Australian Journal of Basic and Applied Sciences, 8(6) April 2014, Pages: 385-389

AENSI Journals

Australian Journal of Basic and Applied Sciences

ISSN:1991-8178

Journal home page: www.ajbasweb.com

On Boostrap Prediction Intervals for Autoregressive Model

[1]T.O. Olatayo, [2]A.A. Akomolafe and [3]N.O.Adeboye

[1]*Department of Mathematical Sciences, Faculty of Science Olabisi Onabanjo University, P.M.B.2002 Ago-Iwoye, Nigeria.*
[2]*Department of Mathematics and Statistics, College of Natural Sciences Joseph Ayo Babalola University, P.M.B.5006, Ikeji-Arakeji, Osun State, Nigeria.*
[3]*Department of Mathematics and Statistics, Faculty of Science Federal Polytechnic,P.M.B. 50 Ilaro, Ogun State, Nigeria.*

ARTICLE INFO

Article history:
Received 25 January 2014
Received in revised form
8 April 2014
Accepted 20 April 2014
Available online 10 May 2014

Keywords:
Bootstrap method, Standard error, bias, bootstrap confidence t interval, prediction error rate, and autoregressive model.

ABSTRACT

Background:Frequently, an estimated mean squared error is the only indicator or yardstick of measuring error in a prediction. However, the statement that the future values falls in an interval with a specified probability is more informative. Prediction intervals have this probabilistic interpretation, which is similar to that of tolerance intervals . Two resampling methods yield prediction intervals that obtain some types of asymptotic invariance to the sampling distribution. The resampling procedure proposed here utilizes the bootstrap method. The bootstrap interval derives from an empirical distribution generated using bootstrap resampling. The bootstrap is a resampling technique whose aim is to gain information on the distribution of an estimator. Objective: The bootstrap method for measures of Statistical accuracy such as standard error, bias, prediction error and to complicated data structures such as autoregressive models are considered. We estimated the parameters and the bootstrap t confidence interval with an autoregressive model fitted to the real data. Results:Bootstrap prediction intervals provide a non parametric measure of the probable error of forecast from a standard linear autoregressive model. Empirical measure prediction error rate motivate the choice of these intervals, which are calculated by an application of the bootstrap methods, to a time series data. Conclution: Bootstrap prediction intervals represent a useful addition to the traditional set of measures to assess the accuracy of forecast. The asymptotic properties of the intervals do not depend upon the sampling distribution, and the bootstrap results suggest that the invariance approximately holds for relatively all sample sizes.

To Cite This Article: T.O. Olatayo, A.A. Akomolafe and N.O.Adeboye., On Boostrap Prediction Intervals for Autoregressive Model.. ***Aust. J. Basic & Appl. Sci.,*** *8(6): 385-389, 2014*

INTRODUCTION

One of the most important objectives in the analysis of a time series economic data is to forecast it future values. Even if the final purpose of time series modeling is the control of a system, it operation is usually based on forecasting, Wei (1989).

The term forecasting is used more frequently in recent time series literature than the term prediction. However, most forecasting results are derived from a general theory of linear prediction developed by Kolmogorov (1941), Wiener (1949), Kalman (1960), Yaglom (1962), and Whittle (1983).

Frequently, an estimated mean squared error is the only indicator or yardstick of measuring error in a prediction. However, the statement that the future values falls in an interval with a specified probability is more informative. Prediction intervals have this probabilistic interpretation, which is similar to that of tolerance intervals , Stine (1982), (1985). Two resampling methods yield prediction intervals that obtain some types of asymptotic invariance to the sampling distribution. The resampling procedure proposed here utilizes the bootstrap method, Efron (1979, 1982). The bootstrap interval derives from an empirical distribution generated using bootstrap resampling. The bootstrap is a resampling technique whose aim is to gain information on the distribution of an estimator.

This article investigates and describes the basis of bootstrap theory as a measure of statistical accuracy such as bias and prediction error with confidence intervals.

Suppose that our data consists of a random sample from an unknown probability distribution F on the real line,

$$X_1, X_2, X_3, \text{------------------}, X\, n \sim F \qquad (1.1)$$

Corresponding Author: T.O. Olatayo, Department of Mathematical Sciences,Faculty of Science Olabisi Onabanjo University, P.M.B.2002 Ago-Iwoye, Nigeria.
E-mail: otimtoy@yahoo.com

We compute the sample mean $\bar{x} = \frac{\sum_{i}^{n} x_i}{n}$ and wonder how accurate it is as an estimate of the true mean $\theta = E_F(x)$. if the second central moment of F is $\mu_2(F) = E_F X^2 - (E_F X)^2$, then the standard error $\sigma(F;n,\bar{x})$,that is the standard deviation of $\bar{x}$ for a sample of size n from distribution F, is

$\sigma(F) = [\mu_2(F)/n]^{1/2}$ (1.2)

The shortened notation $\sigma(F) = \sigma(F;n,\bar{x})$ is allowable because the sample size n and statistics of interest $\bar{x}$, are known, only F being unknown. The standard error is the traditional measure of $\bar{x}$ accuracy. Unfortunately, we cannot actually use (1.2) to assess the accuracy of $\bar{x}$, since we do not know $\mu_2(F)$, but we can use the estimated standard error,

$\bar{\sigma} = [\bar{\mu}_2/n]^{1/2}$ (1.3)

Where $\sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$, the unbiased estimate of $\mu_2(F)$. There is a more obvious way to estimated $\sigma(F)$. Let F^ indicate the empirical probability distribution; F^: probability mass i/n on

$X_1, X_2\ Xn$ (1.4)

Then we can simply replace F by F^ in (1.2), obtaining

$\sigma\hat{} = \sigma(F\hat{}) = [\mu_2\ (F)/n]^{1/2}$ (1.5)

as the estimated standard error for $\bar{x}$. This is the bootstrap estimate.

Since

$\mu_2 = \mu_2\ (F\hat{}) = \Sigma\ (x_i - x)^2\ /n$ (1.6)

$\sigma\hat{}$ is not quite the same as $\bar{\sigma}$, but the difference is too small to be important in most application, Efon and Tibshirani (1986).

Standard errors are crude but useful measure of statistical accuracy. They are frequently used to give approximate confidence intervals for an unknown parameter θ.

$\Theta \in \theta\hat{} \pm \sigma\hat{} Z^{(\alpha)}$ (1.7)

Where $Z^{(\alpha)}$ is the 100 α percentile point of a standard normal variate. The standard interval (1.7) is based on taking literally the large sample normal approximation $(\theta\hat{} - \theta)/\sigma\hat{} \sim N(O,1)$.There many ways to construct bootstrap confidence intervals for any given bootstrap. In this paper we employ the bootstrap t confidence interval proposed by Olatayo (2010). Most common statistical methods were developed in the 1920s and 1930s, when computation was slow and expensive. Now that computation is fast we can hope for and expect change in statistical methodology. This paper discusses one such potential change.

The boostrap estimate of standard error:

Let F be the empirical distribution, putting probability 1/n on each of the observed values of x_i, i= 1,2,-----------------, n. A bootstrap sample is defined to be a random sample of size n drawn from F^, the bootstrap algorithm works by drawing many independent bootstrap samples, evaluating the corresponding bootstrap replications, and estimating the standard error of θ^ by the empirical standard deviation of the replications. The result is called bootstrap estimate of standard error, denoted by $Se\hat{b}$, where B is the number of bootstrap sample used.

The limit of $Se\hat{b}$, as B goes to infinity is the ideal bootstrap estimate of

$$\lim_{B\to\infty} se\hat{b} = se_F = Se_F\left(\hat{\theta}*\right)$$ (2.1)

The fact that $Se\hat{b}$ approaches Se_F as B goes to infinity amounts to saying that an empirical standard deviation as the number of replications grown large. The ideal bootstrap estimate $Se_F\,\hat{\theta}*$ and its approximation $Se\hat{b}$ are sometimes called non parametric bootstrap estimate because they are based on $\hat{F}$, the non parametric estimate of the population F. Efron and Tibshirani (1993).

The bootstrap algorithm for estimating standard error proceeds in three steps.

Select B independent bootstrap samples $X,^1,X^2,X^3$, --- X^B, each consist of n data value drawn with replacement from X.

i Evaluate the bootstrap replication corresponding to each boostrap sample

$$\hat{\theta}*(b) = S\left(X^{*b}\right), b = 1,2,3........B$$ (2.2)

Australian Journal of Basic and Applied Sciences, 8(6) April 2014, Pages: 385-389

ii Estimate the standard error $se_F\left(\hat{\theta}\right)$ by the sample standard deviation of the B replications

$$Se\hat{B} = \left[\sum_{b=1}^{B}\left[\hat{\theta}*\right](b) - \hat{\theta}*(.)^2 \frac{}{B-1}\right]^{\frac{1}{2}} \tag{2.3}$$

where $\hat{\theta}*(.) = \dfrac{\sum_{b=1}^{B}\hat{\theta}*(b)}{B}$

It is easy to see that as $B \to \infty$ -, $Se\hat{b}$ is the ideal bootstrap estimate of $se_F\left(\hat{\theta}\right)$, the bootstrap estimated of standard error.

The Boostrap Estimate Of Bias:

The bias of $\hat{\theta}$=S(X)

As an estimate of θ, is defined to be the difference between the expectation of θ and the value of the parameter θ.

$$bias = bias_F\left(\hat{\theta},\theta\right) = E_F\left[S(X)\right] - t(F)$$

Where the real valued parameter θ=t(F) and the statistic $\hat{\theta}$= s(x)respectively. A large bias is usually an undesirable aspect of an estimator's performance. We resigned to the fact that $\hat{\theta}$ is a variable estimator of θ, but usually we don't want the variability to be overwhelmingly on the low side or on the high side.

In this article we use the bootstrap to asses the bias of any estimator θ= S(x). the bootstrap estimate of bias is defined to be the estimate $bias_F$, we obtain by substituting $\hat{F}$ for F in (2.4)

$$Bias_F = E_{\hat{F}}\left[S\left(X^*\right)\right] - E(F) \tag{2.4}$$

We generate independent bootstrap samples $X^{*1}, X^{*2}, \ldots\ldots\ldots\ldots X^{*B}$ as in section 2.0, evaluate the bootstrap replication $\hat{\theta}*(b) = s\left(X^{*b}\right)$ and approximate the bootstrap expectation $E_{\hat{F}}\left[S\left(X^*\right)\right]$ by the average

$$\hat{\theta}*(.) = \sum_{b=1}^{B}\hat{\theta}*(b)\frac{}{B} = \frac{\sum_{b=1}^{B} S\left(X^{*b}\right)}{B} \tag{2.5}$$

The bootstrap estimate of bias based on the B replication $bia\hat{s}$ B IS in (3.2) with $\hat{\vartheta}*(.)$ substituted for $E_{\hat{F}}\left[S\left(X^*\right)\right]$

$$bia\hat{s}\ \mathrm{B} = \hat{\theta}*(.) - t\left(\hat{F}\right) \tag{2.6}$$

The algorithm in section 2.0 applies exactly to the calculation of (2.6), except that at the last step we calculate $\hat{\theta}*(.) - t\left(\hat{F}\right)$ rather than $Se\hat{B}$ of course we calculate both $Se\hat{B}$ and $bia\hat{s}$ B from the same set of bootstrap replications.

Prediction Intervals from The Boostrap:

Suppose F $\longrightarrow$ X =(X1, X2,---------------------Xn) and require a 1-2α prediction interval for a new observation $Z \approx F$, that is, we would like random variables a(X) and b (X) so that

$$\Pr ob_F\left\{a(X) \le Z \le b(X)\right\} = 1 - 2\alpha \tag{2.7}$$

It is important to note that probability in (2.7) refers to the randomness in both $X_i \approx$ f, and $Z \approx F$

We find a value a(x) so that

$$\Pr ob_F\left\{\frac{\overline{x} - z}{s}\right\} \le \hat{t}(\alpha) = \alpha \tag{2.8}$$

$$S = \frac{\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2}{(n-1)}$$

Our prediction interval is then obtained by pivoting expression (2.8) giving

$$\left(\bar{x}\right) - \hat{t}\left(1-\alpha\right)s, \bar{x} - \hat{t}^{(\alpha)}S \tag{2.9}$$

if we assume that F is standard normal with mean N and unknown variance σ^2, we obtained

$$\hat{t}^{(\alpha)} = t^{(\alpha)}{}_{n-1}\sqrt{1+\frac{1}{n}}$$

where $t^{(\alpha)}{}_{n-1}$ is the α percentile of the t- distribution with n-1 degree of freedom. This differs from a confidence interval for M=E (X) in that the factor $\sqrt{1+\frac{1}{n}}$ appears rather than $\sqrt{\frac{1}{n}}$. The extra “1” account for the variance of the new observation Z.

the bootstrap approach resample $\hat{F} \to X^*$ and $\hat{F} \to Z^*$ independently, and then estimates $\hat{t}^{(\alpha)}$ – by the empirical α^{th} quartile of the values

$$\frac{\bar{x}^* - Z^*}{S^*} \tag{2.10}$$

where $\bar{X}^*$ and S^* are the mean and sample deviation of the bootstrap sample X^*. thus as in (2.9) we have

$$[\hat{\theta} - \hat{S}_{e^*}(\hat{\theta} t^*_{1-\alpha/2}, \hat{\theta} + \hat{S}_{e^*}(\hat{\theta})t^*_{1-\alpha/2}] \tag{2.11}$$

respectively, for bootstrap prediction interval

Table 1:

Methods	USD to NGN 95% Confidence Interval	GBP to NGN 95% Confidence Interval
Standard	[160.0169, 160.4071]	[260.1240 , 260.4692]
Boostrap – t	[159.9819 , 160.4181]	[259.1300 , 260.4692]

RESULTS AND DISCUSSIONS

The data used to illustrate an application of t he bootstrap are the daily exchange rates of US dollars and British pound sterling to Nigeria naira.

The bootstrap samples at step (i/(iii) of section 2 algorithm using proposed truncated geometric bootstrap method by Olatayo (2010),was implemented and the estimate of standard error, bias and prediction interval error were computed.

In order to assess the accuracy of an estimator. Measures of statistical error, such as standard error for USD to NGN is 0.1263 and GBP to NGN is 0.4068,and their bias are 0.001857 and -0.005811 respectively. Both the standard errors and bias are alright, but USD to NGN is more better and preferable to GBP to NGN.

A bootstrap 95% confidence interval for the parameter of the variable is as given above with wider coverage, than normal confidence interval. Thus the goal is to construct a prediction interval for the new set of values of the variables. A bootstrap 95% prediction interval based on our robust estimator is given by Table 1, the interval covers a wider range compared with confident interval. Bootstrap inference based on a robust estimator seems preferable to confidence interval with an inference procedure that only cover limited range for forecasting or prediction purpose.

Autoregressive time series model:

The data are the daily exchange rates of USD (United State of American Dollar) to NGN (Nigerian Naira) and GBP (Great Britain Pound Sterling) to NGN. Let the account for t^{th} day be X_t after centering the data that is replacing x_t by $x_t - \bar{x}$ we fit a first order autoregressive model

$$x_t = \phi x_{t-1} + \xi_t \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.$$

$$\text{where}\,\xi \approx \text{iid N}\left(0,\sigma^2\right).$$

The estimate $\hat{\phi}$ turned out to be 0.1379 for USD to NGN and the estimate $\hat{\phi}$ of GBP to NGN is 0.6758, therefore the model for USD to NGN is

$$x_t = 0.1379\, x_{t-1} + \xi_t$$

and for GBP to NGN will be

$$x_t = 0.675\, x_{t-1} + \xi_t\,.$$

The forecast were generated based on the fitted models, which reveals that the USD to NGN will appreciate better than GBP to NGN in the nearest features.

Conclusion:

Lets pause to consider the suitability or otherwise of the variables in the models analyzed. So far the attempt is estimating the exchange rate of US$ to Nigerian ₦ with that of exchange rate of Great Britain £ to Nigerian ₦ , or put in another form, we are trying to explain a major financial economic factor of Nigeria relative to the US and Great Britain economies. This is an “antipodal” theory of financial economy, which is the measure of economic strength of a nation. From the forecast values, attention must be given to pound sterling, in other to improve its values, whereas, with time Naira to dollars will appreciate.

Bootstrap prediction intervals represent a useful addition to the traditional set of measures to assess the accuracy of forecast. The asymptotic properties of the intervals do not depend upon the sampling distribution, and the bootstrap results suggest that the invariance approximately holds for relatively all sample sizes.

REFERENCES

Efron, B., 1979. Computer and the theory of statistics: thinking the unthinkable, SIAM. Review, 21: 460-480.

Efron, B., 1982. The Jacknife, the bootstrap and other resampling plans. 38 of CBMS-NSF. Regional Conference Series in Applied Mathematics. SIAM.

Efron, B. and R. Tibshirani, 1986. The bootstrap measures for standard errors, confidence intervals, other measures of statistical accuracy. Statistical Science, 1: 54-77.

Efron, B. and R. Tibshirani, 1993. An introduction to the bootstrap. Chapman and Hall / CRC.

Kalman, R.E., 1960. A new approach to linear filtering and prediction problems. Trans ASME. J. Basic Eng., 82: 35-45.

Kolmogorov, A.N., 1941. Interpolation and extrapolation. Bull.Acad.Sci.U.R.S.S., 5: 3-14.

Olatayo, T.O., 2010. On truncated geometric bootstrap method for stochastic time series process. An Unpublished Ph.D. Thesis, Department of Statistcs, University of Ibadan, Ibadan Nigeria.

Stine, R.A., 1982. Prediction intervals for Time Series. Ph.D., Dissertation, Princeton University.

Stine, R.A., 1985. Bootstrap prediction intervals for regression. J. Amer. Statist. Assoc., 80: 1026-1031.

Wei, W.W.S., 1989. Time Series Analysis, Univariate and Multivariate Methods. Addison-Wesley Publishing Company, Inc.

Whittle, P., 1953. Estimation and information in stationary time series. Arkivfur Mathematik, 2: 423.

Wiener, N., 1949. The extrapolation, interpolation and smoothing of stationary Time Series with Engineering Application .New York: Wiley.

Yaglom, A. M., 1962. An introduction to the theory of stationary random functions. Prentice Hall, Inc., Englewood Cliffs, N.J.