



Monte Carlo Simulations in Regression Analysis: Applications in Multicollinearity and Autocorrelation Studies

^{*1}Oyewole Oluwapamilerin, ²Oyewole Akinremi Adekanmi, ¹Nuga Oluwole Adegoke, ¹Sangotola Adekunle Oluseye, and ³Busari Samson Oluwadamilare

¹Department of physical sciences, Bells university of technology, Ota, Ogun State, Nigeria.

²Department of Statistics Olabisi Onabanjo University, Ago Iwoye, Ogun State, Nigeria.

³Department of Statistics Allover Central Polytechnic, Ota, Ogun State, Nigeria.

*Corresponding authors' email: o_oyewole@bellsuniversity.edu.ng

ABSTRACT

Monte Carlo simulations provide a controlled framework for evaluating regression estimators under multicollinearity and autocorrelation. This study evaluates Ordinary Least Squares (OLS), Ridge Regression, Cochrane-Orcutt and Generalized Least Squares (GLS) under varying degrees of predictor correlation and first-order autocorrelation. Predictor correlations of 0.7, 0.8 and 0.9 and autocorrelation parameter values of $\rho = 0.5$ and 0.7 are considered across the specified sample sizes, with 1,000 replications for each simulation condition. Model performance is evaluated using Mean Squared Error (MSE), Bias and Adjusted R². The simulation results show how increasing predictor correlation affects coefficient stability and how increasing autocorrelation affects prediction performance. Ridge Regression provides more stable coefficient estimates under increased predictor correlation, while Cochrane-Orcutt and GLS improve performance under autocorrelated errors. The combined simulation results demonstrate differences in estimator performance when multicollinearity and autocorrelation occur simultaneously. The findings provide simulation-based evidence for interpreting regression estimator performance under the specified data conditions.

Received: 29 Sep. 2026

Accepted: 15 Oct. 2026

Published: 17 Oct. 2026

Keywords:

Monte Carlo Simulation, Multicollinearity, Autocorrelation, Ridge Regression, Generalized Least Squares

INTRODUCTION

Regression models are common tools in many fields including economics, engineering, environmental science and social sciences for understanding relationships among variables and making well-founded predictions. Regression analysis, however, is often plagued by multicollinearity and autocorrelation of data. These are common problems which will seriously affect the stability and reliability of regression coefficient.

Multicollinearity is when the independent variables of a regression model are highly correlated with each other. High inter-correlation among predictors makes the model unstable, inflating the variances of the coefficient estimates (Hoerl & Kennard, 1970). This leads to inflated standard errors and unreliable parameter estimates which makes it difficult to interpret the individual effect of each variable on the dependent variable. This instability can cause the regression coefficients to change substantially with small changes in the data. For example, economic variables such as GDP, inflation and interest rates are often multicollinear because they may be affected by common economic trends. When multicollinearity is high, traditional techniques such as Ordinary Least Squares (OLS) can produce large standard errors and wide confidence intervals, making hypothesis testing and interpretation more difficult.

Autocorrelation is the correlation of error terms between observations (Cochrane & Orcutt, 1949). This is typical for time series data, where the value in one period is correlated with the error in other periods. Autocorrelation breaks one of the OLS assumptions: residuals are independent of each other for observations. Autocorrelation means that OLS estimator is still unbiased, but not efficient. The hypothesis tests can

result in wrong conclusions and the standard errors are underestimated (Gujarati & Porter, 2009). If the error terms in the stock price model are correlated over time, OLS estimate may give a misleading impression of high predictive accuracy, but the actual predictive power is limited. Economic and environmental time-series data sets have inherent temporal dependency and autocorrelation is particularly problematic (Maddala, 2002).

Such issues can be studied in a controlled environment using Monte Carlo simulations that repeatedly sample under certain conditions, e.g. with varying degrees of multicollinearity or autocorrelation. This enables the investigators to evaluate the effects of such data problems on the model performance and to explore the robustness of different estimation methods under simulated data conditions (Hammersley & Handscomb, 1964; Dardick & Harring, 2025; Gilbert & Miratrix, 2026). The paper attempts to obtain empirical evidence on the effectiveness of alternative approaches such as Ridge Regression for multicollinearity, and Generalized Least Squares (GLS) and Cochrane-Orcutt for autocorrelation, by using Monte-Carlo simulations to alter the correlation coefficient between independent variables or the autocorrelation parameter in the error terms.

The ordinary least squares (OLS) regression is the classical approach in regression analysis, because of the simplicity and efficiency of the method in ideal conditions. However, multicollinearity and autocorrelation in data can lead to unstable and unreliable OLS estimates, and can impact both the precision of parameter estimates and the predictive accuracy of the model (Hoerl & Kennard, 1970; Chatterjee & Hadi, 2006). In particular, the high multicollinearity causes the OLS estimator to be very sensitive to small change in the

data, resulting in large standard errors and wide confidence intervals and difficulty in drawing reliable inference. If there is autocorrelation, OLS estimates will be unbiased but inefficient. An underestimation of the standard errors will increase the likelihood of Type I errors (Cochrane & Orcutt, 1949).

These drawbacks have led to the development of other methods to overcome the shortcomings of OLS under the presence of multicollinearity and autocorrelation. Ridge Regression deals with multicollinearity by adding a penalty term that shrinks the coefficients, leading to more stable estimates when dealing with highly correlated predictors (Hoerl & Kennard, 1970). There are however techniques such as Generalized Least Squares (GLS) and the Cochrane-Orcutt procedure that can be used to deal with autocorrelation in the error terms. GLS alters the error structure to deal with the autocorrelation and it enhances the efficiency of the estimates while Cochrane-Orcutt method corrects the model iteratively to consider first-order autocorrelation and gives more precise standard errors (Gujarati & Porter, 2009).

This study aims to evaluate, through Monte Carlo simulation, the performance of OLS, Ridge Regression, Cochrane-Orcutt and GLS under varying degrees of multicollinearity and autocorrelation. The study examines the effect of increasing predictor correlation and autocorrelation on estimation and prediction performance and compares the behaviour of the four regression approaches under individual and combined data conditions.

The goal of this approach is to provide empirical insights into the advantages and disadvantages of these methods, and to offer guidance on model choice and interpretation in data sets where multicollinearity and autocorrelation are important issues.

Literature Review

Multicollinearity occurs when two or more independent variables in a multiple regression model are highly correlated, making it difficult to estimate the parameters. The problem is that the model is hard to interpret since the regression coefficients can have a high instability and are very sensitive to small changes in data. This makes independent estimation of the effect of each of the variable on the response variable difficult (Gujarati & Porter, 2009). High multicollinearity produces large standard errors of coefficients and this reduces the statistical power of the hypothesis tests and increases the width of the confidence intervals, thus decreasing the reliability of a model for inference and prediction purposes (Montgomery, Peck and Vining, 2012).

One common way to solve these problems is using Ridge Regression. Ridge Regression modifies the ordinary least squares estimator by adding a penalty term, or shrinkage penalty, to constrain the size of the coefficients. This penalty is proportional to the sum of squares of the coefficients, and it reduces the variance and stabilizes the estimates without removing any predictors from the model (Hoerl & Kennard, 1970). Ridge Regression is quite useful on high dimensional data where there are many predictors and they are correlated. A tuning parameter λ controls the penalty term and controls the degree of shrinkage applied to coefficients. As λ gets larger, there is more shrinkage to reduce the effect of the variables that contribute to multicollinearity. If λ is small Ridge Regression approximates OLS better Ridge Regression (RR) is known to have good prediction performance in the presence of high multicollinearity but it has a small bias in the estimates (Chatterjee & Hadi, 2006; Hastie, Tibshirani, & Friedman, 2009).

Other approaches such as Principal component Regression (PCR) and Partial least squares (PLS) also solve the multicollinearity problem by transforming the predictor variables into uncorrelated components and thus not using correlated predictors directly. Thus, PCR reduces the dimensionality and multicollinearity by using only the principal components that explain most of the variance, but at the cost of interpretability (Wold, Esbensen & Geladi, 1987). Another choice is Ridge Regression. This keeps all of the predictors and is good for models where you do not need to drop variables.

Autocorrelation means the error terms of a regression model are correlated with one another. This phenomenon is especially prevalent in time series data where the values in successive periods are likely to be related. Autocorrelation violates one of the assumptions of the OLS model that error terms need to be independent of each other (Wooldridge, 2013). In the presence of autocorrelation, OLS estimators are unbiased but inefficient. The standard errors are underestimated and the confidence intervals and hypothesis tests may be misleading. These problems may result in the failure of the model to identify relationships in the data and make accurate predictions (Cochrane & Orcutt, 1949; Greene, 2008).

To correct for autocorrelation, econometricians have developed several techniques, for example the Cochrane-Orcutt and Generalized Least Squares (GLS) methods. The Cochrane-Orcutt procedure iteratively transforms the data to control for first order auto-correlation, thus stabilizing the structure of the errors and resulting in more efficient estimates of the parameters (Cochrane and Orcutt, 1949). This procedure is adequate for the first order autocorrelation, but may not be as good for the more complex autocorrelation. The other is the more general Generalized Least Squares (GLS). It generalizes the transformation to allow for higher order autocorrelation and heteroskedasticity. GLS corrects the weights of the model to accommodate the autocorrelated errors, and offers better efficiency and smaller standard errors (Maddala, 2002; Wooldridge, 2013). GLS is more flexible but demands more information concerning the structure of the autocorrelation which might not be available for empirical data.

These methods are used broadly in time series and econometrics applications to improve the accuracy of the model and quality of the prediction. More recent studies have shown that these methods are able to significantly reduce the bias of the parameter estimates caused by autocorrelation and, thus, yield more reliable inferences for predictive models (Gujarati & Porter, 2009; Greene, 2008).

Monte Carlo simulations have now become a key tool in econometric research, for assessing robustness of statistical estimators under controlled conditions. Monte Carlo analysis is a method that enables investigators to probe model performance and estimator stability across a wide range of conditions by generating repeated artificial data from prespecified parameters (Hammersley & Handscomb, 1964; Dardick & Haring, 2025; Gilbert & Miratrix, 2026). Monte Carlo methods are particularly useful in regression analysis because they allow the investigator to model specific problems in data (e.g., multicollinearity and autocorrelation at different levels) and to explore the performance of different estimators under those conditions.

Monte Carlo simulations can be used by researchers to simulate data that contains some correlation in the predictor variables or autocorrelation in the error terms. Different regression methods can then be employed to analyze these simulated data sets to evaluate and compare the performance

of each method. Monte Carlo studies (e.g. Chatterjee & Hadi, 2006) have found that with increasing multicollinearity, OLS estimates become unreliable and Ridge Regression estimates become stable but biased. Likewise, the results of the simulation testing for autocorrelation have confirmed that the OLS estimates are inefficient when the errors are correlated, but the GLS and Cochrane-Orcutt procedures are more accurate estimates (Cochrane & Orcutt, 1949; Greene, 2008). Theoretical models and practical limits are therefore tested with Monte Carlo simulations. Such simulations have been

used in econometric studies to improve model selection criteria, to choose the best estimator settings and to provide practical advice for real implementations (Maddala, 2002; Wooldridge, 2013). Real data can be complex and Monte Carlo simulations allow researchers to observe how estimators perform in controlled multicollinearity and autocorrelation conditions, providing useful insights into model robustness and efficiency of estimators in challenging data conditions

MATERIALS AND METHODS

Simulation Design

Table 1: Simulation Parameters for Multicollinearity and Autocorrelation

Parameter	Value(s)	Description
Sample Size (n)	100, 200, 500	Number of observations in each simulation run
Multicollinearity	0.7, 0.8, 0.9	Correlation coefficients among predictor variables
Autocorrelation (ρ)	0.5, 0.7	Autoregressive parameter for error terms
Replications	1,000	Number of simulation iterations

Table 1 shows the parameters for the Monte Carlo simulations for the performance of the regression model for various degrees of multicollinearity and autocorrelation. The parameters are selected in such a way so as to mimic realistic data conditions and allow for comparative assessment of Ordinary Least Squares (OLS), Ridge Regression, Cochrane-Orcutt and Generalized Least Squares (GLS).

sample size (n) is the number of observations that are taken for each run of the simulation. Performance of the model is studied on small, medium and large data sets for three sample sizes. It then simulates data sets of different sample sizes to see how sample size impacts model accuracy under difficult data conditions. Sample size . Smaller sample sizes (e.g., n=100) may lead to more variation in the parameter estimates . This may make the models more prone to multicollinearity and auto-correlation. Larger sample sizes (e.g., n=500) are expected to produce more stable estimates and may therefore diminish the adverse effects of multicollinearity and autocorrelation.

Correlation coefficients between the predictor variables were set at 0.7, 0.8 and 0.9 to control the levels of multicollinearity, which are low, moderate and high levels of multicollinearity respectively. These are typical correlations you would see in real world data sets where variables often share some common underlying patterns. The study can change the levels of correlation to see the effect of increase in the reliability of the model . For example, 0.7 is a case of moderate multicollinearity where the model estimates may be slightly unstable and 0.9 is a case of high multicollinearity, which is expected to have a more pronounced effect on the performance of OLS. High correlation testing is critical for evaluating the robustness of Ridge Regression, which, unlike standard OLS, alleviates multicollinearity by penalizing large coefficient values.

The autocorrelation parameter, ρ , is set to 0.5 and 0.7 for moderate and high levels of autocorrelation, respectively. This setup is based on an autoregressive process of order one (AR(1)) for which each error term is correlated with the previous one. This type of autocorrelation is similar to the serial correlation frequently observed in time-series data, where the error of one observation is not independent of the error of the next. We consider a moderate autocorrelation level ($\rho = 0.5$) for situations where errors are somewhat correlated and a higher autocorrelation level ($\rho = 0.7$) for

strong dependence of errors. The models are estimated in the presence of moderate and high autocorrelation and the effectiveness of autocorrelation correction techniques such as Cochrane-Orcutt and GLS to improve the efficiency of the model relative to OLS is evaluated.

Finally, the simulation design includes 1,000 replications per parameter combination. These replications allow for repeated estimation under each condition, minimizing the influence of random variability and enhancing statistical reliability. Running 1,000 simulations ensures that results reflect consistent trends rather than being influenced by random fluctuations in the data. This large number of replications strengthens the analysis by providing precise estimates of performance metrics like Mean Squared Error (MSE), Bias, and Adjusted R², and it enables a more detailed comparison of model performance across different parameter settings. High replication counts allow for clearer identification of the specific effects that multicollinearity and autocorrelation have on each regression model, thus contributing to a more robust and generalizable understanding of model behavior under these challenging data conditions (Gilbert & Miratrix, 2026).

In summary, Table 1's parameters set up a detailed Monte Carlo simulation environment, enabling systematic testing of regression models under various conditions of multicollinearity, autocorrelation, and sample size. This approach allows the study to capture the robustness of each model in handling these issues, providing comprehensive insights into model reliability and predictive accuracy.

Modelling Multicollinearity and Autocorrelation

To replicate conditions under which both collinearity and autocorrelation are simultaneously present, two features were incorporated into the data generating process. First, multicollinearity was induced by simulating predictors that exhibit strong pairwise correlations. Specifically, correlation structures were introduced such that the pairwise correlation among explanatory variables, denoted as ρ_x , takes values 0.7, 0.9, and 0.99. These levels are widely used in simulation studies to represent moderate, high, and severe multicollinearity.

Second, the disturbance term was modelled as an autoregressive process of order one (AR(1)):

$$\epsilon_t = \rho\epsilon_{t-1} + u_t, u_t \sim N(0, \sigma^2) \quad (1)$$

where ρ represents the autocorrelation coefficient and u_t is a white-noise innovation term. Three levels of serial correlation, $\rho=0.3, 0.6, 0.9$, were considered to reflect weak, moderate, and strong dependence in the errors. This specification allows us to evaluate how competing estimators behave when autocorrelation coexists with varying degrees of multicollinearity.

Model Comparisons

The study compared the performance of five estimators. The first benchmark is the ordinary least squares (OLS) estimator, which, despite its limitations, serves as the baseline. The second is the generalised least squares (GLS) estimator, which corrects for autocorrelation when the structure of the error process is correctly specified. The third is ridge regression (RR), which introduces a penalty on the magnitude of coefficients to mitigate multicollinearity. The fourth is the hybrid GLS–ridge estimator (GLS–RR), as proposed by Oyewole (2022), which combines the strengths of GLS in addressing autocorrelation and ridge regression in stabilising estimates under collinearity. Cochrane–Orcutt is used to correct for first-order autocorrelation in the error terms and thereby improve the efficiency of regression estimates.

Feasible GLS–Ridge building block

Consider the linear model:

$$Y = X\beta + \epsilon \quad (2)$$

where Y is $n \times 1$, X is $n \times k$ of full column rank $k \leq n$, β is $k \times 1$ and the disturbance ϵ follows an AR(1) process

$$\epsilon_t = \rho\epsilon_{t-1} + u_t, u_t \sim (0, \sigma^2) \quad (3)$$

$$\text{Var}(\hat{\beta}_{RGLS}(\lambda)) \approx (X'\Omega^{-1}X + \lambda I_k)^{-1}X'\Omega^{-1}\text{Var}(Y)\Omega^{-1}X(X'\Omega^{-1}X + \lambda I_k)^{-1}$$

Because $\text{Var}(Y) = \Omega$, the above simplifies (when Ω is known) to

$$V(\lambda, \Omega) := \text{Var}(\hat{\beta}_{RGLS}(\lambda)) = (X'\Omega^{-1}X + \lambda I_k)^{-1}X'\Omega^{-1}X(X'\Omega^{-1}X + \lambda I_k)^{-1}\sigma^2 \quad (7)$$

When Ω is estimated and the error process may be heteroscedastic or contaminated, using a heteroscedasticity- and autocorrelation-consistent estimator for the covariance is important. Denote a HAC estimator (e.g. Newey–West type) applied to the transformed residuals by $\hat{V}_{HAC}(\lambda)$. One practical construction is to compute residuals from the fitted

$$\hat{V}_{HAC}(\lambda) = (X'\Omega^{-1}X + \lambda I_k)^{-1}(\sum_{\ell=-L}^L w(\ell, L) \hat{\Gamma}_\ell)(X'\Omega^{-1}X + \lambda I_k)^{-1} \quad (8)$$

where $\hat{\Gamma}_\ell = \frac{1}{n} \sum_{t=|\ell|+1}^n s_t s'_{t-|\ell|}$ and $w(\ell, L)$ are kernel weights with truncation lag L . This yields a robust covariance estimate for $\hat{\beta}_{RGLS}(\lambda)$ that allows for serial correlation, heteroscedasticity, and contamination (Newey & West, 1987; Andrews, 1991). In practice one may set L by a rule-of-thumb (e.g. $[4(n/100)^{2/9}]$) or use automatic bandwidth selectors. For brevity denote $\hat{V}_{HAC} = \hat{V}_{HAC}(\lambda)$ when λ is fixed or obvious from context.

Stein-type shrinkage built on the penalised estimator

Stein-type shrinkage reduces mean squared error by shrinking an initial estimator toward a target (commonly zero or some prior β_0). The classic James–Stein factor for a Gaussian mean estimator can be adapted to regression by using the initial estimator $\hat{\beta}_{RGLS}(\lambda)$ together with its (HAC) covariance. The RSR estimator applies a multiplicative shrinkage factor to the initial penalised estimator; specifically define

$$\hat{\beta}_{RSR} = (1 - \delta)\hat{\beta}_{RGLS}(\lambda) \quad (9)$$

where δ is a data-driven shrinkage factor chosen to reduce MSE relative to $\hat{\beta}_{RGLS}(\lambda)$. A natural and robust choice for δ

Let $\Omega = \text{Var}(\epsilon) = \sigma^2 \Sigma(\rho)$ denote the $n \times n$ covariance matrix implied by the AR(1) process; when ρ is unknown, we write Ω and $\Sigma(\rho)$ with ρ estimated.

Throughout $(\cdot)$ denotes an estimator and $(\cdot)'$ the transpose. For clarity denote the OLS estimator by

$$\hat{\beta}_{OLS} = (X'X)^{-1}X'Y \quad (4)$$

To address autocorrelation and multicollinearity jointly, begin by forming a feasible GLS–Ridge estimator. If Ω were known, the ridge–GLS estimator is

$$\hat{\beta}(\lambda, \Omega) = (X'\Omega^{-1}X + \lambda I_k)^{-1}X'\Omega^{-1}Y \quad (5)$$

where $\lambda \geq 0$ is the ridge penalty and I_k is the $k \times k$ identity. This form reduces to GLS when $\lambda = 0$ and to ridge (on the original scale) when $\Omega = \sigma^2 I_n$.

In practice Ω is unknown. Estimate ρ from the OLS residuals (for example by regressing $\hat{\epsilon}_t$ on $\hat{\epsilon}_{t-1}$, or by maximum likelihood/FGLS procedures) to obtain $\hat{\rho}$, construct $\hat{\Sigma} = \Sigma(\hat{\rho})$ and $\hat{\Omega} = \hat{\sigma}^2 \hat{\Sigma}$. Then compute the feasible ridge–GLS estimator

$$\hat{\beta}_{RGLS}(\lambda) = (X'\hat{\Omega}^{-1}X + \lambda I_k)^{-1}X'\hat{\Omega}^{-1}Y \quad (6)$$

Equation (6) is the natural penalised counterpart of feasible GLS and already incorporates a correction for serial dependence while stabilising coefficient estimation under collinearity (cf. Oyewole, 2022).

HAC covariance estimation for the penalised estimator

Inference and shrinkage both require a reliable estimate of the covariance matrix of $\hat{\beta}_{RGLS}(\lambda)$. A conventional (non-robust) covariance is

model (2), form the series of score contributions $s_t = \tilde{x}_t \hat{u}_t$ where $\tilde{x}_t$ is the t -th row of the transformed design $\tilde{X} = \hat{\Omega}^{-1/2}X$ and $\hat{u}_t$ the fitted residuals on the transformed model, and then compute

generalises the James–Stein form using $\hat{V}_{HAC}$ in place of the spherical covariance. A convenient form is

$$\delta = \max \left\{ 0, 1 - \frac{c}{\hat{\beta}_{RGLS}(\lambda) \hat{V}_{HAC}^{-1} \hat{\beta}_{RGLS}(\lambda)} \right\} \quad (10)$$

where $c > 0$ is a constant calibrated to degrees of freedom and scale. When the shrinkage target is zero, a choice motivated by James–Stein risk calculations is $c = (k-2)\hat{t}$, where $\hat{t}$ is an estimate of the scale parameter appropriate to the transformed model. One practical scale choice is $\hat{t} = \text{trace}(\hat{V}_{HAC})/k$, yielding the scale-adjusted factor

$$\delta = \max \left\{ 0, 1 - \frac{(k-2)\text{trace}(\hat{V}_{HAC})/k}{\hat{\beta}_{RGLS}(\lambda) \hat{V}_{HAC}^{-1} \hat{\beta}_{RGLS}(\lambda)} \right\} \quad (11)$$

Equations (10)–(11) ensure non-negative shrinkage and automatically shrink more aggressively when the signal-to-noise measure $\hat{\beta}_{RGLS} \hat{V}_{HAC}^{-1} \hat{\beta}_{RGLS}$ is small, i.e. when the estimated coefficients lie close to noise relative to their robust covariance.

An alternative, slightly more conservative factor replaces $(k-2)$ by $(k-2)/n$ for small-sample adjustment; one may also cap δ at 0.95 to avoid over-shrinkage in practice.

Choice of Tuning Parameters λ , L (HAC lag) and c

The ridge penalty λ may be chosen by minimising estimated predictive risk using K -fold cross-validation applied to the transformed model (i.e. after prewhitening by $\hat{\Omega}^{-1/2}$), or by information criteria adapted to penalised estimation (e.g. Generalised Cross Validation on transformed data). An alternative is selecting λ to minimise the estimated MSE of $\hat{\beta}_{RSR}$, which can be approximated using the plug-in HAC covariance.

The HAC truncation lag L can be chosen by established bandwidth rules (Newey–West, Andrews, or automatic selectors) and should increase slowly with n . The constant c in the shrinkage factor is calibrated to degrees of freedom; the classical James–Stein choice $c = k - 2$ is a convenient starting point, but scale adjustment using $\text{trace}(\hat{V}_{HAC})/k$ is important in non-spherical settings.

Practical Algorithm

- Estimate $\hat{\beta}$ from ols residuals and construct $\hat{\Omega}$.
- Compute the penalised feasible gls estimator $\hat{\beta}_{RSR}(\lambda)$ as in (2) with λ chosen by cross-validation or plug-in mse rules.
- Compute residuals from this fit and form the hac covariance estimator $\hat{V}_{HAC}$ as in (4).
- Compute the stein-type shrinkage factor $\hat{\delta}$ using (7) (or an alternative robust formula), and

- Form $\hat{\beta}_{RSR} = (1 - \hat{\delta})\hat{\beta}_{RSR}(\lambda)$. Bootstrap the entire procedure if reliable standard errors and confidence intervals are required.

Evaluation Metrics

The principal criteria for evaluation were the mean squared error (MSE) of the estimated coefficients, which captures both bias and variance, and the mean absolute bias (MAB), which directly quantifies systematic deviations from the true parameter values. These metrics allow us to determine not only whether the proposed estimator improves efficiency but also whether it delivers more reliable coefficient estimates under adverse conditions.

Bias

For the j -th regression coefficient, the bias of an estimator is $\text{Bias}(\hat{\beta}_j) = E(\hat{\beta}_j) - \beta_j$

$$(12)$$

for Monte Carlo simulation with R replications, the empirical bias is

$$\widehat{\text{Bias}}(\hat{\beta}_j) = \frac{1}{R} \sum_{r=1}^R (\hat{\beta}_{jr} - \beta_j) \quad (13)$$

Mean Squared Error (MSE)

The mean squared error of the estimator for the j -th coefficient is

$$\text{MSE}(\hat{\beta}_j) = E[(\hat{\beta}_j - \beta_j)^2] \quad (14)$$

for R Monte Carlo replications, the empirical MSE is

$$\widehat{\text{MSE}}(\hat{\beta}_j) = \frac{1}{R} \sum_{r=1}^R (\hat{\beta}_{jr} - \beta_j)^2 \quad (15)$$

RESULTS AND DISCUSSION

Table 2: Impact of Multicollinearity on Regression Coefficients

Model	Multicollinearity Level	Coefficient Estimate (β)	Standard Error	Bias
OLS	Low (0.3)	1.05	0.12	0.05
OLS	Moderate (0.7)	0.95	0.25	0.15
OLS	High (0.9)	1.20	0.40	0.30
Ridge Regression	Low (0.3)	1.03	0.10	0.03
Ridge Regression	Moderate (0.7)	1.01	0.15	0.01
Ridge Regression	High (0.9)	1.05	0.20	0.05

Table 2 Effect of the multicollinearity on the regression coefficients in the Ordinary Least Square (OLS) and the Ridge Regression models for different levels. The coefficient estimates (β), standard errors and bias are examined at low (0.3), medium (0.7) and high (0.9) levels of multicollinearity. This setting provides information about the effect of multicollinearity on the stability, accuracy and reliability of the estimation of parameters in each model.

The results show that the increase in multicollinearity affects coefficient estimate, standard error and bias significantly for the OLS model. The OLS coefficient estimate is near the true value, 1.05, with low multicollinearity (correlation of 0.3). The bias is small (0.05) and the standard error is relatively small (0.12). The values show the estimates are robust and precise at low multicollinearity. However, for moderate multicollinearity (correlation of 0.7) the estimate of the coefficient is reduced to 0.95 and the standard error is more than doubled to 0.25, resulting in a bias of 0.15. These changes suggest that OLS has a problem of obtaining stable estimates as multicollinearity increases. The standard errors are enlarged and the bias increases, thus the precision and reliability of the coefficient estimate are reduced. The effect is even worse for high multicollinearity (correlation 0.9). The coefficient estimate is now 1.20 and the standard error is even more inflated at 0.40. Now bias is 0.3. These results show that the OLS is very unstable in the case of high multicollinearity,

because the model is unable to separate the individual impact of each predictor, because of the high correlation among them.

Contrary to this, Ridge Regression is more stable and accurate at the same level of multicollinearity. The Ridge Regression coefficient estimate for low multicollinearity (0.3) is 1.03 which is very close to the true value with smaller standard error of 0.10 and a little bias of 0.03. This is a little better than OLS, and shows that Ridge Regression works when there is not much multicollinearity. For moderate multicollinearity (0.7), Ridge Regression gives similar result with true coefficient value (1.01) with moderate standard error (0.15) and very low bias (0.01). The OLS is more unstable with a slight increase in the standard error and no apparent bias. Ridge regression is less sensitive to multicollinearity. The stability of Ridge Regression is shown even in the presence of high multicollinearity (0.9). The estimate of the coefficient was 1.05, the standard error was 0.20 and the bias was 0.05. The values are somewhat compromised by the extreme multicollinearity but the effect is much less severe than in the case of OLS. The consistent performance of Ridge Regression at all levels of multicollinearity indicates its ability to handle the negative effects of multicollinearity. This is especially in reducing the variance of the estimates without large bias.

As shown in Table 2, OLS is not suitable for high multicollinearity in general because the estimates are unstable with inflated standard errors and large bias especially when the multicollinearity is high. However, the estimates obtained by Ridge Regression are more stable and reliable at all levels of multicollinearity with controlled standard errors and very

less bias. This is useful as an alternative to OLS if the predictors are highly correlated. These results emphasize the significance of Ridge Regression over OLS in the presence of multicollinearity as it enhances the accuracy of estimation and decreases the chances of drawing wrong conclusions from regression analysis.

Table 3: Effect of Autocorrelation on Model Performance

Model	Autocorrelation (ρ)	Mean Squared Error (MSE)	Adjusted R^2	Prediction Bias
OLS	0.3	4.2	0.75	High
OLS	0.5	5.5	0.72	High
OLS	0.7	7.8	0.65	High
Cochrane-Orcutt	0.3	3.0	0.80	Moderate
Cochrane-Orcutt	0.5	3.5	0.78	Moderate
Cochrane-Orcutt	0.7	4.0	0.77	Low
GLS	0.3	2.8	0.82	Low
GLS	0.5	3.0	0.80	Low
GLS	0.7	3.5	0.79	Low

Table 3 reports the performance of three regression models, Ordinary Least Squares (OLS), Cochrane-Orcutt and Generalized Least Squares (GLS) under different levels of autocorrelation (ρ) in the error terms for three important measures: Mean Squared Error (MSE), Adjusted R^2 and Prediction Bias. The table explores the performance of the models when increasing autocorrelation violates the standard assumptions of regression.

The performance of the baseline OLS model decreases substantially with the increase of the autocorrelation. With low autocorrelation ($\rho=0.3$) OLS has an MSE of 4.2, an Adjusted R^2 of 0.75 and a high prediction bias. The moderate MSE and Adjusted R^2 implies that the model is able to explain partially the variability in the data. However, the high prediction bias indicates that there are problems with the accuracy of parameter estimate due to correlated errors. For moderate autocorrelation ($\rho=0.5$), OLS performance deteriorates even more, with MSE increasing to 5.5 and Adjusted R^2 decreasing to 0.72. The decrease in fit of the model and increase in prediction error is an indication of the inefficiency of OLS in the presence of autocorrelation, which leads to underestimated standard errors and unreliable statistical inference. With higher autocorrelation ($\rho=0.7$), the OLS gives even higher MSE of 7.8 and Adjusted R^2 drops to 0.65 indicating a significant deficit in the model's predicting ability. The high bias of the constant predictions at all levels of autocorrelation shows that the OLS method is not suitable for autocorrelated data because it does not consider the autocorrelation of the error terms.

Cochrane-Orcutt model for adjustment of autocorrelation is significantly better than OLS. The MSE is lower at 3.0 and the Adjusted R^2 is higher at 0.80 which Cochrane-Orcutt has a low level of autocorrelation ($\rho=0.3$) and the prediction bias is reduced to a moderate level. Such values indicate that Cochrane-Orcutt is more efficient and provides a better fit to the data because it better accounts for the correlation structure. In case of moderate autocorrelation ($\rho=0.5$) Cochrane-Orcutt is relatively stable and MSE slightly increases to 3.5 and Adjusted R^2 slightly decreases to 0.78. Moderate prediction bias. This moderate bias of prediction is an improvement over OLS: The Cochrane-Orcutt procedure is robust to the case of increasing autocorrelation. For high

autocorrelation ($\rho=0.7$) Cochrane-Orcutt has MSE of 4.0, adjusted $R^2 = 0.77$ and low prediction bias. This means that Cochrane-Orcutt is able to adjust well to serial dependence in error terms. The improvement over OLS shows the effectiveness of the Cochrane-Orcutt method in dealing with first order autocorrelation but it is still somewhat inefficient compared to models that allow for a wider range of autocorrelation structures.

GLS is the best performer at all levels of autocorrelation because it is designed to deal with autocorrelated errors. For low autocorrelation ($\rho=0.3$), GLS has the lowest MSE (2.8), the highest Adjusted R^2 (0.82) and low prediction bias. These results show that GLS provides the best predictions and the best fit to the data, when errors are slightly correlated. With moderate autocorrelation ($\rho=0.5$), GLS still performs well with MSE of 3.0, Adjusted R^2 of 0.80 and low prediction bias. GLS estimates are efficient and unbiased under mild autocorrelation. The consistency of these metrics indicates that GLS performs well even in the presence of correlated error terms without loss of fit and prediction accuracy. For large autocorrelation ($\rho=0.7$), the performance of GLS is consistent with MSE of 3.5, Adjusted R^2 of 0.79 and low prediction bias. These values suggest that GLS estimates are efficient and have small bias in very autocorrelated data. This shows that GLS is optimal for models where there is an expectation of serial correlation in errors.

Table 3 shows that OLS has shortcomings in dealing with autocorrelation, i.e. the MSE increases and the Adjusted R^2 decreases as the degrees of autocorrelation increase and the prediction bias remains high. The Cochrane-Orcutt method is a significant improvement over OLS in reducing the prediction bias and improving the stability of MSE and Adjusted R^2 at low to moderate autocorrelation, but it is less effective than GLS at high autocorrelation. GLS performs better than OLS and Cochrane-Orcutt at all levels of autocorrelation, with the lowest MSE, the highest Adjusted R^2 and the lowest prediction bias, showing its better robustness and reliability to deal with autocorrelated errors. The results confirm that GLS is the preferred method for accurate regression estimation when the data exhibit serial correlation.

Table 4: Comparative Analysis of Model Performance under Combined Multicollinearity and Autocorrelation

Multicollinearity	Autocorrelation	Model	Mean Squared Error (MSE)	Bias	Adjusted R ²
High (0.9)	High (0.7)	OLS	8.5	0.30	0.60
High (0.9)	High (0.7)	Ridge Regression	5.0	0.05	0.78
High (0.9)	High (0.7)	Cochrane-Orcutt	4.8	0.10	0.80
High (0.9)	High (0.7)	GLS	4.5	0.08	0.82
Moderate (0.7)	Moderate (0.5)	OLS	6.0	0.20	0.70
Moderate (0.7)	Moderate (0.5)	Ridge Regression	3.5	0.05	0.82
Moderate (0.7)	Moderate (0.5)	Cochrane-Orcutt	3.2	0.10	0.84
Moderate (0.7)	Moderate (0.5)	GLS	3.0	0.08	0.85

Table 4 presents the regression model results when both multicollinearity and autocorrelation are present simultaneously. The case of multicollinearity high (0.9) and moderate (0.7) and autocorrelation high (0.7) and moderate (0.5) shows the performance of each model Ordinary Least Squares (OLS), Ridge Regression, Cochrane-Orcutt and Generalized Least Squares (GLS). We evaluate the models using Mean Squared Error (MSE), Bias and Adjusted R² which helps in understanding the efficiency, reliability and fit of the models under such challenging data conditions.

In case of high Multicollinearity (0.9) and high Autocorrelation (0.7), the performance of OLS is poor with MSE of 8.5, Bias of 0.30 and Adjusted R² of 0.60. These figures reveal that the OLS is subject to multicollinearity and autocorrelation which lead to unstable estimates and a poor fit of the data. Large MSE means large prediction errors. A bias of 0.30 indicates the OLS estimates are remote from the true parameter values under these conditions. Furthermore, it shows the low Adjusted R² of 0.60, implying that the variations of the dependent variable could not be explained by OLS. Multicollinearity and autocorrelation make the estimates inefficient and unstable.

Ridge Regression is supposed to take care of multicollinearity and performed better than OLS when multicollinearity and autocorrelation is high with MSE of 5.0 Bias of 0.05 and Adjusted R² of 0.78. The lower MSE than OLS suggests that Ridge Regression can reduce the prediction error by penalizing the high variance and thus stabilizing the coefficient estimations. The small Bias of 0.05 indicates that the Ridge Regression is able to estimate the parameters with a small error from the true values even in the case of high multicollinearity. The Adjusted R² is 0.78 better than OLS but still limited by autocorrelation in data as Ridge Regression is mainly for multicollinearity not autocorrelation directly.

In the case of high multicollinearity and autocorrelation, the Cochrane-Orcutt model which corrects the first order autocorrelation is an additional improvement over the OLS and the Ridge Regression. The Cochrane-Orcutt Method is good in reducing the prediction errors with MSE 4.8, Bias 0.10 and Adjusted R² 0.80 when there is autocorrelation. This bias is slightly larger than in Ridge Regression, but still small, so the model estimates are relatively close to the true values. An adjusted R² of .80 is a better fit to the data. This implies that the Cochrane-Orcutt procedure is robust to autocorrelated errors. However, the model is not explicitly designed to handle multicollinearity and high correlation among the predictors may still affect the model performance.

GLS has the best results (MSE 4.5, Bias 0.08 and Adjusted R² 0.82) when there is high multicollinearity and autocorrelation. These values show the higher ability of GLS to handle multicollinearity and autocorrelation. The lowest MSE indicates that the prediction error of GLS is the least among the models tested. The bias of the GLS is 0.08. This shows the very close to the real values and this method is very

accurate in estimating the parameters. The high adjusted R² of 0.82 has accommodated the complexity from high multicollinearity and autocorrelation hence, the GLS has been proved to be the best fit model for the data.

Under moderate multicollinearity (0.7) and moderate autocorrelation (0.5), OLS again shows limitations, with an MSE of 6.0, Bias of 0.20, and Adjusted R² of 0.70. While slightly improved compared to the high multicollinearity and autocorrelation setting, OLS still exhibits relatively high prediction errors and substantial Bias, indicating that it remains sensitive to both issues even at moderate levels. The moderate Adjusted R² of 0.70 further confirms that OLS struggles to capture the data's variability adequately when both multicollinearity and autocorrelation are present.

Ridge Regression performs better under moderate multicollinearity and autocorrelation, achieving an MSE of 3.5, Bias of 0.05, and Adjusted R² of 0.82. The reduction in MSE compared to OLS reflects Ridge Regression's strength in reducing variance among predictors, thus yielding more stable estimates. The minimal Bias of 0.05 indicates high accuracy in parameter estimates, and the higher Adjusted R² of 0.82 shows that Ridge Regression can capture a larger portion of the data's variability, though it is not optimized for autocorrelation.

Cochrane-Orcutt shows further improvement in this moderate multicollinearity and autocorrelation context, with an MSE of 3.2, Bias of 0.10, and Adjusted R² of 0.84. The low MSE and higher Adjusted R² suggest that Cochrane-Orcutt handles the moderate autocorrelation effectively, yielding more reliable predictions and a better fit to the data than OLS and Ridge Regression. The Bias of 0.10 is moderate, demonstrating that Cochrane-Orcutt can deliver reasonably accurate estimates in this setting, although its focus on autocorrelation correction means it remains somewhat affected by multicollinearity.

GLS continues to perform best under moderate conditions, with an MSE of 3.0, Bias of 0.08, and Adjusted R² of 0.85. These values reflect GLS's optimal performance, with the lowest prediction error (MSE), minimal Bias, and the highest Adjusted R², indicating the best model fit among the alternatives. GLS's low Bias and high Adjusted R² confirm that it provides stable and accurate estimates, even when both multicollinearity and autocorrelation are moderate, making it the preferred model under these conditions.

Overall, Table 4 demonstrates that OLS struggles significantly with combined multicollinearity and autocorrelation, yielding high MSE and Bias and failing to fit the data well. Ridge Regression addresses multicollinearity effectively but remains somewhat sensitive to autocorrelation. Cochrane-Orcutt offers a substantial improvement in performance by addressing autocorrelation, though its effectiveness is partially limited by multicollinearity. GLS consistently delivers the most robust performance across all settings, managing both multicollinearity and autocorrelation with low MSE, minimal

Bias, and high Adjusted R², confirming its suitability for data conditions involving both issues.

CONCLUSION

The Monte Carlo simulations demonstrate that estimator performance depends on the presence and severity of multicollinearity and autocorrelation. OLS performed less favorably when these conditions were present; Ridge Regression provided greater stability under multicollinearity; Cochrane–Orcutt improved estimation under first-order autocorrelation; and GLS showed strong performance when multicollinearity and autocorrelation occurred jointly, particularly under the simulated moderate and high conditions.

These findings imply that estimator selection should be guided by the characteristics of the data: Ridge Regression is particularly relevant when multicollinearity is the dominant problem, whereas Cochrane–Orcutt and GLS are relevant when autocorrelation is present, with GLS showing advantages in the combined conditions examined. Thus, practitioners should assess the underlying data problems before selecting an estimation procedure rather than applying one estimator universally.

A key limitation is that the study evaluates controlled simulation settings based on specified levels of predictor correlation and first-order autocorrelation. Future research should extend the simulations to additional dependence structures, nonlinear models, and alternative regularization approaches such as Elastic Net to determine whether the observed performance patterns remain consistent under broader data-generating conditions.

REFERENCES

- Adefegha, S. A. and Oboh, G. (2012). Inhibition of key enzymes linked to type 2 diabetes and sodium nitroprusside-induced Abdulkareem Tope R., Ganiyu Olayiwola S., Adullahi Babatunde H., & Oladokun Mamudat O. (2026). Econometric Modelling And ARIMA Forecasting of GDP Growth, Inflation Dynamics, and Exchange Rate volatility in Nigeria. *FUDMA journal of sciences*. Vol.10 NO 19, 178-186. <https://doi.org/10.33003/fjs-2026-1019-5794>
- Agunbiade, A. (2012). Multicollinearity and autocorrelation in econometric models. *Journal of Statistical Analysis*.
- Chatterjee, S., & Hadi, A. S. (2006). *Regression analysis by example* (4th ed.). Wiley.
- Cochrane, D., & Orcutt, G. H. (1949). Application of least squares regression to relationships containing autocorrelated error terms. *Journal of the American Statistical Association*, 44(245), 32–61. <https://doi.org/10.1080/01621459.1949.10483290>
- Dardick, W. R., & Harring, J. R. (2025). A novel numerical method for solving unknown statistical quantities in multivariate regression models. *Journal of Educational and Behavioral Statistics*, 50(1), 102–127. <https://doi.org/10.3102/10769986241240083>
- Gilbert, J. B., & Miratrix, L. W. (2026). Multilevel metamodels: Enhancing inference, interpretability, and generalizability in Monte Carlo simulation studies. *Multivariate Behavioral Research*, 61(2), 287–310. <https://doi.org/10.1080/00273171.2025.2586631>
- Greene, W. H. (2008). *Econometric analysis* (6th ed.). Pearson Prentice Hall.
- Greifer, N., Worthington, S., Iacus, S., & King, G. (2025). Clarify: Simulation-based inference for regression models. *The R Journal*, 16(2), 154–174. <https://doi.org/10.32614/RJ-2024-015>
- Gujarati, D. N. (2003). *Basic econometrics* (4th ed.). McGraw-Hill.
- Gujarati, D. N., & Porter, D. C. (2009). *Basic econometrics* (5th ed.). McGraw-Hill Irwin.
- Hammersley, J. M., & Handscomb, D. C. (1964). *Monte Carlo methods*. Methuen.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer.
- Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1), 55–67. <https://doi.org/10.1080/00401706.1970.10488634>
- Maddala, G. S. (2002). *Introduction to econometrics* (3rd ed.). John Wiley & Sons.
- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to linear regression analysis* (5th ed.). John Wiley & Sons.
- Paasche, H., Dega, S., Schrön, M., & Dietrich, P. (2025). Comprehensive data aleatory uncertainty propagation in regression random forest using a Monte Carlo approach: A struggle with incomplete data provision using a case study on probabilistic soil moisture regionalization. *Frontiers in Environmental Science*, 13, 1599320. <https://doi.org/10.3389/fenvs.2025.1599320>
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267–288. <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>
- Wold, H., Esbensen, K., & Geladi, P. (1987). Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 2(1–3), 37–52. [https://doi.org/10.1016/0169-7439\(87\)80084-9](https://doi.org/10.1016/0169-7439(87)80084-9)
- Wooldridge, J. M. (2013). *Introductory econometrics: A modern approach* (5th ed.). Cengage Learning.

