

ABSTRACT

Title of Dissertation: FACTORS INFLUENCING THE MIXTURE INDEX
OF MODEL FIT IN CONTINGENCY TABLES
SHOWING INDEPENDENCE

Xuemei Pan, Doctor of Philosophy, 2006

Dissertation directed by: Professor C. Mitchell Dayton
Department of Measurement, Statistics & Evaluation

Traditional methods for evaluating contingency table models based on chi square statistics or quantities derived from them are not attractive in many applied research settings. The two-point mixture index of fit, π^* , introduced by Rudas, Clogg and Lindsay (RCL: 1994) provides a new way to represent goodness-of-fit for contingency tables. This study: (a) evaluated several techniques for dealing with sampling zeros when computing π^* in contingency tables when the independence assumption holds; (b) investigated the performance of the estimate, $\hat{\pi}^*$, in various combinations of conditions, as a function of different sizes of tables, different marginal distributions and different sample sizes; and (c) compared the standard error of $\hat{\pi}^*$ and confidence interval estimated by using a method proposed by RCL, with the “true” standard error based on empirical simulations in various scenarios especially when encountering small sample sizes and π^* close to zero. The goals of this study were achieved by Monte Carlo simulation methods and then were applied to two real data examples. The first is a 6×3 cross-classification of fatal crashes by speed limit and land use with 37,295 cases based

on 2004 USDOT traffic data and the second 4×4 cross-classification of eye color and hair color with 592 cases reported in RCL.

Results suggest that: $\hat{\pi}^*$ is positively biased from zero in a range from 2.98% to 40.86% in the conditions studied when the independence assumption holds. Replacing zero with larger flattening values results in smaller $\hat{\pi}^*$. For each table size, $\hat{\pi}^*$ is smallest for all extremely dispersed row and column marginal distributions. For all extremely and most slightly dispersed marginal distributions tables with small sample size and small table size, using structural zero technique is superior to other sampling zero techniques. The lower bound for $\hat{\pi}^*$ using the RCL method is generally close to the "true" estimate based on empirical parametric simulation. However, under some circumstances, RCL method underestimates the lower bound value even though the magnitude is relatively small and the difference shrinks as the sample size increases. This study will provide guidance for researchers in the use of this important method for interpreting models fit to contingency tables.

FACTORS INFLUENCING THE MIXTURE INDEX OF MODEL FIT
IN CONTINGENCY TABLES SHOWING INDEPENDENCE

By

Xuemei Pan

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2006

Advisory Committee:

Professor C. Mitchell Dayton, Chair
Dr. Jeffery Haring
Professor George Macready
Professor Robert Mislevy
Professor Eric V. Slud

©Copyright by

Xuemei Pan

2006

DEDICATION

This dissertation is dedicated to my parents, Zhenzhong and Yuyu, my sister, Xuesong and most of all, to my husband, Yi for their dedication, encouragement, and support.

ACKNOWLEDGENTS

I am deeply grateful to my advisor, Professor C. Mitchell Dayton for his excellent guidance and unwavering support. I also want to sincerely thank all the faculty members and staff in my department for their continuous help, inspiration and support.

TABLE OF CONTENTS

List of Tables.....	v
List of Figures.....	vi
List of SAS Programs.....	vii
Chapter I: Purpose and Rationale.....	1
Background.....	1
Purpose of Study.....	2
Significance of Study.....	3
Chapter II: Review of Literature.....	4
Mixture Index of Fit	4
EM Algorithm for Calculating the Mixture Index of Fit.....	5
Lower Bound and Standard Error Estimation for two-point Mixture Index.....	9
Sampling Zeros and Structural Zeros.....	10
Chapter III: Methods.....	12
Research Design.....	12
Simulation Details.....	17
Statistical Analysis.....	21
Chapter IV: Results.....	22
Parameter Estimates and Bias.....	22
Lower Bound Comparisons of RCL and “True” Estimates.....	34
Confidence Intervals and Standard Errors.....	35
Examples.....	42
Chapter V: Conclusions and Discussion.....	46
Parameter Estimates and Bias.....	46
Sampling Zero Techniques.....	47
Confidence Interval and Standard Error.....	47
Recommendations.....	48
Implications for Future Research.....	49
Appendix A: Results Tables.....	53
Appendix B: Sample SAS Programs.....	89
References.....	100

LIST OF TABLES

1.	Fictional 4×4 frequency table.....	7
2.	Fatal crashes by speed limit and land use.....	20
3.	Cross-classification of eye color and hair color.....	20
4.	Fit statistics for fatal crashes by speed limit and land use.....	44
5.	$\hat{\pi}^*$ Comparison of sampling zero and structural zero using eye color data.....	45

Appendix A: Results Tables

A1 – A24:	$\hat{\pi}^*$ and $\hat{\pi}^*$ using sampling zero techniques.....	53
A25 – A48:	95% lower bound- $\hat{\pi}_L$	65
A49 – A72:	95% confidence interval ($\hat{\pi}^* - \hat{\pi}_L$) and standard error of $\hat{\pi}^*$	77

LIST OF FIGURES

1.	Histograms for $\hat{\pi}^*$ with small parameter bias (left) and large parameter bias (right).....	16
2.	$\hat{\pi}^*$ for Evenly Distributed Marginals	26
3.	$\hat{\pi}^*$ for Extremely Distributed Marginals	27
4.	$\hat{\pi}^*$ Comparison in $\{P_{1+}= .9, P_{2+}= .1, P_{+1}= .8, P_{+2}= .1, P_{+3}= .1\}$	28
5.	$\hat{\pi}^*$ Comparison in $\{P_{1+}= .9, P_{2+}= .1, P_{+1}= .4, P_{+2}= .4, P_{+3}= .1, P_{+4}= .1\}$	29
6.	$\hat{\pi}^*$ Comparison in $\{P_{1+}= .9, P_{2+}= .1, P_{+1}= .3, P_{+2}= .3, P_{+3}= .1, P_{+4}= .1, P_{+5}= .1, P_{+6}= .1\}$	30
7.	Sampling Zero and Structural Zero Effect on $\hat{\pi}^*$ Comparison in 2×3 table.....	31
8.	Sampling Zero and Structural Zero Effect on $\hat{\pi}^*$ Comparison in 2×4 table.....	32
9.	Sampling Zero and Structural Zero Effect on $\hat{\pi}^*$ Comparison in 6×6 table.....	33
10.	Empirical Simulation Based $\hat{\pi}_L$ With Evenly Distributed Marginals	36
11.	Empirical Simulation Based $\hat{\pi}_L$ With Extremely Distributed Marginals.....	37
12.	Comparison between RCL method and Empirical Simulation method	38
13.	Comparison between RCL method and Empirical Simulation method (Con.).....	39
14.	Confidence Interval of $\hat{\pi}^*$ following empirical simulation method	40
15.	Confidence Interval of $\hat{\pi}^*$ following empirical simulation method (Con.).....	41

LIST OF SAS PROGRAMS

1. Structural Zero Scenario.....89
2. Sampling Zero and Replacing with different flattening constants scenario.....97

CHAPTER I

PURPOSE AND RATIONALE

Background

Traditional methods for evaluating contingency table models based on chi square statistics or quantities derived from them are not attractive in many applied research settings. According to Rudas (1998), “First, when the model is not true, a comparison of the data to what could only be expected if it were is of very little meaning; second, the actual distribution of the statistic may be very different from the reference distribution if some of the underlying assumptions are violated.” In addition, conventional methods are sensitive to sample size; often a model is rejected when fitted to a large data set even though the model may represent a reasonable summary of the data for practical purposes. In sharp contrast to chi-squared tests of fit methods, which rely heavily on size of the table, sample size and actual true probabilities, the mixture index of fit proposed by Rudas, Clogg and Lindsay (RCL: 1994), provides a novel way to represent goodness-of-fit for contingency tables. It has an intuitive rationale that does not assume a simple model that describes the entire population in contrast to the underlying idea of classical significance tests. Also, the new index is not sensitive to sample size in the way in which chi-square-related quantities are. More specifically, it is assumed that there are two components (subgroups) in the population. One of them, of size $1 - \pi$, where model H holds true, describes the fraction of population consistent with model H (e.g., independence); the other one, of size π , which is completely unrestricted, represents the part of the population that is outside of model H. Moreover, RCL introduced an expectation-maximization (EM) algorithm to obtain maximum likelihood estimates of π^* and derived a way to construct a lower-bound estimate of $\hat{\pi}^*$. As summarized by

Dayton (2003), $\hat{\pi}^*$ possesses the following appealing properties: 1. $\hat{\pi}^*$ is always located on the 0, 1 interval; 2. $\hat{\pi}^*$ is unique; 3. $\hat{\pi}^*$ is invariant when frequencies in a contingency table are increased or decreased by a multiplicative constant. Properties and applications of the mixture index of fit are further explored in Clogg, Rudas and Xi (1995), Xi (1996), Clogg, Rudas and Matthews (1997). Furthermore, the two-point mixture index, π^* can be applied when models are fitted to virtually any contingency table (RCL). It also has been applied in differential item functioning (Rudas & Zwick, 1997), latent class analysis (Dayton, 1999), regression models with normal and uniform error structures (Rudas, 1999) and logistic regression model (Verdes & Rudas, 2002).

Purpose of Study

As noted by RCL, there are issues that require further examination and have not been studied in RCL or any other related research. In particular:

- (1) $\hat{\pi}^*$ is positively biased in finite samples; that is, even if H holds so that, in theory, $\pi^* = 0$, $\hat{\pi}^*$ will have expectation greater than zero for finite samples.
- (2) Sampling 0's can greatly affect estimation so it is useful to study the effect of using flattening constants or redefining model H by regarding the sampling zeros as structural zeros.
- (3) Although the estimated lower confidence bound of $\hat{\pi}^*$ introduced by RCL gives inferential information that free of bias, it tends to be problematic when π^* is close to zero or sample size is small; thus a parametric simulation seems to be necessary to examine this measure of precision for $\hat{\pi}^*$. As an aside, SAS code written for this study makes these analyses more accessible to researchers in various disciplines.

Significance of Study

The current study has described the work on the evaluation of the mixture index of fit for contingency tables. It assesses the performance of π^* in various scenarios including sampling zeros and structural zero effects on π^* , the bias of π^* in various combinations of conditions such as different sizes of tables, different marginal distributions and different sample sizes when the independence assumption holds, and the accuracy of the measure of precision for π^* . This study will provide guidance for researchers in the use of this important method for interpreting models fit to contingency tables.

CHAPTER II REVIEW OF LITERATURE

Mixture index of fit

Following Rudas and Zwick (1997), suppose H is a hypothesized probabilistic model for a contingency table and P represent the true. The two-point mixture model is:

$$P = (1 - \pi)\Phi + \pi\psi \quad (2.1)$$

where Φ is the probability distribution comes from H , and ψ is an unspecified probability distribution. The mixture parameter, π , defined on the 0, 1 interval, represents the proportion of the population that does not belong to model H . The model in equation 2.1 is not unique and is true for any model based on any contingency table. The index of fit, π^* , however, is defined as the smallest possible fraction, π for which H holds: $\pi^* = \inf\{\pi \mid P = (1 - \pi)\phi + \pi\psi, \phi \in H\}$.

Consequently, as shown by RCL, π^* is unique and it describes the minimum fraction of frequencies that must be excluded from the contingency table in order that P be fitted exactly by the model for the rest of the cases. It also indicates the residuals associated with the mixture index of fit which are quite different from conventional residuals in chi-squared analyses.

As summarized by Dayton (2003), $\hat{\pi}^*$ possesses the following appealing properties: 1. $\hat{\pi}^*$ is always located on the 0, 1 interval; 2. $\hat{\pi}^*$ is unique; 3. $\hat{\pi}^*$ is invariant when frequencies in a frequency table are increased or decreased by any multiplicative constant. Properties and applications of the mixture index of fit are further explored in Clogg, Rudas and Xi (1995), Xi (1996), Clogg, Rudas and Matthews (1997). Furthermore, the two-point mixture index, π^* can be applied when models are fitted to virtually any

contingency table (RCL). It also has been applied in differential item functioning (Rudas & Zwick, 1997), latent class analysis (Dayton, 1999), regression models with normal and uniform error structures (Rudas, 1999) and logistic regression model (Verdes & Rudas, 2002).

EM algorithm for calculating the mixture index of fit

RCL suggested a procedure that involved repeated application of an EM algorithm (Dempster, Laird, and Rubin, 1977) to compute maximum likelihood estimates for mixture models. Based on RCL, the observed data matrix is $I \times J$ contingency table while the complete (unobserved) table is an $I \times J \times 2$. As above, Φ is defined for the probability distribution designated by H in the first component whereas in the second component it is unspecified (ψ). The two components have distribution $1 - \pi$ and π , which are assumed fixed for the each EM cycle. Let Q_{ijk} , $i=1, \dots, I$, $j=1, \dots, J$, $k=1, 2$, denotes the cell probabilities in the complete data matrix. Note only the marginal Q_{ij+} can be observed: i.e. $Q_{ij+} = f_{ij}/n = P_{ij}$.

Step 1: Let $Q_{ijk}^{(0)}$ represent initial estimates i.e., $Q_{++2}^{(0)} = \pi$ and $Q_{++1}^{(0)} = 1 - \pi$.

Step 2: For the first component Φ , set $Q_{ij1}^{(0)} = (1 - \pi)\hat{P}_{ij}^{(H)}$,

where $\hat{P}_{ij}^{(H)}$ represents the maximum likelihood estimate of P_{ij} under model H (here H is a row-column independence model, $\hat{P}_{ij}^{(H)} = f_{i+}f_{+j}/n^2$).

For the second component, set $Q_{ij2}^{(0)} = \pi(IJ)^{-1}$, $Q_{++1}^{(0)} = 1 - \pi$ and $Q_{++2}^{(0)} = \pi$, meet the requirement for this choice of starting values.

Step 3: At cycle s , the E (expectation) step of the algorithm is defined by

$$g_{ijk}^{(s)} = p_{ij} Q_{ijk}^{(s)} / Q_{ij+}^{(s)}, \text{ for all } i, j \text{ and } k.$$

Step 4: The M (maximization) step is $Q_{ij1}^{(s+1)} = (1 - \pi)Q_{ij1}^{(s)}$, where $Q_{ij1}^{(s)}$ is the maximum likelihood estimate of Φ in cell (i, j) under hypothesis H, i.e. $Q_{ij1}^{(s)} = Q_{i+1}^{(s)}Q_{+j1}^{(s)}$, where $Q_{i+1}^{(s)} = g_{i+1}^{(s)} / (1 - \pi)$ and $Q_{+j1}^{(s)} = g_{+j1}^{(s)} / (1 - \pi)$. For the second component, it turns out to be $Q_{ij2}^{(s+1)} = \pi(g_{ij2}^{(s)} / g_{++2}^{(s)})$.

Step 5: Repeat cycling between the E- and M-steps until a certain predefined criterion attained. For example, the criterion might be that the difference between Q_{ij+} for two successive cycle less than 10^{-5} .

The overall procedure to get estimate $\hat{\pi}^*$ is as follows: (1.) Set the initial estimate, $\hat{\pi}^*$ to zero; (2.) Obtain maximum likelihood estimates of the parameters in the components of the two-point mixture using an expectation-maximization (EM) algorithm as above, and, (3.) Successively increase $\hat{\pi}^*$ by some small increment (e.g., .01 was used in the example below) with re-estimation of the parameters at each step. The value of the likelihood ratio chi-square fit statistic, G^2 , becomes zero (approximately, if the convergence criterion is set to $<10^{-5}$) and the step at which this first happens yields the final estimate of the lack of fit index, $\hat{\pi}^*$ (Dayton, 2003; RCL). In addition, RCL implemented this approach in their FORTRAN program, Mixit, and also it was described in detail by Xi (1994).

To exemplify this procedure, we consider a fictional 4×4 frequency table with a total sample size of 489:

Table 1. Fictional 4×4 frequency table

Col/Row	1	2	3	4	Sum
1	10	20	40	80	150
2	5	10	20	40	75
3	15	30	69	120	234
4	2	4	8	16	30
Sum	32	64	137	256	489

When $\hat{\pi}^*=0$, which, in effect, represents the original frequency table, there is no evidence for lack of independence ($G^2=0.4811$, Chi-Square=0.4813, degrees of freedom=9, P value >0.05) and

$$\hat{P}_{ij}^{(H)} = f_{i+}f_{+j}/n^2 = \begin{Bmatrix} 0.020074 & 0.040147 & 0.085940 & 0.160588 \\ 0.010037 & 0.020074 & 0.042970 & 0.080294 \\ 0.031315 & 0.062629 & 0.134066 & 0.250518 \\ 0.004015 & 0.008029 & 0.017188 & 0.032118 \end{Bmatrix}$$

At Step 2, when $\hat{\pi}^*$ is increased to 0.01,

$$Q_{ij1}^{(0)} = \begin{Bmatrix} 0.019873 & 0.039746 & 0.08508 & 0.158982 \\ 0.009936 & 0.019873 & 0.04254 & 0.079491 \\ 0.031002 & 0.062003 & 0.132725 & 0.248012 \\ 0.003975 & 0.007949 & 0.017016 & 0.031797 \end{Bmatrix}$$

$$Q_{ij2}^{(0)} = \pi(IJ)^{-1} = \begin{Bmatrix} 0.000625 & 0.000625 & 0.000625 & 0.000625 \\ 0.000625 & 0.000625 & 0.000625 & 0.000625 \\ 0.000625 & 0.000625 & 0.000625 & 0.000625 \\ 0.000625 & 0.000625 & 0.000625 & 0.000625 \end{Bmatrix}$$

$$\text{At Step 3, } g_{ij1}^{(0)} = \begin{Bmatrix} 0.020455 & 0.040910 & 0.081819 & 0.163634 \\ 0.010227 & 0.020453 & 0.040907 & 0.081813 \\ 0.030680 & 0.061359 & 0.130879 & 0.245428 \\ 0.004090 & 0.008181 & 0.016361 & 0.032723 \end{Bmatrix}$$

$$\text{And } g_{ij2}^{(0)} = \begin{Bmatrix} 9.64\text{E-}08 & 1.51\text{E-}07 & 7.87\text{E-}17 & 2.73\text{E-}07 \\ 5.84\text{E-}08 & 9.83\text{E-}08 & 1.43\text{E-}16 & 2.53\text{E-}07 \\ 3.06\text{E-}12 & 1.80\text{E-}12 & 0.010243 & 7.89\text{E-}13 \\ 3.15\text{E-}08 & 5.64\text{E-}08 & 4.73\text{E-}16 & 1.85\text{E-}07 \end{Bmatrix}$$

After EM iteration, Step 3 and Step 4, the algorithm attains the convergence criterion and results in

$$Q_{ij1} = \begin{Bmatrix} 0.020290 & 0.040576 & 0.083681 & 0.162299 \\ 0.010143 & 0.020287 & 0.041838 & 0.081145 \\ 0.030969 & 0.061937 & 0.127736 & 0.247743 \\ 0.004057 & 0.008114 & 0.016734 & 0.032455 \end{Bmatrix}$$

$$Q_{ij2} = \begin{Bmatrix} 9.41\text{E-}08 & 1.47\text{E-}07 & 7.68\text{E-}17 & 2.66\text{E-}07 \\ 5.70\text{E-}08 & 9.59\text{E-}08 & 1.39\text{E-}16 & 2.47\text{E-}07 \\ 2.99\text{E-}12 & 1.76\text{E-}12 & 0.009999 & 7.70\text{E-}13 \\ 3.07\text{E-}08 & 5.51\text{E-}08 & 4.62\text{E-}16 & 1.81\text{E-}07 \end{Bmatrix}$$

Since $G^2=0.1022$, which is much greater than zero when $\hat{\pi}^* = .01$, we need to increase $\hat{\pi}^*$ to 0.02 and repeat the above EM cycle. Then, $G^2= 7.435\text{E-}8$ when $\hat{\pi}^* = .02$ and thus, .02 is the estimated mixture index of fit for this table with

$$Q_{ij1} = \begin{Bmatrix} 0.020310 & 0.040657 & 0.081802 & 0.162668 \\ 0.010157 & 0.020328 & 0.040899 & 0.08133 \\ 0.030646 & 0.061334 & 0.123404 & 0.245397 \\ 0.004061 & 0.008128 & 0.016354 & 0.032521 \end{Bmatrix}$$

$$Q_{ij^2} = \begin{Bmatrix} 0.000140 & 0.000242 & 7.60E-08 & 0.00093 \\ 6.83E-05 & 0.000123 & 5.01E-07 & 0.00047 \\ 0.000029 & 1.54E-05 & 0.017699 & 3.54E-06 \\ 2.87E-05 & 5.18E-05 & 5.30E-06 & 0.000199 \end{Bmatrix}$$

As a matter of fact, the independence for row and column would be true if the frequency in cell (3, 3) were 60 instead of 69; thus, $\hat{\pi}^* = 9/489 = .0184$. Note that the algorithm used to find $\hat{\pi}^*$ was rounded to .01, so the value of $\hat{\pi}^* = .02$ calculated above is correct within rounding.

This stepwise computational approach for two-way tables can be applied to virtually any frequency tables (RCL). Also they suggested a 10% value of $\hat{\pi}^*$, as representing “reasonable” fit for an exemplary 4×4 contingency table in their original paper. However, as noted by Dayton (2003), there is no absolute standard for the index that represents suitable fit in real data settings.

In addition to the above-mentioned procedure that involved repeated application of the EM algorithm proposed by RCL, Xi (1994) and Xi and Lindsay (1996) employed nonlinear programming (NLP) techniques to solve the estimation problem for $\hat{\pi}^*$ with respect to optimization. Dayton (2003) further extensively discussed this NLP for calculating $\hat{\pi}^*$.

Lower Bound and Standard error estimation for two-point mixture index

Generally speaking, $\hat{\pi}^*$ may overestimate lack of fit due to random fluctuation of sample data. Therefore, RCL proposed an appropriate lower 95% confidence bound, $\hat{\pi}_L$, based on a G^2 fit statistic equal to 2.70, the 90th percentage point of the one-degree-of-freedom chi-square distribution. Their method is capable of finding the lower bound using the same iterative procedure to compute $\hat{\pi}^*$. (Mixit, the

above-mentioned FORTRAN program in RCL can also be used to find a lower-bound estimate of $\hat{\pi}^*$). Furthermore, based on the definition of π^* , the confidence interval is one-sided due to the fact that all values of $\hat{\pi}$ greater than $\hat{\pi}^*$ would result in models that perfectly fit the observed frequencies (i.e., $G^2=0$ if $\hat{\pi} > \hat{\pi}^*$) (RCL, Dayton 2003). Also, the standard error of π^* can be estimated using re-sampling techniques such as the jackknife (Dayton, 1999, Dayton, 2003). Clogg, Rudas and Xi (1995) advised that the difference of π^* and its lower bound $\hat{\pi}_L$ provides a measure of the effect of sample size on $\hat{\pi}^*$ which means although the point estimator of π^* does not depend on sample size, the confidence interval of π^* , on the contrary, will be shorter for the larger sample than for the smaller sample which seems actually very attractive. (Clogg, Rudas, and Xi, 1995)

In practice we must be careful about using $\hat{\pi}_L$ when π^* close to zero and when the sample size is small. In such cases, they suggested using one of the two methods (RCL):

1. Replace the critical value of the χ^2 statistics with DF degrees of freedom with the original critical value 2.70, or
2. Simulate the null distribution.

The standard error could be derived based on lower 95% confidence bound of a standard normal distribution, i.e. $se = (\hat{\pi}^* - \hat{\pi}_L)/1.645$.

Sampling zeros and Structural zeros

According to RCL, the effect of sampling zeros on $\hat{\pi}^*$ will rely on the structure of the data and the suitability of the model, H, for the data. The occurrence of a sampling zero will force the estimate of the row or the column total in Φ to zero under the

assumption of an independence model. Typically, $\hat{\pi}^*$ will tend to be overestimated by a fraction that is directly related to the smaller of the observed row marginal proportion and the observed column marginal proportion related to the cell with a sampling zero. The sampling zero is an indication of the possibility of an extreme departure from row-column independence in case of relatively large marginal probabilities, and therefore $\hat{\pi}^*$ would be expected to be comparatively large. Rudas and Zwick (1997) replaced zero frequencies with small positive flattening constants in data from a study by Zwick, Thayer and Wingersky (1994) to investigate the sampling zero effect on the performance of π^* . The analysis was conducted with various choices of flattening constants (0.0001, 0.001, 0.01, 0.1 and 0.5). Although they concluded that increases in the flattening values result in decreases in the estimates for π^* , the effect was very small in their example.

Structural zeros, also called logical zeros (Knoke and Burke, 1980), arise when it is logically impossible to observe positive cell counts for particular combinations of row and column variables. To demonstrate structural zeros, a typical example of the logical impossibility of observing male obstetrical patients was presented by Fienberg (1980). In practice, researchers could evaluate the variation in π^* by setting some to-be-ignored cells to structural zeros. Generally, sampling zeros refer to table cells for which the observed frequency is equal to zero but the expected frequency is greater than zero while the expected value is equal to zero in the case of structural zeros. The impact of sampling zeros and structural zeros on π^* can be, under certain circumstances, very large and thus, is worthy of additional study.

CHAPTER III METHODS

Research Design

This study: (a) evaluated several competing techniques for dealing with sampling zeros for the two-point mixture model index, π^* , in contingency tables when the independence assumption holds; (b) investigated the performance of the estimate, $\hat{\pi}^*$, in various combinations of conditions, as a function of different sizes of tables, different marginal distributions and different sample sizes; and (c) compared the standard error of $\hat{\pi}^*$ and confidence interval estimated by use of a method proposed by RCL with the “true” standard error based on empirical simulations in various scenarios especially when encountering small sample sizes and π^* close to zero. These goals were achieved by Monte Carlo methods that simulated a variety of scenarios.

The following aspects of the simulation were implemented:

1. Size of two-way contingency table: 2×2 , 2×3 , 2×4 , 2×6 , 3×3 , 4×4 and 6×6 . These table sizes were chosen because they provided a reasonable range of contingency table sizes in real data settings and are typical of what is found in practice.
2. Marginal distribution: evenly distributed, slightly and extremely dispersed distribution for each different size of tables. Row and column total proportion for 2×2 table:

$$\{P_{1+}=.5, P_{2+}=.5, P_{+1}=.5, P_{+2}=.5\},$$

$$\{P_{1+}=.9, P_{2+}=.1, P_{+1}=.9, P_{+2}=.1\},$$

$$\{P_{1+}=.5, P_{2+}=.5, P_{+1}=.9, P_{+2}=.1\}.$$

2×3 table:

$$\{P_{1+}=.5, P_{2+}=.5, P_{+1}=.8, P_{+2}=.1, P_{+3}=.1\},$$

$$\{P_{1+}=5, P_{2+}=5, P_{+1}=33, P_{+2}=33, P_{+3}=33\},$$

$$\{P_{1+}=9, P_{2+}=1, P_{+1}=8, P_{+2}=1, P_{+3}=1\},$$

$$\{P_{1+}=9, P_{2+}=1, P_{+1}=33, P_{+2}=33, P_{+3}=33\}.$$

2×4 table:

$$\{P_{1+}=5, P_{2+}=5, P_{+1}=25, P_{+2}=25, P_{+3}=25, P_{+4}=25\},$$

$$\{P_{1+}=5, P_{2+}=5, P_{+1}=4, P_{+2}=4, P_{+3}=1, P_{+4}=1\},$$

$$\{P_{1+}=9, P_{2+}=1, P_{+1}=25, P_{+2}=25, P_{+3}=25, P_{+4}=25\},$$

$$\{P_{1+}=9, P_{2+}=1, P_{+1}=4, P_{+2}=4, P_{+3}=1, P_{+4}=1\}.$$

2×6 table:

$$\{P_{1+}=5, P_{2+}=5, P_{+1}=167, P_{+2}=167, P_{+3}=167, P_{+4}=167, P_{+5}=167, P_{+6}=167\},$$

$$\{P_{1+}=5, P_{2+}=5, P_{+1}=3, P_{+2}=3, P_{+3}=1, P_{+4}=1, P_{+5}=1, P_{+6}=1\},$$

$$\{P_{1+}=9, P_{2+}=1, P_{+1}=167, P_{+2}=167, P_{+3}=167, P_{+4}=167, P_{+5}=167, P_{+6}=167\},$$

$$\{P_{1+}=9, P_{2+}=1, P_{+1}=3, P_{+2}=3, P_{+3}=1, P_{+4}=1, P_{+5}=1, P_{+6}=1\}.$$

3×3 table:

$$\{P_{1+}=4, P_{2+}=4, P_{3+}=2, P_{+1}=4, P_{+2}=4, P_{+3}=2\},$$

$$\{P_{1+}=33, P_{2+}=33, P_{3+}=33, P_{+1}=33, P_{+2}=33, P_{+3}=33\},$$

$$\{P_{1+}=33, P_{2+}=33, P_{3+}=33, P_{+1}=4, P_{+2}=4, P_{+3}=2\}.$$

4×4 table:

$$\{P_{1+}=25, P_{2+}=25, P_{3+}=25, P_{4+}=25, P_{+1}=25, P_{+2}=25, P_{+3}=25, P_{+4}=25\},$$

$$\{P_{1+}=4, P_{2+}=4, P_{3+}=1, P_{4+}=1, P_{+1}=4, P_{+2}=4, P_{+3}=1, P_{+4}=1\},$$

$$\{P_{1+}=25, P_{2+}=25, P_{3+}=25, P_{4+}=25, P_{+1}=4, P_{+2}=4, P_{+3}=1, P_{+4}=1\}.$$

6×6 table:

$$\{P_{1+}=167, P_{2+}=167, P_{3+}=167, P_{4+}=167, P_{5+}=167, P_{6+}=167, P_{+1}=167,$$

$P_{+2=}.167, P_{+3=}.167, P_{+4=}.167, P_{+5=}.167, P_{+6=}.167\}$,

$\{P_{1+=}.3, P_{2+=}.3, P_{3+=}.1, P_{4+=}.1, P_{5+=}.1, P_{6+=}.1, P_{+1=}.3, P_{+2=}.3, P_{+3=}.1, P_{+4=}.1, P_{+5=}.1, P_{+6=}.1\}$,

$\{P_{1+=}.167, P_{2+=}.167, P_{3+=}.167, P_{4+=}.167, P_{5+=}.167, P_{6+=}.167, P_{+1=}.3, P_{+2=}.3, P_{+3=}.1, P_{+4=}.1, P_{+5=}.1, P_{+6=}.1\}$.

These marginal distributions were chosen because they represented a reasonable range of different values. Also, the extreme marginal values were used to ensure zero cell frequencies in the observed tables.

3. Sample size for simulated contingency table: 5 per cell, 10 per cell, 20 per cell and 30 per cell. They were chosen because they entailed a practical variety of sample sizes and were large enough to demonstrate a sample size effect on the mixture index of fit.

4. Techniques for zeros cells: treating as sampling zeros, replacing with different small flattening constants (here we used .1, .5 and 1 to represent extremely small, moderately small and small flattening constants range), and redefining model H by regarding the sampling zero as structural zero.

5. In each of the above scenarios, a 95% lower confidence limit based on empirically simulated $\hat{\pi}^*$ s was calculated and compared with the limit derived following RCL.

The complete procedure to estimate $\hat{\pi}^*$ is as follows: (1.) Set the initial estimate, $\hat{\pi}^*$ to, zero; (2.) Obtain maximum likelihood estimates of the parameters in the components of the two-point mixture using an expectation-maximization (EM) algorithm, and (3.) Successively increase $\hat{\pi}^*$ by some small increment (e.g., .01 has been used in this study) with re-estimation of the parameters at each step. The value of the likelihood ratio chi-square fit statistic, G^2 , becomes zero (approximately, since the convergence

criterion was set to 10^{-5}) and the step at which this first happens yields the final estimate, $\hat{\pi}^*$. (RCL, Dayton 2003)

For each table size, sample size and marginal distribution, 1000 frequency tables, were randomly generated based on the specified cumulative distribution to estimate $\hat{\pi}^*$ and the 95% lower bound $\hat{\pi}_L$. For example, for a 2×2 table with sample size of 10 per cell and marginal distribution $\{P_{1+}=.9, P_{2+}=.1, P_{+1}=.9, P_{+2}=.1\}$, the consequent cumulative distribution is $\{0.81, 0.90, 0.99, 1\}$. To generate each of the 1000 simulated data tables, SAS (SAS institute, 2005) generated 40 uniform random numbers on scale 0, 1 and locate them into appropriate cumulative categories according to their positions. (e. g., numbers less than or equal .81 were placed in cell 1, .81; numbers between .81 and .9 in cell 2; numbers between .90 and .99 in cell 3 and the rest in cell 4.) The value of $\hat{\pi}^*$ and associated 95% lower bound $\hat{\pi}_L$ following RCL were obtained for each generated data table, thus for each scenario, 1000 $\hat{\pi}^*$ and 1000 95% lower bound $\hat{\pi}_L$ using RCL method were generated. This was repeated for each of the 96 scenarios. Also for each scenario, four techniques for sampling zeros cells were compared including treating zero cells as sampling zeros, replacing with different small flattening constant (here we used .1, .5 and 1), and redefining model H by regarding a sampling zero as a structural zero.

The mean of the 1000 $\hat{\pi}^*$ values for each scenario was calculated and served as the final parameter estimate; the mean of the 1000 $\hat{\pi}_L$ values was also computed to be the estimate 95% $\hat{\pi}_L$ using RCL method. Since the empirical distribution of $\hat{\pi}^*$ is notably skewed for the generated sets of 1000 $\hat{\pi}^*$ values (see Figure 1 for a histogram for a case

with relatively small bias which shows extreme skewness and for a case relatively large bias for which skewness is still somewhat apparent), we cannot use the regular normal assumption to compute the standard error and confidence interval for $\hat{\pi}^*$. Instead, 50th $\hat{\pi}^*$ value among the 1000 values (i.e., 5th percentage point) was adopted and treated as “true” 95% lower bound based on empirical simulations.

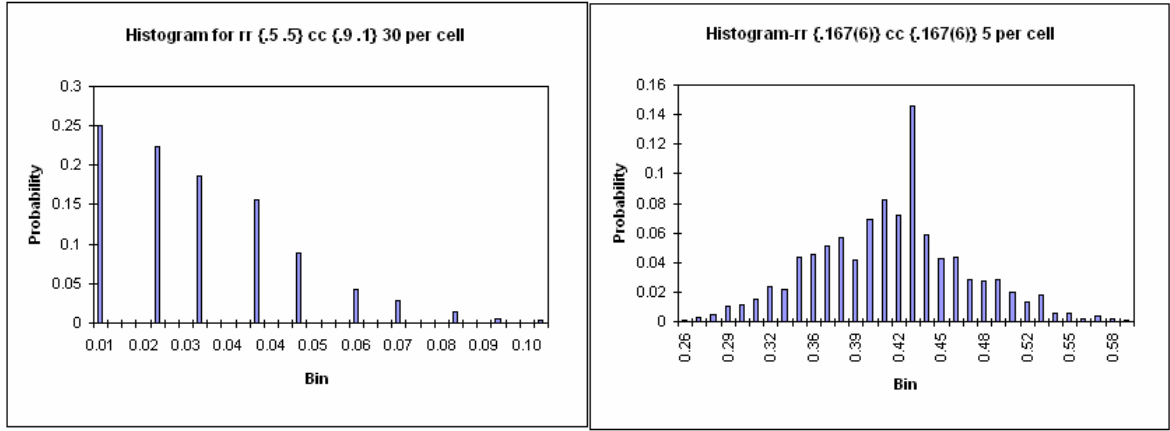


Figure 1: Histograms for $\hat{\pi}^*$ with small parameter bias (left) and large parameter bias (right).

Typically, $\hat{\pi}^*$ will tend to be overestimated by a fraction that is directly related to the smaller of the observed row marginal proportion and the observed column marginal proportion related to the cell with a sampling zero (RCL). As noted above, in practice, researchers could test the π^* variation by setting some to-be-ignored cells to structural zeros to resolve. In this study we focused this issue on any frequency tables with only one structural zero and the procedure using EM based methodology to obtain $\hat{\pi}^*$. The two-point mixture using an expectation-maximization (EM) algorithm which RCL proposed could still be applied to structural zero condition with minor modification.

Step 1: obtain $\hat{\pi}^*$ treating zero cell as sampling zero. In this step the entire row or column with which smaller of observed row marginal proportion and the observed

column marginal proportion would result in zero in the first component, Φ , which is defined as the probability distribution designated by H.

Step 2: pull the proportion back from the second component, ψ , an unspecified probability distribution outside of model H for the entire row or column with zeros in component Φ at step 1.

Step 3: temporally cross out the other column or row that contains the zero cell but not been forced zero at step 1.

Step 4: apply the same EM based procedure in the remaining contingency table while fixing all cell proportions in component 1, Φ and component 2, ψ except the row or column has frequency pulled back in step 2.

Step 5: after iteration converges, subtract original $\hat{\pi}^*$ at step 1 with the sum of the proportion pulled back in Φ and the final value is the estimate of $\hat{\pi}^*$ using structural zero technique.

For the other sampling zero techniques, procedures are same as sampling zeros, just replacing the zero cell with different small flattening constant (.1, .5 and 1) and recall associated $\hat{\pi}^*$.

Simulation Details

The simulation code was written in SAS/IML version 9.1 (SAS institute, 2005). The EM algorithm was used to calculate the mixture index of fit. Each simulation consisted of 1,000 replications with convergence criterion set to 10^{-5} . Data were randomly generated according to cumulative proportion resulted from the different combination scenarios.

The method proceeded in the following manner:

(1) A sample contingency table was randomly generated based on cumulative proportion

resulted from different factors combinations. (table size, sample size and marginal distribution).

(2). EM algorithm based method for mixture index of fit (RCL) was implemented. $\hat{\pi}^*$ and 95% lower bound $\hat{\pi}_L$ were generated and saved in a matrix.

(3). Replicate step 1 and step 2 for 1000 times therefore 1000 $\hat{\pi}^*$ and $\hat{\pi}_L$ were obtained and exported into an external file.

Additionally, if any of 1000 generated contingency tables contains zero cell(s), Replacing with different small flattening constants 0.1, 0.5 and 1 respectively when evaluating performance of $\hat{\pi}^*$ using replacing flattening constants techniques.

The only difference between structural zero and other sampling zero technique procedure is in the above-mentioned step 1. If frequency tables generated by SAS UNIFORM contains 1 or less than 1 frequency zero, it would proceed to step 2 otherwise it would regenerate table until meets the requirement.

Results of the discrepancy measure of $\hat{\pi}^*$ were tabulated, plotted and presented for various combinations of conditions, as a function of different sizes of tables, different marginal distributions and different sample sizes in Appendix A. Comparison outcomes of each competing techniques for the sampling zeros and structural zero effects on π^* were plotted and presented and the standard error differences of $\hat{\pi}^*$ between the method proposed by RCL, and the “true standard error” which based on empirical parametric simulation in various scenarios were also presented in the appendix A.

Fatal crashes by speed limit and land use frequency table

Table 2 presents fatal crashes by speed limit and land use in the United States in 2004 from Traffic Safety Facts 2004: A compilation of Motor Vehicle Crash Data from

the Fatality Analysis Reporting System and the General Estimates System. The National Highway Traffic Safety Administration (NHTSA) provides descriptive statistics about traffic crashes of all severities, from those that result in property damage to those that result in the loss of human life. The Fatality Analysis Reporting System (FARS) became operational in 1975; it keeps track of data on a census of fatal traffic crashes within the 50 states, the District of Columbia, and Puerto Rico. According to NHTSA, a crash must involve a motor vehicle traveling on a traffic way customarily open to the public, and must result in the death of an occupant of a vehicle or nonmotorist within 30 days of crash in order to be included in FARS. The fatal crashes by speed limit and land use table contains 37,295 cases (omitting 958 cases for the unknown speed limit category). There are three categories (Rural, Urban and Unknown) in Land Use variable, and six categories in Speed limit variable (30 mph or less, 35 or 40 mph, 45 or 50 mph, 55mph, 60 mph or higher and no statutory limit). Urban and rural are defined as “an urban area is an area whose boundaries shall be those fixed by responsible state and local officials in cooperation with each other and approved by the Federal Highway Administrations, U. S. Department of Transportation. Such boundaries are established in accordance with the provisions of Title 23 of the United States Code. Urban area boundary information is available from state highway or transportation departments. In the event that boundaries have not been fixed as above for any urban place designated by the Bureau of the Census having a population of 5000 or more, the area within boundaries fixed by the Bureau of the Census shall be an urban area. A rural area is any area which is not within urban areas” (NHTSA).

This data table is used to compare the conclusion using traditional chi-square and

related model fit methods and the mixture index of fit introduced by RCL. More specifically, compare different sampling zero techniques impact on $\hat{\pi}^*$ since there is one zero cell in the contingency table.

Table 2. Fatal crashes by speed limit and land use

Speed Limit	Land Use		
	Rural	Urban	Unknown
30 mph or less	944	2929	27
35 or 40 mph	1951	4463	41
45 or 50 mph	3496	3559	46
55 mph	9646	2121	91
60 mph or higher	5484	2347	27
No statutory limit	92	31	0

Source: USDOT Traffic Safety Facts 2004 (Fatality Analysis Reporting System).

Note: Omit 958 cases for the Unknown Speed Limit category.

Cross-classifying eye color and hair color contingency table

The second example is a cross-classification of eye color and hair color table (Snee, 1974), a 4×4 table with sample size of 592.

Table 3. Cross-classification of eye color and hair color

Eye color	Hair color			
	Black	Brunette	Red	Blonde
Brown	68	119	26	7
Blue	20	84	17	94
Hazel	15	54	14	10
Green	5	29	14	16

Source, Snee (1974) and Diaconis and Efron (1985).

RCL utilized this data example in their paper to study the properties of mixture index of fit. In this study, this example was used to compare the difference between sampling zero and structural zero. The 16 cells were force to be zero one-by-one in turn to study the relationship of data structure and suitability on $\hat{\pi}^*$ as well as the impact of

sampling zero and structural zero on $\hat{\pi}^*$.

Statistical Analysis

In order to effectively compare $\hat{\pi}^*$ and $\hat{\pi}_L$ across difference combinations of scenario, confidence intervals and conventional z test tests for a single proportion, two proportions were adopted (Fleiss, Joseph L., 1981). If a 95% lower confidence bound does not contain zero; we can conclude that there is evidence for some lack of fit of the null model which means two proportions P1 and P2 are significantly differently from each other at significance level .05.

CHAPTER IV RESULTS

Parameter Estimates and Bias

The parameter estimates $\hat{\pi}^*$ from the various simulations are presented in Appendix A where tables are appropriately prefixed along with the estimates under different sampling zero techniques denoted by the column headings.

The convergence of the repeated EM algorithm for estimating the mixture index of fit $\hat{\pi}^*$ works smoothly for the step of each EM cycle at which π is assumed fixed with the criterion for the difference between the maximum likelihood estimates for two successive cycle being 10^{-5} , (here the criterion of 10^{-5} is precise and provides the same results as using a criterion of 10^{-10}). Also, the convergence at the step of the likelihood ratio chi-square fit statistic, G^2 , becoming zero (approximately, if the convergence criterion is set to $<10^{-5}$) by successively increasing $\hat{\pi}^*$ by small increment (e.g., .01) works satisfactorily except that in 6 cells out of total 480 (96 scenarios times 5 sampling zero techniques) did not converge at 10^{-5} but only converged at 10^{-2}).

$\hat{\pi}^*$ is significantly positively ($p < .05$, using conventional z test for proportions) biased from zero by .02298 (2×2 table, slightly dispersed row and column marginals with sample size equals to 30 per cell) to .4086 (6×6 table, evenly dispersed row and column marginals with sample size equals to 5 per cell) for the conditions studied (all table sizes, marginal distributions, and sample sizes) when the independence assumption holds. The magnitude of bias, however, varies across situations. As shown in Figures 2 and 3, for 2×2 , 2×3 , 2×4 , 2×6 tables, when table size increases, $\hat{\pi}^*$ consistently increases for constant sample size (5, 10, 20 and 30 per cell) and marginal distribution (evenly, slightly

and extremely dispersed) with the exceptions of 2×4 extremely dispersed table with sample size 20 and 30 per cell. The same conclusion applies to symmetric tables of sizes 2×2 , 3×3 , 4×4 , 6×6 . In particular, for sample size 5, 10, 20 and 30 per cell evenly dispersed table, $\hat{\pi}^*$ increases from .1252 to .4086; .096 to .3031; .0775 to .2242 and .0668 to .1867 for 2×2 to 6×6 table, respectively. For sample size 5, 10, 20 and 30 per cell extremely dispersed table marginals, $\hat{\pi}^*$ increases from .0598 to .3629; .0568 to .2593; .0476 to .1942 and .0396 to .1626 for a 2×2 table to a 6×6 table, respectively. Moreover, with few exceptions, for each frequency table, as sample size increases, $\hat{\pi}^*$ significantly decreases ($p < .05$, conventional z test for proportions). For each size of contingency table, $\hat{\pi}^*$ is smallest for extremely dispersed row and column marginal distributions while largest for evenly distributed row and column table. The only exception is 2×2 table where a slightly dispersed table contains slightly smaller $\hat{\pi}^*$ values on average than extremely dispersed frequency table, partly due to a convergence problem (used less than .001 instead of otherwise .00001) as shown in Appendix A.

For all two-way tables, replacing zero with larger flattening values results in smaller $\hat{\pi}^*$. For all extremely dispersed and most slightly dispersed (4 out of 6 scenarios) row and column marginal distributions with small sample size (5 per cell) and small table size (2×2 , 2×3 , 2×4 , 3×3) tables, the value of $\hat{\pi}^*$ is smaller using structural zero compared to using sampling zero or any other replacing with positive flattening constants techniques.

Also, Figures 4, 5, and 6 show that for all sample size in 2×2 , 2×4 table, 5, 10, 20 per cell for 2×3 table and 10 per cell for 2×6 tables with extremely dispersed row and column marginal distributions, $\hat{\pi}^*$ is significantly smaller ($p < .05$, conventional z test

for proportions) using structural zero technique compared with sampling zero. However, $\hat{\pi}^*$ values for the different techniques become closer to each other when sample size becomes moderately large (30 per cell) using all the various sampling zero techniques (including structural zero). Note that the techniques of replacing zero cell with flattening constants include virtually any number of simulated zero cells for each table while the structural zero technique used in this study can only accommodate one zero cell per frequency table. Because the number of zero counts and patterns are somewhat different among these techniques, especially when encountering small sample sizes such as 5 per cell and 10 per cell with extremely dispersed marginal distributions for which the percentage of number of two or above sampling zero cells could go up to 86.2% (2×6 small sample size with extremely dispersed marginal table), it might influence the comparison results between structural zero and using sampling zero or any other replacing with small positive flattening constants techniques. However, for comparatively large sample size (20 per cell and 30 per cell) tables with any marginal distributions or evenly dispersed tables with any sample size, the comparison results remain valid.

Notable difference for average $\hat{\pi}^*$ values between sampling zero and structural zero techniques appears for different marginals when sample size is small (5 per cell). Figures 7 to 9 display the structural zero technique influence on $\hat{\pi}^*$ for different marginal distributions with 2×3 , 2×4 and 6×6 table. Again, $\hat{\pi}^*$ is lowest for extremely dispersed marginal distribution, followed by slightly dispersed table and biggest for evenly distributed table and $\hat{\pi}^*$ is consistently smaller when adopting structural zero technique compared to sampling zero. The magnitude of difference, however only significantly smaller ($p < .05$, conventional z test for proportions; sample size equals to 5

per cell) using structural zero for extremely dispersed 2×3 table, slightly dispersed ($\{P_{1+}=.5, P_{2+}=.5, P_{+1}=.8, P_{+2}=.1, P_{+3}=.1\}$) 2×3 table; extremely dispersed 2×4 table, slightly dispersed ($\{P_{1+}=.9, P_{2+}=.1, P_{+1}=.25, P_{+2}=.25, P_{+3}=.25, P_{+4}=.25\}$) 2×4 table and extremely dispersed 6×6 table.

Averaging over results, of the sampling zero techniques that were compared, replacing with larger flattening constants such as 1 and structural zero technique appear to be more superior to the others in the sense that $\hat{\pi}^*$ is consistently smaller. Between these two methods, structural zero technique performs better for all extremely and most slightly dispersed row and column marginal distributions tables with small sample size and small table size while in most other cases replacing with larger flattening constant (i.e., 1) appears to be more superior in relation to parameter bias.

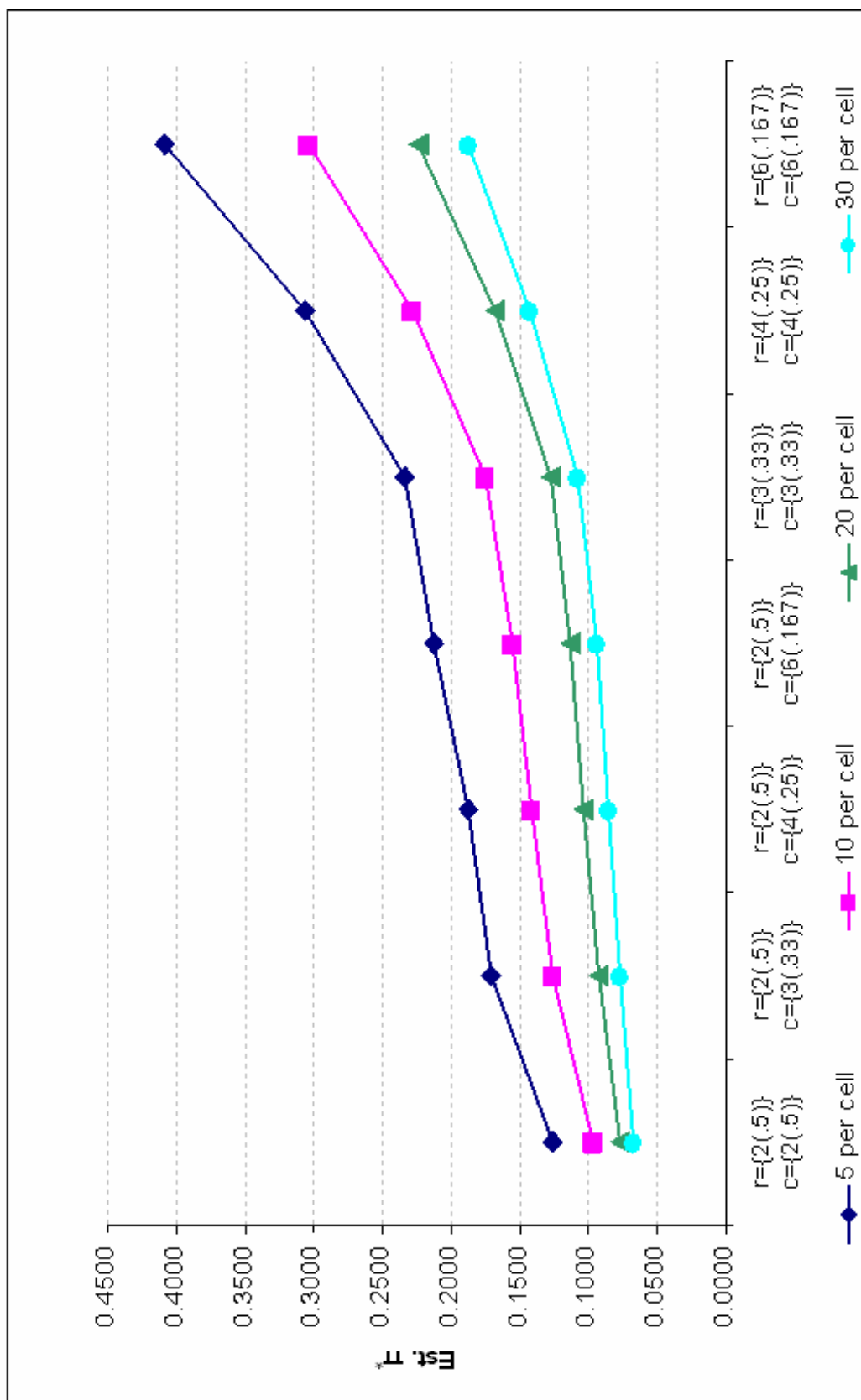


Figure 2. \hat{T}^* for Evenly Distributed Marginals

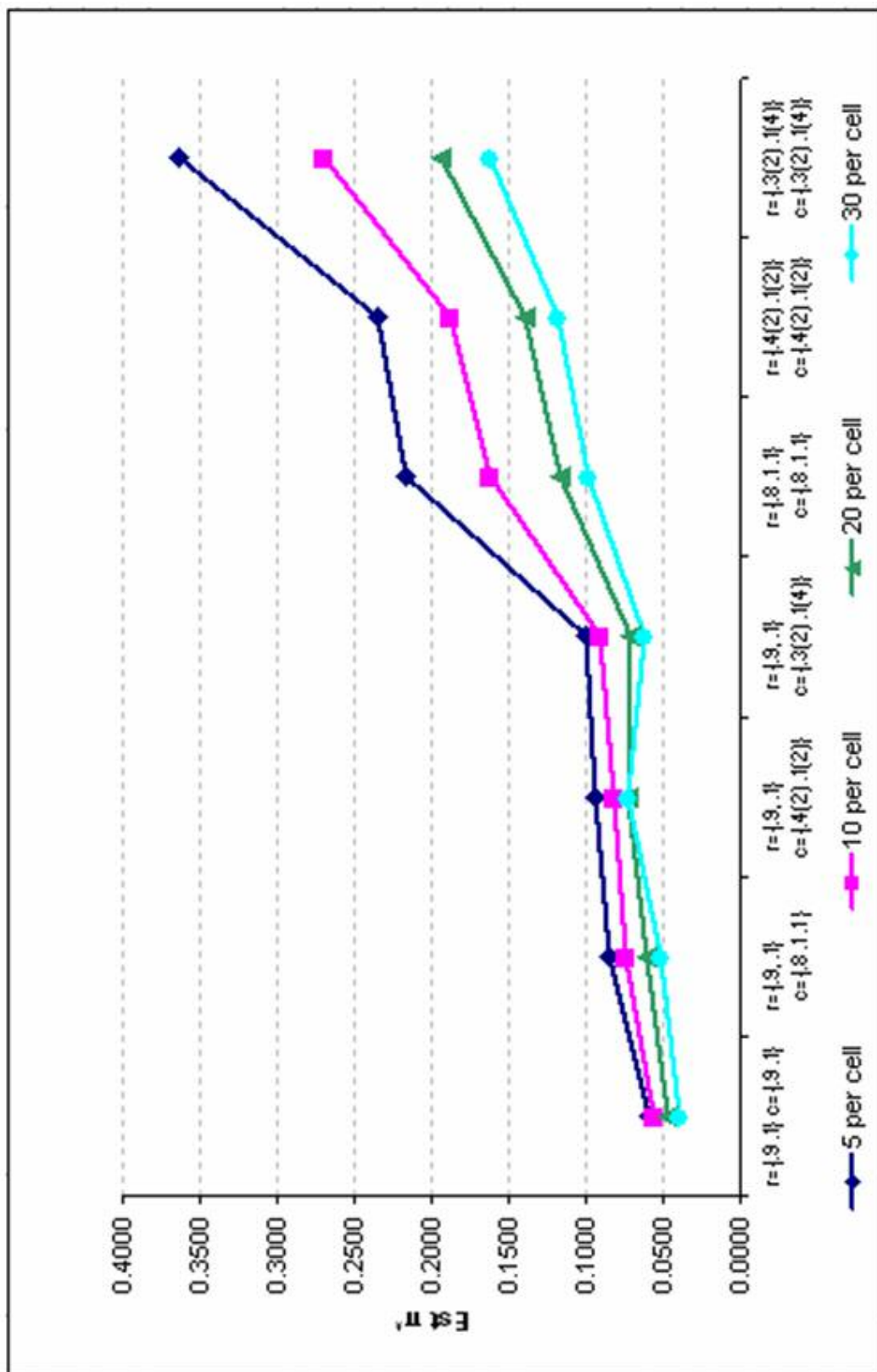


Figure 3. $\hat{\pi}^*$ for Extremely Distributed Marginals

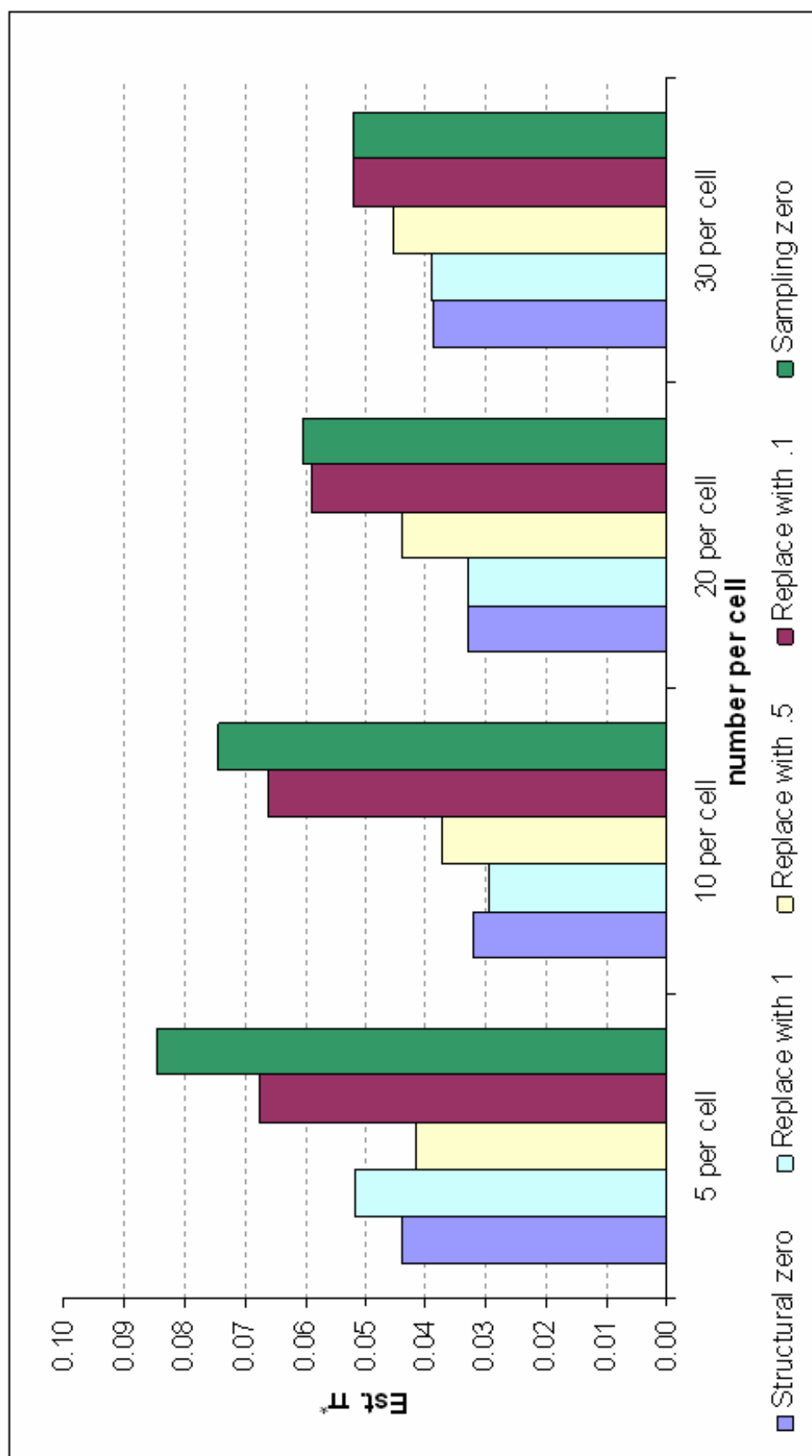


Figure 4. $\hat{\pi}^*$ Comparison in $\{P_{1+}=.9, P_{2+}=.1, P_{4+}=.8, P_{42}=.1, P_{43}=.1\}$

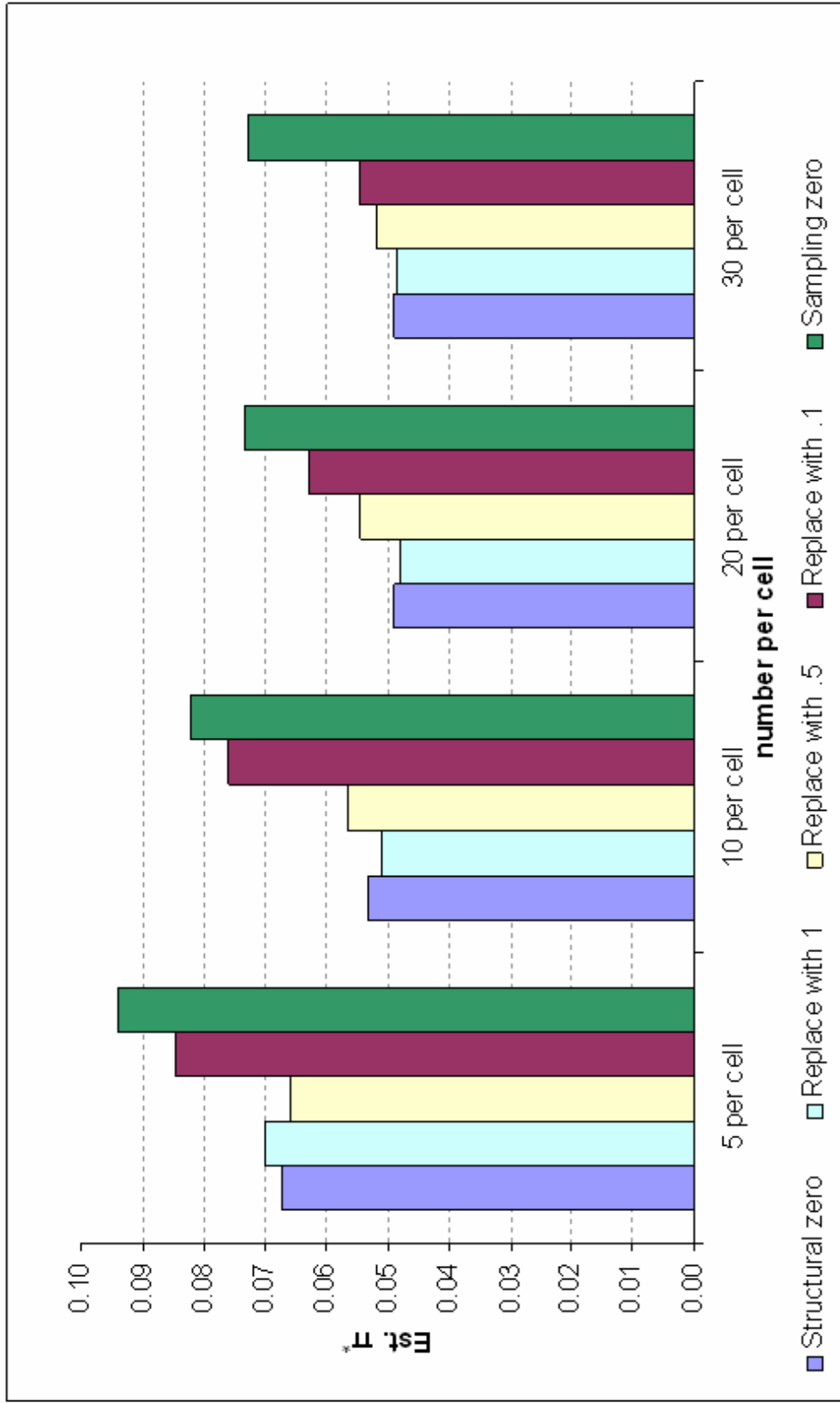


Figure 5. $\hat{\pi}^*$ Comparison in $\{P_{1+}=.9, P_{2+}=.1, P_{3+}=.4, P_{4+}=.1, P_{5+}=.1\}$

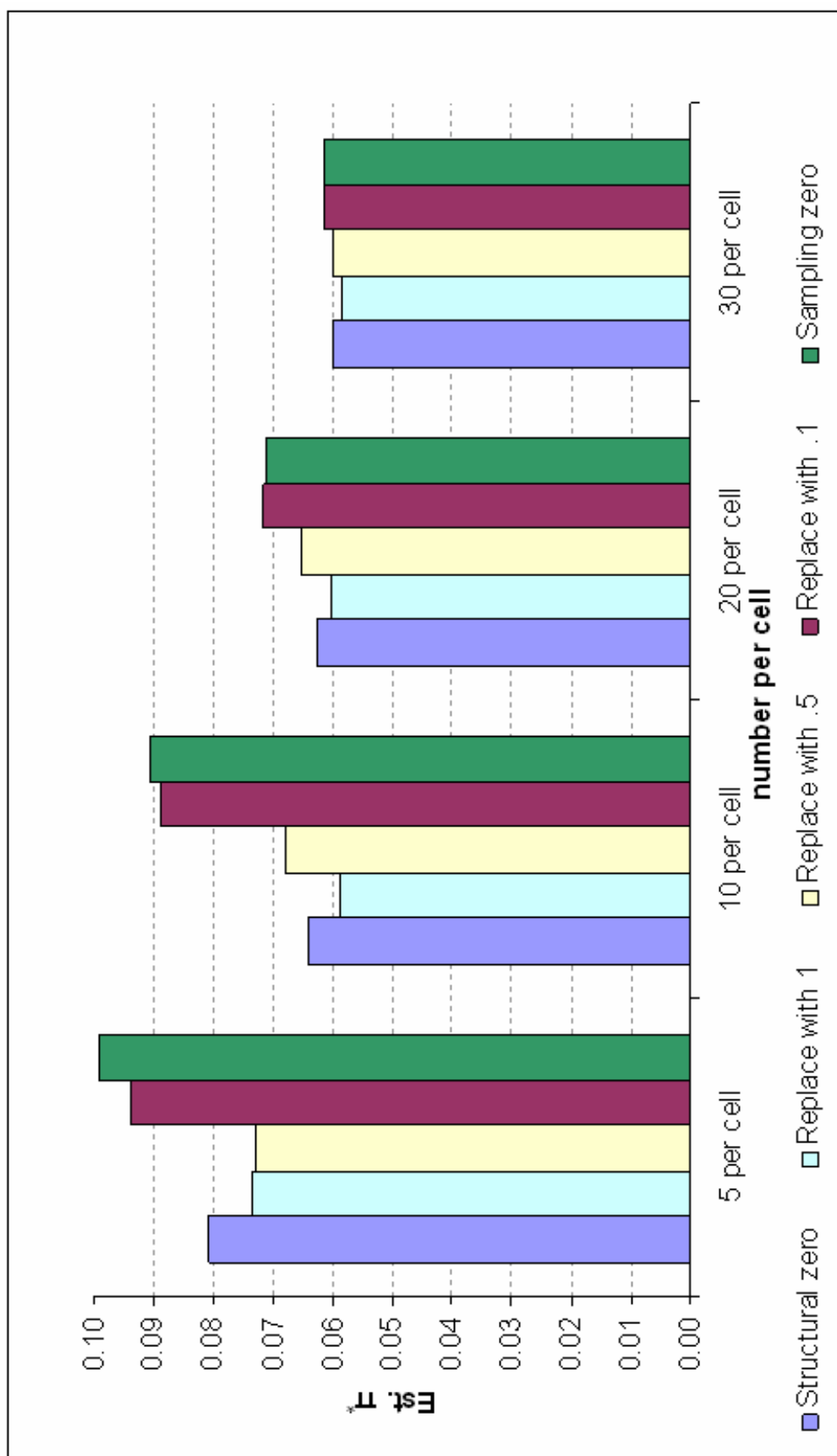


Figure 6. $\hat{\pi}^*$ Comparison in $\{P_1 = .9, P_2 = .1, P_3 = .3, P_4 = .1, P_5 = .1, P_6 = .1\}$

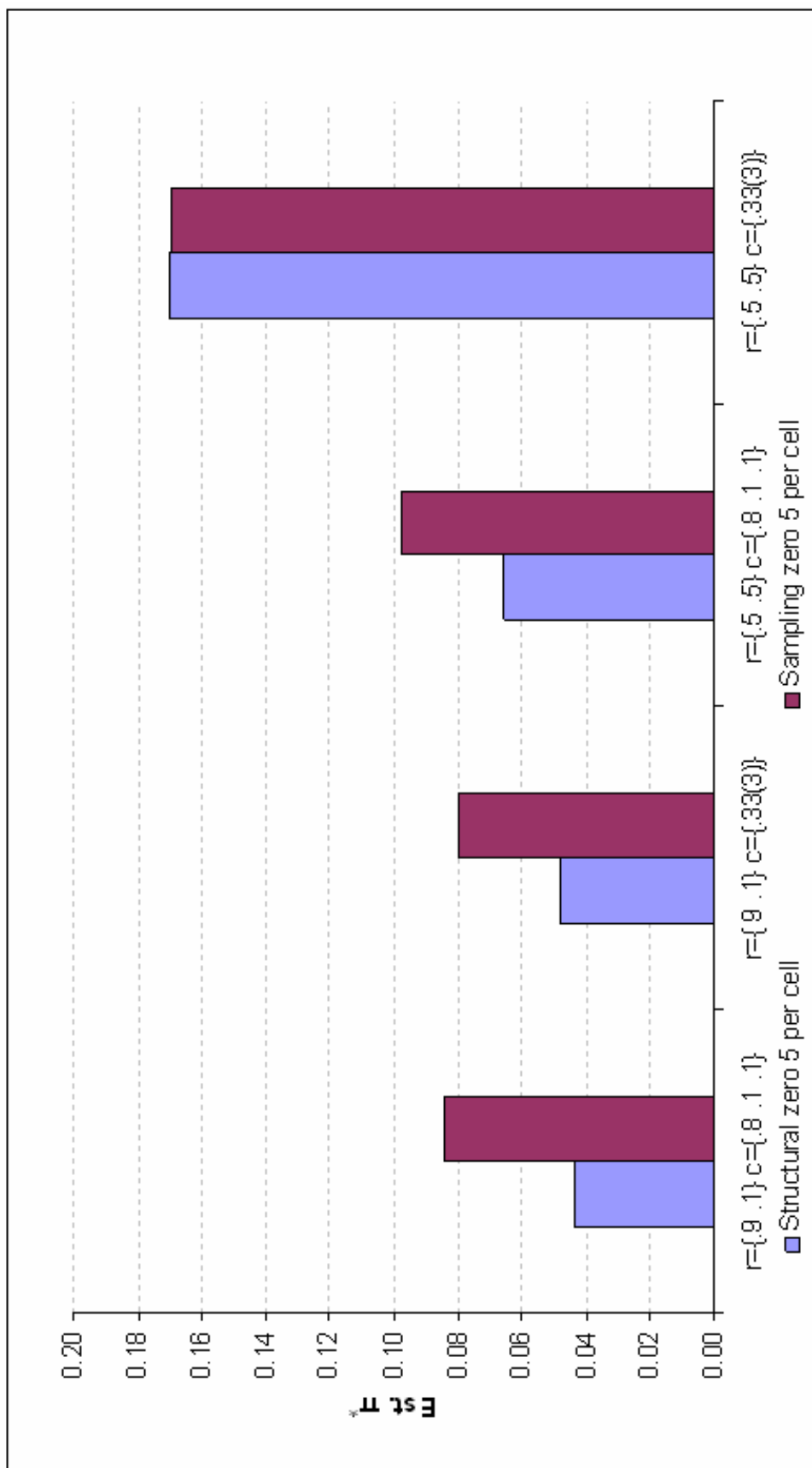


Figure 7. Sampling Zero and Structural Zero Effect on $\hat{\pi}^*$ Comparison in 2×3 table

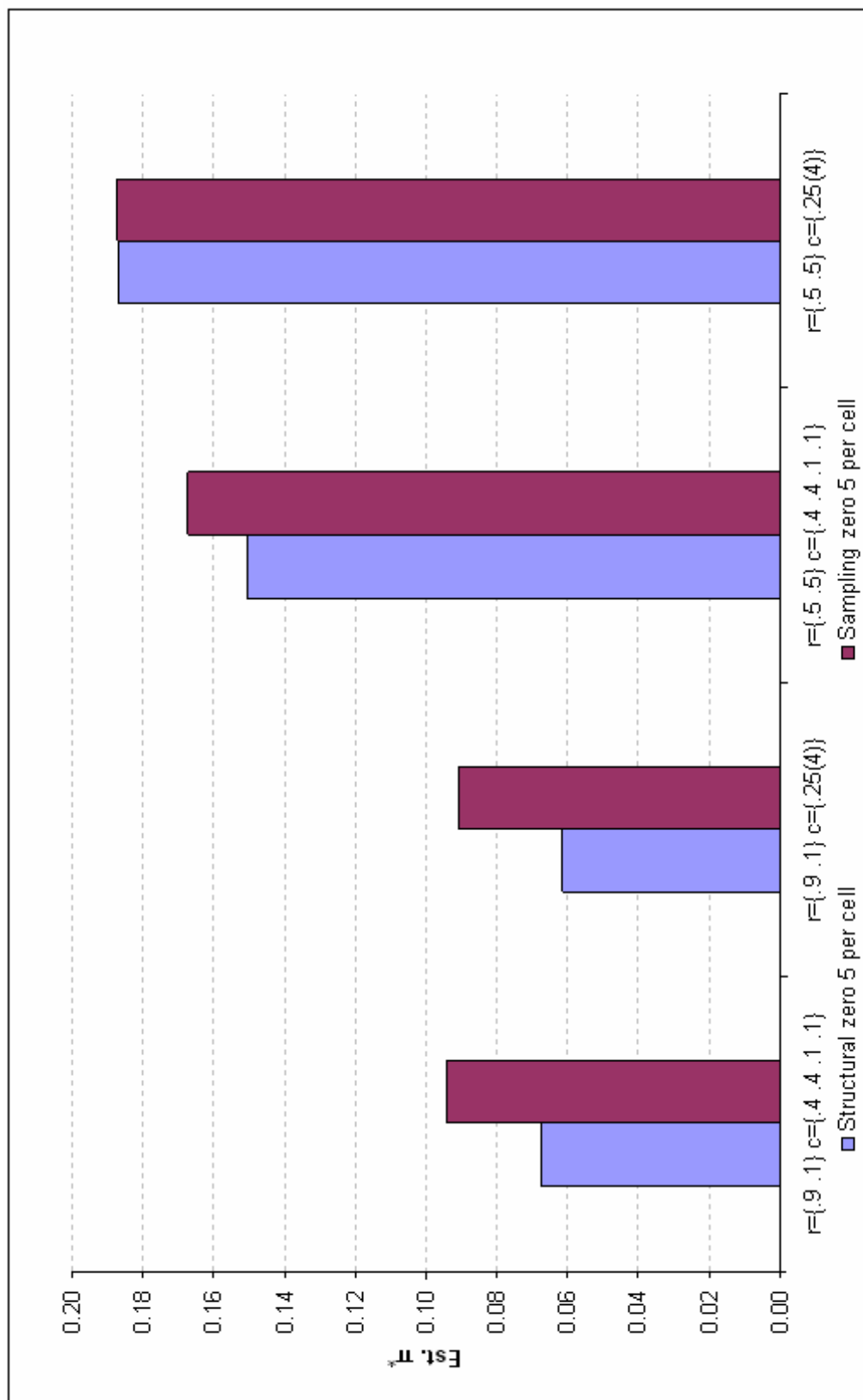


Figure 8. Sampling Zero and Structural Zero Effect on $\hat{\pi}$ Comparison in 2×4 table

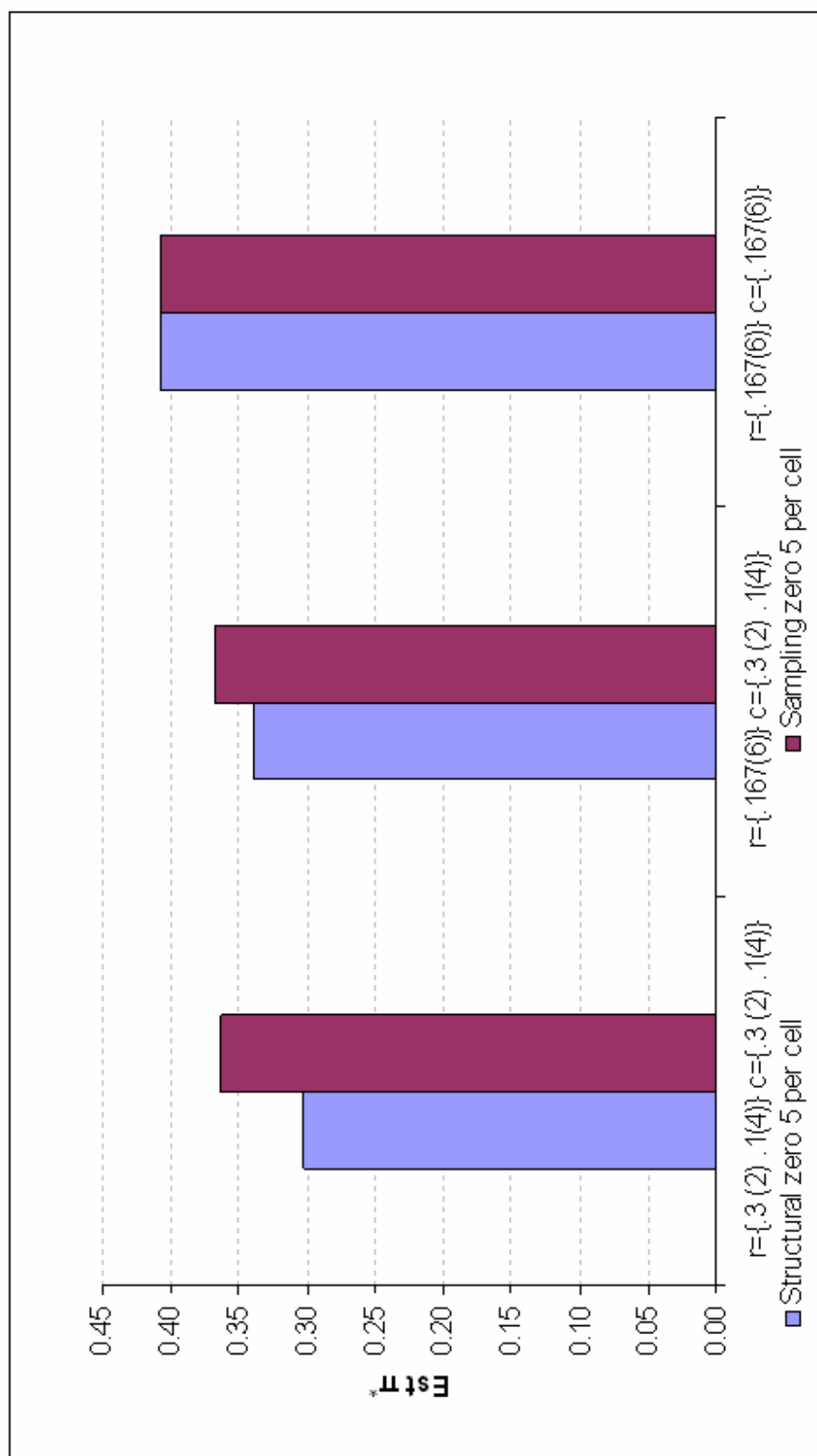


Figure 9. Sampling Zero and Structural Zero Effect on $\hat{\pi}^*$ Comparison in 6×6 table

Lower Bound Comparisons of RCL and “True” Estimates

The 95% lower bound of $\hat{\pi}^*$ using RCL method and empirical based “true” estimates from various simulations are listed in Appendix A where tables are appropriately prefixed along with the estimates under different sampling zero techniques denoted by the column headings.

As shown in Appendix A25 – A48, the 95% lower bound for $\hat{\pi}^*$ using RCL method is generally close to the “true” estimate based on empirical parametric simulation. However, under some circumstances (marked in red in Appendix A, most of them from symmetric frequency tables and any difference of “true” lower bound - RCL method lower bound $< .01$ is due to rounding error), RCL method underestimates the lower bound value even though the magnitude is relatively small and the difference from the “true estimate” decreases as the sample size increases.

Similar to parameter estimators for $\hat{\pi}^*$, the “true” estimates of 95% lower bound of $\hat{\pi}^*$ as shown in Figures 10 and 11, for 2×2 , 2×3 , 2×4 , 2×6 tables, when table size increases, the lower bound of $\hat{\pi}^*$ consistently increases within the same sample size (5, 10, 20 and 30 per cell) and marginal distribution (evenly, slightly and extremely dispersed) with exceptions for 2×3 and 2×4 extremely dispersed tables with sample size 5 (remain nearly unchanged). This conclusion also applies to 2×2 , 3×3 , 4×4 , 6×6 symmetric tables. Additionally, with few exceptions, for each frequency table, as sample size increases, the 95% lower bound decreases.

As displayed in Figures 12 and 13, for each size of contingency table, lower bound $\hat{\pi}^*$ is generally smallest for extremely dispersed row and column marginal distributions, followed by slightly dispersed row and column marginal distributions; while largest for

evenly distributed row and column tables. Different techniques for dealing with sampling zero seem to have no effect on the lower bound of $\hat{\pi}^*$ on either the method following RCL or the “true” lower bound based on empirical simulations.

Confidence Interval and Standard Errors

From Figures 14 and 15, we can see that given the same table size, extremely dispersed row and column marginal distributions consistently provide narrower confidence intervals ($\hat{\pi}^* - \hat{\pi}_L$) than evenly dispersed row and column tables using both RCL method and empirical based “true” estimates. Also, when sample size increases, confidence intervals become narrower for each table size and roughly shrinks to same confidence intervals for different marginal distribution with same table size using both estimation methods.

Standard errors of $\hat{\pi}^*$ in different combination of scenarios are also tabulated in Appendix A. It is apparent that since RCL method underestimates the lower bound of $\hat{\pi}^*$ in many cases and, thus, leads to a higher standard error compared with empirical based “true” lower bound estimates.

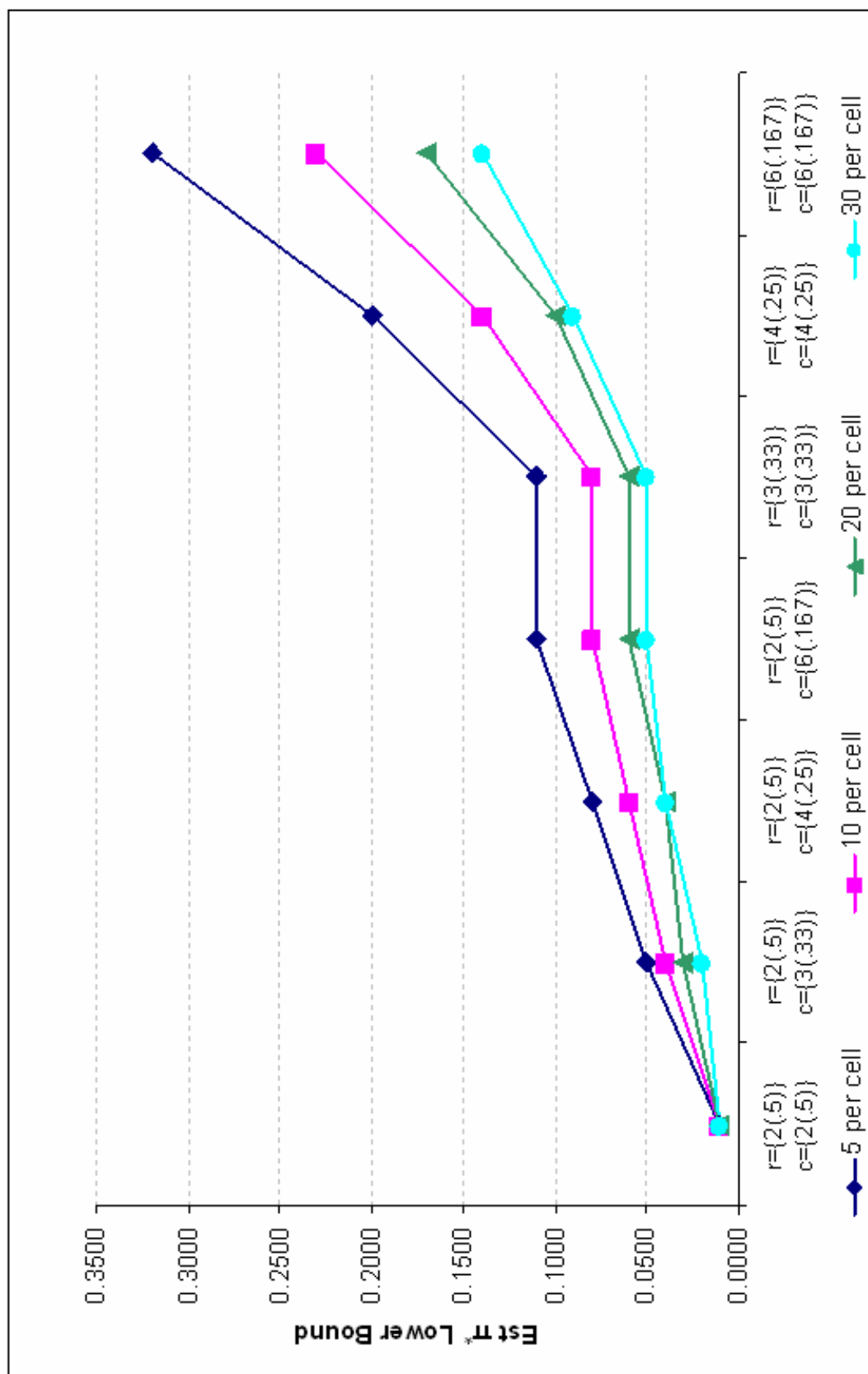


Figure 10. Empirical Simulation Based $\hat{\pi}_T$ With Evenly Distributed Marginals

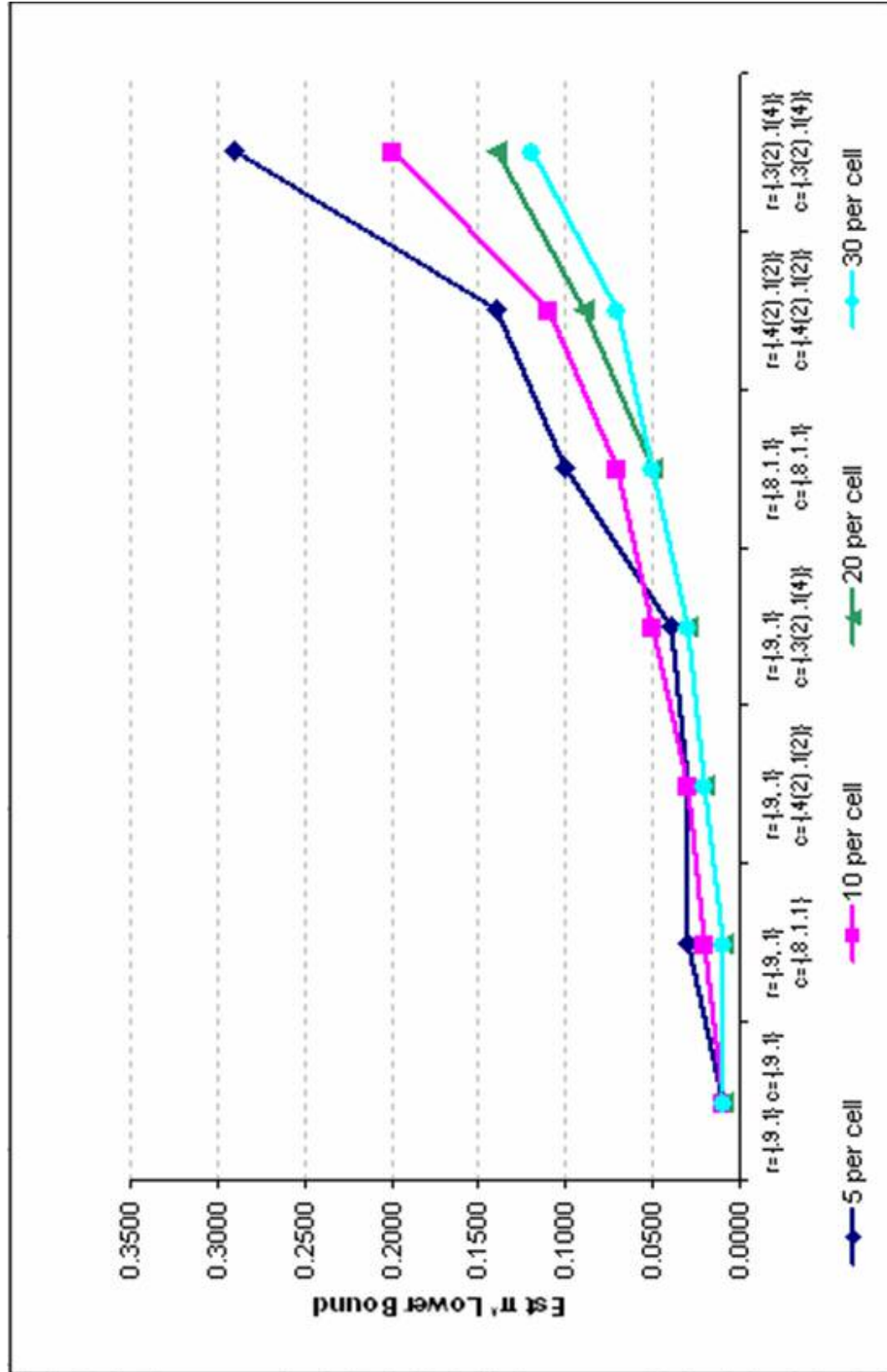


Figure 11. Empirical Simulation Based $\hat{\pi}_1$ With Extremely Distributed Marginals

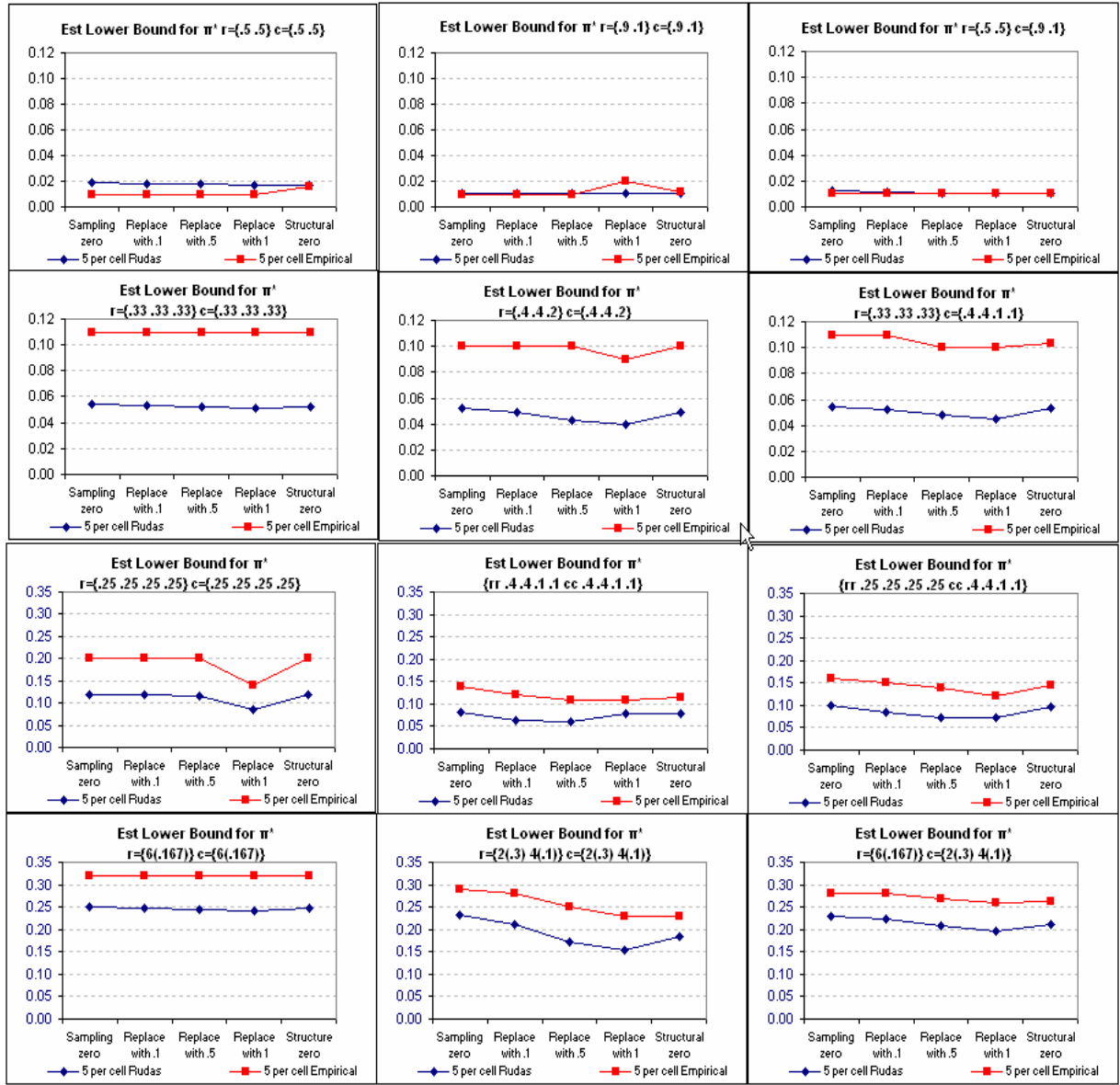


Figure 12. $\hat{\pi}_L$ Comparison between RCL method and Empirical Simulation method

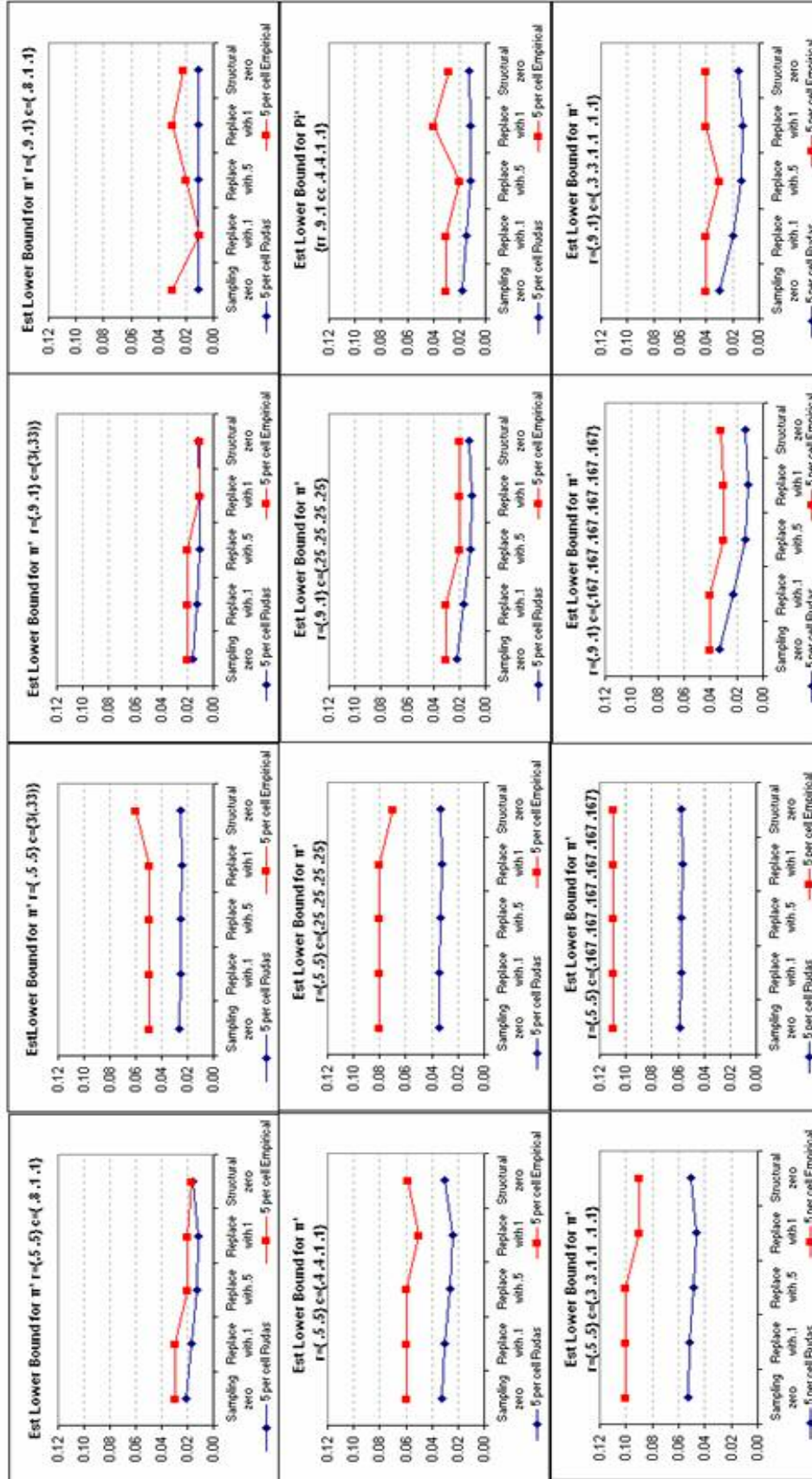


Figure 13. $\hat{\pi}_1$ Comparison between RCL method and Empirical Simulation method (Continued)

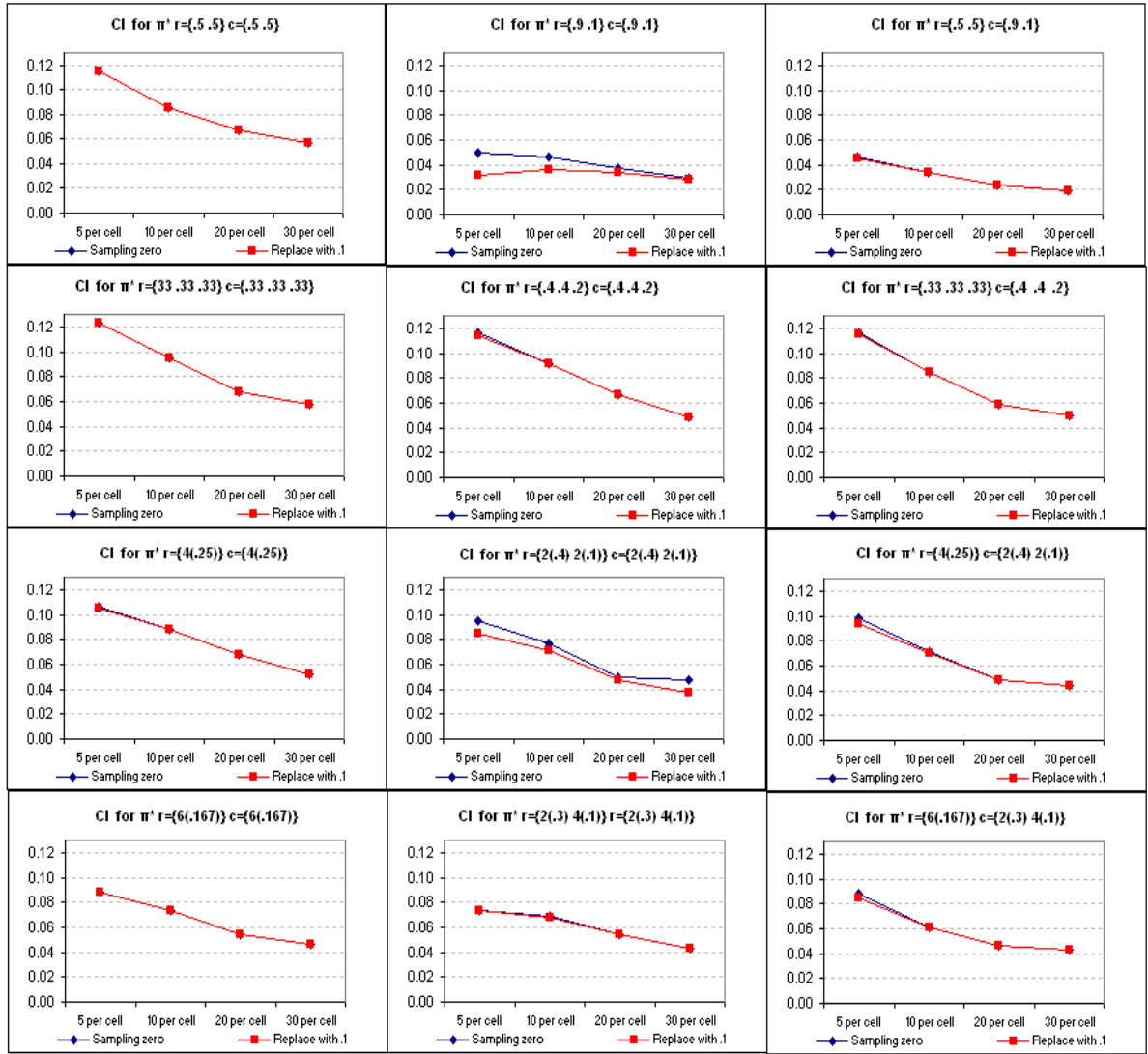


Figure 14. Confidence Interval of $\hat{\pi}^*$ following empirical simulation method

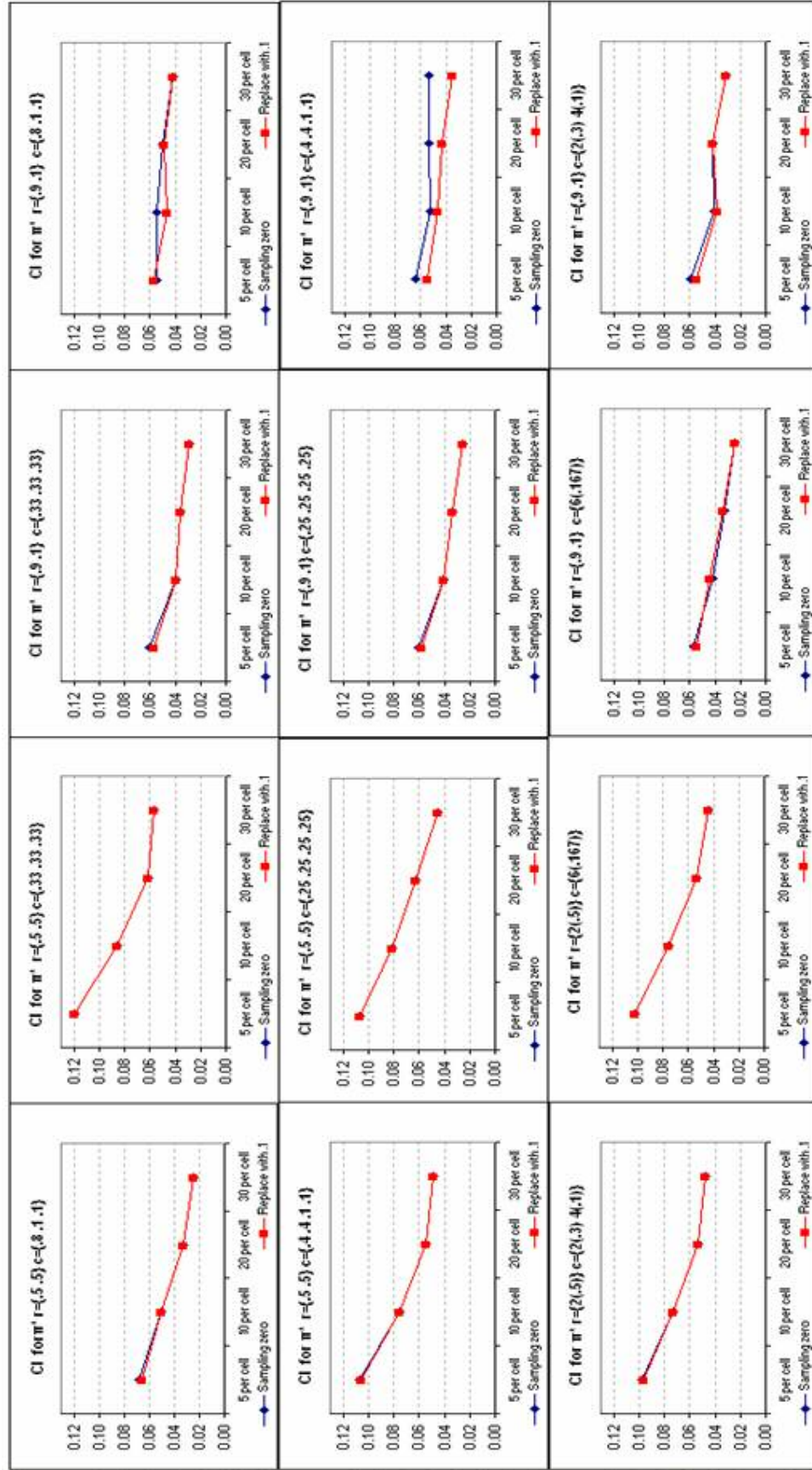


Figure 15. Confidence Interval of \hat{A}^* (Continued)

Examples

The first example is a 6×3 cross-classification of fatal crashes by speed limit and land use with 37,295 cases based on a 2004 USDOT traffic database. This data table is used to compare the results using traditional chi-square and related model fit methods with the mixture index of fit introduced by RCL. More specially, we compare different sampling zero techniques impacts on $\hat{\pi}^*$ since there is a zero cell in the contingency table.

The value of the Pearson Chi-Square statistic is 7200.090, and the likelihood ratio, G^2 statistic is 7600.540 both with degrees of freedom equal to 10 ($P < .01$). Thus, an independence model is not tenable based on these chi-squared tests of fit. As displayed in Table 4, the mixture index of fit $\hat{\pi}^*$ is .294, which means about 29.4% of the total of 37,295 cases (or, 10,965 cases) have to be removed in order to attain perfect model fit. The mixture index of fit provides an interpretation that is consistent with traditional Chi-Square analyses. Furthermore, $\hat{\pi}^*$ only decreases to .293 when replacing sample zero with the flattening constant 0.1 and further reduces to .291 when replacing with .5 and 1 as well as using the structural zero method. The amount of change in $\hat{\pi}^*$, as well as its 95% lower bound using different sampling zero techniques, is extremely small in this example. This occurs due to the very small percentage (0.62%) of Unknown land. In fact, it would not substantially effect $\hat{\pi}^*$, even if the whole column were zeros.

The second example is a cross-classification eye color and hair color in a 4×4 table with sample size of 592 presented by Snee (1974). RCL utilized this data example to study the properties of the mixture index of fit. In this paper, this example is used to compare the difference between sampling zero and structural zero. The 16 cells were set

to zero one-by-one and the results are presented in Table 5. The percentage differences between use of the sampling zero and structural zero techniques range from 5.00% to 40.00%, (note that six of these differences are statistically significant ($p < 0.05$) using conventional z tests for proportions.) The largest percentage reduce in $\hat{\pi}^*$ using structural zero occurs when Hair black and eye color Hazel is set to zero.

Table 4. Fit Statistics for Fatal Crashes by Speed Limit and Land Use

π	Chi-Square	G-Square	π Repl with .1	Chi-Square	G-Square	π Repl with .5	Chi-Square	G-Square	π Repl with 1	Chi-Square	G-Square
0	7200.090	7600.540	0.000	7287.906	7599.937	0.000	7287.426	7599.116	0.000	7287.426	7599.116
0.1	2947.100	3044.340	0.100	2946.904	3043.700	0.100	2946.353	3042.763	0.100	2946.353	3042.763
0.15	1527.210	1578.220	0.150	1527.010	1577.582	0.150	1526.444	1576.641	0.150	1526.444	1576.641
0.2	632.920	655.610	0.200	632.726	654.986	0.200	632.173	654.094	0.200	632.174	654.094
0.25	140.530	142.830	0.250	140.330	142.266	0.250	139.875	141.612	0.250	139.875	141.612
0.26	81.050	82.280	0.260	80.856	81.737	0.260	80.449	81.149	0.260	80.449	81.149
0.27	37.190	37.890	0.270	37.003	37.364	0.270	36.643	36.841	0.270	36.643	36.841
0.28	10.010	10.510	0.280	9.824	9.992	0.280	9.060	9.534	0.280	9.510	9.534
<i>0.286</i>	<i>2.100</i>	<i>2.560</i>	<i>0.286</i>	<i>1.923</i>	<i>2.058</i>	<i>0.285</i>	<i>2.498</i>	<i>2.501</i>	<i>0.285</i>	<i>2.498</i>	<i>2.501</i>
0.29	0.510	0.730	0.290	0.272	0.321	0.290	0.004	0.004	0.290	0.004	0.004
0.294	0.000	0.000	0.293	0.000	0.000	0.291	0.000	0.000	0.291	0.000	0.000
$\pi \geq 0.294$	0.000	0.000	$\pi \geq 0.293$	0.000	0.000	$\pi \geq 0.291$	0.000	0.000	$\pi \geq 0.291$	0.000	0.000

NOTE: 1. Italic figures denote a 95% lower bound for $\hat{\pi}^*$ and associated Chi-Square and G-Square values.

NOTE: 2. Italic and bolded figures denote $\hat{\pi}^*$ and the associated Chi-Square and G-Square values.

Table 5. $\hat{\pi}^*$ Comparison of Sampling Zero and Structural Zero using Eye Color data (each cell manipulated to be zero in turn)

Eye Color	Hair color											
	Black			Brunette			Red			Blonde		
	Observed cell freq	Simpl. zero	Struc. zero	%	Observed cell freq	Simpl. zero	Struc. zero	%	Observed cell freq	Simpl. zero	Struc. zero	%
Brown	68	0.32	0.27	18.52%	119	0.42	0.40	5.00%	26	0.37	0.32	15.63%
Blue	20	0.37	0.31	19.35%	84	0.44	0.38	15.79%	17	0.38	0.31	22.58%
Hazel	15	0.42	0.30	40.00%	54	0.38	0.32	18.75%	14	0.38	0.31	22.58%
Green	5	0.35	0.29	20.69%	29	0.32	0.30	6.67%	14	0.34	0.30	13.33%
									16	0.34	0.29	17.24%

Note: 1. Bold figures denotes significant percentage difference ($p < .05$, conventional z test for proportions) between sampling zero and structural zero techniques.

CHAPTER V

CONCLUSIONS AND DISCUSSION

This study was primarily focused on three areas. The first, the performance of the estimate, $\hat{\pi}^*$, in various combinations of conditions, as a function of different sizes of tables, different marginal distributions and different sample sizes. The second area encompassed evaluating several techniques for dealing with sampling zeros when computing π^* in contingency tables when the independence assumption holds. The third was comparing the standard error of $\hat{\pi}^*$ and confidence interval estimated by use of a method proposed by RCL, with the “true” standard error based on empirical simulations in various scenarios especially when encountering small sample sizes and π^* close to zero.

Parameter Estimates and Bias

In general, when table size increases, $\hat{\pi}^*$ consistently increases for constant sample size and marginal distribution with few exceptions for both symmetric and non-symmetric tables. For each frequency table, as sample size increases, with few exceptions, $\hat{\pi}^*$ decreases as well as its lower bound. For each size of contingency table, $\hat{\pi}^*$ is smallest for extremely dispersed row and column marginal distributions while largest for evenly distributed marginal tables.

$\hat{\pi}^*$ is positively biased from zero by 2.98% to 40.86% in the designs studied (all table sizes, marginal distributions, and sample sizes) when the independence assumption holds. The magnitude of bias, however, varied across different situations. Large table sizes with small sample sizes were heavily biased, 40.86% of $\hat{\pi}^*$ biased from zero when table size is 6×6 with evenly dispersed row and column marginals and small

sample size. However, $\hat{\pi}^*$ is still a reasonable descriptive index for model goodness of fit of contingency tables when independence assumption holds for small table size (2×2 , 2×3 , 2×4 , 3×3 , 2×6) tables with moderately large (or above) sample size (the largest $\hat{\pi}^*$ in such cases was no greater than .1).

Sampling Zero Techniques

As expected, for all two-way contingency tables, replacing zero with a larger flattening values resulted in smaller $\hat{\pi}^*$ as well as its lower bound. For extremely and most slightly dispersed row and column marginal distributions tables with small sample sizes and small table sizes, the value of $\hat{\pi}^*$ is substantially smaller using structural zero compared to using other techniques. Note that the techniques of replacing zero cell with flattening constants include virtually any number of simulated zero cells for each table while structural zero technique using here was based on only one zero cell per table. Because the number of zero counts and patterns are somewhat different between these techniques especially when encountering small sample size, it might affect the comparison between structural zero and using sampling zero or any other replacing with small positive flattening constants techniques.

Different techniques for dealing with sampling zero seem to have no effect on the lower bound of $\hat{\pi}^*$ on either the method following RCL or the “true” lower bound based on empirical simulations.

Confidence Interval and Standard Error

As hypothesized, similar to parameter estimators for $\hat{\pi}^*$, when table size increases, the “true” estimates of 95% lower bound of $\hat{\pi}^*$ consistently increases within the same sample size and marginal distribution for both symmetric and non-symmetric tables

with few exceptions. Additionally, with few exceptions, for each frequency table, as sample size increases, the 95% lower bound decreases. Lower bound of $\hat{\pi}^*$ is generally smallest for extremely dispersed row and column marginal distributions, followed by slightly dispersed row and column marginal distributions; while largest for evenly distributed row and column tables. As a result, given the same table size, extremely dispersed row and column marginal distributions consistently provide narrower confidence interval than evenly dispersed row and column tables.

The 95% lower bound for $\hat{\pi}^*$ using the RCL method is generally close to the "true" estimate based on empirical parametric simulation. However, under many circumstances, the RCL method underestimates the lower bound value although the magnitude is relatively small and the difference from the "true estimate" decreases as the sample size increases.

Recommendations

Among all the sampling zero techniques that were compared in terms of parameter bias, replacing with larger flattening constants such as 1 and structural zero technique appear to perform better in the sense that $\hat{\pi}^*$ is consistently smaller. Between these two techniques, structural zero technique is generally recommended for extremely and slightly dispersed row and column marginal distributions tables with small sample sizes and small table sizes while in other cases replacing with larger flattening constant (i.e., 1) is preferred.

Based on the current findings, RCL standard error estimates were comparatively conservative. In general, it is preferable in practice to use variance estimates that tend to be conservative (i.e., large) rather than liberal (i.e., small). However, it would be

valuable to investigate the standard error of $\hat{\pi}^*$ using re-sampling methods to provide better guidance for users.

Implications for Future Research

1. Evenly distributed, slightly and extremely dispersed marginal distributions for each different size of tables were manipulated in the current study. It would be valuable to investigate more diversified marginal distribution in future studies.

2. As noted, the limitation of structural zero technique with number of zero cells might affect the results when compared with other sampling zero techniques. It would be of interest to investigate structural zero technique applied in two-point mixture index model in contingency tables with more than one zero when the independence assumption holds.

3. In order to attain reasonable execution times for the simulation, in this study, an increment of .01 was adopted to successively increase $\hat{\pi}^*$ when estimating $\hat{\pi}^*$ using an EM algorithm. For very small true values of π^* , it would be necessary to use a value of .001 or even .0001 in order to obtain a more detailed picture, especially for the lower bound of $\hat{\pi}^*$.

4. In future study, it would be beneficial to investigate the standard error of $\hat{\pi}^*$ using other re-sampling methods (e.g., jackknife) and compare with RCL to provide a more concrete guide.

5. The larger value of flattening constants (e.g., 1) might affect the original data structure when sample size of a contingency table is small (e.g., 5 per cell) and thus the results could be slightly influenced. Alternative ways to define the flattening constants such as a percentage to total sample size is of interest in future study.

6. Finally, it would be valuable to evaluate the performance of π^* under conditions where the independence assumption does not hold.

MARGINAL DISTRIBUTION LEGEND FOR APPENDIX A

Case 1: $\{P_{1+}= 5, P_{2+}= 5, P_{+1}= 8, P_{+2}= 1, P_{+3}= 1\}$

Case 2: $\{P_{1+}= 5, P_{2+}= 5, P_{+1}= 33, P_{+2}= 33, P_{+3}= 33\}$

Case 3: $\{P_{1+}= 9, P_{2+}= 1, P_{+1}= 33, P_{+2}= 33, P_{+3}= 33\}$

Case 4: $\{P_{1+}= 9, P_{2+}= 1, P_{+1}= 8, P_{+2}= 1, P_{+3}= 1\}$

Case 5: $\{P_{1+}= 5, P_{2+}= 5, P_{+1}= 25, P_{+2}= 25, P_{+3}= 25, P_{+4}= 25\}$

Case 6: $\{P_{1+}= 5, P_{2+}= 5, P_{+1}= 4, P_{+2}= 4, P_{+3}= 1, P_{+4}= 1\}$

Case 7: $\{P_{1+}= 9, P_{2+}= 1, P_{+1}= 25, P_{+2}= 25, P_{+3}= 25, P_{+4}= 25\}$

Case 8: $\{P_{1+}= 9, P_{2+}= 1, P_{+1}= 4, P_{+2}= 4, P_{+3}= 1, P_{+4}= 1\}$

Case 9: $\{P_{1+}= 5, P_{2+}= 5, P_{+1}= 167, P_{+2}= 167, P_{+3}= 167, P_{+4}= 167, P_{+5}= 167, P_{+6}= 167\}$

Case 10: $\{P_{1+}= 9, P_{2+}= 1, P_{+1}= 167, P_{+2}= 167, P_{+3}= 167, P_{+4}= 167, P_{+5}= 167, P_{+6}= 167\}$

Case 11: $\{P_{1+}= 5, P_{2+}= 5, P_{+1}= 3, P_{+2}= 3, P_{+3}= 1, P_{+4}= 1, P_{+5}= 1, P_{+6}= 1\}$

Case 12: $\{P_{1+}= 9, P_{2+}= 1, P_{+1}= 3, P_{+2}= 3, P_{+3}= 1, P_{+4}= 1, P_{+5}= 1, P_{+6}= 1\}$

Case 13: $\{P_{1+}= 5, P_{2+}= 5, P_{+1}= 5, P_{+2}= 5\}$

Case 14: $\{P_{1+}= 9, P_{2+}= 1, P_{+1}= 9, P_{+2}= 1\}$

Case 15: $\{P_{1+}= 5, P_{2+}= 5, P_{+1}= 9, P_{+2}= 1\}$

Case 16: $\{P_{1+}= 4, P_{2+}= 4, P_{+1}= 4, P_{+2}= 4, P_{+3}= 2\}$

Case 17: $\{P_{1+}= 33, P_{2+}= 33, P_{+1}= 33, P_{+2}= 33, P_{+3}= 33\}$

Case 18: $\{P_{1+}= 33, P_{2+}= 33, P_{+1}= 4, P_{+2}= 4, P_{+3}= 2\}$

Case 19: $\{P_{1+}= 25, P_{2+}= 25, P_{+1}= 25, P_{+2}= 25, P_{+3}= 25\}$

$$P_{+4}=. 25\}$$

Case 20: $\{P_{1+}=. 4, P_{2+}=. 4, P_{3+}=. 1, P_{4+}=. 1, P_{+1}=. 4, P_{+2}=. 4, P_{+3}=. 1, P_{+4}=. 1\}$

Case 21: $\{P_{1+}=. 25, P_{2+}=. 25, P_{3+}=. 25, P_{4+}=. 25, P_{+1}=. 4, P_{+2}=. 4, P_{+3}=. 1,$

$$P_{+4}=. 1\}$$

Case 22: $\{P_{1+}=. 167, P_{2+}=. 167, P_{3+}=. 167, P_{4+}=. 167, P_{5+}=. 167,$

$$P_{6+}=. 167, P_{+1}=. 3, P_{+2}=. 3, P_{+3}=. 1, P_{+4}=. 1, P_{+5}=. 1, P_{+6}=. 1\}.$$

Case 23: $\{P_{1+}=. 167, P_{2+}=. 167, P_{3+}=. 167, P_{4+}=. 167, P_{5+}=. 167, P_{6+}=. 167,$

$$P_{+1}=. 167, P_{+2}=. 167, P_{+3}=. 167, P_{+4}=. 167, P_{+5}=. 167, P_{+6}=. 167\}$$

Case 24: $\{P_{1+}=. 3, P_{2+}=. 3, P_{3+}=. 1, P_{4+}=. 1, P_{5+}=. 1, P_{6+}=. 1,$

$$P_{+1}=. 3, P_{+2}=. 3, P_{+3}=. 1, P_{+4}=. 1, P_{+5}=. 1, P_{+6}=. 1\},$$

APPENDIX A: RESULTS TABLES

Table A1: $\hat{\pi}^$ and $\hat{\pi}^*$ Using Sampling Zero Techniques – Case 1*

{P1+=.5, P2+=.5, P+1=.8, P+2=.1, P+3=.1}	5 per cell		10 per cell		20 per cell		30 per cell	
	Pi Star	No of Zeros	Pi Star	No of Zeros	Pi Star	No of Zeros	Pi Star	No of Zeros
Sampling zero	0.0974	598	0.0700	136	0.0533	4	0.0443	0
Replace with .1	0.0951	598	0.0698	136	0.0533	4	0.0443	0
Replace with .5	0.0851	598	0.0688	136	0.0533	4	0.0443	0
Replace with 1	0.0757	598	0.0677	136	0.0533	4	0.0443	0
Structural zero*	0.0658	522	0.0645	130	0.0542	4	0.0451	0

NOTE: 1. * denotes accommodate only one zero cell each simulated frequency table for structural zero technique.

Table A2: $\hat{\pi}^$ and $\hat{\pi}^*$ Using Sampling Zero Techniques – Case 2*

{P1+=.5, P2+=.5, P+1=.33, P+2=.33, P+3=.33}	5 per cell		10 per cell		20 per cell		30 per cell	
	Pi Star	No of Zeros	Pi Star	No of Zeros	Pi Star	No of Zeros	Pi Star	No of Zeros
Sampling zero	0.1698	28	0.1256	0	0.0921	0	0.0766	0
Replace with .1	0.1697	28	0.1256	0	0.0921	0	0.0766	0
Replace with .5	0.1691	28	0.1256	0	0.0921	0	0.0766	0
Replace with 1	0.1683	28	0.1256	0	0.0921	0	0.0766	0
Structural zero*	0.1700	28	0.1294	0	0.0946	0	0.0786	0

NOTE: 1. * denotes accommodate only one zero cell each simulated frequency table for structural zero technique.

Table A3: $\hat{\pi}^*$ and $\hat{\pi}^*$ Using Sampling Zero Techniques – Case 3

{P1+=.9, P2+=.1, P+1=.33, P+2=.33, P+3=.33}	5 per cell		10 per cell		20 per cell		30 per cell	
	Pi Star	No of Zeros	Pi Star	No of Zeros	Pi Star	No of Zeros	Pi Star	No of Zeros
Sampling zero	0.0800	746	0.0600	311	0.0462	36	0.0390	8
Replace with .1	0.0771	746	0.0594	311	0.0461	36	0.0390	8
Replace with .5	0.0605	746	0.0552	311	0.0459	36	0.0389	8
Replace with 1	0.0502	746	0.0512	311	0.0456	36	0.0389	8
Structural zero*	0.0481	656	0.0481	286	0.0455	36	0.0393	8

NOTE: 1. * denotes accommodate only one zero cell each simulated frequency table for structural zero technique.

Table A4: $\hat{\pi}^*$ and $\hat{\pi}^*$ Using Sampling Zero Techniques – Case 4

{P1+=.9, P2+=.1, P+1=.8, P+2=.1, P+3=.1}	5 per cell		10 per cell		20 per cell		30 per cell	
	Pi Star	No of Zeros	Pi Star	No of Zeros	Pi Star	No of Zeros	Pi Star	No of Zeros
Sampling zero	0.0844 [#]	960	0.0745 [#]	804	0.0604 [#]	507	0.0520 [#]	308
Replace with .1	0.0673	960	0.0660	804	0.0590	507	0.0520	308
Replace with .5	0.0416	960	0.0375	804	0.0441	507	0.0454	308
Replace with 1	0.0517	960	0.0295	804	0.0328	507	0.0391	308
Structural zero*	0.0440	849	0.0321	653	0.0330	424	0.0388	257

NOTE: 1. * denotes accommodate only one zero cell each simulated frequency table for structural zero technique.

NOTE: 2. # denotes convergence criteria G-Square<.01 instead of .00001.

Table A5: $\hat{\pi}^*$ and $\hat{\pi}^*$ Using Sampling Zero Techniques – Case 5

{P1+=.5, P2+=.5, P+1=.25, P+2=.25, P+3=.25, P+4=.25}	5 per cell		10 per cell		20 per cell		30 per cell	
	$\hat{\pi}^*$	No of Zeros	$\hat{\pi}^*$	No of Zeros	$\hat{\pi}^*$	No of Zeros	$\hat{\pi}^*$	No of Zeros
Sampling zero	0.1873	44	0.1414	1	0.1031	0	0.0858	0
Replace with .1	0.1871	44	0.1414	1	0.1031	0	0.0858	0
Replace with .5	0.1865	44	0.1414	1	0.1031	0	0.0858	0
Replace with 1	0.1857	44	0.1412	1	0.1031	0	0.0858	0
Structural zero*	0.1869	44	0.1450	1	0.1060	0	0.0879	0

NOTE: 1. * denotes accommodate only one zero cell each simulated frequency table for structural zero technique.

Table A6: $\hat{\pi}^*$ and $\hat{\pi}^*$ Using Sampling Zero Techniques – Case 6

{P1+=.5, P2+=.5, P+1=.4, P+2=.4, P+3=.1, P+4=.1}	5 per cell		10 per cell		20 per cell		30 per cell	
	$\hat{\pi}^*$	No of Zeros	$\hat{\pi}^*$	No of Zeros	$\hat{\pi}^*$	No of Zeros	$\hat{\pi}^*$	No of Zeros
Sampling zero	0.1673	406	0.1255	49	0.0940	0	0.0782	0
Replace with .1	0.1654	406	0.1255	49	0.0940	0	0.0782	0
Replace with .5	0.1594	406	0.1251	49	0.0940	0	0.0782	0
Replace with 1	0.1543	406	0.1248	49	0.0940	0	0.0782	0
Structural zero*	0.1505	352	0.1272	48	0.0974	0	0.0811	0

NOTE: 1. * denotes accommodate only one zero cell each simulated frequency table for structural zero technique.

Table A7: $\hat{\pi}^*$ and $\hat{\pi}^*$ Using Sampling Zero Techniques – Case 7

	5 per cell		10 per cell		20 per cell		30 per cell	
	$\hat{\pi}^*$	No of Zeros	$\hat{\pi}^*$	No of Zeros	$\hat{\pi}^*$	No of Zeros	$\hat{\pi}^*$	No of Zeros
Sampling zero	0.0908	851	0.0711	441	0.0541	72	0.0460	11
Replace with .1	0.0877	851	0.0705	441	0.0540	72	0.0460	11
Replace with .5	0.0697	851	0.0635	441	0.0534	72	0.0459	11
Replace with 1	0.0582	851	0.0577	441	0.0528	72	0.0458	11
Structural zero*	0.0615	702	0.0570	383	0.0523	71	0.0465	11

NOTE: 1. * denotes accommodate only one zero cell each simulated frequency table for structural zero technique.

Table A8: $\hat{\pi}^*$ and $\hat{\pi}^*$ Using Sampling Zero Techniques – Case 8

	5 per cell		10 per cell		20 per cell		30 per cell	
	$\hat{\pi}^*$	No of Zeros	$\hat{\pi}^*$	No of Zeros	$\hat{\pi}^*$	No of Zeros	$\hat{\pi}^*$	No of Zeros
Sampling zero	0.0942	942	0.0822	722	0.0732	370	0.0730	177
Replace with .1	0.0849	942	0.0762	722	0.0629	370	0.0544	177
Replace with .5	0.0660	942	0.0567	722	0.0545	370	0.0517	177
Replace with 1	0.0703	942	0.0510	722	0.0480	370	0.0487	177
Structural zero*	0.0675	800	0.0530	560	0.0488	281	0.0490	146

NOTE: 1. * denotes accommodate only one zero cell each simulated frequency table for structural zero technique.

Table A9: $\hat{\pi}^*$ and $\hat{\pi}^*$ Using Sampling Zero Techniques –Case 9

{P1+=.5, P2+=.5, P+1=.167, P+2=.167, P+3=.167, P+4=.167, P+5=.167, P+6=.167}								
	5 per cell		10 per cell		20 per cell		30 per cell	
	$\hat{\pi}^*$	No of Zeros	$\hat{\pi}^*$	No of Zeros	$\hat{\pi}^*$	No of Zeros	$\hat{\pi}^*$	No of Zeros
Sampling zero	0.2123	64	0.1552	0	0.1138	0	0.0937	0
Replace with .1	0.2122	64	0.1552	0	0.1138	0	0.0937	0
Replace with .5	0.2114	64	0.1552	0	0.1138	0	0.0937	0
Replace with 1	0.2106	64	0.1552	0	0.1138	0	0.0937	0
Structural zero*	0.2108	63	0.1582	0	0.1161	0	0.0957	0

NOTE: 1. * denotes accommodate only one zero cell each simulated frequency table for structural zero technique.

Table A10: $\hat{\pi}^*$ and $\hat{\pi}^*$ Using Sampling Zero Techniques – Case 10

{P1+=.9, P2+=.1, P+1=.167, P+2=.167, P+3=.167, P+4=.167, P+5=.167, P+6=.167}								
	5 per cell		10 per cell		20 per cell		30 per cell	
	$\hat{\pi}^*$	No of Zeros	$\hat{\pi}^*$	No of Zeros	$\hat{\pi}^*$	No of Zeros	$\hat{\pi}^*$	No of Zeros
Sampling zero	0.0968	938	0.0818	579	0.0628	91	0.0544	13
Replace with .1	0.0951	938	0.0842	579	0.0634	91	0.0544	13
Replace with .5	0.0749	938	0.0709	579	0.0618	91	0.0543	13
Replace with 1	0.0630	938	0.0630	579	0.0609	91	0.0542	13
Structural zero*	0.0744	790	0.0649	484	0.0617	87	0.0553	13

NOTE: 1. * denotes accommodate only one zero cell each simulated frequency table for structural zero technique.

Table A11: $\hat{\pi}^*$ and $\hat{\pi}^*$ Using Sampling Zero Techniques – Case 11

{P1+=.5, P2+=.5, P+1=.3, P+2=.3, P+3=.1, P+4=.1, P+5=.1, P+6=.1}	5 per cell		10 per cell		20 per cell		30 per cell	
	$\hat{\pi}^*$	No of Zeros	$\hat{\pi}^*$	No of Zeros	$\hat{\pi}^*$	No of Zeros	$\hat{\pi}^*$	No of Zeros
Sampling zero	0.1975	278	0.1432	11	0.1029	0	0.0871	0
Replace with .1	0.1968	278	0.1432	11	0.1029	0	0.0871	0
Replace with .5	0.1941	278	0.1432	11	0.1029	0	0.0871	0
Replace with 1	0.1911	278	0.1431	11	0.1029	0	0.0871	0
Structural zero*	0.1872	255	0.1462	11	0.1050	0	0.0890	0

NOTE: 1. * denotes accommodate only one zero cell each simulated frequency table for structural zero technique.

Table A12: $\hat{\pi}^*$ and $\hat{\pi}^*$ Using Sampling Zero Techniques – Case 12

{P1+=.9, P2+=.1, P+1=.3, P+2=.3, P+3=.1, P+4=.1, P+5=.1, P+6=.1}	5 per cell		10 per cell		20 per cell		30 per cell	
	$\hat{\pi}^*$	No of Zeros	$\hat{\pi}^*$	No of Zeros	$\hat{\pi}^*$	No of Zeros	$\hat{\pi}^*$	No of Zeros
Sampling zero	0.0995	980	0.0907	783	0.0712	328	0.0614	126
Replace with .1	0.0941	980	0.0888	783	0.0716	328	0.0614	126
Replace with .5	0.0732	980	0.0680	783	0.0652	328	0.0599	126
Replace with 1	0.0736	980	0.0588	783	0.0604	328	0.0584	126
Structural zero*	0.0809	837	0.0641	582	0.0628	259	0.0599	105

NOTE: 1. * denotes accommodate only one zero cell each simulated frequency table for structural zero technique.

Table A13: $\hat{\pi}^*$ and $\hat{\pi}^*$ Using Sampling Zero Techniques – Case 13

{P1+=.5, P2+=.5, P+1=.5, P+2=.5}	5 per cell		10 per cell		20 per cell		30 per cell	
	$\hat{\pi}^*$	No of Zeros	$\hat{\pi}^*$	No of Zeros	$\hat{\pi}^*$	No of Zeros	$\hat{\pi}^*$	No of Zeros
Sampling zero	0.1252	17	0.0960	0	0.0775	0	0.0668	0
Replace with .1	0.1252	17	0.0960	0	0.0775	0	0.0668	0
Replace with .5	0.1248	17	0.0960	0	0.0775	0	0.0668	0
Replace with 1	0.1243	17	0.0960	0	0.0775	0	0.0668	0
Structural zero*	0.1260	17	0.1015	0	0.0820	0	0.0712	0

NOTE: 1. * denotes accommodate only one zero cell each simulated frequency table for structural zero technique.

Table A14: $\hat{\pi}^*$ and $\hat{\pi}^*$ Using Sampling Zero Techniques – Case 14

{P1+=.9, P2+=.1, P+1=.9, P+2=.1}	5 per cell		10 per cell		20 per cell		30 per cell	
	$\hat{\pi}^*$	No of Zeros	$\hat{\pi}^*$	No of Zeros	$\hat{\pi}^*$	No of Zeros	$\hat{\pi}^*$	No of Zeros
Sampling zero	0.0598 [#]	877	0.0568 [#]	701	0.0476 [#]	489	0.0396 [#]	322
Replace with .1	0.0413	877	0.0463	701	0.0434	489	0.0383	322
Replace with .5	0.0285	877	0.0231	701	0.0272	489	0.0302	322
Replace with 1	0.0420	877	0.0230	701	0.0182	489	0.0223	322
Structural zero*	0.0262	695	0.0235	546	0.0185	356	0.0215	276

NOTE: 1. * denotes accommodate only one zero cell each simulated frequency table for structural zero technique.

NOTE: 2. # denotes convergence criteria G-Square<.01 instead of .00001.

Table A15: $\hat{\pi}^*$ and $\hat{\pi}^*$ Using Sampling Zero Techniques – Case 15

{P1+=.5, P2+=.5, P+1=.9, P+2=.1}	5 per cell		10 per cell		20 per cell		30 per cell	
	$\hat{\pi}^*$	No of Zeros	$\hat{\pi}^*$	No of Zeros	$\hat{\pi}^*$	No of Zeros	$\hat{\pi}^*$	No of Zeros
Sampling zero	0.0559	602	0.0445	239	0.0343	27	0.0298	3
Replace with .1	0.0555	602	0.0444	239	0.0343	27	0.0298	3
Replace with .5	0.0451	602	0.0418	239	0.0342	27	0.0298	3
Replace with 1	0.0385	602	0.0392	239	0.0340	27	0.0298	3
Structural zero*	0.0310	561	0.0351	228	0.0321	28	0.0296	2

NOTE: 1. * denotes accommodate only one zero cell each simulated frequency table for structural zero technique.

Table A16: $\hat{\pi}^*$ and $\hat{\pi}^*$ Using Sampling Zero Techniques – Case 16

{P1+=.4, P2+=.4, P3+=.2, P+1=.4, P+2=.4, P+3=.2}	5 per cell		10 per cell		20 per cell		30 per cell	
	$\hat{\pi}^*$	No of Zeros	$\hat{\pi}^*$	No of Zeros	$\hat{\pi}^*$	No of Zeros	$\hat{\pi}^*$	No of Zeros
Sampling zero	0.2162	242	0.1619	24	0.1163	0	0.0983	0
Replace with .1	0.2140	242	0.1618	24	0.1163	0	0.0983	0
Replace with .5	0.2075	242	0.1614	24	0.1163	0	0.0983	0
Replace with 1	0.2022	242	0.1608	24	0.1163	0	0.0983	0
Structural zero*	0.2015	230	0.1650	24	0.1204	0	0.1023	0

NOTE: 1. * denotes accommodate only one zero cell each simulated frequency table for structural zero technique.

Table A17: $\hat{\pi}^*$ and $\hat{\pi}^*$ Using Sampling Zero Techniques – Case 17

	5 per cell		10 per cell		20 per cell		30 per cell	
	$\hat{\pi}^*$	No of Zeros	$\hat{\pi}^*$	No of Zeros	$\hat{\pi}^*$	No of Zeros	$\hat{\pi}^*$	No of Zeros
Sampling zero	0.2338	50	0.1747	0	0.1277	0	0.1079	0
Replace with .1	0.2336	50	0.1747	0	0.1277	0	0.1079	0
Replace with .5	0.2323	50	0.1747	0	0.1277	0	0.1079	0
Replace with 1	0.2312	50	0.1747	0	0.1277	0	0.1079	0
Structural zero*	0.2347	49	0.1820	0	0.1353	0	0.1145	0

NOTE: 1. * denotes accommodate only one zero cell each simulated frequency table for structural zero technique.

Table A18: $\hat{\pi}^*$ and $\hat{\pi}^*$ Using Sampling Zero Techniques – Case 18

	5 per cell		10 per cell		20 per cell		30 per cell	
	$\hat{\pi}^*$	No of Zeros	$\hat{\pi}^*$	No of Zeros	$\hat{\pi}^*$	No of Zeros	$\hat{\pi}^*$	No of Zeros
Sampling zero	0.2267	165	0.1645	4	0.1187	0	0.0993	0
Replace with .1	0.2258	165	0.1645	4	0.1187	0	0.0993	0
Replace with .5	0.2217	165	0.1645	4	0.1187	0	0.0993	0
Replace with 1	0.2180	165	0.1644	4	0.1187	0	0.0993	0
Structural zero*	0.2190	165	0.1696	5	0.1228	0	0.1030	0

NOTE: 1. * denotes accommodate only one zero cell each simulated frequency table for structural zero technique.

Table A19: $\hat{\pi}^*$ and $\hat{\pi}^*$ Using Sampling Zero Techniques – Case 19

{P1+=.25, P2+=.25, P3+=.25, P4+=.25 P+1=.25, P+2=.25, P+3=.25, P+4=.25}	5 per cell		10 per cell		20 per cell		30 per cell	
	$\hat{\pi}^*$	No of Zeros	$\hat{\pi}^*$	No of Zeros	$\hat{\pi}^*$	No of Zeros	$\hat{\pi}^*$	No of Zeros
Sampling zero	0.3060	70	0.2280	0	0.1684	0	0.1422	0
Replace with .1	0.3046	70	0.2280	0	0.1684	0	0.1422	0
Replace with .5	0.3032	70	0.2280	0	0.1684	0	0.1422	0
Replace with 1	0.2280	70	0.1684	0	0.1422	0	0.1422	0
Structural zero*	0.3092	65	0.2374	0	0.1771	0	0.1493	0

NOTE: 1. * denotes accommodate only one zero cell each simulated frequency table for structural zero technique.

Table A20: $\hat{\pi}^*$ and $\hat{\pi}^*$ Using Sampling Zero Techniques – Case 20

{P1+=.4, P2+=.4, P3+=.1, P4+=.1 P+1=.4, P+2=.4, P+3=.1, P+4=.1}	5 per cell		10 per cell		20 per cell		30 per cell	
	$\hat{\pi}^*$	No of Zeros	$\hat{\pi}^*$	No of Zeros	$\hat{\pi}^*$	No of Zeros	$\hat{\pi}^*$	No of Zeros
Sampling zero	0.2350	937	0.1873	620	0.1402	161	0.1178	38
Replace with .1	0.2052	937	0.1715	620	0.1378	161	0.1174	38
Replace with .5	0.1996	937	0.1607	620	0.1354	161	0.1171	38
Replace with 1	0.1919	937	0.1408	620	0.1178	161	0.1171	38
Structural zero*	0.2027	779	0.1642	529	0.1377	154	0.1207	38

NOTE: 1. * denotes accommodate only one zero cell each simulated frequency table for structural zero technique.

Table A21: $\hat{\pi}^*$ and $\hat{\pi}^*$ Using Sampling Zero Techniques – Case 21

{P1+=.25, P2+=.25, P3+=.25, P4+=.25 P+1=.4, P+2=.4, P+3=.1, P+4=.1}	5 per cell		10 per cell		20 per cell		30 per cell	
	$\hat{\pi}^*$	No of Zeros	$\hat{\pi}^*$	No of Zeros	$\hat{\pi}^*$	No of Zeros	$\hat{\pi}^*$	No of Zeros
Sampling zero	0.2587	707	0.1909	140	0.1391	0	0.1145	0
Replace with .1	0.2442	707	0.1897	140	0.1391	0	0.1145	0
Replace with .5	0.2330	707	0.1882	140	0.1391	0	0.1145	0
Replace with 1	0.1912	707	0.1391	140	0.1145	0	0.1145	0
Structural zero*	0.2375	562	0.1888	136	0.1419	0	0.1166	0

NOTE: 1. * denotes accommodate only one zero cell each simulated frequency table for structural zero technique.

Table A22: $\hat{\pi}^*$ and $\hat{\pi}^*$ Using Sampling Zero Techniques –Case 22

{P1+=.167, P2+=.167, P3+=.167, P4+=.167, P5+=.167, P6+=.167. P+1=.3, P+2=.3, P+3=.1, P+4=.1. P+5=.1, P+6=.1}	5 per cell		10 per cell		20 per cell		30 per cell	
	$\hat{\pi}^*$	No of Zeros	$\hat{\pi}^*$	No of Zeros	$\hat{\pi}^*$	No of Zeros	$\hat{\pi}^*$	No of Zeros
Sampling zero	0.3680	709	0.2710	52	0.0940	0	0.0782	0
Replace with .1	0.3645	709	0.2710	52	0.0940	0	0.0782	0
Replace with .5	0.3512	709	0.2707	52	0.0940	0	0.0782	0
Replace with 1	0.3399	709	0.2702	52	0.0940	0	0.0782	0
Structural zero*	0.3404	573	0.2741	52	0.0974	0	0.0811	0

NOTE: 1. * denotes accommodate only one zero cell each simulated frequency table for structural zero technique.

Table A23: $\hat{\pi}^*$ and $\hat{\pi}^*$ Using Sampling Zero Techniques – Case 23

	5 per cell		10 per cell		20 per cell		30 per cell	
	$\hat{\pi}^*$	No of Zeros	$\hat{\pi}^*$	No of Zeros	$\hat{\pi}^*$	No of Zeros	$\hat{\pi}^*$	No of Zeros
Sampling zero	0.4086	183	0.3031	0	0.2242	0	0.1867	0
Replace with .1	0.4079	183	0.3031	0	0.2242	0	0.1867	0
Replace with .5	0.4054	183	0.3031	0	0.2242	0	0.1867	0
Replace with 1	0.4028	183	0.3031	0	0.2242	0	0.1867	0
Structural zero*	0.4088	173	0.3146	0	0.2337	0	0.1940	0

NOTE: 1. * denotes accommodate only one zero cell each simulated frequency table for structural zero technique.

Table A24: $\hat{\pi}^*$ and $\hat{\pi}^*$ Using Sampling Zero Techniques – Case 24

	5 per cell		10 per cell		20 per cell		30 per cell	
	$\hat{\pi}^*$	No of Zeros	$\hat{\pi}^*$	No of Zeros	$\hat{\pi}^*$	No of Zeros	$\hat{\pi}^*$	No of Zeros
Sampling zero	0.3629	953	0.2693	359	0.1942	15	0.1626	0
Replace with .1	0.3536	953	0.2681	359	0.1942	15	0.1626	0
Replace with .5	0.3175	953	0.2629	359	0.1941	15	0.1626	0
Replace with 1	0.2933	953	0.2579	359	0.1940	15	0.1626	0
Structural zero*	0.3020	752	0.2605	314	0.2001	15	0.1684	0

NOTE: 1. * denotes accommodate only one zero cell each simulated frequency table for structural zero technique.

Table A25: 95% Lower Bound- $\hat{\pi}_L$ - Case 1

{P1+=.5, P2+=.5, P+1=.8, P+2=.1, P+3=.1}	5 per cell		10 per cell		20 per cell		30 per cell	
	RCL	Empirical	RCL	Empirical	RCL	Empirical	RCL	Empirical
Sampling zero	0.021	0.0300	0.0154	0.0200	0.0136	0.0200	0.0124	0.0200
Replace with .1	0.017	0.0300	0.0149	0.0200	0.0136	0.0200	0.0124	0.0200
Replace with .5	0.013	0.0200	0.0139	0.0200	0.0135	0.0200	0.0124	0.0200
Replace with 1	0.011	0.0200	0.0132	0.0200	0.0135	0.0200	0.0124	0.0200
Structural zero	0.016	0.0173	0.0146	0.0200	0.0136	0.0200	0.0124	0.0200

NOTE: 1. Highlighted color denotes $\hat{\pi}_L$ difference between RCL method and empirical simulation method.

Table A26: 95% Lower Bound- $\hat{\pi}_L$ - Case 2

{P1+=.5, P2+=.5, P+1=.33, P+2=.33, P+3=.33}	5 per cell		10 per cell		20 per cell		30 per cell	
	RCL	Empirical	RCL	Empirical	RCL	Empirical	RCL	Empirical
Sampling zero	0.026	0.0500	0.020	0.0400	0.017	0.0300	0.0156	0.0200
Replace with .1	0.026	0.0500	0.020	0.0400	0.017	0.0300	0.0156	0.0200
Replace with .5	0.025	0.0500	0.020	0.0400	0.017	0.0300	0.0156	0.0200
Replace with 1	0.024	0.0500	0.020	0.0400	0.017	0.0300	0.0156	0.0200
Structural zero	0.025	0.0600	0.020	0.0400	0.017	0.0300	0.0156	0.0200

NOTE: 1. Highlighted color denotes $\hat{\pi}_L$ difference between RCL method and empirical simulation method.

Table A27: 95% Lower Bound- $\hat{\pi}_L$ - Case 3

	5 per cell		10 per cell		20 per cell		30 per cell	
	RCL	Empirical	RCL	Empirical	RCL	Empirical	RCL	Empirical
{P1+=.9, P2+=.1, P+1=.33, P+2=.33, P+3=.33}								
Sampling zero	0.0153	0.0200	0.0155	0.0200	0.0129	0.0100	0.0120	0.0100
Replace with .1	0.0127	0.0200	0.0139	0.0200	0.0127	0.0100	0.0120	0.0100
Replace with .5	0.0106	0.0200	0.0117	0.0200	0.0123	0.0100	0.0119	0.0100
Replace with 1	0.0101	0.0100	0.0110	0.0200	0.0120	0.0100	0.0119	0.0100
Structural zero	0.0111	0.0100	0.0115	0.0131	0.0119	0.0165	0.0118	0.0100

NOTE: 1. Highlighted color denotes $\hat{\pi}_L$ difference between RCL method and empirical simulation method.

Table A28: 95% Lower Bound- $\hat{\pi}_L$ - Case 4

	5 per cell		10 per cell		20 per cell		30 per cell	
	RCL	Empirical	RCL	Empirical	RCL	Empirical	RCL	Empirical
{P1+=.9, P2+=.1, P+1=.8, P+2=.1, P+3=.1}								
Sampling zero	0.0110	0.0300	0.0121	0.0200	0.0141	0.0100	0.0148	0.0100
Replace with .1	0.0105	0.0100	0.0107	0.0200	0.0127	0.0100	0.0108	0.0100
Replace with .5	0.0105	0.0200	0.0102	0.0100	0.0102	0.0100	0.0101	0.0100
Replace with 1	0.0107	0.0300	0.0103	0.0100	0.0100	0.0100	0.0101	0.0100
Structural zero	0.0111	0.0217	0.0107	0.0125	0.0110	0.0100	0.0114	0.0100

NOTE: 1. Highlighted color denotes $\hat{\pi}_L$ difference between RCL method and empirical simulation method.

Table A29: 95% Lower Bound- $\hat{\pi}_L$ - Case 5

{P1+=.5, P2+=.5, P+1=.25, P+2=.25, P+3=.25, P+4=.25}	5 per cell		10 per cell		20 per cell		30 per cell	
	RCL	Empirical	RCL	Empirical	RCL	Empirical	RCL	Empirical
Sampling zero	0.0348	0.0800	0.0277	0.0600	0.0218	0.0400	0.0189	0.0400
Replace with .1	0.0345	0.0800	0.0275	0.0600	0.0218	0.0400	0.0189	0.0400
Replace with .5	0.0336	0.0800	0.0276	0.0600	0.0218	0.0400	0.0189	0.0400
Replace with 1	0.0328	0.0800	0.0276	0.0600	0.0218	0.0400	0.0189	0.0400
Structural zero	0.0335	0.0700	0.0276	0.0600	0.0218	0.0400	0.0189	0.0400

NOTE: 1. Highlighted color denotes $\hat{\pi}_L$ difference between RCL method and empirical simulation method.

Table A30: 95% Lower Bound- $\hat{\pi}_L$ - Case 6

{P1+=.5, P2+=.5, P+1=.4, P+2=.4, P+3=.1, P+4=.1}	5 per cell		10 per cell		20 per cell		30 per cell	
	RCL	Empirical	RCL	Empirical	RCL	Empirical	RCL	Empirical
Sampling zero	0.0326	0.0600	0.0246	0.0500	0.0201	0.0400	0.0177	0.0300
Replace with .1	0.0303	0.0600	0.0245	0.0500	0.0201	0.0400	0.0177	0.0300
Replace with .5	0.0260	0.0600	0.0241	0.0500	0.0201	0.0400	0.0177	0.0300
Replace with 1	0.0236	0.0500	0.0238	0.0500	0.0201	0.0400	0.0177	0.0300
Structural zero	0.0307	0.0580	0.0245	0.0500	0.0201	0.0400	0.0177	0.0300

NOTE: 1. Highlighted color denotes $\hat{\pi}_L$ difference between RCL method and empirical simulation method.

Table A31: 95% Lower Bound- $\hat{\pi}_L$ - Case 7

{P1+=.9, P2+=.1, P+1=.25, P+2=.25, P+3=.25, P+4=.25}	5 per cell		10 per cell		20 per cell		30 per cell	
	RCL	Empirical	RCL	Empirical	RCL	Empirical	RCL	Empirical
Sampling zero	0.0223	0.0300	0.0195	0.0300	0.0157	0.0200	0.0139	0.0200
Replace with .1	0.0165	0.0300	0.0167	0.0300	0.0153	0.0200	0.0139	0.0200
Replace with .5	0.0117	0.0200	0.0129	0.0300	0.0143	0.0200	0.0138	0.0200
Replace with 1	0.0107	0.0200	0.0117	0.0200	0.0137	0.0200	0.0136	0.0200
Structural zero	0.0126	0.0200	0.0122	0.0200	0.0134	0.0200	0.0133	0.0200

NOTE: 1. Highlighted color denotes $\hat{\pi}_L$ difference between RCL method and empirical simulation method.

Table A32: 95% Lower Bound- $\hat{\pi}_L$ - Case 8

{P1+=.9, P2+=.1, P+1=.4, P+2=.4, P+3=.1, P+4=.1}	5 per cell		10 per cell		20 per cell		30 per cell	
	RCL	Empirical	RCL	Empirical	RCL	Empirical	RCL	Empirical
Sampling zero	0.0182	0.0300	0.0175	0.0300	0.0172	0.0200	0.0165	0.0200
Replace with .1	0.0147	0.0300	0.0144	0.0300	0.0146	0.0200	0.0147	0.0200
Replace with .5	0.0119	0.0200	0.0122	0.0200	0.0124	0.0200	0.0125	0.0200
Replace with 1	0.0114	0.0400	0.0118	0.0200	0.0120	0.0200	0.0118	0.0200
Structural zero	0.0123	0.0286	0.0125	0.0200	0.0130	0.0200	0.0122	0.0200

NOTE: 1. Highlighted color denotes $\hat{\pi}_L$ difference between RCL method and empirical simulation method.

Table A33: 95% Lower Bound- $\hat{\pi}_L$ - Case 9

{P1+=.5, P2+=.5, P+1=.167, P+2=.167, P+3=.167, P+4=.167, P+5=.167, P+6=.167}	5 per cell		10 per cell		20 per cell		30 per cell	
	RCL	Empirical	RCL	Empirical	RCL	Empirical	RCL	Empirical
Sampling zero	0.0580	0.1100	0.0416	0.0800	0.0325	0.0600	0.0275	0.0500
Replace with .1	0.0578	0.1100	0.0416	0.0800	0.0325	0.0600	0.0275	0.0500
Replace with .5	0.0570	0.1100	0.0416	0.0800	0.0325	0.0600	0.0275	0.0500
Replace with 1	0.0562	0.1100	0.0416	0.0800	0.0325	0.0600	0.0275	0.0500
Structural zero	0.0578	0.1100	0.0416	0.0800	0.0325	0.0600	0.0275	0.0500

NOTE: 1. Highlighted color denotes $\hat{\pi}_L$ difference between RCL method and empirical simulation method.

Table A34: 95% Lower Bound- $\hat{\pi}_L$ - Case 10

{P1+=.9, P2+=.1, P+1=.167, P+2=.167, P+3=.167, P+4=.167, P+5=.167, P+6=.167}	5 per cell		10 per cell		20 per cell		30 per cell	
	RCL	Empirical	RCL	Empirical	RCL	Empirical	RCL	Empirical
Sampling zero	0.0330	0.0400	0.0287	0.0400	0.0201	0.0300	0.0176	0.0300
Replace with .1	0.0230	0.0400	0.0235	0.0400	0.0195	0.0300	0.0176	0.0300
Replace with .5	0.0132	0.0300	0.0165	0.0400	0.0182	0.0300	0.0174	0.0300
Replace with 1	0.0114	0.0300	0.0138	0.0300	0.0174	0.0300	0.0173	0.0300
Structural zero	0.0136	0.0326	0.0146	0.0300	0.0167	0.0300	0.0169	0.0300

NOTE: 1. Highlighted color denotes $\hat{\pi}_L$ difference between RCL method and empirical simulation method.

Table A35: 95% Lower Bound- $\hat{\pi}_L$ - Case 11

{P1+=.5, P2+=.5, P+1=.3, P+2=.3, P+3=.1, P+4=.1, P+5=.1, P+6=.1}	5 per cell		10 per cell		20 per cell		30 per cell	
	RCL	Empirical	RCL	Empirical	RCL	Empirical	RCL	Empirical
Sampling zero	0.0530	0.1000	0.0379	0.0700	0.0284	0.0500	0.0253	0.0400
Replace with .1	0.0519	0.1000	0.0379	0.0700	0.0284	0.0500	0.0253	0.0400
Replace with .5	0.0490	0.1000	0.0378	0.0700	0.0284	0.0500	0.0253	0.0400
Replace with 1	0.0462	0.0900	0.0378	0.0700	0.0284	0.0500	0.0253	0.0400
Structural zero	0.0508	0.0900	0.0379	0.0700	0.0284	0.0500	0.0253	0.0400

NOTE: 1. Highlighted color denotes $\hat{\pi}_L$ difference between RCL method and empirical simulation method.

Table A36: 95% Lower Bound- $\hat{\pi}_L$ - Case 12

{P1+=.9, P2+=.1, P+1=.3, P+2=.3, P+3=.1, P+4=.1, P+5=.1, P+6=.1}	5 per cell		10 per cell		20 per cell		30 per cell	
	RCL	Empirical	RCL	Empirical	RCL	Empirical	RCL	Empirical
Sampling zero	0.0299	0.0400	0.0286	0.0500	0.0237	0.0300	0.0202	0.0300
Replace with .1	0.0199	0.0400	0.0216	0.0500	0.0207	0.0300	0.0192	0.0300
Replace with .5	0.0130	0.0300	0.0143	0.0300	0.0162	0.0300	0.0171	0.0300
Replace with 1	0.0121	0.0400	0.0128	0.0300	0.0143	0.0300	0.0158	0.0300
Structural zero	0.0152	0.0400	0.0141	0.0300	0.0144	0.0300	0.0150	0.0300

NOTE: 1. Highlighted color denotes $\hat{\pi}_L$ difference between RCL method and empirical simulation method.

Table A37: 95% Lower Bound- $\hat{\pi}_L$ - Case 13

{P1+=.5, P2+=.5, P+1=.5, P+2=.5}	5 per cell		10 per cell		20 per cell		30 per cell	
	RCL	Empirical	RCL	Empirical	RCL	Empirical	RCL	Empirical
Sampling zero	0.0186	0.0100	0.0152	0.0100	0.0138	0.0100	0.0131	0.0100
Replace with .1	0.0183	0.0100	0.0152	0.0100	0.0138	0.0100	0.0131	0.0100
Replace with .5	0.0177	0.0100	0.0152	0.0100	0.0138	0.0100	0.0131	0.0100
Replace with 1	0.0172	0.0100	0.0152	0.0100	0.0138	0.0100	0.0131	0.0100
Structural zero	0.0167	0.0160	0.0152	0.0100	0.0138	0.0100	0.0131	0.0100

NOTE: 1. Highlighted color denotes $\hat{\pi}_L$ difference between RCL method and empirical simulation method.

Table A38: 95% Lower Bound- $\hat{\pi}_L$ - Case 14

{P1+=.9, P2+=.1, P+1=.9, P+2=.1}	5 per cell		10 per cell		20 per cell		30 per cell	
	RCL	Empirical	RCL	Empirical	RCL	Empirical	RCL	Empirical
Sampling zero	0.0102	0.0100	0.0102	0.0100	0.0105	0.0100	0.0113	0.0100
Replace with .1	0.0102	0.0100	0.0101	0.0100	0.0102	0.0100	0.0106	0.0100
Replace with .5	0.0101	0.0100	0.0101	0.0100	0.0100	0.0100	0.0101	0.0100
Replace with 1	0.0101	0.0200	0.0101	0.0100	0.0100	0.0100	0.0100	0.0100
Structural zero	0.0102	0.0115	0.0100	0.0100	0.0101	0.0100	0.0103	0.0100

NOTE: 1. Highlighted color denotes $\hat{\pi}_L$ difference between RCL method and empirical simulation method.

Table A39: 95% Lower Bound- $\hat{\pi}_L$ - Case 15

{P1+=.5, P2+=.5, P+1=.9, P+2=.1}	5 per cell		10 per cell		20 per cell		30 per cell	
	RCL	Empirical	RCL	Empirical	RCL	Empirical	RCL	Empirical
Sampling zero	0.0126	0.0100	0.0124	0.0100	0.0114	0.0100	0.0111	0.0100
Replace with .1	0.0114	0.0100	0.0117	0.0100	0.0114	0.0100	0.0111	0.0100
Replace with .5	0.0102	0.0100	0.0109	0.0100	0.0112	0.0100	0.0111	0.0100
Replace with 1	0.0100	0.0100	0.0106	0.0100	0.0111	0.0100	0.0111	0.0100
Structural zero	0.0106	0.0100	0.0105	0.0100	0.0109	0.0100	0.0108	0.0100

NOTE: 1. Highlighted color denotes $\hat{\pi}_L$ difference between RCL method and empirical simulation method.

Table A40: 95% Lower Bound- $\hat{\pi}_L$ - Case 16

{P1+=.4, P2+=.4, P3+=.2, P+1=.4, P+2=.4, P+3=.2}	5 per cell		10 per cell		20 per cell		30 per cell	
	RCL	Empirical	RCL	Empirical	RCL	Empirical	RCL	Empirical
Sampling zero	0.0518	0.1000	0.0370	0.0700	0.0256	0.0500	0.0227	0.0500
Replace with .1	0.0485	0.1000	0.0367	0.0700	0.0256	0.0500	0.0227	0.0500
Replace with .5	0.0430	0.1000	0.0360	0.0700	0.0256	0.0500	0.0227	0.0500
Replace with 1	0.0400	0.0900	0.0355	0.0700	0.0256	0.0500	0.0227	0.0500
Structural zero	0.0494	0.1000	0.0367	0.0700	0.0256	0.0600	0.0227	0.0500

NOTE: 1. Highlighted color denotes $\hat{\pi}_L$ difference between RCL method and empirical simulation method.

Table A41: 95% Lower Bound- $\hat{\pi}_L$ - Case 17

{P1+=.33, P2+=.33, P3+=.33, P+1=.33, P+2=.33, P+3=.33}	5 per cell		10 per cell		20 per cell		30 per cell	
	RCL	Empirical	RCL	Empirical	RCL	Empirical	RCL	Empirical
Sampling zero	0.0543	0.1100	0.0385	0.0800	0.0269	0.0600	0.0242	0.0500
Replace with .1	0.0536	0.1100	0.0385	0.0800	0.0269	0.0600	0.0242	0.0500
Replace with .5	0.0521	0.1100	0.0385	0.0800	0.0269	0.0600	0.0242	0.0500
Replace with 1	0.0507	0.1100	0.0385	0.0800	0.0269	0.0600	0.0242	0.0500
Structural zero	0.0526	0.1100	0.0385	0.0800	0.0269	0.0600	0.0242	0.0500

NOTE: 1. Highlighted color denotes $\hat{\pi}_L$ difference between RCL method and empirical simulation method.

Table A42: 95% Lower Bound- $\hat{\pi}_L$ - Case 18

{P1+=.33, P2+=.33, P3+=.33, P+1=.4, P+2=.4, P+3=.2}	5 per cell		10 per cell		20 per cell		30 per cell	
	RCL	Empirical	RCL	Empirical	RCL	Empirical	RCL	Empirical
Sampling zero	0.0547	0.1100	0.0373	0.0800	0.0270	0.0600	0.0229	0.0500
Replace with .1	0.0525	0.1100	0.0372	0.0800	0.0270	0.0600	0.0229	0.0500
Replace with .5	0.0479	0.1000	0.0372	0.0800	0.0270	0.0600	0.0229	0.0500
Replace with 1	0.0451	0.1000	0.0371	0.0800	0.0270	0.0600	0.0229	0.0500
Structural zero	0.0535	0.1031	0.0372	0.0800	0.0270	0.0600	0.0229	0.0500

NOTE: 1. Highlighted color denotes $\hat{\pi}_L$ difference between RCL method and empirical simulation method.

Table A43: 95% Lower Bound- $\hat{\pi}_L$ - Case 19

{P1+=.25, P2+=.25, P3+=.25, P4+=.25 P+1=.25, P+2=.25, P+3=.25, P+4=.25}								
	5 per cell		10 per cell		20 per cell		30 per cell	
	RCL	Empirical	RCL	Empirical	RCL	Empirical	RCL	Empirical
Sampling zero	0.1192	0.2000	0.0862	0.1400	0.0626	0.1000	0.0529	0.0900
Replace with .1	0.1174	0.2000	0.0862	0.1400	0.0626	0.1000	0.0529	0.0900
Replace with .5	0.1158	0.2000	0.0862	0.1400	0.0626	0.1000	0.0529	0.0900
Replace with 1	0.0862	0.1400	0.0626	0.1000	0.0529	0.0900	0.0529	0.0900
Structural zero	0.1184	0.2000	0.0863	0.1500	0.0626	0.1100	0.0529	0.0900

NOTE: 1. Highlighted color denotes $\hat{\pi}_L$ difference between RCL method and empirical simulation method.

Table A44: 95% Lower Bound- $\hat{\pi}_L$ - Case 20

{P1+=.4, P2+=.4, P3+=.1, P4+=.1 P+1=.4, P+2=.4, P+3=.1, P+4=.1}								
	5 per cell		10 per cell		20 per cell		30 per cell	
	RCL	Empirical	RCL	Empirical	RCL	Empirical	RCL	Empirical
Sampling zero	0.0806	0.1400	0.0687	0.1100	0.0511	0.0900	0.0428	0.0700
Replace with .1	0.0622	0.1200	0.0552	0.1000	0.0478	0.0900	0.0423	0.0700
Replace with .5	0.0599	0.1100	0.0501	0.0900	0.0457	0.0900	0.0417	0.0700
Replace with 1	0.0774	0.1100	0.0525	0.0900	0.0430	0.0700	0.0417	0.0700
Structural zero	0.0776	0.1140	0.0690	0.0950	0.0516	0.0800	0.0426	0.0700

NOTE: 1. Highlighted color denotes $\hat{\pi}_L$ difference between RCL method and empirical simulation method.

Table A45: 95% Lower Bound- $\hat{\pi}_L$ - Case 21

{P1+=.25, P2+=.25, P3+=.25, P4+=.25 P+1=.4, P+2=.4, P+3=.1, P+4=.1}	5 per cell		10 per cell		20 per cell		30 per cell	
	RCL	Empirical	RCL	Empirical	RCL	Empirical	RCL	Empirical
Sampling zero	0.1005	0.1600	0.0729	0.1200	0.0522	0.0900	0.0430	0.0700
Replace with .1	0.0833	0.1500	0.0711	0.1200	0.0522	0.0900	0.0430	0.0700
Replace with .5	0.0729	0.1400	0.0693	0.1200	0.0522	0.0900	0.0430	0.0700
Replace with 1	0.0735	0.1200	0.0522	0.0900	0.0430	0.0700	0.0430	0.0700
Structural zero	0.0971	0.1457	0.0728	0.1200	0.0522	0.0900	0.0430	0.0700

NOTE: 1. Highlighted color denotes $\hat{\pi}_L$ difference between RCL method and empirical simulation method.

Table A46: 95% Lower Bound- $\hat{\pi}_L$ -Case 22

{P1+=.167, P2+=.167, P3+=.167, P4+=.167, P5+=.167, P6+=.167. P+1=.3, P+2=.3, P+3=.1, P+4=.1. P+5=.1, P+6=.1}	5 per cell		10 per cell		20 per cell		30 per cell	
	RCL	Empirical	RCL	Empirical	RCL	Empirical	RCL	Empirical
Sampling zero	0.2306	0.2800	0.1615	0.2100	0.1145	0.1500	0.0939	0.1200
Replace with .1	0.2247	0.2800	0.1615	0.2100	0.1145	0.1500	0.0939	0.1200
Replace with .5	0.2080	0.2700	0.1610	0.2100	0.1145	0.1500	0.0939	0.1200
Replace with 1	0.1956	0.2600	0.1605	0.2100	0.1145	0.1500	0.0939	0.1200
Structural zero	0.2121	0.2610	0.1615	0.2100	0.1145	0.1500	0.0939	0.1200

NOTE: 1. Highlighted color denotes $\hat{\pi}_L$ difference between RCL method and empirical simulation method.

Table A47: 95% Lower Bound- $\hat{\pi}_L$ - Case 23

{P1+=.167, P2+=.167, P3+=.167, P4+=.167, P5+=.167, P6+=.167. P+1=.167, P+2=.167, P+3=.167, P+4=.167. P+5=.167, P+6=.167}	5 per cell		10 per cell		20 per cell		30 per cell	
	RCL	Empirical	RCL	Empirical	RCL	Empirical	RCL	Empirical
Sampling zero	0.2490	0.3200	0.1782	0.2300	0.1293	0.1700	0.1068	0.1400
Replace with .1	0.2481	0.3200	0.1782	0.2300	0.1293	0.1700	0.1068	0.1400
Replace with .5	0.2451	0.3200	0.1782	0.2300	0.1293	0.1700	0.1068	0.1400
Replace with 1	0.2422	0.3200	0.1782	0.2300	0.1293	0.1700	0.1068	0.1400
Structural zero	0.2476	0.3200	0.1782	0.2400	0.1293	0.1800	0.1068	0.1400

NOTE: 1. Highlighted color denotes $\hat{\pi}_L$ difference between RCL method and empirical simulation method.

Table A48: 95% Lower Bound- $\hat{\pi}_L$ - Case 24

{P1+=.3, P2+=.3, P3+=.1, P4+=.1, P5+=.1, P6+=.1. P+1=.3, P+2=.3, P+3=.1, P+4=.1. P+5=.1, P+6=.1}	5 per cell		10 per cell		20 per cell		30 per cell	
	RCL	Empirical	RCL	Empirical	RCL	Empirical	RCL	Empirical
Sampling zero	0.2309	0.2900	0.1594	0.2000	0.1097	0.1400	0.0916	0.1200
Replace with .1	0.2121	0.2800	0.1573	0.2000	0.1096	0.1400	0.0916	0.1200
Replace with .5	0.1727	0.2500	0.1509	0.2000	0.1095	0.1400	0.0916	0.1200
Replace with 1	0.1535	0.2300	0.1455	0.2000	0.1094	0.1400	0.0916	0.1200
Structural zero	0.1843	0.2300	0.1547	0.2000	0.1097	0.1500	0.0916	0.1200

NOTE: 1. Highlighted color denotes $\hat{\pi}_L$ difference between RCL method and empirical simulation method.

Table A49: 95% Confidence Interval ($\hat{\pi}^* - \hat{\pi}_I$) and Standard Error of $\hat{\pi}^*$ - Case 1

	5 per cell				10 per cell				20 per cell				30 per cell			
	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE
{P1+=.5, P2+=.5, P+1=.8, P+2=.1, P+3=.1}																
Sampling zero	0.0768	0.0674	0.0467	0.0410	0.0546	0.0500	0.0332	0.0304	0.0397	0.0333	0.0242	0.0202	0.0319	0.0243	0.0194	0.0147
Replace with .1	0.0780	0.0651	0.0474	0.0396	0.0549	0.0498	0.0334	0.0303	0.0397	0.0333	0.0241	0.0202	0.0319	0.0243	0.0194	0.0147
Replace with .5	0.0723	0.0651	0.0439	0.0396	0.0549	0.0488	0.0334	0.0297	0.0397	0.0333	0.0242	0.0202	0.0319	0.0243	0.0194	0.0147
Replace with 1	0.0642	0.0557	0.0391	0.0338	0.0544	0.0477	0.0331	0.0290	0.0397	0.0333	0.0242	0.0202	0.0319	0.0243	0.0194	0.0147
Structural zero	0.0501	0.0485	0.0304	0.0295	0.0499	0.0445	0.0303	0.0270	0.0407	0.0342	0.0247	0.0208	0.0327	0.0251	0.0199	0.0152

Table A50: 95% Confidence Interval ($\hat{\pi}^* - \hat{\pi}_I$) and Standard Error of $\hat{\pi}^*$ - Case 2

	5 per cell				10 per cell				20 per cell				30 per cell			
	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE
{P1+=.5, P2+=.5, P+1=.33, P+2=.33, P+3=.33}																
Sampling zero	0.1438	0.1198	0.0874	0.0729	0.1051	0.0856	0.0639	0.0520	0.0755	0.0621	0.0459	0.0377	0.0610	0.0566	0.0371	0.0344
Replace with .1	0.1440	0.1197	0.0875	0.0728	0.1051	0.0856	0.0639	0.0520	0.0755	0.0621	0.0459	0.0377	0.0611	0.0566	0.0371	0.0344
Replace with .5	0.1442	0.1191	0.0877	0.0724	0.1051	0.0856	0.0639	0.0520	0.0755	0.0621	0.0459	0.0377	0.0611	0.0566	0.0371	0.0344
Replace with 1	0.1441	0.1183	0.0876	0.0719	0.1051	0.0856	0.0639	0.0520	0.0755	0.0621	0.0459	0.0377	0.0611	0.0566	0.0371	0.0344
Structural zero	0.1452	0.1100	0.0883	0.0669	0.1089	0.0894	0.0662	0.0543	0.0781	0.0646	0.0475	0.0393	0.0630	0.0586	0.0383	0.0356

Table A51: 95% Confidence Interval ($\hat{\pi} - \hat{\pi}_I$) and Standard Error of $\hat{\pi}^*$ - Case 3

	5 per cell			10 per cell			20 per cell			30 per cell		
	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE
{P1+=.9, P2+=.1, P+1=.33, P+2=.33, P+3=.33}												
Sampling zero	0.0647	0.0600	0.0393	0.0364	0.0444	0.0444	0.0270	0.0243	0.0333	0.0362	0.0220	0.0176
Replace with .1	0.0644	0.0571	0.0391	0.0347	0.0455	0.0394	0.0277	0.0240	0.0334	0.0361	0.0220	0.0176
Replace with .5	0.0499	0.0405	0.0303	0.0246	0.0435	0.0352	0.0265	0.0214	0.0336	0.0359	0.0204	0.0176
Replace with 1	0.0401	0.0402	0.0244	0.0244	0.0402	0.0312	0.0244	0.0189	0.0335	0.0356	0.0204	0.0176
Structural zero	0.0369	0.0380	0.0224	0.0231	0.0366	0.0349	0.0223	0.0212	0.0336	0.0290	0.0204	0.0178

Table A52: 95% Confidence Interval ($\hat{\pi} - \hat{\pi}_I$) and Standard Error of $\hat{\pi}^*$ - Case 4

	5 per cell			10 per cell			20 per cell			30 per cell		
	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE
{P1+=.9, P2+=.1, P+1=.8, P+2=.1, P+3=.1}												
Sampling zero	0.0734	0.0544	0.0446	0.0330	0.0624	0.0545	0.0379	0.0331	0.0463	0.0504	0.0281	0.0255
Replace with .1	0.0568	0.0573	0.0345	0.0348	0.0553	0.0460	0.0336	0.0279	0.0463	0.0490	0.0281	0.0255
Replace with .5	0.0311	0.0216	0.0189	0.0131	0.0272	0.0275	0.0165	0.0167	0.0339	0.0341	0.0206	0.0215
Replace with 1	0.0410	0.0217	0.0249	0.0132	0.0192	0.0195	0.0117	0.0118	0.0228	0.0228	0.0139	0.0177
Structural zero	0.0329	0.0223	0.0200	0.0136	0.0214	0.0196	0.0130	0.0119	0.0219	0.0230	0.0133	0.0175

Table A53: 95% Confidence Interval ($\hat{\pi}^* - \hat{\pi}_I$) and Standard Error of $\hat{\pi}^*$ - Case 5

{P1+=.5, P2+=.5, P+1=.25, P+2=.25, P+3=.25, P+4=.25}	5 per cell				10 per cell				20 per cell				30 per cell			
	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE
Sampling zero	0.1525	0.1073	0.0927	0.0652	0.1138	0.0814	0.0892	0.0495	0.0813	0.0631	0.0494	0.0384	0.0689	0.0458	0.0407	0.0278
Replace with .1	0.1526	0.1071	0.0928	0.0651	0.1139	0.0814	0.0893	0.0495	0.0813	0.0631	0.0494	0.0384	0.0689	0.0458	0.0407	0.0278
Replace with .5	0.1529	0.1065	0.0929	0.0647	0.1138	0.0814	0.0892	0.0495	0.0813	0.0631	0.0494	0.0384	0.0689	0.0458	0.0407	0.0278
Replace with 1	0.1530	0.1057	0.0930	0.0643	0.1136	0.0812	0.0891	0.0494	0.0813	0.0631	0.0494	0.0384	0.0689	0.0458	0.0407	0.0278
Structural zero	0.1534	0.1169	0.0933	0.0711	0.1174	0.0850	0.0714	0.0517	0.0842	0.0660	0.0512	0.0401	0.0689	0.0479	0.0419	0.0291

Table A54: 95% Confidence Interval ($\hat{\pi}^* - \hat{\pi}_I$) and Standard Error of $\hat{\pi}^*$ - Case 6

{P1+=.5, P2+=.5, P+1=.4, P+2=.4, P+3=.1, P+4=.1}	5 per cell				10 per cell				20 per cell				30 per cell			
	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE
Sampling zero	0.1347	0.1073	0.0819	0.0652	0.1009	0.0755	0.0613	0.0459	0.0740	0.0540	0.0450	0.0328	0.0605	0.0482	0.0368	0.0293
Replace with .1	0.1351	0.1054	0.0821	0.0641	0.1010	0.0755	0.0614	0.0459	0.0740	0.0540	0.0450	0.0328	0.0605	0.0482	0.0368	0.0293
Replace with .5	0.1334	0.0994	0.0811	0.0604	0.1010	0.0751	0.0614	0.0456	0.0740	0.0540	0.0450	0.0328	0.0605	0.0482	0.0368	0.0293
Replace with 1	0.1307	0.1043	0.0794	0.0634	0.1010	0.0748	0.0614	0.0455	0.0740	0.0540	0.0450	0.0328	0.0605	0.0482	0.0368	0.0293
Structural zero	0.1199	0.0925	0.0729	0.0562	0.1027	0.0772	0.0625	0.0470	0.0774	0.0574	0.0470	0.0349	0.0634	0.0511	0.0386	0.0311

Table A55: 95% Confidence Interval ($\hat{\pi}^* - \hat{\pi}_I$) and Standard Error of $\hat{\pi}^*$ - Case 7

{P1+=.9, P2+=.1, P+1=.25, P+2=.25, P+3=.25, P+4=.25}	5 per cell				10 per cell				20 per cell				30 per cell			
	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE
Sampling zero	0.0686	0.0608	0.0417	0.0370	0.0516	0.0411	0.0313	0.0250	0.0385	0.0341	0.0234	0.0207	0.0321	0.0260	0.0195	0.0158
Replace with .1	0.0712	0.0577	0.0433	0.0351	0.0538	0.0405	0.0327	0.0246	0.0388	0.0340	0.0236	0.0207	0.0321	0.0260	0.0195	0.0158
Replace with .5	0.0580	0.0497	0.0352	0.0302	0.0507	0.0335	0.0308	0.0204	0.0391	0.0334	0.0238	0.0203	0.0322	0.0259	0.0196	0.0158
Replace with 1	0.0475	0.0382	0.0289	0.0232	0.0460	0.0377	0.0280	0.0229	0.0391	0.0328	0.0238	0.0199	0.0322	0.0258	0.0196	0.0157
Structural zero	0.0488	0.0415	0.0297	0.0252	0.0448	0.0370	0.0272	0.0225	0.0389	0.0323	0.0236	0.0196	0.0334	0.0265	0.0201	0.0161

Table A56: 95% Confidence Interval ($\hat{\pi}^* - \hat{\pi}_I$) and Standard Error of $\hat{\pi}^*$ - Case 8

{P1+=.9, P2+=.1, P+1=.4, P+2=.4, P+3=.1, P+4=.1}	5 per cell					10 per cell					20 per cell					30 per cell				
	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE
Sampling zero	0.0760	0.0642	0.0462	0.0391	0.0648	0.0522	0.0394	0.0317	0.0560	0.0532	0.0340	0.0323	0.0564	0.0530	0.0343	0.0322				
Replace with .1	0.0702	0.0549	0.0427	0.0334	0.0618	0.0462	0.0376	0.0281	0.0483	0.0429	0.0293	0.0261	0.0398	0.0344	0.0242	0.0209				
Replace with .5	0.0541	0.0460	0.0329	0.0280	0.0445	0.0367	0.0271	0.0223	0.0420	0.0345	0.0256	0.0210	0.0391	0.0317	0.0238	0.0192				
Replace with 1	0.0589	0.0303	0.0358	0.0184	0.0393	0.0310	0.0239	0.0189	0.0360	0.0280	0.0219	0.0170	0.0370	0.0287	0.0225	0.0175				
Structural zero	0.0552	0.0389	0.0335	0.0237	0.0405	0.0330	0.0246	0.0201	0.0358	0.0288	0.0218	0.0175	0.0368	0.0290	0.0224	0.0176				

Table A57: 95% Confidence Interval ($\hat{\pi}^* - \hat{\pi}_t$) and Standard Error of $\hat{\pi}^*$ - Case 9

{P1+=.5, P2+=.5, P+1=.167, P+2=.167, P+3=.167, P+4=.167, P+5=.167, P+6=.167}	5 per cell					10 per cell					20 per cell					30 per cell				
	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE
Sampling zero	0.1544	0.1023	0.0938	0.0622	0.1136	0.0752	0.0690	0.0457	0.0814	0.0538	0.0495	0.0327	0.0663	0.0437	0.0403	0.0266				
Replace with .1	0.1544	0.1022	0.0939	0.0621	0.1136	0.0752	0.0690	0.0457	0.0814	0.0538	0.0495	0.0327	0.0663	0.0437	0.0403	0.0266				
Replace with .5	0.1544	0.1014	0.0939	0.0617	0.1136	0.0752	0.0690	0.0457	0.0814	0.0538	0.0495	0.0327	0.0663	0.0437	0.0403	0.0266				
Replace with 1	0.1544	0.1006	0.0939	0.0611	0.1136	0.0752	0.0690	0.0457	0.0814	0.0538	0.0495	0.0327	0.0663	0.0437	0.0403	0.0266				
Structural zero	0.1530	0.1008	0.0930	0.0613	0.1166	0.0782	0.0709	0.0476	0.0836	0.0561	0.0508	0.0341	0.0682	0.0457	0.0415	0.0278				

Table A58: 95% Confidence Interval ($\hat{\pi}^* - \hat{\pi}_t$) and Standard Error of $\hat{\pi}^*$ - Case 10

{P1+=.9, P2+=.1, P+1=.167, P+2=.167, P+3=.167, P+4=.167, P+5=.167, P+6=.167}	5 per cell					10 per cell					20 per cell					30 per cell				
	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE
Sampling zero	0.0638	0.0568	0.0388	0.0345	0.0532	0.0418	0.0323	0.0254	0.0427	0.0328	0.0259	0.0199	0.0368	0.0244	0.0223	0.0148				
Replace with .1	0.0721	0.0551	0.0438	0.0335	0.0606	0.0442	0.0369	0.0269	0.0440	0.0334	0.0267	0.0203	0.0368	0.0244	0.0224	0.0148				
Replace with .5	0.0616	0.0449	0.0375	0.0273	0.0544	0.0309	0.0331	0.0188	0.0437	0.0318	0.0265	0.0194	0.0369	0.0243	0.0224	0.0148				
Replace with 1	0.0516	0.0330	0.0314	0.0200	0.0492	0.0330	0.0299	0.0201	0.0435	0.0309	0.0264	0.0188	0.0369	0.0242	0.0225	0.0147				
Structural zero	0.0608	0.0418	0.0369	0.0254	0.0504	0.0349	0.0306	0.0212	0.0450	0.0317	0.0274	0.0193	0.0384	0.0253	0.0234	0.0154				

Table A59: 95% Confidence Interval ($\hat{\pi}^* - \hat{\pi}_I$) and Standard Error of $\hat{\pi}^*$ - Case 11

{P1+=.5, P2+=.5, P+1=.3, P+2=.3, P+3=.1, P+4=.1, P+5=.1, P+6=.1}	5 per cell					10 per cell					20 per cell					30 per cell				
	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE
Sampling zero	0.1445	0.0975	0.0879	0.0593	0.1053	0.0732	0.0840	0.0445	0.0745	0.0529	0.0453	0.0321	0.0618	0.0471	0.0375	0.0286				
Replace with .1	0.1449	0.0968	0.0881	0.0588	0.1053	0.0732	0.0840	0.0445	0.0745	0.0529	0.0453	0.0321	0.0618	0.0471	0.0375	0.0286				
Replace with .5	0.1451	0.0941	0.0882	0.0572	0.1054	0.0732	0.0840	0.0445	0.0745	0.0529	0.0453	0.0321	0.0618	0.0471	0.0375	0.0286				
Replace with 1	0.1449	0.1011	0.0881	0.0614	0.1053	0.0731	0.0840	0.0444	0.0745	0.0529	0.0453	0.0321	0.0618	0.0471	0.0375	0.0286				
Structural zero	0.1364	0.0972	0.0829	0.0591	0.1084	0.0762	0.0859	0.0463	0.0766	0.0550	0.0465	0.0334	0.0637	0.0490	0.0387	0.0298				

Table A60: 95% Confidence Interval ($\hat{\pi}^* - \hat{\pi}_I$) and Standard Error of $\hat{\pi}^*$ - Case 12

{P1+=.9, P2+=.1, P+1=.3, P+2=.3, P+3=.1, P+4=.1, P+5=.1, P+6=.1}	5 per cell					10 per cell					20 per cell					30 per cell				
	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE
Sampling zero	0.0696	0.0595	0.0423	0.0361	0.0622	0.0407	0.0378	0.0248	0.0475	0.0412	0.0289	0.0251	0.0412	0.0314	0.0250	0.0191				
Replace with .1	0.0742	0.0541	0.0451	0.0329	0.0672	0.0388	0.0409	0.0236	0.0509	0.0416	0.0309	0.0253	0.0423	0.0314	0.0257	0.0191				
Replace with .5	0.0602	0.0432	0.0366	0.0262	0.0537	0.0380	0.0327	0.0231	0.0490	0.0352	0.0298	0.0214	0.0428	0.0299	0.0260	0.0182				
Replace with 1	0.0615	0.0336	0.0374	0.0204	0.0461	0.0288	0.0280	0.0175	0.0460	0.0304	0.0280	0.0185	0.0426	0.0284	0.0259	0.0173				
Structural zero	0.0657	0.0409	0.0399	0.0248	0.0500	0.0341	0.0304	0.0207	0.0485	0.0328	0.0295	0.0200	0.0448	0.0299	0.0272	0.0182				

Table A61: 95% Confidence Interval ($\hat{\pi}^* - \hat{\pi}_I$) and Standard Error of $\hat{\pi}^*$ - Case 13

{P1 +=.5, P2 +=.5, P+1= .5, P+2= .5}	5 per cell				10 per cell				20 per cell				30 per cell			
	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE
Sampling zero	0.1067	0.1152	0.0648	0.0701	0.0808	0.0860	0.0491	0.0523	0.0637	0.0675	0.0387	0.0410	0.0538	0.0568	0.0327	0.0346
Replace with .1	0.1069	0.1152	0.0650	0.0700	0.0808	0.0860	0.0491	0.0523	0.0637	0.0675	0.0387	0.0410	0.0538	0.0568	0.0327	0.0346
Replace with .5	0.1071	0.1148	0.0651	0.0698	0.0808	0.0860	0.0491	0.0523	0.0637	0.0675	0.0387	0.0410	0.0538	0.0568	0.0327	0.0346
Replace with 1	0.1071	0.1143	0.0651	0.0695	0.0808	0.0860	0.0491	0.0523	0.0637	0.0675	0.0387	0.0410	0.0538	0.0568	0.0327	0.0346
Structural zero	0.1093	0.1099	0.0664	0.0668	0.0863	0.0915	0.0524	0.0556	0.0682	0.0720	0.0415	0.0438	0.0581	0.0612	0.0353	0.0372

Table 62: 95% Confidence Interval ($\hat{\pi}^* - \hat{\pi}_I$) and Standard Error of $\hat{\pi}^*$ - Case 14

{P1 +=.9, P2 +=.1, P+1= .9, P+2= .1}	5 per cell				10 per cell				20 per cell				30 per cell			
	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE
Sampling zero	0.0496	0.0498	0.0302	0.0303	0.0466	0.0468	0.0283	0.0285	0.0371	0.0376	0.0225	0.0228	0.0283	0.0296	0.0172	0.0180
Replace with .1	0.0311	0.0313	0.0189	0.0190	0.0362	0.0363	0.0220	0.0221	0.0332	0.0334	0.0202	0.0203	0.0277	0.0283	0.0168	0.0172
Replace with .5	0.0184	0.0185	0.0112	0.0113	0.0130	0.0131	0.0079	0.0080	0.0172	0.0172	0.0105	0.0105	0.0202	0.0202	0.0123	0.0123
Replace with 1	0.0319	0.0220	0.0194	0.0134	0.0129	0.0130	0.0078	0.0079	0.0082	0.0082	0.0050	0.0050	0.0123	0.0123	0.0075	0.0075
Structural zero	0.0160	0.0147	0.0097	0.0089	0.0135	0.0135	0.0082	0.0082	0.0084	0.0085	0.0051	0.0052	0.0113	0.0115	0.0069	0.0070

Table A63: 95% Confidence Interval ($\hat{\pi}^* - \hat{\pi}_I$) and Standard Error of $\hat{\pi}$ - Case 15

{P1+=.5, P2+=.5, P+1=.9, P+2=.1}	5 per cell				10 per cell				20 per cell				30 per cell			
	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE
Sampling zero	0.0433	0.0459	0.0263	0.0279	0.0321	0.0345	0.0195	0.0209	0.0229	0.0243	0.0139	0.0148	0.0187	0.0198	0.0114	0.0120
Replace with .1	0.0441	0.0455	0.0268	0.0276	0.0328	0.0344	0.0199	0.0209	0.0229	0.0243	0.0139	0.0148	0.0187	0.0198	0.0114	0.0120
Replace with .5	0.0348	0.0351	0.0212	0.0213	0.0309	0.0318	0.0188	0.0193	0.0230	0.0242	0.0140	0.0147	0.0187	0.0198	0.0113	0.0120
Replace with 1	0.0285	0.0285	0.0173	0.0173	0.0287	0.0292	0.0174	0.0178	0.0228	0.0240	0.0139	0.0146	0.0187	0.0198	0.0113	0.0120
Structural zero	0.0204	0.0210	0.0124	0.0128	0.0246	0.0251	0.0149	0.0153	0.0213	0.0221	0.0129	0.0135	0.0188	0.0196	0.0114	0.0119

Table A64: 95% Confidence Interval ($\hat{\pi}^* - \hat{\pi}_I$) and Standard Error of $\hat{\pi}$ - Case 16

{P1+=.4, P2+=.4, P3+=.2, P+1=.4, P+2=.4, P+3=.2}	5 per cell				10 per cell				20 per cell				30 per cell			
	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE
Sampling zero	0.1644	0.1162	0.0999	0.0706	0.1249	0.0919	0.0759	0.0559	0.0907	0.0663	0.0551	0.0403	0.0756	0.0483	0.0459	0.0293
Replace with .1	0.1655	0.1140	0.1006	0.0893	0.1251	0.0918	0.0761	0.0558	0.0907	0.0663	0.0551	0.0403	0.0756	0.0483	0.0459	0.0293
Replace with .5	0.1645	0.1075	0.1000	0.0854	0.1254	0.0914	0.0762	0.0555	0.0907	0.0663	0.0551	0.0403	0.0756	0.0483	0.0459	0.0293
Replace with 1	0.1623	0.1122	0.0986	0.0882	0.1254	0.0908	0.0762	0.0552	0.0907	0.0663	0.0551	0.0403	0.0756	0.0483	0.0459	0.0293
Structural zero	0.1521	0.1015	0.0925	0.0617	0.1283	0.0950	0.0780	0.0577	0.0949	0.0604	0.0577	0.0367	0.0796	0.0523	0.0484	0.0318

Table A65: 95% Confidence Interval ($\hat{\pi}^* - \hat{\pi}_I$) and Standard Error of $\hat{\pi}^*$ - Case 17

	5 per cell				10 per cell				20 per cell				30 per cell			
	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE
{P1+=.33, P2+=.33, P3+=.33, P+1=.33, P+2=.33, P+3=.33}																
Sampling zero	0.1795	0.1238	0.1091	0.0752	0.1362	0.0947	0.0828	0.0576	0.1008	0.0677	0.0613	0.0412	0.0837	0.0579	0.0509	0.0352
Replace with .1	0.1800	0.1236	0.1094	0.0751	0.1362	0.0947	0.0828	0.0576	0.1008	0.0677	0.0613	0.0412	0.0837	0.0579	0.0509	0.0352
Replace with .5	0.1802	0.1223	0.1096	0.0744	0.1362	0.0947	0.0828	0.0576	0.1008	0.0677	0.0613	0.0412	0.0837	0.0579	0.0509	0.0352
Replace with 1	0.1805	0.1212	0.1097	0.0737	0.1362	0.0947	0.0828	0.0576	0.1008	0.0677	0.0613	0.0412	0.0837	0.0579	0.0509	0.0352
Structural zero	0.1822	0.1247	0.1107	0.0758	0.1436	0.1020	0.0873	0.0620	0.1085	0.0753	0.0659	0.0458	0.0903	0.0645	0.0549	0.0392

Table A66: 95% Confidence Interval ($\hat{\pi}^* - \hat{\pi}_I$) and Standard Error of $\hat{\pi}^*$ - Case 18

	5 per cell				10 per cell				20 per cell				30 per cell			
	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE
{P1+=.33, P2+=.33, P3+=.33, P+1=.4, P+2=.4, P+3=.2}																
Sampling zero	0.1720	0.1167	0.1046	0.0710	0.1273	0.0845	0.0774	0.0514	0.0917	0.0587	0.0558	0.0357	0.0764	0.0493	0.0464	0.0299
Replace with .1	0.1733	0.1158	0.1053	0.0704	0.1273	0.0845	0.0774	0.0514	0.0917	0.0587	0.0558	0.0357	0.0764	0.0493	0.0464	0.0299
Replace with .5	0.1738	0.1217	0.1056	0.0740	0.1273	0.0845	0.0774	0.0513	0.0917	0.0587	0.0558	0.0357	0.0764	0.0493	0.0464	0.0299
Replace with 1	0.1728	0.1180	0.1051	0.0717	0.1273	0.0844	0.0774	0.0513	0.0917	0.0587	0.0558	0.0357	0.0764	0.0493	0.0464	0.0299
Structural zero	0.1656	0.1160	0.1006	0.0705	0.1324	0.0896	0.0805	0.0544	0.0959	0.0628	0.0583	0.0382	0.0802	0.0530	0.0487	0.0322

Table A67: 95% Confidence Interval ($\hat{\pi}^* - \hat{\pi}_I$) and Standard Error of $\hat{\pi}^*$ - Case 19

	5 per cell				10 per cell				20 per cell				30 per cell			
	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE
{P1+=.25, P2+=.25, P3+=.25, P4+=.25, P+1=.25, P+2=.25, P+3=.25, P+4=.25}																
Sampling zero	0.1868	0.1060	0.1135	0.0644	0.1417	0.0880	0.0862	0.0535	0.1058	0.0684	0.0643	0.0416	0.0893	0.0522	0.0543	0.0317
Replace with .1	0.1872	0.1046	0.1138	0.0636	0.1417	0.0880	0.0862	0.0535	0.1058	0.0684	0.0643	0.0416	0.0893	0.0522	0.0543	0.0317
Replace with .5	0.1874	0.1032	0.1139	0.0627	0.1417	0.0880	0.0862	0.0535	0.1058	0.0684	0.0643	0.0416	0.0893	0.0522	0.0543	0.0317
Replace with 1	0.1417	0.0880	0.0862	0.0535	0.1059	0.0684	0.0643	0.0416	0.0893	0.0522	-0.2353	0.0317	0.0893	0.0522	0.0543	0.0317
Structural zero	0.1907	0.1092	0.1159	0.0664	0.1512	0.0874	0.0919	0.0532	0.1145	0.0671	0.0696	0.0408	0.0984	0.0593	0.0586	0.0361

Table A68: 95% Confidence Interval ($\hat{\pi}^* - \hat{\pi}_I$) and Standard Error of $\hat{\pi}^*$ - Case 20

	5 per cell				10 per cell				20 per cell				30 per cell			
	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE
{P1+=.4, P2+=.4, P3+=.1, P4+=.1, P+1=.4, P+2=.4, P+3=.1, P+4=.1}																
Sampling zero	0.1545	0.0950	0.0939	0.0578	0.1185	0.0773	0.0721	0.0470	0.0891	0.0502	0.0541	0.0305	0.0750	0.0478	0.0456	0.0290
Replace with .1	0.1430	0.0852	0.0869	0.0518	0.1162	0.0715	0.0706	0.0434	0.0900	0.0478	0.0547	0.0290	0.0751	0.0474	0.0457	0.0288
Replace with .5	0.1396	0.0896	0.0849	0.0544	0.1106	0.0707	0.0672	0.0430	0.0898	0.0454	0.0546	0.0276	0.0753	0.0471	0.0458	0.0286
Replace with 1	0.1145	0.0819	0.0696	0.0498	0.0883	0.0508	0.0537	0.0309	0.0748	0.0478	0.0455	0.0291	0.0753	0.0471	0.0458	0.0286
Structural zero	0.1251	0.0887	0.0760	0.0539	0.0952	0.0892	0.0579	0.0421	0.0861	0.0577	0.0523	0.0351	0.0781	0.0507	0.0475	0.0308

Table A69: 95% Confidence Interval ($\hat{\pi}^* - \hat{\pi}_I$) and Standard Error of $\hat{\pi}^*$ - Case 21

	5 per cell				10 per cell				20 per cell				30 per cell			
	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE
{P1+=.25, P2+=.25, P3+=.25, P4+=.25, P+1=.4, P+2=.4, P+3=.1, P+4=.1}																
Sampling zero	0.1582	0.0987	0.0962	0.0600	0.1180	0.0709	0.0717	0.0431	0.0869	0.0491	0.0528	0.0298	0.0715	0.0445	0.0435	0.0271
Replace with .1	0.1609	0.0942	0.0978	0.0572	0.1186	0.0697	0.0721	0.0423	0.0869	0.0491	0.0528	0.0298	0.0715	0.0445	0.0435	0.0271
Replace with .5	0.1601	0.0930	0.0973	0.0565	0.1188	0.0682	0.0722	0.0414	0.0869	0.0491	0.0528	0.0298	0.0715	0.0445	0.0435	0.0271
Replace with 1	0.1177	0.0712	0.0716	0.0433	0.0869	0.0491	0.0528	0.0298	0.0715	0.0445	0.0435	0.0271	0.0715	0.0445	0.0435	0.0271
Structural zero	0.1404	0.0917	0.0853	0.0558	0.1159	0.0688	0.0705	0.0418	0.0897	0.0519	0.0545	0.0315	0.0737	0.0466	0.0448	0.0283

Table A70: 95% Confidence Interval ($\hat{\pi}^* - \hat{\pi}_I$) and Standard Error of $\hat{\pi}^*$ - Case 22

	5 per cell				10 per cell				20 per cell				30 per cell			
	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE
{P1+=.167, P2+=.167, P3+=.167, P4+=.167, P5+=.167, P6+=.167, P+1=.3, P+2=.3, P+3=.1, P+4=.1, P+5=.1, P+6=.1}																
Sampling zero	0.1374	0.0880	0.0835	0.0535	0.1095	0.0610	0.0686	0.0371	0.0819	0.0464	0.0498	0.0282	0.0687	0.0426	0.0418	0.0259
Replace with .1	0.1398	0.0845	0.0850	0.0514	0.1095	0.0610	0.0686	0.0371	0.0819	0.0464	0.0498	0.0282	0.0687	0.0426	0.0418	0.0259
Replace with .5	0.1433	0.0812	0.0871	0.0494	0.1097	0.0607	0.0667	0.0369	0.0819	0.0464	0.0498	0.0282	0.0687	0.0426	0.0418	0.0259
Replace with 1	0.1443	0.0799	0.0877	0.0486	0.1097	0.0602	0.0667	0.0366	0.0819	0.0464	0.0498	0.0282	0.0687	0.0426	0.0418	0.0259
Structural zero	0.1283	0.0793	0.0780	0.0482	0.1126	0.0641	0.0685	0.0390	0.0858	0.0503	0.0521	0.0306	0.0718	0.0457	0.0436	0.0278

Table A71: 95% Confidence Interval ($\hat{\pi}^* - \hat{\pi}_I$) and Standard Error of $\hat{\pi}^*$ - Case 23

{P1+=.167, P2+=.167, P3+=.167, P4+=.167, P5+=.167, P6+=.167, P+1=.167, P+2=.167, P+3=.167, P+4=.167, P+5=.167, P+6=.167}	5 per cell					10 per cell					20 per cell					30 per cell						
	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE		
Sampling zero	0.1595	0.0886	0.0970	0.0538	0.1249	0.0731	0.0759	0.0444	0.0949	0.0542	0.0577	0.0329	0.0467	0.0284	0.0799	0.0467	0.0486	0.0284	0.0799	0.0467	0.0486	0.0284
Replace with .1	0.1597	0.0879	0.0971	0.0534	0.1249	0.0731	0.0759	0.0444	0.0949	0.0542	0.0577	0.0329	0.0467	0.0284	0.0799	0.0467	0.0486	0.0284	0.0799	0.0467	0.0486	0.0284
Replace with .5	0.1603	0.0854	0.0974	0.0519	0.1249	0.0731	0.0759	0.0444	0.0949	0.0542	0.0577	0.0329	0.0467	0.0284	0.0799	0.0467	0.0486	0.0284	0.0799	0.0467	0.0486	0.0284
Replace with 1	0.1606	0.0828	0.0977	0.0503	0.1249	0.0731	0.0759	0.0444	0.0949	0.0542	0.0577	0.0329	0.0467	0.0284	0.0799	0.0467	0.0486	0.0284	0.0799	0.0467	0.0486	0.0284
Structural zero	0.1612	0.0888	0.0980	0.0540	0.1364	0.0746	0.0829	0.0454	0.1044	0.0537	0.0635	0.0326	0.0540	0.0328	0.0872	0.0540	0.0530	0.0328	0.0872	0.0540	0.0530	0.0328

Table A72: 95% Confidence Interval ($\hat{\pi}^* - \hat{\pi}_I$) and Standard Error of $\hat{\pi}^*$ - Case 24

	5 per cell					10 per cell					20 per cell					30 per cell						
	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE	RCL based CI	"True" CI	RCL based SE	"True" SE		
{P1+=.3, P2+=.3, P3+=.1, P4+=.1, P5+=.1, P6+=.1, P+1=.3, P+2=.3, P+3=.1, P+4=.1, P+5=.1, P+6=.1}																						
Sampling zero	0.1320	0.0729	0.0803	0.0443	0.1099	0.0693	0.0668	0.0421	0.0845	0.0542	0.0514	0.0330	0.0431	0.0259	0.0710	0.0426	0.0431	0.0259	0.0710	0.0426	0.0431	0.0259
Replace with .1	0.1415	0.0736	0.0860	0.0447	0.1109	0.0681	0.0674	0.0414	0.0846	0.0542	0.0514	0.0329	0.0431	0.0259	0.0710	0.0426	0.0431	0.0259	0.0710	0.0426	0.0431	0.0259
Replace with .5	0.1449	0.0675	0.0881	0.0410	0.1120	0.0629	0.0681	0.0382	0.0846	0.0541	0.0514	0.0329	0.0431	0.0259	0.0710	0.0426	0.0431	0.0259	0.0710	0.0426	0.0431	0.0259
Replace with 1	0.1399	0.0633	0.0850	0.0385	0.1124	0.0579	0.0683	0.0352	0.0846	0.0540	0.0514	0.0328	0.0431	0.0259	0.0710	0.0426	0.0431	0.0259	0.0710	0.0426	0.0431	0.0259
Structural zero	0.1177	0.0720	0.0715	0.0437	0.1058	0.0605	0.0643	0.0368	0.0904	0.0501	0.0550	0.0305	0.0467	0.0294	0.0768	0.0484	0.0467	0.0294	0.0768	0.0484	0.0467	0.0294

APPENDIX B: SAS PROGRAMS

```
/* Structural zero scenario*/
/* This program performs mixture index of fit and 95% lower bound when
encountering zero cell in data table using structural zero technique*/

filename junk dummy;
proc printto log=junk; run;
dm 'log;clear;out;clear;';
options ps=50 nodate nonumber formdlm=' ';
proc iml;
%let rr=6; %let cc=6; /* set table size*/
%let n=180; /*set total sample size=table size* x per cell*/
%let iter=1000; /*number of iterations*/
savepi=J(&iter,4,9); /*matrix to save all 1000 mixture index of fit and
95% lower bound, before and after implementing of structural zero
technique*/
rsum={.3,
      .3,
      .1,
      .1,
      .1,
      .1};
csum={.3 .3 .1 .1 .1 .1}; /* set row and column marginal distributions*/
actual1=rsum*csum;
*print actual1;
/*cum=j(&rr, &cc, 0)*/;
cum=cusum(actual1); /* calculate cumulative proportions */
*print cum;
COUNT0=0;
count1zero=0;
do ii=1 to &iter; /* number of iterations to simulate*/
SKIP:
COUNT0=0;
count=j(&rr, &cc, 0);
cum=cusum(actual1);
do m=1 to &n;
    random=round(uniform(1000), .001); /* Generate n=sample size Uniform
distributed proportions between 0 and 1, rounded to .001*/
    do i=1 to &rr;
        do j=1 to &cc;
            if random<cum[i,j] then
                do;
                    count[i, j]=count[i, j]+1;
/* Allocate cell proportions to associated categories according to
cumulative distributions*/
                    j=&cc;
                    i=&rr;
                end;
            end;
        end;
    end;
    *print count;
    do i=1 to &rr;
        do j=1 to &cc;
            if count[i, j]=0 then do;
```

```

        /*print ii;
        print "Count contains 0";*/
        COUNT0=COUNT0+1;
        *PRINT II;
        if COUNT0>=2 then do;
            ii=ii;
            GOTO SKIP;
/* if table contains more than 1 zero cell, than discard the table and simulate
a new one */
            end;
            /*j=&cc;
            i=&rr;*/
            end;
        end;
        actual=count/sum(count);
        /*print actual;*/
        rowsum=actual[,+]; /*calculate row totals*/
        coloumsum=actual[+,+]; /*calculate column totals*/
        e=(actual[,+]*actual[+,+]); /*calculated expected proportions*/

do i=1 to 100; /* increment =.01 each iteration */
    Pi=i/100;
        Q2=j(&rr,&cc,i/(&rr*&cc*100)); /* initial Q2 matrix */
        Q1=e*(1-i/100); /* initial Q1 matrix */
        QQ2=j(&rr,&cc,i/(&rr*&cc*100));
        g1=j(&rr,&cc,0);
        g2=j(&rr,&cc,0);
        d=j(1,1,1);
    do while (sum(d)>.0001); /* convergence criteria */
        do k=1 to &rr by 1;
            do j=1 to &cc by 1;
                g1[k, j]= actual[k,j]*Q1[k, j]/(Q1[k,j]+Q2[k, j]);
                g2[k, j]= actual[k,j]*Q2[k, j]/(Q1[k,j]+Q2[k, j]);
            end;
        end;
        do k=1 to &rr;
            do j=1 to &cc;
                g1[k, j]=(1-i/100)*g1[k, j]/sum(g1);
                QQ2[k, j]=i/100*g2[k, j]/sum(g2);
            end;
        end;
        QQ1=(g1[,+]*g1[+,+])/(1-i/100); /*QQ1 after EM algorithm*/
        QQ1=(1-i/100)*QQ1/sum(QQ1);

        d=sum(abs((abs(QQ1+QQ2)*sum(count)-abs(Q2+Q1)*sum(count))));
        /*print d;*/
        Q1=QQ1;
        Q2=QQ2;
        /*PRINT Q1;
        PRINT actual;*/
    end;
    /*print Q1;
    print Q2;*/
    ff=(Q1+Q2)*sum(count); /* expected frequencies*/
    /*print ff;*/

```

```

        dd=sum(abs(actual*sum(count)-ff))/2/sum(count);
        chinew=sum((ff-actual*sum(count))##2/ff);      /*calculate
chi-square*/
        do m=1 to &rr;
            do n=1 to &cc;
                if count[m,n]=0 | ff[m,n]=0 then G2[m,n]=0;
                else G2[m,n]=(count[m,n])*(LOG(count[m,n])-LOG(ff[m,n]));
            end;
        end;
        Gsqu=2*sum(G2); /* calculate G-square value */
        chinew=sum((ff-count)##2/ff);
/*PRINT PI;
PRINT GSQU;
PRINT CHINEW;*/

If Gsqu< 2.71& pi<savepi[ii,2]then savepi[ii,2]=pi;
/*Save lower bound of mixture index of fit into second column of savepi
matrix*/ If Gsqu<.00001 then
do;
if pi=1 then GOTO SKIP; X=SUM(COUNT);
/*PRINT X;*/
QQ1=QQ1*sum(count);
** * print QQ1;
QQ2=QQ2*sum(count);
*PRINT QQ2;
savepi[ii,1]=pi; /*save mixture index of fit into first column of savepi
matrix*/
*print ii;
savepi[ii,3]=savepi[ii,1];
savepi[ii,4]=savepi[ii,2];
/*print chinew;*/
i=100;
end;
end;
*print savepi;
*PRINT COUNT0;
p=0;
q=0;
do a=1 to &rr;
    do b=1 to &cc;
        if count[a,b]=0 then do; /* check if any zero cell in data table*/
            countlzero=countlzero+1;
            p=a;
            q=b;
        end;
    end;
end;

if p^=0 then do; CC=QQ1[P,+];
if QQ1[+,q]<=.01 then do;
/* if the column of first component QQ1 sums less than .01 then follow the
below procedure to compute new mixture index of fit and lower bound using
structural zero technique and save in third and fourth column of matrix
savepi*/
CC=QQ1[P,+];
QQ1sum=sum(QQ1);
QQ1[,q]=QQ1[,q]+QQ2[,q]; /*pull the column frequency back from QQ2 for

```

```

column of first component QQ1 adds less than .01*/
if p=1 then count=QQ1[2:&rr,];
else if p=&rr then count=QQ1[1:&rr-1,];
else count=QQ1[1:p-1,]/QQ1[p+1:&rr,];
*PRINT COUNT;
sum=sum(count);
*PRINT SUM;
rowsum=count[,+];
coloumsum=count[+,];
*print sum, rowsum, coloumsum;
expt=rowsum*coloumsum;
e=expt/sum**2;
f=sum(e);
*print f;
*print e;
actual=count/sum;
do i=1 to 100;
Pi=i/100;
*print pi;
    Q2=j(&rr-1,&cc,i/((&rr-1)*&cc*100));
    Q1=e*(1-i/100);
    QQ2=j(&rr-1,&cc,i/((&rr-1)*&cc*100));
    g1=j(&rr-1,&cc,0);
    g2=j(&rr-1,&cc,0);
    d=j(&rr-1,&cc,1);
do while (sum(d)>.00001);
do k=1 to &rr-1 by 1;
do j=1 to &cc by 1;
g1[k, j]= actual[k,j]*Q1[k, j]/(Q1[k,j]+Q2[k, j]);
g2[k, j]= actual[k,j]*Q2[k, j]/(Q1[k,j]+Q2[k, j]);
end;
end;
do k=1 to &rr-1;
do j=1 to &cc;
g1[k, j]=(1-i/100)*g1[k, j]/sum(g1);
QQ2[k, j]=i/100*g2[k, j]/sum(g2);
end;
end;
if q=1 then do;
do k=1 to &rr-1;
do j=2 to &cc;
QQ2[k, j]=0;
end;
end;
end;
if q=&cc then do;
do k=1 to &rr-1;
do j=1 to &cc-1;
QQ2[k, j]=0;
end;
end;
end;
else do;
do k=1 to &rr-1;
do j=1 to q-1;

QQ2[k, j]=0;

```

```

        end;
        do j=q+1 to &cc;
            QQ2[k, j]=0;
        end;
    end;
end;
end;
QQ1=(g1[,+]*g1[+,])/(1-i/100);
QQ1=(1-i/100)*QQ1/sum(QQ1);
d=sum(abs((abs(QQ1+QQ2)*sum-abs(Q2+Q1)*sum)));
ff=(QQ1+QQ2)*sum;
Q1=QQ1;
Q2=QQ2;
end;

do m=1 to nrow(QQ2);
    do n=1 to ncol(QQ2);
        G2=j(nrow(QQ2),ncol(QQ2),0);
        if count[m,n]=0 then G2[m,n]=0;
        else G2[m,n]=(count[m,n])*(LOG(count[m,n])-LOG(ff[m,n]));
    /*compute new G-square using structural zero technique*/
    end;
    end;
    Gsqu=2*sum(G2); /*calculate new G-square statistic*/
    chinew=sum((ff-count)##2/ff); /*calculatate new chi-square
statistic*/
    /* print QQ1;*/
    If Gsqu< 2.71 then do;
        QQ1=QQ1*sum;
        pinew=1-(CC+sum(QQ1))/X;
        if pinew<savepi[ii,4] then savepi[ii,4]=pinew; /*input new lower
bound of mixture index of fit into fourth column of matrix savepi*/
        if savepi[ii,4]<0 then savepi[ii,4]=0;
    end;
    If Gsqu<.00001 & chinew<.00001 then do;
        pinew=1-(CC+sum(QQ1))/X; /*save new mixture index of fit into third
column of matrix savepi*/
        savepi[ii,3]=pinew;
        i=100;
    end;
    *end;
end;
end;

else if QQ1[P,+]<=.01 then do; CC=QQ1[+,Q];
/* if the row of first component QQ1 sums less than .01 then follow the
below procedure to compute new mixture index of fit and lower bound using
structural zero technique and save in third and fourth column of matrix
savepi*/
QQ1sum=sum(QQ1);
QQ1[P,]=QQ1[P,]+QQ2[P,]; /*pull the row frequency back from QQ2 for column
of first component QQ1 adds less than .01*/
*print QQ1;
if q=1 then count=QQ1[,2:&cc];
else if q=&cc then count=QQ1[, 1:&cc-1];
else count=QQ1[,1:q-1]||QQ1[, q+1:&cc];
sum=sum(count);

```

```

*PRINT SUM;
rowsum=count[,+];
coloumsum=count[+,];
*print sum, rowsum, coloumsum;
expt=rowsum*coloumsum;
e=expt/sum**2;
f=sum(e);
*print f;
*print e;
actual=count/sum;
do i=1 to 100;
Pi=i/100;
*print pi;
    Q2=j(&rr,&cc-1,i/(&rr*(&cc-1)*100));
    Q1=e*(1-i/100);
    QQ2=j(&rr,&cc-1,i/(&rr*(&cc-1)*100));
    g1=j(&rr,&cc-1,0);
    g2=j(&rr,&cc-1,0);
    d=j(&rr,&cc-1,1);
do while (sum(d)>.00001);
do k=1 to &rr by 1;
do j=1 to &cc-1 by 1;
g1[k, j]= actual[k,j]*Q1[k, j]/(Q1[k,j]+Q2[k, j]);
g2[k, j]= actual[k,j]*Q2[k, j]/(Q1[k,j]+Q2[k, j]);
end;
end;
do k=1 to &rr;
do j=1 to &cc-1;
g1[k, j]=(1-i/100)*g1[k, j]/sum(g1);
QQ2[k, j]=i/100*g2[k, j]/sum(g2);
end;
end;

if p=1 then do;
do k=2 to &rr;
do j=1 to &cc-1;
QQ2[k, j]=0;
end;
end;
end;
if p=&rr then do;
do k=1 to &rr-1;
do j=1 to &cc-1;
QQ2[k, j]=0;
end;
end;
end;
else do;
do j=1 to &cc-1;
do k=1 to p-1;
QQ2[k, j]=0;
end;
do k=p+1 to &rr;
QQ2[k, j]=0;
end;
end;
end;
end;

```



```

*end;
      QQ1=(g1[,+]*g1[+,])/(1-i/100);
      QQ1=(1-i/100)*QQ1/sum(QQ1);
      d=sum(abs((abs(QQ1+QQ2)*sum-abs(Q2+Q1)*sum)));
      ff=(QQ1+QQ2)*SUM;
      Q1=QQ1;
      Q2=QQ2;
    end;
    do m=1 to nrow(QQ1);
      do n=1 to ncol(QQ1);
        G2=j(nrow(QQ1),ncol(QQ1),0);
        if count[m,n]=0 then G2[m,n]=0;
        else G2[m,n]=(count[m,n])*(LOG(count[m,n])-LOG(ff[m,n]));
      end;
    end;
    *print G2;
    Gsqu=(2*sum(G2)); /*compute new G-square using structural zero
technique*/
    chinew=sum((ff-count)**2/ff); /*compute new chi-square using
structural zero technique*/
    If Gsqu< 2.71 then do;
      QQ1=QQ1*sum;
      pinew=1-(CC+sum(QQ1))/X;
      if pinew<savepi[ii,4] then savepi[ii,4]=pinew; /*input new lower
bound of mixture index of fit into fourth column of matrix savepi*/
      if savepi[ii,4]<0 then savepi[ii,4]=0;
    end;
    If chinew<.00001 & chinew<.00001 then do;
      pinew=1-(CC+sum(QQ1))/X;
      savepi[ii,3]=pinew; /*save new mixture index of fit into third
column of matrix savepi*/
      i=100;
    end;
  end;
end;
end;
end;
pistar=savepi[+,3]/&iter; /*new pi star is the average is 1000
iterations*/
pistarL=savepi[+,4]/&iter; /*new lower bound of pi star is the average
of 1000 iterations*/
b=j(&iter,1,0);
empirank=j(&iter,1,0);
b=savepi[,3];
empirank=b;
b[rank(b),]=empirank;
piempiL=b[50,1]; /*find the 5th percentage number from the 1000 sorted new
pi start array*/
title "SZ rr.3.3.1.1.1.1 cc .3.3.1.1.1.1 &n";
print pistar;
print pistarL;
print piempiL;
print countlzero;
*print countlzero;
filename out 'C:\SAS\SZ\SZ rr.3.3.1.1.1.1 cc .3.3.1.1.1.1 n equal 180';
file out;
  do i=1 to nrow(savepi);

```

```

        do j=1 to ncol(savepi);
            put (savepi[i, j]) 6.3+4 @;    /*output savepi matrix to an external
file*/
        end;
        put;
        end;
    *close file out;

quit;
run;

```

```

/* Sampling zero and replacing with different flattening constants
scenario*/
/* This program performs mixture index of fit and 95% lower bound when
encountering zero cell in data table using different sampling zero
techniques*/

filename junk dummy;
proc printto log=junk; run;
dm 'log;clear;out;clear;';
options ps=50 nodate nonumber formdlm=' ';

%macro Rudas1(nopercell,replwith, rr, cc, iter);
proc printto log=junk; run;
proc iml;count0=0;noofzero=0;nooflzero=0;
savepi=J(&iter,2,9);
rsum={.3,
      .3,
      .1,
      .1,
      .1,
      .1};
csum={.3 .3 .1 .1 .1 .1}; /* set row and column marginal distributions*/
actual1=rsum*csum;
do ii=1 to &iter;
count=j(&rr, &cc, 0);
n=&nopercell*&rr*&cc;
cum=cusum(actual1); count0=0; /*calculate cumulative proportions*/
do m=1 to n;
    random=round(uniform(1000), .001); /* Generate n=sample size Uniform
distributed proportions between 0 and 1, rounded to .001*/
    do i=1 to &rr;
        do j=1 to &cc;
/* Allocate cell proportions to associated categories according to
cumulative distributions*/
            if random<cum[i,j] then
            do;
                count[i, j]=count[i, j]+1;
                j=&cc;
                i=&rr;
            end;
        end;
    end;
end;
do i=1 to &rr;
    do j=1 to &cc;
        if count[i, j]=0 then do;
            count[i, j]=&replwith; /*if zero, then replace with designated
flattening constants*/
            count0=count0+1; /*count number of zeros*/
            *j=&cc;
            *i=&rr;
        end;
    end;
end;
if count0>0 then noofzero=noofzero+1;
if count0=1 then nooflzero=nooflzero+1;
actual=count/sum(count);

```

```

*print count;
/*print actual;*/
rowsum=count[,+];
coloumsum=count[+,];
e=(rowsum*coloumsum)/sum(count)**2;
do i=1 to 100; /* increment =.01 each iteration */
Pi=i/100;
    Q2=j(&rr,&cc,i/(&rr*&cc*100)); /*initial Q1*/
    Q1=e*(1-i/100); /*initial Q1*/
    QQ2=j(&rr,&cc,i/(&rr*&cc*100)); /*initial QQ2*/
    g1=j(&rr,&cc,0);
    g2=j(&rr,&cc,0);
    d=j(1,1,1);
do while (sum(d)>.00001); /*convergence criteria*/
    do k=1 to &rr by 1;
        do j=1 to &cc by 1;
            g1[k, j]= actual[k,j]*Q1[k, j]/(Q1[k,j]+Q2[k, j]);
            g2[k, j]= actual[k,j]*Q2[k, j]/(Q1[k,j]+Q2[k, j]);
        end;
    end;
do k=1 to &rr;
    do j=1 to &cc;
        g1[k, j]=(1-i/100)*g1[k, j]/sum(g1);
        QQ2[k, j]=i/100*g2[k, j]/sum(g2);
    end;
end;
    QQ1=(g1[,+]*g1[+,])/ (1-i/100);
    QQ1=(1-i/100)*QQ1/sum(QQ1);
    d=sum(abs((abs(QQ1+QQ2)*sum(count)-abs(Q2+Q1)*sum(count))));
    ff=(Q1+Q2)*sum(count); /*expected frequency*/
    /*print d;*/
    Q1=QQ1;
    Q2=QQ2;
    /*PRINT Q1;
    PRINT actual;*/
end;
/*print Q1;
print Q2;*/
/*print ff;*/
dd=sum(abs(count-ff))/2/sum(count);
do m=1 to &rr;
    do n=1 to &cc;
        if count[m,n]=0 | ff[m,n]=0 then G2[m,n]=0;
        else G2[m,n]=(count[m,n])*(LOG(count[m,n])-LOG(ff[m,n]));
/*calculate G-square statistic*/
    end;
end;
Gsqu=2*sum(G2);
chnew=sum((ff-count)**2/ff); /*calculate chi-square statistic*/

    If Gsqu< 2.71& pi<savepi[ii,2]then savepi[ii,2]=pi; /*generate 95%
lower bound of mixture index of fit and save into second column of matrix
savepi*/
        If Gsqu<.00001 then
do;
savepi[ii,1]=pi; /* generate mixture index of fit and save into first column
of matrix savepi*/

```

```

/*print chinew;*/
i=100;
end;
    end;
    end;
*print savepi;
pistar=savepi[+,1]/&iter; /*pi star is the average is 1000 iterations*/

pistarL=savepi[+,2]/&iter; /*pi star lower bound is the average is 1000
iterations*/
b=j(&iter,1,0);
empirank=j(&iter,1,0);
b=savepi[,1];
empirank=b;
b[rank(b),]=empirank;
title "rr.3.3.1.1.1.1 cc.3.3.1.1.1.1 nopercell &nopercell repl withh
&replwith iter &iter";
piempiL=b[50,1]; /*find the 5th percentage number from the 1000 sorted new
pi start array*/
print pistar;
print pistarL;
print piempiL;print noofzero;print nooflzero;
*print count0;

filename out "C:\sas\sz\sz rr.3.3.1.1.1.1 cc.3.3.1.1.1.1 nopercell
&nopercell rep &replwith iter &iter";
file out;
    do i=1 to nrow(savepi);
        do j=1 to ncol(savepi);
            put (savepi[i, j]) 6.2 +3 @; /*output savepi matrix to an external
file*/
        end;
        put;
    end;
*close file out;
quit;
run;
proc printto;run;

%mend Rudas1;
%Rudas1(5, 0, 6, 6, 1000)
%Rudas1(5, 0.1, 6, 6, 1000)
%Rudas1(5, 0.5, 6, 6, 1000)
%Rudas1(5, 1, 6, 6, 1000)
%Rudas1(10, 0, 6, 6, 1000)
%Rudas1(10, 0.1, 6, 6, 1000)
%Rudas1(10, 0.5, 6, 6, 1000)
%Rudas1(10, 1, 6, 6, 1000)
%Rudas1(20, 0, 6, 6, 1000)
%Rudas1(20, 0.1, 6, 6, 1000)
%Rudas1(20, 0.5, 6, 6, 1000)
%Rudas1(20, 1, 6, 6, 1000)
%Rudas1(30, 0, 6, 6, 1000)
%Rudas1(30, 0.1, 6, 6, 1000)
%Rudas1(30, 0.5, 6, 6, 1000)
%Rudas1(30, 1, 6, 6, 1000)

```

REFERENCES

- Clogg, C. C., Rudas, T., Xi, L. (1995) A new index of structure for the analysis of models for mobility tables and other cross-classifications. In P. Marsden (ed.) *Sociological Methodology 1995*, 197-222, Blackwell, Oxford.
- Clogg C. C., Rudas T., and Matthews S. (1997) Analysis of model misfit, structure, and local structure in contingency tables using graphical displays based on the mixture index of fit. In Blasius, J., Greenacre, M. (Eds.) *Visualization of Categorical Data*, 425-439, Academic Press.
- Dayton, C. M. (1999). *Latent class scaling analysis*. Thousand Oaks, CA: Sage.
- Dayton, C. M. (2003), Applications and computational strategies for the two-point mixture index of fit. *British Journal of Mathematical and Statistical Psychology*, 56, 1-13.
- Diaconis, P. and Efron, B. (1985) Testing for independence in a two-way contingency table: new interpretations of the chi-square statistic. *The Annals of Statistics*, 13, 845-874.
- Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977) Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Society. B*, 39, 1-22.
- Fleiss, Joseph L. *Statistical Methods for Rates and Proportions*, 2nd Ed. New York: John Wiley & Sons, 1981.
- Fienberg, S. (1980). *The analysis of cross-classified categorical data*, 2nd ed. Boston, MA: MIT Press.

- Knoke, D. and P.J. Burke (1980). Log-Linear Models. Sage Publications, Inc. Newberry Park, California.
- National Highway Traffic Safety Administration. (2006). Traffic safety facts, 2004. Report no. DOT HS-809-919. Washington, DC: US Department of Transportation.
- Rudas, T. (1998). The mixture index of fit. in Ferligoj, A. (ed.) Advances in Methodology, Data Analysis and Statistics, 15-22, University of Ljubljana.
- Rudas, T. (1999). The mixture index of fit and minimax regression. *Metrika*, 50, 163-172.
- Rudas, T., Clogg, C. C., Lindsay, B. G. (1994) A new index of fit based on mixture methods for the analysis of contingency tables. *Journal of the Royal Statistical Society*, B, 56, No 4, 623-639.
- Rudas, T., Zwick, R. (1997) Estimating the importance of differential item functioning. *Journal of Educational and Behavioral Statistics*, 22, 31-45.
- SAS Institute Inc. (2005). SAS User's Guide, Version 9. Cary, NC: SAS Institute Inc.
- Snee, R. (1974) Graphical display of two-way contingency tables. *The American Statistician*. 28, 9-12.
- Verdes, E., Rudas, T. (2002). The π^* index as a new alternative for assessing goodness of fit of logistic regression. In: Haitovsky, y., Lerche, H. R., Ritov, y. (eds.) *Foundations of statistical inference*. Springer, 167-177.
- Xi, L (1994). The mixture index of fit for the independence model in contingency tables. Master of Arts paper, Department of Statistics, Pennsylvania State University.

- Xi, L., & Lindsay, B. G. (1996). A note on calculating the π^* index of fit for the analysis of contingency tables. *Sociological Methods & Research*, 25, 248-259.
- Zwick, R., Rhayer, D. T., & Wingersky, M. (1994). A simulation study of the methods for assessing differential item functioning in computerized adaptive tests. *Applied Psychological Measurement*, 18, 121-140.