Journal of Probability and Statistical Science
24(2), 1-15Aug. 2026

Modeling of Sensitive Behavior Using Additive Data Masking: A Simulation-Based Approach for Privacy-Preserving Regression

Madiha
Quaid-i-Azam University, Pakistan

Muhammad Osama
Quaid-i-Azam University, Pakistan

Syed Muhammad Zeeshan
University of Peshawer, Pakistan

Muhammad Haseeb Ullah
Quaid-i-Azam University, Pakistan

ABSTRACT

It is difficult to model sensitive survey behaviors because of privacy concerns, stigma, or legal risks that prevent honest disclosure. This paper presents a new regression method that integrates additive data masking into Simple Linear Regression while maintaining privacy. In order to protect respondent privacy while maintaining reliable statistical inference, we use Warner's Randomized Response Technique to introduce additive noise into sensitive responses. We show theoretically that even when the dependent variable is masked, the modified regression estimator is still the Best Linear Unbiased Estimator (BLUE) under Gauss-Markov assumptions. A simulation study is conducted across various sample sizes and masking levels. The results from simulation study support the use of randomized masking techniques in sensitive question regression. This method improves data confidentiality without sacrificing the validity of inferences, which make it especially appropriate for studies in the social sciences, criminology, and public health where privacy is crucial. Despite the required privacy protection mechanism, the method guarantees unbiased estimates.

Keywords: Sensitive behavior; additive masking; randomized response technique; privacy preserving regression; BLUE; simulation study.

1. Introduction

In statistics, the Randomize Response Technique plays an important role for conducting accurate research on sensitive variables. To protect respondent anonymity and minimize bias, a binary the Randomize Response Technique design was used to estimate the percentage of a population with an abortion history Warner [1]. Compared to direct questioning, Randomize Response Technique showed less underreporting when used to survey cheating on university exams Greenberg et al. [2]. Estimated the prevalence of cocaine use using the Randomize Response Technique, confirming that the method can produce estimates that are higher and more reliable than those obtained from direct surveys Fox & Tracy [3]. According to a meta-analysis, Randomized Response Technique considerably raises self-reported rates of sensitive behavior when compared to direct questions Lensvelt-Mulders et al. [4]. Overcame significant social desirability bias by estimating the prevalence of undisclosed doping among elite athletes using the Crosswise Randomize Response Technique model Ulrich et al. [5]. Obtained more accurate prevalence estimates by surveying university students about plagiarism using the Unrelated Question Randomize Response Technique Jann et al. [6]. Improved the accuracy of data on HIV risk factors by using the Randomize Response Technique to examine sensitive sexual behaviors (such as having multiple partners) in China Yu et al. [7]. Found that the Crosswise model was the most successful in lowering misreporting when comparing Randomize Response Technique variations to estimate tax evasion rates Höglinger et al. [8]. Randomize Response Technique -

- Received May 24, 2026, in final form August 2026.
- Muhammad Osama (corresponding author) is affiliated with Quaid-i-Azam University, Pakistan
  Email: muhammadosama0846@gmail.com

gathered data on delicate subjects like corruption was used to develop statistical techniques for multivariate regression analysis Blair & Imai [9]. Measured unethical business practices using the Item Count Technique, a variation of the Randomized response technique, and found higher rates than those suggested by direct questions John et al. [10]. Randomized response technique with optional scrambling was used to measure the prevalence of workplace harassment, improving respondent cooperation in work environments Chaudhuri & Christofides [11] estimated Japan's illegal waste disposal rates using the Item Count Technique, a variation of the randomized response technique, and found that compliance gaps were greater than those reported by the government Tsuchiya et al. [12]. Anabolic steroid use among gym patrons was surveyed using the Crosswise Model, which produced noticeably higher usage estimates than direct questioning Petroczi et al. [13]. Randomized response technique was used to estimate immigration law violations in Germany, proving that it is feasible for politically sensitive policy research Coutts et al. [14]. Used Forced Response Randomized response technique to survey benefit fraud in the Netherlands, confirming that social security studies have less social desirability bias Lensvelt-Mulders & Boeije [15]. Randomized response technique was found to be better at minimizing over-reporting bias when compared to Computer-Assisted Self-Interviewing (CASI) for reporting voter turnout Holbrook & Kros nick [16]. Estimated tax non-compliance among independent contractors using Randomized response technique, which exposed systematic underreporting in conventional audits van der Heijden et al. [17]. Created the first methodological underpinnings for sensitive agri-data by pioneering randomized response technique in agricultural economics to survey illicit crop production Raghavarao & Federer [18]. Identified important demographic predictors while maintaining anonymity by using Crosswise RRT to evaluate COVID-19 protocol violations in Pakistan Hussain et al. [19]. RRT was used to measure post-disaster looting behavior following the Fukushima nuclear accident, capturing socially undesirable behaviors that cannot be reported directly Aoki & Sonobe [20]. Using Crosswise RRT, the prevalence of cyberbullying among German youth was measured Enzmann [21]. Measured hiring managers' covert racial bias using ICT Atsusaka & Stevenson [22]. Used randomized response technique to survey vaccine hesitancy in Pakistan and found that religiosity was a significant predictor Hussain & Qureshi [23].

The primary motivation for this research is to create and verify a privacy-preserving regression framework that ensures respondent confidentiality while allowing objective statistical inference on sensitive survey data. This study resolves the crucial conflict between respondent confidentiality and statistical validity in high-stakes fields like criminology and public health by developing a regression framework that protects privacy and allows for objective inference on sensitive survey data using Randomized Response Techniques.

2. Simple Linear Regression with Additive Masking

Simple linear regression model is:

$$Y = \beta_0 + \beta_1 X_{\mathrm{i}} + \varepsilon_{\mathrm{i}}$$

Warner model is:

$$Y = Z_i - w_{\mathrm{i}}$$

Where Z_i = Scrambled response, Y = Sensitive Variable, w_i = Scrambling variable substituting into the Simple linear regression model:

$$Z_{\mathrm{i}} - w_{\mathrm{i}} = \beta_0 + \beta_1 X_{\mathrm{i}} + \varepsilon_{\mathrm{i}}$$
$$Z_{\mathrm{i}} = \beta_0 + \beta_1 X_{\mathrm{i}} + \varepsilon_{\mathrm{i}} + w_{\mathrm{i}}$$

The estimated model is:

$$\hat{Z}_i = \hat{\beta}_0 + \hat{\beta}_1 \mathrm{X}_i$$
$$\because E(w_\mathrm{i}) = 0$$

2.1 Residual Sum of Square (RSS)

The Residual Sum of Squares (RSS) is defined as:

$$RSS = \sum_{i=1}^{n} (Z_\mathrm{i} - \hat{Z}_i)^2$$
$$\mathrm{t} = \sum_{i=1}^{n} (Z_\mathrm{i} - \hat{\beta}_0 - \hat{\beta}_1 \mathrm{X}_i)^2$$

To find the least-squares estimators, we minimize RSS by taking partial derivatives with respect the unknown parameters and setting them to zero. First differentiating w.r.t $\hat{\beta}_0$

$$\frac{\partial RSS}{\partial \beta_0} = -2 \sum_{i=1}^{n} \left(Z_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right) = 0$$

Yielding the first normal equation:

$$n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} X_i = \sum_{i=1}^{n} Z_i \qquad (1)$$

Now differentiating w.r.t $\hat{\beta}_1$

$$\frac{\partial RSS}{\partial \beta_0} = -2 \sum_{i=1}^{n} X_i \left(Z_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right) = 0$$

Yielding the second normal equation:

$$\hat{\beta}_0 \sum_{i=1}^{n} X_i + \hat{\beta}_1 \sum_{i=1}^{n} X_i^2 = \sum_{i=1}^{n} X_i Z_i \qquad (2)$$

The normal equations are

$$n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} X_i = \sum_{i=1}^{n} Z_i$$
$$\hat{\beta}_0 \sum_{i=1}^{n} X_i + \hat{\beta}_1 \sum_{i=1}^{n} X_i^2 = \sum_{i=1}^{n} X_i Z_i$$

In matrix form

$$\begin{bmatrix} n & \sum_{i=1}^{n} X_i \\ \sum_{i=1}^{n} X_i & \sum_{i=1}^{n} X_i^2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} Z_i \\ \sum_{i=1}^{n} X_i Z_i \end{bmatrix}$$

Solving for $\hat{\beta}_1$.

$$\hat{\beta}_1 = \frac{\left[\sum_{i=1}^{n} X_i Z_i - (\frac{1}{n}) \sum_{i=1}^{n} X_i \sum_{i=1}^{n} Z_i\right]}{\left[\sum_{i=1}^{n} X_i^2 - \left(\frac{1}{n}\right) (\sum_{i=1}^{n} X_i)^2\right]}$$

3. BLUE Property

In this section, we derive the BLUE property for the unknown parameters. First, we derive BLUE property for β_1.

3.1 Linearity

An estimator is said to be linear estimator for the corresponding parameter if it can be expressed as a linear function of the dependent variable Z that is , if the estimator is linearly related to the dependent variable Z.

Let

$$x_i = X_i - \bar{X} \text{ and } z_i = Z_i - \bar{Z}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i z_i}{\sum_{i=1}^{n} x_i^2} = \sum_{i=1}^{n} w_i z_i$$

where $w_i = \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} x_i^2}$ is a fixed weight. Therefore $\hat{\beta}_1$ is a linear estimator of β_1.

3.2 Unbiasedness

An estimator, say $\hat{\beta}_1$, is said to be unbiased estimator for the corresponding parameter β_1 $if\ E(\hat{\beta}_1) = \beta_1$. Since

$$\hat{\beta}_1 = \beta_1 + \sum_{i=1}^{n} w_i u_i$$

and using the properties $\sum_{i=1}^{n} w_i = 0, \sum_{i=1}^{n} w_i x_i = 1$, and $E(u_i) = 0$

$$E(\hat{\beta}_1) = \beta_1 + \sum_{i=1}^{n} w_i E(u_i) = \beta_1$$

$\hat{\beta}_1$ is an unbiased estimator of β_1.

3.3 Best (Minimum Variance)

The variance of $\hat{\beta}_1$ is derived as:

$$var(\hat{\beta}_1) = \sigma^2 \sum_{i=1}^{n} w_i^2 = \frac{\sigma^2}{\sum_{i=1}^{n} x_i^2}$$

For any other linear unbiased estimator $\beta_1^* = \sum_{i=1}^{n} d_i Z_i$ where $d_i = w_i + c_i$ with constraints $\sum_{i=1}^{n} c_i = 0$ and $\sum_{i=1}^{n} c_i x_i = 0$

$$var(\beta_1^*) = var(\hat{\beta}_1) + \sigma^2 \sum_{i=1}^{n} c_i^2 \geq var(\hat{\beta}_1)$$

Since $\sigma^2 \sum_{i=1}^{n} c_i^2 \geq 0$, we have $var(\beta_1^*) \geq var(\hat{\beta}_1)$. Therefore $\hat{\beta}_1$ has minimum variance among all linear unbiased estimators. $\hat{\beta}_1$ is the best estimator.

First, we derive BLUE property for β_0.

3.4 Linearity

The OLS estimator for β_0 is:

$$\hat{\beta}_0 = \bar{Z} - \hat{\beta}_1 \bar{X} = \sum_{i=1}^{n} \lambda_i Z_i$$

where $\lambda_i = \frac{1}{n} - \bar{X} w_i$. Hence $\hat{\beta}_0$ is a linear estimator of β_0.

3.5 Unbiasedness

Using $\sum_{i=1}^{n} \lambda_i = 1$ and $\sum_{i=1}^{n} \lambda_i x_i = 0$:

$$E(\hat{\beta}_0) = \beta_0 + \sum_{i=1}^{n} \lambda_i \, E(u_i) = \beta_0$$

$\hat{\beta}_0$ is an unbiased estimator for β_0.

3.6 Best (Minimum Variance)

$$Var(\hat{\beta}_0) = \sigma^2 \sum_{i=1}^{n} \lambda_i^2$$

For any other linear unbiased estimator $\beta_0^* = \sum_{i=1}^{n} e_i Z_i$ with $\sum_{i=1}^{n} e_i = 1$ and $\Sigma e_i x_i = 0$

$$var(\beta_0^*) = \text{var}(\hat{\beta}_0) + \sigma^2 \sum_{i=1}^{n} c_i^2 \geq \text{var}(\hat{\beta}_0)$$

$\hat{\beta}_0$ has minimum variance. $\hat{\beta}_0$ is the best estimator of β_0.

4. Simulation Study

The simulation study is conducted to examine how standard deviation (SD) and changing parameters affect bias and variance inflation in a regression model. In order to observe how more noise impacts bias (deviation from true parameters) and variance inflation (multicollinearity-induced instability), you first maintain the regression coefficients ($\beta 0 = 0.5$ and $\beta 1 = 2$) constant while varying the standard deviation (SD) of the error term. To evaluate the behavior of variance inflation and standard deviation (SD) metrics differently, we then alter β0 and β1 while keeping them constant. In order to determine whether stronger associations lessen or worsen estimation problems, we lastly alter $\beta 1$ (the slope coefficient) to examine how changes in the predictor's effect size interact with variance inflation and bias. By separating out parameter-specific effects, this simulation approach offers insights into the robustness of the model under various data-generating scenarios.

Table 1: Simulation Results Summary ($b_0 = 0.5$, $b_1 = 2$, $\sigma = 1$)

pmask n	**Bias in Slope**			**Variance Inflation**		
	50	100	500	50	100	500
0	-0.003	-0.001	-0.002	1	0.486	0.095
0.2	-0.871	-0.87	-0.889	1.525	0.738	0.146
0.4	-1.206	-1.223	-1.23	1.335	0.643	0.127
0.6	-1.408	-1.405	-1.411	1.149	0.545	0.107
0.8	-1.522	-1.523	-1.524	0.963	0.47	0.092
1	-1.603	-1.6	-1.598	0.859	0.406	0.079

Table 2: Simulation Results Summary ($b_0 = 0.5$, $b_1 = 2$, $\sigma = 7$)

$pmask$ n	**Bias in Slope**			**Variance Inflation**		
	50	100	500	50	100	500
0	-0.022	-0.01	-0.015	1	0.486	0.095
0.2	-0.909	-0.867	-0.89	0.592	0.283	0.055
0.4	-1.187	-1.218	-1.235	0.415	0.196	0.039
0.6	-1.395	-1.402	-1.416	0.316	0.151	0.03
0.8	-1.535	-1.525	-1.528	0.253	0.123	0.024
1	-1.617	-1.597	-1.593	0.222	0.102	0.02

Table 3: Simulation Results Summary ($b_0 = 0.5$, $b_1 = 2$, $\sigma = 15$)

$pmask$ n	**Bias in Slope**			**Variance Inflation**		
	50	100	500	50	100	500
0	-0.048	-0.021	-0.032	1	0.486	0.095
0.2	-0.959	-0.864	-0.892	0.577	0.275	0.053
0.4	-1.16	-1.213	-1.24	0.401	0.189	0.037
0.6	-1.377	-1.399	-1.422	0.302	0.145	0.028
0.8	-1.552	-1.529	-1.533	0.242	0.118	0.023
1	-1.637	-1.593	-1.586	0.211	0.097	0.019

Table 4: Simulation Results Summary ($b_0 = 0.5$, $b_1 = 2$, $\sigma = 22$)

$pmask$ n	**Bias in Slope**			**Variance Inflation**		
	50	100	500	50	100	500
0	-0.07	-0.031	-0.047	1	0.486	0.095
0.2	-1.002	-0.86	-0.893	0.575	0.274	0.053
0.4	-1.137	-1.208	-1.245	0.399	0.188	0.037
0.6	-1.362	-1.396	-1.427	0.3	0.144	0.028
0.8	-1.566	-1.532	-1.537	0.24	0.117	0.023
1	-1.654	-1.589	-1.58	0.21	0.096	0.019

Table 5: Simulation Results Summary ($b_0 = 0.5$, $b_1 = 2$, $\sigma = 29$)

$pmask$	**Bias in Slope**			**Variance Inflation**		
n	50	100	500	50	100	500
0	-0.093	-0.041	-0.062	1	0.486	0.095
0.2	-1.046	-0.857	-0.894	0.574	0.274	0.053
0.4	-1.115	-1.203	-1.25	0.398	0.187	0.037
0.6	-1.346	-1.393	-1.432	0.299	0.143	0.028
0.8	-1.581	-1.535	-1.542	0.239	0.117	0.023
1	-1.672	-1.585	-1.574	0.209	0.096	0.019

Table 6: Simulation Results Summary ($b_0 = 0.7$, $b_1 = 2$, $\sigma = 2$)

$pmask$	**Bias in Slope**			**Variance Inflation**		
n	50	100	500	50	100	500
0	-0.006	-0.003	-0.004	1	0.486	0.095
0.2	-0.877	-0.87	-0.89	0.812	0.39	0.076
0.4	-1.203	-1.222	-1.231	0.63	0.301	0.059
0.6	-1.406	-1.405	-1.412	0.513	0.243	0.048
0.8	-1.524	-1.523	-1.524	0.42	0.204	0.04
1	-1.605	-1.6	-1.598	0.371	0.173	0.034

Table 7: Simulation Results Summary ($b_0 = 1.2$, $b_1 = 2$, $\sigma = 2$)

$pmask$	**Bias in Slope**			**Variance Inflation**		
n	50	100	500	50	100	500
0	-0.006	-0.003	-0.004	1	0.486	0.095
0.2	-0.877	-0.87	-0.89	0.812	0.39	0.076
0.4	-1.203	-1.222	-1.231	0.63	0.301	0.059
0.6	-1.406	-1.405	-1.412	0.513	0.243	0.048
0.8	-1.524	-1.523	-1.524	0.42	0.204	0.04
1	-1.605	-1.6	-1.598	0.371	0.173	0.034

Table 8: Simulation Results Summary ($b_0 = 1.9$, $b_1 = 2$, $\sigma = 2$)

$pmask\ n$	**Bias in Slope**			**Variance Inflation**		
	50	100	500	50	100	500
0	-0.006	-0.003	-0.004	1	0.486	0.095
0.2	-0.877	-0.87	-0.89	0.812	0.39	0.076
0.4	-1.203	-1.222	-1.231	0.63	0.301	0.059
0.6	-1.406	-1.405	-1.412	0.513	0.243	0.048
0.8	-1.524	-1.523	-1.524	0.42	0.204	0.04
1	-1.605	-1.6	-1.598	0.371	0.173	0.034

Table 9: Simulation Results Summary ($b_0 = 2.6$, $b_1 = 2$, $\sigma = 2$)

$pmask\ n$	**Bias in Slope**			**Variance Inflation**		
	50	100	500	50	100	500
0	-0.006	-0.003	-0.004	1	0.486	0.095
0.2	-0.877	-0.87	-0.89	0.812	0.39	0.076
0.4	-1.203	-1.222	-1.231	0.63	0.301	0.059
0.6	-1.406	-1.405	-1.412	0.513	0.243	0.048
0.8	-1.524	-1.523	-1.524	0.42	0.204	0.04
1	-1.605	-1.6	-1.598	0.371	0.173	0.034

Table 10: Simulation Results Summary ($b_0 = 3.7$, $b_1 = 2$, $\sigma = 2$)

$pmask\ n$	**Bias in Slope**			**Variance Inflation**		
	50	100	500	50	100	500
0	-0.006	-0.003	-0.004	1	0.486	0.095
0.2	-0.877	-0.87	-0.89	0.812	0.39	0.076
0.4	-1.203	-1.222	-1.231	0.63	0.301	0.059
0.6	-1.406	-1.405	-1.412	0.513	0.243	0.048
0.8	-1.524	-1.523	-1.524	0.42	0.204	0.04
1	-1.605	-1.6	-1.598	0.371	0.173	0.034

Table 11: Simulation Results Summary ($b_0 = 4.0$, $b_1 = 2$, $\sigma = 2$)

$pmask$	**Bias in Slope**			**Variance Inflation**		
n	50	100	500	50	100	500
0	-0.006	-0.003	-0.004	1	0.486	0.095
0.2	-0.877	-0.87	-0.89	0.812	0.39	0.076
0.4	-1.203	-1.222	-1.231	0.63	0.301	0.059
0.6	-1.406	-1.405	-1.412	0.513	0.243	0.048
0.8	-1.524	-1.523	-1.524	0.42	0.204	0.04
1	-1.605	-1.6	-1.598	0.371	0.173	0.034

Table 12: Simulation Results Summary ($b_0 = 0.5$, $b_1 = 8$, $\sigma = 2$)

$pmask$	**Bias in Slope**			**Variance Inflation**		
n	50	100	500	50	100	500
0	-0.006	-0.003	-0.004	1	0.486	0.095
0.2	-3.472	-3.482	-3.557	4.376	2.13	0.426
0.4	-4.832	-4.892	-4.92	4.163	2.014	0.397
0.6	-5.636	-5.621	-5.643	3.689	1.756	0.343
0.8	-6.083	-6.09	-6.093	3.134	1.535	0.299
1	-6.406	-6.403	-6.396	2.807	1.338	0.261

The simulation results reveal that missing data has a substantial impact on regression analysis, particularly on the estimation of the slope parameter. As the proportion of missing observations increases, the estimated slope becomes increasingly biased downward, indicating that the true relationship between variables is systematically underestimated. This effect is more pronounced when the underlying relationship between the variables is stronger, suggesting that larger slope values are more vulnerable to distortion caused by missing data. Although variations in the error term slightly influence the magnitude of the bias, sample size appears to have little effect on correcting it. The findings further demonstrate that missing data contributes to variance inflation, reducing the stability and reliability of the regression estimates. Variance inflation is especially evident at lower to moderate levels of missingness and is amplified when the true slope is large. However, increasing the sample size helps to reduce this instability, indicating that larger datasets can partially compensate for the negative effects of missing observations. Higher levels of random error also tend to moderate the degree of inflation. In contrast, the intercept parameter remains relatively unaffected throughout the simulation. Changes in the intercept do not significantly alter either the slop bias or the variance inflation, showing that the main consequences of missing data are concentrated on slope estimation rather than the intercept. Overall, the study highlights that missing data can seriously compromise the accuracy and consistency of regression results, particularly in models with strong variable relationships, while larger sample sizes can help mitigate some of these effects.

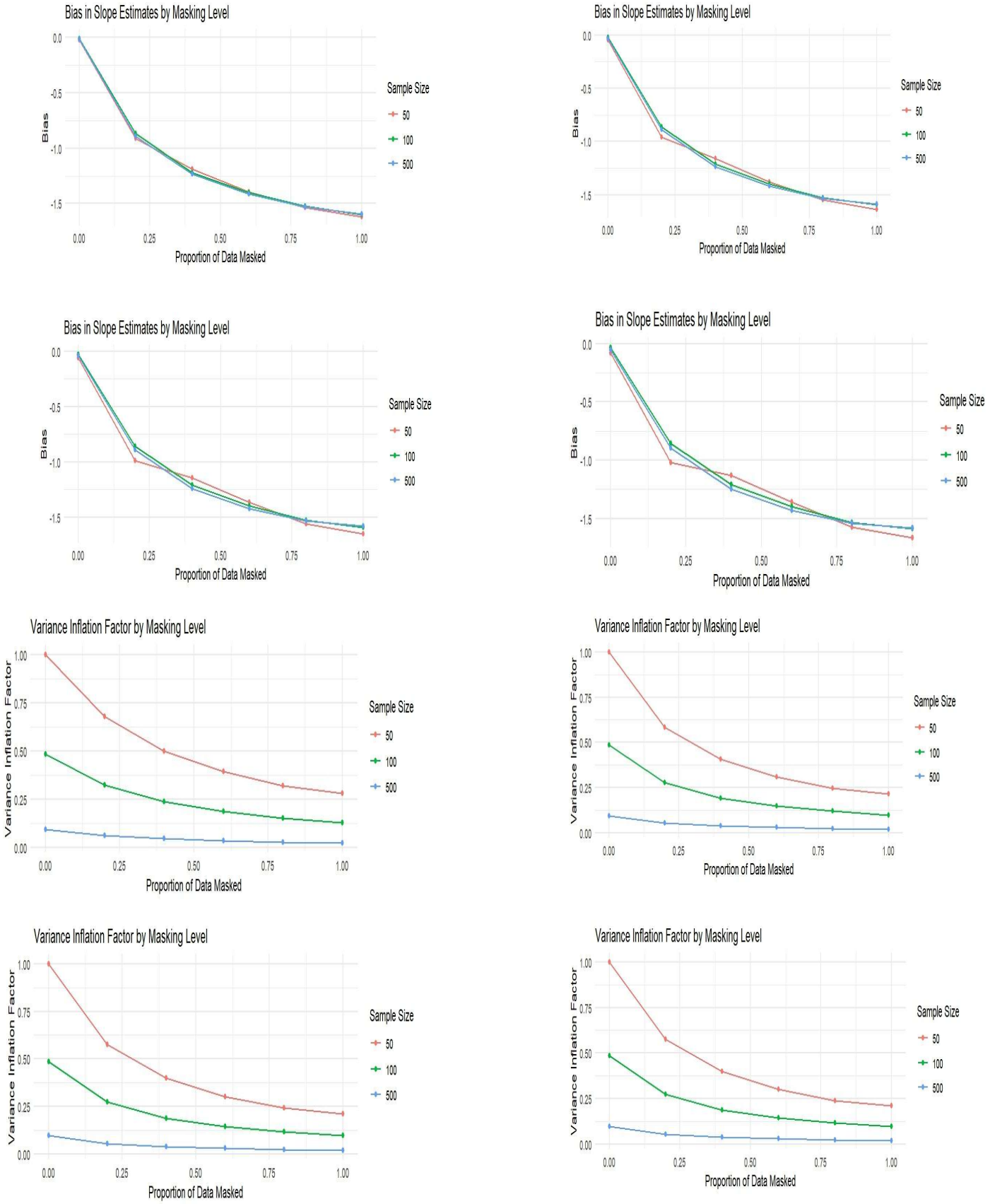


Figure 1. The graphical illustration of the Simulation Results.

5. Conclusion

This study demonstrates that standard linear regression produces biased estimates when analyzing sensitive dependent variables due to respondent misreporting, and proposes replacing the sensitive variable with a Warner randomized response to preserve privacy. Theoretical analysis confirms the modified estimator remains BLUE, and simulations study show that unbiased slope estimates despite introducing increased variance, whereas standard regression exhibits significant, uncorrectable negative bias that worsens with masking and is unaffected by larger sample sizes. The trade-off of higher variance for eliminating bias is preferable, advocating for a paradigm shift toward privacy-preserving Randomized Response Technique like this Warner-integrated regression for valid inference on sensitive data.

References

[1]. Aoki, S., & Sonobe, T. (2021). Measuring post-disaster looting behavior using the randomized response technique. *Journal of Economic Behavior & Organization*, 182, 1–15.
[2]. Atsusaka, Y., & Stevenson, R. T. (2023). Measuring covert racial bias among hiring managers using the item count technique. *Political Behavior*, 45(1), 1–24.
[3]. Blair, G., & Imai, K. (2012). Statistical analysis of list experiments. *Political Analysis*, 20(1), 47–77.
[4]. Chaudhuri, A., & Christofides, T. C. (2013). *Indirect questioning in sample surveys*. Springer.
[5]. Enzmann, D. (2021). Cyberbullying among German youth: Prevalence and correlates using the crosswise model. *Journal of Interpersonal Violence*, 36(13-14).
[6]. Fox, J. A., & Tracy, P. E. (1986). *Randomized response: A method for sensitive surveys*. Sage.
[7]. Greenberg, B. G., Abul-Ela, A. L. A., Simmons, W. R., & Horvitz, D. G. (1969). The unrelated question randomized response model: Theoretical framework. *Journal of the American Statistical Association*, 64(326), 520–539.
[8]. Höglinger, M., Jann, B., & Diekmann, A. (2016). Sensitive questions in online surveys. *Sociological Methods & Research*, 45(3), 516–541.
[9]. Hussain, Z., Qureshi, M. I., & Azeem, M. (2021). Measuring COVID-19 protocol violations using the crosswise randomized response technique in Pakistan. *PLOS ONE*, 16(6), e0253630.
[10]. Hussain, Z., & Qureshi, M. I. (2022). Vaccine hesitancy in Pakistan: A randomized response study. *Social Science & Medicine*, 292, 114552.
[11]. Jann, B., Jerke, J., & Krumpal, I. (2012). Asking sensitive questions using the crosswise model. *Public Opinion Quarterly*, 76(1), 32–49.
[12]. John, L. K., Loewenstein, G., Acquisti, A., & Vosgerau, J. (2018). When and why randomized response techniques (fail to) elicit the truth. *Organizational Behavior and Human Decision Processes*, 148, 101–123.
[13]. Kundu, P., & Bapat, V. (2021). Measuring police bribery using the item count technique: Evidence from India. *World Development*, 138, 105224.

[14]. Lensvelt-Mulders, G. J. L. M., Hox, J. J., van der Heijden, P. G. M., & Maas, C. J. M. (2005). Meta-analysis of randomized response research: Thirty-five years of validation. *Sociological Methods & Research*, 33(3), 319–348.
[15]. Raghavarao, D., & Federer, W. T. (1979). Block total response as an alternative to the randomized response method in surveys. *Journal of the Royal Statistical Society: Series B*, 41(1), 40–45.
[16]. Ulrich, R., et al. (2012). Doping in elite sports: The randomized response technique estimates a 30% prevalence. *PLOS ONE*, 7(7), e42056.
[17]. van der Heijden, P. G. M., van Gils, G., Bouts, J., & Hox, J. J. (2000). A comparison of randomized response, computer-assisted self-interview, and face-to-face direct questioning. *Sociological Methods & Research*, 28(4), 505–537.
[18]. Warner, S. L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, 60(309), 63–69.