

Econometric Institute Report EI-9949/A

Combined Forecasts from Linear and Nonlinear Time Series Models ¹

Nobuhiko Terui

Faculty of Economics, Tohoku University, Japan

and

Herman K. van Dijk

Econometric and Tinbergen Institutes Rotterdam, Erasmus University, The Netherlands

October, 1998 (Revised: November, 1999)

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 model. The methods are applied to data from two kinds of disciplines: the Canadian lynx and sunspot series from the natural sciences, and Nelson-Plosser's U.S. series from economics. 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 and Canadian lynx number series, but it does not uniformly hold for economic time series.

Keywords and phrases

Combining forecasts, ExpAR model, Locally linear(or nonlinear) modeling, Threshold model, Time varying coefficient model

¹This 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 and two anonymous referees for very usefull comments on an earlier draft. Remaining errors are entirely ours.

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 regression method as well as time varying regression method.

There are several reasons to consider the proposed methods. First, 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; p.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.

Second, 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 and c) 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.

Third, 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 note that the combination

of other types of nonlinear models is straightforward.

In our empirical application regarding the Nelson-Plosser series, we first test the linearity of each series. We choose six series which are appropriate to be examined by our approach. 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 $\text{TAR}(p_1, p_2; d, r)$, is defined as

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

where $\{\epsilon_t^{(i)}\}$, $i=1,2$, is the innovation process for each regime. 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 $\text{ExpAR}(p; d, \gamma)$ is defined as

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

where $\{Y'_t\}$ is a mean deleted process and $\{\epsilon_t\}$ is an innovation process. This model can be interpreted as a smoothly switching model between two extreme regimes, according to the magnitude of the amplitude $|Y'_{t-1}|$.

Estimating the TAR and ExpAR models; 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).

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:

$$Y_t = \beta^0 + \beta^l Y_t^l + \beta^t Y_t^t + \beta^e Y_t^e + u_t \quad (3)$$

Time Varying Combination:

$$\begin{aligned} Y_t &= \beta_t^0 + \beta_t^l Y_t^l + \beta_t^t Y_t^t + \beta_t^e Y_t^e + v_t \\ &\equiv \mathbf{X}_t \boldsymbol{\beta}_t + v_t; \quad v_t \sim N(0, \eta^2) \end{aligned} \quad (4)$$

$$\boldsymbol{\beta}_t = \boldsymbol{\beta}_{t-1} + \mathbf{e}_t; \quad \mathbf{e}_t \sim N(\mathbf{0}, \Sigma) \quad (5)$$

where Y_t^l , Y_t^t and Y_t^e 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; $\mathbf{X}_t = [1, Y_t^l, Y_t^t, Y_t^e]$; and $\beta_t = (\beta_t^0, \beta_t^l, \beta_t^t, \beta_t^e)'$.

The time varying combined model - equations (4) and (5) - is a state space model, where equation (4) is the measurement equation which defines the distribution of Y_t , $t \geq 1$; and where equation (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,

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

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

$$\begin{aligned} \hat{\beta}_{t|t-1} &= E(\beta_t | Y_1, \dots, Y_{t-1}), \quad \hat{Y}_{t|t-1} = E(Y_t | Y_1, \dots, Y_{t-1}) \\ \hat{\Sigma}_{t|t-1} &= E[(\beta_t - \hat{\beta}_{t|t-1})(\beta_t - \hat{\beta}_{t|t-1})'], \quad \hat{\sigma}_{t|t-1}^2 = E(Y_t - \hat{Y}_{t|t-1})^2 \end{aligned}$$

are essential. We refer for details to Harvey(1989) and Hamilton(1994). Next, suppose we have T observations (Y_1, Y_2, \dots, Y_T) and we want to determine the optimal inside sample estimator $\hat{\beta}_{t|T} = E(\beta_t | Y_1, \dots, Y_T)$ 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 $\hat{\beta}_{.t|T}$ and $(\hat{\beta}_{.t|T} \pm 1.64\hat{\sigma}_{.t|T})$ are drawn in our figures, where $\hat{\beta}_{.t|T}$ means the smoothing estimate of each β^0 , β^l , β^t , β^e and the standard deviations for the marginal predictors, $\hat{\sigma}_{.t|T}$, are derived from the square root of diagonal elements of $\hat{\Sigma}_{t|T}$.

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.

3 The relative contribution of each model: Within Sample Performance

We make use of two 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 U.S. 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.

3.1 Canadian Lynx and Sunspot Number Series

Using individual models, the Canadian lynx series, for the period 1821-1934, was estimated as a linear autoregressive model AR(11)process [see Tong(1983) and Priestley(1981), p.386]; as a threshold autoregressive process TAR(2; 8, 3)[see Tong and Lim(1980) and Priestley(1988), p.82], and as an exponential autoregressive process ExpAR(1, 11)[see Haggan and Ozaki(1981) and Priestley(1988), p.89].

The sunspot number series observed from the year 1720 through 1989, with the first 221 data used, was individually estimated as linear AR(9)(see Subba Rao and Gabr(1984), p.196 and Tong(1990), p.427); as a TAR(2; 3, 11)(see Tong and Lim(1980); and Tong(1990), p.425); and as an ExpAR(1, 10)(Haggan and Ozaki(1980)).

For both series, figures 1-1 and 1-2 show the observations and their estimates for each identified marginal model and for two kinds of combined models. Table 1 shows the results of estimation using the constant combination method, and figures 2-1 and 2-2 show the values of coefficients of the time varying combination method. Table 2 shows the estimates of the standard errors of each model. Given these results, we draw the following three conclusions:

Canadian Lynx Data

- (1) The models are ordered with increasing order of standard errors as follows: L.T.E.(TV) < L.T.E.(C) < TAR < ExpAR < Linear.
- (2) L.T.E.(C): Only the level estimate is not significant.
- (3) L.T.E.(TV):
 - The level is significant at the 5% level at every data point.
 - The movement of the estimate of the level is similar for each observation. This suggests that yet another model than a linear, a TAR or an ExpAR, may be effective.
 - The estimates of the coefficient for the predictors of the TAR and the ExpAR models are significant at the 5% level at each data point.
 - The estimates of the coefficients for the TAR model become significant at the 5% level after the year 1864.

Sunspot Number Data

- (1) The models are ordered with increasing order of standard errors as follows: L.T.E.(TV) < L.T.E.(C) < TAR < ExpAR < Linear.
- (2) L.T.E.(C): Only the level estimate is not significant.

(3) L.T.E.(TV):

- The level is not significant at a 5% level except for a few points, *i.e.*, 1723 and the years 1807-1819.
- The estimates of the coefficients for the linear model are significant at the 5% level until the year 1762.
- The estimates of the coefficients for the predictors of TAR are significant at the 5% level for every data point.
- The estimates of coefficient for the ExpAR model become significant at the 5% level after the year 1756.

We also considered three other cases of combined forecasts of two models: Linear and TAR; Linear and ExpAR; and TAR and ExpAR. In order to save space we summarize the results as follows: for both data sets the combined forecasts of the three cases with time varying coefficients show better performance than the constant coefficient method. Regarding the Canadian lynx data, the combined Linear and ExpAR model with time varying coefficients, and the combined Linear and TAR model perform well. This suggests that a linear model plays a complementary role for this data set. For the sunspot data, the TAR and ExpAR combined model performs well. The TAR model has more weights at every data point than the ExpAR model. This implies that this data set exhibits explicit nonlinear characteristics, which can be explained more by a TAR model. Furthermore, comparing nonlinear models, we observe mutual complementary roles of these models for the Canadian lynx data, because the TAR model plays an important role in the first part of the period and the ExpAR model become significant in the last period.

3.2 Macroeconomic Time Series

3.2.1 Nonlinearity of the Nelson-Plosser Series

Before making combined forecasts for the Nelson-Plosser 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

$$Y_t = h(Y_{t-1}, Y_{t-2}, \dots, Y_{t-p}) + e_t \quad (6)$$

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 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 Petruccielli and Davis(1986), (iv) the TAR-F test by Tsay(1989) and (v) the New-F test by Tsay(1988).

All of these tests set up, as a null hypothesis, a linear process. Based on the Volterra expansion of (6) around $\mathbf{O} = (0, 0, \dots)'$, we have

$$Y_t = \mu + \sum_{u=1}^{\infty} \psi_u Y_{t-u} + \sum_{u,v=1}^{\infty} \phi_{uv} Y_{t-u} Y_{t-v} + \sum_{u,v,w=1}^{\infty} \phi_{uvw} Y_{t-u} Y_{t-v} Y_{t-w} + \dots + e_t, \quad (7)$$

where

$$\mu = h(\mathbf{O}), \phi_u = \left. \frac{\partial h}{\partial Y_{t-u}} \right|_{\mathbf{O}}, \phi_{uv} = \left. \frac{\partial^2 h}{\partial Y_{t-u} \partial Y_{t-v}} \right|_{\mathbf{O}}, \phi_{uvw} = \left. \frac{\partial^3 h}{\partial Y_{t-u} \partial Y_{t-v} \partial Y_{t-w}} \right|_{\mathbf{O}}, \text{ etc.}$$

The Ori-F and Aug-F tests detect against the nonlinearity of the second and third order polynomials respectively. The CUSUM, TAR-F and New-F tests assume the threshold type nonlinear alternatives;

$$Y_t = \beta_0^{(j)} + \sum_{i=1}^p \beta_i^{(j)} Y_{t-i} + a_t^{(j)} \quad (j = 1, 2), \quad (8)$$

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).

In order to implement the tests, the order p of the autoregression has to be determined for all cases and the value of delay parameter d needs to be specified for the tests (iii), (iv), and (v). We set the maximum of p as 10 and let d run from 1 to 10. 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) .

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.

The p-values of all linearity tests for Nelson-Plosser's annual 14 series are tabulated in table 3. We note the followings results.

- There are several series which are not significantly different from a linear AR process. In particular, RGNP has p-values greater than 5% for all tests.
- The series with strong nonlinearity are NGNP, PRGNP and CPI.

- The annual series of SP500 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.
- 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: Highly nonlinear, Possibly ExpAR, Possibly TAR. First, 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 pick up the following six, possibly, nonlinear series: NGNP, PRGNP, CPI, SP500, IP, VEL.

3.2.2 Combined Models

Linear AR, TAR and ExpAR models are estimated using the six series. Results are summarized in table 4. We note that the estimation of each marginal model was based on the 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 10, 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 data and their estimates of SP500 and NGNP are shown in figures 3-1 and 3-2.

From Table 4, we observe that the nonlinear models improve the fit over a linear model, in particular, they catch up with the sudden change or trough, which could be interpreted as a structural change of the economy. Table 5 shows the result of the constant combination method, and table 6 shows the estimated standard errors of the marginal models and the two combining methods. Figure 4-1 and 4-2 show the movements of the time varying

combination method for the SP500 and the NGNP series. We summarize the conclusions as follows:

Highly Nonlinear Series:

NGNP

1. The ExpAR model dominates the individual models.
2. The order of the estimated standard error of each model is $L.T.E.(TV) < L.T.E.(C) < \text{ExpAR} < \text{TAR} < \text{Linear}$.
3. L.T.E.(C): TAR and ExpAR models are significant.
4. L.T.E.(TV):
 - (a) Linear model is not significant for all data points.
 - (b) The TAR model is significant for all data points and the mean values are gradually increasing.
 - (c) The ExpAR model is significant for all sample periods and the mean values are gradually decreasing.
 - (d) The constant term is not significant for all data points.

PRGNP

1. The TAR model dominates the individual models.
2. The order of the estimated standard error of each model is $L.T.E.(TV) < L.T.E.(C) < \text{TAR} < \text{ExpAR} < \text{Linear}$.
3. L.T.E.(C): The TAR and ExpAR models are significant.
4. L.T.E.(TV):
 - (a) The linear model is significant for the first part until the year 1927.
 - (b) The TAR model is significant for all sample periods and the mean values are gradually decreasing.
 - (c) The ExpAR model is significant for all sample periods and the mean values are gradually increasing.
 - (d) The constant term(level) is insignificant for all sample periods.

CPI

1. The ExpAR model dominates the individual models.
2. The order of the estimated standard error of each model is $L.T.E.(TV) < L.T.E.(C) < \text{ExpAR} < \text{TAR} < \text{Linear}$.
3. L.T.E.(C): Every component is significant.
4. L.T.E.(TV):
 - (a) The linear model is not significant for all sample points.
 - (b) The TAR model is significant for all sample points.
 - (c) The ExpAR model is significant for all sample points.
 - (d) The constant term(level) is significant for 1871-73, 1897-1915 and the 1935-45 years.

ExpAR series:

SP500

1. The ExpAR model dominates the individual models.
2. The order of the estimated standard error of each model is $L.T.E.(TV) < L.T.E.(C) < \text{ExpAR} < \text{TAR} < \text{Linear}$.
3. L.T.E.(C): constant and linear terms are not significant.
4. L.T.E.(TV):
 - (a) The linear model is not significant over all data points.
 - (b) The TAR model is significant for the first 5 years.
 - (c) The ExpAR model keeps significant over all periods.
 - (d) Although constant term(level) is insignificant except for the first 3 years, the mean values are slowly increasing with cyclical movement.

IP

1. TAR model dominates the individual models.
2. The order of estimated standard errors of each models is $L.T.E.(TV) < L.T.E.(C) < \text{TAR} < \text{ExpAR} < \text{Linear}$.
3. L.T.E.(C): Only TAR model is significant.

4. L.T.E.(TV):
 - (a) Linear model is insignificant for all sample periods.
 - (b) TAR model begins significant since 1903 year and its confidence bound is relatively narrow compared with those of ExpAR and Linear models.
 - (c) ExpAR model starts significant since 1906 year.
 - (d) Constant term(level) is insignificant for all sample periods.

TAR series:

VEL

1. The TAR model dominates the individual models.
2. The order of estimated standard errors of each models is L.T.E.(TV) < L.T.E.(C) < TAR < ExpAR < Linear.
3. L.T.E.(C): constant and linear terms are not significant.
4. L.T.E.(TV):
 - (a) The linear model is significant for the first part until 1914.
 - (b) The TAR model is significant for all sample points.
 - (c) The ExpAR model is significant as off the year 1903.
 - (d) The constant term is significant for the first part until the year 1914.

4 Out of Sample Performance

In this section, we compare the out of sample performance of each predictor. The optimal predictor in the sense of minimizing mean squared error criterion is the conditional expectation

$$\tilde{Y}_{t+h} = E\{Y_{t+h} \mid Y_{t-1}, Y_{t-2}, \dots\}, \quad (9)$$

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. We make use of the SK(Skelton) method proposed in Tong and Lim(1980). This method has been applied for TAR models and can be extended in a standard way to ExpAR models. A Bayesian procedure was applied by Geweke and Terui(1991, 1993) and Terui(1992). A formal Bayesian analysis of our method is left for further work.

The multistep forecasts by the SK method for the TAR and the ExpAR models are computed through the following algorithm:

1. if $h \leq d$, use recursive relations just like in the linear process,

Now given T , forecast the h step ahead value as

$$\hat{Y}_{T+h} = \begin{cases} Y_{T+h} & \text{if } h \leq 0 \\ h(\hat{Y}_{T+h-1}, \hat{Y}_{T+h-2}, \dots, \hat{Y}_{T+h-p}) & \text{if } h = 1, 2, \dots \end{cases} \quad (10)$$

2. otherwise, replace $\{Y_{T+h-d}\}$ by the forecast $\{\hat{Y}_{T+h-d}\}$.

After the generation of the marginal predictors, $Y_{T+h}^l, Y_{T+h}^t, Y_{T+h}^e$, by a recursive way for the linear predictors and by the SK method for the TAR and ExpAR predictors, these predictors are the regressor matrix \mathbf{X}_{T+h} . Then the composite forecast with constant coefficients is generated by

$$\begin{aligned} \hat{Y}_{T+h} &= \hat{\beta}^0 + \hat{\beta}^l Y_{T+h}^l + \hat{\beta}^t Y_{T+h}^t + \hat{\beta}^e Y_{T+h}^e \\ &= \mathbf{X}_{T+h} \hat{\beta} \end{aligned} \quad (11)$$

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

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

as well as their variances,

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

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

$$\hat{Y}_{T+h|T} = \mathbf{X}_{T+h} \hat{\beta}_{T+h|T}, \quad \hat{\beta}_{T+h|T} = \hat{\beta}_{T+h-1|T} \quad \text{for } h \geq 1. \quad (12)$$

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

$$\text{RMSE}(h) = \sqrt{\frac{1}{h} \sum_{k=1}^h (\hat{Y}_{t+k} - Y_{t+k})^2}. \quad (13)$$

4.1 Sunspot Number Data and Canadian Lynx Data

First, using the observations up to 1920, we generate multistep ahead predictions for the number of sunspots for the years 1921 through 1989. Table 7 summarizes the results of comparing 25 step ahead predictions. Since the combination method with time varying coefficients has the minimum RMSE in most cases, the $\text{RMSE}(h)$ of each method is divided by the RMSE of the time varying combination method. In table 7 a number less than one means that the predictors up to h step ahead have smaller RMSE than those of the

time varying combination method. The asterisk is attached to the one with the smallest RMSE. In case no asterisk is attached in a row, it means that the best predictor is one based on the combination method with time varying coefficients.

We draw several conclusions from the results in table 7. First, the linear model is the best for the first step; however, it is the worst after that. Secondly, comparing the nonlinear models, the ExpAR model dominates the TAR model after the year 1957. There is no clear difference before that year. Furthermore, the time varying coefficient model dominates the constant coefficient model after the year 1936. There is no clear difference before that. Finally, the order of standard deviations of the prediction errors from small to large is as follows: time varying combination, the constant combination, the ExpAR, the TAR and the linear models.

With respect to marginal models, we observe that the predictors generated by the linear model die out as time proceeds and that the TAR predictors generate an asymmetric cycle. However, the discrepancy of the peaks and troughs between the predictors and the observations becomes larger as the prediction period increases. Further, the ExpAR predictors catch up with the peaks and troughs of observations, but the cycle is symmetric, which is due to the structure of the model. The combined models with time varying coefficients outperform the other models, particularly, as the prediction step proceeds, but the combined models with constant coefficients are producing the best predictors until 14 steps.

Next, for the Canadian lynx number series, the whole data set (1821-1934) has been used in the literature. Here, in order to keep data for a predictive performance, we re-estimate each model by using the first 100 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 $ExpAR(p; d, \gamma)$ models 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 are shown in tables 1 and 2, and we see that there is no great difference between these estimates and the estimates using the whole sample. Further, we did not find any great differences in the movements of time varying coefficients. The estimated standard errors are shown in table 2.

The RMSE for the lynx data are tabulated at the bottom of table 7. In this case, we can see 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 numbers.

4.2 The Nelson-Plosser Data

For the Nelson-Plosser series, the last 10 values (1961-1970) are stored for the evaluation of the predictive performance. Table 8 shows the results of RMSE comparisons between individual marginal models as well as two kinds of combining models. The figures in the table should be interpreted in the same way as table 7. We observe that there is no strong dominance of combined models over marginal models, except for the SP500 and VEL series. Marginal models sometimes show the best forecasting performance, including linear predictors. Particularly, we observe that the combination method with constant coefficients for IP is better than the combination method with time varying coefficients. The ExpAR model for VEL has the best forecasts, and linear forecasts are useful during some forecasting steps for the CPI and NGNP series. These observations are expected since the economic series exhibit structural changes of the economy. Therefore, although uniform dominance of the combination method with time varying coefficients does not always hold for the economic time series examined here, we have some situations where the combination methods produce better forecasts.

From table 8, we have the following observations. First, for the highly nonlinear series(NGNP, CPI, PRGNP), the composite forecasts perform 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, except for the case of the SP500. 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 models is capturing nonlinearities (outlier, non TAR type nonlinearities) which can not be exploited for forecasting. Their observations are consistent with our results. That is, for highly nonlinear series, which show several directions from the linearity, 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.

5 Conclusion

In this paper, we investigate 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.

Clements and Smith(1999) discussed 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 are in line with the observations of these preceding literature.

We end by stating some problems for future research. First, a more theoretical analysis of the proposed method may be investigated, in particular, a Bayesian approach; see Geweke and Terui (1991, 1993) and Terui (1992). Second, an extensive simulation study and the use of other forecasting measures, like forecast encompassing tests, may be investigated. Finally, the use of other nonlinear models such as, for example, bilinear models (Subba Rao and Gabr (1984)) and random coefficient models (Nicholls and Quinn(1982)) is directly possible with our method. Our purpose here is to demonstrate only the usefulness of the principle of combining linear and nonlinear models for forecasting.

References

- Bates, J.M. and C.W.J. Granger(1969) "The combination of forecasts", *Operational Research Quarterly*, **20**, 451-468.
- Cox and Hinkley(1974) *Theoretical Statistics*, Chapman-Hall, London.
- Clements, M.P. and J.Smith(1997) " The performance of alternative forecasting methods for SETAR models", *International Journal of Forecasting*, **13**, 463-475.
- Clements, M.P. and J.Smith(1999) " A monte calro study of the forecasting performance of empirical SETAR models", *Journal of Applied Econometrics*, **14**, 123-141.
- Crane, D.B. and J.R. Crotty(1967) " A two-stage forecasting model: Exponential smoothing and multiple regression", *Management Science*, **13**, B501-507.
- De Gooijer, J.D. and K.Kumar(1992) " Some recent developments in non-linear time seies modelling, testing, and forecasting", *International Journal of Forecasting*, **8**, 135-156.
- Granger, C.W.J. and P. Newbold(1977) *Forecasting Economic Time Series*, Academic Press, London.
- Granger, C.W.J. and R. Ramanathan(1984) " Improved methods of forecasting", *Journal of Forecasting*, **3**, 197-204.
- Granger, C.W.J. and T. Teräsvirta(1993) *Modelling Nonlinear Economic Relationships*, Oxford University Press, New York.
- Geweke, J. and N. Terui(1991) "Threshold autoregressive models for macroeconomic time series: a Bayesian approach", *Proceedings of American Statistical Meeting(Business and Economics Section)*, 42-50.
- Geweke, J.and N. Terui(1993) "Bayesian threshold autoregressive models for nonlinear time series", *Journal of Time Series Analysis*, **14**, 441-451.
- Haggan,V. and T. Ozaki(1980) " Amplitude-dependent exponential AR model fitting for non-linear random vibrations", In *Time Series*, ed. O.D. Anderson, North-Holland, Amsterdam, 57-71.
- Haggan,V. and T. Ozaki(1981) "Modeling non-linear random vibrations using an amplitude-dependent autoregresiive time series model", *Biometrika*, **68**, 189-196.
- Hamilton, J.D.(1994) *Time Series Analysis*, Princeton Univ. Press, New Jersey.

- Harvey, A.C.(1989) *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge Univ. Press, New York.
- Kariya, T. and N. Terui(1997) *Nonlinear Time Series Analysis and Its Applications - Gaussianity Tests and Nonlinear Models -(in Japanese)*, Iwanami Pub., Tokyo.
- Luukonen, R, P.Saikkonen and T. Teräsvirta(1988), "Testing linearity against smooth transition autoregressive models", *Biometrika*, **75**, 491-499.
- Nelson, C.R. and C.I. Plosser (1982), "Trends and random walks in macro-economic time series", *Journal of Monetary Economics*, **10**, 139-162.
- Nicholls, D.F. and B.G. Quinn(1982), *Random coefficient autoregressive models: an introduction*, Springer-Verlag, New York.
- Petrucelli, J. and 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*, Academic Press, London.
- Ozaki, T.(1982) "The statistical analysis of perturbed limit cycle process using nonlinear time series model", *Journal of Time Series Analysis* **3**, 29-41.
- Reinmuth, J.E. and M.P. Geurts(1976) "A multideterministic approach to forecasting", in Makridakis, S. and S.C. Wheelwright eds. *Forecasting*, Vol.12, TIMS Studies in the Management Sciences, North-Holland, New York, 203-211.
- Schotman, P.C. and H.K. van Dijk (1991), "On Bayesian routes to unit routes", in *Journal of Applied Econometrics*, **6**, 387-401.
- Stone(1969) "The role of significant testing: some data with a message", *Biometrika*, **56**, 485-93.
- Tong, H. and K.S. Lim(1980) "Threshold autoregression, limit cycle and cyclical data(with discussion)" *J.Roy.Stat.Soc.*, **B42**, 245-292
- Tong, H.(1983) *Threshold Models in Non-linear Time Series Analysis*, Springer-Verlag, New York.
- Tong, H.(1990) *Non-linear Time Series: A Dynamical System Approach*, Oxford Univ. Press, New York.
- Subba Rao, T. and M.M. Gabr(1984), *An introduction to bispectral analysis and bilinear time series models*, Springer-Verlag, Berlin.

- Terui, N.(1992), "A Bayesian inference on ExpAR models for nonlinear time series", *Working Paper*, Dept. of Statistics, Texas A&M University.
- Terui, N. and T. Kariya(1997a), "Gaussianity and nonlinearity of foreign exchange rate and macroeconomic time series", *Working Paper TI97-004/4*, Tinbergen Institute.
- Terui, N. and T. Kariya(1997b), "Gaussianity and nonlinearity of foreign exchange rate", *IMS Monograph Series No.31(L-1 Statistical Procedures and Related Topics)*, 337-350.
- Terui, N. and T. Kariya(1997c), "Testing Gaussianity and linearity of Japanese stock returns", *Financial Engineering and the Japanese Market*, **4**, 1-30.
- 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.

Table 1: Constant Combining Models

Canadian Lynx: 1821-1934 year(Whole sample)				
Model	Const.	Linear	TAR	ExpAR
L.T.E	-.0256 (.1349) [-.1899]	-.3402 (.1556) [-2.1855]	.4302 (.1377) [3.1245]	.9223 (.2219) [4.1567]
Canadian Lynx: 1821-1920 year				
Model	Const.	Linear	TAR	ExpAR
L.T.E	-.0349 (.0908) [-.3842]	-.3116 (.1942) [-1.6044]	.4743 (.1427) [3.3235]	.8533 (.1726) [4.9428]
Sunspot number				
Model	Const.	Linear	TAR	ExpAR
L.T.E.	-.6526 (1.4870) [-.4389]	-.5319 (.2014) [-2.6425]	.8078 (.1088) [7.4267]	.7328 (.1950) [3.7586]

The numbers show the estimates of coefficients.

Their standard errors and t values are in parenthesis and $[\cdot]$ respectively

Table 2: Estimated Standard Errors for Marginal and Composite Models

Model	Lynx:Whole sample	Lynx:100 samples	Sunspot
Linear	.2870	.1918	14.392
TAR	.1911	.1827	12.436
ExpAR	.1978	.1603	13.561
L.T.E(C)	.1748	.1521	11.982
L.T.E(TV)	.0566	.0761	10.502

Table 3: Nonlinearity Tests: Nelson-Plosser's Series

Variable	Ori-F	Aug-F	TAR-F	New-F	CUSUM
RGNP	0.15517	0.27571	0.07719	0.16726	0.09241
NGNP	0.00037	0.00527	0.02512	0.00127	0.18666
PCRGNP	0.02689	0.25088	0.07406	0.00322	0.10078
IP	0.29993	0.37873	0.08529	0.00655	0.31221
EMP	0.10476	0.04953	0.02986	0.01323	0.04677
UN	0.10010	0.08274	0.13235	0.02331	0.17948
PRGNP	0.00011	0.00002	0.03372	0.00000	0.04200
CPI	0.00014	0.00000	0.02116	0.00023	0.23835
WG	0.10428	0.02679	0.09720	0.00428	0.32098
RWG	0.09704	0.00022	0.02072	0.00980	0.15943
M	0.02202	0.05583	0.04062	0.01178	0.01488
VEL	0.07703	0.10706	0.02028	0.31703	0.60523
BND	0.05406	0.15516	0.06135	0.04360	0.01565
SP500	0.14899	0.24497	0.05048	0.03524	0.56961

The number means the p-value.

Table 4: Estimated Marginal Models: Nelson-Plossor's Series

Linear					
Series	-	-	p	s	AIC
[A] NGNP	-	-	1	.1015	-97.6304
[A] PRGNP	-	-	1	.0690	-136.673
[A] CPI	-	-	2	.0407	-365.337
[B] SP500	-	-	5	.1566	-77.6440
[B] IP	-	-	5	.1048	-165.319
[C] VEL	-	-	1	.0729	-226.550

TAR					
Series	d	r	(p_1, p_2)	s	AIC
[A] NGNP	4	.0855	(4, 10)	.0801	-35.988
[A] PRGNP	3	.0092	(10, 10)	.0436	-55.122
[A] CPI	2	.0098	(3, 2)	.0377	-189.828
[B] SP500	1	.0281	(5, 4)	.1427	-30.6599
[B] IP	1	-.0661	(10, 5)	.0865	-124.392
[C] VEL	3	-.0134	(3, 6)	.0626	-98.1974

ExpAR					
Series	d	γ	p	s	AIC
[A] NGNP	3	1.576	9	.0674	-88.0833
[A] PRGNP	1	4.739	10	.0454	-131.560
[A] CPI	1	.0410	10	.0271	-387.764
[B] SP500	1	.00001	8	.1223	-81.3575
[B] IP	1	.00009	9	.0893	-153.588
[C] VEL	3	.00001	6	.0648	-213.878

Table 5: Constant Combining Models: Nelson-Plossor's Series

Series	Const.	Linear	TAR	ExpAR
[A] NGNP	.0195 (.0129) [1.5091]	-.2682 (.2665) [-1.0064]	.4320 (.1481) [2.9178]	.6789 (.1475) [4.6031]
[A] PRGNP	.0091 (.0077) [1.1841]	-.2841 (.2391) [-1.1883]	.5923 (.1152) [5.1391]	.6098 (.1357) [4.4951]
[A] CPI	.0064 (.0030) [2.1373]	-.4096 (.2034) [-2.0137]	.5225 (.1899) [2.7519]	.9133 (.0942) [9.6960]
[B] SP500	.0278 (.0147) [-1.8875]	-.6427 (.3342) [-1.8692]	.6465 (.2184) [2.9603]	.9058 (.1457) [6.2153]
[B] IP	.0085 (.0158) [.5414]	-.1232 (.3199) [-.3851]	.7139 (.2645) [2.6990]	.3520 (.2506) [1.4046]
[C] VEL	-.0121 (.0235) [-.5145]	-.5131 (1.3060) [-.3929]	.7263 (.2134) [3.4029]	.6224 (.2567) [2.4248]

The number inside parenthesis are standard errors
and the numbers in [·] means t values.

Table 6: Estimated Standard Errors for Marginal and Composite Models

Model	[A] NGNP	[A] PRGNP	[A] CPI	[B] SP500	[B] IP	[C] VEL
Linear	.1015	.0690	.0407	.1566	.1048	.0729
TAR	.0801	.0436	.0377	.1427	.0865	.0626
ExpAR	.0674	.0454	.0271	.1223	.0893	.0648
L.T.E.(C)	.0604	.0344	.0259	.1156	.0854	.0603
L.T.E.(TV)	.0526	.0307	.0235	.1070	.0748	.0450

Table 7: Root Mean Squared Error Comparison for Predictors

Sunspot Number				
Step	Linear	TAR	ExpAR	L.T.E.(C)
1	0.287*	4.827	2.184	2.523
2	2.151	1.423	1.068	0.906*
3	2.999	1.435	1.795	0.910*
4	2.554	1.227	1.831	0.966*
5	2.419	1.532	1.645	1.122
6	1.384	0.872	1.222	0.871*
7	1.560	0.925	1.348	0.902*
8	1.547	0.859	1.396	0.883*
9	2.567	1.598	1.661	1.125
10	2.490	1.285	1.459	0.956*
11	2.388	1.181	1.320	0.871*
12	2.433	1.203	1.332	0.872*
13	2.172	1.162	1.220	0.899*
14	2.148	1.112	1.183	0.878*
15	2.382	1.275	1.506	1.008
16	2.099	1.303	1.502	1.090
17	1.911	1.357	1.385	1.125
18	1.844	1.371	1.338	1.133
19	1.768	1.351	1.287	1.125
20	1.754	1.337	1.273	1.116
21	1.753	1.326	1.263	1.108
22	1.779	1.341	1.269	1.112
23	1.800	1.347	1.285	1.115
24	1.797	1.348	1.283	1.115
25	1.685	1.303	1.254	1.105

Canadian Lynx				
Step	Linear	TAR	ExpAR	L.T.E.(C)
1	1.440	1.399	1.527	0.824*
2	1.380	1.331	1.399	1.038
3	1.584	1.399	1.506	0.976*
4	1.912	1.674	1.784	1.046
5	2.189	1.953	2.025	1.181
6	1.921	1.753	1.813	1.125
7	1.761	1.635	1.688	1.087
8	1.681	1.580	1.617	1.109
9	1.664	1.559	1.598	1.114
10	1.683	1.573	1.617	1.112

The number means each RMSE divided by corresponding RMSE of L.T.E.(TV)

Table 8: Root Mean Squared Error Comparison for Predictors

Step	Linear	TAR	ExpAR	L.T.E.(C)	Linear	TAR	ExpAR	L.T.E.(C)
[A] NGNP				[A] PRGNP				
1	1.109	0.642	1.849	1.254	0.812	3.291	0.868	1.516
2	1.214	1.343	1.516	1.192	0.936	2.688	1.810	1.287
3	0.757	1.022	1.661	1.192	1.535	2.731	2.011	1.235
4	0.837	0.971	1.604	1.177	1.565	2.692	1.986	1.253
5	0.992	1.051	1.577	1.173	1.446	1.952	1.642	1.070
6	1.529	1.065	1.535	1.161	0.908	1.646	1.512	1.002
7	1.528	1.072	1.541	1.163	0.716	1.576	1.551	0.969
8	1.607	1.246	1.524	1.161	1.030	1.631	1.562	0.968
9	0.819	1.292	1.529	1.161	1.360	1.304	1.487	0.923
10	1.578	1.310	1.429	1.139	1.367	1.089	1.494	0.890
[A] CPI				[B] SP500				
1	0.871	0.273	1.643	0.307	1.424	1.769	1.275	1.301
2	0.877	5.920	5.447	1.141	0.962	1.178	0.871	0.981
3	1.139	2.164	1.384	1.160	1.149	1.195	1.123	1.033
4	1.132	2.856	1.681	1.199	1.205	1.204	1.214	1.058
5	0.714	3.414	2.096	1.216	1.229	1.254	1.239	1.081
6	0.720	2.981	2.376	1.100	1.109	1.129	1.144	1.016
7	0.882	3.068	2.587	1.104	1.129	1.156	1.169	1.030
8	0.947	1.794	2.0760	0.981	1.147	1.189	1.183	1.040
9	1.293	1.568	2.032	0.961	1.140	1.182	1.181	1.032
10	1.399	1.335	1.913	0.951	0.954	1.068	0.953	0.950
[B] IP				[C] VEL				
1	2.719	2.428	2.095	1.355	0.935	1.219	1.655	0.822
2	1.741	1.400	2.709	1.112	1.852	0.906	1.107	0.843
3	1.715	1.417	3.229	1.102	1.390	0.657	0.562	0.846
4	0.987	0.893	1.862	0.945	1.502	0.973	0.606	0.846
5	0.997	0.867	1.853	0.955	1.487	0.845	0.543	0.742
6	1.014	0.893	1.752	0.971	1.545	0.764	0.472	0.659
7	1.018	0.906	1.805	0.971	1.407	0.795	0.392	0.649
8	0.972	0.873	1.805	0.960	1.394	0.769	0.377	0.625
9	0.971	0.867	1.831	0.957	1.395	0.754	0.362	0.602
10	1.040	0.957	1.662	0.986	1.395	0.888	0.356	0.693

The number means each RMSE divided by corresponding RMSE of L.T.E.(TV).