Global Journals LaTeX JournalKaleidoscope™
Artificial Intelligence formulated this projection for compatibility purposes from the original article published at Global Journals. However, this technology is currently in beta. *Therefore, kindly ignore odd layouts, missed formulae, text, tables, or figures.*

Simulation-Based Evaluation of Two-Sample Data in Presence of Less-Than-Detectable(LTD) Data in the Classical Domain

Amaresh Das[1]

[1] SOUTHERN UNIVERSITY AT NEW ORLEANS

Received: 15 June 2015 Accepted: 1 July 2015 Published: 15 July 2015

Abstract

A number of methods available for comparing two samples with censored data is evaluated through a simulation-based exercise. The (Geweke-Hajivassilion-Kenne (GHK) simulator is used here. All the methods discussed here can handle the case when there are multiple detection limits. Under the conditions considered in the simulation, the Mann-Whitney/ Wilcoxon method is best in maintain the Type 1 error rate, while still providing sufficient power.

Index terms— censored data, hypothesis testing, parametric and nonparametric methods.

1 Introduction

imited dependent variable modeled are designed to handle samples that have been censored in some way. 1

2 mortality rate

It is common in environmental data analysis to deal with censored data. Censored data most commonly arise in these situations through laboratory analysis of samples with contaminant concentrations that are less than what the analytical method is able to detect reliably. Or suppose a study is conducted to measure the impact of a drug on . In such a study, it may be known that an individual's age at death is at least 75 years. Such a situation could occur if the individual withdrew from the study at age 75, Censoring is a condition in which the value of a measurement or observation is not fully . observed, although one can fit linear regression model or apply normal distribution to data with censored values. 2,3 1 These two terms, truncated and censored are easily confused. A sample has been truncated if some observations that should have been there have been systematically excluded from the sample. For example, a sample of households with income under $1000.000 necessarily excludes all households with incomes over that level. It is not a random sample of all households. If the dependent variable is income, or something correlated with income, results using the truncated sample could potentially be misleading. On the other hand, a sample has been censored if no observations have been systematically excluded, but some of the information contained in them has been suppressed. Think of a 'censor' who made people's mail and blacks out certain parts of it. The recipients still get their mail but parts of it are unreadable. 2 A sample is 'randomly' censored when both the number of censored observations and the censoring levels are random outcomes. This type of censoring commonly arises in medical time-to-event studies. A subject who moves away from the study area before the event of interest occurs has a randomly censored value. The outcome for a subject can be modeled as a pair of random variables, (x -c), where x A sample is singly censored (e.g., singly left censored) if there is only one censoring level t. (Technically, left censored data are singly left censored only if all n uncensored observations are greater than or equal to t, and right-censored data are singly right censored only if all n uncensored observations are less than or equal to t. Otherwise, the data are considered to be multiply censored.) Note Information contributed by a single observed value or a single value censored at a detection limit ranging from 0 to10. The population is assumed to be normal with mean 5 and variance 1.The standard error could have been calculated from the negative inverse of the Fisher information matrix given in Peng (2010).

Multiple censoring commonly occurs with environmental data because detection limits can change over time (e.g., because of analytical improvements), or detection limits can depend on the type of sample or the is the random time to the event and C is the random time until the subject moves away. x is an observed value if x<c and right censored at c if x>c. 3 It is also tempting to use a t-distribution instead of a normal distribution. This is not supported by statistical theory. The derivation of the t distribution is based on independent estimates of the mean and variance. When some observations are censored, the estimated mean and estimated variance are correlated. The magnitude of the correlation depends on the method used to estimate the parameters, the sample size and the number of censored observations. At best, using a t-distribution to calculate a confidence intervalis an ad-hoc method.

background matrix. The distinction between single and multiple censoring is mostly of historical interest. Some older statistical methods are specifically for singly censored samples. Most currently recommended methods can be used with either singly or multiply censored samples, but the implementation is often easier with one censoring level.

There has been a great deal of literature on the subject of estimating population parameters in the presence of censored data. See, for example, Statistical methods for dealing with censored data have a long history in the field of survival analysis and life testing ((Miller, (1981); EPA **??**2005). However a common simulation that has not been addressed adequately in the literature is when two samples are compared for equality of centrality. Two sample comparisons are frequently made in environmental studies. For example, it is common in all studies to compare site metal concentration to background concentration. In groundwater sampling or air monitoring, samples upstream of a suspected source are compared with sample downstream.

The purpose of the paper is to compare a number of techniques used in two sample hypothesis testing. The techniques considered are the Mann/-Whitney/ Wilcox on rank sum test, the Prentice test, the two-sample t-test using simple replacement of lessthan-detectable data by one-half the detection limit and the ML test of a linear model. The methods are compared through simulations However, we begin with a brief review of methods for two sample comparisons in the presence of censored data.

3 II. Two-Sample Comparison

There are a number of techniques available for comparing two samples. They can be broken down into two categories: nonparametric and parametric methods. The nonparametric methods include linear rank statistics, quantile tests, survival analysis techniques and EM algorithm analogues 4 4 It was found that for data sets with less than 70% censored data, the best technique overall for determination of summary statistics was the nonparametric Kaplan-Meier technique, .ROS (robust Order Statistics) and the two substitution methods of assigning one half the detection limit value to censored data or assigning a random number between zero and the detection limit to censored data were adequate alternatives.. The technique of employing all instrument-generated data including numbers below the detection limits was found to be less adequate than the above techniques. At high degrees of censoring (greater than 70% censored data), no technique provided good estimates of summary statistics. Maximum likelihood techniques were found to be far inferior to all other treatments except substituting zero or the detection limit value to censored data.

The parametric methods include t-test, survival analysis techniques discussed by Richards (2012), Singh and Mukhopadhyay(2011), Collet (2003). Rausand (2004) proposed a combination of ROS and likelihood ideas that they called \robust MLE" for log normal data. That is to use maximum likelihood to estimate the mean and standard deviation from log transformed values, impute log-scale values for each censored observation using the MLE's of the mean and standard deviation and exponentiate those imputed values to get imputations on the data scale. Finally, calculate standard deviation using the observed and ted values.ques and maximum likelihood methods. 5 , ,6 ? Linear rank statistics from a general class of methods that involve tests based on linear combinations of the ranks of the two samples. The Mann-Whitney/ Wilcoxon rank sum is an example of a linear rank test statistic. The application of these in the presence of censored data is discussed in Peng (2010).

Quantiles tests involve specific applications of contingency table analysis. The median test is an example of a quantile test. They can be used to test for bimodality in concentration distribution than may correspond to site contamination. The tests are generally not affected by the presence of censored data. There are methods for estimating and constructing confidence intervals for population quantiles or percentiles. The general equation for a 100 **??**1 -) confidence interval for a parameter µ ?using a normal approximation? µ ? ? ?-i < ? µ ? ? - ?-i +

and plugging in the estimated mean and standard error of the mean.

For a normal distribution, the estimated quantiles are functions of the estimated mean and standard deviation. For a lognormal distribution, they are functions of the estimated mean and standard deviation based on the log-transformed observations In the presence of censored observations, population percentiles are estimated using the same formulae used for uncensored data, but the mean and standard error are estimated using censored data formulae. For example, the maximum likelihood estimate of the p'th percentile of a log-normal distribution is where µ ?and ? ?are the ML estimates of µ and ? on the log scale, and p z ? ?is the p'th percentile of a standard normal distribution. The same plug-in approach could be used 5 The mean, µ , ? coefficient of variation, , and standard deviation x ? µ of a lognormal distribution are functions of µ and ? ., By the invariance property of MLE's, the MLE's of µ ,? , and ? are those functions of the MLE's of µ and ? . If some observations are censored, the MLE's of µ and ? are calculated by maximizing the log-likelihood function for censored data, 6 Richards

(2012) proposed a combination of ROS and likelihood ideas that they called \robust MLE" for log normal data. That is to use maximum likelihood to estimate the mean and standard deviation from log transformed values. Impute log-scale values for each censored observation using the MLE's of the mean and standard deviation and exponentiate those imputed values to get imputations on the data scale. Finally, calculate the mean and standard deviation using the observed and the imputed values.

with Kaplan-Meier (KM) or regression on order statistics (ROS) estimates of µ and ? but the estimated percentiles are no longer ML estimates. Very little statistical research has been done on constructing confidence bounds for percentiles when some observations are censored (Helsel (2005))... Survival analysis method are characteristically devoted to handle censored data. However, in survival analysis, the data are typically censored on the right, whereas in environmental research, the data are usually censored on the left. Many of the survival analysis methods can be modified to handle left-censored data. Miller(!981) discusses the application of survival analysis methods to two-sample data.

The two sample situation can be modified by a one way design model in a regression setting. The test for differences between groups can either be implemented as a t-test of the appropriate model coefficient or as an F test of the mean square. In the presence of censored data, maximum likelihood techniques can be used to estimate model parameters.

The EM algorithm provides an iteration solution to the maximum problem. 7 ? Do not use Kaplan-Meier for data with a single detection limit smaller than the smallest observed value. In this situation, Kaplan-Meier is substitution in disguise.

The EM algorithm is discussed and its application to linear models is considered by Aitken (1981) and Wolynetz (1979). The method is not truly the EM algorithm, nor does it provide a maximum likelihhod estimate. Non parametric analogues of the EM algorithm are discussed in **??**chneider and Weissfeld (1986). Tthere have been many studies of the performance of various estimators of the mean and standard deviation of data with below-detection limit observations. Helsel (2012,) summarizes 15 studies and mentions four more. Our interpretation of those studies leads to recommendations similar to Helsel's (2005):

? For small-moderate amounts of censoring (e.g. $< 50\%$), use Robust Order Statistics or Kaplan-Meier, if multiple censoring limits.

? For moderate-large amounts of censoring (e.g. 50% -80%) and small sample sizes (e.g. < 50), use robust ML.

? For very large amounts of censoring (e.g. $> 80\%$), don't try to estimate mean or standard deviation 7 Finding a maximum likelihood solution typically requires taking the derivatives of the likelihood function with respect to all the unknown values viz. the parameters and the latent variables -and simultaneously solving the resulting equations. In statistical models with latent variables, this usually is not possible. Instead, the result is typically a set of interlocking equations in which the solution to the parameters requires the values of the latent variables and vice versa, but substituting one set of equations into the other produces an unsolvable equation.

unless you are extremely sure of the appropriate distribution. Then use ML.

III.

4 Tests Considered

Four methods are used in simulations to assess their ability to test two sample hypotheses in the presence of censored data. Two of test methods are nonparametric, the MWW and the Prentice test. The parametric methods considered are the two sample t test with simple replacement of LTD values by one half the detection limit and the Wolynetz interpretation of the BM algorithm for linear models.

5 Mann-Whitney Wilcoxon Test :

The Mann-Whitney rank sum test also called Wilcoxon test, ranks all the data in both samples and then sums the ranks within each sample. The greater the difference between sums of the ranks, the more likely the two samples have different medians. This test will be denoted by MWW test for the remainder of our paper. The censored values are considered ties and assigned the average rank. MWW tests for equality of the medians between two groups. It assumes that the dispersion is the same in the two groups.

Prentice Test : We have the name Prentice Test to denote the linear rank test in which the ranks are transformed in normal scores. Under the null hypothesis of equal medians, this test does not depend upon the assumed distribution. However, when the null hypothesis is false and there is no censoring, this test is more powerful than the MWW when the two distributions are normal. As with the MWW test, the censored values are considered ties and assigned the average rank. The Prentice test also assumes homogeneous variance.

6 Two-Sample t test :

The two-sample t-test is one of the most commonly used methods for testing the equality of means between two groups when there variances are equal.. In the presence of less-than-detectable data, the censored values are typically replaced with a value between 0 and the detection limit; commonly the replacement value is one-half the detection limit. This test will be denoted by DL/2.

7 Wolynetz's EM algorithm :

The EM algorithm is an iterative method for finding a maximum likelihood estimate. Wolynetz (1979) develops the algorithm for the linear model. However the implementation is not truly the EM algorithm as it replaces individual censored values rather than sufficient statistics as is done in the EM algorithm Thus, it is not clear whether the method results in an MLE. The method involves first replacing the censored values with, say, the detection limit. Then the model parameters are estimated. With these estimates, replacement values for the censored data are recomputed using a maximum likelihood procedure. The procedure iterates between the two steps to convergence.

The linear model used isij i 1 0 ij x y ? ? ? + + =

where i = 0 , 1, j = 1. . . n x i = n and n i is the number of observations in group i. The ij ? are assumed to be independently and identically normally distributed with mean 0 and variance 2 ? To test the equality of group means the treatment effect parameter, i ? is tested to see if it is significantly different from zero. This test will be denoted by WEM.

IV.

8 Simulation Methodology

The simulation used here was first used by Geweke (1989)and later used by Hajivassilion et al **??**1996) and **??**eaane (1993). It is sometimes called GHK simulator. The idea is to directly approximate the probability of a rectangle.. To simplify the presentation, we first consider the bi dimensional case. We have to estimate the probability of a rectangular domainp [D v ?] = p (v ? [1 b , a 1 1] x [2 2 b a ,] where v N ? (0, ?)

We first transform the random term v to get a random vector with a standard normal distribution. The transformation may be chosen as a lower triangular matrix. P[v D ?]] b v a , b v a [p 2 2 1 1 1 1 < < < < =)] b a a a , b a a [p? ? ? ? ? ? < + < < < µ µ µ 11 1 =P[1 a ? µ < < 1 1 , 2 1 2 a ? µ ? µ < + < 2 , say = p [* D ? µ]

In the µ space the domain D * has the form shown in the following figure. [v D ?] is)] - (-) ([ )] (-) ([ ) (p ?* 1 2 * 1 2 1 1 * 1 µ ? ? ? µ ? ? ? ? ? ? ? µ = Indeed we have E) (p ?* 1 µ =) (-) ( ) ( ) (p ? 1 1 * 1] 1 1 [* 1 ? ? ? ? µ ? µ ? ? ? d * 1 µ = ,] 1 , 1 [? ? ?) ( )] - (-) ([* 1 * 1 2 * 2 µ ? µ ? ? ? µ ? ? ? d * 1 µ = p [(D) , * 2 ? µ µ 1 µ

has not been used but it has been introduced in order to prepare the general case.

A random sample of size 20 is selected from a standard normal distribution with 1 , 0 = = ? µ = 0 .The first ten pseudo numbers are assigned to the first group. The second ten their values shifted by addition of the mean of the second group, which took values 0, 0.05 and 1.0; these ten numbers are assigned to the second group. Hence the mean of the first group, 0 µ is always zero and the mean of the second group 1 µ , could take the values 0, 0.5 or 1.0. The variance is constant within groups to meet the assumptions of the tests A set of simulations for each pair of means is created and analyzed at 20% censoring, with another set is simultaneously analyzed at 60 % censoring. The censoring points were determined from the joint density function of the two groups. For each censoring level, a single censoring point is calculated. Any observation falling below that point was censored. If the censoring resulted in either zero or one uncensored value in the entire data set, the sample is not analyzed. The samples were assumed to be from a log-normal distribution, with the values log-transformed to the normal distribution. This assumption has two purposes: First, it allows us to compute one-half the detection limit as DL -log 2, where DL is the detection limit on the normal scale. Second, for the parametric methods, which are testing means on the normal scale, the analogy is to testing medians on the log-normal scale, as the nonparametric methods would do.

The null hypothesis is that of equal medians, the alternative hypothesis is that the second group median is larger than the first. Hence, all tests are against a onesided alternative. All tests are made with 0.05 = ?

. This one-sided alternative is used to mimic the common practice GHK in monitoring of comparing contaminated areas to background. In these circumstances, it is generally, not of interest to test if the concentration in contaminated areas is less than background.

V.

9 Results & Renarks

Table **??** summarizes the results of the simulation. When0 = 1 µ

, the null hypothesis is true and the power is Type 1 error. As can be seen from the does not maintain the ? level at 20 % censoring, the Prentice test is too small at both censoring levels. When 1 µ is greater than zero, the tests with highest power typically are the tests with inflated Type 1 errorrates. The power of the MWW tests is nearly as high as the most powerful tests under most situations. The DL/2method overall does have higher power than the MWW test, yet this power is partially gained by sacrificing the Type 1 error rate at 60 % censoring.

Table **??** shows the bias and the mean squared error (MSE) of the estimates of the model) parameters using the DL/2 and WEM methods. In general, the WEM method has less biased estimates of the parameters, but larger MSE than the DL/2 method. All the methods discussed here .can handle the case when there are multiple

detection .For those who prefer confidence intervals to hypothesis testing, the DL/2 and WEM methods can be used, with coverage results that are expected to be comparable to the power results presented here.

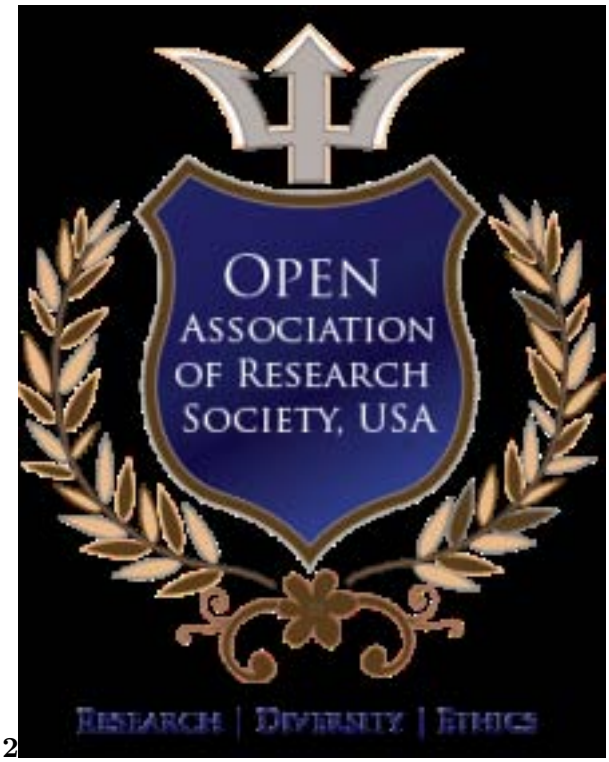
If the sample size is sufficiently large for the sampling distribution of the estimated mean to be close to a normal distribution, one can construct an appropriate confidence interval. Peng (2010) proposed using the bootstrap sampling distribution to assess whether the sample size is sufficiently large.

If the likelihood function is parameterized in terms of the arithmetic mean, the standard error of the estimated mean can be obtained from the negative inverse of the Hessian matrix. However, to obtain an appropriate confidence interval, the sample size needs to be sufficiently large that the distribution of the estimated arithmetic mean is sufficiently close to a normal distribution. If the sample size is large, the difference between quantiles of the normal distribution and quantiles of t distributions is small, so the choice of degrees of freedom is less important. This normal approximation is not recommended when the sample size is small, e.g. n =10 or 20, because the empirical coverage of confidence intervals constructed using will be much smaller than nominal Peng (2010) developed a delta-method approximation to the variance of the arithmetic mean, which can be combined with constructing a confidence interval.

Again, when the sample size is small, e.g. n =10 or 20, the empirical coverage of delta-method confidence Interval is much smaller than nominal (Singh et al(2002) and then this method is not recommended.

10 Table 1 : Summary of the Two -sample Comparison Results

The true mean of the first group 0 µ is zero, the true mean of the second group, 1 µ is shown in the left-most column) 0 (µ ? = 0) 0 - 1 (µ µ ? = 1 1 µ LTD Method N POWER([1]



2

Figure 1: 2 µ

,

only the MWW test maintain the level ?		at
		0.05
for both censoring levels. The DL/2 test nearly maintains		
the level ?		for both censoring levels. The WEM test

Figure 2: Table ,

10 TABLE 1 : SUMMARY OF THE TWO -SAMPLE COMPARISON RESULTS

1.0	20	DL/2	3900	0.6421 (0.0071)	-.0031	.0052.2011
		WEM	3820	0.72330(0.0031).	.0043	.0051.2345
		MWW	1000	0.66332(0.0070)		
		Prentice	1000	0.29451(0.0231)		
0.5	50	DL/2	3888	0.2531(0.3023)	.2141	.0001.0820
		WEM	3780	0.2140(0.3421)	.0361	.0062.2873
		MWW	1000	0.3400(0.2300)		
		Prentice	1000	0.1800(0.2300)		
1.0	20	DL/2	3900	0.6421 (0.0071)	-.0031	.0052.2011
		WEM	3820	0.72330(0.0031).	.0043	.0051.2345
		MWW	1000	0.66332(0.0070)		
		Prentice	1000	0.29451(0.0231)		

Table 1 continued . . .

1 µ	LTD	Method	N	POWER (SE)		
0.0	20	DL/2	4000	0.0467 (0.0020)		
		WEM	4000	0.0713((0.0044)	.	
		MWW	3800	0.0601 (0.0020)		
		Prentice	3800	0.0503 (0.0101)		
0.0	50	DL/2	3980	SE) 0.0500 (0.0000) Bias	SE	MSE
		WEM	3981	0.2500 (0.0011)		

0.0	20	DL/2	4000	MWW	0.0467 (0.0020)	3800	0.0601 (0.0001) -0.011	.0001.0102	
		WEM	4000	Prentice	3800	0.0713((0.0044)	0.0402 (0.0101) .0041	.0002.0102	
0.5	20	MWW	3800	DL/2	0.0601 (0.0020)	4000	0.2441(0.0031)		
		Prentice	3800	WEM	0.0503 (0 .0101)	4000	0.3144(0.0051)		
				MWW	1000		0.4311(0.0231)		
0.0	50	DL/2	3980	Prentice	2000	0.0500 (0.0000)	0.4420(0.0211) 0812	.0024.0831	
0.5	50	WEM	3981	DL/2	0.2500 (0.0011)	3888	0.2531(0.3023) -0231	.0084.2351	
		MWW	3800	WEM	0.0601 (0.0001)	3780	0.2140(0.3421)		
		Prentice	3800	MWW	0.0402 (0.0101)	1000	0.3400(0.2300)		
		-		Prentice	1000		0.1800(0.2300)		
0.5 1.0	20 20	DL/2	4000	DL? 2	0.2441(0.0031)	3888	0.7134(0.0073) -0731	.0003.1003	
		WEM	4000	WEM	0.3144(0.0051)	4000	0.7532(0.0051) -0313	.0013.0001	
		MWW	1000	MWW	0.4311(0.0231)	1000	0.7612(0.0312)		
		Prentice	2000	Prentice	2000	0.4420 (0.0211)	0.56810(.0250)		
1.0	50	DL/2	3890		0.6543(0.0081)				
0.5	50	DL/2	3888	WEM	0.2531(0.3023)	4000	0.6514(0.0077) .2141	.0001.0820	
		WEM	3780	MWW	0.2140(0.3421)	1000	0.7413(0.0054) .0361	.0062.2873	
		MWW	1000	Prentice	1000	0.3400(0.2300)	0.5137(0.0134) -		
		Prentice	1000		0.1800(0.2300)				

Figure 3:

[1]Simulation-Based Evaluation of Two-Sample Data in Presence of Less-Than-Detectable(LTD) Data in the Classical Domain

10 TABLE 1 : SUMMARY OF THE TWO -SAMPLE COMPARISON RESULTS

[Miller R G ()] , Jr Miller R G . 1981. NY: Survival Analysis John Wiley & Sons.

[Richards ()] *A Handbook of Parametric Survival Models for Actuarial Use*, S J Richards . 2012. Scandinavian Actuarial Journal2012. p. .

[Aitken ()] 'A Note on the Regression Analysis of Censored Data'. M Aitken . *Technometrics* 1981. 23 (2) p. .

[Wolynetz ()] 'Algorithm AS 139, Maximum Likelihood Estimation in a Linear Model from Confined and Censored Normal Data'. M Wolynetz . *Applied Statistics* 1979. 28 p. 195.

[Geweke ()] *Efficient Simulation from the Multivariate Normal Distribution Subject to Linear Inequality Constraints and the Evolution of Year*, J Geweke . 1989. 2015.

[Schneider and Weisifeld ()] 'Estimation in Linear Models with Censored Data'. H Schneider , L Weisifeld . *Biometrika* 1986. (73) p. .

[Gillion and Helsel ()] 'Estimation of Distribution Parameters for Censored Trace Level Water Quality Data -Estimation Techniques'. R Gillion , D K Helsel . *Water Resources Research* 1986. (22) p. .

[Peng ()] 'Interval Estimation of Population Parameters Based on Environmental Data with DetectionLimits'. C Peng . *Environmetrics* 2010. (21) p. .

[Collett ()] *Modelling Survival Data in Medical Research*, David Collett . 2003. Boca Raton: Chapman & Hall. (Second ed.)

[Rausand and Hoyland ()] M Rausand , A Hoyland . *System Reliability Theory: Models, Statistical Methods, and Applications*, 2004. John Wiley & Sons.

[Hjivassiliou (ed.) ()] *Simulation Estimation for Panel Data Models with Limited Dependent Variable Models' in Handbook of Statistics in*, Hjivassiliou . G. S. Maddala, C. R. Rao and H. Vinod (ed.) 1993. North Holland, Amsterdam.

[Keane (ed.) ()] *Simulation Estimation for Panel Data Models with Limited Dependent Variable Models' in Handbook of Statistics in*, M Keane . G. S. Maddala, C. R. Rao and H. Vinod (ed.) 1993. North Holland, Amsterdam.

[Singh et al. ()] *Statistical Approaches to Estimate Mean Water Quality Concentrations' Envvironmental Science and Technology*, A Singh , A Singh , R Laci . 2002. 36 p. .

[Singh and Mukhopadhyay ()] 'Survival Analysis in Clinical Trials: Basics and Must Know Areas'. R Singh , K Mukhopadhyay . *Perspect* 2011. 2 (4) p. .