DigitalCommons @WayneState

Journal of Modern Applied Statistical Methods

Volume 7 | Issue 1 Article 11

5-1-2008

Logit Estimation Using Warner's Randomized Response Model

Zawar Hussain
Quaid-i-Azam University, zhlangah@yahoo.com

Javid Shabbir
Quaid-i-Azam University, jsqau@yahoo.com

Journal of Modern Applied Statistical Methods
May, 2008, Vol. 7, No. 1, 140-151

1538 – 9472/08/$95.00

Logit Estimation Using Warner's Randomized Response Model

Zawar Hussain Javid Shabbir
Quaid-i-Azam University

A modified hidden logit estimation procedure is presented based on Warner (1965) randomized response model. Monte Carlo simulations explore the behavior of this estimator and compare its performance with the ordinary logits estimator. Warner's model is more protective and less jeopardizing.

Key words: Logit estimation, randomized response, sensitive character.

Introduction

Binary data have been used quite frequently in econometric modeling. In the early days of econometrics these data were on the explanatory variables named as dummy variables. The development of linear and nonlinear econometrics, now, provided the ways to analyze the discrete dependent variables in regression models. They lead to the probit model and logit model. One of the assumptions in these procedures is that the empirical observations on dichotomous dependent variables are real reflections of the true values of the dependent variable. This is somewhat unrealistic assumption when modeling self-reported data on sensitive topics, such as when survey respondents are asked about embarrassing behavior, or illegal activities. Innocuous questions receive higher response rates than questions on sensitive items, particularly on those involving perceived stigmatizing matters.

Zawar Hussain, Ph. D., is a Lecturer in the Department of Statistics, Quaid-i-Azam. His research interests include randomized response models, unequal probability sampling, Regression analysis, Bayesian estimation, Item count techniques. Email: zhlangah@yahoo.com. Javid Shabbir is Assocaite Professor at the University of Kent, UK, and Head of Department of Statistics, Quaid-i-Azam University. His research interests include randomized response models, unequal probability sampling, Sampling techniques. Email: jsqau@yahoo.com.

The latter often results in either refusal to respond or falsified answers. Due to this, non-response error is introduced and results in the unreliable estimation of population parameters of the interest. The reason of falsification of answer or refusal to answer might be the incentives for the survey respondents in the form of not getting embarrassed or not to be stigmatized. Corstange (2004) noted, "If the problem is that people have incentives to hide their true opinions or behavior from the interviewer, then our science suffers unless we can develop means to nullify these incentives. Survey respondents may not be willing to reveal their true answers to sensitive questions without foolproof guarantees of anonymity – not only from outside observers such as law enforcement or friends and family, but even from the interviewers themselves" (p. 5).

To nullify these adverse incentives, Corstange (2004) discussed changing the wordings of the sensitive question. But changing the statement of the question is actually changing the question and revised statements may not fully deliver the true underlying concept we hope to measure. As a means of guaranteeing anonymity to the respondent, consider Warner's (1965) randomized response model.

The randomized response models originated with Warner (1965), a statistician by discipline, and have since been improved upon by various others. Corstange (2004) stated that surprisingly enough, the procedure was almost entirely unknown among political scientists: other than a few brief research notes published

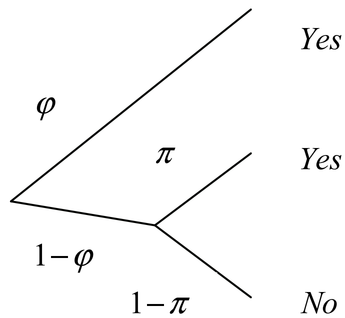


Figure 1. Graphical representation of Corstange (2004) model.

in the late 1970s, randomized response remains relatively terra incognita to the discipline. The reason of this unpopularity of randomized response among the psychologists and politicians might be that, formerly, at best they could estimate population means rather than explanatory models. In other words, they were only able to estimate the proportion of respondents who evaded taxes in the last year without being able to estimate the effects of other characteristics such as family size, race, and number of earning hands, locality, and socio-economic status on tax evasion.

Corstange's (2004) Hidden Logits

The randomized response model used by Corstange (2004) is as follows: Consider the following procedure to a yes/no question where "yes" the sensitive answer is: the respondent flips a coin and does not reveal the result to the interviewer. If the coin comes up *heads,* the respondent answers "yes" unconditionally, but if the coin comes up *tails,* the respondent answers the given yes/no question. Under these conditions, the interviewer does not know – and will never know – whether a "yes" response came as a result of a *heads* or as an answer to the question being asked. Generally, if φ is the probability of an unconditional "yes" *response* (in the example, φ =.50, the probability of getting *heads*) and ($1-\varphi$) is the probability of an actual *answer* (either "yes" or "no"), then we can represent the extensive form of the possible outcomes as in Figure 1.

From the above displayed data generating process

$$prob(yes) = \hat{\pi} = \varphi + (1-\varphi).\pi \qquad (1)$$

On simplification,

$$\pi = \frac{\hat{\pi} - \varphi}{1-\varphi} \qquad (2)$$

In ordinary logit models

$$\ln\left(\frac{\pi_i}{1-\pi_i}\right) = \boldsymbol{x}_i\beta \qquad (3)$$

where $\boldsymbol{x}_i$ is the row vector of observations on explanatory variables and β is the column vector of parameters. From equation (3) we can observe that estimation of the $\beta's$ is not possible because there are not any data on π_i. The only data available are on the explanatory variables and $\hat{\pi}_i$. Therefore, in order to move further express the logit model in terms of information available (i.e., $\hat{\pi}_i$).

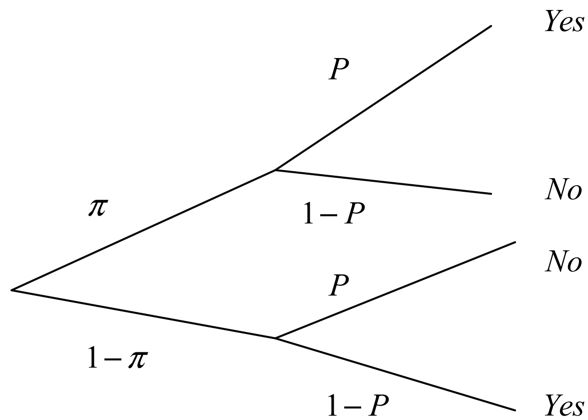


Figure 2 : Graphical representation of Warner's(1965) RRM.

Substituting (2) in (3) and solving for π

$$\hat{\pi}_i = \frac{e^{x_i\beta} + \varphi}{1-\varphi} \qquad (4)$$

Using equation (4) it is possible to estimate our parameters of interst, $\beta's$, and thereafter the logits by maximum likelihood method. As shown by Corstange (2004), setting the derivatives of the likelihood function equal to zero maximizes the likelihood function but that equation cannot be solved analytically. Therefore, it is solved numerically.

Deviation of Modified Hidden Logit

The Warner's (1965) randomized response model provides more privacy and anonymity to the respondents than provided by the randomized response model used by Corstange (shown below). Our Modified hidden logits are based on Warner's (1965) randomized response model. Warner's (1965) randomized response device consists of two complimentary statements, say, A and A^c. The statements A and A^c are presented with probabilities P and 1- P respectively. The respondents are required to select one of the two statements randomly and answer yes or no according to their true status. The extensive form of the outcomes of Warner's device is shown in Fig.2.

The probability of a yes answer in Warner's (1965) device is

$$Prob(yes) = P.\pi + (1-P).(1-\pi) = \theta \text{ (say).} \qquad (5)$$

Then using the steps of equations (2) and (3)

$$\pi = \frac{\theta - (1-2P)}{(2p-1)} \qquad (6)$$

On substituting equation(6) in equation(3)

$$\theta_i = \frac{P.e^{x_i\beta} + (1-P)}{1+e^{x_i\beta}} \qquad (7)$$

For $P = 1.0$ it becomes the ordinary logits derived from direct response.

Because of the interest in the estimation of $\tilde{\beta}'s$, and all the information available is the observed probability of "yes" response, θ_i ,the estimation is conducted using θ_i.

Suppose y_i is a binary random variable taking two values,' 0 '(no) and '1'(yes) with probabilities 1-θ_i and θ_i respectively, then

Table 1.(a)

$N = 1000$	$\tilde{\beta}_i$	Ordinary Logit $P = 1.0$	Modified hidden Logit. $P = 0.10$	Modified hidden Logit. $P = 0.20$	Modified hidden Logit. $P = 0.25$	Modified hidden Logit. $P = 0.40$
x_0	0.0	0.00046 (0.1071)*	-0.004 (0.138)	0.006 (0.193)	0.0078 (0.244)	0.0056 (0.245)
x_1	1.0	1.014 (0.0914)	1.018 (0.132)	1.040 (0.201)	1.075 (0.274)	1.081 (0.3001)
x_2	1.0	1.014 (0.093)	1.019 (0.129)	1.037 (0.2009)	1.0706 (0.272)	1.082 (0.299)
x_3	1.0	1.012 (0.093)	1.018 (0.1302)	1.038 (0.2013)	1.0182 (0.279)	1.034 (0.2987)

Table 1. (b)

N=2000	$\tilde{\beta}_i$	Ordinary Logit $P = 1.0$	Modified hidden Logit. $P = 0.10$	Modified hidden Logit. $P = 0.20$	Modified hidden Logit. $P = 0.30$	Modified hidden Logit. $P = 0.40$
X_0	0.00	-0.0001 (0.070)	-0.001 (0.090)	0.0003 (0.125)	0.0015 (0.200)	0.0016 (0.211)
X_1	1.00	1.008 (0.064)	1.011 (0.092)	1.019 (0.136)	1.051 (0.231)	1.055 (0.223)
X_2	1.00	1.006 (0.064)	1.010 (0.092)	1.018 (0.135)	1.051 (0.228)	1.054 (0.311)
X_3	1.00	1.006 (0.063)	1.011 (0.091)	1.019 (0.136)	1.051 (0.233)	1.049 (0.291)

Table 1. (c)

N=5000	$\tilde{\beta}_i$	Ordinary Logit $P = 1.0$	Modified hidden Logit. $P = 0.10$	Modified hidden Logit. $P = 0.20$	Modified hidden Logit. $P = 0.30$	Modified hidden Logit. $P = 0.40$
X_0	0.00	0.0006 (0.046)	0.0004 (0.059)	0.00004 (0.081)	-0.001 (0.125)	0.0016 (0.192)
X_1	1.00	1.0010 (0.040)	1.001 (0.057)	1.005 (0.082)	1.016 (0.131)	1.025 (0.183)
X_2	1.00	1.002 (0.039)	1.001 (0.056)	1.005 (0.082)	1.017 (0.132)	1.024 (0.194)
X_3	1.00	1.002 (0.040)	1.002 (0.056)	1.007 (0.080)	1.019 (0.132)	1.029 (0.165)

Table 1. (d)

N=10000	$\tilde{\beta}_i$	Ordinary Logit $P=1.0$	Modified hidden Logit. $P=0.10$	Modified hidden Logit. $P=0.20$	Modified hidden Logit. $P=0.30$	Modified hidden Logit. $P=0.40$
X_0	0.00	0.0001 (0.031)	0.0019 (0.042)	0.0013 (0.058)	0.0015 (0.089)	-0.0081 (0.200)
X_1	1.00	1.001 (0.028)	1.001 (0.040)	1.002 (0.057)	1.006 (0.092)	1.061 (0.212)
X_2	1.00	1.001 (0.028)	1.002 (0.038)	1.004 (0.056)	1.008 (0.091)	1.060 (0.199)
X_3	1.00	1.0004 (0.028)	1.0009 (0.040)	1.001 (0.055)	1.004 (0.090)	1.071 (0.187)

given y_i, the likelihood function of β is given by

$$L(\beta / y_i) = \prod_{i=1}^{n} \theta_i^{y_i} (1-\theta_i)^{y_i} \qquad (8)$$

and by taking natural logarithm on both sides

$$\ell = \ln L(\beta / y_i) = \sum_{i=1}^{n} \{y_i . \ln \theta_i + (1-y_i) . \ln(1-\theta_i)\} \qquad (9)$$

The first order derivative of above equation with respect to the parameter vector β is given by

$$\frac{\partial \ell}{\partial \tilde{\beta}} = \qquad (10)$$

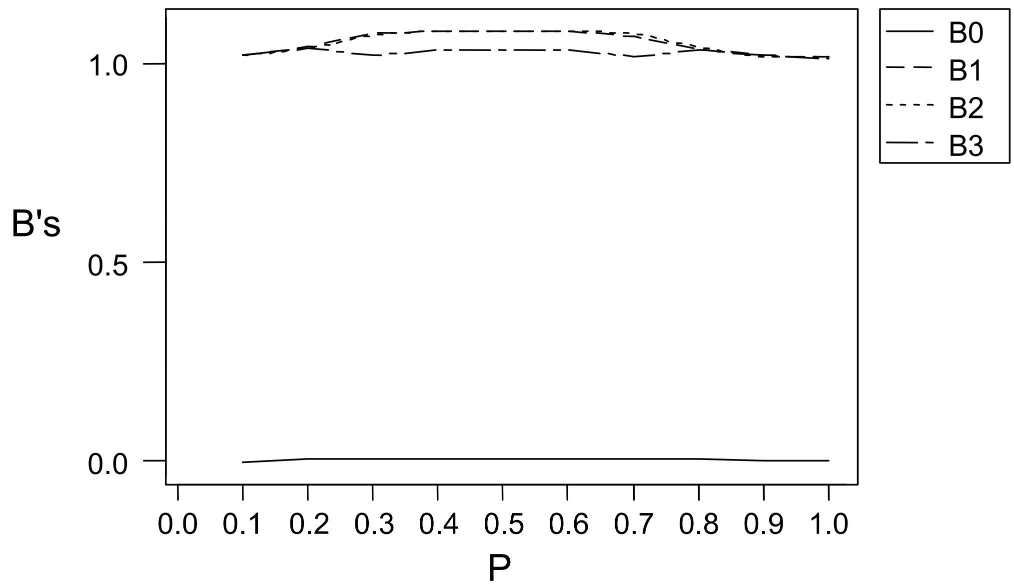
$$\sum_{i=1}^{n} \left\{ \begin{array}{l} y_i (P-1) \left(\dfrac{(3-2P+e^{x_i\tilde{\beta}})}{1-P+Pe^{x_i\tilde{\beta}})(2-P+(1-P)e^{x_i\tilde{\beta}})(1+e^{-x_i\tilde{\beta}})} \right) \\ - \dfrac{(P-1)}{((1-P)e^{x_i\tilde{\beta}}+2-P)(1+e^{-x_i\tilde{\beta}})} \end{array} \right\} x_i$$

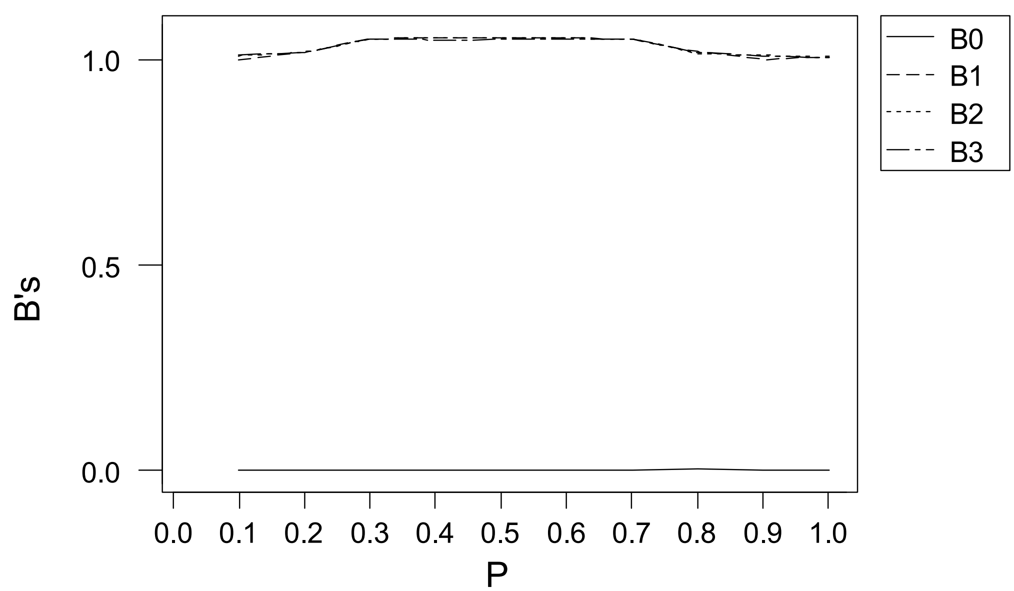
When this equation is set equal to zero it maximizes the log-likelihood function but this equation cannot be solved analytically (see appendix). Therefore, its numerical solution may be obtained.

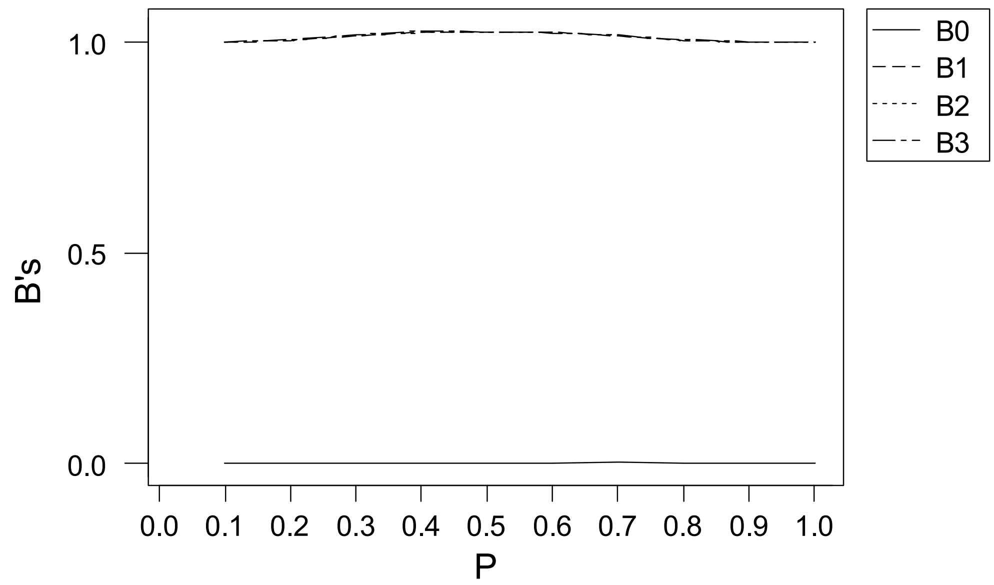
Comparison Of Modified Hidden Logits With Ordinary Logits

For comparison purposes, a small sample simulation study was conducted and results are given in table 1(a).The reason for small sample study is that the properties of consistency, normality and efficiency are well established for all maximum likelihood estimators (Green, 2000, & King, 1998). However, to see the pattern in the variances of β_i results for N= 2000, 5000, and 10000 are presented in Table 1(b,c,d).The data presented here were generated as follows. For each P ,1000, 2000, 5000 and 10000 samples were generated from a three regressors equation with no constant term. For simplicity, each β_i = 1. Also each $x_i \sim U(-3,3)$.

Given the above experimental conditions modified hidden logit return $\hat{b}_i$ that quite closely track the true population parameters β_i. The Table 1 (a, b, c, d) also compare the performance of the modified hidden logit estimator with ordinary logit (when $P=1.0$) at selected levels of P . From the Table 1(a, b, c, d) it is clear that modified hidden logit quite closely track the true $\beta 's$ but at the cost of increased variances.


B0
B1
B2
B3
B's
1.0
0.5
0.0
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
P


B0
B1
B2
B3
B's
1.0
0.5
0.0
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
P



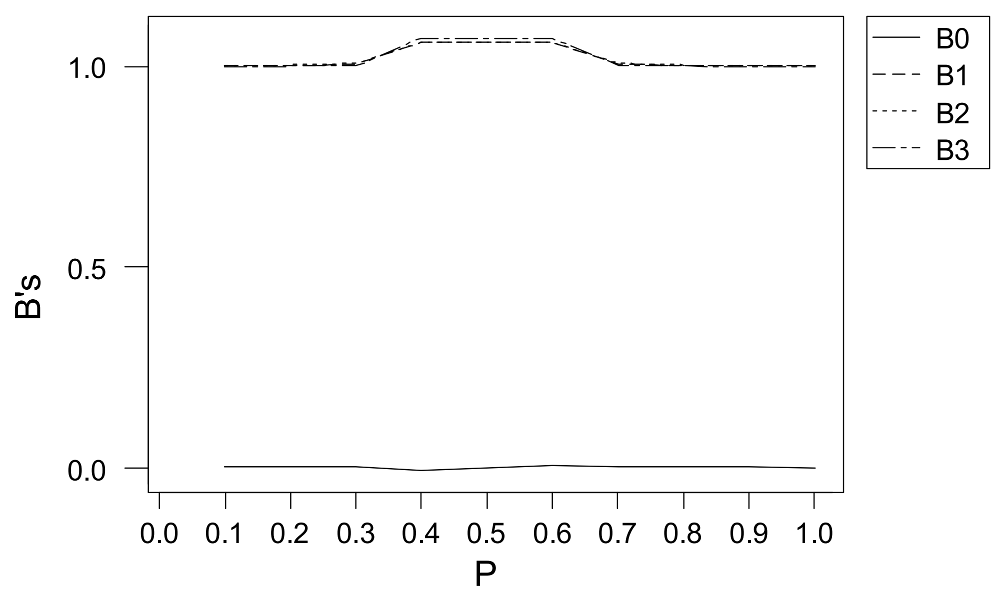


Figure 3. Graphs of $\hat{\beta}_i$'s against P for N = 1000, 2000, 5000 and 10000.

Behavior of $\beta's$ with P

From Figure 3 it is apparent that the modified hidden logit estimates of $\beta_i's$ deviate upward from the true $\beta_i's$ as p moves from 0.0 to 0.5 and then become close to the true parameters $\beta_i's$ as p increases from 0.5 to 1.0. An important point to remember is that the when $P = 0.5$ the estimators of $\beta_i's$ do not exist (as is the case of applying Warner's model to estimate π, the proportion of population with sensitive attribute). It is interesting to note that standard errors of the $\beta_i's$ are symmetric around $P = 0.5$. When the values of P moves away from 0.5 the standard errors of all the $\beta_i's$ decreases.

Respondent's Protection

Three basic concerns in randomized response models are jeopardy, suspicion, and efficiency. Jeopardy is the extent to which an affirmative answer implies the sensitive attribute; that is, the likelihood that the person has the attribute, given a yes response. In forced alternatives (answer either the sensitive or non-sensitive question), jeopardy increases as the probability that sensitive question was asked increases and the percentage of the population with the sensitive character decreases.

Suspicion is the extent to which a negative answer implies the sensitive attribute; that is, the likelihood that a person has the attribute, given a response. In forced alternatives (answer either the sensitive or non-sensitive question), suspicion increases as the probability that the sensitive question was asked decreases and the percentage of population with the non-sensitive character also decreases.

Efficiency is the loss in precision as a result of randomized response technique. It increases as the probability that the sensitive question was asked decreases.

In comparing the randomized response models emphasis has been on the variances. Greenberg, Abul-Ela, Simmons, and Horvitz(1969), Moors(1971), and Dowling and Shachtman(1975) are some of many to be referred. The emphasis on variances amounts to considering the matters from statistician's point of view only. Whereas the respondent's interest would be in the extent to which the different methods provide protection against their privacy. Leysieffer and Warner (1976), and Lanke (1975,76) provided the measures of protection provided by the different methods. Leysieffer and Warner (1976) proposed the natural measure of Jeopardy carried by a response R (either yes $=Y$ or no $= N$), about A and A^c respectively, which are as $g(R/A) = P(R/A)/P(R/A^c)$ and $g(R/A^c) = 1/g(R/A)$, where A and A^c are defined as above. These functions are called jeopardy functions. And the particular response R is jeopardizing if $g(R/A) = 1$.

Lanke(1976) proposed a measure of suspicion defined as

$$\psi = \max\left(P(A/Y), P(A/N)\right), \text{ where}$$

$$P(A/Y) \text{ and } P(A/N)$$

are conditional probabilities of belonging to a sensitive group A given a particular response Y or N, and proposed that a method is more protective for which

$$\psi = \max\left(P(A/Y), P(A/N)\right)$$

is smaller.

These two measures are calculated for both of the randomized response models used by Corstange(2004), and Warner(1965) which are as follows:

(i) For Warner,s model

$$g_w(Y/A) = \frac{P}{1-P}$$

and

$$\psi_w = \max\left(P(A/Y), P(A/N)\right)$$

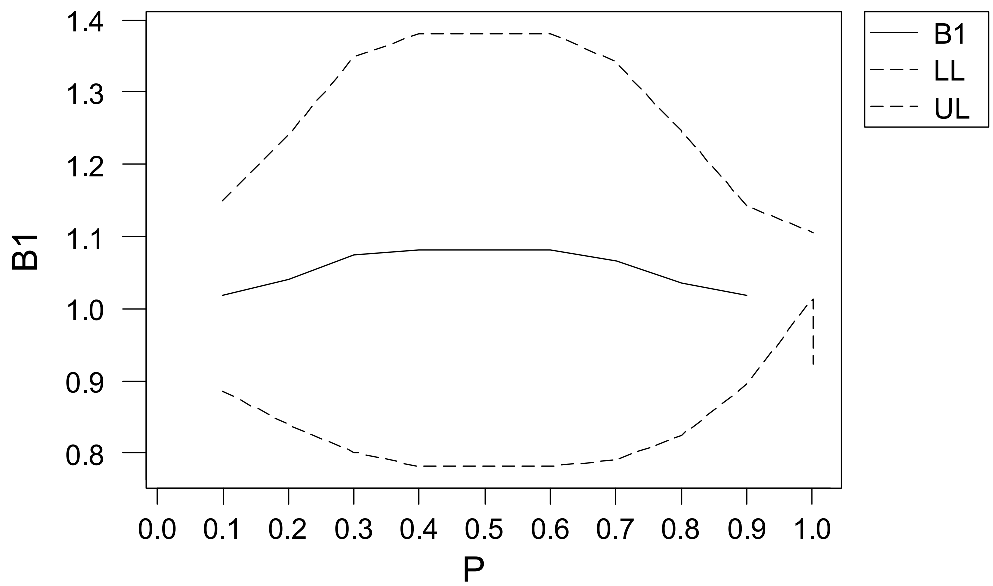
(ii) For Corstange model

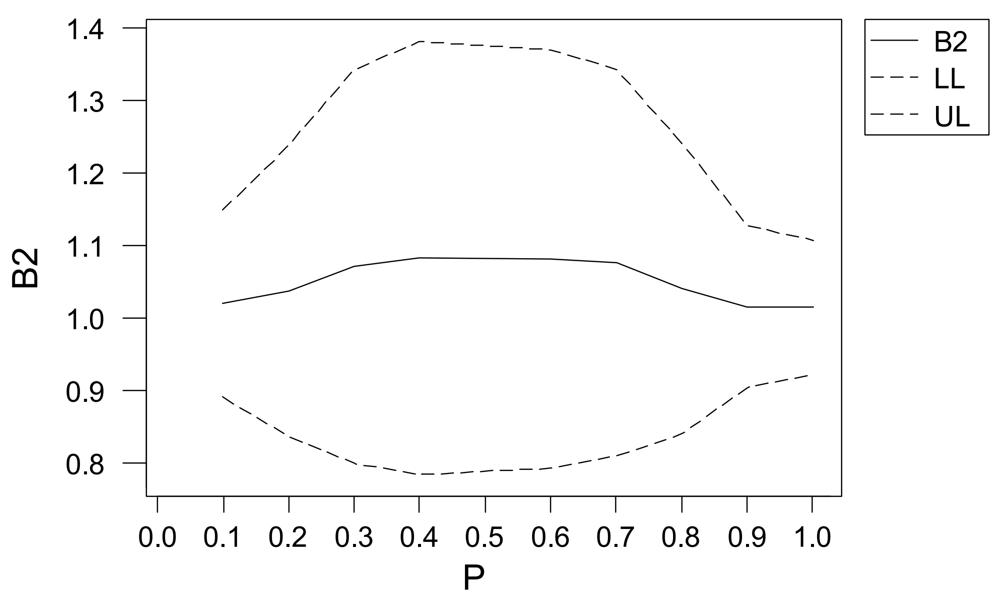
$$g_d(Y/A) = \infty$$

and

$$\psi_d = \max\left(P(A/Y), P(A/N)\right) = 1.$$

It can be seen that $g_w(Y/A) \le g_d(Y/A)$ and $\psi_w \le \psi_d$. It suggests that Warner's model is less jeopardizing and more protective.


B1
LL
UL
1.4
1.3
1.2
1.1
1.0
0.9
0.8
B1
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
P


B2
LL
UL
1.4
1.3
1.2
1.1
1.0
0.9
0.8
B2
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
P

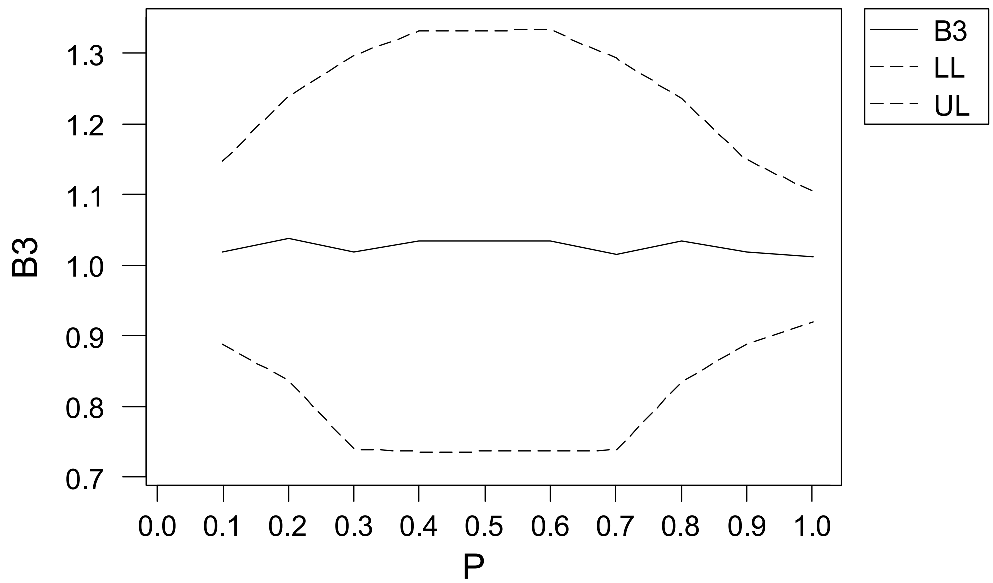


Figure 4. The behavior of standard errors of the $\hat{b}_i$'s with increasing values of *P* for $N = 1000$.

Choice of p

Setting a desirable value of P depends upon the nature of population. As we have just discussed above that that there are three major concerns of using randomized response techniques: jeopardy, suspicion and efficiency. Jeopardy increases with the increase in P and decrease in the proportion of population possessing sensitive character whereas suspicion increases with the decrease in P and the increase in the proportion of population possessing sensitive character. It has been showed that $g_w(Y/A) \le g_d(Y/A)$ and $\psi_w \le \psi_d$,so Warner's randomizing device is superior to that of Corstange's. Form table 2 it is apparent that the larger standard errors of each β_i when P is closer to 0.5. Thus a value farther from 0.5 should be set which seems desirable (e, g.0.3, 0.4, 0.6 or 0.7) and creates a balance between jeopardy and suspicion. In connection with isolated study of comparing the hidden logits based on Corstange (2004)'s model and Warner's(1965) model we suggest to set P at smaller level as it would provide more anonymity and would be less jeopardizing. As far as suspicion is concerned, Warner's (1965) model induces less suspicion for every P. So Warner's model would be a better choice as compared to Corstange (2004) model.

Fig. 4 presents the behavior of standard errors of the estimators $\hat{b}_i, i = 1,2,3.$ for different values of *P*. It can be easily seen that when *P* is closer to 0.5 the standard errors of the estimates are larger and setting *P* closer to 0.5 would induce unreliability in the estimates. Therefore, we suggest setting *P* away from 0.5. The same behavior of standard errors with respect to changes in *P* is observed for other values of *N*.

Discussion

As survey statisticians, our interest in sensitive topics inevitably leads us to ask sensitive questions. As this article shows, however, we must take care when we study such topics, especially when drawing inferences from self-reported, falsifiable answers to questions. By falsifying the true responses respondents get incentives by misrepresenting them. Randomized response as a questioning technique allows us, at least in principle, to nullify these incentives. The estimator developed here allows us to model questions of this nature, and simulations suggest that proceeding in this fashion allows us to draw more valid and more useful inferences about sensitive social issues.

References

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. Available at http://www.umich.edu/~dancorst.

Greene, W. (2000/1993). *Econometric Analysis*. (4th ed.). Upper Saddle River, NJ: Prentice Hall.

King G. (1998/1989. *Unifying political methodology*. Ann Arbor: University of Michigan Press.

Warner, S. (1965). Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias. *Journal of the American Statistical Association, 60*(309), 63-69.

Appendix.
Derivation of equation (4).

For

$$\ln\left(\frac{\pi_i}{1-\pi_i}\right) = x_i\beta$$

put the value of π_i from Warner's model and get:

$$\frac{\left(\frac{\theta_i-(1-P)}{(2P-1)}\right)}{\left(\frac{P-\theta_i}{(2P-1)}\right)} = e^{x_i\beta}$$

$$\frac{(\theta_i-(1-P))}{(P-\theta_i)} = e^{x_i\beta}$$

$$\theta_i-(1-P) = Pe^{x_i\beta} - \theta_i e^{x_i\beta}$$

$$\theta_i = \frac{Pe^{x_i\beta}+(1-P)}{(1+e^{x_i\beta})}$$

first derivative of θ_i with respect to true $\tilde{\beta}$

$$\frac{\partial\theta_i}{\partial\tilde{\beta}} = \frac{\partial\left(\frac{Pe^{x_i\tilde{\beta}}+(1-P)}{1+e^{x_i\tilde{\beta}}}\right)}{\partial\tilde{\beta}}$$

$$\frac{\partial\hat{\pi}}{\partial\tilde{\beta}} = \frac{\partial\left(\frac{Pe^{x_i\tilde{\beta}}}{1+e^{x_i\tilde{\beta}}}\right)}{\partial\tilde{\beta}} + \frac{\partial\left(\frac{1-P}{1+e^{x_i\tilde{\beta}}}\right)}{\partial\tilde{\beta}}$$

$$= \frac{\left(1+e^{x_i\tilde{\beta}}\right)Px_ie^{x_i\tilde{\beta}} - Px_ie^{2x_i\tilde{\beta}}}{\left(1+e^{x_i\tilde{\beta}}\right)^2} + \frac{-x_ie^{x_i\tilde{\beta}}}{\left(1+e^{x_i\tilde{\beta}}\right)^2}$$

$$= \frac{Px_ie^{x_i\tilde{\beta}} - x_ie^{x_i\tilde{\beta}}}{\left(1+e^{x_i\tilde{\beta}}\right)^2}$$

$$\frac{\partial\theta_i}{\partial\tilde{\beta}} = \frac{(P-1)x_ie^{x_i\tilde{\beta}}}{\left(1+e^{x_i\tilde{\beta}}\right)^2}$$

Likelihood function : observed θ_i with respect to true $\tilde{\beta}$.

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

$$\ell = \ln L(\tilde{\beta} / y_i) = \sum_{i=1}^{n}\{y_i.\ln\theta_i + (1-y_i).\ln(1-\theta_i)\}\cdot$$

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

$$\frac{\partial \ell}{\partial \tilde{\beta}}=\sum_{i=1}^{n}\left\{y_i\left(\frac{1+e^{x_i\tilde{\beta}}}{1-P+Pe^{x_i\tilde{\beta}}}\right)\left(\frac{(P-1)x_i e^{x_i\tilde{\beta}}}{(1+e^{x_i\tilde{\beta}})^2}\right)-(1-y_i)\left(\frac{1+e^{x_i\tilde{\beta}}}{(1-P)e^{x_i\beta}+2-P}\right)\left(\frac{(P-1)x_i e^{x_i\tilde{\beta}}}{(1+e^{x_i\tilde{\beta}})^2}\right)\right\}$$

$$\frac{\partial \ell}{\partial \tilde{\beta}}=\sum_{i=1}^{n}\left\{y_i\left(\frac{(P-1)x_i e^{x_i\tilde{\beta}}}{(1-P+Pe^{x_i\tilde{\beta}})(1+e^{x_i\tilde{\beta}})}\right)-(1-y_i)\left(\frac{(P-1)x_i e^{x_i\tilde{\beta}}}{((1-P)e^{x_i\tilde{\beta}}+2-P)(1+e^{x_i\tilde{\beta}})}\right)\right\}$$

$$\frac{\partial \ell}{\partial \tilde{\beta}}=\sum_{i=1}^{n}\left\{y_i x_i\left(\frac{(P-1)}{(1-P+Pe^{x_i\tilde{\beta}})(1+e^{-x_i\tilde{\beta}})}\right)+y_i x_i\frac{(P-1)}{((1-P)e^{x_i\tilde{\beta}}+2-P)(1+e^{-x_i\tilde{\beta}})}-\frac{(P-1)}{((1-P)e^{x_i\tilde{\beta}}+2-P)(1+e^{-x_i\tilde{\beta}})}\right\}$$

$$\frac{\partial \ell}{\partial \tilde{\beta}}=\sum_{i=1}^{n}\left\{y_i x_i(P-1)\left(\frac{(3-2P+e^{x_i\tilde{\beta}})}{1-P+Pe^{x_i\tilde{\beta}})(2-P+(1-P)e^{x_i\tilde{\beta}})(1+e^{-x_i\tilde{\beta}})}\right)-\frac{(P-1)}{((1-P)e^{x_i\tilde{\beta}}+2-P)}x_i(1+e^{-x_i\tilde{\beta}})^{-1}\right\}$$

$$\frac{\partial \ell}{\partial \tilde{\beta}}=\sum_{i=1}^{n}\left\{y_i(P-1)\left(\frac{(3-2P+e^{x_i\tilde{\beta}})}{1-P+Pe^{x_i\tilde{\beta}})(2-P+(1-P)e^{x_i\tilde{\beta}})(1+e^{-x_i\tilde{\beta}})}\right)-\frac{(P-1)}{((1-P)e^{x_i\tilde{\beta}}+2-P)}(1+e^{-x_i\tilde{\beta}})^{-1}\right\}x_i$$