Tatiana Komarova

Extremum sieve estimation in k-out-of-n system

Article (Accepted version) (Refereed)

Original citation:
Komarova, Tatiana (2017) *Extremum sieve estimation in k-out-of-n system*. Communications in Statistics - Theory and Method, 46 (10). pp. 4915-4931. ISSN 0361-0926
DOI: 10.1080/03610926.2015.1091081

Extremum Sieve Estimation in k-out-of-n Systems [1]

Tatiana Komarova[2]

This version: August 2015

ABSTRACT

The paper considers nonparametric estimation of absolutely continuous distribution functions of independent lifetimes of non-identical components in k-out-of-n systems, $2 \leq k \leq n$, from the observed "autopsy" data. In economics, ascending "button" or "clock" auctions with n heterogeneous bidders with independent private values present 2-out-of-n systems. Classical competing risks models are examples of n-out-of-n systems. Under weak conditions on the underlying distributions the estimation problem is shown to be well posed and the suggested extremum sieve estimator is proven to be consistent. The paper considers sieve spaces of Bernstein polynomials which allow to easily implement constraints on the monotonicity of estimated distribution functions.

Keywords: k-out-of-n systems, competing risks, sieve estimation, Bernstein polynomials

[1] I am thankful to the anonymous referee for helpful comments. I am grateful to Xioahong Chen, Oliver Linton, Denis Nekipelov and Elie Tamer for their comments and suggestions. I appreciate feedback from seminar participants at Georgetown University, London School of Economics and Political Science, the University of California at Santa Barbara and the University of Rochester.

[2] Department of Economics, London School of Economics and Political Science; e-mail: t.komarova@lse.ac.uk.

1 Introduction

The paper considers nonparametric estimation of absolutely continuous distribution functions of independent lifetimes in k-out-of-n systems for $2 \leq k \leq n$. Such a system is "alive" if and only if at least k of its components are alive. k-out-of-n systems are often encountered in practice. In economics, ascending "button" or "clock" auctions with n bidders present 2-out-of-n systems. Classical competing risks models are examples of n-out-of-n systems. This paper considers general situations of heterogeneous components – that is, when the lifetimes of different components can have different distributions. However, it is assumed that the distributions of lifetimes have the same (and known) lower support point. The only available data are the "autopsy" data, which give information only on the lifetime of the system and the corresponding fatal set of $n-k+1$ components.

One way to approach the estimation problem would be to impose parametric assumptions on the underlying distributions of components' lifetimes. For instance, the assumption that these distributions are exponential would bring down the estimation task to the task of learning n scalar parameters for n exponential distributions. However, if the underlying distributions are not exponential, then the inference based on the obtained estimates would not be reliable. Such a parametric approach is exploited, for instance, in Meilijson (1994), among many others.

This paper does not make any parametric assumptions and suggests nonparametric estimators of the CDFs of components' lifetimes that are consistent in the uniform metric. The first step in this approach is to write down the system of integral-differential equations that describes the relations between the underlying unknown CDFs of components' lifetimes and the observables. This system is given in (IDE) in section 2. In the second step the distributions of observables are estimated consistently from a given sample. The third step constructs an objective function that represents a distance between the left-hand side and the right-hand in (IDE). Finally, this objective function is minimized when unknown CDFs are represented as unknown functions from a chosen sieve space. It is proven that the operator that maps observable functions into underlying CDFs of components's lifetimes is continuous. This guarantees the well-posedness of the estimation problem and the consistency of described extremum sieve estimators. It is worth noting that this approach works for any $2 \leq k \leq n$ and is easy to implement in practice.

The paper considers spaces of Bernstein polynomials as sieve spaces. In these spaces it is easy to formulate and use the constraints that represent necessary and sufficient conditions for the monotonicity of a function. Monotonicity is of course a desirable property for an estimator of CDF. For a detailed review of sieve estimation methods in econometrics see Chen (2007). Chen (2008) focuses specifically on extremum sieve estimation.

The rest of this paper is organized as follows. Section 2 discusses nonparametric identification of the distributions of components' lifetimes in k-out-of-n systems, $2 \leq k \leq n$, from "autopsy" data. It reviews the identification result from Meilijson (1981) and also shows how to extend the identification result for 2-out-of-n systems established in Komarova (2013) to k-out-of-n systems for any $2 \leq k \leq n$. Section 3 establishes that when the space of underlying distributions of components' lifetimes and the space of distributions of observables are endowed with the uniform metric, the problem of estimating underlying distributions from observables is well-posed. The section also suggests an extremum sieve estimator and proves its consistency. Section 4 illustrates the suggested sieve estimation method in an ascending auction framework by performing estimation in two Monte Carlo experiments. Proofs of propositions, lemmas and theorems are collected in the Appendix.

It is worth mentioning that even though this paper focuses on k-out-of-n systems, this is done mostly for the sake of technical and notational convenience and the presented sieve estimation approach can be extended to a much more general class of coherent systems. The main requirements for such an extension would be the rank condition on the incidence matrix of the coherent system as in Meilijson (1981) together with some convergence and local integrability conditions on some functions of primitives (or, equivalently, of observables) as discussed in section 4 in Komarova (2013).

Related literature

Nonparametric estimation methods of heterogeneous independent lifetimes from autopsy data are considered in Watelet (1990) and Doss, Huffer and Lawson (1997). Watelet's approach is based on rewriting the mathematical model that describes the relation between the unknown underlying CDFs of components' lifetimes and the observable joint distribution of system's lifetime and the fatal set in a form that contains unknown distributions on the left-hand side and some integral expressions on the right-hand side. Then Watelet uses an iterative method to estimate unknown distributions. Importantly, Watelet explores such a procedure only in the simplest case of $k = n$, which is the case of classical competing risks. It is also worth noting that for $2 < k < n$ this procedure cannot work in general. Doss, Huffer and Lawson (1997) suggest a nonparametric Bayesian procedure, which uses mixtures of Dirichlets as priors on the distributions of components' lifetimes.

There is a literature within econometrics and statistics that looks at the classical competing risks models (Roy models) and related models and analyzes the possibility to obtain identification using covariates while allowing for errors to be correlated across risks. E.g., Heckman and Honore (1989) establish under which conditions access to regressors guarantees identifiability for proportional and accelerated failure time models, thus effectively reversing the non-identifiability results in Cox (1962) and Tsiatis (1975). Abbring and Van den Berg (2003) consider identification in the mixed proportional hazards competing risks model under conditions weaker those in Heckman and Honore (1989). Lee and Lewbel (2013) show identification of accelerated failure time competing risks models identified when covariates are independent of latent errors and a certain rank condition is satisfied. This paper does not take that approach by considering models without covariates and with independent risks.

2 Review of identifiability

Consider a system that consists of n components whose lifetimes are mutually independent random variables X_i with distribution functions F_i^*, $i = 1, \ldots, n$. The distribution of i's component is absolutely continuous with respect to the Lebesgue measure, that is,

$$F_i^* \textit{ is absolutely continuous}, \quad i = 1, \ldots, n. \tag{C1}$$

Let t_0 denote the common lower support point for the distributions of the lifetimes: that is,

$$F_i^*(t_0) = 0 \ \ and \ \ F_i^*(t) > 0 \ \ for \ \ t > t_0, \quad i = 1, \ldots, n. \tag{C2}$$

Let t_i stand for the upper support point of the distribution of X_i: that is,

$$F_i^*(t_i) = 1 \ \ and \ \ F_i^*(t) < 1 \ \ for \ \ t < t_i, \quad i = 1, \ldots, n. \tag{C3}$$

Suppose this is a k-out-of-n system, that is, it works as long as at least k of its components are working. The lifetime of this system can be characterized by the so-called fatal sets. In reliability literature a *fatal set* is a subset of components such that the failure of all the components in the subset causes the failure of the system. For a k-out-of-n system with non-atomic component lifetime distributions the collection of fatal sets is the collection of all $(n-k+1)$-element subsets of $\{1, \ldots, n\}$. Denote this collection as $\mathcal{A}$.

This paper considers the case when the only data are observed after the failure of the system and pertain to the system's lifetime Z and a *diagnostic set*, which is the set of components that have failed by time Z and which is revealed during the autopsy. Clearly, $\mathcal{A}$ is the collection of all possible diagnostic sets. To summarize, the following $M \equiv \binom{n}{n-k+1}$ sub-distribution functions are observed: for each $A \in \mathcal{A}$,

$$G_A^*(t) = P(Z \le t, \, A - \text{diagnostic set}), \quad t \ge t_0.$$

A more detailed discussion of such systems (and coherent systems in general) can be found, for instance, in Barlow and Proschan (1975).

For convenience let us assign an order to sets in $\mathcal{A}$ and write this collection as

$$\mathcal{A} = \{A_1, A_2, \ldots, A_{M-1}, A_M\}.$$

Then the collection of observable functions can be written as

$$G_m^*(t) = P(Z \le t, \, A_m - \text{diagnostic set}), \quad t \ge t_0,$$

where $m = 1, \dots, M$. Note that for $t \geq \max_{i \in A_m} t_i$ function G_m is constant:

$$G_m^*(t) = P(A_m \text{ – diagnostic set}), \quad t \geq \max_{i \in A_m} t_i.$$

Unknown underlying distributions F_i^*, $i = 1, \dots, n$, and observable sub-distributions G_m^*, $m = 1, \dots, M$, are related by the following system of M integral-differential equations:

$$G_m^*(t) = \int_{t_0}^{t} \left(\prod_{i \in A_m} F_i^*(s) \right)' \prod_{i \in A_m^c} (1 - F_i^*(s)) \, ds, \quad t \geq t_0, \quad m = 1, \dots, M, \qquad (IDE)$$

where $A_m^c = \{1, \dots, n\} \backslash A_m$. Indeed,

$$\begin{aligned} G_m^*(t) &= P\left(\max_{i \in A_m} X_i \leq t, \max_{i \in A_m} X_i \leq \min_{i \in A_m^c} X_i \right) \\ &= P\left(\max_{i \in A_m} X_i \leq t, \min_{i \in A_m^c} X_i > t \right) + P\left(\max_{i \in A_m} X_i \leq \min_{i \in A_m^c} X_i, \min_{i \in A_m^c} X_i \leq t \right). \end{aligned}$$

Since the value of the density of $\min_{i \in A_m^c} X_i$ at t is equal to $-\left(\prod_{i \in A_m^c} (1 - F_i^*(t)) \right)'$, then

$$\begin{aligned} G_m^*(t) &= \prod_{i \in A_m} F_i^*(t) \prod_{i \in A_m^c} (1 - F_i^*(t)) - \int_{t_0}^{t} \prod_{i \in A_m} F_i^*(s) \left(\prod_{i \in A_m^c} (1 - F_i^*(s)) \right)' ds \\ &= \int_{t_0}^{t} \left(\prod_{i \in A_m} F_i^*(s) \right)' \prod_{i \in A_m^c} (1 - F_i^*(s)) \, ds, \quad t \geq t_0. \end{aligned}$$

The question of identifying marginal distributions of components' lifetimes in coherent systems from the joint distribution of observed "autopsy" data, which is comprised of the lifetime of the system and a diagnostic set, is raised in Meilijson (1981). In Meilijson, under certain restrictions on a coherent system's structure (namely, on the rank of the incidence matrix), the distributions of the components' lifetimes are identified in the case of independent and non-atomic lifetimes that possess the same essential infimum and supremum.

The 1-out-of-n system only discloses the marginal distribution of the maximum of all the components' lifetimes, so the individual marginal distributions cannot be identified. For

k-out-of-n systems with $2 \leq k \leq n$, applying the same techniques as the ones for 2-out-of-n systems in Komarova (2013), one can establish the following identifiability result.

Proposition 2.1. *If $2 \leq k \leq n$, the distribution functions F_i^*, $i = 1, \ldots, n$, that satisfy conditions (C1) and (C2) are identifiable on $[t_0, T]$, where T is the $(n-k+1)$-th order statistic of $\{t_1, \ldots, t_n\}$, from observable functions G_m^*, $m = 1, \ldots, M$, if the following condition holds:*

For each $m = 1, \ldots, M$, the function

$$\left(\sum_{A \in \mathcal{A}} \frac{1}{\prod_{i \in A} F_i^*(t)}\right) \cdot \left(\prod_{i \in A_m} F_i^*(t)\right)' \cdot \sum_{i \in A_m^c} F_i^*(t) \tag{C4}$$

has a finite Lebesgue integral in a right neighborhood of t_0.

The mathematical technique of this identification result is based on establishing that if distribution functions F_i satisfy conditions (C1), (C2) and (C4), then the system of integral-differential equations (IDE) together with the initial conditions

$$F_i^*(t_0) = 0, \quad i = 1, \ldots, n, \tag{IC}$$

has a unique positive solution in a right-hand side neighborhood of t_0.

Let $\mathcal{C}_i$ denote the collection of fatal sets containing i and let $\mathcal{C}_{-i}$ stand for the collection of fatal sets not containing i:

$$\mathcal{C}_i = \{A \in \mathcal{A} \,|\, i \in A\},$$
$$\mathcal{C}_{-i} = \{A \in \mathcal{A} \,|\, i \in A^c\}.$$

Remark 2.2. *Applying techniques similar to those for 2-out-of-n systems in Komarova (2013), it can be shown that conditions (C1) and (C2) imply the following conditions on observable functions:*

1. *G_m^* is absolutely continuous, $m = 1, \ldots, M$.*

2. $G_m^*(t_0) = 0$ *and* $G_m^*(t) > 0$ *for* $t > t_0$, $m = 1, \ldots, M$.

3. If $2 \le k \le n$, *then for any* $i = 1, \ldots, n$,

$$\lim_{t \downarrow t_0} \prod_{A \in \mathcal{C}_i} G_A^*(t) \cdot \prod_{A \in \mathcal{C}_{-i}} G_A^*(t)^{-\frac{n-k}{k-1}} = 0. \tag{2.1}$$

The first two of these conditions are obvious. As for the third one, (IDE) implies in the case of $2 \le k \le n$ that for any $i = 1, \ldots, n$,

$$\lim_{t \downarrow t_0} \frac{\prod_{A \in \mathcal{C}_i} G_A^*(t) \cdot \prod_{A \in \mathcal{C}_{-i}} G_A^*(t)^{-\frac{n-k}{k-1}}}{F_i^*(t)^{\binom{n-1}{n-k}}} = 1.$$

Using (C1) and (C2), we obtain (2.1).

Also, it can be shown that condition (C4) can be equivalently written in terms of observable functions:

For each $m = 1, \ldots, M$, *the function*

$$\left(\sum_{A \in \mathcal{A}} \frac{1}{G_A^*(t)} \right) \cdot g_m(t) \cdot \sum_{i \in A_m^c} \left(\prod_{A \in \mathcal{C}_i} G_A^*(t) \cdot \prod_{A \in \mathcal{C}_{-i}} G_A^*(t)^{-\frac{n-k}{k-1}} \right)^{\binom{n-1}{n-k}^{-1}}$$

has a finite Lebesgue integral in a right neighborhood of t_0.

3 Sieve estimation

Throughout this section it is assumed that $2 \le k \le n$. This section presents an approach to estimating functions F_i^* from a random sample. First, an operator B is defined that maps F_i^* to observable functions G_m^*. It is shown that this operator is Lipschitz and that under weak conditions on the set of possible distributions $F = (F_1, \ldots, F_n)$, its inverse operator B^{-1} is continuous. Sieve estimators of F_i^* are then derived and proved consistent using the properties of B.

3.1 Operator B

Here and throughout the paper T will stand for the $(n-k+1)$-th order statistic of $\{t_1,\ldots,t_n\}$.

For an absolutely continuous function $F=(F_1,\ldots,F_n)^{tr}$ with domain $[t_0,T]$ define the M-dimensional vector function

$$B(F)=(B(F)_1,B(F)_2,\ldots,B(F)_{M-1},B(F)_M)^{tr}$$

as follows:

$$B(F)_m(t)=\int_{t_0}^{t}\left(\prod_{i\in A_m}F_i(s)\right)'\prod_{i\in A_m^c}(1-F_i(s))\,ds,\quad t\in[t_0,T] \tag{3.1}$$

for $m=1,\ldots,M$.

Let Λ be the set of vector functions $F=(F_1,\ldots,F_n)^{tr}$ with domain $[t_0,T]$ satisfying the following conditions:

Conditions (I).

(i) F_i *is absolutely continuous on* $[t_0,T]$, $i=1,\ldots,n$.

(ii) F_i *is increasing on* $[t_0,T]$,

(iii) $F_i(t_0)=0$ *and* $F_i(t)>0$ *for* $t\in(t_0,T]$, $i=1,\ldots,n$.

(iv) $F_i(T)\leq 1$, $i=1,\ldots,n$.

(v) For each $m=1,\ldots,M$, *the function*

$$\left(\sum_{A\in\mathcal{A}}\frac{1}{\prod_{i\in A}F_i(t)}\right)\cdot\left(\prod_{i\in A_m}F_i(t)\right)'\cdot\sum_{i\in A_m^c}F_i(t)$$

has a finite Lebesgue integral in a right neighborhood of t_0.

Let B be defined on Λ. Properties of the image $B(\Lambda)$ are easily deduced from conditions (I): each function $B(F)_m$ is absolutely continuous and increasing on $[t_0,T]$, $B(F)_m(t_0)=0$

and $B(F)_m(t) > 0$ for $t \in (t_0, T]$. The identification result in Proposition 2.1 means that there exists the inverse operator $B^{-1} : B(\Lambda) \to \Lambda$.

Endow the domain Λ and its image $B(\Lambda)$ with the following uniform metric spaces[3]:

$$d(F, \tilde{F}) = \sup_{t \in [t_0, T]} \sqrt{(F(t) - \tilde{F}(t))^{tr}(F(t) - \tilde{F}(t))},$$
$$d(B(F), B(\tilde{F})) = \sup_{t \in [t_0, T]} \sqrt{(B(F)(t) - B(\tilde{F})(t))^{tr}(B(F)(t) - B(\tilde{F})(t))}.$$

Properties of B are important for proving the consistency of the estimators introduced later in this section. Usually, it is easier to obtain desirable properties of B and establish consistency when the space of unknown functions is compact. Compactify Λ by bounding density functions F_i' by the same Lebesgue integrable function:

Condition (II).

$$F_i'(t) \le \phi'(t) \quad a.e. \quad [t_0, T], \quad i = 1, \dots, n,$$

where ϕ is some absolutely continuous increasing function on $[t_0, T]$.

Let Λ_ϕ be the subset of Λ such that all functions F from Λ_ϕ satisfy condition *(II)*. This condition guarantees that Λ_ϕ is relatively compact under the uniform metric. Indeed, for any $F \in \Lambda_\phi$ and $t, \tau \in [t_0, T]$,

$$|F_i(t) - F_i(\tau)| = \left| \int_\tau^t F_i'(s) ds \right| \le |\phi(t) - \phi(\tau)|, \quad i = 1, \dots, n.$$

Because ϕ is absolutely continuous, the last inequality implies that Λ_ϕ is equicontinuous. It is also uniformly bounded because the values of F_i do not exceed 1. According to the Arzela-Ascoli theorem, Λ_ϕ is relatively compact in metric $d(\cdot, \cdot)$.

Note that if $F \in \Lambda_\phi$, then each function $B(F)_m$, satisfies the following condition:

$$B(F)_m'(t) \le k\phi'(t) \quad a.e. \quad [t_0, T], \quad m = 1, \dots, M.$$

[3] $d(F, \tilde{F})$ and $d(B(F), B(\tilde{F}))$ are the compositions of the sup metric in the function space and the Euclidean metric in R^n and R^M, respectively. Therefore, all the axioms in the definition of a metric are satisfied for $d(F, \tilde{F})$ and $d(B(F), B(\tilde{F}))$.

Let $\overline{\Lambda}_\phi$ stand for the closure of Λ_ϕ under metric $d(\cdot,\cdot)$. Because Λ_ϕ is relatively compact, $\overline{\Lambda}_\phi$ is a compact set. To consider operator B on $\overline{\Lambda}_\phi$, we first need to show that B is defined for functions in $\overline{\Lambda}_\phi \backslash \Lambda_\phi$. The proposition below establishes that all functions in $\overline{\Lambda}_\phi$ satisfy conditions *(I)(i), (I)(ii), (I)(iv), (I)(v)* and a slightly modified condition *(I)(iii)*, and also satisfy condition *(II)*.

Proposition 3.1. *If $F = (F_1, \dots, F_n)^{tr} \in \overline{\Lambda}_\phi$, then each F_i, $i = 1, \dots, n$, is absolutely continuous, increasing on $[t_0, T]$, satisfies $F_i(t_0) = 0$, $F_i(T) \leq 1$ and is such that $F_i'(t) \leq \phi'(t)$ a.e. on $[t_0, T]$.*

The proof of Proposition 3.1 is in Appendix.

Functions that are in $\overline{\Lambda}_\phi$ but not in Λ_ϕ are, for instance, those that are equal to 0 in a small right-hand side neighborhood of t_0.

Because all functions in $\overline{\Lambda}_\phi$ are absolutely continuous, operator B can be extended from Λ_ϕ to $\overline{\Lambda}_\phi \backslash \Lambda_\phi$ by applying (3.1) to each $F \in \overline{\Lambda}_\phi \backslash \Lambda_\phi$.

The next proposition implies that B is continuous in metric $d(\cdot,\cdot)$ on $\overline{\Lambda}_\phi$.

Proposition 3.2. *For any $F, \tilde{F} \in \overline{\Lambda}_\phi$,*

$$d(B(F), B(\tilde{F})) \leq M\sqrt{n}\; d(F, \tilde{F}). \tag{3.2}$$

that is, operator B is Lipschitz on $\overline{\Lambda}_\phi$.

The proof of Proposition 3.2 is in Appendix.

Finally, the continuity property of B and the compactness of $B(\Lambda_\phi)$ are used to establish the continuity of B^{-1} on $B(\Lambda_\phi)$.

Proposition 3.3. *B^{-1} is continuous on $B(\Lambda_\phi)$.*

The proof of Proposition 3.3 is in Appendix.

3.2 Estimator

This subsection defines sieve estimators of distribution functions F_i^* and proves their consistency.

Note that $G^* = B(F^*)$, where $F^* = (F_1^*, \ldots, F_n^*)^{tr}$ and $G^* = (G_1^*, \ldots, G_M^*)^{tr}$. Let us choose ϕ in such a way that $F^* \in \Lambda_\phi$.[4]

The next lemma introduces an objective function Q defined at each $F \in \overline{\Lambda}_\phi$ and uses the identification result from section 2 to show that it is uniquely minimized at $F = F^*$.

Lemma 3.4. *F^* is the unique minimizer of*

$$Q(F) = \int_{t_0}^{T} (G^*(t) - B(F)(t))^{tr}(G^*(t) - B(F)(t)) \frac{\sum_{m=1}^{M} {G^*}'_m(t)}{\sum_{m=1}^{M} G_m^*(T)} dt$$

on $\overline{\Lambda}_\phi$.

The proof of Lemma 3.4 is in Appendix.

Note that $\frac{\sum_{m=1}^{M} {G^*}'_m(t)}{\sum_{m=1}^{M} G_m^*(T)}$ is the probability density function of the lifetime of the system on $[t_0, T]$.

The idea of sieve estimation here is to use the sample analog of Q and approximate $\overline{\Lambda}_\phi$ with finite-dimensional spaces. For instance, for each $r = 1, 2, \ldots$, choose base functions $p_{1,r}, \ldots, p_{\gamma(r),r}$ (for example, B-splines with uniform knots or basic Bernstein polynomials) and introduce the set of linear combinations of these functions:

$$\Gamma_r = \{(F_1, \ldots, F_n)^{tr} : F_i(t) = \sum_{l=1}^{\gamma(r)} \alpha_l^i p_{l,r}(t),\, t \in [t_0, T]\}.$$

In this set of functions, consider only those functions that are in Λ_ϕ:

$$\Sigma_r = \Lambda_\phi \cap \Gamma_r.$$

Set Σ_r consists of functions from Γ_r with certain restrictions on coefficients α_l^i. It is relatively compact and, hence, its closure $\overline{\Sigma}_r$ is compact, and $\overline{\Sigma}_r \subset \overline{\Lambda}_\phi$.

Consider a random sample of N observations $\{(z_j, a_j)\}_{j=1}^{N}$, where z_j is the observed lifetime of the system and a_j is the diagnostic set in j's round. Without a loss of generality, assume that $z_j \leq z_{j+1}$, $j = 1, \ldots, N-1$. From the sample, find consistent estimators $\widehat{G}_{m,N}$

[4]For instance, we can assume that $\phi'(t) \geq \sum_{i=1}^{n} {F_i^*}'(t)$.

of G_m, for instance, empirical sub-distribution functions on $[t_0, T]$:

$$\widehat{G}_{m,N}(t) = \frac{1}{N}\sum_{j=1}^{N} 1(z_j \le t)1(a_j = A_m), \quad m = 1, \dots, M.$$

The sample objective function is

$$\widehat{Q}_N(F) = \frac{1}{N}\sum_{j=1}^{N}(\widehat{G}_N(z_j) - B(F)(z_j))^{tr}(\widehat{G}_N(z_j) - B(F)(z_j)),$$

where $\widehat{G}_N = (\widehat{G}_{1,N}, \dots, \widehat{G}_{M,N})^{tr}$. Note that since the lifetime of the system cannot be strictly greater than T, then all z_j belong to $[t_0, T]$ and $\{(z_j)\}_{j=1}^N$ is a random sample from the distribution with density function $\frac{\sum_{m=1}^M {G^*}'_m(t)}{\sum_{m=1}^M G^*_m(T)}$.

Let $r = r(N)$, and define the following estimator of F^*:

$$\widehat{F}_N = \underset{F \in \overline{\Sigma}_{r(N)}}{\operatorname{argmin}}\ \widehat{Q}_N(F).$$

Theorem 3.5 below establishes the consistency of estimator $\widehat{F}_N$ when sets $\overline{\Sigma}_r$ well approximate set $\overline{\Lambda}_\phi$.

Theorem 3.5. *If*

$$\forall(F \in \overline{\Lambda}_\phi)\exists(\widetilde{F} \in \overline{\Sigma}_r) \quad \textit{such that} \quad d(F, \widetilde{F}) \xrightarrow{p} 0 \ \textit{as}\ r = r(N) \to \infty, \tag{3.3}$$

then estimator $\widehat{F}_N$ is consistent:

$$d(\widehat{F}_N, F^*) \xrightarrow{p} 0 \ \textit{as}\ N \to \infty.$$

Condition (3.3) holds if approximating sets are chosen properly – e.g., if base functions $p_{1,r}, \dots, p_{\gamma(r),r}$ are B-splines with uniform knots, Bernstein polynomials or truncated power series.

The proof of Theorem 3.5 is in Appendix.

The next natural question is the rate of convergence of the proposed estimator. It is a difficult question to answer and is left for future research. Most of the results available in the literature relate to convergence rates of M-estimators, usually in the L^2-metric (e.g., see Chen (2007)). The sieve estimator proposed in this paper is of MD-type. One potential approach to finding the rate of convergence of this estimator is to apply results and techniques from Van der Vaart and Wellner (1996) for Z-estimators. In particular, this would require analyzing properties of Fréchet derivatives of Q and Q_N at both F^* and $\widehat{F}_N$ (in the sup metric $d(\cdot,\cdot)$). Another important factor would be the quality of the projection of continuous functions on chosen sieve spaces for a given N. It would be determined by the type of the chosen sieve space and the rate of $r(N)$ as $N \to \infty$. Overall, the rate of convergence in the sup metric $d(\cdot,\cdot)$ results would rely on properties of empirical processes.

4 Monte-Carlo experiment

This section illustrates the suggested sieve estimation method for 2-out-of-3 systems. Suppose that the lifetimes of all three components are distributed on the support $[0,1]$. Consider two Monte-Carlo scenarios.

Scenario 1. For $t \in [0,1]$,

$$F_1^*(t) = t^{\frac{1}{2}}, \quad F_2^*(t) = t^{\frac{2}{3}}, \quad F_3^*(t) = t^{\frac{3}{4}}.$$

All these functions are strictly concave and their derivatives approach ∞ as $t \downarrow 0$. In some sense these functions are not very different in their behavior around $t = 0$.

Scenario 2. For $t \in [0,1]$,

$$F_1^*(t) = \begin{cases} t^{\frac{1}{2}} & \text{if } t \in \left[0, \frac{1}{2}\right], \\ 1 - (\sqrt{2}-1)(1-t)^{\frac{1}{2}} & \text{if } t \in \left(\frac{1}{2}, 1\right]. \end{cases}$$

$$F_2^*(t) = \frac{e^t - 1}{e - 1}, \quad F_3^*(t) = t.$$

The behavior of F_1^* on $[0,1]$ is different from that of F_2^* or F_3^*. The derivative of F_1^*

approaches ∞ as $t \downarrow 0$ or $t \uparrow 1$. Functions F_2^* and F_3^* are infinitely differentiable on $[0,1]$.

Bernstein polynomials. As sieve spaces, consider spaces of Bernstein polynomials. Namely, consider linear sieve spaces with the basic Bernstein polynomials as the base functions. The basic Bernstein polynomials of power r on $[0,1]$ are the following $r+1$ functions:

$$p_{l,r}(t) = \binom{r}{l} t^l (1-t)^{r-l}, \quad l = 0, \ldots, r.$$

The corresponding sieve space is

$$\Gamma_r = \left\{ (F_1, F_2, F_3)^{tr} : F_i(t) = \sum_{l=0}^{r} \alpha_l^i \, p_{l,r}(t), \; t \in [0,1] \right\}.$$

An important property of Bernstein polynomials[5] says that for a continuous function f on $[0,1]$, the relation

$$\lim_{r \to \infty} \sum_{l=0}^{r} f\left(\frac{l}{r}\right) \binom{r}{l} t^l (1-t)^{r-l} = f(t)$$

holds uniformly on $[0,1]$. This property implies that the constraints

$$\alpha_0^i \le \alpha_1^i \le \ldots \le \alpha_{r-1}^i \le \alpha_r^i$$

imposed for each $i = 1,2,3$ guarantee that functions in Γ_r are increasing. Conditions

$$\alpha_0^i = 0 \quad \text{and} \quad \alpha_r^i = 1$$

guarantee that $F_i(0) = 0$ and $F_i(1) = 1$, respectively.

Scenario 1. N = 500.

Table 1 is constructed based on the simulations of outcomes in 500 runs of the system. It shows how often each of the subsets $A_1 = \{2,3\}$ $A_2 = \{1,3\}$ and $A_3 = \{1,2\}$ happens to be the set responsible for the failure of system, or in other words, is the diagnostic set which is discovered during the autopsy. The table also shows the minimum, the maximum

[5] See Lorentz (1986).

and the average lifetime of the system in each of these situations.

	diagnostic	$\min Z$	$\max Z$	Z_{av}
$A_1 = \{2, 3\}$	125 (25%)	0.0146	0.8756	0.3867
$A_2 = \{1, 3\}$	181 (36.6%)	0.0033	0.9438	0.3665
$A_3 = \{1, 2\}$	194 (38.8%)	0.0114	0.9461	0.3553

Table 1. Monte Carlo experiment for Scenario 1 with $N = 500$ rounds. Number of rounds in which each A_m is the diagnostic set discovered during autopsy (*diagnostic*), the minimum lifetime ($\min Z$), the maximum lifetime ($\max Z$), the average lifetime (Z_{av}).

We can think about our 2-out-of-3 system as an observed open ascending auction with three bidders having independent private values. In this auction, bidders hold down a button as the auctioneer raises the price. When the price gets too high for a bidder, she drops out by releasing the button. The auction ends when only one bidder remains. This person wins the object and pays the price at which the auction stopped. The distribution of the lifetime of component i corresponds to the distribution of bidder i's private value. Observing $A_1 = \{2, 3\}$ as a diagnostic set after the failure of the system is equivalent to the case of bidder 3 winning the auction. Analogously, observing $A_2 = \{1, 3\}$ as a diagnostic set is equivalent to the case of bidder 2 winning the auction, and observing $A_1 = \{2, 3\}$ as a diagnostic set is equivalent to the case of bidder 1 winning the auction. The observed lifetime of the system corresponds to the observed transaction price.

From Table 1, the highest observed transaction price in the simulated data is approximately 0.9461 (when bidder 3 wins the auction) and the lowest observed transaction price is 0.0033 (when bidder 2 wins the auction). As we see, bidder 3 wins the auction most often which stems from the fact that the distribution of private value of bidder 3 first-order stochastically dominates that of bidder 1 and that of bidder 2. Bidder 1 wins the auction least often because the distribution of private value of bidder 1 is first-order stochastically dominated by that of bidder 2 and that of bidder 3.

Sieve estimation uses Bernstein polynomials of order 4 with the constraints on the coefficients that guarantee the monotonicity of sieve estimators for F_1^*, F_2^* and F_3^*. These estimators are depicted in Figure 1.

Scenario 2. N = 500.

	diagnostic	$\min Z$	$\max Z$	Z_{av}
$A_1 = \{2, 3\}$	98 (19.6%)	0.0857	0.9635	0.5832
$A_2 = \{1, 3\}$	213 (42.6%)	0.0216	0.9848	0.4467
$A_3 = \{1, 2\}$	189 (37.8%)	0.0288	0.9187	0.4599

Table 2. Monte Carlo experiment for Scenario 2 with $N = 500$ rounds. Number of rounds in which each A_m is the diagnostic set discovered during autopsy (*diagnostic*), the minimum lifetime ($\min Z$), the maximum lifetime ($\max Z$), the average lifetime (Z_{av}).

The distribution of private value of bidder 2 first-order stochastically dominates that of bidder 2. Also, there is no first-order stochastic relationship between the distribution of private value of bidder 1 and those of bidders 2 and 3.

Figure 2 shows the results of sieve estimation of F_i^*, $i = 1, 2, 3$, by Bernstein polynomials of order 4 with monotonicity constraints. As can be seen, the sieve estimator for F_1^* provides a worse approximation of this function around $t = 0$ and $t = 1$ than on the rest of the support because F_1^* has infinite derivative from the right at $t = 0$ and infinite derivative from the left at $t = 1$.

5 Appendix

Proof of Proposition 3.1. Let us start by establishing absolute continuity. Because $F \in \overline{\Lambda}_\phi$, then there exists a sequence $F_q \in \Lambda_\phi$ such that $d(F_q, F) \to 0$ as $q \to \infty$. Take any two points $t_1, t_2 \in [t_0, T]$. Convergence in metric $d(\cdot, \cdot)$ implies point-wise convergence. Therefore, for any $i = 1, \ldots, n$,

$$|F_i(t_1) - F_i(t_2)| = \lim_{q \to \infty} |F_{q,i}(t_1) - F_{q,i}(t_2)| \leq |\phi(t_1) - \phi(t_2)|.$$

The last inequality and the absolute continuity of ϕ imply that each F_i is absolutely continuous.

Because functions $F_{q,i}$ are increasing and converge to F_i point-wise, then F_i is increasing.

Because the values of $F_{q,i}(t_0)$ converge to $F_i(t_0)$, then $F_i(t_0) = 0$.

Because $F_{q,i}(T) \leq 1$ and $F_{q,i}$ converge to F_i point-wise, then $F_i(T) \leq 1$.

Because F_i is absolutely continuous, it can differentiated a.e. on $[t_0, T]$. Let t be a point at which both F_i and ϕ have derivatives. For any fixed h,

$$\frac{F_i(t+h) - F_i(t)}{h} = \lim_{q\to\infty} \frac{F_{q,i}(t+h) - F_{q,i}(t)}{h} \leq \frac{\phi(t+h) - \phi(t)}{h}.$$

Taking the limit as $h \to 0$, we obtain that $F_i'(t) \leq \phi'(t)$.

Proof of Proposition 3.2. Let F, $\tilde{F} \in \overline{\Lambda_\phi}$. For convenience, let us temporarily use the following metric:

$$d_1(F, \tilde{F}) = \sup_{t\in[t_0,T]} \sum_{i=1}^{n} |F_i(t) - \tilde{F}_i(t)|,$$
$$d_1(B(F), B(\tilde{F})) = \sup_{t\in[t_0,T]} \sum_{m=1}^{M} |B(F)_m(t) - B(\tilde{F})_m(t)|.$$

From the definition of B,

$$B(F)_m(t) - B(\tilde{F})_m(t) = \int_{t_0}^{t} \left(\prod_{i\in A_m} F_i(s) - \prod_{i\in A_m} \tilde{F}_i(s) \right)' \prod_{i\in A_m^c} (1 - F_i(s))\, ds$$
$$+ \int_{t_0}^{t} \left(\prod_{i\in A_m} \tilde{F}_i(s) \right)' \left(\prod_{i\in A_m^c} (1 - F_i(s)) - \prod_{i\in A_m^c} \left(1 - \tilde{F}_i(s)\right) \right) ds.$$

Integration by parts gives that

$$\int_{t_0}^{t} \left(\prod_{i\in A_m} F_i(s) - \prod_{i\in A_m} \tilde{F}_i(s) \right)' \prod_{i\in A_m^c} (1 - F_i(s))\, ds = \left(\prod_{i\in A_m} F_i(t) - \prod_{i\in A_m} \tilde{F}_i(t) \right) \prod_{i\in A_m^c} (1 - F_i(t))$$
$$+ \int_{t_0}^{t} \left(\prod_{i\in A_m} F_i(s) - \prod_{i\in A_m} \tilde{F}_i(s) \right) \left(- \prod_{i\in A_m^c} (1 - F_i(s)) \right)' ds,$$

and thus,

$$\left| \int_{t_0}^{t} \left(\prod_{i\in A_m} F_i(s) - \prod_{i\in A_m} \tilde{F}_i(s) \right)' \prod_{i\in A_m^c} (1 - F_i(s))\, ds \right| \leq \sup_{t\in[t_0,T]} \left| \prod_{i\in A_m} F_i(t) - \prod_{i\in A_m} \tilde{F}_i(t) \right|$$
$$\leq \sup_{t\in[t_0,T]} \sum_{i\in A_m} |F_i(t) - \tilde{F}_i(t)|.$$

Also note that

$$\left|\int_{t_0}^{t}\left(\prod_{i\in A_m}\tilde{F}_i(s)\right)'\left(\prod_{i\in A_m^c}(1-F_i(s))-\prod_{i\in A_m^c}\left(1-\tilde{F}_i(s)\right)\right)ds\right|\le\sup_{t\in[t_0,T]}\left|\prod_{i\in A_m^c}(1-F_i(t))-\prod_{i\in A_m^c}\left(1-\tilde{F}_i(t)\right)\right|$$
$$\le\sup_{t\in[t_0,T]}\sum_{i\in A_m^c}|F_i(t)-\tilde{F}_i(t)|.$$

To summarize,

$$\left|B(F)_m(t)-B(\tilde{F})_m(t)\right|\le\sup_{t\in[t_0,T]}\sum_{i=1}^{n}|F_i(t)-\tilde{F}_i(t)|=d_1(F,\tilde{F}),$$

which implies that

$$d_1(B(F),B(\tilde{F}))\le M\ d_1(F,\tilde{F}).$$

Because

$$d_1(F,\tilde{F})\le\sqrt{n}\ d(F,\tilde{F})\ \text{ and }\ d_1(B(F),B(\tilde{F}))\ge d(B(F),B(\tilde{F})),\tag{5.1}$$

then

$$d(B(F),B(\tilde{F}))\le M\sqrt{n}\ d(F,\tilde{F}).$$

Proof of Proposition 3.3. Essentially, the statement of this proposition follows from the fact that if a continuous operator is defined on a compact set and the inverse operator is defined on the image of that set, then the inverse operator is continuous. This result cannot be applied here directly however because even though the inverse operator B^{-1} is clearly defined on $B(\Lambda_\phi)$ it is not defined on the larger set $B(\overline{\Lambda}_\phi)$.

Let $G_0\in B(\Lambda_\phi)$ and $d(G_q,G_0)\to 0$ as $q\to\infty$ for $G_q\in B(\Lambda_\phi)$. Denote $F_0=B^{-1}(G_0)$, $F_q=B^{-1}(G_q)$. Clearly, F_0, $F_q\in\Lambda_\phi$. I want to show that $d(F_q,F_0)\to 0$ as $q\to\infty$. Suppose this is not so and for some $\varepsilon>0$ there exists a subsequence F_{q_l} such that

$$d\left(F_{q_l},F_0\right)>\varepsilon\quad\text{for all } l=1,2,\ldots.\tag{5.2}$$

Notice that the subsequence F_{q_l} is equicontinuous because all functions in it are bounded and

$$|F_{q_l}(t_1) - F_{q_l}(t_2)| \leq |\phi(t_1) - \phi(t_2)|$$

for any $t_1, t_2 \in [t_0, T]$. According to the Arzela-Ascoli theorem, there is a convergent subsequence $F_{q_{l_j}}$. Let $\widetilde{F}$ be the limit of $F_{q_{l_j}}$. Because $\widetilde{F} \in \overline{\Lambda}_\phi$ and B is continuous on $\overline{\Lambda}_\phi$, then

$$d(B(F_{q_{l_j}}), B(\widetilde{F})) \to 0.$$

Thus, $B(\widetilde{F}) = G_0$. Given that on $B(\Lambda_\phi)$ the inverse operator B^{-1} is defined, conclude that $\widetilde{F} = F_0$. Thus, we obtain that $d(F_{q_{l_j}}, F_0) \to 0$, contradicting (5.2). Therefore, $d(F_q, F_0) \to 0$.

Proof of Lemma 3.4. Note that $Q(F^*) = 0$. Because the inverse operator B^{-1} exists on $B(\Lambda_\phi)$, then $B(F) \neq G^*$ and, hence, $Q(F) > 0$ for any $F \in \Lambda_\phi$ such that $F \neq F^*$.

Now consider $F \in \overline{\Lambda}_\phi \backslash \Lambda_\phi$. Since $F \notin \Lambda_\phi$, then some F_i takes value 0 in a right neighborhood of t_0. Without loss of generality assume that $F_1(t) = 0$, $t \in [t_0, t_0+\omega)$. Then for every $m = 1, \ldots, M$, such that $1 \in A_m^c$, we have $B(F)_m(t) = 0$, $t \in [t_0, t_0 + \omega)$. Because $G_m^*(t) > 0$ for $t > t_0$, $m = 1, \ldots, M$ (see Remark 2.2), then obviously $B(F) \neq G^*$.

Proof of Theorem 3.5. To prove this theorem, use lemmas A1 and A2 from Newey and Powell (2003).[6] Consistency will hold if all conditions in Lemma A1 are satisfied. Below these conditions are divided into three groups, as in Newey and Powell (2003).

(i) According to Lemma 3.4, F^* is the unique minimizer of Q on $\overline{\Lambda}_\phi$.

(ii) Set $\overline{\Lambda}_\phi$ is compact. Let me show that Q and $\widehat{Q}_N$ are continuous on $\overline{\Lambda}_\phi$ and

$$\sup_{F \in \overline{\Lambda}_\phi} |\widehat{Q}_N(F) - Q(F)| \xrightarrow{p} 0. \tag{5.3}$$

The continuity of Q and $\widehat{Q}_N$ will follow from the properties of B on $\overline{\Lambda}_\phi$. First, consider Q.

[6]Some theorems from Chen (2007) can also be used to prove this result.

For any $F,\ \widetilde{F} \in \overline{\Lambda}_\phi$

$$|Q(F) - Q(\widetilde{F})| = \left| \int_{t_0}^{T} \left[(G^* - B(F))^{tr}(G^* - B(F)) - (G^* - B(\widetilde{F}))^{tr}(G^* - B(\widetilde{F})) \right] \frac{\sum_{m=1}^{M} G^{*\prime}_m(t)}{\sum_{m=1}^{M} G^*_m(T)}\, dt \right| =$$
$$= \left| \int_{t_0}^{T} \left[\sum_{m=1}^{M} (B(\tilde{F})_m - B(F)_m)(2G^*_m - B(F)_m - B(\widetilde{F})_m) \right] \frac{\sum_{m=1}^{M} G^{*\prime}_m(t)}{\sum_{m=1}^{M} G^*_m(T)}\, dt \right|.$$

For any $t \in [t_0, T]$, $B(F)_m(t) \le 1$ and $G^*_m(t) \le 1$, $m = 1, \ldots, M$, therefore

$$|Q(F) - Q(\widetilde{F})| \le 4 \int_{t_0}^{T} \left[\sum_{m=1}^{M} |B(\widetilde{F})_m - B(F)_m| \right] \frac{\sum_{m=1}^{M} G^{*\prime}_m(t)}{\sum_{m=1}^{M} G^*_m(T)}\, dt.$$

Applying the Cauchy-Schwartz inequality and (3.2),

$$|Q(F) - Q(\widetilde{F})| \le 4\sqrt{M} \int_{t_0}^{T} \left[\sqrt{(B(\widetilde{F}) - B(F))^{tr}(B(\widetilde{F}) - B(F))} \right] \frac{\sum_{m=1}^{M} G^{*\prime}_m(t)}{\sum_{m=1}^{M} G^*_m(T)}\, dt$$
$$\le 4\sqrt{M}\ d(B(F), B(\widetilde{F})) \le 4M\sqrt{Mn}\ d(F, \widetilde{F}).$$

Thus, function Q is Lipschitz and therefore continuous.

Now consider function $\widehat{Q}_N$. Similar to the methods described above,

$$|\widehat{Q}_N(F) - \widehat{Q}_N(\widetilde{F})| \le \frac{1}{N} \sum_{j=1}^{N} \sum_{m=1}^{M} |(\widehat{G}_{N,m}(z_j) - B(F)_m(z_j))^2 - (\widehat{G}_{N,m}(z_j) - B(\widetilde{F})_m(z_j))^2| =$$
$$= \frac{1}{N} \sum_{j=1}^{N} \sum_{m=1}^{M} |(B(\widetilde{F})_m(z_j) - B(F)_m(z_j))(2\widehat{G}_{N,m}(z_j) - B(\widetilde{F})_m(z_j) - B(F)_m(z_j))| \le$$
$$\le \frac{4\sqrt{M}}{N} \sum_{j=1}^{N} \sqrt{(B(\widetilde{F})(z_j) - B(F)(z_j))^{tr}(B(\widetilde{F})(z_j) - B(F)(z_j))} \le$$
$$\le 4\sqrt{M}\ d(B(F), B(\widetilde{F})) \le 4M\sqrt{Mn}\ d(F, \widetilde{F}). \tag{5.4}$$

Property (5.3) will follow from Lemma A2 in Newey and Powell (2003). Indeed, it is clear that

$$\forall (F \in \overline{\Lambda}_\phi) \quad \widehat{Q}_N(F) \xrightarrow{p} Q(F).$$

This fact combined with (5.4) implies (5.3).

(iii) This condition follows from assumption (3.3).

Conditions (i)-(iii) imply the consistency property (3.5).

References

[1] Abbring, J.H., and G.J. Van den Berg (2003). The identifiability of the mixed proportional hazards competing risks model. Journal of the Royal Statistical Society, Series B 65:701–710.

[2] Barlow, R.E., and F. Proschan (1975). *Statistical Theory of Reliability and Life Testing: Probability Models*. New York: Holt, Rinehart and Winston.

[3] Chen, X. (2007). Large Sample Sieve Estimation of Semi-Nonparametric Models. In: Edited by Heckman J.J., Leamer E.E. ed., *Handbook of Econometrics Vol. 6B*. Elsevier B.V.

[4] Chen, X. (2008). Sieve Extremum Estimation. In: Durlauf S.N., Blume L.E., *The New Palgrave Dictionary of Economics*, Second Edition. Palgrave Macmillan.

[5] Cox, D. R. (1962). *Renewal Theory*. London: Methuen.

[6] Doss, H., Huffer, F.W., and K.L. Lawson (1997). Bayesian nonparametric estimation via Gibbs sampling for coherent systems with redundancy. The Annals of Statistics 25:1109–1139.

[7] Heckman, J.J., and B.E. Honore (1989). The identifiability of the competing risks model. Biometrika 76:325–330.

[8] Komarova, T. (2013). A new approach to identifying generalized competing risks models with application to second-price auctions. Quantitative Economics 4:269–328.

[9] Lee, S., and A. Lewbel (2013). Nonparametric identification of accelerated failure time competing risks models. Econometric Theory 29:905–919.

[10] Lorentz, G. G. (1986). *Bernstein polynomials*. Chelsea Publishing Company, New York.

[11] Meilijson, I. (1981). Estimation of the lifetime distribution of the parts from the autopsy statistics of the machine. Journal of Applied Probability 18:829–838.

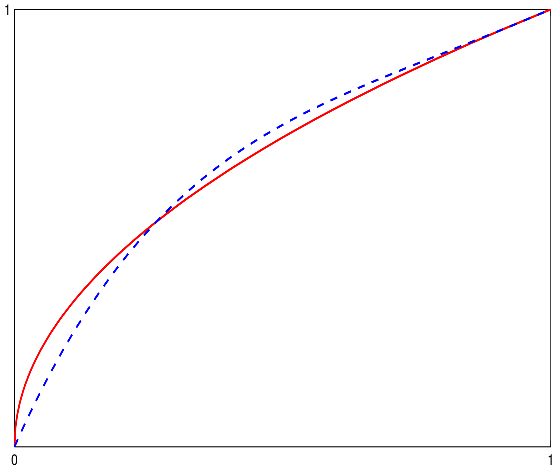
[12] Meilijson, I. (1994). Competing risks on coherent reliability systems: estimation in the parametric case. Journal of the American Statistical Association 89:1459–1464.

[13] Newey, W.K., and J.L. Powell (2003). Instrumental variable estimation of nonparametric models, Econometrica 71:1565–1578.

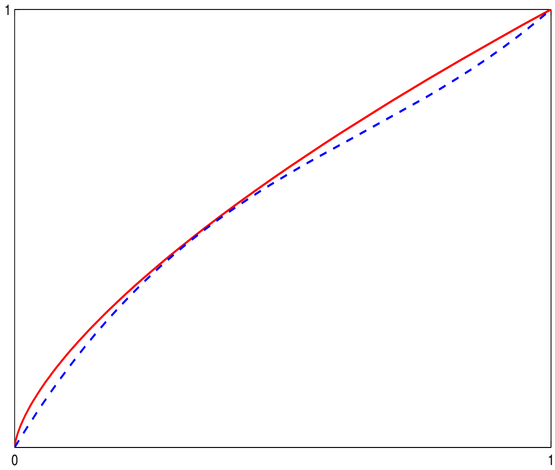
[14] Tsiatis, A. (1975). A nonidentifiability aspect of the problem of competing risks. Proceedings of the National Academy of Sciences 72:20–22.

[15] Van der Vaart, A. W., and J. A. Wellner (1996). *Weak convergence and empirical processes*. Springer-Verlag, New York.

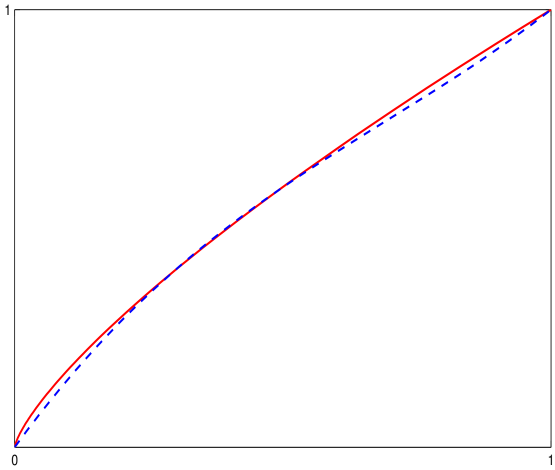
[16] Watelet, L.F. (1990). Nonparametric estimation of component life distributions in Meilijson's competing risks model. Ph.D. thesis, University of Washington, Department of Statistics.



Scenario 1: F_1^* (solid line) and its sieve estimator (dashed line).

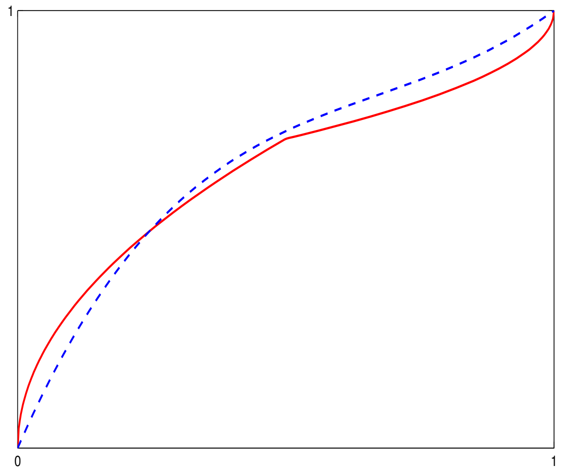


Scenario 1: F_2^* (solid line) and its sieve estimator (dashed line).

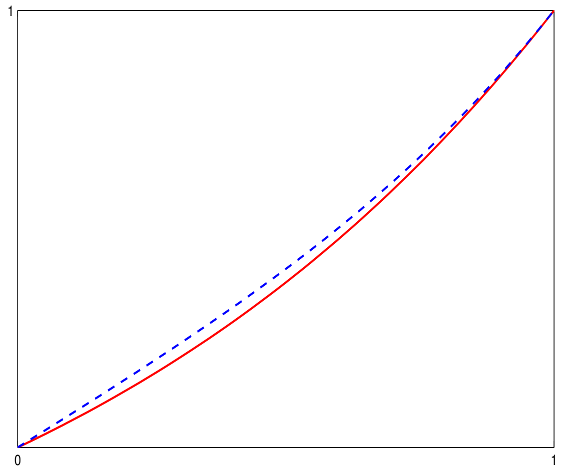


Scenario 1: F_3^* (solid line) and its sieve estimator (dashed line).

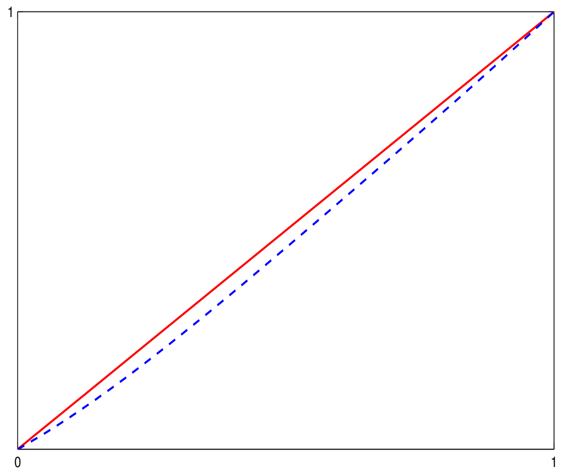
Figure 1. Scenario 1: CDFs F_i^*, $i = 1, 2, 3$, and their sieve estimators by Bernstein polynomials.



Scenario 2: F_1^* (solid line) and its sieve estimator (dashed line).



Scenario 2: F_2^* (solid line) and its sieve estimator (dashed line).



Scenario 2: F_3^* (solid line) and its sieve estimator (dashed line).

Figure 2. Scenario 2: CDFs F_i^*, $i = 1, 2, 3$, and their sieve estimators by Bernstein polynomials.