

QUANTILE AND QUANTILE-FUNCTION ESTIMATIONS UNDER DENSITY RATIO MODEL¹

BY JIAHUA CHEN AND YUKUN LIU²

University of British Columbia and East China Normal University

Population quantiles and their functions are important parameters in many applications. For example, the lower quantiles often serve as crucial quality indices for forestry products. Given several independent samples from populations satisfying the density ratio model, we investigate the properties of empirical likelihood (EL) based inferences. The induced EL quantile estimators are shown to admit a Bahadur representation that leads to asymptotically valid confidence intervals for functions of quantiles. We rigorously prove that EL quantiles based on all the samples are more efficient than empirical quantiles based on individual samples. A simulation study shows that the EL quantiles and their functions have superior performance when the density ratio model assumption is satisfied and when it is mildly violated. An example is used to demonstrate the new method and the potential cost savings.

1. Introduction. Forestry plays a major role in the Canadian economy; maintaining the high quality of wood products is vital economically and socially. We are designing an effective long-term monitoring plan for the quality of forestry products in Canada. Two important quality indices for a piece of lumber are the modulus of elasticity (MOE) and the modulus of rupture (MOR), its strength in terms of elasticity and toughness. The reliability of lumber-based structures may depend heavily on the lower population quantiles of these indices. However, it is costly, time consuming and laborious to obtain these quality measurements. Therefore, efficient estimates of the population quantiles and their functions are important.

The estimation of quantiles based on a single random sample is a well-researched topic. Empirical quantiles have been shown to admit a Bahadur representation [Bahadur (1966); Kiefer (1967); Serfling (1980)], making it simple to study the joint limiting distributions of any number of sample quantiles and their smooth functions. In the presence of auxiliary information, the empirical likelihood [EL; Owen (1988, 2001)] can be utilized to improve efficiency. The Bahadur

Received August 2012; revised April 2013.

¹Supported in part by a Collaborative Research and Development grant from Natural Science and Engineering Research Council of Canada and in-kind contributions from Natural Resources Canada (Canadian Forest Service) through FPIInnovation's national research program.

²Supported by the NNSF of China, 11001083 and 11101156.

MSC2010 subject classifications. Primary 62G20; secondary 62G15.

Key words and phrases. Asymptotic efficiency, Bahadur representation, confidence interval, empirical likelihood.

representation of EL estimators has been established by [Chen and Chen \(2000\)](#). There is also an abundant literature on the Bahadur representation when the samples are not independent or have a time-series structure. See [Wu \(2005\)](#) and [Zhou and Wu \(2009\)](#) for recent examples.

In the targeted application, we potentially have a number of random samples from similar populations, and the combined sample size is large. Even if the size of each random sample is small, the total sample size increases over time. We may also have samples from similar products, such as lumber of various shapes and lengths. If the population distributions have some common features, the pooled information may improve the efficiency of each quantile estimate.

Specifically, we study quantile estimators based on the density ratio model (DRM) of [Anderson \(1979\)](#) and the EL approach, and we focus on investigating their properties. Suppose we have $m + 1$ independent random samples from populations with cumulative distribution and density functions denoted $G_k(x)$ and $g_k(x)$, $k = 0, 1, \dots, m$. The DRM postulates that

$$(1) \quad \log\{g_k(x)/g_0(x)\} = \boldsymbol{\theta}_k^\top \mathbf{q}(x)$$

for some known function $\mathbf{q}(x)$ of dimension d and corresponding unknown vector-valued parameters $\boldsymbol{\theta}_k$. We require the first element of $\mathbf{q}(x)$ to be one so that the first element of $\boldsymbol{\theta}_k$ is a normalization parameter.

In this formulation, the form of $g_0(x)$ is unspecified. Many parametric distribution families including normal and Gamma are special cases of the DRM. [Qin and Zhang \(1997\)](#) showed that the logistic regression model commonly used in case-control studies can be described by the DRM. They studied the EL approach for parameter estimation and for goodness-of-fit tests of the regression model. [Zhang \(2000, 2002\)](#) investigated the EL approach for quantile estimation and goodness-of-fit. [Fokianos et al. \(2001\)](#) used the EL approach under the DRM for a classical one-way analysis-of-variance.

We focus on DRM-based quantile estimation and study its Bahadur representation. We show that the EL quantiles are more efficient than empirical quantiles. The representation is then used to construct confidence intervals for the quantiles and their functions. These results are particularly relevant for the design of a long-term monitoring system for wood products. The finite-sample performance of the new methods is superior to that of the empirical quantiles when the DRM model assumption is valid and when it is mildly violated.

In Section 2, we review the EL approach under the DRM. Section 3 derives the Bahadur representation. In Section 4, we study the asymptotic properties of the new quantile estimates. The finite-sample performance is examined in Section 5, and the proposed quantile estimation is illustrated using lumber data in Section 6. The proofs are given in the [Appendix](#).

2. Empirical likelihood under DRM. Empirical likelihood under the DRM can be found in Qin and Zhang (1997) or Fokianos et al. (2001). Suppose the population distributions, $G_k(x)$, of $m + 1$ random samples of sizes $n_k: \{(x_{kj}: j = 1, \dots, n_k); k = 0, 1, 2, \dots, m\}$ satisfy the DRM (1). The model assumption may also be written

$$dG_k(x) = \exp\{\boldsymbol{\theta}_k^\tau \mathbf{q}(x)\} dG_0(x).$$

If G is discrete, then $dG(x) = G(x) - G(x_-) = P(X = x)$ for the corresponding random variable X . The EL is defined as if the G 's are discrete. Let $p_{kj} = dG_0(x_{kj})$ for all k, j . The EL is defined as

$$L_n(G_0, G_1, \dots, G_m) = \prod_{k,j} dG_k(x_{kj}) = \left\{ \prod_{k,j} p_{kj} \right\} \times \exp\left\{ \sum_{k,j} \boldsymbol{\theta}_k^\tau \mathbf{q}(x_{kj}) \right\},$$

where the product and summation with respect to $\{k, j\}$ are over the full range: $k = 0, \dots, m$ and $j = 1, \dots, n_k$. We set $\boldsymbol{\theta}_0 = 0$ for notational simplicity.

The DRM assumption implies that L_n is also a function of the parameter vector $\boldsymbol{\theta}^\tau = (\boldsymbol{\theta}_1^\tau, \dots, \boldsymbol{\theta}_m^\tau)$ and G_0 . Hence, we may also write its logarithm as

$$\ell_n(\boldsymbol{\theta}, G_0) = \sum_{k,j} \log(p_{kj}) + \sum_{k,j} \boldsymbol{\theta}_k^\tau \mathbf{q}(x_{kj}).$$

The model assumption also implies that, for $r = 0, 1, \dots, m$,

$$(2) \quad \int \exp\{\boldsymbol{\theta}_r^\tau \mathbf{q}(x)\} dG_0(x) = 1.$$

Thus, for any r between 0 and m , $\sum_{k,j} p_{kj} \exp\{\boldsymbol{\theta}_r^\tau \mathbf{q}(x_{kj})\} = 1$, which is naturally accommodated in the EL approach.

Inference on $\boldsymbol{\theta}$ and other aspects of the population distributions is usually carried out by first profiling the EL with respect to $\boldsymbol{\theta}$. That is, we define $\tilde{\ell}_n(\boldsymbol{\theta}) = \max_{G_0} \ell_n(\boldsymbol{\theta}, G_0)$ subject to constraints (2) on G_0 . Technically, we confine the support of G_0 to $\{x_{kj}\}$. The maximum in G_0 is attained when

$$p_{kj} = n^{-1} \left\{ 1 + \sum_{s=1}^m v_s [\exp\{\boldsymbol{\theta}_s^\tau \mathbf{q}(x_{kj})\} - 1] \right\}^{-1},$$

where (v_1, v_2, \dots, v_m) is the solution to $\sum_{k,j} p_{kj} \exp\{\boldsymbol{\theta}_r^\tau \mathbf{q}(x_{kj})\} = 1$ for $r = 1, 2, \dots, m$ and $n = \sum_{k=0}^m n_k$ is the total sample size. The profile log-EL (up to an additive constant) is given by

$$(3) \quad \tilde{\ell}_n(\boldsymbol{\theta}) = - \sum_{k,j} \log \left\{ 1 + \sum_{s=1}^m v_s [\exp\{\boldsymbol{\theta}_s^\tau \mathbf{q}(x_{kj})\} - 1] \right\} + \sum_{k,j} \boldsymbol{\theta}_k^\tau \mathbf{q}(x_{kj}).$$

We may regard $\tilde{\ell}_n(\boldsymbol{\theta})$ as a parametric likelihood and activate the classical likelihood-based statistical inference. This profile likelihood has the same maximum value and point as another function,

$$(4) \quad \ell_n(\boldsymbol{\theta}) = - \sum_{k,j} \log \left[\sum_{r=0}^m \rho_r \exp\{\boldsymbol{\theta}_r^\tau \mathbf{q}(x_{kj})\} \right] + \sum_{k,j} \boldsymbol{\theta}_k^\tau \mathbf{q}(x_{kj}),$$

with $\rho_r = n_r/n$. Because of its simplicity, the literature often regards $\ell_n(\boldsymbol{\theta})$ instead of $\tilde{\ell}_n(\boldsymbol{\theta})$ as the profile likelihood function of $\boldsymbol{\theta}$. In the two-sample situation, Keziou and Leoni-Aubin (2008) found that it is a “dual likelihood.” Its likelihood ratio statistics remain asymptotically chi-square.

We define the maximum EL estimator (MELE) $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ as the maximum point of (3) or equivalently of (4). The asymptotic normality of $\hat{\boldsymbol{\theta}}$ has been established in various situations [Qin and Zhang (1997); Fokianos et al. (2001); Zhang (2002)]. We summarize and extend these results, giving the necessary details as a preparational step. Let $h(x; \boldsymbol{\theta}) = \sum_{k=0}^m \rho_k \exp\{\boldsymbol{\theta}_k^\tau \mathbf{q}(x)\}$, and for $k = 0, 1, \dots, m$, we define

$$h_k(x; \boldsymbol{\theta}) = \rho_k \exp\{\boldsymbol{\theta}_k^\tau \mathbf{q}(x)\} / h(x; \boldsymbol{\theta}).$$

Let $\mathbf{h}(x; \boldsymbol{\theta}) = \{h_0(x; \boldsymbol{\theta}), \dots, h_m(x; \boldsymbol{\theta})\}^\tau$ and define an $(m + 1) \times (m + 1)$ matrix

$$\mathbf{H}(x; \boldsymbol{\theta}) = \text{diag}\{\mathbf{h}(x; \boldsymbol{\theta})\} - \mathbf{h}(x; \boldsymbol{\theta})\mathbf{h}^\tau(x; \boldsymbol{\theta}).$$

When $\boldsymbol{\theta} = \boldsymbol{\theta}^*$, the true value of $\boldsymbol{\theta}$, we may drop $\boldsymbol{\theta}^*$ for notational simplicity. Finally, we use $d\bar{G}(x)$ for $h(x; \boldsymbol{\theta}^*) dG_0(x)$ in the integrations.

THEOREM 2.1. *Suppose we have an independent random sample $\{x_{kj}\}_{j=1}^{n_k}$ from population G_k for $k = 0, 1, \dots, m$. The total sample size $n = \sum_k n_k \rightarrow \infty$, and $\rho_k = n_k/n$ remains a constant (or within the n^{-1} range).*

The population distributions G_k satisfy the DRM (1) with true parameter value $\boldsymbol{\theta}^$ and $\int h(x; \boldsymbol{\theta}) dG_0 < \infty$ in a neighborhood of $\boldsymbol{\theta}^*$. The components of $\mathbf{q}(x)$ are linearly independent and its first element is one.*

Then $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^)$ is asymptotically multivariate normal with mean $\mathbf{0}$ and covariance matrix $\mathbf{W}^{-1} - \mathbf{S}$. Both \mathbf{W} and \mathbf{S} are $md \times md$ block matrices with each block a $d \times d$ matrix, and their (r, s) th blocks are, respectively,*

$$\begin{aligned} \mathbf{W}_{rs} &= \int \mathbf{q}(x)\mathbf{q}^\tau(x) \{h_r(x)\delta_{rs} - h_r(x)h_s(x)\} d\bar{G}(x), \\ \mathbf{S}_{rs} &= (\rho_r^{-1}\delta_{rs} + \rho_0^{-1}) \text{diag}\{1, 0, \dots, 0\}, \end{aligned}$$

where $1 \leq r, s \leq m$ and $\delta_{rs} = 1$ if $r = s$ and 0 otherwise.

The Appendix contains a sketched proof to bridge some gaps. Using the Kronecker product \otimes , we have a tighter expression,

$$(5) \quad \mathbf{W} = \int \mathbf{H}[-1, -1](x) \otimes \{\mathbf{q}(x)\mathbf{q}^\tau(x)\} d\bar{G}(x),$$

where $\mathbf{H}[-1, -1](x)$ is $\mathbf{H}(x)$ with its first row and column removed. This convention is adopted from the statistical software package [R Development Core Team (2011)].

The asymptotic normality is a stepping stone for our main result on the Bahadur representation. It also reveals that the MELE is root- n consistent. The assumption that $\int h(x; \theta) dG_0 < \infty$ in a neighborhood of θ^* implies the existence of the moment generating function of $\mathbf{q}(x)$ and therefore all its finite moments. This fact will be used in our proofs.

3. Bahadur representation and its applications. Given the MELE $\hat{\theta}$, the fitted values of p_{kj} are $\hat{p}_{kj} = \{nh(x_{kj}; \hat{\theta})\}^{-1}$ and the fitted G_r is

$$\begin{aligned} \hat{G}_r(x) &= \sum_{k,j} \hat{p}_{kj} \exp\{\hat{\theta}_r^T \mathbf{q}(x_{kj})\} I(x_{kj} \leq x) \\ &= n_r^{-1} \sum_{k,j} h_r(x_{kj}; \hat{\theta}) I(x_{kj} \leq x) \end{aligned}$$

with $\hat{\theta}_0 = 0$ and $I(A)$ an indicator function of event A . For any $\alpha \in (0, 1)$, define the α -quantile of G_r as $\xi_r = \xi_{r,\alpha} = \inf\{x : G_r(x) \geq \alpha\}$ and its EL-based estimator as

$$(6) \quad \hat{\xi}_r = \hat{\xi}_{r,\alpha} = \inf\{x : \hat{G}_r(x) \geq \alpha\}.$$

We call the $\hat{\xi}_r$ values EL quantiles for simplicity. The asymptotic normality of the EL quantile $\hat{\xi}_r$ is useful for constructing confidence intervals for ξ_r or for testing related hypotheses. Researchers are often interested in smooth functions of quantiles of many populations and/or at several levels. Thus, the multivariate asymptotic behavior is useful, and this calls for the Bahadur representation.

THEOREM 3.1. *Assume the conditions of Theorem 2.1, and that the density function $g_r(x)$ is continuous and positive at $x = \xi_r$. The EL quantile (6) has Bahadur representation*

$$(7) \quad \hat{\xi}_r = \xi_r + \{\alpha - \hat{G}_r(\xi_r)\} / g_r(\xi_r) + O_p(n^{-3/4} \{\log(n)\}^{1/2}),$$

where ξ_r is the α th quantile of $G_r(x)$.

The proof is given in the [Appendix](#). Without the Bahadur representation, it is a daunting task to derive the limiting distribution of functions of EL quantiles such as $\hat{\xi}_{0,\alpha} - \hat{\xi}_{1,\alpha}$. Theorem 3.1 links this task to that of $\{a\hat{G}_0(\xi_{0,\alpha}) - b\hat{G}_1(\xi_{1,\alpha})\}$ with nonrandom constants a and b . The asymptotic properties of \hat{G}_r are simple and easy to use.

THEOREM 3.2. *Assume the same conditions as in Theorem 2.1. For any $0 \leq r_1, r_2, \dots, r_k \leq m$ and an accompanying set of real numbers x_1, x_2, \dots, x_k in the support of $G_0(x)$, $\sqrt{n}\{\hat{G}_{r_j}(x_j) - G_{r_j}(x_j)\}$ are jointly asymptotically k -variate normal with mean $\mathbf{0}$ and covariance matrix $\Omega_{EL} = (\omega_{r_i, r_j}(x_i, x_j))_{1 \leq i, j \leq k}$. The generic form of $\omega_{r_i, r_j}(x_i, x_j)$ is given by*

$$(8) \quad \omega_{rs}(x, y) = \sigma_{rs}(x, y) - (\rho_r \rho_s)^{-1} \{a_{rs}(x \wedge y) - \mathbf{B}_r^T(x) \mathbf{W}^{-1} \mathbf{B}_s(y)\},$$

where $x \wedge y = \min\{x, y\}$,

$$\sigma_{rs}(x, y) = \rho_r^{-1} \delta_{rs} \{G_r(x \wedge y) - G_r(x)G_s(y)\},$$

$$a_{rs}(x) = \int_{-\infty}^x \{\delta_{rs} h_r(t) - h_r(t)h_s(t)\} d\bar{G}(t)$$

and $\mathbf{B}_r(x)$ is a vector of length md with its s th segment (of length d with $s = 1, 2, \dots, m$) being

$$\mathbf{B}_{r,s}(x) = \int_{-\infty}^x \{\delta_{rs} h_r(t) - h_r(t)h_s(t)\} \mathbf{q}(t) d\bar{G}(t).$$

The proof is given in the [Appendix](#). The Bahadur representation (7) and the multivariate asymptotic normality of the \hat{G} 's lead to multivariate asymptotic normality of the EL quantiles. For notational simplicity, we will state the result only for the bivariate case. Let ξ_r be the population quantile at some level α_r of the r th population in the DRM. We similarly define ξ_s at some level α_s . The exact levels α_r and α_s are not important.

THEOREM 3.3. *Assume that the conditions in Theorem 3.1 hold for ξ_r and ξ_s . The centralized EL quantile under the DRM assumption*

$$\sqrt{n}(\hat{\xi}_r - \xi_r, \hat{\xi}_s - \xi_s)$$

is asymptotically bivariate normal with mean $\mathbf{0}$ and covariance matrix

$$(9) \quad \Sigma_{EL} = \begin{pmatrix} \omega_{rr}(\xi_r, \xi_r)/g_r^2(\xi_r) & \omega_{rs}(\xi_r, \xi_s)/\{g_r(\xi_r)g_s(\xi_s)\} \\ \omega_{rs}(\xi_r, \xi_s)/\{g_r(\xi_r)g_s(\xi_s)\} & \omega_{ss}(\xi_s, \xi_s)/g_s^2(\xi_s) \end{pmatrix}.$$

The above result does not restrict the selection of the two populations or the levels of the quantiles. It can be used to conveniently obtain the limiting distributions of smooth functions of the EL quantiles.

4. Efficiency comparison. The EL quantiles are constructed by pooling information from $m + 1$ independent random samples. We trust that they are more efficient than empirical quantiles (hereafter EM) based on single samples. A rigorous proof of this intuitive claim is not simple.

Let $\tilde{G}_r(x)$ be the empirical distribution function based solely on the r th sample. As processes indexed by x , $\sqrt{n}\{\tilde{G}_r(x) - G_r(x)\}$, $r = 0, 1, \dots, m$, are independent

and each converges in distribution to a Gaussian process with covariance function $\sigma_{rr}(x, x)$. Let $\check{\xi}_r$ be the EM quantiles of the r th population at level α_r . Based on the classical Bahadur presentation, with any number of choices in r and α_r , $\{\sqrt{n}(\check{\xi}_r - \xi_r)\}$ are jointly asymptotically multivariate normal with mean 0. In the bivariate case, the asymptotic covariance matrix of $\sqrt{n}(\check{\xi}_r - \xi_r, \check{\xi}_s - \xi_s)$ is given by

$$\Sigma_{EM} = \begin{pmatrix} \sigma_{rr}(\xi_r, \xi_r)/\{g_r^2(\xi_r)\} & \sigma_{rs}(\xi_r, \xi_s)/\{g_r(\xi_r)g_s(\xi_s)\} \\ \sigma_{sr}(\xi_s, \xi_r)/\{g_r(\xi_r)g_s(\xi_s)\} & \sigma_{ss}(\xi_s, \xi_s)/\{g_s^2(\xi_s)\} \end{pmatrix},$$

where $\sigma_{rs}(x, y)$ was given in Theorem 3.2.

Since the EL and EM quantiles are asymptotically unbiased, the efficiency comparison reduces to a comparison of two asymptotic covariance matrices. The following result generalizes Corollary 4.3 of Zhang (2000).

THEOREM 4.1. *For any pair of integers $0 \leq r, s \leq m$ and any quantile levels α_r and α_s , we have $\Sigma_{EM} - \Sigma_{EL} \geq 0$. This conclusion remains true for any number of quantiles.*

5. Inferences on functions of quantiles. In applications such as the wood project, we are interested in the size of ξ_r , $\xi_r - \xi_s$, etc. for various choices of r and s and various levels. Two scenarios are of particular interest. (A) For a specific wood product in a given year, is its quality index above or below the industrial standard? (B) How different are the quality indices for wood products produced in two different years, mills or regions?

(A) and (B) can be addressed through hypothesis tests or the construction of confidence intervals. With the asymptotic normality and favorable efficiency properties of the EL quantiles, the task is simple. We must find a consistent estimate of $\text{var}(\hat{\xi}_r)$ and construct approximate $100(1 - \alpha)\%$ confidence intervals as $\hat{\xi}_r \pm z_{\alpha/2}\sqrt{\hat{\text{var}}(\hat{\xi}_r)}$, where $z_{\alpha/2}$ denotes the $(1 - \alpha/2)$ th quantile of the standard normal distribution. Similarly, approximate confidence intervals for $\xi_r - \xi_s$ are $(\hat{\xi}_r - \hat{\xi}_s) \pm z_{\alpha/2}\sqrt{\hat{\text{var}}(\hat{\xi}_r - \hat{\xi}_s)}$. In both cases, we need effective and consistent estimates of $\text{var}(\hat{\xi}_r)$ and $\text{var}(\hat{\xi}_r - \hat{\xi}_s)$.

With the help of (9), plug-in consistent variance estimators can easily be constructed. Two necessary ingredients are consistent estimations of $\omega_{rs}(x, y)$ and $g_r(x)$. Although $\hat{G}_r(x)$ is discrete, the idea of kernel density estimation can be used to produce a density estimate. Let $K(\cdot) \geq 0$ be a commonly used kernel function such that $\int K(x) dx = 1$, $\int x K(x) dx = 0$ and $\int x^2 K(x) dx < \infty$. For some bandwidth $b > 0$, let $K_b(x) = (1/b)K(x/b)$. Then a kernel estimate of $g_r(x)$ is given by

$$\hat{g}_r(x) = \int K_b(x - y) d\hat{G}_r(y).$$

In the simulation study, we set $K(x)$ to the standard normal density function. We chose the bandwidth b according to the rule of thumb of Deheuvels (1977) and Silverman (1986),

$$b = 1.06n^{-1/5} \min\{\hat{\sigma}, \hat{R}/1.34\}.$$

The above formula is designed for the situation where the density function is estimated based on n independent and identically distributed observations. In our simulation, we regard the fitted $\hat{G}_r(x)$ as a nonrandom distribution function, and compute the standard deviation and inter-quartile range of this distribution as $\hat{\sigma}$ and \hat{R} .

The analytical form of $\omega_{rs}(x, y)$ contains many terms, but it is straightforward to estimate them consistently and sensibly. Let

$$\hat{a}_{rs}(x) = \int_{-\infty}^x \{\delta_{rs}h_r(t; \hat{\theta}) - h_r(t; \hat{\theta})h_s(x; \hat{\theta})\}h(t; \hat{\theta}) d\hat{G}_0(t),$$

and we form $\hat{\mathbf{B}}_r(x)$ and $\hat{\mathbf{W}}$ via

$$\begin{aligned} \hat{\mathbf{B}}_{r,s}(x) &= \int_{-\infty}^x \{\delta_{rs}h_r(t; \hat{\theta}) - h_r(t; \hat{\theta})h_s(x; \hat{\theta})\} \mathbf{q}(t)h(t; \hat{\theta}) d\hat{G}_0(t), \\ \hat{\mathbf{W}}_{rs} &= \int_{-\infty}^{\infty} \{\delta_{rs}h_r(t; \hat{\theta}) - h_r(t; \hat{\theta})h_s(x; \hat{\theta})\} \mathbf{q}(t)\mathbf{q}^{\tau}(t)h(t; \hat{\theta}) d\hat{G}_0(t). \end{aligned}$$

We then form a consistent estimator of $\omega_{rs}(\xi_r, \xi_s)$ as

$$\hat{\omega}_{rs}(\xi_r, \xi_s) = \rho_r^{-1}\delta_{rs}(\alpha_r - \alpha_r\alpha_s) - (\rho_r\rho_s)^{-1}\{\hat{a}_{rs}(\hat{\xi}_r \wedge \hat{\xi}_s) - \hat{\mathbf{B}}_r^{\tau}(\hat{\xi}_r)\hat{\mathbf{W}}^{-1}\hat{\mathbf{B}}_s(\hat{\xi}_s)\},$$

where we have used the facts that $\delta_{rs}G_r(\xi_r \wedge \xi_s) = \delta_{rs}\alpha_r$ and $G_r(\xi_r) = \alpha_r$.

6. Simulation study. We now examine the finite-sample performance of the inference procedures via simulation. Are the EL quantiles $\hat{\xi}$ more efficient than the EM quantiles $\check{\xi}$? The simulation studies shed light on how large the sample must be before the asymptotic result applies. We analyze data sets generated from several sets of populations, which are divided into two groups: those that satisfy the DRM assumption and those that do not.

6.1. Populations satisfying DRM assumptions. Recall that the Gamma and normal distribution families are special DRMs. We choose two sets of distributions from these families with the parameter values specified in Table 1. For the Gamma distributions, the first parameter is the degrees of freedom and the second is the scale. Therefore, the expectation of the first population is 6×1.5 . The parameters for the normal distribution are the mean and variance. The populations have similar means and variances to those seen in applications.

The simulations were carried out with $n_r = 50$ and 2000 repetitions. We examine the performance of $\hat{G}_r(x)$ and $\check{G}_r(x)$ for x set to the quantile levels $\alpha = 5\%$,

TABLE 1
Parameters under DRM

Distributions						$q^r(x)$
$\Gamma(6, 1.5)$	$\Gamma(6, 1.4)$	$\Gamma(7, 1.3)$	$\Gamma(7, 1.2)$	$\Gamma(8, 1.1)$	$\Gamma(8, 1.0)$	$(1, x, \log(x))$
$N(18, 4)$	$N(18, 9)$	$N(20, 6)$	$N(20, 9)$	$N(22, 8)$	$N(22, 10)$	$(1, x, x^2)$

10%, 50%, 90% and 95%. We computed the relative bias, the asymptotic variance and the simulated variance of the EL estimator $\hat{G}_r(x)$. The EM estimator $\check{G}_r(x)$ is unbiased, and its asymptotic variance is $\alpha(1 - \alpha)/\rho_r$. For ease of comparison, we report the ratios of the EM and EL asymptotic variances and the ratios of their simulated variances. We also report the ratios of the mean estimated variances of $\hat{G}_r(x)$ and their corresponding asymptotic variances. The results are presented in Table 2.

There is an efficiency gain in the range of 40% to 70% for the EL estimators in terms of both the theoretical and simulated variances. The variances of the EL estimators are estimated accurately: in the $\hat{\sigma}^2(\hat{G})/\sigma^2(\hat{G})$ column all the entries are close to 1. Finally, the relative biases $B(\hat{G})$ and $B(\check{G})$ are both small.

We now turn to investigating the performance of the EL and EM quantiles for both point and interval estimations. The quantile of a discrete distribution $G(x)$ is not a smooth function, and this puts the EM quantile at a disadvantage. To ensure that the EL quantile had a strong competitor, we modified the EM quantile. We replaced $\check{G}(x)$ by $\check{G}_r(x) - (2n_r)^{-1}$, and we used linear interpolation to calculate this quantile. These modifications do not alter the first-order asymptotics. We continue to use the notation $\hat{\xi}_r$ and $\check{\xi}_r$ for the EL and EM quantiles after these modifications.

The simulation results for the quantile estimates are given in Table 3. The simulated EL variances $\text{var}(\hat{\xi})$ and the mean estimated EL variances $\hat{\sigma}^2(\hat{\xi})$ are both close to the asymptotic variances $\sigma^2(\hat{\xi})$. The results support the asymptotic theory and the viability of the EL variance estimator. The ratio $\text{var}(\hat{\xi})/\sigma^2(\hat{\xi})$ is based on simulated variances and ranges between 1.20 and 1.60. These results indicate an efficiency gain of between 20% and 60% in the EL quantiles. Finally, the relative biases $B(\hat{\xi})$ and $B(\check{\xi})$ are low and within $\pm 3\%$.

The simulation results for the interval estimates of the quantiles and quantile differences are given in Table 4. We compute the average lengths and coverage probabilities of the EL and EM confidence intervals at the 95% level. The coverage probabilities of the EL intervals are almost always closer to the nominal 95%. This advantage is more obvious for the upper-tail quantiles (such as the 95% quantile). Often, the coverage gains of the EL intervals reach 5%, and these intervals are 10% to 20% shorter.

We also conducted simulations for the second group of populations, as shown in Table 1, and for $n_r = 100$. The results are similar and omitted. In conclusion, the EL approach is superior when the model assumptions are satisfied.

TABLE 2
EL and EM distribution estimates. \hat{G} : EL estimate; \check{G} : EM estimate; σ^2 : asymptotic variance; $\hat{\sigma}^2$: average of variance estimate; B: bias of $\sqrt{\alpha(1-\alpha)}$ as percentage; $n_k = 50$

	α	$\sigma^2(\hat{G})$	$\sigma^2(\check{G})/\sigma^2(\hat{G})$	$\text{var}(\check{G})/\text{var}(\hat{G})$	$\hat{\sigma}^2(\hat{G})/\text{var}(\hat{G})$	$B(\hat{G})$	$B(\check{G})$
$\Gamma(6, 1.5)$	0.05	0.176	1.62	1.48	1.03	-0.46	-4.02
	0.10	0.378	1.43	1.38	1.00	-0.28	-3.24
	0.50	1.055	1.42	1.33	0.98	-0.65	-1.92
	0.90	0.382	1.41	1.38	1.01	-0.83	-3.68
	0.95	0.178	1.60	1.58	1.12	-0.45	-4.67
$\Gamma(6, 1.4)$	0.05	0.176	1.62	1.49	1.08	-1.15	-4.37
	0.10	0.374	1.44	1.39	1.03	-0.71	-3.33
	0.50	1.031	1.45	1.37	1.00	-0.31	-2.03
	0.90	0.370	1.46	1.39	1.04	-0.68	-3.20
	0.95	0.172	1.66	1.66	1.13	-0.74	-4.39
$\Gamma(7, 1.3)$	0.05	0.170	1.67	1.50	1.00	-0.27	-4.18
	0.10	0.368	1.47	1.49	1.04	-0.20	-2.74
	0.50	1.034	1.45	1.44	0.98	-0.29	-1.74
	0.90	0.394	1.37	1.39	1.01	-0.24	-3.12
	0.95	0.186	1.53	1.60	1.07	-0.29	-4.33
$\Gamma(7, 1.2)$	0.05	0.171	1.67	1.56	1.06	-0.86	-4.14
	0.10	0.369	1.46	1.39	1.05	-0.78	-3.38
	0.50	1.065	1.41	1.32	0.97	-0.51	-2.33
	0.90	0.371	1.46	1.42	0.99	-0.81	-3.57
	0.95	0.169	1.68	1.60	0.99	-1.13	-4.68
$\Gamma(8, 1.1)$	0.05	0.172	1.65	1.54	1.06	-0.72	-4.26
	0.10	0.372	1.45	1.44	1.05	-0.67	-3.27
	0.50	1.055	1.42	1.43	1.01	-0.25	-1.53
	0.90	0.373	1.45	1.47	1.01	-0.19	-3.56
	0.95	0.176	1.62	1.68	1.08	-0.37	-4.15
$\Gamma(8, 1.0)$	0.05	0.180	1.59	1.55	1.07	-1.04	-4.22
	0.10	0.379	1.42	1.41	0.98	-0.81	-3.44
	0.50	1.041	1.44	1.33	0.95	0.28	-1.63
	0.90	0.364	1.48	1.52	1.09	0.01	-2.99
	0.95	0.164	1.73	1.64	1.09	-0.64	-4.39

6.2. *Performance when model is misspecified.* What happens to the EL approach when the model is misspecified? Fokianos and Kaimi (2006) quantified the effect of choosing an incorrect linear form of $\mathbf{q}(x)$. In general, both the point estimation and the hypothesis tests are adversely affected when the model is misspecified. These findings may have motivated the model selection approach in Fokianos (2007). That is, instead of pre-specifying a known $\mathbf{q}(x)$, one may select $\mathbf{q}(x)$ as a linear combination of a rich class of functions. For instance, let $\mathbf{q}(x) = \{1, \log x, x^{0.5}, x, x^{1.5}, x^2\}^\tau$. The most appropriate $\mathbf{q}(x)$ is then determined

TABLE 3

EL and EM quantiles. ξ : true quantile; $\hat{\xi}$: EL quantile; $\check{\xi}$: EM quantile; σ^2 : asymptotic variance; $\hat{\sigma}^2$: average of variance estimate; B : relative bias as percentage; $n_k = 50$

	α	ξ	$\sigma^2(\hat{\xi})$	$\text{var}(\hat{\xi})$	$\hat{\sigma}^2(\hat{\xi})$	$\text{var}(\check{\xi})/\text{var}(\hat{\xi})$	$B(\hat{\xi})$	$B(\check{\xi})$
$\Gamma(6, 1.5)$	0.05	3.92	71.31	70.38	67.50	1.51	2.19	1.59
	0.10	4.73	69.23	68.28	73.79	1.34	1.08	0.97
	0.50	8.51	83.70	88.50	94.46	1.27	0.45	0.15
	0.90	13.91	298.78	292.15	287.19	1.35	0.23	0.19
	0.95	15.77	473.53	468.31	469.48	1.54	-0.25	-0.32
$\Gamma(6, 1.4)$	0.05	3.66	62.34	62.42	59.15	1.55	2.84	1.94
	0.10	4.41	59.63	60.02	64.02	1.31	1.55	0.88
	0.50	7.94	71.30	75.64	81.58	1.39	0.27	0.28
	0.90	12.98	252.32	248.04	248.28	1.37	0.11	-0.05
	0.95	14.72	397.73	387.44	399.88	1.51	-0.05	-0.41
$\Gamma(7, 1.3)$	0.05	4.27	67.35	66.55	62.81	1.59	1.83	1.07
	0.10	5.06	63.93	64.70	68.40	1.33	0.88	0.50
	0.50	8.67	72.61	76.18	82.87	1.37	0.30	0.04
	0.90	13.69	260.55	254.74	241.91	1.30	-0.08	-0.12
	0.95	15.40	413.83	414.73	395.77	1.52	-0.22	-0.11
$\Gamma(7, 1.2)$	0.05	3.94	57.63	59.87	54.10	1.54	2.32	1.11
	0.10	4.67	54.66	56.15	58.62	1.36	1.32	0.75
	0.50	8.00	63.76	65.99	70.68	1.32	0.30	0.23
	0.90	12.64	208.95	229.25	206.39	1.32	0.24	0.12
	0.95	14.21	320.80	340.92	319.98	1.61	0.13	-0.13
$\Gamma(8, 1.1)$	0.05	4.38	60.62	56.57	54.22	1.66	1.95	1.07
	0.10	5.12	56.23	54.25	58.29	1.33	1.12	0.63
	0.50	8.44	61.08	60.83	67.63	1.40	0.23	-0.10
	0.90	12.95	195.71	198.99	188.42	1.36	-0.15	-0.03
	0.95	14.46	309.32	323.05	295.76	1.47	-0.24	-0.55
$\Gamma(8, 1.0)$	0.05	3.98	52.12	50.42	44.41	1.55	2.48	1.22
	0.10	4.66	47.38	47.27	47.85	1.31	1.35	0.88
	0.50	7.67	49.81	53.38	55.67	1.35	-0.02	0.00
	0.90	11.77	157.93	162.29	155.97	1.33	-0.22	-0.30
	0.95	13.15	238.30	247.86	242.16	1.55	-0.14	-0.32

by selecting a subvector of the current $\mathbf{q}(x)$. Hence, the classical model selection approaches can be used.

Following this lead, we provide a limited study of the impact of misspecification on the quantile estimations. For this purpose, we simulated random samples from a number of Gamma distributions, Weibull distributions, denoted $W(\cdot, \cdot)$, and normal distributions, as shown in Table 5. These populations are chosen to have similar means and variances. We obtained the EL quantile estimates as if they satisfy DRM for some pre-specified but wrong $\mathbf{q}(x)$.

TABLE 4
Confidence intervals for quantile and quantile differences. Nominal level: 95%;
 $\Gamma(6, 1.5) - \Gamma(6, 1.4)$: *differences of $\Gamma(6, 1.5)$ and $\Gamma(6, 1.4)$ quantiles at the given level; $n_k = 50$*

	α	EL					EM				
		5%	10%	50%	90%	95%	5%	10%	50%	90%	95%
$\Gamma(6, 1.5)$	length	1.83	1.93	2.18	3.77	4.73	2.33	2.42	2.66	4.23	5.20
	coverage	93.0	94.8	95.2	90.7	89.5	92.3	96.5	95.9	89.1	84.9
$\Gamma(6, 1.4)$	length	1.71	1.79	2.03	3.49	4.36	2.16	2.24	2.47	3.98	4.87
	coverage	91.8	94.3	95.0	90.9	88.7	91.7	95.0	96.3	89.0	84.3
$\Gamma(7, 1.3)$	length	1.76	1.85	2.03	3.41	4.27	2.25	2.31	2.48	3.88	4.78
	coverage	92.0	93.7	95.5	92.2	90.1	90.0	95.0	96.8	90.1	85.1
$\Gamma(7, 1.2)$	length	1.63	1.71	1.88	3.16	3.92	2.07	2.13	2.31	3.62	4.45
	coverage	91.9	94.3	95.3	92.5	91.3	91.3	95.5	97.4	89.4	85.1
$\Gamma(8, 1.1)$	length	1.64	1.71	1.84	3.02	3.73	2.09	2.12	2.26	3.50	4.24
	coverage	91.5	93.9	95.4	92.0	90.1	91.5	95.0	95.9	90.6	85.7
$\Gamma(8, 1.0)$	length	1.50	1.56	1.67	2.74	3.36	1.91	1.93	2.05	3.15	3.82
	coverage	90.1	93.4	95.3	92.5	91.7	89.3	94.9	96.6	90.0	85.1
$\Gamma(6, 1.5) - \Gamma(6, 1.4)$	length	2.41	2.58	2.92	5.08	6.33	3.23	3.32	3.65	5.95	7.38
	coverage	95.0	95.5	95.5	93.8	94.5	94.5	97.3	96.0	91.8	89.5
$\Gamma(6, 1.5) - \Gamma(7, 1.3)$	length	2.47	2.62	2.91	5.00	6.22	3.29	3.37	3.66	5.87	7.33
	coverage	95.1	96.1	95.3	94.8	95.1	94.1	96.8	96.2	93.2	89.9
$\Gamma(6, 1.5) - \Gamma(7, 1.2)$	length	2.37	2.53	2.82	4.89	6.09	3.17	3.25	3.54	5.69	7.09
	coverage	93.9	95.5	95.8	94.1	93.7	94.2	96.2	96.5	93.4	90.2
$\Gamma(6, 1.5) - \Gamma(8, 1.1)$	length	2.40	2.54	2.79	4.78	5.95	3.18	3.25	3.51	5.61	6.95
	coverage	94.2	95.8	96.3	94.7	93.7	94.4	96.6	96.1	93.2	90.0
$\Gamma(6, 1.5) - \Gamma(8, 1.0)$	length	2.29	2.43	2.71	4.68	5.84	3.06	3.12	3.38	5.39	6.67
	coverage	93.5	95.2	95.9	92.3	91.9	94.1	96.4	96.7	92.7	89.0

As a trade-off between model interpretation and parsimony, we choose $\mathbf{q}(x) = (1, x, \log(1 + |x|), \sqrt{|x|})^\tau$. The remaining settings are the same as before. In Table 6, we report only the biases and mean square errors (mse) of the EL and EM quantiles. The EL quantiles are still uniformly more efficient with the efficiency gains usually above 15%. The variance estimators remain accurate, and the relative biases $B(\hat{\xi})$ and $B(\check{\xi})$ are still negligible.

TABLE 5
Parameters for non-DRM

$\Gamma(16, 0.6)$	$\Gamma(19, 0.5)$	$N(9, 5)$	$N(9.6, 5.6)$	$W(10, 4.5)$	$W(11, 5)$
$\Gamma(16, 0.6)$	$\Gamma(19, 0.5)$	$\Gamma(17.5, 0.5)$	$W(10.5, 4.5)$	$W(10, 4.5)$	$W(11, 5)$

TABLE 6

EL and EM quantiles under model mis-specification. ξ : true quantile; $\hat{\xi}$: EL quantile; $\check{\xi}$: EM quantile; $\hat{\sigma}^2$: average of variance estimate; B: relative bias as percentage; $n_k = 50$

	α	ξ	$mse(\hat{\xi})$	$mse(\check{\xi})/mse(\hat{\xi})$	$\hat{\sigma}(\hat{\xi})/mse(\hat{\xi})$	$B(\hat{\xi})$	$B(\check{\xi})$
$\Gamma(16, 0.6)$	0.05	6.022	61.06	1.23	0.83	1.20	0.56
	0.10	6.681	49.19	1.19	0.95	0.48	0.07
	0.50	9.401	44.14	1.16	1.13	-0.16	-0.18
	0.90	12.775	125.38	1.10	0.94	0.24	-0.04
	0.95	13.858	210.07	1.17	0.92	0.05	-0.27
$\Gamma(19, 0.5)$	0.05	6.221	57.81	1.22	0.78	1.29	0.71
	0.10	6.836	44.47	1.14	0.90	0.77	0.33
	0.50	9.334	34.15	1.20	1.17	0.07	0.10
	0.90	12.378	97.34	1.14	0.97	0.17	-0.00
	0.95	13.346	151.68	1.20	0.98	0.06	-0.40
$N(9, 5)$	0.05	5.322	103.06	1.23	0.85	1.43	0.77
	0.10	6.134	73.47	1.08	0.94	0.17	0.23
	0.50	9.000	37.86	1.20	1.14	0.21	-0.03
	0.90	11.866	68.81	1.18	0.92	-0.26	-0.17
	0.95	12.678	97.39	1.30	0.89	-0.39	-0.39
$N(9.6, 5.6)$	0.05	5.708	116.53	1.25	0.89	0.79	0.49
	0.10	6.567	87.51	1.14	0.90	0.18	0.13
	0.50	9.600	43.71	1.20	1.11	0.37	0.27
	0.90	12.633	76.92	1.22	0.96	-0.06	-0.13
	0.95	13.492	112.45	1.25	0.91	-0.26	-0.30
$W(10, 4.5)$	0.05	3.344	12.09	1.01	0.85	0.89	0.48
	0.10	3.593	6.62	1.01	0.97	0.58	0.20
	0.50	4.338	2.08	1.10	1.11	-0.09	-0.01
	0.90	4.891	1.91	1.17	1.14	-0.01	-0.11
	0.95	5.022	2.17	1.39	1.17	0.13	-0.18
$W(11, 5.0)$	0.05	3.817	11.89	1.15	0.94	0.79	0.25
	0.10	4.075	6.67	1.11	0.94	0.56	0.01
	0.50	4.836	1.99	1.16	1.14	-0.09	-0.03
	0.90	5.394	2.15	1.03	1.18	-0.11	-0.07
	0.95	5.524	2.63	1.14	1.18	0.05	-0.08

The simulation results for the interval estimates of the quantiles and quantile differences are given in Table 7. We compute the average lengths and the coverage probabilities of the EL and EM confidence intervals at the 95% level. The EL confidence intervals are not clearly better. These intervals have better coverage probabilities for the upper quantiles but similar or slightly inferior probabilities for the lower quantiles. The EL intervals are always shorter, and they are more than

TABLE 7
Confidence intervals for quantile and quantile differences under model mis-specification. Nominal level 95%; $n_k = 50$

α		EL					EM				
		5%	10%	50%	90%	95%	5%	10%	50%	90%	95%
$\Gamma(16, 0.6)$	length	1.57	1.52	1.57	2.36	2.91	1.83	1.79	1.77	2.50	3.00
	coverage	87.5	91.8	95.7	91.2	88.9	89.0	94.1	96.5	90.6	86.4
$\Gamma(19, 0.5)$	length	1.48	1.41	1.40	2.10	2.63	1.71	1.64	1.61	2.25	2.66
	coverage	87.9	90.8	95.5	91.9	89.7	88.5	93.0	96.4	91.0	85.2
$N(9, 5)$	length	2.04	1.82	1.47	1.77	2.05	2.22	2.00	1.67	2.02	2.28
	coverage	87.4	92.0	94.6	91.1	89.6	86.2	92.3	96.4	91.7	87.6
$N(9.6, 5.6)$	length	2.13	1.90	1.54	1.91	2.22	2.35	2.12	1.76	2.12	2.35
	coverage	83.5	90.6	96.1	92.4	89.3	84.4	91.3	96.8	92.3	86.6
$W(10, 4.5)$	length	0.66	0.54	0.34	0.33	0.35	0.69	0.57	0.37	0.36	0.37
	coverage	82.3	88.8	95.1	96.1	95.3	84.7	90.8	96.0	95.0	91.6
$W(11, 5.0)$	length	0.68	0.53	0.33	0.35	0.38	0.70	0.59	0.37	0.36	0.37
	coverage	83.3	87.9	93.8	93.7	92.4	83.5	90.9	94.8	94.6	89.9
$\Gamma(16, 0.6) - \Gamma(19, 0.5)$	length	2.14	2.04	2.08	3.15	3.95	2.55	2.45	2.41	3.42	4.14
	coverage	92.8	94.6	95.4	94.5	93.3	93.5	95.8	96.1	94.2	91.3
$\Gamma(16, 0.6) - N(9, 5)$	length	2.59	2.36	2.13	2.96	3.60	2.94	2.71	2.45	3.26	3.87
	coverage	92.4	94.3	95.7	93.4	91.9	91.4	95.3	96.9	94.0	90.9
$\Gamma(16, 0.6) - N(9.6, 5.6)$	length	2.65	2.41	2.17	3.01	3.66	3.05	2.81	2.51	3.33	3.92
	coverage	90.5	93.4	96.5	94.5	93.0	91.1	94.7	96.1	94.4	91.8
$\Gamma(16, 0.6) - W(10, 4.5)$	length	1.73	1.63	1.61	2.39	2.93	1.99	1.89	1.81	2.53	3.03
	coverage	90.1	93.3	95.3	91.6	88.9	91.8	94.7	96.1	91.1	86.1
$\Gamma(16, 0.6) - W(11, 5.0)$	length	1.74	1.62	1.60	2.39	2.94	1.99	1.89	1.81	2.53	3.03
	coverage	89.3	92.9	95.4	91.4	89.0	90.8	94.2	96.3	90.8	86.7

10% shorter in most cases. The simulation results for the second set of populations are similar; they are omitted to save space.

In conclusion, while the model misspecification has a serious impact on the estimation of θ , as shown by Fokianos and Kaimi (2006), the quantile estimations are not as badly affected.

7. Real-data analysis. In this section, we apply our method to lumber data. The data come from tests conducted at an FPInnovations laboratory. They contain the MOE and MOR measurements for lumber produced in 2007 and in 2010 with sample sizes 98 and 282, respectively. We analyze the MOE and MOR characteristics separately. We regard the measurements of each index as two independent random samples from two populations satisfying the DRM assumption.

We use the EL approach to obtain point estimates and confidence intervals for the quantiles and the quantile differences between 2007 and 2010 of each quality

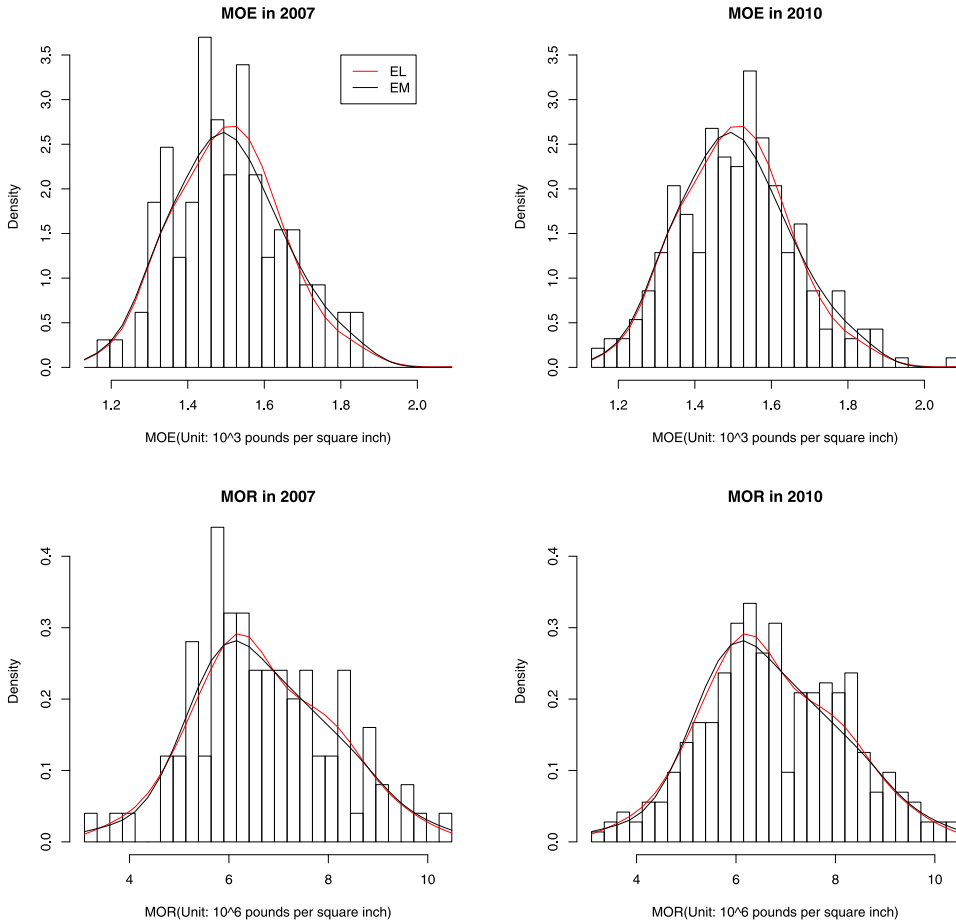


FIG. 1. Histograms of MOE and MOR with EL and EM density estimates.

index. We set $\mathbf{q}(x) = (1, x, \log(1 + |x|), \sqrt{|x|})^T$ as in the second simulation study. Different choices of $\mathbf{q}(x)$ do not markedly change the quantile estimates and confidence intervals, although they may give very different estimates of θ .

Figure 1 presents histograms of the MOE and MOR measurements with the EL and EM density estimates. We computed the EL and EM quantiles, the quantile differences, and their 95% confidence intervals for the 5% level to the 95% level in 5% increments. These point estimates and the confidence limits are connected to obtain the six plots shown in Figure 2. The EL quantiles and confidence limits are much smoother than those of EM. This phenomenon can be explained by the fact that the EL method is designed to use information from all the samples, which leads to less variation. These plots do not indicate that the EL method has sharper confidence limits. In fact, the EL intervals are 10% shorter than the EM intervals for both the quantiles and the quantile differences. In view of the simulation sup-

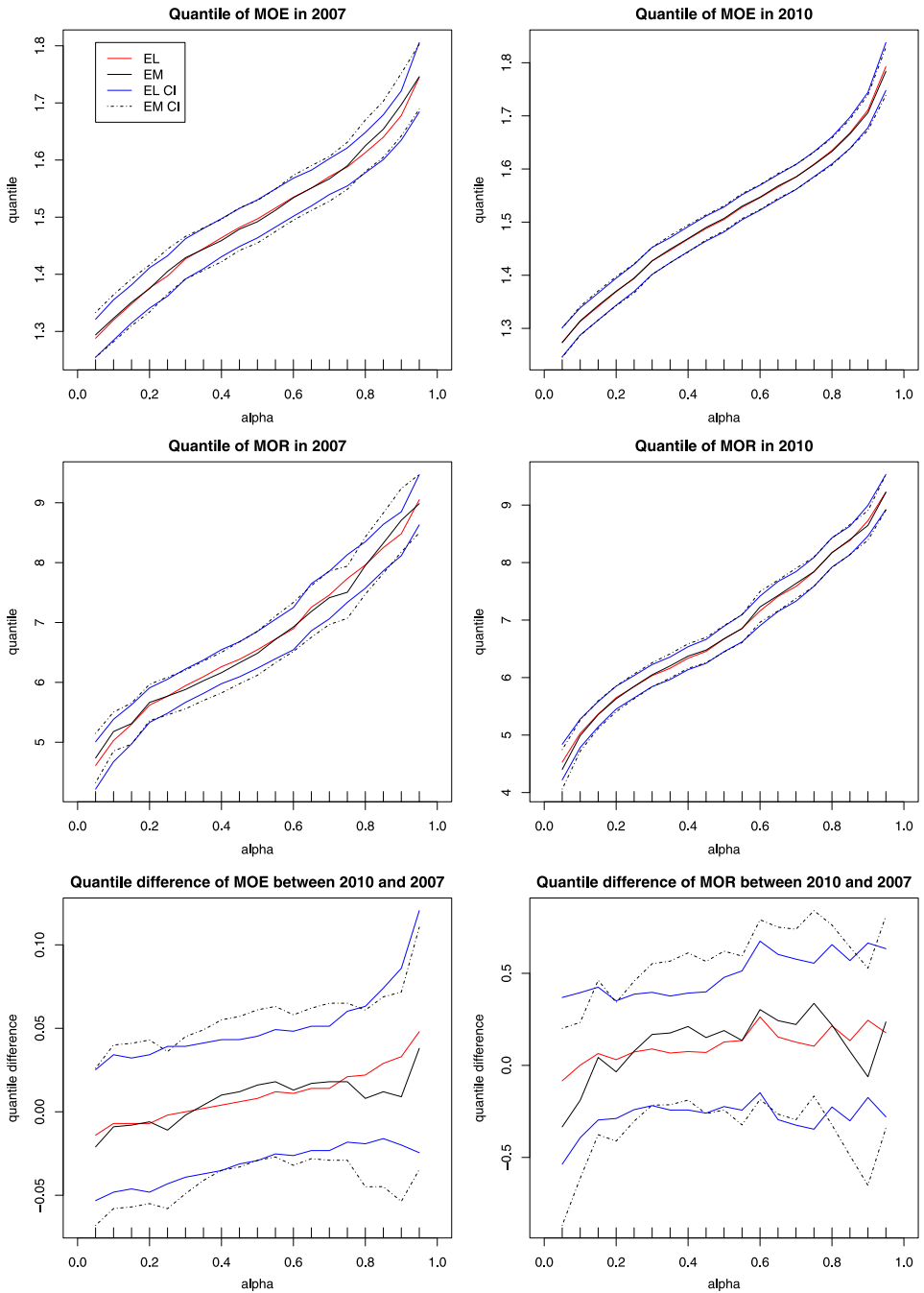


FIG. 2. Point estimates and confidence intervals for quantiles and quantile differences for lumber data.

port for the validity of both the EL and EM approaches, the 10% gain is likely real, and it implies significant cost savings in applications. To save space, we do not include tables of the results.

APPENDIX

A.1. Sketched proof of Theorem 2.1. This theorem mostly summarizes and extends the results in Qin and Zhang (1997), Fokianos et al. (2001) and Zhang (2002). To enable readers to understand the other proofs, we provide some necessary details. Interested readers can contact the authors for more detailed derivations.

LEMMA A.1. Assume the conditions of Theorem 2.1. For any θ such that $\theta = \theta^* + o(n^{-1/3})$, we have

$$\ell_n(\theta) - \ell_n(\theta^*) = (\theta - \theta^*)^\tau \mathbf{Z}_n - \frac{n}{2}(\theta - \theta^*)^\tau \mathbf{W}(\theta - \theta^*) + o_p(1),$$

where the vector \mathbf{Z}_n is

$$\begin{aligned} \mathbf{Z}_{n,r} &= \sum_{j=1}^{n_r} \mathbf{q}(x_{rj}) - \sum_{k,j} h_r(x_{kj}) \mathbf{q}(x_{kj}) \\ (10) \qquad &= \sum_{k,j} [\{\delta_{kr} - h_r(x_{kj})\} \mathbf{q}(x_{kj})], \end{aligned}$$

and $n^{-1/2}\mathbf{Z}_n$ is asymptotic normal with mean $\mathbf{0}$ and a positive definite covariance matrix.

Note that \mathbf{Z}_n is a sum of independent random vectors with finite moments. The mean of each is not zero, but the total is zero. In Theorem 3.2 we defined the r th segment of $\mathbf{B}_k = \mathbf{B}_k(\infty)$ as

$$\begin{aligned} \mathbf{B}_{k,r} &= \rho_k \mathbb{E}\{\delta_{kr} \mathbf{q}(x_{kj}) - h_r(x_{kj}) \mathbf{q}(x_{kj})\} \\ &= \int \{\delta_{kr} h_k(x) - h_r(x) h_k(x)\} \mathbf{q}(x) d\bar{G}(x). \end{aligned}$$

An interesting and useful observation is that for $1 \leq k, r \leq m$, $\mathbf{B}_{k,r} = \mathbf{B}_{r,k}$. From $\sum_{k=0}^m h_k(x) = 1$, we may also verify that

$$\mathbb{E}\{\mathbf{Z}_{n,r}\} = n \sum_{k=0}^m \mathbf{B}_{k,r} = 0.$$

Thus, $n^{-1/2}\mathbf{Z}_n$ is asymptotic normal with mean $\mathbf{0}$ and some variance matrix. This fact together with the form of the quadratic approximation implies that $\ell_n(\theta) - \ell_n(\theta^*)$ is maximized at the $\hat{\theta}$ that satisfies

$$n^{1/2}(\hat{\theta} - \theta^*) = n^{-1/2}\mathbf{W}^{-1}\mathbf{Z}_n + o_p(1).$$

See Hjort and Pollard (1993) for this justification.

The remaining task is to verify that the asymptotic variance of $n^{-1/2}\mathbf{Z}_n$ is given by $\mathbf{W} - \mathbf{WSW}$. This proves Theorem 2.1.

A.2. Proof of Theorem 3.1. The key to the proof is to show a seemingly obvious claim: $\hat{\xi}_r - \xi_r = O_p(n^{-1/2})$. This is an immediate consequence of

$$(11) \quad \hat{G}_r(x) - G_r(x) = O_p(n^{-1/2})$$

uniformly for x in a neighborhood of ξ_r . We now prove (11). Recall that

$$\hat{G}_r(x) = n_r^{-1} \sum_{k,j} h_r(x_{kj}; \hat{\theta}) I(x_{kj} \leq x).$$

Replacing $\hat{\theta}$ in $\hat{G}_r(x)$ by its true value θ^* , we define

$$\tilde{G}_r(x) = n_r^{-1} \sum_{k,j} h_r(x_{kj}; \theta^*) I(x_{kj} \leq x),$$

a sum of independent random variables. From $E\{\tilde{G}_r(x)\} = G_r(x)$, we get

$$E\{\tilde{G}_r(x) - G_r(x)\}^2 = n_r^{-2} \sum_{k,j} \text{var}\{h_r(x_{kj}; \theta^*) I(x_{kj} \leq x)\}.$$

Since $0 \leq h_r(x; \theta^*) \leq 1$, we have $\text{var}\{h_r(x_{kj}; \theta^*) I(x_{kj} \leq x)\} \leq 1$. Hence,

$$\sup_x E\{G_r(x) - \tilde{G}_r(x)\}^2 \leq n/n_r^2 = O(n^{-1}).$$

Since $G_r(x)$ and $\tilde{G}_r(x)$ are distribution functions, the above rate is uniform in x . Hence

$$\sup_x |\tilde{G}_r(x) - G_r(x)| = O_p(n^{-1/2}).$$

Therefore, (11) is implied by $\sup_x |\hat{G}_r(x) - \tilde{G}_r(x)| = O_p(n^{-1/2})$. Note that

$$\tilde{G}_r(x) - \hat{G}_r(x) = n_r^{-1} \sum_{k,j} \{h_r(x_{kj}; \theta^*) - h_r(x_{kj}; \hat{\theta})\} I(x_{kj} \leq x).$$

The partial derivative of $h_r(x; \theta)$ with respect to θ is bounded by $\sqrt{m}\|\mathbf{q}(x)\|$. Thus

$$|\tilde{G}_r(x) - \hat{G}_r(x)| \leq (n\rho_r)^{-1} \|\hat{\theta} - \theta^*\| \sum_{k,j} \sqrt{m}\|\mathbf{q}(x_{kj})\|.$$

The conditions of Theorem 2.1 imply that $\mathbf{q}(x_{kj})$ has finite moments of any order. Thus, $n^{-1} \sum_{k,j} \|\mathbf{q}(x_{kj})\| = O_p(1)$, and subsequently,

$$\begin{aligned} |\tilde{G}_r(x) - \hat{G}_r(x)| &\leq (n\rho_r)^{-1} \|\hat{\theta} - \theta^*\| \sum_{k,j} \sqrt{m}\|\mathbf{q}(x_{kj})\| = O_p(\|\hat{\theta} - \theta^*\|) \\ &= O_p(n^{-1/2}). \end{aligned}$$

This completes the proof of (11).

The classical Bahadur representation was a rate result in the mode of “almost sure.” Our result is stated in terms of “in probability,” and therefore it has a simpler proof. As for the classical case, the representation is equivalent to the following lemma:

LEMMA A.2. *Under the conditions of Theorem 3.1, for any $c > 0$ and $r = 0, 1, \dots, m$, we have*

$$\sup_{x: |x - \xi_r| < cn^{-1/2}} |\{\hat{G}_r(x) - \hat{G}_r(\xi_r)\} - \{G_r(x) - G_r(\xi_r)\}| = O_p(n^{-3/4}(\log(n))^{1/2}).$$

PROOF. We prove this lemma for $r = 0$; the other cases are equivalent. Without loss of generality we assume $x \geq \xi_r$. Note that

$$\begin{aligned} & \{\hat{G}_0(x) - \hat{G}_0(\xi_r)\} - \{\tilde{G}_0(x) - \tilde{G}_0(\xi_r)\} \\ &= n_0^{-1} \sum_{k,j} \{h_0(x_{kj}; \hat{\theta}) - h_0(x_{kj}; \theta^*)\} I(\xi_r < x_{kj} \leq x). \end{aligned}$$

By the mean value theorem and the specific form of $h_0(x; \theta)$, we have

$$|h_0(x_{kj}; \hat{\theta}) - h_0(x_{kj}; \theta^*)| \leq \sqrt{m} \|\mathbf{q}(x_{kj})\| \|\hat{\theta} - \theta^*\|.$$

From $E\{\|\mathbf{q}(x_{kj})\|\} < \infty$, we get $E\{\mathbf{q}(x_{kj}) I(\xi_r < x_{kj} \leq x)\} = O(n^{-1/2})$ and

$$\{\hat{G}_0(x) - \hat{G}_0(\xi_r)\} - \{\tilde{G}_0(x) - \tilde{G}_0(\xi_r)\} = O_p(n^{-1}).$$

With this result, Lemma A.2 is proved if we show that

$$\sup_{x: |x - \xi| < cn^{-1/2}} |[\tilde{G}_0(x) - \tilde{G}_0(\xi_r)] - [G_0(x) - G_0(\xi_r)]| = O_p(n^{-3/4}(\log(n))^{1/2}).$$

Since $\tilde{G}_0(x)$ is a sum of bounded random variables and $E\{\tilde{G}_0(x)\} = G_0(x)$, the result can be proved following Lemma 2.5.4E in Serfling (1980), page 97; we omit the details here. This completes the proof. \square

PROOF OF THEOREM 3.1. We have $\hat{\xi}_r - \xi_r = O_p(n^{-1/2})$ for any ξ_r , and the derivative of G_r is positive and continuous in a neighborhood of ξ_r . Therefore,

$$G_r(\hat{\xi}_r) - G_r(\xi_r) = g_r(\xi_r)(\hat{\xi}_r - \xi_r) + O_p(n^{-1}).$$

By definition, we have $\hat{G}_r(\hat{\xi}_r) = \alpha_r + O(n^{-1})$. Thus, replacing x by $\hat{\xi}_r$, and ξ by ξ_r , the result of Lemma A.2 becomes

$$|\{\alpha - \hat{G}_r(\xi_r)\} - g_r(\xi_r)(\hat{\xi}_r - \xi_r)| = O_p(n^{-3/4}(\log(n))^{1/2}).$$

This is equivalent to the conclusion of the theorem. \square

A.3. Proof of Theorem 3.2. Theorem 3.2 characterizes the asymptotic joint normality of a number of MELE distribution estimates. It is proved by approximating $\hat{G}_r(x)$ by a summation of independent random variables.

By the proof of Theorem 2.1, $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^* = n^{-1}\mathbf{W}^{-1}\mathbf{Z}_n + o_p(1)$. Hence,

$$\begin{aligned}\hat{G}_r(x) &= n_r^{-1} \sum_{k,j} h_r(x_{kj}; \hat{\boldsymbol{\theta}}) I(x_{kj} \leq x) \\ &= n_r^{-1} \sum_{k,j} [h_r(x_{kj}) + \{\dot{h}_r(x_{kj}; \boldsymbol{\theta}^*)\}^\tau (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)] I(x_{kj} \leq x) + o_p(n^{-1/2}) \\ &= n_r^{-1} \sum_{k,j} h_r(x_{kj}) I(x_{kj} \leq x) \\ &\quad + n_r^{-1} \left\{ n^{-1} \sum_{k,j} \dot{h}_r(x_{kj}; \boldsymbol{\theta}^*) I(x_{kj} < x) \right\}^\tau \mathbf{W}^{-1} \mathbf{Z}_n + o_p(n^{-1/2}),\end{aligned}$$

where $\dot{h}_r(x; \boldsymbol{\theta}^*) = \partial h_r(x; \boldsymbol{\theta}) / \partial \boldsymbol{\theta} |_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}$. Working out the expression of $\dot{h}_r(x; \boldsymbol{\theta}^*)$ in terms of $\mathbf{q}(x)$ and $h_r(x)$, and by the law of large numbers, we find that, almost surely,

$$n^{-1} \sum_{k,j} \dot{h}_r(x_{kj}; \boldsymbol{\theta}^*) I(x_{kj} < x) \rightarrow \mathbf{B}_r(x),$$

where $\mathbf{B}_r(x)$ is defined in the theorem. We remark here that $\lim_{x \rightarrow \infty} \mathbf{B}_r(x) = \mathbf{B}_r$; the latter was defined in the proof of Theorem 2.1. Before the final step, we may verify that

$$G_r(x) = n_r^{-1} \mathbf{E} \left\{ \sum_{k,j} h_r(x_{kj}) I(x_{kj} \leq x) \right\}.$$

These preparations enable us to write

$$\begin{aligned}\hat{G}_r(x) - G_r(x) &= n_r^{-1} \sum_{k,j} [h_r(x_{kj}) I(x_{kj} \leq x) - \mathbf{E}\{h_r(x_{kj}) I(x_{kj} \leq x)\}] \\ &\quad + n_r^{-1} \mathbf{B}_r^\tau(x) \mathbf{W}^{-1} \mathbf{Z}_n + o_p(n^{1/2}).\end{aligned}$$

The two leading terms are summations of independent random vectors and both have mean zero. The joint asymptotic normality of $\sqrt{n}\{\hat{G}_r(x) - G_r(x)\}$ and $\sqrt{n}\{\hat{G}_s(y) - G_s(y)\}$ is hence implied. We derive the algebraic expression of $\omega_{rs}(x, y)$ in the next subsection.

A.3.1. Asymptotic covariance $\sigma_{rs}(x, y)$. From the expansion of $\hat{G}_r(x) - G_r(x)$, $\omega_{rs}(x, y)$ is decomposed into four covariances. Using $\text{var}(\mathbf{Z}_n) = n(\mathbf{W} - \mathbf{WSW})$ as shown earlier, we find that one of them is given by

$$\text{Cov}(\mathbf{B}_r^\tau(x) \mathbf{W}^{-1} \mathbf{Z}_n, \mathbf{B}_s^\tau(y) \mathbf{W}^{-1} \mathbf{Z}_n) = n \mathbf{B}_r^\tau(x) (\mathbf{W}^{-1} - \mathbf{S}) \mathbf{B}_s(y).$$

We build another term from the following computations:

$$\begin{aligned} & \text{Cov}\{h_r(x_{kj})I(x_{kj} \leq x), h_s(x_{kj})I(x_{kj} \leq y)\} \\ &= \rho_k^{-1} \int_{-\infty}^{(x \wedge y)} h_r(x)h_s(x)h_k(x) d\bar{G}(x) - \rho_k^{-2}c_{rk}(x)c_{sk}(y), \end{aligned}$$

where $c_{rs}(x) = \int_{-\infty}^x h_r(t)h_s(t) d\bar{G}(t)$. Because $\sum_{k=0}^m h_k(x) = 1$, we get

$$\begin{aligned} & \sum_{k=0}^m n_k \text{Cov}\{h_r(x_{kj})I(x_{kj} \leq x), h_s(x_{kj})I(x_{kj} \leq y)\} \\ &= n \left\{ c_{rs}(x \wedge y) - \sum_{k=0}^m \rho_k^{-1} c_{rk}(x)c_{sk}(y) \right\}. \end{aligned}$$

The last task is the cross-term $\text{Cov}\{h_r(x_{kj})I(x_{kj} \leq x), \mathbf{Z}_n\}$. We break \mathbf{Z}_n into segments $\mathbf{Z}_{n,s}$ and then into centralized $\tilde{\mathbf{Z}}_{n,s}$.

$$\begin{aligned} & \text{Cov}\{h_r(x_{kj})I(x_{kj} \leq x), \mathbf{Z}_{n,s}\} \\ &= \mathbb{E}[h_r(x_{kj})I(x_{kj} \leq x) \{(\delta_{kr} - h_s(x_{kj}))\mathbf{q}^\tau(x_{kj}) - \rho_k^{-1}\mathbf{B}_{k,s}^\tau\}] \\ &= \rho_k^{-1}\mathbf{B}_{k,r}^\tau(x) - \rho_k^{-2}c_{rk}(x)\mathbf{B}_{k,s}^\tau. \end{aligned}$$

Summing over $\{k, j\}$, the first term sums to zero, so we find

$$\text{Cov}\left\{\sum_{k,j} h_r(x_{kj})I(x_{kj} \leq x), \mathbf{Z}_{n,s}\right\} = -n \sum_{k=0}^m \rho_k^{-1} c_{rk}(x)\mathbf{B}_{k,s}^\tau.$$

Next, we assemble $\mathbf{B}_{k,s}$ over s to get \mathbf{B}_k ,

$$\text{Cov}\left\{\sum_{k,j} h_r(x_{kj})I(x_{kj} \leq x), \mathbf{Z}_n\right\} = -n \sum_{k=0}^m \rho_k^{-1} c_{rk}(x)\mathbf{B}_k^\tau.$$

Entering $\mathbf{B}_s^\tau(y)\mathbf{W}^{-1}$ into the second argument of the covariance, we get

$$\text{Cov}\left\{\sum_{k,j} h_r(x_{kj})I(x_{kj} \leq x), \mathbf{B}_s^\tau(y)\mathbf{W}^{-1}\mathbf{Z}_n\right\} = -n \sum_{k=0}^m \rho_k^{-1} c_{rk}(x)\mathbf{B}_k^\tau\mathbf{W}^{-1}\mathbf{B}_s(y).$$

Thus, the covariance between $\sqrt{n}\{\hat{G}_r(x) - G_r(x)\}$ and $\sqrt{n}\{\hat{G}_s(y) - G_s(y)\}$ is given by

$$\begin{aligned} \omega_{rs}(x, y) &= (\rho_r \rho_s)^{-1} \left[c_{rs}(x \wedge y) - \sum_{k=0}^m \rho_k^{-1} c_{rk}(x)c_{sk}(y) \right. \\ &\quad \left. - \sum_{k=0}^m \rho_k^{-1} \mathbf{B}_k^\tau \mathbf{W}^{-1} \{c_{rk}(x)\mathbf{B}_s(y) + c_{sk}(y)\mathbf{B}_r(x)\} \right. \\ &\quad \left. + \mathbf{B}_r^\tau(x)(\mathbf{W}^{-1} - \mathbf{S})\mathbf{B}_s(y) \right]. \end{aligned}$$

Further simplification is possible. We find $c_{r0}(x) + \mathbf{B}_r^\tau(x)\mathbf{W}^{-1}\mathbf{B}_0 = 0$, $c_{rk}(x) + \mathbf{B}_r^\tau(x)\mathbf{W}^{-1}\mathbf{B}_k = \delta_{rk}G_r(x)$ and $\sum_{k=0}^m \rho_k^{-1}\mathbf{W}^{-1}\mathbf{B}_k\mathbf{B}_k^\tau\mathbf{W}^{-1} = \mathbf{S}$. These findings lead to

$$\omega_{rs}(x, y) = (\rho_r \rho_s)^{-1} \{c_{rs}(x \wedge y) + \mathbf{B}_r^\tau(x)\mathbf{W}^{-1}\mathbf{B}_s(y) - \rho_r \delta_{rs} G_r(x) G_s(y)\}.$$

The final expression (8) is obtained by noticing that $c_{rs}(x) = \rho_r \delta_{rs} G_r(x) - a_{rs}(x)$, where $a_{rs}(x)$ was defined in Theorem 3.2.

A.4. Proof of Theorem 4.1. Both the EL and EM quantiles admit Bahadur representations, and it suffices to show the same conclusion for the distribution estimators. For the bivariate case, we denote the asymptotic covariance matrices of the EL and EM distributions as

$$\Omega_{\text{EL}} = \begin{pmatrix} \sigma_{rr}(x, x) & \sigma_{rs}(x, y) \\ \sigma_{rs}(x, y) & \sigma_{ss}(y, y) \end{pmatrix}, \quad \Omega_{\text{EM}} = \begin{pmatrix} \omega_{rr}(x, x) & \omega_{rs}(x, y) \\ \omega_{rs}(x, y) & \omega_{ss}(y, y) \end{pmatrix},$$

where $x = \xi_r$ and $y = \xi_s$ are two population quantiles or two real values. We show that $\Omega_{\text{EM}} - \Omega_{\text{EL}}$ is nonnegative definite by writing it as $D_{11} - D_{12}D_{22}^{-1}D_{21}$, with the D_{ij} being blocks of a nonnegative definite matrix D . By standard matrix theory, the nonnegative definiteness of D implies that of $D_{11} - D_{12}D_{22}^{-1}D_{21}$. The generic element of $\Omega_{\text{EM}} - \Omega_{\text{EL}}$ is $a_{rs}(x, y) - \mathbf{B}_r^\tau(x)\mathbf{W}^{-1}\mathbf{B}_s(y)$, which fits into $D_{11} - D_{12}D_{22}^{-1}D_{21}$ with

$$D_{11} = \begin{pmatrix} a_{rr}(x) & a_{rs}(x \wedge y) \\ a_{rs}(x \wedge y) & a_{ss}(y) \end{pmatrix}, \quad D_{12} = \begin{pmatrix} \mathbf{B}_r^\tau(x) \\ \mathbf{B}_s^\tau(y) \end{pmatrix},$$

$D_{21} = D_{12}^\tau$ and $D_{22} = \mathbf{W}$. We will show that $D = \int \mathbf{U}(z) d\bar{G}(z)$ for some nonnegative definite $\mathbf{U}(z)$ for all z . Then D is nonnegative definite and so is $\Omega_{\text{EM}} - \Omega_{\text{EL}} = D_{11} - D_{12}D_{22}^{-1}D_{21}$.

We now search for such a $\mathbf{U}(z)$. We write

$$a_{rs}(x \wedge y) = \int I(z \leq x)I(z \leq y)\mathbf{H}[r + 1, s + 1](z) d\bar{G}(z),$$

$$\mathbf{B}_r(x) = \int I(z \leq x)\mathbf{H}[-1, r + 1](z) \otimes \mathbf{q}(z) d\bar{G}(z),$$

$$\mathbf{W} = \int \mathbf{H}[-1, -1](z) \otimes \{\mathbf{q}(z)\mathbf{q}^\tau(z)\} d\bar{G}(z).$$

Using the Khatrin–Rao product operator $*$ [Liu and Trenkler (2008)], we find such a $\mathbf{U}(z) = A_1(z) * A_2(z)$ with

$$A_1(z) = P \begin{pmatrix} \mathbf{H}[r + 1, r + 1](z) & \mathbf{H}[r + 1, s + 1](z) & \mathbf{H}[r + 1, -1](z) \\ \mathbf{H}[s + 1, r + 1](z) & \mathbf{H}[s + 1, s + 1](z) & \mathbf{H}[s + 1, -1](z) \\ \mathbf{H}[-1, r + 1](z) & \mathbf{H}[-1, s + 1](z) & \mathbf{H}[-1, -1](z) \end{pmatrix} P,$$

$$A_2(z) = \begin{pmatrix} 1 & 1 & \mathbf{q}^\tau(z) \\ 1 & 1 & \mathbf{q}^\tau(z) \\ \mathbf{q}(z) & \mathbf{q}(z) & \mathbf{q}(z)\mathbf{q}^\tau(z) \end{pmatrix}, \quad P = \begin{pmatrix} I(z \leq x) & 0 & 0 \\ 0 & I(z \leq y) & 0 \\ 0 & 0 & \mathbf{I}_m \end{pmatrix}.$$

The matrix $A_2(z)$ is clearly nonnegative definite for any z . Note that $\mathbf{H}(z)$ is nonnegative definite for any z ; the nonnegative definiteness of $A_1(z)$ is an easy consequence. Since the $*$ product of two nonnegative definite matrices is still nonnegative definite [Lemma 5 of Liu and Trenkler (2008)], we conclude that $\mathbf{U}(z) = A_1(z) * A_2(z)$ is also nonnegative definite for any z . This completes the proof. This proof can easily be extended to the case where more distributions or quantiles are involved.

Acknowledgments. We are grateful to the referees, the Associate Editor, and the Editor for helpful comments.

REFERENCES

- ANDERSON, J. A. (1979). Multivariate logistic compounds. *Biometrika* **66** 17–26. [MR0529143](#)
- BAHADUR, R. R. (1966). A note on quantiles in large samples. *Ann. Math. Statist.* **37** 577–580. [MR0189095](#)
- CHEN, H. and CHEN, J. (2000). Bahadur representations of the empirical likelihood quantile processes. *J. Nonparametr. Stat.* **12** 645–660. [MR1784802](#)
- DEHEUVELS, P. (1977). Estimation non paramétrique de la densité par histogrammes généralisés. *Rev. Statist. Appl.* **25** 5–42. [MR0501555](#)
- FOKIANOS, K. (2007). Density ratio model selection. *J. Stat. Comput. Simul.* **77** 805–819. [MR2409865](#)
- FOKIANOS, K. and KAIMI, I. (2006). On the effect of misspecifying the density ratio model. *Ann. Inst. Statist. Math.* **58** 475–497. [MR2327888](#)
- FOKIANOS, K., KEDEM, B., QIN, J. and SHORT, D. A. (2001). A semiparametric approach to the one-way layout. *Technometrics* **43** 56–65. [MR1819908](#)
- HJORT, N. L. and POLLARD, D. (1993). Asymptotics for minimisers of convex processes. Technical report, Yale Univ.
- KEZIOU, A. and LEONI-AUBIN, S. (2008). On empirical likelihood for semiparametric two-sample density ratio models. *J. Statist. Plann. Inference* **138** 915–928. [MR2384498](#)
- KIEFER, J. (1967). On Bahadur’s representation of sample quantiles. *Ann. Math. Statist.* **38** 1323–1342. [MR0217844](#)
- LIU, S. and TRENKLER, G. (2008). Hadamard, Khatri–Rao, Kronecker and other matrix products. *Int. J. Inf. Syst. Sci.* **4** 160–177. [MR2401772](#)
- OWEN, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **75** 237–249. [MR0946049](#)
- OWEN, A. B. (2001). *Empirical Likelihood*. Chapman & Hall/CRC, New York.
- QIN, J. and ZHANG, B. (1997). A goodness-of-fit test for logistic regression models based on case–control data. *Biometrika* **84** 609–618. [MR1603924](#)
- R DEVELOPMENT CORE TEAM (2011). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York. [MR0595165](#)
- SILVERMAN, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman & Hall, London. [MR0848134](#)
- WU, W. B. (2005). On the Bahadur representation of sample quantiles for dependent sequences. *Ann. Statist.* **33** 1934–1963. [MR2166566](#)
- ZHANG, B. (2000). Quantile estimation under a two-sample semi-parametric model. *Bernoulli* **6** 491–511. [MR1762557](#)

- ZHANG, B. (2002). Assessing goodness-of-fit of generalized logit models based on case-control data. *J. Multivariate Anal.* **82** 17–38. [MR1918613](#)
- ZHOU, Z. and WU, W. B. (2009). Local linear quantile estimation for nonstationary time series. *Ann. Statist.* **37** 2696–2729. [MR2541444](#)

DEPARTMENT OF STATISTICS
UNIVERSITY OF BRITISH COLUMBIA
VANCOUVER, BC
CANADA V6T 1Z2
E-MAIL: jhchen@stat.ubc.ca

DEPARTMENT OF STATISTICS AND ACTUARIAL SCIENCES
SCHOOL OF FINANCE AND STATISTICS
EAST CHINA NORMAL UNIVERSITY
SHANGHAI 200241
CHINA
E-MAIL: ykliu@sfs.ecnu.edu.cn