \beta^e Y^e_t + u_t. \quad (3)$$

Time varying combination:

$$\begin{aligned} Y_t &= \beta^0_t + \beta^l_t Y^l_t + \beta^t_t Y^t_t + \beta^e_t Y^e_t + v_t \\ &\equiv \boldsymbol{X}_t \beta_t + v_t;\ v_t \sim N(0, \eta^2) \end{aligned} \quad (4)$$

$$\beta_t = \beta_{t-1} + e_t;\ e_t \sim N(0, \Sigma) \quad (5)$$

where Y^l_t, Y^t_t and Y^e_t are the mean marginal predictors generated by a linear AR, a TAR, and an ExpAR model, respectively. The process $\{u_t\}$ is a white noise process; $\boldsymbol{X}_t = [1,\ Y^l_t,\ Y^t_t,\ Y^e_t]$; and $\beta_t = (\beta^0_t,\ \beta^l_t,\ \beta^t_t,\ \beta^e_t)'$.

The time varying combined model — Eqs. (4) and (5) — can be interpreted as a state space model, where Eq. (4) is the measurement equation which defines the distribution of Y_t, $t \geq 1$; and where Eq. (5) is the state equation which defines the distribution of β_t for every $t \geq 1$.

In the first step, the filtered state vector and its covariance matrix,

$$\hat{\beta}_{t|t} = E(\beta_t | Y_1, \ldots, Y_t),$$
$$\hat{\Sigma}_{t|t} = E\left[(\beta_t - \hat{\beta}_{t|t})(\beta_t - \hat{\beta}_{t|t})'\right]$$

are evaluated by applying the Kalman filter algorithm, where recursive relations regarding the predictions

$$\hat{\beta}_{t|t-1} = E(\beta_t | Y_1, \ldots, Y_{t-1}),$$
$$\hat{Y}_{t|t-1} = E(Y_t | Y_1, \ldots, Y_{t-1})$$

$$\hat{\Sigma}_{t|t-1} = E\left[(\beta_t - \hat{\beta}_{t|t-1})(\beta_t - \hat{\beta}_{t|t-1})'\right],$$
$$\hat{\sigma}^2_{t|t-1} = E(Y_t - \hat{Y}_{t|t-1})^2$$

are essential. We refer for details to Harvey (1989) and Hamilton (1994).

Next, suppose we have T observations (Y_1, $Y_2, \ldots,$ Y_T) and we want to determine the optimal inside sample estimator $\hat{\beta}_{t|T} = E(\beta_t | Y_1, \ldots, Y_T)$ of β_t, as well as its covariance matrix $\hat{\Sigma}_{t|T}$. Since the recursions regarding these smoothing estimators are available, the trajectories of each $\hat{\beta}_{\cdot t|T}$ and $(\hat{\beta}_{\cdot t|T} \pm 1.64 \hat{\sigma}_{\cdot t|T})$ are drawn in our figures, where $\hat{\beta}_{\cdot t|T}$ means the smoothing estimate of each β^o, β^l, β^t, β^e and the standard deviations for the marginal predictors, $\hat{\sigma}_{\cdot t|T}$, are derived from the square root of diagonal elements of $\hat{\Sigma}_{t|T}$.

The model which combines a linear, a TAR and an ExpAR model is denoted as L.T.E., and we use the notation L.T.E.(C) and L.T.E(TV) for the constant combination and the time varying combination model, respectively. In each model, a constant term is included because multistep forecasts of nonlinear models do not always produce unbiased predictors.

Further, we compare forecasts for the case of time varying weights with different constant parameter models (L.T.E.(TV)) with forecasts of a linear autoregressive model where the parameters are time varying. That is, we make use of

$$Y_t = \gamma_{0t} + \gamma_{1t} Y_{t-1} + \cdots + \gamma_{pt} Y_{t-p} + \epsilon_t \qquad (6)$$

where the time varying coefficients are assumed to follow a random walk. This model is well-known in the literature; see, for example, Cooley and Prescott (1976), Kitagawa (1981) and Harvey (1989). It can be expressed in a state space form and the model is estimated using a Kalman filter and AIC. Extension to other models with time varying parameters, for example, the model by Lundberg, Teräsvirta and van Dijk (2000) is a topic of further research. We comment on this in our conclusions.

3. Data specification and nonlinearity tests on individual series

We make use of three kinds of data sets. The first set consists of the well known Canadian lynx number series and the sunspot number series in the natural sciences. These data sets have played the role of benchmark for measuring the performance of nonlinear time series models. The second data set is Nelson and Plosser's US macroeconomic time series data, see Nelson and Plosser (1982) and for the extended set Schotman and van Dijk (1991). This data set has motivated the discussion regarding deterministic trends and stochastic trends in economic time series. The third data set are 2-monthly series on US unemployment and industrial production. These series have been used in the econometric literature on nonlinear modeling.

3.1. Canadian lynx series and the sunspot series

Using individual models, the Canadian lynx series were described, for the period 1821–1934, by a linear autoregressive model AR(11) process; a threshold autoregressive process TAR(2; 8, 3); and an exponential autoregressive process ExpAR(1, 11) (see Tong & Lim, 1980; Haggan & Ozaki, 1981; Tong, 1983; Priestley, 1988). The sunspot number series, observed for the period 1720–1989, with the first 221 data used, were described by a linear AR(9); a TAR(2; 3, 11); and an ExpAR(1, 10) (see Tong & Lim, 1980; Haggan and Ozaki, 1980; Subba Rao & Gabr, 1984; Tong, 1990). Henceforth, the notation S1 and S2 is used to denote the lynx and sunspot series, respectively.

3.2. US macroeconomic series

We apply several tests for linearity. All tests use a linear model as a null hypothesis and set some specific nonlinear model as alternative.

Let

$$Y_t = h(Y_{t-1}, Y_{t-2}, \ldots, Y_{t-p}) + e_t \qquad (7)$$

be an autoregressive nonlinear time series model, where $\{e_t\}$ is an i.i.d. process with mean zero. If we assume the innovation e_t as Gaussian, the linearity test is equivalent to a test for Gaussianity. We use five well-known linearity tests: (i) the Ori-F test by Tsay (1986), (ii) the Aug-F test by Luukkonenn, Saikkonen and Teräsvirta (1988), (iii) the CUSUM test by Petruccelli and Davis (1986), (iv) the TAR-F test by Tsay (1989) and (v) the New-F test by Tsay (1988). As noted before, all of these tests have a linear process as a null hypothesis. Based on the Volterra expansion of (7) around $\mathbf{O} = (0, 0, \ldots)'$, we have

$$Y_t = \mu + \sum_{u=1}^{\infty} \psi_u Y_{t-u} + \sum_{u,v=1}^{\infty} \phi_{uv} Y_{t-u} Y_{t-v} + \sum_{u,v,w=1}^{\infty} \phi_{uvw} Y_{t-u} Y_{t-v} Y_{t-w} + \cdots + e_t, \qquad (8)$$

where

$$\mu = h(\mathbf{O}),\ \phi_u = \frac{\partial h}{\partial Y_{t-u}}\bigg|_O,$$

$$\phi_{uv} = \frac{\partial^2 h}{\partial Y_{t-u}\partial Y_{t-v}}\bigg|_O,$$

$$\phi_{uvw} = \frac{\partial^3 h}{\partial Y_{t-u}\partial Y_{t-v}\partial Y_{t-w}}\bigg|_O, \text{ etc.}$$

The Ori-F and Aug-F tests can be used to test against nonlinearity of the second and third order polynomials, respectively. The CUSUM, TAR-F and New-F tests assume the threshold type nonlinear alternatives:

$$Y_t = \beta_0^{(j)} + \sum_{i=1}^{p} \beta_i^{(j)} Y_{t-i} + a_t^{(j)} \ (j = 1, 2), \tag{9}$$

where $\{a_t^{(j)}\}$ is the innovation of mean zero and variance σ_j^2. The New-F test covers the most extensive set of alternatives of nonlinearity, including the ExpAR model. Detailed procedures and distributional properties regarding these tests can be found in Granger and Teräsvirta (1993).

Each model is estimated based on the extensive use of AIC; see Tong and Lim (1980) for similar results on the TAR model and see Haggan and Ozaki (1981) for results on the ExpAR model. The maximum values of the autoregressive part of each model was set as 15, and the particular lag order and the nonlinearity parameters (d, r) were chosen by a conditional least squares method and by using the criterion of minimum AIC. The results are summarized in Table 1.

Each of the linearity tests with a different set of (p, d) brings out different results. Following Cox and Hinkley (1974, p. 104) and Stone (1969), we use the most significant result of the test among all the combinations of (p, d).

3.3. The Nelson–Plosser annual macroeconomic series

Nelson and Plosser's 14 series are annual data starting from different years and ending in 1970. The notation and the starting year of the sample periods are as follows: RGNP (real GNP: 1909–), NGNP (nominal GNP: 1909–), PCRGNP (real per capita GNP: 1909–), IP (industrial production: 1860–), EMP (employment: 1890–), UN (unemployment: 1890–), PRGNP (GNP deflator: 1889–), CPI (consumer prices: 1860–), WG (wages: 1900–), RWG (real wages: 1900–), M (money stock: 1889–), VEL (velocity: 1869–), BND (bond yield: 1900–), SP500 (common stock prices: 1871–). All of the series are assumed to be stationary after taking their first difference here.

From the P-values of all linearity tests for Nelson–Plosser's annual 14 series, we note the following results. (The P-values of each test are not tabulated here to save space. A more extensive version of the paper with additional results is available as Terui and van Dijk (1999).)

- Several series are not significantly different from a linear AR process, in particular, RGNP has P-values greater than 5% for all tests.
- Series with strong nonlinearity are NGNP, PRGNP and CPI.
- The SP500 series is not inconsistent with Gaussian disturbances, except for the New-F tests.
- Among the five tests, the New-F test rejects the null hypothesis of linearity most strongly and the CUSUM test rejects the least. We have some similarity of the results between the Ori-F and Aug-F tests.
- The CUSUM test produces different results from other tests, which might be due to its poor power performance reported by Tsay (1988, 1989). Henceforth, we leave the results of the CUSUM test out of our investigation.
- The null hypotheses of linearity for PCRGNP, IP and SP500 are rejected only by the New-F test and this may imply that these series have a bilinear or ExpAR type of nonlinearity.

Table 1
Estimated marginal models: macroeconomic series

Linear						
Series no.	Name	–	–	p	s	AIC
E1	(A) NGNP	–	–	1	0.1015	−97.6304
E2	(A) PRGNP	–	–	1	0.0690	−136.673
E3	(A) CPI	–	–	2	0.0407	−365.337
M1	(A) M-UN	–	–	12	0.0374	−2207.8
M2	(A) M-IP	–	–	12	0.0121	−4123.9
E4	(B) SP500	–	–	5	0.1566	−77.6440
E5	(B) IP	–	–	5	0.1048	−165.319
E6	(C) VEL	–	–	1	0.0729	−226.550
TAR						
Series no.	Name	d	r	(p_1, p_2)	s	AIC
E1	(A) NGNP	4	0.0855	(4, 10)	0.0801	−35.988
E2	(A) PRGNP	3	0.0092	(10, 10)	0.0436	−55.122
E3	(A) CPI	2	0.0098	(3, 2)	0.0377	−189.828
M1	(A) M-UN	1	0.0000	(15, 3)	0.0368	−1225.7
M2	(A) M-IP	2	0.0042	(12, 13)	0.0118	−2097.3
E4	(B) SP500	1	0.0281	(5, 4)	0.1427	−30.6599
E5	(B) IP	1	−0.0661	(10, 5)	0.0865	−124.392
E6	(C) VEL	3	−0.0134	(3, 6)	0.0626	−98.1974
ExpAR						
Series no.	Name	d	γ	p	s	AIC
E1	(A) NGNP	3	1.576	9	0.0674	−88.0833
E2	(A) PRGNP	1	4.739	10	0.0454	−131.560
E3	(A) CPI	1	0.0410	10	0.0271	−387.764
M1	(A) M-UN	1	1.500	13	0.0362	−2220.1
M2	(A) M-IP	1	1.541	15	0.0107	−4273.8
E4	(B) SP500	1	0.00001	8	0.1223	−81.3575
E5	(B) IP	1	0.00009	9	0.0893	−153.588
E6	(C) VEL	3	0.00001	6	0.0648	−213.878

- The result for VEL is significant solely for the TAR-F test and a threshold type nonlinearity might be appropriate for VEL.

Based on the results of the Ori-F, Aug-F, TAR-F and New-F tests, we classify possible nonlinearity into three classes: (A) highly nonlinear, (B) possibly ExpAR, (C) possibly TAR. Results are reported in Table 1. Firstly, we choose the NGNP, PRGNP and CPI series as highly nonlinear, because all tests reject linearity. Next, we select SP500 and IP as possibly ExpAR series, because only the New-F test rejects the linearity. Finally, only VEL can be dealt with as a possible TAR series, because the New-F test does not reject the linearity but the TAR-F test rejects it.

From these observations, we select the following six, possibly, nonlinear series: NGNP, PRGNP, CPI, SP500, IP, VEL. The notation E1, E2, E3, E4, E5, E6 is used for these series.

3.4. Monthly US unemployment and production index

Next, we apply the linearity tests to 2-monthly economic series: the US unemployment rate (January, 1948–January, 2000) and the US industrial production index (January, 1940–January, 2000). Both series are seasonally adjusted. The series are denoted as M-UN and M-IP, respectively. The *P*-values of linearity tests show that these series can be characterized as highly nonlinear. Henceforth, these monthly series are denoted as M1 and M2, respectively.

4. In sample performance

We have six models for each series, that is, four marginal models and two combined models. Table 2 shows the estimated standard errors of three marginal models as well as the two combined models. Figs. 1 and 2 show observations and estimates for the marginal models and for the two combined models for the lynx series and the SP500 series. We observe that the nonlinear models improve the fit over a linear model. In particular, for macroeconomic series, nonlinear models capture an upswing and/or trough. These changes may be interpreted as a structural change of the economy. Figs. 3 and 4 show the movements of $\{\beta_t\}$ over time for the time varying combination method for the lynx and SP500 series, respectively.

We summarize in Table 3 the results of in-sample performance according to three properties, that is, (i) the standard error of the models, (ii) the role of marginal models in the constant combined model, and (iii) the role of marginal models in the time varying combined model.

4.1. Property 1 (standard errors of the models)

According to the order of estimated standard errors of each model, the examined 10 series (S1, S2, E1, E2, E3, E4, E5, E6, M1, and M2) can be classified into the following two groups:

(a) L.T.E(TV) $<$ L.T.E.(C) $<$ TAR $<$ ExpAR $<$ Linear
(b) L.T.E(TV) $<$ L.T.E.(C) $<$ ExpAR $<$ TAR $<$ Linear

Table 2
Estimated standard errors for marginal and composite models

Series no.	Name	Linear	TAR	ExpAR	L.T.E.(C)	L.T.E.(TV)
S1	Lynx(1)	0.2870	0.1911	0.1978	0.1748	0.0566
	Lynx(2)	0.1918	0.1827	0.1603	0.1521	0.0761
S2	Sunspot	14.392	12.436	13.561	11.982	10.502
E1	(A) NGNP	0.1015	0.0801	0.0674	0.0604	0.0526
E2	(A) PRGNP	0.0690	0.0436	0.0454	0.0344	0.0307
E3	(A) CPI	0.0407	0.0377	0.0271	0.0259	0.0235
M1	(A) M-UN	0.0374	0.0368	0.0362	0.0355	0.0345
M2	(A) M-IP	0.0122	0.0118	0.0107	0.0106	0.0104
E4	(B) SP500	0.1566	0.1427	0.1223	0.1156	0.1070
E5	(B) IP	0.1048	0.0865	0.0893	0.0854	0.0748
E6	(C) VEL	0.0729	0.0626	0.0648	0.0603	0.0450

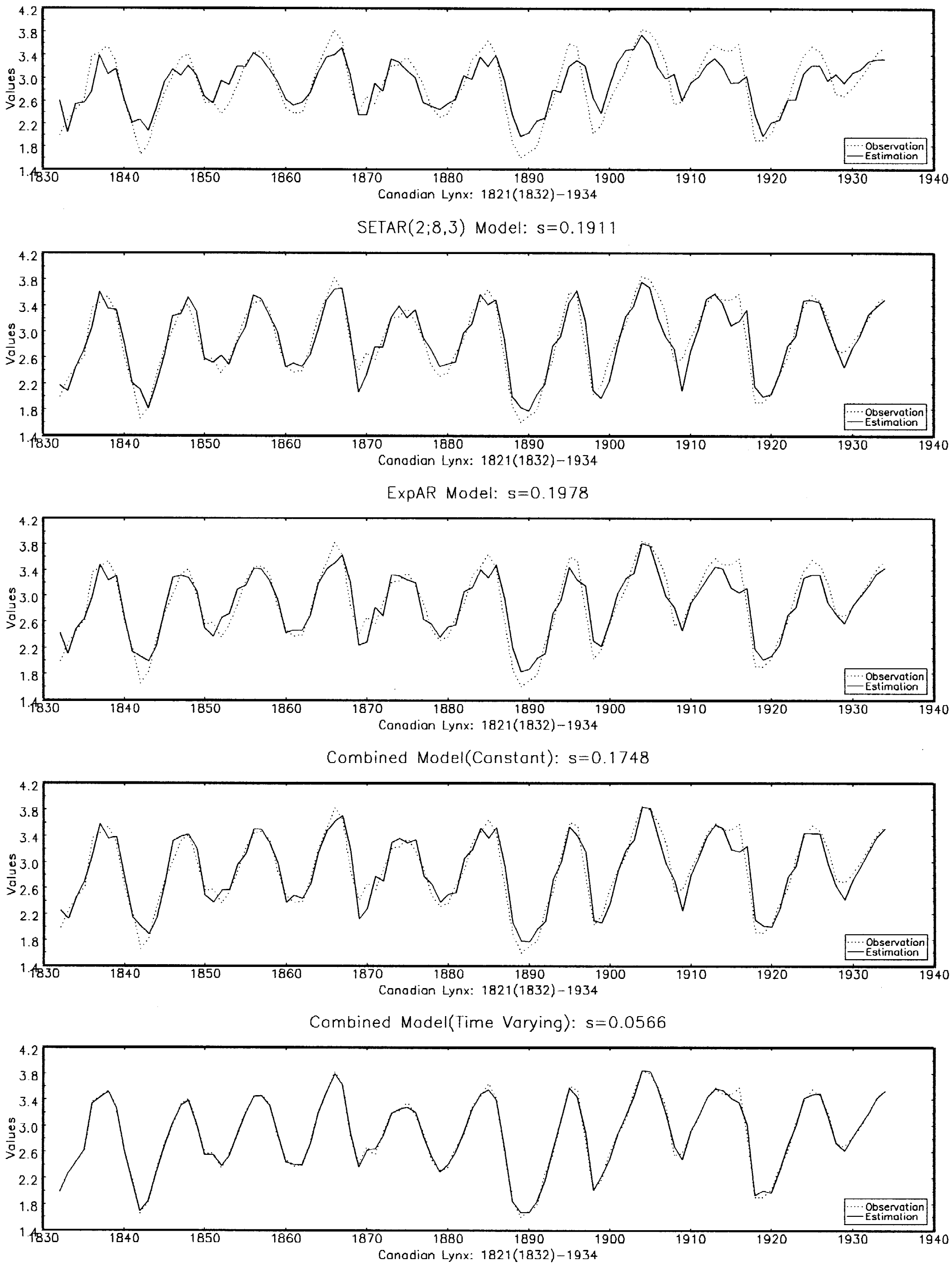

Fig. 1. In-sample performance: Canadian lynx data.

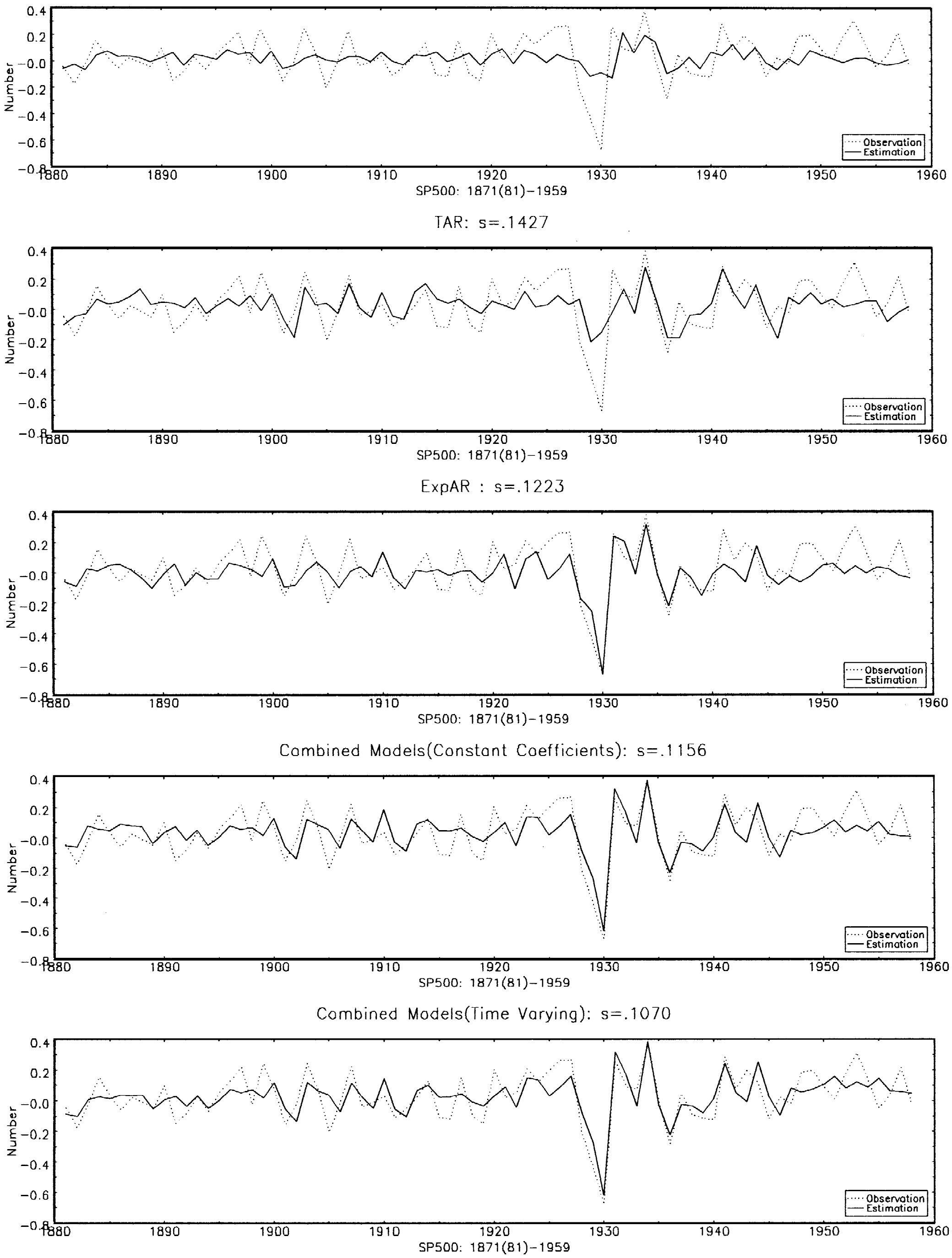

Fig. 2. In-sample performances: SP500.

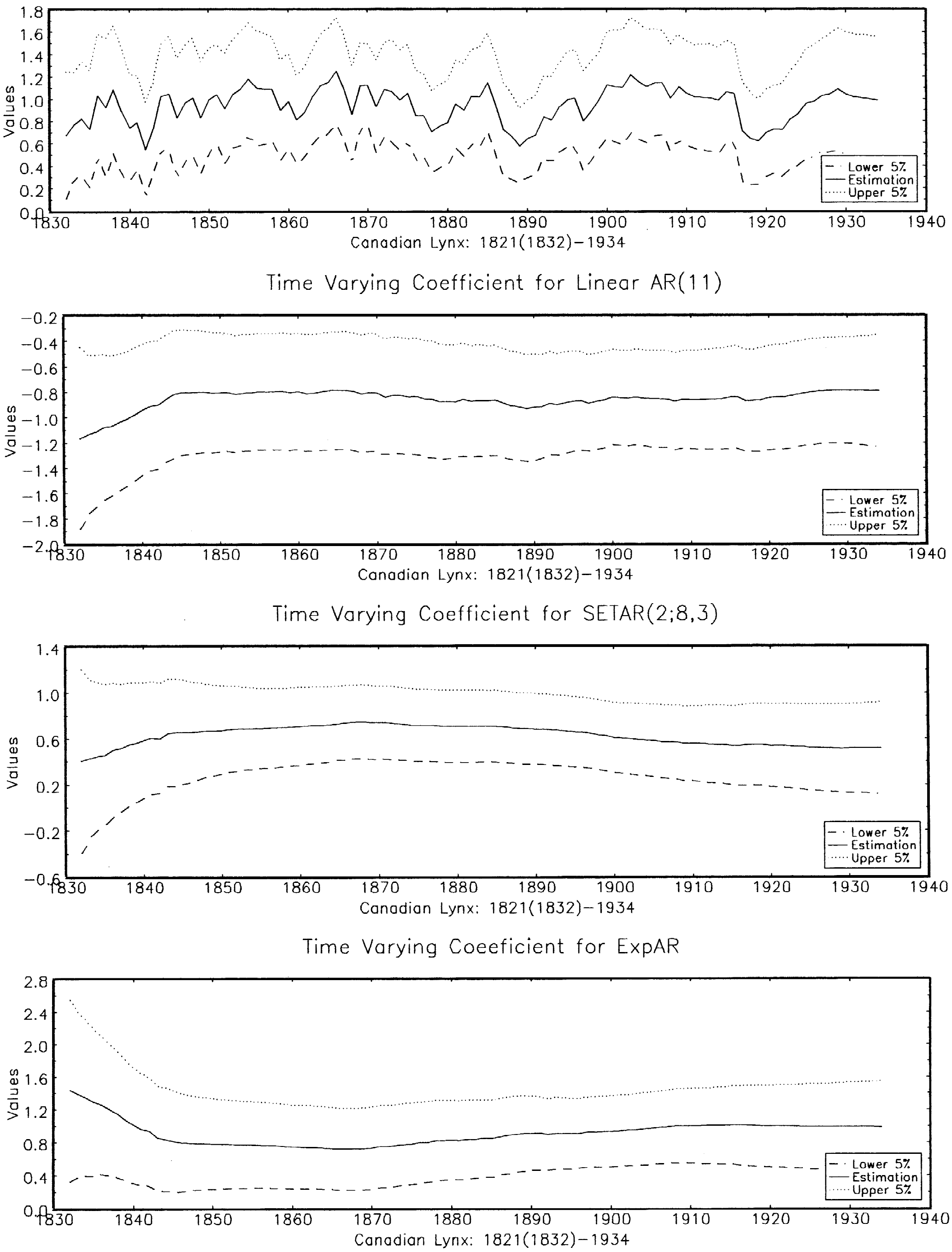

Fig. 3. Contribution of marginal models: Canadian lynx data.

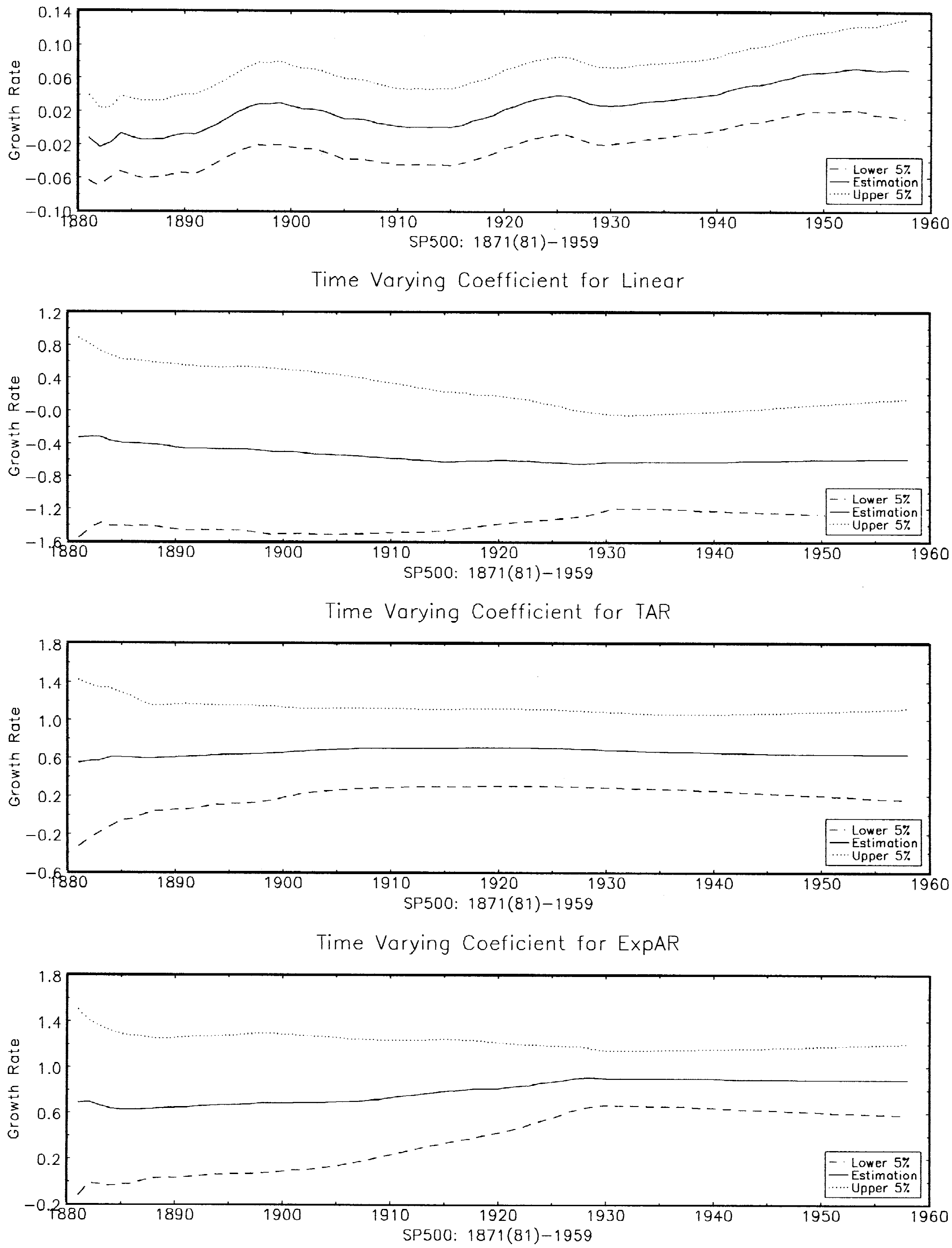

Fig. 4. Contribution of marginal models: SP500.

Table 3
In sample performance

Property 1: standard error (S.E.)	
Order of S.E. of models	Series no.
(a) L.T.V.(TV)<L.T.E.(C)<TAR<ExpAR<Linear	S1, S2, E2, E5, E6
(b) L.T.V.(TV)<L.T.E.(C)<ExpAR<TAR<Linear	E1, E3, M1, M4, E4
Property 2: L.T.E.(C)	
Significance of marginal models	Series no.
(a) Only level not significant	S1, S2, M1, M2
(b) TAR and ExpAR significant	E1, E2, E4, E6
(c) Only TAR significant	E5
(d) All significant	E3
Property 3: L.T.E.(TV)	
Level (const.)	Series no.
(a) Significant at all data points	S1, M2
(b) Significant at some data points	S2, E3, E4, E6
(c) Never significant during all data points	E1, E2, M1, E5
Linear	Series no.
(a) Significant at all data points	S1
(b) Significant at some data points	S2, E2, E6
(c) Never significant during all data points	E1, E3, M1, M2, E4, E5
TAR	Series no.
(a) Significant at all data points	S2, E1, E2, E3, E6
(b) Significant at some data points	S1, M1, M2, E4, E5
(c) Never significant during all data points	–
ExpAR	Series no.
(a) Significant at all data points	S1, E1, E2, E3, E4
(b) Significant at some data points	S2, M1, M2, E5, E6
(c) Never significant during all data points	–

In all cases, the S.E. of L.T.E(TV) is the smallest and that of Linear is the largest and the order between TAR and ExpAR models changes across the series.

4.2. Property 2 (L.T.E.(C): constant combined model)

According to the significance of marginal predictors as well as the level in the constant combined model, 10 series are grouped into four categories: (a) only the level term is not significant; (b) TAR and ExpAR predictors are significant; (c) only the TAR predictor is significant; (d) all predictors are significant.

It is observed that the significances of the level and marginal predictors vary across the series as we can expect.

4.3. Property 3 (L.T.E.(TV): time varying combined model)

The significance of marginal predictors can change over time for a combined model with time varying weights. Corresponding to the significance of the level and of each marginal predictor in the time varying combined model, 10 series are grouped to three categories. (a) The term is significant throughout all data points; (b) the term is significant at some data points; (c) the term is never significant during all data points.

We note that the roles of nonlinear predictors (TAR and ExpAR) are important throughout the sample for all series, and that the level and linear predictor do not contribute much at some periods.

5. Out of sample performance

In this section, we compare the out of sample performance of each predictor generated by the four marginal and two combined models. The optimal predictor in the sense of minimizing the mean squared error criterion is the conditional expectation

$$\tilde{Y}_{t+h} = E\{Y_{t+h}|Y_{t-1},Y_{t-2},\ldots\}. \quad (10)$$

This predictor is, in general, hard to evaluate for nonlinear time series models. Several methods have been proposed in this context. Recently, Clements and Smith (1997) used extensive simulation in order to compare several multistep forecasting methods. They recommend to use the Monte Carlo (MC) method. The multistep forecasts by the MC method for the TAR and the ExpAR models are computed through the following algorithm (see also Franses & van Dijk, 2000).

Define the forecasts of h step ahead as

$$\hat{Y}^{(i)}_{T+h} = h(\hat{Y}^{(i)}_{T+h-1}, \hat{Y}^{(i)}_{T+h-2}, \ldots, \hat{Y}^{(i)}_{T+h-p}) + \eta^{(i)}_h$$
$$i = 1,2,\ldots,M \quad (11)$$

where $\eta^{(i)}_h$ is a pseudo-random number such that, for a TAR model,

$$\eta^{(i)}_h = \begin{cases} \eta^{(i)}_{h(1)} \sim N(0, \hat{\sigma}^2_{(1)}) & \text{for the regime (1)} \\ \eta^{(i)}_{h(2)} \sim N(0, \hat{\sigma}^2_{(2)}) & \text{for the regime (2)} \end{cases} \quad (12)$$

and for an ExpAR model,

$$\eta^{(i)}_h \sim N(0, \hat{\sigma}^2), \quad (13)$$

where $\hat{\sigma}^2_{(i)}$ and $\hat{\sigma}$ mean the estimated standard errors of TAR and ExpAR models, respectively.

Then averaging these forecasts across the $i = 1, 2, \ldots, M$ iterations of the Monte Carlo method produces the predictor

$$Y_{T+h} = \frac{1}{M}\sum_{i=1}^{M} \hat{Y}^{(i)}_{T+h}. \quad (14)$$

After the generation of the marginal predictors, Y^l_{T+h}, Y^t_{T+h}, Y^e_{T+h}, by a recursive way for the linear predictors and by the MC method for the TAR and ExpAR predictors, these predictors are the regressor matrix $\boldsymbol{X}_{T+h}$. Then the composite forecast with constant coefficients is generated by

$$\hat{Y}_{T+h} = \hat{\beta}^0 + \hat{\beta}^l Y^l_{T+h} + \hat{\beta}^t Y^t_{T+h} + \hat{\beta}^e Y^e_{T+h}$$
$$= \boldsymbol{X}_{T+h}\hat{\beta} \quad (15)$$

where $\hat{\beta}$ is the estimated coefficient vector. As for the composite forecast with time varying coefficients, we observe $(Y_1, \ldots, Y_T)$ and we predict β_{T+h} and Y_{T+h}, for $h \geq 1$. That is,

$$\hat{\beta}_{T+h|T} = E(\beta_{T+h}|Y_1, \ldots, Y_T),$$
$$\hat{Y}_{T+h|T} = E(Y_{T+h}|Y_1, \ldots, Y_T)$$

as well as their variances,

$$\hat{\Sigma}_{T+h|T} = E\left[(\beta_{T+h} - \hat{\beta}_{T+h|T})(\beta_{T+h} - \hat{\beta}_{T+h|T})'\right],$$
$$\hat{\sigma}^2_{T+h|T} = E(Y_{T+h} - \hat{Y}_{T+h|T})^2.$$

These quantities are also evaluated by applying the Kalman filter, and the final forms are

$$\hat{Y}_{T+h|T} = \boldsymbol{X}_{T+h}\hat{\beta}_{T+h|T},$$
$$\hat{\beta}_{T+h|T} = \hat{\beta}_{T+h-1|T} \text{ for } h \geq 1. \tag{16}$$

As a measure of predictive performance, we use the root mean squared error (RMSE) of the h step ahead prediction, which is defined, see Franses and van Dijk (2000), as

$$\text{RMSE}(h) = \sqrt{\frac{1}{m}\sum_{j=1}^{m}(\hat{Y}_{t+j|t+j-h} - Y_{t+j})^2}. \tag{17}$$

For all series, we set $M = 2000$ for the MC method. For each sample period, the models are re-estimated. Table 4 summarizes the results of comparing 10 step ahead predictions. Since the combination method with time varying coefficients has the minimum RMSE in many cases, the RMSE(h) of each method is divided by the RMSE of the time varying combination method. The comparison with the time varying AR model, noted as TV-L, is reported in the last column. In Table 4, a number less than one means that the predictors up to h steps ahead have smaller RMSE than those of the time varying combination method.

5.1. Sunspot and lynx series (S1 and S2)

The first and subsequent 221 data points are used to generate multistep ahead prediction for the sunspot series. For the lynx series, the whole data set (1821–1934) has been used. In order to keep data for a predictive performance, we re-estimate each model by leaving the latter 10 observations and we evaluate the succeeding 10 predictions. The identification of each marginal model was conducted by the use of AIC. We set the maximum of the autoregressive order, p, p_1, and p_2 as 15 and we move the delay parameter d from 1 to 5.

The identified AR(p), TAR(p_1, p_2; d, r) and ExpAR(p; d, γ) models, using the first 100 samples, are:

- Linear AR(11); AIC $= -16.418$, $s = 0.19181$.
- TAR(12, 3; 3, 3.328); AIC $= -23.087$, $s = 0.183$.
- ExpAR(12; 3, 3.8); AIC $= -21.167$, $s = 0.160$.

The estimates of the coefficients for the constant combination method exhibit no great difference with estimates using the whole sample. Similarly, we did not find any great differences in the movements of time varying coefficients.

For the sunspot series, S1, the linear model is the best for the first step; however, it is the worst after that. The combined models outperform the other models, particularly, as the prediction step proceeds. The combined model with constant coefficients is producing the best predictors after 10 periods. Note that the forecasts of TV-L are in most cases better than the forecasts of the marginal models. However, they are worse than the forecasts from the combined methods.

For the lynx series, S2, one observes that the time varying combination model shows the best performances during the forecasting periods and the linear model is the worst. Compared among marginal nonlinear models, the TAR model is a little better than the ExpAR model, which is consistent with the values of AIC. For specific marginal models, similar results are observed as for the case of sunspot data. The results for TV-L for the lynx numbers are almost the same as for the sunspot data.

5.2. Macroeconomic series (E1, E2, E3, E4, E5, E6, M1 and M2)

With respect to macroeconomic series, we observe that the combined models, in particular models with varying weights, dominate in many cases over the marginal models. The marginal models show, however, in some cases the best forecasting performance. The ExpAR model for VEL has the best forecasts, and linear forecasts

Table 4
Root mean squared error comparison for predictors

Step	Linear	TAR	ExpAR	L.T.E.(C)	TV-L	Linear	TAR	ExpAR	L.T.E.(C)	TV-L
S1: sunspot number						S2: Canadian lynx				
1	0.285	3.368	2.574	2.698	0.984	1.587	1.457	1.620	0.982	1.026
2	2.452	1.487	1.124	1.006	2.109	1.459	1.423	1.433	0.985	1.023
3	2.214	1.536	1.142	0.922	1.879	1.682	1.432	1.756	0.995	1.145
4	1.534	1.257	1.531	0.932	0.924	1.985	1.654	1.824	1.154	1.348
5	2.859	1.632	1.624	1.241	1.234	1.965	1.968	1.128	1.279	1.259
6	1.384	1.172	1.241	0.986	1.298	2.019	1.697	1.687	1.098	1.572
7	1.597	0.935	1.356	0.978	1.142	1.874	1.653	1.758	1.106	1.236
8	1.496	0.965	1.410	0.910	1.250	1.695	1.576	1.713	1.204	1.456
9	2.652	1.608	1.742	1.135	1.193	1.689	1.756	1.698	1.145	1.369
10	2.542	1.302	1.334	0.952	1.734	1.723	1.856	1.875	1.142	1.765
E1: (A) NGNP						E2: (A) PRGNP				
1	1.120	0.875	1.795	1.198	1.002	0.985	3.698	0.876	1.652	0.982
2	1.215	1.254	1.652	1.202	1.012	0.805	2.254	1.574	1.302	0.812
3	0.985	1.195	1.598	1.198	1.009	1.635	2.445	2.245	1.258	1.112
4	0.652	0.991	1.605	1.168	0.986	1.547	2.694	1.996	1.249	1.256
5	1.025	1.251	1.601	1.142	1.985	1.520	1.823	1.653	1.187	1.220
6	1.634	1.223	1.547	1.165	1.236	0.985	1.687	1.569	1.028	1.309
7	1.632	1.172	1.612	1.174	1.325	0.852	1.652	1.578	0.986	0.937
8	1.578	1.289	1.589	1.169	1.281	0.985	1.679	1.652	1.068	1.095
9	1.352	1.236	1.546	1.201	1.347	1.547	1.502	1.624	0.969	1.235
10	1.652	1.307	1.459	1.168	1.198	1.369	1.129	1.631	0.968	1.195
E3: (A) CPI						M1: (A) M-IP				
1	0.836	0.698	1.574	0.520	1.198	1.503	1.895	1.255	0.968	1.187
2	0.896	3.658	3.521	1.178	1.115	0.997	1.652	1.002	0.954	1.010
3	1.158	2.214	2.024	1.168	1.876	1.256	1.247	1.147	1.133	1.192
4	1.169	2.354	2.875	1.236	1.154	1.255	1.698	1.369	1.658	1.276
5	0.854	1.965	2.254	1.247	0.989	1.347	1.547	1.854	1.741	1.023
6	0.857	1.875	1.965	1.157	0.982	1.209	1.214	1.654	1.256	1.235
7	0.965	2.031	2.247	1.100	1.176	1.258	1.105	1.278	1.020	1.113
8	0.147	1.658	2.360	0.984	0.965	1.157	1.209	1.964	1.470	1.454
9	1.320	1.758	1.877	1.002	1.287	1.196	1.188	1.240	0.987	1.234
10	1.210	1.345	1.965	0.958	1.653	0.974	1.014	1.169	0.965	1.109
M2: (A) M-UN						E4: (B) SP500				
1	1.003	1.073	1.024	0.907	0.989	0.965	0.987	0.987	1.001	0.964
2	1.368	1.020	0.899	1.041	1.212	0.954	1.024	0.871	0.941	0.998
3	0.965	1.214	1.021	1.180	1.165	1.047	1.068	0.957	1.023	0.981
4	1.084	1.854	1.124	1.109	1.176	1.195	1.203	1.024	1.001	1.001
5	0.828	1.025	1.365	1.116	1.098	1.241	1.264	1.147	1.041	0.765
6	1.352	1.254	1.357	1.150	1.221	1.118	1.136	1.144	1.056	0.965
7	1.325	1.068	1.467	1.164	1.343	1.116	1.147	1.179	1.128	1.122
8	1.365	1.294	1.654	1.281	1.236	1.174	1.189	1.196	1.247	1.132
9	1.573	1.328	1.234	1.361	1.348	1.157	1.182	1.281	1.354	0.991
10	1.529	1.285	1.913	1.161	1.453	1.241	1.588	0.954	1.024	1.176

Table 4. Continued

Step	Linear	TAR	ExpAR	L.T.E.(C)	TV-L	Linear	TAR	ExpAR	L.T.E.(C)	TV-L
E5: (B) IP						E6: (C) VEL				
1	1.987	2.354	2.125	1.455	1.086	0.954	1.245	1.742	0.954	0.944
2	1.861	1.454	2.201	1.124	1.073	1.472	0.924	1.125	0.824	0.992
3	1.724	1.214	2.962	1.121	1.281	1.652	0.636	0.654	0.845	0.974
4	1.087	0.987	1.987	0.958	0.916	1.247	0.947	0.687	0.865	1.017
5	0.987	0.857	1.657	0.998	0.901	1.457	0.902	0.554	0.802	0.912
6	1.012	0.903	1.874	0.963	0.964	1.487	0.854	0.502	0.721	0.989
7	1.014	0.812	1.759	0.974	0.876	1.465	0.802	0.403	0.658	0.793
8	0.998	0.874	1.800	0.965	0.962	1.404	0.798	0.398	0.654	0.851
9	0.967	0.789	1.763	0.921	0.915	1.585	0.765	0.357	0.630	1.021
10	1.032	0.863	1.642	0.963	0.921	1.402	0.784	0.364	0.701	0.992

The number means each RMSE divided by corresponding RMSE of L.T.E.(TV). TV-L means time varying linear model.

are useful during some forecasting steps for several series. These observations are expected since the economic series exhibit structural changes of the economy. Therefore, although uniform dominance of the combination methods does not hold for the economic time series, we have many situations where the combination methods produce better forecasts.

From Table 4, we have the following particular observations. Firstly, for the highly nonlinear series (NGNP, CPI, PRGNP, M-IP, M-UN), the composite forecasts perform generally better than the marginal model forecasts. On the other hand, for possibly ExpAR(SP500, IP) and for possibly TAR(VEL) series, the composite forecasts do not produce better forecasts than the marginal forecasts. Note that the SP500 series might belong to the highly nonlinear series because the *P*-value of TAR-F test is 0.05048, which is significant at a little more than 5%.

Clements and Smith (1999) investigated the multistep forecast performances of a number of empirical TAR models that have been proposed in the literature, and they concluded that the TAR models produce better forecasts, unless the TAR forecast model is capturing nonlinearities (outlier, non TAR type nonlinearities) which cannot be exploited for forecasting. Their observations are consistent with our results. That is, for highly nonlinear series, composite forecasts perform better than marginal forecasts. On the other hand, for possibly TAR series and possibly ExpAR series, the marginal forecasts show a relatively better performance than composite forecasts. There are cases where the above statements do not hold, but we interpret these cases as exhibiting nonlinearity caused by outliers or as exhibiting other types of nonlinearity, which do not persist into the future. The comparison with TV-L shows that the combined models behave mostly better than TV-L for the data (A), however, it does not always hold for the data (B) and (C).

Summarizing our results for all 10 series, we conclude that the composite forecasts perform well for highly nonlinear series.

6. Conclusion

In this paper, we investigated combinations of forecasts generated by linear and some nonlinear models using a constant coefficient regression method as well as time varying method. The time varying method makes it possible to provide a locally linear (or nonlinear) model.

It is shown that the combined forecasts

perform well, especially, the method with time varying coefficients dominates marginal forecasts for inside sample performance. This results holds also for out-of-sample performance for the sunspot and the Canadian lynx number series, but does not uniformly hold for economic series. Forecast comparison with the case of a linear autoregressive marginal model with time varying coefficients indicates that combining models is in many cases a better strategy. More research is needed to verify the robustness of this result.

Clements and Smith (1999) state that nonlinear models have an edge in certain states of nature but not in others, and that this can be highlighted by evaluating forecasts conditional on the regime, and they discuss that the lack of forecast gain of nonlinear models over linear models is often explained in terms of a failure of the nonlinearity to persist into the future. De Gooijer and Kumar (1992) report that there is no clear evidence in favor of nonlinear models over linear models in terms of forecast performance. Our results indicate that the inclusion of nonlinear models in a set of models appears a good strategy. However, no uniform dominance is observed for all series.

We end by stating some problems for future research. Firstly, the field of constructing flexible time varying nonlinear time series models is rapidly advancing; see e.g. the models analyzed by Lundberg et al. (2000). Therefore a systematic study on the value of the combination of forecasts of simple models compared with the forecasts of a very flexible nonlinear time series model is of considerable interest. The robustness with respect to outliers and varying volatility should be analyzed in this context. Further, nonlinear models such as, for example, bilinear models (Subba Rao & Gabr, 1984) and random coefficient models (Nicholls & Quinn, 1982) may be considered. Our purpose here is to demonstrate the usefulness of the principle of combining linear and nonlinear models for forecasting. Secondly, a more decision theoretic analysis of the proposed method may be investigated, in particular, a Bayesian approach; see Geweke and Terui (1991, 1993). Thirdly, an extensive simulation study and the use of other forecasting measures, like forecast encompassing tests, may be investigated.

Acknowledgements

This research was started while the first author was visiting the Tinbergen Institute at Erasmus University Rotterdam. The first author acknowledges financial support from NWO (Netherlands Organization for Scientific Research), JSPS (Japan Society for Promotion of Science) and the Japanese Ministry of Education (Scientific Research Grant no. (C)10630020). We thank Timo Teräsvirta, Peter Schotman, Philip Hans Franses and two anonymous referees for very useful comments on earlier drafts. The careful reading by the editor, Prof. Keith Ord, is also much appreciated. Remaining errors are entirely ours.

References

Bates, J. M., & Granger, C. W. J. (1969). The combination of forecasts. *Operational Research Quarterly*, *20*, 451–468.

Cooley, T. F., & Prescott, E. (1976). Estimation in the presence of stochastic parameter variation. *Econometrica*, *44*, 167–184.

Cox, D. R., & Hinkley, D. V. (1974). *Theoretical Statistics*. London: Chapman-Hall.

Clements, M. P., & Smith, J. (1997). The performance of alternative forecasting methods for SETAR models. *International Journal of Forecasting*, *13*, 463–475.

Clements, M. P., & Smith, J. (1999). A Monte Carlo study of the forecasting performance of empirical SETAR models. *Journal of Applied Econometrics*, *14*, 123–141.

Crane, D. B., & Crotty, J. R. (1967). A two-stage forecasting model: exponential smoothing and multiple regression. *Management Science*, *13*, B501–507.

Franses, P. H., & van Dijk, D. (2000). *Non-linear Time Series Models in Empirical Finance*. Cambridge University Press.

De Gooijer, J. D., & Kumar, K. (1992). Some recent developments in non-linear time series modelling, testing, and forecasting. *International Journal of Forecasting*, *8*, 135–156.

Granger, C. W. J., & Ramanathan, R. (1984). Improved methods of forecasting. *Journal of Forecasting*, *3*, 197–204.

Granger, C. W. J., & Teräsvirta, T. (1993). *Modelling Nonlinear Economic Relationships*. New York: Oxford University Press.

Geweke, J., & Terui, N. (1991). Threshold autoregressive models for macroeconomic time series: a Bayesian approach. In *Proceedings of American Statistical Meeting (Business and Economics Section)*, pp. 42–50.

Geweke, J., & Terui, N. (1993). Bayesian threshold autoregressive models for nonlinear time series. *Journal of Time Series Analysis*, *14*, 441–451.

Haggan, V., & Ozaki, T. (1980). Amplitude-dependent exponential AR model fitting for non-linear random vibrations. In Anderson, O. D. (Ed.), *Time Series*. Amsterdam: North-Holland, pp. 57–71.

Haggan, V., & Ozaki, T. (1981). Modeling non-linear random vibrations using an amplitude-dependent autoregressive time series model. *Biometrika*, *68*, 189–196.

Hamilton, J. D. (1994). *Time Series Analysis*. New Jersey: Princeton University Press.

Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. New York: Cambridge University Press.

Kitagawa, G. (1981). A nonstationary time series model and its fitting by a recursive filter. *Journal of Time Series Analysis*, *2*, 103–116.

Lundberg, S., Teräsvirta, T., & van Dijk, D. (2000). *Time-varying smooth transition autoregressive models*. Stockholm School of Economics, SSE/EFI working paper no. 376.

Luukkonen, R., Saikkonen, P., & Teräsvirta, T. (1988). Testing linearity against smooth transition autoregressive models. *Biometrika*, *75*, 491–499.

Nelson, C. R., & Plosser, C. I. (1982). Trends and random walks in macro-economic time series. *Journal of Monetary Economics*, *10*, 139–162.

Nicholls, D. F., & Quinn, B. G. (1982). *Random Coefficient Autoregressive Models: an Introduction*. New York: Springer-Verlag.

Petruccelli, J., & Davis, N. (1986). A portmanteaus test for self-exciting threshold autoregressive-type nonlinearity in time series. *Biometrika*, *73*, 687–694.

Priestley, M. B. (1988). *Nonlinear and Nonstationary Time Series Analysis*. London: Academic Press.

Reinmuth, J. E., & Geurts, M. P. (1976). A multideterministic approach to forecasting. In Makridakis, S., & Wheelwright, S. C. (Eds.), *Forecasting*, *TIMS Studies in the Management Sciences*, vol. 12. New York: North-Holland, pp. 203–211.

Schotman, P. C., & van Dijk, H. K. (1991). On Bayesian routes to unit routes. *Journal of Applied Econometrics*, *6*, 387–401.

Stone, C. (1969). The role of significant testing: some data with a message. *Biometrika*, *56*, 485–493.

Subba Rao, T., & Gabr, M. M. (1984). *An Introduction to Bispectral Analysis and Bilinear Time Series Models*. Berlin: Springer-Verlag.

Terui, N., & Kariya, T. (1997a). Gaussianity and nonlinearity of foreign exchange rate. In Dodge, Y. (Ed.), *L-1 Statistical Procedures and Related Topics*, *IMS Monograph Series*, Vol. No. 31. Springer-Verlag, pp. 337–350.

Terui, N., & Kariya, T. (1997b). Testing Gaussianity and linearity of Japanese stock returns. *Financial Engineering and the Japanese Markets*, *4*, 1–30.

Terui, N., & van Dijk, H. K. (1999). Combined forecasts from linear and nonlinear time series models. In *Report 9949/A*. Econometric Institute, Erasmus University Rotterdam, p. 32.

Tong, H., & Lim, K. S. (1980). Threshold autoregression, limit cycle and cyclical data (with discussion). *Journal of the Royal Statistical Society*, *B42*, 245–292.

Tong, H. (1983). *Threshold Models in Non-linear Time Series Analysis*. New York: Springer-Verlag.

Tong, H. (1990). *Non-linear Time Series: A Dynamical System Approach*. New York: Oxford University Press.

Tsay, R. S. (1986). Nonlinearity tests for time series. *Biometrika*, *73*, 461–466.

Tsay, R. S. (1988). Nonlinear time series analysis of Blowfly population. *Journal of Time Series Analysis*, *9*, 246–263.

Tsay, R. S. (1989). Testing and modeling threshold autoregressive process. *Journal of the American Statistical Association*, *84*, 231–240.

Biographies: Nobuhiko TERUI is Professor in the Graduate School of Economics and Management at Tohoku University. His major interests are in statistical analysis of time series, Bayesian statistical modeling and their applications to economics and marketing research.

Herman K. VAN DIJK is Professor of econometrics and director of the Econometric Institute at Erasmus University Rotterdam. His research interest includes Bayesian inference and decision analysis using simulation methods, dynamic econometric modeling, and neural networks in economics.