COVARIATE ANALYSIS OF BIVARIATE SURVIVAL DATA

by

Leah Bennett

Department of Biostatistics, University of
North Carolina at Chapel Hill, NC.

Institute of Statistics Mimeo Series No. 1899T

April 1992

COVARIATE ANALYSIS OF BIVARIATE SURVIVAL DATA

by

Leah Bennett

A Dissertation submitted to the faculty of The University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor Of Philosophy in the Department of Biostatistics.

Chapel Hill

1992

Approved by :

Advisor

Reader

Reader

Reader

Reader

LEAH ELIZABETH BENNETT. Covariate Analysis of Bivariate Survival Data (Under the direction of Dr. Chirayath Suchindran.)

ABSTRACT

The methods developed in this research are used to analyze the effects of covariates on bivariate survival data when censoring and ties are present. Many existing bivariate survival models make no provision for the inclusion of covariates, nor do they effectively utilize the censored observations in their estimation. The proposed method provides models for bivariate survival data that include differential covariate effects and censored observations. The proposed models are based on an extension of the univariate Buckley-James estimators which replace censored data points by their expected values, conditional on the censoring time and the covariates. For the bivariate situation, since we assume that the outcome times are correlated, it is necessary to determine the expectation of the failure times for one component conditional on the failure or censoring time of the other component.

Two different methods have been developed to estimate these expectations. In the semiparametric approach these expectations are determined from a modification of Burke's estimate of the bivariate empirical survival function. In the parametric approach censored data points are also replaced by their conditional expected values, but instead of using the empirical survival function, the expected values are determined from a specified parametric distribution, such as bivariate exponential or bivariate Weibull. The model estimation will be

based on the revised data set, comprised of uncensored components and expected values for the censored components. Since the covariate parameters depend on the expected failure times which in turn depend on the covariate parameters, an iterative process must be used to estimate both. The variance-covariance matrix for the estimated covariate parameters has also been derived for both the semiparametric and parametric methods.

Data from the Demographic and Health Survey was analyzed by these methods. The two outcome variables are post-partum amenorrhea and breastfeeding; education and parity were used as the covariates. Both the covariate parameter estimates and the variance-covariance estimates for the semiparametric and parametric models will be compared. In addition, a multivariate test statistic was used in the semiparametric model to examine contrasts. The significance of the statistic was determined from a bootstrap distribution of the test statistic.

Acknowledgement

I would like to thank all of my committee members for their helpful comments and thought-provoking suggestions, but especially I would like to thank Dr. Suchindran for his infinite patience and wisdom. The completion of this research would not have been possible without his support and optimism. I also gratefully acknowledge the support of the National Institute of Child Health and Human Development, training grant T32-HD-07237.

The love and understanding exhibited by my friends in my frequently stressful periods is greatly appreciated. Most importantly I would like to thank my family, in particular my dearest mother and sister and my wonderful grandparents, for their constant encouragement and unconditional love. This research is a direct result of their faith in me, and is dedicated to them and the memory of my beloved father.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1. INTRODUCTION

1.1 MOTIVATION AND PURPOSE

The methods of analysis for survival data, while numerous and varied, all have the same general motivation : to estimate the time, or distribution of time, until failure, in some broad sense. Extending these concepts to include the analysis of covariates allows for heterogeneity in the population under study and estimation of the determinants of factors affecting length of life. Thus covariate analysis of survival data can be utilized for descriptive and predictive purposes, i.e., to characterize patterns of survival for the particular data used for estimation, determine the effect of the covariates on length of life, and predict the lifetime of individuals from a similar population.

Until recently the primary focus of research in survival analysis has been on univariate techniques. In the past two decades there has been an increasing emphasis on analyzing survival data with bivariate or even higher dimensional outcomes. The focus of this research is on the analysis of covariates for bivariate survival data when censoring and ties are present. The methods developed here will be applied to a data set to demonstrate the estimation process as well as the rudiments of inference.

1.2 GENERAL ISSUES IN THE DEVELOPMENT OF BIVARIATE SURVIVAL MODELS

Because the field of multivariate survival analysis (bivariate in particular) has been such a popular topic in the past decade, the analytic methods that have been developed are quite numerous and varied. Covariate analysis of bivariate survival data has been approached from many directions, both as an extension of the univariate situation and as a distinct entity. Many difficulties are encountered when univariate methods are directly extended to the bivariate framework because simple mechanisms and concepts in the univariate sense, such as ranking, may be ambiguous.

Many factors have to be considered in developing a model -- whether it is to be based on survival or hazard functions, approached from a parametric or nonparametric perspective, whether estimations are performed using maximum likelihood or least squares methods, and so forth. Other issues to be considered are concerned with how censoring and ties are to be handled. And if covariates are to be included in the model, there is the additional issue of what method is to be used for incorporating them in the model. Several models have been developed but few consider ties, censoring, and covariates; most consider only one or some combination of these factors.

1.3 DESCRIPTION OF BIVARIATE SURVIVAL DATA

There are several types of data that would be suitable for analysis using bivariate survival techniques. The failure time measurements could result from:

i) components in the same system -- such as two organs in the human body or two engines on an airplane -- where the failure of one component places additional strain on the remaining component;

ii) related components -- for example, two members of the same family who have similar genetic makeup -- where the age at death of a parent is strongly correlated with the age at death of the child when certain hereditary diseases are involved;

or,

iii) measurements on the same component -- such as the duration of exposure to an infecting agent and duration of an infectious disease -- where the relationship is often not symmetric, perhaps due to cause and effect.

In most of these situations, it is unquestionable that a dependency between components exists; but quantifying that relationship is not so straightforward. Hence there is a need for analytical techniques to determine the joint survival function as well as to examine the conditional and marginal survival functions and the effect of covariates.

The data to be considered in this work consists of bivariate outcomes resulting from a random process. Either or both component may be right-censored. Though the results of this research may be generalized to include left-censoring and interval censoring as well, only right-censoring is considered here. Censoring may result from several different mechanisms, perhaps the most common being the cessation of the study before the failure time is reached; another cause may be withdrawal from the study for reasons unrelated to the variables of interest (such as relocation). Various other reasons may also be postulated. For most survival analysis techniques, the censoring mechanism is assumed to be independent of the random process generating the failure times. This research will only consider only censoring on the outcome variables. In some situations, particularly with time-dependent covariates, the concomitant variables may also be censored. This type of data requires analytic techniques other than those to be presented in this work, and is not considered here. It is crucial to make a distinction between missing and censored data; missing data provide no information on one or both outcome variables whereas censored data provide partial information.

Information on covariates, sometimes known as concomitant variables, will also be included in the data set; the covariates can be binary, discrete, or continuous. Besides examining the effect of covariates on survival curves, it is often also of interest to consider the correlation of the outcome variables. If the outcomes are known to be independent, it is usually more desirable to perform simpler univariate analyses on each individual lifetime component. With the inclusion of covariates, it may also be desirable to explore the correlation of the covariate parameters.

1.4 BASIC DEFINITIONS

To discuss either univariate or bivariate methods of covariate analysis in any depth requires that some notation and definitions be presented. Let the random variables Y_1 and Y_2 be the failure times of the first and second components, respectively, for some continuous bivariate process. These variables may or may not be independent, though the case of independence is of little interest as univariate techniques could then be used separately on each component. Regardless, the vector of failure times for any observation is independent of every other observation. For a given population, survival is defined as the probability of failing after a particular time point in the univariate case, or after a vector of time points in the bivariate case.

Thus let

$$Y_j = \text{failure time for the } j^{th} \text{ component}, \qquad j = 1, 2; \qquad \text{and}$$

$$S(t_1, t_2) = \Pr(\ Y_1 > t_1 , Y_2 > t_2\ ) .$$

Marginal and conditional probabilities of survival may also be defined, as follows:

$$S_j(t_j) = \Pr(\ Y_j > t_j\ ) ; \qquad \text{and}$$

$$S_{1|2}(t_1 \mid t_2) = \Pr(\ Y_1 > t_1 \mid Y_2 > t_2\ ) , \qquad \text{with}$$

$$S_{2|1}(t_2 \mid t_1) \text{ similarly defined.} \tag{1.1}$$

Much of the research on bivariate survival data makes no provision for the inclusion of censoring. In practice this is often unrealistic; requiring complete observations would necessitate, in many cases, eliminating a substantial portion of the data, despite the fact that there is much information to be gained from partially or even wholly censored observations. The estimated mean failure times would be underestimated and covariate effects would be biased if only complete observations were used. For the purposes of this work, censoring refers to a random process that removes either or both components of an observation from study before the failure time is observed. In other words, if Y_1 is censored at c_1, it is known that the component survived at least until c_1, but it is not known how long it survived past that point.

In mathematical terms:

Y_j^0 = failure time for j^{th} component ($j = 1,2$),

C_j = censoring time for j^{th} component,

$Y_j = \min(Y_j^0, C_j)$ = observed stopping time for j^{th} component, and

$$\delta_j = I(Y_j = Y_j^0) = \begin{cases} 0, & \text{if } j^{th} \text{ component is censored} \\ 1, & \text{if failure time of } j^{th} \text{ component is observed.} \end{cases} \tag{1.2}$$

We assume that the censoring mechanism is independent of survival, though it is not necessary to assume that the two censoring time are independent. Sometimes it may be useful to assume a specific censoring distribution. The method for resolving censored observations further complicates matters: these observations can be ignored, they may be treated as observed values with some weight attached, they may be replaced according to a specified method, or they may be treated by any number of alternative methods. Some of these methods are discussed in more detail in subsequent sections.

1.5 UNIVARIATE METHODS OF COVARIATE ANALYSIS

We briefly review univariate methods to demonstrate how extensions to the bivariate framework arise. In the remainder of this chapter various survival models are presented: the first four are univariate nonparametric models based on regression methods, the last two are related to bivariate parametric methods. Miller and Halpern (1981) present a thorough review of univariate methods, with particular emphasis on the four methods to be described here.

The first method to be considered, perhaps one of the most well-known techniques for univariate survival analysis, is the Cox proportional hazards approach (Cox, 1972), which is based on the assumption that the instantaneous hazard of death at any time, with a given set of covariates, is proportional to the instantaneous hazard at that time with the covariates all set to the baseline value. This can be stated in terms of the hazard function, λ, as follows :

$$\lambda(\mathrm{y}; \mathbf{x}) = \frac{\mathrm{f}(\mathrm{y}; \mathbf{x})}{\mathrm{S}(\mathrm{y}; \mathbf{x})} = \lambda_0(\mathrm{y}) * e^{\mathbf{x}\boldsymbol{\beta}} ;$$

or, equivalently stated in terms of the survival function,

$$\mathrm{S}(\mathrm{y}; \mathbf{x}) = \{ \mathrm{S}(\mathrm{y}; \mathbf{0}) \}^{e^{\mathbf{x}\boldsymbol{\beta}}}, \qquad \text{where}$$

$\mathbf{x} \equiv$ vector of covariates , $\mathrm{Y} \sim \mathrm{f}(\mathrm{y}; \mathbf{x})$

and $$\mathrm{S}(\mathrm{y}; \mathbf{x}) = \int_y^\infty \mathrm{f}(\mathrm{t}; \mathbf{x})\, \mathrm{dt} . \tag{1.3}$$

For the Cox model the estimator of $\boldsymbol{\beta}$ is the value which solves

$$\sum_{i=1}^{n} \delta_{(i)} \left\{ \mathrm{X}_{(i)}\boldsymbol{\beta} - \log \left(\sum_{j \in \Re(y_{(i)})} \mathrm{e}^{X_j \boldsymbol{\beta}} \right) \right\} = 0$$

where $\Re(\mathrm{t})$ = risk set of observations still alive and not censored by t;

where $\Re(t)$ = risk set of observations still alive and not censored by t;

$y_{(1)}, \ldots, y_{(n)}$ = ordered observations;

and $X_{(i)}$ and $\delta_{(i)}$ are the covariates and censoring indicator,

respectively, associated with $y_{(i)}$.

The other three univariate methods to be considered here are all based on the linear regression model, $\mathbf{Y} = \mathbf{X}\beta + \epsilon$, where Y can represent either the failure time or any monotonic transformation of the failure time. If Y represents the logarithm of the observed failure times, then this linear model is the accelerated time model. In the one-dimensional case many regression estimators have been formulated, which for the most part are straightforward.

The first linear model estimator to be considered is the Miller (1976) estimator, which weights the observations according to the following mechanism :

$$\begin{aligned} w_i(\hat{\beta}) &= \hat{F}(e_i) - \hat{F}(e_i^-) \qquad i = 1, \ldots, n \\ &= \begin{array}{c}\text{probability mass assigned to } i^{th} \\ \text{residual by the Kaplan-Meier estimator}\end{array} \end{aligned}$$

Since the Kaplan-Meier estimator increases only with observed failure times and not with censored observation, this weighting scheme actually gives zero weight to censored observations and increases the weight of the observed failure times. The estimation of the covariate parameters requires an iterative solution as the weights are based on these parameters. Thus

$$\hat{\beta}^{(k+1)} = [\,(X - \bar{X}^w)^T \cdot W(\hat{\beta}^{(k)}) \cdot (X - \bar{X}^w)\,]^{-1} \cdot (X - \bar{X}^w)^T \cdot W(\hat{\beta}^{(k)}) \cdot y \quad , \tag{1.4}$$

where $\boldsymbol{W}$ is a diagonal matrix of the weights based on the estimated survival distribution and $\overline{\boldsymbol{X}}^w$ is a matrix of weighted averages for the covariates.

The Buckley-James estimator (1979) also requires an iterative solution. Censored observations at the k^{th} iteration are replaced with their corresponding conditional expected values according to the following scheme:

$$\hat{y}(\hat{\beta}^{(k)}) = \delta_i\, y_i + (1-\delta_i)\cdot \hat{E}[Y_i \mid Y_i > y_i] \ .$$

Thus $\hat{y}_i$ is either the observed failure time or the expected failure time, depending on whether the failure time for the i^{th} observation has been observed or censored. After the censored values have been replaced by the appropriate expected values, based on the Kaplan-Meier empirical survival function estimator, the usual least squares method is used to calculate β.

Thus
$$\hat{\beta}^{(k+1)} = [\ (\boldsymbol{X} - \overline{\boldsymbol{X}})^T \cdot (\boldsymbol{X} - \overline{\boldsymbol{X}})\ ]^{-1} \cdot (\boldsymbol{X} - \overline{\boldsymbol{X}})^T \cdot \boldsymbol{y}\,(\hat{\beta}^{(k)}\,)\ . \qquad (1.5)$$

Note that $\hat{\boldsymbol{y}}(\hat{\beta}^{(k)})$ depends on the estimated survival distribution through $\hat{E}[Y_i \mid Y_i > y_i]$, thus requiring an iterative solution of $\hat{\beta}$.

The Koul-Susarla-Van Ryzin (1981) estimator also replaces the censored outcomes. But unlike the Buckley-James procedure which determines these new values based on the distribution function of the failure times, the Koul-Susarla-Van Ryzin estimator determines these new values from the distribution function of the censoring times, G(y), estimated using Bayesian methods. The new values are determined according to the following scheme :

$$\hat{y}_i = \delta_i\, y_i\ \ [1-\hat{G}\,(y_i)\,]^{-1}\cdot\ I\,(\,y_i \geq M), \quad \text{for some specified cutoff } M;$$

where $n^+(y) = \sum_{j=1}^{n} I\{ y_j \geq y\}$ = number of observations at risk just before y,

and $1 - \hat{G}(t) = \prod_{y_i \leq t} \left\{ \frac{1 + n^+(y_i)}{2 + n^+(y_i)} \right\}^{1-\delta_i}$ = survival function of the censoring times.

(1.6)

Then the usual LSE is used to obtain β. Thus observed values are reweighted while censored values are given zero weight. The advantage of this method is that the estimate of $\hat{\beta}$ is in closed form and the solution requires no iteration.

1.6 CRITICISMS OF UNIVARIATE METHODS

In Miller and Halpern's discussion and comparison of these four techniques for covariate analysis of univariate survival data, it appears that the authors prefer the Cox and Buckley-James estimators over the other two, even over co-author Miller's method, though none of the methods is without difficulties. The Koul-Susarla-Van Ryzin (KSV) estimator assumes that the censoring distribution is invariant to changes in $\mathbf{X}_i$. Frequently censoring is related to one of the covariates, such as age, thus invalidating a basic assumption of the KSV technique. There are also no computational formulas developed for the variance-covariance matrix of the covariate parameters when more than one covariate is used. Furthermore, the technique depends on specifying a cut-off point, which is a rather subjective decision. On the other hand, an appealing characteristic of the KSV estimator is that it requires no iteration.

The Miller estimator is also affected by the censoring pattern of the data. For the estimator of $\hat{\beta}$ to be consistent, the censoring distribution must satisfy

$$G(t; \mathbf{X}) = G(t - \alpha - \mathbf{X}\beta)$$

which Miller and Halpern state does not often occur in practice. Since it involves iterative calculation, the estimator is also prone to convergence difficulties.

The Buckley-James estimator is also iterative, but empirically seems to suffer less frequently from a failure to converge than the Miller estimator. Also, it does not depend on censoring patterns in the data. But their estimate of the variance-covariance matrix for the covariate parameters, while useful in practice, is empirically rather than theoretically based, posing some difficulties in the inference process.

The Cox estimator produces an estimator of $\hat{\beta}$ which is asymptotically normal; and the asymptotic variance-covariance matrix is relatively simple to compute. There are also several suggestions regarding how to modify the Cox model in the presence of ties (Peto (1972), Kalbfleisch and Prentice (1972), and Breslow (1974)). Though the lack of an explicit estimate of β is an issue, the primary drawback to the Cox model is the assumption of proportional hazards which is not necessarily appropriate in all situations. The other three estimators are based on a linear approach, which also may not be appropriate in all situations.

1.7 UNIVARIATE SURVIVAL FUNCTION ESTIMATORS

Both the Miller and Buckley-James methods require estimation of the univariate survival function. Though many methods exist, some of the most commonly used estimation techniques when censoring is present include (i) standard life table estimates of survival and (ii) the Kaplan-Meier product limit estimator (Kaplan and Meier, 1958), which is a series approximation of (iii), the Nelson estimator (Nelson, 1969) :

$$\text{(i)}\quad \hat{S}_{LT}(t) = \prod_{j\,:\,t_j < t} \left(1 - \frac{D_j}{N_j - \frac{W_j}{2}} \right) , \tag{1.7}$$

where D_j is the number of deaths $\in$ $[t_{j-1}, t_j)$; N_j is the number of individuals at risk at t_{j-1}; and W_j is the number of individuals censored $\in$ $[t_{j-1}, t_j)$.

$$\text{(ii)}\quad \hat{S}_{KM}(t) = \prod_{j\,:\,t_j < t} \left(\frac{n_j - d_j}{n_j} \right) , \tag{1.8}$$

where n_j is the number of individuals at risk (i.e., alive and not censored) at t_j , and d_j is the number of deaths at t_j ;

and

$$\text{(iii)}\quad \hat{S}_N(t) = \prod_{j\,:\,t_j < t} e^{-\frac{d_j}{n_j}} \tag{1.9}$$

$$= \prod_{j\,:\,t_j < t} \sum_{i=0}^{\infty} \frac{\left(-\frac{d_j}{n_j} \right)^i}{i\,!} .$$

Taking the first two terms of the infinite series for the Nelson estimator leads to the following approximation of the survival function :

$$\hat{S}_N(t) \doteq \prod_{j\,:\,t_j < t} \left(1 - \frac{d_j}{n_j} \right) \quad \text{, the Kaplan-Meier estimator.}$$

1.8 REVIEW OF METHODS FOR BIVARIATE SURVIVAL DATA

The bivariate models to be discussed include the following method, as suggested by Clayton (1978), which is based on the joint distribution function (or joint survival function) :

$$f(s, t) = \frac{\theta \cdot \frac{da(s)}{ds} \cdot \frac{db(t)}{dt}}{\left(1 + (\theta - 1) \cdot \{a(s) + b(t)\}\right)^{2 + \frac{1}{\theta - 1}}}$$

$$\Rightarrow S(s, t) = \frac{1}{\left(1 + (\theta - 1) \cdot \{a(s) + b(t)\}\right)^{\frac{1}{\theta - 1}}} \tag{1.10}$$

where $a(\cdot)$ and $b(\cdot)$ are nondecreasing functions such that $a(0) = b(0) = 0$. The marginals, $a(\cdot)$ and $b(\cdot)$, are not specified, and θ is an unknown parameter of association, equal to the ratio of age-specific incidence rates. Thus

$$\theta = \frac{\lambda_1(s \mid T = t)}{\lambda_1(s \mid T > t)} = \frac{\lambda_2(t \mid S = s)}{\lambda_2(t \mid S > s)}, \quad \text{with } \lambda \text{, the hazard, as defined in equation (1.3).}$$

If S and T are independent, then θ=1 and the joint distribution function is merely the product of the marginals. In addition to the methods for continuous outcomes, Clayton discusses an analogous method for discrete time data. He also includes provisions for covariates using the following reformulations:

$$a(s) = a_0(s) \cdot e^{\beta^T x} \quad \text{and} \quad b(t) = b_0(t) \cdot e^{\gamma^T z}$$

where x and z are covariates for the first and second components, respectively, and β and γ are the corresponding regression coefficients.

In this specification of the model, different covariate may be used for each of the outcome variables. Though this might provide a better fit for the data and may be more appealing from an intuitive standpoint, this type of a model does not really utilize the dependency of the outcome variables.

Another bivariate model, developed by Oakes (1982), presents a reparameterization of Clayton's model but includes known marginal survival distributions, S_1 and S_2, where :

$$S_1(t_1) = [(\theta\text{-}1)\cdot a(t_1)\ +1]^{-\frac{1}{\theta-1}} \quad \text{and} \quad S_2(t_2) = [(\theta\text{-}1)\cdot b(t_2)\ +1]^{-\frac{1}{\theta-1}} ,$$

with θ , $a(\cdot)$, and $b(\cdot)$ as defined previously in (1.10) .

The joint survival distribution is therefore defined as :

$$S(t_1, t_2) = \frac{1}{\left(\left\{ \frac{1}{S_1(t_1)} \right\}^{\theta-1} + \left\{ \frac{1}{S_2(t_2)} \right\}^{\theta-1} - 1 \right)^{\frac{1}{\theta-1}}} , \tag{1.11}$$

where θ is again unknown, and is still the ratio of age-specific incidence rates. Techniques for estimating θ and inferences based on θ are discussed by both Clayton and Oakes. In both formulations, as $\theta \rightarrow 1^+$ (i.e. independence), the joint survival distribution becomes the product of the marginals. Oakes discusses the case of exponential marginals in much detail, which leads to the following model:

$$S(t_1, t_2) = \frac{1}{\left\{ e^{\alpha(\theta-1)t_1} + e^{\gamma(\theta-1)t_2} - 1 \right\}^{\frac{1}{\theta-1}}}$$

After some algebraic manipulation (see Appendix A), it can be shown that the correlation between the two outcome variables is solely a function of θ, as follows:

$$\text{corr}(t_1, t_2) = \frac{1}{2(1-\theta)^2}\left\{\psi'\left(\frac{1}{2(\theta-1)}\right) - \psi'\left(\frac{\theta}{2(\theta-1)}\right)\right\} - 1 \quad ,$$

where $\psi'(z) = \sum_{n=0}^{\infty} \frac{1}{(n+z)^2}$ is the trigamma function .

Huster (1989) developed an extension of the Clayton-Oakes model to include covariates by allowing the marginal survival functions $S_1(t_1, \beta_1)$ and $S_2(t_2, \beta_2)$ to depend on covariates. The form of these survival functions is not specified; in other words, the inclusion of covariates need not take on a specific form. Huster also examines the use of specific distributions, including the bivariate Weibull and bivariate normal. It should be noted that none of these three models makes any provision for censoring. An advantage of Huster's model is that it allows marginals to come from different families and, as with Clayton's model allowing for covariates, the model does not restrict the covariates to be the same for both components.

Wei, Lin, and Weissfeld (1989) developed a method for covariate analysis with multivariate failure time data by modeling the marginal distribution of each outcome variable using a Cox proportional hazards model. Though this method is intended as a multivariate approach, it really involves the simultaneous estimation of the marginal distributions rather than the estimation of the joint distribution of the distribution of the failure times. The authors developed procedures for hypothesis testing in the multivariate framework. Again, as with the univariate situation, the model assumes proportional hazards for each failure time component.

Though there is potential for bivariate extensions of the four univariate models presented previously, these methods have not all been developed. The work to be discussed in more detail in subsequent chapters deals primarily with a bivariate extension of the Buckley-James estimator; various modifications will be considered using semi-parametric and parametric models. In the univariate case censored data points are replaced by their empirical expected values, conditional on the censoring time and the covariates, determined from the Kaplan-Meier estimate of survival. For the bivariate nonparametric situation, since the outcome times are correlated, it is also necessary to condition the expected value of one component on the failure or censoring time of the other component. Here the empirical estimates are determined from a modification of the Campbell-Földes estimates of survival (due to Burke, 1988), which are conceptually similar to the Kaplan-Meier estimates.

In the parametric situation, censored data points are also replaced by their conditional expected values. But instead of using empirical values, the expected values are based on a specified parametric distribution, with the bivariate exponential and bivariate Weibull distributions being of greatest interest. These methods, as well as estimates of the variance-covariance matrix and inferences based on the estimated parameters, will be discussed in the remainder of this research.

CHAPTER 2. EMPIRICAL SURVIVAL FUNCTION ESTIMATORS

2.1 DISCUSSION OF BIVARIATE SURVIVAL FUNCTION ESTIMATORS

The nonparametric method of bivariate data analysis to be discussed in later chapters requires empirical survival function estimates. The technique selected for estimation of the survival function forms the basis for the covariate analysis, and thus has a significant influence on the conclusions. Though the bivariate estimators that have been suggested are quite varied, many are based, at least in part, on the univariate Kaplan-Meier estimator (1.8) of the survival function. In the remainder of this chapter we describe several estimation methods and discuss the selected method in detail.

Dabrowska (1988), in an attempt to rectify some of the problems encountered with other bivariate survival estimators, such as lack of monotonicity for each coordinate, non-uniqueness, failure to reduce to the standard empirical survival function in the absence of censoring, and path-dependency, developed a bivariate Kaplan-Meier estimate. While this estimator is unique and consistent, unfortunately it may also fail to be monotone in the presence of censoring, which is one of Dabrowska's complaints concerning other estimators.

Another bivariate empirical survival function estimator, developed by Hanley and Parnes (1983), is based on the assignment of probability mass to the various support points in the plane and summing over $(t_1, \infty) \times (t_2, \infty)$ to obtain an estimate for $S(t_1, t_2)$. If the censoring is homogeneous, a specified censoring pattern defined by Hanley and Parnes, then a closed form solution exists for the estimate of the survival function, but this censoring pattern is not very general. An example of homogeneous censoring is when all subjects have equal follow-up times. More realistically, the censoring is heterogeneous. Under this condition a closed-form expression for the estimator does not exist, and the estimate must be found using an appropriate iterative method, such as the Newton-Raphson method or the EM algorithm. In an example discussed by Hanley and Parnes, the convergence was very slow, reaching a solution in the heterogeneous situation in 50 iterations of the EM algorithm.

Tsai, Leurgens and Crowley (1986) proposed a family of estimators for the bivariate survival function, based on a decomposition formula and smoothing techniques. The family of estimators all have a closed-form expression, but they are kernel and bandwidth dependent. The authors claim this is much less of a concern than the path-dependency problem of other estimators. Another drawback to this estimator is that it has a very slow rate of almost sure consistency.

A modified form of the Campbell-Földes bivariate survival function (1980) due to

Burke (1988) satisfies many of the desirable properties and will be used extensively in the present study. The model is described in detail below. Additional modifications to include ties will be presented in Section 4 of this chapter.

2.2 CAMPBELL-FÖLDES BIVARIATE SURVIVAL FUNCTIONS ESTIMATORS

The Campbell-Földes estimators are based on the identity $\hat{S}_1(s, t) = \hat{S}_1(s) \cdot \hat{S}(t \mid s)$.

To discuss the estimators in any more detail entails usage of the following notation:

Y^0_{ij} = failure time for j^{th} component of i^{th} observation $\quad (i = 1, 2, \ldots, n;\ j = 1, 2)$;

C_{ij} = censoring time for j^{th} component of i^{th} observation $\quad (i = 1, 2, \ldots, n;\ j = 1, 2)$;

$Y_{ij} = \min(Y^0_{ij}, C_{ij})$ = observed stopping time (i.e., either time of censoring or failure time) for j^{th} component of i^{th} observation ;

$\mathbf{Y}_i = (Y_{i1}, Y_{i2})$ = vector of observed stopping times ;

$$\delta_{ij} = I(Y_{ij} = Y^0_{ij}) = \begin{cases} 0, & \text{if } j^{th} \text{ component of } i^{th} \text{ observation is censored} \\ 1, & \text{if failure time for } j^{th} \text{ component of } i^{th} \text{ observation is observed} \end{cases} \quad ;$$

$$\alpha_i(s,t) = I\{Y_{i1} \le s,\ Y_{i2} > t,\ \delta_{i1} = 1\} = \begin{array}{c}\text{indicator of the first component failing by} \\ \text{s and the second component surviving t}\end{array} \quad ;$$

$$\gamma_i(s,t) = I\{Y_{i1} > s,\ Y_{i2} \le t,\ \delta_{i2} = 1\} = \begin{array}{c}\text{indicator of the second component failing} \\ \text{by t and the first component surviving s}\end{array} \quad ; \quad \text{and}$$

$$N(s,t) = \sum_{i=1}^{n} I\{Y_{i1} > s,\ Y_{i2} > t\} = \#\text{ pairs at risk at } (s,\ t) \ .$$

The Campbell-Földes survival function estimator is

$$\hat{S}_1(s,t) = \prod_{i=1}^{n} \left\{ \frac{N(y_{i1}\ ,\ 0)}{N(y_{i1}\ ,\ 0) + 1} \right\}^{\alpha_i(s,\ 0)} \cdot \prod_{k=1}^{n} \left\{ \frac{N(s,\ y_{k2})}{N(s,\ y_{k2}) + 1} \right\}^{\gamma_k(s,t)} ,$$

which is the product of the marginal survival for the first component and the conditional survival for the second component given that the first component fails at s.

Thus

$$\boxed{\hat{S}_1(s,t) = \prod_{i=1}^{n}\left\{1 - \frac{1}{N(y_{i1}\ ,\ 0) + 1}\right\}^{\alpha_i(s,\ 0)} \cdot \prod_{k=1}^{n}\left\{1 - \frac{1}{N(s,\ y_{k2}) + 1}\right\}^{\gamma_k(s,t)}}$$

(2.1a)

Similarly, $\hat{S}_2(s,t) = \hat{S}_2(t) \cdot \hat{S}(s \mid t)$, which leads to :

$$\boxed{\hat{S}_2(s,t) = \prod_{i=1}^{n}\left\{1 - \frac{1}{N(0\ ,\ y_{i2}) + 1}\right\}^{\gamma_i(0,\ t)} \cdot \prod_{k=1}^{n}\left\{1 - \frac{1}{N(y_{k1},\ t) + 1}\right\}^{\alpha_k(s,t)}}$$

(2.1b)

As is evident from the two estimators in equations (2.1a) and (2.1b), which are not equivalent, the result will necessarily be dependent on the path chosen. Another disadvantage of this estimator is that the estimated survival distribution, computed from either path, may not be jointly monotonically decreasing. In addition, while this estimator is relatively simple to determine, involving counting only the number of observations satisfying certain criteria, and the computation of α and γ, it requires up to n recalculations at each value of s (for 2.1a) leading to n^2 possible calculations for the survival function estimator for each pair of (s, t) values.

2.3 BURKE BIVARIATE SURVIVAL FUNCTION ESTIMATOR

In an attempt to resolve the difficulty with monotonicity encountered with the Campbell-Földes estimator, Burke (1988) proposed a similar estimator which will be the basis of parameter estimation throughout subsequent sections of this work. Both survival function estimators are based on the concept that the joint survival of two random variables is equal to the product of the marginal survival distribution of one variable and the survival distribution of the other variable conditioned on the first. But, whereas the Campbell-Földes estimator is based on the survival function of the failure times, the Burke estimator is based on the survival function of the censoring times. The Burke survival function estimator is defined as :

$$\boxed{\hat{S}(s, t) = 1 - \hat{F}(s, \infty) - \hat{F}(\infty, t) + \hat{F}(s, t)} \qquad (2.2)$$

where $\hat{F}(s, t) = \frac{1}{n} \cdot \sum_{i=1}^{n} \delta_{i1}\, \delta_{i2} \frac{I\,(Y_{i1} \leq s, Y_{i2} \leq t)}{G_1(y_{i1}, y_{i2})}$ with

$\hat{G}_1(s, t) = \hat{G}_1(s, 0) \cdot \hat{G}(t \mid s)$ = bivariate survival of censoring times ; (2.3)

$\hat{G}_1(s, 0) = \prod_{i=1}^{n} \left\{1 - \frac{1}{N(y_{i1}, 0) + 1}\right\}^{\alpha_i^*(s,\,0)}$ = marginal survival of censoring time for first component ;

$\hat{G}(t \mid s) = \prod_{j=1}^{n} \left\{1 - \frac{1}{N(s, y_{j2}) + 1}\right\}^{\gamma_j^*(s,t)}$

= conditional survival of censoring time for second component, given that first component is censored at s ;

$\alpha_i^*(s,t) = I\{Y_{i1} \leq s, Y_{i2} > t, \delta_{i1} = 0\}$; and

$\gamma_i^*(s,t) = I\{Y_{i1} > s, Y_{i2} \leq t, \delta_{i2} = 0\}$.

The definition of $\hat{G}_1(s, t)$ is similar to $\hat{S}_1(s, t)$ as defined in (2.1a) except that the definitions of α and γ have been altered to include only censored observations rather than observed failures; the definitions of δ, the indicators for censoring, and $N(s, t)$, the number of pairs at risk, are as stated previously. Burke's estimator could also be defined in terms of $\hat{G}_2(s, t) = \hat{G}_2(0, t) \cdot \hat{G}(s \mid t)$; thus the path-dependency problem of the Campbell-Földes estimator was not resolved. To resolve the path-dependency problem, the survival function at each time point could be estimated as the arithmetic mean of $\hat{G}_1(s, t)$ and $\hat{G}_2$ (s, t).

2.4 ADJUSTMENTS OF BURKE'S ESTIMATORS TO INCLUDE TIES

Both the Campbell-Földes estimates and the Burke estimates assume that all outcome times are distinct; thus they are applicable only in the situation where no ties occur. In practical scenarios this is rarely the case, so an extension has been developed here to include ties. The marginal survival function in (2.3), $\hat{G}(s)$, is easily adjusted to account for ties, as the univariate Kaplan-Meier estimator for tied data can be used. But the conditional survival function must also be adjusted, which is a bit more involved. Some additional notation is required, as follows:

m_j = number of distinct observed times of j^{th} component (failure or censored) :

$$\text{ordered } \{ y^*_{(1)j}, \ldots, y^*_{(i)j}, \ldots, y^*_{(m_j)j} \} \quad (j = 1, 2) \;;$$

m = total number of distinct pairs of observed times :

$$\{ (y_{(1)1}, y_{(1)2}), \ldots, (y_{(i)1}, y_{(i)2}), \ldots, (y_{(m)1}, y_{(m)2}) \} \;;$$

$d^*_j(s) = \sum_{i=1}^{n} I\{Y_{ij} = s, \delta_{ij} = 0\}$ = number of j^{th} components censored at s (j = 1, 2);

$$d_1^*(s,t) = \sum_{i=1}^{n} I\{Y_{i1} = s,\ \delta_{i1} = 0,\ Y_{i2} = t\} = \text{number of observations at } (s,\ t) \text{ with } 1^{st} \text{ component censored} \ ;$$

$$d_2^*(s,t) = \sum_{i=1}^{n} I\{Y_{i1} = s,\ Y_{i2} = t,\ \delta_{i2} = 0\} = \text{number of observations at } (s,\ t) \text{ with } 2^{nd} \text{ component censored} \ ;$$

$$n_j(s) = \sum_{i=1}^{n} I\{Y_{ij} \geq s\} = \text{number of } j^{th} \text{ components at risk at } s^- \ ;$$

$$n(s,t) = \sum_{i=1}^{n} I\{Y_{i1} \geq s,\ Y_{i2} \geq t\} = \text{number of units at risk at } (s^-,\ t^-) \ ;$$

$$\alpha_{(i)}^*(s,\ t) = I\{Y_{(i)1} \leq s,\ Y_{(i)2} > t,\ d_1^*(y_{(i)1}) > 0\} \ ;$$

$$\alpha_{(i)}^*(s) = \alpha_{(i)}(s,\ 0) = I\{Y_{(i)1}^* \leq s,\ d_1^*(y_{(i)1}^*) > 0\} \ ;$$

$$\gamma_{(j)}^*(s,\ t) = I\{Y_{(j)1} > s,\ Y_{(j)2} \leq t,\ d_2^*(y_{(j)2}) > 0\} \ ; \quad \text{and}$$

$$\gamma_{(j)}^*(t) = \gamma_{(j)}(0,\ t) = I\{Y_{(j)2}^* \leq t,\ d_2^*(y_{(j)2}^*) > 0\} \ .$$

The joint survival functions for the censoring times, needed for the Burke estimators, allowing for ties, are :

$$\boxed{\begin{aligned} \hat{G}_1(s,\ t) &= \prod_{i=1}^{m_1} \left\{1 - \frac{d_1^*(y_{(i)1}^*)}{n_1(y_{(i)1}^*)}\right\}^{\alpha_{(i)}^*(s)} \cdot \prod_{j=1}^{m} \left\{1 - \frac{\sum_{y_{(j)1} > s} d_2^*(y_{(j)1},\ y_{(j)2})}{n(s,\ y_{(j)2})}\right\}^{\gamma_{(j)}^*(s,t)} \\ \hat{G}_2(s,\ t) &= \prod_{j=1}^{m_2} \left\{1 - \frac{d_2^*(y_{(j)2}^*)}{n_2(y_{(j)2}^*)}\right\}^{\gamma_{(j)}^*(s)} \cdot \prod_{i=1}^{m} \left\{1 - \frac{\sum_{y_{(i)2} > t} d_1^*(y_{(i)1},\ y_{(i)2})}{n(y_{(i)1},\ t)}\right\}^{\alpha_{(i)}^*(s,t)} \end{aligned}} \tag{2.4}$$

Then using the values calculated above, the survival function for the failure times is just as calculated previously :

$$\hat{S}(s, t) = 1 - \hat{F}(s,\infty) - \hat{F}(\infty, t) + \hat{F}(s, t) ,$$

$$\text{where } \hat{F}(s, t) = \frac{1}{n} \sum_{i=1}^{n} \delta_{i1}\, \delta_{i2} \frac{I\,(Y_{i1} \le s,\, Y_{i2} \le t)}{G(y_{i1}, y_{i2})} .$$

Clearly, the survival function estimates provided in (2.4) demonstrate that the estimator is still path-dependent. As suggested previously, replacing the estimate by the mean of the two path-dependent estimates at each point would result in a non-path-dependent estimate.

2.5 COMPARISON OF THE BIVARIATE EMPIRICAL SURVIVAL FUNCTION BASED ON THE CAMPBELL-FÖLDES ESTIMATE AND THE BURKE ESTIMATE

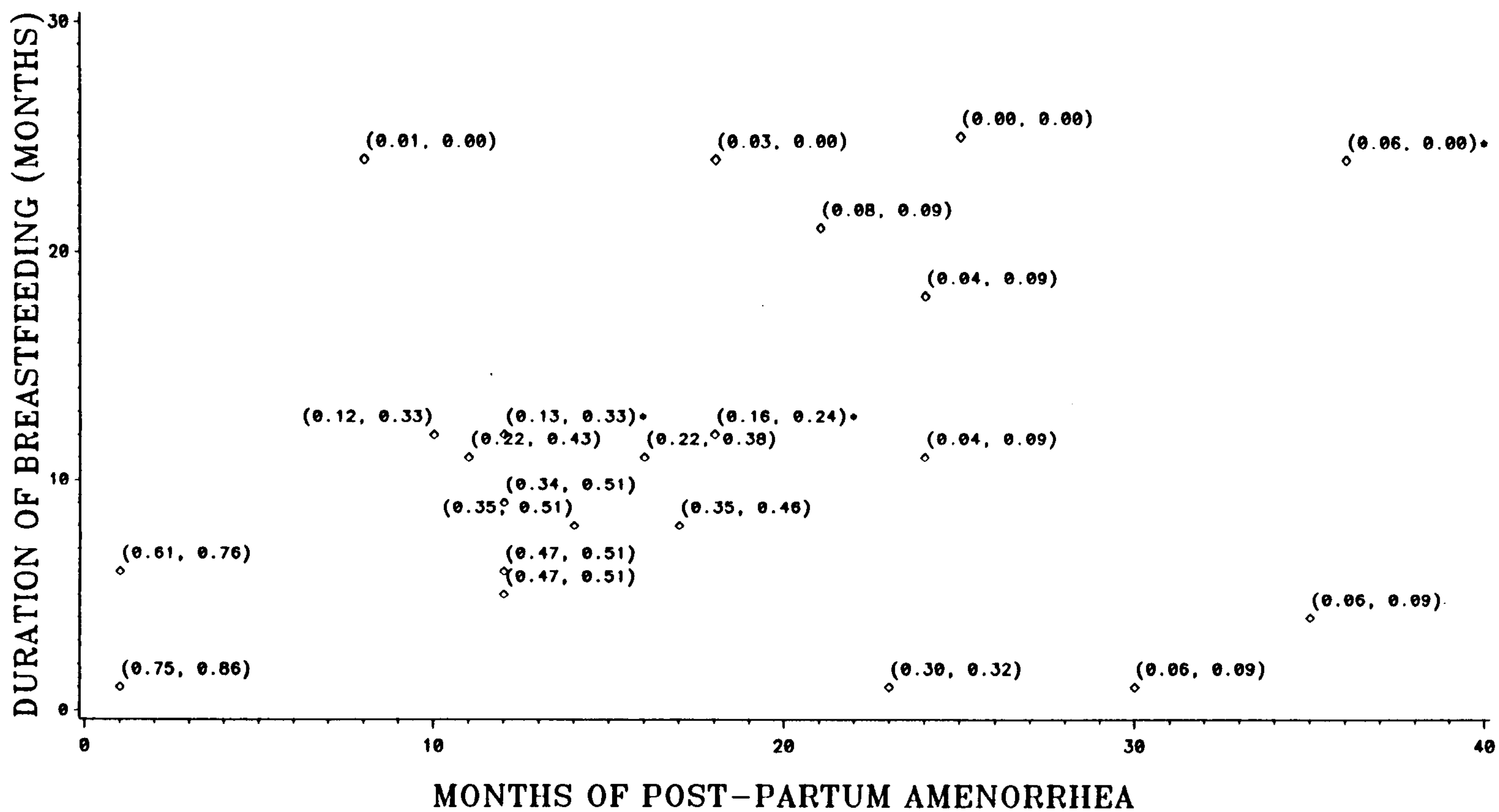
Using a sample of 25 observations on postpartum amenorrhea and breastfeeding obtained from a fertility survey in Ghana (which will also be used for subsequent analyses) yields estimates of the bivariate survival functions found in Table 2.1. From the graph in Fig 2.1 it is clearly demonstrated that there are several points where the Campbell-Földes estimator is not monotonically decreasing though the Burke estimator is.

TABLE 2.1

ESTIMATES OF THE BIVARIATE SURVIVAL FUNCTION WITH TIES IN THE DATA, USING THE CAMPBELL-FÖLDES AND BURKE METHODS OF ESTIMATION

(Y_1, Y_2)	(δ_1, δ_2)	Campbell-Földes estimator of $\hat{S}_1$	Burke estimator of $\hat{S}_1$
(1, 1)	(0, 0)	0.7475	0.8565
(1, 6)	(1, 1)	0.6133	0.7644
(8, 24)	(1, 1)	0.0099	0.0000
(10, 12)	(1, 1)	0.1239	0.3340
(11, 11)	(0, 0)	0.2216	0.4305
(11, 11)	(0, 0)	0.2216	0.4305
(12, 5)	(1, 1)	0.4665	0.5101
(12, 6)	(1, 1)	0.4665	0.5101
(12, 9)	(1, 1)	0.3427	0.5101
(12, 12)	(1, 1)	0.1318	0.3340
(14, 8)	(0, 1)	0.3494	0.5101
(16, 11)	(1, 1)	0.2235	0.3844
(17, 8)	(0, 1)	0.3539	0.4608
(18, 12)	(1, 1)	0.1632	0.2419
(18, 24)	(1, 1)	0.0345	0.0000
(21, 21)	(0, 0)	0.0839	0.0892
(22, 22)	(0, 0)	0.0625	0.0892
(23, 1)	(0, 1)	0.2989	0.3182
(24, 11)	(1, 1)	0.0437	0.0892
(24, 18)	(1, 1)	0.0437	0.0892
(24, 18)	(1, 1)	0.0437	0.0892
(25, 25)	(0, 0)	0.0029	0.0000
(30, 1)	(1, 1)	0.0555	0.0892
(35, 4)	(0, 1)	0.0555	0.0892
(36, 24)	(1, 1)	0.0555	0.0000

FIGURE 2.1. PLOT OF THE ESTIMATED BIVARIATE SURVIVAL FUNCTION WITH TIES IN THE DATA, USING THE CAMPBELL-FOLDES AND BURKE METHODS OF ESTIMATION



THE VECTOR INDENTIFYING EACH POINT IS COMPRISED OF THE CAMPBELL-FOLDES AND THE MODIFIED BURKE ESTIMATES OF THE BIVARIATE SURVIVAL FUNCTION. ASTERISKS INDICATE WHERE THE CAMPBELL-FOLDES ESTIMATE IS NOT MONOTONIC

CHAPTER 3. EXTENSION OF BUCKLEY-JAMES PROCEDURE FROM UNIVARIATE TO BIVARIATE OUTCOMES (SEMI-PARAMETRIC MODEL)

3.1 DESCRIPTION OF MODEL

In this chapter we extend the univariate Buckley-James method described in Chapter 1 to the bivariate framework. Using additional notation and definitions the model for bivariate outcomes can be written as :

$$\mathbf{Y}_i = \mathbf{X}_i \cdot \boldsymbol{\beta} + \epsilon_i \ ,$$

where

Y^O_{ij} = failure time for j^{th} component of i^{th} observation $\quad (i = 1, 2, \ldots, n;\ j = 1, 2)$;

C_{ij} = censoring time for j^{th} component of i^{th} observation $\quad (i = 1, 2, \ldots, n;\ j = 1, 2)$;

$Y_{ij} = \min(Y^O_{ij}, C_{ij})$ = observed stopping time (i.e. failure time or time of censoring)

$\mathbf{Y}_i = (Y_{i1}, Y_{i2})$ = vector of observed stopping times ;

$$\delta_{ij}=I(Y_{ij}=Y^O_{ij}) = \begin{cases} 0, \text{ if } j^{th} \text{ component of } i^{th} \text{ observation is censored} \\ 1, \text{ if failure time for } j^{th} \text{ component of } i^{th} \text{ observation is observed} \end{cases} ;$$

$\boldsymbol{\beta} = \boldsymbol{\beta}_1 \,||\, \boldsymbol{\beta}_2 =$ vertical concatenation of parameters for 1^{st} component with parameters for 2^{nd} component ,

$\epsilon_i \sim f(\cdot,\cdot)$, and

$$S(\mathbf{y}; \mathbf{x}) = \int_{\mathbf{y}}^{\infty} f(\mathbf{t}; \mathbf{x})\, dt \ .$$

Note that $\mathbf{Y}_i$ contains censored observations. In order to apply the Buckley-James procedure, we need to replace censored values with the appropriate expectations.

$$\hat{y}_{ij} = \delta_{ij} \cdot y^O_{ij} + (1-\delta_{ij}) \cdot \delta_{ij'} \cdot \mathrm{E}[\, Y^O_{ij} \mid Y^O_{ij} > c_{ij} \,,\, Y^O_{ij'} = y^O_{ij'} \,;\, X\hat{\beta}]$$

$$+ \; (1-\delta_{ij}) \cdot (1-\delta_{ij'}) \cdot \hat{\mathrm{E}}[\, Y^O_{ij} \mid Y^O_{ij} > c_{ij} \,,\, Y^O_{ij'} > c_{ij'} \,;\, X\hat{\beta}] \;,$$

Then the covariate parameter estimates can be calculated : $\hat{\beta} \; = \; (\mathbf{X}^T \cdot X)^{-1} \cdot X^T \cdot \hat{\mathbf{y}}(\hat{\beta})$

The conditional expectations are derived below.

3.2 ESTIMATION OF EXPECTATIONS AND COVARIATE PARAMETERS

The expected values are calculated according to the following methods:

1) **If both components are censored** (i.e. $\delta_i = (0, 0)$) then the conditional expected value of the pair of failure times for the i^{th} individual is calculated as follows :

$$\mathrm{E}(\, \mathbf{Y}^O_i \mid \mathbf{Y}^O_i > \mathbf{y}_i \,) \quad = \int_{k=0}^{\infty} \mathbf{k} \cdot \mathrm{pr}\,(\, \mathbf{Y}^O_i = \mathbf{k} \mid \mathbf{Y}^O_i > \mathbf{y}_i \,) \;\; \mathbf{dk}$$

$$= \int_{k=y_i}^{\infty} \frac{\mathbf{k} \cdot \mathrm{pr}\,(\, \mathbf{Y}^O_i = \mathbf{k}\,)}{\mathrm{pr}\,(\, \mathbf{Y}^O_i > \mathbf{y}_i)} \quad \mathbf{dk} \;.$$

Using the density of the error terms, $f(\cdot)$, and the corresponding survival function, $S(\cdot)$, the conditional expected value can be rewritten as

$$\mathrm{E}(\, \mathbf{Y}^O_i \mid \mathbf{Y}^O_i > \mathbf{y}_i \,) \quad = \frac{\int_{k=y_i}^{\infty} \mathbf{k} \cdot \mathrm{pr}\,(\, \mathbf{Y}^O_i = \mathbf{k}\,) \;\; \mathbf{dk}}{S(\mathbf{y}_i - \mathbf{X}_i\, \beta)}$$

$$= \frac{\int_{k=y_i}^{\infty} \mathbf{k} \cdot f\,(\, \mathbf{k} - \mathbf{X}_i\, \beta\,) \;\; \mathbf{dk}}{S(\mathbf{y}_i - \mathbf{X}_i\, \beta)}$$

Using a change of variables technique with $r = k - X_i\beta$ ($\Rightarrow$ $dr = dk$) leads to :

$$E(Y_i^O \mid Y_i^O > y_i) = \frac{\left\{ \int_{r = y_i - X_i\beta}^{\infty} (X_i\beta + r) \cdot f(r)\, dr \right\}}{S(y_i - X_i\beta)}$$

$$= \frac{X_i\beta \cdot \{S(y_i - X_i\beta)\} + \left\{ \int_{r = y_i - X_i\beta}^{\infty} r \cdot f(r)\, dr \right\}}{S(y_i - X_i\beta)}$$

$$= X_i\beta + \frac{\left\{ \int_{r = y_i - X_i\beta}^{\infty} r \cdot f(r)\, dr \right\}}{S(y_i - X_i\beta)} .$$

The conditional expected values are written for each component separately are :

$$E\left(\begin{pmatrix} Y_{i1}^O \\ Y_{i2}^O \end{pmatrix} \middle| \begin{pmatrix} Y_{i1}^O \\ Y_{i2}^O \end{pmatrix} > \begin{pmatrix} y_{i1} \\ y_{i2} \end{pmatrix} \right)$$

$$= \begin{pmatrix} X_i\beta_1 + \dfrac{\left\{ \int_{r_1 = y_{i1} - X_i\beta_1}^{\infty} \int_{r_2 = y_{i2} - X_i\beta_2}^{\infty} r_1 \cdot f(r_1, r_2)\, dr_1\, dr_2 \right\}}{S(y_i - X_i\beta)} \\[2ex] X_i\beta_2 + \dfrac{\left\{ \int_{r_1 = y_{i1} - X_i\beta_1}^{\infty} \int_{r_2 = y_{i2} - X_i\beta_2}^{\infty} r_2 \cdot f(r_1, r_2)\, dr_1\, dr_2 \right\}}{S(y_i - X_i\beta)} \end{pmatrix} . \qquad (3.1)$$

The numerator of the first component of (3.1) , $\left\{ \int_{r_1}^{\infty} \int_{r_2}^{\infty} r_1 \cdot f(r_1, r_2)\, dr_1\, dr_2 \right\}$,
can be estimated empirically by $\left\{ \sum_{\{j \,:\, \hat{e}_j \geq \hat{e}_i\}} \hat{e}_{j1} \cdot w_j(\hat{\beta}) \right\}$, a weighted sum of the residuals summed over the appropriate risk set, where

$$w_j(\hat{\beta}) = \hat{F}(\hat{e}_{j1}, \hat{e}_{j2}; \hat{\beta}) - \hat{F}(\hat{e}_{j1}, \hat{e}_{j2}^-; \hat{\beta}) - \hat{F}(\hat{e}_{j1}^-, \hat{e}_{j2}; \hat{\beta}) + \hat{F}(\hat{e}_{j1}^-, \hat{e}_{j2}^-; \hat{\beta})$$

$$= \hat{S}(\hat{e}_{j1}^-, \hat{e}_{j2}^-; \hat{\beta}) - \hat{S}(\hat{e}_{j1}^-, \hat{e}_{j2}; \hat{\beta}) - \hat{S}(\hat{e}_{j1}, \hat{e}_{j2}^-; \hat{\beta}) + \hat{S}(\hat{e}_{j1}, \hat{e}_{j2}; \hat{\beta})$$

= probability mass assigned to $(\hat{e}_{j1}, \hat{e}_{j2})$.

An analogous estimate applies for the numerator of the second component.

Thus $\quad \hat{E}(Y_i^O \mid Y_i^O > y_i) = X_i \hat{\beta} + \dfrac{\left\{ \sum_{\{j \,:\, \hat{e}_j \geq \hat{e}_i\}} \hat{e}_j \cdot w_j(\hat{\beta}) \right\}}{\hat{S}(y_i - X_i \hat{\beta})} \qquad (3.2)$

if both components are censored.

2) **If only one component is censored**, (say Y_{i1}^O), then the conditional expected value of the first component is calculated as :

$$E(Y_{i1}^O \mid Y_{i1}^O > y_{i1} , Y_{i2}^O = y_{i2}) = \int_{k = y_{i1}}^{\infty} \frac{k \cdot pr (Y_{i1}^O = k , Y_{i2}^O = y_{i2})}{pr (Y_{i1}^O > y_{i1} , Y_{i2}^O = y_{i2})} \, dk .$$

Letting $r = k - X_i\beta_1$ leads to:

$$E(Y_{i1}^O \mid Y_{i1}^O > y_{i1} , Y_{i2}^O = y_{i2}) = X_i \beta_1 + \frac{\left\{ \int_{r = y_{i1} - X_i\beta_1}^{\infty} r \cdot f (r, y_{i2} - X_i \beta_2) \, dr \right\}}{pr (Y_{i1}^O > y_{i1} , Y_{i2}^O = y_{i2})} . \tag{3.3}$$

With e_j and $w_j(\hat{\beta})$ defined as before, the numerator of (3.3) can be estimated empirically as:

$$\left\{ \sum_{\{j \,:\, \hat{e}_{j1} \geq \hat{e}_{i1},\ \hat{e}_{j2} = \hat{e}_{i2}\}} \hat{e}_{j1} \cdot w_j(\hat{\beta}) \right\} ,$$

a weighted sum of the residuals, though summed over a different risk set than in (3.2). The denominator of (3.3) can be empirically estimated by $\hat{S}(\hat{e}_{i1}, \hat{e}_{i2}^- ; \hat{\beta}) - \hat{S}(\hat{e}_{i1}, \hat{e}_{i2}; \hat{\beta})$. An analogous formula applies if only Y_{i2}^O is censored.

Thus for each of the four cases of censoring we have the following values for $\hat{\boldsymbol{y}}$:

1) If both components are observed (i.e. $\delta_{i1} = \delta_{i2} = 1$), then neither component has to be estimated and :

$$\boxed{\hat{\boldsymbol{y}}_i\,(\hat{\boldsymbol{\beta}}) = \boldsymbol{y}_i\,.}$$

2) If the first component is censored and the second observed ($\delta_{i1} = 0$ and $\delta_{i2} = 1$), then

$$\hat{y}_{i2}\,(\hat{\boldsymbol{\beta}}) = y_{i2} \quad \text{and}$$

$$\hat{y}_{i1}\,(\hat{\boldsymbol{\beta}}) = \boldsymbol{X}_i\;\hat{\boldsymbol{\beta}}_1 \quad + \quad \left\{ \frac{\displaystyle\sum_{\{j:\, \hat{e}_{j1} \geq \hat{e}_{i1},\, \hat{e}_{j2} = \hat{e}_{i2}\}} [\, \hat{\mathrm{e}}_{j1} \cdot \mathrm{w}_j(\hat{\boldsymbol{\beta}})\,]}{\hat{\mathrm{S}}\,(\,\hat{\mathbf{e}}_{i1},\ \hat{\mathbf{e}}_{\overline{i2}};\ \hat{\boldsymbol{\beta}}) \ - \ \hat{\mathrm{S}}\,(\,\hat{\mathbf{e}}_{i1}, \hat{\mathbf{e}}_{i2}\,;\hat{\boldsymbol{\beta}})} \right\}.$$

3) If the first component is observed and the second censored ($\delta_{i1} = 1$ and $\delta_{i2} = 0$), then

$$\hat{y}_{i1}\,(\hat{\boldsymbol{\beta}}) = y_{i1} \quad \text{and}$$

$$\hat{y}_{i2}\,(\hat{\boldsymbol{\beta}}) = \boldsymbol{X}_i\;\hat{\boldsymbol{\beta}}_2 \quad + \quad \left\{ \frac{\displaystyle\sum_{\{j:\, \hat{e}_{j2} \geq \hat{e}_{i2},\, \hat{e}_{j1} = \hat{e}_{i1}\}} [\, \hat{\mathrm{e}}_{j2} \cdot \mathrm{w}_j(\hat{\boldsymbol{\beta}})\,]}{\hat{\mathrm{S}}\,(\,\hat{\mathbf{e}}_{\overline{i1}}, \hat{\mathbf{e}}_{i2}\,;\hat{\boldsymbol{\beta}}) \ - \ \hat{\mathrm{S}}\,(\,\hat{\mathbf{e}}_{i1},\ \hat{\mathbf{e}}_{i2};\ \hat{\boldsymbol{\beta}})} \right\}.$$

4) And finally if both components are censored ($\delta_{i1} = \delta_{i2} = 0$), both components must be estimated, and

$$\boxed{\hat{y}_i(\hat{\beta}) = X_i\ \hat{\beta} \quad + \quad \left\{ \frac{\sum_{\{j:\ \hat{e}_j \geq \hat{e}_i\}} [\ \hat{e}_j \cdot w_j(\hat{\beta})\ ]}{\hat{S}(\ \hat{e}_{i1},\ \hat{e}_{i2};\ \hat{\beta})} \right\} .}$$

As the covariate parameter estimates depend on the expected failure times, and the expectations of failure times depend on the parameter estimates, an iterative process must be used to solve for $\hat{\beta}$. Thus let

$$\begin{aligned} \hat{\beta}^{(k+1)} &= \text{values of } \hat{\beta} \text{ at } (k+1)^{st} \text{ iteration of algorithm} \\ &= (X^T \cdot X)^{-1} \cdot X^T \cdot \hat{y}(\hat{\beta}^{(k)}) \qquad (k = 0, 1, 2, \ldots) \\ &= (\hat{\beta}_1^{(k+1)} \,||\, \hat{\beta}_2^{(k+1)}) . \end{aligned}$$

Initially all **y** values are treated as uncensored to obtain a starting value of

$$\hat{\beta}^{(0)} = (X^T \cdot X)^{-1} \cdot X^T \cdot y^O .$$

Then the predicted conditional expected values are updated by calculating the conditional expected values based on the covariate parameter estimates from the current iteration. Thus

$$\begin{aligned} \hat{y}_{ij}^{(k)} = \delta_{ij} \cdot y_{ij}^O &+ (1-\delta_{ij}) \cdot \delta_{ij'} \cdot \hat{E}[\ Y_{ij}^O \mid Y_{ij}^O > c_{ij},\ Y_{ij'}^O = y_{ij'}^O ;\ X\hat{\beta}^{(k)}] \\ &+ (1-\delta_{ij}) \cdot (1-\delta_{ij'}) \cdot \hat{E}[\ Y_{ij}^O \mid Y_{ij}^O > c_{ij},\ Y_{ij'}^O > c_{ij'} ;\ X\hat{\beta}^{(k)}] , \qquad k = 1, 2, \ldots ; \end{aligned}$$

and the estimation process is repeated until $||\ \hat{\beta}^{(k+1)} - \hat{\beta}^{(k)}|| \rightarrow \eta$, for η small. In the univariate case the estimate may never approach a limiting value, but instead oscillate between two values. This possibility also exists for the bivariate situation; if so, then a mean of the two values could be used as an estimate.

3.3 VARIANCE ESTIMATION

To obtain the variance-covariance matrix for the covariate parameters, the matrix of covariate parameters must first be transformed into a vector.

Let $\hat{\beta}^* = (\hat{\beta}_1 // \hat{\beta}_2)$

$= \text{vec} (\hat{\beta})$

$=$ horizontal concatenation of columns of $\hat{\beta}$.

In a general multivariate linear model , $\hat{V}(\hat{\beta}^*) = \hat{\Sigma} \otimes (X^T \cdot X)^{-1}$,

$$\text{where} \quad \hat{\Sigma} = \begin{bmatrix} \hat{V}(Y_1) & \hat{Cov}(Y_1, Y_2) \\ \hat{Cov}(Y_1, Y_2) & \hat{V}(Y_2) \end{bmatrix} = \begin{bmatrix} \hat{\sigma}_1^{\,2} & \hat{\sigma}_{12} \\ \hat{\sigma}_{12} & \hat{\sigma}_2^{\,2} \end{bmatrix} .$$

Following the methods of Buckley and James, a reformulation of the variance was developed for the proposed model by taking the censoring into consideration through the use of Δ and Δ_i where $\left\{ \begin{array}{ll} \Delta = \text{diag}(\delta_{j1} \cdot \delta_{j2}) & i = 1,2 \\ \Delta_i = \text{diag}(\delta_{ji}) & j = 1,\ldots,n \end{array} \right\}$. Thus Δ (and Δ_i) are diagonal matrices with the j^{th} diagonal element being 1 only if the failure times of both components (i^{th} component) of the j^{th} pair are (is) observed. Thus

$$\mathbf{V}(\hat{\beta}^*) = \begin{bmatrix} \hat{\sigma}_1{}^2 \cdot (\mathbf{X}^T \Delta_1 \mathbf{X})^{-1} & \hat{\sigma}_{12} \cdot (\mathbf{X}^T \Delta \mathbf{X})^{-1} \\ \hat{\sigma}_{12} \cdot (\mathbf{X}^T \Delta \mathbf{X})^{-1} & \hat{\sigma}_2{}^2 \cdot (\mathbf{X}^T \Delta_2 \mathbf{X})^{-1} \end{bmatrix} \quad ;$$

(3.4)

and the individual variance components are defined as follows:

$$\hat{\sigma}_1{}^2 = \tfrac{1}{n} \sum_{j=1}^{n} \left\{ \delta_{j1} \cdot e_{j1}{}^2 + (1 - \delta_{j1}) \cdot \sum_k \left(\delta_{j2} \cdot u_{jk}^{(1)} + (1 - \delta_{j2}) \cdot w_{jk} \right) e_{k1}{}^2 \right\}$$

$$= \text{average of weighted sum of } e_{j1}{}^2 \quad ;$$

(3.5a)

$$\hat{\sigma}_2{}^2 = \tfrac{1}{n} \sum_{j=1}^{n} \left\{ \delta_{j2} \cdot e_{j2}{}^2 + (1 - \delta_{j2}) \cdot \sum_k \left(\delta_{j1} \cdot u_{jk}^{(2)} + (1 - \delta_{j1}) \cdot w_{jk} \right) e_{k2}{}^2 \right\}$$

$$= \text{average of weighted sum of } e_{j2}{}^2 \quad ;$$

(3.5b)

and

$$\hat{\sigma}_{12} = \tfrac{1}{n} \sum_{j=1}^{n} \left\{ \begin{array}{c} \delta_{j1}\delta_{j2}\, e_{j1} e_{j2} + (1 - \delta_{j1}) \cdot \delta_{j2} \cdot \left(\sum_k u_{jk}^{(1)}\, e_{k1} \right) \cdot e_{j2} + \\ \delta_{j1} \cdot (1 - \delta_{j2}) \cdot e_{j1} \cdot \left(\sum_k u_{jk}^{(2)}\, e_{k2} \right) + (1 - \delta_{j1})(1 - \delta_{j2}) \sum_k w_{jk} \cdot e_{k1} \cdot e_{k2} \end{array} \right\} .$$

$$= \text{average of weighted sum of } e_{j1} \cdot e_{j2} \quad .$$

(3.5c)

The details of these derivations can be found in Appendix B. The appropriate weights are determined on the basis of the empirical survival distribution as follows:

$$w_{jk} = \begin{cases} \left(\dfrac{w_k(\hat{\beta})}{\hat{S}(e_j;\, \hat{\beta})} \right) & \text{for } e_k > e_j \\ 0\,, & \text{otherwise} \end{cases} \quad ;$$

$$u_{jk}^{(1)} = \begin{cases} \left(\dfrac{w_k(\hat{\beta}).}{\hat{S}(e_{j1}, e_{\bar{j2}}; \hat{\beta}) - \hat{S}(e_j; \hat{\beta})} \right) & \text{for } e_{k1} > e_{j1} \text{ and } e_{k2} = e_{j2} \\ 0, \quad \text{otherwise} & \end{cases} \quad ; \quad \text{and}$$

$$u_{jk}^{(2)} = \begin{cases} \left(\dfrac{w_k(\hat{\beta}).}{\hat{S}(e_{\bar{j1}}, e_{j2}; \hat{\beta}) - \hat{S}(e_j; \hat{\beta})} \right) & \text{for } e_{k2} > e_{j2} \text{ and } e_{k1} = e_{j1} \\ 0, \quad \text{otherwise} & \end{cases} \quad .$$

To reduce bias, a correction factor $= \begin{cases} \dfrac{n_{ui}}{(n_{ui} - k)} \\ \dfrac{n_u}{(n_u - k)} \end{cases}$ should be included with the estimated variances and covariance in (1.2), (1.3), and (1.4), where

$$n_{ui} = \sum_{j=1}^{n} \delta_{ji} = \text{number of uncensored } i^{th} \text{ components} \; ;$$

$$n_u = \sum_{j=1}^{n} \delta_{j1} \cdot \delta_{j1} = \text{number of uncensored pairs} \; ; \quad \text{and}$$

$$k = \text{nrow}(\beta) \; .$$

(For the previous derivations, see Appendix B.) The estimated covariance of either component with itself reduces to the estimated variance of the component; in the absence of censoring (i.e., all failure times observed) the variance and covariance estimates in equations (3.5a) - (3.5c) reduce to the usual maximum likelihood estimates.

CHAPTER 4. EXTENSION OF BUCKLEY-JAMES PROCEDURE TO THE BIVARIATE FRAMEWORK (PARAMETRIC MODEL)

4.1 BIVARIATE DISTRIBUTIONS

In the univariate framework the exponential and Weibull distributions are commonly used in the analysis of survival data. Properties of these distributions are examined in the context of analyzing bivariate survival data. For the covariate analysis of bivariate survival data in the parametric situation, instead of replacing censored components using empirical estimates for density in (3.1) and (3.3), estimates of the conditional expected values for censored components are based on specified underlying distributions. Determining which distributions would be of particular interest to examine in more detail led to further examination of the model used in this research, the accelerated life-time model, i.e.,

$$\mathrm{Y}= \log(\mathrm{Time}) = \mathrm{X}\beta+\epsilon \quad \Rightarrow \quad \mathrm{Time} = e^{X\beta} \cdot e^{\epsilon} \ .$$

If e^{ϵ} has an exponential or Weibull distribution, then the marginal distributions of the failure times are also exponential or Weibull, respectively. For example,

(1) Let $e^{\epsilon_1} \sim \exp(1) \quad \Rightarrow \quad f_{e^{\epsilon}}(s) = e^{-s}$.

Using the transformation $\mathrm{Time}_1 = e^{x\beta_1 + \epsilon_1}$ results in the following distribution :

$$f_{Time_1}(t) = e^{-\frac{t}{e^{x\beta_1}}} \cdot \frac{1}{e^{x\beta_1}} \ . \tag{4.1}$$

Thus $\text{Time}_1 \sim \exp(e^{-x\beta_1})$.

(2) Alternatively , let $e^{\epsilon_1} \sim \text{Weibull}(1, \nu) \Rightarrow f_{e^\epsilon}(s) = \nu \cdot s^{\nu - 1} \cdot e^{-s^\nu}$.

Again using the transformation $\text{Time}_1 = e^{x\beta_1 + \epsilon_1}$ leads to :

$$f_{Time_1}(t) = \nu \cdot \left(\frac{1}{e^{x\beta_1}} \right)^{\nu} \cdot s^{\nu - 1} \cdot e^{\left(- \frac{t}{e^{x\beta_1}} \right)^{\nu}} . \tag{4.2}$$

Thus, $\text{Time}_1 \sim \text{Weibull}((e^{-x\beta_1})^\nu, \nu)$.

Clearly, the exponential distribution in (4.1) is just a special case of the Weibull distribution in (4.2) where ν=1. This property is an influential factor in using bivariate distributions with either exponential or Weibull marginals.

There are infinitely many bivariate exponential distributions; unfortunately none exists that simultaneously has exponential marginals and exhibits the lack of memory property of the univariate exponential distribution. For most practical situations the former characteristic is more desirable than the latter; in fact, lack of memory may not be a reasonable assumption about the data under study. Alternatively, if the marginals are exponential then the properties of the distribution are well-known and relatively easy to work with. Therefore the set of bivariate exponential distributions to be considered include only those with exponential marginals. The following two bivariate exponential distributions have exponential marginals and are relatively easy to manipulate (Johnson and Kotz, 1972) .

(i) Gumbel's exponential (I) :

$$f(s,t) = \{ (1 + \theta s)\cdot(1 + \theta t) - \theta \} \cdot e^{-s-t-\theta st} \quad ;$$

(ii) Gumbel's exponential (II) :

$$f(s,t) = \{ 1 + \theta(2e^{-s} - 1)(2e^{-t} - 1) \} \cdot e^{-s-t} \quad ;$$

$$(4.3)$$

The Weibull distributions are obtained from the corresponding exponential distributions by the transformation

$$(w^{\nu}_{Weibull}, z^{\phi}_{Weibull}) = (s_{exponential}, t_{exponential}).$$

(iii) Weibull based on Gumbel's exponential (I) :

$$f(w,z) = \{ (1 + \theta w^{\nu})\cdot(1 + \theta z^{\phi}) - \theta \} \cdot e^{-w^{\nu}-z^{\phi}-\theta w^{\nu}z^{\phi}} \cdot \nu\phi w^{\nu-1} z^{\phi-1} \quad ; \text{ and}$$

(iv) Weibull based on Gumbel's exponential (II) :

$$f(w,z) = \{ 1 + \theta(2e^{-w^{\nu}} - 1)(2e^{-z^{\phi}} - 1) \} \cdot e^{-w^{\nu}-z^{\phi}} \cdot \nu\phi w^{\nu-1} z^{\phi-1} \quad .$$

$$(4.4)$$

The last two distributions will subsequently be referred to as Weibull (I) and Weibull (II).

Since these are the distributions for $(s, t) = (e^{\epsilon_1}, e^{\epsilon_2})$, reparameterizations must be used to obtain the conditional expected moments of ϵ. The Weibull distributions are based on the exponential distributions using $(s^{\nu}_{Weibull}, t^{\phi}_{Weibull}) = (s_{exponential}, t_{exponential})$. The conditional expected values of ϵ and ϵ^2 are easily obtained from the appropriate exponential distributions by the following method.

$$E[\epsilon_1 \mid \epsilon_1 > r_1,\ \epsilon_2 > r_2 ;\ e^{\epsilon} \sim \text{Weibull}] = \frac{\int_{u=r_1}^{\infty} \int_{v=r_2}^{\infty} u \cdot f_{\epsilon}(u, v)\ du\, dv}{S_{\epsilon}(r_1, r_2)}$$

Transforming the variables of integration (e^u, e^v) to (s, t), since the distribution of e^{ϵ} is known, leads to the following formulation of the conditional expectation.

$$E[\epsilon_1 \mid \epsilon_1 > r_1,\ \epsilon_2 > r_2 ;\ e^{\epsilon} \sim \text{Weibull}] = \frac{\int_{s=e^{r_1}}^{\infty} \int_{t=e^{r_2}}^{\infty} \ln(s) \cdot f_{e^{\epsilon}}(s, t)\ ds\, dt}{S_{e^{\epsilon}}(e^{r_1}, e^{r_2})},$$

where the density and survival function are Weibull.

Letting $w = s^{\nu}$ and $z = t^{\phi}$, transforms the Weibull density into the exponential, leading to:

$$E\left[\epsilon_1 \,\middle|\, \begin{matrix}\epsilon_1 > r_1 \\ \epsilon_2 > r_2\end{matrix} ;\ e^{\epsilon} \sim \text{Weibull}\right] = \frac{\int_{w=e^{\nu r_1}}^{\infty} \int_{z=e^{\phi r_2}}^{\infty} \frac{1}{\nu} \cdot \ln(w) \cdot f_{e^{\epsilon}}(w, z)\ dw}{S_{e^{\epsilon}}(e^{\nu r_1}, e^{\phi r_2})}$$

though the density and survival function, on the right-hand side of the equation, are now exponential, rather than Weibull.

Thus $E\left[\epsilon_1 \,\middle|\, \begin{matrix}\epsilon_1 > r_1 \\ \epsilon_2 > r_2\end{matrix} ;\ e^{\epsilon} \sim \text{Weibull}\right] = \frac{1}{\nu} \cdot E\left[\epsilon_1 \,\middle|\, \begin{matrix}\epsilon_1 > \nu \cdot r_1, \\ \epsilon_2 > \phi \cdot r_2\end{matrix} ;\ e^{\epsilon} \sim \text{exponential}\right]$

In an analogous fashion, it can be shown that

$$E\left[\epsilon_2 \,\middle|\, \begin{matrix}\epsilon_1 > r_1 \\ \epsilon_2 > r_2\end{matrix} ;\ e^{\epsilon} \sim \text{Weibull}\right] = \frac{1}{\phi} \cdot E\left[\epsilon_2 \,\middle|\, \begin{matrix}\epsilon_1 > \nu \cdot r_1, \\ \epsilon_2 > \phi \cdot r_2\end{matrix} ;\ e^{\epsilon} \sim \text{exponential}\right]$$

The conditional expectation of either component when the other component is observed can be written similarly. Thus

$$E[\epsilon_1 \mid \begin{smallmatrix}\epsilon_1 > r_1\\ \epsilon_2 = r_2\end{smallmatrix}; e^{\epsilon} \sim \text{Weibull}] = \frac{1}{\nu}\cdot E[\epsilon_1 \mid \begin{smallmatrix}\epsilon_1 > \nu\cdot r_1,\\ \epsilon_2 > \phi\cdot r_2\end{smallmatrix}; e^{\epsilon} \sim \text{exponential}] \ ; \quad \text{and}$$

$$E[\epsilon_2 \mid \begin{smallmatrix}\epsilon_1 > r_1\\ \epsilon_2 = r_2\end{smallmatrix}; e^{\epsilon} \sim \text{Weibull}] = \frac{1}{\phi}\cdot E[\epsilon_2 \mid \begin{smallmatrix}\epsilon_1 > \nu\cdot r_1,\\ \epsilon_2 = \phi\cdot r_2\end{smallmatrix}; e^{\epsilon} \sim \text{exponential}] \ .$$

To demonstrate the relationship between the expected value of the product of the residuals for the exponential and Weibull distributions, the same type of method is used.

$$E[\epsilon_1\cdot\epsilon_2 \mid \begin{smallmatrix}\epsilon_1 > r_1\\ \epsilon_2 > r_2\end{smallmatrix}; e^{\epsilon} \sim \text{Weibull}] = \frac{\int_{u=r_1}^{\infty}\int_{v=r_2}^{\infty} uv\cdot f_{\epsilon}(u,v)\ du\ dv}{S_{\epsilon}(r_1, r_2)} \ .$$

Changing the variables of integration from (u, v) to (s, t) = (e^u, e^v) , so that the Weibull density and survival function can be used, leads to the following reformulation of the conditional expectation of the residual product :

$$E[\epsilon_1\cdot\epsilon_2 \mid \begin{smallmatrix}\epsilon_1 > r_1\\ \epsilon_2 > r_2\end{smallmatrix}; e^{\epsilon} \sim \text{Weibull}] = \frac{\int_{s=e^{r_1}}^{\infty}\int_{t=e^{r_2}}^{\infty} \ln(s)\cdot\ln(t)\cdot f_{e^{\epsilon}}(s, t)\ ds\ dt}{S_{e^{\epsilon}}(e^{r_1}, e^{r_2})}$$

Letting $w = s^{\nu}$ and $z = t^{\phi}$, to transform the Weibull distribution into the exponential, leads to:

$$E[\epsilon_1 \mid \begin{smallmatrix}\epsilon_1 > r_1\\ \epsilon_2 > r_2\end{smallmatrix}; e^{\epsilon} \sim \text{Weibull}] = \frac{\int_{w=e^{\nu r_1}}^{\infty}\int_{z=e^{\phi r_2}}^{\infty} \frac{1}{\nu\phi}\cdot\ln(w)\cdot\ln(z)\cdot f_{e^{\epsilon}}(w, z)\ dw}{S_{e^{\epsilon}}(e^{\nu r_1}, e^{\phi r_2})}$$

where the density and survival function are both exponential.

Thus

$$E[\epsilon_1 \mid \begin{smallmatrix}\epsilon_1 > r_1\\ \epsilon_2 > r_2\end{smallmatrix}; e^{\epsilon} \sim \text{Weibull}] = \frac{1}{\nu\phi}\cdot E[\epsilon_1\cdot\epsilon_2 \mid \begin{smallmatrix}\epsilon_1 > \nu\cdot r_1\\ \epsilon_2 > \phi\cdot r_2\end{smallmatrix}; e^{\epsilon} \sim \text{exponential}]$$

Since all of the necessary moments for the Weibull distributions can be found in terms of the corresponding moments of the exponential distributions, it is only necessary to reparameterize the two exponential distributions to obtain all of the required moments for both the Weibull and exponential distributions. The densities of the residuals, obtained by transforming the densities of the exponential of the residuals found in (4.3), can be written as follows :

Gumbel's exponential (I) :

$$f(\epsilon_1,\epsilon_2) = \{ (1+\theta e^{\epsilon_1})\cdot(1+\theta e^{\epsilon_2}) - \theta \}\cdot e^{\epsilon_1+\epsilon_2-e^{\epsilon_1}-e^{\epsilon_2}-\theta e^{\epsilon_1+\epsilon_2}} \quad ;$$

Gumbel's exponential (II) :

$$f(\epsilon_1,\epsilon_2) = \{ 1+\theta(2e^{-e^{\epsilon_1}}-1)(2e^{-e^{\epsilon_2}}-1) \} \cdot e^{\epsilon_1+\epsilon_2-e^{\epsilon_1}-e^{\epsilon_2}} \quad ;$$

4.2 BIVARIATE PARAMETRIC EXPECTATIONS

As with the semi-parametric situation, the censored components are to be replaced by their expected values, conditional on the time of censoring and the covariate parameters.

$$\begin{aligned} y^*_{ij} &= \delta_{ij}\cdot y^O_{ij} + (1-\delta_{ij})\cdot\delta_{ij'}\cdot \mathrm{E}[Y^O_{ij} \mid Y^O_{ij} > y_{ij}\,,\ Y^O_{ij'} = y^O_{ij'}\,;\, \mathbf{X}\hat{\beta}] \\ &+ (1-\delta_{ij})\cdot(1-\delta_{ij'})\cdot \mathrm{E}[Y^O_{ij} \mid Y^O_{ij} > y_{ij}\,,\ Y^O_{ij'} > y_{ij'}\,;\, \mathbf{X}\hat{\beta}] \\ &\qquad i = 1,\ldots, n \ ;\ j = 1, 2 \ ;\quad j \neq j' \end{aligned}$$

But now these conditional expected values can be determined directly from the underlying distribution. To determine all of the expected values for the censored observations, the following quantities need to be calculated based on each distribution for censoring of the first component, with analogous formulas needed for censoring of the second component.

(i) When only the first component is censored, the conditional expected value of Y^O_{i1} is

$$\mathrm{E}[\, Y^O_{i1} \mid Y^O_{i1} > y_{i1},\, Y^O_{i2} = y_{i2}] = X_i\beta_1 + \frac{\displaystyle\int_{s = y_{i1} - X_i\beta_1}^{\infty} s \cdot f(s\,,\ y_{i2} - X_i\beta_2\,)\, ds}{\displaystyle\int_{s = y_{i1} - X_i\beta_1}^{\infty} f(s\,,\ y_{i2} - X_i\beta_2\,)\, ds} \quad ; \qquad \text{and} \tag{4.5}$$

(ii) When both components are censored, the conditional expected value of Y^O_{i1} is

$$\mathrm{E}[\, Y^O_{i1} \mid Y^O_{i1} > y_{i1},\, Y^O_{i2} > y_{i2}] = X_i\beta_1 + \frac{\displaystyle\int_{s = y_{i1} - X_i\beta_1}^{\infty} \int_{t = y_{i2} - X_i\beta_2}^{\infty} s \cdot f(s, t)\, ds\, dt}{\displaystyle\int_{s = y_{i1} - X_i\beta_1}^{\infty} \int_{t = y_{i2} - X_i\beta_2}^{\infty} f(s, t)\, ds\, dt} \ . \tag{4.6}$$

Letting $r_{i1} = y_{i1} - X_i\beta_1$, and $r_{i2} = y_{i2} - X_i\beta_2$, and using the four distributions presented above, (4.5) becomes as follows:

$$\mathrm{E}\left(Y^O_{i1} \middle| \begin{matrix} Y^O_{i1} > y_{i1},\, Y^O_{i2} = y_{i2}; \\ e^{\epsilon_i} \sim \text{Gumbel's exp(I)} \end{matrix} \right) =$$

$$y_{i1} - \frac{Ei(-e^{r_{i1}} \cdot \{1+\theta e^{r_{i2}}\})}{(1 + \theta e^{r_{i1}}) \cdot e^{-e^{r_{i1}} \cdot (1 + \theta e^{r_{i2}})}} + \frac{\theta}{(1+\theta e^{r_{i1}})(1+\theta e^{r_{i2}})} \quad , \tag{4.7a}$$

where $-Ei(\text{-a}) = \int_{x = a}^{\infty} \frac{e^{-x}}{x}\, dx = -\gamma - \log(a) - \left\{ \sum_{i=1}^{\infty} \frac{(-a)^i}{i \cdot i!} \right\}$

is the exponential integral ; γ is Euler's constant .

Also, $\mathrm{E}\left(Y^O_{i1} \middle| \begin{matrix} Y^O_{i1} > y_{i1},\, Y^O_{i2} = y_{i2}; \\ e^{\epsilon_i} \sim \text{Gumbel's exp(II)} \end{matrix} \right) =$

$$y_{i1} + \frac{[1 - \theta(\, 2e^{-e^{r_{i2}}} - 1)] \cdot -Ei(-e^{r_{i1}}) + \theta(2e^{-e^{r_{i2}}} - 1) \cdot -Ei(-2e^{r_{i1}})}{[1 + \theta(e^{-e^{r_{i1}}} - 1)(2e^{-e^{r_{i2}}} - 1)] \cdot e^{-e^{r_{i1}}}} \quad ; \tag{4.7b}$$

$$\mathbf{E}\left(Y^O_{i1}\,\middle|\, \begin{array}{c} Y^O_{i1} > y_{i1},\, Y^O_{i2} = y_{i2}; \\ e^{\epsilon_i} \sim \text{Weibull (I)} \end{array}\right) =$$

$$y_{i1} + \frac{1}{\nu}\cdot\left(\frac{-Ei(-e^{\nu r_{i1}}\cdot\{1+\theta e^{\phi r_{i2}}\})}{(1+\theta e^{\nu r_{i1}})\cdot e^{-e^{\nu r_{i1}}\cdot(1+\theta e^{\phi r_{i2}})}} + \frac{\theta}{(1+\theta e^{\nu r_{i1}})(1+\theta e^{\phi r_{i2}})} \right) \;;$$

(4.7c)

and
$$\mathbf{E}\left(Y^O_{i1}\,\middle|\, \begin{array}{c} Y^O_{i1} > y_{i1},\, Y^O_{i2} = y_{i2}; \\ e^{\epsilon_i} \sim \text{Weibull (II)} \end{array}\right) =$$

$$y_{i1} + \frac{1}{\nu}\cdot\left(\frac{[1 - \theta(\,2e^{-e^{\phi r_{i2}}} - 1)]\cdot -Ei(-e^{\nu r_{i1}}) + \theta(2e^{-e^{\phi r_{i2}}} - 1)\cdot -Ei(-2e^{\nu r_{i1}})}{[1 + \theta(e^{-e^{\nu r_{i1}}} - 1)(2e^{-e^{\phi r_{i2}}} - 1)]\cdot e^{-e^{\nu r_{i1}}}} \right) \;; \tag{4.7d}$$

Using the transformations $r_{i1} = y_{i1} - \mathbf{X}_i\beta_1$ and $r_{i2} = y_{i2} - \mathbf{X}_i\beta_2$ allows calculation of the expectations of (4.2) , corresponding to each of the four distributions, as follows:

$$\mathbf{E}\left(Y^O_{i1}\,\middle|\, \begin{array}{c} Y^O_{i1} > y_{i1},\, Y^O_{i2} > y_{i2}; \\ e^{\epsilon_i} \sim \text{Gumbel's exp(I)} \end{array}\right) = y_{i1} - \frac{Ei(-e^{r_{i1}}\cdot\{1+\theta e^{r_{i2}}\})}{e^{-e^{r_{i1}}\cdot(1+\theta e^{r_{i2}})}} \;; \tag{4.8a}$$

$$\mathbf{E}\left(Y^O_{i1}\,\middle|\, \begin{array}{c} Y^O_{i1} > y_{i1},\, Y^O_{i2} > y_{i2}; \\ e^{\epsilon_i} \sim \text{Gumbel's exp(II)} \end{array}\right) =$$

$$y_{i1} + \frac{[1 - \theta(\,e^{-e^{r_{i2}}} - 1)]\cdot -Ei(-e^{r_{i1}}) + \theta(e^{-e^{r_{i2}}} - 1)\cdot -Ei(-2e^{r_{i1}})}{[1 + \theta(e^{-e^{r_{i1}}} - 1)(e^{-e^{r_{i2}}} - 1)]\cdot e^{-e^{r_{i1}}}} \;; \tag{4.8b}$$

$$\mathbf{E}\left(Y^O_{i1}\,\middle|\, \begin{array}{c} Y^O_{i1} > y_{i1},\, Y^O_{i2} > y_{i2}; \\ e^{\epsilon_i} \sim \text{Weibull (I)} \end{array}\right) = y_{i1} + \frac{1}{\nu}\cdot\left(\frac{-Ei(-e^{\nu r_{i1}}\cdot\{1+\theta e^{\phi r_{i2}}\})}{e^{-e^{\nu r_{i1}}\cdot(1+\theta e^{\phi r_{i2}})}} \right); \tag{4.8c}$$

$$\mathrm{E}\left(\mathrm{Y}_{i1}^{O}\,\middle|\, \begin{matrix}\mathrm{Y}_{i1}^{O} > y_{i1}, \mathrm{Y}_{i2}^{O} > y_{i2} \\ e^{\epsilon_i} \sim \text{Weibull (II)}\end{matrix}\right) =$$

$$y_{i1} + \frac{1}{\nu}\cdot\left(\frac{[1 - \theta(e^{-e^{\nu r_{i1}}} - 1)]\cdot -Ei(-e^{\nu r_{i1}}) + \theta(e^{-e^{\phi r_{i2}}} - 1)\cdot -Ei(-2e^{\nu r_{i1}})}{[1 + \theta(e^{-e^{\nu r_{i1}}} - 1)(e^{-e^{\phi r_{i2}}} - 1)]\cdot e^{-e^{\nu r_{i1}}}}\right) ; \tag{4.8d}$$

Similar calculations are performed for the second outcome variable under both censoring scenarios.

4.3 BIVARIATE PARAMETRIC VARIANCE CALCULATIONS

To calculate the variance estimates in the parametric situation, in addition to the expected value of the residuals under each censoring scheme, the expected value of squared residuals must also be determined.

$$\hat{\mathrm{V}}(\epsilon_{j1}) = \frac{1}{n}\sum_{j=1}^{n}\left\{\begin{matrix}\delta_{j1}\cdot e_{j1}^{2} + (1-\delta_{j1})\cdot(1-\delta_{j2})\cdot\hat{\mathrm{E}}[\epsilon_{j1}^{2} \mid \delta_{j1}=0, \delta_{j2}=0] \\ + (1-\delta_{j1})\cdot\delta_{j2}\cdot\hat{\mathrm{E}}[\epsilon_{j1}^{2} \mid \delta_{j1}=0, \delta_{j2}=1]\end{matrix}\right\}$$

$$- \left(\frac{1}{n}\cdot\sum_{j=1}^{n}\left\{\begin{matrix}\delta_{j1}\cdot e_{j1} + (1-\delta_{j1})\cdot(1-\delta_{j2})\cdot\hat{\mathrm{E}}[\epsilon_{j1} \mid \delta_{j1}=0, \delta_{j2}=0] \\ + (1-\delta_{j1})\cdot\delta_{j2}\cdot\hat{\mathrm{E}}[\epsilon_{j1} \mid \delta_{j1}=0, \delta_{j2}=1]\end{matrix}\right\}\right)^{2} \tag{4.10}$$

The following quantities need to be calculated based on each distribution for censoring of the first component, with analogous formulas needed for censoring of the second component :

$$\mathrm{E}[(\epsilon_{i1})^2 \mid \mathrm{Y}_{i1} > c_{i1}, \mathrm{Y}_{i2} = y_{i2}] = \frac{\int_{s = y_{i1} - X_i\beta_1}^{\infty} s^2 \cdot f(s, y_{i2} - \mathbf{X}_i\boldsymbol{\beta}_2)\, ds}{\int_{s = y_{i1} - X_i\beta_1}^{\infty} f(s, y_{i2} - \mathbf{X}_i\boldsymbol{\beta}_2)\, ds} ; \quad \text{and} \tag{4.11}$$

$$\mathbf{E}[(\epsilon_{i1})^2 \mid Y_{i1} > c_{i1}, Y_{i2} > c_{i2}] = \frac{\int_{s=y_{i1}-X_i\beta_1}^{\infty} \int_{t=y_{i2}-X_i\beta_2}^{\infty} s^2 \cdot f(s,t)\, ds\, dt}{\int_{s=y_{i1}-X_i\beta_1}^{\infty} \int_{t=y_{i2}-X_i\beta_2}^{\infty} f(s,t)\, ds\, dt} \,. \tag{4.12}$$

Letting $r_1 = y_{i1} - X_i\beta_1$ and $r_2 = y_{i2} - X_i\beta_2$, (4.11) becomes:

$$\mathbf{E}\left((\epsilon_{i1})^2 \middle| \begin{array}{c} Y_{i1} > c_{i1}, Y_{i2} = y_{i2} \\ e^{\epsilon_i} \sim \exp(\mathrm{I}) \end{array} \right)$$

$$= r_{i1}^{\,2} + \frac{-r_{i1}^{\,2} + [\frac{\pi^2}{6} + \{\gamma + \ln(1+\theta e^{r_{i2}})\}^2] - 2r_{i1}\left(-Ei(-e^{r_{i1}}\cdot(1+\theta e^{r_{i2}})) + \gamma + \ln(1+\theta e^{r_{i2}})\right)}{(1+\theta e^{r_{i1}})\cdot e^{-e^{r_{i1}}\cdot(1+\theta e^{r_{i2}})}}$$

$$+ \frac{\left(\sum_{i=0}^{\infty} \frac{2\cdot(-e^{r_{i1}}\cdot(1+\theta e^{r_{i2}}))^{i+1}}{(i+1)!\cdot(i+1)^2}\right) + \frac{2\theta}{(1+\theta e^{r_{i2}})}\left(r_{i1}e^{-e^{r_{i1}}(1+\theta e^{r_{i2}})} - Ei(-(1+\theta e^{r_{i2}})e^{r_{i1}})\right)}{(1+\theta e^{r_{i1}})\cdot e^{-e^{r_{i1}}\cdot(1+\theta e^{r_{i2}})}} \;; \tag{4.13a}$$

$$\mathbf{E}\left((\epsilon_{i1})^2 \middle| \begin{array}{c} Y_{i1} > c_{i1}, Y_{i2} = y_{i2} \\ e^{\epsilon_i} \sim \exp(\mathrm{II}) \end{array} \right)$$

$$= r_{i1}^{\,2} + \frac{-r_{i1}^{\,2} + [\frac{\pi^2}{6} + \gamma^2] + \theta(2e^{-e^{r_{i2}}}-1)\cdot\ln(2)\cdot\{2\gamma + \ln(2)\}}{\left\{1 + \theta(e^{-e^{r_{i1}}}-1)\cdot(2e^{-e^{r_{i2}}}-1)\right\}\cdot e^{-e^{r_{i1}}}}$$

$$+ \frac{\sum_{j=0}^{\infty} \frac{(-e^{r_{i1}})^{(j+1)}}{(j+1)!\cdot(j+1)^2}\cdot\left(2\{1-\theta(2e^{-e^{r_{i2}}}-1)\} - 2^{(ji+2)}\theta(2e^{-e^{r_{i2}}}-1)\right)}{\left\{1 + \theta(e^{-e^{r_{i1}}}-1)\cdot(2e^{-e^{r_{i2}}}-1)\right\}\cdot e^{-e^{r_{i1}}}}$$

$$- \frac{\left(\gamma + \{1-\theta(2e^{-e^{r_{i2}}}-1)\}\cdot -Ei(-e^{r_{i1}}) + \theta(2e^{-e^{r_{i2}}}-1)\cdot[-Ei(-2e^{r_{i1}}) + \ln(2)]\right)\cdot 2r_{i1}}{\left\{1 + \theta(e^{-e^{r_{i1}}}-1)\cdot(2e^{-e^{r_{i2}}}-1)\right\}\cdot e^{-e^{r_{i1}}}} \;; \tag{4.13b}$$

$$\mathbf{E}\left((\epsilon_{i1})^2 \,\middle|\, \begin{matrix} Y_{i1} > c_{i1},\, Y_{i2} = y_{i2} \\ e^{\epsilon_i} \sim \text{Weibull(I)} \end{matrix}\right) = \frac{1}{\nu^2}\cdot \mathbf{E}\left((\epsilon_{i1})^2 \,\middle|\, \begin{matrix} Y_{i1} > \nu c_{i1},\, Y_{i2} = \phi y_{i2} \\ e^{\epsilon_i} \sim \exp(\text{I}) \end{matrix}\right) \quad ; \tag{4.13c}$$

and $$\mathbf{E}\left((\epsilon_{i1})^2 \,\middle|\, \begin{matrix} Y_{i1} > c_{i1},\, Y_{i2} = y_{i2} \\ e^{\epsilon_i} \sim \text{Weibull(II)} \end{matrix}\right) = \frac{1}{\nu^2}\cdot \mathbf{E}\left((\epsilon_{i1})^2 \,\middle|\, \begin{matrix} Y_{i1} > \nu c_{i1},\, Y_{i2} = \phi y_{i2} \\ e^{\epsilon_i} \sim \exp(\text{II}) \end{matrix}\right) \quad . \tag{4.13d}$$

The expectations of (4.12) , corresponding to each of the four distributions, are determined as follows:

$$\mathbf{E}\left((\epsilon_{i1})^2 \,\middle|\, \begin{matrix} Y_{i1} > c_{i1},\, Y_{i2} > c_{i2} \\ e^{\epsilon_i} \sim \exp(\text{I}) \end{matrix}\right) = r_{i1}^2 + \frac{\left\{-r_{i1}^2 + [\frac{\pi^2}{6} + \{\gamma + \ln(1+\theta e^{r_{i2}})\}^2]\right\}}{e^{-e^{r_{i1}} - \theta e^{r_{i1}+r_{i2}}}}$$
$$+ \frac{\left\{-[-Ei(-\{1+\theta e^{r_{i2}}\}\cdot e^{r_{i1}}) + \gamma + \ln(1+\theta e^{r_{i2}})]\cdot 2r_{i1}\right\} + \left\{\sum_{j=0}^{\infty} \frac{2\cdot(-(1+\theta e^{r_{i2}})e^{r_{i1}})^{j+1}}{(j+1)!\cdot(j+1)^2}\right\}}{e^{-e^{r_{i1}} - \theta e^{r_{i1}+r_{i2}}}} \quad ; \tag{4.14a}$$

$$\mathbf{E}\left((\epsilon_{i1})^2 \,\middle|\, \begin{matrix} Y_{i1} > c_{i1},\, Y_{i2} > c_{i2} \\ (e^{\epsilon_{i1}}, e^{\epsilon_{i2}}) \sim \exp(\text{II}) \end{matrix}\right)$$
$$= r_{i1}^2 + \frac{-r_{i1}^2 + [\frac{\pi^2}{6} + \gamma^2] + \theta(e^{-e^{r_{i2}}}-1)\cdot\ln(2)\cdot\{2\gamma + \ln(2)\}}{\{1 + \theta(e^{-e^{r_{i2}}}-1)\cdot(e^{-e^{r_{i1}}}-1)\}\cdot e^{-e^{r_{i1}}}}$$
$$+ \frac{\sum_{j=0}^{\infty} \frac{(-e^{r_{i1}})^{j+1}}{(j+1)!\cdot(j+1)^2}\cdot\left(2\cdot\{1-\theta(e^{-e^{r_{i2}}}-1)\} - 2^{(i+2)}\theta(e^{-e^{r_{i2}}}-1)\}\right)}{\{1 + \theta(e^{-e^{r_{i2}}}-1)\cdot(e^{-e^{r_{i1}}}-1)\}\cdot e^{-e^{r_{i1}}}}$$
$$- \frac{\left(\gamma + \{1-\theta(e^{-e^{r_{i2}}}-1)\}\cdot -Ei(-e^{r_{i1}}) + \theta(e^{-e^{r_{i2}}}-1)\cdot[-Ei(-2e^{r_{i1}}) + \ln(2)]\right)\cdot 2r_{i1}}{\{1 + \theta(e^{-e^{r_{i1}}}-1)\cdot(e^{-e^{r_{i2}}}-1)\}\cdot e^{-e^{r_{i1}}}} \quad ; \tag{4.14b}$$

$$\mathbb{E}\left((\epsilon_{i1})^2 \,\middle|\, \begin{matrix} Y_{i1} > c_{i1},\, Y_{i2} = y_{i2} \\ e^{\epsilon_i} \sim \text{Weibull(I)} \end{matrix} \right) = \frac{1}{\nu^2} \cdot \mathbb{E}\left((\epsilon_{i1})^2 \,\middle|\, \begin{matrix} Y_{i1} > \nu c_{i1},\, Y_{i2} = \phi y_{i2} \\ e^{\epsilon_i} \sim \exp(\text{I}) \end{matrix} \right) \quad ; \tag{4.14c}$$

and

$$\mathbb{E}\left((\epsilon_{i1})^2 \,\middle|\, \begin{matrix} Y_{i1} > c_{i1},\, Y_{i2} = y_{i2} \\ e^{\epsilon_i} \sim \text{Weibull(II)} \end{matrix} \right) = \frac{1}{\nu^2} \cdot \mathbb{E}\left((\epsilon_{i1})^2 \,\middle|\, \begin{matrix} Y_{i1} > \nu c_{i1},\, Y_{i2} = \phi y_{i2} \\ e^{\epsilon_i} \sim \exp(\text{II}) \end{matrix} \right) \quad . \tag{4.14d}$$

The results for the conditional expected value of $(\epsilon_{i2})^2$ are analogous.

4.4 BIVARIATE PARAMETRIC COVARIANCE CALCULATIONS

To calculate the covariance estimates in the parametric situation, in addition to the expected value of the residuals under each censoring scheme (as found in section 4.3), the expected value of the product of residuals must also be determined for each distribution.

$$\hat{\text{Cov}}(\epsilon_{j1}, \epsilon_{j2})$$

$$= \frac{1}{n} \cdot \sum_{j=1}^{n} \left\{ \begin{matrix} \delta_{j1} \cdot \delta_{j2} \cdot e_{j1} \cdot e_{j2} \quad + \\ (1\text{-}\delta_{j1}) \cdot \delta_{j2} \cdot \hat{\mathbb{E}}[\epsilon_{j1} \cdot \epsilon_{j2} \mid \delta_{j1} = 0,\, \delta_{j2} = 1] + \\ \delta_{j1} \cdot (1\text{-}\delta_{j2}) \cdot \hat{\mathbb{E}}[\epsilon_{j1} \cdot \epsilon_{j2} \mid \delta_{j1} = 1,\, \delta_{j2} = 0] + \\ (1\text{-}\delta_{j1}) \cdot (1\text{-}\delta_{j2}) \cdot \hat{\mathbb{E}}[\epsilon_{j1} \cdot \epsilon_{j2} \mid \delta_{j1} = 0,\, \delta_{j2} = 0] \end{matrix} \right\}$$

$$- \left(\frac{1}{n} \cdot \sum_{j=1}^{n} \left\{ \begin{matrix} \delta_{j1} \cdot e_{j1} + (1\text{-}\delta_{j1}) \cdot (1\text{-}\delta_{j2}) \cdot \hat{\mathbb{E}}[\epsilon_{j1} \mid \delta_{j1}=0,\, \delta_{j2}=0] \\ + (1\text{-}\delta_{j1}) \cdot \delta_{j2} \cdot \hat{\mathbb{E}}[\epsilon_{j1} \mid \delta_{j1}=0,\, \delta_{j2}=1] \end{matrix} \right\} \right) \cdot$$

$$\left(\frac{1}{n} \cdot \sum_{j=1}^{n} \left\{ \begin{matrix} \delta_{j2} \cdot e_{j2} + (1\text{-}\delta_{j1}) \cdot (1\text{-}\delta_{j2}) \cdot \hat{\mathbb{E}}[\epsilon_{j2} \mid \delta_{j1}=0,\, \delta_{j2}=0] \\ + \delta_{j1} \cdot (1\text{-}\delta_{j2}) \cdot \hat{\mathbb{E}}[\epsilon_{j2} \mid \delta_{j1}=1,\, \delta_{j2}=0] \end{matrix} \right\} \right) \tag{4.15}$$

The following quantity needs to be calculated based on each distribution :

$$\mathrm{E}[\,(\epsilon_{i1}\cdot\epsilon_{i1})\mid Y_{i1}>c_{i1},\,Y_{i2}>c_{i2}]=\frac{\displaystyle\int_{s=y_{i1}-X_i\beta_1}^{\infty}\int_{t=y_{i2}-X_i\beta_2}^{\infty} s\cdot t\cdot f(s,\,t)\;ds\,dt}{\displaystyle\int_{s=y_{i1}-X_i\beta_1}^{\infty}\int_{t=y_{i2}-X_i\beta_2}^{\infty} f(s,\,t)\;ds\,dt}\,. \tag{4.16}$$

Letting $r_1 = y_{i1} - X_i\beta_1$ and $r_2 = y_{i2} - X_i\beta_2$, (4.16) becomes:

$$\mathrm{E}\left(\epsilon_{i1}\cdot\epsilon_{i2}\,\middle|\,\begin{matrix}Y_{i1}>c_{i1},\,Y_{i2}>c_{i2}\\ (e^{\epsilon_{i1}},\,e^{\epsilon_{i2}})\sim\exp(\mathrm{I})\end{matrix}\right)$$

$$= r_{i1}\cdot r_{i2} \;+\; \frac{r_{i1}\cdot e^{-e^{r_{i1}}}\cdot\text{-}Ei(\text{-}\{1+\theta e^{r_{i2}}\}\cdot e^{r_{i1}}) + r_{i2}\cdot e^{-e^{r_{i2}}}\cdot\text{-}Ei(\text{-}\{1+\theta e^{r_{i1}}\}\cdot e^{r_{i2}})}{e^{-e^{r_{i1}}-e^{r_{i2}}-\theta e^{r_{i1}+r_{i2}}}}$$

$$+\;\frac{Ei(\text{-}e^{r_{i1}})\cdot Ei(\text{-}e^{r_{i2}}) - \displaystyle\sum_{k=1}^{\infty}\sum_{j=1}^{k}\frac{(-e^{r_{i2}})^k}{k\cdot(k-j)!j!}\cdot\theta^j\cdot\Gamma(j\,,\,e^{r_{i1}}) + \sum_{k=1}^{\infty}\frac{(-1)^k}{k}\cdot\Gamma(j\,,\,e^{r_{i1}})}{e^{-e^{r_{i1}}-e^{r_{i2}}-\theta e^{r_{i1}+r_{i2}}}}\;; \tag{4.17a}$$

$$\mathrm{E}\left(\epsilon_{i1}\cdot\epsilon_{i2}\,\middle|\,\begin{matrix}Y_{i1}>c_{i1},\,Y_{i2}>c_{i2}\\ (e^{\epsilon_{i1}},\,e^{\epsilon_{i2}})\sim\exp(\mathrm{II})\end{matrix}\right)$$

$$= r_{i1}\cdot r_{i2} + \frac{\theta\cdot\{Ei(\text{-}e^{r_{i1}}) - Ei(\text{-}2e^{r_{i1}})\}\cdot\{Ei(\text{-}e^{r_{i2}}) - Ei(\text{-}2e^{r_{i2}})\} + Ei(\text{-}e^{r_{i1}})\cdot Ei(\text{-}e^{r_{i2}})}{\left(1+\theta(e^{-e^{r_{i1}}}\text{-}1)\cdot(e^{-e^{r_{i2}}}\text{-}1)\right)\cdot e^{-e^{r_{i1}}-e^{r_{i2}}}} +$$

$$\frac{\theta\left(r_{i1}\,e^{-e^{r_{i1}}}(e^{-e^{r_{i1}}}\text{-}1)\{Ei(\text{-}e^{r_{i2}})\text{-}Ei(\text{-}2e^{r_{i2}})\}+r_{i2}\,e^{-e^{r_{i2}}}(e^{-e^{r_{i2}}}\text{-}1)\{Ei(\text{-}e^{r_{i1}})\text{-}Ei(\text{-}2e^{r_{i1}})\}\right)}{\left(1+\theta(e^{-e^{r_{i1}}}\text{-}1)\cdot(e^{-e^{r_{i2}}}\text{-}1)\right)\cdot e^{-e^{r_{i1}}-e^{r_{i2}}}}\;; \tag{4.17b}$$

$$\mathbf{E}\left(\epsilon_{i1} \cdot \epsilon_{i2} \middle| \begin{array}{c} Y_{i1} > c_{i1},\, Y_{i2} > c_{i2} \\ e^{\epsilon_i} \sim \text{Weibull(I)} \end{array}\right) = \frac{1}{\nu\phi} \cdot \mathbf{E}\left(\epsilon_{i1} \cdot \epsilon_{i2} \middle| \begin{array}{c} Y_{i1} > \nu c_{i1},\, Y_{i2} > \phi c_{i2} \\ e^{\epsilon_i} \sim \exp(\text{I}) \end{array}\right) \;; \tag{4.17c}$$

and $$\mathbf{E}\left(\epsilon_{i1} \cdot \epsilon_{i2} \middle| \begin{array}{c} Y_{i1} > c_{i1},\, Y_{i2} > c_{i2} \\ e^{\epsilon_i} \sim \text{Weibull(II)} \end{array}\right) = \frac{1}{\nu\phi} \cdot \mathbf{E}\left(\epsilon_{i1} \cdot \epsilon_{i2} \middle| \begin{array}{c} Y_{i1} > \nu c_{i1},\, Y_{i2} > \phi c_{i2} \\ e^{\epsilon_i} \sim \exp(\text{II}) \end{array}\right) . \tag{4.17d}$$

CHAPTER 5. APPLICATION OF SEMI-PARAMETRIC AND PARAMETRIC METHODS

The models developed in the previous chapter were used to analyze the outcome variables of duration of post-partum amenorrhea and breastfeeding. A random sample of 100 observations was selected from the Demographic and Health Survey of women in Ghana. All of the women were aged 18-49 and only those with parity > 1 were included in the sample. The covariates examined in this analysis are education recorded in years completed and parity (birth order) of the child. Duration of breastfeeding is the first outcome variable and duration of post-partum amenorrhea is the second outcome varible. There was a high proportion of censored observations : 46% of the sample were either singly or doubly censored. The primary cause of censoring was that the activities, post-partum amenorrhea and/or breastfeeding, were ongoing at the time of the study. The censoring indicator, δ, is 1 only if the activity has ceased (i.e., "failure" has occurred) for each outcome variable.

		δ_1	
		0	1
δ_2	0	24	18
	1	4	54

Almost two-thirds of the women in the sample had three or less years of education, with over half of the sample having none; only 16 had nine or more years of education (i.e., some secondary school or more). Since all first births were excluded from consideration, all of the parity values are two or more; 57% of the women had four or fewer children, while only ten had nine or more births. The outcome variables, including both censored outcomes and failures, ranged from 1 to 35 months for the duration of breast-feeding and from 1 to 36 months for post-partum amenorrhea. When the censored outcomes were excluded the range for the duration of breast-feeding reduced to 1 to 25 months; the range for post-partum amenorrhea was not affected by the exclusion of censored outcomes.

The estimates for the semi-parametric and parametric methods have been calculated using SAS IML. The covariate parameter estimates and standard errors for each of the five models can be found in Table 5.1. For comparison purposes, the parameter estimates were found based on marginal analyses under the assumption of independent outcomes (estimated using SAS PROC LIFEREG) and are also included in Table 5.1.

TABLE 5.1 COVARIATE PARAMETER ESTIMATES FOR EACH MODEL

Model		log(months of breastfeeding)	log(months of amenorrhea)	Standard errors	
empirical survival	intercept	2.7181806	1.8579996	0.1519	0.1746
	education(yrs)	-0.005894	0.0340847	0.0148	0.0192
	parity	-0.006762	0.037571	0.0276	0.0333
Gumbel's exponential (I)	intercept	2.6860659	1.9052683	0.2522	0.2612
	education(yrs)	0.0015026	0.0353857	0.0245	0.0269
	parity	0.005204	0.0428116	0.0459	0.0467
Gumbel's exponential (II)	intercept	2.6227863	1.8180563	0.2720	0.3440
	education(yrs)	0.0036892	0.0353857	0.0264	0.0354
	parity	0.0040728	0.038764	0.0495	0.0615
Weibull (I)	intercept	2.9096196	1.9087483	0.2433	0.1009
	education(yrs)	-0.007279	0.0308122	0.0236	0.0104
	parity	0.0099007	0.0391454	0.0443	0.0180
Weibull (II)	intercept	3.035269	1.8442362	0.2145	0.2127
	education(yrs)	-0.016412	0.0334033	0.0208	0.0213
	parity	0.020281	0.0352661	0.0390	0.0369
exponential assuming independence	intercept	3.2100076	2.3897499	0.3320	0.2921
	education(yrs)	-0.0324428	0.0122253	0.0331	0.0313
	parity	0.0208825	0.0483367	0.0601	0.0526
Weibull assuming independence	intercept	3.0234413	2.4126285	0.1069	0.1859
	education(yrs)	-0.0198274	0.0095677	0.0111	0.0206
	parity	0.0101980	0.0425989	0.0217	0.0336

The estimates in Table 5.1 are for the logarithm of time, so the untransformed estimates for months of breast-feeding and months of post-partum amenorrhea at baseline (excluding the estimates found assuming independence of the outcomes) range from 13.8 to 20.8 months and 6.2 to 6.7 months, respectively. It might be more appropriate to examine an adjusted baseline value, as a parity of 2 is the minimum value within the range of estimation for this analysis; using (parity, education) = (2, 0) as the new set of baseline covariates yields adjusted baseline estimates. The range of adjusted estimates for duration of breast-feeding for the five models is rather wide : 13.9 to 21.7 months; the range of adjusted baseline estimates for post-partum amenorrhea is much narrower in comparison: 6.7 to 7.3 months.

For the two Weibull distributions the estimates of ν and ϕ, functions of the marginal expected values, are (0.833, 0.745) for the Weibull (I) and (0.734, 0.748) for the Weibull (II) distribution. If these two values for each Weibull model were identically 1, then the distributions would reduce to the corresponding exponential distribution. The estimates of these parameters are quite different when calculated using univariate models separately on each outcome : $(\nu, \phi) = (0.339, 0.642)$. Since these values indicate that the Weibull distributions are quite distinct from the exponential distributions, a similar difference in parameter estimates might be expected.

Actually, the intercept and covariate estimates for the duration of post-partum amenorrhea are fairly consistent across models, though the intercepts are quite a bit lower than those found using univariate techniques. On the other hand, the parameter estimate for education using any of the bivariate models is almost three times as large as either estimate from the univariate models. The parameter estimates of parity assuming independent outcomes are fairly consistent with those of the bivariate models.

For this data the choice of underlying distribution between exponential or Weibull has little effect on the estimates for amenorrhea; also the selection of either Gumbel's (I) or Gumbel's (II) seems not to be unduly influential. For this outcome the parametric estimates are quite similar to those found with the semi-parametric model, indicating, at least intuitively, that the exponential distribution provides a reasonable fit to the data.

For the first outcome variable, months of breast-feeding, the underlying distribution appears to have a much more significant affect on the parameter estimates. In fact, even the direction of the effect for the concomitant variables varies across methods, though these effects are quite small and perhaps non-significant. The intercept estimates are also not consistent across models; thus for breast-feeding the choice of model is much more important. And unlike the situation with the second outcome, there is relatively poor agreement between the exponential estimates and those found with the Weibull models, though the two exponential models are in fairly close agreement, as are the two Weibull models. The intercept value found with the semi-parametric model lies between the exponential estimates and the Weibull estimates. Rather surprisingly, the intercept estimate found from the univariate exponential model is not only larger than the corresponding Weibull estimate, but it is also larger than any of the intercept estimates from any of the bivariate models. All of the parameter estimates for the univariate Weibull model are similar to those of the two bivariate Weibull models, but this not the case for the exponential models.

The estimated variance-covariance matrices of $\hat{\beta}$ were calculated for each model according to the methods discussed in Chapters 3 and 4. The standard errors and correlations were determined from the estimated variance-covariance matrices and are presented in Table 5.2. For comparison purposes, the standard errors for the models based on the assumption of independent outcomes are also presented in Table 5.2. The variances are available from PROC

LIFEREG, but the covariances for the parameters within each component, while they could be determined, are not provided. Since separate univariate analyses were run, the covariances for between components parameters do not exist. Therefore the correlations between covariate parameters for the same outcome component exist, but are not available; the correlations between covariate parameters for different outcome components do not exist.

Of the bivariate models the semi-parametric model, in general, provides the smallest estimates of the variance. The Weibull model estimates were smaller than their corresponding exponential models. This was not unexpected; since the exponential distribution is a special case of the Weibull distribution, the Weibull model should provide a better fit to the data. The results from the univariate models are rather surprising : the exponential model estimates are a good bit higher than their bivariate model counterparts, but the univariate Weibull model estimates are, for the most part, substantially lower than their bivariate counterparts. The estimated correlations between covariate parameters for the duration of breastfeeding outcome were in extremely close agreement across all of the bivariate models; the correlations for the post-partum amenorrhea outcome, while similar, were not as consistent across models. The correlations for covariate parameters between the outcome variables were even more variable across models. Generally the semi-parametric model provided correlation estimates with the largest magnitude.

TABLE 5.2 ESTIMATES OF STANDARD ERRORS (DIAGONAL ELEMENTS) AND CORRELATIONS (OFF-DIAGONAL ELEMENTS) FOR COVARIATE PARAMETERS

Empirical survival :

$$\begin{bmatrix} 0.152 & -0.487 & -0.853 & 0.419 & -0.182 & -0.337 \\ & 0.0148 & 0.172 & -0.207 & 0.386 & 0.0659 \\ & & 0.0276 & -0.354 & 0.0609 & 0.399 \\ & & & 0.175 & -0.520 & -0.912 \\ & & & & 0.0192 & 0.191 \\ & & & & & 0.0333 \end{bmatrix}$$

Gumbel's exponential (I) :

$$\begin{bmatrix} 0.252 & -0.487 & -0.854 & 0.252 & -0.117 & -0.217 \\ & 0.0245 & 0.172 & -0.124 & 0.249 & 0.0423 \\ & & 0.0459 & -0.213 & 0.0392 & 0.257 \\ & & & 0.261 & -0.486 & -0.851 \\ & & & & 0.0269 & 0.190 \\ & & & & & 0.0467 \end{bmatrix}$$

Gumbel's exponential (II) :

$$\begin{bmatrix} 0.272 & -0.487 & -0.854 & 0.401 & -0.187 & -0.345 \\ & 0.0264 & 0.172 & -0.198 & 0.396 & 0.0674 \\ & & 0.0495 & -0.339 & 0.0625 & 0.408 \\ & & & 0.344 & -0.486 & -0.851 \\ & & & & 0.0354 & 0.190 \\ & & & & & 0.0615 \end{bmatrix}$$

Weibull (I) :

$$\begin{bmatrix} 0.243 & -0.486 & -0.854 & 0.389 & -0.181 & -0.335 \\ & 0.0236 & 0.172 & -0.192 & 0.384 & 0.0654 \\ & & 0.0443 & -0.329 & 0.0606 & 0.396 \\ & & & 0.101 & -0.486 & -0.851 \\ & & & & 0.0104 & 0.190 \\ & & & & & 0.0180 \end{bmatrix}$$

Weibull (II) :

$$\begin{bmatrix} 0.215 & -0.487 & -0.854 & 0.291 & -0.135 & -0.250 \\ & 0.0208 & 0.172 & -0.144 & 0.287 & 0.0488 \\ & & 0.0390 & -0.246 & 0.0452 & 0.296 \\ & & & 0.213 & -0.486 & -0.851 \\ & & & & 0.0213 & 0.190 \\ & & & & & 0.0369 \end{bmatrix}$$

exponential assuming independence :

$$\begin{bmatrix} 0.332 & & & & & \\ & 0.0331 & & & & \\ & & 0.0601 & & & \\ & & & 0.292 & & \\ & & & & 0.0313 & \\ & & & & & 0.0526 \end{bmatrix}$$

Weibull assuming independence :

$$\begin{bmatrix} 0.107 & & & & & \\ & 0.0111 & & & & \\ & & 0.0217 & & & \\ & & & 0.186 & & \\ & & & & 0.0206 & \\ & & & & & 0.0336 \end{bmatrix}$$

For hypothesis testing, a bootstrap sampling procedure was used. There were 500 replications selected; the 100 observations from the original sample were sampled with replacement using a uniform random number generator in a SAS program. The semi-parametric model was used and parameter estimates are calculated as described previously. The mean and the standard errors for the covariate parameter estimates based on the 500 bootstrap samples are :

$$\bar{\hat{\beta}} = \begin{bmatrix} 2.686 & 1.845 \\ -0.00526 & 0.0350 \\ -0.00238 & 0.0405 \end{bmatrix} \qquad \text{s.e.}(\bar{\hat{\beta}}) = \begin{bmatrix} 0.217 & 0.253 \\ 0.0200 & 0.0290 \\ 0.0404 & 0.0336 \end{bmatrix} .$$

There is a large class of test statistics used for testing general multivariate hypotheses of the following type :

$$\mathbf{H}_0 : \theta = \mathbf{C}\beta\mathbf{U} \qquad \mathbf{H}_1 : \theta \neq \mathbf{C}\beta\mathbf{U}$$

where **C** is a contrast matrix which computes linear combinations of the covariate parameters for each outcome variable and **U** is a contrast matrix which combines the covariate parameters across the outcome variables. A commonly used test statistic for multivariate hypotheses is Wilks Lambda, Λ ; it is a likelihood ratio statistic for data having an underlying normal distribution. The Wilks Lambda statistic is known to have a non-central F-distribution. The statistic is defined below.

$$\Lambda = \frac{|\mathbf{E}|}{|\mathbf{H}+\mathbf{E}|} \qquad \text{where}$$

$$H = \theta' [C(X'X)^{-1}C']^{-1}\theta$$

$$E = U'\Sigma U \cdot (n-r)$$

n = number of observations

r = rank(X) .

In the semi-parametric model, no assumptions are made about the underlying distribution, so the test statistic does not necessarily have an F-distribution. But using the bootstrap samples, a distribution of the Wilks lambda statistic can be obtained, and significance levels determined accordingly. The three primary tests of interest are whether or not the covariate parameters are jointly significant across outcomes and whether or not they are individually significant. It is also possible to perform tests for each outcome individually, but it would not be possible to reduce the bivariate model if the covariate for only one outcome is nonsignificant. The three tests described above can be written as:

1) H_0 : No effect of parity or education on post-partum amenorrhea or breastfeeding

H_1 : At least one of the covariates has an effect on the outcomes

(Thus $C = \begin{Bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{Bmatrix}$ and $U = \begin{Bmatrix} 1 & 0 \\ 0 & 1 \end{Bmatrix}$)

2) H_0 : No effect of education on post-partum amenorrhea or breastfeeding

H_1 : Education has an effect on the outcomes

(Thus $C = \{0\ 1\ 0\}$ and $U = \begin{Bmatrix} 1 & 0 \\ 0 & 1 \end{Bmatrix}$)

3) H_0 : No effect of parity on post-partum amenorrhea or breastfeeding

H_1 : Parity has an effect on the outcomes

(Thus $C = \{0\ 0\ 1\}$ and $U = \begin{Bmatrix} 1 & 0 \\ 0 & 1 \end{Bmatrix}$)

The test statistics corresponding to the hypotheses described above, using the estimates of the covariate parameters and the variance-covariance matrix for the outcomes derived in Chapter 3, are $(\Lambda_1, \Lambda_2, \Lambda_3) = (0.945, 0.976, 0.934)$. And the corresponding significance levels determined from the bootstrap distributions are $(0.25 > \alpha_1 > 0.10,\ \alpha_2 > 0.25,\ \alpha_3 > 0.25)$. Clearly the joint test of both covariate parameters was nonsignificant; technically, further hypotheses should not have been tested, but they were examined here merely for demonstration purposes. While other linear combinations of the covariate parameters could also have been examined, within the context of this dataset there were no other a priori tests of interest.

Since the Wilks Lambda statistic is a likelihood ratio test statistic based on an underlying normal distribution, it is not appropriate for the parametric models unless some asymptotic results can be invoked. Development of a statistic for the parametric models is the subject of future research.

CHAPTER 6. DISCUSSION AND SUGGESTIONS FOR FUTURE RESEARCH

The primary goal of this research has been to extend the univariate Buckley-James procedure for survival analysis to the bivariate framework. The bivariate methods provide information about the relationship of the two outcome variables that cannot be obtained by applying univariate methods independently on each outcome. Besides taking all of the data into consideration in estimation of the covariate parameters, the bivariate methods allow for the estimation of covariance components for both the outcomes and the covariate parameters. Thus it is possible to make inferences concerning both outcome variables simultaneously.

The methods for covariate analysis of bivariate survival data presented in the preceding chapters have resolved several of the difficulties found in existing methods. Ties in the data are included in the model through modification of the empirical bivariate survival function; censored observations are incorporated by replacing them with their conditional expected values; and specification of an underlying parametric distribution is possible by using the appropriate conditional expectations. The variance-covariance matrix for the covariate parameters is still empirically based, rather than the result of some analytical development. The implications on the inference process have yet to be studied.

There are many other issues for further study, such as the choice between the semi-parametric and parametric methods for parameter estimation. The primary consideration is previous knowledge. If there is no prior information about the distributional form of the outcomes or other factors indicating a specific underlying parametric distribution, then the semi-parametric method, a more general approach, would provide a more reasonable course of analysis. Conversely, the semi-parametric method is much more computer intensive; with the availability of supercomputers and more powerful PCs this is not necessarily an influential factor.

A goodness-of-fit statistic, for both the semi-parametric and parametric situations, is a prospect for future research. Questions concerning the choice of underlying distribution could be decided on the basis of such a statistic. A related issue to be considered is the effect of misspecifying the underlying distribution for the parametric model. Misspecification of the model is another issue for further examination, which may result in difficulties such as non-convergence of the covariate parameter estimates or patterns in the residuals. By modifying some existing residual diagnostics for multivariate regression to account for censoring, it may be possible to detect some of the difficulties arising from misspecification of the model.

Another area that deserves some investigation is the examination of properties of the covariate parameter estimates ; such properties include consistency, efficiency, unbiasedness, asymptotic properties, and small sample properties. The estimates have been widely studied when the outcome is univariate, and it is likely that many of the same properties will also hold in the bivariate situation. Properties of the variance-covariance estimates for the covariate parameters are also of interest. Another important issue for future research is the development of a statistical test for hypotheses in the parametric model. Depending on the asymptotic properties of the estimates, the use of existing statistics may be justified.

The univariate Buckley-James procedure determines estimates based on the Kaplan-Meier estimate of the survival function, but there are several bivariate Kaplan-Meier type estimates of the survival function. Whether or not the choice of the empirical bivariate survival function estimator has a substantial effect on the estimates of the covariate parameters in the semi-parametric model is unknown. Additionally, it might be interesting to compare these results with covariate parameter estimates calculated based on other types of empirical survival function estimators, such as those using smoothing techniques.

In the univariate setting, the censoring pattern appears to have little effect on the estimation process for the Buckley-James procedure. The same may not be true for the bivariate setting. Another related issue is how the proportion of censored observations affects the variability of the covariate parameter estimates.

The techniques developed in this research for covariate analysis in the bivariate framework can be generalized in several ways : the definition of censoring could be modified to include not just right-censoring, but also left- and interval-censoring; the bivariate methods could be extended to the multivariate framework; and the concept of covariate censoring, as opposed the censoring of outcome times, could also explored.

APPENDIX A. CALCULATION OF CORRELATION FOR OUTCOME VARIABLES IN CLAYTON'S MODEL

When the marginals are exponential, the joint density in Clayton's model becomes

$$f(t_1, t_2) = \frac{\theta\alpha\gamma \cdot e^{\alpha(\theta-1)t_1} \cdot e^{\gamma(\theta-1)t_2}}{\left\{ e^{\alpha(\theta-1)t_1} + e^{\gamma(\theta-1)t_2} - 1 \right\}^{\frac{1}{\theta-1}+2}}$$

$$f(t_1) = \int_{t_1} \frac{\theta\alpha\gamma \cdot e^{\alpha(\theta-1)t_1} \cdot e^{\gamma(\theta-1)t_2} \, dt_2}{\left\{ e^{\alpha(\theta-1)t_1} + e^{\gamma(\theta-1)t_2} - 1 \right\}^{\frac{1}{\theta-1}+2}} = e^{-\alpha t_1}$$

So $t_1 \sim \exp(\alpha)$ and, similarly, $t_2 \sim \exp(\gamma)$.

The marginal moments for the exponential distribution are known, so the only unknown quantity needed to calculate the correlation of the outcomes is the expectation of the product of the time variables.

$$\mathrm{corr}(t_1, t_2) = \frac{\mathbf{E}(t_1 \cdot t_2) - \mathbf{E}(t_1) \cdot \mathbf{E}(t_2)}{\left(\mathbf{V}(t_1) \cdot \mathbf{V}(t_2)\right)^{\frac{1}{2}}}$$

$$\mathbf{E}(t_1 \cdot t_2) = \int_{t_1=0}^{\infty} \int_{t_2=0}^{\infty} \frac{t_1 \cdot t_2 \cdot \theta\alpha\gamma \cdot e^{\alpha(\theta-1)t_1} \cdot e^{\gamma(\theta-1)t_2}}{\left\{ e^{\alpha(\theta-1)t_1} + e^{\gamma(\theta-1)t_2} - 1 \right\}^{\frac{1}{\theta-1}+2}} \, dt_1 \, dt_2$$

Using the transformations $x = \alpha(\theta\text{-}1)t_1$ and $y = \gamma(\theta\text{-}1)t_2$ leads to

$$\mathbf{E}(t_1 \cdot t_2) = \int_{x=0}^{\infty} \int_{y=0}^{\infty} \frac{x \cdot y \cdot e^{x} \cdot e^{y}}{\left\{ e^{x} + e^{y} - 1 \right\}^{\frac{1}{\theta-1}+2}} \cdot \frac{\theta}{\alpha\gamma(\theta\text{-}1)^4} \, dx \, dy$$

$$= \int_{x=0}^{\infty} \frac{\theta x \cdot e^{x}}{\alpha\gamma(\theta\text{-}1)^4} \cdot \left(\int_{y=0}^{\infty} \frac{y \cdot e^{y}}{\{ e^{x} + e^{y} - 1 \}^{\frac{1}{\theta-1}+2}} \, dy \right) dx$$

$$= \int_{x=0}^{\infty} \frac{\theta x \cdot e^{x}}{\alpha\gamma(\theta\text{-}1)^4} \cdot \left(\frac{-(\theta\text{-}1) y \{ e^{x} + e^{y} \text{-}1 \}^{-\frac{1}{\theta-1}-1}}{\theta} \Bigg|_0^{\infty} + \int_{y=0}^{\infty} \frac{(\theta\text{-}1)}{\theta \cdot \{ e^{x} + e^{y} - 1 \}^{\frac{1}{\theta-1}+1}} \, dy \right) dx$$

$$= \int_{x=0}^{\infty} \frac{x \cdot e^{x}}{\alpha\gamma(\theta\text{-}1)^3} \cdot \left(\int_{y=0}^{\infty} \frac{1}{\{ e^{x} + e^{y} - 1 \}^{\frac{1}{\theta-1}+1}} \, dy \right) dx$$

$$= \frac{1}{\alpha\gamma(\theta\text{-}1)^3} \cdot \int_{y=0}^{\infty} \left(\int_{x=0}^{\infty} \frac{x \cdot e^{x}}{\{ e^{x} + e^{y} - 1 \}^{\frac{1}{\theta-1}+1}} \, dx \right) dy$$

$$= \frac{1}{\alpha\gamma(\theta\text{-}1)^3} \cdot \int_{y=0}^{\infty} \left(- \frac{(\theta\text{-}1) \cdot x}{\{ e^{x} + e^{y} - 1 \}^{\frac{1}{\theta-1}}} \Bigg|_0^{\infty} + \int_{x=0}^{\infty} \frac{(\theta\text{-}1)}{\{ e^{x} + e^{y} - 1 \}^{\frac{1}{\theta-1}}} \, dx \right) dy$$

$$= \frac{1}{\alpha\gamma(\theta\text{-}1)^2} \cdot \int_{y=0}^{\infty} \int_{x=0}^{\infty} \frac{1}{\{ e^{x} + e^{y} - 1 \}^{\frac{1}{\theta-1}}} \, dx \, dy$$

By Oakes (1982) , $\int\int \frac{1}{(e^{u}+e^{v}\text{-}1)^{z}} \, du \, dv = \frac{1}{2} \cdot \left\{ \Psi^1(\tfrac{1}{2}z) \text{ -} \Psi^1(\tfrac{1}{2}z + \tfrac{1}{2}) \right\}$,

where

$$\Psi^1(z) = \sum_{n=0}^{\infty} \frac{1}{(n+z)^2} \; .$$

Therefore

$$\mathbf{E}(t_1 \cdot t_2) = \frac{1}{2\alpha\gamma(\theta\text{-}1)^2} \cdot \left\{ \Psi^1(\tfrac{1}{2(\theta-1)}) \text{ -} \Psi^1(\tfrac{1}{2(\theta-1)} + \tfrac{1}{2}) \right\} .$$

$$\text{corr}(t_1, t_2) = \frac{\mathbf{E}(t_1 \cdot t_2) - \mathbf{E}(t_1) \cdot \mathbf{E}(t_2)}{(\mathbf{V}(t_1) \cdot \mathbf{V}(t_2))^{\frac{1}{2}}} = \frac{\frac{1}{2\alpha\gamma(\theta\text{-}1)^2} \cdot \left\{ \Psi^1(\tfrac{1}{2(\theta-1)}) \text{ -} \Psi^1(\tfrac{1}{2(\theta-1)} + \tfrac{1}{2}) \right\} - \frac{1}{\alpha} \cdot \frac{1}{\gamma}}{(\frac{1}{\alpha^2} \cdot \frac{1}{\gamma^2})^{\frac{1}{2}}}$$

$$= \frac{1}{2(1\text{-}\theta)^2} \left\{ \Psi^1 \left(\frac{1}{2(\theta\text{-}1)} \right) - \Psi^1 \left(\frac{\theta}{2(\theta\text{-}1)} \right) \right\} - 1$$

APPENDIX B. DERIVATION OF VARIANCE ESTIMATORS

1) NONPARAMETRIC VARIANCE CALCULATIONS

In the usual linear models setting :

$$\hat{\beta} = (\mathbf{X}^T \cdot \mathbf{X})^{-1} \cdot \mathbf{X}^T \cdot \mathbf{Y}$$

$$\hat{\beta}^* = (\ \hat{\beta}_1 // \hat{\beta}_2 // \ \dots \ // \hat{\beta}_p)$$

$$= \text{vec} \ (\ \hat{\beta}\ )$$

In a general multivariate linear model $\mathbf{V}(\ \hat{\beta}^*) = \Sigma \otimes (\mathbf{X}^T \cdot \mathbf{X})^{-1}$

$$\text{where} \quad \Sigma = \begin{bmatrix} \hat{\text{V}}(\text{Y}_1) & \hat{\text{Cov}}(\text{Y}_1, \text{Y}_2) \\ \hat{\text{Cov}}(\text{Y}_1, \text{Y}_2) & \hat{\text{V}}(\text{Y}_2) \end{bmatrix} = \begin{bmatrix} \hat{\sigma}_1{}^2 & \hat{\sigma}_{12} \\ \hat{\sigma}_{12} & \hat{\sigma}_2{}^2 \end{bmatrix} .$$

For our model a reformulation was developed to take the censoring into consideration through the use of Δ. Thus

$$\mathbf{V}(\ \hat{\beta}^*) = \begin{bmatrix} \hat{\sigma}_1{}^2 \cdot (\mathbf{X}^T \Delta_1 \ \mathbf{X})^{-1} & \hat{\sigma}_{12} \cdot (\mathbf{X}^T \Delta \ \mathbf{X})^{-1} \\ \hat{\sigma}_{12} \cdot (\mathbf{X}^T \Delta \ \mathbf{X})^{-1} & \hat{\sigma}_2{}^2 \cdot (\mathbf{X}^T \Delta_2 \ \mathbf{X})^{-1} \end{bmatrix} \qquad (1.1)$$

where $\begin{cases} \Delta = \text{diag}(\delta_{j1} \cdot \delta_{j2}) & \\ \Delta_i = \text{diag}(\delta_{ji}) & \begin{matrix} \text{i} = 1,2 \\ \text{j} = 1,\dots,\text{n} \end{matrix} \end{cases}$; and the individual variance components are defined as follows:

$$\hat{\sigma}_1{}^2 = \frac{1}{n}\sum_{j=1}^{n}\left\{ \delta_{j1}\cdot e_{j1}{}^2 + (1-\delta_{j1})\cdot\left(\sum_k (\delta_{j2}\cdot u_{jk}^{(1)} + (1-\delta_{j2})\cdot w_{jk})\cdot e_{k1}{}^2\right)\right\}; \tag{1.2}$$

= weighted average of the squares of residuals for the first component

$$\hat{\sigma}_2{}^2 = \frac{1}{n}\sum_{j=1}^{n}\left\{ \delta_{j2}\cdot e_{j2}{}^2 + (1-\delta_{j2})\cdot\left(\sum_k (\delta_{j1}\cdot u_{jk}^{(2)} + (1-\delta_{j1})\cdot w_{jk})\cdot e_{k2}{}^2\right)\right\}; \quad \text{and} \tag{1.3}$$

= weighted average of the squares of residuals for the second component

$$\hat{\sigma}_{12} = \frac{1}{n}\sum_{j=1}^{n}\left\{\begin{array}{c} \delta_{j1}\delta_{j2}\, e_{j1}e_{j2} + (1-\delta_{j1})\cdot\delta_{j2}\cdot\left(\sum_k u_{jk}^{(1)}\; e_{k1}\right)\cdot e_{j2} + \\ \delta_{j1}\cdot(1-\delta_{j2})\cdot e_{j1}\cdot\left(\sum_k u_{jk}^{(2)}\, e_{k2}\right) + (1-\delta_{j1})(1-\delta_{j2})\sum_k w_{jk}\cdot e_{k1}\cdot e_{k2} \end{array}\right\} \tag{1.4}$$

= weighted average of the product of residuals for the first and second components.

The appropriate weights are determined based on the empirical survival distribution (as determined in equation 2.2) as follows:

$$w_{jk} = \begin{cases} \left(\dfrac{w_k(\hat{\beta})}{\hat{S}(\,e_j;\,\hat{\beta})}\right) & \text{for } e_k > e_j \\ 0\,, & \text{otherwise} \end{cases} \quad ;$$

$$u_{jk}^{(1)} = \begin{cases} \left(\dfrac{w_k(\hat{\beta})}{\hat{S}(\,e_{j1},\, e_{\bar{j2}};\,\hat{\beta}) - \hat{S}(\,e_j;\,\hat{\beta})}\right) & \text{for } e_{k1} > e_{j1} \text{ and } e_{k2} = e_{j2} \\ 0\,, & \text{otherwise} \end{cases} \quad ; \text{ and}$$

$$u_{jk}^{(2)} = \begin{cases} \left(\dfrac{w_k(\hat{\beta})}{\hat{S}(e_{j1}^{-}, e_{j2}; \hat{\beta}) - \hat{S}(e_j; \hat{\beta})} \right) & \text{for } e_{k2} > e_{j2} \text{ and } e_{k1} = e_{j1} \\ 0, \quad \text{otherwise} \end{cases}$$

To reduce bias a correction factor $= \begin{cases} \dfrac{n_{ui}}{(n_{ui} - k)} \\ \dfrac{n_u}{(n_u - k)} \end{cases}$ should be included with the estimated

variances and covariance in (1.2), (1.3), and (1.4) where

$$n_{ui} = \sum_{j=1}^{n} \delta_{ji} = \# \text{ uncensored } i^{th} \text{ components}$$

$$n_u = \sum_{j=1}^{n} \delta_{j1} \cdot \delta_{j1} = \# \text{ uncensored pairs}$$

and $k = \text{nrow}(\beta)$.

The variances are computed following the methods of Schneider and Weissfeld (1986) as they can be determined from the marginal suvival distribution function and other information on that component.

$$\begin{matrix} E(\epsilon_{ji}) = 0 \\ V(\epsilon_{ji}) = \sigma_i^2 \end{matrix} \quad \Rightarrow \quad V(\epsilon_{ji}) = \frac{1}{n} \sum_{j=1}^{n} \epsilon_{ji}{}^2 = \frac{1}{n} \sum_{j=1}^{n} (y_{ji} - X_j\beta_i)^2$$

$$\begin{aligned} \text{where } y_{ji} = \delta_{ji}\, y_{ji}^{(0)} &+ (1 - \delta_{ji}) \cdot (1 - \delta_{ji'}) \cdot E[Y_{ji}^{(0)} \mid Y_{ji}^{(0)} > C_{ji}, Y_{ji'}^{(0)} > C_{ji'}, X_j\hat{\beta}] \\ &+ (1 - \delta_{ji}) \cdot \delta_{ji'} \cdot E[Y_{ji}^{(0)} \mid Y_{ji}^{(0)} > C_{ji}, Y_{ji'}^{(0)} = Y_{ji'}, X_j\hat{\beta}] \end{aligned} \tag{B.1}$$

The square of the outcome times are replaced by $\{y_{ji}\}^2$ which is defined as :

$$\begin{aligned} \{y_{ji}\}^2 = \delta_{ji} \left\{y_{ji}^{(0)}\right\}^2 &+ (1 - \delta_{ji}) \cdot (1 - \delta_{ji'}) \cdot E[\left\{Y_{ji}^{(0)}\right\}^2 \mid Y_{ji}^{(0)} > C_{ji}, Y_{ji}^{(0)} > C_{ji'}, X_j\hat{\beta}] \\ &+ (1 - \delta_{ji}) \cdot \delta_{ji'} \cdot E[\left\{Y_{ji}^{(0)}\right\}^2 \mid Y_{ji}^{(0)} > C_{ji}, Y_{ji}^{(0)} = Y_{ji'}, X_j\hat{\beta}] \end{aligned} \tag{B.2}$$

The estimated variance for the second component is determined similarly.

$$\hat{\mathrm{Cov}}(\epsilon_{j1}, \epsilon_{j2}) = \hat{\sigma}_{12}{}^2 = \frac{1}{n} \cdot \left\{ \sum_{j=1}^{n} (\hat{\epsilon}_{j1} \cdot \hat{\epsilon}_{j2}) \right\}$$

$$= \frac{1}{n} \cdot \sum_{j=1}^{n} \left\{ \begin{array}{c} \delta_{j1} \cdot \delta_{j2} \cdot \hat{\mathrm{E}}[\epsilon_{j1} \cdot \epsilon_{j2} \mid \delta_{j1} = 1, \delta_{j2} = 1] \; + \\ (1\text{-}\delta_{j1}) \cdot \delta_{j2} \cdot \hat{\mathrm{E}}[\epsilon_{j1} \cdot \epsilon_{j2} \mid \delta_{j1} = 0, \delta_{j2} = 1] + \\ \delta_{j1} \cdot (1\text{-}\delta_{j2}) \cdot \hat{\mathrm{E}}[\epsilon_{j1} \cdot \epsilon_{j2} \mid \delta_{j1} = 1, \delta_{j2} = 0] + \\ (1\text{-}\delta_{j1}) \cdot (1\text{-}\delta_{j2}) \cdot \hat{\mathrm{E}}[\epsilon_{j1} \cdot \epsilon_{j2} \mid \delta_{j1} = 0, \delta_{j2} = 0] \end{array} \right\}$$

$$= \frac{1}{n} \cdot \sum_{j=1}^{n} \left\{ \begin{array}{c} \delta_{j1} \cdot \delta_{j2} \cdot \mathrm{e}_{j1} \cdot \mathrm{e}_{j2} \; + \\ (1\text{-}\delta_{j1}) \cdot \delta_{j2} \cdot \hat{\mathrm{E}}[\epsilon_{j1} \mid \delta_{j1} = 0, \delta_{j2} = 1] \cdot \mathrm{e}_{j2} + \\ \delta_{j1} \cdot (1\text{-}\delta_{j2}) \cdot \mathrm{e}_{j1} \cdot \hat{\mathrm{E}}[\epsilon_{j2} \mid \delta_{j1} = 1, \delta_{j2} = 0] + \\ (1\text{-}\delta_{j1}) \cdot (1\text{-}\delta_{j2}) \cdot \hat{\mathrm{E}}[\epsilon_{j1} \cdot \epsilon_{j2} \mid \delta_{j1} = 0, \delta_{j2} = 0] \end{array} \right\} \tag{B.4}$$

Thus

$$\hat{\mathrm{Cov}}(\epsilon_{j1}, \epsilon_{j2}) = \frac{1}{n} \sum_{j=1}^{n} \left\{ \begin{array}{c} \delta_{j1}\delta_{j2}\, \mathrm{e}_{j1}\mathrm{e}_{j2} + (1 - \delta_{j1}) \cdot \delta_{j2} \cdot \left(\sum_k \mathrm{u}_{jk}^{(1)} \; \mathrm{e}_{k1} \right) \cdot \mathrm{e}_{j2} + \\ \delta_{j1} \cdot (1 - \delta_{j2}) \cdot \mathrm{e}_{j1} \cdot \left(\sum_k \mathrm{u}_{jk}^{(2)} \; \mathrm{e}_{k2} \right) + (1 - \delta_{j1})(1 - \delta_{j2}) \sum_k \mathrm{w}_{jk} \cdot \mathrm{e}_{k1} \cdot \mathrm{e}_{k2} \end{array} \right\}$$

In the absence of censoring (i.e., all failure times are observed) the variance and covariance estimates reduce to the usual maximum likelihood estimates.

2) PARAMETRIC VARIANCE CALCULATIONS

The variance and covariance calculations for a specific distribution are similar to the nonparametric situation, but the expected values of residuals, squared residuals and the product of residuals now depend on the specified distribution. Also, due to the nature of the distribution used, while the expected value of the residuals is quite small, it is not zero; this is a result of e^{ϵ} having the exponential or Weibull distribution. Thus to appropriately estimate the variance and covariance, these terms must be included in the calculations.

$$\hat{\mathrm{V}}(\epsilon_{j1}) = \hat{\sigma}_1^{\,2} = \frac{1}{n}\cdot\sum_{j=1}^{n}(\hat{y}_{j1} - \mathbf{X}_j\hat{\boldsymbol{\beta}}_1)^2 - \left(\frac{1}{n}\cdot\sum_{j=1}^{n}(\hat{y}_{j1} - \mathbf{X}_j\hat{\boldsymbol{\beta}}_1)\right)^2$$

$$= \frac{1}{n}\cdot\sum_{j=1}^{n}(\hat{\epsilon}_{j1}^{\,2}) - \left(\frac{1}{n}\cdot\sum_{j=1}^{n}\hat{\epsilon}_{j1}\right)^2$$

$$= \frac{1}{n}\sum_{j=1}^{n}\left\{\begin{array}{c}\delta_{j1}\cdot e_{j1}^{\,2} + (1\text{-}\delta_{j1})\cdot(1\text{-}\delta_{j2})\cdot\hat{\mathrm{E}}[\epsilon_{j1}^{\,2} \mid \delta_{j1}{=}0,\, \delta_{j2}{=}0] \\ + (1\text{-}\delta_{j1})\cdot\delta_{j2}\cdot\hat{\mathrm{E}}[\epsilon_{j1}^{\,2} \mid \delta_{j1}{=}0,\, \delta_{j2}{=}1]\end{array}\right\}$$

$$- \left(\frac{1}{n}\cdot\sum_{j=1}^{n}\left\{\begin{array}{c}\delta_{j1}\cdot e_{j1} + (1\text{-}\delta_{j1})\cdot(1\text{-}\delta_{j2})\cdot\hat{\mathrm{E}}[\epsilon_{j1} \mid \delta_{j1}{=}0,\, \delta_{j2}{=}0] \\ + (1\text{-}\delta_{j1})\cdot\delta_{j2}\cdot\hat{\mathrm{E}}[\epsilon_{j1} \mid \delta_{j1}{=}0,\, \delta_{j2}{=}1]\end{array}\right\}\right)^2 \qquad \text{(B.5)}$$

$$\hat{\mathrm{Cov}}(\epsilon_{j1}, \epsilon_{j2}) = \hat{\sigma}_{12}^{\,2} = \frac{1}{n}\cdot\left\{\sum_{j=1}^{n}(\hat{\epsilon}_{j1}\cdot\hat{\epsilon}_{j2})\right\} - \left\{\left(\frac{1}{n}\cdot\sum_{j=1}^{n}\hat{\epsilon}_{j1}\right)\cdot\left(\frac{1}{n}\cdot\sum_{j=1}^{n}\hat{\epsilon}_{j2}\right)\right\}$$

$$= \frac{1}{n}\cdot\sum_{j=1}^{n}\left\{\begin{array}{c}\delta_{j1}\cdot\delta_{j2}\cdot e_{j1}\cdot e_{j2} + \\ (1\text{-}\delta_{j1})\cdot\delta_{j2}\cdot\hat{\mathrm{E}}[\epsilon_{j1}\cdot\epsilon_{j2} \mid \delta_{j1} = 0,\, \delta_{j2} = 1] + \\ \delta_{j1}\cdot(1\text{-}\delta_{j2})\cdot\hat{\mathrm{E}}[\epsilon_{j1}\cdot\epsilon_{j2} \mid \delta_{j1} = 1,\, \delta_{j2} = 0] + \\ (1\text{-}\delta_{j1})\cdot(1\text{-}\delta_{j2})\cdot\hat{\mathrm{E}}[\epsilon_{j1}\cdot\epsilon_{j2} \mid \delta_{j1} = 0,\, \delta_{j2} = 0]\end{array}\right\}$$

$$- \left(\frac{1}{n}\cdot\sum_{j=1}^{n}\left\{\begin{array}{c}\delta_{j1}\cdot e_{j1} + (1\text{-}\delta_{j1})\cdot(1\text{-}\delta_{j2})\cdot\hat{\mathrm{E}}[\epsilon_{j1} \mid \delta_{j1}{=}0,\, \delta_{j2}{=}0] \\ + (1\text{-}\delta_{j1})\cdot\delta_{j2}\cdot\hat{\mathrm{E}}[\epsilon_{j1} \mid \delta_{j1}{=}0,\, \delta_{j2}{=}1]\end{array}\right\}\right)\cdot$$

$$\left(\frac{1}{n}\cdot\sum_{j=1}^{n}\left\{\begin{array}{c}\delta_{j2}\cdot e_{j2} + (1\text{-}\delta_{j1})\cdot(1\text{-}\delta_{j2})\cdot\hat{\mathrm{E}}[\epsilon_{j2} \mid \delta_{j1}{=}0,\, \delta_{j2}{=}0] \\ + \delta_{j1}\cdot(1\text{-}\delta_{j2})\cdot\hat{\mathrm{E}}[\epsilon_{j2} \mid \delta_{j1}{=}1,\, \delta_{j2}{=}0]\end{array}\right\}\right) \qquad \text{(B.6)}$$

The appropriate expected values needed to obtain the variance and covariance for each distribution can be found in Chapter 4.

APPENDIX C. COMPUTATIONS FOR PARAMETRIC VARIANCE AND COVARIANCE ESTIMATES

1) FUNCTIONS NECESSARY TO CALCULATE EXPECTATIONS

Since the Weibull distribution is obtained from the exponential by the following relationship :

$$(e^{\epsilon_1}, e^{\epsilon_2})_{Weibull} = (e^{\nu \cdot \epsilon_1}, e^{\phi \cdot \epsilon_2})_{exp} \quad ,$$

the expected values for the Weibull distribution are proportional to the corresponding values for the exponential distribution. For example :

$$E\left(\epsilon_1 \middle| \begin{matrix} Y_1 > c_1, Y_2 > c_2 \\ (e^{\epsilon_1}, e^{\epsilon_2}) \sim \text{Weibull(I)} \end{matrix}\right) = \tfrac{1}{\nu} \cdot E\left(\epsilon_1 \middle| \begin{matrix} Y_1 > \nu c_1, Y_2 > \phi c_2 \\ (e^{\epsilon_1}, e^{\epsilon_2}) \sim \exp(\text{I}) \end{matrix}\right)$$

Thus it is only necessary to calculate the expected values for the two exponential distributions. For both the variance estimate and the covariance estimate, it is necessary to obtain expected values of the residuals. In addition, for the variance estimate the expected value of ϵ_i^2 must also be determined. The expected value of $\epsilon_1 \cdot \epsilon_2$ is needed for the covariance estimate.

Many of these calculations involve $\int_k^\infty (\ln(x))^2 \cdot e^{-x}\, dx$ which is determined as follows :

$$\int_0^\infty e^{-\mu x} \cdot (\ln(x))^2 \; dx \;=\; \tfrac{1}{\mu} \cdot [\tfrac{\pi^2}{6} + \{\gamma + \ln(\mu)\}^2]$$

Gradshteyn and Ryzhik (p.574, 4.335(1))

Thus $$\int_k^\infty e^{-\mu x} \cdot (\ln(x))^2 \; dx \;=\; \int_0^\infty e^{-\mu x} \cdot (\ln(x))^2 \; dx \;-\; \int_0^k e^{-\mu x} \cdot (\ln(x))^2 \; dx$$

$$\int_0^k e^{-\mu x} \cdot (\ln(x))^2 \; dx \;=\; \int_0^k \sum_{i=0}^{\infty} \frac{(-\mu x)^i}{i!} \cdot (\ln(x))^2 \; dx$$

$$= \sum_{i=0}^{\infty} \frac{(-\mu)^i}{i!} \cdot \left\{ x^{i+1} \cdot \left(\frac{(\ln(x))^2}{i+1} - \frac{2\ln(x)}{(i+1)^2} + \frac{2}{(i+1)^3} \right) \right\} \Bigg|_0^k$$

Since $\lim\limits_{x \to 0} \; (\ln(x))^2 \cdot x^{i+1} \;=\; \lim\limits_{x \to 0} \; \dfrac{(\ln(x))^2}{x^{-(i+1)}}$.

L'Hospital's rule $\Rightarrow$ $\lim\limits_{x \to 0} \; \dfrac{\frac{2 \cdot \ln(x)}{x}}{-(i+1) \cdot x^{-(i+2)}} \;=\; \lim\limits_{x \to 0} \; \dfrac{2 \cdot \ln(x)}{-(i+1) \cdot x^{-(i+1)}}$

L'Hospital's rule $\Rightarrow$ $\lim\limits_{x \to 0} \; \dfrac{\frac{2}{x}}{(i+1)^2 \cdot x^{-(i+2)}} \;=\; \lim\limits_{x \to 0} \; \dfrac{2}{(i+1)^2 \cdot x^{-(i+1)}}$

$$= \lim_{x \to 0} \; \frac{2 \cdot x^{(i+1)}}{(i+1)^2} \;=\; 0$$

and similarly $\lim\limits_{x \to 0} \; \ln(x) \cdot x^{(i+1)} \;=\; 0$,

then $\lim\limits_{x \to 0} \; \sum\limits_{i=0}^{\infty} \dfrac{(-\mu)^i}{i!} \cdot \left\{ x^{i+1} \cdot \left(\dfrac{(\ln(x))^2}{i+1} - \dfrac{2\ln(x)}{(i+1)^2} + \dfrac{2}{(i+1)^3} \right) \right\} = 0$.

Thus $\displaystyle\int_k^{\infty} e^{-\mu x} \cdot (\ln(x))^2 \; dx \;=\; \int_0^{\infty} e^{-\mu x} \cdot (\ln(x))^2 \; dx \;-\; \int_0^k e^{-\mu x} \cdot (\ln(x))^2 \; dx$

$$= \frac{1}{\mu} \cdot [\frac{\pi^2}{6} + \{\gamma + \ln(\mu)\}^2]$$

$$- \sum_{i=0}^{\infty} \frac{(-\mu)^i}{i!} \cdot \left\{ k^{i+1} \cdot \left(\frac{(\ln(k))^2}{i+1} - \frac{2\ln(k)}{(i+1)^2} + \frac{2}{(i+1)^3} \right) \right\}$$

$$= \frac{1}{\mu} \cdot [\frac{\pi^2}{6} + \{\gamma + \ln(\mu)\}^2] + \sum_{i=0}^{\infty} \frac{(-\mu k)^{i+1}}{(i+1)!} \cdot (\ln(k))^2 \cdot \frac{1}{\mu}$$

$$- \sum_{i=0}^{\infty} \frac{(-\mu k)^{i+1}}{(i+1)! \cdot (i+1)} \cdot 2\ln(k) \cdot \frac{1}{\mu} + \sum_{i=0}^{\infty} \frac{(-\mu k)^{i+1}}{(i+1)! \cdot (i+1)^2} \cdot 2 \cdot \frac{1}{\mu}$$

Therefore

$$\boxed{\begin{aligned} \int_k^{\infty} e^{-\mu x} \cdot (\ln(x))^2\, dx \;=\; & \frac{1}{\mu} \cdot [\frac{\pi^2}{6} + \{\gamma + \ln(\mu)\}^2] + (e^{-\mu k} - 1) \cdot (\ln(k))^2 \cdot \frac{1}{\mu} \\ & + [-\mathbf{E}i(-\mu k) + \gamma + \ln(\mu k)] \cdot 2\ln(k) \cdot \frac{1}{\mu} + \sum_{i=0}^{\infty} \frac{(-\mu k)^{i+1}}{(i+1)! \cdot (i+1)^2} \cdot 2 \cdot \frac{1}{\mu} \end{aligned}}$$

(C.3)

2) EXPECTED VALUE OF RESIDUALS

A. Gumbel's exponential (I)

(i) $$E\left(\epsilon_1 \,\middle|\, \begin{matrix} Y_1 > c_1,\ Y_2 > c_2 \\ (e^{\epsilon_1}, e^{\epsilon_2}) \sim \exp(I) \end{matrix}\right)$$

$$= \frac{\int_{r_1}^{\infty}\int_{r_2}^{\infty} \epsilon_1 \cdot \{(1+\theta e^{\epsilon_1})(1+\theta e^{\epsilon_2})-\theta\} \cdot e^{-\theta e^{\epsilon_1+\epsilon_2} - e^{\epsilon_1} - e^{\epsilon_2} + \epsilon_1 + \epsilon_2}\, \partial\epsilon_1 \partial\epsilon_2}{\int_{r_1}^{\infty}\int_{r_2}^{\infty} \{(1+\theta e^{\epsilon_1})(1+\theta e^{\epsilon_2})-\theta\} \cdot e^{-\theta e^{\epsilon_1+\epsilon_2} - e^{\epsilon_1} - e^{\epsilon_2} + \epsilon_1 + \epsilon_2}\, \partial\epsilon_1 \partial\epsilon_2}$$

where $r_1 = c_1 - X\beta_1 \quad r_2 = c_2 - X\beta_2$

Let $e^{\epsilon_1} = v \quad \Rightarrow \quad \epsilon_1 = \ln(v) \qquad \partial\epsilon_1 = \frac{\partial v}{v}$

$e^{\epsilon_2} = w \quad \Rightarrow \quad \epsilon_2 = \ln(w) \qquad \partial\epsilon_2 = \frac{\partial w}{w}$

$$E\left(\epsilon_1 \,\middle|\, \begin{matrix} Y_1 > c_1,\ Y_2 > c_2 \\ (e^{\epsilon_1}, e^{\epsilon_2}) \sim \exp(I) \end{matrix}\right)$$

$$= \frac{\int_{e^{r_1}}^{\infty} \ln(v) \cdot e^{-v} \left(\{1+\theta v-\theta\} \cdot \int_{e^{r_2}}^{\infty} e^{-(1+\theta v)w} \partial w + \{\theta(1+\theta v)\} \cdot \int_{e^{r_2}}^{\infty} w \cdot e^{-(1+\theta v)w} \partial w \right) \partial v}{\int_{e^{r_1}}^{\infty} e^{-v} \left(\{1+\theta v-\theta\} \cdot \int_{e^{r_2}}^{\infty} e^{-(1+\theta v)w} \partial w + \{\theta(1+\theta v)\} \cdot \int_{e^{r_2}}^{\infty} w \cdot e^{-(1+\theta v)w} \partial w \right) \partial v}$$

$$= \frac{\int_{e^{r_1}}^{\infty} \ln(v) e^{-v} \left(\{1+\theta v-\theta\} \frac{e^{-(1+\theta v)e^{r_2}}}{(1+\theta v)} + \theta(1+\theta v) \left\{ \frac{e^{r_2}}{(1+\theta v)} + \frac{1}{(1+\theta v)^2} \right\} \cdot e^{-(1+\theta v)e^{r_2}} \right) \partial v}{\int_{e^{r_1}}^{\infty} e^{-v} \left(\{1+\theta v-\theta\} \cdot \frac{e^{-(1+\theta v)e^{r_2}}}{(1+\theta v)} + \theta(1+\theta v) \cdot \left\{ \frac{e^{r_2}}{(1+\theta v)} + \frac{1}{(1+\theta v)^2} \right\} \cdot e^{-(1+\theta v)e^{r_2}} \right) \partial v}$$

$$= \frac{\int_{e^{r_1}}^{\infty} \ln(v) \cdot e^{-v} \cdot (1+\theta e^{r_2}) \cdot e^{-(1+\theta v)e^{r_2}} \partial v}{\int_{e^{r_1}}^{\infty} e^{-v} \cdot (1+\theta e^{r_2}) \cdot e^{-(1+\theta v)e^{r_2}} \partial v}$$

$$= \frac{\int_{e^{r_1}}^{\infty} \ln(v) \cdot e^{-v(1+\theta e^{r_2})}\, \partial v}{\int_{e^{r_1}}^{\infty} e^{-v(1+\theta e^{r_2})}\, \partial v}$$

$$= \frac{\frac{1}{(1+\theta e^{r_2})} \cdot \left\{ -\ln(v) \cdot e^{-v(1+\theta e^{r_2})} \Big|_{e^{r_1}}^{\infty} + \int_{e^{r_1}}^{\infty} \frac{1}{v} \cdot e^{-v(1+\theta e^{r_2})} \, \partial v \right\}}{\frac{e^{-e^{r_1}(1+\theta e^{r_2})}}{(1+\theta e^{r_2})}}$$

$$= \frac{\left\{ r_1 \cdot e^{-e^{r_1}(1+\theta e^{r_2})} - Ei(-e^{r_1}(1+\theta e^{r_2})) \right\}}{e^{-e^{r_1}(1+\theta e^{r_2})}}$$

where $-Ei(-b) = \int_b^{\infty} \frac{e^{-x}}{x} dx = -\gamma - \log(b) - \sum_{i=1}^{\infty} \frac{(-b)^i}{i \cdot i!}$
($\gamma = 0.5772157\ldots =$ Euler's constant)

Thus

$$\boxed{E\left(\epsilon_1 \,\middle|\, \begin{array}{c} Y_1 > c_1,\ Y_2 > c_2 \\ (e^{\epsilon_1}, e^{\epsilon_2}) \sim \exp(I) \end{array} \right) = r_1 - \frac{Ei(-e^{r_1}(1+\theta e^{r_2}))}{e^{-e^{r_1}(1+\theta e^{r_2})}}}$$

(ii) $E\left(\epsilon_1 \,\middle|\, \begin{matrix} Y_1 > c_1,\ Y_2 = y_2 \\ (e^{\epsilon_1}, e^{\epsilon_2}) \sim \exp(I) \end{matrix}\right)$

$$= \frac{\int_{r_1}^{\infty} \epsilon_1 \cdot \{(1+\theta e^{\epsilon_1})(1+\theta e^{r_2})-\theta\} \cdot e^{-\theta e^{\epsilon_1 + r_2} - e^{\epsilon_1} - e^{r_2} + \epsilon_1 + r_2}\, \partial\epsilon_1}{\int_{r_1}^{\infty} \{(1+\theta e^{\epsilon_1})(1+\theta e^{r_2})-\theta\} \cdot e^{-\theta e^{\epsilon_1 + r_2} - e^{\epsilon_1} - e^{r_2} + \epsilon_1 + r_2}\, \partial\epsilon_1}$$

where $r_1 = c_1 - X\beta_1 \quad r_2 = y_2 - X\beta_2$

Let $\quad e^{\epsilon_1} = v \quad \Rightarrow \quad \epsilon_1 = \ln(v) \qquad \partial\epsilon_1 = \frac{\partial v}{v}$

$$= \frac{\int_{e^{r_1}}^{\infty} \ln(v) \cdot \{(1+\theta v)(1+\theta e^{r_2})-\theta\} \cdot e^{-\theta v e^{r_2} - v - e^{r_2} + r_2}\, \partial v}{\int_{e^{r_1}}^{\infty} \{(1+\theta v)(1+\theta e^{r_2})-\theta\} \cdot e^{-\theta v e^{r_2} - v - e^{r_2} + r_2}\, \partial v}$$

$$= \frac{\{1+\theta e^{r_2}-\theta\} \cdot \int_{e^{r_1}}^{\infty} \ln(v) \cdot e^{-(1+\theta e^{r_2})v} \partial v + \theta(1+\theta e^{r_2}) \int_{e^{r_1}}^{\infty} \ln(v) \cdot v \cdot e^{-(1+\theta e^{r_2})v} \partial v}{\int_{e^{r_1}}^{\infty} \{(1+\theta v)(1+\theta e^{r_2})-\theta\} \cdot e^{-(1+\theta e^{r_2})v}\, \partial v}$$

$$= \frac{\left\{(1+\theta e^{r_2}-\theta) \cdot \left.\frac{-\ln(v) \cdot e^{-(1+\theta e^{r_2})v}}{(1+\theta e^{r_2})}\right|_{e^{r_1}}^{\infty}\right\} + \left\{(1+\theta e^{r_2}-\theta) \cdot \int_{e^{r_1}}^{\infty} \frac{e^{-(1+\theta e^{r_2})v}}{(1+\theta e^{r_2}) \cdot v}\, \partial v\right\}}{\left\{(1+\theta e^{r_2}-\theta) \cdot \int_{e^{r_1}}^{\infty} e^{-(1+\theta e^{r_2})v}\, \partial v\right\} + \left\{\theta(1+\theta e^{r_2}) \cdot \int_{e^{r_1}}^{\infty} v \cdot e^{-(1+\theta e^{r_2})v}\, \partial v\right\}}$$

$$+ \frac{\theta(1+\theta e^{r_2})\left\{\left.\frac{-\ln(v)\left(v + \frac{1}{1+\theta e^{r_2}}\right)e^{-(1+\theta e^{r_2})v}}{(1+\theta e^{r_2})}\right|_{e^{r_1}}^{\infty} + \int_{e^{r_1}}^{\infty} \frac{e^{-(1+\theta e^{r_2})v} \cdot \left(v + \frac{1}{1+\theta e^{r_2}}\right)}{(1+\theta e^{r_2}) \cdot v}\, \partial v\right\}}{\left\{(1+\theta e^{r_2}-\theta) \cdot \int_{e^{r_1}}^{\infty} e^{-(1+\theta e^{r_2})v}\, \partial v\right\} + \left\{\theta(1+\theta e^{r_2}) \cdot \int_{e^{r_1}}^{\infty} v \cdot e^{-(1+\theta e^{r_2})v}\, \partial v\right\}}$$

$$= \frac{(1+\theta e^{r_2}-\theta)\cdot\left\{\dfrac{r_1\cdot e^{-e^{r_1}\cdot(1+\theta e^{r_2})}}{(1+\theta e^{r_2})}\right\} + (1+\theta e^{r_2}-\theta)\cdot\left\{\displaystyle\int_{e^{r_1}}^{\infty}\frac{e^{-(1+\theta e^{r_2})v}}{(1+\theta e^{r_2})\cdot v}\,\partial v\right\}}{\left\{(1+\theta e^{r_2}-\theta)\cdot\dfrac{e^{-e^{r_1}\cdot(1+\theta e^{r_2})}}{(1+\theta e^{r_2})}\right\} + \left\{\theta(1+\theta e^{r_2})\cdot\dfrac{e^{-e^{r_1}\cdot(1+\theta e^{r_2})}}{(1+\theta e^{r_2})}\cdot\left(e^{r_1}+\dfrac{1}{1+\theta e^{r_2}}\right)\right\}}$$

$$+ \frac{\theta\left\{r_1\cdot\left(e^{r_1}+\dfrac{1}{1+\theta e^{r_2}}\right)\cdot e^{-e^{r_1}\cdot(1+\theta e^{r_2})} + \displaystyle\int_{e^{r_1}}^{\infty} e^{-(1+\theta e^{r_2})v}\,\partial v + \int_{e^{r_1}}^{\infty}\frac{e^{-(1+\theta e^{r_2})v}}{(1+\theta e^{r_2})\cdot v}\,\partial v\right\}}{\left\{(1+\theta e^{r_2}-\theta)\cdot\dfrac{e^{-e^{r_1}\cdot(1+\theta e^{r_2})}}{(1+\theta e^{r_2})}\right\} + \left\{\theta(1+\theta e^{r_2})\cdot\dfrac{e^{-e^{r_1}\cdot(1+\theta e^{r_2})}}{(1+\theta e^{r_2})}\cdot\left(e^{r_1}+\dfrac{1}{1+\theta e^{r_2}}\right)\right\}}$$

$$= \frac{\left\{(1+\theta e^{r_1})\, r_1 e^{-e^{r_1}\cdot(1+\theta e^{r_2})}\right\} + \overbrace{\left\{\displaystyle\int_{e^{r_1}}^{\infty}\frac{e^{-(1+\theta e^{r_2})v}}{v}\,\partial v\right\}}^{-Ei(-e^{r_1}(1+\theta e^{r_2}))} + \left\{\theta\cdot\displaystyle\int_{e^{r_1}}^{\infty} e^{-(1+\theta e^{r_2})v}\,\partial v\right\}}{(1+\theta e^{r_1})\cdot e^{-e^{r_1}\cdot(1+\theta e^{r_2})}}$$

Thus

$$\boxed{\begin{aligned} &E\left(\epsilon_1 \,\middle|\, \begin{matrix} Y_1 > c_1,\ Y_2 = y_2 \\ (e^{\epsilon_1}, e^{\epsilon_2}) \sim \exp(I) \end{matrix}\right) \\ &= r_1 - \frac{Ei(-e^{r_1}(1+\theta e^{r_2}))}{(1+\theta e^{r_1})\cdot e^{-e^{r_1}(1+\theta e^{r_2})}} + \frac{\theta}{(1+\theta e^{r_1})\cdot(1+\theta e^{r_2})} \end{aligned}}$$

B. Gumbel's exponential (II)

(i) $E\left(\epsilon_1 \middle| \begin{array}{l} Y_1 > c_1,\ Y_2 > c_2 \\ (e^{\epsilon_1}, e^{\epsilon_2}) \sim \exp(II) \end{array}\right) =$

$$\frac{\int_{r_1}^{\infty}\int_{r_2}^{\infty} \epsilon_1 \cdot \{1+\theta(2e^{-e^{\epsilon_1}}-1)(2e^{-e^{\epsilon_2}}-1)\} \cdot e^{-e^{\epsilon_1}-e^{\epsilon_2}+\epsilon_1+\epsilon_2}\, \partial\epsilon_1 \partial\epsilon_2}{\int_{r_1}^{\infty}\int_{r_2}^{\infty} \{1+\theta(2e^{-e^{\epsilon_1}}-1)(2e^{-e^{\epsilon_2}}-1)\} \cdot e^{-e^{\epsilon_1}-e^{\epsilon_2}+\epsilon_1+\epsilon_2}\, \partial\epsilon_1 \partial\epsilon_2}$$

where $r_1 = c_1 - X\beta_1 \quad r_2 = c_2 - X\beta_2$

Let $\quad e^{\epsilon_1} = v \quad \Rightarrow \quad \epsilon_1 = \ln(v) \qquad \partial\epsilon_1 = \frac{\partial v}{v}$

$\qquad e^{\epsilon_2} = w \quad \Rightarrow \quad \epsilon_2 = \ln(w) \qquad \partial\epsilon_2 = \frac{\partial w}{w}$

$E\left(\epsilon_1 \middle| \begin{array}{l} Y_1 > c_1,\ Y_2 = y_2 \\ (e^{\epsilon_1}, e^{\epsilon_2}) \sim \exp(II) \end{array}\right)$

$$= \frac{\int_{e^{r_1}}^{\infty} \ln(v) \cdot e^{-v} \left(\{1-\theta(2e^{-v}-1)\} \cdot \int_{e^{r_2}}^{\infty} e^{-w} \partial w + \{2\theta(2e^{-v}-1)\} \cdot \int_{e^{r_2}}^{\infty} e^{-2w} \partial w \right) \partial v}{\int_{e^{r_1}}^{\infty} e^{-v} \left(\{1-\theta(2e^{-v}-1)\} \cdot \int_{e^{r_2}}^{\infty} e^{-w} \partial w + \{2\theta(2e^{-v}-1)\} \cdot \int_{e^{r_2}}^{\infty} e^{-2w} \partial w \right) \partial v}$$

$$= \frac{\int_{e^{r_1}}^{\infty} \ln(v) \cdot e^{-v} \left(\{1-\theta(2e^{-v}-1)\} \cdot e^{-e^{r_2}} + \theta(2e^{-v}-1) \cdot e^{-2e^{r_2}} \right) \partial v}{\int_{e^{r_1}}^{\infty} e^{-v} \left(\{1-\theta(2e^{-v}-1)\} \cdot e^{-e^{r_2}} + \theta(2e^{-v}-1) \cdot e^{-2e^{r_2}} \right) \partial v}$$

$$= \frac{\left(\{1-\theta(e^{-e^{r_2}}-1)\} \cdot \int_{e^{r_1}}^{\infty} \ln(v) \cdot e^{-v} \partial v \ + \ 2\theta(e^{-e^{r_2}}-1) \cdot \int_{e^{r_1}}^{\infty} \ln(v) \cdot e^{-2v} \partial v \right) \cdot e^{-e^{r_2}}}{\left(\{1-\theta(e^{-e^{r_2}}-1)\} \cdot \int_{e^{r_1}}^{\infty} e^{-v} \partial v \ + \ 2\theta(e^{-e^{r_2}}-1) \cdot \int_{e^{r_1}}^{\infty} e^{-2v} \partial v \right) \cdot e^{-e^{r_2}}}$$

$$= \frac{\{1-\theta(e^{-e^{r_2}}-1)\}\cdot[-\ln(v)\cdot e^{-v}\Big|_{e^{r_1}}^{\infty} + \int_{e^{r_1}}^{\infty}\frac{1}{v}\cdot e^{-v}\,\partial v]}{\left(\{1-\theta(e^{-e^{r_2}}-1)\}\cdot e^{-e^{r_1}} + \theta(e^{-e^{r_2}}-1)\cdot e^{-2e^{r_1}}\right)} +$$

$$\frac{\theta(e^{-e^{r_2}}-1)\cdot[-\ln(v)\cdot e^{-2v}\Big|_{e^{r_1}}^{\infty} + \int_{e^{r_1}}^{\infty}\frac{1}{v}\cdot e^{-2v}\,\partial v]}{\left(\{1-\theta(e^{-e^{r_2}}-1)\}\cdot e^{-e^{r_1}} + \theta(e^{-e^{r_2}}-1)\cdot e^{-2e^{r_1}}\right)}$$

$$= \frac{\{1-\theta(e^{-e^{r_2}}-1)\}\cdot[r_1\cdot e^{-e^{r_1}} - Ei(-e^{r_1})] + \theta(e^{-e^{r_2}}-1)\cdot[r_1\cdot e^{-2e^{r_1}} - Ei(-2e^{r_1})]}{\left(1 + \theta(e^{-e^{r_1}}-1)\cdot(e^{-e^{r_2}}-1)\right)\cdot e^{-e^{r_1}}}$$

Thus

$$\boxed{\begin{aligned} &E\left(\epsilon_1 \,\middle|\, \begin{matrix} Y_1 > c_1,\ Y_2 > c_2 \\ (e^{\epsilon_1}, e^{\epsilon_2}) \sim \exp(II) \end{matrix}\right) \\ &= r_1 + \frac{\{1-\theta(e^{-e^{r_2}}-1)\}\cdot[-Ei(-e^{r_1})] + \theta(e^{-e^{r_2}}-1)\cdot[-Ei(-2e^{r_1})]}{\left(1 + \theta(e^{-e^{r_1}}-1)\cdot(e^{-e^{r_2}}-1)\right)\cdot e^{-e^{r_1}}} \end{aligned}}$$

$$\text{(ii) } \mathbf{E}\left(\epsilon_1 \,\middle|\, \begin{matrix} Y_1 > c_1,\ Y_2 = y_2 \\ (e^{\epsilon_1}, e^{\epsilon_2}) \sim \exp(II) \end{matrix}\right) = \frac{e^{-e^{r_2}+r_2} \int_{r_1}^{\infty} \epsilon_1\{1+\theta(2e^{-e^{\epsilon_1}}-1)(2e^{-e^{r_2}}-1)\}e^{-e^{\epsilon_1}+\epsilon_1}\,\partial\epsilon_1}{e^{-e^{r_2}+r_2} \int_{r_1}^{\infty}\{1+\theta(2e^{-e^{\epsilon_1}}-1)(2e^{-e^{r_2}}-1)\}\cdot e^{-e^{\epsilon_1}+\epsilon_1}\,\partial\epsilon_1}$$

where $r_1 = c_1 - X\beta_1 \quad r_2 = y_2 - X\beta_2$

Let $\quad e^{\epsilon_1} = v \quad \Rightarrow \quad \epsilon_1 = \ln(v) \qquad \partial\epsilon_1 = \frac{\partial v}{v}$

$$\mathbf{E}\left(\epsilon_1 \,\middle|\, \begin{matrix} Y_1 > c_1,\ Y_2 = y_2 \\ (e^{\epsilon_1}, e^{\epsilon_2}) \sim \exp(II) \end{matrix}\right) = \frac{\int_{e^{r_1}}^{\infty} \ln(v)\cdot\{1+\theta(2e^{-v}-1)(2e^{-e^{r_2}}-1)\}\cdot e^{-v}\,\partial v}{\int_{e^{r_1}}^{\infty} \{1+\theta(2e^{-v}-1)(2e^{-e^{r_2}}-1)\}\cdot e^{-v}\,\partial v}$$

$$= \frac{\{1-\theta(2e^{-e^{r_2}}-1)\}\cdot\int_{e^{r_1}}^{\infty}\ln(v)\cdot e^{-v}\,\partial v \;+\; 2\theta(2e^{-e^{r_2}}-1)\}\cdot\int_{e^{r_1}}^{\infty}\ln(v)\cdot e^{-2v}\,\partial v}{\{1-\theta(2e^{-e^{r_2}}-1)\}\cdot\int_{e^{r_1}}^{\infty} e^{-v}\,\partial v \;+\; 2\theta(2e^{-e^{r_2}}-1)\}\cdot\int_{e^{r_1}}^{\infty} e^{-2v}\,\partial v}$$

$$= \frac{\{1-\theta(2e^{-e^{r_2}}-1)\}\cdot[-\ln(v)\cdot e^{-v}\big|_{e^{r_1}}^{\infty} + \int_{e^{r_1}}^{\infty}\frac{1}{v}\cdot e^{-v}\,\partial v]}{\left(\{1-\theta(2e^{-e^{r_2}}-1)\}\cdot e^{-e^{r_1}} + \theta(2e^{-e^{r_2}}-1)\cdot e^{-2e^{r_1}}\right)} +$$

$$\frac{\theta(2e^{-e^{r_2}}-1)\cdot[-\ln(v)\cdot e^{-2v}\big|_{e^{r_1}}^{\infty} + \int_{e^{r_1}}^{\infty}\frac{1}{v}\cdot e^{-2v}\,\partial v]}{\left(\{1-\theta(2e^{-e^{r_2}}-1)\}\cdot e^{-e^{r_1}} + \theta(2e^{-e^{r_2}}-1)\cdot e^{-2e^{r_1}}\right)}$$

$$= \frac{\{1-\theta(2e^{-e^{r_2}}-1)\}\cdot[r_1\cdot e^{-e^{r_1}} - Ei(-e^{r_1})] + \theta(2e^{-e^{r_2}}-1)\cdot[r_1\cdot e^{-2e^{r_1}} - Ei(-2e^{r_1})]}{\left(1 + \theta(e^{-e^{r_1}}-1)\cdot(2e^{-e^{r_2}}-1)\right)\cdot e^{-e^{r_1}}}$$

Thus

$$\boxed{\begin{aligned}&\mathrm{E}\left(\epsilon_1 \,\middle|\, \begin{matrix} Y_1 > c_1,\ Y_2 = y_2 \\ (e^{\epsilon_1}, e^{\epsilon_2}) \sim \exp(\mathrm{II}) \end{matrix}\right) \\ &= r_1 \quad + \quad \frac{\{1\text{-}\theta(2e^{-e^{r_2}}\text{-}1)\}\cdot[-\ \boldsymbol{Ei}(\text{-}e^{r_1})] \ + \ \theta(2e^{-e^{r_2}}\text{-}1)\cdot[-\ \boldsymbol{Ei}(\text{-}2e^{r_1})]}{\left(1 + \ \theta(e^{-e^{r_1}}\text{-}1)\cdot(2e^{-e^{r_2}}\text{-}1)\right) \ \cdot e^{-e^{r_1}}}\end{aligned}}$$

3) EXPECTED VALUE OF SQUARED RESIDUALS

A. Gumbel's exponential (I)

(i) $$E\left((\epsilon_1)^2 \middle| \begin{array}{l} Y_1 > c_1,\ Y_2 > c_2 \\ (e^{\epsilon_1}, e^{\epsilon_2}) \sim \exp(I) \end{array} \right)$$

$$= \frac{\int_{r_1}^{\infty}\int_{r_2}^{\infty} (\epsilon_1)^2 \cdot \{(1+\theta e^{\epsilon_1})(1+\theta e^{\epsilon_2})-\theta\} \cdot e^{-\theta e^{\epsilon_1+\epsilon_2} - e^{\epsilon_1} - e^{\epsilon_2} + \epsilon_1 + \epsilon_2}\, \partial\epsilon_1 \partial\epsilon_2}{\int_{r_1}^{\infty}\int_{r_2}^{\infty} \{(1+\theta e^{\epsilon_1})(1+\theta e^{\epsilon_2})-\theta\} \cdot e^{-\theta e^{\epsilon_1+\epsilon_2} - e^{\epsilon_1} - e^{\epsilon_2} + \epsilon_1 + \epsilon_2}\, \partial\epsilon_1 \partial\epsilon_2}$$

where $r_1 = c_1 - \mathbf{X}\boldsymbol{\beta}_1 \quad r_2 = c_2 - \mathbf{X}\boldsymbol{\beta}_2$

Let $$e^{\epsilon_1} = v \quad \Rightarrow \quad \epsilon_1 = \ln(v) \qquad \partial\epsilon_1 = \frac{\partial v}{v}$$
$$e^{\epsilon_2} = w \quad \Rightarrow \quad \epsilon_2 = \ln(w) \qquad \partial\epsilon_2 = \frac{\partial w}{w}$$

$$E\left((\epsilon_1)^2 \middle| \begin{array}{l} Y_1 > c_1,\ Y_2 > c_2 \\ (e^{\epsilon_1}, e^{\epsilon_2}) \sim \exp(I) \end{array} \right)$$

$$= \frac{\int_{e^{r_1}}^{\infty} (\ln(v))^2 \cdot e^{-v} \left(\{1+\theta v-\theta\} \cdot \int_{e^{r_2}}^{\infty} e^{-(1+\theta v)w} \partial w + \{\theta(1+\theta v)\} \cdot \int_{e^{r_2}}^{\infty} w \cdot e^{-(1+\theta v)w} \partial w \right) \partial v}{\int_{e^{r_1}}^{\infty} e^{-v} \left(\{1+\theta v-\theta\} \cdot \int_{e^{r_2}}^{\infty} e^{-(1+\theta v)w} \partial w + \{\theta(1+\theta v)\} \cdot \int_{e^{r_2}}^{\infty} w \cdot e^{-(1+\theta v)w} \partial w \right) \partial v}$$

$$= \frac{\int_{e^{r_1}}^{\infty} (\ln(v))^2 e^{-v} \left(\{1+\theta v-\theta\} \frac{e^{-(1+\theta v)e^{r_2}}}{(1+\theta v)} + \theta(1+\theta v) \left\{ \frac{e^{r_2}}{(1+\theta v)} + \frac{1}{(1+\theta v)^2} \right\} e^{-(1+\theta v)e^{r_2}} \right) \partial v}{\int_{e^{r_1}}^{\infty} e^{-v} \left(\{1+\theta v-\theta\} \cdot \frac{e^{-(1+\theta v)e^{r_2}}}{(1+\theta v)} + \theta(1+\theta v) \cdot \left\{ \frac{e^{r_2}}{(1+\theta v)} + \frac{1}{(1+\theta v)^2} \right\} \cdot e^{-(1+\theta v)e^{r_2}} \right) \partial v}$$

$$= \frac{\int_{e^{r_1}}^{\infty} (\ln(v))^2 \cdot e^{-v} \cdot (1+\theta e^{r_2}) \cdot e^{-(1+\theta v)e^{r_2}} \partial v}{\int_{e^{r_1}}^{\infty} e^{-v} \cdot (1+\theta e^{r_2}) \cdot e^{-(1+\theta v)e^{r_2}} \partial v}$$

$$= \frac{(1+\theta e^{r_2}) \cdot e^{-e^{r_2}} \cdot \int_{e^{r_1}}^{\infty} (\ln(v))^2 \cdot e^{-(1+\theta e^{r_2})v}\, \partial v}{e^{-e^{r_1} - e^{r_2} - \theta e^{r_1+r_2}}}$$

By C.3

$$= \frac{(1+\theta e^{r_2})\left\{\frac{1}{(1+\theta e^{r_2})}\cdot[\frac{\pi^2}{6}+\{\gamma+\ln(1+\theta e^{r_2})\}^2] + \frac{1}{(1+\theta e^{r_2})}\cdot(e^{-(1+\theta e^{r_2})\cdot e^{r_1}}-1)\cdot r_1^2\right\}}{e^{-e^{r_1}-\theta e^{r_1+r_2}}}$$

$$- \frac{(1+\theta e^{r_2})\cdot\left\{\frac{1}{(1+\theta e^{r_2})}\cdot[-\mathrm{E}i(-\{1+\theta e^{r_2}\}\cdot e^{r_1})+\gamma+r_1+\ln(1+\theta e^{r_2})]\cdot 2r_1\right\}}{e^{-e^{r_1}-\theta e^{r_1+r_2}}}$$

$$+ \frac{(1+\theta e^{r_2})\cdot\left\{\frac{1}{(1+\theta e^{r_2})}\cdot\sum_{i=0}^{\infty}\frac{(-(1+\theta e^{r_2})\cdot e^{r_1})^{i+1}}{(i+1)!\cdot(i+1)^2}\cdot 2\right\}}{e^{-e^{r_1}-\theta e^{r_1+r_2}}}$$

Thus

$$\boxed{\begin{aligned}
&E\left((\epsilon_1)^2 \,\middle|\, \begin{matrix} Y_1 > c_1,\ Y_2 > c_2 \\ (e^{\epsilon_1}, e^{\epsilon_2}) \sim \exp(I)\end{matrix}\right) \\
&= r_1^2 + \frac{\left\{r_1^2 + [\frac{\pi^2}{6}+\{\gamma+\ln(1+\theta e^{r_2})\}^2]\right\}}{e^{-e^{r_1}-\theta e^{r_1+r_2}}} \\
&+ \frac{\left\{[-\mathrm{E}i(-\{1+\theta e^{r_2}\}e^{r_1})+\gamma+\ln(1+\theta e^{r_2})]\cdot 2r_1\right\} + \left\{\sum_{i=0}^{\infty}\frac{(-(1+\theta e^{r_2})\cdot e^{r_1})^{i+1}}{(i+1)!\cdot(i+1)^2}\cdot 2\right\}}{e^{-e^{r_1}-\theta e^{r_1+r_2}}}
\end{aligned}}$$

(ii) $$E\left((\epsilon_1)^2 \,\middle|\, \begin{matrix} Y_1 > c_1,\ Y_2 = y_2 \\ (e^{\epsilon_1}, e^{\epsilon_2}) \sim \exp(I) \end{matrix}\right)$$

$$= \frac{\int_{r_1}^{\infty} (\epsilon_1)^2 \cdot \{(1+\theta e^{\epsilon_1})(1+\theta e^{r_2})-\theta\} \cdot e^{-\theta e^{\epsilon_1 + r_2} - e^{\epsilon_1} - e^{r_2} + \epsilon_1 + r_2}\, \partial\epsilon_1}{\int_{r_1}^{\infty} \{(1+\theta e^{\epsilon_1})(1+\theta e^{r_2})-\theta\} \cdot e^{-\theta e^{\epsilon_1 + r_2} - e^{\epsilon_1} - e^{r_2} + \epsilon_1 + r_2}\, \partial\epsilon_1}$$

where $r_1 = c_1\text{-}\mathbf{X}\boldsymbol{\beta}_1 \quad r_2 = y_2\text{-}\mathbf{X}\boldsymbol{\beta}_2$

Let $e^{\epsilon_1} = v \quad \Rightarrow \quad \epsilon_1 = \ln(v) \qquad \partial\epsilon_1 = \frac{\partial v}{v}$

$$E\left((\epsilon_1)^2 \,\middle|\, \begin{matrix} Y_1 > c_1,\ Y_2 = y_2 \\ (e^{\epsilon_1}, e^{\epsilon_2}) \sim \exp(I) \end{matrix}\right)$$

$$= \frac{\int_{e^{r_1}}^{\infty} (\ln(v))^2 \cdot \{(1+\theta v)(1+\theta e^{r_2})-\theta\} \cdot e^{-\theta v e^{r_2} - v - e^{r_2} + r_2}\, \partial v}{\int_{e^{r_1}}^{\infty} \{(1+\theta v)(1+\theta e^{r_2})-\theta\} \cdot e^{-\theta v e^{r_2} - v - e^{r_2} + r_2}\, \partial v}$$

$$= \frac{\{1+\theta e^{r_2}\text{-}\theta\} \cdot \int_{e^{r_1}}^{\infty} (\ln(v))^2 \cdot e^{-(1+\theta e^{r_2})v} \partial v + \{\theta(1+\theta e^{r_2})\} \int_{e^{r_1}}^{\infty} (\ln(v))^2 \cdot v \cdot e^{-(1+\theta e^{r_2})v} \partial v}{\left\{(1+\theta e^{r_2}\text{-}\theta) \cdot \int_{e^{r_1}}^{\infty} e^{-(1+\theta e^{r_2})v}\, \partial v\right\} + \left\{\theta(1+\theta e^{r_2}) \cdot \int_{e^{r_1}}^{\infty} v \cdot e^{-(1+\theta e^{r_2})v}\, \partial v\right\}}$$

$$= \frac{\{1+\theta e^{r_2}\text{-}\theta\} \int_{e^{r_1}}^{\infty} (\ln(v))^2 e^{-(1+\theta e^{r_2})v} \partial v - \theta(1+\theta e^{r_2}) \left(\frac{(\ln(v))^2}{(1+\theta e^{r_2})} \cdot e^{-(1+\theta e^{r_2})v} \left(v + \frac{1}{1+\theta e^{r_2}}\right)\right)\Bigg|}{(1+\theta e^{r_1}) \cdot e^{-e^{r_1} \cdot (1+\theta e^{r_2})}}$$

$$+ \frac{\{\theta(1+\theta e^{r_2})\} \cdot \frac{2}{(1+\theta e^{r_2})} \cdot \left(\int_{e^{r_1}}^{\infty} \ln(v) \cdot e^{-(1+\theta e^{r_2})v}\, \partial v + \frac{1}{1+\theta e^{r_2}} \cdot \int_{e^{r_1}}^{\infty} \frac{\ln(v)}{v} \cdot e^{-(1+\theta e^{r_2})v} \partial v \right)}{(1+\theta e^{r_1}) \cdot e^{-e^{r_1} \cdot (1+\theta e^{r_2})}}$$

$$= \frac{\{1+\theta e^{r_2}-\theta\}\cdot\int_{e^{r_1}}^{\infty}(\ln(v))^2\cdot e^{-(1+\theta e^{r_2})v}\partial v}{(1+\theta e^{r_1})\cdot e^{-e^{r_1}\cdot(1+\theta e^{r_2})}}$$

$$+ \frac{\{\theta(1+\theta e^{r_2})\}\cdot\left(-\frac{(ln(v))^2}{(1+\theta e^{r_2})}\cdot e^{-(1+\theta e^{r_2})v}\cdot\left(v+\frac{1}{1+\theta e^{r_2}}\right)\right)\Bigg|_{e^{r_1}}^{\infty}}{(1+\theta e^{r_1})\cdot e^{-e^{r_1}\cdot(1+\theta e^{r_2})}}$$

$$+ \frac{2\theta\cdot\left(\frac{-\ln(v)\cdot e^{-(1+\theta e^{r_2})v}}{(1+\theta e^{r_2})}\Bigg|_{e^{r_1}}^{\infty} + \frac{1}{(1+\theta e^{r_2})}\cdot\overbrace{\int_{e^{r_1}}^{\infty}\frac{e^{-(1+\theta e^{r_2})v}}{v}\partial v}^{-\mathbf{E}i(-(1+\theta e^{r_2})\cdot e^{r_1})}\right)}{(1+\theta e^{r_1})\cdot e^{-e^{r_1}\cdot(1+\theta e^{r_2})}}$$

$$+ \frac{\frac{2\theta}{(1+\theta e^{r_2})}\left(\frac{(ln(v))^2}{2}\cdot e^{-(1+\theta e^{r_2})v}\Bigg|_{e^{r_1}}^{\infty} + \int_{e^{r_1}}^{\infty}(\ln(v))^2\cdot\frac{(1+\theta e^{r_2})}{2}\cdot e^{-(1+\theta e^{r_2})v}\partial v\right)}{(1+\theta e^{r_1})\cdot e^{-e^{r_1}\cdot(1+\theta e^{r_2})}}$$

$$= \frac{\{1+\theta e^{r_2}\}\left(\frac{1}{(1+\theta e^{r_2})}\cdot[\frac{\pi^2}{6}+\{\gamma+\ln(1+\theta e^{r_2})\}^2] + (e^{-e^{r_1}\cdot(1+\theta e^{r_2})}-1)\cdot r_1{}^2\cdot\frac{1}{(1+\theta e^{r_2})}\right)}{(1+\theta e^{r_1})\cdot e^{-e^{r_1}\cdot(1+\theta e^{r_2})}}$$

$$+ \frac{\{1+\theta e^{r_2}\}\cdot\left([-\mathbf{E}i(-e^{r_1}\cdot(1+\theta e^{r_2})) + \gamma + r_1 + \ln(1+\theta e^{r_2})]\cdot 2r_1\cdot\frac{1}{(1+\theta e^{r_2})}\right)}{(1+\theta e^{r_1})\cdot e^{-e^{r_1}\cdot(1+\theta e^{r_2})}}$$

$$+ \frac{\{1+\theta e^{r_2}\}\cdot\left(\sum_{i=0}^{\infty}\frac{(-e^{r_1}\cdot(1+\theta e^{r_2}))^{i+1}}{(i+1)!\cdot(i+1)^2}\cdot 2\cdot\frac{1}{(1+\theta e^{r_2})}\right) + \theta\cdot r_1{}^2\cdot e^{-e^{r_1}\cdot(1+\theta e^{r_2})}\cdot e^{r_1}}{(1+\theta e^{r_1})\cdot e^{-e^{r_1}\cdot(1+\theta e^{r_2})}}$$

$$+ \frac{2\theta\cdot\left(\frac{r_1\cdot e^{-e^{r_1}\cdot(1+\theta e^{r_2})}}{(1+\theta e^{r_2})} + \frac{-\mathbf{E}i(-(1+\theta e^{r_2})\cdot e^{r_1})}{(1+\theta e^{r_2})}\right)}{(1+\theta e^{r_1})\cdot e^{-e^{r_1}\cdot(1+\theta e^{r_2})}}$$

Thus

$$\mathbf{E}\left((\epsilon_1)^2 \,\middle|\, \begin{array}{l} Y_1 > c_1,\ Y_2 = y_2 \\ (e^{\epsilon_1}, e^{\epsilon_2}) \sim \exp(I) \end{array}\right)$$

$$= r_1^2 + \frac{r_1^2 + [\frac{\pi^2}{6} + \{\gamma + \ln(1+\theta e^{r_2})\}^2] + \left([-\mathbf{E}i(-e^{r_1}\cdot(1+\theta e^{r_2})) + \gamma + \ln(1+\theta e^{r_2})\,]\cdot 2r_1\right)}{(1+\theta e^{r_1})\cdot e^{-e^{r_1}\cdot(1+\theta e^{r_2})}}$$

$$+ \frac{\left(\sum_{i=0}^{\infty} \frac{2\cdot(-e^{r_1}\cdot(1+\theta e^{r_2}))^{i+1}}{(i+1)!\cdot(i+1)^2}\right) + \frac{2\theta}{(1+\theta e^{r_2})}\left(r_1\cdot e^{-e^{r_1}\cdot(1+\theta e^{r_2})} - \mathbf{E}i(-(1+\theta e^{r_2})\cdot e^{r_1})\right)}{(1+\theta e^{r_1})\cdot e^{-e^{r_1}\cdot(1+\theta e^{r_2})}}$$

B. Gumbel's exponential (II)

(i) $\mathbb{E}\left((\epsilon_1)^2 \,\middle|\, \begin{matrix} Y_1 > c_1,\ Y_2 > c_2 \\ (e^{\epsilon_1}, e^{\epsilon_2}) \sim \exp(II) \end{matrix} \right) =$

$$\frac{\int_{r_1}^{\infty}\int_{r_2}^{\infty} (\epsilon_1)^2 \cdot \{1+\theta(2e^{-e^{\epsilon_1}}-1)(2e^{-e^{\epsilon_2}}-1)\} \cdot e^{-e^{\epsilon_1}-e^{\epsilon_2}+\epsilon_1+\epsilon_2}\ \partial\epsilon_1\partial\epsilon_2}{\int_{r_1}^{\infty}\int_{r_2}^{\infty} \{1+\theta(2e^{-e^{\epsilon_1}}-1)(2e^{-e^{\epsilon_2}}-1)\} \cdot e^{-e^{\epsilon_1}-e^{\epsilon_2}+\epsilon_1+\epsilon_2}\ \partial\epsilon_1\partial\epsilon_2}$$

where $r_1 = c_1\text{-}X\beta_1 \quad r_2 = c_2\text{-}X\beta_2$

Let $e^{\epsilon_1} = v \quad \Rightarrow \quad \epsilon_1 = \ln(v) \qquad \partial\epsilon_1 = \frac{\partial v}{v}$

$e^{\epsilon_2} = w \quad \Rightarrow \quad \epsilon_2 = \ln(w) \qquad \partial\epsilon_2 = \frac{\partial w}{w}$

$\mathbb{E}\left((\epsilon_1)^2 \,\middle|\, \begin{matrix} Y_1 > c_1,\ Y_2 > c_2 \\ (e^{\epsilon_1}, e^{\epsilon_2}) \sim \exp(II) \end{matrix} \right)$

$$= \frac{\int_{e^{r_1}}^{\infty} (\ln(v))^2 \cdot e^{-v} \cdot \left(\{1\text{-}\theta(2e^{-v}\text{-}1)\} \cdot \int_{e^{r_2}}^{\infty} e^{-w}\partial w \ + 2\theta(2e^{-v}\text{-}1)\} \cdot \int_{e^{r_2}}^{\infty} e^{-2w}\partial w \right) \partial v}{\int_{e^{r_1}}^{\infty} e^{-v} \cdot \left(\{1\text{-}\theta(2e^{-v}\text{-}1)\} \cdot \int_{e^{r_2}}^{\infty} e^{-w}\partial w \ + 2\theta(2e^{-v}\text{-}1)\} \cdot \int_{e^{r_2}}^{\infty} e^{-2w}\partial w \right) \partial v}$$

$$= \frac{\int_{e^{r_1}}^{\infty} (\ln(v))^2 \cdot e^{-v} \cdot \left(\{1\text{-}\theta(2e^{-v}\text{-}1)\} \cdot e^{-e^{r_2}} \ + \theta(2e^{-v}\text{-}1)\} \cdot e^{-2e^{r_2}} \right) \partial v}{\int_{e^{r_1}}^{\infty} e^{-v} \cdot \left(\{1\text{-}\theta(2e^{-v}\text{-}1)\} \cdot e^{-e^{r_2}} \ + \theta(2e^{-v}\text{-}1)\} \cdot e^{-2e^{r_2}} \right) \partial v}$$

$$= \frac{e^{-e^{r_2}} \cdot \left(\{1\text{-}\theta(e^{-e^{r_2}}\text{-}1)\} \cdot \int_{e^{r_1}}^{\infty} (\ln(v))^2 \cdot e^{-v}\ \partial v \ + \ 2\theta(e^{-e^{r_2}}\text{-}1)\} \cdot \int_{e^{r_1}}^{\infty} (\ln(v))^2 \cdot e^{-2v}\ \partial v \right)}{e^{-e^{r_2}} \cdot \left(\{1\text{-}\theta(e^{-e^{r_2}}\text{-}1)\} \cdot \int_{e^{r_1}}^{\infty} e^{-v}\ \partial v \ + \ 2\theta(e^{-e^{r_2}}\text{-}1)\} \cdot \int_{e^{r_1}}^{\infty} e^{-2v}\ \partial v \right)}$$

By C.3

$$= \frac{\{1-\theta(e^{-e^{r_2}}-1)\}\left([\frac{\pi^2}{6}+\gamma^2]+(e^{-e^{r_1}}-1)r_1^2+[-\mathrm{E}i(-e^{r_1})+\gamma+r_1]2r_1+\sum_{i=0}^{\infty}\frac{2\cdot(-e^{r_1})^{i+1}}{(i+1)!\cdot(i+1)^2}\right)}{\left(\{1-\theta(e^{-e^{r_2}}-1)\}\cdot e^{-e^{r_1}}+\theta(e^{-e^{r_2}}-1)\}\cdot e^{-2e^{r_1}}\right)}$$

$$+\frac{2\theta(e^{-e^{r_2}}-1)\}\cdot\left(\frac{1}{2}\cdot[\frac{\pi^2}{6}+\{\gamma+\ln(2)\}^2]+(e^{-2e^{r_1}}-1)\cdot r_1^2\cdot\frac{1}{2}\right)}{\left(\{1-\theta(e^{-e^{r_2}}-1)\}\cdot e^{-e^{r_1}}+\theta(e^{-e^{r_2}}-1)\}\cdot e^{-2e^{r_1}}\right)}$$

$$+\frac{2\theta(e^{-e^{r_2}}-1)\}\cdot\left([-\mathrm{E}i(-2e^{r_1})+\gamma+r_1+\ln(2)]\cdot 2r_1\cdot\frac{1}{2}+\sum_{i=0}^{\infty}\frac{(-2e^{r_1})^{i+1}}{(i+1)!\cdot(i+1)^2}\cdot 2\cdot\frac{1}{2}\right)}{\left(\{1-\theta(e^{-e^{r_2}}-1)\}\cdot e^{-e^{r_1}}+\theta(e^{-e^{r_2}}-1)\}\cdot e^{-2e^{r_1}}\right)}$$

$$= \frac{\left(\{1-\theta(e^{-e^{r_2}}-1)\}\cdot(e^{-e^{r_1}}-1)+\theta(e^{-e^{r_2}}-1)\cdot(e^{-2e^{r_1}}-1)+2\right)\cdot r_1^2}{\{1+\theta(e^{-e^{r_2}}-1)\cdot(e^{-e^{r_1}}-1)\}\cdot e^{-e^{r_1}}}$$

$$+\frac{[\frac{\pi^2}{6}+\gamma^2]+\theta(e^{-e^{r_2}}-1)\cdot\ln(2)\cdot\{2\gamma+\ln(2)\}}{\{1+\theta(e^{-e^{r_2}}-1)\cdot(e^{-e^{r_1}}-1)\}\cdot e^{-e^{r_1}}}$$

$$+\frac{\sum_{i=0}^{\infty}\frac{(-e^{r_1})^{i+1}}{(i+1)!\cdot(i+1)^2}\cdot\left(2\cdot\{1-\theta(e^{-e^{r_2}}-1)\}-2^{(i+2)}\theta(e^{-e^{r_2}}-1)\}\right)}{\{1+\theta(e^{-e^{r_2}}-1)\cdot(e^{-e^{r_1}}-1)\}\cdot e^{-e^{r_1}}}$$

$$+\frac{\left(\{1-\theta(e^{-e^{r_2}}-1)\}\cdot[-\mathrm{E}i(-e^{r_1})+\gamma]+\theta\cdot(e^{-e^{r_2}}-1)\}\cdot[-\mathrm{E}i(-2e^{r_1})+\gamma+\ln(2)]\right)\cdot 2r_1}{\{1+\theta(e^{-e^{r_2}}-1)\cdot(e^{-e^{r_1}}-1)\}\cdot e^{-e^{r_1}}}$$

Thus

$$\mathbb{E}\left((\epsilon_1)^2 \,\middle|\, \begin{array}{l} Y_1 > c_1,\ Y_2 > c_2 \\ (e^{\epsilon_1}, e^{\epsilon_2}) \sim \exp(\mathrm{II}) \end{array} \right)$$

$$= \quad r_1{}^2 + \frac{r_1{}^2 + [\frac{\pi^2}{6} + \gamma^2] \ + \ \theta(e^{-e^{r_2}}-1) \cdot \ln(2) \cdot \{2\gamma + \ln(2)\}}{\{1 + \theta(e^{-e^{r_2}}-1) \cdot (e^{-e^{r_1}}-1)\} \cdot e^{-e^{r_1}}}$$

$$+ \ \frac{\sum_{i=0}^{\infty} \frac{(-e^{r_1})^{i+1}}{(i+1)! \cdot (i+1)^2} \cdot \left(2 \cdot \{1\text{-}\theta(e^{-e^{r_2}}-1)\} \ - 2^{(i+2)}\theta(e^{-e^{r_2}}-1)\}\right)}{\{1 + \theta(e^{-e^{r_2}}-1) \cdot (e^{-e^{r_1}}-1)\} \cdot e^{-e^{r_1}}}$$

$$+ \ \frac{\left(\gamma + \{1\text{-}\theta(e^{-e^{r_2}}-1)\} \cdot \text{-}\mathrm{E}i(\text{-}e^{r_1}) \ + \theta(e^{-e^{r_2}}-1) \cdot [\text{-}\mathrm{E}i(\text{-}2e^{r_1}) + \ \ln(2)\,]\right) \cdot 2r_1}{\{1 + \ \theta(e^{-e^{r_1}}-1) \cdot (e^{-e^{r_2}}\ -1)\} \cdot e^{-e^{r_1}}}$$

(ii) $E\left((\epsilon_1)^2 \,\middle|\, \begin{matrix} Y_1 > c_1,\ Y_2 = y_2 \\ (e^{\epsilon_1}, e^{\epsilon_2}) \sim \exp(II) \end{matrix}\right) =$

$$\frac{e^{-e^{r_2}+r_2} \int_{r_1}^{\infty} (\epsilon_1)^2 \cdot \{1+\theta(2e^{-e^{\epsilon_1}}-1)(2e^{-e^{r_2}}-1)\} \cdot e^{-e^{\epsilon_1}+\epsilon_1}\, \partial\epsilon_1}{e^{-e^{r_2}+r_2} \int_{r_1}^{\infty} \{1+\theta(2e^{-e^{\epsilon_1}}-1)(2e^{-e^{r_2}}-1)\} \cdot e^{-e^{\epsilon_1}+\epsilon_1}\, \partial\epsilon_1}$$

where $r_1 = c_1 - X\beta_1 \quad r_2 = y_2 - X\beta_2$

Let $\quad e^{\epsilon_1} = v \quad \Rightarrow \quad \epsilon_1 = \ln(v) \qquad \partial\epsilon_1 = \frac{\partial v}{v}$

$$E\left((\epsilon_1)^2 \,\middle|\, \begin{matrix} Y_1 > c_1,\ Y_2 = y_2 \\ (e^{\epsilon_1}, e^{\epsilon_2}) \sim \exp(II) \end{matrix}\right) = \frac{\int_{e^{r_1}}^{\infty} (\ln(v))^2 \cdot \{1+\theta(2e^{-v}-1)(2e^{-e^{r_2}}-1)\} \cdot e^{-v}\, \partial v}{\int_{e^{r_1}}^{\infty} \{1+\theta(2e^{-v}-1)(2e^{-e^{r_2}}-1)\} \cdot e^{-v}\, \partial v}$$

$$= \frac{\{1-\theta(2e^{-e^{r_2}}-1)\} \cdot \int_{e^{r_1}}^{\infty} (\ln(v))^2 \cdot e^{-v}\, \partial v \; + \; 2\theta(2e^{-e^{r_2}}-1) \cdot \int_{e^{r_1}}^{\infty} (\ln(v))^2 \cdot e^{-2v}\, \partial v}{\left\{1 + \theta(e^{-e^{r_1}}-1) \cdot (2e^{-e^{r_2}}-1)\right\} \cdot e^{-e^{r_1}}}$$

By C.3

$$= \frac{\{1-\theta(2e^{-e^{r_2}}-1)\} \cdot \left([\frac{\pi^2}{6} + \gamma^2] \; + \; (e^{-e^{r_1}}-1) \cdot r_1^2\right)}{\left\{1 + \theta(e^{-e^{r_1}}-1) \cdot (2e^{-e^{r_2}}-1)\right\} \cdot e^{-e^{r_1}}}$$

$$+ \frac{\{1-\theta(2e^{-e^{r_2}}-1)\} \cdot \left([-Ei(-e^{r_1}) + \gamma + r_1] \cdot 2r_1 \; + \; \sum_{i=0}^{\infty} \frac{(-e^{r_1})^{(i+1)}}{(i+1)! \cdot (i+1)^2} \cdot 2\right)}{\left\{1 + \theta(e^{-e^{r_1}}-1) \cdot (2e^{-e^{r_2}}-1)\right\} \cdot e^{-e^{r_1}}}$$

$$+ \frac{\theta(2e^{-e^{r_2}}-1) \cdot \left([\frac{\pi^2}{6} + \{\gamma + \ln(2)\}^2] \; + (e^{-2e^{r_1}}-1) \cdot r_1^2\right)}{\left\{1 + \theta(e^{-e^{r_1}}-1) \cdot (2e^{-e^{r_2}}-1)\right\} \cdot e^{-e^{r_1}}}$$

$$+ \frac{\theta(2e^{-e^{r_2}}-1) \cdot \left([-Ei(-2e^{r_1}) + \gamma + r_1 + \ln(2)\,] \cdot 2r_1 + \sum_{i=0}^{\infty} \frac{(-2e^{r_1})^{i+1}}{(i+1)! \cdot (i+1)^2} \cdot 2\right)}{\left\{1 + \theta(e^{-e^{r_1}}-1) \cdot (2e^{-e^{r_2}}-1)\right\} \cdot e^{-e^{r_1}}}$$

$$= \frac{[\frac{\pi^2}{6} + \gamma^2] + \theta(2e^{-e^{r_2}}-1)\cdot \ln(2) \cdot \{2\gamma + \ln(2)\}}{\left\{1 + \theta(e^{-e^{r_1}}-1)\cdot(2e^{-e^{r_2}}-1)\right\}\cdot e^{-e^{r_1}}}$$

$$+ \frac{\left\{(e^{-e^{r_1}}-1)\cdot\left(\{1-\theta(2e^{-e^{r_2}}-1)\} + \theta(2e^{-e^{r_2}}-1)\cdot(e^{-e^{r_1}}+1)\right) + 2\right\} r_1^2}{\left\{1 + \theta(e^{-e^{r_1}}-1)\cdot(2e^{-e^{r_2}}-1)\right\}\cdot e^{-e^{r_1}}}$$

$$+ \frac{\sum_{i=0}^{\infty} \frac{(-e^{r_1})^{(i+1)}}{(i+1)!\cdot(i+1)^2}\cdot\left(2\{1-\theta(2e^{-e^{r_2}}-1)\} + 2\theta(2e^{-e^{r_2}}-1)\cdot(2)^{i+1}\right)}{\left\{1 + \theta(e^{-e^{r_1}}-1)\cdot(2e^{-e^{r_2}}-1)\right\}\cdot e^{-e^{r_1}}}$$

$$+ \frac{\left(\{1-\theta(2e^{-e^{r_2}}-1)\}\cdot[-\mathbf{E}i(-e^{r_1}) + \gamma] + \theta(2e^{-e^{r_2}}-1)\cdot[-\mathbf{E}i(-2e^{r_1}) + \gamma + \ln(2)]\right)\cdot 2r_1}{\left\{1 + \theta(e^{-e^{r_1}}-1)\cdot(2e^{-e^{r_2}}-1)\right\}\cdot e^{-e^{r_1}}}$$

Thus

$$\boxed{\begin{aligned}
&\mathbf{E}\left((\epsilon_1)^2 \,\middle|\, \begin{matrix} Y_1 > c_1,\ Y_2 = y_2 \\ (e^{\epsilon_1}, e^{\epsilon_2}) \sim \exp(\mathrm{II}) \end{matrix}\right) \\
&= r_1^2 + \frac{r_1^2 + [\frac{\pi^2}{6} + \gamma^2] + \theta(2e^{-e^{r_2}}-1)\cdot \ln(2)\cdot\{2\gamma + \ln(2)\}}{\left\{1 + \theta(e^{-e^{r_1}}-1)\cdot(2e^{-e^{r_2}}-1)\right\}\cdot e^{-e^{r_1}}} \\
&+ \frac{\sum_{i=0}^{\infty} \frac{(-e^{r_1})^{(i+1)}}{(i+1)!\cdot(i+1)^2}\cdot\left(2\{1-\theta(2e^{-e^{r_2}}-1)\} - 2^{(i+2)}\theta(2e^{-e^{r_2}}-1)\right)}{\left\{1 + \theta(e^{-e^{r_1}}-1)\cdot(2e^{-e^{r_2}}-1)\right\}\cdot e^{-e^{r_1}}} \\
&+ \frac{\left(\gamma + \{1-\theta(2e^{-e^{r_2}}-1)\}\cdot -\mathbf{E}i(-e^{r_1}) + \theta(2e^{-e^{r_2}}-1)\cdot[-\mathbf{E}i(-2e^{r_1}) + \ln(2)]\right)\cdot 2r_1}{\left\{1 + \theta(e^{-e^{r_1}}-1)\cdot(2e^{-e^{r_2}}-1)\right\}\cdot e^{-e^{r_1}}}
\end{aligned}}$$

4) EXPECTED VALUE OF PRODUCT OF RESIDUALS

A. Gumbel's exponential (I)

$$\mathbf{E}\left(\epsilon_1\cdot\epsilon_2 \;\middle|\; \begin{matrix} Y_1 > c_1,\, Y_2 > c_2 \\ (e^{\epsilon_1},\, e^{\epsilon_2}) \sim \exp(I) \end{matrix}\right)$$

$$= \frac{\int_{r_1}^{\infty}\int_{r_2}^{\infty} \epsilon_1\cdot\epsilon_2\cdot\{(1+\theta e^{\epsilon_1})(1+\theta e^{\epsilon_2})-\theta\}\cdot e^{-\theta e^{\epsilon_1+\epsilon_2} - e^{\epsilon_1} - e^{\epsilon_2} + \epsilon_1 + \epsilon_2}\,\partial\epsilon_1\partial\epsilon_2}{\int_{r_1}^{\infty}\int_{r_2}^{\infty} \{(1+\theta e^{\epsilon_1})(1+\theta e^{\epsilon_2})-\theta\}\cdot e^{-\theta e^{\epsilon_1+\epsilon_2} - e^{\epsilon_1} - e^{\epsilon_2} + \epsilon_1 + \epsilon_2}\,\partial\epsilon_1\partial\epsilon_2}$$

where $r_1 = c_1\text{-}\mathbf{X}\boldsymbol{\beta}_1 \quad r_2 = c_2\text{-}\mathbf{X}\boldsymbol{\beta}_2$

Let $\quad e^{\epsilon_1} = v \quad \Rightarrow \quad \epsilon_1 = \ln(v) \qquad \partial\epsilon_1 = \frac{\partial v}{v}$

$\qquad\; e^{\epsilon_2} = w \quad \Rightarrow \quad \epsilon_2 = \ln(w) \qquad \partial\epsilon_2 = \frac{\partial w}{w}$

$$\mathbf{E}\left(\epsilon_1\cdot\epsilon_2 \;\middle|\; \begin{matrix} Y_1 > c_1,\, Y_2 > c_2 \\ (e^{\epsilon_1},\, e^{\epsilon_2}) \sim \exp(I) \end{matrix}\right)$$

$$= \frac{\int_{e^{r_1}}^{\infty} \ln(v)\cdot e^{-v}\left(\{1+\theta v-\theta\}\cdot\int_{e^{r_2}}^{\infty}\ln(w)\cdot e^{-(1+\theta v)w}\partial w\right)\partial v}{\int_{e^{r_1}}^{\infty} e^{-v}\left(\{1+\theta v-\theta\}\cdot\int_{e^{r_2}}^{\infty} e^{-(1+\theta v)w}\partial w + \{\theta(1+\theta v)\}\cdot\int_{e^{r_2}}^{\infty} w\cdot e^{-(1+\theta v)w}\partial w\right)\partial v} +$$

$$\frac{\int_{e^{r_1}}^{\infty} \ln(v)\cdot e^{-v}\left(\{\theta(1+\theta v)\}\cdot\int_{e^{r_2}}^{\infty}\ln(w)\cdot w\cdot e^{-(1+\theta v)w}\partial w\right)\partial v}{\int_{e^{r_1}}^{\infty} e^{-v}\left(\{1+\theta v-\theta\}\cdot\int_{e^{r_2}}^{\infty} e^{-(1+\theta v)w}\partial w + \{\theta(1+\theta v)\}\cdot\int_{e^{r_2}}^{\infty} w\cdot e^{-(1+\theta v)w}\partial w\right)\partial v}$$

$$= \frac{\int_{e^{r_1}}^{\infty} \ln(v)\cdot e^{-v}\cdot\frac{1+\theta v-\theta}{1+\theta v}\cdot\left(-\ln(w)\cdot e^{-(1+\theta v)w}\Big|_{e^{r_2}}^{\infty} + \int_{e^{r_2}}^{\infty}\frac{e^{-(1+\theta v)w}}{w}\,\partial w\right)\partial v}{\int_{e^{r_1}}^{\infty} e^{-v}\left(\frac{1+\theta v-\theta}{1+\theta v}\cdot e^{-(1+\theta v)e^{r_2}} + \theta\cdot\left(e^{r_2}+\frac{1}{1+\theta v}\right)\cdot e^{-(1+\theta v)e^{r_2}}\right)\partial v} +$$

$$\frac{\int_{e^{r_1}}^{\infty} \ln(v)e^{-v}\cdot\frac{\theta(1+\theta v)}{1+\theta v}\left(-\ln(w)\left\{w+\frac{1}{1+\theta v}\right\}e^{-(1+\theta v)w}\Big|_{e^{r_2}}^{\infty} + \int_{e^{r_2}}^{\infty}\left\{w+\frac{1}{1+\theta v}\right\}\frac{e^{-(1+\theta v)w}}{w}\partial w\right)\partial v}{\int_{e^{r_1}}^{\infty} e^{-v}\left(\frac{1+\theta v-\theta}{1+\theta v}\cdot e^{-(1+\theta v)e^{r_2}} + \theta\cdot\left(e^{r_2}+\frac{1}{1+\theta v}\right)\cdot e^{-(1+\theta v)e^{r_2}}\right)\partial v}$$

$$= \frac{\int_{e^{r_1}}^{\infty} \ln(v)\cdot e^{-v}\cdot\left(r_2\cdot e^{-(1+\theta v)\cdot e^{r_2}} + \int_{e^{r_2}}^{\infty}\frac{e^{-(1+\theta v)w}}{w}\,\partial w\right)\partial v}{\int_{e^{r_1}}^{\infty} e^{-v}\left((1+\theta e^{r_2})\cdot e^{-(1+\theta v)e^{r_2}}\right)\partial v}$$

$$+ \frac{\int_{e^{r_1}}^{\infty} \ln(v)\cdot e^{-v}\cdot\theta\cdot\left(e^{r_2}\cdot r_2\cdot e^{-(1+\theta v)\cdot e^{r_2}} + \int_{e^{r_2}}^{\infty} e^{-(1+\theta v)w}\partial w\right)\partial v}{\int_{e^{r_1}}^{\infty} e^{-v}\left((1+\theta e^{r_2})\cdot e^{-(1+\theta v)e^{r_2}}\right)\partial v}$$

$$= \frac{(1+\theta e^{r_2})\cdot r_2\cdot e^{-e^{r_2}}\int_{e^{r_1}}^{\infty}\ln(v)\cdot e^{-v\cdot(1+\theta e^{r_2})}\,\partial v + \int_{e^{r_1}}^{\infty}\int_{e^{r_2}}^{\infty}\ln(v)\cdot\frac{e^{-v-w-\theta vw}}{w}\,\partial v\,\partial w}{e^{-e^{r_1}-e^{r_2}-\theta e^{r_1+r_2}}}$$

$$+ \frac{\theta\cdot\int_{e^{r_1}}^{\infty}\int_{e^{r_2}}^{\infty}\ln(v)\cdot e^{-v-w-\theta vw}\,\partial v\,\partial w}{e^{-e^{r_1}-e^{r_2}-\theta e^{r_1+r_2}}}$$

$$= \frac{(1+\theta e^{r_2})\cdot r_2\cdot e^{-e^{r_2}}\left(\frac{-\ln(v)}{1+\theta e^{r_2}}\cdot e^{-v\cdot(1+\theta e^{r_2})}\Big|_{e^{r_1}}^{\infty} + \int_{e^{r_1}}^{\infty}\frac{e^{-v\cdot(1+\theta e^{r_2})}}{(1+\theta e^{r_2})\cdot v}\,\partial v\right)}{e^{-e^{r_1}-e^{r_2}-\theta e^{r_1+r_2}}}$$

$$+ \frac{\int_{e^{r_2}}^{\infty}\left(\frac{1+\theta w}{w}\right)\cdot e^{-w}\cdot\left(\int_{e^{r_1}}^{\infty}\ln(v)\cdot e^{-v-\theta vw}\,\partial v\right)\partial w}{e^{-e^{r_1}-e^{r_2}-\theta e^{r_1+r_2}}}$$

$$= \frac{r_1\cdot r_2\cdot e^{-e^{r_1}-e^{r_2}-\theta e^{r_1+r_2}} - r_2\cdot e^{-e^{r_2}}\cdot Ei(-e^{r_1}\cdot\{1+\theta e^{r_2}\})}{e^{-e^{r_1}-e^{r_2}-\theta e^{r_1+r_2}}}$$

$$+ \frac{\int_{e^{r_2}}^{\infty}\left(\frac{1+\theta w}{w}\right)\cdot e^{-w}\cdot\left(\frac{-\ln(v)\cdot e^{-v\cdot(1+\theta w)}}{1+\theta w}\Big|_{e^{r_1}}^{\infty} + \int_{e^{r_1}}^{\infty}\frac{e^{-v\cdot(1+\theta w)}}{v\cdot(1+\theta w)}\,\partial v\right)\partial w}{e^{-e^{r_1}-e^{r_2}-\theta e^{r_1+r_2}}}$$

$$= \frac{r_1 \cdot r_2 \cdot e^{-e^{r_1} - e^{r_2} - \theta e^{r_1 + r_2}} - r_2 \cdot e^{-e^{r_2}} \cdot Ei(-e^{r_1} \cdot \{1+\theta e^{r_2}\})}{e^{-e^{r_1} - e^{r_2} - \theta e^{r_1 + r_2}}}$$

$$+ \frac{r_1 \cdot e^{-e^{r_1}} \cdot \int_{e^{r_2}}^{\infty} \frac{e^{-w(1+\theta e^{r_1})}}{w} + \int_{e^{r_1}}^{\infty}\int_{e^{r_2}}^{\infty} \frac{e^{-v-w-\theta v w}}{v \cdot w} \partial v \, \partial w}{e^{-e^{r_1} - e^{r_2} - \theta e^{r_1 + r_2}}}$$

$$= \frac{r_1 r_2 \cdot e^{-e^{r_1} - e^{r_2} - \theta e^{r_1 + r_2}} - r_1 \cdot e^{-e^{r_1}} Ei(-e^{r_2}\{1+\theta e^{r_1}\}) - r_2 \cdot e^{-e^{r_2}} Ei(-e^{r_1} \cdot \{1+\theta e^{r_2}\})}{e^{-e^{r_1} - e^{r_2} - \theta e^{r_1 + r_2}}}$$

$$+ \frac{\int_{e^{r_1}}^{\infty}\int_{e^{r_2}}^{\infty} \frac{e^{-v-w-\theta v w}}{v \cdot w} \partial v \, \partial w}{e^{-e^{r_1} - e^{r_2} - \theta e^{r_1 + r_2}}}$$

Aside : $$\int_{e^{r_1}}^{\infty}\int_{e^{r_2}}^{\infty} \frac{e^{-v-w-\theta v w}}{v \cdot w} \partial v \, \partial w = \int_{e^{r_1}}^{\infty} \frac{e^{-v}}{v} \cdot \left(\int_{e^{r_2}}^{\infty} \frac{e^{-w(1+\theta v)}}{w} \partial w \right) \partial v$$

$$= \int_{e^{r_1}}^{\infty} \frac{e^{-v}}{v} \cdot -Ei((1+\theta v) \cdot e^{r_2}) \, \partial v$$

$$= \int_{e^{r_1}}^{\infty} \frac{e^{-v}}{v} \cdot \left(-\gamma - \ln((1+\theta v) \cdot e^{r_2}) - \sum_{i=1}^{\infty} \frac{(-(1+\theta v) \cdot e^{r_2})^i}{i \cdot i!} \right) \partial v$$

$$= \left\{ (-\gamma - r_2) \cdot \int_{e^{r_1}}^{\infty} \frac{e^{-v}}{v} \partial v \right\} - \left\{ \int_{e^{r_1}}^{\infty} \frac{e^{-v}}{v} \cdot \ln(1+\theta v) \, \partial v \right\} - \left\{ \int_{e^{r_1}}^{\infty} \frac{e^{-v}}{v} \cdot \sum_{i=1}^{\infty} \frac{(-(1+\theta v) \cdot e^{r_2})^i}{i \cdot i!} \, \partial v \right\}$$

$$= \left\{ (-\gamma - r_2) \cdot -Ei(-e^{r_1}) \right\} - \left\{ \int_{e^{r_1}}^{\infty} \frac{e^{-v}}{v} \cdot \sum_{i=1}^{\infty} \frac{(-v)^i}{i} \partial v \right\} - \left\{ \sum_{i=1}^{\infty} \frac{(-e^{r_2})^i}{i \cdot i!} \cdot \int_{e^{r_1}}^{\infty} \frac{e^{-v}}{v} \cdot (1+\theta v)^i \, \partial v \right\}$$

$$= \left\{ (-\gamma - r_2) \cdot -Ei(-e^{r_1}) \right\} - \left\{ \sum_{i=1}^{\infty} \frac{(-1)^i}{i} \int_{e^{r_1}}^{\infty} e^{-v} \cdot v^{i-1} \, \partial v \right\}$$

$$- \left\{ \sum_{i=1}^{\infty} \frac{(-e^{r_2})^i}{i \cdot i!} \cdot \int_{e^{r_1}}^{\infty} \frac{e^{-v}}{v} \cdot \sum_{j=0}^{i} \binom{i}{j} \cdot (\theta v)^j \, \partial v \right\}$$

$$= \left\{ (-\gamma - r_2) \cdot -Ei(-e^{r_1}) \right\} - \left\{ \sum_{i=1}^{\infty} \frac{(-1)^i}{i} \cdot \Gamma(i, e^{r_1}) \right\}$$

$$- \left\{ \sum_{i=1}^{\infty} \frac{(-e^{r_2})^i}{i \cdot i!} \cdot \int_{e^{r_1}}^{\infty} \frac{e^{-v}}{v} \, \partial v \right\} - \left\{ \sum_{i=1}^{\infty} \frac{(-e^{r_2})^i}{i \cdot i!} \sum_{j=1}^{i} \binom{i}{j} \cdot (\theta)^j \int_{e^{r_1}}^{\infty} e^{-v} \cdot v^{j-1} \, \partial v \right\}$$

$$= \left\{ (-\gamma - r_2) \cdot -Ei(-e^{r_1}) \right\} - \left\{ \sum_{i=1}^{\infty} \frac{(-1)^i}{i} \cdot \Gamma(i, e^{r_1}) \right\}$$

$$- \left\{ \sum_{i=1}^{\infty} \frac{(-e^{r_2})^i}{i \cdot i!} \cdot -Ei(-e^{r_1}) \right\} - \left\{ \sum_{i=1}^{\infty} \frac{(-e^{r_2})^i}{i \cdot i!} \sum_{j=1}^{i} \binom{i}{j} \cdot (\theta)^j \cdot \Gamma(j, e^{r_1}) \right\}$$

$$= \left\{ -Ei(-e^{r_1}) \cdot -Ei(-e^{r_2}) \right\} - \left\{ \sum_{i=1}^{\infty} \frac{(-1)^i}{i} \cdot \Gamma(i, e^{r_1}) \right\} - \left\{ \sum_{i=1}^{\infty} \frac{(-e^{r_2})^i}{i \cdot i!} \sum_{j=1}^{i} \binom{i}{j} \cdot (\theta)^j \cdot \Gamma(j, e^{r_1}) \right\}$$

Therefore

$$\boxed{\begin{aligned} &\mathbf{E}\left(\epsilon_1 \cdot \epsilon_2 \,\middle|\, \begin{matrix} Y_1 > c_1, Y_2 > c_2 \\ (e^{\epsilon_1}, e^{\epsilon_2}) \sim \exp(I) \end{matrix} \right) \\ &= r_1 \cdot r_2 + \frac{-r_1 \cdot e^{-e^{r_1}} \cdot Ei(-e^{r_2} \cdot \{1+\theta e^{r_1}\}) - r_2 \cdot e^{-e^{r_2}} \cdot Ei(-e^{r_1} \cdot \{1+\theta e^{r_2}\})}{e^{-e^{r_1} - e^{r_2} - \theta e^{r_1 + r_2}}} \\ &+ \frac{Ei(-e^{r_1}) \cdot Ei(-e^{r_2}) - \sum_{i=1}^{\infty} \frac{(-1)^i}{i} \cdot \Gamma(i, e^{r_1}) - \sum_{i=1}^{\infty} \frac{(-e^{r_2})^i}{i \cdot i!} \sum_{j=1}^{i} \binom{i}{j} \cdot (\theta)^j \cdot \Gamma(j, e^{r_1})}{e^{-e^{r_1} - e^{r_2} - \theta e^{r_1 + r_2}}} \end{aligned}}$$

B. Gumbel's exponential (II)

$$E\left(\epsilon_1 \cdot \epsilon_2 \middle| \begin{array}{l} Y_1 > c_1, Y_2 > c_2 \\ (e^{\epsilon_1}, e^{\epsilon_2}) \sim \exp(II) \end{array}\right) =$$

$$\frac{\int_{r_1}^{\infty}\int_{r_2}^{\infty} \epsilon_1 \cdot \epsilon_2 \cdot \{1+\theta(2e^{-e^{\epsilon_1}}-1)(2e^{-e^{\epsilon_2}}-1)\} \cdot e^{-e^{\epsilon_1}-e^{\epsilon_2}+\epsilon_1+\epsilon_2}\, \partial\epsilon_1 \partial\epsilon_2}{\int_{r_1}^{\infty}\int_{r_2}^{\infty} \{1+\theta(2e^{-e^{\epsilon_1}}-1)(2e^{-e^{\epsilon_2}}-1)\} \cdot e^{-e^{\epsilon_1}-e^{\epsilon_2}+\epsilon_1+\epsilon_2}\, \partial\epsilon_1 \partial\epsilon_2}$$

where $r_1 = c_1\text{-}X\beta_1 \quad r_2 = c_2\text{-}X\beta_2$

Let $\quad e^{\epsilon_1} = v \quad \Rightarrow \quad \epsilon_1 = \ln(v) \quad \partial\epsilon_1 = \frac{\partial v}{v}$

$\quad e^{\epsilon_2} = w \quad \Rightarrow \quad \epsilon_2 = \ln(w) \quad \partial\epsilon_2 = \frac{\partial w}{w}$

$$E\left(\epsilon_1 \middle| \begin{array}{l} Y_1 > c_1, Y_2 > c_2 \\ (e^{\epsilon_1}, e^{\epsilon_2}) \sim \exp(II) \end{array}\right)$$

$$= \frac{\int_{e^{r_1}}^{\infty} \ln(v) \cdot e^{-v} \left(\{1\text{-}\theta(2e^{-v}\text{-}1)\} \cdot \int_{e^{r_2}}^{\infty} \ln(w) \cdot e^{-w} \partial w + \{2\theta(2e^{-v}\text{-}1)\} \cdot \int_{e^{r_2}}^{\infty} \ln(w) \cdot e^{-2w} \partial w \right) \partial v}{\int_{e^{r_1}}^{\infty} e^{-v} \left(\{1\text{-}\theta(2e^{-v}\text{-}1)\} \cdot \int_{e^{r_2}}^{\infty} e^{-w} \partial w + \{2\theta(2e^{-v}\text{-}1)\} \cdot \int_{e^{r_2}}^{\infty} e^{-2w} \partial w \right) \partial v}$$

$$= \frac{\int_{e^{r_1}}^{\infty} \ln(v) \cdot e^{-v} \cdot \{1\text{-}\theta(2e^{-v}\text{-}1)\} \cdot \left\{ \text{-}\ln(w) \cdot e^{-w} \Big|_{e^{r_2}}^{\infty} + \int_{e^{r_2}}^{\infty} \frac{e^{-w}}{w} \partial w \right\} \partial v}{\int_{e^{r_1}}^{\infty} e^{-v} \left(\{1\text{-}\theta(2e^{-v}\text{-}1)\} \cdot e^{-e^{r_2}} + \theta(2e^{-v}\text{-}1) \cdot e^{-2e^{r_2}} \right) \partial v}$$

$$+ \frac{\int_{e^{r_1}}^{\infty} \ln(v) \cdot e^{-v} \cdot \theta(2e^{-v}\text{-}1) \cdot \left\{ \text{-}\ln(w) \cdot e^{-2w} \Big|_{e^{r_2}}^{\infty} + \int_{e^{r_2}}^{\infty} \frac{e^{-2w}}{w} \partial w \right\} \partial v}{\int_{e^{r_1}}^{\infty} e^{-v} \left(\{1\text{-}\theta(2e^{-v}\text{-}1)\} \cdot e^{-e^{r_2}} + \theta(2e^{-v}\text{-}1) \cdot e^{-2e^{r_2}} \right) \partial v}$$

$$= \frac{\int_{e^{r_1}}^{\infty} \ln(v)\cdot e^{-v}\cdot\{1-\theta(2e^{-v}-1)\}\cdot\left\{r_2\cdot e^{-e^{r_2}} - Ei(-e^{r_2})\right\}\partial v}{\left(\{1-\theta(e^{-e^{r_2}}-1)\}\cdot\int_{e^{r_1}}^{\infty} e^{-v}\,\partial v \;+\; 2\theta(e^{-e^{r_2}}-1)\cdot\int_{e^{r_1}}^{\infty} e^{-2v}\,\partial v\right)\cdot e^{-e^{r_2}}}$$

$$+ \frac{\int_{e^{r_1}}^{\infty} \ln(v)\cdot e^{-v}\cdot\theta(2e^{-v}-1)\}\cdot\left\{r_2\cdot e^{-2e^{r_2}} - Ei(-2e^{r_2})\right\}\partial v}{\left(\{1-\theta(e^{-e^{r_2}}-1)\}\cdot\int_{e^{r_1}}^{\infty} e^{-v}\,\partial v \;+\; 2\theta(e^{-e^{r_2}}-1)\cdot\int_{e^{r_1}}^{\infty} e^{-2v}\,\partial v\right)\cdot e^{-e^{r_2}}}$$

$$= \frac{\left\{(1+\theta)\cdot\left(r_2\cdot e^{-e^{r_2}} - Ei(-e^{r_2})\right) \;-\; \theta\cdot\left(r_2\cdot e^{-2e^{r_2}} - Ei(-2e^{r_2})\right)\right\}\cdot\int_{e^{r_1}}^{\infty}\ln(v)\cdot e^{-v}\,\partial v}{\left(\{1-\theta(e^{-e^{r_2}}-1)\}\cdot e^{-e^{r_1}} \;+\; \theta(e^{-e^{r_2}}-1)\cdot e^{-2e^{r_1}}\right)\cdot e^{-e^{r_2}}}$$

$$+ \frac{\left\{2\theta\cdot\left(-r_2\cdot e^{-e^{r_2}} + Ei(-e^{r_2}) \;+\; r_2\cdot e^{-2e^{r_2}} - Ei(-2e^{r_2})\right)\right\}\cdot\int_{e^{r_1}}^{\infty}\ln(v)\cdot e^{-2v}\,\partial v}{\left(\{1-\theta(e^{-e^{r_2}}-1)\}\cdot e^{-e^{r_1}} \;+\; \theta(e^{-e^{r_2}}-1)\cdot e^{-2e^{r_1}}\right)\cdot e^{-e^{r_2}}}$$

$$= \frac{\left\{(1+\theta)\left(r_2\cdot e^{-e^{r_2}} - Ei(-e^{r_2})\right) - \theta\cdot\left(r_2\cdot e^{-2e^{r_2}} - Ei(-2e^{r_2})\right)\right\}\left\{-\ln(v)\cdot e^{-v}\Big|_{e^{r_1}}^{\infty} + \int_{e^{r_1}}^{\infty}\frac{e^{-v}}{v}\,\partial v\right\}}{\left(1 + \theta(e^{-e^{r_1}}-1)\cdot(e^{-e^{r_2}}-1)\right)\cdot e^{-e^{r_1}-e^{r_2}}}$$

$$+ \frac{\left\{2\theta\left(-r_2\cdot e^{-e^{r_2}} + Ei(-e^{r_2}) + r_2\cdot e^{-2e^{r_2}} - Ei(-2e^{r_2})\right)\right\}\cdot\left\{\frac{-\ln(v)\cdot e^{-2v}}{2}\Big|_{e^{r_1}}^{\infty} + \int_{e^{r_1}}^{\infty}\frac{e^{-2v}}{2\cdot v}\,\partial v\right\}}{\left(1 + \theta(e^{-e^{r_1}}-1)\cdot(e^{-e^{r_2}}-1)\right)\cdot e^{-e^{r_1}-e^{r_2}}}$$

$$= \frac{\left\{(1+\theta)\cdot\left(r_2\cdot e^{-e^{r_2}} - Ei(-e^{r_2})\right) - \theta\cdot\left(r_2\cdot e^{-2e^{r_2}} - Ei(-2e^{r_2})\right)\right\}\cdot\left\{r_1\cdot e^{-e^{r_1}} - Ei(-e^{r_1})\right\}}{\left(1+\ \theta(e^{-e^{r_1}}-1)\cdot(e^{-e^{r_2}}-1)\right)\ \cdot e^{-e^{r_1}-e^{r_2}}}$$

$$+ \frac{\left\{\theta\cdot\left(-r_2\cdot e^{-e^{r_2}} + Ei(-e^{r_2})\ +\ r_2\cdot e^{-2e^{r_2}} - Ei(-2e^{r_2})\right)\right\}\cdot\left\{r_1\cdot e^{-2e^{r_1}} - Ei(-2e^{r_1})\right\}}{\left(1+\ \theta(e^{-e^{r_1}}-1)\cdot(e^{-e^{r_2}}-1)\right)\cdot e^{-e^{r_1}-e^{r_2}}}$$

Thus

$$\boxed{\begin{aligned}
&\mathrm{E}\left(\epsilon_1\cdot\epsilon_2 \,\middle|\, \begin{matrix} Y_1 > c_1,\ Y_2 > c_2 \\ (e^{\epsilon_1}, e^{\epsilon_2}) \sim \exp(\mathrm{II}) \end{matrix}\right) \\
&= r_1\cdot r_2 \ + \ \frac{\theta\cdot\left\{Ei(-e^{r_1}) - Ei(-2e^{r_1})\right\}\cdot\left\{Ei(-e^{r_2}) - Ei(-2e^{r_2})\right\}\ +\ Ei(-e^{r_1})\cdot Ei(-e^{r_2})}{\left(1+\ \theta(e^{-e^{r_1}}-1)\cdot(e^{-e^{r_2}}-1)\right)\ \cdot e^{-e^{r_1}-e^{r_2}}} \ + \\
&\frac{\theta\left(r_1 e^{-e^{r_1}}(e^{-e^{r_1}}-1)\left\{Ei(-e^{r_2}) - Ei(-2e^{r_2})\right\} + r_2\, e^{-e^{r_2}}\cdot(e^{-e^{r_2}}-1)\left\{Ei(-e^{r_1}) - Ei(-2e^{r_1})\right\}\right)}{\left(1+\ \theta(e^{-e^{r_1}}-1)\cdot(e^{-e^{r_2}}-1)\right)\ \cdot e^{-e^{r_1}-e^{r_2}}}
\end{aligned}}$$

REFERENCES

Breslow, N. (1974). Covariance analysis of censored survival data. *Biometrics* **30** : 89-99.

Buckley, Jonathan and James, Ian. (1979). Linear regression with censored data. *Biometrika* **66** : 429-36.

Burke, Murray D. (1988). Estimation of a bivariate distribution function under random censorship. *Biometrika* **75** : 379-82.

Campbell, G. and Földes, A. (1980). Large-sample properties of nonparametric bivariate estimators with censored data. In *Nonparametric Statistical Inference,* Colloq. Math. Soc. Jånos Bolyai 32, Ed. B.B. Gnedenko. M.L. Puri and I. Vincze, 103-21. Amsterdam : North-Holland.

Clayton, D.G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika* **65** : 141-51.

Dabrowska, D.M. (1988). Kaplan-Meier estimate on the plane. *Annals of Statistics* **16** : 1475-89.

Hanley, James A. and Parnes, Milton N. (1983). Nonparametric estimation of a multivariate distribution in the presence of censoring. *Biometrics 39 :* 129-39.

Hougaard, Philip (1987). Modelling multivariate survival. *Scand. Jour. Stat.* **14** : 291-304.

Huster, William J., Brookmeyer, Ron, and Self, Steven G. (1989). Modelling paired survival data with covariates. *Biometrics* **45** : 145-56.

Johnson, Norman L. and Kotz, S. (1972). *Distributions in Statistics : Continuous Multivariate Distributions,* New York : Wiley.

Kalbfleisch, J. and Prentice, R.L. (1972). Discussion on Professor Cox's paper. *Journal of the Royal Statistical Society, Series B* **34** : 215-216.

Kao, Chihwa (1984). On the estimation of censored regression with constant coefficient of variation. *Academia Economic Papers* **12** : 129-34.

Miller, Rupert (1981). *Survival Analysis,* New York: Wiley.

Miller, Rupert and Halpern, Jerry (1981). Regression with censored data. *Technical Report No. 66,* Division of Biostatistics, Stanford University.Dabrowska, D.M. (1988). Kaplan-Meier estimate on the plane. *Annals of Statistics* **16** : 1475-89.

Nelson, W. (1969). Hazard plotting for incomplete failure data. *Journal of Quality Technology* **1** : 27-52.

Oakes, David (1982). A model for association in bivariate survival data. *Journal of the Royal Statistical Society, Series B* **44** : 414-22.

Oakes, David (1989). Bivariate survival models induced by frailties. *Journal of the American Statistical Association* **84** : 487-93.

Peto, R. (1972). Discussion on Professor Cox's paper. *Journal of the Royal Statistical Society, Series B* **34** : 205-207.

Tawn, Jonathan (1988). Bivariate extreme value theory: models and estimation. *Biometrika* **75**, 397-415.

Tsai, W.Y., Leurgens, S., and Crowley, J. (1986). Nonparametric estimation of a bivariate survival function in the presence of censoring. *Annals of Statistics* **14** : 1351-1365.

Wei, L.J., Lin, D.Y., Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modelling marginal distributions. *Journal of the American Statistical Association* **84** : 1065-1073.

Weissfeld, Lisa, and Schneider, Helmut (1987). Inferences based on the Buckley-James procedure. *Communications in Statistics - Theory Meth.* **16** : 1773-87.