Risk journals

Volume 11 Number 1
Fall 2008

The Journal of Risk

The Journal of Risk

Volume 11 Number 1 Fall 2008

Incisive Media Plc, Haymarket House, 28-29 Haymarket, London SW1Y 4RX

The Journal of Risk (3–32) Volume 11/Number 1, Fall 2008

Empirical likelihood for value-at-risk and expected shortfall

Rafet Evren Baysal
Department of Industrial Engineering and Management Sciences, Northwestern University, 2145 Sheridan Road, Evanston, IL 60208-3119, USA; email: e-baysal@northwestern.edu

Jeremy Staum
Department of Industrial Engineering and Management Sciences, Northwestern University, 2145 Sheridan Road, Evanston, IL 60208-3119, USA; email: j-staum@northwestern.edu

When estimating risk measures, whether from historical data or by Monte Carlo simulation, it is helpful to have confidence intervals that provide information about statistical uncertainty. We provide asymptotically valid confidence intervals and confidence regions involving value-at-risk (VaR), conditional tail expectation and expected shortfall (conditional VaR), based on three different methodologies. One is an extension of previous work based on robust statistics, the second is a straightforward application of bootstrapping, and we derive the third using empirical likelihood. We then evaluate the small-sample coverage of the confidence intervals and regions in simulation experiments using financial examples. We find that the coverage probabilities are approximately nominal for large sample sizes, but are noticeably low when sample sizes are too small (roughly, less than 500 here). The new empirical likelihood method provides the highest coverage at moderate sample sizes in these experiments.

1 INTRODUCTION

We want to measure the risk of a given portfolio that has random profits at the end of a predetermined investment period. We can sample from the distribution of the portfolio's profits using Monte Carlo simulation based on a stochastic model of financial markets. Our focus will be on estimating risk measures for our portfolio based on simulated profits and providing information in the form of confidence intervals and regions about the statistical uncertainty of these estimates. We address only this Monte Carlo sampling error in estimating risk, not the model risk that includes errors introduced by using an incorrect model of financial markets and statistical error in estimating the model's parameters from data. We will emphasize moderate Monte Carlo sample sizes, which are appropriate when

This material is based upon work supported by the National Science Foundation under Grants No. DMS-0202958 and DMI-0555485. The authors are grateful for discussions with Dan Apley, Per Mykland, Barry Nelson and Art Owen and to two anonymous referees, who provided references and helped to improve the presentation and substance of this article. The opinions expressed are those of the authors, who are responsible for any errors.

it is computationally expensive to simulate financial scenarios and determine the value of the portfolio in each scenario.

Define V to be the random profit of the given portfolio at a specific investment horizon. The 95% value-at-risk ($\mathrm{VaR}_{95\%}$) of the portfolio is the 95% quantile of the loss $-V$. A related risk measure is the 95% conditional tail expectation ($\mathrm{CTE}_{95\%}$), which is:

$$\mathrm{CTE}_{95\%} = \mathbf{E}[-V \mid -V \geq \mathrm{VaR}_{95\%}]$$

Another closely related risk measure is expected shortfall ($\mathrm{ES}_{95\%}$), which is:

$$\mathrm{ES}_{95\%} = -\frac{1}{0.05}(\mathbf{E}[V\mathbf{1}_{\{V \leq v_{0.05}\}}] + v_{0.05}(0.05 - \Pr[V \leq v_{0.05}]))$$

where $v_{0.05}$ is the lower 5% quantile (Definition 5.2) of the distribution of V. Under continuity conditions on the loss distribution, CTE equals ES (Acerbi and Tasche (2002)). ES always equals conditional value-at-risk (CVaR), which is coherent (Acerbi and Tasche (2002); Rockafellar and Uryasev (2002)). A risk measure is coherent if it satisfies certain axioms of translation invariance, subadditivity, positive homogeneity and monotonicity (Artzner *et al* (1999)). We use the term "expected shortfall" here because ES includes an expectation, which is closely related to simulation, on which we focus, while CVaR is closely associated with a minimization formula due to Rockafellar and Uryasev (2000, 2002).

Our goal is to construct confidence intervals and regions for the above risk measures based on a simulated sample $V_1, \ldots, V_k$ of independent profits with common distribution F_0. Let $V_{[1]}, \ldots, V_{[k]}$ be ascending order statistics. The obvious point estimators of VaR and CTE at the $(1-p)$ level, assuming kp is an integer, are:

$$\widehat{\mathrm{VaR}}_{1-p,k} = -V_{[kp]} \tag{1}$$

$$\widehat{\mathrm{CTE}}_{1-p,k} = -\frac{1}{kp}\sum_{i=1}^{kp} V_{[i]} \tag{2}$$

respectively. Other point estimators are discussed in Section 7.

Here we focus on constructing a confidence interval for ES and a confidence region for VaR and CTE simultaneously. We consider three methods for constructing them. To facilitate comparisons between their error rates, we also compare the three methods' confidence intervals for VaR to a standard confidence interval for VaR. This standard is the binomial confidence interval for a quantile (Clopper and Pearson (1934)), which we summarize in Section 2. In Section 3, we construct confidence intervals and regions by extending results of Yamai and Yoshiba (2002) and Manistre and Hancock (2005) based on the influence function used in robust statistics. Section 4 briefly discusses how to construct them by bootstrapping. The major new results are in Section 5, where we show how to construct them using empirical likelihood (Owen (2001)). In Section 6, we present computer simulation experiments to show that these confidence intervals and regions achieve close to nominal coverage for large sample sizes, but not for moderate sample sizes that are too small. Empirical likelihood provides the highest coverage at moderate sample sizes in these experiments, for the most part.

One contribution of this paper is simply in providing the first test (known to us) of the coverage of confidence regions and intervals involving CTE on financial examples. This provides some guidance about how large the sample size must be before the coverage is adequate, or how low the coverage might be at low sample sizes. Through a non-trivial application of empirical likelihood, we provide a method for generating confidence regions and intervals with higher coverage. The empirical likelihood approach is also useful in enabling risk measurement procedures that can cope with the need to use simulation at two levels: in sampling from a distribution of risky scenarios and in estimating the portfolio loss in each of those scenarios (Lan *et al* (2007)).

2 BINOMIAL CONFIDENCE INTERVALS FOR VALUE-AT-RISK

There is a well-known confidence interval for quantiles (Clopper and Pearson (1934)), and thus VaR, based on the binomial distribution of the number of losses $N(q) := \sum_{i=1}^{k} \mathbf{1}\{-V_i \geq q\}$ that exceed a threshold q. The lower and upper limits of a two-sided confidence interval for VaR_{1-p} with $(1-\alpha)$ nominal coverage probability are respectively:

$$\inf\left\{q \;\middle|\; \sum_{n=N(q)+1}^{k} \binom{k}{n} p^n (1-p)^{k-n} \geq \alpha/2\right\}$$

and:

$$\sup\left\{q \;\middle|\; \sum_{n=0}^{N(q)} \binom{k}{n} p^n (1-p)^{k-n} \geq \alpha/2\right\}$$

The limit of a one-sided upper confidence interval for VaR_{1-p} with $(1-\alpha)$ nominal coverage probability is $\sup\{q \mid \sum_{n=0}^{N(q)} \binom{k}{n} p^n (1-p)^{k-n} \geq \alpha\}$.

These endpoints of the confidence interval equal order statistics of the data sample, ie, quantiles of the empirical distribution function. It is not generally possible to get exactly nominal coverage for the confidence interval because of the discreteness of the empirical distribution function, or, viewed differently, because of the discreteness of the binomial distribution (Agresti and Coull (1998)). Nonetheless, these confidence intervals are often called “exact” because they are related to an exact hypothesis test for the value of the quantile. The justification of these confidence intervals does not involve the convergence of a statistic’s distribution to a limiting distribution as sample size k grows, as do the methods described in later sections.

3 INFLUENCE FUNCTION

The approach based on the influence function in the theory of robust statistics allows us to compute the variances of the asymptotic normal distributions of the estimators in Equations (1) and (2). As Manistre and Hancock (2005, note 6) state, under regularity conditions discussed in Staudte and Sheather (1990):

$$k\,\mathbf{Var}(\widehat{\mathrm{CTE}}_{1-p,k}) \to \frac{\mathbf{Var}(-V \mid -V > \mathrm{VaR}_{1-p}) + p(\mathrm{CTE}_{1-p} - \mathrm{VaR}_{1-p})^2}{(1-p)} \tag{3}$$

$$k\,\mathbf{Var}(\widehat{\mathrm{VaR}}_{1-p,k}) \to \frac{p(1-p)}{f^2(\mathrm{VaR}_{1-p})} \tag{4}$$

$$k\,\mathbf{Cov}(\widehat{\mathrm{CTE}}_{1-p,k}, \widehat{\mathrm{VaR}}_{1-p,k}) \to \frac{p(\mathrm{CTE}_{1-p} - \mathrm{VaR}_{1-p})}{f(\mathrm{VaR}_{1-p})} \tag{5}$$

where $f(\mathrm{VaR}_{1-p})$ is the value of the probability density of the underlying distribution at the quantile. Similar results appear in Yamai and Yoshiba (2002), but complicated by a truncation argument. Yamai and Yoshiba (2002) report confidence intervals for VaR and ES, but not a confidence region for both simultaneously. It also remains to show how to estimate the unknown quantities in Equations (3)–(5) to construct a confidence interval or region.

Manistre and Hancock (2005) propose the following estimates of asymptotic variances and covariances:

$$\widehat{\mathbf{Var}}_k(\widehat{\mathrm{CTE}}_{1-p,k}) = \frac{(kp-1)^{-1}\sum_{i=1}^{kp}(\widehat{\mathrm{CTE}}_{1-p,k} + V_{[i]})^2 + p(\widehat{\mathrm{CTE}}_{1-p,k} + V_{[kp]})^2}{k\,(1-p)} \tag{6}$$

$$\widehat{\mathbf{Var}}_k(\widehat{\mathrm{VaR}}_{1-p,k}) = \frac{p(1-p)}{k\hat{f}^2(-V_{[kp]})} \tag{7}$$

$$\widehat{\mathbf{Cov}}_k(\widehat{\mathrm{CTE}}_{1-p,k}, \widehat{\mathrm{VaR}}_{1-p,k}) = \frac{p(\widehat{\mathrm{CTE}}_{1-p,k} + V_{[kp]})}{k\,\hat{f}(-V_{[kp]})} \tag{8}$$

where $\hat{f}(-V_{[kp]})$ is an estimate of the probability density. Manistre and Hancock (2005) proposed the use of:

$$\hat{f}(-V_{[kp]}) = \frac{\xi}{F_k^{-1}(p) - F_k^{-1}(p-\xi)}$$

where $F_k(x) = (1/k)\sum_{i=1}^{k} 1_{\{V_i \le x\}}$ is the empirical distribution derived from the sample of size k and ξ is chosen to be a small number. Note that the choice of ξ affects the empirical density function estimate $\hat{f}$ considerably, especially for small samples. Hence, we propose to use the kernel method to estimate f via a Gaussian kernel estimator function:

$$\hat{f}_k(-V_{[kp]}) = \frac{1}{kh}\sum_{i=1}^{k} \Phi'\left(\frac{-V_{[kp]} + V_{[i]}}{h}\right) \tag{9}$$

where $h = (4/3k)^{1/5}\sigma$, $\Phi'(u) = (2\pi)^{-1/2}\exp(-u^2/2)$ and the sample standard deviation can be used for σ.

We extend the above results to create confidence intervals and regions. We define $Y := \binom{\widehat{\mathrm{VaR}}_{1-p,k}}{\widehat{\mathrm{CTE}}_{1-p,k}}$ based on a sample of size k. This is asymptotically normal with mean $y_0 := \binom{\mathrm{VaR}_{1-p}}{\mathrm{CTE}_{1-p}}$ and covariance matrix $\mathbf{\Sigma}$ described by Equations (3)–(5). There exists a unique symmetric positive definite matrix $\mathbf{A}$ such that $\mathbf{A}^\top\mathbf{A} = \mathbf{\Sigma}^{-1}$. We define $Z := \mathbf{A}(Y - y_0)$ whose components are independent and asymptotically standard normal. Then, the quadratic form $(Y - y_0)^\top\mathbf{\Sigma}^{-1}(Y - y_0) = \mathrm{Z}^\top\mathrm{Z}$ is distributed asymptotically as χ^2 with two degrees of freedom. Note that the asymptotic formulas (6)–(8) can be used to construct $\widehat{\mathbf{\Sigma}}$ as an estimate of the covariance matrix $\mathbf{\Sigma}$. With probability one, $\widehat{\mathbf{Var}}_k(\widehat{\mathrm{CTE}}_{1-p,k})$ converges to $\mathbf{Var}(\widehat{\mathrm{CTE}}_{1-p,k})$ (Hong (2006)). Weak convergence results for the kernel density estimate $\hat{f}_k$ (9) are given by Silverman (1978) and Chang *et al* (2003). Hence, $\widehat{\mathbf{\Sigma}}$ is a consistent estimator of $\mathbf{\Sigma}$ and by the converging-together lemma of Durrett (1996) an asymptotically valid $(1 - \alpha)$ confidence region for VaR_{1-p} and CTE_{1-p}, is an elliptical region centered at $(\widehat{\mathrm{VaR}}_{1-p,k}, \widehat{\mathrm{CTE}}_{1-p,k})$ and is given by:

$$\{y_0 \mid (Y - y_0)^\top\widehat{\mathbf{\Sigma}}^{-1}(Y - y_0) \le \chi^2_{(2),1-\alpha}\} \tag{10}$$

where $\chi^2_{(2),1-\alpha}$ is the $1 - \alpha$ quantile of the chi-squared distribution with two degrees of freedom. By applying the converging-together lemma to $\widehat{\mathrm{CTE}}_{1-p,k}$ and $\widehat{\mathbf{Var}}_k(\widehat{\mathrm{CTE}}_{1-p,k})$, one can show that where $Z_{1-\alpha/2}$ is the $1 - \alpha/2$ quantile of the standard normal distribution:

$$\{\mu_0 \mid |\widehat{\mathrm{CTE}}_{1-p,k} - \mu_0| \le Z_{1-\alpha/2}\sqrt{\widehat{\mathbf{Var}}_k(\widehat{\mathrm{CTE}}_{1-p,k})}\} \tag{11}$$

is a two-sided confidence interval for CTE_{1-p} (Hong (2006)). Correspondingly:

$$\{\mu_0 \mid \mu_0 \le \widehat{\mathrm{CTE}}_{1-p,k} + Z_{1-\alpha}\sqrt{\widehat{\mathbf{Var}}_k(\widehat{\mathrm{CTE}}_{1-p,k})}\} \tag{12}$$

is a one-sided upper confidence interval for CTE_{1-p}. Likewise:

$$\{q_0 \mid |\widehat{\mathrm{VaR}}_{1-p,k} - q_0| \le Z_{1-\alpha/2}\sqrt{\widehat{\mathbf{Var}}_k(\widehat{\mathrm{VaR}}_{1-p,k})}\} \tag{13}$$

is a two-sided confidence interval for VaR_{1-p} and:

$$\{q_0 \mid q_0 \le \widehat{\mathrm{VaR}}_{1-p,k} + Z_{1-\alpha}\sqrt{\widehat{\mathbf{Var}}_k(\widehat{\mathrm{VaR}}_{1-p,k})}\} \tag{14}$$

is a one-sided upper confidence interval for VaR_{1-p}.

4 BOOTSTRAPPING

The idea of bootstrapping to create confidence intervals for CTE was suggested by Dowd (2005) and Hardy (2006). Bootstrap methods are in general motivated by the need to evaluate the accuracy of an estimate in the absence of distributional assumptions (Chernick (1999)). Shao and Tu (1995) discuss in detail the application of bootstrap methods to hypothesis testing and confidence interval estimation for various statistics including quantiles. The logic behind bootstrapping for quantile estimation is readily applicable to estimating VaR, CTE and ES.

As before, let $V_{[1]} \leq \cdots \leq V_{[k]}$ be order statistics, sorted after sampling profits independently from the common distribution F_0. We assume kp is an integer. To estimate VaR_{1-p} and CTE_{1-p} based on this sample, we compute the obvious estimators previously mentioned:

$$\widehat{\mathrm{VaR}}_{1-p,k} = -V_{[kp]} \quad \text{and} \quad \widehat{\mathrm{CTE}}_{1-p,k} = -\frac{1}{kp}\sum_{i=1}^{kp} V_{[i]}$$

We will denote them by $\hat{q}_k(p)$ and $\hat{\mu}_k(p)$, respectively, to emphasize their dependence on the initial sample of size k. Because kp is an integer, the estimate $\hat{\mu}_k(p)$ of CTE_{1-p} is also an estimate of ES_{1-p}.

In order to assess the uncertainty associated with these estimates, we generate B independently and identically distributed (iid) bootstrap samples by resampling from the empirical distribution function F_k of the initial Monte Carlo sample. For risk management applications, resampling may be considerably faster than generating samples from the original distribution F_0. We denote the bootstrap samples by $\hat{V}^b_{[1]}, \ldots, \hat{V}^b_{[k]}$ for $b = 1, \ldots, B$. From the bth bootstrap sample, we compute the estimates:

$$\hat{\mu}_b(p) = -\frac{1}{kp}\sum_{i=1}^{kp} \hat{V}^b_{[i]} \quad \text{and} \quad \hat{q}_b(p) = -\hat{V}^b_{[kp]}$$

Note that we only need $\hat{V}^b_{[1]}, \ldots, \hat{V}^b_{[kp]}$ to compute $\hat{q}_b(p)$ and $\hat{\mu}_b(p)$, and the bootstrap sample for the first kp order statistics can be generated efficiently by $\hat{V}^b_{[j]} = F_k^{-1}(U_{[j]})$, where $U_{[1]}, \ldots, U_{[k]}$ are the order statistics of an iid sample of size k from the standard uniform distribution. The following algorithm of order $O(kp)$ from Dagpunar (1988) can be used to generate $U_{[1]}, \ldots, U_{[kp]}$:

$$U_{[0]} = 0$$

for $i = 1$ to kp

  generate $V_i \sim \mathrm{Uniform}[0, 1]$

  $U_{[i]} = 1 - (1 - U_{[i-1]})V_i^{1/(kp-i+1)}$

end for

4.1 Bootstrap confidence intervals for VaR and ES

There are various methods for constructing asymptotically valid confidence intervals for VaR_{1-p} and ES_{1-p} from $\hat{q}_1(p), \ldots, \hat{q}_B(p)$ and $\hat{\mu}_1(p), \ldots, \hat{\mu}_B(p)$, such as the bootstrap t, the bootstrap percentile, the bootstrap bias-corrected percentile and the bootstrap bias-corrected/accelerated (BCa) percentile methods (Shao and Tu (1995)). We use the `bootci` function of the MATLAB Statistical Toolbox to construct BCa intervals in our experiments. We set the upper confidence limits of one-sided $100(1-\alpha)\%$ upper confidence intervals to the upper limits of the corresponding two-sided confidence intervals with $(1-2\alpha)$ nominal coverage probability.

4.2 Bootstrap confidence regions for VaR and CTE

Davison and Hinkley (1997) suggest basing a joint bootstrap confidence region for a vector parameter y_0 on the quadratic form:

$$\mathcal{Q} = (Y - y_0)^\top \widehat{\mathbf{\Sigma}}^{-1}(Y - y_0)$$

where Y is an estimate of y_0 and $\widehat{\mathbf{\Sigma}}$ is the estimated covariance matrix of Y. When Y is approximately normal, $\mathcal{Q}$ will be approximately $\chi^2_{(2)}$. Its distribution can be assessed by bootstrapping instead.

As in Section 3, we let:

$$y_0 = \begin{pmatrix} \mathrm{VaR}_{1-p} \\ \mathrm{CTE}_{1-p} \end{pmatrix}, \quad Y = \begin{pmatrix} -V_{[kp]} \\ -\dfrac{1}{kp}\displaystyle\sum_{i=1}^{kp} V_{[i]} \end{pmatrix}$$

and $\widehat{\mathbf{\Sigma}}^{-1}$ be the influence function estimate of the covariance matrix of Y, as in Equations (6)–(8). We calculate:

$$\mathcal{Q}^b = (Y_b - Y)^\top \widehat{\mathbf{\Sigma}}_b^{-1}(Y_b - Y)$$

for each bootstrap sample $b = 1, \dots, B$, yielding an estimate Y_b of CTE_{1-p} and an estimated covariance matrix $\widehat{\mathbf{\Sigma}}_b^{-1}$. We denote the ordered bootstrap values as $\mathcal{Q}^b_{[1]} \leq \cdots \leq \mathcal{Q}^b_{[B]}$. Then a bootstrap confidence region for the vector parameter y_0 is the set:

$$\{y_0 \mid (Y - y_0)^\top \widehat{\mathbf{\Sigma}}^{-1}(Y - y_0) \leq \mathcal{Q}^b_{[B(1-\alpha)]}\} \tag{15}$$

which is similar to Equation (10) but with $\mathcal{Q}^b_{[B(1-\alpha)]}$ replacing $\chi^2_{(2),1-\alpha}$.

5 EMPIRICAL LIKELIHOOD

Empirical likelihood (EL) is a non-parametric method for hypothesis testing (and therefore for confidence region construction) that is similar to the usual parametric likelihood ratio approach, which rejects a hypothesis when its likelihood ratio is too low. The empirical likelihood ratio, instead of being constructed from a parametric family of distributions, considers the family $\mathcal{F}_k := \{F \mid F \ll F_k\}$ of discrete distributions absolutely continuous with respect to the empirical cumulative distribution F_k whose support equals the observed data points. Such a distribution $F \ll F_k$ puts weights (ie, probability mass) $w_1, \dots, w_k$ on order statistics $V_{[1]}, \dots, V_{[k]}$, where the weights must be non-negative and sum to 1. The empirical likelihood of F is $\prod_{i=1}^k w_i$ and the empirical likelihood ratio of F is defined as $R(F) := \prod_{i=1}^k (kw_i)$, since the maximum likelihood member of $\mathcal{F}_k$ is the empirical distribution, F_k, which has all weights equal to $1/k$ and thus has empirical likelihood k^{-k}.

Let $T(\cdot)$ be some statistical functional of the distribution F_0, where F_0 is the true distribution of portfolio profit V. The non-parametric maximum likelihood estimate of $T(F_0)$ is $T(F_k)$ and sets of the form:

$$\{T(F) \mid R(F) \geq r,\ F \in \mathcal{F}_k\} \tag{16}$$

can be used as confidence regions for $T(F_0)$, where r is chosen appropriately to obtain the right asymptotic coverage, as $k \to \infty$ (Owen (1998)).

In particular, the empirical likelihood confidence interval for VaR coincides with the binomial confidence interval of Section 2 (Owen (2001, Section 3.6)).

5.1 A non-parametric confidence region for VaR and CTE

DEFINITION 5.1 For any $0 < p < 1$, any value Q^p such that $\Pr(V \leq Q^p) \geq p$ and $\Pr(V \geq Q^p) \geq 1 - p$ is a p-quantile of F_0 (Owen (2001)).

We defined the 95% VaR of our portfolio as the 95% quantile of the loss given by $Q^{95\%}_{-V}$. Using the above definition, we see that this is equivalent to the negative of the 5% quantile of the profit, which is given by $-Q^{5\%}_V$. Then, the 95% CTE of our portfolio is $\mathbf{E}[-V \mid -V \geq Q^{95\%}_{-V}] = -\mathbf{E}[V \mid V \leq Q^{5\%}_V]$.

Our goal is to construct an empirical likelihood confidence region for $\mathrm{VaR}_{1-p} = -Q^p_V$ and $\mathrm{CTE}_{1-p} = -\mathbf{E}[V \mid V \leq Q^p_V]$ and to provide asymptotic coverage probability results for such confidence regions.

DEFINITION 5.2 The lower and upper p-quantiles of any distribution F are defined as $\nu_p := \inf\{v \mid F(v) \geq p\}$ and $\nu^p := \inf\{v \mid F(v) > p\}$, respectively (Acerbi and Tasche (2002)).

DEFINITION 5.3 The ES at level $1 - p$ of V is defined as:

$$\mathrm{ES}_{1-p} := -p^{-1}(\mathbf{E}[V\mathbf{1}_{\{V \leq \nu_p\}}] + \nu_p(p - \Pr[V \leq \nu_p]))$$

where ν_p is the lower p-quantile of the distribution of V (Acerbi and Tasche (2002)).

Because it is not, in general, uniquely defined, it is not possible to write Q^p_V of Definition 5.1 as a statistical functional $T(F_0)$. This poses a problem for constructing confidence regions of the form (16). However, if F_0 is continuous and strictly increasing at Q^p_V, then Q^p_V is unique and is equal to ν_p. Furthermore, $\Pr(V \leq Q^p_V) = \Pr(V \leq \nu_p) = p$, which by Definition 5.3 implies $\mathrm{ES}_{1-p} = \mathrm{CTE}_{1-p}$. Under this simple restriction on F_0, the empirical likelihood results for M-estimates (Owen (1990)) can be used to construct empirical likelihood confidence regions for VaR_{1-p} and ES_{1-p}.

DEFINITION 5.4 An M-estimate is a statistical functional defined as a root $t = T_\psi(F)$ of:

$$\int \psi(V, t)F(\mathrm{d}V) = 0 \tag{17}$$

where $V \sim F$ (Owen (1990)).

PROPOSITION 5.1 *If F_0 is continuous and strictly increasing at its p-quantile, the functional T_ψ defined by Equation (17) is an M-estimate for the vector* $(\mathrm{VaR}_{1-p}, \mathrm{ES}_{1-p})$ *where the function $\psi : \mathcal{R} \times \mathcal{R}^2 \to \mathcal{R}^2$ is given by:*

$$\psi(V, (q, \mu)) := \left(p - \mathbf{1}_{\{V \leq -q\}}, \mu + \frac{1}{p}V\mathbf{1}_{\{V \leq -q\}}\right)$$

PROOF The unique root of Equation (17) with $F = F_0$ is $(\mathrm{VaR}_{1-p}, \mathrm{ES}_{1-p})$, as follows. First, $\int (p - \mathbf{1}_{\{V \le -q\}}) F_0(\mathrm{d}V) = p - F_0(-q) = 0$ which implies $F_0(-q) = p$. Since we assumed F_0 has a unique p-quantile with $F_0(Q_V^p) = p$, we find $Q_V^p = -q$ and $\mathrm{VaR}_{1-p} = -Q_V^p = q$. Second:

$$\int \left(\mu + \frac{1}{p} V \mathbf{1}_{\{V \le -q\}} \right) F_0(\mathrm{d}V) = \mu + \frac{1}{p} \int_{-\infty}^{-q} V F_0(\mathrm{d}V) = 0$$

which implies $\mu = -\mathbf{E}[V \mid V \le -q]$. Again by uniqueness of $Q_V^p = \nu_p = -q$ and therefore of $\mathrm{ES}_{1-p} = -\mathbf{E}[V \mid V \le \upsilon_p]$, we obtain $\mu = \mathrm{ES}_{1-p}$. □

Note that for $\psi(V, (q, \mu))$ defined as in Proposition 5.1 and $T_\psi(F)$ defined as in Definition 5.4, the set $\{T_\psi(F) \mid F \ll F_k,\, R(F) \ge r\}$ equals the confidence region $\{(q, \mu) \mid \int \psi(V, (q, \mu)) F(\mathrm{d}V) = 0,\, F \ll F_k,\, R(F) \ge r)\}$ for VaR_{1-p} and ES_{1-p} depicted in Figure 1.

PROPOSITION 5.2 *For ψ defined as in Proposition 5.1, if F_0 is continuous and strictly increasing at its p-quantile, and if $V\mathbf{1}_{\{V \le Q_V^p\}}$ is not a constant and $\mathbf{E}[V^2 \mathbf{1}_{\{V \le Q_V^p\}}] < \infty$, then $\{T_\psi(F) \mid F \ll F_k,\, R(F) \ge \exp(-\frac{1}{2}\chi^2_{(2),1-\alpha})\}$ is a confidence region for VaR_{1-p} and CTE_{1-p} with $(1-\alpha)$ asymptotic coverage probability.*

PROOF By Proposition 5.1, $T_\psi(F_0)$ exists and is unique if F_0 is continuous and increasing at ν_p, which we have already assumed. The assumption that $V\mathbf{1}_{\{V \le Q_V^p\}}$ is not a constant and $\mathbf{E}[V^2 \mathbf{1}_{\{V \le Q_V^p\}}] < \infty$ implies that the rank of $\mathbf{Var}[\psi(V, t)]$ is two. Then, we can use Theorem 3 of Owen (1990) to show that $\Pr[T_\psi(F_0) \notin \{T_\psi(F) \mid F \ll F_k,\, R(F) \ge r\}] \to \alpha$ as $k \to \infty$ if we pick $r = \exp(-\frac{1}{2}\chi^2_{(2),1-\alpha})$. □

While computing $\{T_\psi(F) \mid F \ll F_k,\, R(F) \ge r\}$, we must restrict our attention to F within the family $\mathcal{F}_k$ such that for some l, W_l defined by:

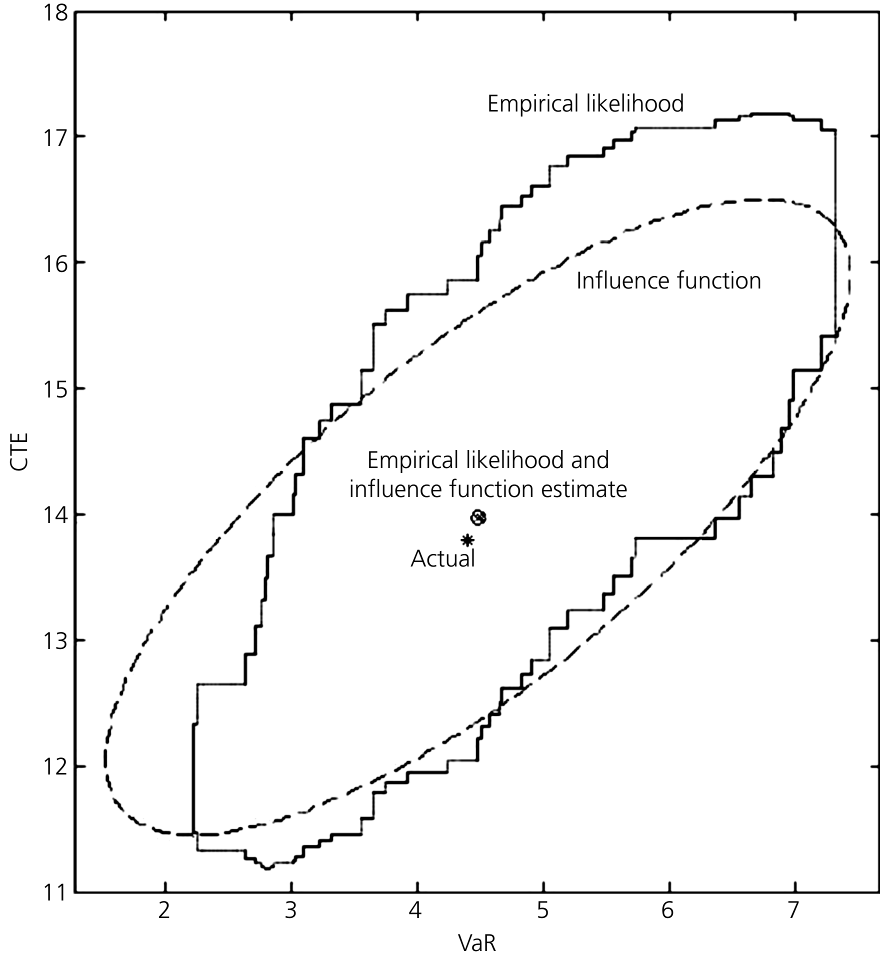
$$W_l := \sum_{i=1}^{l} w_i \tag{18}$$

is equal to p. Otherwise, $T_\psi(F)$ does not exist. It is worth noting that $T_\psi(F)$ is not unique for such $F \in \mathcal{F}_k$ since for any q such that $-q \in [V_{[l]}, V_{[l+1]})$, $(q, -(1/p) \sum_{i=1}^{l} w_i V_{[i]})$ is a root of $\int \psi(V, t)\, \mathrm{d}F(V) = 0$; however, we require only $T_\psi(F_0)$ to be unique.

A confidence region with $(1-\alpha)$ asymptotic coverage probability can be written as:

$$CR_{1-\alpha} = \left\{ t \;\middle|\; \int \psi(V, t)\, \mathrm{d}F(V) = 0,\, F \ll F_k,\, R(F) \ge r \right\}$$

FIGURE 1 Influence function and empirical likelihood confidence regions.



$$
\begin{aligned}
&= \bigcup_{l=1}^{k-1} \left\{ \mu \;\middle|\; \prod_{i=1}^{k} (kw_i) \geq r,\ \sum_{i=1}^{l} w_i = p,\ \mu = -\frac{1}{p} \sum_{i=1}^{l} V_{[i]} w_i,\ w_i \geq 0,\ \sum_{i=1}^{k} w_i = 1 \right\} \\
&\qquad \times (-V_{[l+1]}, -V_{[l]}] \\
&= \bigcup_{l=1}^{k-1} (-V_{[l+1]}, -V_{[l]}] \times \{\mu \mid R_l^{\psi}(\mu) \geq r\}
\end{aligned}
$$

where:

$$
R_l^{\psi}(\mu) := \max_w \left\{ \prod_{i=1}^{k} (kw_i) \;\middle|\; \sum_{i=1}^{l} w_i = p,\ \mu = -\frac{1}{p} \sum_{i=1}^{l} V_{[i]} w_i,\ w_i \geq 0,\ \sum_{i=1}^{k} w_i = 1 \right\}
$$

In the first part of Appendix A we show that this maximum is attained at $\{w_i^*\}_{i=1,\dots,k}$ given by:

$$w_i^* = \begin{cases} \dfrac{p}{l}[1-(V_{[i]}+\mu)\lambda^*]^{-1} & \text{for } i = 1, \dots, l \\ \dfrac{1-p}{k-l} & \text{for } i = l+1, \dots, k \end{cases} \tag{19}$$

where λ^* is the unique solution to:

$$\sum_{i=1}^{l} \frac{V_{[i]}+\mu}{1-(V_{[i]}+\mu)\lambda^*} = 0$$

which can be computed by numerical root finding within the interval:

$$\left[\frac{1-1/l}{\mu+V_{[1]}}, \frac{1-1/l}{\mu+V_{[l]}}\right]$$

By Lemma A.2 in Appendix A, for each l, $R_l^{\psi}(\mu)$ is single peaked at $-(1/l)\sum_{i=1}^{l} V_{[i]}$ and continuous and monotone on either side of this peak. This implies that $I_l^{\psi} := \{\mu \mid R_l^{\psi}(\mu) \geq r\}$ is an interval if it is not empty. Because $\max_{\mu} R_l^{\psi}(\mu) = k^k (p/l)^l [(1-p)/(k-l)]^{k-l}$, I_l^{ψ} is non-empty if:

$$k \log k + l \log \frac{p}{l} + (k-l) \log \frac{1-p}{k-l} \geq -\frac{1}{2}\chi^2_{(2),1-\alpha}$$

Therefore $I_l^{\psi} \subseteq [-V_{[l]}, V_{[1]}]$ can be computed as $I_l^{\psi} = [\mu_l^{lo}, \mu_l^{hi}]$ where:

$$\mu_l^{lo} \text{ is the unique root of } R_l^{\psi}(\mu) = r \text{ in } \left[-V_{[l]}, -\frac{1}{l}\sum_{i=1}^{l} V_{[i]}\right]$$

and:

$$\mu_l^{hi} \text{ is the unique root of } R_l^{\psi}(\mu) = r \text{ in } \left[-\frac{1}{l}\sum_{i=1}^{l} V_{[i]}, -V_{[1]}\right]$$

Finally, we compute $CR_{1-\alpha} = \bigcup_{l=1}^{k-1} (-V_{[l+1]}, -V_{[l]}] \times [\mu_l^{lo}, \mu_l^{hi}]$. Figure 1 compares the shape of such a confidence region to the shape of a confidence region constructed by the influence function approach.

5.2 Non-parametric confidence intervals for ES

Complications arise when we try to compute a confidence interval for CTE even if we restrict our attention to continuous distributions for which CTE is coherent. This is because $\psi(V, (q, \mu))$ is a non-smooth function of (q, μ) and hence theoretical justification is lacking to profile out either component of (q, μ) to obtain a confidence interval for the other. We, therefore, turn our attention to ES for which we can use empirical likelihood theory to compute an asymptotically valid confidence interval. Note that ES is still coherent even if the profit distribution F_0 is not continuous or strictly increasing at Q^p.

Empirical likelihood most naturally produces two-sided confidence intervals, and we will focus on these in this section. We produce one-sided confidence intervals according to the following suggestion of Owen (2001, Section 2.7). Where (L, U) is a two-sided $100(1-2\alpha)\%$ empirical likelihood confidence interval, $(-\infty, U)$ can be used as a one-sided $100(1-\alpha)\%$ confidence interval.

Acerbi and Tasche (2002) show that ES_{1-p} of Definition 5.3 can be represented as a functional T by $T(F_0) := -(1/p)\int_0^p F_0^{-1}(u)\,\mathrm{d}u$, where $F_0^{-1}(u) := \inf\{v \mid F_0(v) \geq u\}$. The empirical likelihood ratio of the hypothesis $\mu = T(F_0)$ is defined as:

$$R(\mu) := \max\left\{\prod_{i=1}^{k}(kw_i) \;\middle|\; T(F) = \mu,\ F \ll F_k\right\}$$

where F has weights $\{w_i\}_{i=1,\ldots,k}$ and with W defined as in Equation (18):

$$\begin{aligned} T(F) &= -\frac{1}{p}\left\{\sum_{i=1}^{l-1}\int_{W_{i-1}}^{W_i} V_{[i]}\,\mathrm{d}u + \int_{W_{l-1}}^{p} V_{[l]}\,\mathrm{d}u\right\} \\ &= -\frac{1}{p}\left\{\sum_{i=1}^{l-1} w_i V_{[i]} + (p - W_{l-1})V_{[l]}\right\} \end{aligned}$$

with l determined by $W_l \geq p$ and $W_{l-1} < p$.

Proposition 5.3 *If* $|F_0^{-1}(u)|$ *is* $O(u^{-\frac{1}{3}+\epsilon})$ *as* $u \to 0$, *for some* $\epsilon > 0$, *then a confidence interval for* ES_{1-p} *with* $100(1-\alpha)\%$ *asymptotic coverage probability is:*

$$\{\mu \mid F \ll F_k,\ R(\mu) \geq \exp(-\tfrac{1}{2}\chi^2_{(1),1-\alpha})\}$$

Proof We start by writing $\mathrm{ES}_{1-p} = T(F_0) = \int_0^1 F_0^{-1}(u)g(u)\,\mathrm{d}u$, where $g(u) = -(1/p)\mathbf{1}_{\{u\leq p\}}$. Note that $T(F)$ produces an L-estimator when we plug in the cumulative distribution function F_k for F. According to Theorem 10.2 of Owen (2001):

$$\Pr[T(F_0) \notin \{\mu \mid F \ll F_k,\ R(\mu) \geq \exp(-\tfrac{1}{2}\chi^2_{(1),1-\alpha})\}] \to \alpha$$

as $k \to \infty$ if for some $c \in (0, \infty)$, some $M \in (0, \infty)$ and some $d \in (1/6, 1/2)$, both:

$$|g(u)| \leq M[u(1-u)]^{1/c-1/2+d} \quad \text{and} \quad |F_0^{-1}(u)| \leq M[u(1-u)]^{-1/c}$$

hold for all $0 < u < 1$. In our case, $g(u) = 0$ for $u > p$, so only the left tail behavior is relevant. That is, we are only concerned with the behavior of F_0^{-1} as $u \to 0$, because our L-estimator uses only values less than the median of the data sample. We will show that:

$$|g(u)| \leq Mu^{1/c-1/2+d} \tag{20}$$

and:

$$|F_0^{-1}(u)| \leq M u^{-1/c} \tag{21}$$

hold for suitable values of c, d and M, given the assumption that $F_0^{-1}(u)$ is $O(u^{-1/3+\epsilon})$ as $u \to 0$.

The interesting case is when losses are unbounded, in which case $\epsilon < 1/3$. Take $c = 1/(1/3 - \epsilon)$ and $d = 1/6 + \epsilon$. Then inequality (21) holds for sufficiently large M by assumption and inequality (20) holds for $M \geq 1/p$ because $1/c - 1/2 + d = 0$ and $|g(u)| = 1/p$ for $u < p$.

If losses are bounded, take M to be the maximum of the bound and $1/p$. Take $c = 3$ and $d = 1/3$. Then inequality (21) holds because $|F_0^{-1}(u)| \leq M \leq M u^{-1/3}$ and inequality (20) holds because $|g(u)| \leq M \leq M u^{-1/6}$. □

We compute $R(\mu)$ by $R(\mu) = \max_{l=1,\dots,k} R_l(\mu)$ where:

$$\begin{aligned} R_l(\mu) &= \sup\left\{ \prod_{i=1}^{k} (k w_i) \;\middle|\; \mu = T(F),\, W_l \geq p,\, W_{l-1} < p,\, W_k = 1,\, w_i \geq 0 \right\} \\ &= \max\{R_l^{\psi}(\mu),\, R_l^{int}(\mu)\} \end{aligned}$$

and $R_l^{int}(\mu)$ is defined as:

$$R_l^{int}(\mu) := \sup\left\{ \prod_{i=1}^{k} (k w_i) \;\middle|\; \mu = T(F),\, W_l > p,\, W_{l-1} < p,\, W_k = 1,\, w_i \geq 0 \right\}$$

We observe that:

$$R_l^{\psi}(\mu) = \max\left\{ \prod_{i=1}^{k} (k w_i) \;\middle|\; \mu = T(F),\, W_l = p,\, W_k = 1,\, w_i \geq 0 \right\}$$

is as defined in the previous section because for $W_l = p$, we obtain $T(F) = -(1/p) \sum_{i=1}^{l} w_i V_{[i]}$ with $W_{l-1} < p$, optimally. As $W_l \to p$ and as $W_{l-1} \to p$, limits of feasible points in this maximization converge to feasible points in the maximizations of the previous section whose optimal values are, respectively, $R_l^{\psi}(\mu)$ and $R_{l-1}^{\psi}(\mu)$. This reasoning shows that:

$$\begin{aligned} R(\mu) &= \max_{l=1,\dots,k} \{\max\{R_l^{\psi}(\mu),\, R_l^{int}(\mu)\}\} \\ &= \max\left\{ \max_{l=1,\dots,k} R_l^{\psi}(\mu),\; \max_{l \in L^{int}(\mu)} R_l^{int}(\mu) \right\} \end{aligned} \tag{22}$$

where $l \in L^{int}(\mu)$ if and only if $R_l^{int}(\mu)$ is attained at an interior solution characterized by $W_l > p$ and $W_{l-1} < p$, since otherwise $R_l^{int}(\mu) = R_l^{\psi}(\mu)$ or $R_{l-1}^{\psi}(\mu)$.

Since we have already found a way to compute $R_l^{\psi}(\mu)$, we need only concern ourselves with interior solutions $R_l^{int}(\mu)$ with $l \in L^{int}(\mu)$ to the following problem:

$$\begin{aligned} \text{maximize} \quad & \prod_{i=1}^{k}(kw_i) \\ \text{subject to} \quad & \mu = -V_{[l]} - \frac{1}{p}\sum_{i=1}^{l-1} w_i(V_{[i]} - V_{[l]}) \\ & W_l > p \text{ and } W_{l-1} < p \\ & W_k = 1 \text{ and } w_i \geq 0 \end{aligned} \tag{23}$$

which we will refer to as Maximization Problem II. It is maximization of a concave objective with linear constraints and non-zero Hessian, so there is an interior solution if and only if there is a solution to the two first-order conditions in two unknowns, which are:

$$W_{l-1}^* - \sum_{i=1}^{l-1} g_i(W_{l-1}^*, \lambda^*) = 0 \tag{24}$$

and:

$$\sum_{i=1}^{l-1} g_i(W_{l-1}^*, \lambda^*)(V_{[l]} - V_{[i]}) - p(\mu + V_{[l]}) = 0 \tag{25}$$

where g_i is a function specifying the optimal weight w_i for $i = 1, \ldots, l-1$. In Appendix B, we show that:

$$g_i(W_{l-1}^*, \lambda^*) := \left[\frac{k-l+1}{1-W_{l-1}^*} + \lambda^*(V_{[l]} - V_{[i]})\right]^{-1}$$

so the optimal weights are:

$$w_i^* = \begin{cases} \left[\dfrac{k-l+1}{1-W_{l-1}^*} + \lambda^*(V_{[l]} - V_{[i]})\right]^{-1} & \text{for } i = 1, \ldots, l-1 \\ \dfrac{1-W_{l-1}^*}{k-l+1} & \text{for } i = l, \ldots, k \end{cases} \tag{26}$$

where λ^* and $W_{l-1}^* \in (p - [(1-p)/(k-l)], p)$ solve the first-order conditions.

We construct a confidence interval with $100(1-\alpha)\%$ asymptotic coverage probability for ES as $CI_{1-\alpha} := \{\mu \mid F \ll F_k, R(\mu) \geq r\}$. By Equation (22), μ is in $CI_{1-\alpha}$ if and only if $R_l^{\psi} \geq r$ for some l or $R_l^{int}(\mu) \geq r$ for some $l \in L^{int}(\mu)$. Then, $CI_{1-\alpha}$ can be computed as:

$$CI_{1-\alpha} = \left(\bigcup_{l=1}^{k} I_l^{\psi}\right) \cup \left(\bigcup_{l=1}^{k} I_l^{int}\right)$$

where we define:

$$I_l^{int} := \{\mu \mid l \in L^{int}(\mu), R_l^{int}(\mu) \geq r\} = \{\mu \mid \mu \in M_l^{int}, R_l^{int}(\mu) \geq r\}$$

and $M_l^{int} = \{\mu \mid l \in L^{int}(\mu)\}$ is the set of μ such that Equations (24) and (25) have an interior solution. We show by Lemma B.2 of Appendix B that M_l^{int} is an open interval whose lower endpoint m_l^{lo} satisfies Equations (24) and (25) with $W_{l-1}^* = p - [(1-p)/(k-l)]$ and whose upper endpoint m_l^{hi} satisfies Equations (24) and (25) with $W_{l-1}^* = p$.

We have already shown how to calculate I_l^{ψ} in Section 5.1 and it remains to compute I_l^{int}. By definition, I_l^{int} is a subset of $M_l^{int} = (m_l^{lo}, m_l^{hi})$, where m_l^{lo} and m_l^{hi} can be found by solving Equation (24) for λ^* with $W_{l-1}^* = p - [(1-p)/(k-l)]$ and $W_{l-1}^* = p$, respectively, and then by solving Equation (25) for μ with these W_{l-1}^* and λ^*. Continuity of R_l^{int} and Lemma B.3 of Appendix B justify the following procedure:

1) If $l \leq kp$: if $R_l^{int}(m_l^{lo}) < r$, then I_l^{int} is empty. Otherwise, the lower endpoint of I_i^l is m_l^{lo} and the upper endpoint of I_l^{int} is the root of $R_l^{int}(\mu) - r = 0$ on (m_l^{lo}, m_l^{hi}).
2) If $kp < l < kp + 1$: the roots of $R_l^{int}(\mu) - r = 0$ on $(m_l^{lo}, T(F_k))$ and $(T(F_k), m_l^{hi})$ are the lower and upper endpoints of I_l^{int}.
3) If $l \geq kp + 1$: if $R_l^{int}(m_l^{hi}) < r$, then I_l^{int} is empty. Otherwise, the upper endpoint of I_i^l is m_l^{hi} and the lower endpoint of I_l^{int} is the root of $R_l^{int}(\mu) - r = 0$ on (m_l^{lo}, m_l^{hi}).

Finally, since both I_l^{int} and I_l^{ψ} are intervals, we compute:

$$CI_{1-\alpha} = \left(\bigcup_{l=1}^{k} I_l^{\psi}\right) \cup \left(\bigcup_{l=1}^{k} I_l^{int}\right) = [\mu^{lo}, \mu^{hi}]$$

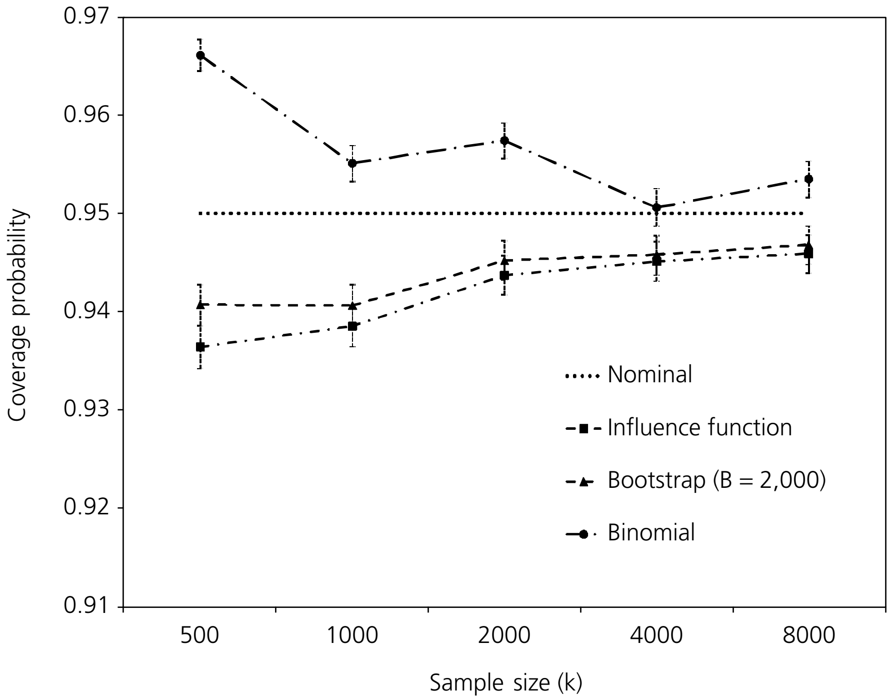
by setting μ^{hi} equal to the maximum of the upper endpoints of I_l^{ψ} and of I_l^{int} and likewise by setting μ^{lo} equal to the minimum of the lower endpoints of I_l^{ψ} and I_l^{int}.

6 EXPERIMENTAL RESULTS

We use the following two examples from Manistre and Hancock (2005) to test the performance of our confidence intervals and regions.

1) *Put option:* the owner of the portfolio has issued an in-the-money European put option and we use Monte Carlo simulation to estimate risk measures of this simple portfolio. The put option matures in 10 years with a strike price of \$110. The current stock price is \$100 and is assumed to follow a lognormal return process with drift 8% and volatility 15%. The continuous discount rate is 6%.
2) *Pareto distribution:* the loss is assumed to have a Pareto distribution, whose tail behavior is similar to that observed in some applications. The Pareto distribution is tractable enough for obtaining closed form expressions for the variance of the CTE estimator. We use Monte Carlo simulation to estimate risk measures for losses generated by a heavy-tailed Pareto distribution with shape and scale parameters set to 2.5 and 25, respectively.

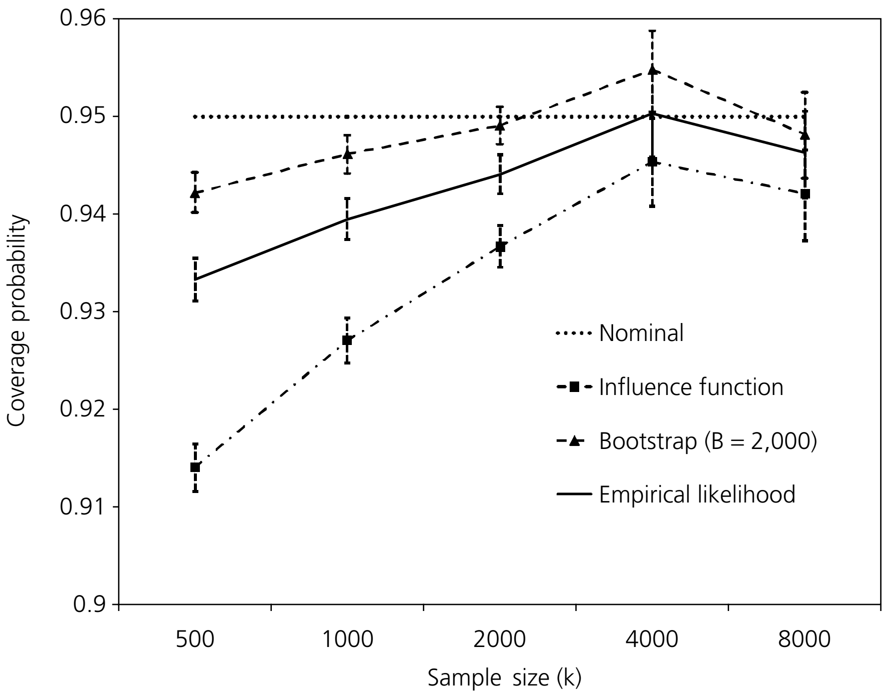
FIGURE 2 Coverage probability of one-sided confidence intervals for VaR: put option; $N = 50{,}000$ macroreplications.



These are simple examples, but the results should be indicative of the coverage we would expect these procedures to provide for similar, larger examples. The simulations reported here do not use variance reduction. It is not straightforward to combine variance reduction techniques, such as those applied to this problem by Manistre and Hancock (2005), with the methods for constructing confidence intervals and regions.

To evaluate the procedures for generating confidence intervals and regions, we run each of them 10,000 or 50,000 times. Each of these N macroreplications contains k simulated losses, where the sample size k is 500, 1,000, 2,000 or more in the experiments whose results are depicted in Figures 2–7. From each macroreplication, we calculate one-sided and two-sided confidence intervals for $ES_{0.95}$ and confidence regions for $VaR_{0.95}$ and $CTE_{0.95}$ at a nominal confidence level of 95% by the influence function, bootstrap and empirical likelihood methods. We also calculate one-sided confidence intervals for $VaR_{0.95}$ at a nominal confidence level of 95% by the binomial, influence function and bootstrap methods. The number of bootstrap samples B we use is either 2,000 or 10,000. We compute, for each sample

FIGURE 3 Coverage probability of one-sided confidence intervals for ES: put option; $N = 50{,}000$ macroreplications.



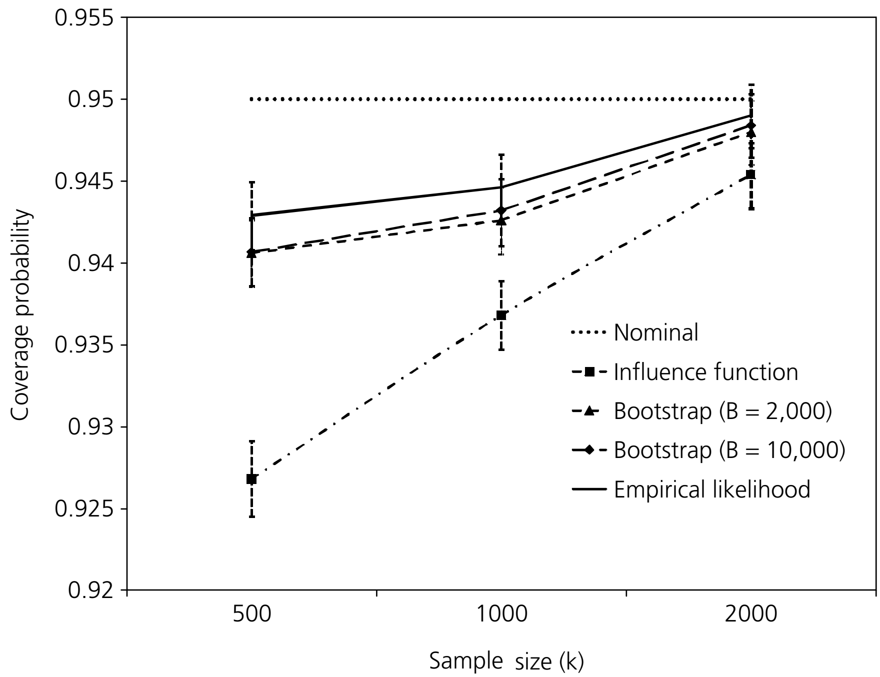
size k, the observed coverage probabilities of confidence intervals or regions:

$$(1 - \hat{\alpha}) := \#\{\text{confidence intervals or regions that include the true value}\}/N$$

where the true values are computed according to the formulas given by Manistre and Hancock (2005). The coverage results for confidence intervals and regions are summarized in Figures 2–7. The error bars in these figures represent 95% binomial confidence intervals for coverage probabilities based on observing N macroreplications, each of which is a success if the true value is included, a failure otherwise.

We first consider the example of selling a put option in the Black–Scholes model. We examine one-sided confidence intervals for VaR in Figure 2 to see how the methods under consideration differ in the well-studied setting of quantile estimation. As has been documented by Agresti and Coull (1998), the one-sided binomial confidence interval show modest overcoverage for sample sizes between 500 and 2,000. The bootstrap and influence function methods show modest undercoverage, but attain coverage above 94% by sample size 4,000. Bootstrapping is slightly better than the influence function method at small sample sizes.

FIGURE 4 Coverage probability of two-sided confidence intervals for ES: put option; $N = 50{,}000$ macroreplications.



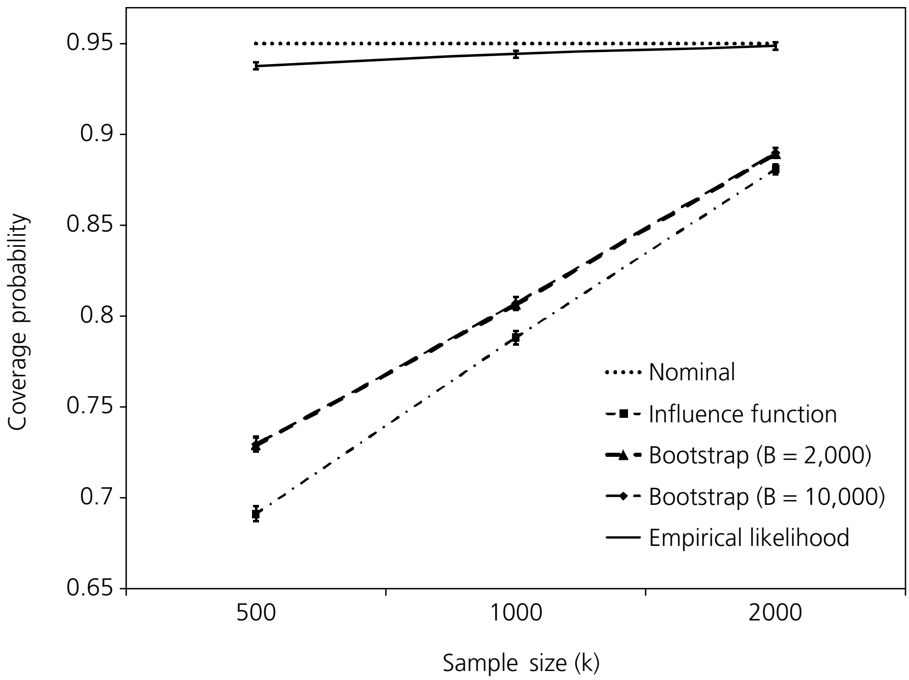
In Figure 3 we turn to one-sided confidence intervals for ES. Again bootstrapping shows modest undercoverage, but for ES it attains nominal coverage by sample size 4,000. Empirical likelihood provides somewhat worse undercoverage until sample size 4,000. The influence function method has the worst undercoverage and has not attained nominal coverage even by sample size 8,000.

Figure 4 shows the coverage of two-sided confidence intervals for ES. The results are qualitatively similar to those for one-sided confidence intervals, but as usual, the two-sided confidence intervals have less undercoverage. Figures 4 and 5 also show that the bootstrap sample size $B = 2{,}000$ that we use elsewhere is adequate: the improvement in coverage created by using a bootstrap sample size of $B = 10{,}000$ is negligible.

In Figure 5 we investigate the coverage of the confidence regions for VaR and CTE. The empirical likelihood method attains nominal coverage by sample size 2,000, while the bootstrap and influence function methods produce disastrous undercoverage at these small sample sizes. We suspect that this deficiency is due to the difficulty of density estimation, resulting in poor covariance matrix estimates.

Figures 6 and 7 portray the results of experiments on the Pareto distribution example, which serve to illustrate how well the methods perform when the loss distribution's tail is heavy instead of light. We focus on one-sided confidence

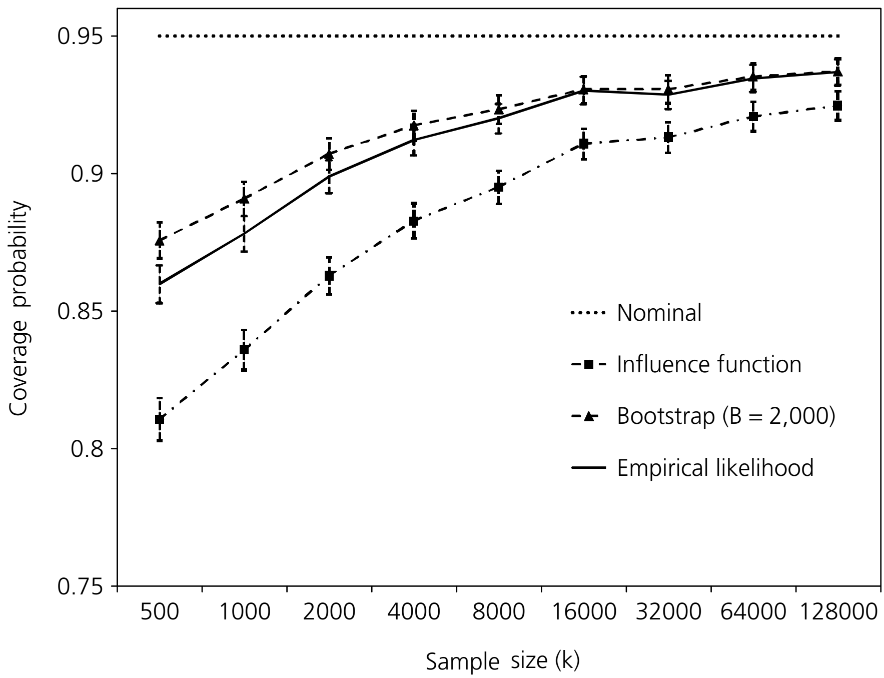
FIGURE 5 Coverage probability of confidence regions for VaR and CTE: put option; $N = 50{,}000$ macroreplications.



intervals for ES in this example. Figure 6 shows that this example is much more challenging. All the methods produce severe undercoverage at small sample sizes, where bootstrapping is slightly better than empirical likelihood, which is in turn much better than the influence function method. At large sample sizes, bootstrapping and empirical likelihood perform similarly. They still undercover somewhat even at a sample size of $k = 128{,}000$, but they are greatly superior to the influence function method.

Considering that confidence intervals fail to produce nearly nominal coverage even for very large sample sizes when the distribution is heavy-tailed, we investigate empirically how quickly the coverage rate converges to the nominal level. Figure 7 is a log–log plot of coverage error, defined as the absolute value of the difference between observed coverage and nominal coverage $|\hat{\alpha} - \alpha|$ against sample size k. For each method, the slope of the curve indicates how quickly the coverage rate converges to the nominal level. For example, the coverage error for one-sided confidence intervals is typically $\mathcal{O}(k^{-1/2})$ when produced by empirical likelihood (Owen (2001, Section 2.7)) and $\mathcal{O}(k^{-1})$ when produced by the BCa bootstrapping method (Owen (2001, Section A.6)). This implies that on a log–log plot of coverage error versus sample size, these methods should yield curves whose

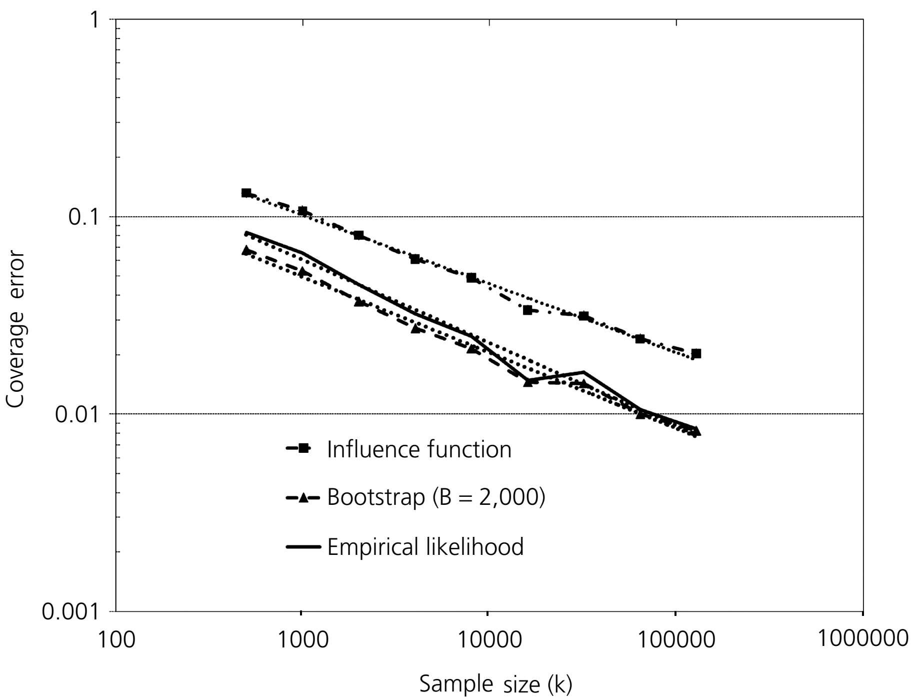
FIGURE 6 Coverage probability of one-sided confidence intervals for ES: Pareto distribution; $N = 10{,}000$ macroreplications.



slopes approach -0.5 and -1, respectively, for large sample size. It is possible to correct empirical likelihood one-sided confidence intervals so that their coverage error is also $\mathcal{O}(k^{-1})$ (Owen (2001, Chapter 13)).

However, far from finding that BCa bootstrapping dominates empirical likelihood asymptotically, we found that as sample size increases, the empirical likelihood method catches up with bootstrapping. Also, the influence function method becomes increasingly uncompetitive. We can see this in Figure 7, where we estimated slopes on the log–log plot of coverage error versus sample size of -0.34 for the influence function method, -0.42 for the empirical likelihood method and -0.38 for bootstrapping, over a range of sample sizes from 500 to 128,000. The slope of -0.42 for empirical likelihood is not too far from the theoretical asymptotic slope of -0.5, but the slope of -0.38 for BCa bootstrapping is far from the typical theoretical asymptotic slope of -1. Of course, for finite sample sizes, the slope may differ from the asymptotic slope as sample size goes to infinity. We conjecture that there is another reason that the slope is far from -1 in Figure 7 for the coverage error of the BCa bootstrap one-sided confidence interval. In this example, the loss distribution is extremely heavy-tailed: the Pareto distribution with shape parameter 2.5 has first and second moments, but no third moment. Because

FIGURE 7 Log–log plot of coverage error of one-sided 95% confidence intervals for ES versus sample size: Pareto distribution; $N = 10{,}000$ macroreplications.



the BCa method is based on a skewness correction, one would not expect it to work if skewness does not exist.

7 CONCLUSIONS AND FUTURE RESEARCH

Based on empirical likelihood, we have developed an asymptotically valid confidence interval for ES and confidence region for VaR and CTE. In Monte Carlo experiments, we found that they have coverage close to nominal for moderate sample sizes: about 1,000 samples in a financial example in which losses are light-tailed and somewhat more in an example in which the loss distribution is Pareto. The confidence interval based on empirical likelihood performed about as well as one based on bootstrapping and better than one based on the influence function. The confidence region based on empirical likelihood performed better than both its competitors.

The confidence intervals and regions discussed here are based on the most straightforward point estimators of VaR and CTE or ES. The most straightforward point estimator of VaR, which is a quantile, is a sample quantile. There is a large literature on quantile estimation which shows that more complicated estimators, such as kernel estimators and the Harrell–Davis estimator, can outperform the sample quantile (Chang *et al* (2003); Sheather and Marron (1990)).

In this study, we have applied the basic version of empirical likelihood, but more advanced versions could be applied to the same problem. Methods such as Bartlett correction can improve the coverage of empirical likelihood confidence intervals (Owen (2001, Chapter 13)). It has been found that smoothed or adjusted empirical likelihood methods can produce confidence intervals for quantiles with improved coverage (Chen and Hall (1993); Zhou and Jing (2003)). It is also possible to apply data tilting methods, which are generalizations of empirical likelihood, to construct confidence intervals for quantiles. Peng and Qi (2006) do this for extreme quantiles by explicitly estimating the tail index of the loss distribution. This method may also be applied to CTE or ES.

As suggested by Dowd (2005), the techniques described here could be applied to any spectral measure of risk (Acerbi and Tasche (2002)) as well as to ES. Another direction for future research is to show how to construct confidence intervals and regions when variance reduction techniques are used in the Monte Carlo sampling. This would yield smaller confidence intervals and regions given the same amount of computational effort.

APPENDIX A MAXIMIZATION PROBLEM I

The problem of computing $R_l^{\psi}(\mu)$ given by:

$$R_l^{\psi}(\mu) = \max\left\{\prod_{i=1}^{k}(kw_i) \;\middle|\; \sum_{i=1}^{l} w_i = p,\ \mu = -\frac{1}{p}\sum_{i=1}^{l} V_{[i]}w_i,\ w_i \geq 0,\ \sum_{i=1}^{k} w_i = 1\right\} \tag{A.1}$$

reduces to solving the following problem referred to as Maximization Problem I:

$$\begin{aligned} \text{maximize} \quad & \sum_{i=1}^{l} \log(kw_i) + (k-l)\log\left(k\frac{1-p}{k-l}\right) \\ \text{subject to} \quad & -\frac{1}{p}\sum_{i=1}^{l} w_i V_{[i]} = \mu \\ & \sum_{i=1}^{l} w_i = p \end{aligned} \tag{A.2}$$

since $W_l = \sum_{i=1}^{l} w_i$ is restricted to be exactly equal to p by (A.1) and in this case $R_l^{\psi}(\mu)$ is achieved by assigning equal weights to the remaining $k-l$ portfolio values $V_{[l+1]}, \ldots, V_{[k]}$.

Note that the first equation in (A.2) can be written as $\sum_{i=1}^{l} w_i(\mu + V_{[i]}) = 0$ by $p\mu + \sum_{i=1}^{l} w_i V_{[i]} = \left(\sum_{i=1}^{l} w_i\right)\mu + \sum_{i=1}^{l} w_i V_{[i]}$.

Since a strictly concave function is maximized on a linear set of equality constraints, the solution to this maximization problem will be found by using the

Lagrangian function:

$$\mathcal{L} = \sum_{i=1}^{l} \log(kw_i) + (k-l)\log\left(k\frac{1-p}{k-l}\right) + \frac{l}{p}\lambda\sum_{i=1}^{l} w_i(V_{[i]}+\mu) + \gamma\left(\sum_{i=1}^{l} w_i - p\right)$$

and the first-order conditions:

$$\frac{\partial\mathcal{L}}{\partial w_i^*} = \frac{1}{w_i^*} + \frac{l}{p}\lambda^*(V_{[i]}+\mu) + \gamma^* = 0 \quad \forall i = 1, \dots, l \tag{A.3}$$

are sufficient.

Using Equations (A.3), we obtain:

$$\sum_{i=1}^{l} w_i^* \frac{\partial\mathcal{L}}{\partial w_i^*} = 0$$

which leads together with constraints in (A.2) to $l + 0 + \gamma^* p = 0$ and hence $\gamma^* = -l/p$. Plugging the value of γ^* back into Equations (A.3), we obtain:

$$w_i^* = \frac{p}{l}[1-(V_{[i]}+\mu)\lambda^*]^{-1} \quad \forall i = 1, \dots, l \tag{A.4}$$

Plugging the values of w_i^* calculated above into the first constraint in (A.2), we obtain:

$$\sum_{i=1}^{l} \frac{V_{[i]}+\mu}{1-(V_{[i]}+\mu)\lambda^*} = 0 \tag{A.5}$$

LEMMA A.1 *If $\mu \in (-V_{[l]}, -V_{[1]})$, then Equation (A.5) is satisfied for some:*

$$\lambda^* \in \left(\frac{l-1}{l(\mu+V_{[1]})}, \frac{l-1}{l(\mu+V_{[l]})}\right)$$

PROOF Define:

$$f_i(\mu,\lambda) := \frac{V_{[i]}+\mu}{1-(V_{[i]}+\mu)\lambda} \quad \text{and} \quad f_i^{\mu}(\lambda) := f_i(\mu,\lambda)$$

Each f_i^{μ} has one discontinuity at $(V_{[i]}+\mu)^{-1}$, where f_i^{μ} is not defined. Therefore $\sum_{i=1}^{l} f_i^{\mu}(\lambda)$ is continuous on $((V_{[1]}+\mu)^{-1}, (V_{[l]}+\mu)^{-1})$. Because the partial derivative:

$$\frac{\partial f_i}{\partial\lambda} = \frac{(V_{[i]}+\mu)^2}{[1-(V_{[i]}+\mu)\lambda]^2}$$

is positive unless $\mu = -V_{[i]}$, in which case it is zero, we can see that $\sum_{i=1}^{l} f_i^{\mu}(\lambda)$ is increasing in λ on $((V_{[1]}+\mu)^{-1}, (V_{[l]}+\mu)^{-1})$. Because:

$$\lim_{\lambda\uparrow(V_{[l]}+\mu)^{-1}} f_l^{\mu}(\lambda) = \infty \quad \text{and} \quad \lim_{\lambda\downarrow(V_{[1]}+\mu)^{-1}} f_1^{\mu}(\lambda) = -\infty$$

there exists $\lambda^* \in ((V_{[1]} + \mu)^{-1}, (V_{[l]} + \mu)^{-1})$ such that $\sum_{i=1}^{l} f_i^{\mu}(\lambda^*) = 0$.

In fact, we can find tighter bounds for λ^* than $(V_{[1]} + \mu)^{-1}$ and $(V_{[l]} + \mu)^{-1}$. Because:

$$w_1^* = \frac{p}{l}[1 - \lambda^*(V_{[1]} + \mu)]^{-1} \le p \quad \text{and} \quad V_{[1]} < -\mu \Rightarrow V_{[1]} + \mu < 0$$

we find:

$$\lambda^* > \frac{l-1}{l(\mu + V_{[1]})}$$

Likewise, using:

$$w_l^* = \frac{p}{l}[1 - \lambda(V_{[l]} + \mu)]^{-1} \le p \quad \text{and} \quad -\mu < V_{[l]} \Rightarrow V_{[l]} + \mu > 0$$

we find:

$$\lambda^* < \frac{l-1}{l(\mu + V_{[l]})}$$

□

LEMMA A.2 *R_l^{ψ} is increasing on $(-V_{[l]}, \mu_l^*)$ and decreasing on $(\mu_l^*, -V_{[1]})$, where μ_l^* is defined as $\mu_l^* := -(1/l)\sum_{i=1}^{l} V_{[i]}$.*

PROOF The empirical likelihood ratio R_l^{ψ} is maximized at μ_l^*, where the solution to Maximization Problem I with $\mu = \mu_l^*$ involves $\lambda^* = 0$, so the optimal weights are p/l for $i = 1, \dots, l$ and are $(1-p)/(k-l)$ for $i = l+1, \dots, k$. Consider some $\mu \in (-V_{[l]}, \mu_l^*)$. Let F_μ, with weights $\{w_i^{\mu}\}_{i=1,\dots,k}$, be the distribution at which $R_l^{\psi}(\mu)$ is attained. Because $\mu < \mu_l^*$, Equations (A.2) and (A.4) imply that $\lambda_\mu^* > 0$ at the solution to Maximization Problem I. This makes the optimal weights $\{w_i^{\mu}\}_{i=1,\dots,l}$ increasing in i. In the trivial case where $V_{[i]}$ is the same for all $i = 1, \dots, l$, the conclusion of the lemma holds; we henceforth assume that there exist $m < n \le l$ such that $V_{[m]} < V_{[n]}$. Because the weights are increasing, for some $\epsilon > 0$, $w_m^{\mu} = w_n^{\mu} - \epsilon$. For any $\mu' \in (\mu, \mu + (\epsilon/2)(V_{[n]} - V_{[m]}))$, let $\delta = (\mu' - \mu)/(V_{[n]} - V_{[m]})$. Construct F' with weights $\{w_i'\}_{i=1,\dots,k}$ such that $F' = F_\mu$ except $w_m' = w_m^{\mu} + \delta$ and $w_n' = w_n^{\mu} - \delta$. Because $\delta \in (0, \epsilon/2)$, $w_m' w_n' > w_m^{\mu} w_n^{\mu}$, so $R(F') > R(F_\mu)$. This leads to the conclusion $R_l^{\psi}(\mu') \ge R(F') > R(F_\mu) = R_l^{\psi}(\mu)$. We have therefore shown that for all $\mu \in (-V_{[l]}, \mu_l^*)$, R_l^{ψ} is increasing on a non-empty open interval whose left endpoint is μ, which in turn proves that R_l^{ψ} is increasing on $(-V_{[l]}, \mu_l^*)$. A similar analysis for $\mu > \mu_l^*$, involving $\lambda_\mu^* < 0$, proves that R_l^{ψ} is decreasing on $(\mu_l^*, -V_{[1]})$. □

APPENDIX B MAXIMIZATION PROBLEM II

Maximization Problem II is more complicated due to the inequality constraints for W_l and W_{l-1}. We write $w_l = p - W_{l-1} + \delta$ where $\delta > 0$ so that the constraint $W_l > p$ is satisfied. The ES constraint can be written as $p(\mu + V_{[l]}) = \sum_{i=1}^{l-1} w_i(V_{[l]} - V_{[i]})$. Since w_l does not appear in the modified ES constraint, $R_l^{int}(\mu)$ will be attained when w_l is as close as possible to $w_{l+1} = \cdots = w_k =$

$(1-W_{l-1})/(k-l+1)$ and this implies $W_{l-1} > p-(1-p)/(k-l)$ since δ is defined to be positive. Then, the problem of finding $R_l^{int}(\mu)$ reduces to solving the following maximization problem:

$$\begin{aligned}
\text{maximize} \quad & \sum_{i=1}^{l-1} \log(kw_i) + (k-l+1)\log\biggl(k\frac{1-W_{l-1}}{k-l+1}\biggr) \\
\text{subject to} \quad & \sum_{i=1}^{l-1} w_i(V_{[l]} - V_{[i]}) = p(\mu + V_{[l]}) \\
& W_{l-1} = \sum_{i=1}^{l-1} w_i \\
& W_{lb} < W_{l-1} < W_{ub}
\end{aligned} \tag{B.1}$$

where:

$$W_{lb} := p - \frac{1-p}{k-l} \quad \text{and} \quad W_{ub} := p$$

The Hessian of the above objective function is an l-dimensional diagonal matrix with:

$$\biggl\{-\frac{1}{w_i^2}\biggr\}_{i,\dots,l-1} \quad \text{and} \quad -\frac{1}{(1-W_{l-1})^2}$$

as the diagonal entries and therefore is negative definite. Since a concave function is maximized subject to linear constraints, there exists a unique global optimum for the above maximization problem. This maximum can be computed by using the Lagrangian function:

$$\begin{aligned}
\mathcal{L} = {} & \sum_{i=1}^{l-1} \log(kw_i) + (k-l+1)\log\biggl(k\frac{1-W_{l-1}}{k-l+1}\biggr) \\
& - \lambda\biggl(\sum_{i=1}^{l-1} w_i(V_{[l]} - V_{[i]}) - p(\mu + V_{[l]})\biggr) - \gamma\biggl(W_{l-1} - \sum_{i=1}^{l-1} w_i\biggr)
\end{aligned}$$

and the first-order conditions:

$$0 = \frac{\partial \mathcal{L}}{\partial w_i^*} = \frac{1}{w_i^*} - \lambda^*(V_{[l]} - V_{[i]}) + \gamma^*, \quad \forall i = 1, \dots, l-1 \tag{B.2}$$

$$0 = \frac{\partial \mathcal{L}}{\partial W_{l-1}^*} = -\frac{k-l+1}{1-W_{l-1}^*} - \gamma^* = 0 \tag{B.3}$$

are sufficient. Using Equations (B.2) and (B.3), we write:

$$w_i^* := g_i(W_{l-1}^*, \lambda^*) = \biggl[\frac{k-l+1}{1-W_{l-1}^*} + \lambda^*(V_{[l]} - V_{[i]})\biggr]^{-1}, \quad \forall i = 1, \dots, l-1 \tag{B.4}$$

and obtain the following system of non-linear equations in two unknowns:

$$W_{l-1}^* - \sum_{i=1}^{l-1} g_i(W_{l-1}^*, \lambda^*) = 0$$

$$\sum_{i=1}^{l-1} g_i(W_{l-1}^*, \lambda^*)(V_{[l]} - V_{[i]}) - p(\mu + V_{[l]}) = 0$$

LEMMA B.1 *The function g_i defined in Equation (19) is strictly decreasing in each of its arguments.*

PROOF The partial derivatives of g_i with respect to W_{l-1} and λ are:

$$\frac{\partial g_i(W_{l-1}, \lambda)}{\partial W_{l-1}} = -\frac{(k-l+1)w_i^2}{(1-W_{l-1})^2} \tag{B.5}$$

$$\frac{\partial g_i(W_{l-1}, \lambda)}{\partial \lambda} = -(V_{[l]} - V_{[i]})w_i^2 \tag{B.6}$$

which are negative everywhere. □

The following lemma provides bounds on μ for which $R_l^{int}(\mu)$ can be found by solving the first-order conditions for Maximization Problem II. Define M_l^{int} as a set which contains μ if and only if Equations (24) and (25) have a solution (W_{l-1}^*, λ^*) with $W_{l-1}^* \in (W_{lb}, W_{ub})$.

LEMMA B.2 *The set M_l^{int} is an interval (m_l^{lo}, m_l^{hi}) such that Equations (24) and (25) can be solved for $\mu = m_l^{lo}$ and $W_{l-1} = W_{lb}$ and for $\mu = m_l^{hi}$ and $W_{l-1} = W_{ub}$. If (W_{l-1}, λ) and $(\tilde{W}_{l-1}, \tilde{\lambda})$ satisfy the first-order conditions (24) and (25) for μ and $\tilde{\mu}$, respectively, while $W_{lb} \le W_{l-1} < \tilde{W}_{l-1} \le W_{ub}$, then $\tilde{\lambda} < \lambda$ and $\tilde{\mu} > \mu$.*

PROOF The first statement follows from the second, whose proof follows. Suppose that a pair (W_{l-1}, λ) with $W_{l-1} < p$ solves Equations (24) and (25) for some μ. This implies that $w_i = g_i(W_{l-1}, \lambda)$ as in Equation (19). If W_{l-1} is increased by δ to $\tilde{W}_{l-1} = W_{l-1} + \delta < p$, then $g_i(\tilde{W}_{l-1}, \lambda) < g_i(W_{l-1}, \lambda)$, $\forall i = 1, \dots, l-1$ and $\sum_{i=1}^{l-1} g_i(\tilde{W}_{l-1}, \lambda) < \sum_{i=1}^{l-1} g_i(W_{l-1}, \lambda) = W_{l-1}$. By Lemma B.1, $\sum_{i=1}^{l-1} g_i(\tilde{W}_{l-1}, \lambda') = W_{l-1}$ can be satisfied only for a unique $\lambda' < \lambda$. Since $V_{[1]}, \dots, V_{[k]}$ are sorted in ascending order, $(V_{[l]} - V_{[i]})$ is decreasing in i and the derivative of $g_i(W_{l-1}, \lambda)$ with respect to λ given in Equation (B.6) is increasing in i. Therefore, for $\lambda' < \lambda$, $i < j$ implies $g_i(\tilde{W}_{l-1}, \lambda') - g_i(\tilde{W}_{l-1}, \lambda) > g_j(\tilde{W}_{l-1}, \lambda') - g_j(\tilde{W}_{l-1}, \lambda)$. If $\sum_{i=1}^{l-1} g_i(\tilde{W}_{l-1}, \lambda') = \sum_{i=1}^{l-1} g_i(W_{l-1}, \lambda)$ is satisfied, then there exists $i_\delta < l-1$ such that $g_i(\tilde{W}_{l-1}, \lambda') \ge g_i(W_{l-1}, \lambda)$ for $i \le i_\delta$ and $g_i(\tilde{W}_{l-1}, \lambda') < g_i(W_{l-1}, \lambda)$ for $i > i_\delta$. We define $\Delta := \sum_{i=1}^{i_\delta} g_i(\tilde{W}_{l-1}, \lambda') - \sum_{i=1}^{i_\delta} g_i(W_{l-1}, \lambda)$ to be the total increase of $w_1, \dots, w_{i_\delta}$ and $\sum_{i=i_\delta+1}^{l-1} g_i(W_{l-1}, \lambda) - \sum_{i=i_\delta+1}^{l-1} g_i(\tilde{W}_{l-1}, \lambda') = -\Delta$ follows from the fact that both the new and original weights add up to W_{l-1}. We then plug the new weights into Equation (25) to find the ES μ' corresponding to

$(\tilde{W}_{l-1}, \lambda')$ by $p(\mu' + V_{[l]}) = \sum_{i=1}^{l-1} g_i(\tilde{W}_{l-1}, \lambda')(V_{[l]} - V_{[i]})$. Note that:

$$
\begin{aligned}
p(\mu' + V_{[l]}) &- p(\mu + V_{[l]}) \\
&= \sum_{i=1}^{k} [g_i(\tilde{W}_{l-1}, \lambda') - g_i(W_{l-1}, \lambda)](V_{[l]} - V_{[i]}) \\
&= \sum_{i=1}^{i_\delta} [g_i(\tilde{W}_{l-1}, \lambda') - g_i(W_{l-1}, \lambda)](V_{[l]} - V_{[i]}) \\
&\quad + \sum_{i=i_\delta+1}^{k} [g_i(\tilde{W}_{l-1}, \lambda') - g_i(W_{l-1}, \lambda)](V_{[l]} - V_{[i]}) \\
&\geq \sum_{i=1}^{i_\delta} [g_i(\tilde{W}_{l-1}, \lambda') - g_i(W_{l-1}, \lambda)](V_{[l]} - V_{[i_\delta]}) \qquad \text{(B.7)} \\
&\quad + \sum_{i=i_\delta+1}^{k} [g_i(\tilde{W}_{l-1}, \lambda') - g_i(W_{l-1}, \lambda)](V_{[l]} - V_{[i_\delta+1]}) \\
&= \Delta(V_{[l]} - V_{[i_\delta]}) - \Delta(V_{[l]} - V_{[i_\delta+1]}) \\
&= \Delta(V_{[i_\delta+1]} - V_{[i_\delta]}) \geq 0
\end{aligned}
$$

which implies $\mu' \geq \mu$. Inequality (B.7) follows since:

$$
(V_{[l]} - V_{[i_\delta]}) \leq (V_{[l]} - V_{[i]}) \quad \text{and} \quad g_i(\tilde{W}_{l-1}, \lambda') - g_i(W_{l-1}, \lambda) \geq 0, \quad \forall i \leq i_\delta
$$

and:

$$
(V_{[l]} - V_{[i_\delta+1]}) \geq (V_{[l]} - V_{[i]}) \text{ and } g_i(\tilde{W}_{l-1}, \lambda') - g_i(W_{l-1}, \lambda) \leq 0, \ \forall i \geq i_\delta + 1
$$

Equation (24) for $\tilde{W}_{l-1}$ becomes $\sum_{i=1}^{l-1} g_i(\tilde{W}_{l-1}, \tilde{\lambda}) = \tilde{W}_{l-1} > W_{l-1}$, which can be satisfied only for a unique $\tilde{\lambda} < \lambda'$. Due to monotonicity, $g_i(\tilde{W}_{l-1}, \tilde{\lambda}) > g_i(\tilde{W}_{l-1}, \lambda')$, $\forall i = 1, \dots, l-1$ and this implies for the ES $\tilde{\mu}$ corresponding to $(\tilde{W}_{l-1}, \tilde{\lambda})$ that:

$$
\begin{aligned}
p(\tilde{\mu} + V_{[l]}) &= \sum_{i=1}^{l-1} g_i(\tilde{W}_{l-1}, \tilde{\lambda})(V_{[l]} - V_{[i]}) \\
&> \sum_{i=1}^{l-1} g_i(\tilde{W}_{l-1}, \lambda')(V_{[l]} - V_{[i]}) \\
&= p(\mu' + V_{[l]}) \\
&\geq p(\mu + V_{[l]})
\end{aligned}
$$

Hence, we have shown that if (W_{l-1}, λ) and $(\tilde{W}_{l-1}, \tilde{\lambda})$ satisfy the first-order conditions (24) and (25) for μ and $\tilde{\mu}$, respectively, while $\tilde{W}_{l-1} > W_{l-1}$, then $\tilde{\mu} > \mu$.

We complete the proof of the lemma by showing that there exist $W_0 \in (W_{lb}, W_{ub})$, λ_0, and μ_0 such that Equations (24) and (25) are solved with $(W_{l-1}, \lambda) = (W_0, \lambda_0)$ and $\mu = \mu_0$; this proves that M_l^{int} is non-empty. Define $f(W_{l-1}, \lambda) := W_{l-1} - \sum_{i=1}^{l-1} g_i(W_{l-1}, \lambda)$, which is increasing in λ by Lemma B.1. For any $W_0 \in (W_{lb}, W_{ub})$, $\lim_{\lambda \to \infty} f(W_0, \lambda) = W_0$. Let λ_{lb} be the solution of $g_1(W_0, \lambda_{lb}) = W_0$. Then $0 < g_i(W_0, \lambda_{lb}) \leq W_0, \forall i > 1$ because the absolute value of the derivative of $g_i(W_0, \lambda)$ with respect to λ given in Equation (B.6) is decreasing in i. Consequently, $f(W_0, \lambda_{lb}) = W_0 - W_0 - \sum_{i=2}^{l-1} g_i(W_0, \lambda) < 0$. Because f is continuous in its second argument over the range $[\lambda_{lb}, \infty)$, there exists λ_0 such that $f(W_0, \lambda_0) = 0$, ie, Equation (24) holds for $W_{l-1} = W_0$ and $\lambda = \lambda_0$. Then μ_0 is chosen to satisfy Equation (25). □

The following lemma justifies the way in which root-finding is used to determine the endpoints of confidence intervals and the rectangles that make up confidence regions.

LEMMA B.3 *If*

- $l \leq kp$: *R_l^{int} is decreasing on (m_l^{lo}, m_l^{hi}) and the supremum of R_l^{int} on (m_l^{lo}, m_l^{hi}) is $R_l^{int}(m_l^{lo})$, where the first-order conditions with $\mu = m_l^{lo}$ are solved at $W_{l-1}^* = W_{lb} = p - (1-p)/(k-l)$ and $\lambda^* < 0$.*
- $kp < l < kp + 1$: *R_l^{int} is increasing on $(m_l^{lo}, T(F_k))$ and decreasing on $(T(F_k), m_l^{hi})$ and the supremum of R_l^{int} on (m_l^{lo}, m_l^{hi}) is $R_l^{int}(T(F_k)) = 1$, where the first-order conditions with $\mu = T(F_k)$ are solved at $W_{l-1}^* = (l-1)/k$ and $\lambda^* = 0$.*
- $l \geq kp + 1$: *R_l^{int} is increasing on (m_l^{lo}, m_l^{hi}) and the supremum of R_l^{int} on (m_l^{lo}, m_l^{hi}) is $R_l^{int}(m_l^{hi})$, where the first-order conditions with $\mu = m_l^{hi}$ are solved at $W_{l-1}^* = W_{ub} = p$ and $\lambda^* > 0$.*

PROOF Consider the set $\mathcal{F}_l^*$ of all points $(W_{l-1}^*, \lambda^*, \mu)$ such that the first-order conditions of Maximization Problem II are satisfied for $W_{l-1}^* \in (0, 1)$. This includes the point $((l-1)/k, 0, T(F_k))$, corresponding to equal weights $w_1^*, \ldots, w_k^* = 1/k$.

This point is feasible if and only if $(l-1)/k \in (W_{lb}, W_{ub})$, that is, $kp < l < kp + 1$, and in this case $R_l^{int}(T(F_k)) = 1$, the largest possible empirical likelihood ratio. It follows from Lemma B.2 that if (W_{l-1}, λ, μ) and $(\tilde{W}_{l-1}, \tilde{\lambda}, \tilde{\mu})$ are in $\mathcal{F}_l^*$ while $\mu < \tilde{\mu} < T(F_k)$, then $W_{l-1} < \tilde{W}_{l-1} < (l-1)/k$ and $\lambda > \tilde{\lambda} > 0$. Because $W_{l-1} < \tilde{W}_{l-1} < (l-1)/k$, the average weight in the tail (ie, w_i for $i = 1, \ldots, l-1$) is "too small", that is, less than $1/k$, in the solution to Maximization Problem II with $\tilde{\mu}$ and it is even less in the solution with μ. Because $\lambda > \tilde{\lambda} > 0$, the weights in the tail are unequal to each other in the solutions to Maximization Problem II with μ or with $\tilde{\mu}$, and they are more distorted in the solution with μ. Both of these effects cause $R_l^{int}(\mu) < R_l^{int}(\tilde{\mu})$. Thus R_l^{int} is increasing on $(-V_{[k]}, T(F_k)]$. By similar reasoning, it is decreasing on $[T(F_k), -V_{[1]})$.

Next consider the case $l \geq kp + 1$. In this case, $((l-1)/k, 0, T(F_k))$ is infeasible because $(l-1)/k \geq p = W_{ub}$. By Lemma B.2, $(m_l^{lo}, m_l^{hi}) \subset (-V_{[k]}, T(F_k)]$ and the conclusion follows.

Finally, consider the case $l \leq kp$, where $((l-1)/k, 0, T(F_k))$ is infeasible because $(l-1)/k \leq W_{lb}$. The conclusion follows in a similar manner from $(m_l^{lo}, m_l^{hi}) \subset (T(F_k), -V_{[1]})$. $\square$

REFERENCES

Acerbi, C. (2002). Spectral measures of risk: a coherent representation of subjective risk aversion. *Journal of Banking and Finance* **26**, 1505–1518.

Acerbi, C., and Tasche, D. (2002). On the coherence of expected shortfall. *Journal of Banking and Finance* **26**, 1487–1503.

Agresti, A., and Coull, B. A. (1998). Approximate is better than exact for interval estimation of binomial proportions. *The American Statistician* **52**(2), 119–126.

Artzner, P., Delbaen, F., Eber, J., and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance* **9**, 203–228.

Chang, Y. P., Hung, M. C., and Wu, Y. F. (2003). Nonparametric estimation for risk in value-at-risk estimator. *Communications in Statistics: Simulation and Computation* **3**, 1041–1064.

Chen, S. X., and Hall, P. (1993). Smoothed empirical likelihood confidence intervals for quantiles. *Annals of Statistics* **21**(3), 1166–1181.

Chernick, M. R. (1999). *Bootstrap Methods: A Practitioner's Guide*. John Wiley & Sons.

Clopper, C. J., and Pearson, E. S. (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika* **26**(4), 404–413.

Dagpunar, J. (1988). *Principles of Random Variate Generation*. Clarendon Press, Oxford.

Davison, A. C., and Hinkley, D. V. (1997). *Bootstrap Methods and their Application*. Cambridge University Press.

Dowd, K. (2005). Estimating risk measures. *Financial Engineering News* **43**, 13.

Durrett, R. (1996). *Probability: Theory and Examples*. Duxbury Press.

Hardy, M. (2006). Simulating VaR and CTE. *Financial Engineering News* **47**, 17.

Hesterberg, T. C., and Nelson, B. L. (1998). Control variates for probability and quantile estimation. *Management Science* **44**(9), 1295–1312.

Hong, L. J. (2006). Estimating value and sensitivities of conditional value-at-risk. Unpublished manuscript, Department of Industrial Engineering and Logistics Management, Hong Kong University of Science and Technology.

Lan, H., Nelson B. L., and Staum, J. (2007). Two-level simulations for risk management. *Proceedings of the 2007 INFORMS Simulation Society Research Workshop*, 102–107.

Manistre, B. J., and Hancock, G. H. (2005). Variance of the CTE estimator. *North American Actuarial Journal* **9**, 129–154.

Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **75**, 237–249.

Owen, A. B. (1990). Empirical likelihood ratio confidence regions. *The Annals of Statistics* **18**, 90–120.

Owen, A. B. (2001). *Empirical Likelihood*. Chapman & Hall/CRC.

Peng, L., and Qi, Y. (2006). Confidence regions for high quantiles of a heavy tailed distribution. *Annals of Statistics* **34**, 1964–1986.

Rockafellar, R. T., and Uryasev, S. (2000). Optimization of conditional value-at-risk. *The Journal of Risk* **2**(3), 21–41.

Rockafellar, R. T., and Uryasev, S. (2002). Conditional value-at-risk for general loss distributions. *Journal of Banking and Finance* **26**, 1443–1471.

Shao, J., and Tu, D. (1995). *The Jackknife and Bootstrap*. Springer.

Sheather, S. J., and Marron, J. S. (1990). Kernel quantile estimators. *Journal of the American Statistical Association* **85**, 410–416.

Silverman, B. W. (1978). Weak and uniform consistency of the kernel estimate of a density function and its derivatives. *The Annals of Statistics* **6**, 177–184.

Staudte, R. G., and Sheather, S. J. (1990). *Robust Estimation and Testing*. John Wiley & Sons.

Yamai, Y., and Yoshiba, T. (2002). Comparative analyses of expected shortfall and value-at-risk: Their estimation error, decomposition, and optimization. *Monetary and Economic Studies* **20**(1), 87–121.

Zhou, W., and Jing, B.-Y. (2003). Adjusted empirical likelihood method for quantiles. *Annals of the Institute of Statistical Mathematics* **55**(4), 689–703.