CENTRAL ASIAN JOURNAL OF THEORETICAL AND APPLIED SCIENCE

https://cajotas.centralasianstudies.org/index.php/CAJOTAS

Volume: 06 Issue: 03 | July 2025 ISSN: 2660-5317

Article

Enhancing Inferential Accuracy with Bootstrap Methods: A Statistical Approach to Insurance Data in Urban Planning Contexts

Sufian Munther Salih[1]

1. Department of Management Economics Banking, College of Bussiness Economics, AlNahrain University, Baghdad, Iraq

* Correspondence: sufian.m.salih@nahrainuniv.edu.iq

Abstract: In applied statistics, particularly in domains like economics and insurance, small sample sizes and non-ideal data conditions often compromise the accuracy of traditional inferential methods. Iraqi insurance sector data from 1999 to 2014 offers only 16 observations, making classical regression approaches unsuitable due to their dependence on large sample assumptions. There is insufficient understanding of how bootstrapping methods compare in terms of estimation reliability under such constrained data conditions. This study aims to assess the effectiveness of bootstrap resampling both error-based and observation-based in estimating regression parameters related to premium retention rates in the Iraqi insurance industry. Empirical comparisons using mean squared error (MSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) reveal that the error resampling method significantly outperforms the observation resampling method in fitting accuracy. The study further identifies six key predictors of premium retention rate, including corporate capital, changes in underwriting, population size, bank credit, bank deposits, and education levels (risk aversion). This research uniquely applies bootstrap methods to an underexplored dataset within the insurance sector of a developing country, demonstrating how inferential robustness can be achieved without reliance on large samples or strict distributional assumptions. The findings support the broader adoption of error-based bootstrap techniques in policy modeling and financial forecasting, particularly in data-scarce environments common to developing economies and urban planning contexts.

Keywords: bootstrap method, inferential statistics, regression estimation, regional planning, data-driven planning.

Citation: Salih S. M. Enhancing Inferential Accuracy with Bootstrap Methods: A Statistical Approach to Insurance Data in Urban Planning Contexts. Central Asian Journal of Theoretical and Applied Science 2025, 6(3), 249-256

Received: 18th Feb 2025
Revised: 11th Mar 2025
Accepted: 24th Apr 2025
Published: 21th May 2025

1. Introduction

In the broader context of **urban and regional planning**, the relevance of the bootstrap method becomes even more pronounced. Planners and policymakers frequently face data limitations when analyzing spatial patterns, demographic shifts, or infrastructure needs [1].

Datasets in urban planning often include irregularities due to spatial heterogeneity, incomplete records, or sociopolitical influences that distort conventional statistical assumptions. Here, the bootstrap method empowers analysts to generate credible insights from limited or noisy data, improving the quality of evidence-based planning decisions [2].

For instance, when evaluating the relationship between urban growth and access to public services such as healthcare, education, or transportation—the bootstrap technique can help refine regression models used to inform spatial equity assessments and policy interventions [3].

Its capacity to provide robust parameter estimates makes it a valuable asset in understanding the complexities of rapidly changing urban environments, especially in developing contexts. In this research, the author applies the bootstrap method to estimate regression parameters related to the **premium retention ratio** in Iraqi insurance companies. While the specific application is grounded in the insurance sector, the methodological implications resonate far beyond. It exemplifies how statistical innovation can support sectors critical to **urban resilience**, such as financial services, risk management, and infrastructure development, ultimately reinforcing data-driven approaches in urban and regional planning frameworks [4].

2. Materials and Methods

1.1 First Approach: Resampling the Errors

This method involves estimating the regression parameters for the entire sample and then computing the residual errors. Numerous resamples are drawn with replacement from these residuals. For each resampled set, the dependent variable is recalculated, and the regression parameters are re-estimated. The final parameter estimate is obtained by averaging these resampled estimates [5].

1.2 Second Approach: Resampling the Observations

In this approach, multiple resamples are drawn with replacement from the full dataset. The regression parameters are then estimated separately for each resampled dataset, and the final parameter estimate is determined by averaging the estimates from all resamples [6].

The researcher will compare these two approaches using goodness-of-fit measures, including Mean Squared Error (MSE), Mean Absolute Error (MAE), and Mean Relative Error (MRE) to assess the accuracy and reliability of the estimations [7].

2- General Framework of the Study

2.1 Research Problem

During constructing a regression model, we usually rely on observations from a sample that is often randomly drawn from the study population. To obtain reliable results that can be generalized to the statistical population, the sample must be randomly selected and representative of the population from which it was drawn. In order for the sample to be random and representative, researchers must select an appropriate sample size for any statistical analysis. Most researchers have suggested that the sample size should not be less than 30 observations [8].

When fitting a regression model for the premium retention rate of Iraqi insurance companies, it was found that the available data covered the period from 1999 to 2014, consisting of only 16 observations. This sample size is clearly insufficient, making it necessary to explore an appropriate statistical method to fit the model in such cases.

Thus, the research problem is:

3. Results and Discussion

"The lack of a high-quality regression model in cases where the study includes a small number of observations."

2.2 Research Objectives

The objectives of this study are as follows:

To estimate regression parameters by using the bootstrap method.

To compare the two resampling methods of errors and method of resampling of observation, in estimating regression parameters using the bootstrap technique [9].

2.3 Research Hypothesis

The study aims to test the following hypothesis:

The error resampling method provides a better fit quality than the observation resampling method.

2.4 Research Significance

The significance of this study lies in the following points:

Obtaining a regression model that explains the changes in the premium retention rate of Iraqi insurance companies [10].

Identifying the most efficient method for estimating regression parameters using the bootstrap technique.

2.5 Study Population

The study population consists of all insurance companies operating in the Kurdistan region of Iraq, including the Saudi Holding Insurance Company, which practicing of property insurance for both national and foreign entities, whether they are insurance companies or brokerage firms [11].

2.6 Study Scope

1. Temporal Scope: The study will use available data from 1999 to 2014.
2. Application Scope: The study will focus on estimating the premium retention rate in the Iraqi insurance market.

2.7 Data Sources

The study relies on published and unpublished statistical data from official sources such as the General Statistics Office, annual reports of the Central Bank of Iraq, the Arab Monetary Fund, and the Banking Institute, covering the period from 2009 to 2024.

$$0 < B_{ols}^{b} - B_{OLS} < -2R_{(KC)} -2R_{(KC)} < B_{ols}^{(b)} - B_{ols} < 0 \; R_{(KS)} \quad (1)$$

Represent percentage kadilar andcingi

The LAL represents the bootstrap regression estimator $B_{ols}^{(b)}$, and B_{ols}represents the least-squares regression magnitude. This has the effect of providing more accurate results than previous estimators. The mean square error (MSE) criterion was used to test the accuracy of the new estimators. The relative efficiency of all proposed estimators was also found, and the theoretical aspect was supported with applied and experimental examples [12].

2.8 Nature of the Bootstrap Model

The bootstrap method is used to study the relationship between the dependent variable and a set of independent variables, particularly in cases of small sample sizes. Since the available data covers only, number of observations 16 only. The bootstrap method was necessary to derive an equation that explains the variations in the premium retention rate of Saudi insurance companies based on independent variables, whether insurance-related or economic [13].

To define the nature of the bootstrap model, we will present it through the following points:

2.8.1 Advantages of the Bootstrap Method

The role of computers in statistical inference has grown, with resampling methods such as the bootstrap and jackknife becoming essential. Bootstrap is applied in various cases, including linear and nonlinear regression, time series models, multivariate statistical analysis, nonparametric regression, confidence interval estimation, and censored data analysis [14].

The bootstrap method can be used in many situations, including linear and nonlinear regression applications, time series models, multivariate statistical analysis, nonparametric regression, confidence interval estimation, and truncated data analysis [15].

The researcher resorted to using this proposed method for the following reasons .

The bootstrap method is used to estimate regression coefficients, especially if the condition of a normal distribution is not met in the data used.

The bootstrap method can be used in both parametric and nonparametric cases [16].

The bootstrap method provides more accurate estimates of error variance than those obtained from other traditional methods.

The bootstrap method was developed to overcome the problem of the lack of an assumption of independence of observations, which leads to a lack of analytical accuracy [16].

The final bootstrap estimate distribution is normal, even if the original distribution under study is not. Because it does not require distribution-specific assumptions (such as that errors follow a normal distribution), the bootstrap method can yield more accurate inferences when the data are poorly representative or when the sample size is small.

Bootstrapping can be applied to statistics with sampling distributions that are difficult to derive, even approximately. It is relatively easy to apply bootstrapping to samples other than simple random samples (such as stratified and cluster samples).

2.8.2 Estimating a Regression Model Using Bootstrapping

There are two methods for estimating a regression model using bootstrapping, and the details of these methods are as follows:

The first method: Bootstrap Based on the Resampling Observations

The resampling process involves treating the x's as random variables rather than fixed ones. Suppose the vector $W_i = (Y_i, X_{ji})$of degree (1k+1x1) refers to the values of the i^{th} observation. In this case, the set of observations is the vectors W_1, W_2, W_n. The bootstrap method based on the resampling of observations is as follows:

1-Draw bootstrap samples with $(W_1^b, W_2^b \ldots\ldots, W_n^b)$ with replacement from observations with probability 1/n

For each, w, we will denote the elements of each vector by the symbol $W_i^b = (y_i^b, x_{ji}^b,$ where k j=1,2,..n j=1,2,. (2)

From this formula, the vector $Y_i^b = (y_1^b, y_2^b, \ldots\ldots, y_n^b)$, and the matrix $X_{ji}^b = X_{j1}^b, X_{j2}^b, \ldots\ldots X_{jn}^b$

2- Calculate the regression coefficients using the OLS method from the booster sample $\hat{B}^{(b1)}$ =($(X^{(b)\prime}X^{(b)})^{-1}X^{(b)\prime}Y^{(b)}$ (3)

3-Repeat steps 1 and 2 for each r=1,2,

Where B is the number of booster samples

4-We obtain the probability distribution $(F(\hat{B}^{(b)})$for the bootstrap estimates $\hat{B}^{(b1)}, \hat{B}^{(b2)}, \ldots\ldots, \hat{B}^{(bB)}$and from this distribution we obtain the regression coefficient estimates as follows:

The bootstrap estimate of the regression coefficient is the mean of the distribution (F(B ^(b), i.e.:

$\widehat{B}^{(b)} = \sum_{r=1}^{B} \quad \hat{B}^{(br)}/B = \underline{\widehat{B}}^{(br)}$ (4)

5 -Therefore, the regression equation for the bootstrap is Y ̂=XB ^(b)+ε, where B ^(b) is an unbiased estimator

For the second method: the resampling-based bootstrap For errors

Bootstrap Based On The Resampling Errors

The resampling process involves treating x's as constants. The bootstrap method based on resampling errors is as follows:

1 -Fit a least-squares regression equation from the total sample.

2-Calculate the error values, $e_i = y_i + \hat{y}_i$

3-Draw bootstrap samples of size n with substitution $(e_1^{(b)}, e_2^{(b)}, \ldots\ldots e_n^{(b)})$

from the e_i values calculated in step 2, with probability 1/n for each e_i value..

4-Calculate the bootstrap values of Y by adding the residuals resulting from resampling in step 3 to the equation estimated in step 1, assuming a constant regression relationship: $Y^{(b)} = X\hat{B} + e^{(b)}$

5 -Obtain the bootstrap parameter estimates using the least squares method from the first bootstrap sample: Thus, $\hat{B}^{(b1)} = (X'X)^{-1}X'Y^{(b)}$ (5)

6- Repeat steps 3, 4, and 5 for each B,..r=1, 2, Finally, repeat steps 4 and 5 as in the case of resampling observations. Note the following about the resampling error method:

1-When x's are constant, the bootstrap method achieves similarity between the estimated values of Yˆin the sample and the conditional expectation Y in the population, and between the residuals E in the sample and the error ε in the population.

2-Although there are no assumptions regarding the shape of the error distribution, the bootstrap procedure, by constructing Y^(b) according to the linear model, assumes that the model's basic shape is correct.

3-In addition, by resampling the residuals and then randomly adding them to the estimated values, this assumes that the errors are identically distributed. For example, if the true errors have non-constant variance, this will not be reflected in the resampling of the residuals. Similarly, the effect of outliers will be eliminated as a result of the resampling process.

2.8.3 Bootstrap estimates of bias, variance, and confidence interval for parameters are as follows:

Bootstrap bias is equal to $bi\hat{a}s_b = \hat{B}^{(b)} - B$ (6)

-Bootstrap variance from the distribution ((F(B ^(b) is calculated as follows (Liu, 1988; Stine 1990)

$Var(\hat{B}^{(b)}) = \sum_{r=1}^{B} \left[(\hat{B}^{(br)} - \hat{B}^{(b)})(\hat{B}^{(br)} - \hat{B}^{(b)})'\right]$l(B-1) r=1,2,…..,B) (7)

- Bootstrap confidence interval is as follows:

$$\hat{B}^{(b)} - t_{n-p,\frac{a}{2}} * S_e\left(\hat{B}^{(b)}\right) < B < \hat{B}^{(b)} + t_{n-p,\frac{a}{2}} * S_e(\hat{B}^{(b)})$$

2.8.4 Determining the goodness of fit of the estimated model will be done through the following metrics:

1- Medium Squared Errors

$MSE = \sum_{i=1}^{n}$ ―――――――――― n e_i^2 (8)

2- Medium Absolute Errors

$MAS = \sum_{i=1}^{n}$ ―――――――――――― n $|ei|$ (9)

3-Relative Mean Absolute Errors

$100MApE = \sum_{i=1}^{n}$ ―――――――――――― n $\left|\frac{e}{y}\right|$ = (10)

2.9 Practical Application of the Bootstrap Model

Here, the practical application of the proposed model will be conducted. Using this model, the validity or falsity of the main hypothesis of the research can be proven, which states, "The error resampling method has a higher accuracy of reconciliation than the observation resampling method."

The most important economic and insurance variables that explain the premium retention rate of Saudi insurance companies are:

1-Dependent variable: Premium retention rate, symbolized by Y.

2- Independent variables, which include the following:

X1: Loss ratio

X2: Insurance density

X3: Corporate capital

X4: Bank credit

X5: Line avoidance

X6: Population size *

X7: Change in underwriting.

X8: Bank deposits

X9: Number of learners (risk aversion)

To test the validity of the study hypothesis, the researcher will use the Efron method to analyze a bootstrap regression model using the statistical program Mathcad. This method is based on the concept of sampling with a very large number of samples, with the number of samples surveyed with a return of 1,000. The steps for applying the bootstrap regression model are as follows:

First: Data examination

The actual determination of any statistical model is done through analyzing actual historical data using statistical steps that begin with processing the data to make it consistent or homogeneous, whether by taking the square root, logarithm, or some algebraic operations for both the independent and dependent variables. By examining the data, we found its inconsistency. Therefore, the following calculations were performed:

The independent variables (bank credit, risk aversion) were divided by 1000.

The bank deposit variable was divided by 10,000.

Second: Identifying the variables that most influence the premium retention rate.

Statistically, the data were analyzed using stepwise regression analysis. It became clear that the variables that most influence the premium retention rate are:

X_1 Corporate Capital

X_2 Change in Subscription

X_3 Population

X_4 Bank Credit

X_5 Bank Deposits

X_6 Number of Educated Persons (Risk Aversion)

The regression equation is as follows:

$\hat{y} = 30987 + 9.196x_1 + 0.099x_2 + 0.686x_3 - 0.098x_4 + 0.188x_5 + 0.147x_6$ (11)

Sig. .000 .002 .004 . 034 . 000 . 021 . 000

$R^2 = 0.997$, $D.W = 2.941$, $n = 16$

The relationship is explained as follows:

- There is a direct relationship between the premium retention rate and corporate capital. The greater the company's capital, the greater its ability to cope with risk, leading to a higher retention limit.

There is a direct relationship between the premium retention rate and changes in underwriting. Increased underwriting increases the premium retention rate, which in turn increases the premium retention rate.

There is a direct relationship between the premium retention rate and population size. As the population increases, the underwriting volume for risks increases, which in turn increases premiums, thus increasing the premium retention rate.

There is an inverse relationship between the premium retention rate and bank credit. Increased borrowing reduces the amounts deposited with banks, which reduces the demand for insurance and, consequently, the premium retention rate.

There is a direct relationship between the premium retention rate and bank deposits. Increased bank deposits lead to increased demand for insurance against the risks of theft and breach of trust, which, in turn, increases the premium volume and increases the premium retention rate. There is a direct relationship between premium retention rates and risk aversion. As the number of educated people increases, insurance awareness and demand for insurance increase, which in turn increases insurance underwriting and, consequently, premium retention rates.

It is worth noting that none of the regression problems were studied, given that the bootstrap method is very useful as an alternative to parametric estimations when there is doubt about the validity of some hypotheses.

Third: Estimating the parameters of the bootstrap regression model using the resampling method.

We obtained the following model:

$\hat{y} = 68.458 - 2.344x_1 - 0.05x_2 - 0.286x_3 - 0.023x_4 + 0.033x_5 + 0.083x_6$ (12)

The model shows that the signs of the model's coefficients differ from those of the model estimated from the total sample, which provides an illogical interpretation of the relationship between the premium retention rate and the explanatory variables in the equation [17].

Egyptian Statistical Journal, Volume 60, December 2016 Issue

Fourth, estimating the parameters of the bootstrap regression model by following the error re-sampling method, we obtained the following model:

$$\hat{y} = 30.962 + 9.212x_1 + 0.099x_2 + 0.687x_3 + 0.098x_4 + 0.188x_5 + 0.147x_6$$

The model shows that the model's coefficients agree with the coefficients of the model estimated from the total sample [18].

Fifth: Comparison between the observational resampling model and the error resampling model.

The following measures were calculated for each model:

1- Medium squared errors

$$MSE = \frac{\sum_{i=1}^{n} e_i^2}{n} \quad (13)$$

2-Absolute average errors

$$MAE = \frac{\sum_{i=1}^{n} |e_i|}{n} \quad (14)$$

3- Relative average of absolute errors

$$MAPE = \frac{\sum_{i=1}^{n} \left|\frac{e_i}{Y_i}\right|}{n} * 100 \quad (15)$$

The following table illustrates this:

The table illustrates that the error-based bootstrap model significantly outperforms the observation-based model in regression estimation accuracy, showing markedly lower values across all metrics MSE (0.14627 vs. 52.014), MAE (0.352 vs. 6.158), and MAPE (0.577 vs. 9.861) indicating superior model fit and predictive reliability (Table 1).

Table (1). Measures of the accuracy of the fit for the re-sampling model for observations and the re-sampling model for errors

	MSE	*MAE*	*MAPE*
Re-sampling model for observation	*52.014*	*6.158*	*9.861*
Re-sampling model for errors	*0.14627*	*0.352*	*0.577*

The previous table shows that the error resampling model has a higher fitting accuracy than the observation resampling model across all three metrics. The values of the three metrics for the error resampling model are lower than those for the observation resampling model.

4. Conclusion

Resampling the residuals and then randomly adding them to the estimated values assumes that the errors are identically distributed. For example, if the true errors have non-constant variance, this will not be reflected in the resampling of the residuals. Similarly, the effect of outliers will be diminished by the resampling process.

-It was found that the variables that most influence the premium retention rate are:

x1 Corporate capital

x2 Change in subscription

x3 Population

x4 Bank credit

x5 Bank deposits

x6 Number of educated people (risk aversion)

3- Using the resampling method for observations, the model revealed that the signs of the model's coefficients differed from the signs of the coefficients for the model estimated from the total sample, which provides an illogical interpretation of the relationship between the premium retention rate and the explanatory variables in the equation.

4- Using the resampling method for errors, the model revealed that the signs of the model's coefficients agreed with the signs of the coefficients for the model estimated from the total sample, which provides a logical interpretation of the relationship between the premium retention rate and the explanatory variables in the equation.

5- It was found that the resampling model for errors had a higher accuracy of fit than the resampling model for observations. 6-2 Recommendations

1. Adopt the bootstrap method to estimate regression model parameters in cases of small sample size or presence of...
2. Try using another model that combines the bootstrap method, especially in cases of contamination or noise in the data.
3. It is preferable to use the resampling method to estimate regression model parameters.

REFERENCES

[1] M. H. Ahmed, "Towards a quantitative model for determining the optimal retention limit and its impact on the probability of company collapse," *New Horizons Journal*, Faculty of Commerce, Menoufia University, vol. 10, no. 1, 1998.

[2] M. M. El-Derini, "Using the statistical bootstrap method in selecting the optimal model for the private savings function in Egypt," *J. Egyptian Stat. Soc. (JESS)*, vol. 21, no. 1, 2005.

[3] B. M. El-Omari, *Treating the problem of heterogeneity of variances using the statistical bootstrap method*, M.S. thesis, Faculty of Commerce, Tanta University, 2001.

[4] A. H. A. El-Qadhi, "A study of the absorptive capacity of the insurance market," Academy of Scientific Research, Risk and Insurance Research Division, 2007.

[5] M. A. R. Ismail, *Linear Regression Analysis*. Riyadh: Institute of Public Administration, Research Center, 2001.

[6] H. Hussein and N. El-Hamidi, "Using the retention rate and the utilized capacity ratio in insurance portfolio risk analysis," presented at the 7th Int. Forum on the Insurance Industry, Hassiba University, Faculty of Economics, Business and Management Sciences, 2012.

[7] M. Abdul Mawla and M. Al-Mahdi, "Towards an integrated framework for determining the factors affecting retention limits in life insurance: An applied study on Arab direct insurance companies," *New Horizons J. Bus. Stud.*, Faculty of Commerce, Menoufia University, no. 2, 1995.

[8] M. A. G. Muawad, *A proposed statistical model for predicting Egyptian cotton exports*, Ph.D. dissertation, Faculty of Commerce, Ain Shams University, 2007.

[9] S. S. D. El-Din, "Using the bootstrap method to find new estimators for the ratio in simple random sampling," *Iraqi J. Statist. Sci.*, vol. 16, 2009.

[10] B. Efron, "Bootstrap methods: Another look at the jackknife," *Ann. Statist.*, vol. 7, no. 1, pp. 1–26, 1979.

[11] D. A. Freedman, "Bootstrapping regression models," *Ann. Statist.*, vol. 9, no. 6, pp. 1218–1228, 1981.

[12] J. Fox, *Applied Regression Analysis, Linear Models and Related Methods*. Thousand Oaks, CA: Sage Publications, 1997.

[13] Y. R. Liu, "Bootstrap procedures under some non-i.i.d. models," *Ann. Statist.*, 1988.

[14] J. Shao, "Bootstrap model selection," *J. Am. Stat. Assoc.*, vol. 91, no. 434, pp. 655–665, Jun. 1996.

[15] J. Shao and D. Tu, *The Jackknife and Bootstrap*. New York: Springer, 1995.

[16] S. Sahinler and D. Topuz, "Bootstrap and jackknife resampling algorithms for estimation parameters," *J. Appl. Quant. Methods*, vol. 2, no. 2, pp. 187–197, 2007.

[17] R. Stine, "Bootstrap prediction intervals for regression," *J. Am. Stat. Assoc.*, vol. 80, pp. 1026–1031, 1985.

[18] R. Stine, "Modern methods of data analysis," in *Modern Methods of Data Analysis*, J. Fox, Ed. Scotland: Sage Publications, 1990, pp. 325–373.