Journal of Applied Sciences

ISSN 1812-5654

science alert

ANSInet
an open access publisher
http://ansinet.com

Journal of Applied Sciences 13 (22): 5186-5191, 2013
ISSN 1812-5654 / DOI: 10.3923/jas.2013.5186.5191

Simple Average for Linear Combined Forecasting Weights Method

Zhenxing Yin
Jiangsu Automation Research Institute, Lianyungang, 222061, Jiangsu, China

Abstract: This algorithm explains the time series forecasting as a measure. At the time, the optimal combined forecasting using each method can be defined as the measurement of the actual strike after the true value problem. It theoretical correlation coefficient estimation bias affects forecasting values. The optimal weights of linear combined forecasting were deduced theoretically. It can be proved that the simple average is the superior weights method of the linear combined forecasting. Particularly, based on robust statistical theory, the superiority of the simple average is proved by mathematical deduction and numerical test.

Key words: Simple average, optimal weights, linear combined forecasting

INTRODUCTION

The combined forecasting more and more attention in recent years. The combined forecasting is appropriate combination of different forecasting methods for utilization of various forecasting methods to provide information, as much as possible to improve prediction accuracy. The basic idea is to first use of a plurality of specific single model to predict and get their predictions and then combine the results of these projections a better result. The simple average performed well compared to models based on past performance data and performance improved in general as more forecasts were included in the consensus (Clemen and Winkler, 1999). So far, the linear combined forecasting weight optimization is a hot spot theory being studied.

COMBINED FORECASTING

Actual complex time series forecasting is a difficult problem. Complex combined forecasting method is the actual time series forecasting a new stage of development (Hibon and Evgeniou, 2005; Eklund and Karlsson, 2007). The linear combined forecasting can be expressed as.

For the time series, t = 1, 2,…, N methods using the predicted results:

$$y_{it} = x_t + \varepsilon_{it},\ i=1,\ 2,\dots\ N,\ t=1,2,\dots,\ T \tag{1}$$

Here, y_{ij} is the i-th test phase forecasting method T a predictive value. ε_{it} is residual that mean value is 0., The correlation of any two residuals between ε_{it} and ε_{jt} coefficient $\rho_{ij} \neq 0$, i, j = 1, 2,…, N, I ≠ j.

In general, it does not assume that ε_{it} is normally distributed. ε_{it} obey arbitrary distribution. It becauses that it is difficult to determine the specific distribution pattern of the noise an forecasting error. Thought assumed to follow a normal distribution is reasonable, when the T is larger.

The results of the linear combined forecasting is:

$$\hat{y}_t = k_1 y_{1t} + k_2 y_{2t} + \cdots + k_N y_{Nt},\ \ t = 1,\ 2,\dots,\ T \tag{2}$$

The optimal weights of linear combined forecasting became find $\{k_1,\ k_2,\ \cdots,\ k_N\}$, that $\hat{\varepsilon}_t$ of the:

$$\hat{y}_t = x_t + \hat{\varepsilon}_t,\ t = 1,\ 2,\dots,\ T \tag{3}$$

is in the minimum variance.

The ε_{it} is the residual of a forecasting result. Its mean may not 0, when the linear combined forecasting is biased. In this case, it can analyze the historical data to estimate the residual mean of ε_{it} and then to be corrected. The variance of the residual ε_{it} generally consists of two parts. One part is the noise in the historical data which can't been forecasted. Another part comes from the forecasting of the defect. Any kind of forecasting method has its own shortcomings, a single method to predict will cause additional errors. The combined forecasting is powerless for the first part factor and it forms predicted limits. It can improve the forecasting methods and use of combined forecasting which can significantly reduce the forecasting error.

IMPACT FACTORS OF THE COMBINED FORECASTING

The main factors affecting combined forecasting accuracy rate is.

- Sample size is limited Table 1 shows the confidence intervals for the population parameter estimate of the normal distribution. Table 2 shows the confidence interval for robust parameter estimation. They show that parameter estimates and confidence intervals are directly related to the sample size n. The length of the confidence interval is inversely proportional to $\sqrt{n}$

  The length of historical data big impacts forecasting accuracy and forecasting reliability which are based on empirical. Although, theoretically it can easily obtain a sufficiently long historical data, however, the actual history of complex systems data length is limited. It is because of the changing nature of such data. Old historical and present situation may have occurred behavioral changes in the nature. They have not been used.
- It unknowns probability distribution function of the forecasting error. In general, it use the normal distribution formula derived sample statistical methods. At this point, the sample quality is very good. But many times, the sample distribution can only approximate a normal distribution, even far away from the normal distribution. Under these circumstances, the robustness of the statistical methods lost and traditional statistical methods can't do anything. Box and Tukey illustrate some common estimation and testing became so bad in some approximation model. At this point, robust estimators are not precise enough (Clemen and Winkler, 1999)
- Smples contained anomalous data. Since, the large number of tests, it is difficult to avoid completely the individual neglect. Hampel noted that the actual data contained in 10% of the abnormal data is a common thing. It makes some classical statistical analysis become worthless, who contain a small amount of data outliers. It can lead to erroneous statistical conclusions. It can use robust statistical estimator to reduce the adverse effects of abnormal data. One way is to design an effective forecasting method to discover outliers in the data. Then put them removed. Another method is to design a method for the statistical estimation of the amount which can reduce the adverse effects of the abnormal sample data statistics estimation

Table 1: Confidence intervals for the population parameter estimate of the normal distribution

Estimated parameters	Known parameters	Confidence intervals
μ	σ	$\bar{X} - u_\alpha \frac{\sigma}{\sqrt{n}}, \bar{X} + u_\alpha \frac{\sigma}{\sqrt{n}}$
	-	$\bar{X} - t_\alpha \frac{S_X}{\sqrt{n-1}}, \bar{X} + t_\alpha \frac{S_X}{\sqrt{n-1}}$
σ	-	$n\frac{S_X^2}{\chi^2_{\alpha/2}}, n\frac{S_X^2}{\chi^2_{1-\alpha/2}}$
	μ	$\frac{\sum(X_i-\mu)^2}{\chi^2_{\alpha/2}}, \frac{\sum(X_i-\mu)^2}{\chi^2_{1-\alpha/2}}$

Table 2: Confidence interval for robust parameter estimation

Location	Scale estimator	Confidence interval
M	d_F	$M - \frac{t_{n-1}(d_F)}{(1.075\sqrt{n})}, M + \frac{t_{n-1}(d_F)}{(1.075\sqrt{n})}$
T_{bi}	S_{bi}	$T_{bi} - \frac{t_{.7(n-1)}S_{bi}}{\sqrt{n}}, T_{bi} + \frac{t_{.7(n-1)}S_{bi}}{\sqrt{n}}$

OPTIMAL WEIGHTS OF LINEAR COMBINED FORECASTING

In general, time series forecasting is divided into three phases. First, predictive equation should be build by the sufficient historical data. Then, the forecast data tests and predicted parity forecasting equation. If the forecasting method is validated pass, it can confirm that the forecasting equation is available. It is combined forecasting that the combined forecasting is a variety of predicted values according to the Eq. 2 weighted synthesis and finally let the Eq. 3 in the minimum variance of $\hat{\varepsilon}_t$. Then, the combined forecasting uses Eq. 2 to predict the future (Tsangari, 2007; Liang *et al*., 2006; Greer, 2005).

Thus seen, the problem can be interpreted as predictive measurement problems. At the time, the optimal combined forecasting using each method can be defined as the measurement of the actual strike after the true value problem. It can study the forecasting problem, by using of statistical methods for analysis of experimental error. Theorem 1 is optimal weights given formula.

Theorem 1: It assumes that using N kinds of forecasting methods. Variance of forecasting varies is $\sigma_i^2, i = 1, 2, \cdots, N$. The correlation coefficient of the forecasting is ρ_{ij}. The optimal combined forecasting coefficients is:

$$\begin{cases} k_1 = (b_{11} + b_{12} + \cdots + b_{1N})[(b_{11} + b_{12} + \cdots + b_{1N}) \\ +(b_{21} + b_{22} + \cdots + b_{2N}) + \cdots + (b_{N1} + b_{N2} + \cdots + b_{NN})]^{-1} \\ k_2 = (b_{21} + b_{22} + \cdots + b_{2N})[(b_{11} + b_{12} + \cdots + b_{1N}) \\ +(b_{21} + b_{22} + \cdots + b_{2N}) + \cdots + (b_{N1} + b_{N2} + \cdots + b_{NN})]^{-1} \\ \cdots \\ k_N = (b_{N1} + b_{N2} + \cdots + b_{NN})[(b_{11} + b_{12} + \cdots + b_{1N}) \\ +(b_{21} + b_{22} + \cdots + b_{2N}) + \cdots + (b_{N1} + b_{N2} + \cdots + b_{NN})]^{-1} \end{cases} \quad (4)$$

It is unbiased and variance is minimal. b_{ij} is the elements of the:

$$B = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nn} \end{bmatrix} =$$

$$A^{-} = \begin{bmatrix} 2\sigma_1^2 & 2\rho_{12}\sigma_1\sigma_2 & \cdots & 2\rho_{1N}\sigma_1\sigma_N \\ 2\rho_{12}\sigma_1\sigma_2 & 2\sigma_2^2 & \cdots & 2\rho_{2N}\sigma_1\sigma_N \\ \vdots & \vdots & \ddots & \vdots \\ 2\rho_{1N}\sigma_1\sigma_N & 2\rho_{2N}\sigma_2\sigma_N & \cdots & 2\sigma_N^2 \end{bmatrix}^{-}$$

Let residuals $\varepsilon_{1t}, \varepsilon_{2t}, \ldots, \varepsilon_{Nt}$ is the mean of 0 and variance is $\sigma_1^2, \sigma_2^2, \ldots, \sigma_N^2$. Thus, the forecasting residuals for each forecasting correlation coefficient is ρ_{ij}, $i \neq j$. $k_1, k_2, \ldots, k_N$ is weights of the linear combined of the forecasting. The mean of linear combined of the forecasting is:

$$\hat{\varepsilon}_t = \sum_1^T k_i \varepsilon_{it},\ t = 1, 2, \ldots, T$$

It is 0. Variance is:

$$\sigma^2 = \sum_1^n k_i^2 \sigma_i^2 + 2\sum_{i<j}\sum k_i k_j \rho_{ij}\sigma_i\sigma_j$$

The optimal weights of linear combined forecasting is $k_1 + k_2 + \cdots + k_N = 1$ and:

$$V = k_1^2\sigma_1^2 + k_2^2\sigma_2^2 + \cdots + k_n^2\sigma_n^2 + 2k_1k_2\rho_{12}\sigma_1\sigma_2 + 2k_1k_3\rho_{13}\sigma_1\sigma_3 + \cdots + 2k_{n-1}k_n\rho_{(n-1)n}\sigma_{n-1}\sigma_n) \quad (5)$$

is the smallest.

Putting the lagrange multiplier into the Eq. 5, it get that:

$$V = k_1^2\sigma_1^2 + k_2^2\sigma_2^2 + \cdots + k_n^2\sigma_n^2 + 2k_1k_2\rho_{12}\sigma_1\sigma_2 + 2k_1k_3\rho_{13}\sigma_1\sigma_3 + \cdots + 2k_{n-1}k_n\rho_{(n-1)n}\sigma_{n-1}\sigma_n)$$

is the smallest:

$$\frac{\partial V}{\partial k_1} = 2k_1\sigma_1^2 + 2k_2\rho_{12}\sigma_1\sigma_2 + \cdots + 2k_n\rho_{1n}\sigma_1\sigma_n - \lambda = 0$$

$$\frac{\partial V}{\partial k_2} = 2k_2\sigma_2^2 + 2k_1\rho_{12}\sigma_1\sigma_2 + \cdots + 2k_n\rho_{2n}\sigma_2\sigma_n - \lambda = 0$$

$$\frac{\partial V}{\partial k_n} = 2k_n\sigma_n^2 + 2k_1\rho_{1n}\sigma_1\sigma_n + \cdots + 2k_{n-1}\rho_{(n-1)n}\sigma_{n-1}\sigma_n - \lambda = 0$$

It can be shown that:

$$\begin{bmatrix} 2\sigma_1^2 & 2\rho_{12}\sigma_1\sigma_2 & \cdots & 2\rho_{1n}\sigma_1\sigma_n \\ 2\rho_{12}\sigma_1\sigma_2 & 2\sigma_2^2 & \cdots & 2\rho_{2n}\sigma_1\sigma_n \\ \vdots & \vdots & \ddots & \vdots \\ 2\rho_{1n}\sigma_1\sigma_n & 2\rho_{1n}\sigma_1\sigma_n & \cdots & 2\sigma_n^2 \end{bmatrix} \times \begin{bmatrix} k_1 \\ k_2 \\ \vdots \\ k_n \end{bmatrix} = \begin{bmatrix} \lambda \\ \lambda \\ \vdots \\ \lambda \end{bmatrix}$$

Let:

$$A = \begin{bmatrix} 2\sigma_1^2 & 2\rho_{12}\sigma_1\sigma_2 & \cdots & 2\rho_{1n}\sigma_1\sigma_n \\ 2\rho_{12}\sigma_1\sigma_2 & 2\sigma_2^2 & \cdots & 2\rho_{2n}\sigma_1\sigma_n \\ \vdots & \vdots & \ddots & \vdots \\ 2\rho_{1n}\sigma_1\sigma_n & 2\rho_{2n}\sigma_2\sigma_n & \cdots & 2\sigma_n^2 \end{bmatrix}$$

Suppose:

$$A = \begin{bmatrix} 2\sigma_1^2 & 2\rho_{12}\sigma_1\sigma_2 & \cdots & 2\rho_{1n}\sigma_1\sigma_n \\ 2\rho_{12}\sigma_1\sigma_2 & 2\sigma_2^2 & \cdots & 2\rho_{2n}\sigma_1\sigma_n \\ \vdots & \vdots & \ddots & \vdots \\ 2\rho_{1n}\sigma_1\sigma_n & 2\rho_{2n}\sigma_2\sigma_n & \cdots & 2\sigma_n^2 \end{bmatrix}$$

is not Singular Matrix, then:

$$B = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nn} \end{bmatrix}$$

is the inverse matrix of A.

Then:

$$\begin{bmatrix} k_1 \\ k_2 \\ \vdots \\ k_n \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nn} \end{bmatrix} \times * \begin{bmatrix} \lambda \\ \lambda \\ \vdots \\ \lambda \end{bmatrix} \quad (6)$$

is the optimal weights of linear combined forecasting.

Putting $k_1 + k_2 + \cdots + k_N = 1$ into the Eq. 6, it gets that theorem 1.

If only two forecasting methods consist of combined forecasting, it get that:

$$\begin{cases} k_1 = \dfrac{\sigma_2^2 - \rho_{12}\sigma_1\sigma_2}{\sigma_2^2 - 2\rho_{12}\sigma_1\sigma_2 + \sigma_1^2} \\ k_2 = \dfrac{\sigma_1^2 - \rho_{12}\sigma_1\sigma_2}{\sigma_2^2 - 2\rho_{12}\sigma_1\sigma_2 + \sigma_1^2} \end{cases} \quad (7)$$

is the optimal weights of linear combined forecasting.

In Bates and Granger (1969) gives the simplest case. This study gives combined forecasting formulas consisting of N forecasting methods.

Corollary 1: When correlation coefficient of every residuals (ε_{it}, ε_{ij}) is $\rho_{ij} = 0$, $i \neq j$, forecast errors are independent. Then, the weights of linear combined forecasting is:

$$k_i = \frac{1/\sigma_i^2}{1/\sigma_1^2 + 1/\sigma_2^2 + \cdots + 1/\sigma_N^2} \quad (8)$$

Corollary 2: When correlation coefficient of every residuals (ε_{it}, ε_{ij}) I s $\rho_{ij} = 0$, $i \neq j$, and every variance is equal, $(\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_N^2)$, forecast errors are independent. Then, the weights of linear combined forecasting is:

$$k_i = \frac{1}{N} \qquad (9)$$

In particular, the above proof is not required for any distribution are established. It gives the optimal weights of linear combining where the residuals of each forecasting method are not independent. The simple average is the optimal weights when the residuals are independent and identical.

OPTIMALITY OF THE SIMPLE AVERAGE

The sample size is limited. It unknowns the probability distribution function of the forecasting error. The smples contained anomalous data. These cases give us unable to confirm the sample variance and correlation coefficient. of equal Eq. 4 is rough and not credible. Therefore, the optimal weights of the linear combined forecasting relying on these two parameters are not accurate. The sample average do not need σ^2_i and ρ_{ij}. It can avoid this limitation, so its effect is often better than other weight calculation method (Koning *et al*., 2005; Makridakis and Hibon, 2000; De Menezes *et al*., 2000).

When correlation coefficient of every residuals $(\varepsilon_{it}, \varepsilon_{ij})$ is $\rho_{ij} = 0$, $i \neq j$ and every variance is not equal 0. Putting two kinds of forecasting into the Eq. 4, it gets that:

$$\varepsilon_{it}, \frac{d\hat{y}}{y} = \left(\frac{\partial k_1}{\partial \sigma_1} d\sigma_1 + \frac{\partial k_1}{\partial \sigma_2} d\sigma_2\right)\frac{y_1}{y} + \left(\frac{\partial k_2}{\partial \sigma_1} d\sigma_1 + \frac{\partial k_2}{\partial \sigma_2} d\sigma_2\right)\frac{y_2}{y}$$
$$= \left(\frac{-2\sigma_1\sigma_2^2}{\sigma_1^2+\sigma_2^2} d\sigma_1 + \frac{2\sigma_1^2\sigma_2}{\sigma_1^2+\sigma_2^2} d\sigma_2\right) y_1 + \left(\frac{2\sigma_1\sigma_2^2}{\sigma_1^2+\sigma_2^2} d\sigma_1 + \frac{2\sigma_1^2\sigma_2}{\sigma_1^2+\sigma_2^2} d\sigma_2\right) y_2 \qquad (10)$$

Equal Eq. 10 shows that σ_1 and σ_2 affect the contained anomalous values.

Figure 1 shows simulation figure of the $d\hat{y}/y$, when σ_1 and σ_2 change. Simulation results from two types of time series forecasting which are AR model and Support Vector Machine model. $\sigma_1 = 24.4478$ and $\sigma_2 = 28.9587$.

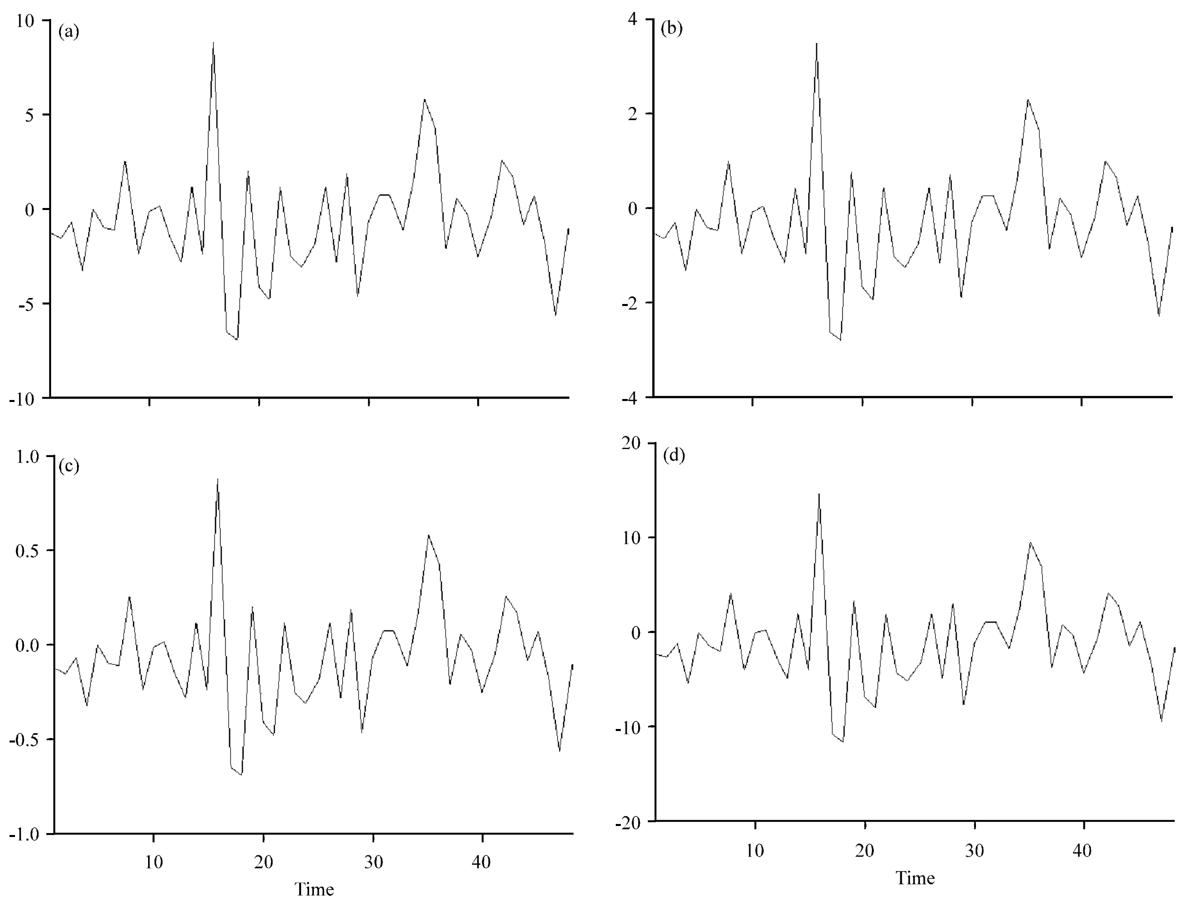

Fig. 1(a-d): Simulation figure of the $d\hat{y}/y$ (a) = 0.1, = -0.2, (b) = 0.1, = -0.2, (c) = 0.01, = -0.02 and (d) = 0.3, = -0.2, (b)

When correlation coefficient of every residuals (ε_{it}, ε_{ij}) is $\rho_{ij} = 0$, $i \neq j$, and every variance is not equal 0. Putting two kinds of forecasting into the Eq. 4.

It gets:

$$\begin{aligned}\frac{d\hat{y}}{y} &= \left(\frac{\partial k_1}{\partial \sigma_1}d\sigma_1 + \frac{\partial k_1}{\partial \sigma_2}d\sigma_2 + \frac{\partial k_1}{\partial \rho_{12}}d\rho_{12}\right)\frac{y_1}{y} \\ &+ \left(\frac{\partial k_2}{\partial \sigma_1}d\sigma_1 + \frac{\partial k_2}{\partial \sigma_2}d\sigma_2 + \frac{\partial k_2}{\partial \rho_{12}}d\rho_{12}\right)\frac{y_2}{y} \\ &= \Big(\frac{\rho_{12}\sigma_1^2\sigma_2 + \rho_{12}\sigma_2^3 - 2\sigma_1\sigma_2^2}{\sigma_2^2 - 2\rho_{12}\sigma_1\sigma_2 + \sigma_1^2} y_1 d\sigma_1 + \frac{2\sigma_1^2\sigma_2 - \rho_{12}\sigma_1\sigma_2^2 - \rho_{12}\sigma_1^3}{\sigma_2^2 - 2\rho_{12}\sigma_1\sigma_2 + \sigma_1^2} y_1 d\sigma_2 \\ &+ \frac{\sigma_1\sigma_2^3 - \sigma_1^3\sigma_2}{\sigma_2^2 - 2\rho_{12}\sigma_1\sigma_2 + \sigma_1^2} y_1 d\rho_{12}\Big) + \Big(\frac{2\sigma_1\sigma_2^2 - \rho_{12}\sigma_1^2\sigma_2 - \rho_{12}\sigma_2^3}{\sigma_2^2 - 2\rho_{12}\sigma_1\sigma_2 + \sigma_1^2} y_2 d\sigma_1 \\ &+ \frac{\rho_{12}\sigma_1\sigma_2^2 + \rho_{12}\sigma_1^3 - 2\sigma_1^2\sigma_2}{\sigma_2^2 - 2\rho_{12}\sigma_1\sigma_2 + \sigma_1^2} y_2 d\sigma_2 + \frac{\sigma_1^3\sigma_2 - \sigma_1\sigma_2^3}{\sigma_2^2 - 2\rho_{12}\sigma_1\sigma_2 + \sigma_1^2} y_2 d\rho_{12}\Big)\end{aligned} \quad (11)$$

Equal Eq. 11 shows that σ_1, σ_2 and ρ_{12} affect the contained anomalous values.

Figure 2 shows simulation figure of the $d\hat{y}/y$, when σ_1, σ_2 and ρ_{12} change. Simulation results from two types of time series forecasting which are AR model and Support Vector Machine model. $\sigma_1 = 24.4478$, $\sigma_2 = 28.9587$ and $\rho_{12} = 0.7$. Figure 2 shows that the contained forecasting value dramatic changes, when σ^2_i and ρ_{ij} change. Then, the contained forecasting value is credible.

The error of the sample average is:

$$\hat{\varepsilon}_t = \sum_{1}^{T} \frac{\varepsilon_{it}}{N},\ t = 1, 2, \ldots, T$$

Variance decreases with the combined forecasting method to participate in growing the number. This

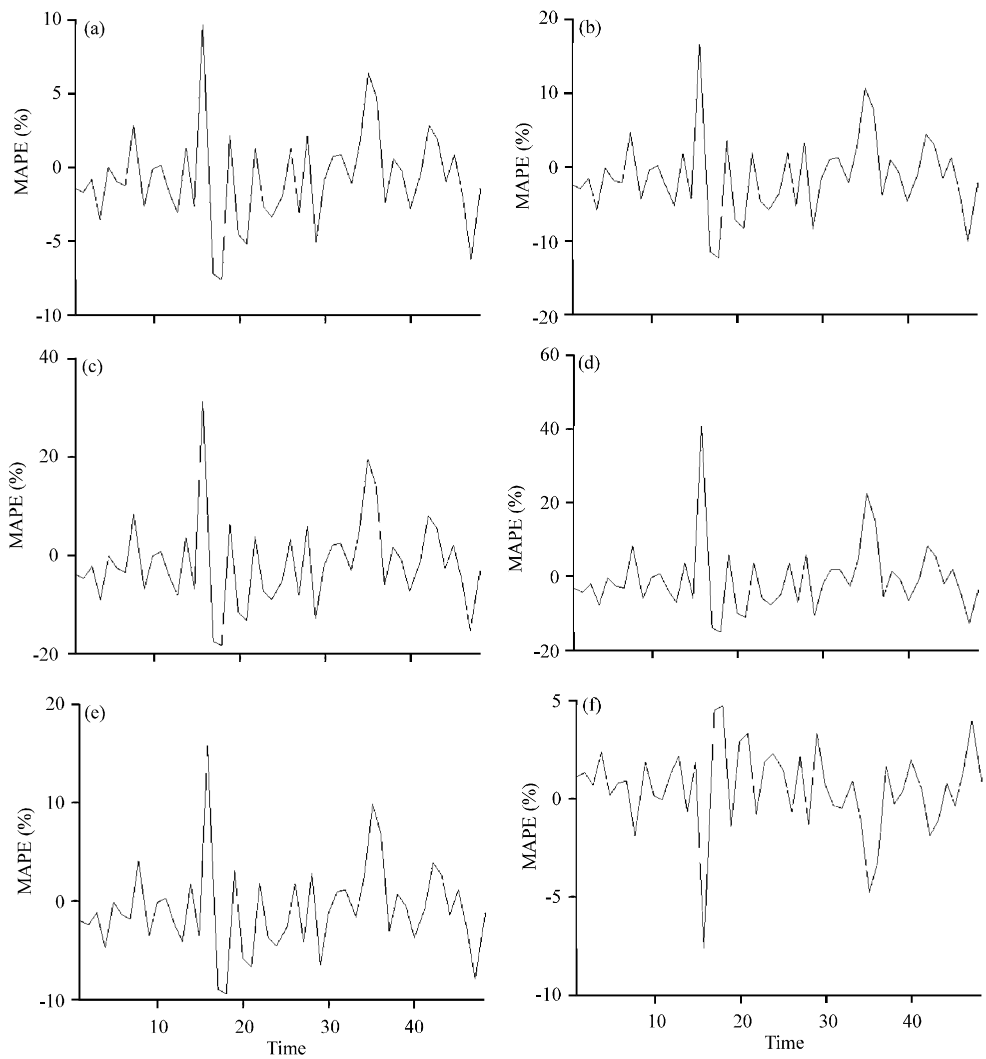

Fig. 2(a-f): Simulation figure of the $d\hat{y}/y$ (a) = 0.1, = 0.1, = 0.1, = -0.2, (b) = 0.5, = -0.1, = 0.1, = -0.2, (c) = 0.7, = 0.1, = 0.1, = -0.02, (d) = 0.9, = 0.1, = 0.1, = -0.2, (e) = 0.7, = 0.3, = 0.1, = -0.2 and (d) = 0.7, = 0.1, = 0.0, = 0.0

includes the white noise forecasting accuracy limits. The combined forecasting method reduces the individual forecasting additional error, but can't improve the formation of white noise error.

CONCLUSION

This study explains the time series forecasting as a measure. At the time, the optimal combined forecasting using each method can be defined as the measurement of the actual strike after the true value problem. Putting result of the mathematical statistician Dickinson into the time series, it 3 types optimal weights of linear combined forecasting. It theoretical solution to the optimal weights of linear combined forecasting and theoretical analysis of variance and correlation coefficient estimation bias affect forecasting values. It proves that the Simple Average is the Superior Weights Method of Linear Combined Forecasting.

ACKNOWLEDGMENT

The study is supported by Natural Science Foundation of Jiangsu (Project No. BK2012237).

REFERENCES

Clemen, R.T. and R.L. Winkler, 1999. Combining probability distributions from experts in risk analysis. Risk Anal., 19: 187-203.

Hibon, M. and T. Evgeniou, 2005. To combine or not to combine: Selecting among forecasts and their combinations. Int. J. Forecasting, 21: 15-24.

Eklund, J. and S. Karlsson, 2007. Forecast combination and model averaging using predictive measures. Econ. Rev., 26: 329-363.

Tsangari, H., 2007. An alternative methodology for combining different forecasting models. J. Applied Stat., 34: 403-421.

Liang, K.Y., J.C. Lee and K.S.H. Shao, 2006. On the distribution of the inverted linear compound of dependent F-variates and its application to the combination of forecasts. J. Applied Stat., 33: 961-973.

Greer, M.R., 2005. Combination forecasting for directional accuracy: An application to survey interest rate forecasts. J. Applied Stat., 32: 607-615.

Koning, A.J., P.H. Franses, M. Hibon and H.O. Stekler, 2005. The M3 competition: Statistical tests of the results. Int. J. Forecasting, 21: 397-409.

Makridakis, S. and M. Hibon, 2000. The M3-Competition: Results, conclusions and implications. Int. J. Forecasting, 16: 451-476.

De Menezes, L.M., D.W. Bunn and J.W. Taylor, 2000. Review of guidelines for the use of combined forecasts. Eur. J. Oper. Res., 120: 190-204.