Earthline Journal of Mathematical Sciences
E-ISSN: 2581-8147
Volume 14, Number 6, 2024, Pages 1213-1228
https://doi.org/10.34198/ejms.14624.12131228

Improved Two Stage Least Square Estimation with Permutation Methods for Solving Endogeneity Problems

C. O. Ikpere[1] and C. O. Aronu[2,*]

[1] Department of Statistics, Delta State Polytechnic, Otefe, Oghara, Delta State, Nigeria

[2] Department of Statistics, Chukwuemeka Odumegwu Ojukwu University, Uli, Anambra State, Nigeria
e-mail: amaro4baya@yahoo.com

Abstract

This study introduces permutation methods for Two Stage Least Square (2SLS) estimation to improve accuracy and address endogeneity in econometric analysis. The research compares traditional 2SLS with newly developed permutation methods, specifically evaluating their efficiency, predictive power, and precision across various datasets. Instrumental variables were used in a two-step process, applying permutation techniques to both the instrumental variable and the estimated endogenous variable. Results indicate that the classic 2SLS method is more efficient and precise in estimating intercepts, showing the least standard deviation in exact p-values (0.2396). However, the Permute $\hat{\boldsymbol{X}}$ method outperformed in estimating endogenous variables, demonstrating higher accuracy and stability with a standard deviation of 0.1121. The study highlights that while traditional 2SLS is reliable for intercept estimation, permutation methods, particularly the Permute $\hat{\boldsymbol{X}}$ method, offer significant improvements in handling endogeneity for endogenous variables. These findings suggest that permutation methods could enhance econometric analyses by providing alternative strategies for more accurate and robust estimates. The study's contribution lies in advancing econometric methodologies by integrating permutation methods into 2SLS estimation, providing empirical evidence of their effectiveness. The research emphasizes the importance of robustness checks and sensitivity analyses in econometric studies, advocating for a tailored approach to selecting estimation methods based on specific data characteristics and research objectives. Future research should further refine these methods and explore their application in diverse econometric contexts.

Received: August 13, 2024; Accepted: September 22, 2024; Published: October 4, 2024

2020 Mathematics Subject Classification: 62J05, 62P20.

Keywords and phrases: permutation methods, two stage least square, endogeneity, instrumental variables, estimation accuracy, efficiency.

*Corresponding author

1. Introduction

The Two-Stage Least Squares (2SLS) method is a widely used econometric technique for estimating parameters in multi-equation models, particularly when errors between equations are uncorrelated. It is especially effective for over-identified or exactly identified equations, operating in a stepwise manner rather than solving all equations simultaneously like the Three-Stage Least Squares method. The TSLS approach addresses endogeneity issues and has significantly advanced econometric modelling and statistical inference, benefiting from technological progress and addressing complex econometric challenges. Recent studies, including those by López-Espín et al. [7], Wilms et al. [10], and Sheikhi et al. [9], have refined the TSLS method by incorporating QR-decomposition and tackling issues like multicollinearity and omitted variable bias. Concurrently, the literature emphasizes the role of permutation techniques in enhancing statistical inference. This study builds on these advancements by integrating permutation approaches into the TSLS framework, aiming to improve the robustness and reliability of TSLS estimations, thereby contributing to econometric modelling both theoretically and practically.

Exploring Enhancements and Applications of Two-Stage Least Squares (TSLS) and Permutation Tests in Econometrics. The two-stage least squares (TSLS) method is a widely used econometric tool for addressing endogeneity in regression models, particularly when dealing with instrumental variables. Over the years, researchers have explored various enhancements to the TSLS approach to improve its efficiency, robustness, and applicability across different contexts. Suppose we look at the advancements and the role of permutation tests as a non-parametric alternative for econometric analysis. López-Espín et al. [7] introduced a method to improve the computational efficiency of TSLS by utilizing QR-decomposition. This approach optimizes the matrix structures within TSLS and leverages multicore platforms for parallel processing, which significantly enhances computing performance. The method is particularly valuable when dealing with large datasets or complex models, where traditional TSLS computations may be cumbersome. Sheikhi et al. [9] developed a bridge estimator to handle issues of sparsity and multicollinearity in TSLS models. This estimator improves the performance of TSLS in scenarios where endogeneity and multicollinearity are present, offering a more robust alternative to conventional techniques. The bridge estimator provides a solution that maintains the integrity of the estimates while effectively managing the challenges posed by sparse data. Wilms et al.

[10] emphasized the dangers of omitted variable bias, especially in psychological studies. They advocated for the use of TSLS and regression discontinuity design (RDD) as tools to mitigate this bias. The authors demonstrated how TSLS, when properly applied, can provide more accurate estimates by addressing endogeneity that arises from omitted variables, thus enhancing the reliability of the study's findings. Joshi and Wooldridge [11] explored the application of TSLS in unbalanced panel data models with instrumental variables. They developed robust Hausman tests for comparing different TSLS estimators, thereby improving panel data analysis. Their work is particularly relevant for studies where the data is unbalanced, and traditional panel data techniques may not be sufficient. Lu et al. [12] applied TSLS to investigate the determinants of foreign direct investment (FDI) inflows in the Commonwealth of Independent States (CIS) countries. By addressing endogeneity through TSLS, they underscored the importance of geographical and economic factors in FDI decisions. Similarly, Khan et al. [13] utilized TSLS to explore the relationships between financial development, economic growth, energy consumption, and carbon emissions, supporting the Environmental Kuznets Curve. Their work highlighted the utility of TSLS in macroeconomic research, particularly when dealing with complex interdependencies among variables. Shin et al. [14] studied the effect of after-school exercise on academic achievement among Korean students. They found that TSLS, unlike ordinary least squares (OLS), provided robust evidence of a positive correlation between exercise and academic performance. This study exemplifies the strength of TSLS in overcoming limitations associated with direct empirical approaches, particularly in educational research.

Moreover, permutation tests have gained prominence as a flexible and reliable non-parametric statistical technique in econometrics. They offer an alternative to traditional parametric tests, especially when the assumptions of normality or homoscedasticity are violated. Phipson and Smyth [15] defined permutation tests as randomization tests that estimate the null distribution of a test statistic by permuting the class labels of the data. The appeal of permutation tests lies in their minimal assumptions—they only require that the observations be independent and uniformly distributed under the null hypothesis. Fisher [16] introduced the concept, and Dwass [17] extended it by suggesting that calculations could be restricted to a randomly selected subset of permutations. Anderson and Legendre [18] proposed a permutation test for partial regression coefficients in linear models. Their study showed that different permutation approaches yielded varied results, emphasizing the importance of selecting an appropriate method based on the specific characteristics of the data. Ortner et al. [19] examined the use of permutation tests for detecting structural changes in time series data. They pointed out that traditional methods, such as those relying on F-statistics, often assume independent and normally distributed residuals. In contrast, permutation tests offer a more reliable alternative, especially when these assumptions do not hold. Pesarin and Salmaso [8] argued that the non-parametric combination approach is effective for multivariate testing problems that are difficult to address using parametric methods. This approach is particularly advantageous when dealing with non-normal or categorical variables, where traditional parametric tests may not be applicable. Neuhaus [20] proposed a permutation test based on a studentized statistic to address the issue of controlling type 1 error in cases where the underlying distribution is not identical under the null hypothesis. Janssen [6] further demonstrated that permutation tests could provide asymptotically valid inferences even when population distributions have different variances.

Based on the review, it was found that the TSLS approach in econometrics has been enhanced through various methodological advancements, including QR decomposition, bridge estimators, and robust panel data techniques. These developments have broadened the applicability of TSLS across different fields, from macroeconomic studies to educational research. On the other hand, permutation tests offer a non-parametric alternative that is flexible and robust, particularly when traditional parametric assumptions are not met. Together, these techniques provide powerful tools for econometric analysis, enabling researchers to tackle complex problems with greater precision and reliability.

Hence, the study focuses on enhancing the Two-Stage Least Squares (2-SLS) estimation method, widely used in econometrics for parameter estimation in multi-equation models, particularly when dealing with over-identified or accurately identified equations. Despite its significant contributions to statistical inference and econometric modelling, there is still potential for improvement. Recent research has highlighted shifts in the application and methodology of 2-SLS, with a growing interest in using permutation methods to bolster statistical inference. The study aims to develop permutation techniques to improve the precision, efficiency, and predictive power of the 2-SLS estimator, particularly in addressing endogeneity issues. By integrating these techniques, the research seeks to enhance the robustness, credibility, and empirical relevance of 2-SLS across various economic scenarios, contributing to both theoretical frameworks and real-world applications in econometric modelling. The study objectives include comparative analysis, efficiency improvement, and performance evaluation of these permutation methods.

2. Methods

2.1. Source of data collection for the study

The study uses secondary data sourced from published journals and the Central Bank of Nigeria (CBN) statistical bulletin. The datasets include: Finance Dataset: Contains daily returns on selected stocks, spanning January 5, 1993, to January 30, 2009. It includes market factors like risk-free rate and market portfolio returns. Value of Stocks Dataset: Quarterly data from 1960 to 1977, covering stock values, monetary base, and production capacity. Cigar Dataset: Panel data on cigarette consumption in the U.S. from 1963 to 1992, detailing price, population, and income. Gasoline Consumption Dataset: Panel data from 1960 to 1978, including gasoline consumption and income. These datasets are valuable for analyzing financial, economic, and consumption patterns.

2.2. The proposed two stage least square estimation with permutation

In the theory of econometrics, instrumental variables (IVs) are usually employed in a two-step approach to estimate coefficients when endogeneity is present. IVs are variables that have no correlation with the error term (u) but a correlation with the endogenous predictor (X).

Suppose a multi-equation econometric model is described by the system of its structural equation:

$$YA + XB + U = 0, \tag{1}$$

where

$Y =$ is an $n \times m$ data matrix of m endogenous variables in n observations;

$X =$ is an $n \times k$ data matrix of k exogenous variables in n observation;

$A =$ is an $m \times m$ full rank of unknown parameters or coefficients associated with Y;

$B =$ is a $k \times k$ matrix of unknown parameter or coefficients associated with X;

$U =$ is an $n \times m$ matrix of (unobserved) error.

The structural parameters are the components of A and B. Because U is known to be frequently associated with Y, which is stochastic, the Gauss-Markov conditions for the application of the Ordinary Least Squares (OLS) are violated, making it impossible to estimate the parameters in the columns of A and B using the OLS. The system of equations (3.1) is first converted into the reduced form equations rather than being used directly with the OLS. The reduced form equations describe Y in terms of X only. We can post-multiply the system of equations (1) by A^{-1} to obtain:

$$YAA^{-1} + XBA^{-1} + UA^{-1} = 0,$$

$$Y = XP + E, \tag{2}$$

where

$$P = -BA^{-1},$$

$$E = -\,UA^{-1}.$$

The system of reduced form equations (2) may be estimated using the OLS since X is constant (non-stochastic) and is unable to be associated with E. Consequently, the OLS estimates P, the matrix of the reduced form coefficients is estimated as:

$$\hat{P} = [X_{,}X]^{-1} X_{,}Y. \tag{3}$$

We can then obtain:

$$\hat{Y} = X\hat{P}. \tag{4}$$

Then in each equation where any endogenous variable $Y_j \subset Y$ appears as an explanatory variables, Y_j is replaced by $\hat{Y}_j$. As a result of this substitution, the explanatory variables in the corresponding equation are no longer stochastic or associated with the error term, making the equation suitable for OLS estimation. The estimations of the parameters in this converted equation may be easily obtained by applying the OLS (again) to it.

2.2.1. The permute the instrumental variable Z in the 2SLS estimator

Suppose we consider permuting the instrumental variable Z in the first step of the 2SLS method and then regress the endogenous parameter on its instruments and on the remaining exogenous variables in the model.

Stage 1: First-Stage Regression

In this stage, the IV is permuted and the permuted variable Z^p is used to estimate the relationship between the endogenous variable X and the permuted instrumental variable (Z^p). The equation for the first stage is:

$$X = \pi_0 + \pi_1 Z^p + \varepsilon, \tag{5}$$

where X is the endogenous variable, Z^p is the permuted instrumental variable, ε represents the error term in the first stage, and π_0 and π_1 are the coefficients to be estimated.

In the second step, the fitted values from the estimated endogenous regressor are plugged into the original regression and regressed.

Stage 2: Second-Stage Regression

In the second stage, we use the predicted values of X from the first stage (denoted as $\hat{X}$) in the main regression model to estimate the coefficients of interest. The equation for the second stage is:

$$Y = \beta_0 + \beta_1 \hat{X} + u, \tag{6}$$

where Y is the dependent variable, $\hat{X}$ is the predicted value of X from the first stage based on permutation of the instrumental variable, u is the error term in the main regression, and β_0 and β_1 are the coefficients of interest.

The main reasons for permuting the instrumental variable Z include possible endogeneity, bias from missing variables, or issues with weak instruments in the conventional 2SLS (Two-Stage Least Squares) estimate. The goal of the study is to investigate the instrumental variable approach's resilience and sensitivity to various instrument configurations or structures by permuting Z. Below are specific justifications for this permutation strategy:

i. Robustness Check: When compared to the conventional 2SLS estimation, permuting Z acts as a robustness check. We can determine if the results are sensitive to changes in the composition or properties of the instrumental variable and examine the stability of the estimates by adding permutations.

ii. Addressing Weak Instrument Concerns: By permuting the instrumental variable, we may investigate whether different specifications or combinations of Z result in more accurate and efficient estimations in situations where Z might be a weak instrument. Through experimenting with various combinations, we might be able to find Z setups that alleviate weak instrument issues.

iii. Enhancing Instrumental Variable Relevance: By permuting Z, we may assess the instrumental variable's exogeneity and relevance in various permutations. We may ascertain whether particular permutations improve the validity and relevance of the instrument in resolving endogeneity problems by evaluating how changes in Z affect the first and second-stage estimations.

2.2.2. The permute the $\hat{X}$ in the 2SLS Estimator

Suppose we consider permuting the predicted variable of X ($\hat{\boldsymbol{X}}$) in the second step of the 2SLS method and then regress the endogenous parameter and on the remaining exogenous variables in the model.

Stage 1: First-Stage Regression

In this stage, the IV is not permuted remains unchanged Z and is used to estimate the relationship between the endogenous variable X. The equation for the first stage is:

$$X = \pi_0 + \pi_1 Z + \varepsilon, \tag{7}$$

where, X is the endogenous variable, Z is the instrumental variable, ε represents the error term in the first stage, and π_0 and π_1 are the coefficients to be estimated.

However, in the second step, the fitted values from the estimated endogenous regressor are plugged into the original regression and regressed.

Stage 2: Second-Stage Regression

In the second stage, permute the predicted values of X from the first stage (denoted as $\hat{X}^p$) and use in the main regression model to estimate the coefficients of interest. The equation for the second stage is:

$$Y = \beta_0 + \beta_1 \hat{X}^p + u, \tag{8}$$

where, Y is the dependent variable, $\hat{X}^p$ is the permuted predicted value of X from the first stage, u is the error term in the main regression, and β_0 and β_1 are the coefficients of interest.

Given equation (8), to estimate the vector of unknown parameters β's, the least squares estimator for β, can be obtained by minimizing the sum of squared errors (SSE), which is given by:

$$SSE = (y - \hat{X}^p \beta)^\tau (y - \hat{X}^p \beta). \tag{9}$$

To find the β that minimizes SSE, take the derivative of SSE with respect to β and set it equal to zero:

$$\frac{\partial SSE}{\partial \beta} = -2\left(\hat{X}^{p}\right)^{\tau}\left(y - \hat{X}^{p}\beta\right) = 0.$$

Solving for β:

$$= \left(\hat{X}^{p}\right)^{\tau}\left(y - \hat{X}^{p}\beta\right) = 0.$$

Distribute $\left(\hat{X}^{p}\right)^{\tau}$ into the parentheses:

$$= \left(\hat{X}^{p}\right)^{\tau} y - \left(\hat{X}^{p}\right)^{\tau}\hat{X}^{p}\beta = 0.$$

Rearrange the equation:

$$\left(\hat{X}^{p}\right)^{\tau} y = \left(\left(\hat{X}^{p}\right)^{\tau}\hat{X}^{p}\right)\beta.$$

To isolate β, multiply both sides by the inverse of $\left(\hat{X}\prime\hat{X}\right)$ assuming that $\left(\hat{X}\prime\hat{X}\right)$ is invertible:

$$\left(\hat{X}^{p}\right)^{\tau} y\left(\left(\hat{X}^{p}\right)^{\tau}\hat{X}^{p}\right)^{-1} = \beta.$$

So, the least squares estimator for β is:

$$\hat{\beta} = \left(\left(\hat{X}^{p}\right)^{\tau}\hat{X}^{p}\right)^{-1}\left(\hat{X}^{p}\right)^{\tau} y. \tag{10}$$

The equation presented as equation (10) gives the least squares estimates for the parameters β in the model based on the permutation of the predicted variable X in the second stage.

The decision to permute the predicted variable $\hat{\boldsymbol{X}}$ in the 2SLS estimator arises from a desire to explore the robustness, sensitivity, and potential biases associated with the instrumental variable approach, particularly in addressing endogeneity issues. Below are specific justifications for this permutation strategy:

i. Robustness Evaluation: Permuting $\hat{\boldsymbol{X}}$ serves as a robustness check to assess the stability and consistency of the 2SLS estimator's results. By introducing variations in $\hat{\boldsymbol{X}}$, we can evaluate whether the estimated coefficients and relationships remain consistent across different configurations or specifications of the predicted variable, providing insights into the reliability of the instrumental variable approach.

ii. Addressing Model Misspecification: In scenarios where the model may suffer from misspecification or omitted variable bias, permuting $\hat{\boldsymbol{X}}$ allows us to examine how changes in the predicted variable impact the estimation results. By exploring alternative configurations or structures of $\hat{\boldsymbol{X}}$, we can identify potential biases, outliers, or influential observations that may affect the validity and reliability of the 2SLS estimator.

iii. Exploratory Analysis: Permuting $\hat{\boldsymbol{X}}$ provides an exploratory framework to analyze the sensitivity of the instrumental variable approach to variations in the predicted variable. By systematically varying $\hat{\boldsymbol{X}}$, we can identify patterns, trends, or specific configurations that influence the estimation results, thereby enhancing our understanding of the underlying relationships and dynamics captured by the model.

Enhancing Model Efficiency: In some instances, permuting $\hat{\boldsymbol{X}}$ may improve the efficiency, precision, or predictive power of the 2SLS estimator by exploring alternative specifications or transformations of the predicted variable. By assessing how changes in $\hat{\boldsymbol{X}}$ impact the estimation outcomes, we can potentially identify configurations that enhance the model's performance and robustness against endogeneity concerns.

This study evaluates the performance of proposed permutation methods for the Two-Stage Least Squares (2-SLS) Estimator against the traditional 2-SLS based on absolute bias, efficiency, and variance. Absolute bias measures the robustness of an estimator, assessing its deviation from the true parameter value, with a robust estimator showing minimal bias even with outliers or assumption violations. Efficiency, another key metric, compares how well two estimators utilize data to minimize variance and mean squared error (MSE). An estimator with a lower variance is deemed more efficient. The study uses these metrics to determine the effectiveness of permutation methods in improving the precision and reliability of 2-SLS estimations in econometric analysis

3. Results

Assessing the performance of the classic 2SLS and the proposed Permute Z and Permute $\hat{\boldsymbol{X}}$ for the intercept and endogenous variable.

Table 1: Exact p-value and standard deviation of exact p-value for the classic 2SLS and the proposed Permute Z and Permute $\hat{\boldsymbol{X}}$ for the intercept

Name of Data	Dimension	2SLS	Permute Z	Permute $\hat{\boldsymbol{X}}$
Finance	4012 × 24	0.699	0.135	0
value of stocks	71 × 5	0	0.801	1
Cigar	1380 × 9	0	0.793	0.5
Gasoline	342 × 6	0.3542	0.522	0.5
Parity	1768 × 9	0	0.814	1
Grunfeld	220 × 5	0	0.794	1
CigarettesB	46 × 3	0	0.704	0.929
PSID1982	595 × 12	0.2867	0.02	0.614
Wage	3000 × 11	0	0.938	1
Export_Import	30 × 4	0	0.805	1
Standard Deviation		0.239611	0.31202	0.34162

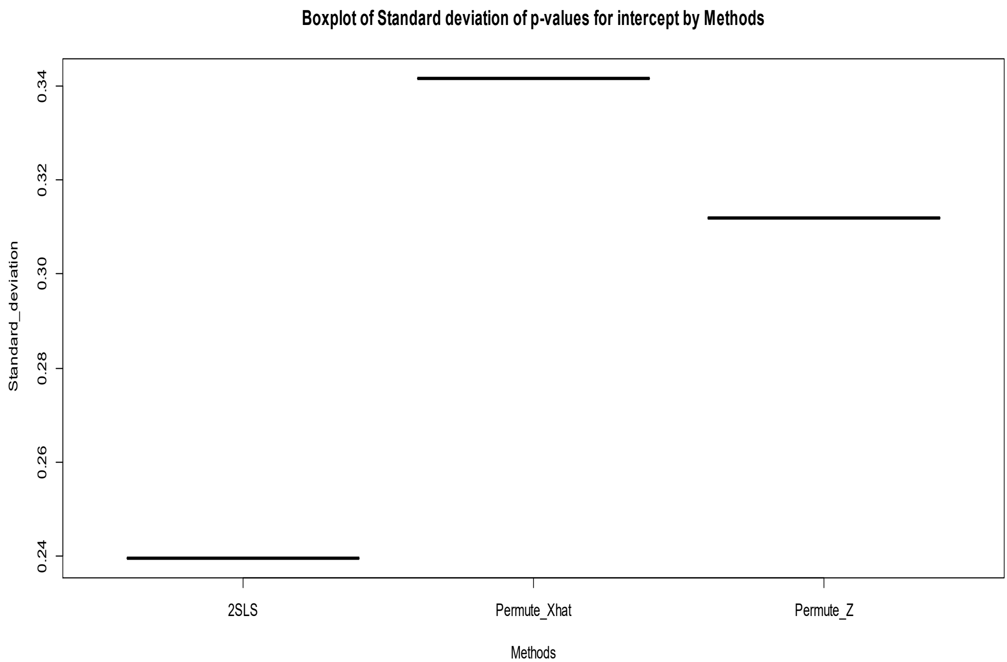

Figure 1: The box plot of the standard deviation of exact p-value for the classic 2SLS and the proposed Permute Z and Permute $\hat{\boldsymbol{X}}$ for the intercept.

The result presented in Table 1 and Figure 1 shows the distribution of the standard deviation of the exact p-value for the classic 2SLS and the proposed Permute Z and Permute $\hat{\boldsymbol{X}}$ for the intercept. The 2SLS method was found to be the most efficient, precise, and potentially more reliable among the three methods for the estimation of the intercept since it recorded the least standard deviation of the exact p-value across the 10 different datasets with varied dimensions with $2SLS = 0.2396$, Permute $Z = 0.3120$, and Permute $\hat{\boldsymbol{X}} = 0.3416$. Hence, these findings suggest that the classical 2SLS method is more accurate and stable in estimating the intercept.

Table 2: Exact p-value and standard deviation of exact p-value for the classic 2SLS and the proposed Permute Z and Permute $\hat{\boldsymbol{X}}$ for the endogenous variable

Name of Data	**Dimension**	**2SLS**	**Permute $\boldsymbol{Z}$**	**Permute $\hat{\boldsymbol{X}}$**
Finance	4012 × 24	0.473	0.661	0.207
value of stocks	71 × 5	0	0.34	0
Cigar	1380 × 9	0	0.423	0
Gasoline	342 × 6	0	0.327	0.309
Parity	1768 × 9	0	0.297	0
Grunfeld	220 × 5	0	0.335	0
CigarettesB	46 × 3	0.2624	0.501	0.151
PSID1982	595 × 12	0.2858	0.001	0.114
Wage	3000 × 11	0	0.152	0
Export_Import	30 × 4	0	0.325	0
Standard Deviation		0.1732	0.1789	0.1121

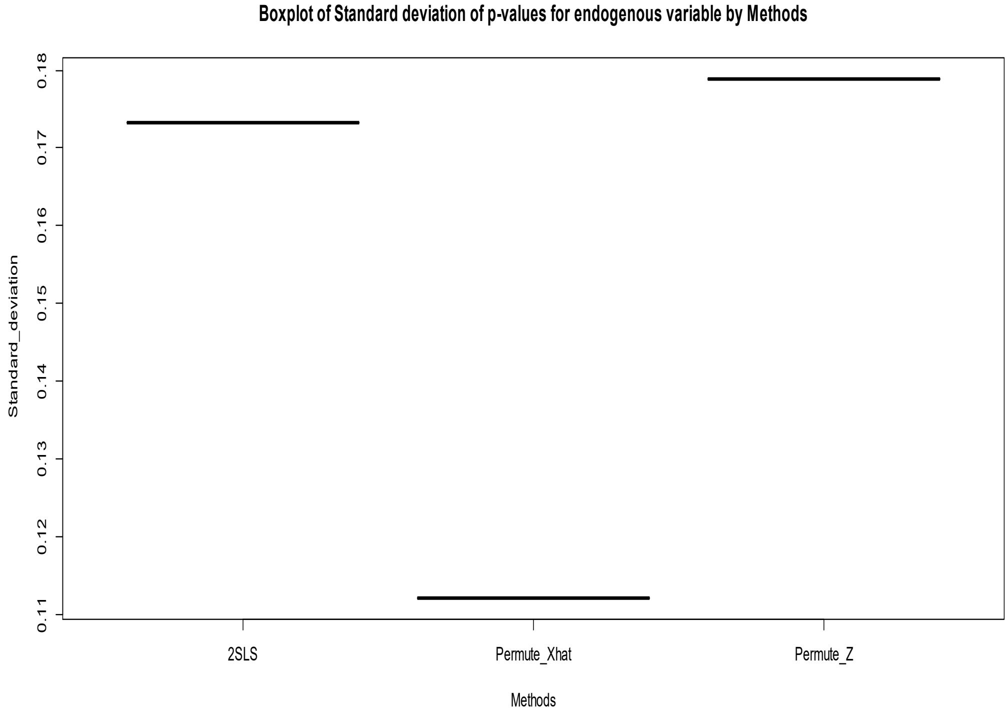

Figure 2: The box plot of the standard deviation of exact p-value for the classic 2SLS and the proposed Permute Z and Permute $\hat{\boldsymbol{X}}$ for the endogenous variable.

The result presented in Table 2 and Figure 2 shows the distribution of the standard deviation of the exact p-value for the classic 2SLS and the proposed Permute Z and Permute $\hat{\boldsymbol{X}}$ for the intercept. The Permute $\hat{\boldsymbol{X}}$ method was found to be the most efficient, precise, and potentially more reliable among the three methods for the estimation of the endogenous variable since it recorded the least standard deviation of the exact p-value across the 10 different datasets with varied dimensions with Permute $\hat{\boldsymbol{X}} = 0.1121$, $2SLS = 0.1732$, and Permute $Z = 0.1732$. Hence, these findings suggest that the Permute $\hat{\boldsymbol{X}}$ is more accurate and stable in estimating the endogenous variable.

4. Conclusion

This study developed permutation methods for Two Stage Least Square (2SLS) Estimation and evaluated their effectiveness in addressing endogeneity issues in econometric analysis. Through a comprehensive analysis of various datasets and comparison with traditional 2SLS methods, several key findings emerged by expressing the efficacy and reliability of permutation methods in econometric estimation. It was found that between permutation methods and traditional 2SLS there exist nuanced differences in efficiency, precision, and reliability across different estimation scenarios. While the permutation methods showed promise in estimating endogenous variables, the classic 2SLS method was able to demonstrate higher accuracy and stability in estimating intercepts.

The evaluation of the permutation methods indicated their potential to improve the efficiency and predictive power of 2SLS estimators, particularly in addressing weak instrument concerns and robustness checks. Hence, by permuting instrumental variables or predicted variables in 2SLS estimation, insights into the sensitivity of estimation results to changes in instrument configurations and identifying more accurate and reliable estimation strategies are gained. Hence, the outcome of the study was able to contribute to advancing knowledge of permutation techniques for 2SLS estimation and emphasizes the significance of choosing the right estimation techniques by the particular goals and features of the data. It also emphasizes the necessity of sensitivity studies and robustness checks to guarantee the accuracy and consistency of estimation results in econometric research.

References

[1] Anderson, M. J., & Legendre, P. (1999). An empirical comparison of permutation methods for tests of partial regression coefficients in linear models. *Journal of Statistical Computation and Simulation*, *62*(3), 271-303. https://doi.org/10.1080/00949659908811936

[2] Aronu, C. O., & Nworuh, G. E. (2019a). Permutation methods for Chow test analysis: An alternative for detecting structural break in linear models. *Advances in Science, Technology and Engineering Systems Journal*, *4*(4), 12-20. https://doi.org/10.25046/aj040402

[3] Aronu, C. O., & Nworuh, G. E. (2019b). Determining the relatively efficient test statistic measure between the traditional Chow test and the Milek permutation test for detecting structural break in linear models. *International Journal of Latest Technology in Engineering, Management & Applied Science (IJLTEMAS)*, *8*(6), 25-28.

[4] Blum, J. R., Kiefer, J., & Rosenblatt, M. (1961). Distribution free tests of independence based on the sample distribution function. *The Annals of Mathematical Statistics*, *32*(2), 485-498. https://doi.org/10.1214/aoms/1177705055

[5] Fisher, R. A. (1935). *The design of experiments* (3rd ed.). Oliver & Boyd.

[6] Janssen, A. (1997). Studentized permutation tests for non-i.i.d. hypotheses and the generalized Behrens-Fisher problem. *Statistics & Probability Letters*, *36*(1), 9-21. https://doi.org/10.1016/S0167-7152(97)00043-6

[7] López-Espín, J. J., Navas-Palomino, F. J., & Martínez, V. (2012). A parallel implementation of the QR-decomposition for the two-stage least squares method. *Computer Physics Communications*, *183*(11), 2332-2342.

[8] Pesarin, F., & Salmaso, L. (2006). *Permutation tests for complex data: Theory, applications, and software*. Wiley.

[9] Sheikhi, A., Bahador, F., & Arashi, M. (2022). On a generalization of the test of endogeneity in a two stage least squares estimation. *Journal of Applied Statistics*, *49*(3), 709-721. https://doi.org/10.1080/02664763.2020.1837084

[10] Wilms, I., Croux, C., & Van den Poel, D. (2021). Correcting for omitted variable bias in regression models with a regression discontinuity design. *Journal of Econometrics*, *222*(2), 539-556.

[11] Joshi, R., & Wooldridge, J. M. (2019). Correlated random effects models with endogenous explanatory variables and unbalanced panels. *Annals of Economics and Statistics*, *134*, 243-268. https://doi.org/10.15609/annaeconstat2009.134.0243

[12] Lu, W., Kasimov, I., Karimov, I., & Abdullaev, Y. (2019). Foreign direct investment, natural resources, economic freedom, and sea-access: Evidence from the commonwealth of independent states. *Sustainability*, *12*(8), 3135. https://doi.org/10.3390/su12083135

[13] Khan, S., Peng, Z., & Li, Y. (2019). Energy consumption, environmental degradation, economic growth and financial development in globe: Dynamic simultaneous equations panel analysis. *Energy Reports*, *5*, 1089-1102. https://doi.org/10.1016/j.egyr.2019.08.004

[14] Shin, K., You, S., & Kim, M. (2021). A comparison of two-stage least squares (TSLS) and ordinary least squares (OLS) in estimating the structural relationship between after-school exercise and academic performance. *Mathematics*, *9*(23), 3105. https://doi.org/10.3390/math9233105

[15] Phipson, B., and Smyth, G. K. (2010). Permutation p-values should never be zero: calculating exact p-values when permutations are randomly drawn. *Statistical Applications in Genetics and Molecular Biology*, *9*(1), Article 39, 1-12. https://doi.org/10.2202/1544-6115.1585

[16] Fisher, R. A. (1935). *The design of experiments* (3rd ed.). Oliver & Boyd, London.

[17] Dwass, M. (1957). Modified randomization tests for nonparametric hypotheses. *Ann. Math. Statist.*, *28*, 181-187. https://doi.org/10.1214/aoms/1177707045

[18] Legendre, P. (2000). Comparison of permutation methods for the partial correlation and partial mantel tests. *J. Statist. Comput. Simulation*, *67*, 37-73. https://doi.org/10.1080/00949650008812035

[19] Ortner, T., Filzmoser, P., & Endel, G. (2015). Identifying structural changes in Austrian social insurance data. *IFAC-PapersOnLine*, *48*(1), 115-120. https://doi.org/10.1016/j.ifacol.2015.05.152

[20] Neuhaus, G. (1993). Conditional rank tests for the two-sample problem under random censorship. *Annals of Statistics*, *21*, 1760-1779. https://doi.org/10.1214/aos/1176349396