ISSN: 1525-3066

Center for Policy Research
Working Paper No. 68

COX-McFADDEN PARTIAL AND MARGINAL LIKELIHOODS FOR THE PROPORTIONAL HAZARD MODEL WITH RANDOM EFFECTS

Jan Ondrich

Center for Policy Research
Maxwell School of Citizenship and Public Affairs
Syracuse University
426 Eggers Hall
Syracuse, New York 13244-1020
(315) 443-3114 | Fax (315) 443-1081
e-mail: ctrpol@syr.edu

August 2005

$5.00

Up-to-date information about CPR's research projects and other activities is available from our World Wide Web site at **www-cpr.maxwell.syr.edu.** All recent working papers and Policy Briefs can be read and/or printed from there as well.

CENTER FOR POLICY RESEARCH – Spring 2005

Timothy Smeeding, Director
Professor of Economics & Public Administration

Associate Directors

Margaret Austin
Associate Director,
Budget and Administration

Douglas Holtz-Eakin
Professor of Economics
Associate Director, Center for Policy Research

Douglas Wolf
Professor of Public Administration
Associate Director, Aging Studies Program

John Yinger
Professor of Economics and Public Administration
Associate Director, Metropolitan Studies Program

SENIOR RESEARCH ASSOCIATES

Pablo Beramendi……………..……Political Science
Dan Black..Economics
Lloyd Blanchard………………Public Administration
Stacy Dickert-Conlin............................Economics
William Duncombe .... Public Administration
Gary Engelhardt ....Economics
Deborah Freund ....Public Administration
Madonna Harrington Meyer ....Sociology
Christine Himes......................................Sociology
William C. Horrace ....Economics
Bernard Jump ....Public Administration
Duke Kao ....Economics
Eric Kingson ....Social Work
Thomas Kniesner ....Economics
Jeff Kubik ....Economics
Andrew London .... Sociology
Len Lopoo .... Public Administration
Jerry Miner .... Economics
John Moran .... Economics
Jan Ondrich... Economics
John Palmer .... Public Administration
Lori Ploutz-Snyder........................ Exercise Science
Jeff Racine .... Economics
Grant Reeher ....Political Science
Stuart Rosenthal .... Economics
Ross Rubenstein..................... Public Administration
Margaret Usdansky…………………………Sociology
Michael Wasylenko .... Economics
Janet Wilmoth .... Sociology

GRADUATE ASSOCIATES

Anna Amirkhanyan.................Public Administration
Megan Bahns………………………………Sociology
Sonali Ballal………………….. Public Administration
Dana Balter ....Public Administration
Sami Beg………………………Public Administration
Yong Chen ....Economics
Seokjoon Choi……………………………Economics
Christopher Cunningham ....Economics
Ana Dammert..Economics
Mike Eriksen ....Economics
Ying Fang………………………………… Sociology
Katie Fitzpatrick………………………......Economics
Jose Galdo ....Economics
Alexandre Genest…………….Public Administration
Andrew Hanson…………………………...Economics
Yue Hu…………………………………......Economics
Hatice Karahan……………………………Economics
Kristina Lambright ……………Public Administration
Jessica Lee……………………Public Administration
Liqun Liu……………………………………..Economics
Long Liu……………………………………Economics
Joseph Marchand………………………....Economics
Emily NaPier……………………………….Sociology
Emily Pas …………………………………Economics
Kristenne Robison…………………………Sociology
Cynthia Searcy………………..Public Administration
Claudia Smith……………………….........Economics
Mark Thomas…………………..Public Administration
Jeff Thompson .... Economics
Wen Wang……………………..……………Economics
Bo Zhao……………………………..……….Economics

STAFF

Kelly Bogart ....Administrative Secretary
Martha Bonney..... Publications/Events Coordinator
Karen Cimilluca........... Librarian/Office Coordinator
Kim Desmond …………….Administrative Secretary
Kati Foley……………..Administrative Assistant, LIS
Kitty Nasto…………………Administrative Secretary
Candi Patterson......................Computer Consultant
Mary Santy……..…………Administrative Secretary

ABSTRACT

In survival analysis, Cox's name is associated with the partial likelihood technique that allows consistent estimation of proportional hazard scale parameters without specifying a duration dependence baseline. In discrete choice analysis, McFadden's name is associated with the generalized extreme-value (GEV) class of logistic choice models that relax the independence of irrelevant alternatives assumption. This paper shows that the mixed class of proportional hazard specifications allowing consistent estimation of scale and mixing parameters using partial likelihood is isomorphic to the GEV class. Independent censoring is allowed and I discuss approximations to the partial likelihood in the presence of ties. Finally, the partial likelihood score vector can be used to construct log-rank tests that do not require the independence of observations involved.
(*JEL* C14, C41)

Keywords: proportional hazard, random effects, partial likelihood, GEV class

COX-McFADDEN PARTIAL AND MARGINAL LIKELIHOODS FOR THE PROPORTIONAL HAZARD MODEL WITH RANDOM EFFECTS

1. INTRODUCTION

This paper examines the problem of incorporating random effects in a proportional hazard model, leaving the baseline hazard unspecified. It shows that the set of models that support partial likelihood estimation of the hazard scale coefficients can be made isomorphic to the set of generalized extreme-value models developed by McFadden (1978). An interesting aspect of the proof is the application of a multivariate extension of a theorem proved by Sergei Bernstein in 1928 for the univariate case. This extension provides a means to check whether a given multivariate function can be the likelihood function for a sample of durations, marginal on group-specific random effects.

Cox (1972, 1975) develops the proportional hazard model of durations and suggests estimation using a partial likelihood approach. Contributions to the partial likelihood are provided at each failure time by the subset of the sample at risk immediately before the failure time. The partial likelihood approach has the advantage of being baseline-free: duration-dependence parameters, frequently viewed as nuisance parameters, do not have to be estimated. For researchers interested in duration dependence, the duration baseline can be recovered in a second step. The case for partial likelihood was strengthened with the later finding that partial likelihood estimation is equivalent to rank-information marginal likelihood estimation.

The introduction of stratified partial likelihood estimation (see Chamberlain 1985, Gross and Huber 1987, Andersen, Borgan, Gill, and Keiding 1993, and Ridder and Tunali 1999) allows for models with group-specific fixed effects and group-specific duration baselines. Group-specific duration baselines can be recovered in a second stage, but group-specific fixed effects and the coefficients of covariates invariant within groups cannot be recovered. Stratified partial likelihood estimation, therefore, does not allow hazard prediction.

This paper investigates a class of models for baseline-free partial likelihood or rank-information marginal likelihood with random effects that allows hazard prediction and the estimation of coefficients of covariates invariant within groups. The model draws heavily on the previous work of Hougaard (1986a, 1986b) and the analysis of McFadden (1978) generalizing the multinomial logit model.

In the absence of group-specific fixed or random effects the mathematical form of the partial likelihood or rank-information marginal likelihood contributions is identical to that of the individual log-likelihood contributions for the multinomial logit model, proposed by Luce (1959) to estimate the probability that an item is selected from a choice set of alternatives. McFadden (1974) presents a formal econometric analysis of the multinomial logit model. The model assumes that the stochastic utility of each choice is the sum of a deterministic component and an extreme-value error term. The model has the property that the log-odds of any two choices are independent of the availability or attributes of other alternatives. While the independence of irrelevant alternatives (IIA) property simplifies the econometric estimation, it is an undesirable feature in choice settings in which alternatives have close substitutes. McFadden (1978) outlines a generalization of the multinomial logit model that allows the IIA property to be relaxed.

McFadden's generalization of the multinomial logit model introduces a class of multivariate extreme-value distributions (called generalized extreme-value or simply

GEV) defined by imposing restrictions on the negative of the log of the copula of the distribution. (Copulas and negative log copulas are defined in section 2 of this paper.) Four restrictions are imposed on the negative log copula: sign alternation of partial derivatives, non-negativity, an infinite limit as any argument limits to infinity, and homogeneity of degree one. As will be demonstrated, the key step for incorporating random effects in a baseline-free partial likelihood or rank-information marginal likelihood framework is the use of McFadden's negative log copula to model the joint hazard function of the durations in the sample.

Section 2 describes the multinomial logit model, the IIA property, and the GEV class of models developed by McFadden. Section 3 presents Cox's proportional hazard model and the two main propositions of this study. Proposition 1 states that any non-negative multivariate function with appropriately alternating partial derivatives is a joint survivor function marginal on group-specific random effects. Proposition 2 states that the additional properties regarding infinite limits and homogeneity of degree one make the partial likelihood and rank-information marginal likelihood baseline-free. Proposition 1 is proved in section 4. Proposition 2 is proved for the partial likelihood case in section 5 and for the rank-information marginal likelihood case in section 6. In section 7 the case of tied data is discussed. The recovery of the baseline hazard is described in section 8. Section 9 discusses the construction of log-rank tests with dependent observations. Section 10 presents examples of Cox-McFadden random-effects models with unspecified baselines. Section 11 summarizes the paper. An appendix discusses asymptotic inference.

2. THE MULTINOMIAL LOGIT, IIA, AND THE GEV MODEL

The discrete choice model specification that is used most often in applied econometric applications is the multinomial logit model. The multinomial logit model provides a simple closed form for the choice probabilities; in contrast, the calculation of the choice probabilities in the multinomial probit model requires multivariate integration that can only be accomplished through numerical approximation. The likelihood function for the multinomial logit specification is globally concave, which eases the computational burden of obtaining maximum likelihood estimates.

In the multinomial logit model, the probability that an individual chooses choice i from a choice set C consisting of J choices is given by

$$P(i \mid C, \mathbf{Z}, \boldsymbol{\beta}) = e^{\mathbf{Z}_i \boldsymbol{\beta}} / \sum_{j \in C} e^{\mathbf{Z}_j \boldsymbol{\beta}} \quad , \tag{2.1}$$

where $\mathbf{Z}_j$ is a K-vector of explanatory variables describing the attributes of alternative j (perhaps interacted or moderated by the characteristics of the decision-maker), $\mathbf{Z} = (\mathbf{Z}_1, \ldots, \mathbf{Z}_J)$ gives the attributes of C, and $\boldsymbol{\beta}$ is a K-vector of taste parameters.

The multinomial logit model is characterized by the independence of irrelevant alternatives (IIA) property, namely, the ratio of probabilities (relative odds) of choosing any two alternatives is independent of the availability of a third alternative:

$$P(i \mid C, \mathbf{Z}, \boldsymbol{\beta}) = P(i \mid C_0, \mathbf{Z}, \boldsymbol{\beta}) P(C_0 \mid C, \mathbf{Z}, \boldsymbol{\beta}) , \tag{2.2}$$

where $i \in C_0 \subseteq C$ and

$$P(C_0 \mid C, \mathbf{Z}, \boldsymbol{\beta}) = \sum_{j \in C_0} P(j \mid C, \mathbf{Z}, \boldsymbol{\beta}) . \tag{2.3}$$

A famous example has a commuter choosing between a car and a bus for a commute. When he is late for work, which happens randomly one-third of the time, he drives (choice *A*); otherwise he chooses a bus. There are two bus companies, a red bus company and a blue bus company, indistinguishable but for color. When he is not late and is waiting for a bus, the first bus to arrive is equally likely to be blue (choice *BB*) or red (choice *RB*). From this information it is clear that with choice set $C = \{A, RB, BB\}$

$$P(A) = P(RB) = P(BB) = 1/3 \quad . \tag{2.4}$$

Now suppose that the blue bus company suspends operations. The choice set becomes $C_0 = \{A, RB\}$, which has probability 2/3 by equations (2.3) and (2.4). With choice set $C_0 = \{A, RB\}$, the multinomial logit model predicts that

$$P(A) = P(RB) = 1/2 \quad , \tag{2.5}$$

using equations (2.2) and (2.4). But this prediction is not likely to be validated. The commuter will continue to choose the car whenever he is late, 1/3 of the time, and the red bus, 2/3 of the time, whenever he is not.

It is clear from this example that models that have the IIA property are inadequate in describing choice from a set of alternatives that have different degrees of substitutability or complementarity. The red bus and blue bus are perfect substitutes, whereas the car and the red bus (or the car and the blue bus) are not. Several studies (McFadden, Train, and Tye 1977, Hausman and McFadden 1984, Small and Hsiao 1985, and McFadden 1987) discuss methods of testing whether IIA is violated in a given econometric application. The next step is choosing an alternative model, preferably one with closed forms for the choice probabilities.

This problem was resolved by McFadden (1978) making using of results derived by Williams (1977) and Daly and Zachary (1978) on the compatibility of a given probabilistic choice model with utility maximization.[1] McFadden's solution is given in the following theorem.

Theorem 1 (McFadden):

Suppose $M(\theta_1, \ldots, \theta_J)$ is a function defined on the non-negative real numbers with the following four properties:

1) alternating distinct partials, i.e., for any distinct $\{j_1, \ldots, j_Q\}$ from the choice set $\{1, \ldots, J\}$, the Qth partial $\partial^Q M / \partial\theta_{j_1} \ldots \partial\theta_{j_Q}$ is non-negative if Q is odd and non-positive if Q is even.
2) non-negativity;
3) infinite limits, i.e., $\lim_{\theta_i \to \infty} M(\theta_1, \ldots, \theta_J) = \infty$, $i = 1, \ldots, J$; and
4) homogeneity of degree one.

Then, the probabilities

$$(2.6) \quad P(i \mid C, \mathbf{Z}, \boldsymbol{\beta}) = e^{\mathbf{Z}_i\boldsymbol{\beta}} \frac{\partial M(e^{\mathbf{Z}_1\boldsymbol{\beta}}, \ldots, e^{\mathbf{Z}_J\boldsymbol{\beta}})}{\partial e^{\mathbf{Z}_i\boldsymbol{\beta}}} / M(e^{\mathbf{Z}_1\boldsymbol{\beta}}, \ldots, e^{\mathbf{Z}_J\boldsymbol{\beta}}), \quad i = 1, \ldots, J$$

---

1. A probabilistic choice model is compatible with utility maximization if and only if the choice probabilities sum to unity, are non-negative, translation invariant, integrable, i.e., $\partial P(i \mid C, \mathbf{Z}, \boldsymbol{\beta}) / \partial \mathbf{Z}_j\boldsymbol{\beta} = \partial P(j \mid C, \mathbf{Z}, \boldsymbol{\beta}) / \partial \mathbf{Z}_i\boldsymbol{\beta}$ for all $i, j \in C$, and their implied distribution function is well-defined, i.e., $(-1)^{J-1}\partial^{J-1} P(i \mid C, \mathbf{Z}, \boldsymbol{\beta}) / \partial \mathbf{Z}_1\boldsymbol{\beta} \ldots [\partial \mathbf{Z}_i\boldsymbol{\beta}] \ldots \partial \mathbf{Z}_J\boldsymbol{\beta}$ exists and is nonnegative and continuous for all $i \in C$ (see Daly and Zachary 1979, or Börsch-Supan 1987).

define a probabilistic choice model on the choice set $\{1,\ldots,J\}$ that is consistent with utility maximization.

The function M is McFadden's negative log copula. A copula is a function that assigns the value of the joint distribution function to each n-tuple of values of the marginal distributions. (Andersen 2004 uses copulas to construct a two-stage semi-parametric estimator for multivariate failure-time data.) I define a negative log copula to be a function that assigns the value of the negative log of the joint distribution function to each n-tuple of values of the negative log of the marginal distributions. McFadden's negative log copula will be shown to play a crucial role in specifying the baseline-free partial likelihood and rank-information marginal likelihood that is consistent with group-specific random effects.

3. THE PROPORTIONAL HAZARD MODEL AND TWO PROPOSITIONS

The duration or failure time T of a stochastic process is its random age at termination or failure. The assumption in this study is that durations are continuous random variables: they possess an absolutely continuous distribution function $F(t)$. The distribution function is non-defective, i.e., $F(\infty)=1$, and has density $f(t)$. The unitary complement of the distribution function of a continuous duration

$$\begin{aligned} S(t) &\equiv P(T \geq t) \\ &= 1 - F(t). \end{aligned} \tag{3.1}$$

is its survivor function. The survivor function represents the probability that the process survives up to age t, and only fails at time t or later.

One of the fundamental concepts in the analysis of continuous durations is the hazard rate, denoted by *h* and defined by:

(3.2) $$h(t) \equiv f(t)/(1-F(t)).$$

The quantity $h(t)dt$ represents the probability that the process fails in the interval $[t, t+dt)$ conditional on survival to age *t*. It is well known that for a specific $h(t)$, the survivor function and density are given by:

(3.3) $$\begin{aligned} S(t) &= \exp(-\int_0^t h(u)du) \\ &\equiv \exp(-H(t)) \end{aligned}$$

and

(3.4) $$f(t) = h(t)\exp(-\int_0^t h(u)du).$$

For a sample of *N* spells, Cox's proportional hazard specification assigns to spell *i* a hazard rate of the form:

(3.5) $$\begin{aligned} h_i(t \mid \mathbf{Z}, \boldsymbol{\beta}) &= \exp(\mathbf{Z}_i \boldsymbol{\beta}) h_0(t) \\ &\equiv \theta_i h_0(t), \end{aligned}$$

where $\mathbf{Z}_i$ is the covariate vector for spell *i*, $\mathbf{Z} = (\mathbf{Z}_1, \ldots, \mathbf{Z}_N)$, $\boldsymbol{\beta}$ is the coefficient vector, and h_0 is the (unspecified) baseline hazard rate. (I will assume that the covariate vector is time-invariant, i.e., it does not change with process age. The principal results of this study allow time-varying covariates, as discussed in the appendix, Ondrich 2005.) The integrated baseline hazard rate is defined by

(3.6) $$H_0(t) \equiv \int_0^t h_0(u)du,$$

so that the survivor function for spell *i* can be written simply as $S_i(t) = \exp(-\theta_i H_0(t))$. In a proportional hazard model the hazard elasticity with respect to any continuous positive covariate depends only on the value of the covariate and its coefficient, and does not require additional knowledge of the process age *t*.

The sample survivor function for the sample of *N* spells is defined as:

(3.7) $$S(u_1, \ldots, u_N \mid \mathbf{Z}, \boldsymbol{\beta}) \equiv P(T_1 \geq u_1, \ldots, T_N \geq u_N) .$$

It will also be necessary to define marginal survivor functions. The marginal survivor function of a subset of the *N* sample spells is derived from the sample survivor function by setting $u_i = 0$ for all i not in the subset. Alternatively, denote the subset by *A* and for each *i* define Y_i^A if i is an element of *A*. Then, letting $\mathbf{u}$ be the vector $(u_1, \ldots, u_N)$, the marginal survivor function is given by:

(3.8) $$S_A(\mathbf{u} \mid \mathbf{Z}, \boldsymbol{\beta}) = S(Y_1^A u_1, \ldots, Y_N^A u_N \mid \mathbf{Z}, \boldsymbol{\beta}) .$$

Of particular interest will be the marginal survivor function $S_{R(t)}(t\,\boldsymbol{\iota} \mid \mathbf{Z}, \boldsymbol{\beta})$, for which $\mathbf{u}$ is the constant vector $t\,\boldsymbol{\iota}$, where $\boldsymbol{\iota}$ is the *N*-dimensional unitary vector, and the subset of interest is the *risk set* at time *t*, denoted $R(t)$, the subset of sample spells that empirically survive to age *t*. Note that $R(0)$ is the complete sample of durations.

If the *N* sample spells are statistically independent, the sample survivor function is:

(3.9) $$S(\mathbf{u} \mid \mathbf{Z}, \boldsymbol{\beta}) = \exp(-\sum_{i=1}^{N} \theta_i H_0(u_i)) ,$$

defined on non-negative real *N*-tuples $\mathbf{u}$. The main proposition in this study involves samples of statistically dependent (mixed) spells for which the survivor function takes the form:

(3.10) $$S(\mathbf{u} \mid \mathbf{Z}, \boldsymbol{\beta}) = \exp(-M(\theta_1 H_0(u_1), \ldots, \theta_N H_0(u_N))),$$

again defined on the non-negative real N-tuples $\mathbf{u}$. The function M stands for McFadden's negative log copula.

In equation (3.10), $\bar{\theta}_i(u_i) \equiv \theta_i H_0(u_i)$ has two possible interpretations, one of which must be chosen. The first is the unmixed individual integrated hazard. This equals the negative log of the unmixed individual survivor function, that is, the negative log of the individual survivor function when the value of its multiplicative random effect equals unity. The second interpretation is the mixed individual integrated hazard. This is the negative log of the mixed individual survivor function. The mixed individual survivor function is the individual survivor function with the multiplicative random effect integrated out. The mixed individual survivor function has also been called the marginal individual survivor function, because it is the survivor function marginal on the random effects, but I will reserve the term marginal individual survivor function for the functions in equation (3.8).

The function $\bar{\theta}_i(u_i)$ is given the second interpretation. The unmixed individual integrated hazard is denoted by $\lambda_i(u_i)$. It is clear that for each M in equation (3.10) there exists M^* such that

(3.11) $$M^*(\lambda_1(u_1), \ldots, \lambda_N(u_N)) = M(\bar{\theta}_1(u_1), \ldots, \bar{\theta}_N(u_N)) \quad .$$

The functions M and M^* satisfying equation (3.11) are said to be associated.

To simplify the notation further, let $\boldsymbol{\theta} \equiv (\theta_1, \ldots, \theta_N)$, $\mathbf{H}_0(\mathbf{u}) \equiv (H_0(u_1), \ldots, H_0(u_N))$, and the indicators Y_i^A be defined as before. Now define

(3.12) $$M(\boldsymbol{\theta}, \mathbf{H}_0(\mathbf{u}), A) = M(Y_1^A \theta_1 H_0(u_1), \ldots, Y_N^A \theta_N H_0(u_N)) .$$

Then, for all $\mathbf{u}$ and A,

(3.13) $$S_A(\mathbf{u} \mid \mathbf{Z}, \boldsymbol{\beta}) = \exp(-M(\boldsymbol{\theta}, \mathbf{H}_0(\mathbf{u}), A))\,.$$

If $\mathbf{u}$ is a constant vector and $M(\overline{\theta}_1(u_1), \ldots, \overline{\theta}_N(u_N))$ is homogeneous of degree one, then:

(3.14) $$M(\boldsymbol{\theta}, \mathbf{H}_0(t\,\boldsymbol{\iota}), A) = H_0(t) M(\boldsymbol{\theta}, \boldsymbol{\iota}, A)\,,$$

so that for all constant vectors $\mathbf{u}$ and sets A,

(3.15) $$S_A(t\,\boldsymbol{\iota} \mid \mathbf{Z}, \boldsymbol{\beta}) = \exp(-H_0(t) M(\boldsymbol{\theta}, \boldsymbol{\iota}, A))\,.$$

The preceding results on sample survivor functions and marginal survivor functions will be useful in proving the main propositions of this study, which I now present.

Proposition 1:

Suppose $M^*(\lambda_1, \ldots, \lambda_N)$ is a non-negative function defined on the non-negative real numbers possessing alternating partials, i.e.,

1) for any non-negative vector of integers $(q_1, \ldots, q_N)$, the Qth partial, where $Q = \sum_{i=1}^{N} q_i$, of $M^*(\lambda_1, \ldots, \lambda_N)$, $\partial^Q M^*(\lambda_1, \ldots, \lambda_N) / \partial \lambda_1^{q_1} \ldots \partial \lambda_N^{q_N}$, is non-negative if Q is odd and non-positive if Q is even.

Then the sample survivor function $\exp(-M^*)$ is consistent with a random-effects specification.

Proposition 2:

Suppose that the non-negative function $M(\overline{\theta}_1, \ldots, \overline{\theta}_N)$ associated with $M^*(\lambda_1, \ldots, \lambda_N)$ from Proposition 1 has the following properties:

2) infinite limits, i.e., $\lim_{\overline{\theta}_i \to \infty} M(\overline{\theta}_1, \ldots, \overline{\theta}_N) = \infty$, $i = 1, \ldots, N$;

and

3) homogeneity of degree one.

Suppose again that the sample survivor function is given by equation (3.10), and ties in the data, i.e., two durations with the same age, occur with probability zero. Then the partial likelihood and rank-information marginal likelihood are baseline-free and the probability that spell i in risk set $R(t)$ fails at time t is given by

$$P(i \mid R(t), \mathbf{Z}, \boldsymbol{\beta}) = \theta_i \frac{\partial M(\boldsymbol{\theta}, \boldsymbol{\iota}, R(t))}{\partial \theta_i} / M(\boldsymbol{\theta}, \boldsymbol{\iota}, R(t)) . \tag{3.16}$$

Proposition 1 will be proved in the next section. The condition on alternating partials for M^* in Proposition 1 is for any partial derivative and not just for distinct partials. In fact, the condition on alternating partials for M^* implies that M has alternating distinct partials. To see this, write λ_i as a function of $\overline{\theta}_i$ and note that

$$\begin{aligned} \overline{\theta}_i &= M^*(0, \ldots, 0, \lambda_i(\overline{\theta}_i), 0, \ldots, 0) \\ &= \vartheta_i(\lambda_i(\overline{\theta}_i)) \end{aligned} \tag{3.17}$$

for $i = 1, \ldots, N$. Therefore, ϑ_i and λ_i are inverse functions, and by the chain rule:

$$\vartheta_i'(\lambda_i(\overline{\theta}_i))\lambda_i'(\overline{\theta}_i) = 1 \tag{3.18}$$

for each i. By Proposition 1, $\vartheta_i' > 0$ and therefore $\lambda_i' > 0$ for each i. It follows that for any distinct $\{j_1, \ldots, j_Q\}$:

$$\begin{aligned} &\partial^Q M(\overline{\theta}_{j_1}, \ldots, \overline{\theta}_{j_Q}) / \partial \overline{\theta}_{j_1} \ldots \partial \overline{\theta}_{j_Q} \\ &= \partial^Q M^*(\lambda_{j_1}, \ldots, \lambda_{j_Q}) / \partial \lambda_{j_1} \ldots \partial \lambda_{j_Q} \left(\prod_{i=1}^{Q} \frac{d\lambda_{j_i}}{d\overline{\theta}_{j_i}} \right). \end{aligned} \tag{3.19}$$

In equation (3.19), the sign of the left-hand side is the same as the sign of $\partial^Q M^*(\lambda_{j_1}, \ldots, \lambda_{j_Q}) / \partial\lambda_{j_1} \ldots \partial\lambda_{j_Q}$. Thus, Proposition 1 implies that *M* has alternating distinct partials. But because Proposition 1 requires more than this, the class of models generated by Propositions 1 and 2 is, strictly speaking, a subclass of McFadden's original GEV class. However, all examples presented in McFadden (1978) also belong to the subclass, and to simplify the discussion, I will equate the subclass class and the GEV class.

The proof of Proposition 2 will be completed in sections 5 and 6. Note first that I deal only with non-negative functions *M* because a negative value for *M* implies that the survivor function, which is a probability, can exceed unity. Furthermore, properties 2)-3) in Proposition 2 are, in fact, necessary and sufficient for the partial likelihood and rank-information marginal likelihood to be baseline-free. The necessity of property 2) follows from the requirement that the sample survivor function be non-defective to ensure that the integral of the rank-information marginal likelihood is unity.

A non-defective sample survivor function is one that assigns a zero probability to all events in which any spell survives to infinity. In other words, zero is the limiting probability for any limiting sequence of events for which the age of a given spell limits to infinity. The maintained assumption is that each mixed individual survivor function is non-defective: it limits to zero as age limits to infinity. Equivalently, the mixed individual integrated hazard limits to infinity with age. Thus, zero is the limiting probability for any limiting sequence of events for which the mixed individual integrated hazard of a given spell limits to infinity. It follows that *M* limits to infinity with the mixed individual integrated hazard of any spell.

The necessity and sufficiency of property 3) will be demonstrated for the partial likelihood estimator in section 5 and for the rank-information marginal likelihood in section 6.

4. THE LAPLACE TRANSFORM, COMPLETE MONOTONICITY, AND RANDOM EFFECTS

In this section I prove Proposition 1, which states that alternating partials of the negative log of the sample survivor function are sufficient for the sample survivor function to be consistent with a random-effects specification.

The starting point is to specify a vector of non-negative spell-specific random effects, $\mathbf{v} = (\nu_1, \ldots, \nu_N)$, that is orthogonal to the covariate matrix $\mathbf{Z}$. Denote the joint distribution function of $\mathbf{v}$ by $\Omega(\nu_1, \ldots, \nu_N \mid \mathbf{Z}, \boldsymbol{\beta})$. The vector $\mathbf{v}$ captures the effect of unobserved variables that determine the sample survivor function. Conditional on $\mathbf{v}$, $\mathbf{Z}$, and $\boldsymbol{\beta}$, the sample survivor function is the product of N independent spell survivor functions and is written as:

$$(4.1) \qquad S(\mathbf{u} \mid \mathbf{v}, \mathbf{Z}, \boldsymbol{\beta}) = \exp(-\sum_{i=1}^{N} \nu_i \lambda_i(u_i)).$$

Unfortunately, the survivor function in equation (4.1) cannot be the basis for a partial likelihood or rank-information marginal likelihood since the ν_i 's are unobserved. The unobserved effects must be integrated out of the sample survivor function over their joint distribution:

$$(4.2) \qquad S(\mathbf{u} \mid \mathbf{Z}, \boldsymbol{\beta}) = \int_{R^{N+} \cup \{\mathbf{0}\}} \exp(-\sum_{i=1}^{N} \nu_i \lambda_i(u_i))\, d\Omega(\nu_1, \ldots, \nu_N \mid \mathbf{Z}, \boldsymbol{\beta}).$$

The sample survivor function $S(\mathbf{u} \mid \mathbf{Z}, \boldsymbol{\beta})$ is an example of a multivariate Laplace transform. Relevant results on univariate Laplace transforms that are straightforward to extend to the multivariate case are found in Feller (1971). In the univariate case, if G_1 is a univariate distribution function concentrated on $R^+ \cup \{0\}$, the Laplace transform ω_1 of G_1 is defined as:

(4.3) $$\omega_1(\lambda) = \int_{R^+ \cup \{0\}} \exp(-\lambda s)\, dG_1(s)\,,\ \lambda \geq 0.$$

Analogously, in the multivariate case, if G_J is a J-variate distribution function concentrated on $R^{J+} \cup \{\mathbf{0}\}$ and $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_J)$, then the Laplace transform ω_J of G_J is defined as:

(4.4) $$\omega_J(\boldsymbol{\lambda}) = \int_{R^{J+} \cup \{\mathbf{0}\}} \exp(-\sum_{i=1}^{J} \lambda_i s_i)\, dG_J(s_1, \ldots, s_J)\,, \quad \lambda_i \geq 0, \quad i = 1, \ldots, J.$$

Feller (1971) shows that distinct distribution functions have distinct Laplace transforms, and he discusses the convergence of the integral in equation (4.3). If the integral converges for $\lambda > a$, then the function ω_1 defined for $\lambda > a$ is called the Laplace transform of G_1. In the present context, I deal only with Laplace transforms that are defined for all $\lambda \geq 0$ (actually, since my chief concern is with the multivariate case, I deal only with J-variate Laplace transforms that are defined for the region $\{\boldsymbol{\lambda} \geq \mathbf{0}\}$.) The reason for dealing only with these is that λ_i corresponds to the unmixed individual integrated hazard, which is non-negative, but should not otherwise be bounded from below *a priori*.

Feller (1971) also proves a theorem on the convergence of sequences of univariate Laplace transforms. I will need a multivariate version of this theorem, which I state without proof. The proof of the theorem is a straightforward extension of the proof in Feller (1971) for the univariate case.

Theorem 2 (Continuity Theorem):

For $n = 1, 2, \dots$ let G_J^n be a J-variate distribution function with Laplace transform φ_J^n.

If $G_J^n \to G_J$ where G_J is a possibly defective distribution with transform φ_J, then

$\varphi_J^n(\boldsymbol{\lambda}) \to \varphi_J(\boldsymbol{\lambda})$ for non-zero and non-negative $\boldsymbol{\lambda}$.

Conversely, if the sequence $\{\varphi_J^n(\boldsymbol{\lambda})\}$ converges for each non-zero and non-negative $\boldsymbol{\lambda}$ to a limit $\varphi_J(\boldsymbol{\lambda})$, then φ_J is the transform of a possibly defective distribution function G_J,

and $G_J^n \to G_J$.

The limit G_J is not defective if and only if $\varphi_J(\boldsymbol{\lambda}) \to 1$ as $\boldsymbol{\lambda} \to \mathbf{0}$.

The next step is to define the property of complete monotonicity. Feller (1971) defines a (non-negative) univariate function φ_1 to be completely monotone if it possesses derivatives $\frac{d^n \varphi_1}{d\lambda^n}$ of all orders and $(-1)^n \frac{d^n \varphi_1(\lambda)}{d\lambda^n} \geq 0$. Bernstein (1928) proves that a univariate function φ_1 on $[0, \infty)$ is the Laplace transform of a probability distribution G_1 if and only if it is completely monotone and $\varphi_1(0) = 1$. Feller (1971) calls Bernstein's theorem a beautiful theorem. While it may be difficult to justify such an attribution objectively, the multivariate version of Bernstein's beautiful theorem that I present below is the critical step in demonstrating that McFadden's class of GEV models for discrete choice can be made isomorphic to the set of models that incorporate random effects in the Cox proportional hazard model with an unspecified baseline.

I define the (non-negative) *J*-variate function $\varphi_J(\lambda_1,\ldots,\lambda_J)$ to be completely monotone if for any non-negative vector of integers $(q_1,\ldots,q_J)$, the *Q*th partial, where $Q=\sum_{i=1}^{J} q_i$, of $\varphi_J(\lambda_1,\ldots,\lambda_J)$, $\partial^Q \varphi_J(\lambda_1,\ldots,\lambda_J)/\partial\lambda_1^{q_1}\ldots\partial\lambda_J^{q_J}$, is non-negative if *Q* is even and non-positive if *Q* is odd. It is important to note that this definition of multivariate complete monotonicity is not the same as the alternating partials property in Proposition 1. In Proposition 1, the function M^* has first partials which are positive, while for the completely monotone function, first partials are negative. However, M^* does have completely monotone first partials.

Theorem 3:

A function φ_J on $R^{J+}\bigcup\{\mathbf{0}\}$ is the Laplace transform of a *J*-variate distribution G_J if and only if it is completely monotone and $\varphi_J(\mathbf{0})=1$.

Proof:

Note first that if G_J is a *J*-variate probability distribution and φ_J is its Laplace transform, then $\varphi_J(\mathbf{0})=1$ and φ_J possesses partial derivatives of all orders. Furthermore, if

$Q=\sum_{i=1}^{J} q_i$,

$$(-1)^Q \partial^Q \varphi_J(\lambda_1,\ldots,\lambda_J)/\partial\lambda_1^{q_1}\ldots\partial\lambda_J^{q_J} = \int_{R^{J+}\bigcup\{\mathbf{0}\}} \exp(-\sum_{i=1}^{J}\lambda_i s_i)\left(\prod_{i=1}^{J} s_i^{q_i}\right) dG_J(s_1,\ldots,s_J) \geq 0 \quad . \tag{4.5}$$

I have therefore proved the "only if" part.

To prove the "if" part, assume $\varphi_J(s_1,\ldots,s_J)$ (with $\varphi_J(\mathbf{0})=1$) to be completely monotone and consider the substitution $s_i = n - ne^{-\lambda_i/n}, i=1,\ldots,J$, for fixed $n>0$ and positive λ_i. Define

(4.6) $$\varphi_J^n(\lambda_1,\ldots,\lambda_J)=\varphi_J(n-ne^{-\lambda_1/n},\ldots,n-ne^{-\lambda_J/n}).$$

Taylor-expanding the right-hand side of (4.6) around the *J*-vector equal to *n* for each component yields

(4.7) $$\begin{aligned}\varphi_J^n(\lambda_1,\ldots,\lambda_J) &= \sum_{Q=0}^{\infty}\left(\sum_{i=1}^{J}-ne^{-\lambda_i/n}\frac{\partial}{\partial s_i}\right)^Q \varphi_J(n,\ldots,n)\\ &= \sum_{q_1=0}^{\infty}\cdots\sum_{q_J=0}^{\infty}\frac{(-n)^Q}{Q!}\frac{\partial^Q\varphi_J(n,\ldots,n)}{\partial s_1^{q_1}\ldots\partial s_J^{q_J}}\prod_{i=1}^{J}e^{-q_i\lambda_i/n},\end{aligned}$$

where $\left(\sum_{i=1}^{J}-ne^{-\lambda_i/n}\frac{\partial}{\partial s_i}\right)^Q$ in the first line is the identity operator when $Q=0$, and in the second line, $Q=\sum_{i=1}^{J}q_i$. From the second line, $\varphi_J^n(\lambda_1,\ldots,\lambda_J)$ is the Laplace transform of a distribution attributing probability mass $\frac{(-n)^Q}{Q!}\frac{\partial^Q\varphi_J(n,\ldots,n)}{\partial s_1^{q_1}\ldots\partial s_J^{q_J}}$ to the point $(\frac{q_1}{n},\ldots,\frac{q_J}{n})$ (where for each $i=1,\ldots,J$, $q_i=0,1,2,\ldots$). Now $\varphi_J^n(\boldsymbol{\lambda})\to\varphi_J(\boldsymbol{\lambda})$ as $n\to\infty$. Therefore, by the Continuity Theorem and the fact that $\varphi_J(\mathbf{0})=1$, $\varphi_J(\boldsymbol{\lambda})$ is the Laplace transform of a non-defective distribution G_J.

I have now shown that e^{-M^*} is a sample survivor function consistent with a random-effects specification if and only if it is completely monotone. If, in addition, $M^*(\mathbf{0})\to 0$, the joint distribution of random effects $\Omega(\nu_1, \ldots, \nu_N \mid \mathbf{Z}, \boldsymbol{\beta})$ is non-defective. To prove Proposition 1, I now need to show that e^{-M^*} is completely monotone if M^* has completely monotone first partials. This again is the multivariate version of a theorem in Feller (1971).

The proof is by induction on *Q*. It is obviously true for $Q = 1$. Any *Q*th partial derivative is the sum of *i* terms of the form $\chi_i e^{-M^*}$, where χ_i is the product of integral positive powers of partials of M^*. Thus, any (*Q* + 1)th partial derivative is the sum of i terms of the form $\chi_i D_j e^{-M^*} + e^{-M^*} D_j \chi_i$, where D_j is the operator for the partial derivative with respect to the *j*th argument. Clearly, $\chi_i D_j e^{-M^*}$ is of opposite sign to $\chi_i e^{-M^*}$ because M^* has completely monotone first partials. On the other hand, $D_j \chi_i$ is evaluated by the chain rule, and is of opposite sign to χ_i, because taking a partial derivative of any integral positive power of a partial of M^* involves a sign change if M^* has completely monotone first partials. Proposition 1 is now proved.

5. PARTIAL LIKELIHOOD

Cox (1975) develops the partial likelihood method for inference in models containing a large, possibly infinite, number of the nuisance parameters. In the context of the proportional hazard model, the coefficient vector $\boldsymbol{\beta}$ represents the parameter of interest and the baseline hazard $h_0(t)$ is characterized by a (possibly infinite-dimensional) set of nuisance parameters labeled $\boldsymbol{\psi}$. When $\boldsymbol{\psi}$ is finite-dimensional and the form of the baseline hazard is known, it may be possible to construct the likelihood function and jointly maximize $\boldsymbol{\beta}$ and $\boldsymbol{\psi}$. In other situations it may be possible to condition on a sufficient statistic for $\boldsymbol{\psi}$ and use the resulting conditional distribution for inference about $\boldsymbol{\beta}$.

Unfortunately, when $\boldsymbol{\psi}$ is infinite-dimensional or when the likelihood function for $\boldsymbol{\beta}$ and $\boldsymbol{\psi}$ is complex, neither approach, joint maximization of the likelihood function with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\psi}$ or the computation of conditional distributions given a sufficient statistic, may be feasible. The method of partial likelihood attempts to overcome this obstacle by constructing the likelihood function and decomposing it into two parts.

Let $\mathbf{X}$ be a random vector with density $g_{\mathbf{X}}(\mathbf{x} \mid \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\psi})$. In the case of an analysis of spells, $\mathbf{X}$ might be the vector $(X_1, \ldots, X_N)$, where $X_i = \min(T_i, U_i)$, T_i is the failure time, U_i is an uninformative censoring time, and T_i and U_i are independent. Spell data for observation i are censored at age t if it is not known that $T_i = t$ and t is the greatest age for which it is known that $T_i \geq t$. In this case $U_i = t$. Censoring is uninformative if, at each age t, the probability that a spell is censored in $[t, t+dt)$, given $R(t)$ and $R(t+) = \bigcup_{h>0} R(t+h)$, does not depend on $\boldsymbol{\beta}$ (see Kalbfleisch and Prentice, 1980; Arjas and Haara, 1984; and Fleming and Harrington, 1991). Thus, censoring is informative whenever the distribution of U_i depends on $\boldsymbol{\beta}$, even if U_i and T_i are independent (Fleming and Harrington, 1991).

Fleming and Harrington (1991) motivate the idea underlying Cox's partial likelihood by pointing out that in some applications, the likelihood can be written as the product of conditional likelihood and marginal likelihood:

(5.1) $$g_{\mathbf{X}}(\mathbf{x} \mid \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\psi}) = g_{\mathbf{W}|\mathbf{V}}(\mathbf{w} \mid \mathbf{v}, \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\psi})\, g_{\mathbf{V}}(\mathbf{v} \mid \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\psi}),$$

where $\mathbf{x}' = (\mathbf{w}', \mathbf{v}')$. If one of the factors on the right-hand side of equation (5.1) does not depend on $\boldsymbol{\psi}$, then it can be used for inference about $\boldsymbol{\beta}$, with the simplification compensating for the loss of efficiency. Cox assumes that there exists a one-to-one transformation from $\mathbf{X}$ into $\mathbf{W}^{(N)}, \mathbf{V}^{(N)}$, where $\mathbf{W}^{(i)} = (\mathbf{W}_1, \ldots, \mathbf{W}_i)$ and $\mathbf{V}^{(i)} = (\mathbf{V}_1, \ldots, \mathbf{V}_i)$. Then:

(5.2) $$g_{\mathbf{X}}(\mathbf{x} \mid \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\psi}) = \prod_{i=1}^{N} g_{\mathbf{W}_i|\mathbf{W}^{(i-1)}, \mathbf{V}^{(i)}}(\mathbf{w}_i \mid \mathbf{w}^{(i-1)}, \mathbf{v}^{(i)}, \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\psi})$$
$$\bullet \prod_{i=1}^{N} g_{\mathbf{V}_i|\mathbf{W}^{(i-1)}, \mathbf{V}^{(i-1)}}(\mathbf{v}_i \mid \mathbf{w}^{(i-1)}, \mathbf{v}^{(i-1)}, \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\psi}),$$

where $\mathbf{W}^{(0)} = \mathbf{V}^{(0)} = \{\varnothing\}$. When the first product on the right-hand side of equation (5.2) does not depend on $\boldsymbol{\psi}$, Cox calls it the partial likelihood for $\boldsymbol{\beta}$ and suggests inference based on its maximization. Wong (1986) derives regularity conditions for the consistency and asymptotic normality of the partial likelihood estimator. In the context of duration analysis, $\mathbf{W}^{(i)}$ contains the sample information on failure times and $\mathbf{V}^{(i)}$ contains the sample censoring information. Note that when censoring is uninformative, the second product in equation (5.2) is unlikely to contain substantial information about $\boldsymbol{\beta}$. Fleming and Harrington (1991) provide the following construction of the partial likelihood for duration analysis.

Suppose there are L observed failure times:

(5.3) $$0 = T_0^o < T_1^o < \ldots < T_L^o < T_{L+1}^o = \infty,$$

and let (i) be the anti-rank, the label for the spell failing at T_i^o, i.e., $T_{(i)} = T_i^o$. Note that the covariate vectors for the L spells that fail are $\mathbf{Z}_{(1)}, \ldots, \mathbf{Z}_{(L)}$. Then:

$$\mathbf{W}_i = \{(i)\}. \tag{5.4}$$

Suppose further that there are n_i spells censored at or after T_i^o but before T_{i+1}^o, at the ordered times $T_{i1}^o, \ldots, T_{in_i}^o$. Let (i, j) be the label for the spell censored at T_{ij}^o, so that the covariate vectors associated with these n_i spells are $\mathbf{Z}_{(i,1)}, \ldots, \mathbf{Z}_{(i,n_i)}$. Then:

$$\mathbf{V}_i = \{T_{i+1}^o, \{T_{ij}^o, (i, j) \mid j = 1, \ldots, n_i\}\}. \tag{5.5}$$

The partial likelihood is:

$$\prod_{i=1}^{L} P(\mathbf{W}_i = \{(i)\} \mid \mathbf{W}^{(i-1)}, \mathbf{V}^{(i)}, \mathbf{Z}, \boldsymbol{\beta}). \tag{5.6}$$

This is the probability that spell (i) fails at $T_i^o = t_i$, given that there is exactly one failure at t_i and the risk set $R(t_i)$ survives to t_i:

(5.7)

$$\frac{P(T_{(i)} \in [t_i, t_i + dt) \mid \{T_l \geq t_i \mid l \in R(t_i)\}, \mathbf{Z}, \boldsymbol{\beta}) \prod_{k \in R(t_i) - \{(i)\}} P(T_k \notin [t_i, t_i + dt) \mid \{T_l \geq t_i \mid l \in R(t_i)\}, \mathbf{Z}, \boldsymbol{\beta})}{\sum_{j \in R(t_i)} P(T_j \in [t_i, t_i + dt) \mid \{T_l \geq t_i \mid l \in R(t_i)\}, \mathbf{Z}, \boldsymbol{\beta}) \prod_{k \in R(t_i) - \{j\}} P(T_k \notin [t_i, t_i + dt) \mid \{T_l \geq t_i \mid l \in R(t_i)\}, \mathbf{Z}, \boldsymbol{\beta})}.$$

When the N sample spells are independent,

$$P(\mathbf{W}_i = \{(i)\} \mid \mathbf{W}^{(i-1)}, \mathbf{V}^{(i)}, \mathbf{Z}, \boldsymbol{\beta}) = e^{\mathbf{Z}_{(i)}\boldsymbol{\beta}} / \sum_{j \in R(t_i)} e^{\mathbf{Z}_j \boldsymbol{\beta}}. \tag{5.8}$$

To prove that the partial likelihood estimator is baseline-free when the N sample spells are not statistically independent and the sample survivor function is given by equation (3.10), two lemmas are required. The first lemma describes the first partials of homogeneous functions.

Lemma 1:

If $M(\theta_1, \ldots, \theta_N)$ is homogeneous of degree *k*, then $M^{[i]}(\theta_1, \ldots, \theta_N) = \frac{\partial M}{\partial \theta_i}(\theta_1, \ldots, \theta_N)$ is homogeneous of degree *k*-1 for $i = 1, \ldots, N$.

Proof:

$$
\begin{aligned}
t^k M^{[i]}(\theta_1, \ldots, \theta_N) &= t^k \frac{\partial M}{\partial \theta_i}(\theta_1, \ldots, \theta_N) \\
&= \frac{\partial M}{\partial \theta_i}(t\theta_1, \ldots, t\theta_N) \\
&= \frac{\partial M}{\partial (t\theta_i)}(t\theta_1, \ldots, t\theta_N) \cdot \frac{d(t\theta_i)}{d\theta_i} \\
&= tM^{[i]}(t\theta_1, \ldots, t\theta_N).
\end{aligned}
$$

Hence, $t^{k-1} M^{[i]}(\theta_1, \ldots, \theta_N) = M^{[i]}(t\theta_1, \ldots, t\theta_N)$. *Q.E.D.*

The second lemma is frequently known as Euler's Theorem (see Friedman 1971).

Lemma 2 (Euler's Theorem):

If $M(\theta_1, \ldots, \theta_N)$ is homogeneous of degree *k*, then

$$kM(\theta_1, \ldots, \theta_N) = \sum_{i=1}^{N} \theta_i \frac{\partial M}{\partial \theta_i}(\theta_1, \ldots, \theta_N). \tag{5.9}$$

Proof:

Define the function $\bar{M}$ as follows:

(5.10) $$\bar{M}(\theta_1, \ldots, \theta_N, t) = t^{-k} M(t\theta_1, t\theta_2, \ldots, t\theta_N) .$$

Since M is homogeneous of degree k, $\bar{M}$ does not depend on t and $\frac{\partial \bar{M}}{\partial t} = 0$ for all $(\theta_1, \ldots, \theta_N)$ and all $t > 0$. Hence, we are done if we show that, when $t = 1$,

(5.11) $$\frac{\partial \bar{M}}{\partial t} = kM(\theta_1, \ldots, \theta_N) - \sum_{i=1}^{N} \theta_i \frac{\partial M}{\partial \theta_i}(\theta_1, \ldots, \theta_N) .$$

Applying the product rule of differentiation to the right-hand side of (5.10) yields:

(5.12) $$\frac{\partial \bar{M}}{\partial t} = kt^{-k-1} M(t\theta_1, \ldots, t\theta_N) - \sum_{i=1}^{N} \frac{\partial M}{\partial t\theta_i}(t\theta_1, \ldots, t\theta_N) \cdot \frac{dt\theta_i}{dt} \quad .$$

Setting $t = 1$ gives the desired result. *Q.E.D.*

Now note that in the statistically dependent case, the *j*th term in the denominator of equation (5.7) can be written as the difference of two conditional survivor functions:

(5.13)

$$S_{R(t_i)}((t_i + dt)\boldsymbol{\iota} - (dt)\boldsymbol{\iota}_j \mid \{T_l \geq t_i \mid l \in R(t_i)\}, \mathbf{Z}, \boldsymbol{\beta}) - S_{R(t_i)}((t_i + dt)\boldsymbol{\iota} \mid \{T_l \geq t_i \mid l \in R(t_i)\}, \mathbf{Z}, \boldsymbol{\beta}),$$

where $\boldsymbol{\iota}_j$ is the *j*th row of the N x N identity matrix. The first conditional survivor function in (5.13) is equal to

(5.14) $$S_{R(t_i)}((t_i + dt)\boldsymbol{\iota} - (dt)\boldsymbol{\iota}_j \mid \mathbf{Z}, \boldsymbol{\beta}) / S_{R(t_i)}(t_i\boldsymbol{\iota} \mid \mathbf{Z}, \boldsymbol{\beta}) ,$$

while the second equals

(5.15) $$S_{R(t_i)}((t_i + dt)\boldsymbol{\iota} \mid \mathbf{Z}, \boldsymbol{\beta}) / S_{R(t_i)}(t_i\boldsymbol{\iota} \mid \mathbf{Z}, \boldsymbol{\beta}) .$$

The difference therefore equals

(5.16) $$-S^{[j]}_{R(t_i)}(t_i\boldsymbol{\iota} \mid \mathbf{Z}, \boldsymbol{\beta}) dt / S_{R(t_i)}(t_i\boldsymbol{\iota} \mid \mathbf{Z}, \boldsymbol{\beta}) ,$$

where $S^{[j]}_{R(t_i)}$ represents the partial derivative of the survivor function with respect to its *j*th argument. Since

(5.17) $$S_{R(t_i)}(\mathbf{u} \mid \mathbf{Z}, \boldsymbol{\beta}) = \exp(-M(Y_1^{R(t_i)}\theta_1 H_0(u_1), \ldots, Y_N^{R(t_i)}\theta_N H_0(u_N))) \, ,$$

the derivative equals

(5.18)

$$S^{[j]}_{R(t_i)}(\mathbf{u} \mid \mathbf{Z}, \boldsymbol{\beta}) = -Y_j^{R(t_i)}\theta_j h_0(u_j) M^{[j]}(Y_1^{R(t_i)}\theta_1 H_0(u_1), \ldots, Y_N^{R(t_i)}\theta_N H_0(u_N)) S_{R(t_i)}(\mathbf{u} \mid \mathbf{Z}, \boldsymbol{\beta}) \, .$$

Because $M^{[j]}$ is homogeneous of degree zero:

(5.19) $$S^{[j]}_{R(t_i)}(t_i \boldsymbol{\iota} \mid \mathbf{Z}, \boldsymbol{\beta}) = -Y_j^{R(t_i)}\theta_j h_0(t_i) M^{[j]}(Y_1^{R(t_i)}\theta_1, \ldots, Y_N^{R(t_i)}\theta_N) S_{R(t_i)}(t_i \boldsymbol{\iota} \mid \mathbf{Z}, \boldsymbol{\beta})$$

by Lemma 1, and (5.13) becomes

(5.20)

$$-S^{[j]}_{R(t_i)}(t_i \boldsymbol{\iota} \mid R(t_i), \mathbf{Z}, \boldsymbol{\beta}) dt \,/\, S_{R(t_i)}(t_i \boldsymbol{\iota} \mid \mathbf{Z}, \boldsymbol{\beta}) = Y_j^{R(t_i)}\theta_j h_0(t_i) M^{[j]}(Y_1^{R(t_i)}\theta_1, \ldots, Y_N^{R(t_i)}\theta_N) dt \, ,$$

which represents the *j*th term in the denominator of (5.7). Expression (5.7) becomes

(5.21) $$P(\mathbf{W}_i = \{(i)\} \mid \mathbf{W}^{(i-1)}, \mathbf{V}^{(i)}, \mathbf{Z}, \boldsymbol{\beta}) = \frac{Y_{(i)}^{R(t_i)}\theta_{(i)} \dfrac{\partial M}{\partial \theta_{(i)}}(\boldsymbol{\theta}, \boldsymbol{\iota}, R(t_i))}{\sum_{j=1}^{N} Y_j^{R(t_i)}\theta_j M^{[j]}(\boldsymbol{\theta}, \boldsymbol{\iota}, R(t_i))} \, ,$$

since the presence of the set-inclusion indicators Y permits the summation from 1 to N in the denominator. Now, Euler's Theorem, the fact that M is homogeneous of degree one, and the fact that spell (i) is in $R(t_i)$ allow the simplification:

(5.22) $$P(\mathbf{W}_i = \{(i)\} \mid \mathbf{W}^{(i-1)}, \mathbf{V}^{(i)}, \mathbf{Z}, \boldsymbol{\beta}) = \frac{\theta_{(i)} \dfrac{\partial M}{\partial \theta_{(i)}}(\boldsymbol{\theta}, \boldsymbol{\iota}, R(t_i))}{M(\boldsymbol{\theta}, \boldsymbol{\iota}, R(t_i))} \, ,$$

giving the same form as the probability in (3.15). The proof of Proposition 2 for the partial likelihood is complete.

6. RANK-INFORMATION MARGINAL LIKELIHOOD

When censoring is uninformative and failure times and censoring times are statistically independent, it seems reasonable to conclude that the second product in equation (5.2) provides little information about $\boldsymbol{\beta}$, and maximizing the first product with respect to $\boldsymbol{\beta}$ will yield an estimator close to the maximum likelihood estimator. Unfortunately, no formal proof of this has ever been provided. The discovery that maximization of the rank-information marginal likelihood yields the partial likelihood estimator when spells are independent was important, because the marginal likelihood function is a proper likelihood function to which the usual asymptotic theory of maximum likelihood directly applies. In this section it will be shown that whenever the joint survivor function has the form specified in Proposition 2, the partial likelihood and rank-information marginal likelihood estimators are identical.

Initially, it is assumed that the sample spells are uncensored. Let $T_i,\ i=1,\ldots,N$, represent the failure times of the N sample spells. Further, let $T_0^o < T_1^o < \ldots < T_N^o$ be the ordered failure times and let (i) denote the anti-rank, the label of the spell failing at T_i^o. Construct two vectors, $\mathbf{O} = (T_1^o, \ldots, T_N^o)$, the vector of order statistics, and $\mathbf{r} = ((1), \ldots, (N))$, the vector of rank statistics. Note that the vector of sample failure times, $\mathbf{T} = (T_1, \ldots, T_N)$ can be reconstructed from knowledge of $\mathbf{O}$ and $\mathbf{r}$.

Kalbfleisch and Prentice (1980) present an example in which $N = 4$ and $\mathbf{T} = (5,17,12,15)$. The vector of order statistics for this data is $\mathbf{O} = (5,12,15,17)$ and the vector of rank statistics is $\mathbf{r} = (1,3,4,2)$. If the value of the jth component of $\mathbf{r}$ equals i, then T_i is the jth smallest sample failure time, with value given by the jth component of $\mathbf{O}$.

The fact that the vector of rank statistics carries the sample information about $\boldsymbol{\beta}$ when the baseline hazard rate h_0 is completely unspecified can be demonstrated by a simple argument. The hazard rate for duration i, T_i, in the proportional hazard model was specified in equation (3.5) by $h_i(t \mid \mathbf{Z}, \boldsymbol{\beta}) = \exp(\mathbf{Z}_i \boldsymbol{\beta}) h_0(t)$. For all i, define $V_i = g^{-1}(T_i)$, where g is an arbitrary element of G, the group of differentiable and strictly increasing transformations of $(0, \infty)$ into $(0, \infty)$. Then, given $\mathbf{Z}$ and $\boldsymbol{\beta}$, the hazard rate for V_i is given by $\exp(\mathbf{Z}_i \boldsymbol{\beta}) h_0^*(v)$, where $h_0^*(v) = h_0(g(v)) g'(v)$. This shows that when the baseline hazard is unspecified, the vector of order statistics can be modified arbitrarily as long as the vector of rank statistics is unchanged, and the problem of inference about $\boldsymbol{\beta}$ has not changed. The estimation problem for $\boldsymbol{\beta}$, given an unspecified baseline, is invariant to (continuous) monotonic transformations of duration.

The estimation of $\boldsymbol{\beta}$ can therefore be based on the rank-information marginal likelihood, i.e., the marginal likelihood of $\mathbf{r}$. As in the discussion of the partial likelihood, sample values of the random ordered failure times are $(T_1^o, \ldots, T_N^o) = (t_1, \ldots, t_N)$. When sample spells are independent and, the marginal likelihood of $\mathbf{r}$ is given by

$$
\begin{aligned}
(6.1) \qquad P(\mathbf{r} = ((1), \ldots, (N)) \mid \mathbf{Z}, \boldsymbol{\beta}) &= P(T_{(1)} < \ldots < T_{(N)} \mid \mathbf{Z}, \boldsymbol{\beta}) \\
&= \int_0^\infty \int_{t_1}^\infty \cdots \int_{t_{N-1}}^\infty \prod_{i=1}^{N} f(t_i \mid \mathbf{Z}_{(i)}, \boldsymbol{\beta}) dt_N \cdots dt_1 .
\end{aligned}
$$

When sample spells are dependent, the density for T_i must also be conditioned on A_i, the event $\{T_{(j)} > t_i \mid j = i+1, \ldots, N\}$, where A_N is the null event. Therefore, in the case of dependent spells,

$$(6.2) \qquad P(\mathbf{r} = ((1), \ldots, (N)) \mid \mathbf{Z}, \boldsymbol{\beta}) = P(T_{(1)} < \ldots < T_{(N)} \mid \mathbf{Z}, \boldsymbol{\beta})$$

$$= \int_0^\infty \int_{t_1}^\infty \cdots \int_{t_{N-1}}^\infty \prod_{i=1}^N f(t_i \mid A_i, \mathbf{Z}_{(i)}, \boldsymbol{\beta}) dt_N \cdots dt_1 \,.$$

The multiple integral in equation (6.2) is evaluated recursively, as given by:

(6.3)

$$\int_0^\infty f(t_1 \mid A_1, \mathbf{Z}_{(1)}, \boldsymbol{\beta}) \left[\cdots \left[\int_{t_{N-2}}^\infty f(t_{N-1} \mid A_{N-1}, \mathbf{Z}_{(N-1)}, \boldsymbol{\beta}) \left[\int_{t_{N-1}}^\infty f(t_N \mid A_N, \mathbf{Z}_{(N)}, \boldsymbol{\beta}) dt_N \right] dt_{N-1} \right] \cdots \right] dt_1 \,.$$

It is required to prove that (6.3) equals

$$(6.4) \qquad \prod_{i=1}^N \frac{\theta_{(i)} M^{[(i)]}(\boldsymbol{\theta}, \boldsymbol{\iota}, R(t_i))}{M(\boldsymbol{\theta}, \boldsymbol{\iota}, R(t_i))},$$

where the superscript $[(i)]$ denotes the partial derivative with respect to the argument given by the anti-rank (i). I will prove that the rank-information marginal likelihood equals

$$(6.5) \qquad \left[\prod_{i=1}^N \frac{\theta_{(i)} M^{[(i)]}(\boldsymbol{\theta}, \boldsymbol{\iota}, R(t_i))}{M(\boldsymbol{\theta}, \boldsymbol{\iota}, R(t_i))} \right] \left(S_0(0) \right)^{M(\boldsymbol{\theta}, \boldsymbol{\iota}, R(0))},$$

where $S_0(t)$ is the baseline survivor function $\exp(-H_0(t))$. The result follows from the fact that $S_0(0) = 1$.

The proof is by induction on the number of integrations performed. Because A_N is the null event, the first integration is simply the probability that the (N)th spell survives to t_{N-1}:

(6.6) $$\left[\frac{\theta_{(N)} M^{[(N)]}(\boldsymbol{\theta}, \boldsymbol{\iota}, R(t_N))}{M(\boldsymbol{\theta}, \boldsymbol{\iota}, R(t_N))}\right]\left(S_0(t_{N-1})\right)^{M(\boldsymbol{\theta}, \boldsymbol{\iota}, R(t_N))}.$$

Note that the expression in brackets equals one by virtue of Euler's theorem and the fact that only the final one of the N arguments of $M(\theta_1, \ldots, \theta_N)$ is nonzero when the risk set is $R(t_N)$.

The induction hypothesis is that the result for the first j integrations is:

(6.7) $$\left[\prod_{i=N-j+1}^{N} \frac{\theta_{(i)} M^{[(i)]}(\boldsymbol{\theta}, \boldsymbol{\iota}, R(t_i))}{M(\boldsymbol{\theta}, \boldsymbol{\iota}, R(t_i))}\right]\left(S_0(t_{N-j})\right)^{M(\boldsymbol{\theta}, \boldsymbol{\iota}, R(t_{N-j+1}))}.$$

The proof is complete if I show that the result after $j+1$ integrations is:

(6.8) $$\left[\prod_{i=N-j}^{N} \frac{\theta_{(i)} M^{[(i)]}(\boldsymbol{\theta}, \boldsymbol{\iota}, R(t_i))}{M(\boldsymbol{\theta}, \boldsymbol{\iota}, R(t_i))}\right]\left(S_0(t_{N-j-1})\right)^{M(\boldsymbol{\theta}, \boldsymbol{\iota}, R(t_{N-j}))}.$$

Therefore, it must be shown that:

(6.9) $$\int_{t_{N-j-1}}^{\infty} f(t_{N-j} \mid A_{N-j}, \mathbf{Z}_{(N-j)}, \boldsymbol{\beta})\left(S_0(t_{N-j})\right)^{M(\boldsymbol{\theta}, \boldsymbol{\iota}, R(t_{N-j+1}))} dt_{N-j}$$
$$= \frac{\theta_{(N-j)} M^{[(N-j)]}(\boldsymbol{\theta}, \boldsymbol{\iota}, R(t_{N-j}))}{M(\boldsymbol{\theta}, \boldsymbol{\iota}, R(t_{N-j}))}\left(S_0(t_{N-j-1})\right)^{M(\boldsymbol{\theta}, \boldsymbol{\iota}, R(t_{N-j}))}.$$

The first task is to evaluate $f(t_{N-j} \mid A_{N-j}, \mathbf{Z}_{(N-j)}, \boldsymbol{\beta})$. Note that the probability that spell i survives to u_i given that spell j survives to u_j for $j \neq i$ is given by:

(6.10) $$\frac{S(u_1, \ldots, u_{i-1}, u_i, u_{i+1}, \ldots, u_N \mid \mathbf{Z}, \boldsymbol{\beta})}{S(u_1, \ldots, u_{i-1}, 0, u_{i+1}, \ldots, u_N \mid \mathbf{Z}, \boldsymbol{\beta})}.$$

Hence, the probability that spell $(N-j)$ survives to t_{N-j} given that the remaining spells in risk set $R(t_{N-j})$ exceed t_{N-j} is given by:

(6.11) $$S_{R(t_{N-j})}(t_{N-j}\boldsymbol{\iota} \mid \mathbf{Z},\boldsymbol{\beta}) / S_{R(t_{N-j+1})}(t_{N-j}\boldsymbol{\iota} \mid \mathbf{Z},\boldsymbol{\beta}).$$

The density $f(t_{N-j} \mid A_{N-j}, \mathbf{Z}_{(N-j)}, \boldsymbol{\beta})$ is obtained by deriving the numerator in (6.10) with respect to argument $(N-j)$ and changing the sign:

(6.12) $$f(t_{N-j} \mid A_{N-j}, \mathbf{Z}_{(N-j)}, \boldsymbol{\beta}) = -\frac{S_{R(t_{N-j})}^{[(N-j)]}(t_{N-j}\boldsymbol{\iota} \mid \mathbf{Z},\boldsymbol{\beta})}{S_{R(t_{N-j+1})}(t_{N-j}\boldsymbol{\iota} \mid \mathbf{Z},\boldsymbol{\beta})}.$$

Since the denominator on the right-hand side of (6.12) equals $\left(S_0(t_{N-j})\right)^{M(\boldsymbol{\theta},\boldsymbol{\iota},R(t_{N-j+1}))}$, the integral in (6.9) equals

(6.13) $$-\int_{t_{N-j-1}}^{\infty} S_{R(t_{N-j})}^{[(N-j)]}(t_{N-j}\boldsymbol{\iota} \mid \mathbf{Z},\boldsymbol{\beta})dt_{N-j}.$$

The partial derivative inside the integral of (6.13) equals

(6.14) $$\theta_{(N-j)}M^{[(N-j)]}(\boldsymbol{\theta},\boldsymbol{\iota},R(t_{N-j}))h_0(t_{N-j})\exp(-H_0(t_{N-j})M(\boldsymbol{\theta},\boldsymbol{\iota},R(t_{N-j}))).$$

Substituting (6.14) into (6.13), multiplying inside the integral by $M(\boldsymbol{\theta},\boldsymbol{\iota},R(t_{N-j}))$ and outside the integral by its reciprocal yields

(6.15)

$$-\frac{\theta_{(N-j)}M^{[(N-j)]}(\boldsymbol{\theta},\boldsymbol{\iota},R(t_{N-j}))}{M(\boldsymbol{\theta},\boldsymbol{\iota},R(t_{N-j}))}\int_{t_{N-j-1}}^{\infty} h_0(t_{N-j})M(\boldsymbol{\theta},\boldsymbol{\iota},R(t_{N-j}))\exp(-H_0(t_{N-j})M(\boldsymbol{\theta},\boldsymbol{\iota},R(t_{N-j})))dt_{N-j}.$$

The integrand in (6.15) equals

(6.16) $$\frac{d\left(S_0(t_{N-j})\right)^{M(\boldsymbol{\theta},\boldsymbol{\iota},R(t_{N-j}))}}{dt_{N-j}},$$

and therefore the integral in (6.15) equals $-\left(S_0(t_{N-j-1})\right)^{M(\boldsymbol{\theta},\boldsymbol{\iota},R(t_{N-j}))}$. Substituting this expression into (6.15) yields the right-hand side of equation (6.9). The proof is complete for the case of no censoring.

When sample spells can be censored, the data vector for the ith spell is $(X_i, \delta_i, \mathbf{Z}_i)$, where again $X_i = \min(T_i, U_i)$ for uninformative censoring time U_i independent of T_i, and δ_i is the censoring indicator equal to one when $X_i = U_i < T_i$ and zero otherwise. Let $X_1^o < \ldots < X_N^o$ represent the ordered observation times, and define $\mathbf{O}^* = (X_1^o, \ldots, X_N^o)$. Let $\mathbf{r}^* = ((1)^*, \ldots, (N)^*)$ denote the vector of corresponding anti-ranks, and let $\boldsymbol{\delta}^* = (\delta_{(1)^*}, \ldots, \delta_{(N)^*})$ denote the vector of ordered censoring indicators. Just as in the uncensored case where $\mathbf{T} = (T_1, \ldots, T_N)$ can be reconstructed from knowledge of $(\mathbf{O}, \mathbf{r})$, here in the case where spells may be censored $(\mathbf{X}, \boldsymbol{\delta})$, where $\mathbf{X} = (X_1, \ldots, X_N)$ and $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_N)$, can be reconstructed from knowledge of $(\mathbf{O}^*, \mathbf{r}^*, \boldsymbol{\delta}^*)$. As an example, suppose $\mathbf{X} = (5, 17, 12, 15)$ and $\boldsymbol{\delta} = (0, 1, 1, 0)$. Then, $\mathbf{O}^* = (5, 12, 15, 17)$, $\mathbf{r}^* = (1, 3, 4, 2)$, and $\boldsymbol{\delta}^* = (0, 1, 0, 1)$. If the value of the jth component of $\mathbf{r}^*$ equals i, then X_i is the jth smallest sample failure time, with value given by the jth component of $\mathbf{O}^*$. Similarly, the ith component of $\boldsymbol{\delta}$ equals one if and only if the value of the jth component of $\mathbf{r}^*$ equals i and the value of the jth component of $\boldsymbol{\delta}^*$ equals one. The value of the ith component of $\boldsymbol{\delta}$ equals zero otherwise.

Kalbfleisch and Prentice (1980) explain that some modification to the rank-information marginal likelihood is necessary in the presence of general uninformative independent censoring. The censored model will not in general possess the group invariance properties. When censoring occurs in the sample, the rank information is incomplete. In the example above, the vector of rank statistics, $\mathbf{r}$, is known to be either $(1,3,4,2)$, $(1,4,3,2)$, or $(1,4,2,3)$; more generally, it seems reasonable to estimate $\boldsymbol{\beta}$ using the marginal likelihood that the vector of rank statistics is one of those observationally possible. Doing so ignores the exact time of censoring, but the invariance property of the uncensored model demonstrates that the time between failures is irrelevant. Therefore, in a model with L failures, the marginal likelihood in (6.2) is adjusted as follows:

$$(6.17) \qquad P(\mathbf{r} = ((1),\ldots,(L)) \mid \mathbf{Z},\boldsymbol{\beta}) = P(T_{(1)} < \ldots < T_{(L)} \mid \mathbf{Z},\boldsymbol{\beta})$$

$$= \int_0^\infty \int_{t_1}^\infty \cdots \int_{t_{L-1}}^\infty \prod_{i=1}^{L} f(t_i \mid A_i, \mathbf{Z}_{(i)}, \boldsymbol{\beta}) dt_L \cdots dt_1 \,.$$

It is clear from the demonstration in the uncensored case that the marginal likelihood equals:

$$(6.18) \qquad \prod_{i=1}^{L} \frac{\theta_{(i)} M^{[(i)]}(\boldsymbol{\theta}, \boldsymbol{\iota}, R(t_i))}{M(\boldsymbol{\theta}, \boldsymbol{\iota}, R(t_i))} \quad .$$

The proof of Proposition 2 for the rank-information marginal likelihood in the presence of censoring is complete.

7. TIES IN THE DATA

Although durations are continuous, the recording of durations will always involve some measurement error, and ties may result. This is problematic because both the partial likelihood and rank-information marginal likelihood require the data to be completely rank-ordered. To incorporate tied data into the analysis, the same approach can be used as in the case of censoring.

Suppose that there are m_i spells ($m_i \geq 1$) at each of the L ordered observed failure times, t_i, where $\sum_{i=1}^{L} m_i = N$. Assuming the ties to result from the grouping of durations in the continuous model, the information available on the rank vector is incomplete. While it is known that the ranks of spells failing at t_i are less than failing at t_j whenever $i < j$, the ranks of the m_i spells failing at t_i cannot be known. The rank-information marginal likelihood in this case should specify the likelihood that the rank vector is one of those possible.

In their discussion of the case of independent spells, for which $M(\theta_1, \ldots, \theta_N) = \sum_{i=1}^{N} \theta_i$, Kalbfleisch and Prentice (1980) point out that the calculation can be simplified somewhat by recognizing that the ranks assigned to the m_i spells failing at t_i do not depend on the ranks assigned to the m_j spells failing at t_j. The sum then becomes the product of L weighted sums. Let Ξ_i be the set of permutations of the labels of the m_i spells failing at t_i and let $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_{m_i})$ be an element of Ξ_i. As before, $R(t_i)$ is the risk set at time t_i. Define $R(t_i, \xi^r)$ to be the set difference $R(t_i) - \{\xi_1, \ldots, \xi_{r-1}\}$ and $D(t_i) = R(t_i) - R(t_i+)$ to be the set of spells failing at t_i.

Then, the marginal likelihood for $\boldsymbol{\beta}$ can be expressed as

$$\text{(7.1)} \qquad \prod_{i=1}^{L} \left(\prod_{j \in D(t_i)} \theta_j \sum_{\boldsymbol{\xi} \in \Xi_i} \left(\prod_{r=1}^{m_i} \left[\sum_{l \in R(t_i, \xi^r)} \theta_l \right]^{-1} \right) \right) .$$

Because the summation in (7.1) is over all permutations of labels of the tied spells, the computation of (7.1) may be burdensome if there are a large number of ties at any failure time. When the number of spells failing at each t_i is small relative to the number of spells in the corresponding risk set $R(t_i)$, Peto (1972) and Breslow (1974) claim that (7.1) can be approximated using

$$\prod_{i=1}^{L}\left(\frac{\prod_{j\in D(t_i)}\theta_j}{\left(\sum_{l\in R(t_i)}\theta_l\right)^{m_i}}\right) \quad . \tag{7.2}$$

Efron (1977) suggests an alternative approximation to (7.1) that takes into account that distinct summations $\sum_{l\in R(t_i,\xi^r)}\theta_l$ in (7.1) will have greater multiplicity the lower is the value of r:

$$\prod_{i=1}^{L}\left(\frac{\prod_{j\in D(t_i)}\theta_j}{\prod_{r=1}^{m_i}\left(\left(\sum_{l\in R(t_i)}\theta_l\right)-\frac{(r-1)}{m_i}\left(\sum_{l\in D(t_i)}\theta_l\right)\right)}\right) \quad . \tag{7.3}$$

Kalbfleisch and Prentice (1980) suggest using a semi-parametric model formed by grouping failure times whenever the ratio of m_i to the size of the risk set $R(t_i)$ is high for any failure time (see Prentice and Gloeckler 1978, and Meyer 1990).

In the more general case in which sample spells are dependent, the marginal likelihood for $\boldsymbol{\beta}$ becomes

$$\prod_{i=1}^{L}\left(\prod_{j\in D(t_i)}\theta_j M^{[j]}(\boldsymbol{\theta},\boldsymbol{\iota},R(t_i))\sum_{\xi\in\Xi_i}\left(\prod_{r=1}^{m_i}\left[M(\boldsymbol{\theta},\boldsymbol{\iota},R(t_i,\xi^r))\right]^{-1}\right)\right). \tag{7.4}$$

When the number of spells failing at each t_i is small relative to the number of spells in the corresponding risk set $R(t_i)$, (7.4) can be approximated using

(7.5) $$\prod_{i=1}^{L}\left(\frac{\prod_{j\in D(t_i)}\theta_j M^{[j]}(\boldsymbol{\theta},\boldsymbol{\iota},R(t_i))}{\left(M(\boldsymbol{\theta},\boldsymbol{\iota},R(t_i))\right)^{m_i}}\right) .$$

Finally, the following alternative approximation to (7.4) takes into account that distinct summations $M(\boldsymbol{\theta},\boldsymbol{\iota},R(t_i,\xi^r))$ in (7.4) will have greater multiplicity the lower is the value of r:

(7.6) $$\prod_{i=1}^{L}\left(\frac{\prod_{j\in D(t_i)}\theta_j M^{[j]}(\boldsymbol{\theta},\boldsymbol{\iota},R(t_i))}{\prod_{r=1}^{m_i}\left(M(\boldsymbol{\theta},\boldsymbol{\iota},R(t_i))-\frac{(r-1)}{m_i}\left(\sum_{l\in D(t_i)}\theta_l M^{[l]}(\boldsymbol{\theta},\boldsymbol{\iota},R(t_i))\right)\right)}\right) .$$

8. RECOVERING THE BASELINE

Breslow (1972) develops a methodology for recovering the duration baseline from the partial likelihood estimates for a sample of independent spells. Breslow explains that the Kaplan-Meier estimate can be derived in a maximum likelihood framework by assuming that the hazard is constant between successive observed failure times:

(8.1) $$h_0(t)=\rho_i\ ,\quad t_{i-1}<t\le t_i\ ,\quad i=1,\ldots,L\ .$$

He notes that this approach is used by Grenander (1956) to derive maximum likelihood estimates for the monotone hazard class. Breslow next adopts the convention of considering all censored spells as censored at the previous uncensored failure time. Breslow’s estimator for λ_i is the maximum likelihood estimator for the resulting likelihood (see Kalbfleisch and Prentice 1980):

(8.2) $$\prod_{i=1}^{L}h_0(t_i)^{m_i}\left(\prod_{j\in D(t_i)}\theta_j\right)\exp(-\int_0^{t_i}h_0(u)du\sum_{j\in\Omega(t_i)}\theta_j)\quad ,$$

where $\Omega(t_i)$ is the set of spells either failing or censored at t_i. Substituting in from (8.1) and rearranging terms gives

$$\prod_{i=1}^{L} \rho_i^{m_i} \left(\prod_{j \in D(t_i)} \theta_j \right) \exp(-\rho_i (t_i - t_{i-1}) \sum_{j \in R(t_i)} \theta_j) \quad . \tag{8.3}$$

Since $\theta_j = \exp(\mathbf{Z}_j \boldsymbol{\beta})$, the maximum likelihood estimator of ρ_i for any value of $\boldsymbol{\beta}$ is therefore

$$\hat{\rho}_i = \frac{m_i}{(t_i - t_{i-1}) \sum_{j \in R(t_i)} \exp(\mathbf{Z}_j \boldsymbol{\beta})} \quad , \tag{8.4}$$

and the estimate of the cumulative baseline hazard $H_0(t) = \int_0^t h_0(u)du$, evaluated at t_i, is

$$\hat{H}_0(t_i) = \sum_{l=1}^{i} \frac{m_l}{\sum_{j \in R(t_l)} \exp(\mathbf{Z}_j \boldsymbol{\beta})} \quad . \tag{8.5}$$

The estimators in (8.4) and (8.5) can both be evaluated at the value of $\boldsymbol{\beta}$ that maximizes the rank-information marginal likelihood (corrected for ties).

When spells are dependent, the likelihood becomes

$$\prod_{i=1}^{L} \rho_i^{m_i} \left(\prod_{j \in D(t_i)} \theta_j M^{[j]}(\boldsymbol{\theta}, \boldsymbol{\iota}, R(t_i)) \right) \exp(-\rho_i (t_i - t_{i-1}) M(\boldsymbol{\theta}, \boldsymbol{\iota}, R(t_i))) \quad . \tag{8.6}$$

The maximum likelihood estimator of ρ_i for any value of $\boldsymbol{\beta}$ is

$$\hat{\rho}_i = \frac{m_i}{(t_i - t_{i-1}) M(\boldsymbol{\theta}, \boldsymbol{\iota}, R(t_i))} \quad , \tag{8.7}$$

and the estimate of the cumulative baseline hazard evaluated at t_i is

$$\hat{H}_0(t_i) = \sum_{l=1}^{i} \frac{m_l}{M(\boldsymbol{\theta}, \boldsymbol{\iota}, R(t_l))} \quad . \tag{8.8}$$

9. THE CONSTRUCTION OF LOG-RANK STATISTICS WHEN OBSERVATIONS ARE DEPENDENT

It is frequently important to determine whether two or more samples have been drawn from populations with identical survivor functions. If the available data are to be used efficiently in such a determination, the attempt should be made to construct a statistical test that summarizes differences in the survivor functions over the entire sample period and not just at a point in time. One of the first tests to do this with uncensored data was the log-rank test. The subsequent discovery that the log-rank test can be derived from score function tests based on the marginal and partial (log-) likelihoods led to more general tests that allowed for censoring. All of this work was in the context of independent spells. When spells are dependent, the new marginal and partial likelihood estimators described in sections 5 and 6 can be used to develop log-rank tests. The development of these log-rank tests is the subject of this section.

A convenient starting point is a review of the construction of log-rank tests from score statistics for the marginal and partial likelihoods when spells are independent. The presentation follows closely the analysis by Kalbfleisch and Prentice (1980).

The first step is the derivation of the score vector for the parameter vector $\boldsymbol{\beta}$. There are four cases to consider. In the first case, there are no ties and no censoring. In this case, for a sample of L spells, the score vector for $\boldsymbol{\beta}$ is given by

(9.1) $$\mathbf{U}_{\boldsymbol{\beta}}(\boldsymbol{\beta}) = \sum_{i=1}^{L}\left(\mathbf{Z}_{(i)} - \frac{\sum_{j=i}^{L}\mathbf{Z}_{(j)}\theta_{(j)}}{\sum_{j=i}^{L}\theta_{(j)}}\right).$$

(In preceding sections, covariate vectors $\mathbf{Z}$ are row vectors to avoid the necessity of writing transposes. Starting with this section and continuing through the appendix, covariate vectors are columns.)

The second case continues to assume no ties but allows for independent censoring. There are several ways to write the score vector in this case. Perhaps the simplest uses the indicators $Y_j^{R(t_i)}$ that state whether spell j is in the risk set $R(t_i)$. In a sample of N spells, $N-L$ of which are censored, the score vector can be written as

(9.2) $$\mathbf{U}_{\boldsymbol{\beta}}(\boldsymbol{\beta}) = \sum_{i=1}^{L}\left(\mathbf{Z}_{(i)} - \frac{\sum_{j=1}^{N}Y_j^{R(t_i)}\mathbf{Z}_j\theta_j}{\sum_{j=i}^{N}Y_j^{R(t_i)}\theta_j}\right).$$

Another way to write the score vector for this case uses the labels $(i,1)$ through (i,n_j) to denote the ordered censored spells that are in risk set $R(t_i)$ but not in $R(t_{i+1})$. Using these labels, the score vector can be written as

(9.3) $$\mathbf{U}_{\boldsymbol{\beta}}(\boldsymbol{\beta}) = \sum_{i=1}^{L}\left(\mathbf{Z}_{(i)} - \frac{\sum_{j=i}^{L}(\mathbf{Z}_{(j)}\theta_{(j)} + \sum_{l=1}^{n_i}\mathbf{Z}_{(j,l)}\,\theta_{(j,l)}}{\sum_{j=i}^{L}(\theta_{(j)} + \sum_{l=1}^{n_i}\theta_{(j,l)})}\right).$$

It might seem that the analysis is problematic when ties occur in the data. Computing the partial likelihood estimator requires the evaluation of all permutations of possible (strict) orderings of the sample durations given the observed data. Moreover, the Breslow and Efron solutions are only approximations to the true partial likelihood that may not be appropriate to the calculation of the log-rank statistic. It turns out that the simplified covariate vectors required for the log-rank statistic ensure that the score vector for $\boldsymbol{\beta}$ from the true partial likelihood can always be calculated. Nonetheless, Kalbfleisch and Prentice find it insightful to present the score vector for $\boldsymbol{\beta}$ from the Breslow solution and I shall do the same.

The third case allows ties but not censoring. I assume that there are L distinct failure times $t_1, \ldots, t_L$ and m_i failures at time t_i. For $j = 1, \ldots, m_i$, define $\mathbf{Z}_{(i)}^{j}$ to be covariate vector for the j the failure at t_i, $\mathbf{Z}_{(i)S} = \sum_{j=1}^{m_i} \mathbf{Z}_{(i)}^{j}$ and $\theta_{(i)}^{j} = \exp\{-\mathbf{Z}_{(i)}^{j}{}'\boldsymbol{\beta}\}$. Then the score vector in the third case can be written as

$$\text{(9.4)} \qquad \mathbf{U}_{\boldsymbol{\beta}}(\boldsymbol{\beta}) = \sum_{i=1}^{L} \left(\mathbf{Z}_{(i)S} - m_i \frac{\sum_{j=i}^{L} \sum_{l=1}^{m_i} \mathbf{Z}_{(j)}^{l} \theta_{(j)}^{l}}{\sum_{j=i}^{L} \sum_{l=1}^{m_i} \theta_{(j)}^{l}} \right) .$$

The final case allows both ties and censoring. Again I assume L distinct failure times $t_1, \ldots, t_L$ and m_i failures at time t_i. There are also n_i censoring times $t_{i1}, \ldots, t_{in_i}$ greater than or equal to t_i but less than t_{i+1}. There are m_{ij} spells censored at time t_{ij}. The vectors $\mathbf{Z}_{i}^{j}$ and $\mathbf{Z}_{iS}$, as well as θ_{i}^{j}, are defined as in the third case. Corresponding definitions are required for the censoring times. Define $\mathbf{Z}_{ij}^{l}$ to be the covariate vector for the l th spell censored at time t_{ij}, $\mathbf{Z}_{ijS} = \sum_{l=1}^{m_{ij}} \mathbf{Z}_{ij}^{l}$, and $\theta_{ij}^{l} = \exp\{-\mathbf{Z}_{ij}^{l}{}'\boldsymbol{\beta}\}$. The score vector in this final case can be written as

$$\text{(9.5)} \qquad \mathbf{U}_{\boldsymbol{\beta}}(\boldsymbol{\beta}) = \sum_{i=1}^{L} \left(\mathbf{Z}_{(i)S} - m_i \frac{\sum_{j=i}^{L} \left(\left(\sum_{l=1}^{m_j} \mathbf{Z}_{j}^{l} \theta_{j}^{l} \right) + \sum_{p=1}^{n_j} \sum_{l=1}^{m_{jp}} \mathbf{Z}_{jp}^{l} \theta_{jp}^{l} \right)}{\sum_{j=i}^{L} \left(\left(\sum_{l=1}^{m_j} \theta_{j}^{l} \right) + \sum_{p=1}^{n_j} \sum_{l=1}^{m_{jp}} \theta_{jp}^{l} \right)} \right) .$$

Testing the hypothesis that $\boldsymbol{\beta} = \mathbf{0}$ involves replacing θ's by unity in equations (9.1) through (9.5). For the case without ties or censoring, the score statistic becomes

$$\mathbf{U}_{\boldsymbol{\beta}}(\mathbf{0}) = \sum_{i=1}^{L}\left(\mathbf{Z}_{(i)} - (L-i+1)^{-1}\sum_{j=i}^{L}\mathbf{Z}_{(j)}\right), \tag{9.6}$$

which can be re-written in the following elegant form:

$$\mathbf{U}_{\boldsymbol{\beta}}(\mathbf{0}) = \sum_{i=1}^{L}\mathbf{Z}_{(i)}\left(1 - \sum_{j=1}^{i}(L-j+1)^{-1}\right). \tag{9.7}$$

The score statistic is linear in the ranked covariate vectors and is therefore a simple example of a linear rank statistic. For the case with both ties and censoring, the score statistic has the form:

$$\mathbf{U}_{\boldsymbol{\beta}}(\mathbf{0}) = \sum_{i=1}^{L}\left(\mathbf{Z}_{iS} - m_i' \sum_{j=i}^{L}(\mathbf{Z}_{jS} + \sum_{p=1}^{n_j}\mathbf{Z}_{jpS})\right), \tag{9.8}$$

where $m_i' = m_i \Big/ \sum_{i=1}^{L}(m_j + \sum_{p=1}^{n_j} m_{jp})$.

The log-rank test tests whether $s+1$ populations labeled $0, 1, 2, \ldots, s$ have identical survivor functions. It arises as a special case of the hypothesis test for $\boldsymbol{\beta} = \mathbf{0}$ by defining $\mathbf{Z}_i = (Z_{1i}, \ldots, Z_{si})'$, where Z_{ji} equals one or zero according to whether or not individual i is drawn from population j. For the most general case that allows for both ties and censoring, the log-rank statistic can be written

$$\mathbf{U}_{\boldsymbol{\beta}}(\mathbf{0}) = \mathbf{O} - \mathbf{E}, \tag{9.9}$$

where $\mathbf{O} = \mathbf{Z}_{1S} + \ldots + \mathbf{Z}_{LS}$ is a vector giving the observed number of failures, and

$$\mathbf{E} = m_i' \sum_{i=1}^{L}\sum_{j=i}^{L}\left(\mathbf{Z}_{jS} + \sum_{p=1}^{n_j}\mathbf{Z}_{jpS}\right) \tag{9.10}$$

is a vector representing "expected" failures.

Kalbfleisch and Prentice (1980) point out that $\mathbf{E}$ is not exactly the number of expected failures

> but is rather the sum over failure times of the conditional expected number of failures in each sample, the expectation being under the null hypothesis and, at each time, being conditional upon the total number of failures at that time. (p.80)

Kalbfleisch and Prentice explain that since the elements of $\mathbf{E}$ are themselves random variables, it is clear that $\mathbf{E}$ can represent the vector of expected failures only in an informal sense.

Letting $\mathbf{V}_{\boldsymbol{\beta}}$ represent the asymptotic variance matrix obtained from the true partial likelihood incorporating censoring and ties, the log-rank test statistic $\mathbf{U}_{\boldsymbol{\beta}}(\mathbf{0})'\mathbf{V}_{\boldsymbol{\beta}}^{-1}\mathbf{U}_{\boldsymbol{\beta}}(\mathbf{0})$ is asymptotically χ_s^2 under the null hypothesis. Kalbfleisch and Prentice note that the asymptotic variance matrix obtained from Breslow's approximation to the partial likelihood tends to overestimate the score statistic variance. Using the asymptotic variance matrix from the Breslow approximation results in a lower value of the test statistic and therefore leads to a more conservative test.

The log-rank test is a non-parametric test when the observations are independent. When observations are dependent, however, it turns out that parameters relating to the mixing distribution (the parameters incorporated in M) have to be estimated. The partial likelihood estimates of these parameters can be used for this purpose.

In presenting the construction of the log-rank test when observations are dependent, I will present only the Breslow approximation to the partial likelihood (see section 7 and especially equation (7.5)), which allows for ties and censoring. The case without ties can be recovered by setting $m_i = 1,\ i = 1,\ldots,L$ and recognizing that each set $D(t_i)$ will contain the single element i.

From equation (7.5), the log of the partial likelihood is given by

$$(9.11)\ L_p = \sum_{i=1}^{L}\left[\left(\sum_{j\in D(t_i)} \log\left(\theta_j M^{[j]}(\boldsymbol{\theta},\boldsymbol{\iota},R(t_i)\right)\right) - m_i \log \sum_{j\in R(t_i)} \theta_j M^{[j]}(\boldsymbol{\theta},\boldsymbol{\iota},R(t_i))\right].$$

Defining

$$(9.12) \qquad \mathbf{Z}_j^* = \mathbf{Z}_j + \sum_{l\in R(t_i)} Z_l \theta_l M^{[j,l]}(\boldsymbol{\theta},\boldsymbol{\iota},R(t_i)) \big/ M^{[j]}(\boldsymbol{\theta},\boldsymbol{\iota},R(t_i))$$

and

$$(9.13) \qquad P_j = \theta_j M^{[j]}(\boldsymbol{\theta},\boldsymbol{\iota},R(t_i)) \big/ M(\boldsymbol{\theta},\boldsymbol{\iota},R(t_i)),$$

the score vector for $\boldsymbol{\beta}$ can be written as

$$(9.14) \qquad \mathbf{U}_{\boldsymbol{\beta}}(\boldsymbol{\beta}) = \sum_{i=1}^{L}\left(\sum_{j\in D(t_i)} \mathbf{Z}_j^*\right) - m_i\left(\sum_{j\in R(t_i)} P_j \mathbf{Z}_j^*\right).$$

The score statistic testing the null hypothesis that $\boldsymbol{\beta} = \mathbf{0}$ is then given by

$$(9.15) \qquad \mathbf{U}_{\boldsymbol{\beta}}(\boldsymbol{\beta}) = \sum_{i=1}^{L}\left(\sum_{j\in D(t_i)} \mathbf{Z}_j^{**}\right) - m_i\left(\sum_{j\in R(t_i)} P_j^* \mathbf{Z}_j^{**}\right),$$

where

$$(9.16) \qquad \mathbf{Z}_j^{**} = \mathbf{Z}_j + \sum_{l\in R(t_i)} \mathbf{Z}_l M^{[j,l]}(\boldsymbol{\iota},\boldsymbol{\iota},R(t_i)) \big/ M^{[j]}(\boldsymbol{\iota},\boldsymbol{\iota},R(t_i))$$

and

$$(9.17) \qquad P_j^* = M^{[j]}(\boldsymbol{\iota},\boldsymbol{\iota},R(t_i)) \big/ M(\boldsymbol{\iota},\boldsymbol{\iota},R(t_i))$$

are, respectively, equations (9.12) and (9.13) with all instances of $\theta_p,\ p = 1,\ldots,N$ evaluated at unity.

As in the case of independent observations, the log-rank test arises here as a special case of the hypothesis test for $\boldsymbol{\beta} = \mathbf{0}$. All instances of $\mathbf{Z}_j(\mathbf{Z}_\ell)$ comprising $\mathbf{Z}_j^{**}$ in equation (9.15) are binary vectors with i th element equal to unity, $i = 1, \ldots, s$ if and only if the j th (lth) individual is drawn from population i. Finally, note that the test can only be regarded as semi-parametric because the evaluation of M and its first and second partials in (9.16) and (9.17) involve parameters of the mixing distribution.

The score statistic when observations are dependent again has a representation of the form given in equation (9.9). However, the terms $\mathbf{O}$ and $\mathbf{E}$ no longer correspond to observed and "expected" failures, but instead to observed and "expected" failures corrected for the dependence of observations (through clustering or grouping of the data).

Letting $\mathbf{V}_{\boldsymbol{\beta}}$ represent the asymptotic variance matrix for $\boldsymbol{\beta}$ from the true partial likelihood, the log-rank test statistic $\mathbf{U}_{\boldsymbol{\beta}}(\mathbf{0})'\mathbf{V}_{\boldsymbol{\beta}}^{-1}\mathbf{U}_{\boldsymbol{\beta}}(\mathbf{0})$ is asymptotically χ_s^2 under the null. (See Ondrich 2005 for a discussion of the asymptotic normality of the score vector.) It should be noted that the variance matrix for $\boldsymbol{\beta}$ from the true partial likelihood will be more complicated than in the case of independent observations, and using the asymptotic variance matrix resulting from Breslow's approximation to the true partial likelihood may become expedient. For reasons states previously, it is likely that this substitution will result in a more conservative test. An analysis of these conjectures would appear to be a fruitful area of research.

10. EXAMPLES OF COX-McFADDEN MODELS

I start this section by presenting several examples of Cox-McFadden models. Actually, all examples that I present have already appeared in the epidemiological survival analysis or econometric discrete choice literatures, but have not been previously applied to partial likelihoods.

I assume that the sample of N individuals can be divided into G independent groups or clusters. Group g is composed of N_g individuals and is associated with its own negative log copula M_g. Each M_g satisfies the conditions of Propositions 1 and 2, and therefore $\sum_{g=1}^{G} M_g$ also satisfies these conditions.

The first example comes from Hougaard (1986a, 1986b) and results from positive stable mixing. Suppose $X_1, X_2, \ldots, X_n, \ldots$ are independent and identically distributed. Their common distribution is stable if, for each n, there exists a constant c_n such that $c_n X_1$ and $\sum_{i=1}^{n} X_i$ follow the same distribution. Any stable distribution has constants c_n of the form $n^{1/\alpha}$, where the characteristic exponent $\alpha \in (0,2]$. Normal distributions have $\alpha = 2$ and are the only stable distributions with finite variance. The positive stable distributions (having support on the positive real numbers) all have $\alpha \in (0,1)$ and have Laplace transforms (apart from scaling factors) of the form $\omega(\lambda) = \exp(-\lambda^{\alpha})$, for $\lambda \geq 0$.

If group g shares a common positive stable random effect with characteristic exponent α, then

$$M_g(\theta, \ldots, \theta_{N_g}) = \left(\sum_{i=j}^{N_g} \theta_i^{1/\alpha} \right)^{\alpha}, \tag{10.1}$$

which satisfies the conditions of Propositions 1 and 2. Econometricians will recognize this functional form from McFadden (1978).

Feller (1971) shows that if X_1 and X_2 are independent stable distributions with characteristic exponents α_1 and α_2 $(\alpha_2 < 1)$, then $X_1 X_2^{1/\alpha}$ is stable with characteristic exponent $\alpha_1\alpha_2$. Therefore, if X_1 and X_2 are both positive stable, $X_1 X_2^{1/\alpha}$ is positive stable as well. Hougaard (1986b) uses this to construct a nested frailty model in which three siblings share a family effect and the twins share a "twin" effect.

Sastry (1997) analyzes a nested frailty (using gamma distributions) for child survival in Brazil , where the data are clustered at both the family and community levels. Following Hougaard and using positive stable distributions to construct the nests, the negative log copula for community *g* composed of individuals *j*, each a member of a family *i*, is given by:

(10.2)
$$M_g = \left(\sum_{i \in g} \left(\sum_{j \in i} \theta_{ij}^{1/\alpha_2} \right)^{\alpha_2/\alpha_1} \right)^{\alpha_1},$$

where $\alpha_1 \geq \alpha_2$. In a discrete choice context, McFadden (1978) presents a two-tiered hierarchy that is identical. Hierarchies with more than two tiers can be easily constructed, and other non-nested models are possible.

11. CONCLUSIONS

Cox (1972, 1975) develops the proportional hazard model of durations and suggests estimation using a partial likelihood approach. Contributions to the partial likelihood are provided at each failure time by the subset of the sample at risk immediately before the failure time. The partial likelihood approach has the advantage of being baseline-free: duration-dependence parameters, frequently viewed as nuisance parameters, do not have to be estimated. For researchers interested in duration dependence, the duration baseline can be recovered in a second step.

This paper examines the problem of incorporating random effects in a proportional hazard model, leaving the baseline hazard unspecified. It shows that the set of models that support partial likelihood estimation of the hazard scale coefficients can be made isomorphic to the set of GEV models developed by McFadden (1978). A multivariate extension of a theorem proved by Sergei Bernstein (1928) is used in the proof. This extension provides a means to check whether a given multivariate function can be the likelihood function for a sample of durations, marginal on group-specific random effects.

The partial likelihoods allow independent censoring and I discuss approximations to the partial likelihoods in the presence of ties. The partial likelihood score vector can be used to construct semi-parametric log-rank tests that do not require the independence of observations involved.

An appendix on asymptotic inference (Ondrich 2005) can be found at http://faculty.maxwell.syr.edu/jondrich. This appendix makes three contributions. First, the theory of multiplicative intensity models supports the incorporation of time-varying covariates. Second, $\sqrt{G}$ -consistency and asymptotic normality of the scale, mixing and baseline parameters follow directly from the previous work of Andersen and Gill (1982) for the partial likelihood with independent observations. With independent observations the partial likelihood is globally concave, which is not the case here. However, the results carry over to the case of dependent observations if one considers a compact set of parameter values containing the true values, over which set the partial likelihood is strictly concave. Third, because the data are clustered, the inverse of the information matrix is not the appropriate asymptotic variance matrix. The correct asymptotic variance matrix for the partial likelihood estimator of scale and mixing parameters $\boldsymbol{\gamma}$ is $\mathscr{I}^{-1}(\boldsymbol{\gamma})\mathscr{O}(\boldsymbol{\gamma})\mathscr{I}^{-1}(\boldsymbol{\gamma})$, where $\mathscr{O}(\boldsymbol{\gamma})$ is *G* times the limit, across independent groups *g*, of the sample mean of the outer product of scores, and $\mathscr{I}(\boldsymbol{\gamma})$ is the information matrix.

References

Andersen, Elisabeth Wreford. 2004. "Composite Likelihood and Two-Stage Estimation in Family Studies." *Biostatistics* 5 (1): 15-30.

Andersen, Per Kragh, and Richard D. Gill. 1982. "Cox's Regression Model for Counting Processes: A Large Sample Study." *Annals of Statistics* 10 (4) (December): 1100-1120.

Andersen, Per Kragh, Ørnulf Borgan, Richard D. Gill, and Niels Keiding. 1993. *Statistical Models Based on Counting Processes.* Springer Series in Statistics New York: Springer-Verlag.

Arjas, Elja, and Pentti Haara. 1984. "A Marked Point Process Approach to Censored Failure Data with Complicated Covariates." *Scandinavian Journal of Statistics* 11 (4): 193-209.

Bernstein, Sergei. 1928. "Sur les fonctions absolument monotones." *Acta Mathematica* 51: 1-66.

Börsch-Supan, Axel. 1987. "Econometric Analysis of Discrete Choice: With Applications on the Demand for Housing in the U.S. and West Germany." Lecture Notes in Economics and Mathematical Systems No. 296. Berlin: Springer-Verlag.

Breslow, N. 1974. "Covariance Analysis of Censored Survival Data." *Biometrics* 30 (1) (March): 89-99.

Chamberlain, Gary. 1985. “Heterogeneity, Omitted Variable Bias, and Duration Dependence.” Chapter One In *Longitudinal Analysis of Labor Market Data*, edited by James J. Heckman and Burton Singer. Cambridge: Cambridge University Press, 3-38.

Cox, D. R. 1975. “Partial Likelihood.” *Biometrika* 62 (2) (August): 269-276.

Cox, D. R. 1972. “Regression Models and Life-Tables.” *Journal of the Royal Statistical Society.Series B (Methodological)* 34 (2): 187-220.

Daly, Andrew J., and Stanley Zachary. 1978. “Improved Multiple Choice Models.” Chapter 10 In *Determinants of Travel Choice*, edited by David A. Hensher and Quasim Dalvi. Farnborough, England: Saxon House, Teakfield Limited, 335-357.

Efron, Bradley. 1977. “The Efficiency of Cox’s Likelihood Function for Censored Data.” *Journal of the American Statistical Association* 72 (359) (September): 557-565.

Feller, William. 1971. *An Introduction to Probability Theory and Its Applications.* Second edition.A Wiley publication in mathematical statistics New York: Wiley.

Fleming, Thomas R., and David P. Harrington. 1991. *Counting Processes and Survival Analysis.* New York: John Wiley & Sons.

Grenander, Ulf. 1956. “On the Theory of Mortality Measurement, I and II.” *Skandinavisk Aktuarietidskrift* 39: 70-96; 125-153.

Gross, Shulamith T., and Catherine Huber. 1987. “Matched Pair Experiments: Cox and Maximum Likelihood Estimation.” *Scandinavian Journal of Statistics* 14: 27-41.

Hausman, Jerry, and Daniel. McFadden. 1984. “Specification Tests for the Multinomial Logit Model.” *Econometrica* 52 (5) (September): 1219-1240.

Hougaard, Philip. 1986a. “Survival Models for Heterogeneous Populations Derived from Stable Distributions.” *Biometrika* 73 (2): 387-396.

Hougaard, Philip. 1986b. “A Class of Multivariate Failure Time Distributions.” *Biometrika* 73 (3): 671-678.

Kalbfleisch, J. D., and Ross L. Prentice. 1980. *The Statistical Analysis of Failure Time Data.*Wiley Series in Probability and Mathematical Statistics New York: Wiley.

McFadden, Daniel, Kenneth Train, and William B. Tye. 1977. “An Application of Diagnostic Tests for the Independence from Irrelevant Alternatives Property of the Multinomial Logit Model.” In *Transportation Research Record No. 637: Forecasting Passenger and Freight Travel* Washington, DC: Transportation Research Board, National Academies of Science, 39-46.

McFadden, Daniel. 1974. “Conditional Logit Analysis of Qualitative Choice Behavior.” Chapter 4 In *Frontiers in Econometrics*, edited by Paul Zarembka. New York: Academic Press.

McFadden, Daniel. 1978. “Modelling the Choice of Residential Location.” Paper No. 25 In *Spatial Interaction Theory and Planning Models*, edited by Anders Karlqvist, Lars Lundqvist, Folke Snickars, and Jörgen W. Weibull. Amsterdam: North Holland, 75-96.

McFadden, Daniel. 1987. “Regression-Based Specification Tests for the Multinomial Logit Model.” *Journal of Econometrics* 34 (1/2) (January-February): 63-82.

Meyer, Bruce D. 1990. “Unemployment Insurance and Unemployment Spells.” *Econometrica* 58 (4) (July): 757-782.

Nelson, Roger B. 1999. *An Introduction to Copulas.* Lecture Notes in Statistics, Volume 139. New York: Springer-Verlag.

Ondrich, Jan. 2005. “Asymptotic Inference for Cox-McFadden Partial and Marginal Likelihoods.” Mimeo, Center for Policy Research, Syracuse University.

Peto, Richard. 1972. “Contribution to the discussion of ‘Regression Models and Life-Tables” by D. R. Cox.” *Journal of the Royal Statistical Society, Series B* 34 (2): 205-207.

Prentice, Ross L., and Lynn A. Gloeckler. 1978. “Regression Analysis of Grouped Survival Data with Application to Breast Cancer Data.” *Biometrics* 34 (1) (March): 57-67.

Ridder, Geert, and Insan Tunali. 1999. “Stratified Partial Likelihood Estimation.” *Journal of Econometrics* 92: 193-232.

Sastry, Narayan. 1997. “A Nested Frailty Model for Survival Data, with an Application to the Study of Child Survival in Northeast Brazil.” *Journal of the American Statistical Association* 92 (438) (June): 426-435.

Small, Kenneth A., and Cheng Hsiao. 1985. “Multinomial Logit Specification Tests.” *International Economic Review* 26 (3) (October): 619-627.

Williams, Huw. 1977. “On the Formation of Travel Demand Models and Economic Evaluation Measures of User Benefit.” *Environment and Planning* 9A (3): 285-344.

Wong, Wing Hung. 1986. “Theory of Partial Likelihood.” *Annals of Statistics* 14 (1) (March): 88-123.