# REGRESSION WITH VOLATILE ERRORS IN THE PRESENCE OF MEASUREMENT ERRORS

Anna Thomas[1], Nimitha John[2]

•

[1,2]Department of Statistics & Data Science
CHRIST(Deemed to be University)
Bengaluru, Karnataka, India
anna.thomas@res.christuniversity.in[1], nimitha.john@christuniversity.in[2]

**Abstract**

*This study explores the estimation and testing of regression models with volatile errors when measurement errors are present. The presence of measurement error in models with heteroscedastic disturbances, such as those following an autoregressive conditional heteroscedasticity (ARCH) or Generalized ARCH (GARCH) structure, can lead to biased estimates and misleading inferences. To address this, we develop an estimation framework that accounts for both heteroscedasticity and mismeasured observations, ensuring consistent and asymptotically normal parameter estimates. We estimate the parameters using corrected score estimation and weighted linear regression, which effectively mitigate the impact of measurement error and hetroscedasticity. Additionally, we perform a Likelihood Ratio (LR) test to assess the significance of measurement errors in regression models with volatile errors. Through Monte Carlo simulations, we analyze the bias and efficiency of traditional estimators and demonstrate the robustness of our proposed approach. Finally, the methodology is applied to real-life economic and financial data, illustrating its practical relevance and effectiveness in empirical research. The findings contribute to improving statistical inference in models where measurement error and volatility coexist, ensuring more reliable and accurate parameter estimation.*

**Keywords:** regression, volatile errors, measurement error, corrected score, likelihood ratio test

## 1. Introduction

The Simple Linear Regression (SLR) model is a fundamental statistical technique used to model and analyze the relationship between an explanatory variable and one response variable. The model relies on several key assumptions: (i) Linearity, means the relationship between explanatory and response variable is linear; (ii) Independence, where the residuals are not correlated with each other; (iii) Normality, requiring the residuals to follow a normal distribution; and (iv) Homoscedasticity (constant variance), which states that the variance of errors remains the same across all values of explanatory variable. When the assumption of homoscedasticity (constant variance) is violated, the residuals exhibit heteroscedasticity, where the variability of response variable changes across different values of an explanatory variable. This can result in biased and inefficient estimates of standard errors, thereby compromising the reliability of hypothesis testing. Measurement error in regression analysis arises when the recorded values of variables differ from their true values due to imprecise measurement. These errors can arise from various sources, such as instrument limitations, data recording mistakes, or respondent misreporting. When measurement error affects the independent variables (explanatory variables), it may result in parameter estimates that are both biased and inconsistent.

In Literature various techniques are existing for determining the coefficients of a regression model when measurement error is present. The presence of measurement error may influence either the independent variable, the dependent variables, or both. A linear model was proposed to account for measurement errors, and the impact of these errors on the ordinary least squares estimators was examined [4]. A corrected score function was introduced to adjust for the effects of measurement error on parameter estimation [7]. A comprehensive framework for the errors-in-variables problem was provided, covering both linear and nonlinear models, and addressing functional and structural cases with dependent measurement errors [11]. It explores the asymptotic bias in M-estimators due to measurement error and proposes a bias-reduced estimator, particularly in the context of generalized linear models, supported by asymptotic analysis and Monte Carlo simulations. The effect of measurement errors in the linear regression model with autocorrelated errors was analyzed using a Lagrange Multiplier (LM) test [9].

Evidence of ARCH(1) errors in the context of spurious regressions is provided in the study [1]. The study demonstrates how exact predictive densities can be obtained in ARCH linear models using Monte Carlo integration with importance sampling, enhancing computational efficiency through the use of exact likelihood functions and antithetic acceleration [5]. The issue of heterogeneity in error variance in regression and analysis of variance is addressed, with an iterative method proposed to estimate appropriate weights from residuals [6]. The application of Weighted Least Squares regression in forecasting is studied [12]. A test for a linear regression model with ARCH disturbances is developed using the Information Matrix (IM) test framework [2]. The likelihood ratio test for detecting the presence of measurement error is examined [3]. A dynamic model with ARCH errors where the underlying process is latent and subject to additive measurement error was investigated by [10]. Corrected score estimation in regression models with autocorrelated errors is studied [13]. The aforementioned articles focus on the impact of measurement error in linear regression models. However, to the best of our knowledge, the estimation and testing of regression models with volatile (heteroscedastic) errors in the presence of measurement error have not been explored in the existing literature. Therefore, the objective of this study is to estimate the parameters of such a model using the corrected score estimation method and to test the presence of measurement error through the likelihood ratio test. So in this paper, our objective is to estimate the parameters of regression with volatile errors in the presence of measurement error using corrected score estimation, then testing the effect of measurement error by using Likelihood ratio test.

The remainder of the paper is organized as follows: Section 2 presents the methodology for estimating regression parameters under heteroscedastic errors with measurement error. Section 3 discusses the proposed approach for parameter estimation and testing the presence of measurement error. Section 4 outlines the algorithm employed in the study. Section 5 describes the simulation study and data analysis conducted based on the proposed estimation and testing procedures. Finally, Section 6 concludes the study with key findings.

## 2. Methodology

### 2.1. Corrected Score Estimation

The corrected score function is designed to address the estimation bias introduced by measurement errors, particularly in the explanatory variables. This approach involves modifying the conventional score function so that the resulting parameter estimates remain consistent despite the presence of measurement error. The score function itself is defined as the derivative of the log-likelihood function with respect to the parameters.

According to [8], a corrected score function is one whose expectation, taken with respect to the distribution of the measurement error, equals the conventional score function derived from the unobserved true explanatory variables. This property ensures that the corrected score function remains unbiased in expectation. The effectiveness of this method hinges on the existence of such a score function that satisfies this unbiasedness condition. Mathematically, the corrected score function satisfies the following expectation identity:

$$E[l^*(\boldsymbol{\beta}, X, Y)] = l(\boldsymbol{\beta}, Z, Y) \tag{1}$$

The likelihood function based on the true model is expressed as $L(\boldsymbol{\beta}, Z, Y)$, where $\boldsymbol{\beta}$ ={$\beta_0, \beta_1$} are the parameters to be estimated. In this context, Z represents the true explanatory variable, X is the observed explanatory variable affected by measurement error, and Y is the response variable. The corresponding log-likelihood function is given by $l(\boldsymbol{\beta}, Z, Y)$, and the associated score function is denoted by $U(\boldsymbol{\beta}, Z, Y)$. The corrected log-likelihood function, which accounts for measurement error, is expressed as $l^*(\boldsymbol{\beta}, X, Y)$, and it depends on the observed variables X and Y.

Assume that U represents the score function derived from the true model.

$$U(\boldsymbol{\beta}, Z, Y) = \frac{\partial l(\boldsymbol{\beta}, Z, Y)}{\partial \beta}, \ \beta = \{\beta_0, \beta_1\} \tag{2}$$

Thus, the corrected score function $U^*$ takes the form,

$$U^*(\boldsymbol{\beta}, X, Y) = \frac{\partial l^*(\boldsymbol{\beta}, X, Y)}{\partial \beta} \tag{3}$$

The value of $\beta_0$ and $\beta_1$ is obtained by solving the above score functions.

## 2.2. Weighted Simple Linear Regression

The Weighted Simple Linear Regression (WSLR) model is an extension of the standard simple linear regression model that accounts for heteroscedasticity (non-constant variance) in the error terms by assigning different weights to observations. This method is particularly useful when the variability of the response variable differs across levels of the explanatory variable, ensuring that observations with lower variance have greater influence on the estimated regression parameters.

The standard simple linear regression model is given by:

$$Y_t = \beta_0 + \beta_1 Z_t + \epsilon_t, \quad t = 1, 2, \ldots, n \tag{4}$$

where:

- $Y_t$ is the response variable,
- $X_t$ is the explanatory variable,
- $\beta_0$ and $\beta_1$ are the regression coefficients,
- $\epsilon_t$ represents the error term, which is assumed to have a constant variance ($\sigma_\epsilon^2$) in ordinary least squares (OLS).

However, when heteroscedasticity is present, the error variance varies across observations:

$$\text{Var}(\epsilon_t) = \sigma_{\epsilon t}^2 \quad t = 1, 2, \ldots, n \tag{5}$$

To address this, weighted least squares (WLS) is used, assigning a weight $w_t = \frac{1}{\sigma_\epsilon^2}$ to each observation. The weighted regression model is then:

$$Y_t = \beta_0 + \beta_1 Z_t + \epsilon_t, \quad \text{with } \epsilon_t \sim N(0, \sigma_\epsilon^2) \tag{6}$$

The weighted least squares (WLS) method minimizes the weighted sum of squared residuals:

$$S(\beta_0, \beta_1) = \sum_{i=1}^{n} w_t (Y_t - \beta_0 - \beta_1 Z_t)^2 \tag{7}$$

The estimated regression coefficients are:

$$\hat{\beta}_1 = \frac{\sum w_t (Z_t - \bar{Z}_w)(Y_t - \bar{Y}_w)}{\sum w_t (Z_t - \bar{Z}_w)^2} \tag{8}$$

$$\hat{\beta}_0 = \bar{Y}_w - \hat{\beta}_1 \bar{Z}_w \tag{9}$$

where $\bar{Z}_w$ and $\bar{Y}_w$ are the weighted means of $Z_t$ and $Y_t$, respectively.

## 3. Proposed Methodology

### 3.1. Estimation of Parameters

Parameter estimation for a simple linear regression model in the presence of measurement error was discussed by [13]. Examine a simple linear regression model

$$Y_t = \beta_0 + \beta_1 Z_t + \varepsilon_t \qquad t = 1, 2, ...n \tag{10}$$

Where $\varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon^2)$

Suppose $Z_t$ is subject to a measurement error, So the observed explanatory variable is given by,

$$X_t = Z_t + U_t \tag{11}$$

Where $U_t \sim \mathcal{N}(0, \sigma_U^2)$

Thus, the observed regression model incorporating measurement error in the explanatory variable is given by,

$$Y_t = \beta_0 + \beta_1 X_t - \beta_1 U_t + \varepsilon_t \tag{12}$$

The log likelihood function corresponding to the actual regression model is,

$$l(\boldsymbol{\beta}, Z, Y) = -\frac{1}{2} n \log(2\pi) - n \log(\sigma) + \frac{1}{2\sigma^2} \sum (Y_t - (\beta_0 + \beta_1 Z_t))^2 \tag{13}$$

The effectiveness of the corrected score method relies on the existence of a score function that satisfies the required property.

$$E[l^*(\boldsymbol{\beta}, X, Y)] = l(\boldsymbol{\beta}, Z, Y)$$

By using this property the obtained corrected log likelihood function is given as

$$l^*(\boldsymbol{\beta}, X, Y) = -\frac{1}{2n} \log(2\pi) - n \log(\sigma) - \frac{1}{2\sigma^2} \left( \sum (Y_t - (\beta_0 + \beta_1 X_t))^2 - \beta_1^2 \sigma_U^2 \right) \tag{14}$$

The Corrected score estimates can be obtained by solving the score functions $\frac{\partial l^*}{\partial \beta_0} = 0$ and $\frac{\partial l^*}{\partial \beta_1} = 0$. Thus corrected score estimates are

$$\hat{\beta_0} = \frac{\sum Y_t - \beta_1 \sum X_t}{n} \tag{15}$$

and

$$\hat{\beta_1} = \frac{\sum Y_t X_t - \frac{1}{n} \sum Y_t \sum X_t}{\sum X_t^2 - n\sigma_U^2 - \frac{(\sum X_t)^2}{n}} \tag{16}$$

The model parameters have been obtained using the corrected score procedure. Now, we incorporate the additional assumption of non-constant variance of $\varepsilon_t$.
Assume that $\varepsilon_t$ has non-constant variance, i.e.,

$$\varepsilon_t \sim \text{ARCH}(1)$$

$$\varepsilon_t = \sqrt{h_t}\, \eta_t \tag{17}$$

where

$$\eta_t \sim N(0, 1)$$

The variance of ARCH(1) is given by:

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 \tag{18}$$

where $\alpha_0 > 0$ and $\alpha_1 \geq 0$

Substitute this into the regression equation:

$$Y_t = \beta_0 + \beta_1 (X_t - U_t) + \sqrt{h_t}\,\eta_t \tag{19}$$

Expanding further,

$$Y_t = \beta_0 + \beta_1 X_t - \beta_1 U_t + \sqrt{h_t}\,\eta_t \tag{20}$$

We can estimate the parameters of eqn (20) by using weighted Generalized linear estimation. The weights ($w_t$) are inversely proportional to the variance of residuals

$$w_t = \frac{1}{h_t} \tag{21}$$

Substituting weights into eqn (19) we obtain,

$$\sqrt{w_t} Y_t = \sqrt{w_t}\beta_0 + \beta_1 \sqrt{w_t}(X_t - U_t) + \sqrt{w_t}\,\eta_t \tag{22}$$

The final obtained model is:

$$Y_t^* = \beta_0^* + \beta_1 (X_t - U_t)^* + \eta_t \tag{23}$$

where $Y_t^* = \sqrt{w_t} Y_t$ and $(X_t - U_t)^* = \sqrt{w_t}(X_t - U_t)$.

## 3.2. Testing of Measurement Error

To test the presence of measurement error in the regression model given in eqn (10) we can use the likelihood ratio test. The likelihood ratio test compares the goodness-of-fit of two nested models by evaluating the ratio of their likelihoods. This study aims to investigate the existence of measurement errors. To accomplish this, we examine the problem of validating the null hypothesis that there is no measurement error $H_0 : \sigma_U^2 = 0$ against the alternative hypothesis, which suggests the existence of measurement error $H_1 : \sigma_U^2 \neq 0$. The total variance of error in the model given in eqn (12) is expressed as $\beta_1^2 \sigma_U^2 + h_t$. The log-likelihood function is:

$$L = \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi(h_t + \beta_1^2 \sigma_U^2)}} \exp\left( -\frac{(Y_t - \beta_0 - \beta_1 X_t)^2}{2(h_t + \beta_1^2 \sigma_U^2)} \right) \tag{24}$$

On expanding logarithm:

$$L(H_1) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{n} \log(h_t + \beta_1^2 \sigma_U^2) - \frac{1}{2}\sum_{t=1}^{n} \frac{(Y_t - \beta_0 - \beta_1 X_t)^2}{h_t + \beta_1^2 \sigma_U^2} \tag{25}$$

Under the null hypothesis, $H_0 : \sigma_U^2 = 0$, we have:

$$\hat{\beta_0} = \frac{\sum Y_t - \beta_1 \sum X_t}{n} \tag{26}$$

$$\hat{\beta}_1^* = \frac{\sum Y_t X_t - \frac{1}{n}\sum Y_t \sum X_t}{\sum X_t^2 - \frac{1}{n}(\sum X_t)^2} \tag{27}$$

The log-likelihood under $H_0$ becomes:

$$L(H_0) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{n} \log(h_t) - \frac{1}{2}\sum_{t=1}^{n} \frac{(Y_t - \beta_0 - \beta_1^* X_t)^2}{h_t} \tag{28}$$

To test statistic of the Likelihood ratio test is given as,

$$\Lambda = -2\left(\ln L(H_0) - \ln L(H_1)\right) \tag{29}$$

Substituting the values on test statistic we obatin:

$$\begin{aligned}\Lambda = -2\Bigg[ & -\frac{n}{2}\log(2\pi) + \frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{n}\log(h_t) \\ & + \frac{1}{2}\sum_{t=1}^{n}\log(h_t + \beta_1^2\sigma_U^2) - \frac{1}{2}\sum_{t=1}^{n}\frac{(Y_t - \beta_0 - \beta_1^* X_t)^2}{h_t} \\ & + \frac{1}{2}\sum_{t=1}^{n}\frac{(Y_t - \beta_0 - \beta_1 X_t)^2}{h_t + \beta_1^2\sigma_U^2}\Bigg]\end{aligned} \tag{30}$$

Simplifying, we obtain:

$$\Lambda = \frac{1}{2}\sum_{t=1}^{n}\left[\log\left(\frac{h_t + \beta_1^2\sigma_U^2}{h_t}\right) - \frac{(Y_t - \beta_0 - \beta_1^* X_t)^2}{h_t} + \frac{(Y_t - \beta_0 - \beta_1 X_t)^2}{h_t + \beta_1^2\sigma_U^2}\right] \tag{31}$$

$$\begin{aligned}\Lambda = -2\Bigg[ & -\frac{1}{2}\sum_{t=1}^{n}\log(h_t) + \frac{1}{2}\sum_{t=1}^{n}\log(h_t + \beta_1^2\sigma_U^2) \\ & - \frac{1}{2}\sum_{t=1}^{n}\frac{(Y_t - \beta_0 - \beta_1^* X_t)^2}{h_t} + \frac{1}{2}\sum_{t=1}^{n}\frac{(Y_t - \beta_0 - \beta_1 X_t)^2}{h_t + \beta_1^2\sigma_U^2}\Bigg]\end{aligned} \tag{32}$$

Further Simplifying we obtain,

$$\begin{aligned}\Lambda = -\sum_{t=1}^{n}\Bigg[ & \log(h_t + \beta_1^2\sigma_U^2) - \log(h_t) \\ & + \frac{(Y_t - \beta_0 - \beta_1 X_t)^2}{h_t + \beta_1^2\sigma_U^2} - \frac{(Y_t - \beta_0 - \beta_1^* X_t)^2}{h_t}\Bigg]\end{aligned} \tag{33}$$

$$\begin{aligned}\Lambda = -\sum_{t=1}^{n}\Bigg[ & \frac{\log(h_t + \beta_1^2\sigma_U^2)}{h_t} - \frac{\log(h_t)}{h_t} \\ & + \frac{(Y_t - \beta_0 - \beta_1 X_t)^2}{h_t + \beta_1^2\sigma_U^2} - \frac{(Y_t - \beta_0 - \beta_1^* X_t)^2}{h_t}\Bigg]\end{aligned} \tag{34}$$

After simplification the final test statistic obtained is,

$$\begin{aligned}\Lambda = \sum_{t=1}^{n}\Bigg[ & \log\left(\frac{h_t}{h_t + \beta_1^2\sigma_U^2}\right) + \frac{(Y_t - \beta_0 - \beta_1^* X_t)^2}{h_t} \\ & - \frac{(Y_t - \beta_0 - \beta_1 X_t)^2}{h_t + \beta_1^2\sigma_U^2}\Bigg]\end{aligned} \tag{35}$$

## 4. Algorithm

The following is a step-by-step summary of the algorithm used to implement a simple linear regression model with arch errors in the presence of measurement errors:

**Step 1:** Specification of a model with measurement error in independent variable and non-constant variance in the random term.

$$Y_t = \beta_0 + \beta_1 Z_t + \varepsilon_t \tag{36}$$

**Step 2:** Give initial values for $\beta_0$, $\beta_1$ and generate Arch(1) errors.
**Step 3:** Apply simple linear regression model and estimate the parameters by using the corrected score estimation method.
**Step 4:** Obtain the residuals and confirm the heteroscedasticity in the error term of the regression model using estimates obtained from step 3.
**Step 5:** Estimate Arch(1) model parameters from residuals $\varepsilon_t$.
**Step 6:** For each $\varepsilon_t$, calculate conditional standard deviation $\sqrt{h_t}$.

$$\sqrt{h_t} = \sqrt{\alpha_0 + \alpha_1 \varepsilon_{t-1}^2} \tag{37}$$

**Step 7:** Find the weights using the eqn (21)
**Step 8:** Fixing the heteroscedasticity and the parameters are re-estimated by using weighted linear regression in the presence of measurement errors.
**Step 9:** Check for the assumption of constant variance for the residuals.
**Step 10:** Finally, compute the mean squared error (MSE) of the estimated parameters to assess the model's accuracy and also compare the estimates with the estimates obtained by the assumption of heteroscedasticity in the given model.
**Step 11:** After estimating the model parameters, the next step is to test whether there is a presence of measurement error in regression with volatile errors.
**Step 12:** Calculate the Likelihood ratio test statistic by using the eqn 26 and repeat the above steps 100 times
**Step 13:** The Proportion of rejections is computed as,

$$\text{Empirical Power (EP)} = \frac{(LRT > \chi_\alpha, i = 1, 2, ..., MC)}{MC} \tag{38}$$

## 5. Numerical Analysis

### 5.1. Simulation Study

#### 5.1.1 Estimation of Parameters

A simulation study was conducted to evaluate the performance of regression models with non-constant errors in the presence of measurement error using the model described below:

$$Y_t = \beta_0 + \beta_1 Z_t + \varepsilon_t \tag{39}$$

In this model, the error term $\varepsilon_t$ follows an Arch process of order 1. Specifically, we simulated ARCH(1) errors with $\alpha_0 = 1$ and $\alpha_1 = 0.9$. The model parameters used for the simulation were $\beta_0 = \{1, 2, 3\}$ and $\beta_1 = \{2, 3, 4\}$.

The parameters of the regression with ARCH(1) errors in the presence of measurement error is estimated using the corrected score estimation procedure. Simulations were conducted across varying sample sizes n = (300, 600,900,1200), with each configuration repeated over 100 iterations to obtain robust results. The table below presents the mean estimates of $\beta_0$ and $\beta_1$, along with their corresponding Mean Square Errors (MSE). This approach allowed us to assess the accuracy and precision of the parameter estimates for different model configurations under the influence of both measurement error and heteroscedasticity. The results obtained are summarized in the given below Table 1

**Table 1:** *The Average Estimates and Corresponding MSE of Corrected Score Estimation by with and without incorporating heteroscedasticity*

| Sample size | Incorporating Heteroscedasticity | | Without Incorporating Heteroscedasticity | |
|---|---|---|---|---|
| | $\hat{\beta_0}(MSE)$ | $\hat{\beta_1}(MSE)$ | $\hat{\beta_0}(MSE)$ | $\hat{\beta_1}(MSE)$ |
| | | $\beta_0 = 2, \beta_1 = 4$ | | |
| 300 | 1.9999(0.0031) | 4.0013(0.0034) | 2.2077(5.1745) | 3.9902 (3.0078) |
| 600 | 2.0009(0.0017) | 3.9978(0.0017) | 2.0826(5.2319) | 4.1340 (6.6574) |
| 900 | 2.0036(0.0013) | 3.9930(0.0013) | 2.0742(3.0544) | 4.0040 (4.6263) |
| 1200 | 1.9966(0.0007) | 3.9970(0.0008) | 1.7876(14.7981) | 4.3513 (6.3708) |
| | | $\beta_0 = 1, \beta_1 = 3$ | | |
| 300 | 0.9973(0.0038) | 3.0007(0.0029) | 0.4530(19.4281) | 2.9542 (0.9038) |
| 600 | 0.9952(0.0020) | 2.9974(0.0014) | 1.0460(1.7673) | 3.0616 (1.0953) |
| 900 | 0.9963(0.0010) | 2.9966(0.0013) | 1.2664(4.3442) | 2.8738 (6.1777) |
| 1200 | 1.0003(0.0009) | 2.9998(0.0008) | 1.4492(12.5775) | 3.2866 (4.5420) |
| | | $\beta_0 = 3, \beta_1 = 2$ | | |
| 300 | 3.0035(0.0032) | 2.0018(0.0040) | 2.9819(0.9208) | 1.7956 (7.1151) |
| 600 | 2.9910(0.0020) | 2.0014(0.0018) | 3.2916(3.0898) | 2.0660 (8.3690) |
| 900 | 2.9985(0.0011) | 1.9989(0.0010) | 2.7898(40.2556) | 1.4937 (18.5692) |
| 1200 | 2.9982(0.0008) | 2.0006(0.0010) | 2.7490(12.8247) | 1.5380 (10.8179) |

Table 1 gives the average estimates and corresponding MSE of both with and without incorporating heteroscedasticity. As the sample size increases, the MSE of estimates decreases. The estimates obtained by incorporating heteroscedasticity are consistently closer to the true parameter values compared to the estimates obtained without incorporating heteroscedasticity. The MSE is significantly lower when heteroscedasticity is incorporated, indicating more precise estimates. When heteroscedasticity is not incorporated, the estimates show higher bias, and the MSE increases, especially for smaller sample sizes. This highlights the negative impact of ignoring heteroscedasticity in the model. This indicates that in regression with non-constant variance and the presence of measurement error, incorporating heteroscedasticity leads to more accurate and reliable estimates.

### 5.1.2 Testing of measurement error

The model parameters are estimated by using both with and without incorporating heteroscedasticity in regression model in the presence of measurement error. After estimating the parameters, we have to analyze whether there is a presence of measurement error by using the Likelihood ratio test. In order to determine the power efficiency for the specified model, we conducted 1000 experiments and examined the test's power effectiveness for different sample sizes n = (200,500,1000,1500). We then averaged the results over the repetitions. R Studio was utilized to perform all of the computations.

The power of the LRT test by incorporating hetroscedasticity and measurement error in regression model at a 5% significance level are given in Table 2.

**Table 2:** *The power of the LRT test*

| | | | | | | |
|---|---|---|---|---|---|---|
| $n = 200$ | | | | | | |
| $\sigma_U$ | 0 | 0.01 | 0.05 | 0.1 | 0.5 | 1 |
| EP | 0 | 0 | 0.19 | 0.44 | 0.91 | 1 |
| $n = 500$ | | | | | | |
| $\sigma_U$ | 0 | 0.01 | 0.05 | 0.1 | 0.5 | 1 |
| EP | 0 | 0 | 0 | 0.71 | 0.99 | 1 |
| $n = 1000$ | | | | | | |
| $\sigma_U$ | 0 | 0.01 | 0.05 | 0.1 | 0.5 | 1 |
| EP | 0 | 0 | 0.11 | 0.51 | 1 | 1 |
| $n = 1500$ | | | | | | |
| $\sigma_U$ | 0 | 0.01 | 0.05 | 0.1 | 0.5 | 1 |
| EP | 0 | 0 | 0.31 | 0.7 | 1 | 1 |

From Table 2 it is seen that, as the sample size increases, the power of the LRT improves across all levels of $\sigma_U$. This indicates that larger sample sizes provide better detection of measurement error. The power of the LRT increases with higher values of $\sigma_U$. For smaller $\sigma_U$, the power is close to 0, particularly for smaller sample sizes. However, for larger $\sigma_U$ values, the power approaches 1, demonstrating the test's effectiveness in identifying measurement error when the variance of measurement error increases. The graphical representation of above Table 2 is given in Figure 1.

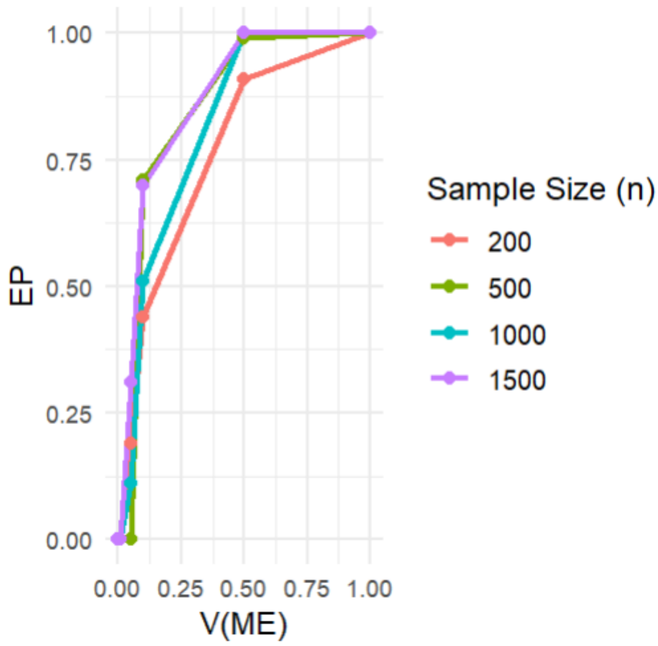


**Figure 1:** *Plot of EP of LRT test*

## 5.2. Data Analysis

The proposed methodology is applied to the Airquality dataset. The airquality dataset in R is a inbuilt dataset that contains daily air quality measurements collected in New York from May to September 1973. It consists of 153 observations and six variables: Ozone, Solar.R, Wind, Temp, Month, and Day. The Ozone variable represents the ozone concentration in parts per billion, while Solar.R measures solar radiation in Langley. The Wind variable records wind speed in miles per hour, and Temp represents temperature in degrees Fahrenheit. The Month and Day variables indicate the date of the recorded observation.

In our analysis, we have taken Ozone as the dependent variable and Wind as the independent variable. Initially, we fitted the ordinary linear regression model and checked the nonconstant variance using the bptest test. The p-value obtained for bptest is 0.005954, it shows the evidence of non constant variance. Then by using corrected score estimation we have find out the $\beta_0$ & $\beta_1$ parameters. Since there is a non constant variance in the residuals of corrected score estimation, we have fitted an ARCH(1) model to the residuals. Then we have fixed the hetroscedasticity and restimate the regression parameters using weighted linear regression model. Here our first objective is to test whether there exists a measurement error in this airquality data. The Likelihood ratio test statistic obtained is 409.1567 which is greater than the critical value 3.84 at 0.05 significance level. Since the test statistic value is greater than critical value, we reject the null hypothesis that there is no measurement error in this airquality data. That is by using LRT we have got that, there is exists some amount of measurement error in this airquality data.

Since there is a measurement error in this data, our next objective is to estimate the parameters of regression with nonconstant variance in the presence of measurement error. Initially we estimate the parameters using ordinary linear regression model and checked the nonconstant variance using the bptest test. The p-value obtained for bptest is 0.005954, it shows the evidence of non constant variance. The variance of measurement error is estimated by using newton raphson method and the obtained variance of measurement error is 1.597. Then by using corrected score estimation we have find out the $\beta_0$ & $\beta_1$ parameters. The obtained $\beta_0$ & $\beta_1$ parameters are 107.3532 & -6.565037. Since the residuals have nonconstant variance, we fit the Arch(1) model to it. Then, the regression coefficients are re-estimated by using weighted linear regression model. The $\beta_0$ & $\beta_1$ parameters obtained is 99.041 & -5.729.

The Box-Pierce test is used to check for autocorrelation in the residuals of a fitted model. In this case, the test was applied to the residuals of weighted linear regression model. Since the p-value (0.05179) is greater than 0.05, we fail to reject the null hypothesis at the 5% significance level. This suggests that there is no strong statistical evidence of autocorrelation in the residuals. The acf and pacf plot of residuals are given in Figure 2 and 3 simultaneously.

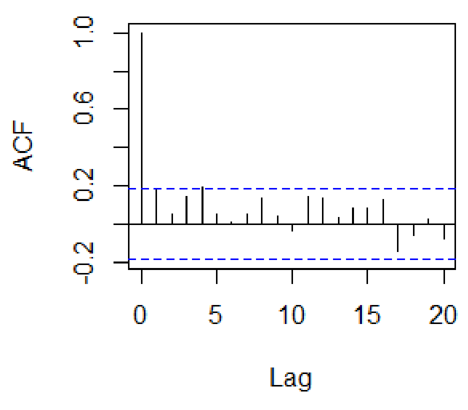


**Figure 2:** *ACF plot*

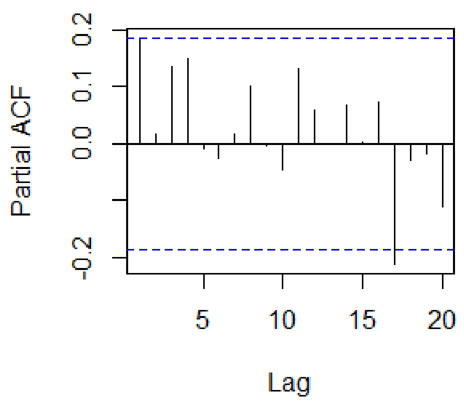


**Figure 3:** *PACF plot*

The Breusch-Pagan test is used to check for heteroscedasticity (non-constant variance) in the residuals of a regression model. Since the p-value (0.8099) is much greater than 0.05, we fail to reject the null hypothesis, which states that the residuals have constant variance (homoscedasticity). This suggests that there is no significant evidence of heteroscedasticity in the model, meaning that the assumption of constant variance is likely satisfied. The Kolmogorov-Smirnov (KS) test is used to check whether the residuals of the model follow a normal distribution. Since the p-value

(0.3549) is much greater than 0.05, we fail to reject the null hypothesis, which states that the residuals follow a normal distribution. This suggests no significant evidence against normality, meaning the residuals are approximately normally distributed. Hence, the normality assumption for the model appears to be satisfied. The normal QQ plot and histogram of residuals are given in Figure 4 and 5 simultaneously.

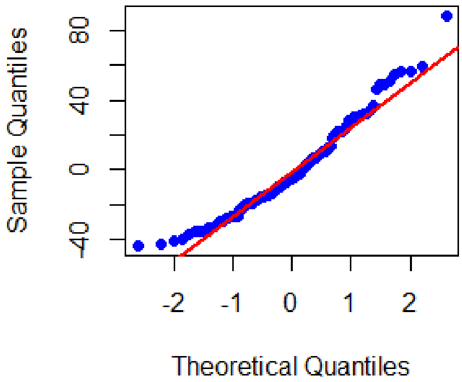


**Figure 4:** *QQ plot*

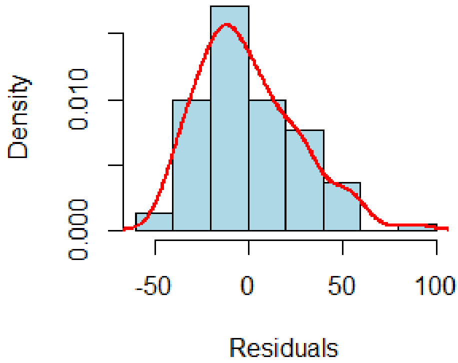


**Figure 5:** *Histogram*

## 6. Conclusion

This study addresses the challenges of estimation and testing in regression models with volatile errors when measurement errors are present. By incorporating corrected score estimation and weighted linear regression, we develop a robust estimation framework that ensures consistency and asymptotic normality of parameter estimates, even in the presence of heteroscedastic disturbances. The Likelihood Ratio (LR) test provides an effective approach to detecting measurement error and assessing its impact on regression models with volatile errors. Monte Carlo simulations demonstrate that ignoring hetroscedasticity and measurement error can lead to significant biases in estimation, while our proposed methods offer improved accuracy and efficiency. The application to real-life economic and financial data further validates the practical utility of the proposed approach. Overall, this study contributes to the broader literature on measurement error in regression with heteroscedastic disturbances, providing a reliable methodology for improving statistical inference.

## References

[1] Christos Agiakloglou. Evidence of arch (1) errors in the context of spurious regressions. *Communications in Statistics-Simulation and Computation*, 38(9):1803–1810, 2009.

[2] Anil K Bera and Xiao-Lei Zuo. Specification test for a linear regression model with arch process. *Journal of statistical planning and inference*, 50(2):283–308, 1996.

[3] Michel Broniatowski, Jana Jurečková, and Jan Kalina. Likelihood ratio testing under measurement errors. *Entropy*, 20(12):966, 2018.

[4] Wayne A Fuller. *Measurement error models*. John Wiley & Sons, 2009.

[5] S Hossain and Melody Ghahramani. Shrinkage estimation of linear regression models with garch errors. *Journal of Statistical Theory and Applications*, 15(4):405–423, 2016.

[6] John Mandel. Estimation of weighting factors in linear regression and analysis of variance. *Technometrics*, 6(1):1–25, 1964.

[7] Tsuyoshi Nakamura. Corrected score function for errors-in-variables models: Methodology and application to generalized linear models. *Biometrika*, 77(1):127–137, 1990.

[8] Tsuyoshi Nakamura. Corrected score function for errors-in-variables models: Methodology and application to generalized linear models. *Biometrika*, 77(1):127–137, 1990.

[9] Himanshu Pokhriyal and N Balakrishna. Testing for measurement error in regression with autoregressive innovations. *Communications in Statistics-Simulation and Computation*, 52(5):1834–1848, 2023.

[10] Mustafa Salamh and Liqun Wang. Identifiability and estimation of autoregressive arch models with measurement error. In *Advances and Innovations in Statistics and Data Science*, pages 235–255. Springer, 2012.

[11] Leonard A Stefanski. The effects of measurement error on parameter estimation. *Biometrika*, 72(3):583–592, 1985.

[12] O Sulaimon Mutiu. Application of weighted least squares regression in forecasting. *Int. J. Recent. Res. Interdiscip. Sci*, 2(3):45–54, 2015.

[13] Anna Thomas, Nimitha John, Lizan Meryl Pereira, and Himanshu Pokhriyal. Corrected score estimation of regression with autocorrelation under measurement errors. In *International Conference on Communication and Computational Technologies*, pages 179–189. Springer, 2024.