Proceedings of the Faculty of Economics in East Sarajevo
Year 2024, Issue 29, pp. 11-22
Received: 22th August 2024

UDC 636.082:519.852
DOI: 10.7251/ZREFIS2429011R
Original Scientific Paper

Zarrukh Rakhimov

Westminster International University in Tashkent,
Tashkent,
Uzbekistan

✉ zrakhimov@wiut.uz

HANDLING HETEROSCEDASTICITY IN LINEAR MODELS: HUBER-WHITE STANDARD ERRORS VS BOOTSTRAP CONFIDENCE INTERVALS

РЈЕШАВАЊЕ ХЕТЕРОСКЕДАСТИЧНОСТИ У ЛИНЕАРНИМ МОДЕЛИМА: ХУБЕР-ВАЈТОVЕ СТАНДАРДНЕ ГРЕШКЕ У ОДНОСУ НА БУДСТРАР ИНТЕРВАЛ ПОВЈЕРЕЊА

Summary: *Heteroscedasticity are one of the several violations of the assumptions of OLS. If no remedy applied, residuals with non-constant variance can lead to inaccurate and biased results. Academia has suggested a wide range of remedies to tackle with heteroscedastic residuals. In this study, we suggest another approach, bootstrapping of dataset to construct our confidence intervals. In order to compare the outcome, we look at Huber-White method and look at its performance against bootstrap intervals. Results indicate that bootstrap intervals perform equally well as Huber-White based confidence intervals. This indicates that bootstrap method, similar to Huber-White approach, can be a good remedy for heteroscedasticity.*

Keywords: *heteroscedasticity, homoscedasticity, linear model, confidence Interval, bootstrap, Huber-White, residuals, accuracy*

JEL Classification: *C15, C87, E22*

Резиме: *Хетероскедастичност предтавља једну од нарушених претпоставки ОЛС-а. Ако се не примјени одговарајуће рјешење, резидуали са не-константном варијансом могу довести до нетачних и пристрасних оцјена параметара. Академска заједница је предложила низ решења за решавање овог проблема. У овој студији предлажемо другачији приступ, поновљено узорковање података (бутстрап) како бисмо конструисали интервале повјерења. Посматрамо Хубер-Wајтове методу и њену примјену у поређењу са боотстрап интервалима како бисмо упоредили резултате. Резултати показују да боотстрап интервали дају једнако добре резултате као интервали повјерења засновани на Хубер-Wајтовој методи. То имплицира да боотстрап метода, слично Хубер-Wхите приступу, може бити добро решење за хетероскедастичност.*

Кључне ријечи: *хетероскедастичност, хомоскедастичност, линеарни модел, интервал поверења, бутстрап, Хубер-Вхите, резидуали, тачност*

ЈЕЛ касификација: *C15, C87, E22*

INTRODUCTION

Linear models often derived using the Ordinary Least Squares (OLS) method, is one of the most popular methods of deriving impact of one variable on another and predicting future outcomes. OLS approach is currently used in wide range of sciences such as economics, finance, psychology, sociology to name just a few. The popularity of OLS approach comes from its simplicity, good efficiency and ease of interpretation, although it is almost an approximation to real life relationships. Yet, linear models require a set of theoretical assumptions to be satisfied in order to have accurate unbiased efficient point and intervals estimates. The core assumptions are homoscedasticity, no autocorrelation, and normality of residuals and violation of any of them might lead to inefficient or even biased confidence interval estimates. As a result of such violations, researchers and practitioners might do wrong conclusions when evaluating impact of one variable on another, affecting decisions in fields of medicine, economics and engineering. To cite an example, in studies of income and its determinants, which is very common subject of research in economics and finance, income dataset very often violate homoscedasticity assumptions. When no remedy is applied, OLS estimations of coefficients or their confidence intervals are usually inefficient and often biased.

Situations when homoscedasticity is not satisfied is also known as heteroscedasticity. This implies that variances of the error term are not constant across observations. Presence of heteroscedasticity can distort standard errors and make confidence intervals less accurate. In practice, it means that studies that have heteroscedastic dataset can draw incorrect conclusions about variable significance or effect sizes when no remedy is applied to mitigate heteroscedasticity. One of the techniques widely used in practice is the Huber-White standard error correction. Huber-White is a mathematical transformation, which makes OLS estimations more robust to heteroscedasticity without altering the functional form of the model. Another method of handling heteroscedasticity is suggested in this study known as bootstrapping. Bootstrap confidence intervals have no distributional assumptions, which makes this approach less sensitive to non-constant variances.

This paper looks into the practical implications of heteroscedasticity on OLS estimates and investigates how both the Huber-White robust intervals and bootstrap methods can serve as ways to handle heteroscedasticity. By comparing these approaches with heteroscedastic data, we want to shed some light into how practitioners can select the best method for their specific case, balancing statistical robustness with computational efficiency.

1. LITERATURE REVIEW

Bootstrap method is a resampling method of a given dataset to build a sampling distribution of a specific statistic. Bootstrapping has become popular because it has proven to provide reliable inferences in many cases even when underlying assumptions are not satisfied. This also applied to cases of heteroscedastic residuals which is first discussed in papers of Efron (1979). Since then, theoretical foundations have been concentrated on justifying validity and efficiency of bootstrap confidence intervals with non-constant variance of errors (Davison and Hinkley 1997).

In the context of linear models, there have been primarily two types of bootstrapping used for estimating point and interval estimates, bootstrapping residuals and bootstrapping pairs (Chernick and LaBudde 2011).

Bootstrapping residuals: This method of bootstrapping was first introduced by Efron (1982). Imagine we have the following model

$$Y_i = g_i(\beta) + e_i, \quad \text{for} \quad i=1,2,\ldots,n$$

where $g_i(\beta)$ is a function with a known form. To estimate β , we minimize distance between our true dependent variable Y_i and estimated function $g_i(\beta)$. These distances are expressed in terms of residuals $\widehat{e_i} = Y_i - g_i(\widehat{\beta})$. The idea behind Wild bootstrap is to take the distribution of residuals each having probability of *1/n* for *i=1,2,….,n* and sample *n* times from this distribution to get bootstrap sample of residuals which can be denoted as $(e_1, e_2, e_3, \ldots, e_n)$. Afterwards, bootstrap dependent variable can be generated using $Y_i^* = g_i(\widehat{\beta}) + e_i^*$. Now, as we have our bootstrap dataset, we use simple OLS method to estimate β^*. We repeat the above procedure *B* times to get a distribution of β_j^* estimates for *j=1,2,….,B*. One can get standard deviation of β^* to build bootstrap confidence intervals.

Bootstrapping pairs: bootstrapping pairs is a rather simple but powerful approach proposed first by Freedman (1981). Under this approach, we resample independent and dependent variables from the original sample which results in a bootstrap sample. We then use usual OLS method to estimate β^* from the bootstrap sample. This procedure is repeated B times in order to get distribution of coefficients β_j^* estimates for *j=1,2,….,B*. This distribution in turn can give bootstrap standard deviation.

Efron and Tibshirani (1986) conclude that two approaches are equivalent when the model is correctly specified, but they can perform differently when the sample is small. Flachaire (2003) compared bootstrapping residuals and bootstrapping pairs when the model is correctly specified and when heteroscedasticity is present in the linear models. Flachaire (2003) concludes that when a proper transformation to the residual term is applied (wild bootstrap), residuals bootstrap performs better than bootstrapping pairs. Chernick and LaBudde (2011) conclude however that bootstrapping vectors are less sensitive to violations of model assumptions and can still perform well if those assumptions are not met. This can be explained by the fact that the vector method does not depend on model structure while bootstrapping residuals do.

Other approaches are stationary bootstrap (Politis and Roman 1994), and the percentile-t bootstrap (Diciccio and Efron 1992) each used under different scenarios of non-constant variance of the residuals.

On comparing Huber-White robust confidence intervals and bootstrap confidence interval, there has been numerous studies carried out and concluded that each approaches has advantages in specific scenarios. Long and Ervin (2000) claid that Huber-White standard errors are rather quick to estimate and perform well when data is sufficiently large and heteroscedastic behavior is not severe. Yet, Huber-White robust estimates can depict inconsistent results when sample is small or heteroscedasticity is arousing from model complexity and varies a lot (MacKinnon and White 1985).

In contrast, Efron and Tibshirani (1994) claim that in extreme cases of heteroscedasticity or small samples, bootstrap can be more suitable. Cameron and Trivedi (2005) in their paper also conclude that bootstrap confidence intervals resulted in more accurate intervals than robust standard errors in above conditions, arguing that bootstrap is better at capturing the true sampling distribution in the presence of heteroscedasticity and other non-normal errors.

Yet, many studies (e.g. Davidson and MacKinnon, 2004) conclude that bootstapping comes at a computational cost especially when sample is large and resampling will take time and computing power. In contrast, transformational remedies such as Huber-White standard errors do not suffer from this shortcoming, as it is computationally easy to implement Huber-White approach.

2. LINEAR REGRESSION MODELS

First of all, let's look into how linear models are built and how coefficients as well as their intervals are estimated. As mentioned earlier, the linear model evaluates the impact of one or more variables (explanatory variables) to another variable (explained or dependent variable). This is done by estimating coefficients of estimates of each explanatory variable. For instance, imagine that we want to evaluate whether your year of education affects your income and by how much. If we build our simple OLS model where income is dependent "Y" variable, and year of education is "X_1" explanatory variable, then coefficient of "years of educations" ($\beta_1$1) shows the size and direction (positive or negative) of the impact.

$$Y = \beta_0 + \beta_1 * X_1 + e$$

Where

$Y -$ dependent variable,

β_0 – intercept,

$\beta_1 -$ coefficient of first explanatory variable

$X_1 -$ explanatory or independent variable

$e -$ error or residual term

The above model is the simplest one variable example of linear regression and usually most studies take into account more explanatory variables that will improve the model (there are metrics to evaluate whether a model is improving or not, e.g. adj. R squared, AIC, MSE).

Estimation of coefficients in the above model is done with the method of least squares commonly known as OLS (ordinary least squares). Least squares estimate of β_1 is given by:

$$\widehat{\beta_1} = \frac{\sum_{i=1}^{n} (X_i - \underline{X})(Y_i - \underline{Y})}{\sum_{i=1}^{n} (X_i - \underline{X})^2}$$

where

$n -$ number of observations

$X_i -$ value of the independent variable for the i-th observation

$Y_i -$ value of the dependent variable for the i-th observation

$\underline{X} -$ mean of the independent variable X

$\underline{Y} -$ mean of the independent variable Y

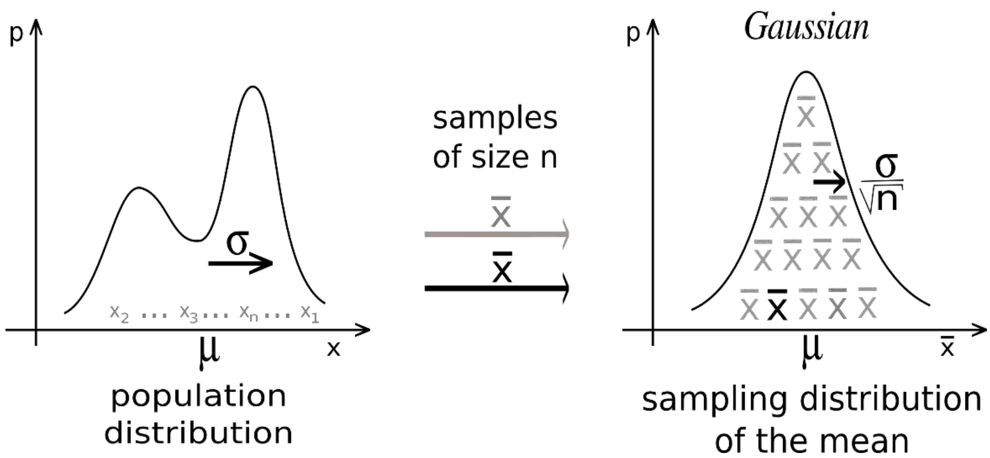
3. TRADITIONAL CONFIDENCE INTERVALS

Researchers are often interested not only in point estimates of coefficient, but also interval estimations. This is because point estimates of coefficients are always an approximation to true population value. In contrast, interval estimations, commonly known as confidence intervals, have a set of advantages. Firstly, it gives a range of values where true population value can be located. Secondly, confidence intervals will indicate whether the true population parameter might be equal to 0. In other words, whether the effect of that specific explanatory/independent variable to dependent variable is insignificant. Currently, all statistical softwares provide both point and interval estimates by default. Below, we will look at the theoretical side of building confidence intervals of coefficients of linear models.

Central Limit Theorem

Central Limit theorem (CLM) is the core concept of statistics that is employed also in building confidence intervals. The theory says that irrespectful of the true population dataset, if one derives many sample averages from many samples generated from the same population, then the distribution of sample averages is approximately normal (also referred as Gaussian, see graph below) (Lind et al, 1967). The midpoint of resulting distribution of sample averages will be equal to the true population mean (see Figure 1). This is a very strong finding that can also be applied in confidence interval construction.

Figure 1: Central Limit Theorem



Source: generated by the author

In practice, we often cannot take many samples from the same population and very often left to work with only one sample. Nevertheless, one can still make some estimation regarding the population value (e.g. mean, coefficient) using the central limit theorem even when the distribution of the population dataset is not known.

Confidence interval based on CLT

Consider we have only one sample from the population data. Firstly, we can estimate the sample coefficient using the method of ordinary least squares (discussed in previous chapter). Afterwards, we can estimate standard error of the estimated coefficient using the following formula also arising from the method of least squares.

$$se\left(\hat{\beta}_1\right) = \frac{s}{\sqrt{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}}$$

where

s – standard deviation of the residuals (residual standard error)

n – number of observations

$X_i -$ value of the independent variable for the i-th observation
$\underline{X} -$ mean of the independent variable X

As distribution of $\widehat{\beta}_1$ coefficient is approximately normal distribution based on central limit theorem, we employ properties of standard normal distribution (z-distribution) and build 90%, 95% or 99% confidence intervals.

$$\widehat{\beta}_1 \pm z_{\frac{\alpha}{2}} * se\left(\widehat{\beta}_1\right)$$

where
$\widehat{\beta}_1$ - is sample coefficient estimate
$z_{\frac{\alpha}{2}}$ – is a value from the standard normal distribution the give an area of $\frac{\alpha}{2}$
$se(\widehat{\beta}_1)$ - sample variance of the coefficient

The above interval estimation is interpreted in the following way. 97% interval indicates that if we construct 100 confidence intervals from 100 random samples generated from the true population, then 97 of those confidence intervals will contain true population coefficient β_1. Also, employing this confidence interval you can verify whether population coefficient is insignificant. If estimated confidence interval contains zero, then one can suspect that the true population parameter can be equal to zero (Gujarati, 2004)

However, one can see that estimation of the standard error of the same coefficient depends on the normality of the residual term. In the presence of heteroscedasticity, standard deviation of the error term can be inflated which will result in inaccuracies in confidence interval constructions using the CLT approach (Gujarati, 2004).

Heteroscedasticity can arise from various sources, such as:

1. Omitted variables
2. Measurement error
3. Non-linearity of the relationship of dependent and independent variable
4. Outliers
5. Residual variance that deviates with time
6. Endogeneity
7. Model misspecification

If no remedy is applied to heteroscedasticity in residuals, it will make the standard error of the residuals biased and can lead to wrong conclusions in hypothesis testing. Academia suggested a set of way on how heteroscedasticity, such transforming variables, weighted least squares, including important variables and many others (Greene 2021)

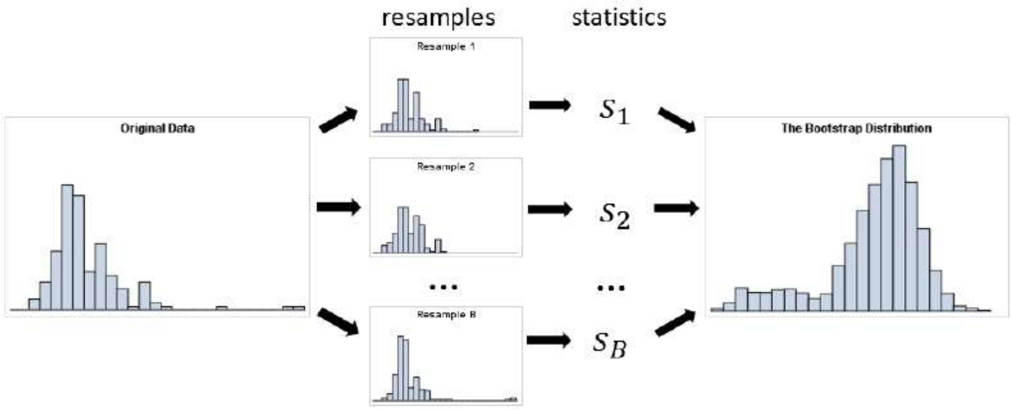
Below, we suggest another way, bootstrap, of handling heteroscedasticity in residuals for construction of our confidence intervals for coefficients.

4. BOOTSTRAP CONFIDENCE INTERVAL ESTIMATION

In the first place, it is necessary to explain the concept of bootstrapping. Bootstrap is a relatively easy resampling technique that can offer alternative ways of building confidence intervals. Bootstrap implies selecting one sample and generating many other different samples from this single original sample and estimating your parameter of interest in each newly created sample. Under the bootstrap approach, the original sample is considered as a population and we generate many other samples (known as bootstrap samples) out of it. When a large number of bootstrap samples are created, we estimate sample parameters (e.g. coefficient) from every bootstrap sample. Consequently, we will have a distribution of bootstrap sample estimates.

This distribution of bootstrap sample estimates can be used to construct our confidence intervals. For example, if we want to construct a 95 percent interval, we take 2.5th and 97.5th percentiles from bootstrap distribution. Figure 2 explains visually the method of bootstrapping.

Figure 2: Bootstrap distribution generated from original sample



Source: generated by the author

Bootstrap in case of heteroscedasticity

Consider a sample that indicates presence of heteroscedasticity as a result of measurement of error in some data points. Earlier, we discussed that heteroscedasticity can lead to a bias if we employ traditional OLS based methods of building confidence intervals. This is because standard errors of residuals will be affected by heteroscedastic residuals. This in turn influences standard errors of estimated coefficients which is used for building traditional confidence intervals. There are a set of advantages to this approach over traditional methods. Firstly, if sample size is smaller than 30 if we remove outliers from the original dataset, bootstrap interval estimation can still be derived. In contrast, traditional methods required sample size to be larger than 30 for estimates to be reliable enough. Secondly, by removing samples that contain potential outliers, our distribution of estimates should not be influenced by extreme outliers. Lastly, bootstrap distributions of estimates do not have any assumptions of true distribution of population dataset.

In contrast, bootstrap confidence intervals do not rely on standard deviation of residuals. It generally has no assumptions on the distribution of the coefficient which serves as its biggest advantage over the traditional approach.

5. HUBER-WHITE ROBUST STANDARD ERRORS

The Huber-White approach has turned into one of the commonly used methods to derive robust standard errors in regression models where residual do not have constant variance. As mentioned earlier, one of the assumptions of traditional OLS model is that the error term have a constant variance, also known as homoscedasticity, across all levels of the independent variables. When homoscedasticity is not satisfied, confidence intervals and hypothesis tests might be inaccurate or even biased. Huber (1967) and White (1980) suggested so called Huber-White method, which has proven to be a good remedy to cases of heteroscedasticity and give robust standard errors to build confidence intervals.

Theoretical Foundation of the Huber-White Method

In linear models, the mathematical estimation of $\widehat{\beta_1}$ coefficients assumes that error terms e have constant variance across observations. Mathematically speaking:

$$Var(e) = \sigma^2$$

Yet, if homoscedasticity is not satisfied, the derived variance of $\widehat{\beta_1}$ may be inaccurate and biased. The Huber-White approach resolved this problem by deriving the variance-covariance matrix to adjust for unequal error variances.

Formula of the variance of $\widehat{\beta_1}$ in OLS models is:

$$Var(\widehat{\beta_1}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}\,\mathbf{X}\,(\mathbf{X}'\mathbf{X})^{-1}$$

where $\boldsymbol{\Sigma}$ is the variance-covariance matrix of the residuals, often assumed as $\sigma^2 I$ (a constant times an identity matrix). When residuals are heteroscedastic, $\boldsymbol{\Sigma}$ changes across different X values. The Huber-White approached instead calculates $\boldsymbol{\Sigma}$ as:

$$\widehat{\Sigma}_{HC} = diag(\hat{e}_i^2)$$

where $\hat{e}_i^2$ is the squared error terms for each observation. As the result, Huber-White methods estimates variance of the coefficient using the following formula:

$$Var_{HB}(\widehat{\beta_1}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\left(\sum_{i=1}^{n} x_i x_i' \hat{e}_i^2\right)\mathbf{X}\,(\mathbf{X}'\mathbf{X})^{-1}$$

Here, x_i is the vector of independent variables for each observation, weighted by their residual variances.

This method has proven to be appropriate in handling OLS models with heteroscedasticity and resulting variances are unbiased and more accurate, especially in cross sectional or panel datasets.

6. SIMULATION

In order to evaluate performance of bootstrap confidence intervals when heteroscedasticity is present, it is necessary to carry out a simulation of a linear model. Simulation is necessary for two reasons. First, we need to know the true population coefficient β_1 and in practice we rarely know the true population parameter. Secondly, we need to evaluate performance of estimated confidence intervals in presence of heteroscedasticity. Although real data can have heteroscedasticity of residuals, we do not know the true form of residuals distribution. For these two reasons we need to model our linear model with heteroscedastic residuals. We select the simplest form of linear model with one explanatory variable that is correlated with the error term.

$$Y = \beta_0 + \beta_1 * X1 + \boldsymbol{\varepsilon}$$

where

$X1 \sim N(5, 4)$

$\boldsymbol{\varepsilon} \sim N(0, X1/2)$

where intercept (β_0) and β_1 are defined by us. Independent variables (X_1) come from normal distribution with mean of 5 and standard deviation of 4. Error term ($\boldsymbol{\varepsilon}$) is simulated following the approach suggested by Flachaire (2003) yet the form of heteroscedasticity is slightly different. Under this scenario, error term is correlated with explanatory variable and its variance grow as the value of $X1$ grows.

We check the performance of bootstrap confidence intervals in different sample sizes. Thus, we have a first sample size of 30 and then we increase it by 10 observations up to 200 observations. All of the simulations are carried out in R software.

We take the following steps for simulation of linear model with heteroscedasticity with different sample sizes

Step 1: set intercept $\beta_0 = 4$ and coefficient $\beta_1 = 5$

Step 2: Set sample size to n=30

Step 3: generate *X1 ~ N(5, 4)* starting with sample size n

Step 4: generate Y with $Y = \beta_0 + \beta_1 * X1 + \boldsymbol{\varepsilon}$, where $\boldsymbol{\varepsilon} \sim N(0, \frac{X1}{2})$

Step 5: estimate confidence intervals using traditional method without remedy, Huber-White robust intervals and bootstrap methods in repeated simulations (1000 times). Here we construction 95 percent confidence intervals

Step 6: evaluate how many times (out of 1000), true parameters were within estimated OLS intervals with and without remedy and bootstrap confidence intervals

Step 7: repeat step 2 to step 8 by adding 10 observations to sample size (n=n+10). Finish when sample size reaches 200 observations

Traditional and bootstrap confidence intervals estimations are discussed in above sections. For traditional intervals, we use the following formula which is estimated in any statistical package when we construct our linear model.

$$\widehat{\beta_1} \pm t_{\frac{\alpha}{2}} * se\left(\widehat{\beta_1}\right)$$

Bootstrap confidence intervals are built taking values in certain percentiles of parameter distributions that were generated as a result of bootstrapping.

7. RESULTS

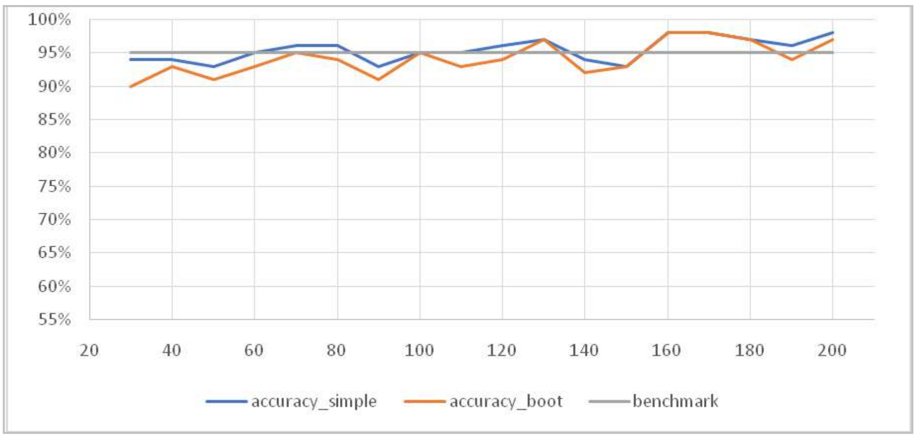
We will look into two results of the simulation in this part. First is with homoscedastic residuals and second is with the presence of heteroscedasticity. In present of heteroscedasticity, we compare traditional confidence intervals without remedy and bootstrap intervals. Afterwards, we build confidence intervals using Huber-White standard errors and check how they perform against bootstrap. In all these comparison, we also look at how estimated intervals change as we change our sample size.

Correctly specified model

First of all, we want to see how traditional CLT based and bootstrap confidence intervals perform when no violations of OLS assumptions are present. We expect that both approaches will do relatively good work in building interval estimates. In other words, for 95 percent confidence intervals, we expect true parameters to fall within estimated intervals at least 95 per cent of cases.

The first graph below shows often true coefficients fall within estimated confidence intervals built using traditional and bootstrap methods. One can see that both methods are doing relatively well, that is constructed intervals are containing true coefficient at least The chart clearly shows that both traditional and bootstrap confidence intervals contain true parameter in 90-100 percent of the cases which is expected outcomes (see Figure 3).

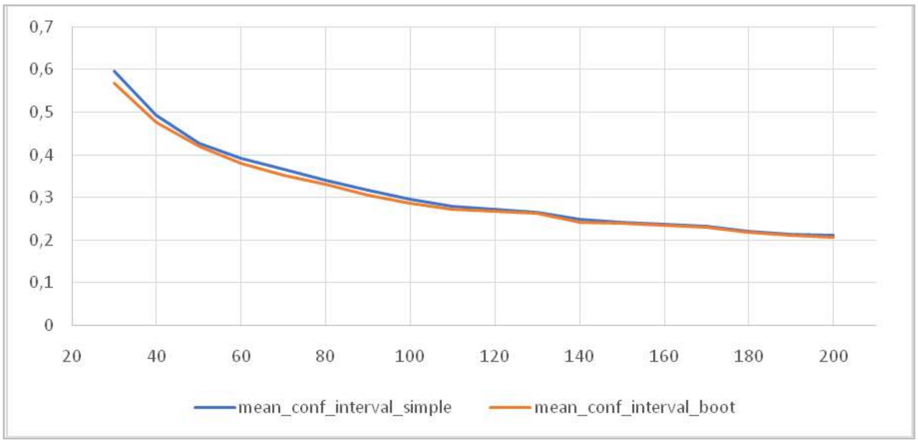
Figure 3: Accuracy of confidence intervals when model is correctly specified: traditional confidence intervals vs bootstrap confidence intervals



Source: generated by the author

Bootstrap confidence intervals contain true coefficients more often compared to traditional OLS intervals. This is explained in the second graph which shows that bootstrap intervals are larger in width compared to OLS intervals across all sample sizes (see Figures 4).

Figure 4: Size or width of confidence intervals when model is correctly specified: traditional confidence intervals vs bootstrap confidence intervals

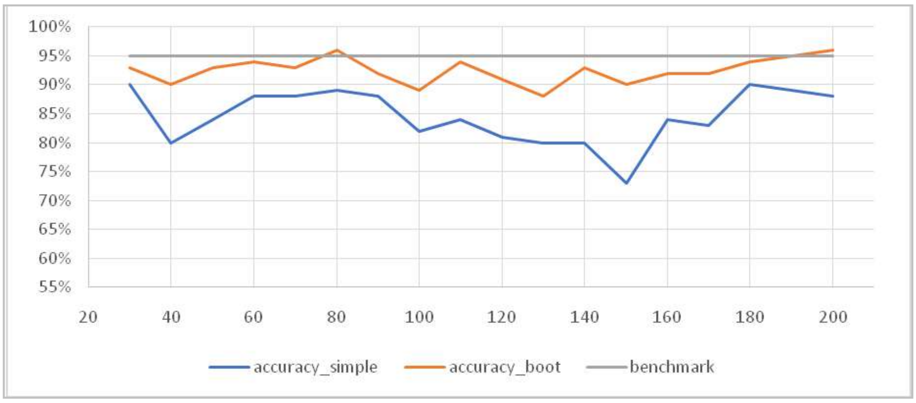


Source: generated by the author

Misspecified model: case of heteroscedastic residuals

As explained in previous chapter, we introduce heteroscedasticity by making our variance of residuals equal to $X\mathbf{1}$ */2* This will make variance of error term be dependent of values of $X\mathbf{1}$ and thus make the error term heteroscedastic. In other words, the larger the explanatory variable, the larger the variance of the error term becomes. Here, we again plot the two graphs to check accuracy and width of traditional OLS confidence intervals compared to bootstrap ones.

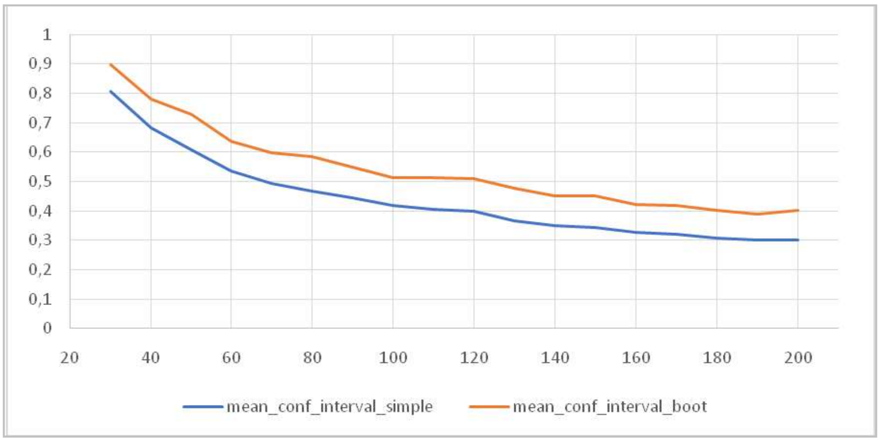
Figure 5: Accuracy of confidence intervals when residuals of the model are heteroscedastic and not remedy is applied: traditional confidence interval without remedy vs bootstrap confidence interval



Source: generated by the author

One can clearly see from the Figure 5 in Appendix that neither of the approaches are reaching expected 95 per cent coverage of confidence intervals. Yet, accuracy of bootstrap confidence intervals are much higher than that of traditional intervals. To put in other words, approximately 90 per cent or more bootstrap confidence intervals contain true population coefficient across different sample sizes. In contrast, traditional confidence intervals' accuracy are below 80 per cent which clearly indicates that confidence intervals are highly influenced by non-constant variance of residuals which distorts standard deviation of estimated coefficient. Higher coverage of bootstrap intervals are explained by the fact that the width of bootstrap intervals are wider compared to traditional ones (see Figure 6).

Figure 6: Size or width of confidence intervals in presence of heteroscedasticity: traditional confidence interval without remedy vs bootstrap confidence interval

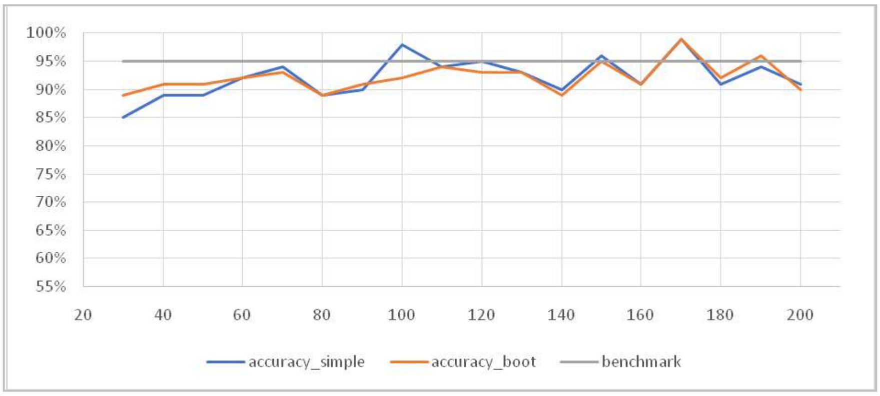


Source: generated by the author

As pointed out in the literature review part, Chernick and LaBudde (2011) claim that bootstrap intervals constructed using bootstrapping pairs are less sensitive to violations of model assumptions, which is also justified in the current simulation.

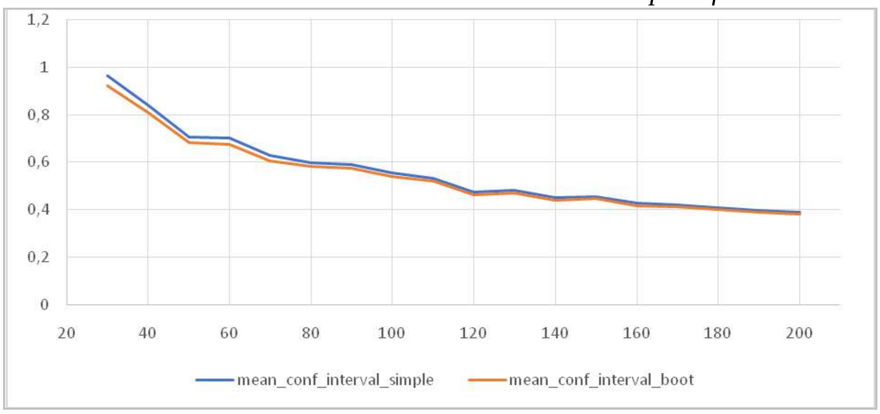
Now we will have a look at how bootstrap confidence intervals contrast with traditional intervals when we apply Huber-White method. In our specific form of variance of the error term (X1/2), Huber-White robust intervals perform almost the same as bootstrap confidence intervals in term of accuracy (see Figure 7). Size of confidence intervals of both methods are also almost the same, although bootstrap intervals are slightly narrower compared to Huber-White intervals.

Figure 7: Accuracy of confidence intervals when residuals of the model are heteroscedastic and remedy is applied: traditional confidence interval with Huber-White robust standard errors vs bootstrap confidence interval



Source: generated by the author

Figure 8: Size or width of confidence intervals in presence of heteroscedasticity: traditional confidence interval with Huber-White robust standard errors vs bootstrap confidence interval



Source: generated by the author

To sum up all the above-mentioned simulation results, traditional confidence intervals can result in misleading inferences when residuals have non-constant variance and no remedy is applied. In such cases, researcher apply different transformation, such as Huber-White method, in order to make our coefficient estimates robust to heteroscedasticity. Here, we suggested and tested bootstrap confidence intervals as another approach of handling heteroscedasticity arguing that bootstrapping does not have any distributional assumptions. Simulations discussed above indicate that bootstrap confidence intervals can work very well to build our intervals estimations and make inferences. Bootstrapping has accuracy at expected 95 per cent and performed as good as Huber-White robust intervals, while bootstrap interval had narrower intervals.

8. LIMITATIONS OF THE STUDY

Although our paper sheds some light in new ways of handling heteroscedasticity in OLS models using bootstrap confidence intervals, there are some limitations in this method. Firstly, aligned with many papers in bootstrapping, this method is computer intensive and resampling of large dataset can be costly. In our simulation study, we also required some computing power and time to get results of the bootstrap confidence intervals while Huber-White methods was delivering faster and less computer intensive outcomes. Thus, readers are recommended to apply Huber-White method in cases when it helps produce more accurate and reliable estimates or when sample data is relatively large. Secondly, in this simulation study, we considered only one type of heteroscedasticity when variance of the error term is equal to $\frac{X1}{2}$. Researchers are highly encouraged to consider other forms of heteroscedasticity and compare bootstrap confidence intervals with widely known remedies such as Huber-White robust method. Lastly, we selected Huber-White method for comparison and it has resulted in similar outcomes as bootstrap method. Areas for further research could be to compare bootstrap intervals with other methods of handling heteroscedastic residuals in linear models.

CONCLUSION

In this paper, we carried out a simulation study of building bootstrap confidence intervals in linear models when variance of residuals is not constant. We first looked at existing literature on this topic and then looked at the theoretical side of linear models with heteroscedasticity. We explained that traditional confidence intervals might be biased when heteroscedasticity is present in data and therefore suggested using bootstrapping pairs for building confidence intervals, which do not have any assumptions of residual distribution. Afterwards, we compared performance bootstrap intervals with one the widely used techniques to handle heteroscedasticity, Huber-White method.

When model is correctly specified and residuals are homoscedastic, both traditional confidence intervals and bootstrap intervals have intervals with good accuracy and almost the same size which is aligned with the theory. Given that bootstrap is computer intensive, traditional confidence intervals are more preferable when all assumptions of OLS are met. Yet, when heteroscedasticity is introduced and not remedy is applied, bootstrap confidence intervals are showing superb performance with accuracy around 95 per cent which traditional intervals are highly inaccurate. Lastly, we compared bootstrap confidence intervals with one of the widely known techniques to get heteroscedasticity robust intervals known as Huber-White approach. Results indicate that the both approaches are promising and very similar in accuracy and size or width of the intervals. This bring us to conclude that bootstrap approach performs equally well as widely known Huber-White methods and researchers can use both approaches interchangeably.

Yet, it is important to be note the limitations of this study. First, one should be aware that bootstrap could be computer intensive, especially when dataset is large. Secondly, we investigated only one type of heteroscedasticity while many other forms of non-constant variance of residuals are present in practice. Lastly, we picked Huber-White approach for comparison and did not consider other remedies that researchers are encouraged to look in future studies.

REFERENCES

Cameron, A. Colin, and Pravin K. Trivedi. 2005. *Microeconometrics: Methods and Applications*. Cambridge: Cambridge University Press.

Chernick, Michael R., and Robert A. LaBudde. 2014. *An Introduction to Bootstrap Methods with Applications to R*. Hoboken, NJ: John Wiley & Sons.

Chernozhukov, Victor, and Han Hong. 2003. "An MCMC Approach to Classical Estimation." *Journal of Econometrics* 115 (2): 293–346.

Davidson, Russell, and James G. MacKinnon. 2004. *Econometric Theory and Methods*. New York: Oxford University Press, 2004.

Davison, Anthony C., and David V. Hinkley. 1997. *Bootstrap Methods and Their Applications*. Cambridge: Cambridge University Press.

DiCiccio, Thomas, and Bradley Efron. 1992. "More Accurate Confidence Intervals in Exponential Families." *Biometrika* 79: 231–45.

Efron, Bradley. 1982. *The Jackknife, the Bootstrap and Other Resampling Plans*. Philadelphia: SIAM.

Efron, Bradley, and Robert Tibshirani. 1986. "Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy." *Statistical Science* 1: 54–77.

Efron, Bradley. 1979. "Bootstrap Methods: Another Look at the Jackknife." *The Annals of Statistics* 7(1): 1–26.

Fan, Yanqin, and Qi Li. 2004. "A Consistent Model Specification Test Based on the Kernel Density Estimation." *Econometrica* 72 (6): 1845–58.

Flachaire, Emmanuel. 2005. "Bootstrapping Heteroscedastic Regression Models: Wild Bootstrap vs Pairs Bootstrap." *Computational Statistics and Data Analysis* 49 (2): 361–76.

Freedman, David A. 1981. "Bootstrapping Regression Models." *The Annals of Statistics* 9: 1218–22.

Greene, William H. 2021. *Econometric Analysis*. 8th ed. Hoboken, NJ: Pearson.

Gujarati, Damodar N., Dawn C. Porter, and Sangeetha Gunasekar. 2012. *Basic Econometrics*. New York: McGraw-Hill Higher Education.

Horowitz, Joel L., and Marianthi Markatou. 1996. "Semiparametric Estimation of Regression Models for Panel Data." *Review of Economic Studies* 63(1): 145–68.

Huber, Peter J. 1967. "The Behavior of Maximum Likelihood Estimates Under Nonstandard Conditions." In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1:221–33. Berkeley, CA: University of California Press.

James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2023. *An Introduction to Statistical Learning*. 2nd ed. New York: Springer.

Lind, Douglas A., William G. Marchal, and Samuel A. Wathen. 1967. *Statistical Techniques in Business and Economics*. 2nd ed. New York: McGraw-Hill.

Liu, Runze Y. 1988. "Bootstrap Procedures Under Some Non-i.i.d. Models." *The Annals of Statistics* 16: 1696–1708.

Long, J. Scott, and Laurie H. Ervin. 2000. "Using Heteroscedasticity-Consistent Standard Errors in the Linear Regression Model." *The American Statistician* 54(3): 217–24.

MacKinnon, James G., and Halbert White. 1985. "Some Heteroskedasticity-Consistent Covariance Matrix Estimators with Improved Finite Sample Properties." *Journal of Econometrics* 29(3): 305–25.

Politis, Dimitris, and Joseph Romano. 1994. "The Stationary Bootstrap." *Journal of the American Statistical Association* 89 (428): 1303–12.

White, Halbert. 1980. "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity." *Econometrica* 48 (4): 817–38.

Wu, C. F. Jeff. 1986. "Jackknife, Bootstrap, and Other Resampling Plans in Regression Analysis (with Discussion)." *The Annals of Statistics* 14: 1261–1350.

This journal is open access - The journal's content is licensed under Creative Commons Attribution - NonCommercial - NoDerivs 4.0.