Estimation of two ordered bivariate mean residual life functions

Javier Rojo*, Musie Ghebremichael

Statistics Department, MS-138, Rice University, Houston, TX 77005, USA

Received 8 March 2004
Available online 25 May 2005

Abstract

Situations occur frequently in which the mean residual life (mrl) functions of two populations must be ordered. For example, if a mechanical device is improved, the mrl function for the improved device should not be less than that of the original device. Also, mrl functions for medical patients should often be ordered depending on the status of concomitant variables. This paper proposes non-parametric estimators of the bivariate mrl function under a mrl ordering. The estimators are shown to be asymptotically unbiased, strongly uniformly consistent and weakly convergent to a bivariate Gaussian process. The estimators are shown to be the projections, in a sense to be made precise, of the empirical mrl function onto an appropriate convex set of mrl functions. In the one-sample problem, the new estimators dominate the empirical mrl function in terms of risk with respect to a wide class of loss functions.

© 2005 Elsevier Inc. All rights reserved.

AMS 1991 subject classification: 62G20; 62E20; 62G30; 62G32; 60G15; 60G70; 60F15

Keywords: Survival function; Mean residual life function; Mean residual life ordering; Asymptotic properties; Admissibility; Loss function

1. Introduction

When assessing the length of remaining life for patients diagnosed with terminal diseases, the mean residual lifetime function, although mathematically equivalent to the survival
function, provides a more informative assessment. That is, a terminal patient finds a statement such as “On average you have \( x \) more years to live” more informative than one that evaluates his or her probability of living beyond a certain time point.

In this paper, we consider nonparametric estimators of the bivariate mean residual life (mrl) function under order restrictions. Many situations arise when bivariate lifetimes are of interest. For example, times to death or times to initial contraction of a disease may be of interest for litter pairs of rats or for twin studies in humans. The times to a certain level of deterioration or the times to reaction to a treatment may be of interest in pairs of lungs, kidneys, breasts, eyes or ears of humans. Because of the dependence between the event times, we use the bivariate mrl function.

Let \( X = (X_1, X_2) \) be a random vector representing the lifetimes of two individuals. That is, \( X \) is a random vector in the first quadrant \( Q = \{ (x_1, x_2) : x_i \geq 0, i = 1, 2 \} \) of \( \mathbb{R}^2 \). Let \( \underline{x} = (x_1, x_2) \) be a vector of non-negative real numbers. Denote by \( S(x_1, x_2) \) the joint survival function of \( X \). Then the bivariate mrl function at age \( \underline{x} \) is given by

\[
e(\underline{x}) = E [ X - \underline{x} \mid X > \underline{x} ] = (e_1(x_1, x_2), e_2(x_1, x_2)), \tag{1.1}
\]

where

\[
e_1(x_1, x_2) = E [ X_1 - x_1 \mid X > \underline{x} ] = \frac{\int_{x_1}^{\infty} S(u, x_2) du}{S(x_1, x_2)}, \tag{1.2}
\]

and

\[
e_2(x_1, x_2) = E [ X_2 - x_2 \mid X > \underline{x} ] = \frac{\int_{x_2}^{\infty} S(x_1, v) dv}{S(x_1, x_2)} \tag{1.3}
\]

for all \((x_1, x_2)\) for which \( S(x_1, x_2) > 0 \). Otherwise define \( e(\underline{x}) = 0 \). Arnold and Zahedi [1] provided the inversion formula to express the joint survival function in terms of the multivariate mrl function. They also derived the relation between the multivariate mrl function and the hazard gradient and provided a characterization of multivariate lack-of-memory in terms of the multivariate mrl function. From their inversion formula, it is possible to obtain the following relationship between the bivariate mrl function and the survival function:

\[
S(x_1, x_2) = \frac{e_2(x_1, 0)e_1(0, 0)}{e_2(x_1, x_2)e_1(x_1, 0)} \exp \left( - \int_{0}^{x_2} \frac{dv}{e_2(x_1, v)} - \int_{0}^{x_1} \frac{du}{e_1(u, 0)} \right) \\
= \frac{S_1(x_1)e_2(x_1, 0)}{e_2(x_1, x_2)} \exp \left( - \int_{0}^{x_2} \frac{dv}{e_2(x_1, v)} \right) \\
= \frac{S_2(x_2)e_1(0, x_2)}{e_1(x_1, x_2)} \exp \left( - \int_{0}^{x_1} \frac{du}{e_1(u, x_2)} \right), \tag{1.4}
\]

where the last identity follows by symmetry. Let \((X_{1j}, X_{2j})\), \( j = 1, 2, \ldots, n \) be independent and identically distributed (iid) random vectors representing the lifetimes of \( n \) pairs of individuals and let \( S_n(x_1, x_2) \) denote the empirical bivariate survival function. The bivariate empirical mrl function at a point \((x_1, x_2)\) is then given by \( \hat{e}_n(x_1, x_2) \)
and later appeared also in Kulkarni and Rattihalli [7]:

When \( X_i \) follows that the bias of the empirical estimator is squared error and renders it inadmissible with respect to a wide class of loss functions. It is demonstrated that the new estimator is the projection of the empirical bivariate mean residual lifetime estimation of a bivariate mrl function when it is bounded above by a known mrl function. The main goal of this paper is to provide estimators for the bivariate mean residual lifetime function when it is bounded above by a known mrl function.

The expected value of the empirical mrl estimator was derived by Zahedi [16] and later appeared also in Kulkarni and Rattihalli [7]:

\[
E \left[ \hat{e}_{i,n}(x_1, x_2) \right] = \left\{ 1 - \left[ 1 - S(x_1, x_2) \right]^n \right\} e_i(x_1, x_2). \tag{1.7}
\]

It follows that the bias of the empirical estimator is \(-e_i(x_1, x_2) \left( 1 - S(x_1, x_2) \right)^n\) and hence \(\hat{e}_{i,n}(x_1, x_2)\) is asymptotically unbiased, with bias decaying exponentially to zero as \(n \to \infty\). When \(E[X_i^2] < \infty\), Zahedi [16] also provided the variance of \(\hat{e}_{i,n}(x_1, x_2)\):

\[
\text{Var} \left[ \hat{e}_{i,n}(x_1, x_2) \right] = \left\{ 1 - S(x_1, x_2) \right\}^n \left\{ 1 - \left[ 1 - S(x_1, x_2) \right]^n \right\} e_i^2(x_1, x_2)
+ \text{Var} \left[ X_i - x_i \mid (X_1, X_2) \geq (x_1, x_2) \right] \sum_{j=1}^{n} \frac{1}{j} B(n, j, S(x_1, x_2)), \tag{1.8}
\]

where \(B(n, j, S(x_1, x_2)) = \binom{n}{j} [S(x_1, x_2)]^j [1 - S(x_1, x_2)]^{n-j}\). Therefore, \(\text{Var} \left[ \hat{e}_{i,n}(x_1, x_2) \right] \to 0\) as \(n \to \infty\) when \(E[X_i^2] < \infty\).

Let \(F(x)\) be a distribution function on \(\mathbb{R}^2\) with corresponding mrl function \(e(x)\), and let \(F_n(x)\) be the bivariate empirical distribution function on \(\mathbb{R}^2\) with corresponding bivariate mrl function \(\hat{e}_n(x)\). Zahedi [16, Theorem 5.2.3], shows that the estimator \(\hat{e}_n(x)\) is a pointwise strongly consistent estimator of \(e(x)\). Kulkarni and Rattihalli [7] demonstrated the strong uniform consistency of \(\hat{e}_n(x_1, x_2)\) on bounded rectangles. They also showed the weak convergence of the suitably normalized bivariate mrl process to a bivariate Gaussian process with mean 0 and a certain covariance matrix \(\Sigma\).

The main goal of this paper is to provide estimators for the bivariate mean residual lifetime function when a partial ordering between mean residual functions obtains. The finite sample properties of these estimators as well as their asymptotic distributions are delineated. The organization of the paper is as follows: Section 2 considers the problem of nonparametric estimation of a bivariate mrl function when it is bounded above by a known mrl function. It is demonstrated that the new estimator is the projection of the empirical bivariate mean residual function onto a convex set of mrl functions in a sense to be made precise later. As a consequence, the new estimator dominates the empirical mrl function in terms of mean squared error and renders it inadmissible with respect to a wide class of loss functions.
Section 3 considers the same problem except that the bounding function is an unknown mrl function. In Sections 2 and 3, the asymptotic theory of the estimators is developed. The asymptotic results show that the new estimator is asymptotically unbiased, strongly uniformly consistent on bounded rectangles and converges weakly to a bivariate Gaussian process. The simulation results illustrate that, in the case considered in Section 2, the new estimator has uniformly smaller mean squared error than the empirical estimator for all distributions from which we simulated. These simulation results are as expected given the theoretical results given here. In addition, the simulations show that for small \( n \) and large times the new estimator has smaller absolute bias than the empirical estimator. The “mirror image” problem where the mrl function of interest is bounded from below by another mrl function which can be known or unknown follows easily from the theoretical results given here. In addition, the simulations show that for small \( n \) and large times the new estimator has smaller absolute bias than the empirical estimator.

In the sequel, \( \| x_1, x_2 \| \) will denote the norm of \( (x_1, x_2) \) defined by \( \| (x_1, x_2) \| = \max \{ |x_1|, |x_2| \} \). Moreover, for functions \( f = (f_1, f_2) \) with \( f_i : \mathbb{R}^2_+ \rightarrow \mathbb{R}_+ \) where \( \mathbb{R}^2_+ = \{ (x, y) : x \geq 0 \text{ and } y \geq 0 \} \) and \( \mathbb{R}_+ = \{ x : x \geq 0 \} \), we will define \( \| f \|_p \) by \( \| f \|_p = \left( \int_0^\infty \int_0^\infty \max(f_1^p, f_2^p) \, dx \, dy \right)^{\frac{1}{p}} \) for \( p \geq 1 \). All the technical details have been relegated to an appendix.

2. The one-sample problem

Suppose that \( X = (X_1, X_2) \) and \( Y = (Y_1, Y_2) \) are random vectors with finite means representing lifetimes of two populations with distribution functions \( F(x_1, x_2) \) and \( G(y_1, y_2) \); survival functions \( \overline{F}(x_1, x_2) \) and \( \overline{G}(y_1, y_2) \); and mrl functions \( e(x_1, x_2) = (e_1(x_1, x_2), e_2(x_1, x_2)) \) and \( m(y_1, y_2) = (m_1(y_1, y_2), m_2(y_1, y_2)) \), respectively. Let \( \mathcal{X} = \{ (x_1, x_2) : \overline{F}(x_1, x_2) > 0 \text{ and } \overline{G}(x_1, x_2) > 0 \} \). For vectors \( a = (a_1, a_2) \) and \( b = (b_1, b_2) \), \( a \leq b \) will denote the order in \( \mathbb{R}^2 \) defined by \( a_1 \leq b_1 \) and \( a_2 \leq b_2 \); \( a \wedge b = \min(a, b) = (a_1 \wedge b_1, a_2 \wedge b_2) \).

Consider the problem of estimating \( e(x_1, x_2) \) when \( e(x_1, x_2) \leq m(x_1, x_2) \) for all \( (x_1, x_2) \), and where \( m(x_1, x_2) \) is a known mean residual function. Motivated by the work of Rojo [11,12] and Rojo and Ma [13], we propose as an estimator of \( e(x_1, x_2) \) the following:

\[
e_n^*(x_1, x_2) = \hat{e}_n(x_1, x_2) \wedge m(x_1, x_2), \tag{2.1}
\]

where \( \hat{e}_n(x_1, x_2) \) is the bivariate empirical mrl function defined by (1.5) and (1.6). Thus, our estimator modifies the empirical mrl function only when it violates the restriction, in which case, it replaces the empirical mrl function by the “benchmark” mrl function \( m(x_1, x_2) \). As a consequence of this, it is not difficult to show that \( e_n^* \) is the projection of the empirical estimator, in a way to be made precise later, onto the set of bivariate mrl functions bounded above by \( m(x_1, x_2) \). It is not obvious that \( e_n^* \) is necessarily a mrl function. That is, it is not clear that there exists a survival function \( S_n^* \) for which (1.4) holds with \( e_1(x, y), e_2(x, y) \) replaced by \( e_{n1}^*(x, y), e_{n2}^*(x, y) \). Nair and Nair [10] and Kulkarni and Rattihalli [6] provided necessary and sufficient conditions for a nonnegative function \( \tilde{m}(x, y) \) to be a proper mrl function. The latter authors note that Nair and Nair’s results...
are incomplete as their conditions are not necessary and sufficient. Both of these results, however, require certain smoothness conditions on the mrl function which do not hold in the present context. Nevertheless, defining \(S^*_n(x, y)\) as in (1.4), with \(e(x, y)\) replaced by \(e_n^*(x, y)\) gives a proper survival function. This fact can be shown by checking directly that \(S^*_n(x, y)\) thus defined is nonincreasing in each coordinate, assigns positive mass to every rectangle, and goes to zero as \(x, y \to \infty\), with \(S^*_n(0, 0) = 1\).

Returning our attention to \(e_n^*\), it is clear that \(e_n^*\) is more negatively biased than \(\hat{e}_n\). Nevertheless, \(e_n^*\) has uniformly smaller mean square error than \(\hat{e}_n\) as stated in the next result, whose proof hinges on the fact that for every \((x_1, x_2)\) and for every \(n\), \(e_n^*\) is closer to \(e\) than \(\hat{e}_n\) is to \(e\).

**Theorem 2.1.** For every \((x_1, x_2)\), the restricted estimator \(e_n^*(x_1, x_2)\) has smaller mean squared error than the empirical mrl estimator \(\hat{e}_n(x_1, x_2)\). That is,

\[
E \left[ \sum_{j=1}^{2} (e_{j,n}^* - e_j)^2 \right] \leq E \left[ \sum_{j=1}^{2} (\hat{e}_{j,n} - e_j)^2 \right].
\]

An examination of the proof of Theorem 2.1 immediately shows that the empirical mrl function \(\hat{e}_n(x_1, x_2)\) is rendered inadmissible with respect to any loss function of the form \(L(e, \hat{e}) = v(\|e - \hat{e}\|)\) with \(v(0) = 0\) and \(v(x)\) nondecreasing on \((0, \infty)\), since it is dominated in risk by the estimator \(e_n^*(x_1, x_2)\). As it turns out, \(e_n^*(x_1, x_2)\) is the projection of the empirical mrl onto the convex set of mrl functions \(k(x_1, x_2)\) such that \(k(x_1, x_2) \leq m(x_1, x_2)\). The interpretation of our estimator as a projection onto an appropriate convex set is provided by the following result:

**Theorem 2.2.** Let \(A\) be the convex set of all mrl functions bounded above by a known mrl function \(m(x_1, x_2)\). That is, let \(A = \{k(x_1, x_2) : k\ is\ a\ mrl\ and\ k(x_1, x_2) \leq m(x_1, x_2)\ for\ all\ (x_1, x_2) \in \mathcal{X}\}\). Let \(e_n^*(x_1, x_2) = \hat{e}_n(x_1, x_2) \land m(x_1, x_2)\). Then,

(i) For any \(k \in A\), \(\sup_{0 \leq (x_1, x_2) \in \mathcal{X}} \|e_n^*(x_1, x_2) - \hat{e}_n(x_1, x_2)\| \leq \sup_{0 \leq (x_1, x_2) \in \mathcal{X}} \|k(x_1, x_2) - \hat{e}_n(x_1, x_2)\|\).

(ii) For \(p \geq 1\) and for all \(k \in A\), \(\|e_n^* - \hat{e}_n\|_p \leq \|k - \hat{e}_n\|_p\).

Theorem 2.1 attests to the superiority of the estimator defined in (2.1) when compared to the empirical mrl function for finite sample sizes. We next concentrate on the asymptotic properties of \(e_n^*\). It turns out that \(e_n^*\) is asymptotically unbiased, uniformly and strongly consistent, and converges weakly to a bivariate Gaussian process. We first discuss the bias of \(e_n^*\). In what follows let \(e_1(x, y) = E[X - x|X > x, Y > y]\), \(e_2(x, y) = E[Y - y|X > x, Y > y]\), \(\sigma_1^2(x, y) = E[(X - e_1(x, y))^2|X > x, Y > y]\), and \(\sigma_2^2(x, y) = E[(Y - e_2(x, y))^2|X > x, Y > y]\), the latter two assumed finite for all \(x, y\). The asymptotic unbiasedness of \(e_n^*\) is a direct consequence of the dominated convergence theorem, but under the additional assumptions of \(\sigma_1^2(x, y)\) and \(\sigma_2^2(x, y)\) being finite, a bound on the difference of the biases of \(\hat{e}_n\) and \(e_n^*\) can be obtained.
Theorem 2.3. The restricted estimator in (2.1) is asymptotically unbiased. That is

$$E\left[\hat{e}_n^*(x_1, x_2)\right] \rightarrow e(x_1, x_2) \quad \text{as } n \rightarrow \infty$$

for all \((x_1, x_2) \in \mathcal{X}\). Moreover, when \(\sigma_1^2(x, y)\) and \(\sigma_2^2(x, y)\) are finite for any \(x, y\),

$$E[\hat{e}_{1n}(x, y)] - E[\hat{e}_{2n}(x, y)] \leq 2\sigma_2^2(x, y)/(n+1)S(x, y)(m_1(x, y) + x - e_1(x, y))$$

and

$$E[\hat{e}_{2n}(x, y)] - E[\hat{e}_{2n}(x, y)] \leq 2\sigma_2^2(x, y)/(n+1)S(x, y)(m_2(x, y) + y - e_2(x, y))$$

so that Bias\((e_n^*)\) \(\rightarrow 0\) exponentially as well.

Thus, although \(e_n^*\) is more negatively biased than \(\hat{e}_n\), the bias of \(e_n^*\) also goes to zero. The following result states the strong uniform convergence of \(e_n^*\) on closed rectangles:

Theorem 2.4. The restricted estimator in (2.1) is strongly uniformly consistent on any finite rectangle. That is, for fixed \((b_1, b_2) \in \mathcal{X}\) and \(D = [0, b_1] \times [0, b_2]\),

$$\sup_{(x_1, x_2) \in D} \left\| e_n^*(x_1, x_2) - e(x_1, x_2) \right\| \rightarrow 0$$

with probability one as \(n \rightarrow \infty\).

Kulkarni and Rattihalli [7] demonstrated that the process

$$\left\{ \sqrt{n} \left[ \hat{e}_n(x_1, x_2) - e(x_1, x_2) \right], (x_1, x_2) \in \mathcal{X} \right\}$$

converges weakly to a bivariate Gaussian process with mean vector \(0\) and some covariance matrix \(\Sigma\). Since the estimator \(e_n^*(x_1, x_2)\) is pointwise closer to \(e(x_1, x_2)\) than \(\hat{e}_n(x_1, x_2)\) is to \(e(x_1, x_2)\) for every \(n\), it is expected that \(e_n^*\) suitably normalized can also be shown to converge weakly to a Gaussian process. For that purpose, define \(Z_n^*\) by

$$Z_n^*(x_1, x_2) = \sqrt{n} \left[ e_n^*(x_1, x_2) - e(x_1, x_2) \right] = \sqrt{n} \left[ \hat{e}_n(x_1, x_2) \wedge m(x_1, x_2) - e(x_1, x_2) \right].$$

In what follows weak convergence is denoted by \(\Rightarrow\). The following result establishes the weak convergence of \(Z_n^*\) to a bivariate Gaussian process.

Theorem 2.5. Let \((X_1, X_2)\) and \((Y_1, Y_2)\) have continuous mean residual lifetime functions given by \(e(x_1, x_2)\) and \(m(x_1, x_2)\), respectively, and let \(\bar{F}(x, y)\) and \(\bar{G}(x, y)\) denote their corresponding continuous survival functions. Define \(\mathcal{X} = \{(x, y) : \bar{F}(x, y) > 0\text{ and }\bar{G}(x, y) > 0\}\) and let \(e(x_1, x_2) \leq m(x_1, x_2)\) where \(m(x_1, x_2)\) is known. Let \((Z_1(x_1, x_2), x_1 > 0, x_2 > 0)\) denote the bivariate Gaussian process obtained in Kulkarni and Rattihalli [7] as the weak limit of the bivariate empirical mrl process.

(i) If \(e < m\) on \(\mathcal{X}\), then \(Z_n^* \Rightarrow Z\) on \(\mathcal{X}\).
(ii) If \(e = m\) on \(\mathcal{X}\), then \(Z_n^* \Rightarrow Z \wedge 0\) on \(\mathcal{X}\).
(iii) If for \(i = 1\) or \(i = 2\), \(e_i(x_0, 0) = m_i(x_0, 0)\) for some \((x_0, 0) \in \mathcal{X}\) and \(e_i(x, y) < m_i(x, y)\) for all \((x, y)\) in the line segment \((x_0, 0) + (1 - z)(x_1, y_1)\), \(0 < z < 1\) for some \((x_1, y_1) \in \mathcal{X}\), then \(\left\{ e_n^*(x, y), (x, y) \in \mathcal{X} \right\}\) does not converge weakly with rate \(\sqrt{n}\).
Note that part (ii) of the above theorem provides the tools for testing the null hypothesis that $e = m$ against the alternative that $e < m$. To test this hypothesis, let

$$D_n^- = \max \left( \sup_{(x_1, x_2) \in \mathcal{X}} [m_1(x_1, x_2) - e_{1,n}^*(x_1, x_2)], \sup_{(x_1, x_2) \in \mathcal{X}} [m_2(x_1, x_2) - e_{2,n}^*(x_1, x_2)] \right)$$

and reject the hypothesis that $e = m$ if $D_n^- > K_\alpha$ where $K_\alpha$ is the $(1 - \alpha) \times 100$ quantile of the distribution of $\max \{ \sup_{(x_1, x_2) \in \mathcal{X}} \max(Z_1, 0), \sup_{(x_1, x_2) \in \mathcal{X}} \max(Z_2, 0) \}$ where $Z_i, i = 1, 2$, is the limiting Gaussian process derived from $\sqrt{n} \left[ \hat{e}_i - e_i \right]$.

As a final remark in this section, it is also possible to find applications where the mean residual function of interest is bounded from below by a known mean residual function $m(x, y)$. With obvious changes in notation, (e.g. changing $\wedge$ to $\vee$), all the results in this section may be repeated, almost verbatim, to apply in the case of $e(x, y) \geq m(x, y)$, Ghebremichael [3]. These results, however, will not be included here.

### 3. The two-sample problem

In this section, we consider the problem of nonparametric estimation of a bivariate mrl function $e$ when it is bounded above by another unknown mrl function $m$. Let $(X_{1i}, X_{2i}), i = 1, \ldots, n_1$ and $(Y_{1j}, Y_{2j}), j = 1, \ldots, n_2$, be random vectors with finite means representing the lifetimes of two populations with distribution functions $F(x, y)$ and $G(x, y)$; survival functions $S^e(x, y)$ and $S^m(x, y)$; and mrl functions $e_1(x, y)$ and $e_2(x, y)$, respectively. Given $(X_{1i}, X_{2i}), i = 1, \ldots, n_1$ and $(Y_{1j}, Y_{2j}), j = 1, \ldots, n_2$ it is of interest to estimate $e$ or $m$, or both, subject to the restriction that $e \leq m$. Consider the empirical survival functions $S_{n_1}^e$ and $S_{n_2}^m$. One possible approach is to define the “pooled” survival function $S^* = (n_1 S_{n_1}^e + n_2 S_{n_2}^m) / (n_1 + n_2)$ and obtain the corresponding empirical mrl function defined by $e^* = (e_1^*, e_2^*)$ where

$$e_1^*(x, y) = \int_x^\infty \left[ \frac{n_1 S_{n_1}^e(u, y) + n_2 S_{n_2}^m(u, y)}{n_1 S_{n_1}^e(x, y) + n_2 S_{n_2}^m(x, y)} \right] du$$

and similarly

$$e_2^*(x, y) = \frac{w_1(x, y) \hat{e}_1(x, y) + w_2(x, y) \hat{m}_1(x, y)}{w_1(x, y) + w_2(x, y) \hat{m}_2(x, y)}$$

where $\hat{e} = (\hat{e}_1(x, y), \hat{e}_2(x, y))$ and $\hat{m} = (\hat{m}_1(x, y), \hat{m}_2(x, y))$ are the empirical mrl functions corresponding to $S_{n_1}^e$ and $S_{n_2}^m$, respectively, and

$$w_1(x, y) = \frac{n_1 S_{n_1}^e(x, y)}{n_1 S_{n_1}^e(x, y) + n_2 S_{n_2}^m(x, y)}, \quad w_2(x, y) = \frac{n_2 S_{n_2}^m(x, y)}{n_1 S_{n_1}^e(x, y) + n_2 S_{n_2}^m(x, y)}.$$
Unfortunately, although this approach still provides better estimators than the empirical mrl function in terms of mean squared error, by choosing a different set of weights \( w_i, i = 1, 2 \), one can improve on the resulting estimators derived from (3.1)–(3.3). Note that, alternatively, we could define our estimators for \( e \) and \( m \) as follows:

\[
\begin{align*}
    e^{**}(x, y) &= \min(\hat{e}(x, y), \hat{m}(x, y)), \\
    m^{**}(x, y) &= \max(\hat{e}(x, y), \hat{m}(x, y)).
\end{align*}
\]  

(3.4)

It turns out that both \( e^{**}(x, y) \) and \( m^{**}(x, y) \) are uniformly strongly consistent on closed and bounded rectangles when both \( n_1 \) and \( n_2 \) → ∞ as stated precisely in the following theorem:

**Theorem 3.1.** Let \( e^{**} \) and \( m^{**} \) be defined as in (3.4) and suppose that \( e(x, y) \leq m(x, y) \) for all \( (x, y) \). Then, with probability one,

\[
\begin{align*}
    \lim_{n_1 \to \infty} \lim_{n_2 \to \infty} \sup_{0 \leq (x_1, x_2) \in \mathcal{X}} \|e^{**}(x_1, x_2) - e(x_1, x_2)\| &= 0, \text{ and} \\
    \lim_{n_1 \to \infty} \lim_{n_2 \to \infty} \sup_{0 \leq (x_1, x_2) \in \mathcal{X}} \|m^{**}(x_1, x_2) - e(x_1, x_2)\| &= 0.
\end{align*}
\]

One drawback of the estimators defined by (3.4) is that although they are strongly uniformly consistent on closed and bounded rectangles when both \( n_1, n_2 \to \infty \), \( e^{**} \) fails to be consistent if \( n_2 \to \infty \) and \( n_1 \) remains bounded. (A similar phenomenon has been observed by Rojo [12] in the estimation of stochastically ordered distribution functions).

To circumvent the above problems, we propose estimators similar to those suggested by (3.1)–(3.3), except that we pool the empirical mean residual lifetime functions using weights \( w_i = \frac{n_i}{n_1 + n_2} \), \( i = 1, 2 \), so that our “benchmark” mrl function is now defined by the pooled estimate

\[
\hat{e}_p(x, y) = w_1 \hat{e}(x, y) + w_2 \hat{m}(x, y)
\]

and our estimators of \( e \) and \( m \) are given, respectively, by

\[
\begin{align*}
    \hat{e}^*(x, y) &= \min(\hat{e}_p(x, y), \hat{e}(x, y)), \\
    \hat{m}^*(x, y) &= \max(\hat{e}_p(x, y), \hat{m}(x, y)).
\end{align*}
\]

(3.5)

(3.6)

These estimators have the property, as will be demonstrated, that \( \hat{e}^*(x, y) \) converges to \( e(x, y) \) almost surely and uniformly on closed bounded rectangles when \( n_1 \to \infty \); similarly, \( \hat{m}^*(x, y) \) is strongly uniformly consistent on closed bounded rectangles when \( n_2 \to \infty \). In addition, because of the simplicity of the weights \( w_i, i = 1, 2 \), the finite sample and asymptotic theories of these estimators are obtained. Results from simulation work, not presented here, suggest that the estimators defined by (3.1)–(3.3) are easily dominated in mean squared error by estimators (3.5) and (3.6). For these reasons, we focus our attention on the estimators \( \hat{e}^* \) and \( \hat{m}^* \) defined by (3.5) and (3.6). As in Section 2, one issue that arises here is whether convex combinations of mrl functions continue to be mrl functions. In the present context, that is, the case of discrete probability distributions, it is possible to define a discrete probability measure supported by the combined sample of size \( n_1 + n_2 \).
For any \( F \) or any \( k(x, y) \) from the mrl function and the empirical mrl onto the convex set of functions \( S(x, y) \) given by (1.1)–(1.3), is equal to the convex combination of the original mrl functions. This can be done by showing that since the mass associated to the point \( (x, y) \) is given by \( S(x, y) + S(x^-, y^-) - S(x^-, y) - S(x, y^-) \), this mass can be obtained directly from the mrl function and the \( S(u, v), u \geq x, v \geq y \).

It follows immediately from (3.5) and (3.6) that \( \hat{e}^* \leq \hat{e}_p \leq \hat{m}^* \) and, for \( i = 1, 2 \),

\[
\hat{e}^*_i (x, y) = \hat{e}_i (x, y)I\{\hat{e}_i (x, y) \leq \hat{m}_i (x, y)\} + \hat{e}_{ip} (x, y) I\{\hat{e}_i (x, y) > \hat{m}_i (x, y)\},
\]

(3.7)

\[
\hat{m}^*_i (x, y) = \hat{m}_i (x, y)I\{\hat{m}_i (x, y) \geq \hat{e}_i (x, y)\} + \hat{e}_{ip} (x, y) I\{\hat{m}_i (x, y) < \hat{e}_i (x, y)\},
\]

(3.8)

This representation leads immediately to the study of the finite-sample and asymptotic properties of our estimators.

The following two theorems show that \( \hat{e}^* (x, y) \) is the projection of the empirical mrl onto the convex set of functions \( k(x, y) \) such that \( k(x, y) \leq \hat{e}_p (x, y) \) and \( \hat{m}^* (x, y) \) is the projection of the empirical mrl onto the convex set of functions \( k(x, y) \) such that \( k(x, y) \geq \hat{e}_p (x, y) \). These results parallel the results obtained in the one-sample case.

**Theorem 3.2.** Let \( A \) be the set of all mrl functions bounded above by the mrl function \( \hat{e}_p (x_1, x_2) \). That is, let \( A = \{k(x_1, x_2); k(x_1, x_2) \leq \hat{e}_p (x_1, x_2) \text{ for all } (x_1, x_2) \in \mathcal{X}\} \). Let \( \hat{e}^* (x_1, x_2) = \hat{e}_p (x_1, x_2) \wedge \hat{e} (x_1, x_2) \). Then,

(i) For any \( k \in A \), \( \sup_{0 \leq (x_1, x_2) \in \mathcal{X}} \|\hat{e}^* (x_1, x_2) - \hat{e} (x_1, x_2)\| \leq \sup_{0 \leq (x_1, x_2) \in \mathcal{X}} \|k(x_1, x_2) - \hat{e} (x_1, x_2)\| \).

(ii) For any \( k \in A \) and for \( p \geq 1 \), \( \|\hat{e}^* - \hat{e}\|_p \leq \|k - \hat{e}\|_p \).

Similar results hold for the estimator \( \hat{m}^* \) as given in the following theorem:

**Theorem 3.3.** Let \( B \) be the set of all mrl functions bounded below by the mrl function \( \hat{e}_p (x_1, x_2) \). That is, let \( B = \{h(x_1, x_2); h(x_1, x_2) \geq \hat{e}_p (x_1, x_2) \text{ for all } (x_1, x_2) \in \mathcal{X}\} \). Let \( \hat{m}^* (x_1, x_2) = \hat{m} (x_1, x_2) \vee \hat{e}_p (x_1, x_2) \). Then

(i) For any \( h \in B \), \( \sup_{0 \leq (x_1, x_2) \in \mathcal{X}} \|\hat{m}^* (x_1, x_2) - \hat{m} (x_1, x_2)\| \leq \sup_{0 \leq (x_1, x_2) \in \mathcal{X}} \|h(x_1, x_2) - \hat{m} (x_1, x_2)\| \).

(ii) For any \( h \in B \) and for \( p \geq 1 \), \( \|\hat{m}^* - \hat{m}\|_p \leq \|h - \hat{m}\|_p \).

We now turn our attention to the asymptotic properties of \( \hat{e}^* \) and \( \hat{m}^* \). Asymptotic unbiasedness and normality as well as strong uniform consistency of \( \hat{e}^* \) and \( \hat{m}^* \) are established in the following results:

**Theorem 3.4.** The estimators \( \hat{e}^* \) and \( \hat{m}^* \) are asymptotically unbiased. That is,

\[
E[\hat{e}^* (x_1, x_2)] \rightarrow e(x_1, x_2) \quad \text{as } n_1 \rightarrow \infty \quad \text{and}
\]

\[
E[\hat{m}^* (x_1, x_2)] \rightarrow m(x_1, x_2) \quad \text{as } n_2 \rightarrow \infty \quad \text{for all } (x_1, x_2) \in \mathcal{X}.
\]
The strong uniform convergence of both \( \hat{e} \) and \( \hat{m} \) on closed bounded rectangles follows almost immediately from representations (3.7) and (3.8) and the fact that \( \hat{e} \) and \( \hat{m} \) converge strongly and uniformly to \( e \) and \( m \), respectively, on closed bounded rectangles. To establish the strong uniform convergence, let \((b_1, b_2) \in \mathcal{X}\) be fixed, and let \( D = [0, b_1] \times [0, b_2] \).

**Theorem 3.5.** The estimators \( \hat{e} \) and \( \hat{m} \) are uniformly strongly consistent on \( D \). That is:

\[
\sup_{(x_1, x_2) \in D} \| \hat{e}(x_1, x_2) - e(x_1, x_2) \| \to 0 \quad \text{as} \quad n_1 \to \infty \quad \text{and} \\
\sup_{(x_1, x_2) \in D} \| \hat{m}(x_1, x_2) - m(x_1, x_2) \| \to 0 \quad \text{as} \quad n_2 \to \infty.
\]

Shifting attention to the asymptotic behavior of the mrl processes, define the bivariate restricted mrl processes \( Z^*_e \) and \( Z^*_m \) by

\[
Z^*_e(x_1, x_2) = \sqrt{n_1} \left[ \hat{e}(x_1, x_2) - e(x_1, x_2) \right] = \sqrt{n_1} \left[ \hat{e}(x_1, x_2) - \hat{e}_p(x_1, x_2) - e(x_1, x_2) \right],
\]

and

\[
Z^*_m(x_1, x_2) = \sqrt{n_2} \left[ \hat{m}(x_1, x_2) - m(x_1, x_2) \right] = \sqrt{n_2} \left[ \hat{m}(x_1, x_2) - \hat{m}_p(x_1, x_2) - m(x_1, x_2) \right].
\]

Similar to the one-sample case, it is possible to obtain the weak convergence of the processes defined through (3.9) and (3.10).

The following theorem provides the asymptotic theory of the mrl processes defined through \( \hat{e} \) and \( \hat{m} \). In what follows, let \( C_1(x, y) \) and \( C_2(x, y) \) represent the covariance functions obtained by plugging into (11a)–(11e) of [7], the quantities corresponding to the mrl functions \( e \) and \( m \).

**Theorem 3.6.** Let \( S^e(x, y) \) and \( S^m(x, y) \) be continuous survival functions and let \( \mathcal{X} = \{(x, y): S^e(x, y) > 0 \text{ and } S^m(x, y) > 0\} \). Let \( \{Z_e(x, y) \text{ and } Z_m(x, y) x > 0, y > 0\} \) denote mean zero Gaussian processes with covariance functions \( C_1(x, y) \) and \( C_2(x, y) \). Suppose that \( \frac{n_2}{n_1} \to \alpha \) with \( 0 \leq \alpha \leq \infty \) and \( e(x, y) \leq m(x, y) \) for all \((x, y) \in \mathcal{X}\) where both \( e \) and \( m \) are unknown.

(i) Suppose that \( e \leq m \) on \( \mathcal{X} \). Then

- if \( n_1 \to \infty \) with \( \alpha = 0 \), \( Z^*_e \Rightarrow Z_e \); if \( n_1 \to \infty \) with \( \alpha = \infty \), \( Z^*_e \Rightarrow Z^+_e \),
- if \( n_2 \to \infty \) with \( \alpha = \infty \), \( Z^*_m \Rightarrow Z_m \); if \( n_2 \to \infty \) with \( \alpha = 0 \), \( Z^*_m \Rightarrow Z^+_m \),

where \( f^+ = \max(0, f) \), \( f^- = \min(0, f) \).

(ii) If \( e < m \) on \( \mathcal{X} \), then as \( n_1 \to \infty \) (\( n_2 \to \infty \)) \( Z^*_e \Rightarrow Z_e \) (\( Z^*_m \Rightarrow Z_m \)).

(iii) If \( e = m \) on \( \mathcal{X} \) with \( 0 < \alpha < \infty \), then

\[
Z^*_e \Rightarrow \min \left( Z_e, \frac{\sqrt{\alpha}}{1 + \alpha} Z_m + \frac{1}{1 + \alpha} Z_e \right) \text{ on } \mathcal{X} \quad \text{and} \\
Z^*_m \Rightarrow \max \left( Z_m, \frac{\sqrt{\alpha}}{1 + \alpha} Z_e + \frac{\alpha}{1 + \alpha} Z_m \right) \text{ on } \mathcal{X}.
\]
(iv) If for sequences \( \{ \hat{c}_i \} \) that pairs and the denominator is the product-limit estimator for \( S(x, y) \) and the numerator is the empirical estimator for the survival function of the observed distributed censoring times with survival function \( G(t) = P(C \geq t) \), suppose that the two sequences \( \{(X_i, Y_i)\}_{i=1}^{n} \) and \( \{(C_i)\}_{i=1}^{n} \) are independent. In the univariate censorship model, the \( (X_i, Y_i) \) are censored on the right by the single censoring variable \( C_i \), so that we observe the random vectors \( (\tilde{X}_i, \tilde{Y}_i, \delta_i^x, \delta_i^y) \), \( i = 1, 2, \ldots, n \), where \( \tilde{X}_i = X_i \land C_i \), \( \tilde{Y}_i = Y_i \land C_i \), \( \delta_i^x = I(X_i \leq C_i) \) and \( \delta_i^y = I(Y_i \leq C_i) \). The survival function of the observed pairs \( \{ \tilde{X}_i, \tilde{Y}_i \} \) is \( S(x, y)G(x \lor y) \), which is a simple consequence of the independence between \( (X, Y) \) and \( C \). Thus, under univariate censoring, it is natural to estimate the survival function \( S(x, y) \) by

\[
\hat{S}_n(x, y) = \frac{n^{-1} \sum_{i=1}^{n} I(\tilde{X}_i \geq x, \tilde{Y}_i \geq y)}{\tilde{G}(x \lor y)},
\]

where the numerator is the empirical estimator for the survival function of the observed pairs and the denominator is the product-limit estimator for \( G(.) \). Lin and Ying [8] showed that \( \hat{S}_n(x, y) \) is strongly consistent, and upon proper normalization, converges weakly to a zero mean Gaussian process for all \((x, y) \in [0, \tau]^2 \), where \( \tau \) satisfies \( S(\tau, \tau)G(\tau) > 0 \). Using \( \hat{S}_n(x, y) \), one can estimate the bivariate mrl function \( e(x, y) = (e_1(x, y), e_2(x, y)) \) by \( \hat{e}_n(x, y) = (\hat{e}_{1,n}(x, y), \hat{e}_{2,n}(x, y)) \) where

\[
\hat{e}_{i,n}(x, y) = \begin{cases} 
\hat{S}_n(x, y) & \text{if } i = 1, \\
\hat{S}_n(x, y) & \text{if } i = 2.
\end{cases}
\]

Jeong et al. [4] showed weak uniform consistency and weak convergence of \( \hat{e}_n(x, y) \) on bounded rectangles under the assumption that both \( \sqrt{n} \int_{\tilde{X}_n} \Delta_S(u, y)du \) and \( \sqrt{n} \int_{\tilde{Y}_n} \Delta_S(x, v)dv \) converge to zero in probability, where \( \tilde{X}_* = \max(\tilde{X}_1, \tilde{X}_2, \ldots, \tilde{X}_n) \) and \( \tilde{Y}_* = \max(\tilde{Y}_1, \tilde{Y}_2, \ldots, \tilde{Y}_n) \).

Now suppose that \( e(x, y) \leq m(x, y) \) for all \( (x, y) \in [0, \tau]^2 \) for some known bivariate mrl function \( m(x, y) \). As an estimator of \( e(x, y) \) under the constraint \( e(x, y) \leq m(x, y) \), we propose

\[
e_n^*(x, y) = \min(\hat{e}_n(x, y), m(x, y)),
\]

where \( \hat{e}_n(x, y) \) is the estimator of \( e(x, y) \) defined by (4.2). Define, the process

\[
Z_n^*(x, y) = \sqrt{n} \left\{ [e_n^*(x, y) - e(x, y)], (x, y) \in [0, \tau]^2 \right\}.
\]
The estimator $e_n^c$ possesses properties similar to those demonstrated for the estimator $e_n^*$ in Section 2. As the proof for the following results are similar to the proofs for the results in Section 2, they will be omitted.

**Theorem 4.1.** Let $e_n^c$ be the estimator defined by (4.3)

(i) $|e_n^c(x, y) - e(x, y)| \leq |\tilde{e}_n(x, y) - e(x, y)|$ for all $x, y$.
(ii) As a consequence of (i), $e_n^c$ is weakly uniformly consistent on bounded rectangles.
(iii) $\tilde{e}_n$ is inadmissible with respect to the class of loss functions $L(e, \tilde{e}_n) = v(\|\tilde{e}_n - e\|)$, with $v(0) = 0$ and $v(x)$ increasing in $x$, $\tilde{e}_n$ being dominated in risk by $e_n^c$.
(iv) The estimator $e_n^c$ is the projection of $\tilde{e}_n$ onto the convex set of mean residual loss functions $\{k(x, y) \leq m(x, y)\}$.
(v) Let $\{Z(x, y), (x, y) \in [0, \tau]^2\}$ denote the bivariate Gaussian process obtained in [4] as the weak limit of the bivariate empirical mrl process under univariate censoring.
(a) If $e < m$ on $[0, \tau]^2$, then $Z_n^c \Rightarrow Z$ on $[0, \tau]^2$.
(b) If $e = m$ on $[0, \tau]^2$, then $Z_n^c \Rightarrow Z \wedge 0$ on $[0, \tau]^2$.
(c) If for $i = 1$ or 2, $e_i(x_0, y_0) = m_i(x_0, y_0)$ for some $(x_0, y_0) \in [0, \tau]^2$ and $e_i(x, y) < m_i(x, y)$ for all $(x, y) \in [0, \tau]^2$ in the line segment $z(x_0, y_0) + (1 - z)(x_1, y_1)$, $0 < z < 1$ for some $(x_1, y_1) \in \mathcal{X}$, then $Z_n^c$ does not converge weakly.

In the two-sample case with censored data, similar arguments can be used to develop estimators analogous to the estimators and results obtained in Section 3.

**5. Simulation studies**

Simulation studies were carried out to examine the properties of the proposed estimators as a function of various sample sizes 15, 30 and 45. Each simulation consisted of a series of 10,000 trials. Several bivariate distributions were used for the simulation study:

1. Gumbel: $f(x, y; \theta) = [(1 + \theta x)(1 + \theta y) - \theta] \exp(-x - y - \theta xy)$ where $x, y > 0$; $0 \leq \theta \leq 1$.
2. Pareto: $f(x_1, x_2, \theta_1, \theta_2, a) = a(a + 1)(\theta_1 \theta_2)^{(a+1)}(\theta_2 x_1 + \theta_1 x_2 - \theta_1 \theta_2)^{-(a+2)}$ where $x_1 \geq \theta_1 > 0$, $x_2 \geq \theta_2 > 0$, $a > 0$.
3. Morgenstern [5,9]: $f(u_1, u_2, \alpha) = 1 + \alpha(2u_1 - 1)(2u_2 - 1)$ where $0 \leq u_1, u_2 \leq 1$, $-1 \leq \alpha \leq 1$.
4. Sarmanov [14]: $f(x_1, x_2, \alpha) = e^{-x_1} e^{-x_2} \{1 + \alpha(2e^{-x_1} - 1)(2e^{-x_2} - 1)\}$ where $0 \leq x_1, x_2 < \infty$, $-1 \leq \alpha \leq 1$.

The parameters of the distributions were chosen to yield $e(x, y) \leq m(x, y)$. Simulation results are plotted in Appendix B for the ratio of the mean squared error of the restricted estimator to the empirical estimator. In this paper, we have only included the surface plots of the bivariate Gumbel ($\theta_1 = 1$, $\theta_2 = 0.5$)—Fig. 3—, and Morgenstern ($x_1 = 0.5$, $x_2 = 0.6$)—Fig. 4—distributions, and only for sample sizes 15 and 45. The reader may find additional simulation results, including estimated biases, at [http://www.stat.rice.edu/~jrojo](http://www.stat.rice.edu/~jrojo).
It is clear from the simulations that the new estimators dominate the empirical mrl function uniformly for all the cases examined in the simulation study. This conclusion is not surprising in the one-sample case in view of Theorem 2.1. Although the “mirror” image problem where \( e(x, y) \geq m(x, y) \) was not considered in detail in this paper, this problem being similar to the problem treated in Section 2, simulations were also run for this case and are also presented in Appendix B.

The gain in mean squared error in the case of the estimator under the restriction that \( e(x, y) \geq m(x, y) \) is substantially more than the gain obtained in connection with the problem discussed in Section 2. What is truly surprising is the uniformity of the gain in the two-sample problem. It is easy to understand that knowing \( m(x, y) \) provides a “benchmark” that our estimators take advantage of to calibrate themselves. But the simulation work shows that even in the absence of this benchmark, the estimators can still beat the empirical mrl function. A more detailed description of the plots is given in Appendix B.

6. Application to a real data set

This section illustrates the estimators using a data set from a Diabetic retinopathy study. The 197 patients in this data set were a 50% random sample of the patients with high-risk diabetic retinopathy as defined by the Diabetic Retinopathy Study (DRS). Diabetic retinopathy is a complication associated with diabetes mellitus consisting of abnormalities in the microvasculature within the retina of the eye. It is the leading cause of new cases of blindness in patients under 60 years of age in the United States and is the major cause of visual loss elsewhere in many industrialized countries. The study begun in 1971 to study the effectiveness of laser photocoagulation in delaying the onset of blindness in patients with diabetic retinopathy. Patients with diabetic retinopathy in both eyes and visual acuity of 20/100 or better in both eyes were eligible for the study. One eye of each patient was randomly selected for treatment and the other eye was observed without treatment. For each eye, the event of interest was the time from initiation of treatment to the time when visual acuity dropped below 5/200 two visits in a row (call it “blindness”). Thus there is a built-in lag time of approximately 6 months (visits were every 3 months). Survival times in this data set are therefore the actual time to blindness in months, minus the minimum possible time to event (6.5 months). Diabetes can be classified into two general groups by the age at the onset: juvenile (\(<20\) years) and adult diabetes.

In the DRS study censoring was caused by death, dropout, or the end of the study. In this paper, attention is focused on the uncensored cases. For each uncensored case \( i \), the survival times of the treated \( X_i \) and untreated \( Y_i \) eyes are given in Tables 1 and 2.

Let \( e_1(x, y) \) and \( e_2(x, y) \) be the mean residual lifetime functions corresponding to the adult and juvenile onset diabetes, respectively. It seems natural to assume that the mrl for the juvenile diabetes be longer than the mrl of the adult diabetes. We calculated both the empirical estimators \( \hat{e}_1(x, y) \) and \( \hat{e}_2(x, y) \) as well the estimators under mrl ordering \( e_1^*(x, y) \), \( e_2^*(x, y) \) of \( e_1(x, y) \) and \( e_2(x, y) \). From the empirical estimators, it is observed that there are some points \((x, y)\) for which the ordering is reversed, that is, sometimes \( \hat{e}_1(x, y) \) falls above
<table>
<thead>
<tr>
<th>Patient, $i$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_i$</td>
<td>38.57</td>
<td>1.33</td>
<td>21.9</td>
<td>13.87</td>
<td>48.3</td>
<td>9.9</td>
<td>8.3</td>
</tr>
<tr>
<td>$Y_i$</td>
<td>30.83</td>
<td>5.77</td>
<td>25.63</td>
<td>25.8</td>
<td>5.73</td>
<td>9.9</td>
<td>8.3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Patient, $i$</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_i$</td>
<td>12.2</td>
<td>33.63</td>
<td>27.6</td>
<td>1.63</td>
<td>1.57</td>
<td>4.97</td>
<td>9.87</td>
</tr>
<tr>
<td>$Y_i$</td>
<td>4.1</td>
<td>33.63</td>
<td>63.33</td>
<td>38.47</td>
<td>13.83</td>
<td>12.93</td>
<td>24.43</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Patient, $i$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_i$</td>
<td>6.9</td>
<td>1.63</td>
<td>13.83</td>
<td>35.53</td>
<td>14.8</td>
<td>6.2</td>
<td>22</td>
<td>1.7</td>
</tr>
<tr>
<td>$Y_i$</td>
<td>20.17</td>
<td>10.27</td>
<td>5.67</td>
<td>5.90</td>
<td>33.9</td>
<td>1.73</td>
<td>30.2</td>
<td>1.7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Patient, $i$</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_i$</td>
<td>43.03</td>
<td>6.53</td>
<td>42.17</td>
<td>48.43</td>
<td>9.6</td>
<td>7.6</td>
<td>1.8</td>
<td>9.9</td>
</tr>
<tr>
<td>$Y_i$</td>
<td>1.77</td>
<td>18.7</td>
<td>42.17</td>
<td>14.3</td>
<td>13.33</td>
<td>14.27</td>
<td>34.57</td>
<td>21.57</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Patient, $i$</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
<th>21</th>
<th>22</th>
<th>23</th>
<th>24</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_i$</td>
<td>13.77</td>
<td>0.83</td>
<td>1.97</td>
<td>11.3</td>
<td>30.4</td>
<td>19</td>
<td>5.43</td>
<td>46.63</td>
</tr>
<tr>
<td>$Y_i$</td>
<td>13.77</td>
<td>10.33</td>
<td>11.07</td>
<td>2.1</td>
<td>13.97</td>
<td>13.80</td>
<td>13.57</td>
<td>42.43</td>
</tr>
</tbody>
</table>

$\hat{e}_2(x, y)$. Figs.1 and 2 show the surface plots of the empirical and the restricted estimators, respectively, where it is evident that the new estimators modify the empirical mrl function only in the region where the order is violated.

7. Concluding remarks

In this paper we have proposed estimators for two mrl functions, $e_1$ and $e_2$, under the order restriction that $e_1 \leq e_2$ when $e_2$ is known or unknown. We have proved that they are strongly uniformly consistent and asymptotically unbiased. We have also showed their weak convergence on $\mathbb{R}^2_+$. Simulation studies were carried out for various sample sizes and various bivariate distributions, and some of the results are presented here. The simulation results indicate that the proposed estimators are superior to the empirical estimators in terms of mean squared error.
Fig. 1. The empirical mlr function for the DRS data: juvenile (top) and adult (bottom).

Fig. 2. The restricted estimators for the DRS data: juvenile (top) and adult (bottom).
Acknowledgments

Thanks to an anonymous referee for a careful reading of an earlier version of this manuscript. The comments were very helpful in improving the reading of the paper. Thanks also to an associate editor and the Editor for guidance in improving the paper. The work of Javier Rojo was supported in part by NSF Grant DMS-0244184, and NSA Grant MDA904-03-1-0063. The work of Musie Ghebremichael was supported in part by NSF Cooperative Agreement No. HRD-9817555.

Appendix A. Proofs

Proof of Theorem 2.1. Let

\[ Q_1 = \{ (x_1, x_2) : \hat{\epsilon}_{1,n}(x_1, x_2) > m_1(x_1, x_2), \hat{\epsilon}_{2,n}(x_1, x_2) \leq m_2(x_1, x_2) \} \]
\[ Q_2 = \{ (x_1, x_2) : \hat{\epsilon}_{2,n}(x_1, x_2) > m_2(x_1, x_2), \hat{\epsilon}_{1,n}(x_1, x_2) \leq m_1(x_1, x_2) \} \]
\[ Q_{12} = \{ (x_1, x_2) : \hat{\epsilon}_{1,n}(x_1, x_2) > m_1(x_1, x_2), \hat{\epsilon}_{2,n}(x_1, x_2) > m_2(x_1, x_2) \} \]
\[ Q = \{ (x_1, x_2) : \hat{\epsilon}_{1,n}(x_1, x_2) \leq m_1(x_1, x_2), \hat{\epsilon}_{2,n}(x_1, x_2) \leq m_2(x_1, x_2) \} . \]

Then

\[ |e_{1,n}^*(x_1, x_2) - e_1(x_1, x_2)| = |m_1(x_1, x_2) - e_1(x_1, x_2)| I_{Q_1 \cup Q_{12}} + |\hat{\epsilon}_{1,n}(x_1, x_2) - e_1(x_1, x_2)| I_{Q_2 \cup Q} \leq |\hat{\epsilon}_{1,n}(x_1, x_2) - e_1(x_1, x_2)| I_{Q_1 \cup Q_{12}} + |\hat{\epsilon}_{1,n}(x_1, x_2) - e_1(x_1, x_2)| I_{Q_2 \cup Q} = |\hat{\epsilon}_{1,n}(x_1, x_2) - e_1(x_1, x_2)| . \] (A.1)

A similar inequality holds for \( e_{2,n}^* \). The result then follows immediately. □

Proof of Theorem 2.2. Let \( A = \{ k(x_1, x_2) : k \) is a mrl and \( k(x_1, x_2) \leq m(x_1, x_2) \) for all \( (x_1, x_2) \in \mathcal{X} \} \). We prove (i) first. Note that

\[
e_{i,n}^*(x_1, x_2) - \hat{\epsilon}_n(x_1, x_2) = \begin{cases} 0 & \text{if } \hat{\epsilon}_n(x_1, x_2) \leq m(x_1, x_2), \\ \hat{\epsilon}_n(x_1, x_2) \land m(x_1, x_2) - \hat{\epsilon}_n(x_1, x_2) & \text{otherwise.} \end{cases}
\]

Now if \( \hat{\epsilon}_n(x_1, x_2) \not\leq m(x_1, x_2) \), then \( \hat{\epsilon}_n(x_1, x_2) \land m(x_1, x_2) < \hat{\epsilon}_n(x_1, x_2) \) with \( \hat{\epsilon}_n(x_1, x_2) \land m(x_1, x_2) \in A \). Moreover, letting \( e_{i,n}^*(x_1, x_2) = (e_{1,n}^*(x_1, x_2), e_{2,n}^*(x_1, x_2)) = \hat{\epsilon}(x_1, x_2) \land m(x_1, x_2) \), it follows easily that for all \( k(x_1, x_2) = (k_1(x_1, x_2), k_2(x_1, x_2)) \in A \),

\[ |e_{i,n}^*(x_1, x_2) - \hat{\epsilon}_{i,n}(x_1, x_2)| \leq |k_i(x_1, x_2) - \hat{\epsilon}_{i,n}(x_1, x_2)| \text{ for } i = 1, 2. \]

Therefore, for each \( k \in A \),

\[ \sup_{0 \leq (x_1, x_2) \in \mathcal{X}} \|e_{i,n}^*(x_1, x_2) - \hat{\epsilon}_n(x_1, x_2)\| \leq \sup_{0 \leq (x_1, x_2) \in \mathcal{X}} \|k(x_1, x_2) - \hat{\epsilon}_n(x_1, x_2)\|. \]

The proof of (ii) follows immediately from the above arguments. □
Proof of Theorem 2.3. It follows from the definition that \( e_n^*(x_1, x_2) \leq m(x_1, x_2) \) with \( m(x_1, x_2) < \infty \) for all \((x_1, x_2) \in \mathcal{X}\). It follows from Theorem 1 of [7], that \( \hat{e}_n(x_1, x_2) \to e(x_1, x_2) \) with probability one, and as a consequence, \( e_n^*(x_1, x_2) \to e(x_1, x_2) \) almost surely for each \((x_1, x_2) \in \mathcal{X}\). Hence, by the dominated convergence theorem

\[
E \left[ e_n^*(x_1, x_2) \right] \to e(x_1, x_2) \quad \text{as} \quad n \to \infty
\]

for all \((x_1, x_2) \in \mathcal{X}\). To prove the second part, as in [7], let \( S_k \) be all subsets of size \( k \) of \( \{1, \ldots, n\} \), and for each \( f_k \in S_k \), define \( A_{f_k} = \{(X_j > x, Y_j > y), j \in f_k; (X_j \leq x \text{ or } Y_j \leq y), j \notin f_k\}. \) Then,

\[
E[\min(\hat{e}_1n, m_1)] = \int_0^{m_1} P(\hat{e}_1 > t)dt = \sum_{k=1}^{n} \sum_{f_k \in S_k} \int_0^{m_1} P(\hat{e}_1 > t|A_{f_k})P(A_{f_k})dt.
\]

Therefore,

\[
E[\hat{e}_1n] - E[\min(\hat{e}_1n, m_1)]
\]

\[
= \sum_{k=1}^{n} \sum_{f_k \in S_k} P(A_{f_k}) \int_{m_1}^{\infty} P(\hat{e}_1 > t|A_{f_k})dt
\]

\[
= \sum_{k=1}^{n} \sum_{f_k \in S_k} P(A_{f_k}) \int_{m_1}^{\infty} P(\sum_{j \in f_k} (X_j - e_1(x, y)) > k(t + x - e_1(x, y))|A_{f_k}) dt
\]

\[
\leq \sum_{k=1}^{n} \sum_{f_k \in S_k} P(A_{f_k}) \frac{\sigma_1^2(x, y)}{k(m_1(x, y) + x - e_1(x, y))}
\]

\[
= \frac{\sigma_1^2(x, y)}{(m_1(x, y) + x - e_1(x, y))S(x, y)} \sum_{k=1}^{n} \binom{n}{k} \frac{S(x, y)^k(1 - S(x, y))^{n-k}}{k}
\]

\[
\leq \frac{2\sigma_1^2(x, y)}{(n + 1)(m_1(x, y) + x - e_1(x, y))S(x, y)},
\]

where the first inequality follows from Chebyshev’s inequality followed by integration. A similar proof works for the case of \( E[\hat{e}_2n] - E[\hat{e}_n^*] \). \(\square\)

Proof of Theorem 2.4. Since the consistency of \( e_n^* \) is equivalent to the consistency of both marginals, it is sufficient to prove the theorem for one of the marginals. The proof follows immediately from (A.1) and Theorem 2 of [7]. \(\square\)

Proof of Theorem 2.5. (i) Define \( Z_n(x_1, x_2) = \sqrt{n}(\hat{e}_n(x_1, x_2) - e(x_1, x_2)) \) and consider

\[
Z_n^*(x_1, x_2) = \sqrt{n}(e_n^*(x_1, x_2) - e(x_1, x_2)) = \sqrt{n}((\hat{e}_n(x_1, x_2) - e(x_1, x_2)) \wedge \sqrt{n}(m(x_1, x_2) - e(x_1, x_2)). \quad (A.2)
\]

Suppose first that \( \inf_{(x_1, x_2) \in \mathcal{X}} (m(x_1, x_2) - e(x_1, x_2)) > 0 \). Then, \( \sqrt{n}(m(x_1, x_2) - e(x_1, x_2)) \) converges uniformly to \( \infty \) for \( i = 1, 2 \) and, since \( \sqrt{n}(\hat{e}_i, n(x_1, x_2) - e_i(x_1, x_2)) = O_p, i = 1, 2, P(\sup_{(x_1, x_2) \in \mathcal{X}} \| Z_n^*(x_1, x_2) - Z_n(x_1, x_2) \| > \varepsilon) \to 0 \) as \( n \to \infty \), and hence \( \{ Z_n^*(x_1, x_2), (x_1, x_2) \in \mathcal{X} \} \Rightarrow \{ Z(x_1, x_2), (x_1, x_2) \in \mathcal{X} \} \).
Now consider the case when $e < m$ but $\inf_{(x_1,x_2) \in \mathcal{X}} (m_1(x_1, x_2) - e(x_1, x_2)) > 0$ for $i = 1$ or 2 or both. The proof hinges on the same idea as in the case that $\inf_{(x_1,x_2) \in \mathcal{X}} (m_1(x_1, x_2) - e(x_1, x_2)) > 0$ except that we apply it to an increasing sequence of closed and bounded $T_i$, with $T_i \uparrow \mathcal{X}$. Consider a sequence $x_n \downarrow 0$, and define $T_i = \{ (x_1, x_2) \in \mathcal{X} : F(x_1, x_2) \geq x_i \}$ for each $i = 1, 2, \ldots$. Since the $T_i$’s are closed and bounded $\inf_{(x_1,x_2) \in T_i} (m_1(x_1, x_2) - e(x_1, x_2)) > 0$.

By the previous arguments $Z^*_n \mid T_i \Rightarrow Z \mid T_i$ for each $i = 1, 2, \ldots$ where $Z^*_n \mid T_i$ denotes the restriction of $Z^*_n$ to $T_i$. The fact that $Z^*_n \Rightarrow Z$ then follows immediately from the results in Chapter 1.6 in [15].

(ii) If $e(x_1, x_2) = m(x_1, x_2)$ for all $(x_1, x_2) \in \mathcal{X}$, then $\sqrt{n}(e^*_n - e) = \sqrt{n}(\min(\hat{e}_n - e, 0))$. Since the map $\min(\hat{e}_n - e, 0)$ is continuous, by the continuous mapping theorem $Z^*_n \Rightarrow Z \wedge 0$.

(iii) If $Z^*_n \Rightarrow Z$, then by the continuous mapping theorem, the projection mappings $e^*_{i,n}(x, y) = \sqrt{n} \{ \hat{e}_{i,n}(x, y) \wedge m_1(x, y) - e_i(x, y) \}$ for $i = 1, 2$ must converge weakly to the projection mappings of $Z$. It is now shown that under the conditions of the theorem under (iii), this is not possible. Without loss of generality, suppose that $e_1(x_0, y_0) = m_1(x_0, y_0)$ for some $(x_0, y_0) \in \mathcal{X}$, and $e_1(x, y) < m_1(x, y)$ for all $(x, y) \in (x_0, x_0 + \delta_1] \times (y_0, y_0 + \delta_2]$ where $\delta_1 \geq 0, \delta_2 \geq 0$ and $\delta_1 \vee \delta_2 \geq 0$. We show that $e^*_n$ is not tight, see e.g. Billingsley [2], on $[x_0, x_0 + \delta_1] \times [y_0, y_0 + \delta_2]$ and hence cannot converge weakly, which in turn implies that $e^*_n$ does not converge weakly. The proof follows Rojo [11].

Suppose then that $\delta_1 > 0$, and for small $\gamma > 0$ consider

$$\sup_{(s_0, y_0) \leq (s_1, y_0) \leq (s_0 + \gamma, y_0)} |e^*_{1,n}(s_1, y_0) - e^*_{1,n}(s_0, y_0)|$$

$$= \|e^*_{1,n}(s_1, y_0) - e^*_{1,n}(s_0, y_0)\|_{(s_0 + \gamma, y_0)}$$

$$= \sqrt{n} \|\hat{e}_{1,n}(s_1, y_0) \wedge m_1(s_1, y_0) - e_1(s_1, y_0) - (\hat{e}_{1,n}(s_0, y_0) - e_1(s_0, y_0))\|_{(s_0 + \gamma, y_0)}.$$ 

Now, for $s_0 = x_0 + \min(\delta_1, \gamma)$, eventually with probability one, $m_1(s_0, y_0) > \hat{e}_{1,n}(s_0, y_0)$.

Therefore, eventually with probability one,

$$\sqrt{n} \|e^*_{1,n}(s_1, y_0) - e^*_{1,n}(s_0, y_0)\|_{(s_0 + \gamma, y_0)} \geq \sqrt{n} \|\hat{e}_{1,n}(s_0, y_0) - e_1(s_0, y_0) - (\hat{e}_{1,n}(s_0, y_0) - e_1(s_0, y_0))\|_{(s_0 + \gamma, y_0)}.$$

Therefore

$$\lim_{n \to \infty} P \left\{ \sqrt{n} \|e^*_{1,n}(s_1, y_0) - e^*_{1,n}(s_0, y_0)\|_{(s_0 + \gamma, y_0)} \geq \varepsilon \right\}$$

$$\geq \lim_{n \to \infty} P \left\{ \sqrt{n} \max(0, e_1(s_0, y_0) - \hat{e}_{1,n}(s_0, y_0)) + |\hat{e}_{1,n}(s_0, y_0) - e_1(s_0, y_0)| \geq \varepsilon \right\}$$

$$\geq \lim_{n \to \infty} P \left\{ (\max(0, \sqrt{n}(e_1(s_0, y_0) - \hat{e}_{1,n}(s_0, y_0)) + \sqrt{n}(\hat{e}_{1,n}(s_0, y_0) - e_1(s_0, y_0))) \geq \varepsilon \right\}$$

$$\geq \lim_{n \to \infty} P \left\{ \sqrt{n}(\hat{e}_{1,n}(s_0, y_0) - e_1(s_0, y_0)) \geq \varepsilon \right\} = 1 - \Phi\left(\frac{\varepsilon}{\sigma^2}\right),$$
where \( \sigma^2_n = \text{Var}(Z(s_0, y_0)) \), and \( \Phi \) denotes the standard normal distribution. It follows that \( \{ e_n^*(x_1, x_2), (x_1, x_2) \in X \} \) is not tight. □

**Proof of Theorem 3.1.** Only the proof for \( e^*(x, y) \) is shown as the proof for \( m^*(x, y) \) follows exactly as that for \( e^*(x, y) \). The strong uniform convergence of \( e^*(x, y) \) on bounded rectangles follows immediately from the strong uniform convergence of the empirical mrl function, \( \hat{e}(x, y) \), on bounded rectangles and the following inequality applied to each component of \( e^*(x, y) \). Let \( i = 1, 2 \). Then

\[
|e_i^*(x, y) - e_i(x, y)| = |\min(\hat{e}_i(x, y), \hat{m}_i(x, y)) - \min(e_i(x, y), m_i(x, y))| \leq \min(|\hat{e}_i(x, y) - e_i(x, y)|, |\hat{m}_i(x, y) - m_i(x, y)|). \tag{A.3}
\]

The result follows immediately from the above inequalities. □

**Proof of Theorem 3.2.** Only (i) will be proven. The proof of (ii) follows immediately from the arguments used in (i). Let \( \hat{e}^* = (\hat{e}_1^*, \hat{e}_2^*) \) and \( \hat{e}_p = (\hat{e}_{1p}, \hat{e}_{2p}) \). The result follows immediately if it can be shown that \( |\hat{e}_i^*(x_1, x_2) - \hat{e}_i(x_1, x_2)| \leq |\hat{e}_{ip}(x_1, x_2) - \hat{e}_i(x_1, x_2)| \) for \( i = 1, 2 \). Now, from Eq. (3.4) we know that, for \( i = 1, 2 \),

\[
\hat{e}_i^*(x_1, x_2) - \hat{e}_i(x_1, x_2) = \begin{cases} \hat{e}_{ip}(x_1, x_2) - \hat{e}_i(x_1, x_2) & \text{if } \hat{e}_i(x_1, x_2) \geq \hat{m}_i(x_1, x_2), \\ 0 & \text{otherwise}. \end{cases}
\]

But when \( \hat{e}_i(x_1, x_2) \geq \hat{m}_i(x_1, x_2), \hat{m}_i(x_1, x_2) \leq \hat{e}_{ip}(x_1, x_2) \leq \hat{e}_i^*(x_1, x_2) \). Since \( k(x_1, x_2) \leq \hat{e}_p(x_1, x_2) \), this implies that for \( i = 1, 2 \),

\[
|e_i^*(x_1, x_2) - \hat{e}_i(x_1, x_2)| = |\hat{e}_{ip}(x_1, x_2) - \hat{e}_i(x_1, x_2)| \leq |k_i(x_1, x_2) - \hat{e}_i(x_1, x_2)|.
\]

Hence,

\[
\sup_{0 \leq (x_1, x_2) \in X} \|\hat{e}^*(x_1, x_2) - \hat{e}(x_1, x_2)\| \leq \sup_{0 \leq (x_1, x_2) \in X} \|k(x_1, x_2) - \hat{e}(x_1, x_2)\|
\]

for all \( k(x_1, x_2) \in A \).

**Proof of Theorem 3.3.** The proof is the same as the previous proof. □

**Proof of Theorem 3.4.** The proof proceeds by showing uniform integrability, and almost sure convergence to zero, of an appropriate sequence of statistics. Consider,

\[
E[\hat{e}_1^*(x, y)] = E \left[ \min \left( \hat{e}_1(x, y), \frac{n_1}{n_1 + n_2} \hat{e}_1(x, y) + \frac{n_2}{n_1 + n_2} \hat{m}_1(x, y) \right) \right] = E[\hat{e}_1(x, y)] + \frac{n_2}{n_1 + n_2} E \left[ \min \left( 0, \hat{m}_1(x, y) - \hat{e}_1(x, y) \right) \right]. \tag{A.4}
\]
If \( n_1 \to \infty \) while \( n_2 \) remains bounded, \( E[\hat{e}_1(x, y)] \to e_1(x, y) \) while the last term in (A.4) converges to zero. On the other hand, if both \( n_1 \) and \( n_2 \to \infty \), since \( \hat{m}_1(x, y) - \hat{e}_1(x, y) \) converges to \( m_1(x, y) - e_1(x, y) \geq 0 \) with probability one so that \( \min(0, \hat{m}_1(x, y) - \hat{e}_1(x, y)) \) converges to zero with probability one, it is enough to show that the sequence \( \min(0, \hat{m}_1 - \hat{e}_1) \) is uniformly integrable. This follows easily after writing
\[
E \left[ \left| \min(0, \hat{m}_1(x, y) - \hat{e}_1(x, y)) \right| I_{\{ \min(0, \hat{m}_1 - \hat{e}_1) > 0 \}} \right] 
= E \left[ \max(0, \hat{e}_1 - \hat{m}_1) I_{\{ \max(0, \hat{e}_1 - \hat{m}_1) > 0 \}} \right]
\]
and noticing that \( \hat{e}_1 - \hat{m}_1 \) converges almost surely to \( e_1 - m_1 \leq 0 \). Therefore, eventually, with probability one, \( I_{\{ \max(0, \hat{e}_1 - \hat{m}_1) > 0 \}} = 0 \) and the result follows. Similar arguments yield the result for the asymptotic unbiasedness of \( \hat{m}^* \).

**Proof of Theorem 3.5.** Since the strong uniform consistency of the bivariate mrl vector is equivalent to the strong uniform consistency of both marginals, it is sufficient to prove the theorem for the marginals. Here, we will show the proof for the first component of the mrl vector, the proof for the second component is exactly the same.

Let \( D = [0, b_1] \times [0, b_2] \) where \( b_1, b_2 > 0 \), and consider
\[
\sup_{(x,y) \in D} \left| \hat{e}_1^*(x, y) - e_1(x, y) \right| 
= \sup_{(x,y) \in D} \left| \min(\hat{e}_1(x, y), \hat{e}_{1}\nu(x, y)) - e_1(x, y) \right|
\]
\[
= \sup_{(x,y) \in D} \left| \min(\hat{e}_1(x, y), \hat{e}_{1}(x, y)) - e_1(x, y), \frac{n_1}{n_1 + n_2} (\hat{e}_1(x, y) - e_1(x, y)) + \frac{n_2}{n_1 + n_2} (\hat{m}_1(x, y) - e_1(x, y)) \right|
\]
Now let \( n_1 \to \infty \) and \( \frac{n_2}{n_1 + n_2} \to \alpha, 0 \leq \alpha \leq 1 \). Since, \( \hat{e} \) and \( \hat{m} \) are strongly and uniformly consistent on \( D \) for \( e \) and \( m \), respectively, it follows that \( \left| \min(\hat{e}_1(x, y) - e_1(x, y), \frac{n_1}{n_1 + n_2} (\hat{e}_1(x, y) - e_1(x, y)) + \frac{n_2}{n_1 + n_2} (\hat{m}_1(x, y) - e_1(x, y)) \right| \) converges, uniformly on \( D \), with probability one to \( \min(0, \alpha(m_1(x, y) - e_1(x, y))) = 0 \), and the result follows. A similar proof yields the result for \( m^* \). □

**Proof of Theorem 3.6.** (i) Note that
\[
Z^*_e = \sqrt{n_1}(\hat{e} - e) + \sqrt{n_1} \min \left( 0, \frac{n_2}{n_1 + n_2} \{ (\hat{m} - m) + (m - e) + (\hat{e} - e) \} \right).
\]
Consider first the case with \( \alpha = 0 \). Since \( \sup_{(x,y) \in \mathcal{X}} \sqrt{n_2} \{ \hat{m} - m \} = O_p, \sqrt{n_1/n_2} (n_2/(n_1 + n_2)) \to 0, \sup_{(x,y) \in \mathcal{X}} \sqrt{n_1} \{ \hat{e} - e \} = O_p, \) and \( (m - e) \geq 0 \), it follows that the last term in (A.5) converges uniformly with probability one to 0. Therefore, \( Z^*_e \Rightarrow Z_e \). If \( \alpha = \infty \), a similar analysis yields the result that \( Z^*_e \Rightarrow Z_e + \min(0, -Z_e) = Z_e^* \). The results for the process \( Z^*_m \) are obtained in a similar fashion.
(ii) Let $K_{n_1,n_2}(x, y) = \sqrt{n_1/(n_1+n_2)} \min \left(0, (\hat{m} - m) + (m - e) + (e - \hat{e}) \right)$. Note that because of (i), it is enough to consider the case where $0 < \xi < \infty$. The proof exploits the expression for $K_{n_1,n_2}(x, y)$. Suppose first that $\inf_{(x,y)\in \mathcal{X}} (m(x, y) - e(x, y)) > 0$ so that $\sqrt{n_1n_2/(n_1 + n_2)}(m - e)$ converges to $(\infty, \infty)$ uniformly on $\mathcal{X}$. The result follows immediately after observing that, for $0 < \xi < \infty$, $\sup_{(x,y)\in \mathcal{X}} \sqrt{n_1/n_2(n_2/(n_1 + n_2))} \|\sqrt{n_1}(\hat{e} - e)\| = O_p$ and therefore, with probability one, $\sup_{(x,y)\in \mathcal{X}} \left\|K_{n_1,n_2}(x, y)\right\| \to 0$. It follows that $Z_\epsilon^* \Rightarrow Z_\epsilon$.

If, on the other hand, $\inf_{(x,y)\in \mathcal{X}} (m(x, y) - e(x, y)) = 0$, we proceed as in the proof of Theorem 2.5. Choose $\beta_k \downarrow 0$ and define $T_k = \{(x, y) \in \mathcal{X} : (x, y) \in \mathcal{X} : S(x, y) \leq \beta_k \}$ and $S(x, y) = \sup_{(x,y)\in \mathcal{X}} \left\|K_{n_1,n_2}(x, y)\right\|$. Arguments similar to those used above yield the result that $Z_\epsilon^*|_{\{T_k\}} \Rightarrow Z|_{\{T_k\}}$. The weak convergence of $Z_\epsilon^*$ to $Z_\epsilon$ follows. Similar arguments show that $Z^*_m \Rightarrow Z^*_m$.

(iii) Only the proof for $Z^*_m$ is given here, the proof for $Z^*_m$ following from similar arguments. Consider the independent bivariate processes $\{\hat{W}_1(x, y), (x, y) \in \mathcal{X}\} = \{\sqrt{n_1}(\hat{e}(x, y) - e(x, y), (x, y) \in \mathcal{X}\}$ and $\{\hat{W}_2(x, y), (x, y) \in \mathcal{X}\} = \{\sqrt{n_2}(\hat{m}(x, y) - m(x, y), (x, y) \in \mathcal{X}\}$; each process converging weakly to its appropriate Gaussian process limit $W_i$. It follows that $\hat{W}_1(x, y), \hat{W}_2(x, y) \in \mathcal{X}$ \Rightarrow $\{W_1(x, y), W_2(x, y) ; (x, y) \in \mathcal{X}\}$. Since the map $\min(f, g) \rightarrow \min(f, af + bg)$ is continuous, it follows from the continuous mapping theorem that $Z^*_e \Rightarrow \min \left(Z_e, \sqrt{\frac{\epsilon}{1+\epsilon}} Z_m + \frac{1}{1+\epsilon} Z_e \right)$. Similar arguments yield the result for $Z^*_m$.

(iv) Without loss of generality we suppose that $e(x_0, y_0) = m(x_0, y_0)$ for some $(x_0, y_0) \in \mathcal{X}$ and $e(x_0, y) < m(x_0, y)$ for all $y \in (y_0, y_1)$. Since weak convergence of $Z^*_e$ implies weak convergence of the projection mappings, it is enough to show that the process $\{\sqrt{n_1}(\hat{e}_1(x, y) - e_1(x, y), (x, y) \in \mathcal{X}\}$ cannot converge weakly. But this follows immediately from arguments similar to those used in the proof of Theorem 2.5. \(\square\)

Appendix B. Figures

The surface plots for the ratio of the mean squared errors, MSE(estimators under the restriction)/MSE(Empirical mrl function) for two distributions and two-sample sizes are displayed in Figs. 3 and 4. The results presented here are representative of the many simulations performed. As sample size, closeness of mrl functions to each other, and tail-heaviness of the underlying distributions, all impact the mean squared error performance of the estimators of interest, we proceeded to consider a wide variety of distributions that offered a diversity of situations that will allow for a thorough assessment of the mean squared error performance of the proposed estimators when compared to the empirical mrl function.

Each plot has four levels which correspond to the estimators developed in Sections 2 and 3, and the estimator for $e(x, y)$ in the case that $e(x, y) \geq m(x, y)$ with $m$ known. In this case, the estimator of interest is given by $\max(e_n(x, y), m(x, y)))$. The first, second, and third levels of the plots are the surface plots of the estimators defined in (2.1), (3.5), and (3.6), respectively. The fourth level corresponds to the surface plot of the ratios of mean.
squared error of the estimator $\max(e_n, m)$ to the mean squared error of the empirical. As observed in these plots, the new estimators uniformly dominate the empirical mrl function $e_n$ in terms of mean squared error.
Fig. 4. Ratios of mean squared errors for two Morgenstern distributions with $\alpha_1 = 0.5$ and $\alpha_2 = 0.6$. 
References