International Journal of Development Mathematics Vol 3 Issue 2 Page No. 362 - 379

INTERNATIONAL JOURNAL OF DEVELOPMENT MATHEMATICS

ISSN: 3026-8656 (Print) | 3026-8699 (Online)

journal homepage: https://ijdm.org.ng/index.php/Journals

Evaluating Efficiency of Hybrid Estimators in Exact and Over – Identified Simultaneous Equations

Alademomi Aladesuyi[a,*], Olatayo O. Alabi[b] and Abimbola H. Bello[b]

[a]Department of Statistics, The Federal Polytechnic, Ado Ekiti, Ekiti State, Nigeria

[b]Department of Statistics, Federal University of Technology, Akure, Ondo State, Nigeria

A R T I C L E I N F O

Article history:
Received 02 March 2026
Received in revised form 15 May 2026
Accepted 27 May 2026

Keywords:
Multicollinearity, Principal Component Regression, Simultaneous Equation Models, Monte Carlo Simulation, Identification Problem

MSC 2020 Subject classification:
62J05, 62H25

A B S T R A C T

Multicollinearity poses a significant threat to the reliability of classical estimators in Simultaneous Equation Models (SEMs). This problem leads to inflated variances and unstable coefficient estimates. Conventional SEM estimators such as Two-Stage Least Squares (2SLS), Three–Stage Least Squares (3SLS), Full Information Maximum Likelihood (FIML), Indirect Least Squares (ILS) and Limited Information Maximum Likelihood (LIML) do not inherently address collinearity among regressors. This study develops and evaluates hybrid estimators that integrate Principal Component Analysis (PCA) with classical SEM estimators (PCR-2SLS, PCR-3SLS, PCR-FIML, PCR-ILS, PCR-LIML) to mitigate multicollinearity. A Monte Carlo simulation is conducted under varying sample sizes and multicollinearity levels. Performance is assessed using the Mean Squared Error (MSE) criterion. Results demonstrate that for over–identified equations, the FIML-PCR estimator consistently outperforms all competitors across all sample sizes and collinearity levels. For exact identified equations, 3SLS-PCR and FIML-PCR are generally preferred. The study concludes that hybrid PCR–SEM estimators, particularly FIML-PCR, offer substantial improvements in estimation efficiency under multicollinearity and are recommended for applied econometric modelling.

1. Introduction

Multicollinearity defined as the presence of high linear interdependencies among predictor variables, remains a pervasive challenge in econometric modelling particularly within the framework of Simultaneous Equation Models (SEMs). Unlike single-equation regression models, SEMs consist of a system of interdependent structural equations where endogenous variables appear on both sides of the equations, making identification, estimation and inference considerably more complex (Schmidt, 2005; Alabi, 2016a). In such systems, the simultaneous determination of variables amplifies the adverse effects of multicollinearity, which include inflated standard errors, unstable coefficient estimates, and reduced predictive precision. Classical SEM estimators, such as Two–Stage Least Squares (2SLS), Three-Stage Least Squares (3SLS), Full Information Maximum Likelihood (FIML), Indirect Least Squares (ILS) and Limited Information Maximum Likelihood (LIML), assume well-conditioned design matrices and independent

* Corresponding author. Tel.: +2347082876059
E-mail address: aaladesuyi@gmail.com (Aladsuyi, A.)
https://doi.org/10.62054/ijdm/0302.23

instruments. However, when multicollinearity is severe, these estimators produce unreliable inferences and inconsistent parameter estimates (Olubusoye, 2001; Johnson *et al.* 2010; Alabi and Oyejola, 2015; Alabi, 2016b).

The problem of multicollinearity in SEMs has been examined by several authors, yet its systematic mitigation remains underexplored. In linear regression, ridge regression and Principal Component Regression (PCR) are well-established techniques for reducing collinearity (Fayose and Ayinde, 2019; Aladesuyi *et al.* 2025). PCR transforms correlated regressors into orthogonal components, effectively eliminating linear dependencies by discarding components associated with small eigenvalues (Alabi *et al.* 2025). This approach has been extensively validated in single–equation contexts but has not been fully adapted to the simultaneous equation framework. The application of PCR to SEMs is not straightforward because the structural system involves endogenous variables, instrumental variables and cross-equation error correlations. Conventional SEM estimators do not automatically account for multicollinearity in the structural system; they are designed primarily to address endogeneity not multicollinearity. Consequently, parameter estimates may remain unbiased but inefficient, with hypothesis tests becoming unreliable due to inflated standard errors (Alabi, 2016a; Alabi, 2016b; Cankaya and Eker, 2025).

Recent advances have attempted to address these limitations through Bayesian extensions (Garnier-Villarreal and Jorgensen, 2024) and penalized Generalized Method of Moments (GMM) approaches, but these methods focus largely on robustness to heteroscedasticity and non–normality rather than explicit multicollinearity correction (Okeke *et al.* 2025). A notable gap exists in the literature regarding hybrid estimators that integrate PCR logic with SEM estimators. Specifically, no comprehensive Monte Carlo study has systematically compared the performance of PCR-augmented versions of 2SLS, 3SLS, FIML, ILS and LIML under controlled multicollinearity conditions. This gap is critical and will be addressed in this study.

Standard 2SLS estimates may become erratic and FIML may fail to converge. Therefore, there is an urgent need for estimators that retain the consistency properties of SEM estimators while gaining efficiency through collinearity reduction.

This paper develops and evaluates PCR–augmented versions of five classical SEM estimators: PCR–2SLS, PCR–3SLS, PCR–FIML, PCR–ILS and PCR–LIML. The hybrid approach proceeds by first applying PCR to the matrix of all exogenous and instrumental variables to obtain orthogonal principal components. These components, rather than the original correlated regressors are then used as instruments or directly in the structural equations. The number of components retained is determined based on eigenvalue thresholds, typically retaining components with eigenvalues greater than one or cumulative variance above 80%. This transformation ensures that the estimation procedure operates in a space free of linear dependencies.

A comprehensive Monte Carlo simulation was designed to assess the finite-sample performance of these hybrid estimators. The data generation process follows established procedures (Ayinde, 2007; Alabi, 2019), with a four–equation linear SEM featuring both exact identified and over–identified structural equations.

2. Literature Review

Multicollinearity arises when two or more predictor variables in a regression model are highly linearly related. In simultaneous equation models, this problem is exacerbated because endogenous variables are functions of multiple exogenous factors, leading to near–linear dependencies among instruments and regressors (Fayose *et al.* 2023a; Fayose *et al.* 2023b). The consequences of multicollinearity include: (i) inflated standard errors of coefficient estimates, (ii) reduced statistical power, (iii) unstable estimates sensitive to small changes in data, (iv) difficulty in identifying individual variable effects, and (v) poor out–of–sample predictive performance. Contrary to common belief, multicollinearity does not introduce bias in ordinary least squares estimates, but it renders them imprecise and unreliable for hypothesis testing (Leamer, 1973).

In SEMs, detection is more complex due to the presence of both structural and reduced-form equations. Researchers often examine the condition number of the instrument matrix; values above 30 suggest severe multicollinearity (Judge *et al.* 1985).

2.1 Principal Component Analysis and Principal Component Regression

Principal Component Analysis (PCR) is a dimensionality reduction technique that transforms a set of correlated variables into a smaller set of uncorrelated linear combinations called principal components. These components are ordered such that the first component captures the maximum variance, the second captures the next maximum orthogonal variance and so on (De Jong and Kiers, 1992). When applied to regression, PCR involves regressing the output variable on a subset of the principal components typically discarding those associated with small eigenvalues. This discarding process reduces multicollinearity because the retained components are orthogonal by construction. Studies have shown that PCR can substantially reduce MSE compared to ordinary least squares under collinearity, although it introduces some bias (Cankaya and Eker, 2025). However, the application of PCR to SEMs is not merely a direct extension; care must be taken to preserve the structural interpretation of parameters and to avoid discarding components that, while explaining little variance in the regressors, may be highly correlated with the endogenous variables.

2.2 Classical Simultaneous Equation Estimators

The literature on SEM estimation is extensive. Two–Stage Least Squares (2SLS) is the workhorse estimator for over-identified equations. It first regresses each endogenous regressor on all exogenous variables (instruments) and then uses the predicted values in the structural equation. While consistent, 2SLS is not fully efficient under general error covariance structures (Schmidt, 2005). Three–Stage Least Squares (3SLS) improves upon 2SLS by incorporating cross–equation error correlations through a generalized least squares adjustment, making it more efficient when errors are contemporaneously correlated. Full Information Maximum Likelihood (FIML) estimates all equations simultaneously under multivariate normality, achieving asymptotic efficiency but at the cost of higher computational complexity and sensitivity to specification errors (Alabi, 2019). Limited Information Maximum Likelihood (LIML) offers a compromise by estimating single equations with limited system information and it has been shown to have better finite–sample properties than 2SLS under weak instruments. Indirect Least Squares (ILS) is applicable only to exact identified equations, deriving structural parameters from reduced–form coefficients.

2.3 Hybrid and Penalized Approaches in SEMs

The integration of regularization techniques with SEMs is a nascent but growing area. Ridge–based SEM estimators have been proposed, where a ridge parameter is added to the instrument cross–product matrix to stabilize estimates (Alabi *et al.* 2025). However, ridge regression does not orthogonalize the regressors; it merely shrinks coefficients. In contrast, PCR orthogonalization directly addresses the source of collinearity. A few studies have explored PCR in the context of instrumental variables, but systematic evaluations in SEMs are lacking. Okeke *et al.* (2025) examined penalized GMM approaches for robust estimation but did not focus specifically on multicollinearity. Aladesuyi *et al.* (2025) assessed the role of significant roots in linear regression under multicollinearity, highlighting the importance of eigenvalue–based methods. No comprehensive simulation study has compared PCR–augmented versions of 2SLS, 3SLS, FIML, ILS and LIML under controlled conditions. This study fills that gap by providing a rigorous Monte Carlo evaluation of five hybrid PCR–SEM estimators across varying sample sizes and multicollinearity levels by using MSE as the primary criterion.

3. Methods

Following Aladesuyi *et al.* (2026), we consider a linear simultaneous equation system of the form:

$$\begin{aligned}
y_{i1} &= \beta_{12}y_{i2} + \beta_{13}y_{i3} + \beta_{14}y_{i4} + \gamma_{14}x_{i4} + u_{i1} && \text{(i)} \\
y_{i2} &= \beta_{23}y_{i3} + \gamma_{21}x_{i1} + \gamma_{23}x_{i3} + u_{i2} && \text{(ii)} \qquad (1) \\
y_{i3} &= \beta_{31}y_{i1} + \beta_{34}y_{i4} + \gamma_{31}x_{i1} + \gamma_{32}x_{i2} + u_{i3} && \text{(iii)} \\
y_{i4} &= \beta_{41}y_{i1} + \beta_{42}y_{i2} + \gamma_{42}x_{i2} + u_{i4} && \text{(iv)}
\end{aligned}$$

where y_{i1} is an output variable, i = 1,2,3,4, x_{i1}, x_{i2}, x_{i3} and x_{i4} are the input variables

$$\begin{bmatrix} u_{i1} \\ u_{i2} \\ u_{i3} \\ u_{i4} \end{bmatrix} \sim N(0,\Sigma), where\ i = 1,2,3,\ldots,n$$

and $\beta_{12}, \beta_{13}, \beta_{14}, \beta_{23}, \beta_{31}, \beta_{34}, \beta_{41}, \beta_{42}, \gamma_{14}, \gamma_{21}, \gamma_{23}, \gamma_{31}, \gamma_{32}$ and γ_{42} are the structural parameters of the model. Equations (i) and (iii) are exact identified while equations (ii) and (iv) are over identified by both order and rank conditions.

$$y_{i1} - \beta_{12}y_{i2} - \beta_{13}y_{i3} - \beta_{14}y_{i4} + 0x_{i1} + 0x_{i2} + 0x_{i3} - \gamma_{14}x_{i4} = u_{i1} \qquad \text{(i)}$$

$$0y_{i1} + y_{i2} - \beta_{23}y_{i3} + 0y_{i4} - \gamma_{21}x_{i1} + 0x_{i2} - \gamma_{23}x_{i3} + 0x_{i4} = u_{i2} \qquad \text{(ii)} \qquad (2)$$

$$-\beta_{31}y_{i1} + 0y_{i2} + y_{i3} - \beta_{34}y_{i4} - \gamma_{31}x_{i1} - \gamma_{32}x_{i2} + 0x_{i3} + 0x_{i4} = u_{i3} \qquad \text{(iii)}$$

$$-\beta_{41}y_{i1} - \beta_{42}y_{i2} + 0y_{i3} + y_{i4} + 0x_{i1} - \gamma_{42}x_{i2} + 0x_{i3} + 0x_{i4} = u_{i4} \qquad \text{(iv)}$$

This can be written in matrix form as:

$$\beta y_i + \Gamma x_i = u_i \qquad (3)$$

where

$$\beta = \begin{bmatrix} 1 & -\beta_{12} & -\beta_{13} & -\beta_{14} \\ 0 & 1 & -\beta_{23} & 0 \\ -\beta_{31} & 0 & 1 & -\beta_{34} \\ -\beta_{41} & -\beta_{42} & 0 & 1 \end{bmatrix} \qquad \Gamma = \begin{bmatrix} 0 & 0 & 0 & -\gamma_{14} \\ -\gamma_{21} & 0 & -\gamma_{23} & 0 \\ -\gamma_{31} & -\gamma_{32} & 0 & 0 \\ 0 & -\gamma_{42} & 0 & 0 \end{bmatrix} \quad y_i = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix}$$

$$x_1 = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \text{ and } u_i = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix}$$

Now, from equation (3),

$$\beta^{-1}\beta y_i = \beta^{-1}\Gamma x_i = \beta^{-1}u_i \qquad (4)$$

3.1 Data generation of input variables

Data generation was carried out following Ayinde (2007), Alabi (2019) and Aladesuyi *et al.* (2026). The equations are given as:

$$x_1 = \mu_1 + \sigma_1 z_1$$

$$x_2 = \mu_2 + \rho_{12}\sigma_2 z_1 + \left(\sqrt{m_{22}}\right)z_2$$

$$x_3 = \mu_3 + \rho_{13}\sigma_3 z_1 + \frac{m_{23}}{\sqrt{m_{22}}} z_2 + \left(\sqrt{n_{33}}\right)z_3$$

$$x_4 = \mu_3 + \rho_{14}\sigma_4 z_1 + \frac{m_{24}}{\sqrt{m_{22}}} z_2 + \frac{n_{34}}{\sqrt{n_{33}}} z_3 + \sqrt{0_{44}}\, z_4$$

where $m_{22} = \sigma_2^2(1-\rho_{12}^2)$ $m_{23} = \sigma_2\sigma_3\left(\rho_{23} - \rho_{12}\rho_{13}\right)$ $m_{24} = \sigma_2\sigma_4\left(\rho_{24} - \rho_{12}\rho_{14}\right)$,

$$m_{33} = \sigma_3^2\left(1-\rho_{13}^2\right), \quad m_{44} = \sigma_4^2\left(1-\rho_{14}^2\right),\ 0_{44} = n_{44} - \frac{n_{34}^2}{n_{33}}, \quad n_{44} = m_{44} - \frac{m_{24}^2}{m_{22}},$$

$$n_{34} = m_{34} - \frac{m_{23}m_{24}}{m_{22}}, \quad n_{33} = m_{33} - \frac{m_{23}^2}{m_{22}}$$

And $Z_i \sim N(0,1)$, i = 1, 2, 3, 4 and $|\rho_{ij}| < 1$ is the value of correlation between the two variables i and j (Ayinde, 2007; Alabi, 2019).

The input variables were generated to be normally distributed with mean zero and unity variance. i.e. $x \sim N(0,1)$, i = 1,2,3,4 and $|\rho_{ij}| < 1$ is the value of correlation between the two variables i and j. Also in the study, $\rho_{12} = \rho_{13} = \rho_{14} = \rho_{23} = \rho_{24} = \rho_{34} = \rho =$ 0.3, 0.6, 0.9 and 0.99 was adopted to represent a situation with the presence multicollinearity between the input variables in the model.

3.2 Methods of generating output variables

Following Ayinde (2007), Alabi (2019), and Aladesuyi *et al.* (2026), equation (4) was used to generate the output variables by taking the true values of the parameters as:

$$\beta = \begin{bmatrix} 1 & -\beta_{12} & -\beta_{13} & -\beta_{14} \\ 0 & 1 & -\beta_{23} & 0 \\ -\beta_{31} & 0 & 1 & -\beta_{34} \\ -\beta_{41} & -\beta_{42} & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & -3.2 & -0.2 & -1.2 \\ 0 & 1 & -3.8 & 0 \\ -1.6 & 0 & 1 & -1.0 \\ -2.8 & -2.2 & 0 & 1 \end{bmatrix}$$

$$\Gamma = \begin{bmatrix} 0 & 0 & 0 & -\gamma_{14} \\ -\gamma_{21} & 0 & -\gamma_{23} & 0 \\ -\gamma_{31} & -\gamma_{32} & 0 & 0 \\ 0 & -\gamma_{42} & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & -0.8 \\ -3.0 & 0 & -1.3 & 0 \\ -1.5 & -0.6 & 0 & 0 \\ 0 & -2.6 & 0 & 0 \end{bmatrix}$$

3.3 Generation of correlated error term

The mean of the error terms is zero and unit variance. i.e. 1 as adopted by (Ayinde, 2007; Alabi, 2019). The correlation values are $\lambda_{12} = \lambda_{13} = \lambda_{14} = \lambda_{23} = \lambda_{24} = \lambda_{34} = \lambda$ and $|\lambda_{ij}|$ < 1 is the value of correlation between the two error terms i and j. In this study, $\lambda_{12} = \lambda_{13} = \lambda_{14} = \lambda_{23} = \lambda_{24} = \lambda_{34} = \lambda = 0$ was adopted to represent a situation in which no correlation exists between the error terms in the model.

3.4 PCR Simultaneous Equation Model Estimators

To mitigate the adverse effects of multicollinearity among the exogenous and instrumental variables, Principal Component Analysis (PCR) was integrated into each of the classical simultaneous equation estimators considered in this study. The objective of the transformation is to replace a highly correlated set of instruments with a smaller set of orthogonal principal components while preserving most of the information contained in the original variables.

Let $Z = (X_1, X_2, \dots, X_p)$ denote the $n \times p$matrix of standardized exogenous and instrumental variables. The variables were first standardized to have zero mean and unit variance. The sample correlation matrix of $Z$is given by $R = \frac{1}{n-1} Z'Z$.

The eigenvalue decomposition of $R$is $R = A\Lambda A'$, where $\Lambda = \operatorname{diag}(\lambda_1, \lambda_2, \dots, \lambda_p)$ contains the ordered eigenvalues $\left(\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p\right)$ and $A = (a_1, a_2, \dots, a_p)$ is the corresponding matrix of eigenvectors.

The principal component scores are then obtained as $T = ZA$, where $T = (T_1, T_2, \dots, T_p)$ represents a set of mutually orthogonal components. Because the components are uncorrelated by construction, $\operatorname{Cov}(T_i, T_j) = 0, i \neq j$, thus eliminating multicollinearity among the transformed regressors.

Component Retention Criterion

Only the dominant principal components were retained for estimation. Consistent with the procedure stated in the Introduction, component selection was based on either:

1. **Kaiser's criterion**, retaining components with eigenvalues greater than one: $\lambda_j > 1$,

   or

2. **Cumulative variance criterion**, retaining the smallest number $k$of components satisfying $\frac{\sum_{j=1}^{k} \lambda_j}{\sum_{j=1}^{p} \lambda_j} \geq 0.80$.

(5)

The retained component matrix is therefore $T_k = (T_1, T_2, \dots, T_k)$, where $k < p$.

The matrix T_k replaces the original exogenous and instrumental variable matrix in all subsequent estimation procedures.

Construction of PCR–2SLS

For the PCR–2SLS estimator, the retained principal components are used as instruments in the first stage. Consider the structural equation:

$$y = X\beta + u, \tag{6}$$

where Xcontains endogenous regressors.

In the first stage, each endogenous regressor is projected onto the retained principal component space:

$$X = T_k \Pi + V. \tag{7}$$

The fitted values are obtained as

$$\hat{X} = T_k(T_k'T_k)^{-1}T_k'X. \quad (8)$$

In the second stage, the structural parameters are estimated by ordinary least squares using the predicted endogenous variables:

$$\hat{\beta}_{PCR-2SLS} = (\hat{X}'\hat{X})^{-1}\hat{X}'y. \quad (9)$$

Because the instruments T_k are orthogonal, the resulting estimator is less sensitive to multicollinearity than conventional 2SLS.

Construction of PCR–3SLS

The PCR–3SLS estimator begins with the PCR-transformed instruments T_k and obtains fitted endogenous variables using the same first-stage procedure as PCR–2SLS. The system of equations is then estimated jointly using Generalized Least Squares (GLS) while accounting for contemporaneous correlations among equation disturbances. Thus, PCR–3SLS combines PCR-based collinearity reduction with the efficiency gains of system estimation.

Construction of PCR–FIML

For PCR–FIML, the retained principal components replace the original exogenous variables in the likelihood function. The full system is estimated simultaneously under the assumption of multivariate normal disturbances. The orthogonal nature of the retained components improves the conditioning of the information matrix and enhances numerical stability in the optimization process.

Construction of PCR–LIML

The PCR–LIML estimator applies the LIML procedure using T_k as the instrument matrix. The retained principal components provide a set of orthogonal instruments, reducing the instability often associated with severe multicollinearity and weak instrument configurations.

Construction of PCR–ILS

For exactly identified equations, the reduced-form parameters are first estimated using the retained principal components. The structural parameters are subsequently recovered through the usual indirect least-squares transformation. Replacing the original regressors with orthogonal components improves the precision of the reduced-form estimates from which the structural coefficients are derived.

Adjustment for Endogeneity

The PCR transformation is applied only to the matrix of exogenous variables and instruments and not to the endogenous variables themselves. Consequently, the orthogonalization process addresses multicollinearity without altering the identification structure of the simultaneous equation system. Endogeneity is handled through the underlying SEM estimator (2SLS, 3SLS, FIML, LIML or ILS), while PCR serves exclusively as a preprocessing step that produces a well-conditioned instrument matrix. Therefore, the hybrid estimators retain the consistency properties of the original SEM estimators while improving efficiency through variance reduction.

The resulting hybrid estimators are denoted as PCR–2SLS, PCR–3SLS, PCR–FIML, PCR–ILS and PCR–LIML, respectively. These estimators were subsequently evaluated under varying sample sizes and levels of multicollinearity using the Mean Squared Error (MSE) criterion (Judge, *et al.* 1985; De Jong and Kiers, 1992; Cankaya and Eker, 2025).

3.5 Criteria for assessment

Using Monte Carlo simulations, we evaluate the characteristics of estimators at various stages of multicollinearity, sample sizes and instrument strength. Mean Squared Error (MSE) serves as the performance criterion. The MSE is defined as:

$$MSE(\hat{\beta}_{ij}) = \frac{1}{R}\sum_{i=1}^{R}(\hat{\beta}_{ij} - \beta_{ij})^2, \text{ i = 1, 2, ……, R, j = 1, 2, ….., where R = 1000.} \qquad (10)$$

In the study, $\hat{\beta}_{ij}$ is ith element of the estimator β in the jth replication which gives the estimate of β_i. β_i are the true value of the parameter previously mentioned. The mean square error values were obtained for each of the parameters of the equation of the estimation methods in the entire sample sizes at all the levels of correlation values for normally distributed input variables via R programming language.

Summary/Procedures to identify the best estimators under MSE criterion

i. At each level of multicollinearity, error variance and lambda, the MSE produced by each estimator were ranked.

ii. The number of occasions on which each estimator recorded the least MSE was evaluated under varying levels of multicollinearity, error variance and lambda values.

iii. An estimator is best if it has highest number of counts.

4. Results

In this section, the performances of the estimators were compared in the absence and presence of multicollinearity respectively. The simulation results are presented in tabular-form for PCR–SEM.

Table 1: Performance of the estimators based on principal component under the MSE criterion when there is no multicollinearity in the model

N	Estimator	Exact Identified			Over – Identified		
		Eq1	Eq3	Total	Eq2	Eq4	Total
10	PCR-2SLS	1	0	1	0	2	2
	PCR-3SLS	4	6	10	0	0	0
	PCR-FIML	**7**	**4**	**11**	**11**	**8**	**19**
	PCR-ILS	0	1	1	0	1	1
	PCR-LIML	0	1	1	1	1	2

20	PCR-2SLS	0	0	0	1	1	2
	PCR-3SLS	7	3	10	0	0	0
	PCR-FIML	**5**	**9**	**14**	**10**	**9**	**19**
	PCR-ILS	0	0	0	1	2	3
	PCR-LIML	0	0	0	0	0	0
30	PCR-2SLS	0	0	0	1	1	2
	PCR-3SLS	4	5	9	1	0	1
	PCR-FIML	**8**	**7**	**15**	**9**	**9**	**18**
	PCR-ILS	0	0	0	0	1	1
	PCR-LIML	0	0	0	1	1	2
50	PCR-2SLS	1	1	2	0	2	2
	PCR-3SLS	**4**	**10**	**14**	1	0	1
	PCR-FIML	7	0	7	**8**	**8**	**16**
	PCR-ILS	0	1	1	3	2	5
	PCR-LIML	0	0	0	0	0	0
100	PCR-2SLS	0	0	0	1	1	2
	PCR-3SLS	3	3	6	0	0	0
	PCR-FIML	**9**	**9**	**18**	**10**	**8**	**18**
	PCR-ILS	0	0	0	0	1	1
	PCR-LIML	0	0	0	1	2	3
200	PCR-2SLS	0	0	0	0	1	1
	PCR-3SLS	**8**	**7**	**15**	0	0	0
	PCR-FIML	4	5	9	**12**	**9**	**21**
	PCR-ILS	0	0	0	0	1	1
	PCR-LIML	0	0	0	0	1	1
300	PCR-2SLS	1	1	2	0	2	2
	PCR-3SLS	4	5	9	0	0	0
	PCR-FIML	**6**	**6**	**12**	**11**	**8**	**19**
	PCR-ILS	0	0	0	1	1	2
	PCR-LIML	1	0	1	0	1	1

Source: Computed from simulated results

Note: From Table 1, at each sample size level, the summed ranks were further added over the equations and preferred estimators under both exact and over identified equations were bolded. Table 2 gives the summary of the preferred estimators.

Table 2: Summary of the preferred estimators under MSE criterion at different multicollinearity level (ρ) in the model

Sample Sizes	Exact Identified	Over Identified
10	PCR-FIML	PCR-FIML
20	PCR-FIML	PCR-FIML
30	PCR-FIML	PCR-FIML
50	PCR-3SLS	PCR-FIML
100	PCR-FIML	PCR-FIML
200	PCR-3SLS	PCR-FIML
300	PCR-FIML	PCR-FIML

Source: Table 1

Note: From Table 2, the following can be observed. For the exact identified equation, PCR-FIML is the preferred estimator at most sample sizes, except for a few instances where PCR-3SLS performs better. For over identified equations, PCR-FIML is the preferred estimator in all sample sizes.

Table 3: Performance of the estimators based on principal component under the MSE criterion at different multicollinearity level (ρ) in the model

N	Estimator	Exact Identified			Over – Identified		
		Eq1	Eq3	Total	Eq2	Eq4	Total
10	PCR-2SLS	6	1	7	4	12	16
	PCR-3SLS	**27**	**25**	**52**	4	3	7
	PCR-FIML	18	31	49	**38**	**39**	**77**
	PCR-ILS	5	1	6	8	2	10
	PCR-LIML	4	2	6	6	4	10
20	PCR-2SLS	2	2	4	8	6	14
	PCR-3SLS	**23**	**31**	**54**	5	1	6
	PCR-FIML	**28**	**26**	**54**	**38**	**39**	**77**
	PCR-ILS	2	1	3	5	9	14
	PCR-LIML	5	0	5	4	5	9
30	PCR-2SLS	6	0	6	4	7	11
	PCR-3SLS	24	28	52	1	3	4
	PCR-FIML	**22**	**31**	**53**	**41**	**39**	**80**
	PCR-ILS	5	0	5	8	8	16
	PCR-LIML	3	1	4	6	3	9
50	PCR-2SLS	5	2	7	5	2	7
	PCR-3SLS	**38**	**31**	**69**	4	3	7
	PCR-FIML	10	24	34	**41**	**40**	**81**
	PCR-ILS	4	2	6	4	6	10
	PCR-LIML	3	1	4	6	9	15
100	PCR-2SLS	4	2	6	4	6	10
	PCR-3SLS	**33**	**29**	**62**	5	5	10
	PCR-FIML	15	27	42	**37**	**36**	**73**
	PCR-ILS	3	1	4	10	6	16
	PCR-LIML	5	1	6	4	7	11
200	PCR-2SLS	4	0	4	6	5	11
	PCR-3SLS	16	20	36	2	3	5
	PCR-FIML	**30**	**40**	**70**	**43**	**37**	**80**
	PCR-ILS	6	0	6	2	9	11
	PCR-LIML	4	0	4	7	6	13
300	PCR-2SLS	6	1	7	3	5	8
	PCR-3SLS	17	28	45	2	5	7
	PCR-FIML	**26**	**31**	**57**	**42**	**37**	**79**
	PCR-ILS	4	0	4	2	7	9
	PCR-LIML	7	0	7	11	6	17

Source: Computed from simulated results

Note: From Table 3, at each sample size level, the summed ranks were further added over the equations and preferred estimators under both exact and over identified equations were bolded. Table 4 gives the summary of the preferred estimators.

Table 4: Summary of the preferred estimators under MSE criterion at different multicollinearity level (ρ) in the model

Sample Sizes	Exact Identified	Over Identified
10	PCR-3SLS	PCR-FIML

20	PCR-3SLS / PCR-FIML	PCR-FIML
30	PCR-FIML	PCR-FIML
50	PCR-3SLS	PCR-FIML
100	PCR-3SLS	PCR-FIML
200	PCR-FIML	PCR-FIML
300	PCR-FIML	PCR-FIML

Source: Table 3

Note: From Table 4, the following are observed about the preferred estimators under Mean Square Error criterion. For the exact identified equations, PCR-3SLS or PCR-FIML or both estimators are generally preferred. For over identified equations, PCR-FIML is preferred estimator at all different sample size

Table 5: Performance of the estimators based on principal component under the MSE criterion at different multicollinearity levels (ρ) in the model

ρ	n	Estimator	Exact Identified			Over – Identified		
			Eq1	Eq3	Total	Eq2	Eq4	Total
0.8	10	PCR-2SLS	2	1	3	1	2	3
		PCR-3SLS	**4**	**5**	**9**	1	0	1
		PCR-FIML	**5**	**4**	**9**	**6**	**9**	**15**
		PCR-ILS	1	1	2	3	0	3
		PCR-LIML	0	1	1	1	1	2
	20	PCR-2SLS	0	1	1	2	2	4
		PCR-3SLS	**6**	**6**	**12**	1	0	1
		PCR-FIML	5	5	10	**6**	**9**	**15**
		PCR-ILS	1	0	1	2	1	3
		PCR-LIML	0	0	0	1	0	1
	30	PCR-2SLS	2	0	2	1	0	1
		PCR-3SLS	**5**	**7**	**12**	0	0	0
		PCR-FIML	5	5	10	**9**	**9**	**18**
		PCR-ILS	0	0	0	2	2	4
		PCR-LIML	0	0	0	0	1	1
	50	PCR-2SLS	0	1	1	2	1	3
		PCR-3SLS	**8**	**6**	**14**	0	0	0
		PCR-FIML	3	5	8	**10**	**9**	**19**
		PCR-ILS	1	0	1	0	2	2
		PCR-LIML	0	0	0	0	0	0
	100	PCR-2SLS	0	2	2	1	0	1
		PCR-3SLS	**7**	**6**	**13**	2	0	2
		PCR-FIML	5	3	8	**6**	**10**	**16**
		PCR-ILS	0	1	1	3	1	4
		PCR-LIML	0	0	0	0	1	1
	200	PCR-2SLS	2	0	2	2	0	2
		PCR-3SLS	4	6	10	1	0	1
		PCR-FIML	**5**	**6**	**11**	**7**	**8**	**15**
		PCR-ILS	1	0	1	1	2	3
		PCR-LIML	0	0	0	1	2	3
	300	PCR-2SLS	2	0	2	0	1	1
		PCR-3SLS	4	1	5	2	0	2
		PCR-FIML	**4**	**11**	**15**	**8**	**8**	**16**
		PCR-ILS	1	0	1	0	2	2

		PCR-LIML	1	0	1	2	1	3
0.9	10	PCR-2SLS	1	0	1	1	0	1
		PCR-3SLS	**6**	**8**	**14**	1	0	1
		PCR-FIML	4	3	7	**7**	**9**	**16**
		PCR-ILS	0	0	0	1	1	2
		PCR-LIML	1	1	2	2	2	4
	20	PCR-2SLS	0	1	1	1	1	2
		PCR-3SLS	**5**	**6**	**11**	0	0	0
		PCR-FIML	**6**	**5**	**11**	**9**	**8**	**17**
		PCR-ILS	0	0	0	1	2	3
		PCR-LIML	1	0	1	1	1	2
	30	PCR-2SLS	1	0	1	0	1	1
		PCR-3SLS	2	3	5	0	0	0
		PCR-FIML	**7**	**8**	**15**	**10**	**8**	**18**
		PCR-ILS	2	0	2	0	2	2
		PCR-LIML	0	1	1	2	1	3
	50	PCR-2SLS	1	0	1	1	0	1
		PCR-3SLS	**7**	**5**	**12**	0	0	0
		PCR-FIML	2	4	6	**10**	**10**	**20**
		PCR-ILS	2	2	4	1	1	2
		PCR-LIML	0	1	1	0	1	1
	100	PCR-2SLS	2	0	2	0	0	0
		PCR-3SLS	**6**	**5**	**11**	0	0	0
		PCR-FIML	3	6	9	**9**	**8**	**17**
		PCR-ILS	0	0	0	2	2	4
		PCR-LIML	1	1	2	1	2	3
	200	PCR-2SLS	0	0	0	0	0	0
		PCR-3SLS	4	4	8	1	0	1
		PCR-FIML	**6**	**8**	**14**	**8**	**9**	**17**
		PCR-ILS	1	0	1	1	2	3
		PCR-LIML	1	0	1	2	1	3
	300	PCR-2SLS	1	0	1	0	0	0
		PCR-3SLS	6	1	7	0	0	0
		PCR-FIML	**3**	**11**	**14**	**8**	**11**	**19**
		PCR-ILS	1	0	1	1	0	1
		PCR-LIML	1	0	1	3	1	4
0.95	10	PCR-2SLS	0	0	0	1	2	3
		PCR-3SLS	**6**	**5**	**11**	1	1	2
		PCR-FIML	3	7	10	**8**	**8**	**16**
		PCR-ILS	2	0	2	0	0	0
		PCR-LIML	1	0	1	2	1	3
	20	PCR-2SLS	1	0	1	0	1	1
		PCR-3SLS	1	4	5	0	0	0
		PCR-FIML	**8**	**7**	**15**	**10**	**8**	**18**
		PCR-ILS	1	1	2	0	1	1
		PCR-LIML	1	0	1	2	2	4
	30	PCR-2SLS	2	0	2	0	3	3
		PCR-3SLS	**6**	**5**	**11**	0	1	1
		PCR-FIML	3	7	10	**8**	**7**	**15**
		PCR-ILS	0	0	0	1	0	1
		PCR-LIML	1	0	1	3	1	4
	50	PCR-2SLS	2	1	3	1	0	1

		PCR-3SLS	**8**	**8**	**16**	1	0	1
		PCR-FIML	0	3	3	**7**	**8**	**15**
		PCR-ILS	0	0	0	2	0	2
		PCR-LIML	2	0	2	1	4	5
	100	PCR-2SLS	1	0	1	1	1	2
		PCR-3SLS	**5**	**8**	**13**	1	1	2
		PCR-FIML	5	4	9	**8**	**7**	**15**
		PCR-ILS	0	0	0	2	2	4
		PCR-LIML	1	0	1	0	1	1
	200	PCR-2SLS	0	0	0	1	1	2
		PCR-3SLS	1	6	7	0	0	0
		PCR-FIML	**10**	**6**	**16**	**9**	**8**	**17**
		PCR-ILS	0	0	0	0	3	3
		PCR-LIML	1	0	1	2	0	2
	300	PCR-2SLS	1	1	2	2	1	3
		PCR-3SLS	1	4	5	0	1	1
		PCR-FIML	**8**	**7**	**15**	**8**	**7**	**15**
		PCR-ILS	1	0	1	1	1	2
		PCR-LIML	1	0	1	1	2	3
0.99	10	PCR-2SLS	2	0	2	0	3	3
		PCR-3SLS	**7**	**4**	**11**	1	2	3
		PCR-FIML	2	8	10	**9**	**7**	**16**
		PCR-ILS	0	0	0	2	0	2
		PCR-LIML	1	0	1	0	0	0
	20	PCR-2SLS	0	0	0	0	0	0
		PCR-3SLS	**7**	**5**	**12**	1	1	2
		PCR-FIML	4	7	11	**10**	**8**	**18**
		PCR-ILS	0	0	0	1	3	4
		PCR-LIML	1	0	1	0	0	0
	30	PCR-2SLS	0	0	0	2	2	4
		PCR-3SLS	**4**	**8**	**12**	0	2	2
		PCR-FIML	4	4	8	**8**	**7**	**15**
		PCR-ILS	2	0	2	2	1	3
		PCR-LIML	2	0	2	0	0	0
	50	PCR-2SLS	1	0	1	1	0	1
		PCR-3SLS	**9**	**9**	**18**	0	3	3
		PCR-FIML	1	3	4	**9**	**6**	**15**
		PCR-ILS	1	0	1	1	0	1
		PCR-LIML	0	0	0	1	3	4
	100	PCR-2SLS	1	0	1	2	1	3
		PCR-3SLS	**6**	**5**	**11**	0	4	4
		PCR-FIML	2	7	9	**8**	**5**	**13**
		PCR-ILS	1	0	1	1	0	1
		PCR-LIML	2	0	2	1	2	3
	200	PCR-2SLS	1	0	1	1	2	3
		PCR-3SLS	6	3	9	0	3	3
		PCR-FIML	**2**	**9**	**11**	**9**	**6**	**15**
		PCR-ILS	2	0	2	0	0	0
		PCR-LIML	1	0	1	2	1	3
	300	PCR-2SLS	2	0	2	0	2	2
		PCR-3SLS	**4**	**11**	**15**	0	4	4
		PCR-FIML	4	1	5	**10**	**4**	**14**

		PCR-ILS	0	0	0	0	1	1
		PCR-LIML	2	0	2	2	1	3
0.999	10	PCR-2SLS	1	0	1	1	5	6
		PCR-3SLS	4	3	7	0	0	0
		PCR-FIML	**4**	**9**	**13**	**8**	**6**	**14**
		PCR-ILS	2	0	2	2	1	3
		PCR-LIML	1	0	1	1	0	1
	20	PCR-2SLS	1	0	1	5	2	7
		PCR-3SLS	**4**	**10**	**14**	3	0	3
		PCR-FIML	5	2	7	**3**	**6**	**9**
		PCR-ILS	0	0	0	1	2	3
		PCR-LIML	2	0	2	0	2	2
	30	PCR-2SLS	1	0	1	1	1	2
		PCR-3SLS	**7**	**5**	**12**	1	0	1
		PCR-FIML	3	7	10	**6**	**8**	**14**
		PCR-ILS	1	0	1	3	3	6
		PCR-LIML	0	0	0	1	0	1
	50	PCR-2SLS	1	0	1	0	1	1
		PCR-3SLS	6	3	9	3	0	3
		PCR-FIML	**4**	**9**	**13**	**5**	**7**	**12**
		PCR-ILS	0	0	0	0	3	3
		PCR-LIML	1	0	1	4	1	5
	100	PCR-2SLS	0	0	0	0	4	4
		PCR-3SLS	**9**	**5**	**14**	2	0	2
		PCR-FIML	0	7	7	**6**	**6**	**12**
		PCR-ILS	2	0	2	2	1	3
		PCR-LIML	1	0	1	2	1	3
	200	PCR-2SLS	1	0	1	2	2	4
		PCR-3SLS	1	1	2	0	0	0
		PCR-FIML	**7**	**11**	**18**	**10**	**6**	**16**
		PCR-ILS	2	0	2	0	2	2
		PCR-LIML	1	0	1	0	2	2
	300	PCR-2SLS	0	0	0	1	1	2
		PCR-3SLS	**2**	**11**	**13**	0	0	0
		PCR-FIML	7	1	8	**8**	**7**	**15**
		PCR-ILS	1	0	1	0	3	3
		PCR-LIML	2	0	2	3	1	4

Source: Computed from simulated results.

Note: From Table 5, at each sample size level, the summed ranks were further added over the equations and preferred estimators under both exact and over identified equations were bolded. Table 6 gives the summary of the preferred estimators.

Table 6: Summary of the preferred estimators under MSE criterion at different multicollinearity level (ρ) in the model

ρ	n	Exact Identified	Over Identified
0.8	10	PCR-3SLS, PCR-FIML	PCR-FIML
	20	PCR-3SLS	PCR-FIML
	30	PCR-3SLS	PCR-FIML

	50	PCR-3SLS	PCR-FIML
	100	PCR-3SLS	PCR-FIML
	200	PCR-FIML	PCR-FIML
	300	PCR-FIML	PCR-FIML
0.9	10	PCR-3SLS	PCR-FIML
	20	PCR-3SLS, PCR-FIML	PCR-FIML
	30	PCR-FIML	PCR-FIML
	50	PCR-3SLS	PCR-FIML
	100	PCR-3SLS	PCR-FIML
	200	PCR-FIML	PCR-FIML
	300	PCR-FIML	PCR-FIML
0.95	10	PCR-3SLS	PCR-FIML
	20	PCR-FIML	PCR-FIML
	30	PCR-3SLS	PCR-FIML
	50	PCR-3SLS	PCR-FIML
	100	PCR-3SLS	PCR-FIML
	200	PCR-FIML	PCR-FIML
	300	PCR-FIML	PCR-FIML
0.99	10	PCR-3SLS	PCR-FIML
	20	PCR-3SLS	PCR-FIML
	30	PCR-3SLS	PCR-FIML
	50	PCR-3SLS	PCR-FIML
	100	PCR-3SLS	PCR-FIML
	200	PCR-FIML	PCR-FIML
	300	PCR-3SLS	PCR-FIML
0.999	10	PCR-FIML	PCR-FIML
	20	PCR-3SLS	PCR-FIML
	30	PCR-3SLS	PCR-FIML
	50	PCR-FIML	PCR-FIML
	100	PCR-3SLS	PCR-FIML
	200	PCR-FIML	PCR-FIML
	300	PCR-3SLS	PCR-FIML

Source: Table 5

Note: From Table 6, in the over identified equations, the results indicate that PCR-FIML is generally preferred for all sample sizes at different multicollinearity levels. Whereas, in exact identified equations, PCR-3SLS or PCR-FIML or both estimators are generally preferred.

5. Discussion

This study develops hybrid estimators integrating Principal Component Analysis (PCR) with classical simultaneous equation model (SEM) to mitigate multicollinearity. A Monte Carlo simulation evaluates their performance under varying sample sizes and collinearity levels using mean squared error (MSE). Results reveal that for over – identified equations, PCR-FIML consistently outperforms all competitors across all scenarios. For exact identified equations, PCR-3SLS and PCR-FIML are generally preferred. The hybrid PCR–SEM estimators particularly PCR-FIML substantially improve estimation efficiency in the presence of multicollinearity.

6. Conclusion

The study concludes that hybrid PCR–SEM estimators offer a robust solution to multicollinearity in simultaneous equation models. PCR-FIML emerges as the most efficient estimator for over – identified equations while PCR-3SLS and PCR-FIML are recommended for exact identified equations. These estimators reduce inflated variances and unstable coefficient estimates inherent in classical SEM methods. Datasets with high collinearity should adopt these hybrid approaches to enhance reliability and inference accuracy.

Conflict of Interest

No conflict of interest was declared by the authors.

References

Alabi, O. O. & Oyejola, B. A. (2015). Assessment of Some Simultaneous Equation Estimation Techniques with Normal and Uniformly Distributed Exogenous Variables, *Journal of Applied Mathematics*, 6: 1902 – 1912. http://dx.doi.org/10.2015.611167.

Alabi, O. O. (2016a). Assessment of Some Simultaneous Equation Estimation Techniques under Uniformly Distributed Exogenous Variables with Correlated Error Terms, *Journal of Nigeria Association of Mathematical Physics*, 36, 203 – 2147.

Alabi, O. O. (2016b). Assessment of Some Simultaneous Equation Estimation Techniques under Normally and Uniformly Correlated Exogenous Variables. Unpublished Ph.D. Thesis University of Ilorin, Ilorin, Nigeria.

Alabi, O. O. (2019). Assessment of Estimation Techniques of Simultaneous Equation Model with Multicollinearity Problem under Normally and Uniformly Distributed Exogenous Variables, *FUW Trends in Science and Technology Journal*, 4(1), 103 – 117.

Alabi, R. E., Alabi, O. O., Ojo, O. O. & Fayose, T. S. (2025). Hybrid Estimators for Solving Multicollinearity in a Gaussian Linear Regression Model Based on Ridge – PCR Estimators. *PSSN Conference Proceedings, 1 – 20.*

Aladesuyi, A., Alabi, O. O. & Bello, A. H. (2026). Hybrid Principal Component – Based Estimators for Multicollinearity Issues in Simultaneous Equation Models. *Benin Journal of Physical Sciences* (in press).

Aladesuyi, A., Ayinde, K. & Fayose, T. S. (2025). Assessing the Role of Significant Roots in Parameter Estimation of Linear Regression Models under Multicollinearity. *Tech – Sphere Journal of Pure and Applied Sciences (TSJPAS),* 2(1), 1 – 16. https://doi:/10.5281/zenodo.15470100.

Ayinde, K. (2007). Equation to Generate Normal Variates with Desired Intercorrelations Matrix. *International Journal of Statistics and System,* 2(2), 99 – 111.

Çankaya, S. & Eker, S. (2025). Ridge vs PCR in Regression with Multicollinearity. *Turkish Journal of Agriculture – Food Science and Technology*, 2(34), 101 – 127.

De Jong, S. & Kiers, H. A. L. (1992). Principal Covariates Regression: Part I. *Chemometrics and Intelligent Laboratory Systems*, 14(1–3), 155 – 164.

Fayose, T. S. & Ayinde, K. (2019). Different Forms Biasing Parameter for Generalized Ridge Regression Estimator. *International Journal of Computer Applications,* 181, 21 – 29.

Fayose, T. S., Ayinde K. & Alabi, O. O. (2023a). M Robust Weighted Ridge Estimator in Linear Regression Model. *African Scientific Reports,* 2(123), 1 – 28.

Fayose, T. S., Ayinde K., Alabi, O. O. & Bello, A. H. (2023b). Robust Weighted Ridge Regression Based on S – Estimator. *African Scientific Reports,* 2(126), 1 – 28.

Garnier-Villarreal, M. & Jorgensen, T. D. (2024). Evaluating Local Model Misspecification with Modification Indices in Bayesian SEM. Structural Equation Modelling: *A Multidisciplinary Journal*, 31(6), 932 – 946.

Johnson, T. L., Ayinde, K. & Oyejola, B. A. (2010). Effect of Correlations and Equation Identification Status on Estimators of a System of Simultaneous Equation Model, *Electronic Journal of Applied Statistical Analysis*, 3(2), 115 – 125.

Judge, G. G., Griffiths, W. E., Hill, R. C., Lűtkepohl, H. & Lee, T. C. (1985). *The Theory and Practice of Econometrics*. (2nd ed.), Wiley.

Leamer, E. E. (1973). Multicollinearity: A Bayesian Interpretation. *Review of Economics and Statistics*, 55(3), 371 – 380.

Olubusoye, E. A. (2001). The Consequences of the Violation of the Assumption of Zero Correlation between Pairs of Random Stochastic Terms used in Monte Carlo Experiments. An Unpublished Ph.D. Thesis submitted to the Department of Statistics, University of Ibadan, Nigeria.

Okeke, N. C., Olanrewaju, S. O. & Mohammed, Z. A. (2025). Robust Estimation in SEMs Addressing Multicollinearity and Heteroscedasticity through Adaptive Penalized GMM Techniques. *African Journal of Mathematics and Statistics Studies*, 8(3), 73 – 95.

Schmidt, J. S. (2005). Econometrics, Published by McGraw – Hill International Edition.