

## Weighted Ranking Procedure for Combining Univariate Time Series Models

Sampson Ankrah<sup>\*</sup>, B.L. Peiris<sup>1</sup> and R.O. Thattil<sup>1</sup>

Postgraduate Institute of Agriculture  
University of Peradeniya  
Sri Lanka

**ABSTRACT.** *This paper extends the standard approach of combining forecast by proposing weights which are based on ranking the performance of forecast accuracy measures of models. These weights became necessary due to the problems associated with the Akaike weights, equal weights and forecast from the ‘best’ model selected by the minimum AICc value; which are pointed out in this study. According to a selection criterion, five models were fitted to the simulated dataset with two different sample sizes, n=25 and n=200. The results revealed that the mean squared forecast error (MSFE) from the combined forecast of the proposed weights (weighted ranking procedure) outperformed all other approaches that were investigated in this study. Furthermore, the three combined forecast approaches consistently outperformed the forecast from the best model selected by the minimum AICc. Thus, we recommend the use of the weighted ranking procedure in combining models.*

**Keywords:** *Akaike weight, forecast, information criteria, model selection, weighted ranking*

### INTRODUCTION

In time series analysis, one major interest is to be able to forecast the future values of a series from a ‘best’ model. This is to say that before forecasting, one is faced with a challenge of choosing the ‘best’ model among a variety of candidate models. The selection procedures of the ‘best’ model have several difficulties. Usually, one has to go through a series of evaluations in order to get the “best” model. The obvious difficulty is that there is no objective guideline for the choice of the size of the various tests involved in the selection procedures. Thus, little or no information is known about the errors associated with these procedures (i.e., after conducting a series of evaluations).

Our preliminary analysis and available literature indicated that, the model preferred by a test or information criterion does not necessarily perform better than other competing models in terms of prediction risk. In addition, one major drawback with model selection is its instability. Zou *et al.* (2004) argued that with a small or moderate number of observations, models close to each other are usually hard to distinguish and the model selection criteria are usually quite close to each other. Thus, a slight change of the data may result in the choice of a different model.

The unstable nature of model selection criteria often may inflate variability in the estimation or prediction. The instability of model selection has been recognized in statistics and related literature (Breiman, 1996). Chatfield, 2004; Hoeting *et al.* 1999; have used the term ‘model

---

<sup>1</sup> Department of Crop Science, Faculty of Agriculture, University of Peradeniya, Peradeniya, Sri Lanka

<sup>\*</sup> Corresponding author: sampson.ankrah@yahoo.com

uncertainty' to capture the difficulty in identifying the correct model. In addressing this challenge, combining forecast was introduced over the past three decades (Bates and Granger, 1969; Clemen, 1989) and various methods have been proposed. Thus, when there is a substantial uncertainty in finding the 'best' model, alternative method, such as a combined model should be considered.

Most often, the following weighting schemes have been distinguished: equal weights, Akaike weights, optimized and constrained weights, and Bayesian weights. However, these authors often do not focus on deriving a weighted model but rather weighted forecasting. Thus, this study does two different things relating to combining models: (1) evaluate two most commonly used conventional weighting schemes and propose a new weighted procedure based on ranking competing models by their respective forecast accuracy measure performance, and (2) derive the weighted model based on the new proposed weights.

## METHODOLOGY

### Weights for combining forecast

In this section, we briefly discuss the Akaike weights, equal weights and the proposed weights, called weighted ranking (WR).

#### Akaike weights

The Akaike weights was proposed by Akaike (1974). The procedures of the Akaike weights are as follows:

##### a. Ranking alternative models

The AICc values, of the entire set of models is rescaled, such that the model with the minimum AICc has a value of 0. Thus, information criterion values can be rescaled as simple differences,

$$\Delta_i = AICc_i - AICc_{min} \quad (1)$$

where  $AICc_i$  is the individual AICc of each alternative model and  $AICc_{min}$  is the best model or model with the minimum AICc. The  $\Delta_i$  allows a ranking of the models from best to worst; the larger the  $\Delta_i$ , the less plausible is model  $i$ .

##### b. Defining the Akaike weights

The simple transformation  $\exp\left(-\frac{1}{2}\Delta_i\right)$  results in the (*discrete*) likelihood of model  $i$ , given the data  $L(g_i|x)$ . These are functions in the same sense that  $L(\theta|x, g_i)$  is the likelihood of the parameters  $\theta$ , given the data  $(x)$  and the model  $(g_i)$ . These likelihoods are very useful; for example, the evidence ratio for model  $i$  versus model  $j$  is

$$\frac{L(g_i | x)}{L(g_j | x)} \quad (2)$$

It is convenient to normalize these likelihoods such that they sum to 1, as

$$w_i = \frac{\exp\left(-\frac{1}{2}\Delta_i\right)}{\sum_{i=1}^R \exp\left(-\frac{1}{2}\Delta_r\right)} \quad (3)$$

where  $w_i$  is the akaike weight,  $R$  is the number of the entire set of models (Burnham and Anderson, 2002 ). Our interest is with the Akaike weights.

This weight makes use of the Akaike information criterion. However, as pointed out earlier, these information criteria performances are unstable at times. In other words, they wrongly select models at times, and models selected by them as ‘best’ models do not mean they forecast well. Thus, the estimation of weights based on any of these criteria will likely exhibit instability at times. In other words, since these information criteria do not always select the ‘best’ model, any estimation from them makes the results unstable or uncertain.

### Equal weights

Equal weights for combining forecast is a straightforward approach compared to other weighting approaches. Here, forecast values from the entire set of models are averaged by using any kind of mean (*e.g., trimmed, arithmetic, etc*).

One basic problem with this weighting approach is that, the forecast performance of different models are not the same; some models outperform others. Thus, it will be conceptually wrong to consider the forecast performance of all models as same.

### Proposed weights

To overcome the above problems, we propose a weight whose estimation is not based on information criteria and the assumption of equality in forecast performance is desirable. We, therefore, propose a weight based on ranking the forecast or predictive performance of all competing models.

The basis of the weighted ranking procedure is that, each competing model has the potential of relatively predicting the future value of a series, since the true model is unknown. Thus, we allow each model in the competing set of models to forecast. We therefore rank each model based on their predictive performance, by ranking the model with the lowest forecast accuracy giving the highest rank to that model while the best model gets the lowest rank. The weighted ranking procedure is given below:

1. Fit a set of competing models to a dataset. Here, it is appropriate that the selected models of the set have close distance, meaning difference in the information criterion of the respective models should be small. Thus, the guideline for including a model to the set of competing models is that, the information criterion difference of a model and the best model should be less than 4 (*i.e.,  $\Delta < 4$* ).

2. Forecast each model in the entire set of models based on the ‘out-of-sample’ or ‘in-of-sample’ data.
3. Calculate their respective forecast accuracy measure, e.g., MSFE, MAPE etc.
4. Rank models in the entire set by their forecast accuracy measure. Thus, the lowest forecast accuracy measure model receives the highest rank.
5. Sum the ranks and respectively divide the individual rank by the total of the ranks to get the corresponding model weights.

Thus, we can express the proposed weight as:

$$w_i = \frac{\psi_i}{\sum_{i=1}^s \psi_i}, \quad i = 1, 2, \dots, s \tag{4}$$

where  $\psi_i$  is the rank for model  $i$  forecast accuracy measure, MSFE; and  $\sum_{i=1}^s \psi_i$  is the sum of ranks of forecast accuracy measure, MSFE, in the entire set of models ( $s =$  last model in the entire set).

**Formula for combining forecast values**

For period  $h + 1$ , the forecast values are:

$$\begin{aligned} \text{Model 1 : } & F_{h+1} \rightarrow F_{h+1,1} \times w_1 \\ \text{Model 2 : } & F_{h+1} \rightarrow F_{h+1,2} \times w_2 \\ \text{Model } n : & F_{h+1} \rightarrow F_{h+1,n} \times w_n \end{aligned}$$

In general, forecast value for each model is:

$$F_{h+1,i} \times w_i, \quad i = 1, 2, 3, \dots, n$$

where  $i$  is the respective model.

Thus, overall forecast value for a particular period is:

$$f_{h+1} = \sum_{i=1}^R w_i f_{h+1,i} \tag{5}$$

where  $R$  is the number of models in the entire set.

**Combining ARIMA models**

In the available literature, the interest of researchers, usually, is to combine forecast values and not combining models. Thus, in this study we present an appropriate way of combining models.

**Univariate ARMA (p,q) model**

A time series  $\{x_t, t = 0, \pm 1, \pm 2, \dots\}$  is ARMA ( $p, q$ ) and is stationary

$$x_t = \phi_1 x_{t-1} + \dots + \phi_p x_{t-p} + w_t + \theta_1 w_{t-1} + \dots + \theta_q w_{t-q} \tag{6}$$

with  $\phi_p \neq 0, \theta_q \neq 0$  and  $\sigma_w^2 > 0$ . The parameters  $p$  and  $q$  are called the autoregressive and the moving average orders, respectively.

### Weighted ARIMA model

Thus, ARIMA  $(p,d,q)$  model in equation (6) can be combined with appropriate weights as:

$$x_t = \phi_1^* x_{t-1} + \dots + \phi_i^* x_{t-i} + w_t + \theta_1^* w_{t-1} + \dots + \theta_j^* w_{t-j} \quad (7)$$

where  $\phi_i^*$ ,  $i = 1, \dots, p$  and  $\theta_j^*$ ,  $j = 1, \dots, q$  are the “weighted” parameter estimates of autoregressive and moving average processes, respectively.

### Definition of weighted parameter estimates

a). For the Autoregressive model

$$\phi_i^* = \frac{\sum_{i=1}^R w_i I_i(g_i) \hat{\phi}_{i,h}}{\sum_{i=1}^R w_i I_i(g_i)}, \quad (8)$$

where

$$I_i(g_i) = \begin{cases} 1, & \text{if } \phi_i \text{ is in the model } g_i, \\ 0, & \text{otherwise} \end{cases}$$

Here,  $\hat{\phi}_{i,h}$  denotes the estimator of  $\hat{\phi}_i$  based on model  $g_h$ .

b). For the Moving Average Model

$$\theta_j^* = \frac{\sum_{i=1}^R w_i I_i(g_i) \hat{\theta}_{j,h}}{\sum_{i=1}^R w_i I_i(g_i)}, \quad (9)$$

Here,  $\hat{\theta}_{j,h}$  denotes the estimator of  $\hat{\theta}_j$  based on model  $g_h$ .

Similarly, we can define drift or intercept as in (8) and (9).

### Simulation study for combining models

Data were simulated from a true model using R software. Several models were then fitted to the data. However, since the true model is known, we ignore it from the competing models. Here, we consider all models as approximation of the true model. The justification is that, in the real world data, the true model of a data set is not known. Thus, without the true model, the model with the minimum information criterion is considered as the best model.

Datasets are generated for samples of  $N = 35$  (*considered as small sample*) and  $N = 210$  (*considered as large sample*). A ‘seed’ was set so that the same data set were produced for the large sample size. However, for the purpose of cross-validation, 10 data-points were removed from each sample sets. Thus, the sample size for model fitting were  $n = 25$  and  $n = 200$ . The rationale behind these two different sample sizes is that, we want to know how the proposed method performs with an increase in sample size.

## RESULTS AND DISCUSSION

## Akaike weights

Based on the guideline indicated earlier, five models were identified and fitted to both data set of sample sizes  $n=25$  and  $n=200$ . The results of all necessary computations are given in Table 1.

**Table 1. Derived Akaike weight for  $n = 25$**

Model	AICc	$\Delta$	$L(g_i x)$	$w_i$
1. <b>(1,1,0)*</b>	154.22	0	1	0.27
2. <b>(2,1,0)</b>	154.42	0.20	0.9048	0.24
3. <b>(1,1,2)</b>	154.96	0.74	0.6907	0.18
4. <b>(2,1,1)</b>	155.12	0.90	0.6376	0.17
5. <b>(0,1,2)</b>	155.55	1.33	0.5143	0.14
R = 5			<b>3.7474</b>	<b>1</b>

\*is the 'best' model

In Table 1, the weight of the 'best' model (which is model 1, (1, 1, 0)) is quite similar to model 2, and even its difference with models 3, 4, and 5 are considerably smaller. In other words, all the models in the entire set look good as approximation of the true model. Thus, in this case, it is not appropriate to consider only the 'best' model for inference by neglecting equally good models. This is the basis of combining forecast or models.

**Table 2. Derived Akaike weight when  $N=200$**

Model	AICc	$\Delta$	$L(g_i x)$	$w_i$
1. <b>(2,1,1)*</b>	582.6	0	1	0.32
2. <b>(1,1,1)</b>	583.82	1.22	0.5434	0.17
3. <b>(3,1,1)</b>	583.48	0.88	0.6440	0.21
4. <b>(2,1,2)</b>	583.88	1.28	0.5273	0.16
5. <b>(3,1,0)</b>	584.31	1.71	0.4253	0.14
6. <b>(1,1,0)#</b>	<b>587.04</b>	<b>4.44</b>	----	----
R = 5			<b>3.14</b>	<b>1</b>

\*is the 'best' model

# Model 6, (1,1,0) was the 'best' model when sample size was,  $n = 25$ . However, it could not meet the selection requirement into the set of competing models when sample size,  $n = 200$ . Thus, model 6, (1,1,0), cannot be included into the set of competing models when sample size,  $n = 200$ .

In Table 2, the weight of the 'best' model is relatively higher as compared to models 2, 4, and 5; but the different with model 3 is not that much.

Since the difference between the 'best' model weight and other competing or alternative models is very small; instead of using only the 'best' model for inference, it will be appropriate to add the other four alternative models to construct a single weighted model (*composite model*) for proper inference. This will lead to increase in precision.

### Equal weight

There are five (5) models in the entire set, thus the equal weight will be 0.2 for each model.

### Proposed weight (*weighted ranking procedure*)

Based on the selection criterion indicated earlier, five models were identified and fitted to both dataset of sample size  $n = 25$  and  $n = 200$ . However, it should be noted that, several models were fitted to the data but for lack of space, we report models which meet the selection criterion. We got the forecast values for each model for 10 horizons. The out-of-sample forecast accuracy measure, mean square forecast error (MSFE), was computed on each model. These MSFE of the individual models were ranked from lowest to highest. The results are given in Tables 3 and 4.

**Table 3. Deriving the weighted ranking procedure when  $n = 25$**

Model	MSFE	Rank	Weights
1. (1,1,0)	68.92	3	0.20
2. (2,1,0)	71.86	2	0.13
3. (1,1,2)*	60.09	5	0.33
4. (2,1,1)	64.33	4	0.27
5. (0,1,2)	72.51	1	0.07
<b>Total</b>		15	1

\*'best' model in terms of forecast performance

In Table 3, the distributions of weights are very different from the Akaike weights. Here, the 'best' model (i.e., (1, 1, 0)) selected by the Akaike information criterion was the third 'best' model in terms of forecasting performance. This confirms the fact that 'best' model does not always give better forecast.

**Table 4. Deriving the weighted ranking procedure when  $n = 200$**

Model	MSFE	Rank	Weights
1. (2,1,1)	83.93	3	0.20
2. (1,1,1)	84.23	1	0.07
3. (3,1,1)*	83.66	4	0.27
4. (2,1,2)	83.98	2	0.13
5. (3,1,0)	82.23	5	0.33
<b>Total</b>		15	1

\*'best' model in terms of forecast performance

In Table 4, the distributions of weights are also very different from the Akaike weights. Here, the 'best' model (i.e., (2,1,1)) selected by the Akaike information criterion was ranked as the third 'best' model in terms of forecasting performance by the ranking weight. Thus, we expect that our weighted ranking approach to out-perform the Akaike weight.

### Overall performance of approaches

The purpose of this section is to compare the performance of combining forecasts using Akaike weights, equal weights, and our proposed weights and with individual forecast obtained from the ‘best’ model selected by the AICc; with respect to different horizon. Here, we define horizon as the distant from one observation in time to the other. Mean square forecast errors (MSFE) for sample  $n = 25$  are given in Table 5.

**Table 5. Mean Square Forecast Error (MSFE) for sample  $n = 25$**

Horizon	$W_{\text{rank}}$	$W_{\text{akaike}}$	$W_{\text{equal}}$	$\text{Min}_{\text{AICc}}$
1	2.784	2.974	2.833	3.848
2	9.990	10.448	10.162	12.344
3	14.299	15.113	14.867	17.099
4	14.596	15.622	15.437	17.248
5	27.234	28.840	28.646	30.505
6	45.808	48.085	47.881	49.619
7	73.782	76.848	76.629	78.144
8	125.917	130.081	129.831	131.095
9	158.889	163.696	163.439	164.237
10	179.755	184.965	184.707	185.044
<b>Total (MSFE)</b>	<b>65.306</b>	<b>67.667</b>	<b>67.443</b>	<b>68.918</b>

#  $\text{Min}_{\text{AICc}}$  is the MSFE from ‘best’ model selected by minimum AICc,  $W_{\text{equal}}$  is the MSFE from equal weights,  $W_{\text{akaike}}$  is MSFE from Akaike weights and  $W_{\text{rank}}$  is the MSFE from weighted ranking.

Mean square forecast errors (MSFE) for sample,  $n=200$  are given in Table 6.

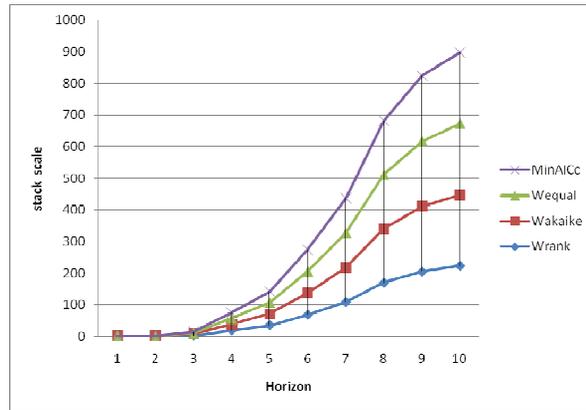
**Table 6. Mean Square Forecast Error (MSFE) for sample  $n = 200$**

Horizon	$W_{\text{rank}}$	$W_{\text{akaike}}$	$W_{\text{equal}}$	$\text{Min}_{\text{AICc}}$
1	0.299	0.294	0.301	0.271
2	0.283	0.271	0.272	0.254
3	3.459	3.443	3.443	3.426
4	18.612	18.665	18.652	18.650
5	35.227	35.399	35.357	35.467
6	68.184	68.511	68.436	68.679
7	108.721	109.233	109.114	109.548
8	169.742	170.486	170.309	170.989
9	205.154	206.062	205.841	206.717
10	223.542	224.566	224.314	225.332
<b>Total (MSFE)</b>	<b>83.321</b>	<b>83.693</b>	<b>83.604</b>	<b>83.933</b>

In Table 5 and 6, the weighted ranking procedure measured the lowest MSFE, indicating that our proposed weighted ranking procedure outperforms the other approaches of weights combined forecasts (i.e., equal weight and Akaike weight) and the individual forecast

obtained from the ‘best’ model selected by minimum AICc. The second performed approach is the equal weighting method, followed by Akaike weighting method and lastly the ‘best’ model selected by minimum AICc; and this is depicted in Figure 1. Such performance indicates the dominance of combined forecasts.

Generally, in all the approaches, forecast horizon increases with MSFE, and this is evident in Figure 1. This confirms that time series forecast does not perform well with high horizon, thus, it’s always appropriate to forecast smaller or shorter horizon. It should be noted that the smaller the MSFE value the better the model.



**Fig 1. Performance of weighted approaches and best model as horizon increases**

MinAICc is the MSFE from ‘best’ model( Wrank) selected by minimum AICc, Wequal is the MSFE from equal weights, Wakaike is the MSFE from Akaike weights and Wrank is the MSFE from weighted ranking.

**Combining models using Weighted Ranking procedure**

Available literature often neglects the aspect of combining models. Thus, in this section, we show how to combine models using the weighted ranking procedure. It should be noted that, the approach can be applied when using any other weighting technique. We will illustrate this technique by using models for which when sample size,  $n = 200$ .

**Weighted parameter estimates**

The computation of the weighted parameter estimates of the models are presented in Table 7. It should be noted that, when  $n = 25$ , the ‘best’ model was (1,1,0); however it could not make it into the set of competing models when  $n = 200$ , according to the selection criterion.

**Table 7. Weighted parameter estimates, when N = 200**

Model	Original parameter (2)	Weights (3)	New paramters [ 2*3]
(2,1,1)	$\hat{\phi}_1 = 0.1913$	0.20	$\phi_{1,1}^* = 0.0383$
	$\hat{\phi}_2 = 0.2544$		$\phi_{2,1}^* = 0.0509$
	$\hat{\theta}_1 = 0.7154$		$\theta_{1,1}^* = 0.1431$
(1,1,1)	$\hat{\phi}_1 = 0.4729$	0.07	$\phi_{1,2}^* = 0.0331$
	$\hat{\theta}_1 = 0.4586$		$\theta_{1,2}^* = 0.0321$
(3,1,1)	$\hat{\phi}_1 = 0.3847$	0.27	$\phi_{1,3}^* = 0.1039$
	$\hat{\phi}_2 = 0.0470$		$\phi_{2,3}^* = 0.0127$
	$\hat{\phi}_3 = 0.1218$		$\theta_{3,3}^* = 0.0329$
	$\hat{\theta}_1 = 0.5083$		$\theta_{1,3}^* = 0.0137$
(2,1,2)	$\hat{\phi}_1 = 0.5391$	0.13	$\phi_{1,4}^* = 0.0701$
	$\hat{\phi}_2 = 0.1098$		$\phi_{2,4}^* = 0.0143$
	$\hat{\theta}_1 = 0.3571$		$\theta_{1,4}^* = 0.0464$
	$\hat{\theta}_2 = -0.1891$		$\theta_{2,4}^* = -0.0246$
(3,1,0)	$\hat{\phi}_1 = 0.8624$	0.33	$\phi_{1,5}^* = 0.2846$
	$\hat{\phi}_2 = -0.3465$		$\phi_{2,5}^* = -0.1144$
	$\hat{\phi}_3 = 0.2047$		$\theta_{3,5}^* = 0.0676$

Applying equation (2.8) and (2.9), we achieve the following weighted parameter estimates for the autoregressive and moving average models, respectively.

$$\phi_{1,r}^* = 0.52994(0.1774)$$

$$\phi_{2,r}^* = -0.0393(0.1590)$$

$$\phi_{3,r}^* = 0.1674(0.08405)$$

$$\theta_{1,r}^* = -0.5355(0.2219)$$

$$\theta_{2,r}^* = -0.02458(0.0213)$$

Note that the values in the parentheses are the respective standard errors. Thus, the combined or weighted model is ARIMA (3,1,2) but with different parameter estimates. It's defined as:

$$(1 - 0.52994B + 0.0393B^2 - 0.1674B^3) \nabla Z_t = (1 - 0.5355B + 0.02458B^2) a_t$$

## CONCLUSION

The problems or challenges associated with the Akaike weights, equal weights for combining forecasts and forecast from a single 'best' model selected by the minimum AICc value are pointed out in this study. Thus, an approach that can minimize or handle these challenges will be useful. We therefore, proposed a weight based on ranking procedure to combine forecast. The results revealed that the MSFE from the combined forecast of the proposed weight (*weighted ranking procedure*) outperformed all other approaches that were investigated in this study. Furthermore, the three combined forecast approaches consistently outperformed the forecast from the 'best' model selected by minimum AICc. We therefore, recommend the use of weighted ranking procedure for combining models.

## REFERENCES

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6), 716 - 723.
- Bates, J., and Granger, C. (1969). "The Combination of Forecasts," *Operational Research Quarterly*, 20, 451 - 468.
- Box, G. and Jenkins, G. (1994). *Time Series Analysis, Forecasting and Control*. John Wiley & Sons: New Jersey, USA.
- Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24, 123 - 140.
- Burnham, K.P., Anderson, D.R., (2002). *Model selection and multimodel inference: a practical information-theoretic approach*, 2<sup>nd</sup> Ed, Springer-Verlag New York, Inc.
- Chatfield, C. (2004). *The Analysis of Time Series: An Introduction*. Chapman & Hall/CRC: London, UK.
- Clemen, R.T. (1989). "Combining Forecasts: A Review and an Annotated Bibliography," *International Journal of Forecasting*, 5, 559 - 583.
- Hoeting, J.A., Madigan D., Raftery A.E. and Volinsky, C.T. (1999). Bayesian Model Averaging: A Tutorial. *Statistical Science*, 14 (4), (Nov., 1999), pp. 382-401
- Mees, A. (2000). *Nonlinear Dynamics and Statistics*, ed. Mees, A. I. (Birkhauser, Boston).
- Nakamura, T. & Small, M. (2006). "Modelling nonlinear time series using improved least squares method," *Int. J. Bifurcation and Chaos*, 16, 445 - 464.
- Stock, J.H., and Watson, M.W. (2002). Macroeconomic Forecasting Using Diffusion Indexes, *Journal of Business and Economic Statistics*, 20(2), 147 - 162.
- Zou H. and Yang Y. (2004). Combining Time Series Models for Forecasting, *International Journal of Forecasting*, 20.