# AN EVALUATION OF RESIDUALS FOR THE COX MODEL

William E. Barlow, University of Southern California
Shan-Pin C. Fanchiang, University of Southern California

Abstract

Residuals are often helpful in detecting model departures. This paper compares two residuals for the Cox regression model, the Schoenfeld (1982) score residuals and the Barlow and Prentice (1988) integrated residuals, in their ability to detect nonproportional hazards. Simulation results indicate that the Schoenfeld residuals may have greater specificity. The integrated residual does estimate the empirical influence function, however.

## 1. Introduction

The Cox (1972) model is widely used for analyzing the relationship of survival time to prognostic factors. Departure from the assumed model can lead to inappropriate inferences, therefore it is important to examine these assumptions. The utilization of residuals in testing assumptions is most applicable in parametric model fitting, though residuals for the semi-parametric Cox model may still be informative. These residuals may indicate model departures due to mismodelled covariates or failure of the proportional hazards assumption. It is also useful to compute estimates of the empirical influence function, the resulting change in the estimated coefficients if an individual observation is deleted.

Schoenfeld (1982) proposed residuals corresponding to the score component at each failure time. Barlow and Prentice (1988) suggested another residual, the integrated residual, which is specific to an individual rather than a failure time and can accomodate time-dependent covariates or a general relative risk form. They also estimate the influence of the observation on the estimates. The Schoenfeld and integrated residuals are compared in this simulation study. The ability of the residuals to detect departures from the proportional hazards assumption is investigated empirically.

## 2. Residuals for the Cox Model

Assume there are m unique failures among the n individuals. Let $z_i$ be the covariate vector for individual i, $Y_i(t)$ indicate whether that individual was at risk at time t, and $dN_i(t)$ indicate whether individual i fails at time t or not. Both $Y(t)$ and $dN(t)$ are indicator processes which take only value one or zero at every time point.

The proportional hazards model assumes the hazard is $\lambda(z;t) = \lambda_0(t)\ r(z^T\beta)$ where $\lambda_0(t)$ is an arbitrary baseline hazard and r is the relative risk function. Typically, the risk form will be $\exp(z^T\beta)$ without time-dependent covariates which is the form considered here, though the more general form is considered in Barlow and Prentice (1988). The conditional probability that it was individual i who failed at failure time $t_j$, conditional on the risk set at that time is

$$p_i(t) = \frac{Y_i(t_j)\ e^{z_i^T\beta}}{\sum_{k=1}^{n} Y_k(t_j)\ e^{z_k^T\beta}} = \frac{Y_i(t_j)\ \lambda(z_i;t_j)}{\sum_{k=1}^{n} Y_k(t_j)\ \lambda(z_k;t_j)}$$

The conditional expectation of the covariate at time j is

$$E(t_j) = \sum_{k=1}^{n} Y_k(t_j)\ p_k(t_j)\ z_k$$

Define the "residual component" for the person i at time $t_j$ to be

$$c_i(t_j) = Y_i(t_j)\ [dN_i(t_j)-p_i(t_j)]\ [z_i-E(t_j)]$$

so that it is the product of (1) the discrepancy between the difference in the observed failure status and the conditional probability of failure and (2) the difference in the observed covariate and its conditional expectation. In practice, an estimate of $\beta$ must be used in the formulas.

If these residual components are summed over individuals at a fixed failure time $t_j$ one obtains the Schoenfeld residual. This residual will have mean zero and known variance function since it is the contribution to the score at each failure time. The number of Schoenfeld residuals is limited to the m unique failure times.

If the residuals are summed over time points for a single individual one obtains the Barlow and Prentice integrated residual. The integrated residual will have mean zero and known variance function since it is derived using martingale results. There are n integrated residuals with a contribution at each failure time that the individual is at risk. If the baseline hazard is small then the residuals will be dominated by the failure contributions.

The influence function measures the sensitivity of an estimate to a particular observation. It is estimated by the empirical influence function, the discrepancy between the overall $\hat{\beta}$ and $\hat{\beta}_{-i}$ which is the value obtained when the ith observation is deleted (Reid, 1983). Both Cain and Lange (1984) and Reid and Crepeau (1985) independently suggest an estimate of the empirical influence function for the standard Cox model with exponential relative risk form and fixed covariates. The integrated residual is an unstandardized form of these estimates for the exponential risk form with fixed covariates. Since the integrated residual is an estimator of the actual empirical influence function, it is expected that the correlation between the two should be close to unity.

3. Example

Consider an example with four observed failure times and one covariate, i.e. $(z,t)$ is (2,50), (1,25), (5,24), and (4,13). A $\hat{\beta}$ of 0.5493 is obtained. Table 1 illustrates the $p_i(t)$ and expected covariate values for each failure time. As an example of the computation the residual component for the failure at time $t_1$ is (4 - 4.14983) (1 - 0.30695) = -0.10384. Table 2 exhibits the matrix of $c_i(t_j)$ for the example dataset and the residuals. Both the Schoenfeld and integrated residuals have mean zero. The integrated residuals compare well to the actual empirical influence function obtained by deleting each observation in turn. A Pearson correlation coefficient of 0.9633 is ascertained between the actual empirical influence function and the integrated residual.

4. Computation in the SAS® System

Cox regression analysis is available in the SUGI Supplemental Library, Version 5. Both COXREGR and PHGLM procedures fit the Cox regression model. For our purpose we find the PHGLM procedure more suitable. The model assumes an exponential risk form with fixed covariates. Though PHGLM computes the Schoenfeld residuals and performs a test of the proportional hazards assumption, it cannot output the calculated residuals without further data manipulation. One option outputs an estimated hazard function at each time point for each individual at risk at that time which makes it possible to compute the residuals.

The conditional probability of failure can be ascertained by means of the hazards, $\hat{\lambda}(t)$. After outputting the entire hazard function and the covariate matrix, one can calculate the failure event indicator, $dN_i(t)$, and $z_i\,\hat{\lambda}_i(t_j)$ used to compute the expected value of the covariate. Then, manipulating the data matrix by using PROC SUMMARY in Base SAS software, one can obtain $\sum_{k=1}^{n} Y_k(t_j)\hat{\lambda}_k(t_j)$ and $\sum_{k=1}^{n} Y_k(t_j)z_k\hat{\lambda}_k(t_j)$. Finally, the residual component, $c_i(t_j)$, is calculated through the following formula:

$$\left\{ z_i - \frac{\sum_{k=1}^{n} Y_k(t_j)z_k\hat{\lambda}_k(t_j)}{\sum_{k=1}^{n} Y_k(t_j)\hat{\lambda}_k(t_j)} \right\} \left\{ dN_i(t_j) - \frac{\hat{\lambda}_i(t_j)}{\sum_{k=1}^{n} Y_k(t_j)\hat{\lambda}_k(t_j)} \right\}$$

The Schoenfeld and the integrated residuals result from the PROC SUMMARY procedure summed over the individuals at each observed time point t, and summed over the observed time points for each individual, i, respectively. Explicit computation of the baseline hazard and relative risk is not required. This procedure is very cumbersome and good programmers should be able to find a much easier method.

5. Simulation Studies

5.1 Objectives

The objectives of this simulation study were to evaluate the following:

(1) The ability of the integrated residual to correctly estimate the empirical influence function when the proportional hazard assumption is correct.
(2) The degree of randomness the Schoenfeld and integrated residuals show when the proportional hazards assumption is correct.
(3) The degree of nonrandomness the Schoenfeld and integrated residuals show when the proportional hazards assumption is not correct.

To meet the first objective, a Pearson correlation coefficient between the integrated residual and the actual empirical influence function was computed. Calculation of the actual empirical influence function requires that each observation be deleted in turn and the model refit. A very high correlation coefficient is expected if the integrated residual is a good estimator of the empirical influence function.

To evaluate the second and third objectives, the Pearson correlation coefficients were computed between the rank of the failure time and each of the Schoenfeld and integrated residuals. A strong and significant correlation indicates a linear trend between the residuals and the rank of the failure time points. The lag-1 serial correlation (Draper and Smith, 1981) within the Schoenfeld and integrated residuals were also computed. Significant serial correlation coefficients may indicate nonrandomness or curvilinear trend of the residuals.

If the proportional hazards assumption holds, no trend or pattern of the residuals should be observed. Ideally, the characteristic of randomness should exist within both types of residuals. Therefore, low and nonsignificant correlation coefficients were expected for both the Schoenfeld and integrated residuals. By comparison, if the proportional hazards assumption was not met, then any trends, patterns, or nonrandomness within the residuals should have been observed. Consequently, significant correlations between the rank of the failure time and the residuals as well as significant serial correlations within residuals were expected.

5.2 Simulated Data

Data were generated from a survival distribution with a single continuous covariate. The hazards functions were from both proportional and nonproportional hazards distributions. All models assumed no censoring and no tied failure times. When the data were derived from a nonproportional hazards model, a Cox regression model was fit in order to produce residuals which hopefully indicated the failure of the proportional hazards assumption to hold.

Failure time data from a Weibull distribution was generated such that systematically varying the parameters resulted in different proportional and nonproportional hazards models. The usual Weibull hazard function is the following:

$$\lambda(t) = \alpha\gamma(\gamma t)^{\alpha-1}$$

The model was generalized such that both the shape parameter, $\alpha$, and the scale parameter, $\gamma$, were functions of a single covariate, x. Let $\alpha = e^{\alpha_0+\alpha_1 x}$ and $\gamma = e^{\gamma_0+\gamma_1 x}$. If the parameter $\alpha_1$ in the shape parameter $\alpha$ was zero, then the model was proportional hazards, otherwise it was nonproportional. As a result, altering the value of $\alpha_0$, $\alpha_1$, $\gamma_0$, and $\gamma_1$ could yield different hazards functions. Since the Cox model used the ranks of the failure times, no change of $\gamma_0$ in the scale parameter $\gamma$ would influence the rank of the failure time. Therefore, $\gamma_0$ was fixed at zero. All models were generated with a covariate from a normal distribution with mean zero and variance one.

The PHGLM procedure allows fitting of the Cox model and subsequent evaluation of results so this simulation was conducted in SAS. Using nested macros, two hundred replications were performed for each simulation. Within each replication, thirty distinct failure times were generated and fitted with the Cox model using the PHGLM procedure. In addition, the program calculated the Schoenfeld residuals, integrated residuals, and the necessary correlations as well.

The computation of the empirical influence function and the correlation was performed only for one of the exponential hazard models since the primary interest was to evaluate the estimate of the empirical influence function when the assumption of proportional hazards was correct.

The program for the simulation study is available upon request.

5.3 Simulation Results

The first objective assesses the efficacy of the integrated residuals to estimate the actual empirical influence function when the proportional hazard assumption is correct. The mean correlation coefficient was 0.9831 with a standard error of 0.0015. There is excellent agreement between the integrated residual and the actual empirical influence function.

The second objective concerns the behavior of the residuals when the true model is proportional hazards. This occurs in the Weibull distribution when $\alpha_1=0$ in the shape parameter. The results of these models are shown in Figure 1. The integrated residuals were correlated with the ranks of the failure time in all cases. When the failure time did not depend on x, i.e. $\gamma_1=0$, the magnitude of the correlation was the lowest. The Schoenfeld showed low correlations in all cases. Results from the serial correlation also showed greater correlation for the integrated residuals than for the Schoenfeld residuals.

Nonproportional hazard models with $\alpha_1=1$ and $\alpha_0=0$ in the shape parameter were simulated. One would expect that the correlation among the residuals and the ranks of the failure time to be high and significant so that the incorrect proportional hazard assumption could be detected. Figure 2 shows a plot of these correlations. In general the integrated residual showed a stronger correlation with failure time rank than the Schoenfeld residual.

Figure 3 illustrates the relationship between the correlation and the shape parameter $\alpha_1$ when the scale parameter $\gamma_1$ was fixed at zero. The correlations for both residuals increased when $\alpha_1$ increased. The correlations were the lowest when the model correctly assumed proportional hazards. It appears that the correlations were monotonically related to the shape parameter $\alpha_1$ when the scale parameter $\gamma_1$ was fixed, and that the correlations were stronger for the integrated residual than for the Schoenfeld residual.

6. Discussion

In conclusion, the integrated residual was a good estimator of the empirical influence function resulting in considerable computational efficiency. Therefore, those observations which impact the parameter estimates strongly could be identified easily. The Schoenfeld residual correctly showed a random pattern when the true model was proportional hazards, but lacked power under certain nonproportional hazards situations. The integrated residual was more likely to be correlated with the rank of the failure time and also tended to be serially correlated in both the proportional and nonproportional hazard models used in this study. The shape parameter seemed to provide a steady prediction of changes for the mean correlations between both residuals and the rank of the failure time in the models with Weibull failure time distribution.

It appears that neither type of residual would detect an incorrect proportional hazard assumption with precision and confidence. The simulated data did not allow censoring or tied failure times and did not vary the sample size which limits the generalizability of the results.

## References


Barlow, W.E., & Prentice, R.L. (1988). Residuals for relative risk regression models. *Biometrika*, 75, 65-74.

Cain, K.C., Lange, N.T. (1984). Approximate case influence for the proportional hazards regression model with censored data. *Biometrics*, 40, 493-499.

Cox, D.R. (1972) Regression models and life tables (with discussion). *Journal of the Royal Statistical Society (B)*, 34, 187-220.

Draper, N. & Smith, H. (1981) *Applied Regression Analysis*. New York: John Wiley and Sons.

Reid, N. (1983) Influence functions. In *Encyclopedia of Statistical Sciences*, 4, Ed. S. Kotz, N. L. Johnson, and C. Read, pp. 117-119. New York: Wiley.

Reid N, & Crepeau H. (1985). Influence functions for proportional hazards regression. *Biometrika*, 72, 1-9.

Schoenfeld, D. (1982) Partial residuals for the proportional hazards regression model. *Biometrika*, 69, 239-241.

SAS Institute Inc. *SUGI Supplemental Library User's Guide, Version 5 Edition. Cary, NC: SAS* Institute Inc., 1986.


Appendix: SAS program to compute residuals.

```
** ---------------------------------------- **;
** Program Description:                     **;
** 1.read in data.                          **;
**   time=failure or censoring time         **;
**   delta=failure indicator                **;
**   cov=covariate                          **;
** 2.fit phglm for Cox model,               **;
**   compute beta estimate & variance.      **;
** 3.macro <IandS>, compute residuals.      **;
**   I: Integrated Residual.                **;
**   S: Schoenfeld Residual.                **;
** 4.merge residuals and parameter estimates.**;
**                  William E. Barlow;      **;
**                  Shan-Pin Fanchiang 5/88 **;
** ---------------------------------------- **;

   data exp;
     input time delta cov ;
     rep=1; ** rep: used for merging results **;
    cards;
50 1 2
25 1 1
13 1 4
24 1 5
  ;
     proc sort; by descending time;

** ---------------------------------------- **;
** Fit Cox Model & Compute Residuals        **;
** ---------------------------------------- **;
    data maindata; set exp exp(in=L);
      tim=time;
      id=0; ** id:(see SAS SUGI pp.450) **;
      if L then do; time=.; id=1; end;
        proc phglm data=maindata out=full
                 outp=predata  nt= 8;
           event  delta;
           model time=cov;
           id id;

** ---------------------------------------- **;
** Get beta estimate & variance             **;
** ---------------------------------------- **;
 data betatemp; merge
   full(obs=1 rename=(cov=beta_hat))
   full(firstobs=2 rename=(cov=beta_var));
   rep=1;

** ---------------------------------------- **;
** Compute S.Residual & I.Residual          **;
** ---------------------------------------- **;
 data scurve; set predata(firstobs= 5);
   drop loglog_s survival id;
   if hazard =0 or time > tim then delete;
   dn=(time=tim);  ** dn=dN_i(t) **;
   hx=hazard*cov;

   proc summary nway;
     class time;
       var hx hazard;
       output out=summary sum=shx shazard;
   proc sort data=scurve; by time;

 data datares; merge
   scurve(drop=hx)
   summary(drop=_type_ _freq_); by time;
   expxbeta=exp(xbeta);
   hazard_o=hazard/expxbeta;
     ** hazard_o=baseline hazards **;
   Ci=(dn-hazard/shazard)*(cov-shx/shazard);
   E_xt=shx/shazard;  ** E_xt=Ê(β̂,t) **;
   Pi_t=hazard/shazard; ** Pi_t=p̂_i(t) **;
   t=time;

   %macro IandS(classv,resid,res);
     proc summary nway idmin data=datares;
       class &classv;
        var Ci;
         id tim rep cov;
       output out=&resid sum=&res;
   %mend;

   %IandS(t,  Sresid,S);
   %IandS(tim,Iresid,I);

** ---------------------------------------- **;
** Output original data, residuals, & beta  **;
** ---------------------------------------- **;
data expfinal; merge
  betatemp
  sresid(drop=_type_  _freq_ cov tim)
  iresid(drop=_type_ _freq_ tim);
  by rep;
     proc print;
```

Table 1
Example Data Set
Conditional Probability of Failure

| Person $i$ | Covariate $x_i$ | Conditional Probability of Failure $p_i(t_1)$ | $p_i(t_2)$ | $p_i(t_3)$ | $p_i(t_4)$ |
|---|---|---|---|---|---|
| 1 | 4 | 0.3070 | 0.0000 | 0.0000 | 0.0000 |
| 2 | 5 | 0.5317 | 0.7671 | 0.0000 | 0.0000 |
| 3 | 1 | 0.0591 | 0.0852 | 0.3660 | 0.0000 |
| 4 | 2 | 0.1023 | 0.1476 | 0.6340 | 1.0000 |
| Expected covariate | | 4.1498 | 4.2162 | 1.6340 | 2.0000 |

Table 2
Example Data Set
Illustration of Residuals

| Person | Residual Component | | | | Integrated Residual |
|---|---|---|---|---|---|
| $i$ | $c_i(t_1)$ | $c_i(t_2)$ | $c_i(t_3)$ | $c_i(t_4)$ | $w_i$ |
| 1 | -0.1038 | 0.0000 | 0.0000 | 0.0000 | -0.1038 |
| 2 | -0.4521 | 0.1825 | 0.0000 | 0.0000 | -0.2695 |
| 3 | 0.1861 | 0.2741 | -0.4019 | 0.0000 | 0.0583 |
| 4 | 0.2200 | 0.3272 | -0.2321 | 0.0000 | 0.3151 |
| $s(t_j)$ | -0.1498 | 0.7383 | -0.6340 | 0.0000 | |

$s(t_j)$: Schoenfeld Residual

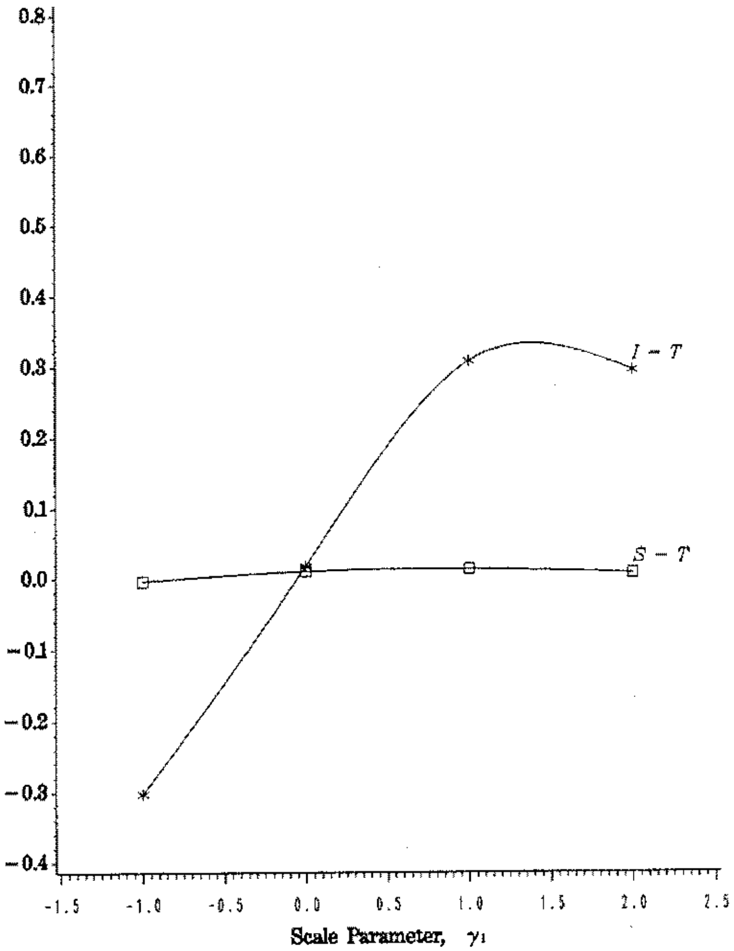
**Figure 1**

**Mean Correlations and Scale Parameter**

**Mean Correlations among Residuals and Ranks of the Failure Time**

**Scale Parameter of Weibull Hazard Function, $\gamma_1$**

**Proportional Hazards, $\alpha_1=0$**




Note: Residuals: $S$=Schoenfeld, $I$=Integrated, $T$=Failure Time Rank

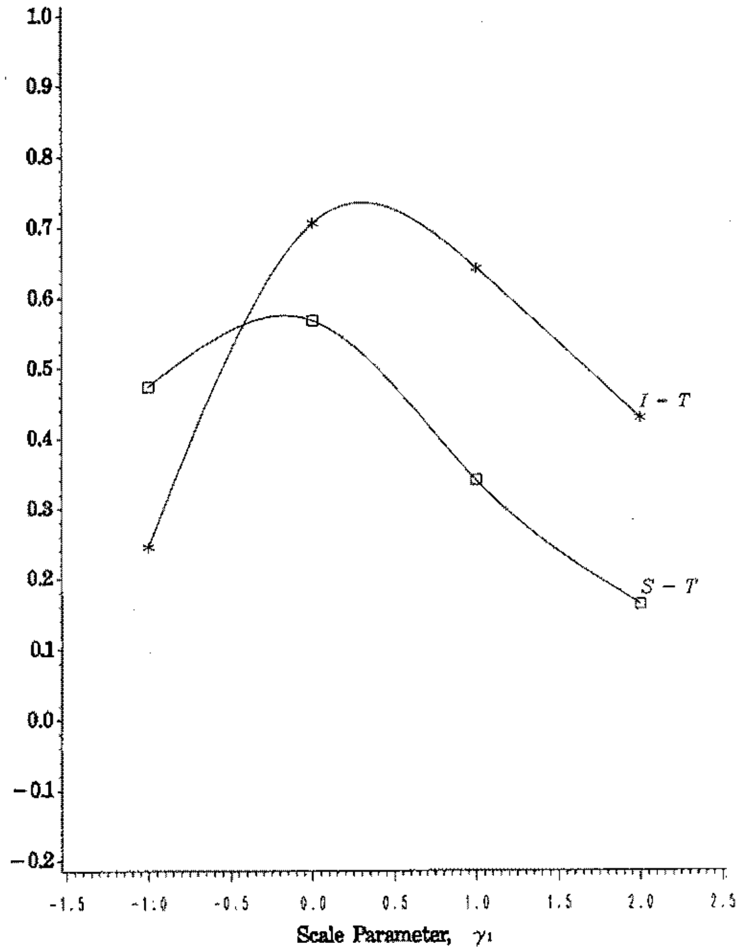
**Figure 2**

**Mean Correlations and Scale Parameter**

**Mean Correlations among Residuals and Ranks of the Failure Time**

**Scale Parameter of Weibull Hazard Function, $\gamma_1$**

**Nonproportional Hazards, $\alpha_1 = 1$**




Note: Residuals: $S$=Schoenfeld, $I$=Integrated, $T$=Failure Time Rank

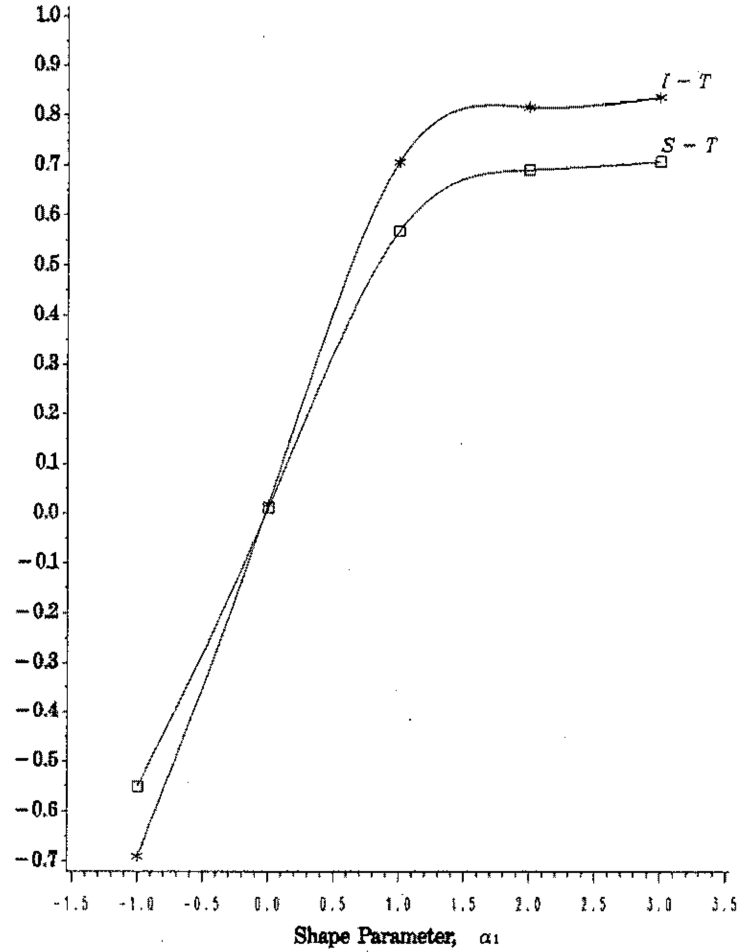
**Figure 3**

**Mean Correlations and Shape Parameter**

**Mean Correlations among Residuals and Ranks of the Failure Time**

**Shape Parameter of Weibull Hazard Function, $\alpha_1$**

**Scale Parameter, $\gamma_1 = 0$**




Note: Residuals: $S$=Schoenfeld, $I$=Integrated, $T$=Failure Time Rank