Statistica Sinica **13**(2003), 903-926

ESTIMATION OF DISTRIBUTION FUNCTIONS UNDER SECOND ORDER STOCHASTIC DOMINANCE

Javier Rojo Jiménez and Hammou El Barmi

Rice University and City University of New York

Abstract: The concept of stochastic ordering as introduced by Lehmann (1955) plays a major role in the theory and practice of statistics, and a large body of existing statistical work concerns itself with the problem of estimating distribution functions F and G under the constraint that $F(x) \leq G(x)$ for all x. Nevertheless in economic theory, the weaker concept of second order stochastic dominance plays a prominent role in the general framework of analyzing choice under uncertainty by considering maximization of expected utilities. More specifically, an investment portfolio B with random return Y dominates an investment portfolio A with random return X if and only if $E(U(Y)) \geq E(U(X))$ for all increasing and concave utility functions U. This condition can be seen to be equivalent to the condition that the distribution functions of X and Y are ordered according to the second order stochastic dominance requirement. Here, a family of strongly uniformly consistent estimators for the survival functions under a second order stochastic dominance constraint is proposed in the one-sample and the two-sample problems. In the one-sample problem the new family of estimators dominate the empirical distribution function with respect to a certain class of loss functions. The asymptotic distributions of the estimators are explored and the new estimators are compared, via simulations, in terms of Mean Squared Error (MSE) with the empirical distribution. The case of right-censored data is also considered. Stocks and bonds data from 1810−1989 are used to illustrate the estimators.

Key words and phrases: Bonds, expected utility, investment portofolio, stocks, weak convergence.

1. Introduction

The concept of stochastic ordering as proposed by Lehmann (1955) has found important applications in the theory and applications of statistics. Lehmann and Rojo (1992) have considered characterizations of stochastic ordering in terms of a maximal invariant with respect to the group of monotone transformations. Shaked and Shanthikumar (1994) provide a recent and thorough review of the literature on stochastic ordering. In the context of the problem of estimating F and G subject to the stochastic ordering constraint, we mention the work of Dykstra (1982), Lo (1987), Rojo and Ma (1996) and Rojo (1995).

In economic theory, however, the weaker concept of second order stochastic dominance (SSD) plays a major role in developing a general framework for studying choice under uncertainty by the establishment of some criteria to select one option over another. One such criterion involves the maximization of expected utility over possible options. (See, e.g., Hadar and Russell (1971) and Porter and Carey (1974) and, more recently, Levy (1992) and Kijima and Ohnishi (1996)). Specifically, let $U(\cdot)$ denote an increasing utility function which expresses the intensity of preferences for the investor. Then, faced with a decision to choose among n different investments, which generate the random amounts of money $Z_1, ., Z_n$, the investor prefers the investment for which $E(U(Z_i))$ is a maximum. A risk averter is characterized as having a concave utility function. Several criteria to order investment portfolios have been proposed. This paper focuses on the following criterion (see, e.g., Porter and Carey (1974)). A risk averter, and hence every risk averter, prefers the investment portfolio B which generates a random return of money Y with distribution G, to the investment portfolio A which generates a random return of money X with distribution F if and only if

$$\int_{-\infty}^{x} F(t)dt \geq \int_{-\infty}^{x} G(t)dt \text{ for all } x. \tag{1.1}$$

That is, if X denotes the random return of the investment portfolio A, while Y denotes the random return of the investment portfolio B, and X and Y have distribution functions given by F and G respectively, then, portfolio B is preferable to portfolio A, by every risk averter, if and only if $E(U(X)) \leq E(U(Y))$, for all increasing and concave utility functions U. Moreover, this criterion is equivalent to criterion (1.1). (See, e.g., Levy (1992)).

The relationship between F and G defined as (1.1) will be denoted $F <_{SSD} G$, and if X and Y are random variables with respective distributions F and G, we will also write $X <_{SSD} Y$. This order ($<_{SSD}$) has been employed by Joy and Porter (1974) to test the performance of mutual funds relative to the market as measured by the Dow-Jones Industrial Average. Applications of the ordering $<_{SSD}$ have also appeared in the diversification of independent portfolios. (See, e.g., Hong and Herk (1996) and references therein.)

El Barmi (1993) considered the problem of finding the maximum likelihood estimators of F and G when F and G satisfy (1.1) and F and G are discrete distributions, but the nonpararametric maximum likelihood estimators for general distribution functions satisfying (1.1) have not been derived. Deshpande and Singh (1985) constructed a test based on the empirical distribution function to test for second order stochastic dominance in the one-sample problem. Other references on testing for second order stochastic dominance include McFadden (1998), and Schmid and Trede (1998).

The purpose of this paper is to propose a family of new estimators in the one- and two-sample problems when F and G satisfy (1.1) and are arbitrary distribution functions. The organization of the paper is as follows. Section 2 considers the one-sample problem : strongly uniformly consistent estimators are provided and their weak convergence is discussed. For a certain class of loss functions, the new estimators dominate the empirical distribution function in terms of risk. The two-sample problem is considered in Section 3 : a family of strongly uniformly consistent estimators for F and G is proposed and their asymptotic distribution theory is established. Section 4 considers the case of censored data by replacing the empirical distribution functions F_n and G_m by their Kaplan-Meier counterparts. Section 5 discusses a family of new estimators in the case where the inequality in (1.1) is an identity for some values of x. Section 6 illustrates the estimators by examining a data set (Taylor (1997)) on total annual rate of return for bonds and stocks in The United States from 1810−1989. A simulation is presented in section 7 to compare the new family of estimators with the empirical distribution function in terms of Mean Squared Error. The simulation work suggests that the proposed estimators behave uniformly better than the empirical distribution function for a large class of examples. Technicalities have been relegated to an appendix.

2. One Sample Problem

Let $X_1, \ldots, X_n$ be a random sample from F and suppose that G is known, F and G satisfying (1.1). It is clear that in many situations the empirical distribution function F_n cannot satisfy (1.1) and therefore alternative estimators must be sought. For that purpose, define

$$H(x) = \int_{-\infty}^{x} F(t)dt. \tag{2.1}$$

Motivated by the work of Lo (1987), Rojo and Ma (1996) and Rojo (1995), define

$$H_n(x) = \int_{-\infty}^{x} F_n(t)dt, \tag{2.2}$$

$$\hat{H}_n(x) = \max(H_n(x),\ \int_{-\infty}^{x} G(t)dt). \tag{2.3}$$

It is easy to see that, under (1.1),

$$|\hat{H}_n(x) - H(x)| \le |H_n(x) - H(x)| \text{ for all } x. \tag{2.4}$$

Define now

$$\hat{F}_n(x) = \lim_{h\to 0^+} \frac{\hat{H}_n(x+h) - \hat{H}_n(x)}{h}, \tag{2.5}$$

so that $\hat{F}_n$ is the right derivative of $\hat{H}_n$. Note that $\hat{F}_n$ is well defined since $\hat{H}_n$ is the maximum of two differentiable functions.

Consider loss functions $L(\delta, F) = \upsilon(|\int_{-\infty}^{x} \delta(t)dt - \int_{-\infty}^{x} F(t)dt|)$, where $\upsilon(0) = 0$, and υ is nondecreasing on $(0, \infty)$. It follows from (2.4) that $\hat{F}_n$ dominates F_n as an estimator of F with respect to the class of loss functions $L(\cdot, \cdot)$.

The estimator $\hat{F}_n$ defined by (2.5) behaves better than F_n in terms of Mean Squared Error for many examples. Now define $x_0 \equiv -\infty$,

$$x_1 = \inf\{y \geq -\infty : \int_{-\infty}^{y} F_n(t)dt > \int_{-\infty}^{y} G(t)dt\},$$
$$x_2 = \inf\{y > x_1 : \int_{-\infty}^{y} F_n(t)dt < \int_{-\infty}^{y} G(t)dt\}.$$

Having defined $x_1, \ldots, x_{2k}$, let

$$x_{2k+1} = \inf\{y > x_{2k} : \int_{-\infty}^{y} F_n(t)dt > \int_{-\infty}^{y} G(t)dt\},$$
$$x_{2k+2} = \inf\{y > x_{2k+1} : \int_{-\infty}^{y} F_n(t)dt < \int_{-\infty}^{y} G(t)dt\},$$

and stop as soon as $x_j = +\infty$ for some j.

Thus, $\int_{-\infty}^{y} F_n(t)dt \leq \int_{-\infty}^{y} G(t)dt$ on the intervals $[x_{2k-2}, x_{2k-1}]$, while the inequality is reversed on the intervals $[x_{2k-1}, x_{2k}]$. It follows from (2.4), and (2.5) that

$$\hat{H}_n(x) = \sum_{j=1}^{\infty} \int_{-\infty}^{x} G(t) I_{\{[x_{2j-2}, x_{2j-1})\}} dt + \sum_{j=1}^{\infty} \int_{-\infty}^{x} F_n(t) I_{\{[x_{2j-1}, x_{2j})\}} dt, \tag{2.6}$$

$$\hat{F}_n(x) = \sum_{j=1}^{\infty} F_n(x) I_{\{[x_{2j-1}, x_{2j})\}} + \sum_{j=1}^{\infty} G(x) I_{\{[x_{2j-2}, x_{2j-1})\}}, \tag{2.7}$$

where, $I_{\{[\infty,\infty)\}} = 0$. It can be shown that $\hat{H}_n(x)$ is absolutely continuous and convex. As a result, $\hat{H}_n(x) = \int_{-\infty}^{x} \hat{F}_n(t)dt$ for all x, and $\hat{F}_n$ is nondecreasing. It is not difficult to see that $\hat{F}_n(x) \to 1$ as $x \to \infty$, and $\hat{F}_n(x) \to 0$ as $x \to -\infty$ (Rojo and El Barmi (2001)). Thus, $\hat{F}_n(x)$ is a legitimate distribution function that satisfies (1.1). Alternatively, one may write

$$\hat{F}_n(x) = F_n(x) I_{\{\int_{-\infty}^{x} F_n(t)dt \geq \int_{-\infty}^{x} G(t)dt\}} + G(x) I_{\{\int_{-\infty}^{x} F_n(t)dt < \int_{-\infty}^{x} G(t)dt\}}. \tag{2.8}$$

In this paper, attention will be focused on the case where (1.1) is a strict inequality for every x. A discussion on the strong uniform convergence and asymptotic distribution of similar estimators when (1.1) is an identity for some

x may be found at the end of the paper. A detailed account of these estimators when (1.1) is an identity for some x may be found in Rojo and El Barmi (2001).

Now if $\int_{-\infty}^{\infty}(F(x)(1-F(x)))^{1/2}dx < \infty$, then $\int_{-\infty}^{\infty}|F_n(x)-F(x)|dx = O_p(n^{-1/2})$, (see e.g., Serfling (1980)), and hence $I_{\{\int_{-\infty}^{x}F_n(t)dt<\int_{-\infty}^{x}G(t)dt\}}$ converges to zero in probability. It follows that $\hat{F}_n$ is weakly consistent for $F(x)$ when $\int_{-\infty}^{\infty}(F(x)(1-F(x)))^{1/2}dx < \infty$.

Under additional conditions, a stronger result is possible. The following conditions turn out to be sufficient for the strong uniform consistency of $\hat{F}_n$.

$$F \text{ has support } (-\infty,\infty) \text{ and for some } \eta > 1.5, F(x) = O((-x)^{-4-\eta}) \text{ as } x \to -\infty. \tag{2.9}$$

$$F \text{ is continuous with } 1-F(x) > 0 \text{ for every } x > 0. \tag{2.10}$$

$$F(x) < 1 \text{ for all } x, \text{ and for some } 0 < \delta < 1/2, E|X|^{2/(1-\delta)} < \infty. \tag{2.11}$$

Theorem 2.1. *Let $\hat{F}_n$ be defined by* (2.7), *or equivalently by* (2.8), *and suppose that $F <_{SSD} G$, where G is a known distribution function and F is continuous. Suppose that* (1.1) *holds strictly for every x. Then any of* (2.9), (2.10), *or* (2.11) *are sufficient for $\hat{F}_n$ to be strongly uniformly consistent for F.*

In terms of the asymptotic distribution, similar arguments show that under either (2.9), (2.10), or (2.11)

$$\sqrt{n}(\hat{F}_n(x)-F(x)) \xrightarrow{D} N(0,\ F(x)(1-F(x))). \tag{2.12}$$

A stronger result than (2.12) is possible. Weak convergence of the process $\{\sqrt{n}(\hat{F}_n(x)-F(x)),\ 0<x<1\}$, where $\hat{F}_n$ is defined by (2.8), can be demonstrated when (1.1) is a strict inequality for every x. This is the content of the following theorem which assumes that $F(x)=x,\ 0<x<1$, and G has support on (0,1). The general case follows as an easy corollary.

Theorem 2.2. *Suppose $F(x)=x,\ 0<x<1$, and suppose that G is a distribution function with support on $(0,1)$ and satisfying the condition* (1.1) *with strict inequality for every x. Let $\hat{F}_n$ be defined by* (2.8). *Then, the process $\{\sqrt{n}(\hat{F}_n-F),\ 0<x<1\}$ converges weakly to W°, where W° represents standard Brownian Bridge.*

In the general case, suppose that (1.1) still holds with strict inequality for every x, and consider

$$\begin{aligned}\sqrt{n}(\hat{F}_n(x)-F(x)) &= \sqrt{n}(F_n(x)-F(x))I_{\{\int_{-\infty}^{x}F_n(t)dt\geq\int_{-\infty}^{x}G(t)dt\}}\\ &\quad+\sqrt{n}(G(x)-F(x))I_{\{\int_{-\infty}^{x}F_n(t)dt<\int_{-\infty}^{x}G(t)dt\}}.\end{aligned} \tag{2.13}$$

Now, under any of the conditions (2.9), (2.10), or (2.11) on F it follows that $P\{\int_{-\infty}^x F_n(t)dt < \int_{-\infty}^x G(t)dt,\ i.o.\} = 0$. Therefore, the finite-dimensional distributions of the process $\{\sqrt{n}\ (\hat{F}_n(x) - F(x)),\ -\infty < x < \infty\}$ converge to the finite-dimensional distributions of $\{B(t),\ -\infty < t < \infty\}$, where $B(t)$ represents Brownian motion with $E(B(t)) = 0$ and $\mathrm{Cov}(B(s), B(t)) = \min(F(s), F(t)) - F(s)F(t)$.

Thus, to show that $\{\sqrt{n}(\hat{F}_n(x) - F(x)),\ -\infty < x < \infty\}$ converges weakly to $\{B(t),\ -\infty < t < \infty\}$, it is enough to prove tightness. However, arguments similar to the ones used to prove tightness in the previous theorem immediately guarantee tightness. This is summarized in the following.

Theorem 2.3. *Let F satisfy one of the conditions* (2.9), (2.10), *or* (2.11), *and suppose F and G satisfy* (1.1) *with strict inequality for every x, G known. Let $\hat{F}_n$ be defined by* (2.7) *or equivalently by* (2.8). *Then, the process $\{\sqrt{n}(\hat{F}_n(x) - F(x)),\ -\infty < x < \infty\}$ converges weakly to $\{B(x),\ -\infty < x < \infty\}$, where $B(x)$ denotes Brownian motion with $E(B(x)) = 0$ and Cov$(B(s), B(t)) = \min(F(s), F(t)) - F(s)F(t)$.*

In the next section, attention will be shifted to the case where G is also unknown.

3. Two Sample Problem

In this section, let $X_1, \ldots, X_n$ be a random sample from F, and let $Y_1, \ldots, Y_k$ be an independent random sample from G, where F and G satisfy (1.1).

It is of interest to estimate F and G subject to (1.1). For that purpose, define $\hat{G}_k(x) = G_k(x)$ as an estimator of G and

$$H^*_{n,k}(x) = \max\{\int_{-\infty}^x F_n(t)dt, \int_{-\infty}^x G_k(t)dt\}, \tag{3.1}$$

$$\hat{F}_{n,k}(x) = F_n(x) I_{\{\int_{-\infty}^x F_n(t)dt \geq \int_{-\infty}^x G_k(t)dt\}} + G_k(x) I_{\{\int_{-\infty}^x F_n(t)dt < \int_{-\infty}^x G_k(t)dt\}}, \tag{3.2}$$

where $G_k(\cdot)$ denotes the empirical cumulative distribution function based on $Y_1, \ldots, Y_k$. Clearly, $\hat{F}_{n,k}$ is the result of simply replacing G in (2.8) by the empirical distribution function G_k. It can be shown that $H^*_{n,k}$ is absolutely continuous, with $H^*_{n,k}(x) = \int_{-\infty}^x \hat{F}_{n,k}(t)dt$, so that $\int_{-\infty}^x \hat{F}_{n,k}(t)dt \geq \int_{-\infty}^x G_k(t)dt$. Moreover, using arguments similar to those used to prove that the estimator $\hat{F}_n$ is a distribution function, it can be shown that $\hat{F}_{n,k}$ is also a distribution function.

Now note that for each fixed x, the event $\{\int_{-\infty}^x F_n(t)dt < \int_{-\infty}^x G_k(t)dt\}$ is equal to the event $\{\int_{-\infty}^x (G_k(t) - G(t))dt + \int_{-\infty}^x (F(t) - F_n(t))dt > \int_{-\infty}^x (F(t) -$

$G(t))dt\}$. It follows that

$$\begin{aligned}
&|\hat{F}_{n,k}(x)-F(x)| \\
&\le |F_n(x)-F(x)|I_{\{\int_{-\infty}^x F_n(t)dt\ge\int_{-\infty}^x G_k(t)dt\}}+|G_k(x)-F(x)|I_{\{\int_{-\infty}^x F_n(t)dt<\int_{-\infty}^x G_k(t)dt\}} \\
&\le |F_n(x)-F(x)|I_{\{\int_{-\infty}^x F_n(t)dt\ge\int_{-\infty}^x G_k(t)dt\}} \\
&\quad +|G_k(x)-F(x)|(I_{\{\int_{-\infty}^x (G_k(t)-G(t))dt>h(x)/2\}}+I_{\{\int_{-\infty}^x (F(t)-F_n(t))dt>h(x)/2\}}),
\end{aligned}$$

where $h(x)=\int_{-\infty}^x(F(t)-G(t))dt>0$ for all x, whenever (1.1) holds with strict inequality for all x. Now if F satisfies any of (2.9), (2.10), or (2.11) then, as $n\to\infty$, $P(\int_{-\infty}^x(F(t)-F_n(t))dt> h(x)/2,\ i.o.)=0$, as demonstrated in the appendix. Similar assumptions on G then yield the result that $P(\int_{-\infty}^x(G_k(t)-G(t))dt>h(x)/2,\ i.o.)=0$. It follows from the above arguments that as $n,k\to\infty$, eventually with probability one for each x,

$$|\hat{F}_{n,k}(x)-F(x)|\le|F_n(x)-F(x)|I_{\{\int_{-\infty}^x F_n(t)dt\ge\int_{-\infty}^x G_k(t)dt\}}\le|F_n(x)-F(x)|. \tag{3.3}$$

The strong uniform consistency then follows from a lemma in Chung (1974).

Theorem 3.1. *Suppose that F and G satisfy* (1.1) *with strict inequality for all x, and suppose that F and G satisfy any of the conditions* (2.9), (2.10), *or* (2.11). *Let $\hat{F}_{n,k}$ be defined by* (3.2) *with F continuous. Then* $\lim_{n\to\infty}\lim_{k\to\infty}\sup_x|\hat{F}_{n,k}-F(x)|=0$.

We now discuss the weak convergence of the process $\{\sqrt{n}(\hat{F}_{n,k}(x)-F(x)),\ -\infty<x<\infty\}$. In what follows, $\lim_{n,k\to\infty}$ denotes the iterated limit $\lim_{n\to\infty}\lim_{k\to\infty}$.

Theorem 3.2. *Let F and G satisfy* (1.1) *with strict inequality for every x, and suppose that F satisfies one of the conditions* (2.9), (2.10), *or* (2.11), *similarly G. Then the process $\{\sqrt{n}(\hat{F}_{n,k}(x)-F(x)),\ -\infty<x<\infty\}$ converges weakly to $\{B(t),\ -\infty<t<\infty\}$ as $n,k\to\infty$, where $B(t)$ represents Brownian motion with $E(B(t))=0$ and* $\mathrm{Cov}\,(B(t),B(s))=\min(F(t),\ F(s))-F(t)F(s)$.

4. The Case of Censored Data

This case is handled by replacing the empirical distribution functions by the corresponding Kaplan-Meier estimators. Let $X_1,\ldots,X_n$ be a random sample from the distribution F. In the same set-up as in Csörgő and Horváth (1983), another random sample $Y_1,\ldots,Y_n$ with (left-continuous) distribution function H censors on the right the distribution F. As a result, the observations available consist of the pairs $(Z_i,\ \delta_i), i=1,\ldots,n$, where Z_i is the minimum of X_i and Y_i and δ_i is the indicator function of the event $(X_i\le Y_i)$. Let $\hat{F}_n$ denote the Kaplan-Meier estimator of F. For a probability distribution F^*, define $T_{F^*}=\inf\{t:$

$F^*(t) = 1\}$. The distribution W of Z_i is given by $1-W(t) = (1-F(t))(1-H(t))$ for each t. It is assumed throughout this section that

$$F \text{ and } H \text{ do not have jumps in common.} \tag{4.1}$$

Let $T^* = \min(T_F, T_H)$. Under (4.1), the strong uniform convergence of the Kaplan-Meier estimator $\hat{F}_n$ on $(-\infty, T^*]$ holds, as demonstrated by Stute and Wang (1993), if and only if

$$\text{either} \quad F\{T^*\} = 0 \quad \text{or} \quad F\{T^*\} > 0 \quad \text{but} \quad H(T^{*-}) < 1. \tag{4.2}$$

Under somewhat stricter conditions on F and H, the weak convergence of the process $\{\sqrt{n}(\hat{F}_n(x) - F(x)),\ -\infty < x < T\}$, for some T, has been proven by Breslow and Crowley (1974). Henceforth it is assumed, for clarity, that F and G are survival distributions. The results presented below in Theorems 4.1 and 4.2, however, hold more generally under conditions similar to (2.9) and (2.11). Consider the one-sample problem first, and define

$$\hat{F}_n^*(x) = \hat{F}_n(x) I_{\{\int_0^x \hat{F}_n(t)dt > \int_0^x G(t)dt\}} + G(x) I_{\{\int_0^x \hat{F}_n(t)dt \le \int_0^x G(t)dt\}}. \tag{4.3}$$

It can be seen that the estimator $\hat{F}_n^*$ defined by (4.3) is a legitimate distribution function.

Theorem 4.1. *Suppose that* (4.1) *and* (4.2) *hold. Then* $\sup_{0 \le x \le T^*} |\hat{F}_n^*(x) - F(x)| \to 0$ *with probability one. Moreover, if F and H are continuous and $T < T^*$ with $W(T) < 1$, then the process $\{\sqrt{n}(\hat{F}_n^*(x) - F(x)),\ 0 < x < T\}$ for $T < T^*$ converges weakly to the zero mean Gaussian process Z^* specified by Breslow and Crowley.*

In the two-sample problem, let $X_1, \ldots, X_n$ be a random sample from a distribution F and let $X_1^*, \ldots, X_n^*$ be a random sample from a distribution H_1 which censors F on the right, independent from $X_1, \ldots, X_n$. Also let $Y_1, \ldots, Y_k$ be a random sample from a distribution G and let $Y_1^*, \ldots, Y_k^*$ be a random sample independent from $Y_1, \ldots, Y_k$, from a distribution H_2 which censors G on the right. Let $T^* = \min(T_F, T_{H_1})$ and let $T_2^* = \min(T_G, T_{H_2})$. Let $\hat{F}_n$ and $\hat{G}_k$ denote the Kaplan-Meier estimators of F and G respectively. Suppose that H_1 and H_2 are left-continuous with F and G continuous. Let $T^* = \min(T_1^*, T_2^*)$. Define, on $(0, T^*)$,

$$\hat{F}_{n,k}^*(x) = \hat{F}_n(x) I_{\{\int_0^x \hat{F}_n(t)dt > \int_0^x \hat{G}_k(t)dt\}} + \hat{G}_k(x) I_{\{\int_0^x \hat{F}_n(t)dt \le \int_0^x \hat{G}_k(t)dt\}}. \tag{4.4}$$

The following result follows along the lines of the proof of Theorem 4.1, proof omitted.

Theorem 4.2. *Let $T < T^*$, and suppose F and G are continuous with H_1 and H_2 left-continuous. Let $\hat{F}^*_{n,k}$ be defined by* (4.4). *Then* $\lim_{n,k\to\infty}\sup_{0<t\le T^*}|\hat{F}^*_{n,k}(t)-F(t)| = 0$. *Moreover, the process $\{\sqrt{n}(\hat{F}^*_{n,k}(t) - F(t)), 0 < t < T\}$ converges weakly as $n, k \to \infty$ to the zero mean Gaussian process $\{Z^*(t), 0 < t < T\}$ as specified in Breslow and Crowley* (1974).

5. Extensions and Generalizations

The estimator $\hat{F}_n$ defined by (2.8) behaves better than F_n in terms of Mean Squared Error for many examples except that, on occasion, it may behave poorly in the left tail. This occurs because $\hat{F}_n$ will estimate F as G for all x in a neighborhood of $-\infty$. Moreover, the theory outlined in the previous sections for the case where (1.1) is a strict inequality for all x does not go through if (1.1) is an equality for some values of x. We briefly consider a family of estimators which satisfy (1.1) and contain $\hat{F}_n$ as a special case. The asymptotic theory for this family of new estimators parallels the theory discussed in previous sections, but will not be considered here. See Rojo and El Barmi (2001) for a detailed account. Some of the simulation results presented in Section 7 include a study of the estimators in this section and the example in the following section illustrates their use. Define

$$H^*_n(x) = \max\Big(\int_{-\infty}^{x} m_n(t)dt,\ \int_{-\infty}^{x} G(t)dt\Big), \tag{5.1}$$

where $m_n(t) = \min(F_n(t) + \gamma_n(t),\ 1)$, $\gamma_n(t) = \varepsilon_n r(t)$ where $r(t) > 0$ and $\int_{-\infty}^{x} r(t)dt < \infty$ for all x, $r(t)$ nondecreasing and continuous, and $\varepsilon_n \ge 0$ is a sequence of real numbers with $\varepsilon_n \to 0$, and where ε_n may be appropriately selected so that the asymptotic theory goes through. As our family of estimators, define

$$F^*_n(x) = \lim_{h\to 0^+}\frac{H^*_n(x+h) - H^*_n(x)}{h}. \tag{5.2}$$

Note that when $\varepsilon_n = 0$ for all n, $m_n(t) = F_n(t)$ and hence $H^*_n = \hat{H}_n$ and $F^*_n = \hat{F}_n$. Our simulation work will also show that F^*_n has smaller Mean Squared Error than both F_n and $\hat{F}_n$ in various examples. As in Section 2, we may define the intervals $[x_{2k-2}, x_{2k-1}]$ as those where $\int_{-\infty}^{y} m_n(t)dt \le \int_{-\infty}^{y} G(t)dt$, while the inequality is reversed on the intervals $[x_{2k-1}, x_{2k}]$. Then it follows from (5.1), and (5.2) that

$$H^*_n(x) = \sum_{j=1}^{\infty}\int_{-\infty}^{x} G(t)I_{\{[x_{2j-2},x_{2j-1})\}}dt + \sum_{j=1}^{\infty}\int_{-\infty}^{x} m_n(t)I_{\{[x_{2j-1},x_{2j})\}}dt, \tag{5.3}$$

$$F^*_n(x) = \sum_{j=1}^{\infty} m_n(x)I_{\{[x_{2j-1},x_{2j})\}} + \sum_{j=1}^{\infty} G(x)I_{\{[x_{2j-2},x_{2j-1})\}}, \tag{5.4}$$

where $I_{\{[\infty,\infty)\}} = 0$. It can be shown that $H_n^*(x)$ is absolutely continuous and convex. As a result, $H_n^*(x) = \int_{-\infty}^{x} F_n^*(t)dt$ for all x and F_n^* is nondecreasing. Since $r(t) > 0$ with $\int_{-\infty}^{x} r(t)dt < \infty$ for all x, it follows that $r(x) \to 0$ as $x \to -\infty$. Therefore $F_n^*(x) \to 0$ as $x \to -\infty$. It is also easy to see that $F_n^*(x) \to 1$ as $x \to \infty$. By (5.4) and the right continuity of $m_n(t)$ and $G(t)$, the right continuity of F_n^* follows. Thus, F_n^* is a legitimate distribution function which satisfies (1.1).

In case G is unknown, G may be replaced by G_k – the empirical distributon based on a sample from G – in both (5.3) and (5.4). Similarly, in the case of censored data, both F_n and G_k are replaced by their Kaplan-Meier counterparts. One consequence of the slight perturbation introduced into $\hat{F}_n$ by $\varepsilon_n r(t)$ to obtain F_n^*, is the separation of $\int_{-\infty}^{x} F_n(t)dt$ from $\int_{-\infty}^{x} G(t)dt$ for the values of x for which equality holds in (1.1). Then the asymptotic theory goes through as in Sections 2−4. Details about the strong uniform convergence and asymptotic distribution of the estimators when (1.1) is an equality for some values of x may be found in Rojo and El Barmi (2001).

6. Stocks and Bonds Annual Return Rate

It is well-established in the financial literature (e.g., Siegel (1995)), that stocks have outperformed bonds by providing higher returns in the past 200 years. In this section the estimators proposed in previous sections are applied to financial data on annual total return of stocks and bonds by decade from 1810-1989. The data comes from Global Financial Data (www.globalfindata.com), with permission. The table below summarizes annual total return of stocks and bonds by decade. Decade i indicates the decade starting during the year $1810 + i * 10$.

Table 1. Stocks/Bonds total return by decade.

Decade	1	2	3	4	5	6	7	8	9
Bonds	6.41	5.92	6.25	4.89	5.35	6.65	7.96	5.53	3.92
Stocks	2.68	5.31	4.53	6.73	0.45	15.73	7.58	6.72	5.45
Decade	10	11	12	13	14	15	16	17	18
Bonds	2.60	2.23	5.84	4.20	2.49	0.73	1.96	4.32	13.26
Stocks	9.62	4.69	13.86	-0.17	9.57	18.53	8.17	6.75	16.64

Investors are routinely advised that stocks not only outperform bonds over the long-term but, during the past century, stocks have been less volatile than bonds over extended periods of time, and they have been less likely to lose your money during any given decade.

Therefore a risk averter should invest in stocks rather than bonds. However, the data does not obey the second order stochastic dominance restriction between stocks and bonds as shown by Figure 1. Note that (1.1) is not satisfied for Total Annual Return in a neighborhood of zero. Figure 2 illustrates the new estimators satisfying the restriction that bonds $<_{SSD}$ stocks. The empirical distribution functions for bonds and stocks are plotted together with three estimators for F − the distribution function for bonds. The three estimators illustrated in Figure 2 were computed by using (5.4) with $\varepsilon_n = 0, \varepsilon_n = e^{-n}$, $\varepsilon_n = 1/n^2$. In all three cases, $r(t)$ was set to one. Note that the new estimators modify the empirical distribution function for bonds only in that region where the restriction is violated. Thus the new estimator for the distribution function for bonds with $\varepsilon_n = 0$ agrees with its empirical distribution function for Total Annual Return values greater than 3.525, while agreeing with the empirical distribution function for stocks for values of Total Annual Return smaller than 3.524. When $\varepsilon_n = e^{-n}$, the new estimator takes on the values $F_n + e^{-n}$ except on the interval from -1.699 to 3.525 where it agrees with the empirical distribution function G_k. Finally, the new estimator with $\varepsilon_n = 1/n^2$ takes on the values $F_n + 1/n^2$ except on the interval from -0.1198 to 3.402, where it agrees with G_k.

7. Simulation Work

With the purpose of studying the Mean Squared Error (MSE) behavior of the proposed estimators, and comparing this behavior to that of the empirical distribution function, simulations were performed for several examples. Ten thousand replications were run for each experiment. Note that (2.4) suggests that the estimator defined by (2.5) may be uniformly better than the empirical distribution function in the one-sample case. This provides the motivation for a closer examination of the Mean Squared Error properties of the estimators defined by (2.7), or equivalently (2.8), and (5.4) in the one-sample problem, and the estimators defined by (3.2) in the case of the two-sample problem. The case of censored data is considered as well. We compare the MSE of the Kaplan-Meier estimator with the new estimators defined by (4.3) and (4.4). Some simulation results are also included for the case where (1.1) does not hold strictly for all x. In these cases the estimator defined by (5.4), and its counterpart in the two-sample problem, were used for the simulations for various choices of ε_n: $\varepsilon_n = 0$, $\varepsilon_n = e^{-n}$, and $\varepsilon = 1/n^2$.

In the case of nonnegative random variables we used $r(t) = 1$ and, in the case of distributions with R^1 as their support, we used $r(t) = e^t I_{\{t<0\}} + I_{\{t>0\}}$. In the one-sample problem, since G is known, we have selected $r(t) = G(t)$ in a few cases. As the Mean Squared Error behavior typically depends on the tail behavior of the underlying distributions, probability distributions with various degrees of tail-heaviness were considered. (See, e.g., Rojo (1988), Rojo (1992) and Rojo (1996) for a method of classifying distributions by tail behavior). The examples presented here are typical of the results obtained. Interested readers may request all simulation results from the first author. The results suggest that the new estimators uniformly dominate the empirical distribution function in terms of MSE in all the examples considered in the one-sample case. In the two-sample case, the estimators compare well with the empirical distribution function (Kaplan-Meier) except in a neighborhood of zero. Comparing the new estimators among themselves, simulations show that, although there is no uniform winner, the new estimator with $\varepsilon_n = e^{-n}$ gives good results overall, and in some cases dominates uniformly the estimators with $\varepsilon_n = 0$ and $\varepsilon = 1/n^2$. The examples considered for the case of censored data similarly show that the estimator with $\varepsilon_n = e^{-n}$ has good overall results while being the best choice in some of the examples, although the improvement over the case with $\varepsilon_n = 0$ is not substantial, as expected. All in all, computational ease, availability of asymptotic distributional theory, and performance in terms of MSE, the new estimators seem to fare rather well. The simulation results are presented in Figure 3. The graphs show the ratio of the Mean Squared Errors of the new estimators to the Mean Squared Error of the empirical distribution function. In all cases, the solid black line indicates the ratio of MSE of the new estimator to the Empirical (or Kaplan-Meier estimator) when $\varepsilon_n = 0$. The red line indicates the ratio of the MSE of the new estimator, with $\varepsilon_n = e^{-n}$, to the MSE of the empirical distribution. The dashed line indicates the ratio of the MSE of the new estimator with $\varepsilon_n = 1/n^2$ to the MSE of the empirical distribution function.

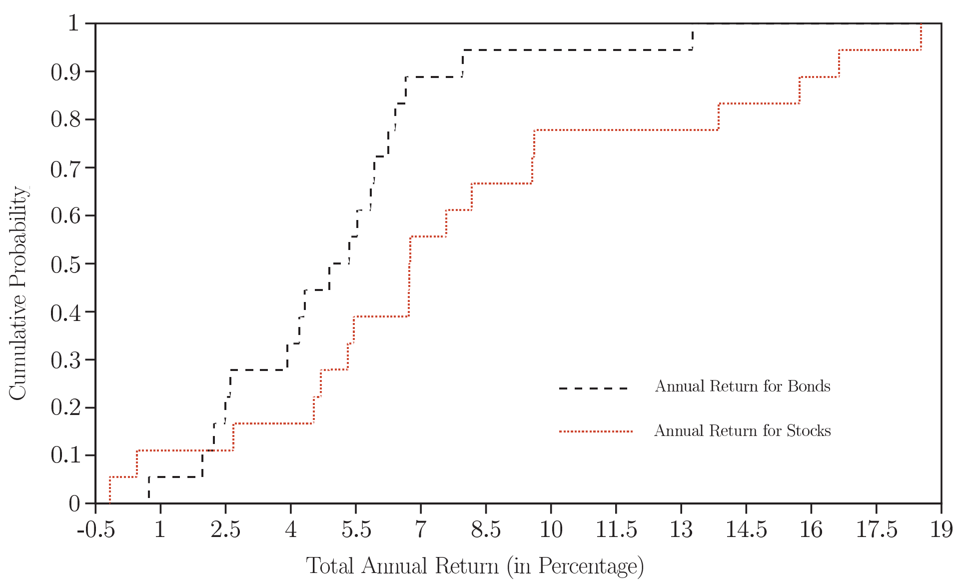

Figure 1. Empirical annual total return for stocks and bonds: 1810−1989.

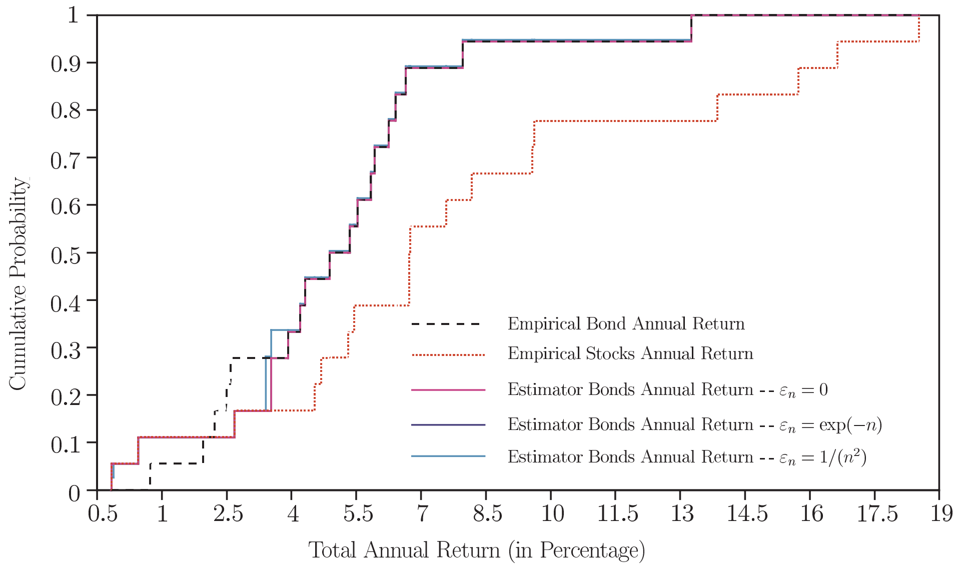

Figure 2. New estimators for annual total return for bonds: 1810−1989.

Table 2. Summary of distributions, sample sizes and functions $r(t)$.

Label	F-Censor Dist.	G-Censor Dist.	Sample Size	$r(t)$
1s-short-30-2	$1-1/(1+x)^7$	Exp(6)	30	1
1s-short-10-2G	″	″	10	G
2s-short-10-10-2	″	″	10,10	1
2s-short-30-30-2	″	″	30,30	1
Pareto-1s-30	Pareto, $\theta = 1.5$	Pareto, $\theta = 1.0$	30	1
Exp-1s-10	Exp, $\lambda = 1.5$	Exp, $\lambda = 1.0$	10	1
Exp-2s-50-30	″	″	50, 30	1
Normal-2s-30-30	$\mathrm{N}(0, 1.21)$	$\mathrm{N}(0, 1.0)$	30, 30	$e^t I_{\{t<0\}}$ $+I_{\{t>0\}}$
t-2s-30-30	t dist, 10 df	t dist, 4 df	30, 30	″
Dexp-1s-10	Double Exp, $\lambda = 1.5$	Double Exp, $\lambda = 1.0$	10	″
Dexp-2s-30-30	″	″	30, 30	″
cexp1s1030	Exp(1.5) - - Exp(1/6)	Exp(1.0)	10	1
cexp1s2510	Exp(1.5) - - Exp(1/2)	Exp(1.0)	10	1
cexp2s103030	Exp(1.5) - - Exp(1/6)	Exp(1.0) - - Exp(1/3)	30, 30	1
cexp2s103050	Exp(1.5) - - Exp(1/6)	Exp(1.0) - - Exp(1/3)	30, 50	1
cde2s3050	Dble Exp(1.5) - - Exp(3/13)	Dble Exp(1.0) - - Exp(1/4)	30, 50	$e^t I_{\{t<0\}}$ $+I_{\{t>0\}}$
cnormal2s3050	$N(0, 1.21)$ - - $1-\exp(-x^2)$	$N(0, 1)$ - - $1-\exp(-x^2)$	30, 50	″

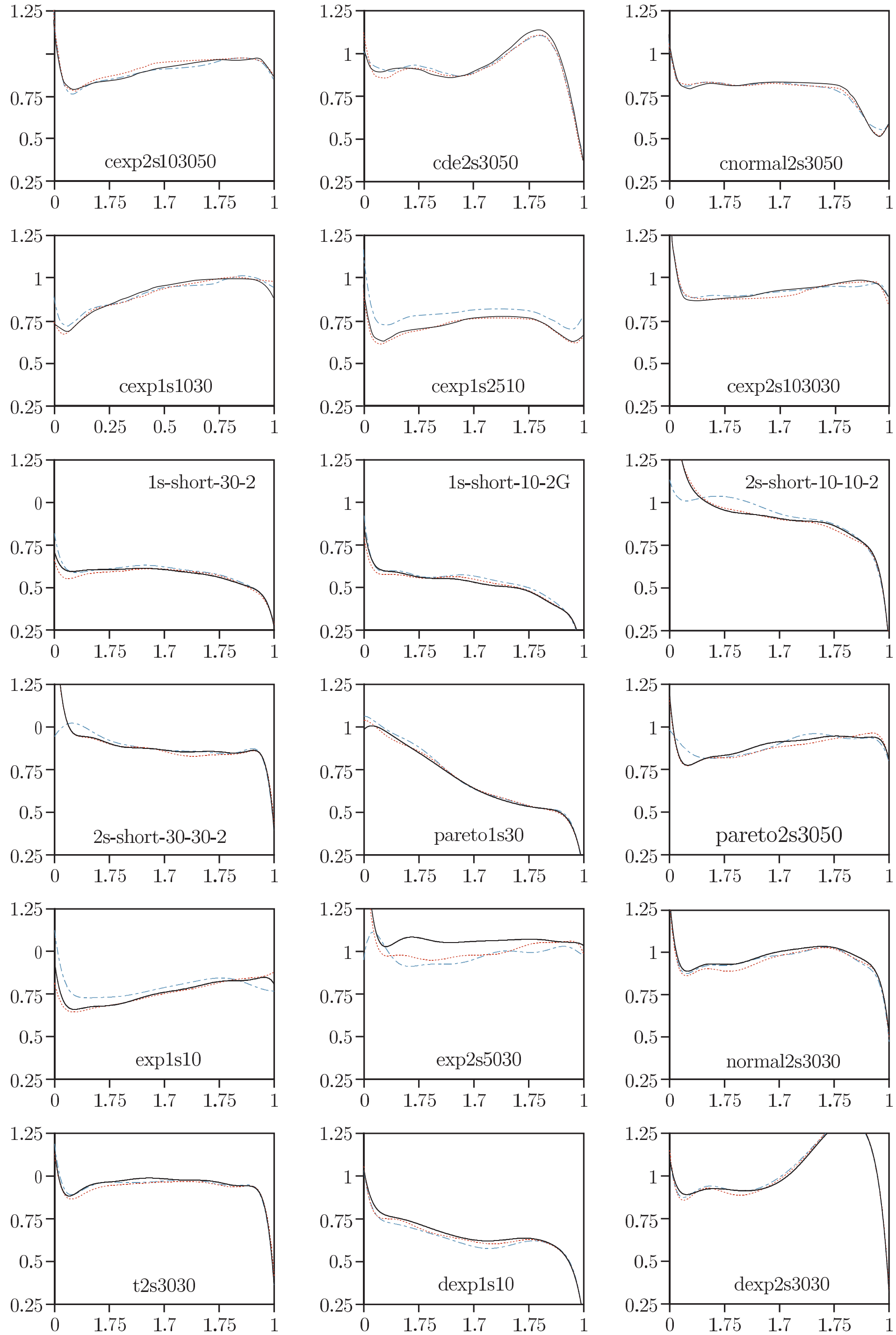

Figure 3. Simulation results for the censored and uncensored data cases.

Table 2 summarizes the distributions, sample sizes, and the parameters used for the simulations. In all examples the distribution F is the one being estimated. The table refers to the distribution function given by $F(x) = 1-x^{-\theta}$, $x > 1$ as the Pareto distribution with parameter $\theta > 0$. A "c" in front of the label indicates a censored case.

Acknowledgements

This research was supported in part by the NIH through the BioStatistical Laboratory, of which the first author was the Founding Director, by grant no. 5G12RR08124-05, and by NSF through grant no. RII-8802973. Support from NASA through Grant NRA 00-OEOP-1 is also gratefully acknowledged. Some of this work was conceived while the first author visited the Statistics Department at the University of California at Berkeley and during a visit of Dr. El Barmi to the BioStatistical Laboratory. A revision of this work was completed while the first author visited the Centro de Investigaciones Matemáticas (CIMAT) in Guanajuato, México, and while visiting the Statistics Department at SMU. The authors thank all institutions for their support. The Editor, an associate editor, and the referees made valuable comments which substantially improved the presentation.

Appendix

Proof of Theorem 2.1. Pointwise convergence with probability one is demonstrated first. Suppose that (2.9) holds. It follows from (2.8) that for each x,

$$|\hat{F}_n(x) - F(x)| = |F_n(x) - F(x)| I_{\{\int_{-\infty}^x F_n(t)dt \geq \int_{-\infty}^x G(t)dt\}} + |G(x) - F(x)| I_{\{\int_{-\infty}^x F_n(t)dt < \int_{-\infty}^x G(t)dt\}}. \tag{A.1}$$

The first term in the right side of (A.1) is bounded above by $|F_n(x) - F(x)|$, and hence converges to zero with probability one for each x. It follows that to show almost sure pointwise consistency of $\hat{F}_n$, it is enough to show that $I_{\{\int_{-\infty}^x F_n(t)dt < \int_{-\infty}^x G(t)dt\}}$ converges to zero with probability one for each x. Thus, it is enough to show that the probability of $A(x) = \{\int_{-\infty}^x F_n(t)dt < \int_{-\infty}^x G(t)dt, i.o.\}$ equals zero for every x. Note that

$$A_n(x) = \{\int_{-\infty}^x F_n(t)dt < \int_{-\infty}^x G(t)dt\} = \{\int_{-\infty}^x (F(t) - F_n(t))dt > \int_{-\infty}^x (F(t) - G(t))dt\}.$$

Therefore, $w \in \overline{\lim} A_n(x) = A(x)$ if and only if $\int_{-\infty}^x (F(t) - F_n(w,t))dt \geq \int_{-\infty}^x (F(t) - G(t))dt$ infinitely often. Here we have emphasized the dependence of F_n on w for clarity. Now consider

$$P\{\int_{-\infty}^x (F(t) - F_n(t))dt \geq h(x)\}, \tag{A.2}$$

where $h(x) = \int_{-\infty}^{x}(F(t) - G(t))dt > 0$. It follows from (A.2) that we only need to show

$$\sum_{n=1}^{\infty} P\{\int_{-\infty}^{x}(F(t) - F_n(t))dt \geq h(x)\} < \infty. \tag{A.3}$$

Let $B_n(x) = \{\int_{-\infty}^{x}(F(t) - F_n(t))dt \geq h(x)\}$, and write

$$\sum_{n=1}^{\infty} P\{B_n(x)\} = \sum_{n=1}^{\infty} P\{B_n(x), x < X_{(1)}\} + \sum_{n=1}^{\infty} P\{B_n(x), x \geq X_{(1)}\}.$$

The second sum above is bounded above by $\sum_{n=1}^{\infty} P\{x < X_{(1)}\} = \sum_{n=1}^{\infty}(\bar{F}(x))^n < \infty$. Therefore we need only to consider the third sum. Hence write, for $x > X_{(1)}$,

$$\int_{-\infty}^{x}(F(t) - F_n(t))dt \leq \int_{-\infty}^{X_{(1)}}(F(t) - F_n(t))dt + (x - X_{(1)}) \sup_x |F_n(x) - F(x)|.$$

Choose δ so that $3/(18 + 4\eta) < \delta < \eta/(2\eta + 8)$, and set $c = 1/2 - \delta$ and $l = (1 + c)/(4 + \eta)$. Then $0 < \delta < 1/2$ and, eventually, with probability one,

$$\begin{aligned}\int_{-\infty}^{X_{(1)}}(F(t) - F_n(t))dt &\leq (\sup_x |F(x) - F_n(x)|)^{1-l} \int_{-\infty}^{X_{(1)}} |F_n(t) - F(t)|^l dt \\ &\leq K(\sup_x |F(x) - F_n(x)|)^{\frac{3+\eta-c}{4+\eta}} \int_{-\infty}^{X_{(1)}} \frac{dt}{(-t)^{1+c}} \\ &= K(\sup_x |F(x) - F_n(x)|)^{\frac{3+\eta-c}{4+\eta}} /c(-X_{(1)})^c.\end{aligned}$$

Therefore eventually, with probability one, $\int_{-\infty}^{X_{(1)}}(F(t) - F_n(t))dt \leq (\sup_x |F(x) - F_n(x)|)^{\frac{3+\eta-c}{4+\eta}}(-X_{(1)})$. Hence, and abbreviating "infinitely often" to i.o.,

$$\begin{aligned}&P(\int_{-\infty}^{x}(F(t) - F_n(t))dt \geq h(x), x \geq X_{(1)}, i.o.) \\ &\leq P((\sup_y |F(y) - F_n(y)|)^{\frac{3+\eta-c}{4+\eta}}(-X_{(1)}) \geq h(x)/2, x \geq X_{(1)},\ i.o.) \\ &\quad +P((x - X_{(1)}) \sup_y |F(y) - F_n(y)|) \geq h(x)/2, i.o.).\end{aligned} \tag{A.4}$$

Now, by the conditions on δ and η, $n^c(\sup_x |F_n(x) - F(x)|)^{\frac{3+\eta-c}{4+\eta}}$ converges to zero with probability one. Also eventually, with probability one, $(x - X_{(1)}) \sup_y |F(y) - F_n(y)| \leq -2X_{(1)} \sup_y |F(y) - F_n(y)|$, and $n^c \sup_x |F_n(x) - F(x)|$ converges to zero with probability one. Therefore, to show that the right side of (A.4) is zero, it is enough to show that $P(-X_{(1)}n^{\delta-1/2} \geq h(x)/2,\ i.o.) = 0$. For that purpose, consider

$$\begin{aligned} P(-X_{(1)}n^{\delta-1/2} \geq h(x)/2) &= P(X_{(1)} < -n^{1/2-\delta}h(x)/2) \\ &= 1-(1-F(-n^{1/2-\delta}h(x)/2))^n \\ &= \sum_{j=0}^{n-1}(1-F(-n^{1/2-\delta}h(x)/2))^j F(-n^{1/2-\delta}h(x)/2) \\ &\leq nF(-n^{1/2-\delta}h(x)/2). \end{aligned}$$

Since $F(x) = O(|-x|^{-4-\eta})$ as $x \to -\infty$, $nF(-n^{1/2-\delta}h(x)/2) = nO((n^{1/2-\delta}h(x)/2)^{-4-\eta}) = M2^{4+\eta}n^{4\delta-\eta/2+\eta\delta-1}/(h(x))^{4+\eta}$, where $M < \infty$. Since $4\delta-\eta/2+\eta\delta-1 < -1$, it follows that $\sum_{n=1}^{\infty} P(-X_{(1)}n^{\delta-1/2} > h(x)/2) < \infty$, and hence the probability that $-X_{(1)}n^{\delta-1/2} > h(x)/2$ i.o. is zero. Therefore, $\overline{\lim}_{n\to\infty} A_n$ has probability zero, and strong pointwise convergence follows. The uniform convergence then follows from a lemma in Chung (1974) at p.133.

If (2.10) holds, the proof of (A.3) is immediate since, by a result of (DKW) Dvoretzky, Kiefer, and Wolfowitz (1956),

$$P\Big\{\int_0^x (F(t)-F_n(t))dt > h(x)\Big\} \leq P\{x \sup_t |F_n(t)-F(t)| > h(x)\} \leq e^{-2n(h(x)/x)^2}.$$

It follows that, $\sum_{n=1}^{\infty} P\{\sup_t |F_n(t)-F(t)| > h(x)\} \leq C\sum_{n=1}^{\infty} e^{-2n(h(x)/x)^2} < \infty$ and, as a consequence, $P\{\int_0^x F_n(t)dt < \int_0^x G(t)dt\ \ i.\ o.\} = 0$. Therefore, $\hat{F}_n(x) \to F(x)$ with probability one, for any $x > 0$. The result then follows by a lemma in Chung (1974) at p.133.

If (2.11) holds, it follows from (A.2) that it is enough to show $P\{\int_{-\infty}^x (F(t)-F_n(t))dt > h(x),\ i.o.\} = 0$. Note that $\int_{-\infty}^x (F(t)-F_n(t))dt = (1/n)\sum_{i=1}^n X_i I_{\{X_i \leq x\}} - E(XI_{\{X\leq x\}}) + x(F(x)-F_n(x))$. Thus it is enough to show that for every x, $P(|E(XI_{\{X\leq x\}}) - (1/n)\sum_{i=1}^n X_i I_{\{X_i\leq x\}}| \geq h(x)/2,\ i.o.) = 0$ and $P(|x(F(x)-F_n(x))| \geq h(x)/2,\ i.o.) = 0$. It follows from the DKW (1956) inequality that $P(|F(x)-F_n(x)| \geq h(x)/(2|x|),\ i.o.) = 0$ for every x. By a result due to Katz (see, e.g., Shorack and Wellner (1986, p.84)), if $E|X|^r < \infty$, then for $a > 1/2$ with $ra > 1$, *and* $S_n = \sum_{i=1}^n X_i$, where $X_1, X_2, \ldots$ are i.i.d. random variables,

$$\sum_{n=1}^{\infty} n^{ra-2}P(\frac{|S_n|}{n} > n^{a-1}\varepsilon) < \infty \text{ for all } \varepsilon > 0.$$

Setting $r = 2/(1-\delta), a = 1-\delta$, it follows that $P(|E(XI_{\{X\leq x\}}) - (1/n)\sum_{i=1}^n X_i I_{\{X_i\leq x\}}| \geq h(x)/2\ \ i.o.) = 0$, and hence, invoking the lemma at p.133 of Chung (1974), $\hat{F}_n$ is uniformly strongly consistent.

Proof of Theorem 2.2. Write $\sqrt{n}(\hat{F}_n(x) - F(x)) = \sqrt{n}(F_n(x) - x) I_{\{\int_0^x F_n(t)dt > \int_0^x G(t)dt\}} + \sqrt{n}(G(x) - x)I_{\{\int_0^x F_n(t)dt \leq \int_0^x G(t)dt\}}$. The fact that $P\{\int_0^x F_n(t)dt \leq \int_0^x G(t)dt,\ i.o.\} = 0$ for each fixed x, implies that the finite dimensional distributions of $\{\sqrt{n}(\hat{F}_n(x) - x),\ 0 < x < 1\}$ equal the finite-dimensional distributions of $\{\sqrt{n}(F_n(x) - x), 0 < x < 1\}$ eventually and with probability one, and these in turn converge to the finite-dimensional distributions of the process $\{W^\circ(x),\ 0 < x < 1\}$.

The result will follow if the sequence of stochastic processes $\{\sqrt{n}(\hat{F}_n(t) - t) : 0 < t < 1\}$ is tight. Following Billingsley (1968), p.137, define $W_n(t) = \sqrt{n}(\hat{F}_n(t) - t)$, $0 < t < 1$. Since $W_n(0) = 0$, it is enough to show that for every $\varepsilon^* > 0$, and $\eta > 0$, there exists a $0 < \delta^* < 1$, and an integer n_0 such that

$$P\{\sup_{t \leq s \leq t+\delta^*} |W_n(s) - W_n(t)| \geq \varepsilon^*\} \leq \eta\delta^* \tag{A.5}$$

for all $n \geq n_0$ and each $0 < t < 1$, where $t + \delta^*$ is replaced by 1 if $t + \delta^* > 1$. Now consider

$$\begin{aligned}
&\sup_{t \leq s \leq t+\delta^*} |\sqrt{n}(\hat{F}_n(s) - s) - \sqrt{n}(\hat{F}_n(t) - t)| \\
\leq &\sup_{t \leq s \leq t+\delta^*} |\sqrt{n}(F_n(s) - s)I_{\{\int_0^s F_n(x)dx > \int_0^s G(x))dx\}} \\
&-\sqrt{n}(F_n(t) - t)I_{\{\int_0^t F_n(x)dx > \int_0^t G(x)dx\}}| \\
&+ \sup_{t \leq s \leq t+\delta^*} \sqrt{n}|G(t) - t|I_{\{\int_0^t F_n(x)dx \leq \int_0^t G(x)dx\}} \\
&+ \sup_{t \leq s \leq t+\delta^*} \sqrt{n}|G(s) - s|I_{\{\int_0^s F_n(x)dx \leq \int_0^s G(x)dx\}}.
\end{aligned} \tag{A.6}$$

The function $g : \sqrt{n}(F_n(t) - t) \to \sqrt{n}(F_n(t) - t)I_{\{\int_0^t F_n(x)dx > \int_0^t G(x)dx\}}$ is continuous with respect to the sup norm and, since the sequence $\{\sqrt{n}(F_n(t) - t), 0 < t < 1\}$ satisfies (A.5), then the sequence $\{\sqrt{n}(F_n(t) - t)I_{\{\int_0^t F_n(x)dx > \int_0^t G(x)dx\}},\ 0 < t < 1\}$ also satisfies (A.5). Therefore, to prove tightness of $\{W_n(t),\ 0 < t < 1\}$, it is enough to show that the last two terms on the right side of the inequality (A.6) converge to zero in probability.

In fact, a stronger result holds. Note that

$$\begin{aligned}
&P\{\sup_{t \leq s \leq t+\delta^*} \sqrt{n}|G(t) - t|I_{\{\int_0^t F_n(x)dx \leq \int_0^t G(x)dx\}} > \varepsilon^*\} \\
&= P\{\sqrt{n}|G(t) - t|I_{\{\int_0^t F_n(x)dx \leq \int_0^t G(x)dx\}} > \varepsilon^*\} \\
&\leq P\{\int_0^t F_n(x)dx \leq \int_0^t G(x)dx\} = P\{\int_0^t (x - F_n(x))dx \geq \int_0^t (x - G(x))dx\}.
\end{aligned}$$

Since $\int_0^t (x-F_n(x))dx \le t\sup_{0\le x\le t}|x-F_n(x)|$ and $\int_0^t (x-G(x))dx > 0$ for every t, in fact $\sup_{t\le s\le t+\delta^*}\sqrt{n}|G(t)-t|I_{\{\int_0^t F_n(x)dx \le \int_0^t G(x)dx\}} \to 0$ with probability one. Similarly,

$$
\begin{aligned}
P\{\sup_{t\le s\le t+\delta^*}&\sqrt{n}|G(s)-s|I_{\{\int_0^s F_n(x)dx\le\int_0^s G(x)dx\}} > \varepsilon^*\}\\
&\le P\{\sup_{t\le s\le t+\delta^*}(\int_0^s G(x)dx - \int_0^s F_n(x)dx) \ge 0\}\\
&= P\{\sup_{t\le s\le t+\delta^*}\{-\int_0^s (x-G(x))dx + \int_0^s (x-F_n(x))dx\} \ge 0\}\\
&\le P\{\sup_{t\le s\le t+\delta^*}\int_0^s (G(x)-x)dx + \sup_{t\le s\le t+\delta^*}\int_0^s (x-F_n(x))dx \ge 0\}\\
&= P\{\sup_{t\le s\le t+\delta^*}\int_0^s (x-F_n(x))dx \ge \inf_{t\le s\le t+\delta^*}\int_0^s (x-G(x))dx\}. \qquad \text{(A.7)}
\end{aligned}
$$

Since $\int_0^s (x-G(x))dx$ is continuous on $[t,\ t+\delta^*]$, $\inf_{t\le s\le t+\delta^*}\int_0^s (x-G(x))dx = \int_0^{s^*}(x-G(x))dx$ with $\int_0^{s^*}(x-G(x))dx > 0$ for some s^* in $[t,\ t+\delta^*]$. Therefore the last term in the string of expressions given by (A.7) is bounded above by $P\{\sup_{t\le s\le t+\delta^*}\int_0^s (x-F_n(x))dx \ge \int_0^{s^*}(x-G(x))dx\}$ and, since $\sup_{t\le s\le t+\delta^*}\int_0^s (x-F_n(x))dx$ converges to zero with probability one, it follows that the last term in the right side of (A.6) converges to zero in probability. Thus the process $\{\sqrt{n}(\hat{F}_n(t)-t),\ 0<t<1\}$ converges weakly to $\{W^\circ(t),\ 0<t<1\}$ where W° denotes Brownian Bridge.

Proof of Theorem 2.3. This proof follows immediately from the proof of Theorem 2.2 and hence is not included here.

Proof of Theorem 3.1. The proof follows immediately from the discussion leading to (3.3) and the proof of Theorem 2.1.

Proof of Theorem 3.2. For fixed x as $n,k\to\infty$, the proof of the pointwise strong consistency of $\hat{F}_{n,k}(x)$ shows that eventually, with probability one, $\hat{F}_{n,k}(x) = F_n(x)$. This fact in turn implies that the finite-dimensional distributions of the process $\{\sqrt{n}(\hat{F}_{n,k}(x)-F(x)),\ -\infty<x<\infty\}$ converge to the finite-dimensional distributions of $\{B(t),-\infty<t<\infty\}$. It remains to prove the tightness of the process $\{\sqrt{n}(\hat{F}_{n,k}(x)-F(x)),-\infty<x<\infty\}$. Note that

$$
\begin{aligned}
\sqrt{n}(\hat{F}_{n,k}(t)-F(t)) &= \sqrt{n}(F_n(t)-F(t)) + \sqrt{n}(\hat{F}_{n,k}(t)-F_n(t))\\
&= \sqrt{n}(F_n(t)-F(t)) + \sqrt{n}(G_k(t)-F_n(t))I_{\{\int_{-\infty}^x F_n(t)dt\le\int_{-\infty}^x G_k(t)dt\}}.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
&\sup_{t\le s\le t+\delta}|\sqrt{n}(\hat{F}_{n,k}(s)-F(s))-\sqrt{n}(\hat{F}_{n,k}(t)-F(t))|\\
&\le \sup_{t\le s\le t+\delta}|\sqrt{n}(F_n(s)-F(s))-\sqrt{n}(F_n(t)-F(t))|\\
&\quad+\sup_{t\le s\le t+\delta}|\sqrt{n}(G_k(t)-F_n(t))|I_{\{\int_{-\infty}^t F_n(x)dx\le\int_{-\infty}^t G_k(x)dx\}}\\
&\quad-\sqrt{n}(G_k(s)-F_n(s))I_{\{\int_{-\infty}^s F_n(x)dx\le\int_{-\infty}^s G_k(x)dx\}}|.
\end{aligned}
$$

Since $\{\sqrt{n}(F_n(t)-F(t)),\ -\infty<t<\infty\}$ is tight, to prove tightness for $\{\sqrt{n}(\hat{F}_{n,k}(t)-F(t)),\ -\infty<t<\infty\}$ it is enough to prove that

$$
\begin{aligned}
\lim_{n\to\infty}\lim_{k\to\infty}\sup_{t\le s\le t+\delta}&|\sqrt{n}(G_k(t)-F_n(t))I_{\{\int_{-\infty}^t F_n(x)dx\le\int_{-\infty}^t G_k(x)dx\}}\\
&-\sqrt{n}(G_k(s)-F_n(s))I_{\{\int_{-\infty}^s F_n(x)dx\le\int_{-\infty}^s G_k(x)dx\}}|=0
\end{aligned}
$$

with probability one. To see this, note that

$$
\begin{aligned}
&\sup_{t\le s\le t+\delta}|\sqrt{n}(G_k(t)-F_n(t))I_{\{\int_{-\infty}^t F_n(x)dx\le\int_{-\infty}^t G_k(x)dx\}}\\
&\quad-\sqrt{n}(G_k(s)-F_n(s))I_{\{\int_{-\infty}^s F_n(x)dx\le\int_{-\infty}^s G_k(x)dx\}}|\\
&\le 2\sqrt{n}I_{\{\int_{-\infty}^t F_n(x)dx\le\int_{-\infty}^t G_k(x)dx\}}+2\sqrt{n}\sup_{t\le s\le t+\delta}I_{\{\int_{-\infty}^s F_n(x)dx\le\int_{-\infty}^s G_k(x)dx\}}. \qquad (A.8)
\end{aligned}
$$

Now, for fixed t, previous arguments yield the result that $\int_{-\infty}^t F_n(x)dx>\int_{-\infty}^t G_k(x)dx$ eventually with probability one, and therefore the first term on the right side of (A.8) is zero eventually and with probability one.

Consider now the second term on the right side of (A.8):

$$
\begin{aligned}
&P(2\sqrt{n}\sup_{t\le s\le t+\delta}I_{\{\int_{-\infty}^s F_n(x)dx\le\int_{-\infty}^s G_k(x)dx\}}\ge\varepsilon)\\
&=P(\sup_{t\le s\le t+\delta}(\int_{-\infty}^s F_n(x)dx-\int_{-\infty}^s G_k(x)dx)\le 0)\le P(\int_{-\infty}^t F_n(x)dx\le\int_{-\infty}^t G_k(x)dx).
\end{aligned}
$$

It follows that eventually, with probability one,

$$
\begin{aligned}
&\sup_{t\le s\le t+\delta}|\sqrt{n}(G_k(t)-F_n(t))I_{\{\int_{-\infty}^t F_n(x)dx\le\int_{-\infty}^t G_k(x)dx\}}\\
&\quad-\sqrt{n}(G_k(s)-F_n(s))I_{\{\int_{-\infty}^s F_n(x)dx\le\int_{-\infty}^s G_k(x)dx\}}|=0.
\end{aligned}
$$

Therefore, the sequence $\{\sqrt{n}(\hat{F}_{n,k}-F)\}_{n,k=1}^{\infty}$ is tight and weak convergence follows.

Proof of Theorem 4.1. We prove strong uniform convergence first. Note that for each $x < T^*$,

$$\begin{aligned}|\hat{F}_n^*(x) - F(x)| &= |\hat{F}_n(x) - F(x)| I_{\{\int_0^x \hat{F}_n(t)dt > \int_0^x G(t)dt\}} \\ &\quad + |G(x) - F(x)| I_{\{\int_0^x \hat{F}_n(t)dt \le \int_0^x G(t)dt\}} \\ &\le |\hat{F}_n(x) - F(x)| + I_{\{\int_0^x \hat{F}_n(t)dt \le \int_0^x G(t)dt\}}. \end{aligned} \tag{A.9}$$

Now for each $\varepsilon > 0$, eventually,

$$P(I_{\{\int_0^x \hat{F}_n(t)dt \le \int_0^x G(t)dt\}} \ge \varepsilon) \le P(x \sup_{0 \le t \le x} |F(t) - \hat{F}_n(t)| \ge \int_0^x (F(t) - G(t))dt). \tag{A.10}$$

Since $\int_0^x (F(t) - G(t))dt > 0$ and $\sup_{0 \le t \le x} |F(t) - \hat{F}_n(t)| \to 0$ with probability one, it follows that $P\{x \sup_{0 \le t \le x} |F(t) - \hat{F}_n(t)| \ge \int_0^x (F(t) - G(t))dt,\ i.o.)\} = 0$, and therefore $I_{\{\int_0^x \hat{F}_n(t)dt \le \int_0^x G(t)dt\}} = 0$ eventually with probability one. It follows from (A.9) that $|\hat{F}_n^*(x) - F(x)| \to 0$ with probability one. One more application of a Lemma of Chung (1974) yields the desired result.

Turning our attention to the weak convergence of $\{\sqrt{n}(\hat{F}_n^*(t) - F(t), -\infty < t < \infty\}$, let $0 = t_1 < t_2 < \cdots < t_k = T$, and consider the random vectors with components $(A_n(t_1), \ldots, A_n(t_k))$ and $(Z_n^*(t_1), \ldots, Z_n^*(t_k))$, where $Z_n^*(t) = \sqrt{n}(\hat{F}_n(t) - F(t))$, and $A_n(t) = \sqrt{n}(\hat{F}_n^*(t) - F(t)) = \sqrt{n}(\hat{F}_n(t) - F(t)) I_{\{\int_0^t \hat{F}_n(s)ds > \int_0^t G(s)ds\}} + (G(t) - F(t)) I_{\{\int_0^t \hat{F}_n(s)ds \le \int_0^t G(s)ds\}}$. As argued in the proof of the strong uniform convergence of $\hat{F}_n^*$, $I_{\{\int_0^t \hat{F}_n(s)ds \le \int_0^t G(s)ds\}}$ equals zero eventually with probability one. Therefore eventually, with probability one, $(A_n(t_1), \ldots, A_n(t_k)) = (Z_n^*(t_1), \ldots, Z_n^*(t_k))$. Therefore, the finite-dimensional distributions of $\{\sqrt{n}(\hat{F}_n^*(t) - F(t)), 0 < t < T\}$ converge to the finite-dimensional distributions of the process Z^*. It remains to prove tightness of the sequence $\{\sqrt{n}(\hat{F}_n^*(t) - F(t)), 0 < t < T\}$. For that purpose, let $\varepsilon > 0$, and $\eta > 0$. It is enough to show that there exists $\delta > 0$, and an integer n_0 such that

$$P(\sup_{t \le s \le t+\delta} |A_n(t) - A_n(s)| \ge \varepsilon) \le \eta\delta, \tag{A.11}$$

for all $n \ge n_0$, and each $0 < t < T - \delta$. Note that

$$\begin{aligned} &\sup_{t \le s \le t+\delta} |\sqrt{n}(\hat{F}_n^*(t) - F(t)) - \sqrt{n}(\hat{F}_n^*(s) - F(s))| \\ &\le \sup_{t \le s \le t+\delta} |\sqrt{n}(\hat{F}_n(t) - F(t)) - \sqrt{n}(\hat{F}_n(s) - F(s))| \\ &\quad + \sup_{t \le s \le t+\delta} |\sqrt{n}(\hat{F}_n(t) - F(t)) I_{\{\int_0^t \hat{F}_n(x)dx \le \int_0^t G(x)dx\}} \end{aligned}$$

$$
\begin{aligned}
&-\sqrt{n}(\hat{F}_n(s)-F(s))I_{\{\int_0^s \hat{F}_n(x)dx\le\int_0^s G(x)dx\}}| \\
&+\sup_{t\le s\le t+\delta}|\sqrt{n}(G(s)-F(s))I_{\{\int_0^s \hat{F}_n(x)dx\le\int_0^s G(x)dx\}} \\
&+\sqrt{n}|G(t)-F(t)|I_{\{\int_0^t \hat{F}_n(x)dx\le\int_0^s G(x)dx\}}. \qquad \text{(A.12)}
\end{aligned}
$$

An argument similar to arguments used previously, immediately shows that the last term on the right side of (A.12) is zero eventually with probability one. Also, for any $\varepsilon^* > 0$,

$$
\begin{aligned}
&P(\sup_{t\le s\le t+\delta} I_{\{\int_0^s \hat{F}_n(x)dx\le\int_0^s G(x)dx\}} > \varepsilon^*,\ i.o.) \\
&= P(\sup_{t\le s\le t+\delta}(\int_0^s G(x)dx-\int_0^s \hat{F}_n(x)dx)\ge 0,\ i.o.) \\
&= P(\sup_{t\le s\le t+\delta}(\int_0^s (G(x)-F(x))dx-\int_0^s(\hat{F}_n(x)-F(x))dx\ge 0,\ i.o.) \\
&\le P(\sup_{t\le s\le t+\delta}(\int_0^s (G(x)-F(x))dx+s\sup_{0\le x\le s}|\hat{F}_n(x)-F(x)|)\ge 0,\ i.o.) \\
&\le P((t+\delta)\sup_{0\le s\le T}|\hat{F}_n(x)-F(x)|\ge -a\ i.o.)=0,
\end{aligned}
$$

where $a=\sup_{t\le s\le t+\delta}\int_0^s(G(x)-F(x))<0$.

Therefore, eventually with probability one, the second and third terms on the right side of (A.12) are zero so that eventually with probability one,

$$
\begin{aligned}
&\sup_{t\le s\le t+\delta}|\sqrt{n}(\hat{F}_n^*(t)-F(t))-\sqrt{n}(\hat{F}_n^*(s)-F(s)| \\
\le &\sup_{t\le s\le t+\delta}|\sqrt{n}(\hat{F}_n(t)-F(t))-\sqrt{n}(\hat{F}_n(s)-F(s))|.
\end{aligned}
$$

It follows that, since $\{\sqrt{n}(\hat{F}_n(t)-F(t)),\ 0<t<T\}$ satisfies (7.11) then so does $\{\sqrt{n}(\hat{F}_n^*(t)-F(t)),\ 0<t<T\}$, and weak convergence follows.

References

Billingsley, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

Breslow, N. and Crowley, J. (1974). A large sample study of the life table and product limit estimators under random censorship. *Ann. Statist.* **2**, 437-453.

Chung, K. L. (1974). *A Course in Probability Theory*. Academic Press, New York.

Csörgő, S. and Horváth, L. (1983). The rate of strong uniform consistency for the product-limit estimator. *Z. Wahrsch verw. Gebiete* **62**, 411-426.

Deshpande, J. V. and Singh, H. (1985). Testing for second order stochastic dominance. *Comm. Statist. Theory Method* **14**, 887-893.

Dvoretzky, A., Kiefer, J. and Wolfowitz, J. (1956). Asymptotic minimax character of the sample distribution functions and of the classical multinomial estimator. *Ann. Math. Statist.* **27**, 642-669.

Dykstra, R. L. (1982). Maximum Likelihood Estimation of the survival functions of two stochastically ordered random variables. *J. Amer. Stat. Assoc.* **77**, 621-628.

El Barmi H. (1993). Inference under convex constraints. Unpublished Ph.D. dissertation, Dept. of Statistics, Univ. of Iowa.

Hadar, J. and Russell, W. R. (1971). Stochastic dominance and diversification. *J. Econom. Theory* **3**, 288-305.

Hong, C. S. and Herk, L. F. (1996). Incremental risk aversion and diversification preference. *J. Econom. Theory* **70**, 180-200.

Joy, 0. M., and Porter, R. B. (1974). Stochastic dominance and mutual fund performance. *J. Financial Quantitative Analy.* **9**, 25-31.

Kijima, M. and Ohnishi, M. (1996). Portfolio selection problems via the bivariate characterization of stochastic dominance relations. *Math. Finance* **6**, 237-277.

Lehmann, E. L. (1955). Ordered families of distributions. *Ann. Math. Statist.* **26**, 399-419.

Lehmann, E. L. and Rojo, J. (1992). Invariant directional orderings. *Ann. Statist.* **20**, 2100-2110.

Levy, H. (1992). Stochastic dominance and expected utility: survey and analysis. *Management Sci.* **38**, 555-593.

Lo, S. H. (1987). Estimation of distribution functions under order restrictions. *Statist. Decisions* **5**, 251-262.

McFadden, D. (1989). Testing for stochastic dominance. In *Studies in Economics of Uncertainty* (Edited by T. Fombay and T. K. Seo). New York.

Porter, R. B. and Carey, K. (1974). Stochastic dominance as a risk analysis criterion. *Decisions Sci.* **5**, 10-21.

Rojo, J. (1988). On the concept of tail-heaviness. Technical report 175, Department of Statistics, University of California, Berkeley.

Rojo, J. (1992). A pure-tail ordering based on the ratio of the quantile functions. *Ann. Statist.* **20**, 570-579.

Rojo, J. (1996). On tail categorization of probability laws. *J. Amer. Statist. Assoc.* **91**, 378-384.

Rojo, J. and Ma, Z. (1996). On the estimation of stochastically ordered survival functions. *J. Statist. Comput. Simulation* **55**, 1-21.

Rojo, J. (1995). On the weak convergence of certain estimators of stochastically ordered survival functions. *J. Nonparametr. Statist.* **4**, 349-363.

Rojo, J. and El Barmi, H. (2001). Estimation of distribution functions under second order stochastic dominance. Tech. report 01-1, Statistics Department, Rice University, Houston, TX.

Schmid, F. and Trede, M. (1998). A Kolmogorov-type test for second order stochastic dominance. *Statist. Probab. Lett.* **37**, 183-193.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. John Wiley and Sons, New York.

Shaked, M. and Shanthikumar, J. G. (1994). *Stochastic Orders and Their Applications*. Academic Press, San Diego.

Shorack, G. R. and Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics*. John Wiley and Sons, San Diego.

Stute, W. and Wang, J. L. (1993). The strong law under random censorship. *Ann. Stat.* **21**, 1591-1607.

Taylor, B. (1997). Global Financial Data. www.globalfindata.com.

Department of Statistics, MS-138, Rice University, 6100 Main Street, Houston, TX 77005, U.S.A.

E-mail: jrojo@stat.rice.edu

Baruch College, City University of New York, New York, NY 10010, 17 Lexington Ave., U.S.A.

E-mail: Hammou_ElBarmi@baruch.cuny.edu

(Received December 1999; accepted April 2003)