Combining Forecasts via Simulations

Yuzhi Cai,* *Swansea University, United Kingdom*
Neville Davies, *Plymouth University, United Kingdom*
Wan-Kai Pang, *The Hong Kong Polytechnic University, China*

Abstract

Combining forecasts can be based on different data or different methods or both. In practice, many different data sets can be difficult to obtain. In this paper we propose a combining forecasts method that uses both simulated and observed time series. Several combining methods have been studied and the combined forecasts have been compared with those obtained from the model fitted to the observed time series only. Simulation studies and applications of the proposed method to real data sets show that on average the combined forecasts are more accurate than those obtained without combining.

Key words: Combining forecasts, simulation, marginal distribution, autocorrelation structure, weighted average.

1   Introduction

Forecasting for time series is always an important issue in time series analysis. One approach to forecasting is to fit a model to an observed time series and then to forecast using the fitted model, see for example, Box et al. (2006), Tong (1990) and Cai (2003, 2005, 2007, 2010). Another approach to forecasting is to combine forecasts that are obtained from different methods or different data or both. Such an approach is very appealing in that combining can reduce error arising from faulty assumptions, bias or mistakes in data. Hence, combining forecast methods have been used in many areas. For example, Batchelor and Dua (1995) examined forecasts of real GNP, inflation, corporate profits, and unemployment for forecast horizons of 6, 12 and 18 months ahead. They found the Mean Square Error (MSE) decreased by an average of 9.2% by combining any two economists' forecasts. Kumar et al. (2003) studied the prediction of Typhoon tracks and intensity by combining forecasts from different models. Winkler and Poses (1993) examined physicians' predictions of survival for 231 patients who were admitted to an intensive care unit. Different methods have also been proposed for combining forecasts. Clemen (1989) conducted a comprehensive review of the evidence and found equal weighting to be accurate for many

---

*Address for correspondence: Dr Yuzhi Cai, College of Business, Economics and Law, Swansea University, Swansea SA2 8PP, United Kingdom. Email: y.cai@swansea.ac.uk

types of forecasting. McNees (1992) found that the mean seemed superior when the forecasting accuracy was measured by root MSE (RMSE), while the median seemed better if the Mean Absolute Error (MAE) was used. Given different forecasts, Armstrong (2001) suggested that combining them using equal weights usually performs better. Bunn and Wright (1991) discussed issues on combining judgemental forecasts and statistical forecasts. McCracken and Clark (2008) derived the MSE-minimizing weights for combining the restricted and unrestricted forecasts. Wang and Nie (2008) argued that choosing the method which has the best forecasting result as the combining forecasts model can greatly enhance the forecast effectiveness. Armstrong and Collopy (1992) evaluated measures for making comparisons of errors across time series. They recommended to use the Median Absolute Percentage Error (MdAPE) for selecting the most accurate methods if reasonable number of series available.

In practice, it is usually very difficult to obtain several observed time series over a period of time, which makes it difficult to combine forecasts based on different data, and which motivated our research. Cai et al. (2008) proposed a simulation method that enables us to generate many independent time series which mimic the observed time series to a high standard. The method has been successfully applied to sea condition time series in coastal engineering such as wave height and wave period etc. We expect that the simulated time series should also help in forecasting for the following reasons. The autocorrelation structure and the marginal distribution are two main important features of a time series. The simulated time series obtained from the method of Cai et al. (2008) have very similar autocorrelation structure and marginal distribution to those of the observed time series but not exactly the same. Therefore, the simulated time series can be viewed as data coming from different sources. As Armstrong (2001) pointed out, using different sources of data can add useful information into a combined forecast and may also adjust for biases. We will see later in the paper that the simulated time series do work very well.

The proposed combining forecasts method consists of four steps: The first step of the method is to use a modified version of the simulation method of Cai et al. (2008) to obtain simulated time series. We will use a different marginal distribution from that used by Cai et al. (2008), which we believe is more suitable for the real data sets considered in this paper. The second step of the method is to fit a time series model to each simulated and the observed time series, so we have a sequence of fitted time series models. The third step is to use each fitted model to obtain forecasts. The fourth step of the method considers how to combine different forecasts obtained from different models so that a final conditional forecast can be produced. Therefore, the combined forecasts contain information from both different sources and different models. We will see that the developed forecasting method does increase the forecasting accuracy.

The arrangement of the paper is as follows. In Section 2, we present the forecasting method. The simulation studies are given in Section 3, and the applications to real time series can be found in Section 4. Finally, some comments and conclusions are given in Section 5.

2 The method

In practice, usually only one observed time series over a period of time is available. However, if we could generate simulated time series which possess necessary properties that the observed time series has, then we could obtain forecasts based on both the simulated and the observed time series. A final conditional forecast could then be produced by combining those forecasts in a proper way. This is the basic idea of the forecasting method presented below.

Let y_t $(t = 1, 2, \ldots, n)$ be an observed time series and let $F(y)$ be the marginal distribution defined by

$$F(y) = G_1(y)I_{[y \leq \mu]} + [(1 - \beta)G_2(y) + \beta]I_{[y > \mu]}, \tag{1}$$

where

$$G_1(y) = 1 - e^{-(y/\lambda)^\alpha}, \ \lambda > 0, \ \alpha > 0, \ y > 0$$

is the Weibull distribution function, and

$$G_2(y) = 1 - \left[1 + \frac{\xi(y - \mu)}{s}\right]^{-1/\xi}, \quad y > \mu, \ 1 + \frac{\xi(y - \mu)}{s} > 0$$

is the generalized Pareto distribution function (GPD), μ is a proper threshold value and $P(Y \leq \mu) = \beta$.

The marginal distribution defined by (1) is a mixture of the Weibull distribution and the GPD, which is different from that used by Cai et al. (2008). Most of the economics and financial time series take positive values and show many extreme values over the observed period of time. Therefore, a normal distribution may not be appropriate for describing the marginal distribution of such time series. Although G_1 could be the empirical distribution of the observed data, taking G_1 as a continuous distribution function makes the calculations easier. More importantly, we have found that this marginal distribution performs well for many economics and financial time series in practice, such as the dividend yields time series considered later in this paper. Note that for other types of data different marginal distributions may be used.

The forecasting method consists of the following four steps (a) to (d).

(a) Use a modified method of Cai et al. (2008) to obtain k time series y_{it} $(i = 1, \ldots, k,$ $t = 1, \ldots, m)$, where m is the length of simulated time series. To achieve this, we first define a base process and estimate the parameters involved by using the observed time series. The base process takes the form

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \cdots + \phi_p z_{t-p} + u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \cdots + \theta_q u_{t-q},$$

where p and q are the order of the base process, $\phi_1, \ldots, \phi_p$ and $\theta_1, \ldots, \theta_q$ are the parameters to be estimated, and u_t are independently identically distributed (i.i.d.) normal random variables with mean zero and variance σ^2. Then we simulate k independent time series z_{it} $(i = 1, \ldots, k, t = 1, \ldots, m)$ from the base process. Finally the simulated series z_{it} are transformed properly to the required time series y_{it} such that they have the same autocorrelations up to a lag, say, L and the same marginal distribution as those of the observed

3

time series. For details, see Cai et al. (2008). In this paper we use the marginal distribution given by (1), so y_{it} can be obtained as follows: let $y_{it} = G_1^{-1}(\Phi(z_{it}))$ if $\Phi(z_{it}) \leq \beta$, and let $y_{it} = G_2^{-1}(\Phi(z_{it}))$ otherwise, where $\Phi(\cdot)$ is the standard normal distribution function, and $G_1^{-1}(x)$ and $G_2^{-1}(x)$ are the inverse functions of the Weibull distribution and the GPD respectively. Specifically, we have

$$G_1^{-1}(x) = \lambda \left(-\ln(1-x) \right)^{1/\alpha}, \quad G_2^{-1}(x) = w + \left[\left(\frac{1-x}{1-\beta} \right)^{-\xi} - 1 \right] \frac{s}{\xi}.$$

(b) Fit a time series model to each of the simulated and the observed time series. For illustration purposes, the fitted models are in the class of $ARIMA$ models, but any other types of time series models can also be used if required.

(c) Use each fitted model to obtain forecasts up to ℓ-step ahead, where $\ell = 1, \ldots, M$.

(d) Combine forecasts in a proper way. In this paper we propose the following methods for combining forecasts. Method 1 is an unequally weighted average method, Method 2 is a modified version of a commonly used method, and Method 3 is an equally weighted average method. The details about each method are given below.

Method 1. Let AIC_i be the Akaike information criterion (AIC) value, which is a measure of the relative goodness of fit of a statistical model, and let w_i be the weight of the ith model ($i = 1, \ldots, k$). Then if all the AIC values are negative, let $w_i = AIC_i / \sum_{j=1}^{k} AIC_j$; If all the AIC values are positive, let $w_i = (1/AIC_i) / \sum_{j=1}^{k} (1/AIC_j)$; If some of the AIC values are positive and some are negative, let o_i be the rank of AIC_i such that a smaller AIC corresponds to a larger rank. Then $w_i = o_i / \sum_{j=1}^{k} o_j$. Finally, we believe the observed time series contains more information than the simulated time series. Therefore, the final combined forecasts are given by

$$\tilde{y}_{n+\ell} = \frac{1}{2} \left(\sum_{i=1}^{k} w_i y_{n+\ell}^{(i)} + \hat{y}_{n+\ell} \right), \quad \ell = 1, \ldots, M, \tag{2}$$

where $y_{n+\ell}^{(i)}$ is the ℓ-step ahead forecast obtained by using the ith simulated time series, and $\hat{y}_{n+\ell}$ is the ℓ-step ahead forecast obtained by using the observed time series.

Method 2. This method uses the AIC weights defined by (see, for example, Burnham and Anderson, 2002)

$$w_i = e^{-0.5*AIC_i} / \sum_{j=1}^{k} e^{-0.5*AIC_j}, \quad i = 1, \ldots, k.$$

Similar to Method 1, the final combined forecast can be calculated by using (2).

Method 3. This method uses equal weights. So the final ℓ-step ahead combined forecast is the average value of those ℓ-step ahead forecasts obtained from both the simulated and observed time series.

Note that, in Method 1 and Method 2, a model with a smaller AIC value always corresponds to a larger weight, hence will make a bigger contribution to the final forecasts. The rationale behind this is that models with smaller values of AIC are favored on the basis of fit and parsimony.

4

It is worth mentioning that in the literature of time series analysis, it is also possible to use some re-sampling schemes to obtain simulated time series, such as bootstrap methods. Härdle et al. (2003) gave a review on this topic. However, those types of simulation methods are very different from our method. One of the main differences is that those methods cannot simultaneously guarantee the similarities of the autocorrelation structures and the marginal distributions between the observed and simulated time series, but our method can. Therefore, when treating simulated time series as data coming from different sources, our simulated time series are of more required properties. However, a thorough investigation on different forecasting methods needs to be carried out in the future.

3 Simulation studies

Many simulation studies have been conducted in order to investigate the performance of the forecasting method. Each time we have similar results. To save space, we only show two simulation studies in this section.

3.1 Simulation study 1

Consider a data generating process

$$y_t = 0.1y_{t-1} + 0.4y_{t-2} + 0.02y_{t-3} + \epsilon_t, \tag{3}$$

where ϵ_t are i.i.d. with density function given by

$$f(\epsilon) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\epsilon^{a-1}(1-\epsilon)^{b-1}, \quad a > 0, \ b > 0,$$

where $a = 0.1$ and $b = 2$, which were chosen arbitrarily. We used the statistical software R to simulate a time series of length 550 from model (3). In the following we only used the first 500 values for obtaining forecasts but leave the last 50 values for checking the forecasting accuracy. So the last 50 values will not be used to produce combined forecasts. We call this time series the observed time series and show it in Figure 1(a) (darker curve). It is seen that the time series has many large values. Hence the marginal distribution should be skewed to the right.

By applying the developed method to the observed time series, we obtained the estimated marginal distribution (1) with

$$G_1(y) = 1 - e^{-(y/0.0837)^{0.8856}}, \quad G_2(y) = 1 - \left[1 - \frac{0.3494(y - 0.3)}{0.2256}\right]^{1/0.3494}$$

for $y > 0.3$, $1 - \frac{0.3494(y-0.3)}{0.2256} > 0$ and $\beta = P(y \le 0.3) = 0.93$.

The estimated base process is given by

$$z_t = \sum_{j=1}^{10} \hat{\phi}_j z_{t-j} + u_t,$$

5

Table 1: Estimated parameter values of the base processes

	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\phi}_3$	$\hat{\phi}_4$	$\hat{\phi}_5$	
Simulation 1	0.0197	0.4269	0.0033	0.0023	-0.0054	
Simulation 2	0.3935	0.2039	0.0557	0.0927	0.0307	
	$\hat{\phi}_6$	$\hat{\phi}_7$	$\hat{\phi}_8$	$\hat{\phi}_9$	$\hat{\phi}_{10}$	$\hat{\sigma}^2$
Simulation 1	-0.0546	-0.0144	-0.0470	0.1099	0.0143	0.8040
Simulation 2	-0.0968	0.0131	0.0461	0.0292	0.0109	0.5923

where $\hat{\phi}_j$ can be found in Table 1.

We generated 1001 independent time series z_{it} ($i = 1, \ldots, 1001$, $t = 1, \ldots, 500$) each of length 500 from the estimated base process and then transformed them into the y_{it} processes. For convenience, the first simulated time series y_{1t} was used for comparing the marginal distributions of the observed and the simulated time series. Figure 1(a) also shows the simulated time series y_{1t} (lighter curve). It is seen that both time series look very similar. Figure 1(b) checks the marginal distributions: the histogram corresponds to the observed time series and the continuous curve corresponds to the estimated density function of the simulated time series y_{1t}. Obviously, they are also in a good agreement with each other. This is further confirmed by Figure 1(e) which shows the QQ-plot of the observed and the simulated time series. To check the auto-correlation structure, we produced the autocorrelation function (acf) plot of the observed time series up to lag 10 in Figure 1(c). Then, we calculated the average acf values over the remaining 1000 simulated time series and plotted the average acf in Figure 1(d). Clearly, on average, the auto-correlation structure of the simulated time series is very similar to that of the observed time series.

To check the forecasting performance of the method, we first simulated another 100 time series each of length 500 from the estimated base process. Then, by using the statistical software R, we obtained the best fitted ARIMA model to each of the 100 simulated and the observed time series and predicted future values up to 50 steps ahead. Finally, we combined the forecasts by using different methods presented in step (d) of the developed forecasting method. The accuracy of the combined forecasts were measured by RMSE and MdAPE shown in Table 2, where Method 4 corresponds to the results obtained from the model fitted to the observed time series only. It is seen that, under RMSE, the combined forecasts obtained from Method 1 and Method 2 are at least as good as those obtained from Method 4, while those from Method 3 are worse. On the other hand, under MdAPE, all the MdAPE values for this simulation study are significantly less than that for Method 4, suggesting that combining does improve the accuracy of forecasts in this study.

3.2   Simulation study 2

A time series of length 550 was simulated from the model

$$y_t = 0.7y_{t-1} + 0.1y_{t-2} + \epsilon_t - 0.3\epsilon_{t-1}, \tag{4}$$

where ϵ_t are i.i.d random variables following an exponential distribution with rate 2. Again we leave the last 50 values for checking the forecasting accuracy only. Similarly, we call
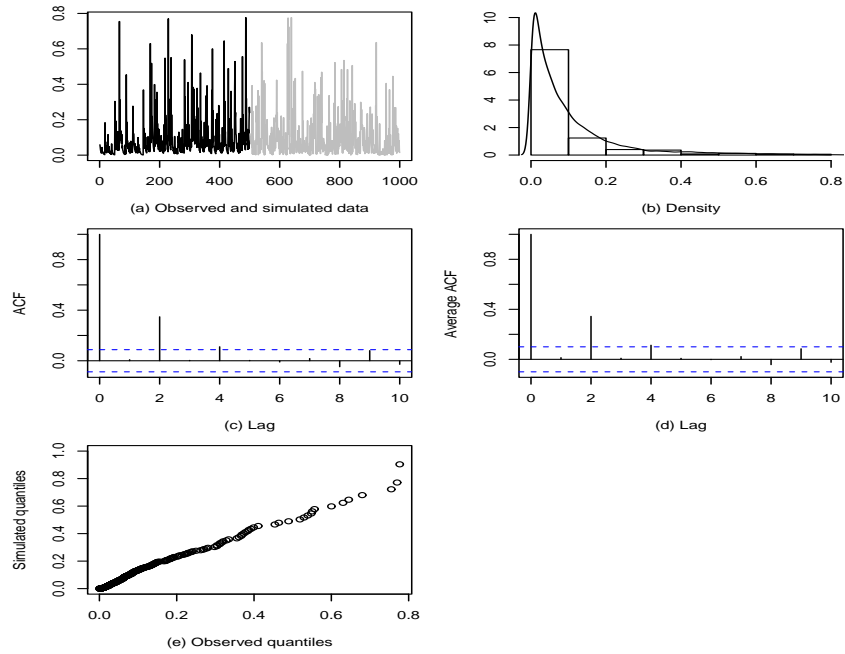
Figure 1: (a) Observed (darker curve) and simulated (lighter curve) time series. (b) Marginal distributions. (c) (d) auto-correlation structures. (e) QQ-plot of the observed and the simulated time series in Simulation study 1.

Table 2: RMSE and MdAPE values between forecasts and actual observed data in the simulation studies

	RMSE			
	Method 1	Method2	Method 3	Method 4
Simulation study 1	0.143	0.143	0.145	0.143
Simulation study 2	0.670	0.666	0.673	0.671
	MdAPE			
	Method 1	Method2	Method 3	Method 4
Simulation study 1	94.66%	83.90%	86.9%	108.7%
Simulation study 2	22.58%	22.67%	22.96%	22.64%

this series the observed time series and show it in Figure 2(a) (darker curve).

The estimated marginal distribution of the observed time series is given by (1), where

$$G_1(y) = 1 - e^{-(y/2.0019)^{2.8494}}, \quad G_2(y) = 1 - \left[1 - \frac{0.3660(y-3)}{0.7773}\right]^{1/0.3660}$$

for $y > 3$, $1 - \frac{0.3660(y-3)}{0.7773} > 0$ and $\beta = P(y \leq 3) = 0.952$.

The same base process as that in Simulation study 1 was used with the estimated parameters also shown in Table 1. Figure 2 checks the marginal distributions and the average auto-correlation structure of the simulated time series, which shows that they are very similar to those of the observed time series.
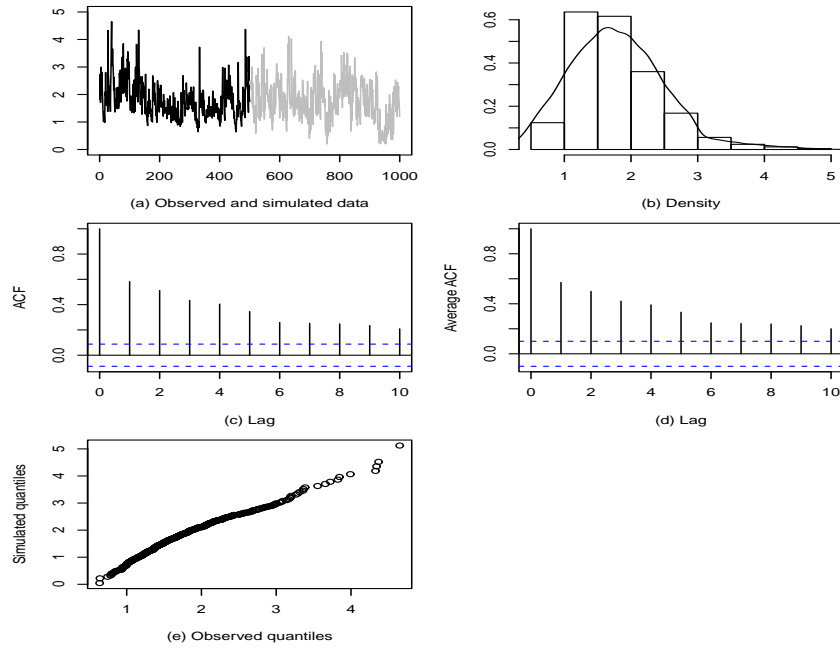

Figure 2: (a) Observed (darker) and simulated (lighter) time series. (b) Marginal distributions. (c) (d) auto-correlation structures. (e) QQ-plot of the observed and the simulated time series in Simulation study 2.

Table 2 also shows the performance of the combined forecasts in this simulation study. Again, under RMSE, methods 1 and 2 perform better than methods 3 and 4, while under MdAPE, all methods performed similarly, but Method 1 is the best and Method 2 is only 0.03% larger than that of Method 4. So according to MdAPE, Method 1 is the best in this simulation study.

In summary, the simulation studies show that the simulated time series are very similar to the observed time series and can be used for producing combined forecasts. In general the combined forecasts obtained from Method 1 and Method 2 are better than those obtained from Method 3 and those obtained from the observed time series only.

Table 3: Estimated parameter values of the base processes for HSI and FIN time series

			HSI time series			
$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\phi}_3$	$\hat{\phi}_4$	$\hat{\phi}_5$	$\hat{\phi}_6$	$\hat{\phi}_7$
0.6422	0.1296	0.1171	0.0370	0.0536	−0.0691	−0.1411
$\hat{\phi}_8$	$\hat{\phi}_9$	$\hat{\phi}_{10}$	$\hat{\phi}_{11}$	$\hat{\phi}_{12}$	$\hat{\phi}_{13}$	$\hat{\phi}_{14}$
−0.1157	0.1059	0.1200	−0.0115	0.0487	−0.0523	−0.0497
$\hat{\phi}_{15}$	$\hat{\phi}_{16}$	$\hat{\phi}_{17}$	$\hat{\phi}_{18}$	$\hat{\phi}_{19}$		
−0.0293	−0.020	0.0340	0.0299	0.0162		
			FIN time series			
$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\phi}_3$	$\hat{\phi}_4$	$\hat{\phi}_5$	$\hat{\phi}_6$	$\hat{\phi}_7$
0.4993	0.2107	0.1014	0.0614	-0.0168	−0.0433	0.0009
$\hat{\phi}_8$	$\hat{\phi}_9$	$\hat{\phi}_{10}$	$\hat{\phi}_{11}$	$\hat{\phi}_{12}$	$\hat{\phi}_{13}$	$\hat{\phi}_{14}$
0.0676	0.0645	−0.0788	−0.0340	0.0061	0.0389	−0.0184
$\hat{\phi}_{15}$	$\hat{\phi}_{16}$	$\hat{\phi}_{17}$	$\hat{\phi}_{18}$	$\hat{\phi}_{19}$		
0.0227	−0.0476	0.0895	0.0074	−0.0406		

4 Applications

In this section, we apply the forecasting method to the dividend yields series of the Hang Seng Index (HSI) and its two sub-indexes: Finance (FIN) and Commerce & Industry (C&I). HSI is one of the earliest stock market indexes in Hong Kong and has become the most widely quoted indicator of the performance of the Hong Kong stock market. The time series are the monthly data recorded from 1986 to 2002 and were retrieved from the database in the website of the Hang Seng index Services Limited. The time series plots of the observed series are given in Figure 3.

Note that in the following, we only used the first 192 values to obtain forecasts and leave the last 12 observed values for checking the forecasting accuracy.

The estimated marginal distribution for HSI has

$$G_1(y) = 1 - e^{-(y/0.0459)^{3.3470}}, \quad G_2(y) = 1 - \left[1 - \frac{0.4444(y - 0.06)}{0.0159}\right]^{1/0.4444}$$

for $y > 0.06$, $1 - \frac{0.4444(y-0.06)}{0.0159} > 0$ and $\beta = P(Y \leq 0.06) = 0.8333$.

The base process is estimated by

$$z_t = \sum_{i=1}^{19} \hat{\phi}_i z_{t-i} + u_t, \tag{5}$$

where u_t are i.i.d. $N(0, 0.3235)$, and the estimated parameters are given in Table 3.

The estimated marginal distribution for FIN has

$$G_1(y) = 1 - e^{-(y/0.0397)^{4.4550}}, \quad G_2(y) = 1 - \left[1 - \frac{1.1262(y - 0.05)}{0.0101}\right]^{1/1.1262}$$

for $y > 0.05$, $1 - \frac{1.1262(y-0.05)}{0.0101} > 0$ and $\beta = P(Y \le 0.05) = 0.9375$.

The base process takes the same form as that given by (5), where u_t are i.i.d. $N(0, 0.3463)$, and the estimated parameters are also given in Table 3.

For C&I series, the estimated marginal distribution has

$$G_1(y) = 1 - e^{-(y/0.0335)^{2.9977}}, \quad G_2(y) = 1 - \left[1 - \frac{0.5655(y - 0.047)}{0.0049}\right]^{1/0.5655}$$

for $y > 0.047$, $1 - \frac{0.5655(y-0.047)}{0.0049} > 0$ and $\beta = P(Y \le 0.047) = 0.9375$.

The base process is estimated by

$$z_t = 0.5253 z_{t-1} + 0.3885 z_{t-2} + u_t,$$

where u_t are i.i.d. $N(0, 0.2227)$.

We simulated a time series from each of the base processes and transformed them into the required time series. Figure 3 shows three observed time series (dark curves) followed by the simulated time series (grey curves). It also shows the QQ-plot of the simulated and the observed time series. Obviously, the marginal distributions of the observed and the simulated time series are very similar.
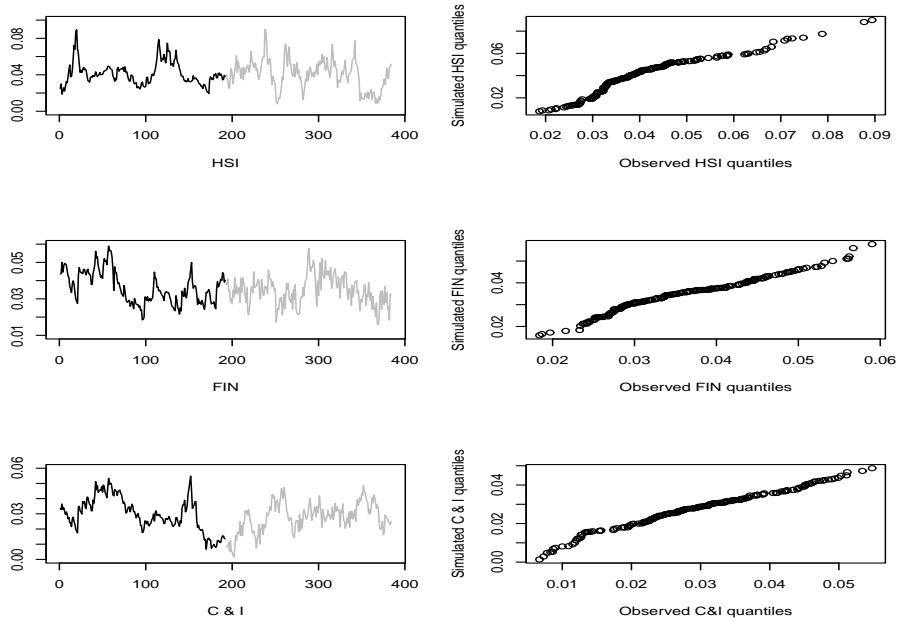

Figure 3: Observed (darker curve) and simulated (lighter curve) HSI, FIN and C&I time series and the corresponding QQ-plots.

To check the autocorrelation structure of the simulated and the observed time series, we obtained 1000 simulated time series each of size 192 for each of the HSI, FIN and C&I time series, calculated the acf up to lag 10 for each simulated time series, obtained the average acf over the 1000 simulated time series, and calculated the acf of the observed series. The
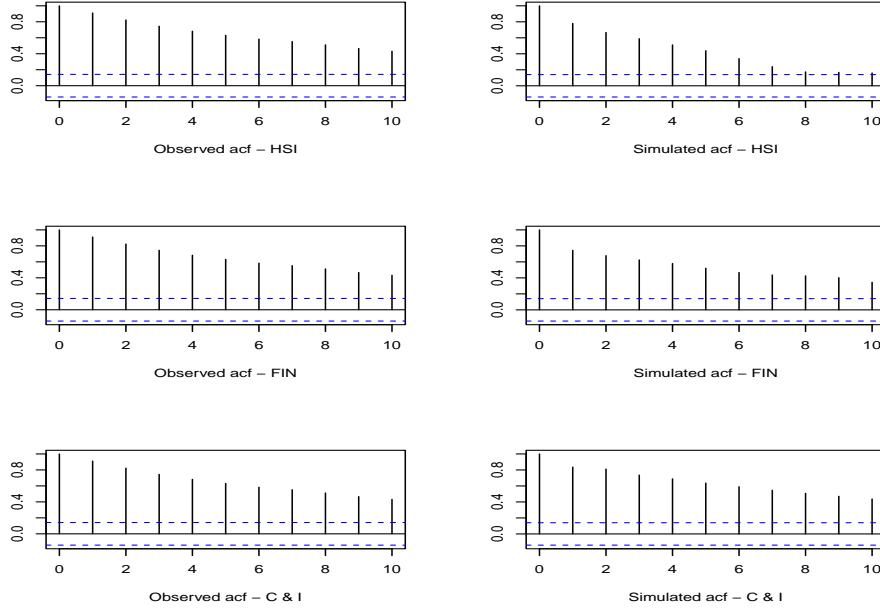
Figure 4: Autocorrelation structures of the simulated and the observed HSI, FIN and C&I time series.

results are shown in Figure 4. Clearly, on average, the autocorrelation structures of the simulated and the observed time series are also very similar.

To check the forecasting performance, we simulated three sets of time series for each of the observed time series. Each set contains 200 time series, and each time series in the first set has length $m = 192$ (i.e. the length of the observed time series), in the second set has $m = 300$ and in the third set $m = 400$. Then we fitted a sequence of $ARIMA(p, d, q)$ models with $p, q = 0, 1, 2, 3, 4$ and $d = 1$ to each time series. By using the best fitted models for each time series we obtained forecasts up to 12-step ahead. These forecasts were then combined using the methods given in Section 2 to produce final conditional forecasts for each observed time series. Finally, the combined forecasts were compared with the actual observed data according to RMSE and MdAPE. The results are given in Table 4, where the row named "Average" provides the column average RMSE and MdAPE for each method among different observed time series.

Table 4 shows that on average, the performance of the method depends on the value of m. It is seen that the accuracy does not increase as the value of m increases and the best results correspond to $m = 300$ in these applications. This is because when m is much larger than the observed time series, more extreme values would be expected in the simulated time series. In these applications, the forecasts obtained from the simulated time series were based on the ARIMA models which are not appropriate for time series with many extreme values. Therefore, we would expect that when m is much larger than the observed time series, other time series models should be used, and we leave this for further investigations in the future.

Table 4 also shows that on average the forecasts obtained from Method 4 are worse than the combined forecasts obtained by using methods 1 and 2 when $m \leq 300$. Note that

11

Table 4: RMSE and MdAPE values between the actual observed data and the forecasts obtained from the best fitted models for HSI, FIN and C&I time series.

	$m = 192$							
	RMSE				MdAPE			
Methods	1	2	3	4	1	2	3	4
HSI	0.008	0.010	0.009	0.009	14.09%	20.09%	17.31%	16.76%
FIN	0.008	0.007	0.010	0.007	16.15%	14.60%	20.95%	15.10%
C&I	0.004	0.003	0.009	0.004	18.50%	8.86%	50.87%	20.96%
Average	0.007	0.007	0.009	0.007	16.25%	14.52%	29.71%	17.61%
	$m = 300$							
	RMSE				MdAPE			
Methods	1	2	3	4	1	2	3	4
HSI	0.008	0.009	0.007	0.009	14.21%	17.58%	13.74%	16.76%
FIN	0.007	0.005	0.009	0.007	15.97%	9.97%	20.48%	15.10%
C&I	0.004	0.002	0.010	0.004	19.23%	11.50%	52.68%	20.96%
Average	0.006	0.005	0.009	0.007	16.47%	13.02%	28.97%	17.61%
	$m = 400$							
	RMSE				MdAPE			
Methods	1	2	3	4	1	2	3	4
HSI	0.008	0.011	0.008	0.009	15.32%	21.00%	15.23%	16.76%
FIN	0.008	0.008	0.010	0.007	16.47%	17.49%	21.30%	15.10%
C&I	0.004	0.008	0.011	0.004	18.34%	41.27%	57.64%	20.96%
Average	0.007	0.009	0.010	0.007	16.71%	26.59%	31.39%	17.67%

among different combining methods, the average performances of Methods 1 and Method 2 are better than others in both the simulation studies and the applications. This is because Method 3 uses equal weights, so both good models and bad models make equal contributions to the combined forecasts. In our case, we used the simulated and the observed time series, and we know the simulated time series do not contain all the information that the observed time series has. Therefore, it is reasonable to expect that the model based on the observed time series should make bigger contributions to the combined forecasts. Hence, using unequally weighted methods should be more appropriate here.

To check the effects of the models used for producing forecasts, we also carried out the following investigation. Consider the HSI time series first. Three sets of simulated time series based on the observed HSI were obtained. Each set contains 100 time series and each series has length $m = 192, 300$ and 400 for the first, second and third set respectively. For each simulated time series, we fitted a sequence of ARIMA(p, d, q) models, where $d = 1$, $p, q = 0, 1, \ldots, 4$. Then all fitted models were used to obtain combined forecasts up to 12-steps ahead for HSI, which were then compared with the actual observed values via RMSE and MdAPE. Finally, the average RMSE and MdAPE were calculated for different methods and different m values respectively. Similar work was done for FIN and C&I series. Table 5 shows the RMSE and MdAPE values, where the row named "Average" corresponds to the column average of the RMSE and MdAPE values, i.e. the average among the observed time series, which further confirmed that the average performances of methods 1 and 2 are much better than that of Method 3 and that of the method based on the observed time series only. Furthermore, all the column average RMSE values in Table 5 are less than those in Table 4, but no significant improvements for the average MdAPE values. This may suggest that using the best model or using all the fitted models in the same class of time series models may not help significantly in improving the forecasting accuracy, and this is also in good agreement with Collopy and Armstrong's (1989) argument, that is, when combining, the models should be as different as possible.

It is worth emphasizing that the above fitted $AR(19)$ and $AR(2)$ models are not the models for the original observed HSI, FIN and C&I data, but they are the base processes which were used to capture the autocorrelation structure and marginal distribution of the observed time series. The $AR(19)$ and $AR(2)$ processes were also used to obtain time series which will be transformed to the final simulated time series that we need. The order of the base process is decided by comparing the autocorrelation structures and the marginal distributions of the simulated and the observed time series in order to ensure that the simulated time series is of high quality. There is no need to check whether all the parameters of the base process are significant or not. For C&I process, we found that a much lower order of the base process works fine. Finally, we choose to forecast up to 12 steps ahead because the original time series are not long and 12 steps cover a whole year ahead forecasts.

5   Conclusions and discussion

In this paper we propose a combining forecasts method that uses both simulated and observed time series. The simulation studies and applications to real time series show that the forecasting method is able to produce more accurate forecasts compared with those obtained by using the observed time series only. This is because the simulated time series can

Table 5: Average RMSE and MdAPE values for different methods and different lengths of the simulated time series

	$m = 192$							
	RMSE				MdAPE			
Methods	1	2	3	4	1	2	3	4
HSI	0.008	0.008	0.006	0.009	14.04%	15.42%	11.50%	16.76%
FIN	0.007	0.007	0.008	0.007	16.06%	15.48%	17.51%	15.10%
C&I	0.004	0.003	0.012	0.004	21.14%	14.56%	64.52%	20.96%
Average	0.006	0.006	0.009	0.007	17.08%	15.15 %	31.18%	17.61%
	$m = 300$							
	RMSE				MdAPE			
Methods	1	2	3	4	1	2	3	4
HSI	0.008	0.008	0.007	0.009	14.36%	14.82%	12.28%	16.76%
FIN	0.007	0.006	0.008	0.007	16.02%	11.24%	17.49%	15.10%
C&I	0.004	0.002	0.011	0.004	19.27%	11.17%	60.71%	20.96%
Average	0.006	0.005	0.007	0.007	16.55%	12.41%	30.16%	17.61%
	$m = 400$							
	RMSE				MdAPE			
Methods	1	2	3	4	1	2	3	4
HSI	0.008	0.011	0.008	0.009	15.52%	20.54%	14.39%	16.76%
FIN	0.008	0.007	0.009	0.007	16.50%	16.07%	18.51%	15.10%
C&I	0.004	0.003	0.010	0.004	17.33%	11.16%	56.82%	20.96%
Average	0.007	0.007	0.009	0.007	16.45%	15.92 %	29.91%	17.61%

be treated as new data coming from different sources, and this is also because individual forecasts are obtained from different models. Therefore, the combined forecasts contain information from both different sources and different models, which increases the forecasting accuracy.

As the method requires that the marginal distribution does not depend on time, it is not appropriate to apply the method to any observed time series that do not satisfy this condition. On the other hand, the method can be used to time series with any auto-correlation structures. We found that the marginal distribution proposed in the paper is a suitable candidate for the HSI, FIN and C&I time series. However, for other types of time series, the method for obtaining the simulated time series needs to be modified so that high quality of simulated time series can be obtained.

In this paper we only showed the results corresponding to the base process which takes an AR form of time series models. The order of the base process can be determined according to the QQ-plots and the acf plots such as those shown in Figure 3 and Figure 4. Our experience with the method shows that after only several test runs an appropriate order of the base process can be chosen. In practice, we often know the maximum lag L value of the auto-correlation function that we are interested in. In this case, we could choose this maximum lag value as the order of the base process.

Our study suggests that the performance of the developed method depends both on the length of the simulated time series and on the models used for producing forecasts. We feel that the optimal length of simulated time series is certainly worth further investigation in the future.

It is worth mentioning that the performance of the developed forecasting method also depends on the quality of the simulated time series, while the quality of the simulated time series depends on the quality of the estimated autocorrelation structure and the estimated marginal distribution of the observed data, which requires suitable length of observed series. Therefore, if the observed time series are very short, for example, with only 8 observations or fewer, the developed method should not be used, but the damped trend and no-change models could be considered. See, for example, Collopy and Armstrong (1989).

Graefe et al. (2010) pointed out that forecasts from different methods that use different data may lead to bracketing and low correlations of errors, and in such situations, the gains from combining can be substantial. However, few studies directly analyzed this ex ante condition. One of the difficulties in carrying out such direct analysis might be the availability of the required data sets. This paper suggests that the simulated time series can be used for this purpose and that combining does help in improving forecasting accuracy. However, a thorough analysis on this ex ante condition by using simulated time series needs to be carried in the future.

15

References

[1] Armstrong JS. 2001. *Principles of forecasting: a handbook for researchers and practitioners*. Kluwer Academic Publishing.

[2] Armstrong JS and Collopy F. 1992 . Error measures for generalizing about forecasting methods: empirical comparisons. *International Journal of Forecasting* **8**, 69-80.

[3] Batchlor D, Dua P. 1995. Forecaster diversity and the benefits of combining forecasts. *Management Science* **41**: 68-75.

[4] Box G, Jenkins GM, Reinsel G. 2006. *Time Series Analysis: Forecasting & Control*, 4th Edition, Wiley.

[5] Bunn D, Wright G. 1991. Interaction of judgemental and statistical forecasting methods: issues and analysis. *Management Science* **37**: 501-518.

[6] Burnham KP, Anderson DR. 2002. *Model selection and multimodel inference: a practical information-theoretic approach*. Second edition. Springer-Verlag, New York, USA.

[7] Cai Y. 2003. Convergence Theory of a Numerical Method for Solving Chapman Kolmogorov Equation. *SIAM Journal on Numerical Analysis* **40**: 2337-2351.

[8] Cai Y. 2005. A Forecasting Procedure for Nonlinear Autoregressive Time Series Models. *The Journal of Forecasting* **24**: 335-351.

[9] Cai Y. 2007. A Quantile Approach to US GNP. *Economic Modelling* **24**: 969-979.

[10] Cai Y. 2010. Forecasting for quantile self exciting threshold autoregressive time series models. *Biometrika*, 97, 199 - 208.

[11] Cai Y, Gouldby B, Hawkes P, Dunning P. 2008. Statistical Simulation of Flood Variables: Incorporating Short-Term Sequencing. *Journal of Flood Risk Management* **1**: 1-10.

[12] Clemen RT. 1989. Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting* **5**: 559-583.

[13] Collopy, F and Armstrong, JS. 1989. Toward Computer-Aided Forecasting Systems: Gathering, Coding, And Validating The Knowledge. *DSS-89 Transactions: Ninth International Conference On Decision Support Systems*, ed. George R. Widmeyer, Institute of Management Sciences, 1989, 103-119.

[14] Graefe, A, Cuzn, AG, Jones, RJ and Armstrong JS. 2010. Combining Forecasts: An Application U.S. Presidential Elections.

http://uwf.edu/govt/documents/I.B.20-CuzanandGraefeArmstrongJones-2010-Combining %20forecasts-An%20application%20to%20U.S.%20Presidential%20 Elections-March%2016, %202010.pdf

[15] Härdle W, Horowitz J, Kreiss JP. 2003. Bootstrap Methods for Time Series. *International Statistical Review* **71**: 435-459.

[16] Kumar TSV, Krishnamurti TN, Fiorno Mn Nagata M. 2003. Numerical prediction of Typhoon tracks and intensity using a multimodel superensemble. *Monthly Weather Review* **131**: 574-583.

[17] McCracken M, Clark T. 2008. Combining forecasts from nested models. *FRB of Kansas City Research Working Paper*.

[18] McNees SK. 1992. The uses and abuses of "consensus" forecasts. *Journal of Forecasting* **11**: 703-710.

[19] Tong H. 1990. *Non-linear time series*. Oxford: Clarendon Press.

[20] Wang W, Nie S. 2008. The Performance of Several Combining Forecasts for Stock Index. *Future Information Technology and Management Engineering, 2008. FITME '08. International Seminar on*. 450-455.

[21] Winkler RL, Poses RM. 1993. Evaluating and combining physicians' probabilities of survival in an intensive care unit. *Management Science* **39**: 1526-1543.