Journal of Policy Research, 8(4), 447-461.
https://jprpk.com
https://doi.org/10.5281/zenodo.7726022

A Comparative Study of Modified Hidden Logits Using Randomized Response Techniques

Asma Halim[1], Irshad Ahmad Arshad[2], Summaira Haroon[3], Waqas Shair[4]

Abstract
The survey sampling is one of the driving and most extensively used technique to collect the data about individual's behaviors, beliefs, views and opinions on a certain matter or topic. We aim to acquire flawless and reliable responses while collecting data. This aim is not achieved in such cases, when we are dealing with sensitive or socially stigmatized variables. Frequently respondents give elusive or false or non-responses about sensitive questions. In such sensitive or stigmatized characteristics, we use randomized response techniques (RRT). In current article using Mangat and Singh (1990) randomized response model, a modified hidden logit estimation procedure is presented. The proposed logit estimation procedure is also compared with ordinary logits and Corstange (2004) randomized response model. We detect that modified hidden logit estimates for Mangat and Singh (1990) are closer to the true parametric values as compare to the higher values of p and T and show elevated precision. The akaike and schwarz information criterion are renowned measures to model selection that favors more parsimonious models over more complex models. This study is also conducted for checking best model selection. This paper has a great contribution towards application and estimation of logistic models when sensitive or stigmatized issues are under consideration.
Keywords: Logistic Regression, Sensitive Issues, Randomized Response, Hidden Logit Estimation, Sensitive Characteristics

1. Introduction

1.1. Survey Sampling

Survey sampling is broadly used procedure in different public or private sectors for data collection. The data collection is the main step in Survey Sampling. Data is collected regarding certain characteristic of population which can be qualitative or quantitative in nature. Important point to be noted is that the main thing we aspire in data collection is to get accurate, consistent and reliable results. To get reliable or accurate findings is not a problem when we are working on common or general topics but it becomes thorny to get hold of true responses when we are bearing some personal issues for survey, in mind. If our characteristic of interest is the presence or absence of socially undesirable characteristic then respondent wants to keep privacy and is always unwilling to share his/her personal information. Such a variable or characteristic is considered as sensitive or stigmatized. When survey consists of highly private and personal questions, the respondents feel reluctant to provide accurate information and our survey yields erroneous, incorrect and fuzzy results.
In Survey Sampling, it is obvious that there arises a bias, while we are dealing with sensitive variables. This type of bias is particularly referred as Social Desirability Bias (SDB) when variable of interest makes respondents socially stigmatized. SDB is stated as a bias which arises when respondent conceals or hides the true response of a sensitive or highly controversial issue due to the fear that if he/she discloses the right information so that make him/her socially undesirable or stigmatized. Sensitive issues can be like frauds, use of illegal drugs or alcoholic beverages, extra marital affairs, illegal income, evading income tax and a savings in forms of prize bonds. In such cases we require a better technique for truthful data collection, where respondent feel confident in reporting his answers. In such cases most widely used technique is randomized response.

1.2. Randomized Response (RR)

It is clear from all above examples and discussions that direct questioning technique for obtaining information regarding sensitive issues will end us up into inaccurate and false results and sometimes we have on-responses. For the above discussed example of sensitive issues, the direct questioning technique will be totally useless because in all these cases respondents will be reluctant to provide accurate information regarding specific sensitive issues due to fear of being penalized by authority or fear of being socially stigmatized. So, finally respondent either refuse to answer or provide intentionally false answer, which results in biased estimates. To reduce this response bias, there should be some other survey methodology which has the strong potential to handle these sensitive variables.
We know that the possession of sensitive attribute results in biased and inaccurate results and our usual survey methods which are applied in case of un-harmful / innocuous questions are not appropriate to apply in case of sensitive issues and if applied then there is a great chance of un-true results. So, it was desired to find some other methods that raise the true response rate and reduce the response bias. To satisfy this need analysts chipped away at finding some strategy, which will be effectively managing sensitive issues. Perhaps the most popular strategy among these techniques is randomized response technique. Randomized response (RR) is a survey technique used to gather information about sensitive issues and the aim is to protect the privacy of respondent, which in result reduce non-response or intentionally misreporting. More formally it can

[1]Corresponding Author, Ph. D Scholar, Allama Iqbal Open University, Islamabad, Pakistan; Assistant Professor, Department of Business Administration, Iqra University, Islamabad, Pakistan
[2]Professor, Department of Statistics, Allama Iqbal Open University, Islamabad, Pakistan
[3] Senior Lecturer, Department of Business Studies, Bahria University, Islamabad, Pakistan
[4]Lecturer, Minhaj University Lahore, Pakistan

be defined as a technique which reduces the response error in survey of sensitive issues by providing respondents foolproof guarantees of anonymity.

Warner (1965) is pioneer, who introduced RR procedure to deal with sensitive variables. He introduced this technique to find the proportion of sensitive/ socially stigmatized variables. This technique is very useful to reduce Social Desirability Response Bias (SDB) or evasive answer bias up to great degree. At one end it can be useful to enhance the cooperation between interviewer and interviewee to work together and at other end, to keep up privacy and confidentiality level of interviewee.

A major measure of improvements and variations of Warner's RRT have been suggested by various investigators. Greenberg et al., (1969), Mangat and Singh (1990), Mangat (1994), Mahmood et al., (1998), Christofides (2003), Kim and Warde (2004) and Singh and Tarray (2014) are some of the many to be cited. Narjis and Shabbir (2021) presented another two-stage RRT model to find the commonness of sensitive characteristic. The utility of presented two-stage RR model under stratification is likewise investigated. Hsieh et al., (2022) presented an odd model for two-stage RR data and applied inverse probability weighting and various other strategies to manage covariates regarding missing values. Singh and Singh (2021) and Singh et al., (2022) presented their work on using negative binomial and Poisson distribution as a randomization device.

1.3. Logistic Regression

While working with regression models, we have the exceptionally realize as a primary concern to obtain the values of parameter of the concerned model. For this specific motivation behind assessment of parameters, we have numerous techniques like maximum likelihood (ML), ordinary least squares (OLS), and Generalized least square (GLS). The OLS is an integral technique when factors are continuous, yet in the dichotomous variable, OLS isn't suitable. Where in all dichotomous variable scenarios, the responses are either in 'yes' or 'no' (Shair & Majeed, 2020). One procedure to assess these sorts of issues is logistic technique. Logistic Regression has been used in many research areas especially in different fields of Psychological or Behavioral studies, like Exercise and Sports studies (Capel (1986)), Pediatric Psychology (Friedrich et al. (1986)), Psychopathology (Clark and Beck (1991)), Community Psychology (Hedeker et al. (1994)), Human Genetics (Waldman et al. 1999) and many others. In such studies, the response variable is more likely to be sensitive (Shair et al. 2022).

1.4. Logit Estimation in RR

We have discussed that logistic regression is applied in case binary variables but when our response dichotomous variable is related to any of the sensitive characteristics, ordinary logit technique is not good to apt. Corstange (2004, 2009) proposed a procedure, known as hidden logit for the estimation of parameters in case of sensitive or socially stigmatized variables. Hidden logit model is modified form of ordinary logit and hold-back the effect of randomizing device (Shair & Anwar, 2023). The hidden procedure is to model the true probability of yes answer π as a function of some explanatory variables X. In ordinary logits we know that odd ratio is:

$$\ln\left(\frac{\pi}{1-\pi}\right)=X\beta \qquad (1)$$

Generally, if π is the probability of a "yes" response, so we will solve our model for π and replace it in ordinary logits. Finally, the hidden logit model can be obtained in terms of X and β. Here values of β are the population parameters and we have to estimate them. Using same equation, we can find parameters of interest and then the logits by ML procedure by setting the derivatives of the likelihood function equal to zero. The RR model employed by Corstange (2004 & 2009) is as following: The respondent is asked to toss a coin if head appears, he/she is requested to report 'yes' irrespective of his actual status, but if tail appears they have to answer a yes/no question. If φ is the probability of unconditional 'yes' and p is the true proportion of respondent saying 'yes', then the probability of a 'yes' response is given as:

$$p(yes)=\theta=\varphi+(1-\varphi)\,p \qquad (2)$$

For the derivation purpose modified hidden logits, we will interchange φ with π, so the probability of a 'yes' response is given as:

$$p(y)=\theta=\pi+(1-\pi)\,p \qquad (2)$$

Solving (2) for the value of π

$$\pi=\frac{\theta-p}{1-p} \qquad (3)$$

Now replacing value of π in ordinary logits (1) and solving for θ.

$$\ln\left[\frac{\theta-p}{1-p}\bigg/1-\frac{\theta-p}{1-p}\right]=X_i\beta$$

$$\ln\left[\frac{\theta-p}{1-p}\bigg/\frac{1-p-\theta+p}{1-p}\right]=X_i\beta$$

$$\frac{\theta-p}{1-\theta}=e^{X_i\beta}$$

$$\theta-p=e^{X_i\beta}(1-\theta)$$

$$\theta-p=e^{X_i\beta}-e^{X_i\beta}\theta$$

$$\theta+e^{X_i\beta}\theta=e^{X_i\beta}+p$$

$$\theta(1+e^{X_i\beta})=e^{X_i\beta}+p$$

$$\theta=\frac{e^{X_i\beta}+p}{1+e^{X_i\beta}} \tag{4}$$

We can see that above equation becomes the ordinary logits derived from direct responses for $p = 0$. Consider a binary variable 'y_i' which takes value '1' showing 'yes' response and value '0' showing 'no' response, with probabilities θ_i and $1-\theta_i$ respectively. Then the likelihood function of β given y_i is:

$$L(\beta|y_i)=\prod_{i=1}^{n}\theta_i^{y_i}(1-\theta_i)^{1-y_i},$$

$$\ln\left[L\left(\beta|y_i\right)\right]=l=\sum_{i=1}^{n}\left\{y_i\ln\theta_i+\left(1-y_i\right)\ln\left(1-\theta_i\right)\right\} \tag{5}$$

The first derivative of this likelihood function w.r.to β is done as under

$$\frac{\partial\ln L}{\partial\beta}=\sum_{i=1}^{n}\left\{y_i\frac{\partial\ln\theta_i}{\partial\beta}+\left(1-y_i\right)\frac{\partial\ln\left(1-\theta_i\right)}{\partial\beta}\right\}$$

$$=\sum_{i=1}^{n}\left\{y_i\frac{1}{\theta_i}\frac{\partial\theta_i}{\partial\beta}+\left(1-y_i\right)\frac{1}{\left(1-\theta_i\right)}(-1)\frac{\partial\theta_i}{\partial\beta}\right\} \tag{6}$$

Let's first find $\frac{\partial\theta_i}{\partial\beta}$

$$\frac{\partial\theta_i}{\partial\beta}=\frac{\left(1+e^{X_i\beta}\right)\frac{\partial}{\partial\beta}\left(e^{X_i\beta}+p\right)-\left(e^{X_i\beta}+p\right)\frac{\partial}{\partial\beta}\left(1+e^{X_i\beta}\right)}{\left(1+e^{X_i\beta}\right)^2}$$

$$=\frac{\left(1+e^{X_i\beta}\right)e^{X_i\beta}X_i-\left(e^{X_i\beta}+p\right)e^{X_i\beta}X_i}{\left(1+e^{X_i\beta}\right)^2}$$

$$=\frac{e^{X_i\beta}X_i+e^{2X_i\beta}X_i-e^{2X_i\beta}X_i+pe^{X_i\beta}X_i}{\left(1+e^{X_i\beta}\right)^2}$$

$$\frac{\partial\theta_i}{\partial\beta}=\frac{(1-p)X_ie^{X_i\beta}}{\left(1+e^{X_i\beta}\right)^2} \tag{7}$$

Replacing values of $\frac{\partial\theta_i}{\partial\beta}$ and θ in (6)

$$= \sum_{i=1}^{n}\left[y_i \frac{1}{\left(\frac{e^{X_i\beta}+p}{1+e^{X_i\beta}}\right)} \frac{(1-p)X_i e^{X_i\beta}}{\left(1+e^{X_i\beta}\right)^2} + (1-y_i)\frac{1}{\left(1-\left(\frac{e^{X_i\beta}+p}{1+e^{X_i\beta}}\right)\right)}(-1)\frac{(1-p)X_i e^{X_i\beta}}{\left(1+e^{X_i\beta}\right)^2}\right]$$

$$= \sum_{i=1}^{n}\left[y_i \frac{(1-p)X_i e^{X_i\beta}}{\left(e^{X_i\beta}+p\right)\left(1+e^{X_i\beta}\right)} - \frac{(1-y_i)(1-p)e^{X_i\beta}X_i}{(1-p)(1+e^{X_i\beta})}\right]$$

$$= \sum_{i=1}^{n}\left[y_i \frac{(1-p)X_i e^{X_i\beta}}{\left(e^{X_i\beta}+p\right)\left(1+e^{X_i\beta}\right)} - \frac{(1-y_i)e^{X_i\beta}X_i}{1+e^{X_i\beta}}\right]$$

$$= \sum_{i=1}^{n}\left[y_i \frac{(1-p)X_i e^{X_i\beta}}{\left(e^{X_i\beta}+p\right)\left(1+e^{X_i\beta}\right)} - \frac{e^{X_i\beta}X_i}{1+e^{X_i\beta}} + \frac{y_i e^{X_i\beta}X_i}{1+e^{X_i\beta}}\right]$$

$$= \sum_{i=1}^{n}\left[y_i \left\{\frac{(1-p)X_i e^{X_i\beta} + (e^{X_i\beta}+p)e^{X_i\beta}X_i}{\left(e^{X_i\beta}+p\right)\left(1+e^{X_i\beta}\right)}\right\} - \frac{e^{X_i\beta}X_i}{1+e^{X_i\beta}}\right]$$

$$= \sum_{i=1}^{n}\left[y_i \left\{\frac{X_i e^{X_i\beta} - pX_i e^{X_i\beta} + e^{2X_i\beta}X_i + pe^{X_i\beta}X_i}{\left(e^{X_i\beta}+p\right)\left(1+e^{X_i\beta}\right)}\right\} - \frac{e^{X_i\beta}X_i}{1+e^{X_i\beta}}\right]$$

$$= \sum_{i=1}^{n}\left[y_i X_i \left\{\frac{e^{X_i\beta}(1+e^{X_i\beta})}{\left(e^{X_i\beta}+p\right)\left(1+e^{X_i\beta}\right)}\right\} - \frac{e^{X_i\beta}X_i}{1+e^{X_i\beta}}\right]$$

$$= \sum_{i=1}^{n}\left[y_i \left\{\frac{e^{X_i\beta}}{\left(e^{X_i\beta}+p\right)}\right\} - \frac{1}{1+e^{-X_i\beta}}\right]X_i$$

$$\frac{\partial \ln L}{\partial \beta} = \sum_{i=1}^{n}\left[y_i \left\{\frac{e^{X_i\beta}}{\left(p+e^{X_i\beta}\right)}\right\} - \left(1+e^{-X_i\beta}\right)^{-1}\right]X_i \qquad (8)$$

By taking first derivative equivalent to 0, maximizes the function, yet its answer can't be acquire analytically, so we solve this expression numerically. Hussain and Shabbir (2008) utilized Warner (1965) RRT to measure the hidden logits. Hussain et al., (2011) utilized the Mangat (1994) RRT to measure the improved logits using Corstange (2004) methodology and it is compared with Hussain and Shabbir (2008) logit estimation at equal privacy protection. Cruff et al., (2016) projected a review of different regression methods for RR data in handbook of Statistics. Chang et al., (2021) also applied logistic regression to explore the effects of covariates on a stigmatized characteristic for the two-stage RRT of Huang (2004). Hsieh and Perri (2020) proposed a few theoretical and observational advances by providing the strategy for analyzing the elements that impact two stigmatized variates, when information is gathered by RRT.

2. Proposed methodology

2.1. Modified hidden logit using Mangat and Singh (1990) RRT

Mangat and Singh (1990) have anticipated a new randomizing technique and they have utilized two randomizing devices named R_1 and R_2. They examined a model when respondents are making honest revealing utilizing two stage RR. Researchers have then expanded this technique when there is not exactly honest revealing at respondent side. They utilized a sample of size n without replacement and people are asked to utilize R_1 having following proclamations like: or go to R_2, addressed with probabilities *T* and *1-T* respectively. The R_2 containing following explanations like: I belong to sensitive group *A*, or I don't belong to sensitive group *A*, addressed with probabilities *p* and *1-p*. Here for R_1 respondent should answer 'yes' or 'no' as indicated by proclamation and the genuine status he has. R_2 is same like Warner (1965) RRT. The likelihood of a 'yes' response is then given by

$$P(yes) = \theta = T\pi + (1-T)\left\{p\pi + (1-p)(1-\pi)\right\}. \qquad (9)$$

Solving (9) for π

$$\theta = T\pi + (1-T)p\pi + (1-T)(1-p)(1-\pi)$$

$$\theta = [T + (1-T)p - (1-T)(1-p)]\pi + (1-T)(1-p) \quad \theta - (1-T)(1-p) = [T + p - pT - \{1-T-p+pT\}]\pi$$

$$\theta - (1-T)(1-p) = [2T + 2p - 2pT - 1]\pi$$

$$\pi = \frac{\theta - (1-T)(1-p)}{2T + 2p - 2pT - 1} \tag{10}$$

Now replacing value of π in ordinary logits (1) and solving for θ

$$\ln\left(\frac{\theta - (1-T)(1-p)}{2T+2p-2pT-1} \Big/ 1 - \frac{\theta - (1-T)(1-p)}{2T+2p-2pT-1} \right) = X_i\beta$$

$$\ln\left(\frac{\theta - (1-T)(1-p)}{2T+2p-2pT-1} \Big/ \frac{2T+2p-2pT-1-\theta+1-T-p+pT}{2T+2p-2pT-1} \right) = X_i\beta$$

$$\ln\left(\frac{\theta - (1-T)(1-p)}{2T+2p-2pT-1} \Big/ \frac{T+p-pT-\theta}{2T+2p-2pT-1} \right) = X_i\beta$$

$$\theta - (1-T)(1-p) = (T+p-pT-\theta)e^{X_i\beta}$$

$$\theta - (1-T)(1-p) = (T+p-pT)e^{X_i\beta} - \theta e^{X_i\beta}$$

$$\theta(1+e^{X_i\beta}) = (T+p-pT)e^{X_i\beta} + (1-T)(1-p)$$

$$\theta = \frac{(T+p-pT)e^{X_i\beta} + (1-T)(1-p)}{(1+e^{X_i\beta})} \tag{11}$$

For p =1, and T = 1, equation (11) becomes the ordinary logits resulting from direct responses. Using the likelihood function of β given y_i, the first derivative of likelihood function w.r.to β is shown as under:

$$\frac{\partial\theta_i}{\partial\beta} = \frac{\left(1+e^{X_i\beta}\right)[(T+p-pT)e^{X_i\beta}]X_i - [(T+p-pT)e^{X_i\beta} + (1-T)(1-p)]e^{X_i\beta}X_i}{\left(1+e^{X_i\beta}\right)^2}$$

$$\frac{\partial\theta_i}{\partial\beta} = \frac{(T+p-pT)e^{X_i\beta} + (T+p-pT)e^{2X_i\beta} - (T+p-pT)e^{2X_i\beta} - (1-T)(1-p)e^{X_i\beta}}{\left(1+e^{X_i\beta}\right)^2} X_i$$

$$\frac{\partial\theta_i}{\partial\beta} = \frac{[(T+p-pT) - (1-T)(1-p)]e^{X_i\beta}}{\left(1+e^{X_i\beta}\right)^2} X_i$$

$$\frac{\partial\theta_i}{\partial\beta} = \frac{[T+p-pT-1+T+p-pT]e^{X_i\beta}}{\left(1+e^{X_i\beta}\right)^2} X_i$$

$$\frac{\partial\theta_i}{\partial\beta} = \frac{[2T+2p-2pT-1]\ e^{X_i\beta}}{\left(1+e^{X_i\beta}\right)^2} X_i \tag{12}$$

Replacing values of $\frac{\partial\theta_i}{\partial\beta}$ and θ in $\frac{\partial \ln L}{\partial\beta}$

$$\frac{\partial \ln L}{\partial\beta} = \sum_{i=1}^{n}\left\{ y_i \frac{1}{\theta_i}\frac{\partial\theta_i}{\partial\beta} + (1-y_i)\frac{1}{(1-\theta_i)}(-1)\frac{\partial\theta_i}{\partial\beta} \right\}$$

$$= \sum_{i=1}^{n}[y_i \frac{1}{\left(\frac{(T+p-pT)e^{X_i\beta} + (1-T)(1-p)}{(1+e^{X_i\beta})}\right)} \frac{[2T+2p-2pT-1]\ e^{X_i\beta}}{\left(1+e^{X_i\beta}\right)^2} X_i$$

$$+(1-y_i)\frac{1}{\left(1-\left(\frac{(T+p-pT)e^{X_i\beta} + (1-T)(1-p)}{(1+e^{X_i\beta})}\right)\right)}(-1)\frac{[2T+2p-2pT-1]\ e^{X_i\beta}}{\left(1+e^{X_i\beta}\right)^2} X_i]$$

$$= \sum_{i=1}^{n}[y_i \frac{[2T+2p-2pT-1]\ e^{X_i\beta}}{[(T+p-pT)e^{X_i\beta} + (1-T)(1-p)]\left(1+e^{X_i\beta}\right)} X_i$$

$$-(1-y_i)\frac{[2T+2p-2pT-1]\ e^{X_i\beta}}{[(1+e^{X_i\beta}) - (T+p-pT)e^{X_i\beta} - (1-T)(1-p)]\left(1+e^{X_i\beta}\right)} X_i]$$

$$= \sum_{i=1}^{n} [\frac{(2T+2p-2pT-1)e^{X_i\beta} X_i}{\left(1+e^{X_i\beta}\right)} \{\frac{y_i}{[(T+p-pT)e^{X_i\beta}+(1-T)(1-p)]} + \frac{y_i}{1+[1-(T+p-pT)]e^{X_i\beta}-(1-T)(1-p)}$$

$$-\frac{1}{1+[1-(T+p-pT)]e^{X_i\beta}-(1-T)(1-p)}\}]$$

$$= \sum_{i=1}^{n} [\frac{(2T+2p-2pT-1)e^{X_i\beta} X_i}{\left(1+e^{X_i\beta}\right)} \{y_i \frac{1+e^{X_i\beta}}{[(T+p-pT)e^{X_i\beta}+(1-T)(1-p)][1+(1-T-p+pT)e^{X_i\beta}-(1-T)(1-p)]}$$

$$-\frac{1}{1+(1-T-p+pT)e^{X_i\beta}-(1-T)(1-p)}\}]$$

$$\frac{\partial \ln L}{\partial \beta} = \sum_{i=1}^{n} [\frac{(2T+2p-2pT-1)e^{X_i\beta} X_i}{[1+(1-T-p+pT)e^{X_i\beta}-(1-T)(1-p)]} \{\frac{y_i}{[(T+p-pT)e^{X_i\beta}+(1-T)(1-p)]} - \frac{1}{\left(1+e^{X_i\beta}\right)}\}] \quad (13)$$

By taking first derivative equivalent to 0, maximizes the function, yet its answer can't be acquire analytically, so we solve this expression numerically.

2.2. Material

We have used the 'Eviews' programming software for all the numerical analysis. Monte carlo simulations are done using Gauss-Newton method and for every *p* from 0.1- 0.9, and odd values of T, samples of sizes 1000 are generated from a three regressors equation with no constant term. Uniform distribution is used for generation of samples of size 1000 for X_i variate. For simplicity, we can say that logit estimation is performed taking starting values of β= (0, 1, 1, 1) and each X_i is following uniform distribution with parameter -3 and 3. Through this simulation we find estimates of *β's (b's)*, their standard errors (*s.e)* and AIC/SIC for Corstange (2004) and Mangt and Singh (1994) for different values of *p* and *T*

3. Results and Discussions

3.1. Estimates of *β's (b's)*, *s.e(b's)* and AIC/SIC for Corstange (2004) RRT.

Table 1: Estimates of *β's (b's)* for different values of *p*

b's/p	b_0	b_1	b_2	b_3
0	-0.16888	1.138353	1.06685	1.10611
0.1	-0.16281	1.164716	1.02763	1.0995
0.2	-0.14621	1.074617	0.97167	1.05347
0.3	-0.30193	1.183877	1.00889	1.14532
0.4	-0.04816	1.156393	0.90514	0.94355
0.5	-0.06679	1.221259	1.00295	0.95676
0.6	-0.03061	1.109165	0.89037	0.84872
0.7	0.172082	1.362857	0.95295	0.88377
0.8	-1.10519	9.782248	7.97214	7.02703
0.9	0.284891	1.698229	1.20852	1.76406

Table 2: *s.e(b's)* for different values of *p*

s.e(b's)/p	b_0	b_1	b_2	b_3
0	0.07756	0.08761	0.0861	0.08408
0.1	0.09578	0.1099	0.10348	0.10329
0.2	0.11186	0.12074	0.11604	0.11677
0.3	0.1408	0.1591	0.14615	0.15254
0.4	0.15212	0.17465	0.15216	0.14906
0.5	0.18286	0.21983	0.19691	0.18197
0.6	0.20678	0.23007	0.20483	0.18995
0.7	0.26785	0.34846	0.27547	0.25306
0.8	1.4287	8.27226	6.82195	6.09238
0.9	0.07756	0.08761	0.0861	0.08408

Table 3: AIC and SIC against different values of *p*

p	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
AIC	0.5725	0.7853	0.9184	0.9841	0.9959	0.9543	0.8528	0.6506	0.3914
SIC	0.5922	0.8049	0.9381	1.0037	1.0155	0.9740	0.8724	0.6702	0.4110

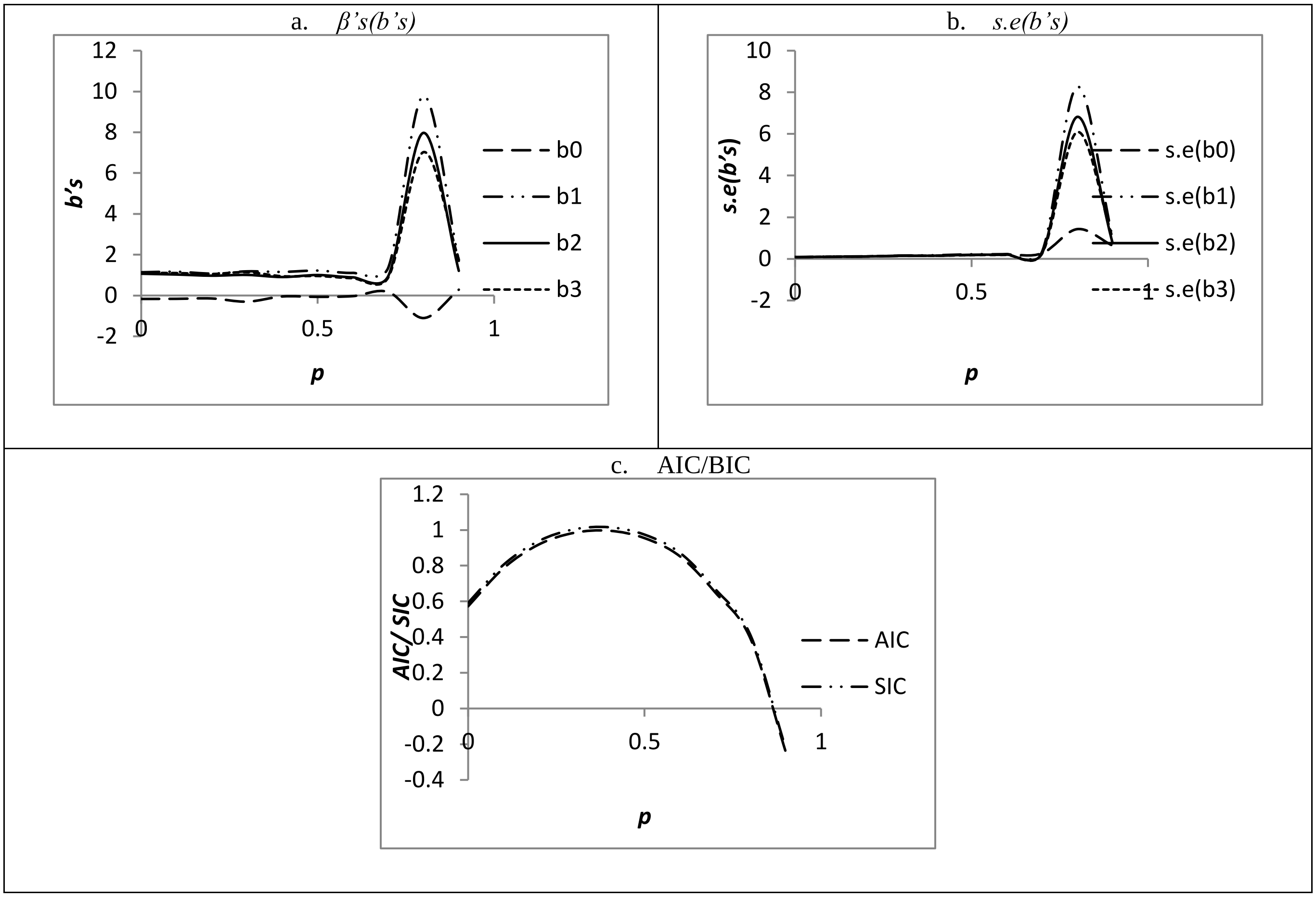

Figure 1: Estimates of *β's(b's), s.e(b's),* AIC/SIC against different values of *p*

4. Discussion on results

Tables 1-2 compares the performance of the modified hidden logit estimates with the estimates of ordinary logit for different values of *p*. From Figure 1(a-b), we can notice that for the above experimental conditions the modified hidden logit estimates and their standard errors (S.E) are moving away from ordinary logits for higher values of *p*. It can be seen that as the value of *p* increases, a negatively skewed pattern can be seen in values of *β's (b's)* and *s.e(b's)* We can also notice that the estimated values of modified hidden logit for Corstange (2004) RR design taking $p = 0.1$,becomes very near to the estimates of ordinary logits. The estimates of ordinary logits are shown against $p = 0$. From the figure 1(b) it is depicted that as the value of *p* increases, the *s.e(b's)* also increases. The *s.e(b's)* becomes the least for $p = 0$ which are the values S.E for ordinary logits. As a summary we can say that the *s.e(b's)* are moving towards the standard errors of ordinary logits for lower values of *p* and near to the standard errors of ordinary logits for $p = 0.1$. The AIC and SIC are calculated in table 3 for logit estimation using Corstange (2004) RRT taking different values of *p.* The figure 1(c) demonstrates that for *p* below 0.4, values of AIC and SIC are increasing and start decreasing *p* as 0.5 and above. We can also notice that AIC and SIC values are least for 0.9.

4.1. Estimates of β's (b's), s.e(b's), and AIC/BIC for different values of p and T for Mangat and Singh (1990) RRT

In current section, we estimate the modified hidden logit model for Mangat and Singh (1990) RRT for different values of *p* by fixing odd values of *T*. Table 4-6 and figure 2-5 (a-e), show the estimates of *β's (b's),* their S.E and AIC/SIC and their graphical presentation respectively for different values of *p* while *T* is fixed. Point to be noted is that, for all the tables of estimates and SE, the values against $p = 1$ are the values of ordinary logits.

Table 4: Estimates of β's (b's) for different values of p when T is odd

T		P									
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
0.1	b_0	-0.3327	-0.4925	-0.8604	-5.7472	8.1818	0.3714	0.1366	0.0060	-0.0152	0.0195
	b_1	0.7803	0.7047	0.5468	-1.2376	22.1450	1.3177	1.1500	1.0890	1.0009	0.9565
	b_2	0.9900	1.0489	1.2911	36.5650	0.6213	0.4359	0.7639	0.8244	0.8538	0.8325
	b_3	0.9020	0.8795	0.7156	19.0722	18.4375	1.0151	0.9295	0.9627	0.8848	0.8855
0.3	b_0	-0.8105	-5.4961	30.6839	45.8133	0.3735	0.2550	0.2225	0.1667	0.1648	0.0777
	b_1	2.0737	11.6376	11.0249	54.2886	1.2830	1.0855	1.0469	1.0510	1.0345	1.0773
	b_2	2.2632	18.0736	14.4182	38.2762	1.1437	0.9485	0.9461	0.9232	0.9204	0.9766
	b_3	2.2451	17.9727	20.2562	44.1005	0.8999	0.8926	0.9077	0.9233	0.9302	0.9520
0.5	b_0	33.2134	0.5251	0.2633	0.2551	0.1698	0.1698	0.1761	0.1544	0.1512	0.1547
	b_1	28.7082	0.7060	0.6677	0.7989	0.8497	0.8497	0.8386	0.9224	0.9615	1.0146
	b_2	-5.0307	0.1320	0.4346	0.7187	0.8091	0.8091	0.8389	0.8809	0.8735	0.9045
	b_3	4.5744	0.4572	0.6494	0.7789	0.9111	0.9111	0.8805	0.9551	0.9652	1.0090
0.7	b_0	-0.0205	-0.0765	-0.1070	-0.0522	-0.0001	-0.0271	-0.0129	-0.0496	-0.1061	-0.0896
	b_1	1.0200	1.0311	0.9283	0.9986	1.0161	1.0390	1.0452	1.0302	1.0386	0.9986
	b_2	1.1786	1.1607	1.0407	1.1046	1.1467	1.1832	1.1410	1.1261	1.1109	1.0586
	b_3	0.9508	0.9404	0.8900	1.0133	1.0655	1.0990	1.0867	1.0851	1.0645	1.0175
0.9	b_0	-0.1251	-0.1180	-0.1100	-0.1045	-0.1086	-0.1008	-0.0998	-0.0970	-0.0936	-0.1002
	b_1	1.0455	1.0371	1.0311	0.9998	0.9749	0.9713	0.9583	0.9264	0.9371	0.9285
	b_2	1.0406	1.0263	1.0282	0.9998	0.9739	0.9772	0.9643	0.9320	0.9411	0.9274
	b_3	1.0142	1.0097	1.0046	0.9771	0.9513	0.9414	0.9332	0.9025	0.9060	0.9069

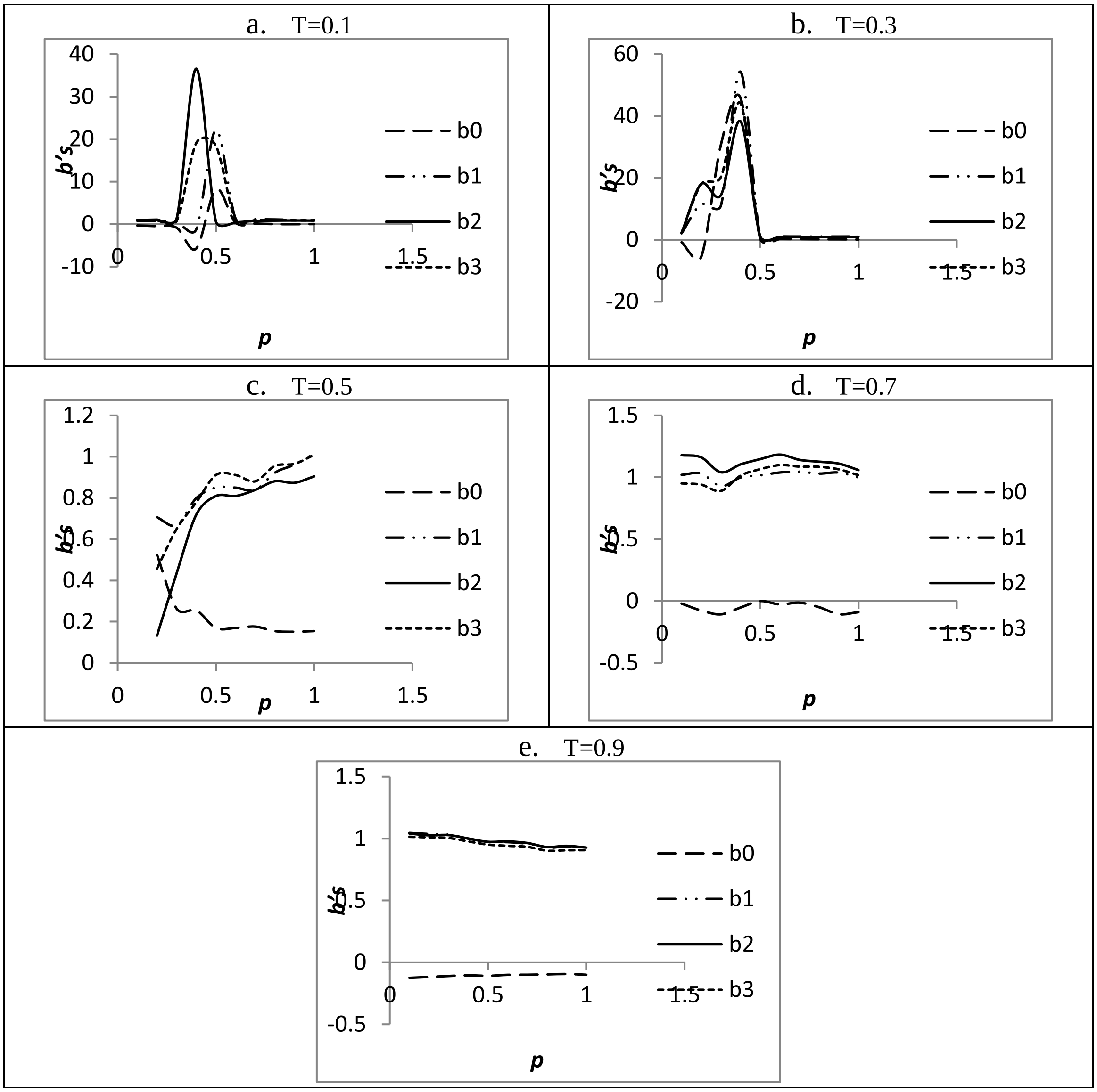

Figure 2: Estimates of *β's (b's)* against different values of *p* when *T* is odd

Table 5: s.e (b's) for different values of p when T is odd

T		P									
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
0.1	$s.e(b_0)$	0.1598	0.2466	0.5193	39.0504	34.2730	0.4183	0.2279	0.1522	0.1057	0.0739
	$s.e(b_1)$	0.1359	0.1907	0.3091	9.3556	88.1510	0.4950	0.2565	0.1665	0.1083	0.0732
	$s.e(b_2)$	0.1584	0.2435	0.5078	258.1015	3.6923	0.2791	0.1990	0.1398	0.0987	0.0679
	$s.e(b_3)$	0.1516	0.2216	0.3561	133.6729	72.5796	0.4220	0.2257	0.1552	0.1018	0.0710
0.3	$s.e(b_0)$	0.7614	19.2822	744.9627	182.4418	0.4091	0.2474	0.1771	0.1328	0.1015	0.0756
	$s.e(b_1)$	1.1681	40.9811	272.9713	217.5677	0.4792	0.2606	0.1830	0.1386	0.1048	0.0810
	$s.e(b_2)$	1.2835	62.6860	351.4991	152.0934	0.4502	0.2416	0.1740	0.1291	0.0985	0.0773
	$s.e(b_3)$	1.2484	62.2894	493.1158	176.5989	0.3739	0.2297	0.1673	0.1276	0.0981	0.0747
0.5	$s.e(b_0)$	149.3713	0.4714	0.3078	0.2493	0.1608	0.1608	0.1312	0.1105	0.0925	0.0766
	$s.e(b_1)$	129.5684	0.3531	0.2425	0.2186	0.1488	0.1488	0.1202	0.1079	0.0929	0.0798
	$s.e(b_2)$	24.2287	0.2592	0.2085	0.2073	0.1447	0.1447	0.1204	0.1049	0.0875	0.0740
	$s.e(b_3)$	21.0472	0.2970	0.2308	0.2071	0.1495	0.1495	0.1191	0.1066	0.0901	0.0771
0.7	$s.e(b_0)$	0.2357	0.2047	0.1748	0.1562	0.1394	0.1252	0.1098	0.0979	0.0872	0.0764
	$s.e(b_1)$	0.2446	0.2130	0.1678	0.1586	0.1441	0.1315	0.1156	0.1017	0.0906	0.0771
	$s.e(b_2)$	0.2773	0.2374	0.1863	0.1752	0.1613	0.1484	0.1266	0.1116	0.0981	0.0829
	$s.e(b_3)$	0.2366	0.2029	0.1660	0.1628	0.1513	0.1390	0.1207	0.1072	0.0937	0.0794
0.9	$s.e(b_0)$	0.1091	0.1050	0.1010	0.0961	0.0919	0.0882	0.0850	0.0810	0.0776	0.0745
	$s.e(b_1)$	0.1143	0.1094	0.1049	0.0975	0.0914	0.0875	0.0835	0.0776	0.0749	0.0714
	$s.e(b_2)$	0.1165	0.1111	0.1070	0.0998	0.0937	0.0901	0.0861	0.0801	0.0773	0.0734
	$s.e(b_3)$	0.1123	0.1078	0.1033	0.0963	0.0903	0.0860	0.0824	0.0766	0.0737	0.0707

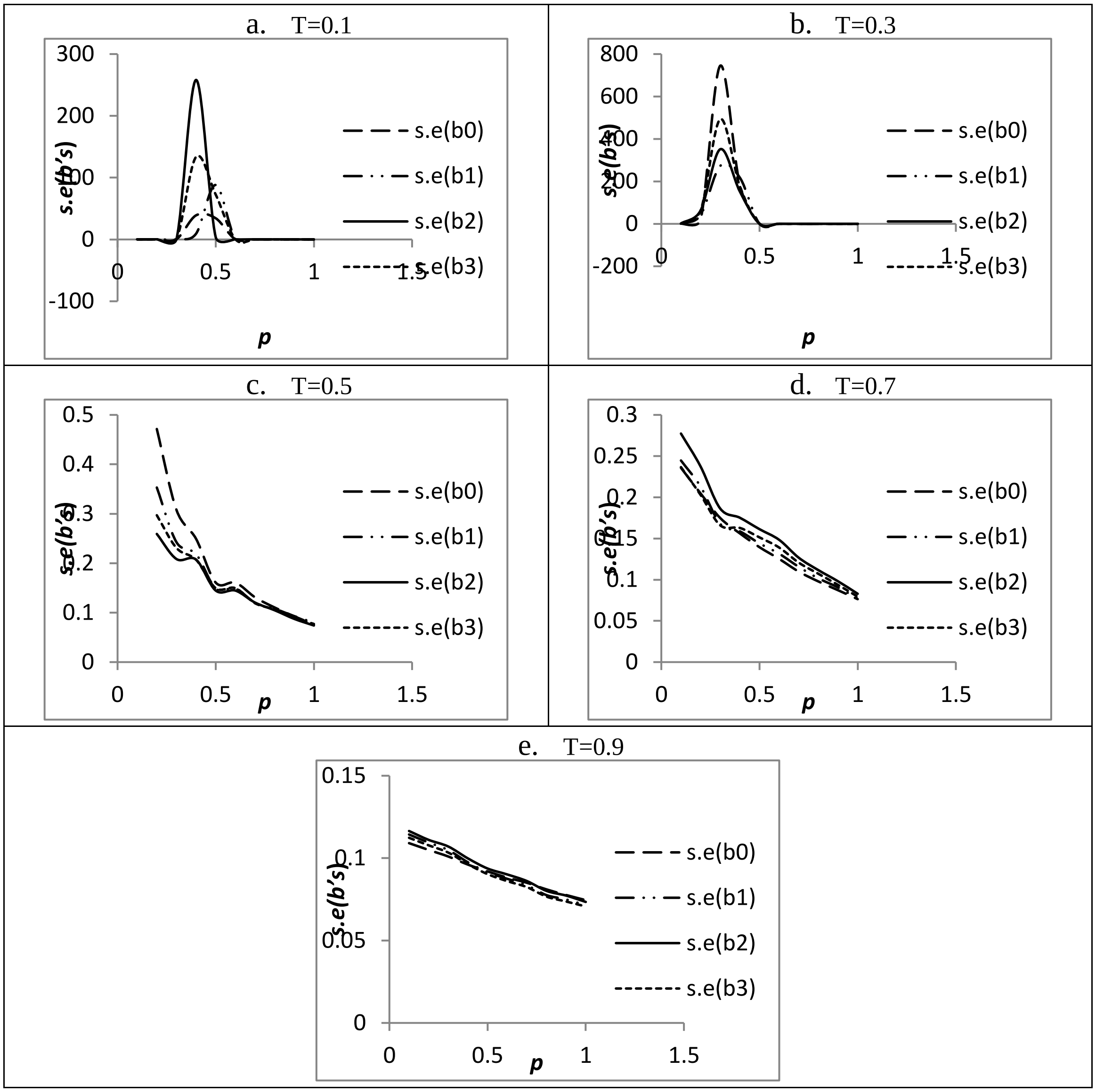

Figure 3: ***s.e (b's)*** **for different values of** ***p*** **when** ***T*** **is odd.**

Table 6: AIC/SIC for different values of p when T is odd

T		P									
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
0.1	AIC	1.2458	1.3471	1.4133	1.4482	1.4437	1.4136	1.3339	1.2051	1.0187	0.7229
	SIC	1.2655	1.3668	1.4329	1.4679	1.4633	1.4332	1.3535	1.2248	1.0383	0.7425
0.3	AIC	1.4213	1.4488	1.4571	1.4295	1.3996	1.3334	1.2400	1.1113	0.9345	0.6256
	SIC	1.4409	1.4685	1.4767	1.4492	1.4192	1.3530	1.2596	1.1309	0.9541	0.6452
0.5	AIC	1.4485	1.4438	1.4175	1.3763	1.2347	1.2347	1.1421	1.0027	0.8708	0.6624
	SIC	1.4681	1.4634	1.4371	1.3959	1.2544	1.2544	1.1617	1.0224	0.8904	0.6821
0.7	AIC	1.3184	1.2789	1.2592	1.1624	1.0876	1.0104	0.9114	0.8283	0.7293	0.6299
	SIC	1.3380	1.2986	1.2788	1.1821	1.1073	1.0300	0.9310	0.8479	0.7490	0.6495
0.9	AIC	0.8798	0.8601	0.8344	0.8078	0.7871	0.7529	0.7342	0.7100	0.6581	0.6223
	SIC	0.8994	0.8798	0.8540	0.8274	0.8068	0.7725	0.7538	0.7296	0.6778	0.6419

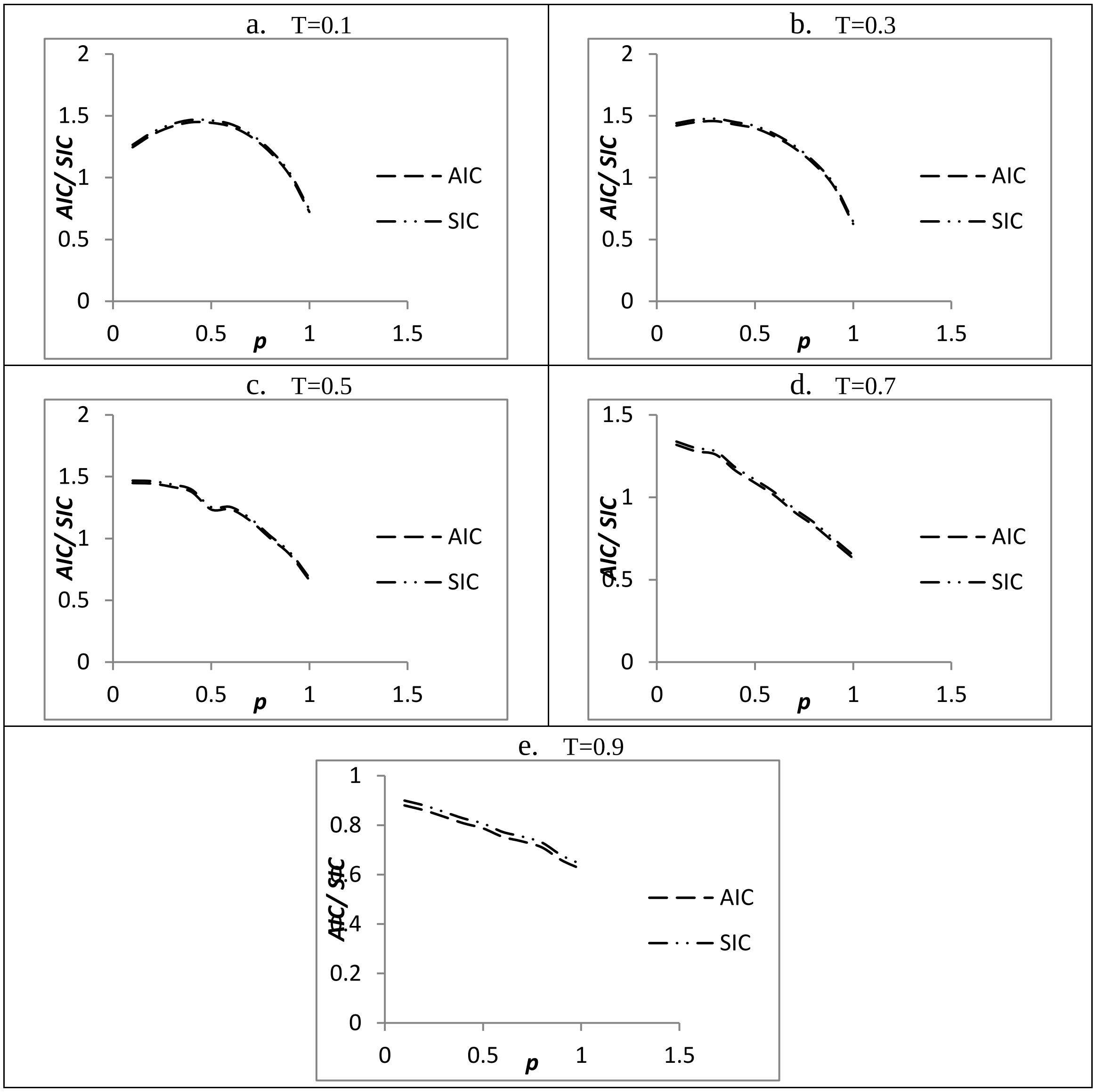

Figure 4: AIC/SIC against different values of p when T is odd.

4.2. Discussion and Comparison

Table 4 and figure 2(a-e), depict the estimates of modified hidden logit for the above experimental conditions. We can notice that the modified hidden logit is resulting in the estimates, which are quite close to ordinary logits for higher values of *T* and p. Taking lower values of *T*, the estimates of modified hidden logit model for Mangat and Singh (1990) RRT are highly deviating from usual values of ordinary logits for p = 0.4 - 0.5. But when p is greater than 0.5, so estimates are moving towards the ordinary logits. We can see that when p increases the estimates decrease and transfer towards ordinary logits. The estimates of ordinary logits are shown against p = 1 in all tables. Table 5 and figure 3(a-e), depict the SE of estimates of modified hidden logits. We can notice that the standard errors are quite close to ordinary logits for higher values of *T* and *p*. Taking T = 0.1, the S.D of estimates of modified hidden logit model for Mangat and Singh (1990) RRT are very high as compared to the SE of ordinary logits for p = 0.4 - 0.5. But when p is greater than 0.5, SE of estimates are moving towards the SE of ordinary logits. One can see that if p and T increase, the SE of estimates of modified hidden logit model for Mangat and Singh (1990) RRT keeps on decreasing and approaching to the SE of estimates of ordinary logits. From table 6 and Figure 4 (a-e), it can be seen that the values of AIC and SIC are increasing for low values of p and T, but begin decreasing for p = 0.5 and above for all values of T. We can also notice that AIC and SIC values are least for p and T as 0.9 which shows a good model fit.

5. Conclusion and Recommendation

The findings of this research depict the estimates of modified hidden logit and compare the performance of these estimates with ordinary logit for two RRT's. As a summary, we can say that the estimates are moving towards the original values for every increase in *p* and *T* and very near to ordinary logits for taking *p* and *T* as 0.9 for Mangat and Singh (1994) RRT. We discover that the estimates of hidden logit move toward to true parametric values as the *p* and *T* increases. Also, we examine that the standard errors of estimates decrease as *p* and *T* increases, and is least for ordinary logit taking *p* and *T* as 1. Also, we observe that hidden logit estimates for Mangat and Singh (1990) are closer to the true parametric values as compare to Corstange (2004) and they show raise in accuracy. So modified hidden logit estimation using Mangat and Singh (1990) is more appropriate to obtain true estimates of population proportion in case of sensitive characteristics. We know that a lesser AIC/SIC value designates a good fit. So in our proposed model, AIC and SIC values are decreasing for higher values of *p* and *T*. Hence we can conclude that for higher values of p and *T* Mangat and Singh (1990) gives best estimates. This paper has an extraordinary assurance towards application and estimation of logistic models when one is dealing with sensitive attributes. We recommend that the proposed methodology of modified hidden logits can be extended and applied to more RR models.

References

Capel, S.A., (1986). Psychological and organizational factors related to burnout in athletic trainers. *Research Quarterly for Exercise and Sport*, 57, 321-328.

Chang, P.C., Pho, K.H., Lee, S.M., Li, C.S., (2021). Estimation of parameters of logistic regression for two-stage randomized response technique. *Computational Statistics*, 36, 2111–2133.

Christofides, T.C., (2003). A generalized randomized response technique. *Metrika*, 57, 195-200.

Clark, D.A. and Beck, A.T., (1991). Personality factors in dysphoria: A psychometric refinement of Beck's Sociotropy-Autonomy Scale. *Journal of Psychopathology and Behavioral Assessment,* 13, 369–388.

Corstange, D. (2004). Sensitive questions, truthful responses? Randomized response and hidden logit as a procedure to estimate it." 2004 Annual Meeting of the American Political Science Association September2 - September 5, 2004.

Corstange, D., (2009). Sensitive Questions, Truthful Answers? Modeling the List Experiment with LISTIT. *Political Analysis*, 17, 45-63.

Cruff, M.J.L.F., Böckenholt, U., Van der Heijden, P.G.M., Frank, L.E., (2016). Chapter 18 - A Review of Regression Procedures for Randomized Response Data, Including Univariate and Multivariate Logistic Regression, the Proportional Odds Model and Item Response Model, and Self-Protective Responses, in Handbook of Statistics. Edited by A. Chaudhuri, T.C., Christofides, and Rao C.R., Amsterdam, the Netherlands, Elsevier. pp. 287-315.

Friedrich, W.N., Urquiza, A.J., Beilke, R.L., (1986). Behavior problems in sexually abused young children. *Journal of Pediatric Psychology*, 11(1), 47-57.

Greenberg, B.G., Latif, A.A., Ela, A., Simmons, W.R., Horvitz, D.G., (1969). The unrelated question randomized response model: theoretical framework. *Journal of the American Statistical Association*, 64, 520-539.

Hedeker,D., McMahon, S.D., Jason, L.A., Salina, D., (1994). Analysis of clustered data in community psychology: with an example from a worksite smoking cessation project. *American Journal of Community Psychology*, 22, 595-615.

Hsieh, S.H., Lee, S.M., Li, C.S., (2022). A two-stage multilevel randomized response technique with proportional odds models and missing covariates. *Sociological Methods & Research*, 51(1), 439-461.

Hsieh, S.H., Perri P.F., (2020). A Logistic Regression Extension for the Randomized Response Simple and Crossed Models: Theoretical Results and Empirical Evidence. *Sociological Methods & Research*, 51(3), 1244-1281.

Huang. K.C., (2004). A survey technique for estimating the proportion and sensitivity in a dichotomous finite population. *Statistica Neerlandica*, 58(1), 75-82.

Hussain, Z., Shabbir, J., (2008). Logit estimation using Warner's randomized response model. *Journal of Modern Applied Statistical Methods*, 7 (1), 140-151.

Hussain, Z., Anjum, S., Shabbir, J., (2011). Improved logit estimation through Mangat randomized response model. *International Journal of Business and Social Sciences,* 2(5), 179-188.

Kim, J., Warde, W.D., (2005). A Mixed Randomized Response Model. *Journal of Statistical Planning Inference,* 133(1), 1-11.

Mahmood, M., Singh, S., Horn, S., (1998). On the confidentiality guaranteed under randomized response sampling: a comparison with several new techniques. *Biometrical Journal*, 40, 237-242.

Mangat, N.S., (1994). An improved randomized response strategy. *Journal of Royal Statistical Society, B*, 56(1), 93-95.

Mangat, N.S., Singh, R., (1990). An alternative randomized response procedure. *Biometrika*, 77, 439-442.

Narjis, G., Shabbir, J., (2021). An Improved Two-stage Randomized Response Model for Estimating the Proportion of Sensitive Attribute. *Sociological Methods and Research,* 50(2), 1-21.

Shair, W., Majeed, M.T., (2020). Labor Market Outcomes of Non-migrant Members in Response to Remittances: Evidence from Provincial capital of Punjab and Khyber Pakhtunkhawa (KPK). *Review of Socio-Economic Perspectives*. 5, 1–22.

Shair, W., Waheed, A., Kamran, M. M., & Kubra, N. (2022). Digital Divide in Pakistan: Barriers to ICT Usage among the Individuals of Pakistan. *Journal of Economic Impact*, *4*(3), 196-204.

Shair, W., & Anwar, M. (2023). Effect of internal and external remittances on expenditure inequality in Pakistan. *Cogent Economics & Finance*, *11*(1), 2178121.

Singh, C., Singh, G. N., Kim, J.M., (2021). A randomized response model for sensitive attribute with privacy measure using Poisson distribution. *Ain Shams Engineering Journal*, 12(4), 4051-4061.

Singh, G., Singh, C., (2022). Proficient randomized response model based on blank card strategy to estimate the sensitive parameter under negative binomial distribution, *Ain Shams Engineering Journal*, 13(5).

Singh, H.P., Tarray, T.A., (2014). A stratified Mangat and Singh's optional randomized response model using proportional and optimal allocation. *Statistica*, 74(1), 65–83.

Waldman, I.D., Robinson, B.F., Rowe, D.C., (1999). A logistic regression based extension of the TDT for continuous and categorical traits. *Annals of Human Genetics*, 63, 329-340.

Warner, S.L., (1965). Randomized response: a survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, 60, 63-69.