Bank of Canada Banque du Canada

Working Paper 2001-12 / Document de travail 2001-12

Evaluating Linear and Non-Linear Time-Varying Forecast-Combination Methods

by

Fuchun Li and Greg Tkacz

ISSN 1192-5434

Printed in Canada on recycled paper

Bank of Canada Working Paper 2001-12

July 2001

Evaluating Linear and Non-Linear Time-Varying Forecast-Combination Methods

by

Fuchun Li and Greg Tkacz

Department of Monetary and Financial Analysis
Bank of Canada
Ottawa, Ontario, Canada K1A 0G9
gtkacz@bankofcanada.ca

Contents

Acknowledgements

Thanks to Robert Amano, Scott Hendry, Christoph Schleicher, and Christian Zimmermann, as well as seminar participants at the Bank of Canada and at the 2001 meetings of the Society for Nonlinear Dynamics and Econometrics (Atlanta), for several useful comments and suggestions.

Abstract

This paper evaluates linear and non-linear forecast-combination methods. Among the non-linear methods, we propose a nonparametric kernel-regression weighting approach that allows maximum flexibility of the weighting parameters. A Monte Carlo simulation study is performed to compare the performance of the different weighting schemes. The simulation results show that the non-linear combination methods are superior in all scenarios considered. When forecast errors are correlated across models, the nonparametric weighting scheme yields the lowest mean-squared errors. When no such correlation exists, forecasts combined using artificial neural networks are superior.

JEL classification: C53, C14, E27
Bank classification: Econometric and statistical methods

Résumé

Les auteurs de l'étude évaluent les méthodes linéaires et non linéaires de combinaison des prévisions. Entre autres formules de pondération non linéaires, ils proposent une technique d'estimation non paramétrique par la méthode du noyau qui offre une souplesse maximale en matière de pondération. Afin de comparer l'efficacité de différentes formules de pondération, ils procèdent à une simulation de Monte-Carlo. Les résultats obtenus montrent que les méthodes de combinaison non linéaires sont supérieures aux autres dans le cas de tous les scénarios envisagés. Lorsque les erreurs de prévision sont corrélées d'un modèle à l'autre, c'est la formule de pondération non paramétrique qui produit les erreurs quadratiques moyennes les plus faibles. Si les erreurs ne sont pas corrélées, les prévisions combinées à l'aide de réseaux neuronaux artificiels s'avèrent les meilleures.

Classification JEL : C53, C14, E27
Classification de la Banque : Méthodes économétriques et statistiques

1. Introduction

By definition, economic models are simplifications of reality. They help us understand economic phenomena and are often used to predict the future paths of key variables. Since there are many competing views of the world, there are, of course, many types of models that can help explain it. Some models may prove to be superior to others over a certain time frame or regime, but seldom is a model reliable for all states of the economy. Random disturbances, be they economic, political, or natural, ensure that no model can prove infallible.

To guard against uncertainty, economic forecasters often monitor multiple models that are built around different paradigms, so that if a disturbance occurs at least one model may capture it. This risk-minimizing behaviour is akin to the portfolio-diversification strategy of investors who spread their investments across several stocks or industries to minimize the impact of any downturn in a given area of economic activity.

When providing policy advice, it is often useful to consolidate forecasts from competing models into a single forecast by assigning weights to them. The major practical difficulty is to determine how the weights should be chosen. Surveys of the forecast-combination literature by Clemen (1989), Granger (1989), and Diebold and Lopez (1996) reveal that many methods have been proposed. The simplest approach is to assign constant weights to each forecast using a simple average, so that each forecast would be assigned the same weight; see, for example, Makridakis and Winkler (1983). Another simple approach is to choose the weights that minimize the sum of squared forecast errors, as proposed by Granger and Ramanathan (1984).

One problem with the above approaches is that the weights remain fixed over time. This may prove inadequate, especially in economic data, where changes in policy regimes may induce structural changes in the pattern of forecast errors of the different models, thereby altering the relative effectiveness of each model over time. Logically, since forecasters are constantly trying to improve their models, this would suggest that the relative performances of the different models are also changing over time.

The possible usefulness of time-varying combination weights has long been recognized in the forecast-combination literature, and several authors have proposed different ways to estimate them. For example, Bates and Granger (1969) suggest estimating time-varying weights by using moving subsets of the data. This approach has been used by Clemen and Winkler (1986). More recently, Diebold and Pauly (1987) have demonstrated that weight functions can be specified as polynomial functions of time. They argue that forecast errors can be greatly reduced through the systematic combination of forecasts when serial correlation exists in the forecast errors.

Although the aforementioned papers are successful in relaxing the assumption of constant-weight functions, the functional form of the weight function is still assumed to be known, apart from a finite number of unknown parameters. If the forecaster's objective is to minimize forecast errors, there is no reason to believe that a linear combination of the different forecasts will prove to be optimal. To this end, it should be productive to investigate the usefulness of non-linear weighting functions, of which the linear form is simply a special case.

In this paper, we compare the performance of several linear and non-linear methods that can be used to assign weights to forecasts produced from different models under several different assumptions regarding the structure of forecast errors. This exercise is similar to that performed by Diebold and Pauly (1987), who assessed the performance of linear combining methods under the assumption that the forecast errors were independently distributed.

Among the non-linear methods, we use the artificial neural network (ANN) approach suggested by Donaldson and Kamstra (1996), in addition to proposing a nonparametric kernel-regression time-varying (NPTV) weighting approach. The NPTV allows the form of the weight function to be determined by the forecast errors, which should allow maximum flexibility in situations of structural change of unknown form. The proposed nonparametric estimator for the unknown weight function is obtained by modifying the method of Robinson (1989). Our simulation results reveal that the non-linear methods are superior to linear methods. This superiority is more pronounced when we introduce serial and cross-correlations into the forecasting errors of the individual forecasting models, which are realistic features in time-series data.

This paper is organized as follows. Section 2 presents a nonparametric procedure to combine individual forecasts. Section 3 describes a Monte Carlo simulation study performed to evaluate this and other combination methods. Section 4 performs an empirical application to combine forecasts of Canadian GDP growth to demonstrate the use of the different forecast-combining methods. Section 5 concludes and offers suggestions for future research.

2. Combining Forecasts Using Nonparametric Weights

This section introduces a nonparametric estimator that can be used to estimate the combination weight function. Let $f_t^1, f_t^2, \ldots f_t^m$ denote m competing forecasts of y_t made at time t-1 such that when $t = 1, \ldots, T$ we have

$$y_t = \beta_1(t) f_t^1 + \beta_2(t) f_t^2 + \ldots + \beta_m(t) f_t^m + \zeta_t \ , \qquad (1)$$

where ζ_t is an error term with a zero mean.

One interesting issue is whether the weight functions are known functions of time *t*. If they are specified as known functions of *t* and of several finitely unknown parameters, the specification reduces to a standard one of parametric specification. For example, Diebold and Pauly (1987) use a deterministic non-linear (polynomial) function in *t*. An alternative approach to the specification of the weight function is to assume that it is generated by a finite-parameter stochastic model; e.g., an AR(1) was used in Sessions and Chatterjee (1989). Again, after some manipulation, the forecast-combining problem there is of a parametric type.

As alternatives to the above, we propose instead not to impose any unnecessary restrictions on the weight function, but rather allow the forecast errors to dictate the form of the weight function. Furthermore, to capture any dynamics in the forecast variable y_t not captured by the *m* individual forecasts, we allow for serially correlated disturbances ζ_t in the combining regression (equation 1). This kind of specification for the weight function leads us to approach the estimation of the weights in a nonparametric fashion. The implementation of one-step-ahead forecasts depends on the estimates of $\beta_1(T+1), \beta_2(T+1), \ldots, \beta_m(T+1)$ for the given data $\{y_t\}_{t=1}^{T}, \{f_t^1\}_{t=1}^{T+1}, \{f_t^2\}_{t=1}^{T+1}, \ldots, \{f_t^m\}_{t=1}^{T+1}$ at time $T+1$. Robinson (1989) provides a nonparametric estimation approach of $\beta_1(t), \beta_2(t), \ldots, \beta_m(t)$ at $t = 1, 2, \ldots, T$. However, the nonparametric estimation procedure proposed in Robinson cannot be used to estimate $\beta_1(T+1), \beta_2(T+2), \ldots, \beta_m(T+1)$, since the observation y_{T+1} is not available at time *T*. Therefore, a modification is required to estimate $\beta_1(T+1), \beta_2(T+1), \ldots, \beta_m(T+1)$. We propose the following kernel-estimation procedure:

$$\hat{\beta}(T+1) = \begin{bmatrix} \hat{\beta}_1(T+1) \\ \vdots \\ \hat{\beta}_m(T+1) \end{bmatrix} = \left(\Sigma_{t=1}^{T+1} K_{(T+1),\, t} f_t f_t'\right)^{-1} \Sigma_{t=1}^{T} K_{(T+1),\, t} f_t y_t, \quad (2)$$

where $K_{i,\,j} = K\left(\frac{i-j}{(T+1)h}\right)$, $f_t = [f_t^1, \ldots, f_t^m]'$ and kernel function, $K(\cdot)$, is a real-valued function heavily concentrated around the origin, and h is a smoothing parameter and depends on $T+1$.

As in Robinson (1989), we assume that $\beta_1(\tau)$, $\beta_2(\tau)$, ... , $\beta_m(\tau)$ satisfy a Lipschitz condition of order δ, $\delta_0 < \delta \le 1$, where δ_0 is some positive constant. The smoothing parameter h is required to satisfy $h \to 0$, $(T+1)h \to \infty$, and $(T+1)h^{1+2\delta} \to 0$.

The consistency and asymptotic normality of $\hat{\beta}(T+1)$ for $\beta(T+1)$ can be verified in the following manner,

$$\begin{aligned}\hat{\beta}(T+1) = (\Sigma_{t=1}^{T+1} K_{(T+1),t} f_t f_t')^{-1} \Sigma_{t=1}^{T+1} K_{(T+1),t} f_t y_t \\ -(\Sigma_{t=1}^{T+1} K_{(T+1),t} f_t f_t')^{-1} (K(0) f_t y_t)\end{aligned}. \quad (3)$$

However, we know from equation (15.8) in Robinson (1989) that the second term on the right-hand side in (3) is $O_p\left(\frac{1}{(T+1)h}\right)$. Therefore, the leading term is the first term, which is consistent and asymptotically normally distributed, as in Robinson (1989).

3. Simulation Study

The purpose of this section is to design and perform a Monte Carlo simulation experiment to study the performance of linear and non-linear forecast-combining methods. We consider five common linear methods: the combination of forecasts using a simple average (SA), ordinary least squares (OLS), non-negative restricted least squares (NRLS), equality restricted least squares (ERLS), and weighted least squares with polynomial weights (WLSP). In addition to the nonparametric estimator in section 2, we consider a second non-linear combination method based on ANN, proposed by Donaldson and Kamstra (1996).

3.1 Design

In this simulation study, we wish to forecast a time series, y_t, over the next S periods in one-period steps, with T denoting the number of available in-sample observations, using information available at period $T+s$, where $s = 0, ..., S-1$, and three different values of S are considered: S = 20, 50, 100. Such a forecast is denoted as $\hat{y}_{T+s+1|T+s}$. The mean-squared error (MSE) is therefore

$$MSE = \frac{\Sigma_{s=0}^{S-1}(\hat{y}_{T+s+1|T+s} - y_{T+s})^2}{S}. \quad (4)$$

The performance of different combined forecasts is compared for three different time series, which are generated from

$$y_t = \rho y_{t-1} + \eta_t , \tag{5}$$

$$y_t = \theta\eta_t + \eta_{t-1} , \tag{6}$$

$$y_t = \begin{cases} \rho_1 y_{t-1} + \eta_t, \text{ if } y_{t-1} < \tau \\ \rho_2 y_{t-1} + \eta_t, \text{ if } y_{t-1} \geq \tau \end{cases} . \tag{7}$$

Equations (5), (6), and (7) represent, respectively, AR(1), MA(1), and SETAR(1, 1) processes. Calibrating the models, we set $\rho, \theta, \rho_1,$ and ρ_2 to $0.9, 0.9, 0.6, 0.4$, respectively. This ensures the stationarity of the AR and MA models, while the SETAR model has rapid mean-reversion above the threshold τ and slow mean-reversion below it. η_t is assumed to be distributed as $N(0, 1)$. The threshold and starting values are assumed to be 0.0. The models selected are standard choices in the forecasting of stationary processes.

To assess the performance of forecast-combination methods, we generate 1080 observations. We discard the first 880 observations to eliminate any start-up effects. The number of replications for each combined method is set to 500. We use *Matlab* 5 running on a Sparc Ultra 10 for this exercise.

In this Monte Carlo experiment, we focus on the case of three individual forecasts, $f_t^1, f_t^2,$ and f_t^3, which will be presented later. The combined forecasts with constant weights are based on the following linear regression model:

$$y_t = \beta_1 f_t^1 + \beta_2 f_t^2 + \beta_3 f_t^3 + v_t , \; t = 1, 2, \ldots , T, \tag{8}$$

where T observations on y_t are regressed on the T observations of the three forecasts. The general form of the combining forecasts with constant weights is

$$\hat{y}_{T+1} = \hat{\beta}_1 f_{T+1}^1 + \hat{\beta}_2 f_{T+1}^2 + \hat{\beta}_3 f_{T+1}^3 . \tag{9}$$

$\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3$ are obtained by, respectively, a simple average where $\hat{\beta}_1 = \hat{\beta}_2 = \hat{\beta}_3 = \frac{1}{3}$, OLS, ERLS, and NRLS, which are all well-documented estimators. We also obtain combined forecasts using WLSP as follows:

$$y_t = P^1(t) f_t^1 + P^2(t) f_t^2 + P^3(t) f_t^3 + v_t , \tag{10}$$

where $P^i(t) = p_0^i + p_1^i t$, $i = 1, 2, 3$, and the parameters p_0^i and p_1^i are estimated by the weighted least squares approach. Therefore, the combined forecast with polynomial weights is

$$\hat{y}_{T+1} = \hat{P}^1(T+1) f_{T+1}^1 + \hat{P}^2(T+1) f_{T+1}^2 + \hat{P}^3(T+1) f_{T+1}^3 . \tag{11}$$

For the non-linear combination methods, the nonparametric combined forecast based on (1) is

$$\hat{y}_t = \hat{\beta}_1(T+1)f^1_{T+1} + \hat{\beta}_2(T+1)f^2_{T+1} + \hat{\beta}_3(T+1)f^3_{T+1}, \tag{12}$$

where $\hat{\beta}_1(T+1)$, $\hat{\beta}_2(T+1)$, and $\hat{\beta}_3(T+1)$ are given by (2). The ANN combined forecast is based on the following regression expression:

$$y_t = \alpha_0 + \Sigma^3_{k=1}\alpha_k G(f_t\gamma_k) + \nu_t, \tag{13}$$

where

$$G(f_t\gamma_k) = (1 + \exp[-(\gamma_{0,k} + \gamma_{1,k}f^1_t + \gamma_{2,k}f^2_t + \gamma_{3,k}f^3_t)])^{-1}. \tag{14}$$

To estimate the parameters $\{\alpha_k\}^3_{k=0}$ and $\{\gamma_{0,k}, \gamma_{1,k}, \gamma_{2,k}, \gamma_{3,k}\}^3_{k=1}$ in equation (13), the following sum of squared deviations between the output and the network is minimized:

$$SSD = \Sigma^T_{t=1}[y_t - (\alpha_0 + \Sigma^3_{k=1}\alpha_k G(f_t\gamma_k))]^2. \tag{15}$$

Equation (15) is estimated using back-propagation, which updates the parameter values until we achieve the pre-specified convergence level. The combined forecast based on ANN is thus

$$\hat{y}_{T+1} = \hat{\alpha}_0 + \Sigma^3_{k=1}\hat{\alpha}_k G(f_{T+1}\hat{\gamma}_k). \tag{16}$$

In the estimation of the weights and ANN parameters in regression models (1), (8), (10), and (13), and for each combined forecast, we use the first 100 observations on y and 3 individual forecasts to estimate unknown parameters, and then produce the one-step-ahead combined forecast $\hat{y}_{100+1}$. We next update our data set by adding the one-step-ahead combined forecast and dropping the first observation, while keeping the sample size constant at 100. We then re-estimate the parameters and produce the one-step-ahead forecast $\hat{y}_{100+2}$. This recursive updating and one-step-ahead, out-of-sample forecasting procedure is repeated until the one-step-ahead, out-of-sample forecast $\hat{y}_{100+S}$ is produced.

Furthermore, in the estimation of the weight function in our nonparametric combined forecast, we must choose the kernel function and window width, or smoothing parameter (h). The kernel function is chosen as the standard Gaussian density function, $K(x) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)$, which is continuously differentiable of any order. The choice of the smoothing parameter is set at $h = T^{-0.6}$, which satisfies $h \to 0$, $(T+1)h \to \infty$, and $(T+1)h^{1+2\delta}$. Our experiments show that the nonparametric kernel estimator is sensitive to the choice of h in that different values of h generate different standard deviations for the sampling distribution of the nonparametric estimator. Whether the above admissible smoothing parameter represents the achievable optimal rate of convergence is unknown to us. However, statistically, a small *h* tends to correspond to a small bias in $\hat{\beta}(T+1)$ and a large *h* to a small variance.[1] It is thus clear that further research is required concerning the optimal choice of the smoothing parameter *h*. This research is beyond the scope of this paper, and hence it is not pursued in this study.

Suppose we have three different individual forecasts, f_t^1, f_t^2, and f_t^3, which are generated as

$$\begin{aligned} f_t^1 &= 1 + y_t + e_t^1 \\ f_t^2 &= 1 + y_t + e_t^2 \\ f_t^3 &= 1 + y_t + e_t^3, \end{aligned} \tag{17}$$

where each forecast, f_t^i, $i = 1, 2,$ and 3, is equal to the true realized value y_t plus a one-step-ahead forecast error, which consists of a systematic bias equal to one, and the non-systematic error e_t^i. We use (5), (6), and (7) to generate the y_t. The results, however, are not highly sensitive to this choice, and thus we present only the results using the AR(1) process (5). To examine the effects of forecast errors on the performance of alternative methods used to combine forecasts, we consider the five cases described below.

Case 1: It is assumed that the one-step-ahead forecast errors are free of serial correlation and that no covariance exists between them. Their respective variances are assumed to be constant, but different, throughout the sample: $var(e_t^1) = \sigma_1 = 0.5$, $var(e_t^2) = \sigma_2 = 1$, and $var(e_t^3) = \sigma_3 = 2$, respectively.

---

1. This trade-off leads to a possible choice of *h* by $\mathrm{argmin}_h CV(h)$, where

$$CV(h) = \sum_{t=1}^{T} [y_t - \hat{\beta}_1(t) - \hat{\beta}_2(t) - \ldots - \hat{\beta}_m(t)]^2$$

and

$$\begin{bmatrix} \hat{\beta}_1(t) \\ \vdots \\ \hat{\beta}_m(t) \end{bmatrix} = \left(\Sigma_{j=1}^T K_{t,j} f_j f_j'\right)^{-1} \Sigma_{t=1}^T K_{t,j} f_t y_t.$$

Unfortunately, we cannot obtain the analytical expression for $\mathrm{argmin}_h CV(h)$.

Case 2: The forecast errors are still assumed to be free of serial correlation and to have zero covariance, and we have again $\sigma_1 = 0.5$, and $\sigma_2 = 1$ throughout the sample, but we assume that the forecast errors of f_t^3 are heteroscedastic. In particular, $\sigma_3 = 2$ for t = 1, ... , 100, and it begins to increase linearly until it achieves a value of 15 at t = 150.

Case 3: We allow for the possibility of time-varying variances and covariance among the forecast errors. We have $\sigma_1 = 0.5$, $\sigma_2 = 1$, and $\sigma_3 = 2$, and the covariance is set to zero at $t = 1, \ldots, 100$, but σ_1 and σ_2 begin to change, respectively, to 3 and 4 after t = 100. σ_3 grows linearly from 1 at t = 101 to 5 at t = 120, and it decreases linearly to 1 at t = 150.

To generate varying correlated individual forecast errors, let $u_t^i = 1 + e_t^i$ for i = 1, 2, 3. Then, in vector form, the forecast errors in equation (17) can be written as

$$U_t = \begin{bmatrix} u_t^1 \\ u_t^2 \\ u_t^3 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + \begin{bmatrix} e_t^1 \\ e_t^2 \\ e_t^3 \end{bmatrix}.$$

Thus, the covariance of U_t can be obtained as

$$\Sigma_{U_t} = Cov\begin{bmatrix} u_t^1 \\ u_t^2 \\ u_t^3 \end{bmatrix} = Cov\begin{bmatrix} e_t^1 \\ e_t^2 \\ e_t \end{bmatrix} = \begin{bmatrix} \sigma_{1t}^2 & \sigma_{12t} & \sigma_{13t} \\ \sigma_{12t} & \sigma_{2t}^2 & \sigma_{23t} \\ \sigma_{13t} & \sigma_{32t} & \sigma_{3t}^2 \end{bmatrix} \qquad (18)$$

where $\sigma_{it}^2 = E((e_t^i)^2)$, $\sigma_{ijt}^2 = E(e_t^i e_t^j)$ $i, j = 1, 2, 3$. We consider a version of Σ_t as follows:

$$\Sigma_t = \begin{bmatrix} \sigma_{1t} & \sigma_{12t} & \sigma_{13t} \\ \sigma_{12t} & \sigma_{2t} & 0 \\ \sigma_{13t} & 0 & \sigma_{3t} \end{bmatrix}$$

where $\sigma_{12t} = 0$, and $\sigma_{13t} = 0$ for t = 1, 2, . . . 100, but from t = 101, σ_{12t} and σ_{13t} begin to change linearly, respectively, by reaching 8 and 1 for t = 101, . . . , 120, and decreasing linearly to 0.4 and 0.05 at t = 150.

To generate e_t^1, e_t^2, and e_t^3, we transform the three-dimensional standard normal random variable N_t by

$$S = \begin{bmatrix} S_{1t} & S_{12t} & S_{13t} \\ 0 & S_{22t} & 0 \\ 0 & 0 & S_{33t} \end{bmatrix}. \tag{19}$$

By solving the equation $\Sigma_{U_t} = Cov(S_t N_t)$ we can obtain the solutions of $S_{11t}, S_{12t}, S_{13t}, S_{22t}$, and S_{33t} as follows:

$$S_{11t} = \left[\sigma_1^2 - \left(\frac{\sigma_{12t}}{\sigma_{22}}\right)^2 - \left(\frac{\sigma_{13t}}{\sigma_{33t}}\right)^2\right]^{\frac{1}{2}}, S_{12t} = \frac{\sigma_{12t}}{\sigma_{22}} \text{, and} \tag{20}$$

$$S_{13t} = \frac{\sigma_{13t}}{\sigma_{33t}}, S_{22t} = \sigma_{22}, S_{33t} = \sigma_{33t} . \tag{21}$$

Therefore, we have $e_t = S_t N_t$, where $e_t = \begin{bmatrix} e_t^1 & e_t^2 & e_t^3 \end{bmatrix}'$.

Case 4: We allow for serial correlation in the forecast errors, while the variances are constant and covariances are set to zero. The data-generating process (DGP) of each individual forecast error is given by its transition density function:

$$f(s, x; t, y) = \frac{1}{\sqrt{2\pi(1 - e^{-2(t-s)})}} \exp\left(\frac{(y - xe^{-(t-s)})^2}{2(1 - e^{-2(t-s)})}\right). \tag{22}$$

The original values of the DGP are drawn directly from their marginal densities, which are the standard normal densities. The sample interval is fixed at *t*-*s* = 1.

Case 5: We allow for both serial correlation and covariances. We transform the three-dimensional variable by *S* (equation 19) such that the forecast errors have the same covariance as in Case 3, but with serially correlated forecast errors.

3.2 Results

Table 1 reports MSEs for each of the combining methods, each of the individual forecasts, and the five different cases reflecting the different forms of the forecast-error structures. The true DGP is an AR(1). The sample size for the forecasting exercise is set to $S = 50$ in all cases.[2]

Several noteworthy features emerge from these results. First, in Cases 1 and 2, where the errors are assumed to be uncorrelated, it is clear that the ANN combining method yields the lowest MSEs, but only in the constant-variance case can we claim that the MSE is significantly lower than the alternative methods. Of the linear combining methods, the NRLS approach marginally yields the lowest errors.

Second, when forecast errors across models are correlated (Cases 3 and 5), the nonparametric combining method dominates, although we cannot claim that its MSEs are significantly lower than all alternatives at the usual significance levels (this is deduced by examining the standard errors). The performance of the ANN combining methodology is noticeably worse when cross-correlations are introduced, likely reflecting the difficulty in training the network to recognize the pattern of correlation present.

Third, taking a simple average of the individual forecasts does poorly when the forecast errors are uncorrelated. Recall that in Case 1 the individual forecasts are uncorrelated, and in addition to a systematic bias of 1.0 for f^1, f^2, and f^3, the prediction errors have $\sigma_1 = 0.5$, $\sigma_2 = 2$, and $\sigma_3 = 1$, for all t. Thus the expected MSE of the SA is $var\left(1 + \frac{1}{3}(e_t^1 + e_t^2 + e_t^3)\right) = 1.5833$. The MSE of the simple average of f^1, f^2, and f^3 in Table 1 is very close to the expected MSE, while, in general, MSEs of f_t^1, f_t^2, and f_t^3 are somewhat below or above their expected Case 1 MSEs of 1.250, 2.000, and 5.000, respectively. Note that in Cases 2 and 3 the MSEs of individual forecast 3 are also very close to their respective values of

$$1 + \frac{1}{50}\left(\sum_{t=1}^{50}\left[1 + \frac{4}{19}(t-1)\right]^2\right) = 88.91 \text{ and}$$

$$1 + \frac{1}{50}\sum_{t=1}^{20}\left[1 + \frac{4}{19}(t-1)\right]^2 + \frac{1}{50}\sum_{t=21}^{50}\left[5 - \frac{4}{30}(t-20)\right]^2 = 11.15.$$

All other forms of combined forecasts usually outperform the individual forecasts.

2. Results for $S = 20$ and $S = 100$ are not presented, but are available from the authors. They are qualitatively similar to those for $S = 50$.

When serial correlation is introduced in Cases 4 and 5, it is apparent that the non-linear methods continue to dominate, although the ANN again does relatively poorly when cross-correlations are introduced. Serial correlation in forecasting models can be quite common in the forecasting of economic time series, thus these two cases are probably the most realistic for the applied researcher. Our main finding suggests that the choice of forecast-combining technique appears to depend upon whether forecast errors across models are correlated. If they are, the NPTV method should be used; otherwise, an ANN combined forecast would prove to be superior.

4. Empirical Application

The variable forecasted in this paper is the growth rate of real GDP in Canada in 1992 dollars. The data are available from 1947Q1 to 2000Q1 and can be obtained from Statistics Canada's CANSIM database. The forecasts of the growth rates of real GDP are obtained using the simple random walk model, single exponential smoothing model, and AR(1) model. The quarterly growth rate is computed as $\Delta y_t = 100 \times \log\left(\frac{y_t}{y_{t-1}}\right)$.

Data from the first quarter of 1947 to the last quarter of 1972 are used to estimate the parameters of the simple random walk model, single exponential smoothing model, and AR(1) model. We then produce the one-step-ahead, out-of-sample individual forecasts for the first quarter of 1973. Next we update our data set, while keeping the sample size constant, by adding the first quarter of 1973 and dropping the first quarter from 1947. We then re-estimate the individual models and produce one-step-ahead, out-of-sample forecasts for the second quarter of 1973. This recursive updating and one-step-ahead, out-of-sample forecasting procedure is repeated until one-step-ahead, out-of-sample individual forecasts of GDP are produced for each quarter from the first quarter of 1973 to the first quarter of 2000. These constitute the individual out-of-sample forecasts that are used to combine forecasts in the empirical study.

The next step is to divide the individual out-of-sample forecasts into two subsamples: 1973Q1 to 1997Q1 (100 observations) and 1997Q2 to 2000Q1 (12 observations). We use the data from the first subsample to estimate weight functions for every combined forecast to obtain one-step-ahead, out-of-sample combined forecasts for the second quarter of 1997. Then we update our information set by one quarter to obtain new weight functions and new one-step-ahead, out-of-sample combined forecasts for the third quarter of 1997. This procedure is recursively repeated until we have obtained combined one-step-ahead, out-of-sample forecasts for each of our five combining models for the period 1997Q2 to 2000Q1. To evaluate the performance of the nonparametric combining method, we need to specify the smoothing parameter, h. It is chosen according to $h = T^{-\lambda}$, where $T = 100$ and λ is a positive constant, and $\lambda > 1/(1 + 2\delta)$, for $1/3 < \delta \leq 1$. We choose $\lambda = 0.6, 0.7$.

Table 2 reports the MSEs for each of the individual forecasts and each combined forecast for the out-of-sample period 1997Q2 to 2000Q1. Canada experienced relatively constant and robust growth over this period. Based on the MSE, the combined forecasts are more accurate than any of the individual forecasts included in the combination, except for the performance of the WLSP combined forecast, which is noticeably worse than the other combining methods. The non-linear combined forecasts are superior, with the ANN-combined forecast dominating. The performance of the nonparametric combined forecast depends on the value of the smoothing parameter, as stated earlier (λ values are in parentheses). The range of its MSE is roughly 0.17 to 0.19, which is slightly lower than the linearly combined forecasts.

5. Conclusion

This paper has evaluated, via a simulation study, the performance of different forecast-combining methods. This type of study was previously conducted by Diebold and Pauly (1987), but they focused on linear forecast-combination methods only and restricted their attention to individually independent forecast errors. Donaldson and Kamstra (1996) have since proposed a non-linear, ANN combining method, and we propose here a nonparametric, kernel-regression combining method, comparing both to the traditional linear methods.

Our results demonstrate that non-linear combined forecasts outperform linear combined forecasts in all cases considered, even when serial correlation and heteroscedasticity are introduced to the forecast errors. When forecast errors are dependent across models, the NPTV procedure that we propose appears to dominate; when no such dependence exists, the ANN approach is superior.

However, the improved performance in some cases is not statistically significant at the usual significance levels, indicating that applied forecasters should use their judgment to determine whether the costs of the additional complexity of the non-linear combined forecasts outweigh the benefits of improved forecast accuracy. As computer software increasingly incorporates additional non-linear estimation features, non-linear combination methods are likely to prove more popular.

References

Bates, J.M. and C.W.J. Granger. 1969. "The Combination of Forecasts." *Operations Research Quarterly* 20: 451–68.

Clemen, R.T. 1989. "Combining Forecasts: A Review and Annotated Bibliography." *International Journal of Forecasting* 5: 559–81.

Clemen, R.T. and R.L. Winkler. 1986. "Combining Economic Forecasts." *Journal of Business and Economic Statistics* 4: 39–46.

Diebold, F.X. and J.A. Lopez. 1996. "Forecast Evaluation and Combination." In *Handbook of Statistics 14*, G.S. Maddala and C.R. Rao (eds). Amsterdam: Elsevier Science.

Diebold, F.X. and P. Pauly. 1987. "Structural Change and the Combination of Forecasts." *Journal of Forecasting* 6: 21–40.

Donaldson, R.G. and M. Kamstra. 1996. "Forecast Combining with Neural Networks." *Journal of Forecasting* 15: 49–61.

Granger, C.W.J. 1989. "Combining Forecasts—Twenty Years Later." *Journal of Forecasting* 8: 167–73.

Granger, C.W.J. and R. Ramanathan. 1984. "Improved Methods of Forecasting." *Journal of Forecasting* 3: 197–204.

Makridakis, S. and R.L. Winkler. 1983. "Averages of Forecasts: Some Empirical Results." *Management Science* 29: 987–96.

Robinson, P.M. 1989. "Nonparametric Estimation of Time-Varying Parameters." In *Statistical Analysis and Forecasting of Economic Structural Change*, P. Hack (ed). Berlin: Springer-Verlag.

Sessions, D.N. and S. Chaterjee. 1989. "The Combining of Forecasts using Recursive Techniques with Non-Stationary Weights." *Journal of Forecasting* 8: 239–54.

Table 1: Mean-squared errors

Forecast-combining methods	Disturbance structure				
	Case 1	Case 2	Case 3	Case 4	Case 5
Linear combinations					
SA	1.625 (0.256)	11.08 (2.859)	5.840 (1.287)	1.336 (0.255)	5.985 (1.565)
OLS	0.575 (0.106)	0.823 (0.467)	2.058 (0.517)	0.839 (0.214)	2.076 (0.557)
NRLS	0.572 (0.104)	0.793 (0.474)	2.007 (0.486)	0.835 (0.208)	2.016 (0.514)
ERLS	1.250 (0.153)	1.553 (0.511)	6.936 (1.406)	1.372 (0.270)	6.932 (1.492)
WLSP	1.026 (0.272)	1.291 (0.505)	1.721 (0.620)	2.458 (2.027)	9.049 (6.696)
Non-linear combinations					
NPTV	0.579 (0.108)	0.681 (0.201)	*1.426* *(0.317)*	0.841 (0.215)	*1.499* *(0.343)*
ANN	*0.213* *(0.056)*	*0.434* *(0.257)*	2.219 (0.404)	*0.303* *(0.068)*	2.227 (0.569)
Individual forecasts					
f_t^1	1.277 (0.167)	1.277 (0.167)	10.07 (2.029)	1.978 (0.504)	10.06 (1.95)
f_t^2	2.095 (0.341)	2.095 (0.341)	17.27 (3.575)	2.021 (0.515)	17.05 (4.14)
f_t^3	5.028 (0.962)	89.75 (23.868)	11.55 (2.531)	1.983 (0.456)	11.20 (3.78)

Notes: SA = simple average, OLS = ordinary least squares, NRLS = non-negative restricted least squares, ERLS = equality restricted least squares, WLSP = weighted least squares with polynomial weights, NPTV = nonparametric time-varying, ANN = artificial neural network. Case 1 = constant variance, no covariance, no serial correlation; Case 2 = non-constant variance, no covariance, no serial correlation; Case 3 = non-constant variance, positive covariance, no serial correlation; Case 4 = constant variance, no covariance, serial correlation; Case 5 = constant variance, positive covariance, serial correlation. 1000 replications performed. Sample size S = 50 for each replication. Standard errors in parentheses. Lowest MSEs are italicized.

Table 2: Mean-squared errors of combined Canadian GDP forecasts

Forecast-combining method	Mean-squared errors (S = 12 quarters)
Linear combinations	
SA	0.246
OLS	0.276
NRLS	0.267
ERLS	0.276
WLSP	1.735
Non-linear combinations	
NPTV	0.179 (0.60) 0.170 (0.70)
ANN	0.005
Individual forecasts	
Random walk	2.764
Exponential smoothing	1.252
AR(1)	2.986

Note: λ in parentheses.

Bank of Canada Working Papers
Documents de travail de la Banque du Canada

Working papers are generally published in the language of the author, with an abstract in both official languages. ***Les documents de travail sont publiés généralement dans la langue utilisée par les auteurs; ils sont cependant précédés d'un résumé bilingue*****.**

2001

2001-11	Gaining Credibility for Inflation Targets	J. Yetman
2001-10	The Future Prospects for National Financial Markets and Trading Centres	C. Gaa, S. Lumpkin, R. Ogrodnik, and P. Thurlow
2001-9	Testing for a Structural Break in the Volatility of Real GDP Growth in Canada	A. Debs
2001-8	How Rigid Are Nominal-Wage Rates?	A. Crawford
2001-7	Downward Nominal-Wage Rigidity: Micro Evidence from Tobit Models	A. Crawford and G. Wright
2001-6	The Zero Bound on Nominal Interest Rates: How Important Is It?	D. Amirault and B. O'Reilly
2001-5	Reactions of Canadian Interest Rates to Macroeconomic Announcements: Implications for Monetary Policy Transparency	T. Gravelle and R. Moessner
2001-4	On the Nature and the Stability of the Canadian Phillips Curve	M. Kichian
2001-3	On Commodity-Sensitive Currencies and Inflation Targeting	K. Clinton
2001-2	Exact Non-Parametric Tests for a Random Walk with Unknown Drift under Conditional Heteroscedasticity	R. Luger
2001-1	The Elements of the Global Network for Large-Value Funds Transfers	J.F. Dingle

2000

2000-23	The Application of Artificial Neural Networks to Exchange Rate Forecasting: The Role of Market Microstructure Variables	N. Gradojevic and J. Yang
2000-22	Une analyse empirique du lien entre la productivité et le taux de change réel Canada-É-U	D. Dupuis et D. Tessier
2000-21	Les effets réels du cours des actions sur la consommation	L. Pichette
2000-20	Steps in Applying Extreme Value Theory to Finance: A Review	Y. Bensalah
2000-19	Le modèle USM d'analyse et de projection de l'économie américaine	R. Lalonde

Copies and a complete list of working papers are available from:
Pour obtenir des exemplaires et une liste complète des documents de travail, prière de s'adresser à :

Publications Distribution, Bank of Canada
234 Wellington Street, Ottawa, Ontario K1A 0G9
E-mail: publications@bankofcanada.ca
Web site: http://www.bankofcanada.ca

Diffusion des publications, Banque du Canada
234, rue Wellington, Ottawa (Ontario) K1A 0G9
Adresse électronique : publications@banqueducanada.ca
Site Web : http://www.banqueducanada.ca