ACTUARIAL RESEARCH CLEARING HOUSE
1992 VOL. 1

SEMIPARAMETRIC ESTIMATION OF WARRANTY COSTS

by

Edward W. Frees
University of Wisconsin - Madison

School of Business and Department of Statistics
1155 Observatory Drive
Madison, Wisconsin 53706

June 1991

Keywords and Phrases: Residuals, U-statistics, renewal function, nonparametric estimation

ABSTRACT

A large class of financial models of costs associated with warranties involve functions that arise in renewal theory. The simplest interesting example is the renewal function which, for the purposes of this article, may be interpreted as the expected number of failures by a specific warranty duration. To estimate the renewal function, cross-sectional regression data is assumed to be available. The data is assumed to be semiparametric in that the observation may be expressed as a known function of a vector of covariates, a vector of unknown parameters and an unknown error term while no distributional assumptions are made on the error terms. An estimator of the renewal function is constructed and conditions are provided so that it is asymptotically normal, after suitable standardization. This estimator is a special case of a class of statistics introduced here called residual-based infinite order U-statistics. U-statistics are complex averages over functions of observations. In this article, the observations are allowed to be residuals from a complex regression model. By appealing to this large class of statistics, several other parameters of interest in warranty analysis and related fields may be estimated. Convolutions of distribution functions and a discounted renewal function are discussed in this article. Using cross-sectional regression data to estimate characteristics of a stochastic process, such as the renewal function, requires strong assumptions. In fact, an interesting aspect of this article is that this estimation can be accomplished using reasonable models for the contracts and data collection.

Semiparametric Estimation of Warranty Costs

1. Introduction

An important aspect of marketing a product is the quality of the product, both real and perceived. Because of marketing pressures, warranties are routinely provided by manufacturers to a consumer on purchase of a product. A warranty is an agreement to repair or replace a purchased product upon failure. Thus, the financial obligation of the manufacturer is realized only upon failure of the product which is a contingent, or random, event. General accounting procedures dictate that a liability, or reserve, be established to meet this obligation. In the United States, this is part of the Financial Accounting Standards Board's *Statement of Financial Accounting Standards No. 5; Accounting for Contingencies*. Estimating the expected liability of a warranty is the motivation which drives the development this paper. Of course, this is only one aspect of warranty analysis which, roughly speaking, is the subject of financial implications of the reliability of products. If one takes actuarial science to be the quantitative study of financial implications of contingent events, then warranty analysis is a subset of actuarial science. Warranty analysis would not fall under a more traditional definition of actuarial science, the quantitative study of financial security systems. See Taylor (1986) for a description of a warranty system from an insurance company perspective.

There are number of important financial summary measures in warranty analysis. The simplest interesting example is directly related to the renewal function, as follows. Consider a sequence of i.i.d. random variables, $Y_1, Y_2, \ldots$, that represents successive lifetimes of a product. Under a free replacement policy of duration 'T', the product is immediately replaced upon failure up to and including 'T' units of time after initial purchase of the product. In this context, the renewal function evaluated at time 'T' is the expected number of replacements and is defined by

$$H(T) = E \sum_{k=1}^{\infty} I(Y_1 + \ldots + Y_k \leq T), \qquad (1.1)$$

where I( ) is the indicator of a set and E denotes expectation. Assuming the cost per replacement is either fixed or can be modeled by an independent stationary process, the expected warranty cost is the expected number of replacements, H(T), times the expected cost per replacement. Several additional summary measures have been discussed in Blischke and Scheuer (1975, 1981), Mamer (1982, 1987), Nguyen and Murthy (1984) and Frees (1988) and are considered in section 3. These include models of the variability of costs and models which incorporate the time value of money and other economic aspects of the warranty agreement. The important point is that the sequence $\{Y_1, Y_2, \ldots\}$ and summary measures as in (1.1) comprise a model used to determine financial implications of the warranty contract.

To estimate summary measures such as the renewal function in (1.1), clearly a desirable form of data is to have identical and independent copies of the stochastic process $\{Y_1, Y_2, \ldots\}$. With this data, even if dependencies exist among the observations of each process, one could still consistently estimate H(T). When one does not have this desirable form of the data, reliable estimates can still be achieved by making stronger assumptions on the distribution of the process. In particular, in Frees (1986a,b), I showed how to estimate the renewal function assuming that observations $Y_1, \ldots, Y_n$ are i.i.d. Building on this work, Crowell and Sen (1989) have announced the extension of this work where the (i.i.d.) data are gathered sequentially, Schneider et al (1990) have discussed efficient computational algorithms and Schneider et al (1991) have discussed extension to the censored data case. In Frees (1989), the class of parameters estimated was substantially generalized to handle, as special cases, other financial summary measures briefly alluded to above. For example, suppose that $\{h_k\}$ is a sequence of known functions where h_k maps k-dimensional Euclidean space into the real line. Assume that sufficient conditions exist so that the parameter

$$\tau = E \sum_{k=1}^{\infty} h_k(Y_1, \ldots, Y_k), \tag{1.2}$$

is well-defined, e.g., $h_k(Y_1, \ldots, Y_k) = I(Y_1 + \ldots + Y_k \leq T)$. This is an extension of the concept of unbiased estimators called U-statistics to sequences that are possibly unbounded and are called infinite order U-statistics. For example, in the context of warranty analysis, defining $h_k(Y_1, \ldots, Y_k) = \exp(-\delta(Y_1 + \ldots + Y_k))I(Y_1 + \ldots + Y_k \leq T)$ means that τ can be interpreted as the expected number of renewals discounted at the rate of interest δ. See, for example, Mamer (1987) for a discussion of this parameter.

There are a number of ways of collecting data to approximate the above summary measures. In this paper, I relax the stringent assumption that the observations are i.i.d. and assume, instead, that the data are cross-sectional, or regression, data. One traditional formulation is to assume that each observation (Y_i, X_i) follows the nonlinear regression model

$$Y_i = g_i(\theta, X_i) + e_i, \qquad i=1, \ldots, n. \tag{1.3}$$

Here, Y_i is the response or dependent random element, X_i is the covariate or independent element, θ is a p-dimensional vector of unknown parameters, $\{e_i\}$ is an i.i.d. sequence of unobserved random elements, and $\{g_i\}$ is a sequence of known functions. In this paper, I also consider a more general formulation due to Cox and Snell (1968). Here, $\{G_i\}$ and $\{R_i\}$ are sequences of known functions satisfying the relationships,

$$Y_i = G_i(\theta, e_i), \tag{1.4}$$

and

$$e_i = R_i(\theta, Y_i). \tag{1.5}$$

The model in (1.4) and (1.5) is semiparametric in the sense that while the dependent element Y_i is parameterized by θ, no parametric assumptions are made concerning the distribution of $\{e_i\}$. Among other smoothness conditions on (1.4) and (1.5), it is assumed that sufficient conditions are available so that root-n consistent estimates of θ are available; see assumption A1 below. The model outlined in (1.4) and (1.5) includes the nonlinear regression model in (1.3) and other important special cases. To handle some examples in multivariate regression below, assume that $\{Y_i\}$ and $\{e_i\}$ are q-dimensional random vectors. This formulation, using an estimate $\hat{\theta}$ of θ, allows us to define the residuals as $r_i = R_i(\hat{\theta}, Y_i) = R_i(\hat{\theta}, G(\theta, e_i))$. Now, the sequence $\{r_1, \ldots, r_n\}$ is only approximately i.i.d. However, the fact that $r_i \approx e_i$ suggests that one can use the residuals in calculating a statistic $T_n = T_n(r_1, \ldots, r_n)$ and that the distribution of $T_n(r_1, \ldots, r_n)$ will be nearly the same as that of $T_n(e_1, \ldots, e_n)$. Quantifying this proximity has been the subject of considerable attention in the literature. In Section 2, I extend this literature by quantifying the proximity in the case that the statistic is an infinite order U-statistic. The results of this section may be of independent interest and thus are self-contained.

To tie together the model of financial implications and the model for the data collection formally, begin with $\{e_{ij}, i=1,2,\ldots \; j=1,2,\ldots\}$, a rectangular array of mean zero i.i.d. random variables with finite variance. Assume that we observe $\{(Y_{i1}, X_i)\}$ and that the observations follow model (1.4) and (1.5). The interest is in estimating a summary measure of the stochastic process $\{Y_{o,1}, Y_{o,2}, Y_{o,3}, \ldots\}$. As an example, consider the renewal function in (1.1) and regression data model in (1.3). The goal is to estimate

$$\begin{aligned} H_o(\theta, T) &= E \sum_{k=1}^{\infty} I(Y_{o,1} + \ldots + Y_{o,k} \leq T) \\ &= E \sum_{k=1}^{\infty} I(e_1 + \ldots + e_k + k\, g_o(\theta, X_o) \leq T). \end{aligned} \tag{1.6}$$

In Section 3, the results of Section 2 are used to show that the residuals of the regression modeling, together with $\hat{\theta}$, can be used to estimate $H_o(\theta, T)$. The intuition is that the covariate X_o is specified and hence the regression function is known up to the vector of parameters θ. The distribution of the errors is assumed to be common to all observations sampled and hence can, in principle, be reliably estimated. Some readers may wish to focus on the illustrations of renewal function estimators that are presented in Section 4.

Using cross-sectional regression data to estimate characteristics of a stochastic process requires strong assumptions. In fact, an interesting aspect of this paper is that this estimation can be accomplished using reasonable models for the contracts and data collection.

2. Residual-Based Infinite Order U-Statistics

The main contribution of this section is to extend the residual-based estimation work of Sukhatme (1958) and Randles (1982, 1984) to the case of infinite order U-statistics (Frees, 1989). It is convenient to present the results for a regression structure that is more general than considered by Sukhatme and Randles.

Let $\{h_k\}$ be a sequence of kernels where h_k is of order k and is indexed by $\lambda \in R^p$, i.e., $h:R^{qk} \times R^p \to R$. Let $\{c_{nk}\}$ be a triangular array of constants such that

$$h_n^*(e_1, \ldots, e_n; \lambda) = \Sigma_{k=1}^{n} c_{nk}\, h_k(e_1, \ldots, e_k; \lambda).$$

Thus, the kernel h_n^* is not symmetric in its arguments. The parameter of interest is

$$\tau = \lim_{n\to\infty} \tau_n = \lim_{n\to\infty} E\, h_n^*(e_1, \ldots, e_n; \theta)$$

where the limits are assumed to exist. The estimator of τ investigated in this section is

$$U_n^*(\hat{\theta}) = (n!)^{-1}\, \Sigma_\alpha\, h_n^*(r_{\alpha_1}, \ldots, r_{\alpha_n}; \hat{\theta}). \tag{2.1}$$

Here, Σ_α means the sum over n! permutations of $\{1,2, \ldots, n\}$ of the form $\alpha = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$. For example, it turns out that defining $h_k(e_1, \ldots, e_k; \lambda) = I(e_1 + \ldots + e_k + k\, g_o(\lambda, X_i) \le T)$ and $c_{nk} \equiv 1$ implies that $U_n^*(\hat{\theta})$ is a useful estimate of $H_o(\theta, T)$, defined in (1.6). This and further examples are explored in Section 3. The statistic $U_n^*(\hat{\theta})$ would be a U-statistic, and thus have several known properties (cf., Serfling (1980) and Sen (1981)), except for three details. First, the estimated parameter $\hat{\theta}$ is included in the evaluation of the statistic. Second, the statistic is evaluated using residuals in lieu of i.i.d. elements. Third, the statistic may be of infinite order. That is, if there exists a finite m such that $c_{nk} = 0$ for all $n > m$, then h_n^* is said to be a finite order kernel and infinite order, otherwise. The traditional U-statistic set-up is to require $c_{nk} \equiv 1$ for all $n \le m$. The finite order case where c_{nk} is not constant in n can be handled by straightforward projection and triangular array techniques (cf., Shapiro and Hubert, 1979). The purpose of this section is to explore the proximity of $U_n^*(\hat{\theta})$ to the U-statistic evaluated at θ and using the i.i.d. errors $\{e_i\}$.

As noted by Randles (1984), the average of a function of the residuals in (2.1) can be related to the weighted average of a function of i.i.d. errors, as follows. First, define the perturbed errors $e_i(\lambda) = R_i(\lambda, G_i(\theta, e_i))$ and note that $e_i(\theta) = e_i$. Now define the weighted kernel

$$h_{k,\alpha}(e_{\alpha_1}, \ldots, e_{\alpha_k}; \lambda_1, \lambda_2) = h_k(e_{\alpha_1}(\lambda_1), \ldots, e_{\alpha_k}(\lambda_1); \lambda_2) \tag{2.2}$$

Since $r_i = e_i(\hat{\theta})$, we have that the fixed kernel h_k evaluated using residuals equals the weighted kernel $h_{k,\alpha}$ evaluated using i.i.d. errors, i.e.,

$$h_k(r_{\alpha_1}, \ldots, r_{\alpha_k}; \lambda) = h_{k,\alpha}(e_{\alpha_1}, \ldots, e_{\alpha_k}; \hat{\theta}, \lambda)$$

for each $\lambda \in R^p$. This observation is true using the broad Cox and Snell formulation in (1.4) and (1.5) as well as the simpler model in (1.3) investigated by Randles (1984).

With the notation

$$U_n(\lambda_1, \lambda_2) = \Sigma_{k=1}^n c_{nk} (n!)^{-1} \Sigma_\alpha h_{k,\alpha}(e_{\alpha_1}, \ldots, e_{\alpha_k}; \lambda_1, \lambda_2)$$

and

$$\tau_n(\lambda_1, \lambda_2) = E\, U_n(\lambda_1, \lambda_2),$$

we have $U_n^*(\hat{\theta}) = U_n(\hat{\theta}, \hat{\theta})$ and $\tau_n = \tau_n(\theta, \theta)$.

THEOREM 2.1 : Under the regularity conditions in assumptions A1 - A3 below, we have

$$n^{1/2} (U_n(\hat{\theta}, \hat{\theta}) - \tau_n(\hat{\theta}, \hat{\theta}) - (U_n(\theta, \theta) - \tau_n(\theta, \theta))) \to_p 0.$$

Remarks: The proof of Theorem 2.1 is in the Appendix. In many applications, it turns out that $n^{1/2}(\tau_n(\hat{\theta}, \hat{\theta}) - \tau_n(\theta, \theta)) = o_p(1)$ so that $U_n^*(\hat{\theta})$ inherits the properties of the U-statistic $U_n(\theta, \theta)$. This depends on whether the gradient of τ_n at θ eventually vanishes and is made precise in Corollary 2.1 below. Theorem 2.1 is an extension of Theorem A.9 of Randles (1984) and Theorem 2.8 of Randles (1982) to handle infinite order U-statistics and multivariate, generalized residuals.

Prior to stating the regularity conditions, I first collect some useful notation. For $\lambda \in R^p$, define $|\lambda|$ to be the maximum of the coordinates of λ and, for $d > 0$, define the cube $K(\lambda, d) = \{\lambda^* \in R^p : |\lambda^* - \lambda| \leq d\}$. Let $\{\epsilon_d\}$ be a sequence of constants such that $\epsilon_d \to 0$ as $d \to 0$. Let $\{\gamma_{i,k}\}$, $i=1,2$, be sequences of constants such that $\sup_n \Sigma_k k^{i-1} |c_{n,k}| \gamma_{i,k} < \infty$, $i=1,2$.

Assumptions

A1. Assume that $n^{1/2}(\hat{\theta} - \theta) = O_p(1)$, that is, there exists a sufficiently large M so that $P(\hat{\theta} \notin K(\theta, M n^{-1/2})) \to 0$.

Assume, for each $\lambda_1, \lambda_2 \in K(\theta, M n^{-1/2})$, that

A2. $$E \sup \{ |h_{k,\alpha}(e_{\alpha_1}, \ldots, e_{\alpha_k}; \lambda_1^*, \lambda_2^*) - h_{k,\alpha}(e_{\alpha_1}, \ldots, e_{\alpha_k}; \lambda_1, \lambda_2)| : \lambda_1^* \in K(\lambda_1, d), \lambda_2^* \in K(\lambda_2, d)\} \leq d\, \gamma_{1,k},$$

and

A3. $$E \sup \{ |h_{k,\alpha}(e_{\alpha_1}, \ldots, e_{\alpha_k}; \lambda_1^*, \lambda_2^*) - h_{k,\alpha}(e_{\alpha_1}, \ldots, e_{\alpha_k}; \lambda_1, \lambda_2)|^2 : \lambda_1^* \in K(\lambda_1, d), \lambda_2^* \in K(\lambda_2, d)\} \leq \epsilon_d\, \gamma_{2,k}^2 .$$

In the finite order case, Randles observed (1982, Lemma 2.6 and 1984, Condition A.6) that A2 and a uniformly bounded kernel implies A3. In the infinite order case, it is interesting to note that A2 and the fact that h_k is uniformly bounded does not necessarily imply A3.

The application, and relevance, of the assumptions can be best illustrated by considering the simplest example in U-statistic theory, estimating the variance.

Example 2.1. Estimating the Error Variance.

Consider data from the multiplicative error model,

$$Y_i = g_i(\theta)\, e_i \tag{2.3}$$

and initially assume the goal is to estimate the variance of the transformed additive model

$$Y_i^* = g_i^*(\theta) + e_i^*$$

where $Y_i^* = \log Y_i$, $g_i^* = \log g_i$ and $e_i^* = \log e_i$. To estimate $\sigma^2 = \mathrm{Var}\,(e_i^*)$, the well known (cf. Serfling, 1980, page 173) unbiased kernel of order 2 is $h_2(x,y) = (x-y)^2/2$. This is sufficient for the kernel h_n^*, taking $h_k = 0$ for $k \neq 2$ and $c_{n,2} = 1$. For simplicity, now drop the asterisk notation. Let $\hat{\theta}$ be an estimate of θ satisfying A1 and define the residuals $r_i = Y_i - g_i(\hat{\theta})$. Thus, using (2.1), an estimate of σ^2 is $U_n^*(\hat{\theta}) =$ $\binom{n}{2}^{-1} \Sigma_{i<j} (r_i - r_j)^2/2 = (n-1)^{-1} \Sigma_k (r_i - \bar{r}_n)^2$, where $\bar{r}_n = n^{-1} \Sigma_k r_k$.

Now, with (2.2) and the perturbed errors, $e_i(\lambda) = Y_i - g_i(\lambda) = e_i + g_i(\theta) - g_i(\lambda)$, we have $h_{2,\alpha}(e_{\alpha_1}, e_{\alpha_2}; \lambda, \lambda_2) = (e_{\alpha_1}(\lambda) - e_{\alpha_2}(\lambda))^2 / 2$. Require that g_i be uniformly Lipschitz in a neighborhood of θ, more specifically, for some positive constant C, $\sup_i \mid g_i(\lambda_1) - g_i(\lambda_2) \mid \le C \mid \lambda_1 - \lambda_2 \mid$ for all $\lambda_1, \lambda_2 \in K(\theta, M\, n^{-1/2})$. It is straight forward to check that this is sufficient for A2 and A3. Further, since

$$E\, h_{2,\alpha}(e_{\alpha_1}, e_{\alpha_2}; \lambda, \lambda_2) = \sigma^2 + (g_{\alpha_1}(\theta) - g_{\alpha_1}(\lambda) - g_{\alpha_2}(\theta) + g_{\alpha_2}(\lambda))^2 / 2,$$

we have

$$\begin{aligned} \tau_n(\lambda, \lambda) - \tau_n(\theta, \theta) &= \binom{n}{2}^{-1} \Sigma_{i<j} (g_i(\theta) - g_i(\lambda) - g_j(\theta) + g_j(\lambda))^2 / 2 \\ &= (n-1)^{-1}\ \Sigma_i (g_i(\theta) - g_i(\lambda))^2 - (n\,(n-1))^{-1}\, \Sigma_i (g_i(\theta) - g_i(\lambda)) \\ &\le O(\mid \theta - \lambda \mid^2 + n^{-1} \mid \theta - \lambda \mid). \end{aligned}$$

Hence, with A1,

$$\begin{aligned} n^{1/2} (\tau_n(\hat{\theta}, \hat{\theta}) - \tau_n(\theta, \theta)) &= O(n^{1/2} (\hat{\theta}-\theta)^2 + n^{-1/2} (\hat{\theta}-\theta)\,) = O_p(n^{-1/2}) \\ &= o_p(1). \end{aligned}$$

Thus, $U_n^*(\hat{\theta})$ inherits the asymptotic first order properties of U_n. From, for example, Serfling (1980, page 192), assuming finite fourth moments of the errors and the uniform Lipschitz conditions on g_t, we have
$n^{1/2}(U_n^*(\hat{\theta}) - \sigma^2) \to_D N(0, E\, e^4 - \sigma^4)$. ■

Example 2.2. Seemingly Unrelated Regressions

Now consider the model

$$Y_{it} = X_{it}\,\theta_t + e_{it}, \qquad i=1, \dots, n. \qquad t=1, \dots, q. \tag{2.4}$$

This is the linear version of model (1.3) where I write $Y_i = (Y_{i,1}, \dots, Y_{i,q})'$ to suggest that the model (2.4) can be viewed as <u>separate equations</u> (in t) rather than multivariate regression. It is well-known, especially in the econometric literature (cf., Schmidt, 1976), that parameter efficiency is improved by combining equations and using generalized least squares (GLS) in lieu of ordinary least squares (OLS) estimates of $\theta = (\theta_1, \dots, \theta_q)'$. Specifically, assume that $\{e_{1,t}, \dots, e_{n,t}\}$ are i.i.d. for each t and define Σ to be a q x q matrix whose $(i,j)^{th}$ element is $\Sigma_{i,j} = \mathrm{Cov}(e_{1,i}, e_{1,j})$. One procedure for estimating θ_t is to use ordinary least squares for each equation. It is known that OLS for estimating θ reduces to using OLS for each equation for estimating θ_t. Computationally, this is simpler than using generalized least squares. Further, in the case that Σ is a scalar multiple of the identity matrix, OLS estimates are as efficient as GLS estimates. Thus, it is of interest to estimate Σ based on OLS estimators to see if the more complex GLS calculations are warranted.

With the notation $e_i = (e_{i,1}, \dots, e_{i,q})'$, define the kernel $h_{2,i,j}(e_1, e_2) = (e_{1,i} - e_{2,i})(e_{1,j} - e_{2,j}) / 2$. Since $E\, h_{2,i,j}(e_1, e_2)$ equals $\Sigma_{i,j}$, this kernel serves as our unbiased estimator for the $(i,j)^{th}$ element. In example 2.1, the case of i=j was considered and the calculations for i≠j are virtually the same. To this end, let $\hat{\theta}_{OLS}$ be the least square estimate of θ and define r_i to be the corresponding vector of residuals. Define the residual based estimate of $\Sigma_{i,j}$ to be $S_{i,j}(r) = \binom{n}{2}^{-1} \sum_{s<t} h_{2,i,j}(r_s, r_t)$ and define the unobserved estimator $S_{i,j}(e)$ similarly. As in Example 2.1, assuming bounded covariates, it can be checked that

$$S_{i,j}(r) - S_{i,j}(e) = o_p(n^{-1/2}), \tag{2.5}$$

and thus $S_{i,j}(r)$ inherits the asymptotic properties of $S_{i,j}(e)$. Denote S to be the matrix whose $(i,j)^{th}$ element is $S_{i,j}$. To see the applications of this result, recall that there are several statistics available for testing the null hypothesis H_o: $\Sigma = \sigma^2 I_q$, where σ^2 is a scalar and I_q is a q x q identity matrix. Well-known examples include Bartlett's (1954) statistic det $(S)^n / \Pi_{i=1}^q (S_{i,i})^n$ and the Lagrange Multiplier statistic $(\frac{q}{2})^{-1} \Sigma_{i<j} S_{ij}^2$ in, for example, Breusch and Pagan (1980). Both of these examples are continuous transforms of $\{S_{i,j}\}$, say M(S), and have well-known asymptotic properties when calculated using i.i.d. observations. With (2.5) and the continuous mapping theorem, it is straightforward to check that M(S(r)) will inherit the asymptotic properties of M(S(e)).■

To apply the main result under broad settings, listed below are additional regularity conditions that hold in several important cases.

A4. For sufficiently large n, assume that the gradient $\nabla\tau_n(\theta, \theta)$ exists, is finite and satisfies $\sup_n |\tau_n(\lambda, \lambda) - (\tau_n(\theta, \theta) + (\lambda - \theta)' \nabla\tau_n(\theta, \theta))| = o(|\lambda - \theta|)$.

A5. Assume that $U_n(\theta, \theta) = \tau_n(\theta, \theta) + \Sigma_k h_{1,n}(e_k) + o_p(n^{-1/2})$, where $h_{1,n}(x) = E\{U_n(\theta, \theta) - \tau_n(\theta, \theta) \mid e_1 = x\}$ a.s.

A6. Assume that there exists a known function ψ such that $\hat{\theta} = \theta + \Sigma_k Z_{kn} \psi(e_k) + o_p(n^{-1/2})$ where $\{Z_{kn}\}$ is a triangular array of known vectors.

COROLLARY 2.1 Under the regularity conditions A1 - A6, we have

$$U_n^*(\hat{\theta}) = \tau_n(\theta, \theta) + \Sigma_{k=1}^n \{Z_{kn}' \nabla\tau_n(\theta, \theta) \psi(e_k) + h_{1,n}(e_k)\} + o_p(n^{-1/2}).$$

Remarks: The proof of Corollary 2.1 is immediate from Theorem 2.1 and is omitted. Corollary 2.1, and the usual triangular array central limit theorems, immediately yield the limiting distribution for $U_n^*(\hat{\theta})$. Assumption A4 is a mild smoothness requirement on the <u>expected value</u> of the kernel, not the kernel itself. Sufficient conditions for A5 were provided by Hoeffding (1948) in the finite order case and by Frees (1989) in the infinite order case. The expansion in A6 is standard, see Yohai and Maronna (1979, Theorem 3.1) for linear model M-estimates which include maximum likelihood estimates and Welsh (1987, Theorem 1) for linear model L-estimates. Wu (1981, Theorem 5) establishes A6 for nonlinear least squares.

I close this section with an example of Corollary 2.1 using finite order kernels. Examples using infinite order kernels can be found in Section 3.

Example 2.1. Estimating the Error Variance - Continued.

Now consider estimating the variability of the untransformed, multiplicative error model (2.3). The motivation is that it is the variance of e that is directly linked to the variance of the observations, not the variance of the transformed errors. Define the perturbed errors $e_i(\lambda) = Y_i / g_i(\lambda)$ and, for an estimate $\hat{\theta}$ of θ, define the residuals $r_i = e_i(\hat{\theta})$. From (2.2), we have

$$h_{2,\alpha}(e_{\alpha_1}, e_{\alpha_2}; \lambda, \lambda_2) = (e_{\alpha_1} g_{\alpha_1}(\theta) / g_{\alpha_1}(\lambda) - e_{\alpha_2} g_{\alpha_2}(\theta) / g_{\alpha_2}(\lambda))^2 / 2.$$

To satisfy A2 and A3, require as before that g_i is uniformly Lipschitz and also require that g_i is uniformly bounded away from zero. This is sufficient to satisfy A2 and A3. The proof is straightforward and is omitted.

To check A4, we have

$$\begin{aligned} \tau_n(\lambda, \lambda) &= \binom{n}{2}^{-1} \Sigma_{i<j} E\,(e_i\, g_i(\theta) / g_i(\lambda) - e_j\, g_j(\theta) / g_j(\lambda))^2 / 2 \\ &= \sigma^2/n \;\; \Sigma_i \,(\, g_i(\theta) / g_i(\lambda)\,)^2 \end{aligned}$$

and thus,

$$\nabla\tau_n(\theta, \theta) = \sigma^2/n \;\; \Sigma_i\, g_i(\theta)^2 \;\; \partial/\partial\lambda\, (g_i(\lambda))^{-2}|_{\lambda=\theta} \; = -2\,\sigma^2/n \;\; \Sigma_i \;\; \nabla g_i(\theta) / g_i(\theta)\,.$$

Thus, assuming $g_i(\lambda) = g_i(\theta) + (\lambda - \theta)' \nabla g_i(\theta) + o(|\lambda - \theta|)$, we have

$$|\tau_n(\lambda, \lambda) - (\tau_n(\theta, \theta) + (\lambda - \theta)' \nabla\tau_n(\theta, \theta))\,|$$

$$\leq \sigma^2/n \;\; \Sigma_i \,|\, (\, g_i(\theta) / g_i(\lambda)\,)^2 - 1 \;\; + 2\,(\lambda - \theta)' \nabla g_i(\theta) / g_i(\theta)\,| \;\; = o(|\lambda - \theta|),$$

uniformly in n, which is sufficient for A4.

Assumption A5 is essentially a central limit theorem for traditional U-statistics and is satisfied as above. For this example, it turns out (cf., Serfling, 1980, p.182) that $h_{1,n}(x) = ((x\text{-}E\,e)^2\text{-}\sigma^2)/(2n)$.

Assumption A6 is satisfied by appealing to the usual central limit theorems for nonlinear regression after transforming to the additive model. For example, Wu (1981, p.509) provides conditions so that A6 holds with $\psi(e_i) = \log(e_i)$ and $Z_{kn} = (\,\Sigma_i \;\; \nabla g_i(\theta)\, \nabla g_i(\theta)'\,)^{-1}\, \nabla g_k(\theta)$. ■

3. Regression - Based Renewal Function Estimation

In this section, the general theory of Section 2 is applied to several summary measures related to the renewal function that are useful in warranty analysis and another parameter of importance in actuarial science, the probability of ruin. For simplicity, only the regression data model in (1.3) with additive errors is explicitly considered in this section. It turns out that the limiting distribution of the resulting nonparametric estimator is particularly appealing when the kernel can be expressed as a function of the partial sums of the stochastic process of the financial model.

3.1. Convolution and Renewal Function Estimation

Consider the data model in (1.3). Assume that we are interested in the summary measures with characteristics X_o and that the corresponding regression function at that point is $g_o(\theta, X_o) = g_o(\theta)$. In this subsection, I discuss the estimation of the parameter $\tau_n = \sum_{k=1}^{n} c_{nk} F^{(k)}(T - k\, g_o(\theta))$, where $F^{(k)}(T - k\, g_o(\theta)) = P(e_1 + \ldots + e_k + k\, g_o(\theta) \le T)$ is the k-fold convolution of the distribution function evaluated at $T - k\, g_o(\theta)$. When $c_{nk} = 1$ for some fixed k and is zero otherwise, for all n, then τ_n is the k-fold convolution. When c_{nk} is identically equal to one, then τ_n is essentially the renewal function in (1.6).

I begin by discussing an estimate of the k-fold convolution. For an estimate $\hat{\theta}$ of θ satisfying A1, define the residuals $r_i = Y_i - g_i(\hat{\theta})$. Let $h_k(e_1, \ldots, e_k; \lambda) = I(e_1 + \ldots + e_k \le T - k\, g_o(\lambda))$ and define

$$\hat{F}_n^{(k)}(\hat{\theta}, T) = (n!)^{-1} \Sigma_\alpha\, h_k(r_{\alpha_1}, \ldots, r_{\alpha_k}; \hat{\theta})$$

to be the estimate of the k-fold convolution $F^{(k)}(T - kg_o(\theta))$. An interesting interpretation begins with the fact that $F^{(k)}(T - kg_o(\theta))$ can be expressed as $P(Y_1 + \ldots + Y_k \le T)$. If we define $\hat{Y}_{o,i} = g_o(\hat{\theta}) + r_i$, then $\{ \hat{Y}_{o,i} \}$ can be thought of as the bootstrap distribution of the dependent variable at X_o. Freedman (1981) provides an introduction to the use of the bootstrap technique in the linear model set-up. Thus, the estimator of the k-fold convolution is the same as in Frees (1986) except using the bootstrap distribution in lieu of the unobserved i.i.d. random variables. Thus, we can also express the estimate of the k-fold convolution as

$$\hat{F}_n^{(k)}(\hat{\theta}, T) = \binom{n}{k}^{-1} \Sigma_{c(k)}\, I(\hat{Y}_{o,i_1} + \ldots + \hat{Y}_{o,i_k} \le T). \tag{3.1}$$

Here, the notation $\Sigma_{c(k)}$ means sum over all distinct combinations $\{i_1, \ldots, i_k\}$ of $\{1,2,\ldots, n\}$. With this notation, the estimator of τ_n is

$$U_n^*(\hat{\theta}) = (n!)^{-1} \Sigma_\alpha \Sigma_{k=1}^n c_{nk} \; h_k(r_{\alpha_1}, \ldots, r_{\alpha_k}; \hat{\theta})$$

$$= \Sigma_{k=1}^n c_{nk} \hat{F}_n^{(k)}(\hat{\theta}, T).$$

As in Section 2, a useful quantity that is intermediate between τ_n and $U_n^*(\hat{\theta})$ is $\tau_n(\hat{\theta}, \hat{\theta})$, where

$$\tau_n(\lambda, \lambda) = \Sigma_{k=1}^n c_{nk} \binom{n}{k}^{-1} \Sigma_{c(k)} F^{(k)}(T - \Sigma_{j=1}^k (g_{i_j}(\lambda) - g_{i_j}(\theta) - g_o(\lambda))). \tag{3.2}$$

To establish the limiting distribution of these estimates, we need some mild smoothness assumptions on the distribution and regression functions. Similar to Section 2, I assume

L1. The regression function is uniformly Lipschitz of order one in a neighborhood of θ, i.e., for some positive constant C, $\sup_i \mid g_i(\lambda_1) - g_i(\lambda_2) \mid \leq C \mid \lambda_1 - \lambda_2 \mid$ where λ_1, λ_2 are in some neighborhood of θ.

L2. The density of $F^{(k)}$, $f^{(k)}$, exists at $T\text{-}kg_o(\theta)$ and satisfies

(i) $\sup_n \Sigma_k k^2 \, |c_{n,k}| \, f^{(k)}(T\text{-}kg_o(\theta)) < \infty$ and

(ii) $F^{(k)}(T\text{-}kg_o(\theta)+\epsilon) = F^{(k)}(T\text{-}kg_o(\theta)) + \epsilon \, f^{(k)}(T\text{-}kg_o(\theta)) + o(\epsilon)$, uniformly in k.

THEOREM 3.1 Assume that A1, L1 and L2 hold for the data model in (1.3). Then, for a fixed T and X_o,

$$n^{1/2} (U_n(\hat{\theta}, \hat{\theta}) - \tau_n(\hat{\theta}, \hat{\theta}) - \{\Sigma_{k=1}^n c_{nk} (F_n^{(k)}(\theta, T) - F^{(k)}(\theta, T)) \}) \to_p 0,$$

where $F_n^{(k)}(\theta, T) = \binom{n}{k}^{-1} \Sigma_{c(k)} I(e_{i_1} + \ldots + e_{i_k} + kg_o(\theta) \leq T)$ is a U-statistic of order k.

PROOF:

The proof is immediate from Theorem 3.1, after checking A2 and A3. To prove A2, recall the inequality, $\mid I(x \leq y) - I(x \leq z) \mid \leq I(|x-(y+z)/2| \leq |y-z|/2)$ for real numbers x,y and z. Thus,

$$\left| h_{k,\alpha}(e_{\alpha_1}, \ldots, e_{\alpha_k}; \lambda_1^*, \lambda_2^*) - h_{k,\alpha}(e_{\alpha_1}, \ldots, e_{\alpha_k}; \lambda_1, \lambda_2) \right|$$

$$= \Big| I(\Sigma_{j=1}^k (e_{\alpha_j} + g_{\alpha_j}(\theta) - g_{\alpha_j}(\lambda_1^*) \leq T - kg_o(\lambda_2^*))$$
$$- I(\Sigma_{j=1}^k (e_{\alpha_j} + g_{\alpha_j}(\theta) - g_{\alpha_j}(\lambda_1) \leq T - kg_o(\lambda_2)) \Big|$$

$$\leq I\Big(\ \Big| \textstyle\sum_{j=1}^{k} (e_{\alpha_j} + g_{\alpha_j}(\theta) - (g_{\alpha_j}(\lambda_1^*) + g_{\alpha_j}(\lambda_1) - g_o(\lambda_2^*) - g_o(\lambda_2))/2) - T \Big|$$
$$\leq \quad \Big| \textstyle\sum_{j=1}^{k} (g_{\alpha_j}(\lambda_1^*) - g_{\alpha_j}(\lambda_1) - g_o(\lambda_2^*) + g_o(\lambda_2)) \Big| \Big)$$

$$\leq I\Big(\ \Big| \textstyle\sum_{j=1}^{k} (e_{\alpha_j} + g_{\alpha_j}(\theta) - (g_{\alpha_j}(\lambda_1^*) + g_{\alpha_j}(\lambda_1) - g_o(\lambda_2^*) - g_o(\lambda_2))/2) - T \Big| \leq 2kCd \Big)$$

$$\leq I\Big(\ \textstyle\sum_{j=1}^{k} (g_{\alpha_j}(\lambda_1^*) - g_o(\lambda_2^*))/2 - 2kCd$$
$$\leq \textstyle\sum_{j=1}^{k} (e_{\alpha_j} + g_{\alpha_j}(\theta) + (g_o(\lambda_2) - g_{\alpha_j}(\lambda_1))/2) - T$$
$$\leq \textstyle\sum_{j=1}^{k} (g_{\alpha_j}(\lambda_1^*) - g_o(\lambda_2^*))/2 + 2kCd \Big).$$

Define $\eta_U(\alpha_j) = \sup \{g_{\alpha_j}(\lambda_1^*): \lambda_1^* \in K(\lambda_1, d)\}$, $\eta_L(\alpha_j) = \inf \{g_{\alpha_j}(\lambda_1^*): \lambda_1^* \in K(\lambda_1, d)\}$, and similarly define $\eta_U(o)$ and $\eta_\lambda(o)$ using 'o' for α_j and λ_2 for λ_1. By Assumption L1, we have $| \ \eta_U(\alpha_j) - \eta_L(\alpha_j) \ | \leq Cd$. Thus,

$$E \sup \ \Big| h_{k,\alpha}(e_{\alpha_1}, \ldots, e_{\alpha_k} ; \lambda_1^*, \lambda_2^*) - h_{k,\alpha}(e_{\alpha_1}, \ldots, e_{\alpha_k} ; \lambda_1, \lambda_2) \Big|$$

$$\leq \ E\, I\Big(\ \textstyle\sum_{j=1}^{k} (\eta_L(\alpha_j) - \eta_L(o))/2 - 2kCd$$
$$\leq \textstyle\sum_{j=1}^{k} (e_{\alpha_j} + g_{\alpha_j}(\theta) + (g_o(\lambda_2) - g_{\alpha_j}(\lambda_1))/2) - T$$
$$\leq \textstyle\sum_{j=1}^{k} (\eta_U(\alpha_j) - \eta_U(o))/2 + 2kCd \Big).$$

$$= F^{(k)}\Big(\textstyle\sum_{j=1}^{k} (\eta_U(\alpha_j) - \eta_U(o))/2 + 2kCd + T - \sum_{j=1}^{k} (g_{\alpha_j}(\theta) + (g_o(\lambda_2) - g_{\alpha_j}(\lambda_1))/2)\Big)$$
$$- F^{(k)}\Big(\textstyle\sum_{j=1}^{k} (\eta_L(\alpha_j) - \eta_L(o))/2 - 2kCd + T - \sum_{j=1}^{k} (g_{\alpha_j}(\theta) + (g_o(\lambda_2) - g_{\alpha_j}(\lambda_1))/2)\Big)$$

$$= F^{(k)}(T-kg_o(\theta))$$
$$+ (1+o(1))f^{(k)}(T-kg_o(\theta))\big((\textstyle\sum_{j=1}^{k} (g_{\alpha_j}(\theta)-(\eta_U(\alpha_j)+g_{\alpha_j}(\lambda_1))/2) + k(g_o(\lambda_2) - \eta_U(o))/2 + 2kCd\big)$$
$$- \{ (F^{(k)}(T-kg_o(\theta))$$
$$+ (1+o(1))f^{(k)}(T-kg_o(\theta))\big((\textstyle\sum_{j=1}^{k} (g_{\alpha_j}(\theta)-(\eta_L(\alpha_j)+g_{\alpha_j}(\lambda_1))/2) + k(g_o(\lambda_2) - \eta_L(o))/2 - 2kCd\big)\}$$

$$\leq 6\ kCd\ f^{(k)}(T-kg_o(\theta))\ (1 + o(1)).$$

By L1, the o(1) does not depend on i. By L2, this is sufficient for A2. The proof for A3 is similar and is omitted. ■

I now present illustrations of Theorem 3.1 for the convolution and renewal function cases. Another example of Theorem 3.1, not discussed here, is the variance of the number of renewals in, for example, Frees (1988). For simplicity, assume the multiple linear regression model for the data,

$$Y_i = X_i' \beta + \epsilon_i \tag{3.3}$$

Also assume some stability of the covariates sequence,

$$n^{-1} \Sigma_{i=1}^{n} X_i \to \mu_X \tag{3.4}$$

and

$$n^{-1} \Sigma_{i=1}^{n} (X_i - \mu_X)(X_i - \mu_X)' \to \Sigma_X \tag{3.5}$$

where Σ_X is a positive definite matrix. Relations (3.3) - (3.5) are the usual ones made in applied econometrics, cf., Schmidt (1976), and are akin to assuming that the covariates are drawn from a random sample.

COROLLARY 3.2 - CONVOLUTION ESTIMATOR For the data model in (3.1) - (3.3), assume that the covariates are bounded and that L2 holds. Let $\hat{\theta}$ be the least squares estimate and assume that the errors have finite third moments. Then, for a fixed T, X_o and k,

$$n^{1/2} (\hat{F}_n^{(k)}(\hat{\theta}, T) - F^{(k)}(\theta, T)) \to_D N(0, AVAR_1)$$

where the asymptotic variance is $AVAR_1 = f^{(k)}(T\text{-}k X_o' \theta)^2 \sigma^2 (\mu_X - X_o)' \Sigma^{-1} (\mu_X - X_o)$ $+ k^2 \, Var(F^{(k-1)}(T\text{-}k X_o' \theta\text{-}e)) + 2k \, f^{(k)}(T\text{-}k X_o' \theta) \mu_X' \Sigma^{-1} (\mu_X - X_o) \, Cov(e, F^{(k)}(T\text{-}k X_o' \theta\text{-}e))$.

PROOF OF COROLLARY 3.2:

I appeal to Corollary 2.1. Assumptions A1 and A6, with $Z_{kn} = n^{-1}\Sigma^{-1} X_k$ and $\psi(x) = x$, are satisfied by the usual triangular array central limit theorem. Assumptions A2 and A3 are satisfied as in the proof of Theorem 3.1. Assumption A4 is immediate from L2, where $\nabla \tau_n(\theta, \theta) = f^{(k)}(T\text{-}k X_o' \theta) (X - X_o) + o_p(n^{-1/2})$. Assumption A5 is standard in U-statistics theory, see for example Frees (1986, equation (3.3)), where $h_{1n}(e) = k/n \, (F^{(k-1)}(T\text{-}k X_o' \theta\text{-}e) - F^{(k)}(T\text{-}k X_o' \theta))$. Thus, from Corollary 2.1, we have

$$\begin{aligned}\hat{F}_n^{(k)}(\hat{\theta}, T) = F^{(k)}(\theta, T) &+ n^{-1} \Sigma_{i=1}^{n} f^{(k-1)}(T\text{-}k X_o' \theta) X_i' \Sigma^{-1} (X - X_o) e_i \\ &+ k/n \, \Sigma_{i=1}^{n} (F^{(k-1)}(T\text{-}k X_o' \theta\text{-}e_i) - F^{(k)}(T\text{-}k X_o' \theta)) + o_p(n^{-1/2}).\end{aligned}$$

This, Slutsky's Theorem, (3.4), (3.5) and the usual central limit theorem are sufficient for the proof. ■

For the renewal function, let $m = m(n)$ be an integer such that $m \to \infty$ as $n \to \infty$. Define the estimate of the renewal function

$$\hat{H}_{o,n}(\hat{\theta}, T) = \Sigma_{k=1}^{m} \hat{F}_n^{(k)}(\hat{\theta}, T). \tag{3.6}$$

The following is the result which motivated this study.

COROLLARY 3.3 - RENEWAL FUNCTION ESTIMATOR Under the assumptions of Corollary 3.2 and Theorem 3.1 of Frees (1986),

$$n^{1/2} (\hat{H}_{o,n}(\hat{\theta}, T) - H_o(\theta, T)) \to_D N(0, AVAR_2)$$

where $H_o(\theta, T)$ is the renewal function defined in (1.6) and $AVAR_2 =$
$(\Sigma_{k=1}^{\infty} f^{(k)}(T\text{-}k\, X_o'\, \theta))^2\, \sigma^2\, (\mu_X\text{-}X_o)'\Sigma^{-1}(\mu_X\text{-}X_o) + \Sigma_{r=1}^{\infty}\ \Sigma_{s=1}^{\infty}\ rs\ \mathrm{Cov}(F^{(r-1)}(T\text{-}r\, X_o'\, \theta\text{-}e), F^{(s-1)}(T\text{-}s\, X_o'\, \theta\text{-}e))$
$+\ 2(\Sigma_{k=1}^{\infty} f^{(k)}(T\text{-}k\, X_o'\, \theta))\, \mu_X'\Sigma^{-1}(\mu_X\text{-}X_o)\ \ \Sigma_{k=1}^{\infty}\ k\ \mathrm{Cov}(e, F^{(k)}(T\text{-}k\, X_o'\, \theta\text{-}e)).$

The proof of Corollary 3.3 is similar to Corollary 3.2 and is omitted. Further discussion and numerical examples of Corollary 3.3 can be found in Section 4. Consider the case $p=1$ and X as identically constant for the model in (3.3). In this case, $\{Y_i\}$ is an i.i.d. sequence. Here, the asymptotic variance reduces to $AVAR_2 = \Sigma_{r=1}^{\infty}\ \Sigma_{s=1}^{\infty}\ rs\ \mathrm{Cov}(F^{(r-1)}(T\text{-}e), F^{(s-1)}(T\text{-}e))$. A consistent estimator for $AVAR_2$ was established in Theorem 4.1 of Frees (1986). More generally, consistent estimation of the asymptotic variance is a difficult problem. The usual jackknife estimate seems to only capture the middle term in the expression of $AVAR_2$. The problem is complex since, even if $X_o'\, \theta$ may be rescaled to equal zero, one must still estimate the renewal density, $\Sigma_{k=1}^{\infty} f^{(k)}(T)$, using residuals from a preliminary regression fit. Estimation of this asymptotic variance is an interesting possibility for future research.

3.2. Partial Sums

A setting which includes the renewal function estimator of Subsection 3.1 but still restrictive enough to achieve easily interpretable results occurs when each kernel is a function of the sum of observations. I illustrate this setting by considering the discounted renewal function, an important parameter in warranty analysis.

Now, suppose we are interested in summary measures of the form

$$\tau = E\, \Sigma_{k=1}^{\infty}\, s_k(Y_{o,1} + \ldots + Y_{o,k}) \tag{3.7}$$

where $\{Y_{o,1}, Y_{o,2}, \ldots\}$ is the unobserved stochastic process of interest, $\{Y_{o,i}\}$ are independent and follow $Y_{o,i} = X_{o,i}'\beta + e_{o,i}$, and $\{s_k\}$ is a sequence of known functions. Define $S_k(x) = E\, s_k(e_1 + \ldots + e_k + x)$ and let $\nabla S_k(x)$ be the corresponding derivative with respect to x.

As in subsection 3.2, assume the data follows the multiple linear regression model in (3.3)-(3.5). Similar to (3.1), the residual-based infinite order U-statistic estimate is

$$U_n^*(\hat{\theta}) = \Sigma_{k=1}^{m} \binom{n}{k}^{-1} \Sigma_{c(k)}\, s_k(\hat{Y}_{o,i_1} + \ldots + \hat{Y}_{o,i_k}). \tag{3.7}$$

where $\{\hat{Y}_{o,i_1}, \ldots, \hat{Y}_{o,i_k}\}$ is the empirical, or bootstrap, distribution at X_o. The expression in (3.7) is useful for computing the statistic. Similar to (3.2), the quantity useful for computing the asymptotic distribution is

$$\tau_n(\lambda, \lambda) = \Sigma_{k=1}^{m} \binom{n}{k}^{-1} \Sigma_{c(k)}\, S_k(\, \Sigma_{j=1}^{k} X_{i_j} (\lambda - \theta) - X_o\, \lambda)$$

and thus

$$\nabla\tau_n(\theta, \theta) = \Sigma_{k=1}^{m} \nabla S_k(-k\, X_o\, \theta)\, (\overline{X} - X_o).$$

For example, assuming $s_k(x) = e^{-\delta x}\, I(x \leq T)$, then τ in (3.7) is the discounted renewal function evaluated at T, cf., Mamer (1987). This yields $S_k(x) = \int e^{-\delta(e+x)}\, I(e \leq T - x)\, dF^{(k)}(e)$ and thus $\nabla S_k(x) = -\delta\, S_k(x) - e^{-\delta T} f^{(k)}(T - x)$.

4. Illustrative Renewal Function Calculations

To illustrate the calculations for the renewal function in Subsection 3.1 and the discounted version in Subsection 3.2, consider the fictitious data set in Table 1. This data represents experience of failures of a type of photocopy machine from each of 20 randomly selected offices. Here, $\{Y_i\}$ represents the time to failure of the photocopy machine from initial machine installation and $\{X_{1i}\}$ is a measure of the amount of use on a monthly basis, called USAGE. The variable $\{X_{2i}\}$, or TYPE, is an indicator as to the predominant type of user in an office; $X_{2i} = 1$ indicates that the i^{th} office is staffed with primarily professional users, $X_{2i} = 0$ indicates the presence of primarily clerical workers. For this data, the linear regression model with p=3 was fit. Using the data in Table 1, the fitted regression equation turns out to be

$$\hat{Y} = 55.9 - 0.1984 \text{ USAGE} + 14.3 \text{ TYPE}. \tag{4.1}$$

The data, with two superimposed fitted regression lines, appear in Figure 1.

TABLE 1. ILLUSTRATIVE DATA

Office	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
MONTHS	10	22	28	30	22	50	20	48	10	0	18	40	38	30	14	4	20	44	64	56
USAGE	246	120	166	175	151	85	238	31	295	277	272	92	104	224	305	290	210	164	68	124
TYPE	1	0	1	1	0	1	1	0	1	0	1	0	0	1	1	0	0	1	1	1
Bootstrap Values at USAGE = 250 and TYPE = 0	9.2	10.5	11.3	15.1	16.7	17.2	17.6	18.9	18.9	19.7	22.3	23.08	23.4	24.8	24.9	26.3	26.4	26.9	27.9	31.0

Now, suppose that we wish to estimate the expected number of failures by T = 12 months for an office composed of primarily professional workers (TYPE = 0) and with the usage variable at 250. From equation (4.1), the expected time until failure is

$$\hat{Y}_o = 55.9 - 0.1984\,(250) + 14.3\,(0) = 6.3.$$

From standard linear model theory and the above data, it is easy to check that the standard deviation associated with this fit is approximately 2.8. Now, consider a well-known renewal theoretic bound

$$t / \mu - 1 \leq H(t) \leq t / \mu + \sigma^2 / \mu^2 \tag{4.2}$$

which is valid for $t \geq 0$ provided that σ^2 is finite, cf., Carlson and Nerman (1986). Replacing parameters by estimates, one might use the midpoint of the bound in (4.2) to suggest an estimate of H(t). In this case, the estimate turns out to be

Illustrative Data

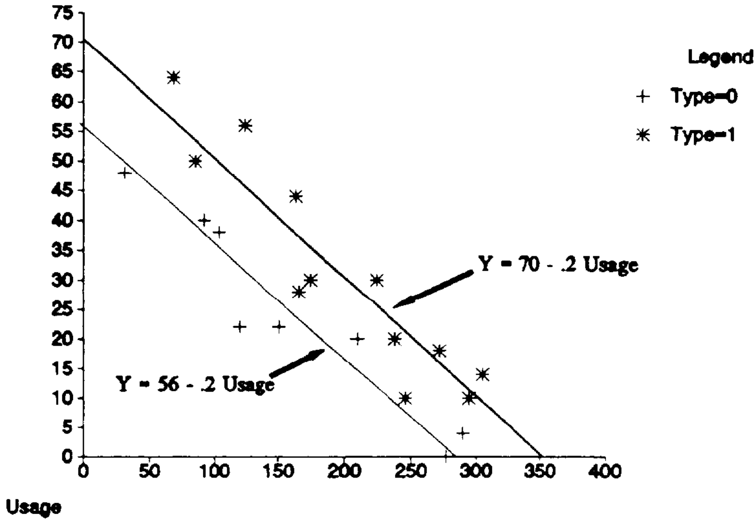

FIGURE 1. The lower fitted regression line corresponds to failure times experienced by clerical workers (Type = 0) and the upper fitted regression line corresponds to professionals (Type = 1).

$$\hat{H}(T) = T/\hat{\mu} + \hat{\sigma}^2 / (2\,\hat{\mu}^2) - 1/2 \tag{4.3}$$

which is approximately $12/6.3 + 2.8^2/(2\ 6.3^2) - 1/2 = 1.51$ in this case. The corresponding estimated bound is $(T/\hat{\mu} - 1,\ T/\hat{\mu} + \hat{\sigma}^2 / \hat{\mu}^2)$, or (.905, 2.103). This type of approximation was discussed in Frees and Nam (1988) for the special case of i.i.d. data. For that situation, the approximation does well as $t \rightarrow \infty$, but performs poorly when T is less than the mean.

As an alternative approximation, consider the semiparametric estimator defined in (3.6). To construct this estimator, one first fits the regression equation as in (4.1) to get $\{r_i\}$, the vector of residuals. For USAGE=250 and TYPE=0, we then construct the bootstrap distribution of observations $\{\hat{Y}_{o,i}\}$, where

$$\hat{Y}_{o,i} = \hat{Y}_o + r_i = 6.3 + r_i.$$

Values of $\{\hat{Y}_{o,i}\}$ are included in Table 1. Estimates of the k-fold convolutions can then be constructed using (3.1). Analogous to Frees (1986), this estimate is the average over all possible evaluations of

$I(\hat{Y}_{o,i_1} + \ldots + \hat{Y}_{o,i_k} \leq T)$. In Table 2 is an example of calculations for various values of k, the level of convolution. Using these values, the semiparametric estimate of the renewal function is $\hat{H}_{o,n}(\hat{\theta}, 12) = .7500 + .4421 + \ldots + .0009 = 1.62$.

TABLE 2. CONVOLUTION ESTIMATES FOR USAGE = 250 AND TYPE = 0

k	1	2	3	4	5	6	7	8
$\hat{F}_n^{(k)}(\hat{\theta}, 12)$	.7500	.4421	.2421	.1119	.0485	.0174	.0050	.0009

In Table 3 is a comparison of the renewal function estimators for various values of USAGE, TYPE, the EXPected fitted value of Y ($\hat{Y}_o$) and DURATION (T). These estimates are based on the illustrative data in Table 1. The USAGE was selected to represent offices with high (USAGE=250) and low (USAGE=100) levels of activity. The warranty duration of T=12 and 24 months were selected to show the effects of duration levels that are a fraction of the mean and greater than the mean. Interpret the situation where the warranty is a fraction of the mean to be an insurance against an "unlikely" failure while the situation where the warranty is greater than the mean is more of a product service contract. In the latter situation, approximations such as in (4.3) seem to fair well, especially given the ease of computing such bounds. In the former situation, the semiparametric estimators seem to provide qualitatively more appealing approximations. See Frees (1986) and Frees and Nam (1988) for a more complete discussion of this issue in the i.i.d. case.

Also in Table 3 is an illustration of the estimation of discounted renewal function estimates. The discounted renewal function is

$$\tau = E \sum_{k=1}^{\infty} \exp(-\delta(Y_{o,1} + \ldots + Y_{o,k})) \quad I(Y_{o,1} + \ldots + Y_{o,k} \leq T)$$

The estimates were computed using (3.7) with $s_k(x) = e^{-\delta x} I(x \leq T)$ and, as above, m=8. As anticipated, the discounted renewal function is smaller for larger values of δ. Further, the larger the warranty duration (T), the greater the effect of δ on the estimated renewal function. While the methods of this article produce asymptotically (as $n \to \infty$) consistent estimates of the discounted renewal function, no simple reliable approximations analogous to (4.3) seem to be available.

TABLE 3. RENEWAL FUNCTION ESTIMATES							
TYPE	USAGE	EXP Y	DURATION	$\hat{H}(T)$	δ		
		$(\hat{Y}_0)$	(T)		0%	1%	2%
0	250	6.4	12	1.514	1.618	1.540	1.470
			24	3.428	3.483	3.104	2.783
0	100	36.1	12	-0.166	0	0	0
			24	0.169	0	0	0
			36	0.499	0.500	0.367	0.269
			48	0.835	1.000	0.698	0.489
1	250	20.7	12	0.087	0.150	0.135	0.122
			24	0.669	0.666	0.560	0.472
1	100	50.4	12	-0.260	0	0	0
			24	-0.022	0	0	0
			36	0.216	0	0	0
			48	0.454	0.350	0.226	0.146

5 Concluding Remarks

In this paper I have studied the use of cross-sectional regression data in semiparametric of costs arising from some simple warranty contracts. The main device was the residual-based infinite order U-statistics introduced in Section 2. This formulation suggests a number of possible extensions. Extensions to the estimation of costs arising from more complex renewal policies should be straightforward using the theory developed in Section. One review of such policies may be found in Frees and Nam (1988). Extensions to other renewal theoretic type measures should be feasible. For example, for the i.i.d. case, the probability of ruin parameter was investigated by Frees (1986c). Extensions to other types of sampling schemes would also prove of interest. For example, Blischke and Scheuer (1975) cite the importance of censored data in warranty analysis.

REFERENCES

Bartlett, M. S. (1954), "A note on multiplying factors for various Chi-squared approximations," *Journal of the Royal Statistical Society B* 16, 296-298.

Bickel, P.J. (1975), "One-step Huber estimates in the linear model," *Journal of the American Statistical Association* 70, 428-434.

Blischke, W. R. and Scheuer, E. M. (1975), "Calculations of the cost of warranty policies as a function of estimated life distributions, *Naval Research Logistics Quarterly* 22, 681-696.

Blischke, W. R. and Scheuer, E. M. (1981), "Application of renewal theory in analysis of the free-replacement warranty," *Naval Research Logistics Quarterly* 28, 193-205.

Breusch, T. S. and Pagan, A. R. (1980), "The Lagrange multiplier test and its applications to model specification in econometrics," *Review of Economic Studies* 47, 239-253.

Carlson, H. and Nerman, O. (1986) "An alternative proof of Lorden's renewal inequality," *Advances in Applied Probability* 18, 1015-1016.

Cox, D. R. and Snell, E. J. (1968), "A general definition of residuals," *Journal of the Royal Statistical Society*, Series B, vol. 30, pp. 248-275.

Crowell, J. I. and Sen, P. K. (1989), "Sequential nonparametric estimation of the renewal function at a point," *IMS Bulletin* 18, no. 3, Abstract No. 210-46.

Freedman, D. A. (1981), "Bootstrapping regression models," *Annals of Statistics* 9, 1218-1238.

Frees, E. W. (1986a), "Warranty analysis and renewal function estimation," *Naval Research Logistics Quarterly* 33, 361-372.

Frees, E. W. (1986b), "Nonparametric renewal function estimation," *Annals of Statistics* 14, 1366-1378.

Frees, E. W. (1986c), "Nonparametric estimation of the probability of ruin," *ASTIN Bulletin* 16S, 81-90.

Frees, E. W. (1988), "On estimating the cost of a warranty," *Journal of Business and Economic Statistics* 6, 79-86.

Frees, E. W. (1989), "Infinite order U-statistics," *Scandinavian Journal of Statistics* 16, 29-44.

Frees, E. W., Nam, S. (1988), "Approximating expected warranty costs", *Management Science* 34, 1441-1449.

Hoeffding, W. (1948), "A class of statistics with asymptotically normal distribution," *Annals of Mathematical Statistics* 19, 193-225.

Mamer, J. W. (1982), "Cost analysis of pro rata and free replacement warranties," *Naval Research Logistics Quarterly* 29, 345-356.

Mamer, J. W. (1987), "Discounted and per unit costs of product warranty," *Management Science* 33, 916-930.

Nguyen, D. G. and Murthy, D. N. P. (1984), "Cost analysis of warranty policies," *Naval Research Logistics Quarterly* 31, 525-541.

Randles, R. H. (1982), "On the asymptotic normality of statistics with estimated parameters," *Annals of Statistics* 10, 462-474.

Randles, R. H. (1984), "On tests applied to residuals," *Journal of the American Statistical Association* 79, 349-354.

Schmidt, P. (1976), *Econometrics*, Marcel Dekker, New York.

Schneider, H., Lin, B. S., and O'Cinneide, C. (1990), "Comparison of nonparametric estimators for the renewal function," *Applied Statistics* 39, 55-61.

Schneider, H., Lin, B. S., and Tan, C. (1991), "Comparison of renewal function estimators for censored data," submitted to *Journal of Statistical Computation and Simulation*.

Sen P. (1981). *Sequential Nonparametrics: Invariance Principles and Statistical Inference*. Wiley, New York.

Serfling, R. J. (1980), *Approximations Theorems of Mathematical Statistics*, Wiley, New York.

Shapiro, C. P. and Hubert, L. (1979), "Asymptotic normality of permutation statistics derived from weighted sums of bivariate functions," *Annals of Statistics* 7, 788-794.

Sukhatme, B. V. (1958), "Testing the hypotheses that two populations differ only in location," *Annals of Mathematical Statistics* 29, 60-78.

Taylor, J. M. (1986), "Extended warranty insurance," *Journal of the Institute of Actuaries Student Society* 29, 1-24.

Wu, C. F. J. (1981), "Asymptotic theory of nonlinear least squares estimation," *Annals of Statistics* 9, 501-513.

Yohai, V. J. and Maronna, R. A. (1979), "Asymptotic behavior of M-estimators for the linear model," *Annals of Statistics* 7, 258-268.

APPENDIX

Proof of Theorem 2.1

Different aspects of the method of the proof can be found in Sukhatme (1958), Bickel (1975), Randles (1982, 1984) and Frees (1989), although the details in this paper are different. Because these different aspects are scattered over a number of papers, an outline of the proof is provided here.

To begin, I first use (2.2) to define a centered version of $h_{k,\alpha}$,

$$W_{k,\alpha}(e_{\alpha_1}, \ldots, e_{\alpha_k}; \lambda_1, \lambda_2) \equiv h_{k,\alpha}(e_{\alpha_1}, \ldots, e_{\alpha_k}; \lambda_1, \lambda_2) - E\, h_{k,\alpha}(e_{\alpha_1}, \ldots, e_{\alpha_k}; \lambda_1, \lambda_2)$$
$$- (h_{k,\alpha}(e_{\alpha_1}, \ldots, e_{\alpha_k}; \theta, \theta) - E\, h_{k,\alpha}(e_{\alpha_1}, \ldots, e_{\alpha_k}; \theta, \theta)).$$

Thus,

$$Q_n(\lambda_1, \lambda_2) \equiv n^{1/2}\,(U_n(\lambda_1, \lambda_2) - \tau_n(\lambda_1, \lambda_2) - (U_n(\theta, \theta) - \tau_n(\theta, \theta)))$$

$$= n^{1/2}\,(n!)^{-1} \Sigma_\alpha \Sigma_{k=1}^{n} c_{nk}\, W_{k,\alpha}(e_{\alpha_1}, \ldots, e_{\alpha_k}; \lambda_1, \lambda_2). \quad \text{(A.1)}$$

Now, define $M(n) = M\, n^{-1/2}$, where M is the large constant in assumption A1. Then, sufficient for the Theorem is

$$\sup \{ |Q_n(\lambda_1, \lambda_2)| : \lambda_1 \in K(\theta, M(n)), \lambda_2 \in K(\theta, M(n))\} \rightarrow_p 0. \quad \text{(A.2)}$$

This is true, since for $\epsilon > 0$,

$$P(\ |Q_n(\hat{\theta}, \hat{\theta})| > \epsilon\ ) \leq P\ (\sup \{\ |Q_n(\lambda_1, \lambda_2)|\ : \lambda_1 \in K(\theta, M(n)), \lambda_2 \in K(\theta, M(n))\} > \epsilon\ )$$
$$+\ P\ (\hat{\theta} \notin K(\theta, M(n))).$$

and the second term on the right hand side tends to zero by assumption A1.

Now, as in Bickel (1975), partition the cube $K(\theta, M(n))$ with a mesh of size δ, to be specified later. Thus, a typical cube in the partition is of the form $(\theta + j_1\, \delta M(n), \ldots, \theta + j_p\, \delta M(n))$ where $j_i = 0, \pm 1, \pm 2, \ldots,$

$\pm([1/\delta]+1)$ and $[.]$ is the greatest integer function. There are $N(\delta) = ([1/\delta]+1)^p$ such cubes. Let $K_1, K_2, \ldots, K_{N(\delta)}$ be some ordering of the cubes and for the i^{th} cube, K_i, let $\theta + \eta_i n^{-1/2}$ denote the position of its lower left hand vertex.

Now, define the mutually exclusive sets

$$E_i = \{ \sup \{|Q_n(\lambda)| : \lambda \in \cup_{j<i} K_j \} < \epsilon \leq \sup \{|Q_n(\lambda)| : \lambda \in K_i \} \}.$$

Here, I use only one parameter λ for simplicity. Thus,

$$P (\sup \{ |Q_n(\lambda)| : \lambda \in K(\theta,M(n))\} > \epsilon) = P (\cup_{i\leq N(\delta)} E_i)$$

$$= P (\cup_{i\leq N(\delta)} (E_i \cap (\{|Q_n(\theta + \eta_i n^{-1/2})| > \epsilon/2\} \cup \{|Q_n(\theta + \eta_i n^{-1/2})| \leq \epsilon/2\})))$$

$$\leq P (\max_{i\leq N(\delta)} \{|Q_n(\theta + \eta_i n^{-1/2})| > \epsilon/2\}) + \Sigma_{i\leq N(\delta)} P(E_i \cap \{|Q_n(\theta + \eta_i n^{-1/2})| \leq \epsilon/2\})$$

$$\leq P (\max_{i\leq N(\delta)} \{|Q_n(\theta + \eta_i n^{-1/2})| > \epsilon/2\}) + \Sigma_{i\leq N(\delta)} P(\sup \{|Q_n(\lambda) - Q_n(\theta + \eta_i n^{-1/2})| : \lambda\in K_i\} > \epsilon/2). \tag{A.3}$$

Sufficient for the first term on the right hand side of (A.3) to tend to zero is

$$\text{Var} (Q_n(\theta + \lambda n^{-1/2})) \to 0, \tag{A.4}$$

for $\lambda \in R^p$, since there are only a finite number of terms in the maximum. Now, let α, β be two permutations of $\{1,2, \ldots, n\}$. From (2.2), consider Cov $\{W_{r,\alpha}(e_{\alpha_1}, \ldots, e_{\alpha_r} ; \theta + \lambda n^{-1/2}), W_{s,\beta}(e_{\beta_1}, \ldots, e_{\beta_s} ; \theta + \lambda n^{-1/2})\}$. If $r+s\leq n$, the number of such covariance terms having zero elements in common is $n! \, (n-r)! \, (n-s)! / (n-r-s)!$. Thus, the proportion of terms having zero elements is $\binom{n-r}{s} / \binom{n}{s}$. For these terms, the covariance term is zero. For other terms, use assumption A3 and Chebyshev's inequality to bound the covariance yielding

$$\text{Cov} \{W_{r,\alpha}(e_{\alpha_1}, \ldots, e_{\alpha_r} ; \theta + \lambda n^{-1/2}), W_{s,\beta}(e_{\beta_1}, \ldots, e_{\beta_s} ; \theta + \lambda n^{-1/2})\} \leq C \, \epsilon_{M(n)} \, \gamma_{2,r} \, \gamma_{2,s}$$

for some positive constant C. Thus,

$$n\ \mathrm{Var}(\ (n!)^{-1} \Sigma_\alpha\ \Sigma_{k=1}^{n} c_{nk}\, W_{k,\alpha}(e_{\alpha_1}, \ldots, e_{\alpha_k}\ ;\ \theta + \lambda\, n^{-1/2}))$$

$$= n\ \ (n!)^{-2}\, \Sigma_\alpha\, \Sigma_\beta\ \ \Sigma_{r=1}^{n}\ \ \Sigma_{s=1}^{n} c_{nr}\, c_{ns}\, \mathrm{Cov}(W_{r,\alpha},\ W_{s,\beta}\,)$$

$$\leq n\ \ \Sigma_{r=1}^{n}\ \ \Sigma_{s=1}^{n} |c_{nr}\ c_{ns}|\ \{\ I(r+s \leq n/2)\ (1 - \tbinom{n-r}{s} / \tbinom{n}{s}) + I(r+s > n/2)\}\ (C\ \epsilon_{M(n)}\ \gamma_{2,r}\ \gamma_{2,s})$$

$$\leq C\ \epsilon_{M(n)}\ \{\ \Sigma_{r+s \leq n/2}\ |c_{nr}\ c_{ns}|\ 2\ r\ s\ \gamma_{2,r}\ \gamma_{2,s} +\ \ \Sigma_{r+s > n/2}\ |c_{nr}\ c_{ns}|\ \gamma_{2,r}\ \gamma_{2,s}\ \}$$

$$\to 0, \tag{A.5}$$

by the requirement that $\sup_n \Sigma_k\ k\ |c_{n,k}|\ \gamma_{2,k} < \infty$. The fact that $n\ (1 - \binom{n-r}{s} / \binom{n}{s}) \leq 2\ rs$ when $r+s \leq n/2$ can be established after several lines of routine algebra. This is sufficient to establish (A.4).

I now establish

$$P(\ \sup\ \{|Q_n(\lambda) - Q_n(\theta + \eta_i\, n^{-1/2})|: \lambda \in K_i\}\ >\ \epsilon/2) \to 0. \tag{A.6}$$

To this end, define

$$H_{k,\alpha}(e_{\alpha_1}, \ldots, e_{\alpha_k};\ K_i) = \sup\{|W_{k,\alpha}(e_{\alpha_1},\ldots,e_{\alpha_k};\ \lambda) - W_{k,\alpha}(e_{\alpha_1},\ldots,e_{\alpha_k}; \theta + \eta_i\, n^{-1/2})|: \lambda\ \in\ K_i\}.$$

Now,

$$\sup\ \{|Q_n(\lambda) - Q_n(\theta\ +\ \eta_i\, n^{-1/2})|: \lambda \in K_i\}$$

$$\leq\ \ \sup\ \{\ n^{1/2}\ |(n!)^{-1}\ \Sigma_\alpha\ \Sigma_{k=1}^{n}\ c_{nk}\ (W_{k,\alpha}(e_{\alpha_1},\ldots,e_{\alpha_k};\ \lambda) - W_{k,\alpha}(e_{\alpha_1},\ldots,e_{\alpha_k};\ \theta\ +\ \eta_i\, n^{-1/2})|: \lambda \in K_i\}$$

$$\leq\ \ n^{1/2}\ (n!)^{-1}\ \Sigma_\alpha\ \Sigma_{k=1}^{n}\ |c_{nk}|\ H_{k,\alpha}(e_{\alpha_1},\ldots,e_{\alpha_k};\ K_i). \tag{A.7}$$

With assumption A2, we have

$$E\, n^{1/2} (n!)^{-1} \Sigma_\alpha \Sigma_{k=1}^{n} |c_{nk}|\, H_{k,\alpha}(e_{\alpha_1},\ldots,e_{\alpha_k}; K_j) \leq 2\, M\, \delta\, \Sigma_{k=1}^{n} |c_{nk}|\, \gamma_{1,k}.$$

Thus, by the requirement that $\sup_n \Sigma_k |c_{n,k}|\, \gamma_{1,k} < \infty$, one can pick the mesh size δ such that $M\, \delta\, \Sigma_{k=1}^{n} |c_{nk}|\, \gamma_{1,k} < \epsilon/4$ for all n. Hence, by (A.7) and the Markov inequality

$$P(\sup \{ | Q_n(\lambda) - Q_n(\theta + \eta_j n^{-1/2}) | : \lambda \in K_j\} > \epsilon/2)$$

$$\leq P\, (n^{1/2} (n!)^{-1} \Sigma_\alpha \Sigma_{k=1}^{n} |c_{nk}|\, (H_{k,\alpha}(e_{\alpha_1},\ldots,e_{\alpha_k}; K_j) - E\, H_{k,\alpha}(e_{\alpha_1},\ldots,e_{\alpha_k}; K_j)) > \epsilon/4)$$

$$\leq 16/\epsilon^2\, \mathrm{Var}\, (n^{1/2} (n!)^{-1} \Sigma_\alpha \Sigma_{k=1}^{n} |c_{nk}|\, (H_{k,\alpha}(e_{\alpha_1},\ldots,e_{\alpha_k}; K_j) - E\, H_{k,\alpha}(e_{\alpha_1},\ldots,e_{\alpha_k}; K_j)))$$

$$\to 0,$$

similarly to (A.5). This is sufficient for (A.6) and hence the result. ∎