PROPOSED SHRUNKEN VARIANCE ESTIMATOR FOR SURVIVAL FUNCTION

Qamruz Zaman[1], Muhammad Iqbal[1], Siraj ud Din[2], Fazl-e-Raziq[3] & Habib Nawaz[4]
1Department of Statistics, University of Peshawar.
[2]Department of Business Administration, Gomal University, Dera Ismail Khan
[3]Post Graduate College, Mardan.
[4]Institute of Management Sciences, Bannu University.

ABSTRACT
One of the main purposes of statistics is to estimate parameters, and their reliability. Kaplan-Meier survival function is most commonly used technique for reliability. Several functions variance estimators have been proposed for this purpose, where each has its own advantages and disadvantages. To overcome some of the weaknesses of conventional methods, new variance estimators are proposed. The proposed variance estimators along with their alternative variance estimators were applied in Simulation studies. In addition, the estimators are also applied to the real data sets (Heart Transplant data set and Thallassaemia data set). The performances of the proposed variance estimators are remarkable as compared to its alternative variance estimators.

KEYWORDS: Kaplan-Meier Survival Function, Censored Observation, Variance Estimator

INTRODUCTION

Survival analysis is one of the widely used techniques in medical statistics; its importance also arises in diverse fields such as medicine, engineering, epidemiology, biology, economics, physics, public health and demography. In survival analysis the standard error or variances are computed from the Kaplan-Meier survival function (1958). The commonly used estimators are Greenwood (1926), Peto (1977), Zhao (2005) and Borkowf (2005) Variances. But there are some shortcomings in these estimators, due to which this research work was conducted.

The Greenwood variance estimator is one of the old estimators used for decades. The Greenwood variance estimator tends to underestimate the variability in the tails of the distribution (Peto, 1977). Also if the last observation in a data set is event, the Greenwood estimator becomes infinite. For overcoming this drawback, Peto has recommended an alternative variance estimator for the Kaplan-Meier survival function. But both the Greenwood and Peto variance estimators are only calculated for the events, in case of censoring both fail to calculate.

According to Zhao (1996) and Borkowf (2005), the Greenwood and Peto variance estimators can substantially underestimate the true variance in both left and right tails of the survival distribution, not only in large censored but also with moderately censored data. Peto's variance estimator performs poorly for small sample sizes when S(t) < 0.2 or > 0.8. Also another drawback, according to Borkowf (2005) is that the effective sample size of Peto's variance can numerically exceed the original sample size in certain cases. Especially for first few events and the effective sample size is also not an integer.

The Zhao (1996) also proposed estimators for the Kaplan-Meier survival function and called it homogenetic variance estimator. The homogenetic variance estimator overestimates when there are no ties. Bias in homogenetic variance estimator is less than that of Greenwood and Peto variance estimators (Zhao, 1996). The Borkowf (2005) has proposed estimators for the Kaplan-Meier survival function namely regular hybrid variance estimator. Then shrinking the observed Kaplan-Meier survival function towards 0.5 and proposed variance named it adjusted hybrid variance estimator. The regular hybrid standard error underestimates the target standard deviation at and before median for small sample sizes. For heavy censoring the mean estimated adjusted hybrid standard error curves are less accurate in the right tail of the survival distribution in comparison to simulated standard deviations. For moderate censoring the adjusted hybrid standard error overestimates after median. Furthermore, for heavy censoring all the hybrid variance estimators overestimates for small sample sizes but underestimate for large sample sizes (Borkowf, 2005). For minimizing these shortcomings new estimators are very essential.

The survival time data means the data regarding time to the occurrence of a certain event. Event is predefined and may be a death, failure, relapse, remission or some other response of an individual in a data set. The survival times measures the time to an event. The times are considered from the beginning of the study or from an origin time "t_o", to the occurrence time of a particular event.

In some cases, the analysis of data is normally up to a time after which some subjects may still remains alive. Further some subjects may have moved away due to personal or other reasons, and thus lost in follow up. In both the above cases, the subject's survival is known for some time, known as censoring times. The exact survival times of the subjects are not known therefore are censored. Otherwise, the subjects are considered as an event times. If the event time of an individual is regarded as a random variable *T*, then the survival function *S*(*t*), is defined as the probability that an individual survives until *t* (or beyond) is

$$S(t) = P[T \geq t]$$

Proposed Variance Estimators

The shrunken Kaplan-Meier survival function idea can be used differently, to achieve much better results. We study many different options for the shrunken, and the reliable among them is given by

$$\hat{S}_P(t) = \left(1 - \frac{1}{\sqrt{n}}\right)\hat{S}(t) + \frac{1}{(2n)^2}$$

To estimate a variance estimator for this new survival function, we proceed as

$$Var\{\hat{S}_P(t)\} = Var\left\{\left(1 - \frac{1}{\sqrt{n}}\right)\hat{S}(t) + \frac{1}{(2n)^2}\right\}$$

So,

$$Var\{\hat{S}_P(t)\} = \left(1 - \frac{1}{\sqrt{n}}\right)^2 Var\{\hat{S}(t)\}$$

Therefore, applying Greenwood (1926) concept, the variance of the proposed estimator is

$$\hat{V}ar\{\hat{S}_P(t)\} = (\hat{S}(t))^2\left(1 - \frac{1}{\sqrt{n}}\right)^2 \sum_{j=1}^{k}\left\{\frac{e_j}{n_j(n_j - e_j)}\right\}$$

Another proposed variance estimator for shrunken survival function using Peto's estimator, is given by

$$\hat{V}ar\{\hat{S}_P(t)\} = (\hat{S}(t))^2\left(1 - \frac{1}{\sqrt{n}}\right)^2 \left[1 - \hat{S}(t)\right] \Big/ n_j$$

These proposed estimators will give better results, especially performs better at the tails of the distributions, when the Greenwood and the Peto's variance estimators not performing well. Some times over-estimates at the tails.

SIMULATIONS

Simulation is one of the resampling techniques, widely used in statistics for comparing different estimators. Here the simulation method is used for comparison of survival functions and also for comparison of variance estimators.

Stepwise Simulation Procedure for Standard Errors Comparison:

i) Number of repetitions (500 times)

ii) Random sample of sizes, *n* = 25, 50, 100 and 500 from different survival distributions.

iii) Compute the Standard Error Estimators at the selected points.

iv) The results are saved in matrix form.

For comparison, survival times are selected from Weibull, Exponential and Lognormal distributions, while for the censoring a Uniform distribution is selected. The survival times and censoring times are then combined, to have a sample of sizes (n = 25, 50, 100, 500). Each data set of size n is repeated 500 times.

Here for generating survival times the distributions are Exponential with parameter 0.5, 1.0 and 1.5. The Weibull distribution is used for survival time, with first parameter fixed at one and the second parameter having values 0.5, 1.5. **S**tandard of errors of the four estimators are computed from survival function. Each standard error is divided into three equal parts, that is; quartiles. These three quartiles are computed 500 times and then means are computed for each quartile and compared, the results are shown in Table 1, Table 2 and Figure 1.

Table 1: Mean of the Quarters, computed from the Greenwood, Proposed Greenwood, Peto's and proposed Peto

Standard Errors. Different sample of sizes n = 25, 50, 100, 500, with different censoring percentage levels that is; 15, 25, 45, 65 and 85 are used. For censoring, Uniform distribution is used. While for survival data, Weibull distribution is used.

Survivor	censored	Percent	n	Greenwood SE			Proposed GW SE			Peto SE			Proposed Peto SE		
				MEAN			MEAN			MEAN			MEAN		
				Q1	Q2	Q3	Q1	Q2	Q3	Q1	Q2	Q3	Q1	Q2	Q3
W(1,1.5)	U(0,10.3)	15	25	0.0755	0.0921	0.1005	0.0604	0.0736	0.0804	0.0679	0.0877	0.0974	0.0543	0.0701	0.0779
			50	0.0520	0.0647	0.0711	0.0446	0.0555	0.0610	0.0494	0.0637	0.0707	0.0424	0.0547	0.0607
			100	0.0366	0.0458	0.0503	0.0330	0.0412	0.0452	0.0360	0.0460	0.0508	0.0324	0.0414	0.0457
			500	0.0163	0.0204	0.0225	0.0156	0.0195	0.0214	0.0167	0.0209	0.0230	0.0160	0.0200	0.0220
W(1,1.5)	U(0,6)	25	25	0.0773	0.0949	0.1043	0.0618	0.0759	0.0834	0.0703	0.0908	0.1016	0.0563	0.0726	0.0813
			50	0.0542	0.0673	0.0738	0.0465	0.0578	0.0634	0.0519	0.0669	0.0747	0.0445	0.0574	0.0641
				0.0386	0.0478	0.0522	0.0347	0.0430	0.0470	0.0382	0.0488				

100		0.0540	0.0344	0.0439	0.0486				
		0.0178	0.0216	0.0234	0.0170	0.0206	0.0223	0.0185	0.0227
500		0.0246	0.0177	0.0217	0.0235				
W(1,1.5)	U(0,2.87)	0.0746	0.0987	0.1142	0.0597	0.0789	0.0914	0.0701	0.0958
45	25	0.1130	0.0561	0.0767	0.0904				
		0.0530	0.0709	0.0821	0.0455	0.0609	0.0705	0.0519	0.0720
50		0.0858	0.0446	0.0618	0.0736				
		0.0380	0.0510	0.0589	0.0342	0.0459	0.0530	0.0382	0.0533
100		0.0639	0.0344	0.0480	0.0575				
		0.0190	0.0244	0.0270	0.0182	0.0233	0.0258	0.0197	0.0265
500		0.0309	0.0188	0.0254	0.0296				
W(1,1.5)	U(0,1.43)	0.0616	0.0931	0.1205	0.0493	0.0745	0.0964	0.0590	0.0929
65	25	0.1224	0.0472	0.0743	0.0979				
		0.0460	0.0669	0.0885	0.0395	0.0574	0.0760	0.0460	0.0703
50		0.0968	0.0395	0.0604	0.0831				
		0.0334	0.0484	0.0644	0.0300	0.0435	0.0579	0.0343	0.0525
100		0.0741	0.0308	0.0473	0.0667				
		0.0167	0.0239	0.0312	0.0159	0.0229	0.0298	0.0176	0.0272
500		0.0385	0.0168	0.0260	0.0367				
W(1,1.5)	U(0,0.52)	0.0412	0.0682	0.1045	0.0329	0.0546	0.0836	0.0401	0.0679
85	25	0.1060	0.0320	0.0543	0.0848				
		0.0320	0.0521	0.0781	0.0275	0.0448	0.0670	0.0321	0.0549
50		0.0864	0.0275	0.0471	0.0742				
		0.0235	0.0377	0.0580	0.0212	0.0340	0.0522	0.0242	0.0416
100		0.0686	0.0218	0.0375	0.0617				
		0.0111	0.0177	0.0274	0.0106	0.0169	0.0261	0.0118	0.0205
500		0.0353	0.0113	0.0196	0.0337				
W(1,0.5)	U(0,3.59)	0.0752	0.0917	0.1002	0.0601	0.0734	0.0801	0.0671	0.0868
15	25	0.0964	0.0536	0.0694	0.0771				
		0.0526	0.0648	0.0710	0.0452	0.0556	0.0609	0.0496	0.0635
50		0.0703	0.0425	0.0546	0.0604				
		0.0369	0.0458	0.0502	0.0332	0.0412	0.0451	0.0361	0.0458
100		0.0504	0.0325	0.0412	0.0454				
		0.0154	0.0198	0.0222	0.0147	0.0189	0.0212	0.0157	0.0202
500		0.0226	0.0150	0.0193	0.0216				
W(1,0.5)	U(0,2)	0.0781	0.0951	0.1037	0.0624	0.0760	0.0830	0.0703	0.0905
25	25	0.1008	0.0563	0.0724	0.0806				
		0.0551	0.0678	0.0737	0.0473	0.0582	0.0633	0.0523	0.0671
50		0.0742	0.0449	0.0576	0.0637				
		0.0397	0.0483	0.0523	0.0357	0.0434	0.0470	0.0392	0.0493
100		0.0538	0.0353	0.0443	0.0485				

	0.0178 0.0215 0.0233 0.0170 0.0204 0.0223 0.0186 0.0226
500	0.0244 0.0178 0.0216 0.0233
W(1,0.5) U(0,0.96)	0.0768 0.0995 0.1142 0.0614 0.0796 0.0914 0.0714 0.0961
45 25	0.1128 0.0571 0.0768 0.0902
	0.0551 0.0722 0.0826 0.0473 0.0619 0.0709 0.0536 0.0729
50	0.0861 0.0460 0.0626 0.0739
	0.0403 0.0526 0.0594 0.0363 0.0473 0.0535 0.0403 0.0549
100	0.0643 0.0363 0.0494 0.0578
	0.0212 0.0256 0.0273 0.0203 0.0245 0.0261 0.0220 0.0283
500	0.0315 0.0210 0.0270 0.0300
W(1,0.5) U(0,0.47)	0.0661 0.0949 0.1227 0.0529 0.0759 0.0981 0.0637 0.0947
65 25	0.1236 0.0509 0.0758 0.0988
	0.0485 0.0696 0.0909 0.0416 0.0597 0.0780 0.0487 0.0734
50	0.0987 0.0418 0.0630 0.0847
	0.0356 0.0512 0.0670 0.0320 0.0461 0.0603 0.0366 0.0558
100	0.0767 0.0330 0.0502 0.0691
	0.0179 0.0255 0.0330 0.0171 0.0244 0.0315 0.0190 0.0293
500	0.0409 0.0181 0.0280 0.0390
W(1,0.5) U(0,0.171)	0.0435 0.0726 0.1102 0.0345 0.0581 0.0882 0.0422 0.0727
85 25	0.1107 0.0337 0.0581 0.0885
	0.0345 0.0551 0.0828 0.0296 0.0473 0.0711 0.0347 0.0584
50	0.0911 0.0298 0.0502 0.0782
	0.0251 0.0391 0.0602 0.0226 0.0351 0.0542 0.0259 0.0432
100	0.0713 0.0233 0.0389 0.0642
	0.0113 0.0181 0.0281 0.0108 0.0173 0.0269 0.0119 0.0207
500	0.0356 0.0113 0.0198 0.0340

Table 2: Mean of the Quarters, computed from the Hybrid, Adjusted Hybrid and proposed shrunken Standard Errors. Different sample of sizes *n* = 25, 50, 100, 500, with different censoring percentage levels that is; 15, 25, 45, 65 and 85 are used. For censoring, Uniform distribution is used. While for survival data, Weibull distribution is used.

Survivor	Percent censored	censored n	Hybrid SE MEAN Q1	Hybrid SE MEAN Q2	Hybrid SE MEAN Q3	Adjusted Hybrid SE MEAN Q1	Adjusted Hybrid SE MEAN Q2	Adjusted Hybrid SE MEAN Q3	Proposed Shrunken SE MEAN Q1	Proposed Shrunken SE MEAN Q2	Proposed Shrunken SE MEAN Q3
W(1,1.5) U(0,10.3)	15	25	0.0806	0.0953	0.1026	0.0827	0.0961	0.1028	0.0892	0.0975	0.1021
		50	0.0569	0.0671	0.0723	0.0576	0.0674	0.0724	0.0612	0.0682	0.0721
		100	0.0408	0.0477	0.0514	0.0410	0.0478	0.0514	0.0428	0.0482	0.0511
		500	0.0187	0.0214	0.0229	0.0187	0.0214	0.0229	0.0189	0.0214	0.0229
W(1,1.5) U(0,6)	25	25	0.0834	0.1010	0.1078	0.0853	0.1016	0.1080	0.0946	0.1021	0.1062
		50	0.0595	0.0719	0.0764	0.0601	0.0721	0.0764	0.0661	0.0725	0.0755
		100	0.0208	0.0234	0.0242	0.0208	0.0234	0.0242	0.0213	0.0233	0.0241
		500	0.0208	0.0233	0.0242	0.0208	0.0233	0.0242	0.0213	0.0232	0.0241
W(1,1.5) U(0,2.87)	45	25	0.0782	0.1064	0.1224	0.0810	0.1071	0.1225	0.1003	0.1126	0.1207
		50	0.0550	0.0757	0.0871	0.0560	0.0759	0.0871	0.0676	0.0796	0.0864
		100	0.0397	0.0541	0.0618	0.0400	0.0542	0.0618	0.0462	0.0561	0.0615
		500	0.0196	0.0258	0.0281	0.0198	0.0258	0.0281	0.0207	0.0260	0.0280
W(1,1.5) U(0,1.44)	65	25	0.0670	0.1021	0.1349	0.0723	0.1041	0.1354	0.1002	0.1191	0.1396
		50	0.0483	0.0727	0.0964	0.0499	0.0733	0.0965	0.0661	0.0820	0.0988
		100	0.0345	0.0525	0.0691	0.0350	0.0527	0.0691	0.0441	0.0573	0.0702
		500	0.0176	0.0260	0.0325	0.0176	0.0260	0.0325	0.0194	0.0268	0.0326
W(1,1.5) U(0,0.53)	85	25	0.0420	0.0732	0.1199	0.0553	0.0822	0.1263	0.0973	0.1193	0.1579
		50	0.0302	0.0549	0.0902	0.0348	0.0573	0.0916	0.0615	0.0785	0.1075

	0.0229	0.0392	0.0643	0.0241	0.0400	0.0648
100	0.0392	0.0517	0.0734			
	0.0104	0.0180	0.0293	0.0105	0.0180	0.0293
500	0.0142	0.0207	0.0312			
W(1,0.5) U(0,3.57)	0.0791	0.0947	0.1022	0.0813	0.0955	0.1024
15 25	0.0886	0.0971	0.1019			
	0.0560	0.0667	0.0721	0.0568	0.0670	0.0722
50	0.0609	0.0680	0.0720			
	0.0400	0.0473	0.0511	0.0402	0.0474	0.0511
100	0.0424	0.0480	0.0510			
	0.0179	0.0211	0.0228	0.0179	0.0211	0.0228
500	0.0184	0.0213	0.0228			
W(1,0.5) U(0,2)	0.0829	0.1010	0.1083	0.0848	0.1017	0.1085
25 25	0.0946	0.1025	0.1067			
	0.0586	0.0719	0.0766	0.0593	0.0721	0.0766
50	0.0660	0.0726	0.0757			
	0.0413	0.0510	0.0540	0.0414	0.0511	0.0540
100	0.0457	0.0512	0.0536			
	0.0184	0.0229	0.0241	0.0184	0.0229	0.0241
500	0.0195	0.0229	0.0240			
W(1,0.5) U(0,0.96)	0.0773	0.1057	0.1223	0.0801	0.1065	0.1224
45 25	0.0999	0.1126	0.1209			
	0.0541	0.0748	0.0867	0.0551	0.0751	0.0867
50	0.0670	0.0790	0.0860			
	0.0380	0.0530	0.0612	0.0384	0.0530	0.0612
100	0.0451	0.0553	0.0609			
	0.0169	0.0237	0.0274	0.0169	0.0237	0.0274
500	0.0185	0.0242	0.0274			
W(1,0.5) U(0,0.48)	0.0645	0.0994	0.1335	0.0704	0.1016	0.1342
65 25	0.0993	0.1180	0.1391			
	0.0466	0.0709	0.0952	0.0483	0.0716	0.0954
50	0.0652	0.0808	0.0980			
	0.0326	0.0502	0.0676	0.0332	0.0505	0.0677
100	0.0429	0.0556	0.0691			
	0.0146	0.0226	0.0304	0.0146	0.0226	0.0304
500	0.0169	0.0237	0.0307			
W(1,0.5) U(0,0.174)	0.0370	0.0703	0.1198	0.0524	0.0804	0.1265
85 25	0.0969	0.1194	0.1591			
	0.0298	0.0541	0.0897	0.0344	0.0565	0.0912
50	0.0614	0.0783	0.1078			
	0.0229	0.0397	0.0646	0.0241	0.0404	0.0650

100	0.0392 0.0520 0.0735
	0.0104 0.0180 0.0294 0.0105 0.0181 0.0295
500	0.0143 0.0208 0.0313

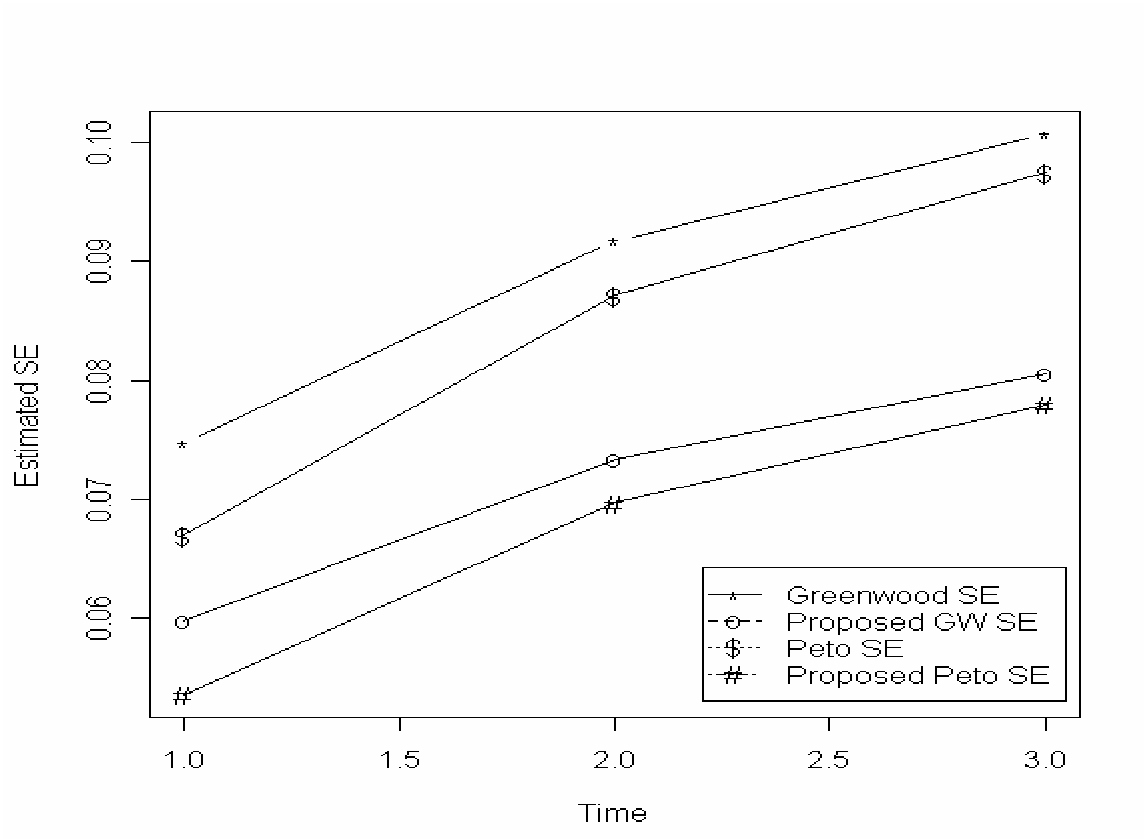

Figure1: Showing the curves for Greenwood, Proposed Greenwood, Peto's and Proposed Peto Standard Error. The three points are the Quartiles selected from the Standard Error.

In the above Tables, we see that when sample size *n* increases, the result for means of all the standard errors estimators becomes smaller as expected. By comparison, at all the censoring levels and for all the sample sizes, the proposed standard errors results are better from both the Greenwood and Peto's standard error. The results of both the proposed estimators are almost the same.

Here the result patterns are same for all censoring levels and for the sample sizes *n* = 25, 50, 100 and 500. As censoring levels and sample sizes increases, results for the means of quartiles are decreases respectively.

In second part, we compare the Hybrid, adjusted hybrid and proposed shrunken standard errors. The means are calculated for the three quartiles of these three estimators (Table 1). In above Tables, we see that when sample size *n* increases, the result for means of all the standard errors estimators become smaller as expected. By comparing, the proposed shrunken standard error results with both the Hybrid and the adjusted hybrid standard errors at 3rd quartile, the results are better from the later. Here the result patterns are same for all the censoring levels and for the sample sizes *n* = 25, 50, 100 and 500. As censoring levels and sample sizes increases, the result for the means of quartiles decreases respectively.

REAL DATA ANALYSIS

In this part of the study, the Stanford Heart Transplant data set (Kalbflesch and Printice, 1980) is reanalyzed. We applied the Greenwood, Peto's standard error methods and compared these conventional methods of standard error with the proposed Greenwood and proposed Peto's standard error. This data set consists of 103 patients and the time observed is in days. 69 patients received transplants; here the censoring rate is 27%.

The result of the proposed Greenwood standard error is significant as compared to the Greenwood standard errors. To confirm, let see at the Table 3 and figure 2, the values at day first for Greenwood standard error is 0.0097 and for proposed Greenwood is 0.0087. We compared the standard error estimators at all the survival time points in a data set. At the last survival time point that is; day 1799, the value for Greenwood standard error is 0.0493, and for Proposed Greenwood is 0.0444. On the other hand, the Proposed Peto's standard error results are significant as compared to the conventional Peto's standard error.

Table: 3 The estimates of Kaplan-Meier and proposed survival functions, along with Standard Errors of Greenwood, proposed Greenwood, Peto's and proposed Peto estimators, applied to the Heart Transplant data set.

				Survival functions		Standard Errors			
time	e_i	c_i	n_i	$\hat{S}(t)$	$\hat{S}_P(t)$	SE_G	SE_{PG}	SE_P	SE_{PP}
1	1	0	103	0.9900	0.89273	0.0097	0.0087	0.0096	0.0086
2	3	0	102	0.9610	0.78112	0.0190	0.0172	0.0187	0.0169
3	3	0	99	0.9320	0.68283	0.0247	0.0224	0.0244	0.0220
5	2	0	96	0.9120	0.60274	0.0278	0.0251	0.0275	0.0248
6	2	0	94	0.8930	0.53181	0.0304	0.0274	0.0301	0.0271
8	1	0	92	0.8830	0.47421	0.0316	0.0285	0.0314	0.0283
9	1	0	91	0.8730	0.42280	0.0327	0.0295	0.0325	0.0293
11	0	1	90	0.8730	0.38115	0.0327	0.0295	0.0325	0.0293
12	1	0	89	0.8630	0.33974	0.0338	0.0305	0.0337	0.0304
16	3	0	88	0.8340	0.29583	0.0367	0.0331	0.0361	0.0326
17	1	0	85	0.8240	0.26355	0.0375	0.0338	0.0374	0.0337
18	1	0	84	0.8140	0.23476	0.0383	0.0346	0.0382	0.0344
21	2	0	83	0.7950	0.20653	0.0398	0.0359	0.0394	0.0356
28	1	0	81	0.7850	0.18389	0.0405	0.0366	0.0404	0.0364
30	1	0	80	0.7750	0.16370	0.0412	0.0372	0.0410	0.0370
31	0	1	79	0.7750	0.14757	0.0412	0.0372	0.0410	0.0370
32	1	0	78	0.7650	0.13133	0.0412	0.0377	0.0419	0.0378
35	1	0	77	0.7550	0.11686	0.0425	0.0383	0.0425	0.0383
36	1	0	76	0.7450	0.10396	0.0430	0.0388	0.0431	0.0388
37	1	0	75	0.7350	0.09247	0.0436	0.0393	0.0436	0.0393
39	1	0	74	0.7250	0.08223	0.0441	0.0398	0.0441	0.0398
39	0	1	73	0.7250	0.07413	0.0441	0.0398	0.0441	0.0398
40	2	0	72	0.7050	0.06497	0.0451	0.0407	0.0451	0.0406
43	1	0	70	0.6950	0.05773	0.0456	0.0412	0.0458	0.0413
45	1	0	69	0.6850	0.05129	0.0460	0.0415	0.0462	0.0417
50	1	0	68	0.6750	0.04556	0.0465	0.0419	0.0466	0.0420
51	1	0	67	0.6650	0.04046	0.0468	0.0423	0.0470	0.0423
53	1	0	66	0.6550	0.03592	0.0472	0.0426	0.0473	0.0426
58	1	0	65	0.6450	0.03188	0.0475	0.0429	0.0476	0.0429
61	1	0	64	0.6350	0.02829	0.0478	0.0432	0.0479	0.0432
66	1	0	63	0.6250	0.02510	0.0481	0.0434	0.0482	0.0434
68	2	0	62	0.6040	0.02190	0.0486	0.0439	0.0482	0.0435
69	1	0	60	0.5940	0.01941	0.0489	0.0441	0.0488	0.0440
72	2	0	59	0.5740	0.01690	0.0492	0.0444	0.0487	0.0439
77	1	0	57	0.5640	0.01497	0.0494	0.0446	0.0493	0.0444
78	1	0	56	0.5540	0.01325	0.0495	0.0447	0.0494	0.0445
80	1	0	55	0.5440	0.01173	0.0496	0.0448	0.0495	0.0446
81	1	0	54	0.5340	0.01038	0.0497	0.0449	0.0496	0.0447
85	1	0	53	0.5240	0.00918	0.0498	0.0449	0.0496	0.0447
90	1	0	52	0.5140	0.00812	0.0499	0.0450	0.0496	0.0448
96	1	0	51	0.5040	0.00717	0.0499	0.0450	0.0497	0.0448
100	1	0	50	0.4940	0.00634	0.0499	0.0450	0.0496	0.0447
102	1	0	49	0.4830	0.00559	0.0499	0.0450	0.0496	0.0447
109	0	1	48	0.4830	0.00504	0.0499	0.0450	0.0496	0.0447

110	1	0	47	0.4730	0.00445	0.0499	0.0450	0.0501	0.0451
131	0	1	46	0.4730	0.00401	0.0499	0.0450	0.0501	0.0451
149	1	0	45	0.4630	0.00353	0.0498	0.0450	0.0505	0.0456
153	1	0	44	0.4520	0.00311	0.0498	0.0450	0.0504	0.0455
165	1	0	43	0.4420	0.00274	0.0498	0.0449	0.0503	0.0453
180	0	1	42	0.4420	0.00247	0.0497	0.0449	0.0503	0.0453
186	1	0	41	0.4310	0.00217	0.0496	0.0448	0.0507	0.0457
188	1	0	40	0.4200	0.00191	0.0495	0.0448	0.0506	0.0456
207	1	0	39	0.4090	0.00168	0.0493	0.0447	0.0504	0.0454
219	1	0	38	0.3980	0.00147	0.0492	0.0445	0.0501	0.0452
263	1	0	37	0.3880	0.00129	0.0492	0.0444	0.0499	0.0449
265	0	1	36	0.3880	0.00116	0.0498	0.0444	0.0499	0.0449
285	2	0	35	0.3650	0.00099	0.0486	0.0440	0.0492	0.0444
308	1	0	33	0.3540	0.00086	0.0483	0.0438	0.0496	0.0447
334	1	0	32	0.3430	0.00075	0.0480	0.0436	0.0492	0.0443
340	1	0	31	0.3320	0.00066	0.0480	0.0433	0.0488	0.0440
340	0	1	30	0.3320	0.00059	0.0487	0.0433	0.0488	0.0440
342	1	0	29	0.3210	0.00051	0.0477	0.0430	0.0491	0.0443
370	0	1	28	0.3210	0.00046	0.0477	0.0430	0.0491	0.0443
397	0	1	27	0.3210	0.00042	0.0477	0.0430	0.0491	0.0443
427	0	1	26	0.3210	0.00037	0.0477	0.0430	0.0491	0.0443
445	0	1	25	0.3210	0.00034	0.0477	0.0430	0.0491	0.0443
482	0	1	24	0.3210	0.00030	0.0477	0.0430	0.0491	0.0443
515	0	1	23	0.3210	0.00027	0.0477	0.0430	0.0491	0.0443
545	0	1	22	0.3210	0.00025	0.0478	0.0430	0.0491	0.0443
583	1	0	21	0.3050	0.00021	0.0478	0.0431	0.0556	0.0501
596	0	1	20	0.3050	0.00019	0.0478	0.0431	0.0556	0.0501
630	0	1	19	0.3050	0.00017	0.0478	0.0431	0.0556	0.0501
670	0	1	18	0.3050	0.00015	0.0472	0.0431	0.0556	0.0501
675	1	0	17	0.2870	0.00013	0.0485	0.0435	0.0589	0.0531
733	1	0	16	0.2690	0.00011	0.0485	0.0437	0.0576	0.0519
841	0	1	15	0.2690	0.00010	0.0487	0.0437	0.0576	0.0519
852	1	0	14	0.2500	0.00008	0.0487	0.0439	0.0579	0.0522
915	0	1	13	0.2500	0.00007	0.0487	0.0439	0.0579	0.0522
941	0	1	12	0.2500	0.00006	0.0483	0.0439	0.0579	0.0522
979	1	0	11	0.2270	0.00005	0.0493	0.0445	0.0603	0.0544
995	1	0	10	0.2050	0.00004	0.0498	0.0445	0.0578	0.0521
1032	1	0	9	0.1820	0.00003	0.0488	0.0441	0.0549	0.0495
1141	0	1	8	0.1820	0.00003	0.0482	0.0441	0.0549	0.0495
1321	0	1	7	0.1820	0.00003	0.0482	0.0441	0.0549	0.0495
1386	1	0	6	0.1510	0.00002	0.0492	0.0444	0.0571	0.0514
1400	0	1	5	0.1510	0.00002	0.0492	0.0444	0.0571	0.0514
1407	0	1	4	0.1510	0.00001	0.0492	0.0444	0.0571	0.0518
1571	0	1	3	0.1510	0.00001	0.0492	0.0444	0.0571	0.0514
1586	0	1	2	0.1510	0.00001	0.0492	0.0444	0.0571	0.0514
1799	0	1	1	0.1510	0.00001	0.0492	0.0444	0.0571	0.0514

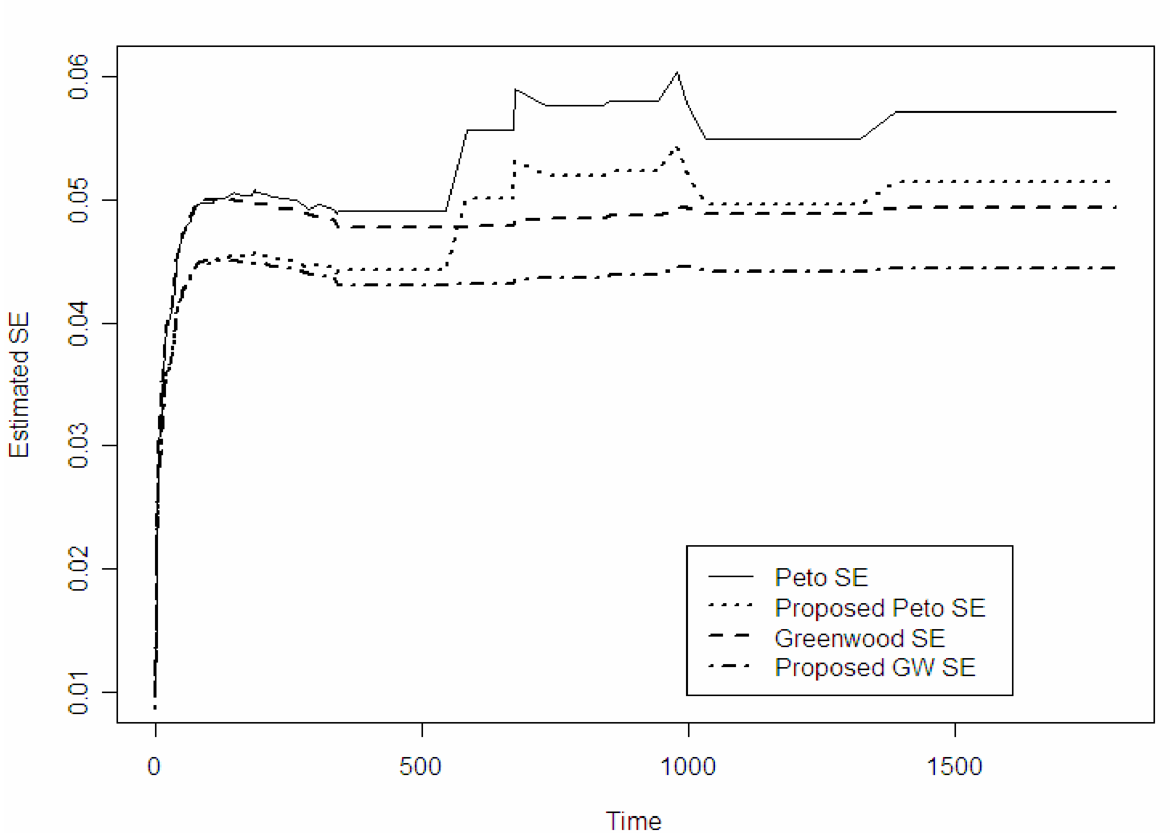

Figure 2: Showing the Standard Errors of Peto, Proposed Peto, Greenwood and Proposed Greenwood Estimators. The Estimators are applied to the real data set of Heart Transplant.

Thalassaemia Data Set Analysis

In last part of this analysis, the Thalassaemia data set (conducted by Fatimid Foundation (1981) Peshawar branch, used for analysis by Zaman and Pfeifer (2005)) is reanalyzed. The Borkowf's adjusted standard error and the proposed shrunken standard error are compared.

The data retrieved from 320 Thalassaemia patients. Thalassaemia is an inherited blood disease; the survival times taken are in months. Out of these 320 patients, 219 events (died) and 101 censored (still alive or migrated) at that time.

The results are analyzed in Table 4 with figure 3, for both methods and some mixed results are observed. The results in early part of the data set are significant for Borkowf's adjusted standard error, and in the last part of the data set are significant for the proposed shrunken standard error.

Table: 4 The estimates of Shrunken Kaplan-Meier and proposed survival functions, along with the Standard Errors of Hybrid and proposed estimators, applied to the real data set of Thallassaemia.

				Survival functions		Standard Errors	
time	e_i	c_i	n_i	$\hat{S}_{SKM}(t)$	$\hat{S}_P(t)$	SE_H	SE_{PH}
3	1	0	320	0.9953	0.9411	0.0038	0.0132
4	1	0	319	0.9922	0.9382	0.0049	0.0135
5	1	0	318	0.9890	0.9352	0.0058	0.0138
6	1	0	317	0.9859	0.9322	0.0066	0.0140
8	1	0	316	0.9828	0.9293	0.0073	0.0143
9	1	0	315	0.9797	0.9263	0.0079	0.0146
15	1	0	314	0.9766	0.9234	0.0084	0.0149
16	1	0	313	0.9735	0.9204	0.0090	0.0151
20	0	1	312	0.9735	0.9204	0.0090	0.0151
21	1	0	311	0.9703	0.9175	0.0095	0.0154
22	2	1	310	0.9641	0.9116	0.0104	0.0159
23	0	1	307	0.9641	0.9116	0.0104	0.0159
25	2	0	306	0.9578	0.9056	0.0113	0.0164
27	1	0	304	0.9547	0.9026	0.0117	0.0166
30	1	0	303	0.9515	0.8997	0.0121	0.0169
32	1	1	302	0.9484	0.8967	0.0124	0.0171
34	1	0	300	0.9452	0.8937	0.0128	0.0173
35	1	0	299	0.9420	0.8907	0.0131	0.0175
36	2	0	298	0.9357	0.8847	0.0138	0.0180
37	1	0	296	0.9326	0.8817	0.0141	0.0182
39	1	0	295	0.9294	0.8787	0.0144	0.0184
46	1	0	294	0.9263	0.8757	0.0147	0.0186
48	1	0	293	0.9231	0.8728	0.0150	0.0187
50	3	0	292	0.9136	0.8638	0.0158	0.0193
52	2	2	289	0.9073	0.8578	0.0164	0.0197
53	3	0	285	0.8978	0.8488	0.0171	0.0202
55	1	0	282	0.8946	0.8458	0.0173	0.0204
56	3	1	281	0.8851	0.8367	0.0180	0.0209
58	2	0	277	0.8787	0.8307	0.0184	0.0212
59	1	0	275	0.8755	0.8277	0.0187	0.0213
60	1	0	274	0.8723	0.8247	0.0189	0.0215
61	3	0	273	0.8628	0.8156	0.0194	0.0219
62	2	0	270	0.8564	0.8096	0.0198	0.0222
64	1	0	268	0.8532	0.8065	0.0200	0.0223
66	2	1	267	0.8468	0.8005	0.0204	0.0226
67	2	0	264	0.8404	0.7944	0.0207	0.0229
68	3	0	262	0.8308	0.7853	0.0212	0.0232
69	2	0	259	0.8244	0.7793	0.0215	0.0235
70	3	0	257	0.8148	0.7702	0.0220	0.0238
71	4	0	254	0.8020	0.7580	0.0226	0.0242
72	3	1	250	0.7924	0.7489	0.0230	0.0246

73	4	1	246	0.7795	0.7368	0.0235	0.0250
74	4	0	241	0.7666	0.7245	0.0240	0.0254
75	1	1	237	0.7634	0.7215	0.0242	0.0255
76	2	1	235	0.7569	0.7153	0.0244	0.0257
77	1	2	232	0.7536	0.7123	0.0246	0.0259
78	1	2	229	0.7503	0.7091	0.0248	0.0260
79	4	3	226	0.7371	0.6966	0.0254	0.0265
80	2	2	219	0.7304	0.6902	0.0257	0.0267
81	5	3	215	0.7134	0.6742	0.0263	0.0272
82	3	0	207	0.7031	0.6644	0.0266	0.0274
83	1	0	204	0.6997	0.6611	0.0266	0.0275
84	4	1	203	0.6859	0.6481	0.0270	0.0278
85	0	1	198	0.6859	0.6481	0.0271	0.0279
87	5	0	191	0.6578	0.6215	0.0278	0.0284
86	3	3	197	0.6755	0.6382	0.0274	0.0282
88	0	2	186	0.6578	0.6215	0.0279	0.0285
89	3	0	184	0.6471	0.6114	0.0281	0.0287
90	5	2	181	0.6293	0.5945	0.0285	0.0290
91	3	1	174	0.6185	0.5843	0.0287	0.0291
93	7	1	170	0.5931	0.5602	0.0291	0.0294
94	1	2	162	0.5894	0.5567	0.0292	0.0295
95	4	0	159	0.5746	0.5427	0.0294	0.0296
96	4	2	155	0.5599	0.5287	0.0296	0.0298
97	5	2	149	0.5411	0.5110	0.0298	0.0299
99	3	0	142	0.5297	0.5002	0.0299	0.0299
100	3	1	139	0.5183	0.4894	0.0300	0.0300
101	0	2	135	0.5183	0.4894	0.0301	0.0301
102	2	1	133	0.5105	0.4820	0.0301	0.0301
103	1	3	130	0.5066	0.4783	0.0303	0.0303
104	2	1	126	0.4986	0.4707	0.0304	0.0303
105	1	0	123	0.4946	0.4669	0.0304	0.0303
107	3	1	122	0.4824	0.4554	0.0304	0.0304
108	2	1	118	0.4743	0.4477	0.0305	0.0304
110	2	1	115	0.4661	0.4399	0.0305	0.0305
111	1	0	112	0.4619	0.4360	0.0305	0.0305
112	1	0	111	0.4578	0.4321	0.0305	0.0305
113	0	1	110	0.4578	0.4321	0.0306	0.0306
114	1	0	109	0.4536	0.4281	0.0306	0.0306
115	1	2	108	0.4494	0.4241	0.0307	0.0307
116	2	0	105	0.4409	0.4161	0.0307	0.0306
117	2	0	103	0.4323	0.4080	0.0307	0.0306
118	2	0	101	0.4238	0.3999	0.0307	0.0306
119	1	0	99	0.4195	0.3959	0.0307	0.0306
120	3	1	98	0.4067	0.3837	0.0307	0.0306
122	0	3	94	0.4067	0.3837	0.0309	0.0307
124	2	1	91	0.3978	0.3753	0.0309	0.0308

125	1	1	88	0.3933	0.3710	0.0310	0.0308
126	4	2	86	0.3751	0.3538	0.0310	0.0308
127	1	4	80	0.3704	0.3493	0.0312	0.0310
128	2	2	75	0.3606	0.3400	0.0313	0.0311
129	0	1	71	0.3606	0.3400	0.0314	0.0311
130	1	1	70	0.3555	0.3352	0.0314	0.0312
131	1	0	68	0.3503	0.3302	0.0314	0.0311
132	1	2	67	0.3451	0.3253	0.0315	0.0312
133	2	1	64	0.3343	0.3151	0.0315	0.0312
134	3	4	61	0.3180	0.2996	0.0316	0.0313
135	2	1	54	0.3062	0.2885	0.0316	0.0313
137	1	1	51	0.3003	0.2829	0.0317	0.0313
138	0	2	49	0.3003	0.2829	0.0318	0.0315
139	3	2	47	0.2812	0.2648	0.0318	0.0315
140	2	0	42	0.2679	0.2522	0.0317	0.0314
142	1	2	40	0.2612	0.2459	0.0318	0.0314
143	1	1	37	0.2542	0.2393	0.0318	0.0314
145	2	1	35	0.2398	0.2256	0.0318	0.0314
146	3	0	32	0.2174	0.2044	0.0316	0.0312
149	1	1	29	0.2100	0.1974	0.0316	0.0312
150	0	1	27	0.2100	0.1974	0.0317	0.0313
151	1	0	26	0.2020	0.1898	0.0316	0.0312
153	1	0	25	0.1939	0.1822	0.0316	0.0312
155	2	1	24	0.1779	0.1670	0.0315	0.0311
156	1	0	21	0.1695	0.1591	0.0315	0.0310
161	0	1	20	0.1695	0.1591	0.0315	0.0311
163	1	0	19	0.1607	0.1507	0.0315	0.0310
164	1	0	18	0.1518	0.1423	0.0314	0.0309
165	2	0	17	0.1341	0.1256	0.0313	0.0308
166	1	0	15	0.1253	0.1172	0.0312	0.0307
168	1	0	14	0.1165	0.1088	0.0311	0.0306
169	0	1	13	0.1165	0.1088	0.0312	0.0307
171	0	1	12	0.1165	0.1088	0.0312	0.0308
173	1	1	11	0.1060	0.0989	0.0312	0.0307
174	0	2	9	0.1060	0.0989	0.0314	0.0309
175	0	1	7	0.1060	0.0989	0.0314	0.0310
176	1	2	6	0.0886	0.0824	0.0315	0.0310
177	1	1	3	0.0596	0.0549	0.0315	0.0310
213	1	0	1	0.0015	2.4e-06	0.0314	0.0309

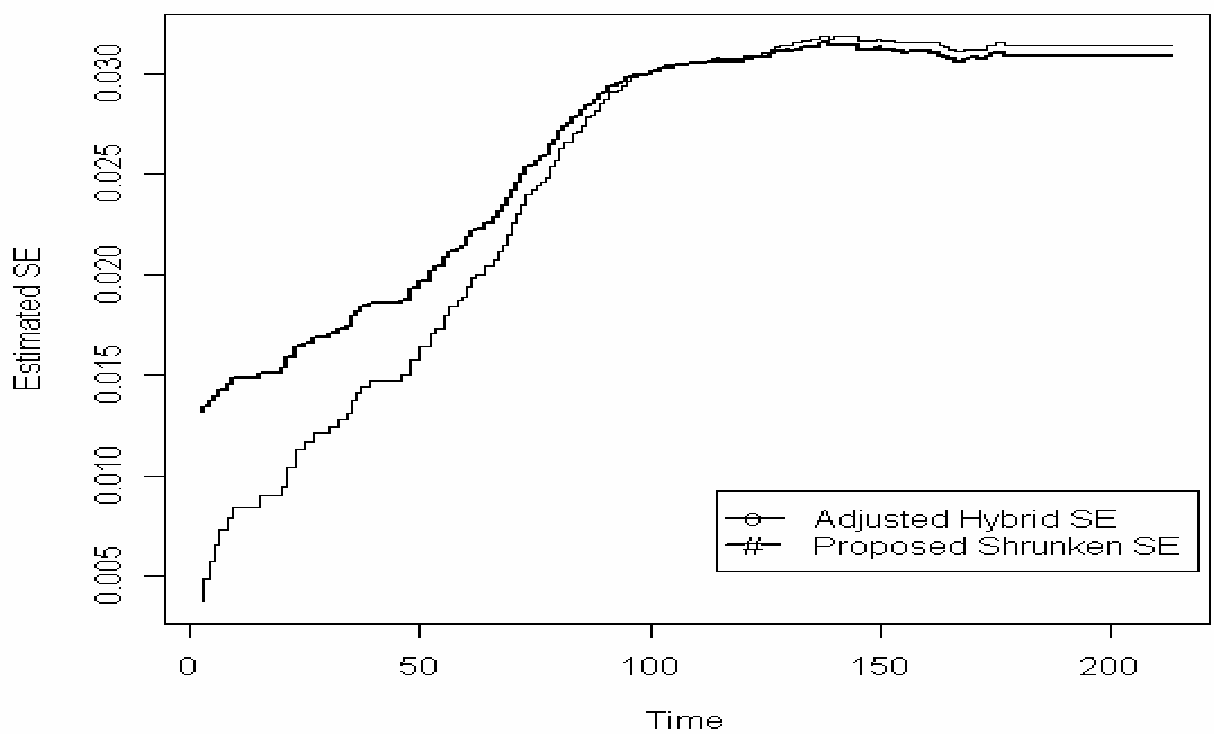

Figure 3: Showing the Standard Errors of Adjusted Hybrid and Proposed Shrunken Estimator. The Estimators are applied to the Thallassaemia data set.

CONCLUSIONS

A modification in the Kaplan-Meier survival function is made, for improving the estimator's performance. For comparison simulation are used to see the performances of the proposed estimator. Therefore, the Kaplan-Meier, shrunken and proposed shrunken survival functions are compared. The results of the simulations clearly support the proposed shrunken survival function compared to the Kaplan-Meier and shrunken survival function estimators.

For comparison Simulations is used, also the estimators are applied to the real data sets. In simulations different sample of sizes n = 25, 50, 100, 200 and 500 are used. Different percentage levels 15, 25, 45, 65 and 85 are used for censoring. Three points are selected for comparisons that are the quarters (Q_1, Q_2 and Q_3).

In next section estimators compared is the Greenwood to the proposed Greenwood variance estimator. Proposed Peto variance estimator is compared to the Peto's variance estimator. Applying simulation compares estimators and variances are applied to the Stanford Heart Transplant data set for comparisons. The results are far better for the proposed variance estimators compared to their respective variance estimators.

In last part of this section, the hybrid and adjusted hybrid variance estimators are compared with the proposed shrunken variance estimator. For comparing these variance estimators simulations are used. The estimators are also applied to the real data set of Thallassaemia. Results for the proposed shrunken variance estimator are better at third quartile than its alternative variance estimator, when simulation is used. When applied to real data set the results are almost same for both the proposed and hybrid variance estimators.

REFERENCES

Borkowf, C. B. (2005). A simple hybrid variance estimator for the Kaplan-Meier survival function. *Statistics in Medicine*, Vol. 24; 827-851.

Fatimid Foundation, A symbol of hope Pakistan(1981). *A Charitable Organization*. Established.

Greenwood, M. (1926). *A report on the natural duration of cancer*. In Reports on Public Health and Medical Subjects, Vol. 33; 1-26.

Kalbfleisch, J. D. & Prentice R. L. (1980). *Statistical Analysis of Failure Time Data*. Wiley:.

Kaplan, E. L. & Meier, P. (1958). Non Parametric Estimation from Incomplete Observations. *Journal of the American Statistical Association*, Vol. 53(282); 457-481.

Peto, R. & Pike, M. C., Armitage, P. et al. (1977). Design and analysis of randomized clinical trials requiring prolonged observation of each patient. Part II: Analysis and Examples. *British Journal of Cancer*, Vol. 35(1); 1-39.

Zaman, Q. & Pfeifer, K. P. (2005). Modified bootstrap survival function. *Sri-Lankan Journal of Applied Statistics*, Vol. 6; 43-51.

Zhao, G. (1996) The homogenetic estimate for the variance of survival rate. *Statistics in Medicine*, Vol. 15; 51-60.