VOL. 6, NO. 5, MAY 2011 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences


www.arpnjournals.com

A BOOTSTRAP TEST FOR EQUALITY OF MEAN ABSOLUTE ERRORS

Naveen Kumar Boiroju, Ramu Yerukala, M. Venugopala Rao and M. Krishna Reddy
Department of Statistics, Osmania University, Hyderabad, India
E-Mail: nanibyrozu@gmail.com

ABSTRACT

In this paper, we develop a bootstrap test procedure for testing of equality of mean absolute errors of two alternative time series models. Applicability of the bootstrap test is explained using two numerical examples and the results compared with the Sign test and Die bold-Mariano Test.

Keywords: time series model, prediction performance evaluation, bootstrap test, mean absolute error, mean absolute percent error.

1. INTRODUCTION

Prediction has drawn a considerable amount of attention for decades, particularly in the field of economics and finance. Forecasts of variables are useful, not only to know the future path of the economy but also for choosing the most proper specification among empirical models. The evaluation of forecasting accuracy via measures of point estimates of out-of-sample predictive errors is well-established practice in the time series analysis. The mean absolute error (MAE), root mean square error (RMSE) and mean absolute percent error (MAPE) are often used for evaluating forecasting performance of a time series model. The usual practice is to choose the model which has a lower accuracy measure among alternative forecasting models. Test for equal predictive ability, in general setting, were proposed by Die bold and Mariano (DM test) [2] and West [7], where the framework of the latter can accommodate the situation where forecasts involve estimated parameters. Harvey, Leybourne and New bold [5] suggested a modification of the Die bold-Mariano test that leads to better small sample properties for testing the equality of mean squared error. In this paper, we discuss on testing of equality of mean absolute errors of two different forecasting models for the same time series. A bootstrap test procedure is developed for testing of equality of mean absolute errors of two competing forecasting models and the results compared with Sign test and Die bold-Mariano (1995) test.

Mean absolute error is an error statistic that averages the distances between each pair of actual (Z_t) and fitted forecast $(\hat{Z}_t)$data points. MAE is calculated by taking the average of the absolute errors and is most appropriate when the cost of forecast errors is proportional to the absolute size of the forecast errors. MAE is given by $MAE = \frac{1}{N}\sum_{t=1}^{N}\left|Z_t - \hat{Z}_t\right| = \frac{1}{N}\sum_{t=1}^{N}\left|e_t\right|$.

Suppose, $(e_{1,t}, e_{2,t})$, t= 1, 2, …m are h-step out-of-sample forecast errors of models 1 and 2, respectively. Taking MAE as a measure of prediction loss, the loss differential from the two models can be expressed as $d_t = \left|e_{1t}\right| - \left|e_{2t}\right|$, t = 1, 2, … m. Sign test and Diebold-Mariano (DM) test is used to test the null hypothesis $E(d_t) = 0 = \theta_0$. Sign test and DM test procedures explained in Section 2. Section 3, introduces the concept of bootstrap, and proposes a bootstrap test of equal prediction accuracy. Section 4, contains the data and empirical results.

2. REVIEW OF FORECAST TESTS

One possibility to test the null hypothesis that there is no qualitative difference between the forecasts from the two models is to use the sign test statistic $S' = \sum_{j=1}^{m} I(d_j > 0)$ which has the binomial distribution with parameters m and $\frac{1}{2}$. Under the null hypothesis the indicator function I (A) equals 1 if the event A occurs and zero otherwise. Significance may be assessed using a table of the cumulative binomial distribution. For large values of m, (m>10) the standardized version of the sign test statistic is asymptotically standard normal

$$S = \frac{S' - \frac{m}{2}}{\sqrt{\frac{m}{4}}} = \frac{2}{\sqrt{m}}\sum_{j=1}^{m}\left[\left(I(d_j > 0) - \frac{m}{2}\right]\right. \overset{a}{\sim} N(0,1) \quad (2.1)$$

If S is significantly large, then one can reject the null hypothesis of forecast equivalence.

Since sign test compares only the relative magnitude of the prediction errors, Diebold and Mariano developed a statistic, which compares the absolute magnitude by testing whether the average loss differential $\bar{d} = \frac{1}{m}\sum_{t=1}^{m} d_t$ is significantly different from zero. DM test statistic is given by

$$DM = \frac{\bar{d}}{\sqrt{Var(\bar{d})}} \sim N(0,1) \quad (2.2)$$

where $\bar{d}$ refers to the sample mean of d_t and $Var(\bar{d})$ is a consistent estimator for the long-run variance of d_t. Assuming h-step ahead forecast exhibit dependence or the forecast errors are serially correlated up to the order h-1. DM test uses a non-parametric method of estimation using the uniform kernel equipped with bandwidth (h-1). Its estimator assumes the form $Var(\bar{d}) = \frac{1}{m}\sum_{k=-(h-1)}^{(h-1)} \hat{\gamma}_k$, where $\hat{\gamma}_k$ denotes the sample auto covariance of d_t with lag k. Finally, under the null hypothesis, this test statistic follows a limiting standard normal distribution [2, 5].

3. BOOTSTRAP TEST FOR EQUALITY OF MEAN ABSOLUTE ERRORS

The bootstrap, introduced by Efron (1979), is a computer - intensive method for estimating the distribution of an estimator or test statistic by resampling the data at hand. It treats the data as if they were the population. In fact, under mild regularity conditions, the bootstrap generally yields and approximation to the sampling distribution of an estimator or test statistics that is at least as accurate as the approximation obtained from traditional first - order asymptotic theory. In many instances the sampling distribution of a statistic may not be analytically available, while the bootstrap, on the other hand, obtains the resampling from the sample at hand [3, 4].

In bootstrap framework, there are two approaches for testing a hypothesis, on based on confidence intervals, and the other direct hypothesis testing. Direct bootstrap hypothesis test requires drawing a sample of the statistic of interest from an empirical distribution under the restrictions specified the null hypothesis, and then the achieved significance level, or the p-value, can be calculated by comparing the observed statistic based on the original data and the sample of the statistic based on bootstrap samples. Usually a hypothesis can be tested by constructing an appropriate confidence set. However, direct bootstrap hypothesis test is sometimes easier when constructing a confidence set is complicated [1, 6].

Let d_t; $t = 1, 2, \cdots, m$ denote the difference of absolute forecast errors of two models, where m is the number of forecasts generated by each model. The bootstrap test procedure explained in the following steps:

Step 1: Let the available sample of loss differential is $(d_1, d_2, \cdots, d_m)$.

Step 2: Draw B (5000) bootstrap samples of size m from the available sample and b[th] - bootstrap sample is given by $d_b^* = (d_{1,b}^*, d_{2,b}^*, \cdots, d_{m,b}^*),\ b = 1, 2, \cdots, B$

Step 3: Compute the mean of b[th] - bootstrap sample and is given by $\bar{d}_b^* = \frac{1}{m}\sum_{t=1}^{m} d_{t,b}^*$; $b = 1, 2, \cdots, B$.

Step 4: Form the sampling distribution and compute the order statistics $\bar{d}_{(L)}^*$ *and* $\bar{d}_{(U)}^*$ where $L = \max\{0, [B\alpha]\}$ $U = \min\{[B(1-\alpha)], B\}$, $0 < \alpha < 1$ and $[x]$ is integer part of x.

Step 5: If the null hypothetical value $\theta_0 = 0 \in (d_{(L)}^*, d_{(U)}^*)$ then accept the null hypothesis that the two models has equal mean absolute errors. Otherwise, reject the null hypothesis.

4. EMPIRICAL STUDY

In this Section, we apply the above three tests to test the equality of mean absolute errors of two competing forecasting models at $\alpha = 0.05$. Two numerical examples are given to assess the bootstrap test by comparing it with sign test and DM test.

Example 1:

The following (Table-1) are out-of-sample forecasting errors (e_{1t}, e_{2t}) (t=1, 2, . . ., 20) of two competing models 1 and 2 respectively.

Table-1. Forecasting errors of two time series models.

e_{1t}	e_{2t}	d_t	e_{1t}	e_{2t}	d_t
-2.23	0.01	2.22	-1.06	-1.42	-0.36
-2.05	1.87	0.18	-0.71	0.26	0.45
1.16	-0.39	0.77	0.69	3.68	-2.99
-1.01	1.12	-0.11	-2.74	-1.07	1.67
0.90	1.14	-0.24	1.11	0.40	0.71
3.40	-1.63	1.77	-0.64	-2.53	-1.89
1.83	0.21	1.62	-2.25	-1.32	0.93
-1.65	1.07	0.58	-1.46	1.94	-0.48
-0.92	2.18	-1.26	-2.07	-2.07	0.00
-1.04	0.33	0.71	2.41	-1.64	0.77

For the given data, we have $\bar{d} = 0.25$, MAE_1=1.57, MAE_2=1.31. To test the equality of forecasting performance of the two models, we use sign test, DM test and proposed bootstrap test. For the given data, Sign test statistic (S) is 1.34 and its critical value is 1.96 at 5% level of significance. It is observed that the DM test statistic is 0.89 and its critical value at 5% level of significance is 1.96. Since both test statistic values are less than the critical value at 5% level of significance, we do not reject the null hypothesis of equal prediction accuracy and we may conclude that there is no significant difference between mean absolute errors of the two forecasting models.

For the same data, we have applied bootstrap procedure as explained in Section 3. It is observed that $d_{(L)}^* = -0.32$ and $d_{(U)}^* = 0.77$ at $\alpha = 0.05$.

Since $\theta_0 = 0 \in (-0.32\ \ 0.77)$, therefore we accept the

null hypothesis and we may conclude that there is no significant difference between the mean absolute errors of the two forecasting models.

Example 2:

The following (Table-2) are (e_{1t}, e_{2t}) (t=1, 2, …, 30) out-of-sample forecasting errors of two models.

Table-2. Forecasting errors of two time series models.

e_{1t}	e_{2t}	d_t	e_{1t}	e_{2t}	d_t	e_{1t}	e_{2t}	d_t
0.98	0.40	0.58	-0.78	-0.75	0.03	2.12	1.65	0.47
-2.95	-2.05	0.90	-0.89	-0.48	0.41	-0.16	-0.11	0.05
0.96	0.11	0.85	0.78	0.51	0.27	-1.59	-1.56	0.03
-0.95	-0.81	0.14	-0.87	-0.33	0.54	-0.98	-0.86	0.12
0.86	0.62	0.24	-1.84	-0.45	1.39	0.18	0.29	-0.11
-0.42	-0.32	0.10	0.64	0.77	-0.13	2.40	2.79	-0.39
2.40	2.95	-0.55	-1.08	-0.52	0.56	-2.28	-2.25	0.03
-0.45	-0.20	0.25	2.82	1.86	0.96	-0.12	0.48	-0.36
2.10	2.56	-0.46	1.12	1.32	-0.20	-1.92	-0.46	1.46
-0.45	-1.76	-1.31	1.40	1.54	-0.14	-1.70	-0.04	1.66

For the given data, we have $\bar{d} = 0.25$, MAE_1=1.03, MAE_2=1.27. For the given data, Sign test statistic in absolute value is 2.92 and its critical value is 1.96 at 5% level of significance. It is observed that the DM test statistic is 2.13 and its critical value at 5% level of significance is 1.96. Since both test statistic values are greater than the critical value at 5% level of significance, we reject the null hypothesis of equal prediction accuracy and we may conclude that there is a significant difference between mean absolute errors of the two forecasting models.

For the same data, we have applied bootstrap procedure as explained in section 3. It is observed that $d^*_{(L)} = -0.47$ and $d^*_{(U)} = -0.05$ at $\alpha = 0.05$.

Since $\theta_0 = 0 \notin (-0.47 \ -0.05)$, therefore we reject the null hypothesis and we may conclude that there is a significant difference between the mean absolute errors of the two forecasting models. Since MAE_1 is less than the MAE_2, we select the model 1 for forecasting the future values.

5. CONCLUSION

The results of bootstrap test are very similar to those of the Sign test and DM test. This paper proposes a test of equal prediction accuracy using bootstrap method that does not rely on specific distributional assumptions. Bootstrap methods have many potential applications in time series analysis, especially when the sample size is limited, in which case traditional asymptotic theories may not provide good approximations. Extending the bootstrap testing framework that allows for serially correlated and heteroscedastic errors should be an interesting direction of further research.

REFERENCES

[1] Becher H., Hall P., Wilson S.R. 1993. Bootstrap hypothesis testing procedures. Biometrics. 49(4): 1268-1272.

[2] Diebold F.X. and Mariano R.S. 1995. Comparing predictive accuracy. Journal of Business and Economic Statistics. 13: 253-263.

[3] Efron B. 1979. Bootstrap methods: Another look at the jackknife. Annals of Statistics. 7: 1-26.

[4] Efron B. and Tibshirani R.J. 1993. An introduction to the bootstrap. Chapman and Hall, New York.

[5] Harvey D.I., Leybourne S.J., and Newbold P. 1997. Testing the equality of prediction mean squared errors. International Journal of Forecasting. 13: 281-291.

[6] Tibshirani R. 1992. Bootstrap hypothesis testing (Letter to the Editor). Biometrics. 48: 969-970.

[7] West K.D. 1996. Asymptotic inference about predictive ability. Econometrica. 64: 1067-1084.