On the Relative Efficiency of the Proposed Reparametized Randomized Response Model

Adepetun, A.O. and Adebola, F.B.
Department of Statistics, Federal University of Technology, PMB 704, Akure, Ondo State, Nigeria.
Phone: +2348038380448, +2347068110449
Email: akinolaoladiran@yahoo.com, femi_adebola@yahoo.com

Abstract

In this paper, we proposed a new reparametized Randomized Response Model by incorporating a third answer option “Undecided” into the Randomized Response Model developed by Hussain-Shabbir[6]. The relative efficiency as well as the variance of the newly proposed reparametized Randomized Response Model over the existing Randomized Response Model was established when data were obtained through the randomized response model proposed by Hussain and Shabbir [6].

Keywords: Reparametized Randomized Response Model (RRRM), Relative Efficiency, Privacy Protection, Stigmatized Characters, Randomization Device

1. Introduction

The problem of determination of the total population of a stigmatized quantitative variable is famous in sampling theory. Warner [11] was the first to put forward a popular method to determine the proportion of stigmatized characters such as an induced abortions, use of drugs etc., through a randomization device like a deck of cards, spinners etc. such that the respondents’ privacy should be protected. At present, Warner’s Randomized Response Model (RRM) has been extended by many researchers. Greenberg et al.[5], Mangat and Singh [8], Mangat [9], Singh et al.[10], Christofides [4], Kim and Warde [7], Adebola and Adepetun [1], Adebola and Adepetun [2], Adebola and Adepetun [3] are some of the many to be referenced. In sections to follow, we present the derivation of the existing Hussain and Shabbir’s Randomized Response Model, Proposed Reparametized Randomized Response Model and thereafter its relative efficiency over the existing one.

2. Derivation of Hussain and Shabbir’s Randomized Response Model

Hussain and Shabbir [6] put forward a Randomized Response Model (RRM) based on the random use of one of the two randomization devices \(X_1\) and \(X_2\). In design, the two randomization devices \(X_1\) and \(X_2\) are the same as that of Warner’s device but with different probabilities of choosing the stigmatize question. The basic idea behind this suggestion is to reduce considerably the suspicion among the respondents by providing them choice to randomly select the randomization device itself. Consequently, respondents may reveal their true status. A simple random sample with replacement (SRSW) sampling is assumed to choose a sample of size \(n\). Let \(\alpha\) and \(\beta\) be any two positive real numbers carefully selected such that \(q = \frac{\alpha}{\alpha + \beta}, (\alpha \neq \beta)\) is the probability of using \(X_1\), where \(X_1\) consists of the two statements of Warner’s device but with pre-assigned probabilities \(P_1\) and \(1 - P_1\) and \(1 - q = \frac{\beta}{\alpha + \beta}\) is the probability of using \(X_2\), where \(X_2\) consists of the two statements of Warner’s device also with pre-assigned probabilities \(P_2\) and \(1 - P_2\) respectively. For the \(i^{th}\) respondent, the probability of a “yes” response is given by

\[
P(\text{yes}) = \phi = \frac{\alpha}{\alpha + \beta}[P_1(1 - P_2)(1 - \pi)] + \frac{\beta}{\alpha + \beta}[P_2(1 - P_2)(1 - \pi)]
\]  

(2.1)
By expanding and simplifying equation (2.1), we have

\[ \varphi = \frac{\alpha [P_1 \pi + 1 - \pi - P_1 + P_1 \pi] + \beta [P_2 \pi + 1 - \pi - P_2 + P_2 \pi]}{\alpha + \beta} \]

\[ = \frac{2\alpha P_1 \pi - \alpha \pi + \alpha - P_1 \alpha + 2\beta P_2 \pi - \beta \pi + \beta - P_2 \beta}{\alpha + \beta} \]

\[ = \frac{\pi [2\alpha P_1 - \alpha + 2\beta P_2 - \beta] + \alpha + \beta - P_1 \alpha - P_2 \beta}{\alpha + \beta} \] (2.2)

Substituting \( P_1 = 1 - P_2 \) into the equation (2.2) in line with Hussain-Shabbir’s, we have

\[ \varphi = \frac{\pi [2\alpha (1 - P_2) - \alpha - \beta + 2\beta P_2] + \alpha + \beta - \alpha (1 - P_2) - P_2 \beta}{\alpha + \beta} \]

\[ = \frac{\pi [2\alpha - 2P_2 \alpha - \alpha - \beta + 2\beta P_2] + \alpha + \beta - \alpha + P_2 \alpha - P_2 \beta}{\alpha + \beta} \]

\[ = \frac{\pi [\alpha - \beta - 2P_2 \alpha + 2\beta P_2] + \beta + P_2 \alpha - P_2 \beta}{\alpha + \beta} \]

\[ = \frac{\pi [\alpha - \beta (1 - 2P_2)] + \beta + P_2 \alpha - P_2 \beta}{\alpha + \beta} \]

\[ = \frac{\pi [(\alpha - \beta)(1 - 2P_2)] + \beta + P_2 \alpha - P_2 \beta}{\alpha + \beta} \] (2.3)

To provide the equal privacy protection in both the randomization devices \( X_1 \) and \( X_2 \), we put \( P_2 = 1 - P_1 \) into the equation (2.3), obtained:

\[ \varphi = \frac{\pi [(\alpha - \beta)(1 - 2(1 - P_1))] + P_2 \alpha + P_1 \beta}{\alpha + \beta} \]

\[ = \frac{\pi [(\alpha - \beta)(2P_1 - 1)] + P_1 \beta + P_2 \alpha}{\alpha + \beta} \] (2.4)

Hence,

\[ \pi = \frac{\varphi (\alpha + \beta) - P_1 \beta - P_2 \alpha}{(2P_1 - 1)(\alpha - \beta)}, P_1 \neq 1/2, \alpha \neq \beta \] (2.5)

The unbiased moment estimator of true probability of yes response (response rate) \( \pi \) is given by

\[ \hat{\pi} = \frac{\hat{\varphi} (\alpha + \beta) - P_1 \beta - P_2 \alpha}{(2P_1 - 1)(\alpha - \beta)} \] (2.6)
where $\hat{\Theta} = \frac{Y}{n}$ and $y$ is the number of respondents reporting a “yes” answer.

When $P_1 = 1 - P_2$, the variance of the estimator is given then by

$$\text{Var}(\hat{r}) = \sqrt{\frac{\hat{\Theta}(\alpha + \beta) - P_1\beta - P_2\alpha}{(2P_1 - 1)(\alpha - \beta)}} = \frac{(\alpha + \beta)^2(\hat{\Theta})}{n(2P_1 - 1)^2(\alpha - \beta)^2}$$

(2.7)

$$\phi^2 = \sqrt{\frac{2\alpha P_1\pi - \alpha P_1 + \alpha - P_1\alpha}{\alpha + \beta}}$$

$$= \frac{\alpha + \beta}{2\alpha P_1\pi - \alpha P_1 + \alpha - P_1\alpha}$$

Further simplification of the above equation gives

$$\phi^2 = \frac{4\alpha^2 P_1^2\pi^2 - 2\alpha^2 P_1\pi^2 + 2\alpha^2 P_1\pi - 2\alpha^2 P_1^2\pi + 4\alpha \beta P_1^2\pi^2 - 2\alpha^2 P_1^2\pi}{\alpha + \beta}$$

(2.8)

$$Hence, we have$$

$$\phi^2 = \frac{4\alpha^2 P_1^2\pi^2 + 2\alpha^2 P_1\pi^2 + 2\alpha^2 P_1\pi - 4\alpha \beta P_1^2\pi^2 + 4\alpha \beta P_1^2\pi^2}{\alpha + \beta}$$

Further simplification of the above equation gives

$$\phi^2 = \frac{-4\alpha^2 P_1^2\pi^2 + 4\alpha^2 P_1\pi^2 + 4\alpha^2 P_1\pi^2 - 8\alpha \beta P_1 P_2^2\pi^2 + 4\alpha \beta P_1^2\pi^2}{\alpha + \beta}$$

(2.9)

Hence, we have
Factoring and setting $P_2 = 1 - P_1$ in the numerator of the second term of the above equation, we have

$$(P_2\alpha + P_1\beta)(P_1\alpha + P_2\beta) = \alpha^2P_1 - \alpha^2P_1^2 - 2\alpha\beta P_1 P_2 + \alpha\beta P_1 + \alpha\beta P_2 + \beta^2P_2 - \beta^2P_2^2 \quad (*)$$

Hence, we have

$$V(\hat{\alpha})_{conv} = \frac{\pi(1 - \pi)}{n} + \frac{(P_2\alpha + P_1\beta)(P_1\alpha + P_2\beta)}{n(2P_1 - 1)^2(\alpha - \beta)^2(\alpha + \beta)^2}$$

**3. The Proposed Reparametrized Randomized Response Model**

Despite the success achieved by many authors in developing an efficient Randomized Response Models (RRMs), the developed Models only considered a dichotomous option of “yes” and “no” response. In view of this, we propose a new Reparametrized Randomized Response Model (RRM) that will be based on the random use of one of the three randomization devices, $X_1, X_2$, and $X_3$. In design, the three randomization devices $X_1, X_2$, and $X_3$ are identical to that of Warner’s device but with different probabilities of selection. In addition to and proposed earlier by Hussain and Shabbir, we introduce $\delta$, a positive real number such that $\frac{\alpha}{\alpha + \beta + \delta}$, $\alpha \neq \beta \neq \delta$ is the probability of using $X_1$, where $X_1$ consists of the two statements of Warner’s device and the new introduce device also with preset probabilities $P_1, P_2$, and $P_3$ respectively. By using Hussain and Shabbir’s probability of a “yes” response for the $i^{th}$ respondent, the probability of a “yes” response when the third option “undecided” is included is given by

$$Q(\text{yes}) = \varphi = \frac{\alpha}{\alpha + \beta + \delta}[P_1\pi + (1 - P_1)(1 - \pi)] + \frac{\beta}{\alpha + \beta + \delta}[P_2\pi + (1 - P_2)(1 - \pi)] + \frac{\delta}{\alpha + \beta + \delta}[P_3\pi + (1 - P_3)(1 - \pi)] \quad (3.1)$$

In order to ensure equal privacy protection in the three randomization devices $X_1, X_2$, and $X_3$, we put $P_1 = 1 - P_2 - P_3$ into equation (3.1), to get

$$\pi = \varphi(\alpha + \beta + \delta) - [(\alpha + \beta + \delta) - P_1\alpha - P_2\beta - P_3\delta] \quad (3.2)$$

Hence, the unbiased sample estimate of $\pi$ is given as
\[ \hat{\phi} = \frac{\hat{\phi}(\alpha + \beta + \delta) - [(\alpha + \beta + \delta) - P_1\alpha - P_2\beta - P_3\delta]}{2P_1\alpha + 2P_2\beta + 2P_3\delta - \alpha - \beta - \delta} \]  

(3.3)

Where \( \hat{\phi} = \frac{x}{n} \) and \( x \) is the number of respondents reporting a “yes” answer when \( P_1 = 1 - P_2 - P_3 \). The variance of the estimator is given then by

\[ V(\hat{\phi}) = \frac{\pi(4\alpha^2P_1^2 - 4\alpha^2P_1 - 4\alpha\beta P_1 + 8\alpha\beta P_2 - 4\alpha P_1)}{n[2P_1(\alpha - \delta) + 2P_2(\beta - \delta) - (\alpha + \beta - \delta)]^2(\alpha + \beta + \delta)^2} \]

(3.4)

On simplification of equation (3.4), we have

\[ V(\hat{\phi})_{prop} = \frac{\pi(1 - \pi)}{n} + \frac{(P_1\alpha + P_2\beta + P_3\delta)(P_1\alpha + P_2\beta + P_3\delta)}{n[2P_1(\alpha - \delta) + 2P_2(\beta - \delta) - (\alpha + \beta - \delta)]^2(\alpha + \beta + \delta)^2} \]

(3.5)

4. **Relative Efficiency Comparison**

Here, we show that the new Reparametized Randomized Response Model (RRRM) is more efficient than the existing one via both relative efficiency and variance approach by adopting the data used by Hussain and Shabbir [6]. In what follows, the proposed Reparametized Randomized Response Model (RRRM) is more efficient than Hussain and Shabbir [6] dichotomous Randomized Response Model (RRM) if we have

\[ \text{Relative Efficiency (RE)} = \frac{\text{Variance of the proposed (RRRM)}}{\text{Variance of the existing (RRM)}} < 1 \]  

(4.1)

Or if

\[ \text{RE} = \frac{\pi(1 - \pi)}{n} + \frac{(P_1\alpha + P_2\beta + P_3\delta)(P_1\alpha + P_2\beta + P_3\delta)}{n[2P_1(\alpha - \delta) + 2P_2(\beta - \delta) - (\alpha + \beta - \delta)]^2(\alpha + \beta + \delta)^2} < 1 \]

(4.2)

The condition given in (4.2) is true, for \( P_1, P_2, P_3, \) and \( \pi \) ranging from 0.1 to 0.9 if \( \alpha \) and \( \beta, \alpha \) and \( \beta, \delta \) and \( \beta \) and \( \delta \) differ from each other by at least 9 where \( \alpha, \beta, \) and \( \delta \) are any suitable real numbers.
Table 4.1: Comparison between existing RRM and proposed RRRM when $P_1 = 0.6$, $P_2 = 0.3$, $P_3 = 0.1$, $\pi = 0.5$, $\alpha = 20$, $\beta = 11$, $\delta = 2$, for varying sample sizes (n)

<table>
<thead>
<tr>
<th>n</th>
<th>$P_1$</th>
<th>$P_2$</th>
<th>$P_3$</th>
<th>$\pi$</th>
<th>$\alpha$</th>
<th>$\beta$</th>
<th>$\delta$</th>
<th>Conventional Variance (RRM) (eqn. 2.10)</th>
<th>Proposed Variance (RRRM) (eqn.3.5)</th>
<th>Relative Efficiency (RE) (eqn.4.1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>0.6</td>
<td>0.3</td>
<td>0.1</td>
<td>0.5</td>
<td>20</td>
<td>11</td>
<td>2</td>
<td>0.00624</td>
<td>0.00546</td>
<td>0.87500</td>
</tr>
<tr>
<td>100</td>
<td>0.6</td>
<td>0.3</td>
<td>0.1</td>
<td>0.5</td>
<td>20</td>
<td>11</td>
<td>2</td>
<td>0.00312</td>
<td>0.00273</td>
<td>0.87500</td>
</tr>
<tr>
<td>150</td>
<td>0.6</td>
<td>0.3</td>
<td>0.1</td>
<td>0.5</td>
<td>20</td>
<td>11</td>
<td>2</td>
<td>0.00208</td>
<td>0.00182</td>
<td>0.87500</td>
</tr>
<tr>
<td>200</td>
<td>0.6</td>
<td>0.3</td>
<td>0.1</td>
<td>0.5</td>
<td>20</td>
<td>11</td>
<td>2</td>
<td>0.00156</td>
<td>0.00136</td>
<td>0.87179</td>
</tr>
<tr>
<td>500</td>
<td>0.6</td>
<td>0.3</td>
<td>0.1</td>
<td>0.5</td>
<td>20</td>
<td>11</td>
<td>2</td>
<td>0.000624</td>
<td>0.000546</td>
<td>0.87500</td>
</tr>
</tbody>
</table>

Table 4.2: Comparison between existing RRM and proposed RRRM when $P_1 = 0.4$, $P_2 = 0.4$, $P_3 = 0.2$, $\pi = 0.4$, $\alpha = 20$, $\beta = 11$, $\delta = 2$, for varying sample sizes (n)

<table>
<thead>
<tr>
<th>n</th>
<th>$P_1$</th>
<th>$P_2$</th>
<th>$P_3$</th>
<th>$\pi$</th>
<th>$\alpha$</th>
<th>$\beta$</th>
<th>$\delta$</th>
<th>Conventional Variance (RRM) (eqn.2.10)</th>
<th>Proposed Variance (RRRM) (eqn.3.5)</th>
<th>Relative Efficiency (RE) (eqn.4.1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>0.4</td>
<td>0.4</td>
<td>0.2</td>
<td>0.4</td>
<td>20</td>
<td>11</td>
<td>2</td>
<td>0.00579</td>
<td>0.00484</td>
<td>0.83592</td>
</tr>
<tr>
<td>100</td>
<td>0.4</td>
<td>0.4</td>
<td>0.2</td>
<td>0.4</td>
<td>20</td>
<td>11</td>
<td>2</td>
<td>0.00289</td>
<td>0.00242</td>
<td>0.83737</td>
</tr>
<tr>
<td>150</td>
<td>0.4</td>
<td>0.4</td>
<td>0.2</td>
<td>0.4</td>
<td>20</td>
<td>11</td>
<td>2</td>
<td>0.00193</td>
<td>0.00161</td>
<td>0.83420</td>
</tr>
<tr>
<td>200</td>
<td>0.4</td>
<td>0.4</td>
<td>0.2</td>
<td>0.4</td>
<td>20</td>
<td>11</td>
<td>2</td>
<td>0.00145</td>
<td>0.00121</td>
<td>0.83448</td>
</tr>
<tr>
<td>500</td>
<td>0.4</td>
<td>0.4</td>
<td>0.2</td>
<td>0.4</td>
<td>20</td>
<td>11</td>
<td>2</td>
<td>0.000579</td>
<td>0.000484</td>
<td>0.83592</td>
</tr>
</tbody>
</table>
Table 4.3: Comparison between existing RRM and proposed RRRM when \( P_1 = 0.2, \)

\( P_2 = 0.5, P_3 = 0.3, \pi = 0.7, \alpha = 20, \beta = 11, \delta = 2, \) for varying sample sizes (n)

<table>
<thead>
<tr>
<th>n</th>
<th>( P_1 )</th>
<th>( P_2 )</th>
<th>( P_3 )</th>
<th>( \pi )</th>
<th>( \alpha )</th>
<th>( \beta )</th>
<th>( \delta )</th>
<th>conventional variance (RRM) (eqn.2.10)</th>
<th>Proposed Variance (RRRM) (eqn.3.5)</th>
<th>Relative Efficiency (RE) (eqn.4.1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>0.2</td>
<td>0.5</td>
<td>0.3</td>
<td>0.7</td>
<td>20</td>
<td>11</td>
<td>2</td>
<td>0.00428</td>
<td>0.00421</td>
<td>0.98364</td>
</tr>
<tr>
<td>100</td>
<td>0.2</td>
<td>0.5</td>
<td>0.3</td>
<td>0.7</td>
<td>20</td>
<td>11</td>
<td>2</td>
<td>0.00214</td>
<td>0.00211</td>
<td>0.98598</td>
</tr>
<tr>
<td>150</td>
<td>0.2</td>
<td>0.5</td>
<td>0.3</td>
<td>0.7</td>
<td>20</td>
<td>11</td>
<td>2</td>
<td>0.00143</td>
<td>0.00140</td>
<td>0.97902</td>
</tr>
<tr>
<td>200</td>
<td>0.2</td>
<td>0.5</td>
<td>0.3</td>
<td>0.7</td>
<td>20</td>
<td>11</td>
<td>2</td>
<td>0.00107</td>
<td>0.00105</td>
<td>0.98131</td>
</tr>
<tr>
<td>500</td>
<td>0.2</td>
<td>0.5</td>
<td>0.3</td>
<td>0.7</td>
<td>20</td>
<td>11</td>
<td>2</td>
<td>0.000428</td>
<td>0.000421</td>
<td>0.98364</td>
</tr>
</tbody>
</table>

Table 4.4: Comparison between existing RRM and proposed RRRM when \( P_1 = 0.15, \)

\( P_2 = 0.6, P_3 = 0.25, \pi = 0.8, \alpha = 20, \beta = 11, \delta = 2, \) for varying sample sizes (n)

<table>
<thead>
<tr>
<th>n</th>
<th>( P_1 )</th>
<th>( P_2 )</th>
<th>( P_3 )</th>
<th>( \pi )</th>
<th>( \alpha )</th>
<th>( \beta )</th>
<th>( \delta )</th>
<th>conventional variance (RRM) (eqn.2.10)</th>
<th>Proposed Variance (RRRM) (eqn.3.5)</th>
<th>Relative Efficiency (RE) (eqn.4.1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>0.15</td>
<td>0.6</td>
<td>0.25</td>
<td>0.8</td>
<td>20</td>
<td>11</td>
<td>2</td>
<td>0.00327</td>
<td>0.00321</td>
<td>0.98165</td>
</tr>
<tr>
<td>100</td>
<td>0.15</td>
<td>0.6</td>
<td>0.25</td>
<td>0.8</td>
<td>20</td>
<td>11</td>
<td>2</td>
<td>0.00163</td>
<td>0.00161</td>
<td>0.98773</td>
</tr>
<tr>
<td>150</td>
<td>0.15</td>
<td>0.6</td>
<td>0.25</td>
<td>0.8</td>
<td>20</td>
<td>11</td>
<td>2</td>
<td>0.00109</td>
<td>0.00107</td>
<td>0.98165</td>
</tr>
<tr>
<td>200</td>
<td>0.15</td>
<td>0.6</td>
<td>0.25</td>
<td>0.8</td>
<td>20</td>
<td>11</td>
<td>2</td>
<td>0.00082</td>
<td>0.000803</td>
<td>0.97927</td>
</tr>
<tr>
<td>500</td>
<td>0.15</td>
<td>0.6</td>
<td>0.25</td>
<td>0.8</td>
<td>20</td>
<td>11</td>
<td>2</td>
<td>0.000327</td>
<td>0.000321</td>
<td>0.98165</td>
</tr>
</tbody>
</table>
Figure 4.1: Comparison between existing RRM and proposed RRRM when $P_1 = 0.6,$

$P_2 = 0.3, P_3 = 0.1, \pi = 0.5, \alpha = 20, \beta = 11, \delta = 2,$ for varying sample sizes ($n$)

Figure 4.2: Comparison between existing RRM and proposed RRRM when $P_1 = 0.4,$

$P_2 = 0.4, P_3 = 0.2, \pi = 0.4, \alpha = 20, \beta = 11, \delta = 2,$ for varying sample sizes ($n$)
Figure 4.3: Comparison between existing RRM and proposed RRRM when $P_1 = 0.2$,

$P_2 = 0.5, P_3 = 0.3, \pi = 0.7, \alpha = 20, \beta = 11, \delta = 2$, for varying sample sizes $(n)$

Figure 4.4: Comparison between existing RRM and proposed RRRM when $P_1 = 0.15$,

$P_2 = 0.6, P_3 = 0.25, \pi = 0.8, \alpha = 20, \beta = 11, \delta = 2$, for varying sample sizes $(n)$

Note:

$\text{Var}(\pi)$ in the figures above represents variance for both existing and proposed Models as obtained in equations 2.10 and 3.5 respectively.
…….conventional variance (equation 2.10)

_____proposed variance (equation 3.5)

5. Conclusion
In this study, the derivation of Hussain and Shabbir’s Randomized Response Model was presented. The relative efficiency as well as the variance of our proposed Reparametized Randomized Response Model (RRRM) over that of the existing one was established by adopting the data used by Hussain and Shabbir [6]. It was obvious in the results on Tables and Figures that the proposed Model is indeed more efficient than the existing one.

References