Statistics and Applications {ISSN 2452-7395(online)}
Vol. 15, Nos. 1&2, 2016 (New Series), pp 37-62

An Empirical Estimator In Randomized Response Sampling

Michael Lee Johnson, Stephen A. Sedory and Sarjinder Singh
Department of Mathematics, Texas A&M University-Kingsville
Kingsville, TX 78363, USA

Received: January 16, 2017; Revised February 03, 2017; Accepted: February 09, 2017

Abstract

In this paper, an open question in the field of randomized response sampling is raised. An empirical estimator of the population proportion is suggested as a possible solution to the open question. By means of simulation studies the proposed empirical estimator is shown to be more efficient than the Warner (1965) and the Greenberg et al. (1969) estimators.

Key words: Estimation of proportion, sensitive characteristics, bias and relative efficiency.

1. Introduction

In this paper, an open question is suggested in the field of survey sampling. A possible solution, not necessarily the best solution is also suggested. The collection of data through personal interview surveys on sensitive issues such as induced abortions, drug abuse, and family income is a serious issue; see for example Fox and Tracy (1986), Gjestvang and Singh (2006), Chaudhuri (2011) and Chaudhuri and Christofides (2013). Warner (1965) considered the case where the respondents in a population Ω can be divided into two mutually exclusive groups: one group with stigmatizing/ sensitive characteristic A and the other group without it. For estimating π, the proportion of respondents in the population Ω belonging to the sensitive group A, a simple random sample s_1 of n_1 respondents is selected using with replacement sampling from the population. For collecting information on the sensitive characteristic, Warner (1965) made use of a randomization device. One such device is a deck of cards with each card bearing one of the following two statements:

(i) "I belong to group A", and (ii) "I do not belong to group A".

The statements (i) and (ii) occur with relative frequencies P and $(1-P)$, respectively, in the deck. Each respondent in the sample s_1 is asked to select a card at random from the well-shuffled deck. Without showing the card to the interviewer, the interviewee answers the question, "Is the statement true for you?" The number of people x_1 that answer "yes" is binomially distributed with parameters $\theta_1 = P\pi + (1-P)(1-\pi)$ and n_1. For large sample sizes, the maximum likelihood estimator of π for $P \neq 0.5$ and is given by:

$$\hat{\pi}_w = \frac{\hat{\theta}_1 - (1-P)}{2P-1} \tag{1.1}$$

Corresponding Author: Sarjinder Singh
E-mail: sarjinder.singh@tamuk.com

where $\hat{\theta}_1 = x_1/n_1$ is the observed proportion of 'yes' answers. Lee, Sedory and Singh (2013) have shown that if sample size is sufficiently large, then there is a very rare chance the estimator in (1.1) can take an inadmissible value outside the interval [0, 1]. The estimator $\hat{\pi}_w$ in (1.1) is unbiased for π and the variance of the estimator $\hat{\pi}_w$ is given by:

$$V(\hat{\pi}_w) = \frac{\pi(1-\pi)}{n_1} + \frac{P(1-P)}{n_1(2P-1)^2} \tag{1.2}$$

Note that in the Warner (1965) model, the two statements relate to groups that are perfectly negatively associated with each other; that is, one group is the complement of the other group in a population of interest. However, it is intuitively evident that to protect the confidentiality of a respondent it is not necessary for the two statements to be complementary. For example, one may use two unrelated statements (I belong to group A/ I belong to group Y). In fact, it is enough to make use of unrelated non-sensitive characteristics in a randomization device, as suggested by Greenberg *et al.* (1969). They proposed the pioneer unrelated questions model which is famous with the name of Simmons' unrelated question model. To our knowledge no solid evidence is available why this model had been names as Simmons' unrelated question model by Greenberg et al. (1969) or it came from Horvitz, Shah and Simmons (1967) work as Simmons' model. Assume a sample s_2 of n_2 respondents is selected by using simple random and with replacement sampling from a population Ω. In their model, each respondent in the sample s_2 should answer one of two questions, which are not related to each other. For example: a respondent is given a deck having cards with two types of questions: (i) Do you belong to group A?, and (ii) Do you belong to group Y?, with known relative frequencies T and $(1-T)$, respectively. Membership in group *Y* is assumed to be independent of membership in group *A*. Again let π be the true proportion of respondents in the population Ω who possesses sensitive characteristic *A*. Also let π_Y be the true proportion of respondents in the population Ω who possesses non-sensitive characteristic *Y*. This method also ensures the privacy of respondents during a face-to-face survey. In the Simmons' unrelated question model, the true probability of a 'yes' answer θ_2 is given by:

$$\theta_2 = T\pi + (1-T)\pi_Y . \tag{1.3}$$

Assuming π_Y is known, an unbiased estimator of π, due to Greenberg *et al.* (1969), is given by:

$$\hat{\pi}_G = \frac{\hat{\theta}_2 - (1-T)\pi_Y}{T} . \tag{1.4}$$

The variance of the estimator $\hat{\pi}_G$ is given by:

$$V(\hat{\pi}_G) = \frac{\theta_2(1-\theta_2)}{n_2 T^2} . \tag{1.5}$$

In the next section, we raise an open question for those who are working in the area of randomized response sampling.

2. An Open Question

Assume one company selected a sample s_1 of n_1 respondents from a population Ω by using the Warner (1965) model. Another company selected an independent sample s_2 of

n_2 respondents from the same population Ω by using the Greenberg *et al.* (1969) (or Simmons') unrelated question model. Later both companies found that they have common interest in estimating the population proportion π of a sensitive attribute A in the same population Ω. Both companies worked together and found that the value of the population proportion of the unrelated characteristic π_Y cannot be exactly known. There is no time and budget to estimate π_Y by doing another independent survey. For example, assume a new election is coming very soon. Now an open question arises, "Can the information from both samples s_1 and s_2 be used to improve the pioneer Warner (1965) estimator $\hat{\pi}_w$ defined in (1.1) assuming the exact value of π_Y remains unknown?" The answer to this open question is not obvious. The new estimator proposed below is an attempt to address such a situation.

3. Proposed Empirical Estimator

We suggest the following empirical estimator of the population proportion π as:

$$\hat{\pi}_e = \alpha\,\hat{\theta}_1 + \beta\hat{\theta}_2, \tag{3.1}$$

where α and β are constants empirically chosen based on certain criterions of interest to an investigator, such as reasonable relative bias and/or minimum mean squared error. We have the following theorems:

Theorem 3.1. The empirical estimator $\hat{\pi}_e$ is an inconsistent estimator of the population proportion π.

Proof. It is easy to verify that the bias $B(\hat{\pi}_e)$ in the estimator $\hat{\pi}_e$ is given by:

$$B(\hat{\pi}_e) = ((2P-1)\alpha + \beta T - 1)\pi + \alpha(1-P) + \beta(1-T)\pi_y, \tag{3.2}$$

which shows that the bias is not a decreasing function of the values of the sample sizes n_1 and n_2 for given experience based values of α and β. Thus the proposed empirical estimator is an inconsistent estimator of the population proportion π.

Theorem 3.2. The mean squared error of the proposed empirical estimator $\hat{\pi}_e$ is given by:

$$\begin{aligned}\mathrm{MSE}(\hat{\pi}_e) = {} & \alpha^2 \frac{1}{n_1}\left[(2P-1)^2\pi(1-\pi) + P(1-P)\right] + \beta^2 \frac{(T\pi + (1-T)\pi_y)}{n_2}\left[1 - T\pi - (1-T)\pi_y\right] \\ & + \left[((2P-1)\alpha + \beta T - 1)\pi + \alpha(1-P) + \beta(1-T)\pi_y\right]^2. \end{aligned} \tag{3.3}$$

Proof. Note that both samples s_1 and s_2 are independent, thus the variance of the proposed estimator $\hat{\pi}_e$ is given by:

$$V(\hat{\pi}_e) = \alpha^2 V(\hat{\theta}_1) + \beta^2 V(\hat{\theta}_2)$$

$$= \alpha^2 \frac{\theta_1(1-\theta_1)}{n_1} + \beta^2 \frac{\theta_2(1-\theta_2)}{n_2}. \tag{3.4}$$

By the definition of mean squared error, we have:

$$\mathrm{MSE}(\hat{\pi}_e) = V(\hat{\pi}_e) + \{B(\hat{\pi}_e)\}^2 \tag{3.5}$$

On using (3.2) and (3.4) in (3.5), we get (3.3).

4. How to Apply the Proposed Empirical Estimator

Note that the mean squared error of the proposed empirical estimator $\hat{\pi}_e$ is minimum if the values of the empirical constants α and β satisfy the ratio given below:

$$\frac{\alpha}{\beta}=\frac{n_1(1-\theta_2)}{n_2(1-\theta_1)}. \tag{4.1}$$

Thus for a given value of β, we can compute the value of α that would result in the minimum mean squared error of the proposed estimator, $\hat{\pi}_e$. Note that any arbitrary value of β can be used to yield the minimum mean squared error, but there may be trouble with the percent relative bias. A grid search was done for different choices of the empirical constants α and β that could take care of both the percent relative bias and mean squared error. The percent relative bias in the proposed estimator $\hat{\pi}_e$ is given by:

$$\mathrm{RB}(\hat{\pi}_e)=\frac{B(\hat{\pi}_e)}{\pi}\times 100\% \tag{4.2}$$

The percent relative efficiency of the proposed empirical estimator $\hat{\pi}_e$ with respect to the Warner (1965) estimator $\hat{\pi}_w$ is defined as:

$$\mathrm{RE}(1)=\frac{V(\hat{\pi}_w)}{\mathrm{MSE}(\hat{\pi}_e)}\times 100\% \tag{4.3}$$

The percent relative efficiency of the proposed empirical estimator $\hat{\pi}_e$ with respect to the Greenberg et al. (1969) estimator $\hat{\pi}_G$ is defined as:

$$\mathrm{RE}(2)=\frac{V(\hat{\pi}_G)}{\mathrm{MSE}(\hat{\pi}_e)}\times 100\% \tag{4.4}$$

Following Cochran (1963), choices of α and β were sought such that the absolute percent relative bias was less than 10% and the percent relative efficiencies, RE(1) and RE(2) were higher than 100%.

Three situations could arise: (i) The sample sizes n_1 and n_2 can be of equal sizes, (ii) the sample size n_1 can be greater than the sample size n_2, and (iii) the sample size n_1 can be smaller than the sample size n_2. The values of π and π_Y could be any real numbers between 0 and 1. Recall that our interest is to estimate the value of π regardless of the value of π_Y. For each value of π, we found optimum values of α and β a range of all values of $\pi_Y \in (0,1)$ under the three situations where the ratio of two sample sizes n_1 and n_2 is one, half or double: $n_1/n_2 = 1.0, 0.5$ or 2.0. In particular, we considered $n_1 = 250$ and 500 as well as $n_2 = 250$ and 500. A summary of those detailed results is presented in the Table 4.1 (Appendix-A). Table 4.1 also provides the frequency, mean, standard deviation, minimum, median and maximum value of each parameter when the proposed estimator has been found to be more efficient than the Warner (1965) and Greenberg et al. (1969) estimators and has absolute RB less than 10%. In Table 4.1, for example if $n_1 = 250$and $n_2 = 250$, there are 4652 values of the parameters where all three criterions are met; that is, absolute value of the percent RB is less than 10%, RE(1)>100% and RE(2)>100%. The RE(1) has a minimum value of 352.0%, maximum value of 13284.7% with a median value of 902.5%, and the

RE(2) has a minimum value of 101.1%, maximum value of 3536.3% with a median efficiency of 226.2%. The average values of RE(1) and RE(2) are 1454.2% and 343.3% with respective standard deviations 1783.3% and 382.9%. No doubt the percent RE(1) and RE(2) values are highly appreciable. In this study the value of α is between 0.1453 and 0.6677 and the value of β in the range between 0.1370 and 0.8390. Similarly a practice in choosing values of α and β is required if one wishes to use the proposed empirical estimator. In the same way, other results in Table 4.1 can be interpreted. The detailed results are given in Appendix-B via 20 pictorial presentations in Figure. 4.1 through Figure 4.20.

5. Practical Empirical Estimator

We suggest a new practical empirical estimator defined as:

$$\hat{\pi}_e^* = \hat{\alpha}\,\hat{\theta}_1 + \hat{\beta}\hat{\theta}_2 \tag{5.1}$$

where $\hat{\alpha}$ and $\hat{\beta}$ are determined utilizing information from both samples s_1 and s_2 by using the following relation:

$$\frac{\hat{\alpha}}{\hat{\beta}} = \frac{n_1(1-\hat{\theta}_2)}{n_2(1-\hat{\theta}_1)} \tag{5.2}$$

Note that $\hat{\theta}_1$ and $\hat{\theta}_2$ are known proportions of 'yes' answers from given samples s_1 and s_2 of sizes n_1 and n_2, respectively. Thus for any arbitrary practicable guess of $\hat{\beta}$, we can obtain the value of $\hat{\alpha}$ from the information in the both sample s_1 and s_2.

In the next section, we provide a simulation study which is more appropriate than a real survey in several ways. For example, in a real survey, a respondent: (i) may not be able to understand a randomized response procedure, (ii) could report untruthfully after using a randomization device, and (iii) might refuse to respond to a sensitive question, etc. The simulation considered in the next section is free from such difficulties and hence removes sources of distortions when comparing different randomized response models.

6. Simulation Study As Good As A Real Survey

Assume one company selected a sample s_1 of $n_1 = 250$ respondents from a population Ω of interest by using the Warner (1965) model with $P = 0.7$. Let $\pi = 0.1$ be the proportion of the sensitive characteristic to be estimated in the population Ω. Obviously the probability of a 'yes' answer in a single trial is computed as: $\theta_1 = P\pi + (1-P)(1-\pi) = 0.34$. We used the subroutine CALL RNBIN (NITR, N1, TH1, IR1) to generate a response from a binomial distribution with parameter $n_1 = 250$ and $\theta_1 = 0.34$. The observed number of 'yes' responses denoted as IR1 were then transformed to find a simulated estimate $\hat{\pi}_w(1) = (\text{IR1}/n_1 - (1-P))/(2P-1)$ of π by using the Warner (1965) estimator. We repeated the same procedure $\Theta = 2000 = \text{NITR(iterations)}$ times and obtained 2000 estimates of π as: $\hat{\pi}_w(j) = (\text{IR1}(j)/n_1 - (1-P))/(2P-1)$, for $j = 1,2,3,....,2000$. Assume another independent company selected a sample s_2 of $n_2 = 250$ respondents from the same population Ω of interest by using the Greenberg *et. al.* (1969) model with $\pi_y = 0.2$ (assumed known), and $T = 0.75$. Again let $\pi = 0.1$ be the proportion of the same sensitive characteristic to be estimated from the same population Ω. Obviously, the probability of a 'yes' answer in a single trial is computed as: $\theta_2 = T\pi + (1-T)\pi_y = 0.125$. We used the subroutine CALL RNBIN (NITR, N2, TH2, IR2) to generate a response from binomial distribution with parameter $n_2 = 250$ and $\theta_2 = 0.125$. The observed number of 'yes' responses, denoted as IR2, were then transformed to find a simulated estimate $\hat{\pi}_G(1) = (\text{IR2}/n_2 - (1-T)\pi_y)/T$ of π by using the Greenberg *et al.* (1969) model. We repeated the same experiment $\Theta = 2000$ times and obtained 2000 estimates of π as: $\hat{\pi}_G(j) = (\text{IR2}(j)/n_2 - (1-T)\pi_y)/T$, for $j = 1,2,3,....,\Theta$. We computed the empirical variances of the Warner (1965) and Greenberg et al. (1969) estimator over the $\Theta = 2000$ iterations as:

$$V^*(\hat{\pi}_w) = \frac{1}{\Theta}\sum_{j=1}^{\Theta}(\hat{\pi}_w(j) - \pi)^2 \tag{6.1}$$

and
$$V^*(\hat{\pi}_G) = \frac{1}{\Theta}\sum_{j=1}^{\Theta}(\hat{\pi}_G(j) - \pi)^2 \tag{6.2}$$

For the given values of $n_1 = 250$, $n_2 = 250$, and $\hat{\beta} = 0.090$ (say, in particular), we computed the value of $\hat{\alpha} = \hat{\beta}\dfrac{n_1(1 - IR2(j))}{n_2(1 - IR1(j))}$.

For 2000 different $\hat{\alpha}$ values, the jth empirical estimator of π was:

$$\hat{\pi}_e^*(j) = \hat{\alpha}\frac{\text{IR1}(j)}{n_1} + \hat{\beta}\frac{\text{IR2}(j)}{n_2},\ j = 1,2,...,\Theta = 2000. \tag{6.3}$$

Next we computed the mean squared error of the proposed empirical practical estimator $\hat{\pi}_e^*$ as:

$$\text{MSE}^*(\hat{\pi}_e^*) = \frac{1}{\Theta}\sum_{j=1}^{\Theta}(\hat{\pi}_e^*(j) - \pi)^2 . \tag{6.4}$$

We also computed the value of the percent relative bias (RB*) in the proposed practical empirical estimator as:

$$\text{RB}^*(\hat{\pi}_e) = \frac{\frac{1}{\Theta}\sum_{j=1}^{\Theta}\hat{\pi}_e^*(j) - \pi}{\pi} \times 100\% \tag{6.5}$$

The percent relative efficiencies of the proposed practical estimator over the Warner (1965) and the Greenberg *et al.* (1969) estimators are, respectively, computed as:

$$\text{RE}^*(1) = \frac{V^*(\hat{\pi}_w)}{\text{MSE}^*(\hat{\pi}_e^*)} \times 100\% \tag{6.6}$$

and
$$\text{RE}^*(2) = \frac{V^*(\hat{\pi}_G)}{\text{MSE}^*(\hat{\pi}_e^*)} \times 100\% \tag{6.7}$$

Note the use of "*" on the top of RE* and RB* values, because these are simulated values based on 2000 iterations. For a given value of $\pi = 0.1$ (say), we did a grid search for the different values of β in the range -1 to +1 with a step of 0.005 and for all vales of π_y in the range of 0.1 to 0.9 with a step of 0.1. The grid search was based on the criterion that the absolute percent RB* value be less than 10% (Cochran, 1963), and that the RE*(1) and RE*(2)

values be higher than 100%. A summary of the detailed results is given in Table 6.1 (Appendix-A). The detailed results obtained are presented in Figure 6.1 through Figure 6.9 in Appendix-C. The results in Table 6.1 are very easily interpretable. For example, in the case $\pi = 0.1$, if one makes a good guess of $\hat{\beta}$ between 0.090 and 0.330 and computes the value of $\hat{\alpha}$, which is likely to be between 0.1867 and 0.2777 assuming the unknown value of π_y is 0.1, then the value of $RE^*(1)$ is likely to be between 1886% and 5822% with a median $RE^*(1)$ value of 3751%; the value of $RE^*(2)$ lies between 146.4% and 806.7% with a median value of 322.9%. The percent relative bias expected to be between -9.85% and 9.88%. Now if for $\pi = 0.1$, if one makes a good guess of $\hat{\beta}$ between 0.090 and 0.310 and computes the value of $\hat{\alpha}$ which is likely to be between 0.1659 and 0.2748 assuming the unknown value of π_y is 0.2, then the value of $RE^*(1)$ is likely to be between 1822% and 6597% with a median $RE^*(1)$ value of 3936%; the value of $RE^*(2)$ lies between 459.6% and 984.8% with a median value of 459.6%. The percent relative bias is likely to be expected between -8.67% and 9.87% and similar results for true for the value of π_y between 0.1 and 0.9. Note that the values of the percent relative efficiencies $RE^*(1)$ and $RE^*(2)$ are very much appreciable for the values of π between 0.1 and 0.6, however there are several situations where the required criterions are not met for π between 0.7 and 0.9. Further recall that that Greenberg et al. (1969) have suggested using their proposed unrelated question model when π is very close to π_y. Thus the proposed estimator $\hat{\pi}_e^*$ is likely to more efficient than the both the Warner (1965) and the Greenberg et al. (1969) estimator if π is likely to be close to the guessed value π_y. In general, the value of π is likely to be less than 0.5 because of the sensitive nature of the characteristic under study. We conclude that the proposed empirical estimator can help the companies to estimate efficiently by pooling information from both samples. We shall also keep our eyes open as to whether someone comes up with a better solution to the open question raised in this paper.

Acknowledgements

The authors are thankful to the Dr. Hukum Chandra and a referee for the comments on the original version of the manuscript.

References

Chaudhuri, A. (2011). *Randomized Response and Indirect Questioning Techniques in Surveys.* Chapman and Hall/CRC.

Chaudhuri, A. and Christofides, T.C. (2013). *Indirect Questioning in Sample Surveys*. Springer.

Cochran, W.G. (1963). *Sampling Techniques.* John Wiley and Sons.

Fox, J.A. and Tracy, P.E. (1986). *Randomized response: A method for sensitive surveys*. SAGE Publications.

Gjestvang, C. R. and Singh, S. (2006). A new randomized response model. *Journal of Royal Statistical Society*, **B68**, 523-530.

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

Horvitz, D.G., Shah, B.V., and Simmons, W.R. (1967). The unrelated question randomized response model. *In 1967 Social Statistics Section Proceedings of the A.S.A., ASA*, 65-72.

Lee, Cheon-Sig, Sedory, S.A. and Singh, S. (2013). Simulated minimum sample size for various randomized response models. *Communications in Statistics: Simulation and Computation*, **42(4)**, 771-789.

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

Appendix-A

Table 4.1: Search results for different values of the parameters considered in the investigation.

n_1	n_2	Parameter	freq	Mean	StDev	Minimum	Median	Maximum
250	250	π	4652	0.4909	0.2305	0.1	0.5	0.9
		π_y	4652	0.4983	0.2573	0.1	0.5	0.9
		α	4652	0.4421	0.1113	0.1453	0.4462	0.6675
		β	4652	0.4822	0.1606	0.1370	0.4820	0.8390
		RB(%)	4652	-0.097	4.410	-9.982	-0.103	9.999
		RE(1)	4652	1454.2	1783.3	352.0	902.5	13284.7
		RE(2)	4652	343.3	382.9	100.1	226.2	3536.3
	500	π	4616	0.4413	0.2339	0.1	0.4	0.9
		π_y	4616	0.4939	0.2567	0.1	0.5	0.9
		α	4616	0.2874	0.0783	0.0997	0.2862	0.4508
		β	4616	0.5951	0.1869	0.1880	0.6070	0.9230
		RB(%)	4616	-0.114	3.925	-9.995	-0.107	9.972
		RE(1)	4616	2437.9	2599.0	695.3	1546.2	19281.7
		RE(2)	4616	284.6	281.7	100.1	193.4	2566.3
500	250	π	3287	0.5131	0.2309	0.1	0.5	0.9
		π_y	3287	0.5072	0.2572	0.1	0.5	0.9
		α	3287	0.6011	0.1492	0.1887	0.6157	0.8881
		β	3287	0.3428	0.1331	0.0880	0.3350	0.7130
		RB(%)	3287	-0.074	4.432	-9.988	-0.096	9.999
		RE(1)	3287	919.0	1283.8	175.4	560.0	10220.9
		RE(2)	3287	434.31	549.31	100.1	273.4	5441.4
	500	π	3615	0.4652	0.2361	0.1	0.4	0.9
		π_y	3615	0.5005	0.2572	0.1	0.5	0.9
		α	3615	0.4306	0.1149	0.1453	0.4356	0.6600
		β	3615	0.4637	0.1661	0.1370	0.4600	0.8340
		RB(%)	3615	-0.069	3.893	-9.977	-0.076	9.999
		RE(1)	3615	1434.9	1726.9	346.3	891.4	13261.5
		RE(2)	3615	336.4	370.7	100.1	221.3	3530.1

Table 6.1: Good guess of the estimator $\hat{\beta}$ and the computed values of $\hat{\alpha}$, RB*(%), RE*(1) and RE*(2) for all $\pi_y \in [0.1, 0.9]$ and given values of $\pi \in [0.1, 0.9]$ for all possible four pairs of sample sizes n_1 and n_2 with values of 250 and 500.

π							π_y				
0.1			0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
		Freq	31	29	29	27	26	26	25	25	24
		Min.	0.1867	0.1659	0.1511	0.1458	0.1405	0.1217	0.1143	0.1036	0.0969
	Alpha	Med.	0.2250	0.2247	0.2130	0.1906	0.1877	0.1786	0.1715	0.1553	0.1508
		Max.	0.2777	0.2748	0.2770	0.2627	0.2585	0.2632	0.2213	0.2303	0.2319
		Min.	0.090	0.090	0.090	0.090	0.090	0.090	0.090	0.090	0.090
	Beta	Med.	0.185	0.185	0.180	0.170	0.170	0.165	0.165	0.160	0156
		Max.	0.330	0.310	0.295	0.280	0.270	0.255	0.245	0.235	0.225
		Min.	1886	1822	1980	1997	2265	1995	2481	2332	2466
	$RE^*(1)$	Med.	3751	3936	4277	4583	4840	5493	5459	6061	5977
		Max.	5822	6597	7674	8696	9921	10856	12178	13765	15147
		Min.	146.4	215.7	247.1	330.1	330.1	436.9	491.9	557	637
	$RE^*(2)$	Med.	322.9	459.6	576.0	747.1	747.1	1058.2	1213.8	1445	1569
		Max.	806.7	984.8	1212.4	1600.2	1771.3	2143.8	2433.9	2773	3037
		Min.	-9.85	-8.67	-9.53	-9.54	-8.49	-9.14	-8.68	-9.68	-8.54
	RB^*	Med.	-0.33	0.60	0.13	-0.49	0.33	0.10	-0.25	-0.63	0.34
		Max.	9.88	9.87	9.58	8.90	9.90	9.59	8.85	8.90	9.51
0.2	Freq		44	44	46	46	45	44	44	42	41

		Min.	0.2698	0.2308	0.2311	0.2098	0.2099	0.1848	0.1737	0.1610	0.1513
	Alpha	Med.	0.3790	0.3689	0.3314	0.3131	0.3132	0.2980	0.2800	0.2732	0.2490
		Max.	0.4919	0.4975	0.4838	0.4346	0.4384	0.4253	0.4039	0.4127	0.4058
		Min.	0.1550	0.1550	0.1550	0.1550	0.1550	0.1600	0.1600	0.1600	0.1600
	Beta	Med.	0.3050	0.3000	0.2975	0.2925	0.2900	0.2875	0.2800	0.2750	0.2700
		Max.	0.5000	0.4800	0.4700	0.4550	0.4400	0.4250	0.4100	0.3950	0.3850
		Min.	579.3	470.0	499.0	543.8	522.0	572.0	518.0	542.0	634.0
	$RE^{*}(1)$	Med.	1288.7	1382.2	1381.2	1431.0	1464.0	1589.0	1638.0	1774.0	1879.0
		Max.	2160.4	2529.5	2798.3	3186.7	3504.0	3896.0	4344.0	4778.0	5519.0
		Min.	102.8	104.1	105.9	108.8	124.5	132.2	155.7	185.1	171.9
	$RE^{*}(2)$	Med.	184.5	221.3	250.9	270.7	328.2	363.1	408.3	456.6	500.5
		Max.	404.7	500.4	554.4	611.6	643.0	764.8	875.3	985.8	1026.8
		Min.	-9.295	-8.567	-9.404	-9.748	-9.764	-9.707	-9.423	-9.819	-9.203
	RB^{*}	Med.	-0.420	0.234	0.245	-0.216	-0.232	0.418	-0.083	-0.209	0.071
		Max.	9.383	9.399	9.153	9.588	9.986	9.730	9.606	9.189	9.246
0.3	Freq		41	44	49	50	51	50	52	54	51
		Min.	0.3343	0.3214	0.2885	0.2683	0.2536	0.2375	0.2186	0.1957	0.1899
	Alpha	Med.	0.4635	0.4683	0.4290	0.4148	0.4042	0.4042	0.3510	0.3376	0.3262
		Max.	0.6446	0.6382	0.6441	0.5787	0.5617	0.5617	0.5328	0.5350	0.5352
		Min.	0.2050	0.2050	0.2100	0.2100	0.2150	0.2150	0.2150	0.2150	0.2200
	Beta	Med.	0.3800	0.3750	0.3750	0.3750	0.3700	0.3675	0.3675	0.3625	0.3600
		Max.	0.5950	0.5800	0.5700	0.5600	0.5450	0.5300	0.5250	0.5100	0.5000
		Min.	283.5	250.7	214.8	249.3	263.6	264.2	293.7	252.6	319.0
	$RE^{*}(1)$	Med.	750.2	770.2	829.1	854.5	862.7	921.2	930.5	912.2	1039.0
		Max.	1326.0	1504.0	1670.0	1879.1	2047.9	2378.9	2595.4	2929.3	3310.0
		Min.	100.38	102.70	101.64	101.74	102.2	109.6	101.1	105.7	111.4
	$RE^{*}(2)$	Med.	147.87	161.45	176.00	192.28	216.3	235.7	252.0	269.1	271.8
		Max.	257.89	297.22	340.15	364.54	413.8	418.3	451.4	462.5	538.2
		Min.	-8.615	-9.351	-9.634	-9.992	-9.430	-9.793	-9.352	-9.900	-9.213
	RB^{*}	Med.	-0.451	-0.530	-0.141	0.011	-0.178	-0.390	0.068	-0.110	-0.092
		Max.	7.455	8.261	9.729	9.729	9.927	9.247	8.695	9.777	9.215
0.4	Freq		33	37	40	42	46	48	49	50	50
		Min.	0.3582	0.3350	0.3265	0.3014	0.2639	0.2547	0.2455	0.2261	0.2124
	Alpha	Med.	0.5476	0.5333	0.4979	0.4866	0.4457	0.4336	0.4020	0.3895	0.3740
		Max.	0.7243	0.7240	0.7947	0.7027	0.6977	0.6345	0.6777	0.5663	0.6881
		Min.	0.2550	0.2550	0.2550	0.2600	0.2600	0.2650	0.2650	0.2700	0.2750
	Beta	Med.	0.4400	0.4400	0.4375	0.4425	0.4375	0.4350	0.4350	0.4350	0.4325
		Max.	0.6650	0.6550	0.6450	0.6350	0.6250	0.6150	0.6050	0.5900	0.5800
		Min.	215.9	208.1	191.8	190.1	185.9	190.6	173.8	189.0	218.3
	$RE^{*}(1)$	Med.	522.0	553.8	592.0	645.3	681.2	699.2	701.1	744.1	766.5
		Max.	972.9	1089.1	1207.0	1340.7	1511.0	1735.9	1915.7	2154.4	2443.8
		Min.	102.60	102.81	102.05	104.30	102.52	103.14	100.02	101.00	102.40
	$RE^{*}(2)$	Med.	125.27	139.82	149.74	157.16	171.84	182.79	184.90	196.20	221.40
		Max.	204.79	206.13	239.00	235.40	255.92	291.45	299.95	335.10	373.60
		Min.	-6.157	-7.044	-7.863	-7.191	-8.284	-7.925	-8.928	-8.538	-9.002
	RB^{*}	Med.	-0.494	-0.303	-0.831	0.035	-0.081	-0.298	-0.163	-0.171	-0.331
		Max.	5.516	5.979	6.370	6.917	7.482	7.654	8.073	7.995	7.290
0.5	Freq		17	26	29	34	37	37	40	42	44
		Min.	0.4101	0.3462	0.3296	0.3221	0.2746	0.2881	0.2677	0.2080	0.2122
	Alpha	Med.	0.5897	0.5930	0.5458	0.5273	0.4919	0.4661	0.4367	0.4014	0.3894
		Max.	0.8078	0.7842	0.7927	0.8063	0.7617	0.7453	0.7152	0.7300	0.6636
		Min.	0.3000	0.3000	0.3050	0.3100	0.3150	0.3200	0.3250	0.3300	0.3350
	Beta	Med.	0.4950	0.4950	0.4950	0.5025	0.5000	0.5000	0.5025	0.5000	0.5025
		Max.	0.7150	0.7050	0.7000	0.6950	0.6950	0.6750	0.6700	0.6600	0.6500
		Min.	206.2	181.4	186.6	194.9	183.8	192.2	196.6	200.5	190.2
	$RE^{*}(1)$	Med.	401.9	430.4	467.9	506.5	545.1	573.7	651.7	681.6	721.1
		Max.	780.0	862.9	960.2	1097.4	1211.2	1370.7	1564.9	1839.7	2064.4
		Min.	101.01	100.92	101.72	102.27	103.43	107.40	102.23	100.35	101.63
	$RE^{*}(2)$	Med.	113.60	120.06	127.75	137.14	143.28	159.55	163.81	169.82	178.14
		Max.	128.36	154.82	173.93	167.25	184.81	197.92	247.43	251.80	273.74
		Min.	-3.518	-5.186	-5.229	-5.188	-5.806	-5.736	-5.649	-6.767	-6.243
	RB^{*}	Med.	-0.376	-0.778	-0.517	-0.416	-0.607	-0.611	-0.267	-0.348	-0.078
		Max.	2.864	2.902	4.221	4.298	4.931	4.680	5.640	5.110	6.270
0.6	Freq		1	6	17	22	26	29	31	32	33

		Min.	0.7031	0.4459	0.3305	0.3438	0.3085	0.2860	0.2569	0.2376	0.2010
	Alpha	Med.	0.7031	0.7195	0.5863	0.5132	0.4984	0.4567	0.4363	0.4293	0.3729
		Max.	0.7031	0.8719	0.8363	0.7530	0.8034	0.7900	0.7555	0.7007	0.6372
		Min.	0.3500	0.3500	0.3550	0.3650	0.3700	0.3800	0.3850	0.3950	0.4050
	Beta	Med.	0.3500	0.4525	0.5500	0.5550	0.5600	0.5600	0.5650	0.5675	0.5700
		Max.	0.3500	0.7550	0.7450	0.7450	0.7400	0.7400	0.7300	0.7200	0.7150
		Min.	172.19	175.00	175.9	187.9	189.6	180.7	180.3	179.6	234.5
	$RE^*(1)$	Med.	172.19	266.00	381.6	418.1	459.6	506.6	556.0	601.4	689.1
		Max.	172.19	706.00	812.7	924.6	1041.1	1220.3	1400.7	1657.7	1886.6
		Min.	109.38	100.63	101.31	101.44	102.33	101.18	101.44	104.09	101.12
	$RE^*(2)$	Med.	109.38	104.50	109.37	114.87	123.20	129.20	139.71	143.32	148.56
		Max.	109.38	113.03	125.95	139.96	148.00	170.92	185.13	208.02	233.12
		Min.	0.1384	-2.144	-2.820	-3.369	-3.869	-4.239	-3.761	-4.271	-4.579
	RB^*	Med.	0.1384	-0.961	-0.015	-0.262	-0.297	-0.014	-0.246	0.031	-0.364
		Max.	0.1384	0.671	1.352	1.924	2.443	3.154	3.308	3.764	4.067
0.7	Freq		-	-	1	4	9	15	19	21	21
		Min.	-	-	0.3686	0.3549	0.2555	0.2496	0.2467	0.1984	0.1773
	Alpha	Med.	-	-	0.3686	0.3735	0.3283	0.4126	0.4000	0.3742	0.3204
		Max.	-	-	0.3686	0.4038	0.5699	0.6068	0.5218	0.4953	0.4459
		Min.	-	-	0.7950	0.7800	0.6100	0.6100	0.6100	0.6150	0.6200
	Beta	Med.	-	-	0.7950	0.7875	0.7750	0.6300	0.6350	0.6400	0.6450
		Max.	-	-	0.7950	0.7950	0.7900	0.7900	0.7850	0.7800	0.7750
		Min.	-	-	715.06	775.22	380.40	410.00	395.80	400.0	488.0
	$RE^*(1)$	Med.	-	-	715.06	789.53	785.50	449.00	507.80	587.4	687.0
		Max.	-	-	715.06	817.46	928.40	1135.4	1327.3	1599.3	1944.0
		Min.	-	-	113.74	102.74	102.06	105.35	101.54	102.78	109.13
	$RE^*(2)$	Med.	-	-	113.74	105.28	112.97	113.00	118.16	124.66	142.40
		Max.	-	-	113.74	117.01	123.62	155.11	160.62	193.39	211.28
		Min.	-	-	-0.225	-1.158	-1.699	-2.118	-2.605	-2.650	-2.703
	RB^*	Med.	-	-	-0.225	-0.239	-0.251	-0.123	-0.472	-0.146	-0.381
		Max.	-	-	-0.225	0.717	1.634	1.728	2.024	2.193	2.388
0.8	Freq		-	-	-	-	3	5	8	13	15
		Min.	-	-	-	-	0.2877	0.2471	0.2147	0.1737	0.1527
	Alpha	Med.	-	-	-	-	0.2931	0.2925	0.2529	0.2841	0.2745
		Max.	-	-	-	-	0.3180	0.3270	0.3943	0.4526	0.3872
		Min.	-	-	-	-	0.8350	0.8250	0.7000	0.7000	0.7100
	Beta	Med.	-	-	-	-	0.8400	0.8350	0.8275	0.8150	0.7250
		Max.	-	-	-	-	0.8450	0.8450	0.8450	0.8450	0.8400
		Min.	-	-	-	-	884.6	928.2	447.00	467.0	561.0
	$RE^*(1)$	Med.	-	-	-	-	887.3	1057.1	1102.0	1037.0	677.0
		Max.	-	-	-	-	932.5	1105.3	1380.0	1746.0	2307.0
		Min.	-	-	-	-	104.00	101.84	101.24	100.00	100.64
	$RE^*(2)$	Med.	-	-	-	-	105.89	118.15	114.45	108.72	115.55
		Max.	-	-	-	-	107.89	123.52	138.00	163.68	189.43
		Min.	-	-	-	-	-0.671	-1.399	-1.651	-1.714	-1.877
	RB^*	Med.	-	-	-	-	-0.115	-0.339	0.155	0.117	0.041
		Max.	-	-	-	-	0.563	1.043	1.563	1.877	1.688
0.9	Freq		-	-	-	-	1	3	4	5	5
		Min.	-	-	-	-	0.2468	0.2125	0.1672	0.1401	0.1174
	Alpha	Med.	-	-	-	-	0.2468	0.2250	0.1989	0.1628	0.1456
		Max.	-	-	-	-	0.2468	0.2251	0.2166	0.2017	0.1648
		Min.	-	-	-	-	0.9050	0.9000	0.8950	0.8900	0.8900
	Beta	Med.	-	-	-	-	0.9050	0.9050	0.9025	0.9000	0.9000
		Max.	-	-	-	-	0.9050	0.9100	0.9100	0.9100	0.9100
		Min.	-	-	-	-	969.95	1146.7	1500.4	1515.0	2002.0
	$RE^*(1)$	Med.	-	-	-	-	969.95	1222.4	1566.0	1566.0	3007.0
		Max.	-	-	-	-	969.95	1245.6	1677.4	1677.4	3809.0
		Min.	-	-	-	-	105.87	108.54	122.84	106.20	113.60
	$RE^*(2)$	Med.	-	-	-	-	105.87	114.26	129.82	132.60	171.50
		Max.	-	-	-	-	105.87	117.86	142.64	171.60	224.30
		Min.	-	-	-	-	0.1679	-0.335	-0.802	-1.362	-1.327
	RB^*	Med.	-	-	-	-	0.1679	0.286	0.005	-0.209	-0.196
		Max.	-	-	-	-	0.1679	0.711	0.738	0.817	0.997

Appendix-B

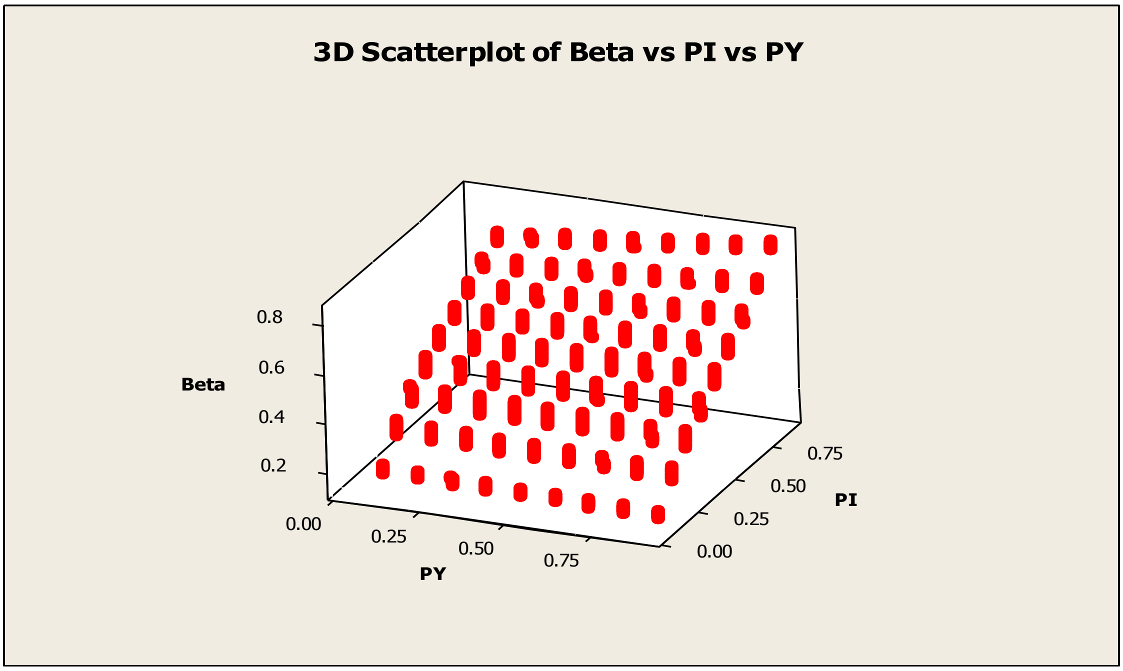

Figure 4.1: A set of choice of values of β for equal sample sizes $n_1 = n_2 = 250$ for different combinations of values of π and π_y in the range (0.1, 0.9) with a step of 0.1.

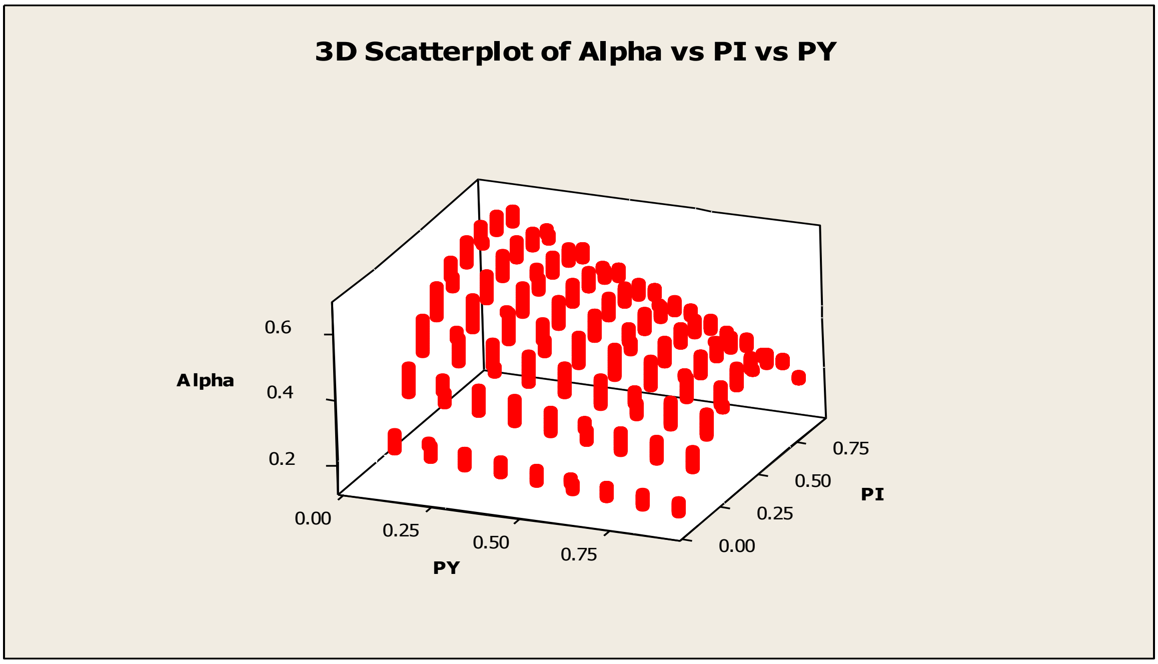

Figure 4.2: A set of choice of values of α for equal sample sizes $n_1 = n_2 = 250$ for different combinations of values of π and π_y in the range (0.1, 0.9) with a step of 0.1.

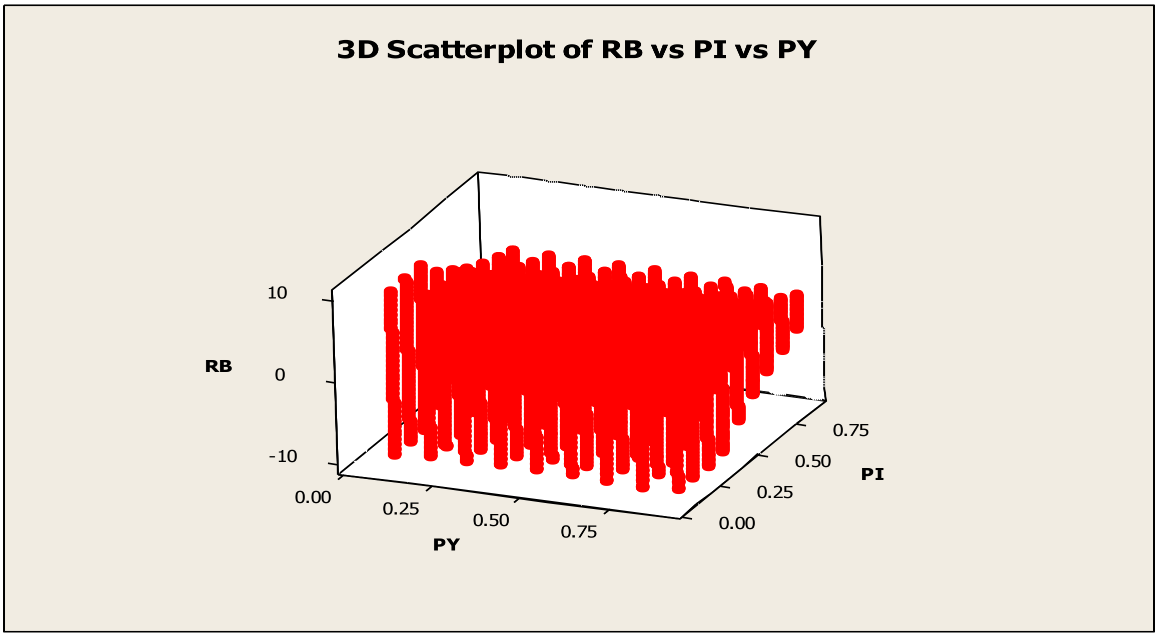

Figure 4.3: Percent relative bias (RB) for equal sample sizes $n_1 = n_2 = 250$ for different combinations of values of π and π_y in the range (0.1, 0.9) with a step of 0.1 and different choices of α and β.

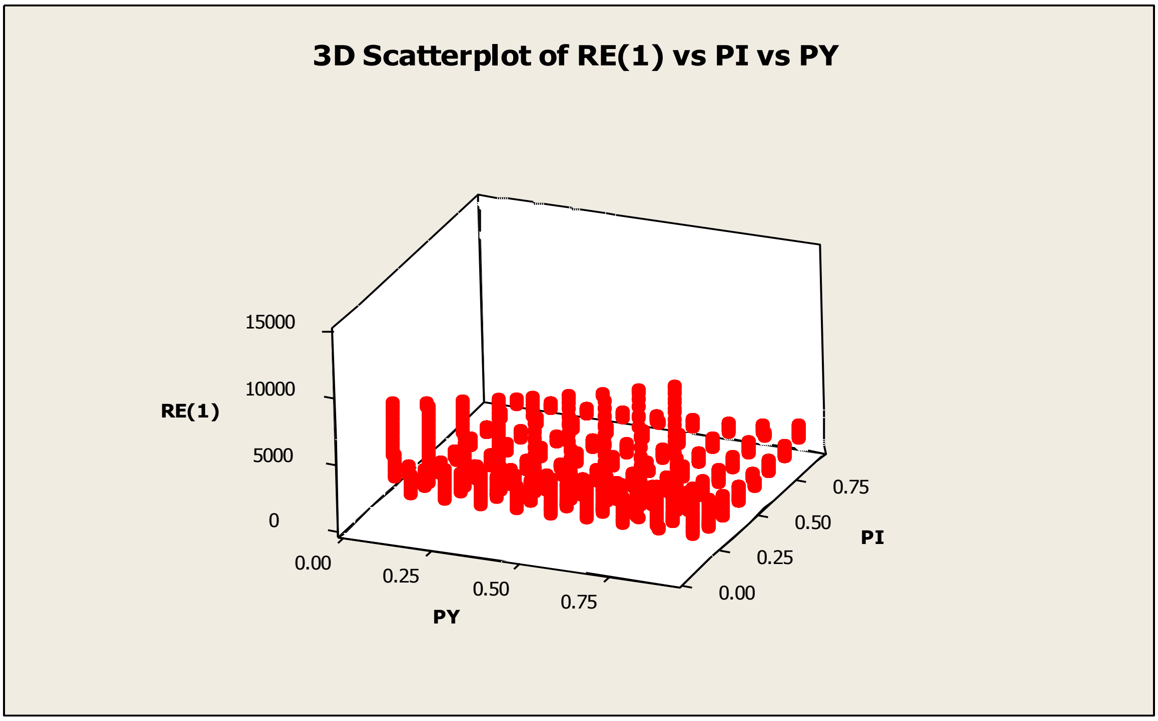

Figure 4.4: Percent relative efficiency RE(1) for equal sample sizes $n_1 = n_2 = 250$ for different combinations of values of π and π_y in the range (0.1, 0.9) with a step of 0.1 and different choices of α and β.

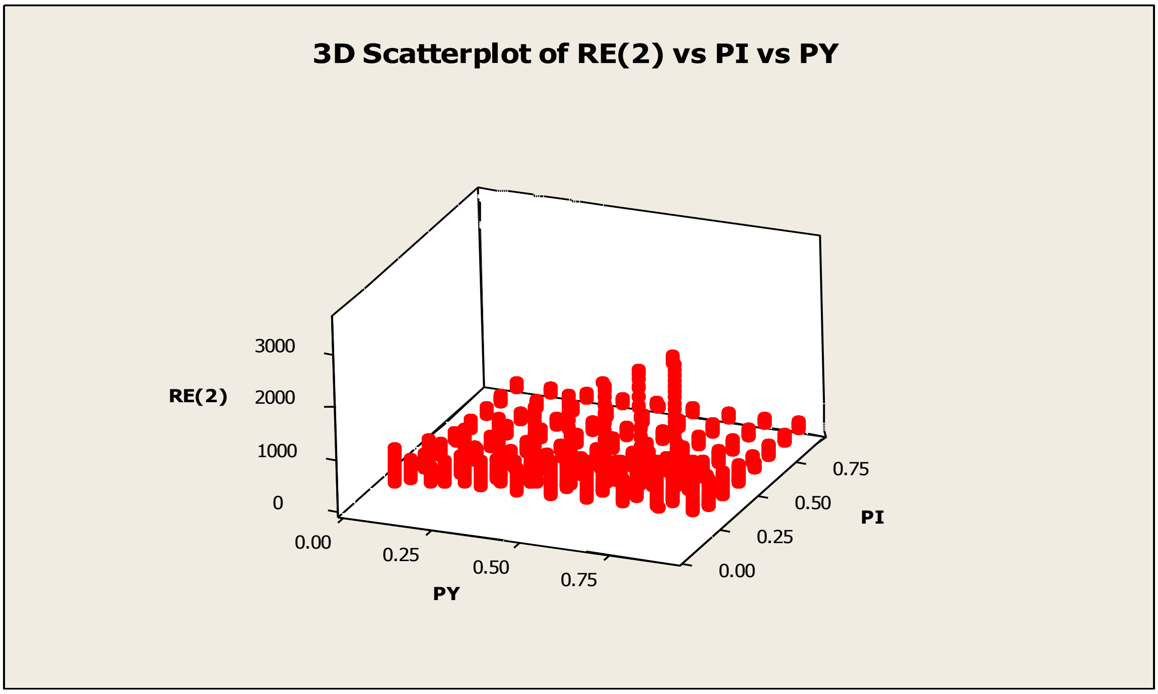

Figure 4.5: Percent relative efficiency RE(2) for equal sample sizes $n_1 = n_2 = 250$ for different combinations of values of π and π_y in the range (0.1, 0.9) with a step of 0.1 and different choices of α and β.

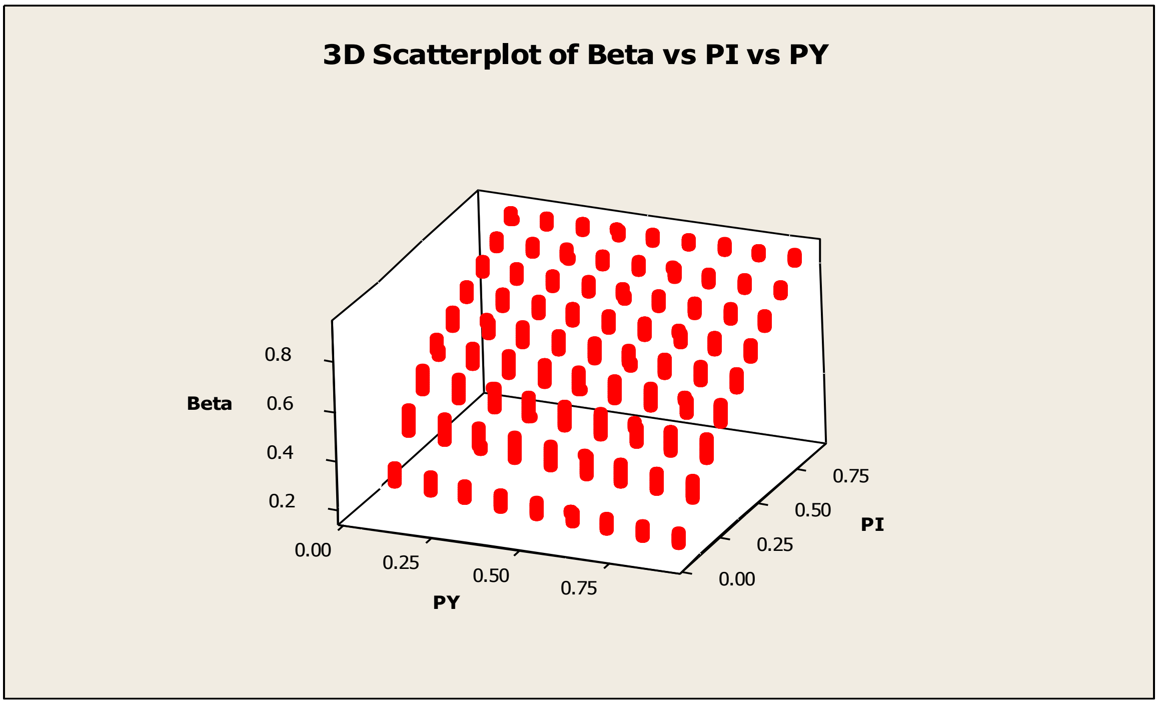

Figure 4.6: A set of choice of values of β for equal sample sizes $n_1 = 250$ and $n_2 = 500$ for different combinations of values of π and π_y in the range (0.1, 0.9) with a step of 0.1.

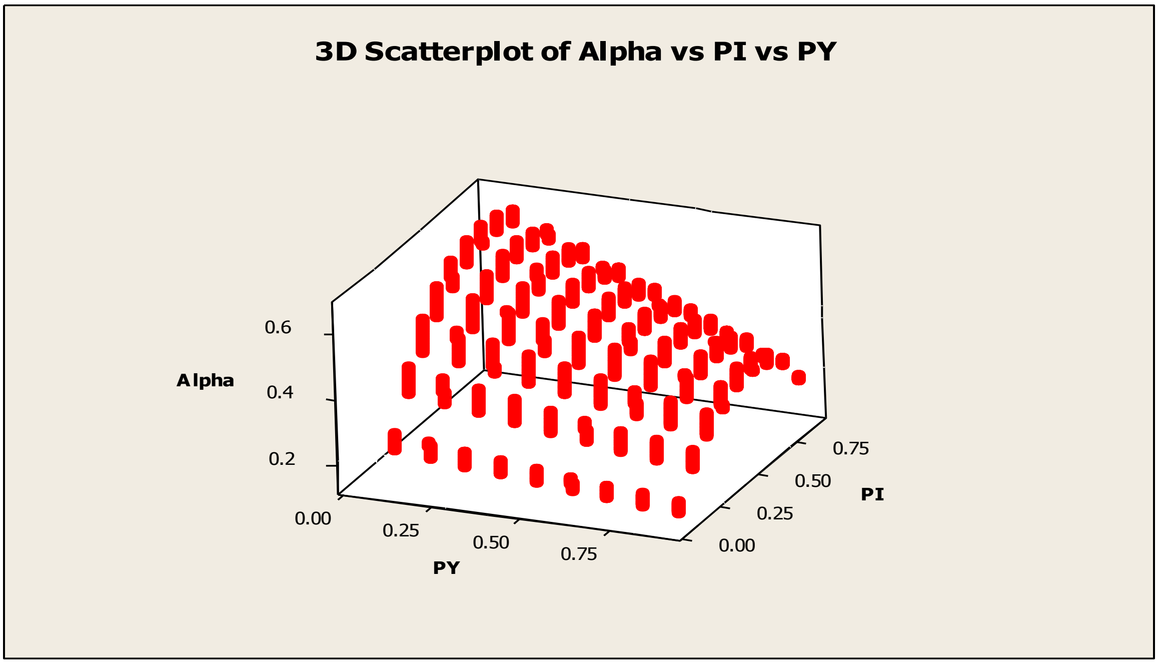

Figure 4.7: A set of choice of values of α for equal sample sizes $n_1 = 250$ and $n_2 = 500$ for different combinations of values of π and π_y in the range (0.1, 0.9) with a step of 0.1.

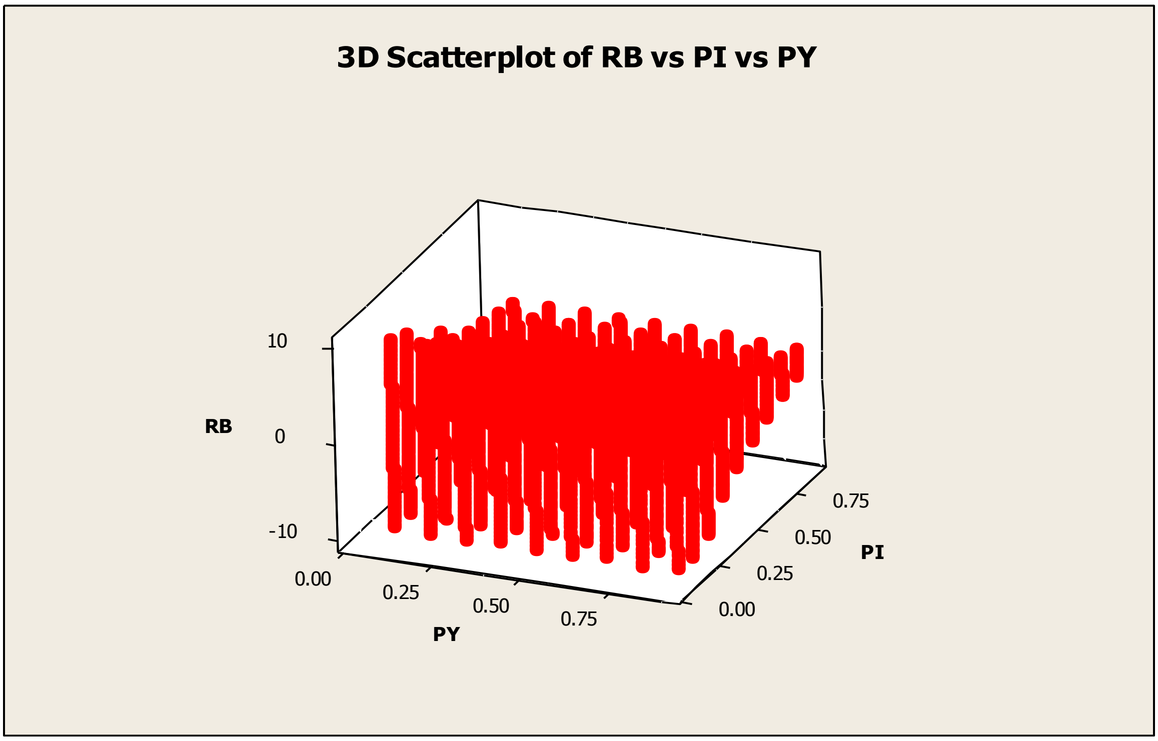

Figure 4.8: Percent relative bias (RB) for equal sample sizes $n_1 = 250$ and $n_2 = 500$ for different combinations of values of π and π_y in the range (0.1, 0.9) with a step of 0.1 and different choices of α and β.

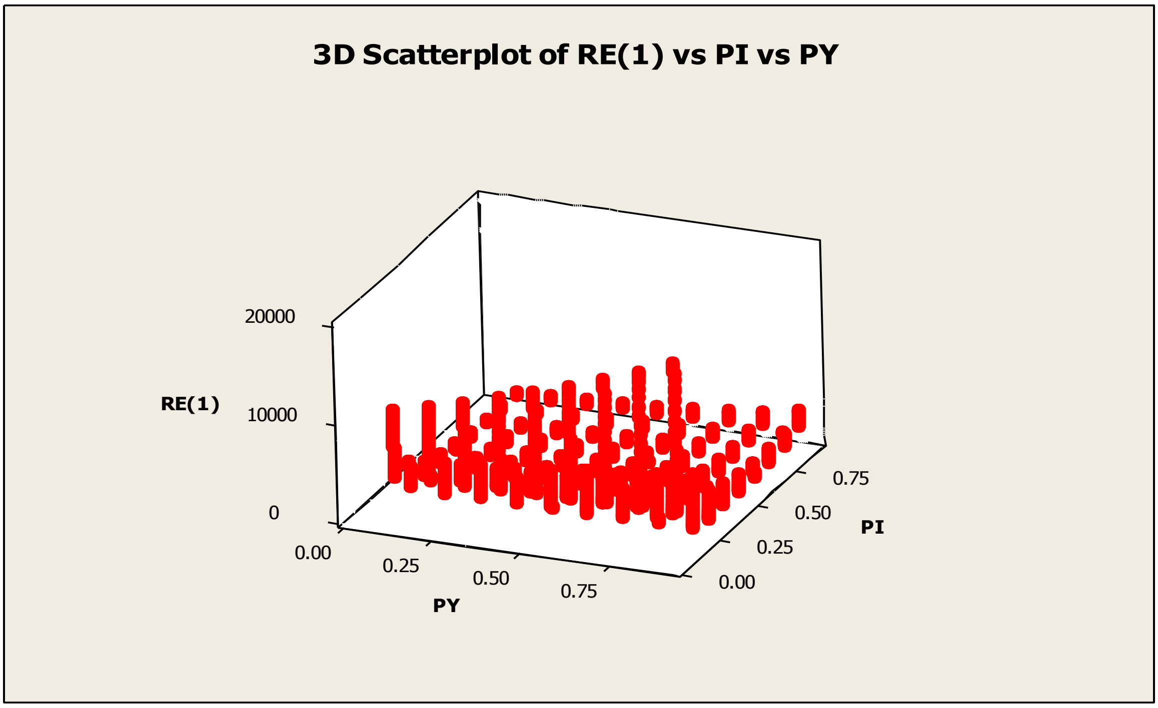

Figure 4.9: Percent relative efficiency RE(1) for equal sample sizes $n_1 = 250$ and $n_2 = 500$ for different combinations of values of π and π_y in the range (0.1, 0.9) with a step of 0.1 and different choices of α and β.

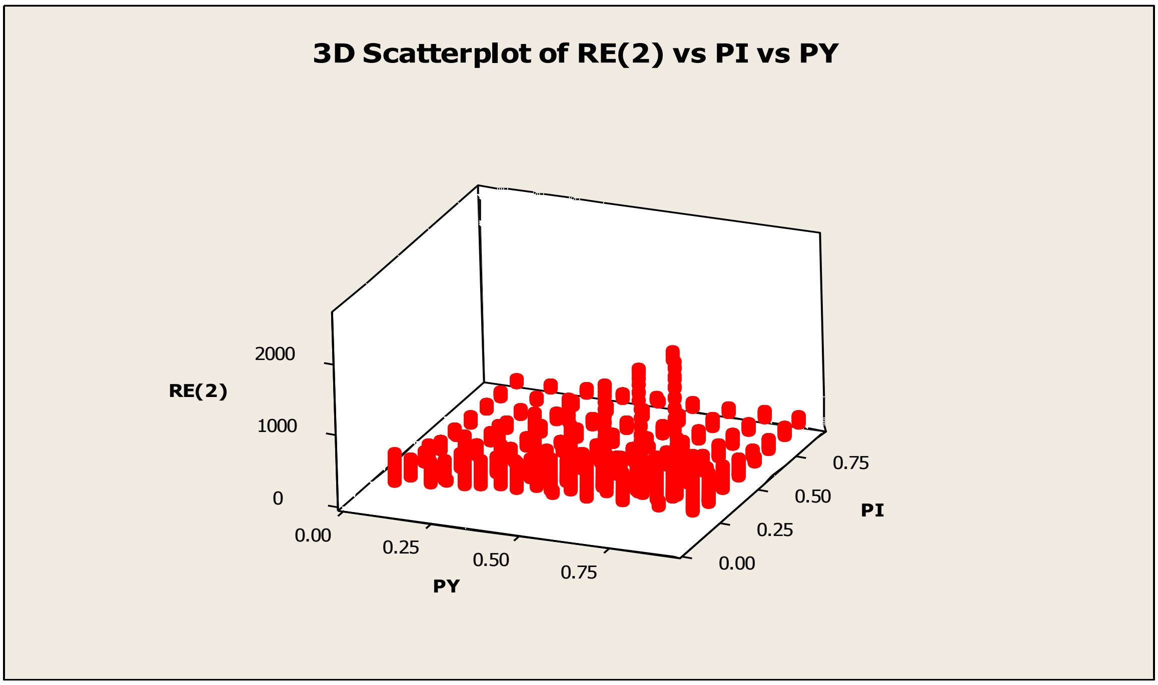

Figure 4.10: Percent relative efficiency RE(2) for equal sample sizes $n_1 = 250$ and $n_2 = 500$ for different combinations of values of π and π_y in the range (0.1, 0.9) with a step of 0.1 and different choices of α and β.

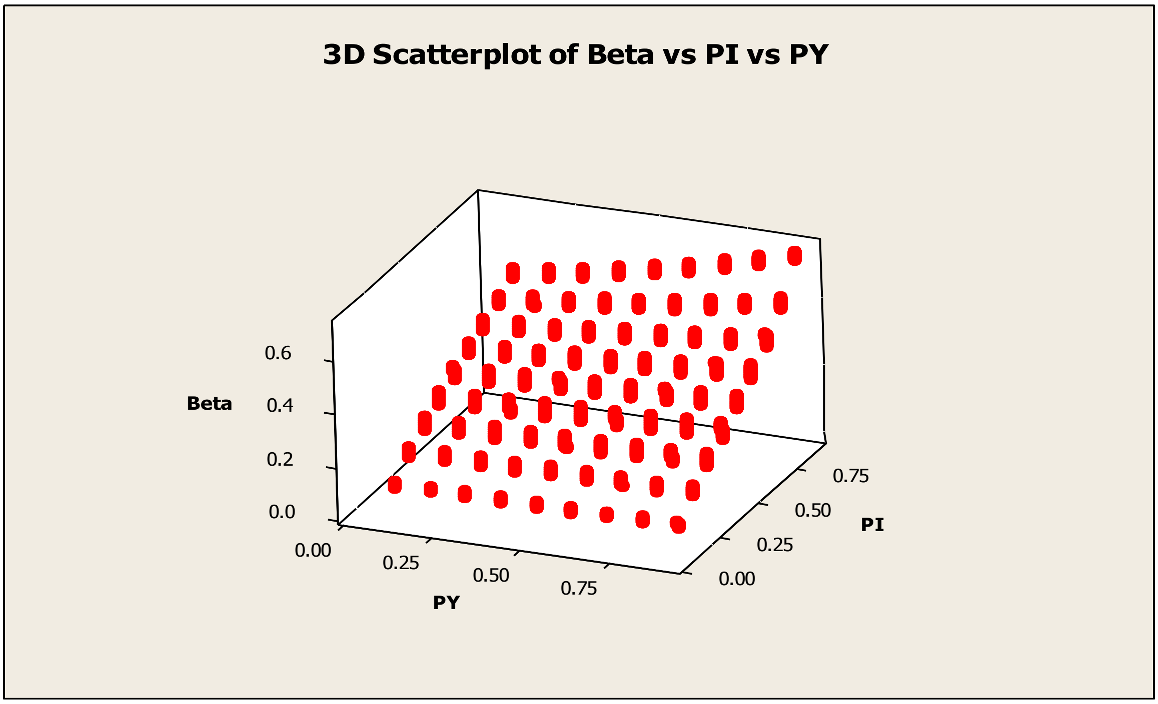

Figure 4.11: A set of choice of values of β for equal sample sizes $n_1 = 500$ and $n_2 = 250$ for different combinations of values of π and π_y in the range (0.1, 0.9) with a step of 0.1.

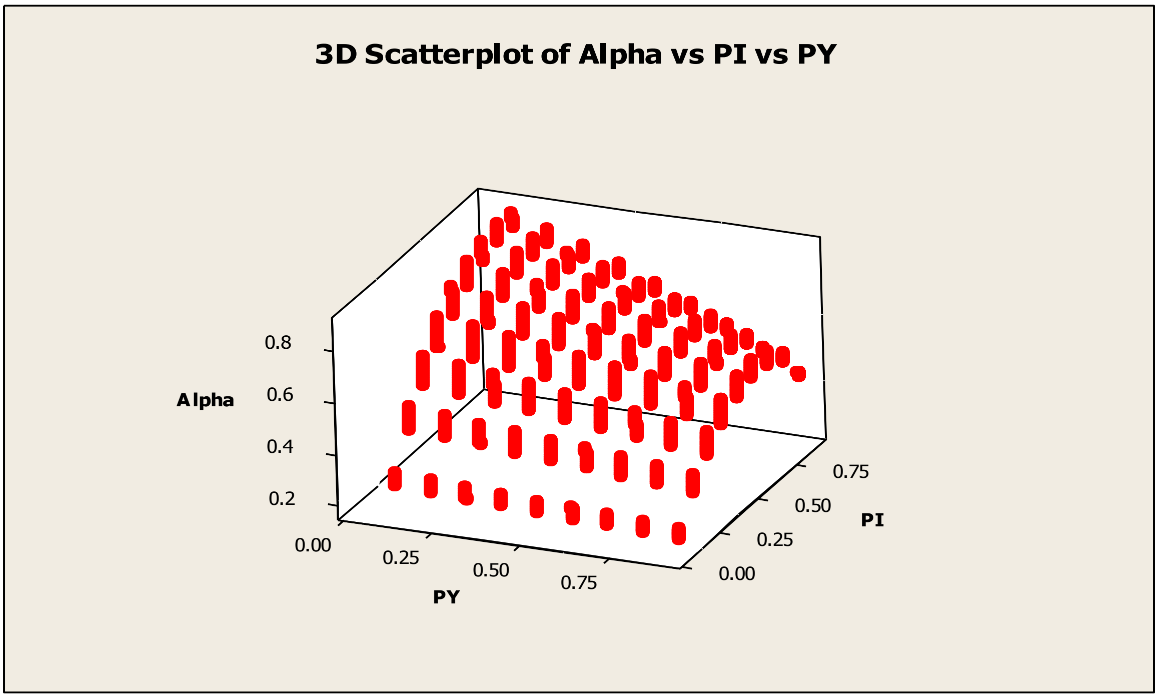

Figure 4.12: A set of choice of values of α for equal sample sizes $n_1 = 500$ and $n_2 = 250$ for different combinations of values of π and π_y in the range (0.1, 0.9) with a step of 0.1.

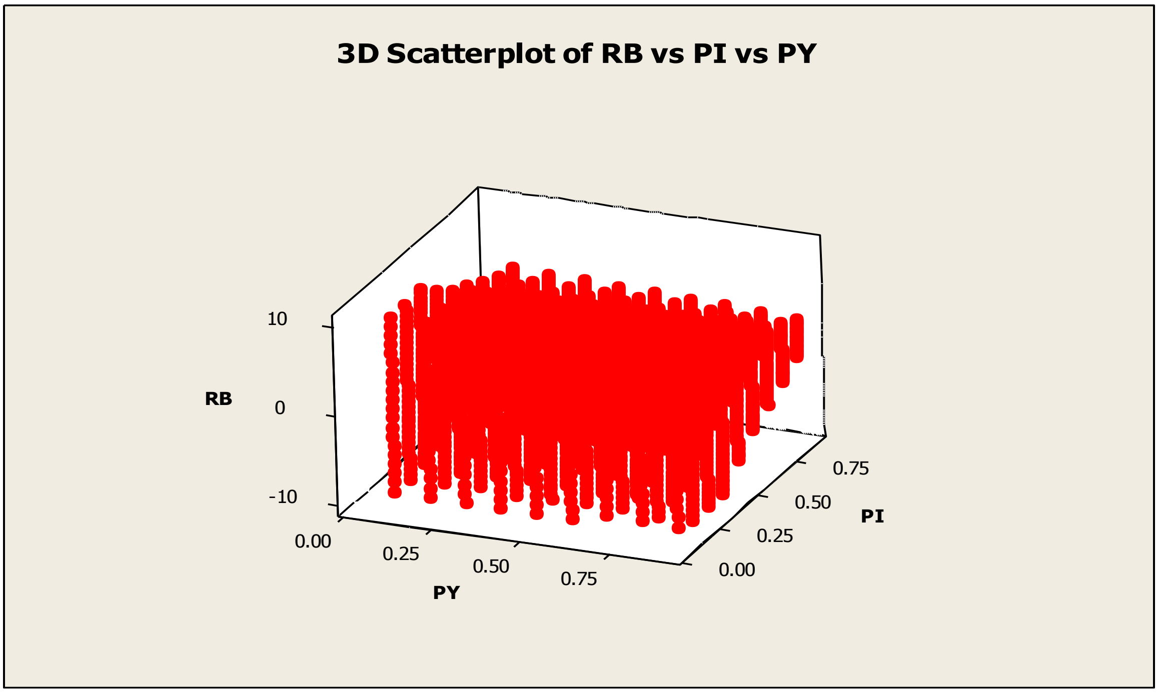

Figure 4.13: Percent relative bias (RB) for equal sample sizes $n_1 = 500$ and $n_2 = 250$ for different combinations of values of π and π_y in the range (0.1, 0.9) with a step of 0.1 and different choices of α and β.

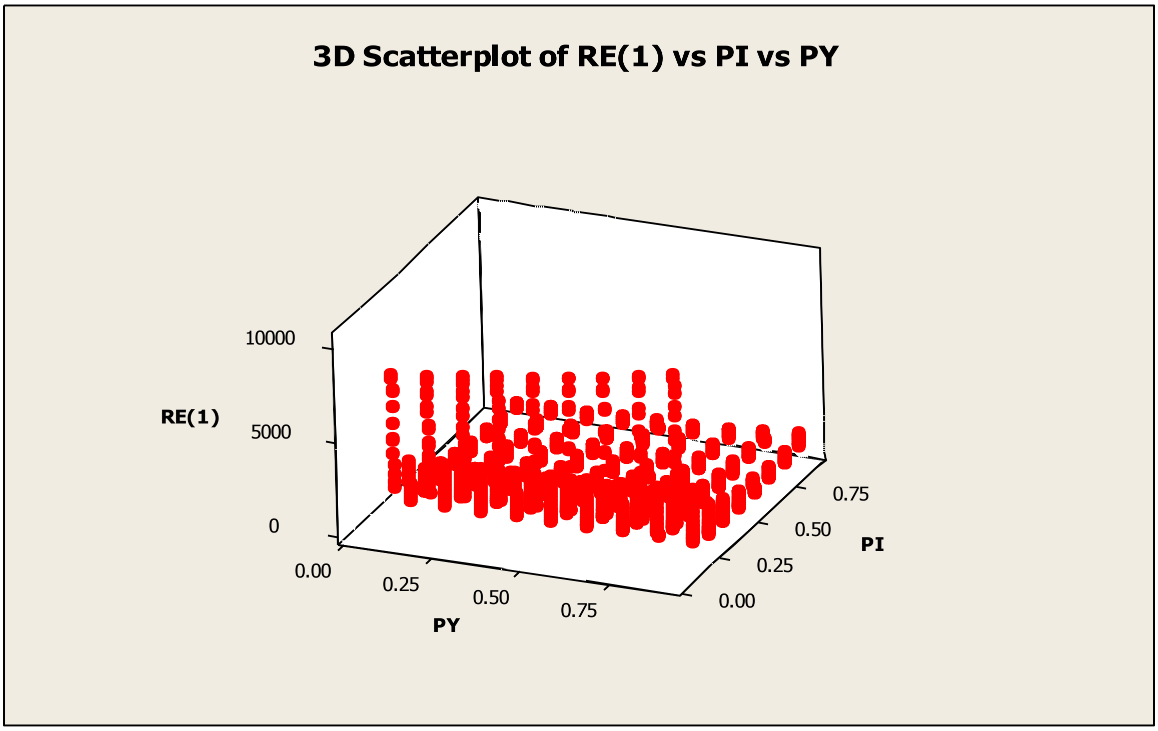

Figure 4.14: Percent relative efficiency RE(1) for equal sample sizes $n_1 = 500$ and $n_2 = 250$ for different combinations of values of π and π_y in the range (0.1, 0.9) with a step of 0.1 and different choices of α and β.

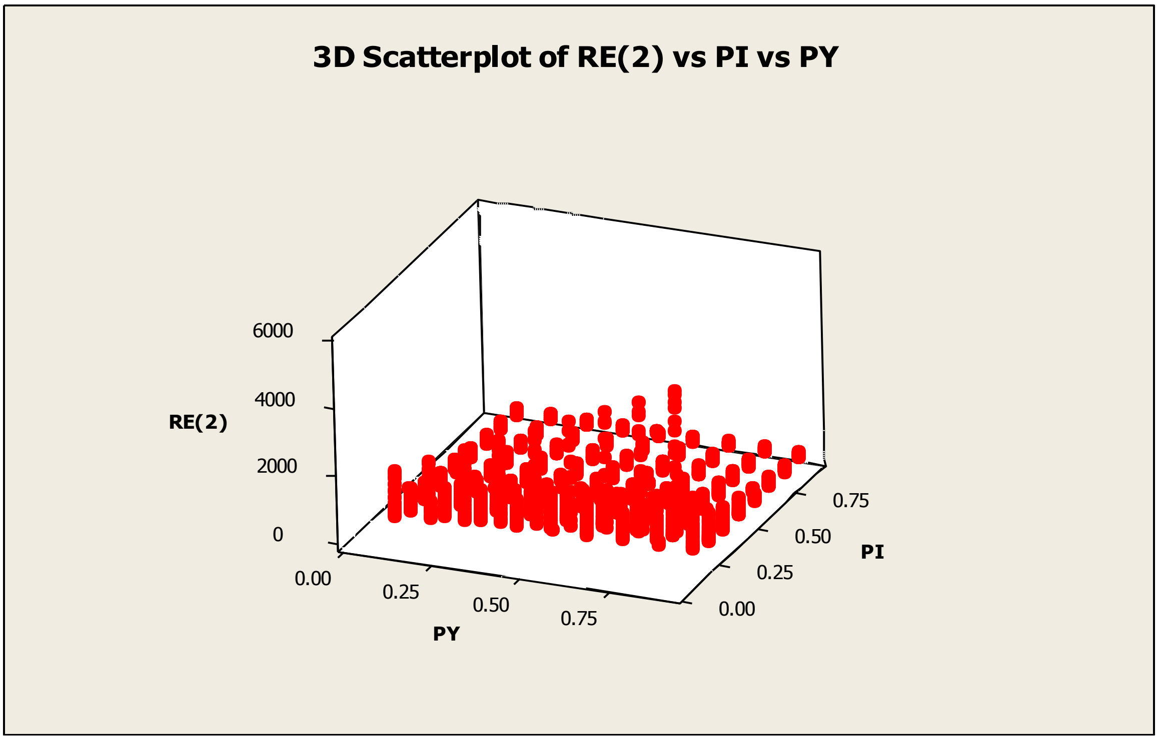

Figure 4.15: Percent relative efficiency RE(2) for equal sample sizes $n_1 = 500$ and $n_2 = 250$ for different combinations of values of π and π_y in the range (0.1, 0.9) with a step of 0.1 and different choices of α and β.

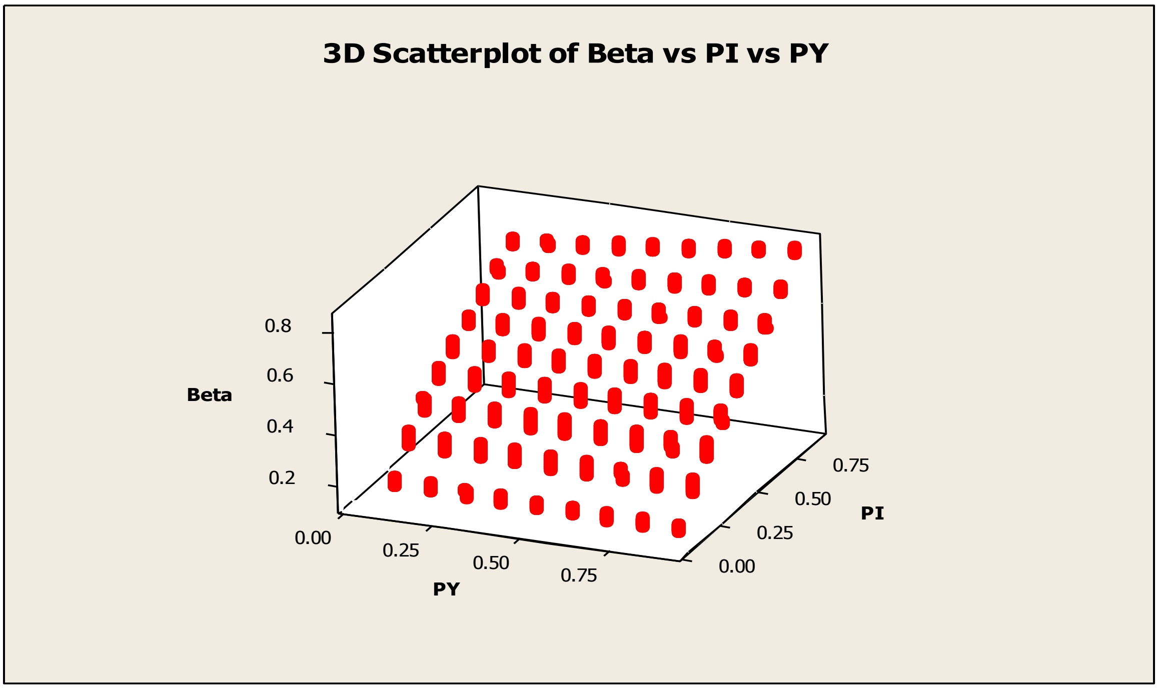

Figure 4.16: A set of choice of values of β for equal sample sizes $n_1 = n_2 = 500$ for different combinations of values of π and π_y in the range (0.1, 0.9) with a step of 0.1.

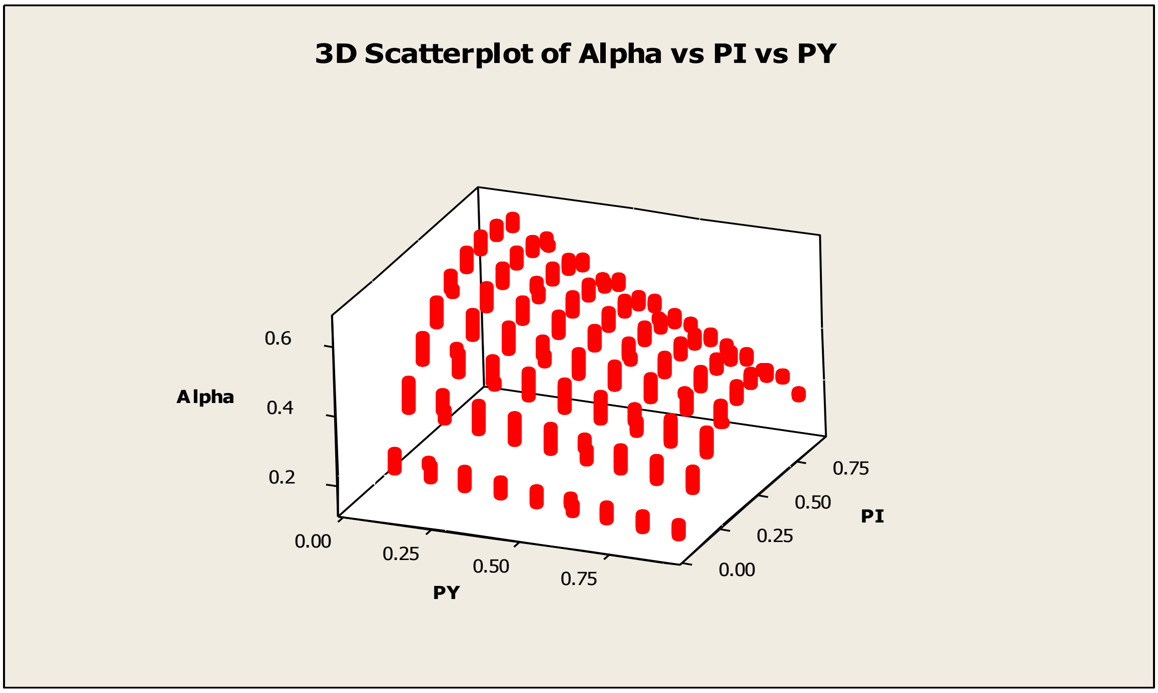

Figure 4.17: A set of choice of values of α for equal sample sizes $n_1 = n_2 = 500$ for different combinations of values of π and π_y in the range (0.1, 0.9) with a step of 0.1.

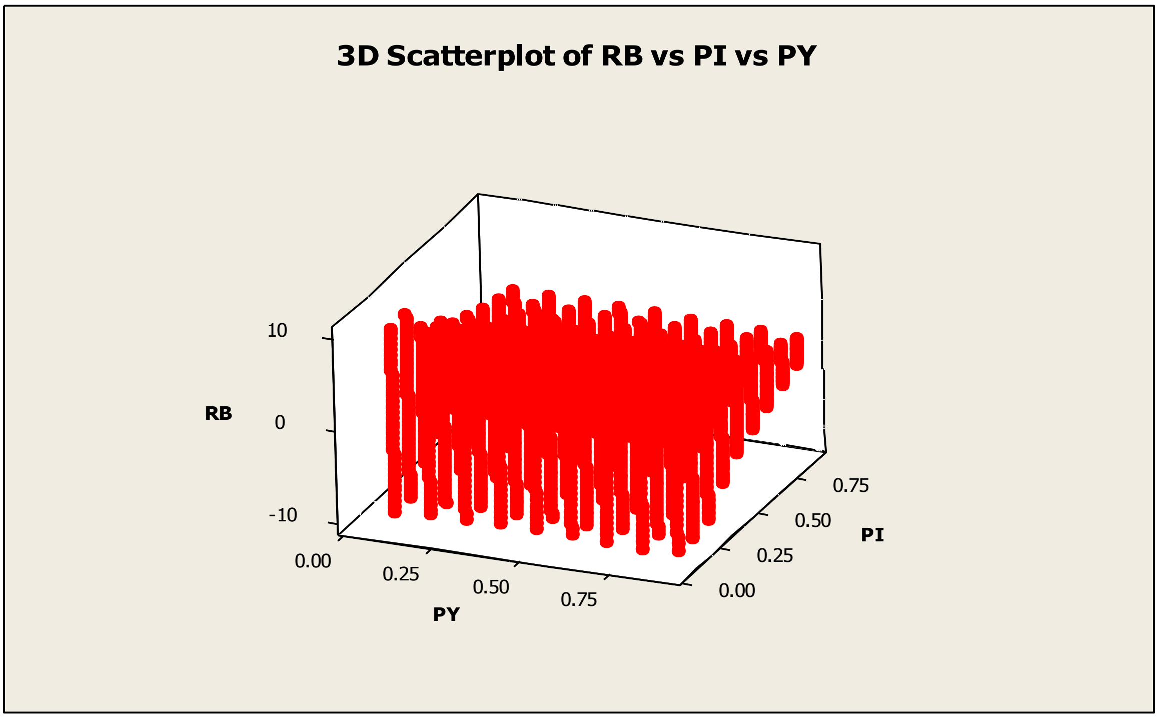

Figure 4.18: Percent relative bias (RB) for equal sample sizes $n_1 = n_2 = 500$ for different combinations of values of π and π_y in the range (0.1, 0.9) with a step of 0.1 and different choices of α and β.

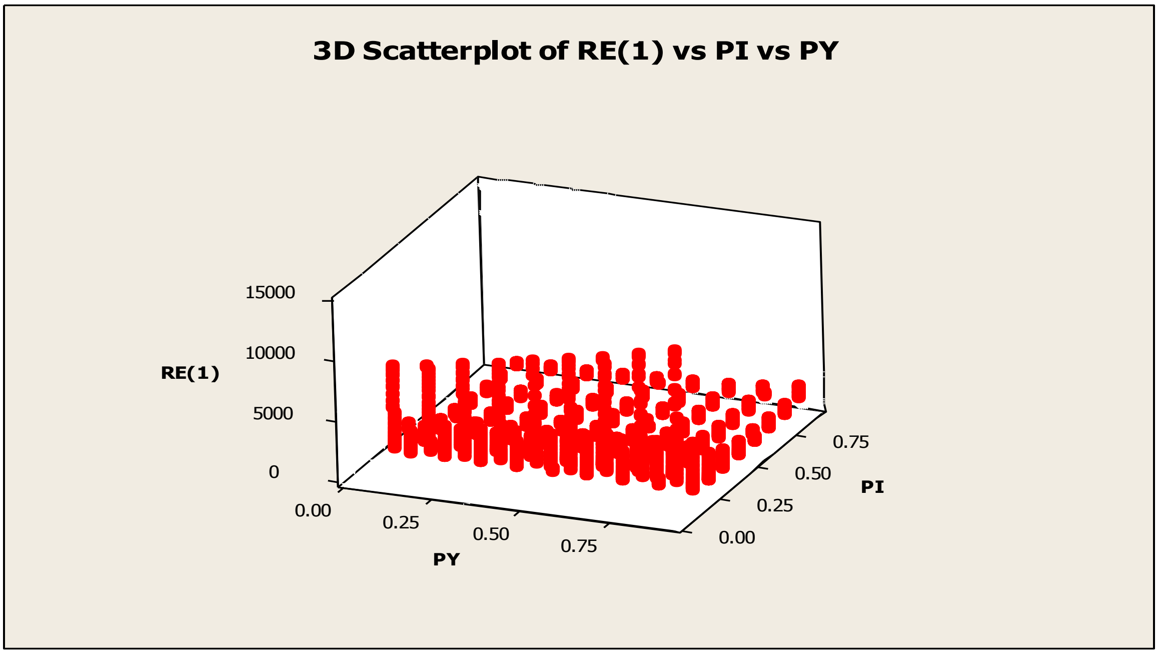

Figure 4.19: Percent relative efficiency RE(1) for equal sample sizes $n_1 = n_2 = 500$ for different combinations of values of π and π_y in the range (0.1, 0.9) with a step of 0.1 and different choices of α and β.

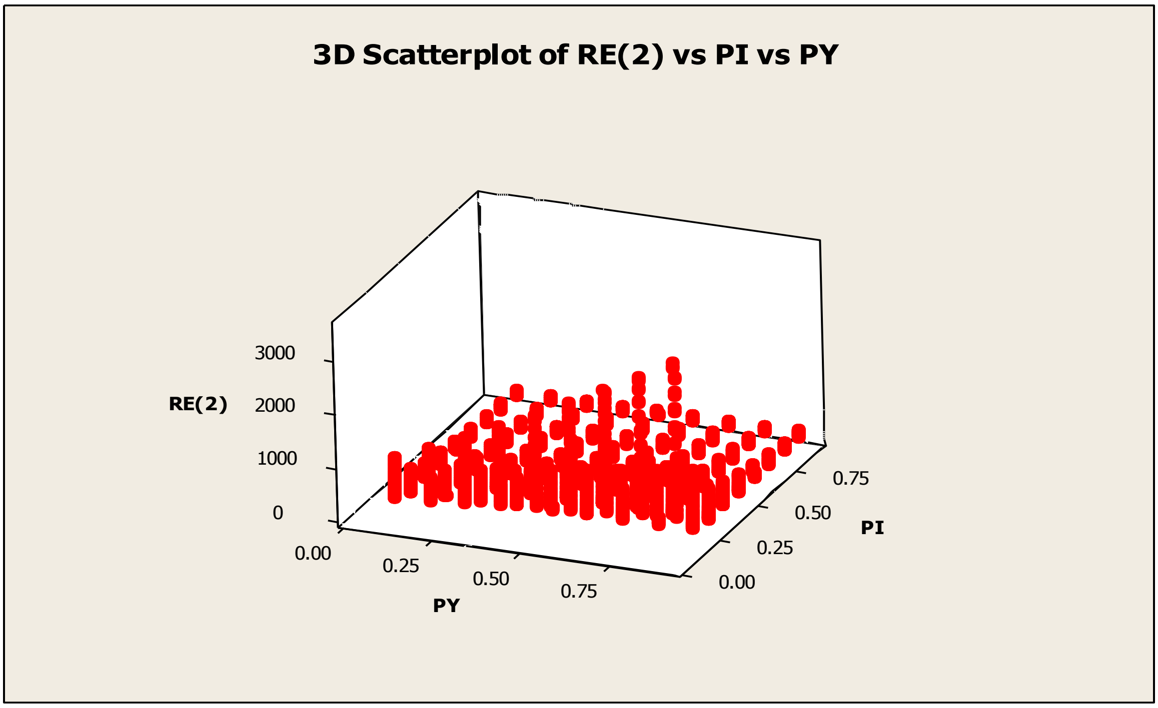

Figure 4.20: Percent relative efficiency RE(2) for equal sample sizes $n_1 = n_2 = 500$ for different combinations of values of π and π_y in the range (0.1, 0.9) with a step of 0.1 and different choices of α and β.

Appendix-C

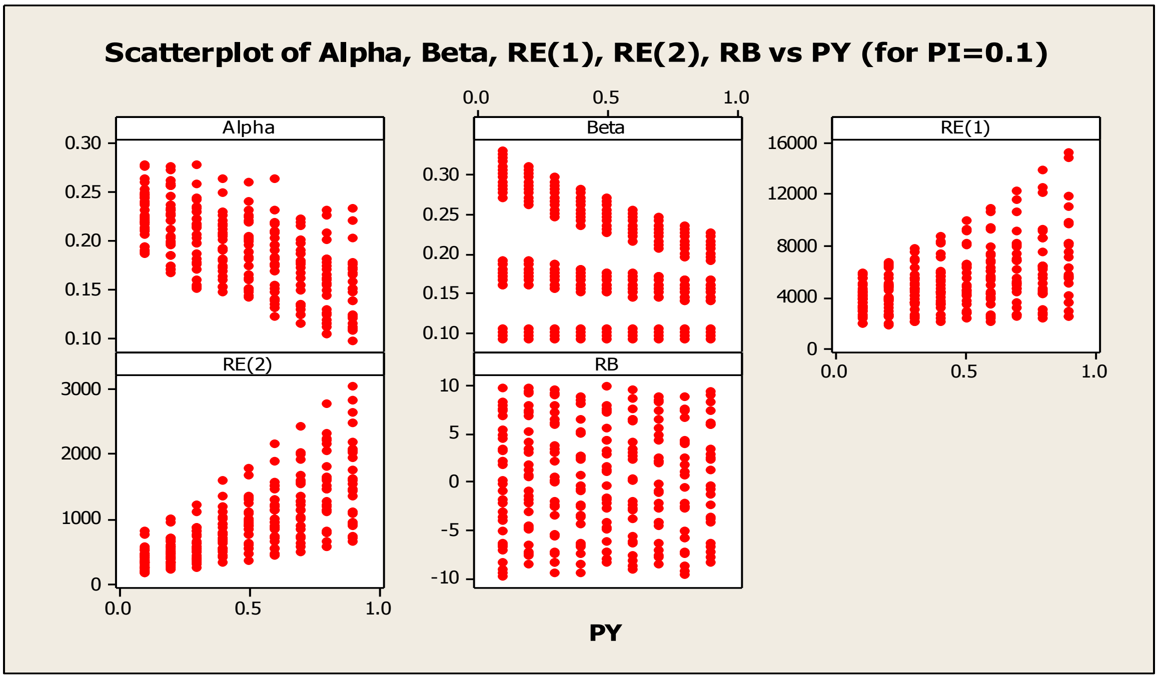

Figure 6.1: Good guess of the estimate $\hat{\beta}$ and the computed values of $\hat{\alpha}$, RB*(%), RE*(1) and RE*(2) for all $\pi_y \in [0.1, 0.9]$ and given value of $\pi = 0.1$ for all possible four pairs of sample sizes n_1 and n_2 with values of 250 and 500.

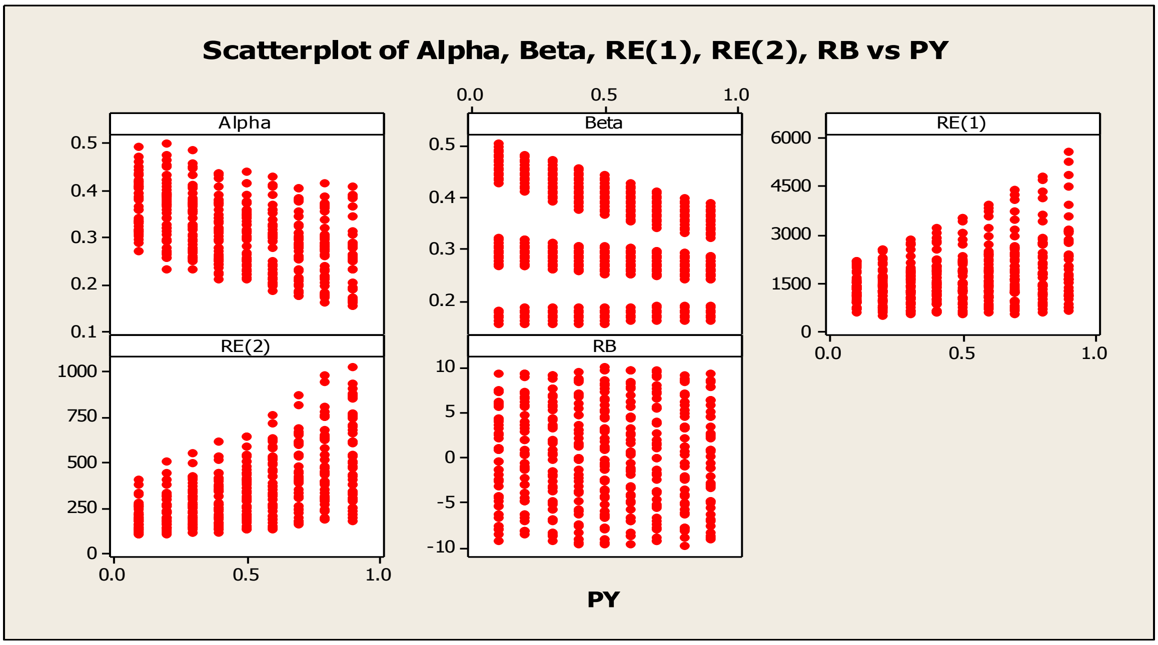

Figure 6.2: Good guess of the estimate $\hat{\beta}$ and the computed values of $\hat{\alpha}$, RB*(%), RE*(1) and RE*(2) for all $\pi_y \in [0.1, 0.9]$ and given value of $\pi = 0.2$ for all possible four pairs of sample sizes n_1 and n_2 with values of 250 and 500.

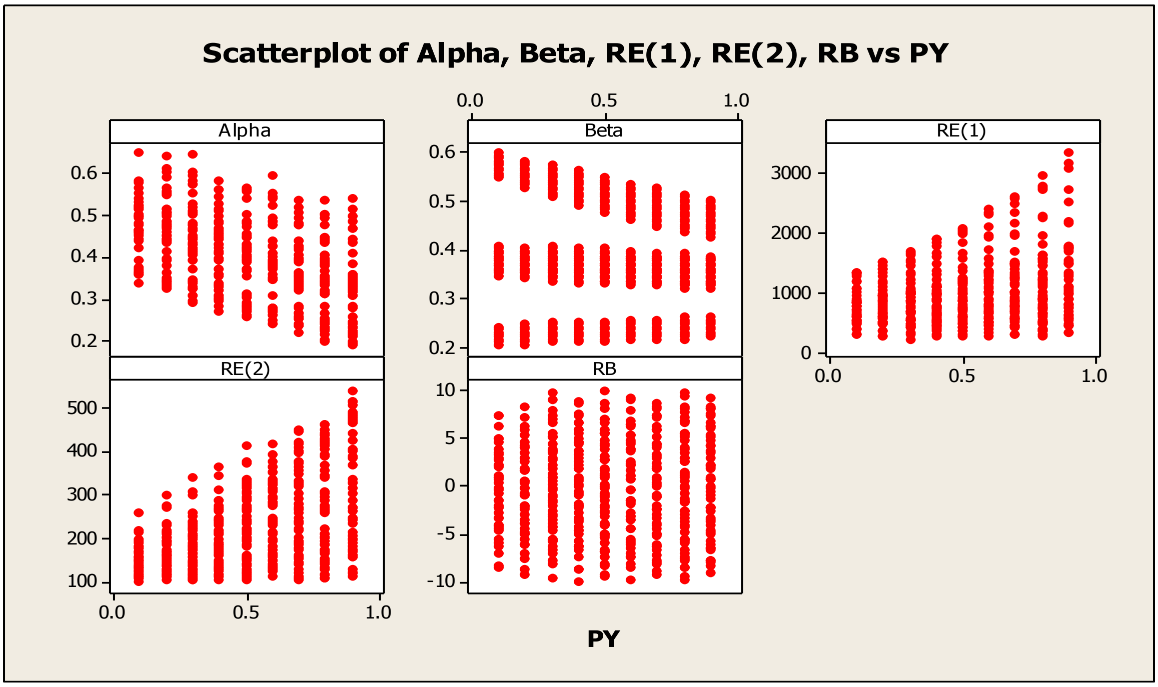

Figure 6.3: Good guess of the estimate $\hat{\beta}$ and the computed values of $\hat{\alpha}$, RB*(%), RE*(1) and RE*(2) for all $\pi_y \in [0.1, 0.9]$ and given value of $\pi = 0.3$ for all possible four pairs of sample sizes n_1 and n_2 with values of 250 and 500.

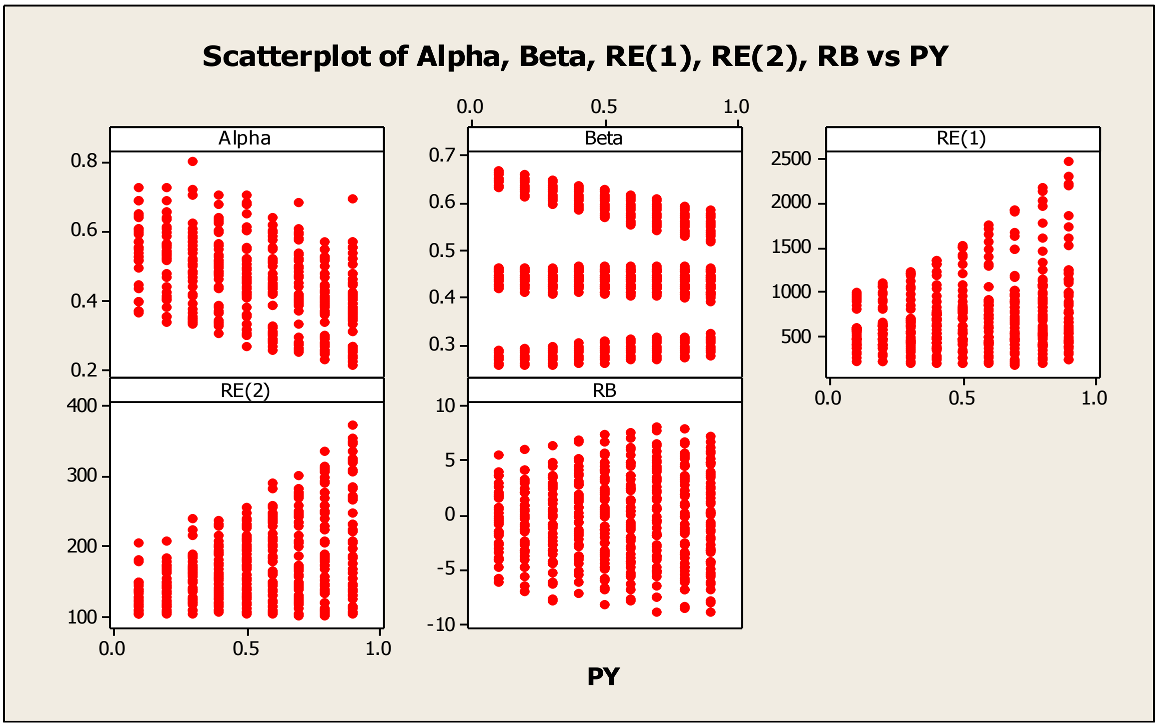

Figure 6.4: Good guess of the estimate $\hat{\beta}$ and the computed values of $\hat{\alpha}$, RB*(%), RE*(1) and RE*(2) for all $\pi_y \in [0.1, 0.9]$ and given value of $\pi = 0.4$ for all possible four pairs of sample sizes n_1 and n_2 with values of 250 and 500.

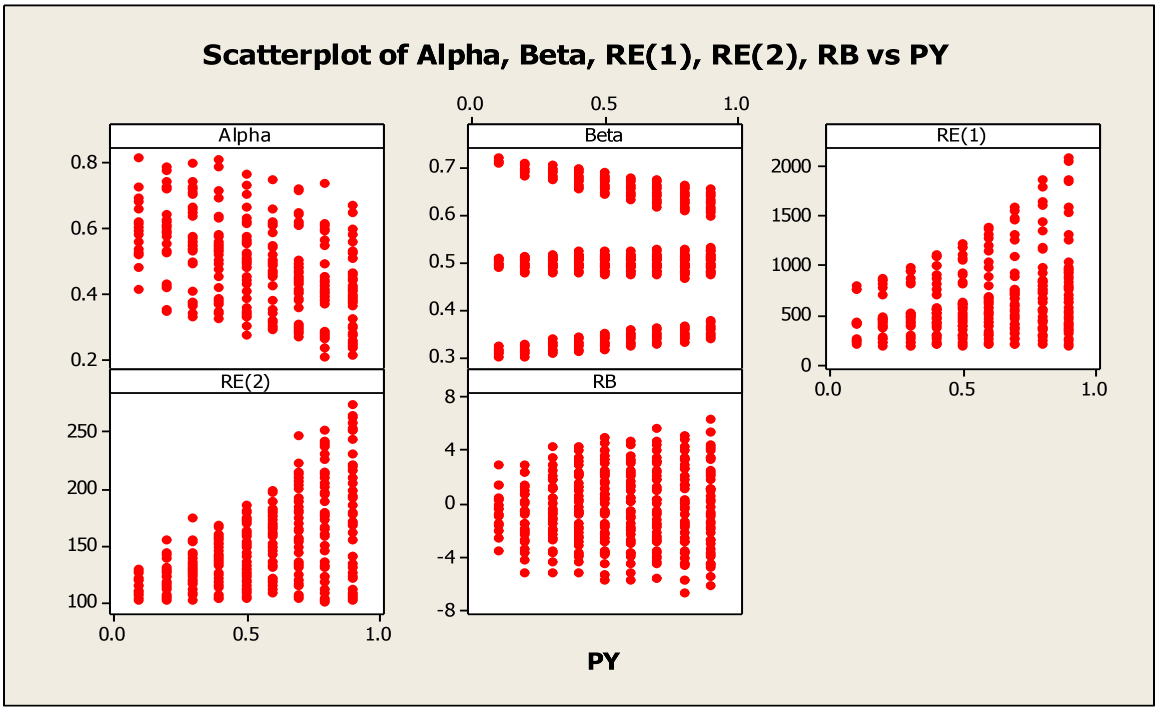

Figure 6.5: Good guess of the estimate $\hat{\beta}$ and the computed values of $\hat{\alpha}$, RB*(%), RE*(1) and RE*(2) for all $\pi_y \in [0.1, 0.9]$ and given value of $\pi = 0.5$ for all possible four pairs of sample sizes n_1 and n_2 with values of 250 and 500.

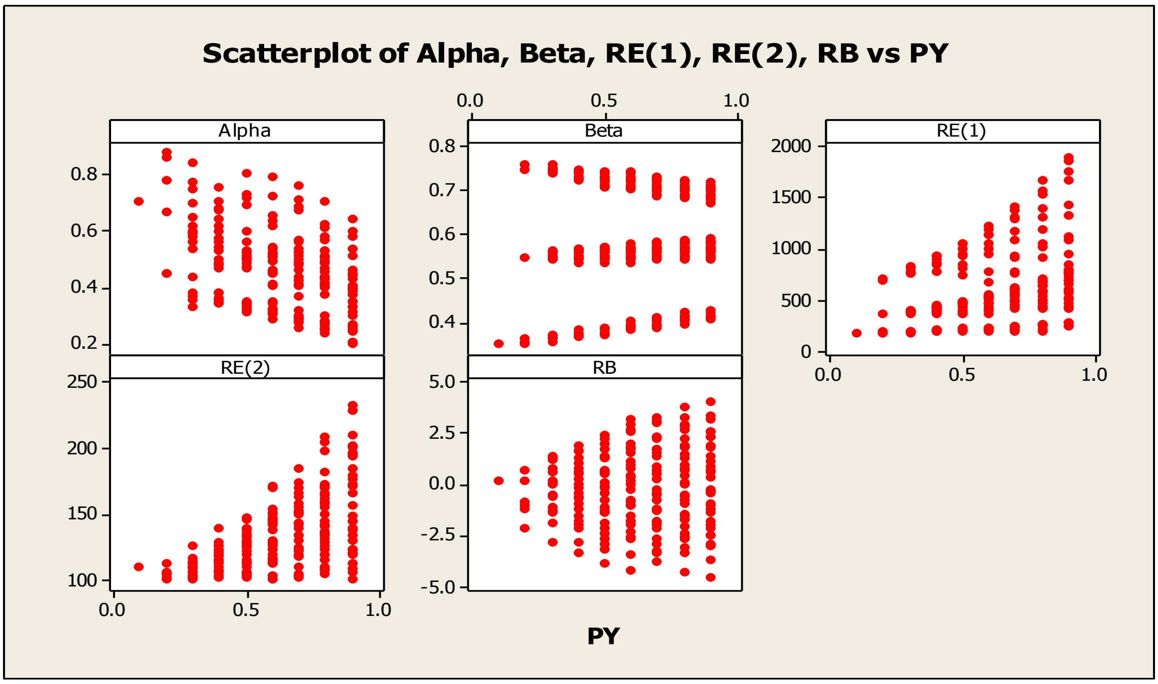

Figure 6.6: Good guess of the estimate $\hat{\beta}$ and the computed values of $\hat{\alpha}$, RB*(%), RE*(1) and RE*(2) for all $\pi_y \in [0.1, 0.9]$ and given value of $\pi = 0.6$ for all possible four pairs of sample sizes n_1 and n_2 with values of 250 and 500.

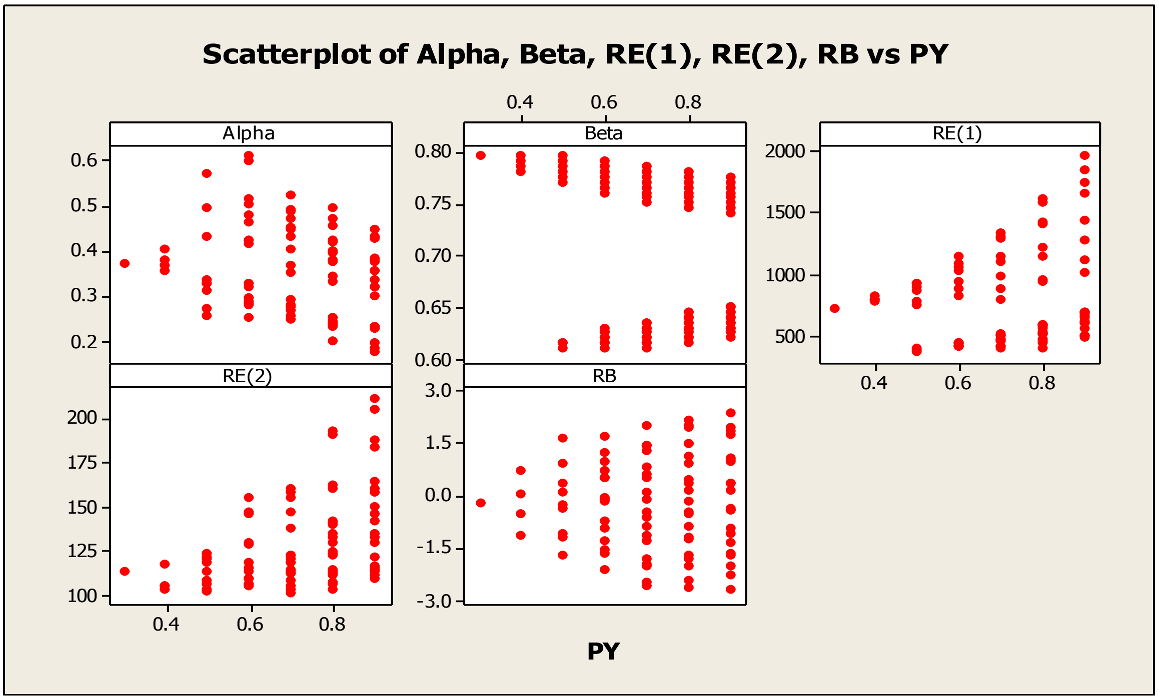

Figure 6.7: Good guess of the estimate $\hat{\beta}$ and the computed values of $\hat{\alpha}$, RB*(%), RE*(1) and RE*(2) for all $\pi_y \in [0.1, 0.9]$ and given value of $\pi = 0.7$ for all possible four pairs of sample sizes n_1 and n_2 with values of 250 and 500.

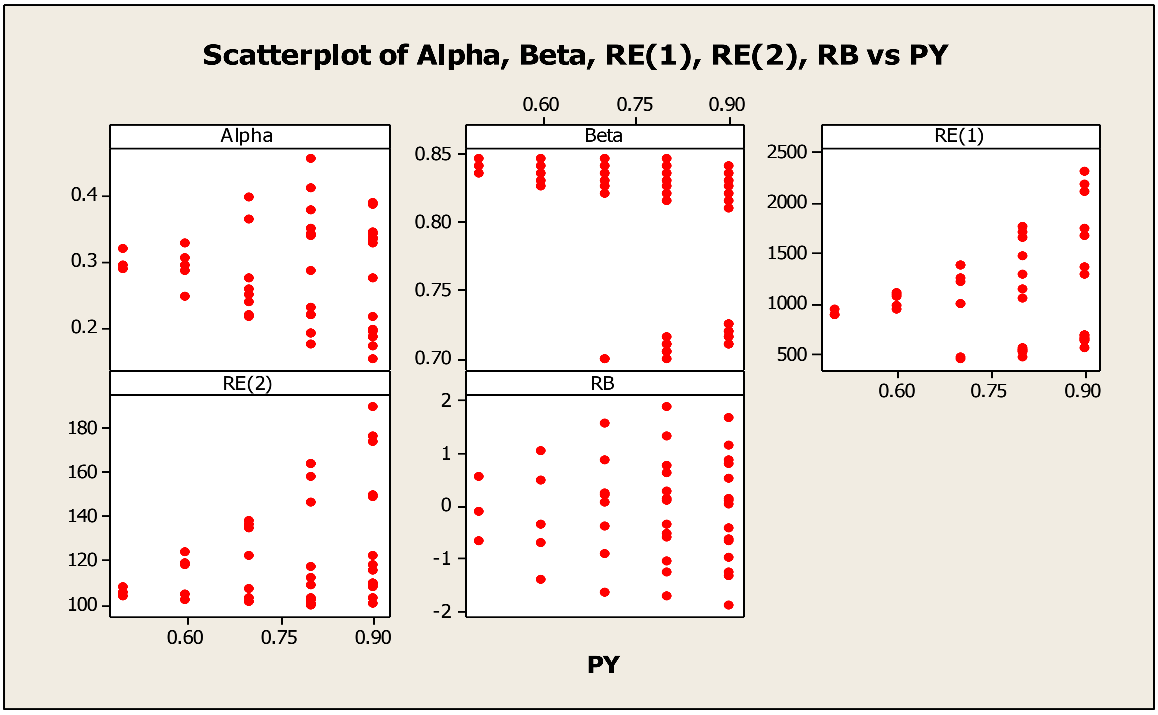

Figure 6.8: Good guess of the estimate $\hat{\beta}$ and the computed values of $\hat{\alpha}$, RB*(%), RE*(1) and RE*(2) for all $\pi_y \in [0.1, 0.9]$ and given value of $\pi = 0.8$ for all possible four pairs of sample sizes n_1 and n_2 with values of 250 and 500.

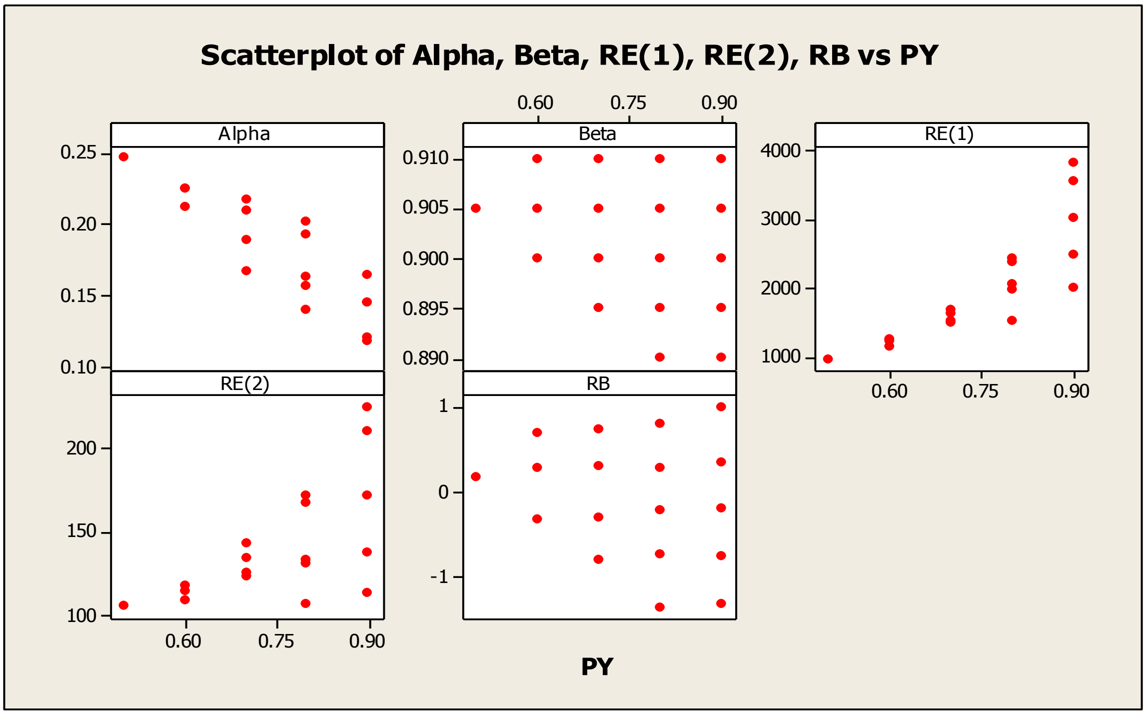

Figure 6.9: Good guess of the estimate $\hat{\beta}$ and the computed values of $\hat{\alpha}$, RB*(%), RE*(1) and RE*(2) for all $\pi_y \in [0.1, 0.9]$ and given value of $\pi = 0.9$ for all possible four pairs of sample sizes n_1 and n_2 with values of 250 and 500.