Artificially Augmented Samples, Shrinkage, and Mean Squared Error Reduction*

Yannis G. YATRACOS

An inequality is provided that determines when shrinkage reduces the mean squared error (MSE) of an unbiased estimate. Artificially augmented samples are then used to obtain, among others, shrinkage estimates of the population's variance and covariance, which improve the unbiased estimates for all parameter values and for all probability models with marginals having finite second moments, and alternative jackknife estimates that complement the usual jackknife estimates in reducing the MSE.

KEY WORDS: Augmented samples; Bias; Jackknife; Mean squared error; Multiple imputation; Shrinkage; U-statistics; Variance estimation.

1. INTRODUCTION

The estimation of a population's variance, σ^2, and covariance, $\sigma_{X,Y}$, is the stuff of statistical folkore. Often the unbiased estimate, s_n^2, of σ^2 is used, but for some probability models, shrinkage estimates, cs_n^2, reduce the mean squared error (MSE) of s_n^2 for every σ, $0 < c < 1$, with n the sample size. So far, there is no shrinkage estimate with a smaller MSE than s_n^2 that applies universally, for all values of σ, for all probability models and every sample size n (Stein 1964; Brown 1968; Arnold 1970; Lehmann 1983, p. 113). Such an estimate is provided in this article. A similar situation holds for the unbiased estimate of $\sigma_{X,Y}$.

For an unbiased estimate t_n of a parameter θ with real values, an increase in the sample size n has usually the same effect as a successful shrinkage; both decrease the MSE. Questions arise as to whether by artificially augmenting the sample, an estimate $\tilde{t}_n$ can be obtained with a smaller MSE than t_n for every θ value, and as to whether $\tilde{t}_n$ is a shrinkage estimate. In this work it is seen that for some parameters, this is indeed so, and that even more is true; $\tilde{t}_n^k$, the average of the values of t_{n+k} on artificially augmented samples, turns out to be a shrinkage estimate that has a smaller MSE than t_n not only for all θ values, but also for all probability models, $1 \le k < n$.

In particular, in variance estimation, the average of $s_{n+1}^2(X_1, \ldots, X_n, X_i)$, $i = 1, \ldots, n$, turns out to be a shrinkage estimate because for the U-statistic kernel $h(x_1, x_2)$, which determines σ^2 and s_n^2, $h(x, x) = 0$. The obtained estimate, $\frac{(n+2)(n-1)}{n(n+1)} s_n^2$, has a smaller MSE than s_n^2 for all values of σ, $n \ge 2$ and for all probability models with finite second moments. The same shrinkage coefficient, $\frac{(n+2)(n-1)}{n(n+1)}$, is obtained when averaging the values $t_{n+1,2}(X_1, \ldots, X_n, X_i)$, $i = 1, \ldots, n$, of a U-statistic $t_{n,2}$ with symmetric kernel of order $m = 2$ vanishing at the diagonal, like, for example, those determined by $\sigma_{X,Y}$, Kendall's τ, and Gini's index g. The shrinkage estimate of $\sigma_{X,Y}$ also has a smaller MSE than the corresponding U-statistic for all covariance values, $n \ge 2$ and for all probability models with marginals having finite second moments. However, additional assumptions are needed for a similar result to hold when estimating either τ or g.

The results are presented for a U-statistic $t_{n,m}$ with a symmetric kernel of order $m \ge 2$ that vanishes when two arguments are repeated, and the shrinkage coefficients $c_{\delta_n,k,m}$ are obtained using $(n+k)$ artificially augmented samples, $1 \le k < n$, where δ_n is a positive number that can be chosen to increase with n. For n large, our analysis suggests that shrinkage coefficients are to be obtained from $(n + k_n)$-augmented samples for the bias and the MSE improvement to slowly decrease to 0 as n increases. Shrinkage coefficients are also obtained that are used to reduce the MSE of some other unbiased estimates.

An alternative $(n+1)$ jackknife estimate, $\tilde{t}_n^J$, is also provided, which together with the usual $(n - 1)$ jackknife estimate, t_n^J, has the potential to reduce the MSE of a biased estimate, t_n, of θ. This is contrary to results on unaugmented jackkknife procedures (Shao and Tu 1995, sec. 2.5, p. 70, l. 3–5). For example, when the population's mean is unknown, the estimate $\tilde{t}_n^J = \frac{n-1}{n+1} s_n^2$ of σ^2 improves $t_n^J = s_n^2$ and $t_n = \frac{n-1}{n} s_n^2$ for various models. When t_n is smooth and n is large, conditions are provided that determine when $\tilde{t}_n^J, t_n^J$, and $\tilde{t}_n^1$ [the average of $t_{n+1}(X_1, \ldots, X_n, X_i)$, $i = 1, \ldots, n$] have smaller MSE than t_n. It is expected that similar results will hold for $(n+k)$ jackknife estimates.

In Section 2 a sufficient condition is provided for a shrinkage estimate to reduce the MSE of an unbiased estimate, t_n, of θ for all θ values and a family, $\mathcal{F}$, of probability models. In Section 3 the basis of the motivation to use $(n+k)$ artificially augmented samples and $\tilde{t}_n^k$ is presented. In Section 4 $\tilde{t}_n^k$ is used to obtain shrinkage estimates that improve the MSE of some U-statistics and other unbiased estimates. Finally, in Section 5, the alternative $(n+1)$ jackknife estimate $\tilde{t}_n^J$ is proposed and studied.

2. SHRINKAGE AND MSE REDUCTION

Let $X_1, \ldots, X_n$ be a sample from an unknown cumulative distribution function F in a known class $\mathcal{F}$ of models, and let $t_n(X_1, \ldots, X_n)$ be an unbiased estimate of the unknown model parameter $\theta \in \Theta$ $(\subseteq R)$ with finite second moment; θ may be, for example, the mean of F. The MSE of the shrinkage estimate $c_n t_n$, $0 < c_n < 1$, is minimized when $c_n(\theta, F) = \theta^2 / E t_n^2 = (1 + \frac{\text{var}(t_n)}{\theta^2})^{-1}$. Because $c_n(\theta, F)$ often depends on θ and F, this approach does not yield a universal shrinkage coefficient c_n that minimizes the MSE of t_n for every $\theta \in \Theta$ and for every

Yannis G. Yatracos is Professor, Department of Statistics and Applied Probability, The National University of Singapore, Singapore 117546 (E-mail: *yatracos@stat.nus.edu.sg*). The author thanks the referees for useful suggestions that served to improve the presentation of the article, particularly the referee whose comments encouraged the addition of Remark 5 in Section 4.4. The author also thanks the associate editor, who provided a very thoughtful report, allowed the expansion of a previous version of the manuscript, encouraged the use of augmented samples, and made valuable suggestions on the presentation of this work. Last but not least, the author thanks Frank Samaniego, the editor, for encouragement, advice on the presentation of this work, and its prompt processing.

*This article is dedicated to the memory of Lucien LeCam.

Journal of the American Statistical Association
December 2005, Vol. 100, No. 472, Theory and Methods
DOI 10.1198/016214505000000321

$F \in \mathcal{F}$ when $\mathcal{F}$ consists of more than one model. An alternative goal is to determine shrinkage coefficients that reduce the MSE of t_n for every $\theta \in \Theta$ and for every $F \in \mathcal{F}$. These coefficients are selected from the set $[\sup_{\Theta,\mathcal{F}} c_n(\theta, F), 1)$ that is nonempty if $\inf_{\Theta,\mathcal{F}} \frac{\text{var}(t_n)}{\theta^2} > 0$ because $\sup_{\Theta,\mathcal{F}} c_n(\theta, F) = (1 + \inf_{\Theta,\mathcal{F}} \frac{\text{var}(t_n)}{\theta^2})^{-1}$; $\sup_{\Theta,\mathcal{F}}$ (resp. $\inf_{\Theta,\mathcal{F}}$) denotes $\sup_{\theta\in\Theta,F\in\mathcal{F}}$ (resp. $\inf_{\theta\in\Theta,F\in\mathcal{F}}$).

We now characterize the shrinkage coefficients that reduce the MSE of t_n for a given θ and F.

Lemma 1.

$$E(c_n t_n - \theta)^2 < E(t_n - \theta)^2 = \text{var}(t_n) \quad \text{iff} \quad \frac{1 - c_n}{1 + c_n} < \frac{\text{var}(t_n)}{\theta^2}. \tag{1}$$

Proof. Use the relation

$$E(c_n t_n - \theta)^2 = c_n^2 \,\text{var}(t_n) + (1 - c_n)^2 \theta^2.$$

From (1), it follows that when F is the true model, the unbiased estimate t_n can be improved with shrinkage for every $\theta \in \Theta$ if $\inf_{\theta\in\Theta} \frac{\text{var}(t_n)}{\theta^2}$ is bounded below by some known positive constant L_F that depends on F and n. This occurs when, for example, the Fisher information $I_{X_1}(\theta) = M/\theta^2, M > 0$, and the Cramer–Rao inequality holds for t_n at the model F. t_n can be improved with shrinkage for every $\theta \in \Theta$ and for every $F \in \mathcal{F}$ if $\inf_{\Theta,\mathcal{F}} \frac{\text{var}(t_n)}{\theta^2}$ is bounded below by some known positive constant L that depends on n.

In (1) the lower bound $\frac{1-c_n}{1+c_n}$ is a decreasing function of c_n that should be suitably chosen; it should be large enough to cause moderate bias and for (1) to hold for every $\theta \in \Theta$ and every $F \in \mathcal{F}$, with the corresponding MSE reduction to slowly decrease to 0 as n increases.

The estimate $\hat{c}_n = (1 + \frac{\hat{V}_n}{t_n^2})^{-1}$ of $c_n(\theta, F)$ will not reduce the MSE of t_n for each n and each F, because $\hat{V}_n$ may not be a satisfactory estimate of $\text{var}(t_n)$ for all models F. This can be observed in large samples when $\theta = \sigma^2$ and $t_n = s_n^2$. Let μ_k and m_k be the kth-order central moments of the population and of the sample, $k \geq 1$, and let $a_n \sim b_n$ denote $\lim_{n\longrightarrow\infty} \frac{a_n}{b_n} = 1$. $\text{var}(s_n^2) \sim \frac{\mu_4-\mu_2^2}{n}$ can be estimated by $\hat{V}_n = \frac{m_4-m_2^2}{n}$, $E(\hat{V}_n - \frac{\mu_4-\mu_2^2}{n}) = O(n^{-2})$ (see, e.g., Serfling 1980, pp. 69–71), and $\text{var}(\hat{V}_n) \sim 56\sigma^8/n^3$ (Stuart and Ord 1994, p. 369). For normal models, $\text{var}(s_n^2) \sim 2\sigma^4/n$ can be estimated by $\hat{V}'_n = 2s_n^4/n$. Recall that if U_n follows a χ_n^2 distribution, then $EU_n^m = \prod_{j=1}^m [n + 2(j - 1)]$, $\text{var}(U_n) = 2n$, $\text{var}(U_n^2) = 8n(n + 2)(n + 3)$, and, therefore, $E\hat{V}'_n - 2\frac{\sigma^4}{n} = 2\frac{\text{var}(s_n^2)}{n} = \frac{4\sigma^4}{n(n-1)} = O(n^{-2})$ and $\text{var}(V'_n) = \frac{4\sigma^8 \,\text{var}(U_{n-1})}{n^2(n-1)^4} \sim \frac{32\sigma^8}{n^3}$. Thus, for large n, $\text{var}(\hat{V}_n) > \text{var}(\hat{V}'_n)$, the MSE of $\hat{V}_n$ is larger than that of $\hat{V}'_n$, and $\hat{c}_n$ may underestimate $\sup_{\Theta,\mathcal{F}} c_n(\theta, F)$.

3. ESTIMATES BASED ON ARTIFICIALLY AUGMENTED SAMPLES

Pseudovalues of an estimate t_n of θ are used to, for example, estimate its variance or to obtain a new estimate with reduced bias or when data are missing. These pseudovalues are usually obtained by evaluating either t_n on B bootstrap samples (Efron 1979), t_{n-k} on $(n - k)$-reduced samples (Quenouille 1956), or t_n on samples obtained with multiple-imputation methods (Rubin 1987).

The class $\mathcal{A}_{n,k}$ of the $(n + k)$ artificially augmented samples consists of the samples $\mathbf{X} = (X_1, \ldots, X_n, X_{n+1} = X_{j_1}, \ldots, X_{n+k} = X_{j_k})$, $1 \leq j_1 < \cdots < j_k \leq n$, and the pseudovalues $t_{n+k}(\mathbf{X}), \mathbf{X} \in \mathcal{A}_{n,k}$, are used to define the estimate

$$\tilde{t}_n^k = \binom{n}{k}^{-1} \sum_{\mathbf{X}\in\mathcal{A}_{n,k}} t_{n+k}(\mathbf{X}), \qquad 1 \leq k < n. \tag{2}$$

$\mathcal{A}_{n,k}$, $t_{n+k}(\mathbf{X}), \mathbf{X} \in \mathcal{A}_{n,k}$, and $\tilde{t}_n^k$ can all be thought of in terms of multiple imputation for a sample with size $(n+k)$ and k missing observations.

The proposition that follows encourages the use both of $(n + k)$-augmented samples and of the estimate $\tilde{t}_n^k$ in (2). The use of B bootstrap $(n + k)$-augmented samples is discouraged due to the additional randomization introduced by finite resampling (Yatracos 2002).

Proposition 1. Let $\hat{F}_n, \hat{F}_{n-k,i_1,\ldots,i_k}$, and $\hat{F}_{n+k,i_1,\ldots,i_k}$ denote the empirical cumulative distributions of the original sample $\{X_1, \ldots, X_n\}$, $\{X_1, \ldots, X_n\} - \{X_{i_1}, \ldots, X_{i_k}\}$, and $\{X_1, \ldots, X_n, X_{i_1}, \ldots, X_{i_k}\}$, $1 \leq k < n$, $1 \leq i_l \neq i_m \leq n$. Then it holds that

$$\sup_x \left|\hat{F}_{n+k,i_1,\ldots,i_k}(x) - \hat{F}_n(x)\right| < \sup_x \left|\hat{F}_{n-k,i_1,\ldots,i_k}(x) - \hat{F}_n(x)\right|. \tag{3}$$

Proof. Let I denote the indicator function. Then (3) follows from the relations

$$\begin{aligned}\hat{F}_{n+k,i_1,\ldots,i_k}(x) &= \frac{n}{n+k}\hat{F}_n(x) + \frac{1}{n+k}\sum_{j=1}^k I(X_{i_j} \leq x) \\ &= \hat{F}_n(x) + \frac{1}{n+k}\sum_{j=1}^k \left[I(X_{i_j} \leq x) - \hat{F}_n(x)\right]\end{aligned}$$

and

$$\begin{aligned}\hat{F}_{n-k,i_1,\ldots,i_k}(x) &= \frac{n}{n-k}\hat{F}_n(x) - \frac{1}{n-k}\sum_{j=1}^k I(X_{i_j} \leq x) \\ &= \hat{F}_n(x) - \frac{1}{n-k}\sum_{j=1}^k \left[I(X_{i_j} \leq x) - \hat{F}_n(x)\right].\end{aligned}$$

4. SHRINKAGE OF *U*–STATISTICS

4.1 *U*-Statistics and Augmented Samples

For a symmetric kernel $h(x_1, x_2, \ldots, x_m)$ of degree m, such that $Eh(X_1, X_2, \ldots, X_m) = \theta$, the U-statistic of θ and the $(n + k)$-augmented sample estimate (2) are

$$t_{n,m} = \binom{n}{m}^{-1} \sum_{1\leq i_1<\cdots<i_m\leq n} h(X_{i_1}, \ldots, X_{i_m}) \tag{4}$$

and

$$\tilde{t}_{n,m}^k = \binom{n}{k}^{-1} \binom{n+k}{m}^{-1} \times \sum_{\mathbf{X}\in\mathcal{A}_{n,k}} \sum_{1\leq i_1<\cdots<i_m\leq n} h(X_{i_1}, \ldots, X_{i_m}). \tag{5}$$

4.2 Shrinkage Estimates

Let V_m denote the class of symmetric kernels of degree $m \geq 2$ that vanish when two of the arguments in the kernel are repeated. When $h \in V_m$, $\tilde{t}^k_{n,m}$ turns out to be a shrinkage estimate of $t_{n,m}$; the shrinkage coefficients are obtained in Proposition 2.

Proposition 2. Let $t_{n,m}$ be as in (4) with $h \in V_m, l = \min\{k, m\}$.

a. The $(n+k)$-augmented sample estimate is

$$\tilde{t}^k_{n,m} = c_{n,k,m} t_{n,m} = \left[\binom{n+k}{m}^{-1} \sum_{j=0}^{l} \binom{k}{j} \binom{n-j}{m-j} \right] t_{n,m}. \tag{6}$$

b. The $(n+1)$-augmented sample estimate is

$$\tilde{t}^1_{n,m} = c_{n,1,m} t_{n,m} = \left[1 - \frac{m^2 - m}{n(n+1)} \right] t_{n,m}, \tag{7}$$

the corresponding lower bound in (1) is

$$\frac{1 - c_{n,1,m}}{1 + c_{n,1,m}} = \frac{m^2 - m}{2n(n+1) - m^2 + m}, \tag{8}$$

and it holds that

$$c_{n,1,m} \leq c_{n,1,2}, \qquad m \geq 2. \tag{9}$$

c. When $m = 2$, the $(n+k)$-augmented sample estimate is

$$\tilde{t}^k_{n,2} = c_{n,k,2} t_{n,2} = \left[1 - \frac{2k}{(n+k)(n+k-1)} \right] t_{n,2}, \tag{10}$$

the corresponding lower bound in (1) is

$$\frac{1 - c_{n,k,2}}{1 + c_{n,k,2}} = \frac{k}{(n+k)(n+k-1) - k}, \tag{11}$$

and it holds that

$$0 < c_{n,k,2} \leq c_{n,k-1,2} \leq c_{n,1,2}, \qquad 2 \leq k < n. \tag{12}$$

Proof. To prove part a, let $\Delta_{n,k,m} = \sum_{j=0}^{l} \binom{k}{j} \binom{n-j}{m-j}$. In (5), $\binom{n}{k} \binom{n+k}{m} \tilde{t}^k_{n,m}$ has $\Delta_{n,k,m} \binom{n}{k}$ nonvanishing terms and equals

$$\binom{n}{k} \frac{\Delta_{n,k,m}}{\binom{n}{m}} \sum_{1 \leq i_1 < i_2 < \cdots < i_m \leq n} h(X_{i_1}, X_{i_2}, \ldots, X_{i_m}). \tag{13}$$

From (5) and (13), (6) follows. Proofs for parts b and c follow from the proof of part a.

Remark 1.

a. For the estimate (6), note, for example, that when $n = 3$, $m = 2$, $k = 2$, and $l = 2$ then $\tilde{t}^2_{3,2} = .8 t_{3,2}$.

b. Equation (12) and Section 2 explain why $c_{n,1,2}$ is used when $m = 2$. From (10), it follows that, for n large, the bias and the MSE reduction decrease more slowly to 0 when using $c_{n,k_n,2}$ with k_n increasing (see also Sec. 4.4, Remark 5).

For $1 \leq j \leq m$, let

$$h_j(x_1, \ldots, x_j) = E\left[h(X_1, \ldots, X_m) | X_1 = x_1, \ldots, X_j = x_j \right],$$
$$\zeta_j = \mathrm{var}[h_j(X_1, \ldots, X_j)].$$

Then it holds that (see, e.g., Serfling 1980, p. 183)

$$\binom{n}{m} \mathrm{var}(t_{n,m}) = \sum_{j=1}^{m} \binom{m}{j} \binom{n-m}{m-j} \zeta_j \tag{14}$$

and

$$0 \leq \zeta_1 \leq \cdots \leq \zeta_m = \mathrm{var}\, h(X_1, \ldots, X_m). \tag{15}$$

Proposition 3. Let $t_{n,m}$ be as in (4) with $h \in V_m$.

a. The following statements are sufficient for the shrinkage estimate $c_{\delta_n,1,m} t_{n,m}$ to have smaller MSE than $t_{n,m}$ for every $\theta \in \Theta$; δ_n is determined by (18) and can be chosen to increase with n.

1. There is a known constant $L_F > 0$ such that

$$\binom{n}{m} \inf_{\theta \in \Theta} \frac{\mathrm{var}(t_{n,m})}{\theta^2} > L_F. \tag{16}$$

2. There is a known constant $L_F > 0$ and j_0, $1 \leq j_0 \leq m$, such that

$$\inf_{\theta \in \Theta} \frac{\zeta_{j_0}}{\theta^2} > L_F. \tag{17}$$

b. When $m = 2$, either (16) or (17) is sufficient for the shrinkage estimate $c_{\delta_n,k,2} t_{n,2}$ to have smaller MSE than $t_{n,2}$ for every $\theta \in \Theta$; δ_n is determined by (19) and can be chosen to increase with n.

Proof. For part a1, use (1) and choose an increasing sequence δ_n such that for every n, it holds that

$$\binom{n}{m} \frac{1 - c_{\delta_n,1,m}}{1 + c_{\delta_n,1,m}} = \binom{n}{m} \frac{m^2 - m}{2\delta_n^2 + 2\delta_n - m^2 + m} < L_F. \tag{18}$$

The proof of part a2 follows from part a1, because from (14) and (15), it holds that

$$\binom{n}{m} \mathrm{var}(t_{n,m}) \geq \zeta_m \geq \cdots \geq \zeta_1.$$

For part b, to prove sufficiency of (16), use (1) and choose an increasing sequence δ_n such that for every n, it holds that

$$\binom{n}{2} \frac{1 - c_{\delta_n,k,2}}{1 + c_{\delta_n,k,2}} = \binom{n}{2} \frac{k}{(\delta_n + k)(\delta_n + k - 1) - k} < L_F. \tag{19}$$

Sufficiency of (17) follows as for part a1.

Remark 2. When (17) holds for $j_0 \leq m - 1$, (15) implies that it also holds for $j_0 = m$.

Corollary 1. If $L = \inf_{F \in \mathcal{F}} L_F$ is positive, then the estimate $c_{\delta_n,1,m} t_{n,m}$ (resp. $c_{\delta_n,k,2} t_{n,2}$) obtained using L instead of L_F in (18) [resp. (19)] has smaller MSE than $t_{n,m}$ (resp. $t_{n,2}$) for all θ's and for all models $F \in \mathcal{F}$.

4.3 Applications

The kernels for the population variance and covariance, Kendall's τ, and Gini's index are $\frac{(x_1-x_2)^2}{2}$, $\frac{(x_1-x_2)(y_1-y_2)}{2}$, $\operatorname{sign}((x_1-x_2)(y_1-y_2))$, and $|x_1-x_2|^\gamma, \gamma>0$. The ordering of the coefficients $c_{n,k,2}$ in Proposition 2c and (1) suggest using $(n+1)$-augmented sample estimates when n is small; $\delta_n = n^r$ is used herein.

4.3.1 The Population Variance σ_X^2 and the Population Covariance σ_{XY}. Let $(X_1,Y_1),\dots,(X_n,Y_n)$ be a two-dimensional sample with joint cumulative distribution function F; $\mu_X = EX_1, \mu_Y = EY_1, \sigma_X^2 = \operatorname{var}(X_1), \sigma_Y^2 = \operatorname{var}(Y_1), \sigma_{X,Y} = E(X_1-\mu_X)(Y_1-\mu_Y)$, and $\mu_{2,2} = E(X_1-\mu_X)^2(Y_1-\mu_Y)^2$. Let $t_{n,2} = \frac{1}{n-1}\sum_{i=1}^n (X_i-\bar X)(Y_i-\bar Y)$ be the U-statistic estimating $\sigma_{X,Y}$; $\bar X$ and $\bar Y$ are the averages of the X's and of the Y's. It holds that (Lee 1990, p. 14)

$$\operatorname{var}(t_{n,2}) = \frac{\mu_{2,2}}{n} - \frac{(n-2)\sigma_{X,Y}^2 - \sigma_X^2\sigma_Y^2}{n(n-1)} = \frac{(n-1)(\mu_{2,2}-\sigma_{X,Y}^2) + \sigma_{X,Y}^2 + \sigma_X^2\sigma_Y^2}{n(n-1)}, \quad (20)$$

and because

$$n(n-1)\frac{1-c_{n,1,2}}{1+c_{n,1,2}} = \frac{n^2-n}{n^2+n-1} < 1 < \frac{(n-1)(\mu_{2,2}-\sigma_{X,Y}^2) + \sigma_{X,Y}^2 + \sigma_X^2\sigma_Y^2}{\sigma_{X,Y}^2}, \quad (21)$$

it follows from Corollary 1 that $\tilde t_{n,2}^1 = c_{n,1,2}t_{n,2}$ has a smaller MSE than $t_{n,2}$ for all values of $\sigma_{X,Y}$ and for any model F with $E_F X_1^2 < +\infty$ and $E_F Y_1^2 < +\infty$, $n \ge 2$.

When μ_X is unknown, σ_X^2 is usually estimated by $s_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i-\bar X)^2$ even though $\frac{n-1}{n+1}s_n^2$ has a smaller MSE for normal populations but is not admissible (Stein 1964). For nonnormal populations, s_n^2 may have smaller MSE than either $\frac{n-1}{n+1}s_n^2$ or $\hat\sigma^2 = \frac{n-1}{n}s_n^2$. Using (20) and (21), it follows that $\tilde s_n^2 = c_{n,1,2}s_n^2 = \frac{(n+2)(n-1)}{n(n+1)}s_n^2$ has a smaller MSE than $t_{n,2} = s_n^2$ for all values of σ and for any model F with finite second moment and $n \ge 2$.

4.3.2 Kendall's τ. Let $t_{n,2} = \frac{2}{n(n-1)}\sum_{1\le i<j\le n}\operatorname{sign}(X_i - X_j)(Y_i-Y_j)$ be the U-statistic estimating $\tau = P[(X_2-X_1)(Y_2-Y_1)>0] - P[(X_2-X_1)(Y_2-Y_1)<0]$, $\tau \in [-1+\epsilon, 1-\epsilon]$, $0<\epsilon<1$. Then there is positive integer $r = r(\epsilon)$ such that $\tilde t_{n,2}^1 = c_{n^r,1,2}t_n$ has a smaller MSE than t_n for $n \ge 2$. This follows from Proposition 3a, because it holds that (Lee 1990, p. 14)

$$\operatorname{var}(t_{n,2}) = \frac{2}{n(n-1)} \times \big(2(n-2)\operatorname{var}\big(E\big[\operatorname{sign}(X_1-X_2)(Y_1-Y_2)| X_1=x_1, Y_1=y_1\big]\big) + 1 - \tau^2\big)$$

and that

$$\frac{n(n-1)}{2}\frac{\operatorname{var}(t_{n,2})}{\tau^2} > \frac{n(n-1)}{2}\operatorname{var}(t_{n,2}) > 1-\tau^2 > 1-(1-\epsilon)^2.$$

4.3.3 Gini's Index, g. Let $t_{n,2} = \frac{2}{n(n-1)}\sum_{1\le i<j\le n}|X_i - X_j|^\gamma$ be the U-statistic estimating $g = E|X_1-X_2|^\gamma \in \mathcal{G}, \gamma>0$. Assume that

$$\inf_{g\in\mathcal{G}} \frac{\operatorname{var}|X_1-X_2|^\gamma}{g^2} > L_F > 0.$$

Then there is a positive integer $r = r(L_F)$ such that $\tilde t_{n,2}^1 = c_{n^r,1,2}t_{n,2}$ has a smaller MSE than t_n for $n \ge 2$. This follows from Proposition 3a, as in the previous example, because it holds that (Serfling 1980, p. 183)

$$\operatorname{var}(t_{n,2}) = \frac{2}{n(n-1)}\big(2(n-2)\operatorname{var}E(|X_1-X_2|^\gamma | X_1 = x_1) + \operatorname{var}|X_1-X_2|^\gamma\big).$$

4.4 Additional Remarks

Remark 3. Equation (1), Proposition 3, and the results in Section 4.3 motivate the use of the shrinkage coefficient $c_{n^r,1,2}$ for any unbiased estimate t_n of θ for which $n^m \inf_{\theta\in\Theta}\frac{\operatorname{var}(t_n)}{\theta^2} > L_F > 0$, with L_F known. r satisfies the inequality $\frac{1}{n^{2r-m}+n^{r-m}-n^{-m}} < L_F$ for $n \ge 2$, and $\tilde t_n = c_{n^r,1,2}t_n$ dominates t_n for all θ values, $n \ge 2$. When L_F is not known, the shrinkage estimate asymptotically improves the unbiased estimate. For example, when estimating the mean μ of a distribution with $\bar X$ and the variance σ^2 is unknown, it follows from (1) and (11) with $k=1$ that $\frac{(n-1)(n+2)}{n(n+1)}\bar X$ dominates $\bar X$ if $\frac{\mu^2}{\sigma^2} < \frac{n^2+n-1}{n}$, that is, if $\frac{\mu^2}{\sigma^2}$ is not "very large," which holds for n large.

Remark 4. When $c_{n^r,1,2}t_n$ is used instead of t_n, the amount of MSE reduction increases as the variance of t_n increases, and can be substantial irrespective of the sample size n. For example, in variance estimation for normal models, it holds that (Lehmann 1983, p. 113)

$$E\left[\tilde c\sum_{i=1}^n (X_i-\bar X)^2 - \sigma^2\right]^2 = \sigma^4[(n^2-1)\tilde c^2 - 2(n-1)\tilde c + 1], \quad (22)$$

and therefore the MSE reduction due to shrinkage is proportional to σ, which can take any positive value.

Remark 5. Rather than using $(n+k)$-augmented samples to obtain a shrinkage coefficient for $t_{n,m}$, a referee suggested finding $L>0$ such that $\inf_{\Theta,\mathcal{F}}\frac{\operatorname{var}(t_{n,m})}{\theta^2} > L$ holds, then solve the equation $\frac{1-c}{1+c} = L$ to obtain the shrinkage estimate $ct_{n,m}$ that dominates $t_{n,m}$ for all $F\in\mathcal{F}$. However, the determination of a lower bound L is not straightforward, and the MSE reduction achieved with $ct_{n,m}$ may rapidly decrease to 0 as n increases. For example, in covariance estimation, it follows from (20) that for each model F, it holds that

$$\frac{\operatorname{var}(t_{n,2})}{\sigma_{X,Y}^2} = \frac{1}{n(n-1)} + g(F,\sigma_{X,Y},n),$$

and it is not clear whether $\inf_{\Theta,\mathcal{F}} g(F,\sigma_{X,Y},n) > 0$ such that one can choose $L = \frac{1}{n(n-1)}$. In variance estimation, for the normal model with unknown mean, it holds that $\inf_{\Theta,\mathcal{F}}\frac{\operatorname{var}(s_n^2)}{\sigma^4} > \frac{2}{n(n-1)}$, and one can choose $L_j = \frac{j}{n(n-1)}, j=1,2$. The solution of the equation $\frac{1-c}{1+c} = L_j$ is $c_{j,n} = \frac{n^2-n-j}{n^2-n+j}, j=1,2$, but ob-

Table 1. Comparing Shrinkage Estimates cs_n^2 (normal model)

	n = 5			n = 10			n = 15			n = 20		
c values	c	$L=\frac{1-c}{1+c}$	MSE	c	$L=\frac{1-c}{1+c}$	MSE	c	$L=\frac{1-c}{1+c}$	MSE	c	$L=\frac{1-c}{1+c}$	MSE
Best ($c=\frac{n-1}{n+1}$)	.6667	.1999	.3333	.8182	.1	.1818	.875	.0667	.125	.9048	.0502	.0952
Unbiased ($c=1$)	1	0	.5	1	0	.2222	1	0	.1429	1	0	.1053
$c_{1,n}=\frac{n^2-n-1}{n^2-n+1}$	.9048	.05	.4184	.978	.011	.2132	.9905	.0044	.1404	.9948	.0022	.1044
$c_{2,n}=\frac{n^2-n-2}{n^2-n+2}$	.8182	.1001	.3677	.9565	.0221	.2053	.9811	.0094	.1379	.9895	.0051	.1032
$c_{n,k=1,2}$	.9333	.0346	.4399	.9818	.0091	.2146	.9917	.0044	.1404	.9952	.0022	.1044
$c_{n,k=2,2}$	.9048	.05	.4184	.9697	.0156	.2097	.9853	.0073	.139	.9913	.0041	.1036
$c_{n,k=3,2}$	.8929	.0566	.41	.9615	.0198	.2068	.9804	.0101	.1376	.9881	.006	.1029
$c_{n,k=4,2}$	.8889	.0589	.4073	.956	.0226	.205	.9766	.0115	.1369	.9855	.007	.1026

serve that $c_{2,2}=0$. Table 1 shows, for $n=5$, 10, 15, and 20 and for various c values, the corresponding value of $L=\frac{1-c}{1+c}$ and the MSE of the shrinkage estimate cs_n^2 obtained from (22) when $\sigma^2=1$. As n increases, the bias of cs_n^2 and the associated MSE improvement both decrease fast to 0, $c\in\{c_{n,k,2},c_{j,n};k=1,\dots,4,j=1,2\}$.

From (10), $1-c_{n,k,2}\sim\frac{2k}{n^2}$, and thus, for n large, larger bias and MSE reduction can be achieved using coefficients $c_{n,k_n,2}$. One may choose, for example, $k_n=\gamma n$, $0<\gamma<1$, to obtain $\frac{\operatorname{var}(s_n^2)-E(c_{n,k_n,2}s_n^2-\sigma^2)^2}{\operatorname{var}(s_n^2)}\sim\frac{2}{n}\frac{\gamma}{(1+\gamma)^2}[2-\frac{\gamma}{(1+\gamma)^2}]$.

5. AUGMENTED SAMPLES AND THE JACKKNIFE

5.1 Jackknife Estimates and Pseudovalues

The $(n-1)$ jackknife estimate t_n^J (Quenouille 1956) aims to reduce the bias of the estimate t_n of θ, and is the average of the pseudovalues $nt_n-(n-1)t_{n-1,i}$, $i=1,\dots,n$,

$$t_n^J=nt_n-\frac{1}{n}\sum_{i=1}^{n}(n-1)t_{n-1,i}$$

$$=t_n+(n-1)\left(t_n-\frac{\sum_{i=1}^n t_{n-1,i}}{n}\right); \qquad (23)$$

$t_{n-1,i}=t(X_1,\dots,X_{i-1},X_{i+1},\dots,X_n)$, $i=1,\dots,n$.

Equation (3) suggests using $(n+1)$-augmented samples to obtain the pseudo values

$$t_n+(n+1)(t_{n+1,i}-t_n)=(n+1)t_{n+1,i}-nt_n, \qquad (24)$$

whose average,

$$\tilde t_n^J=n^{-1}\sum_{i=1}^{n}[(n+1)t_{n+1,i}-nt_n]$$

$$=t_n+(n+1)\left(\frac{\sum_{i=1}^n t_{n+1,i}}{n}-t_n\right), \qquad (25)$$

is an alternative $(n+1)$ jackknife estimate; $t_{n+1,i}=t_{n+1}(X_1,\dots,X_n,X_i)$, $i=1,\dots,n$.

Note that in (23) and (25), the t_n corrections $(n-1)(t_n-\frac{\sum_{i=1}^n t_{n-1,i}}{n})$ and $(n+1)(\frac{\sum_{i=1}^n t_{n+1,i}}{n}-t_n)$ may have opposite signs, and thus $\tilde t_n^J$ may increase the bias of t_n.

It should be mentioned that Hinkley (1978) and Beran (1984) used $(n+k)$-augmented samples, $k=1,2$, to study the properties of t_n^J but not for the purpose of deriving estimates. Cabrera and Fernholz (1999) proposed a "target" estimate that, under model regularity conditions, has smaller bias and MSE than t_n.

5.2 MSE Reduction With t_n^J and $\tilde t_n^J$

From (23) [resp. (25)], it follows that t_n^J (resp. $\tilde t_n^J$) has a smaller MSE than t_n iff

$$E(t_n^J-t_n)^2+2(n-1)E(t_n-\theta)\left(t_n-\frac{\sum_{i=1}^n t_{n-1,i}}{n}\right)<0 \qquad (26)$$

$$\Bigg[\text{resp.}\quad E(\tilde t_n^J-t_n)^2 + 2(n+1)E(t_n-\theta)\left(\frac{\sum_{i=1}^n t_{n+1,i}}{n}-t_n\right)<0\Bigg]. \qquad (27)$$

Because

$$E(t_n-\theta)\left(t_n-\frac{\sum_{i=1}^n t_{n-1,i}}{n}\right)\quad\text{and}\quad E(t_n-\theta)\left(\frac{\sum_{i=1}^n t_{n+1,i}}{n}-t_n\right) \qquad (28)$$

may have opposite signs, only one of (26) and (27) may hold. This is confirmed in the following example and for smooth functionals in Section 5.4.

Example 1. Let $X_1,\dots,X_n$ be a sample from a normal distribution with unknown mean μ and variance σ^2, $\theta=\sigma^2$, and $t_n=\hat\sigma^2$, the maximum likelihood estimate. Because t_n has a smaller MSE than s_n^2 for every σ, (26) does not hold; s_n^2 is also the "target" estimate of σ^2 (Cabrera and Fernholz 1999, sec. 4.1, p. 1093, l. 6 and 7). Among all estimates of σ^2 with form $c\sum_{k=1}^n(X_k-\bar X_n)^2$, $\frac{1}{n+1}\sum_{k=1}^n(X_k-\bar X_n)^2$ minimizes the MSE and is the minimum risk-equivariant estimate (Lehmann 1983, p. 113). Because $\tilde t_n^J=\frac{1}{n+1}\sum_{k=1}^n(X_k-\bar X_n)^2$, (27) holds.

Indeed, for any sample $X_1,\dots,X_n,X_{n+1}$, it holds that $\bar X_{n+1}-\bar X_n=\frac{1}{n+1}(X_{n+1}-\bar X_n)$, implying that

$$(n+1)t_{n+1}=nt_n+\frac{n}{n+1}(X_{n+1}-\bar X_n)^2. \qquad (29)$$

For the $(n+1)$-augmented sample $X_1,\dots,X_n,X_{n+1}=X_i$, it follows from (29) that

$$(n+1)t_{n+1,i}=nt_n+\frac{n}{n+1}(X_i-\bar X_n)^2,$$

and therefore,

$$\tilde t_n^J=n^{-1}\sum_{i=1}^{n}[(n+1)t_{n+1,i}-nt_n]=\frac{1}{n+1}\sum_{i=1}^{n}(X_i-\bar X_n)^2.$$

$\tilde{t}_n^J$ has a larger bias than t_n, because $Et_n = \sigma^2 - \frac{\sigma^2}{n}$ and $E\tilde{t}_n^J = \sigma^2 - \frac{2\sigma^2}{n+1}$.

For the variance $\sigma^2 = \frac{m}{m-2}$ of the T_m distribution, estimates of the MSE based on 1,000 simulations indicate that $\tilde{t}_n^J$ has a smaller MSE than both t_n and t_n^J, $3 \le n \le 30$, $m = 3, 10, 20, 30$, and that (26) does not hold. The graph of the results is presented in Figure 1 and Remark 7 in the next section provides the explanation.

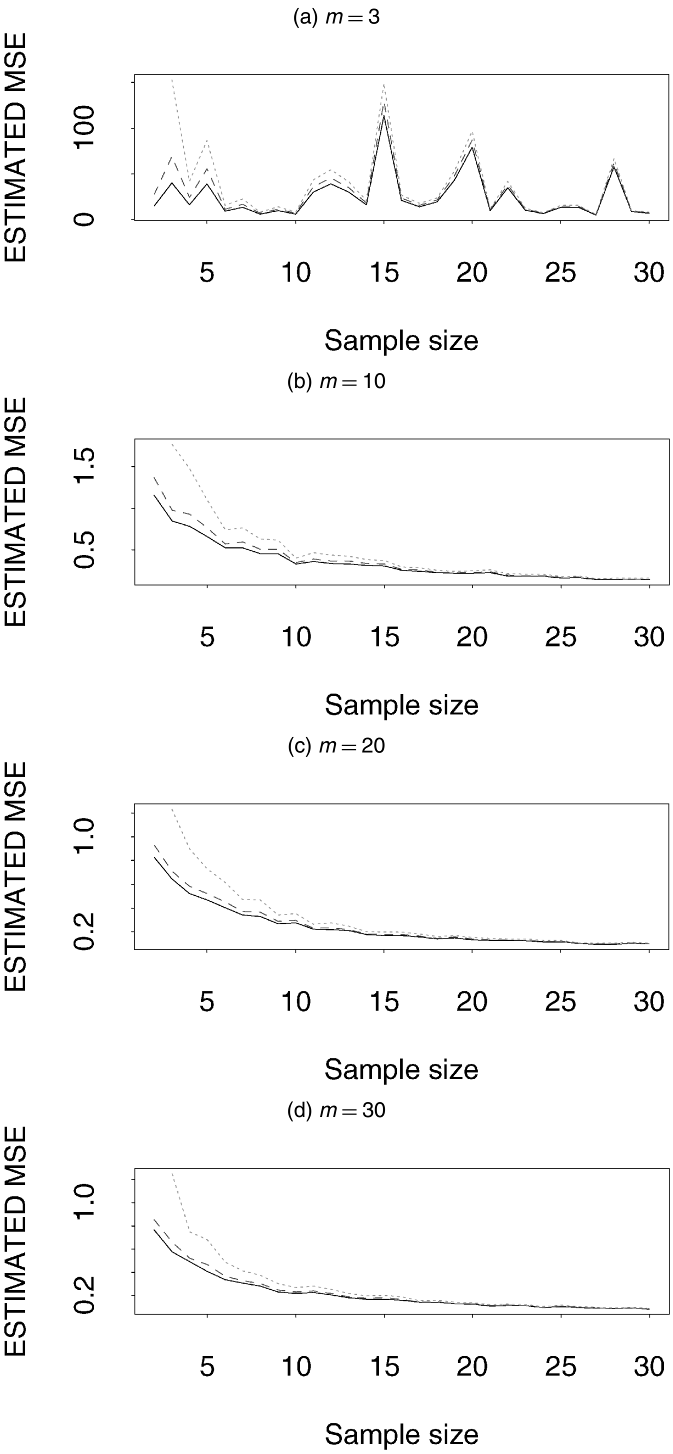

Figure 1. Variance Estimation: Comparing the MSE of t_n, t_n^J, and $\tilde{t}_n^J$ for the T_m Distribution. The estimated MSE of $\tilde{t}_n^J$, t_n^J, and of t_n are, respectively, the solid curve, the dotted curve, and the dashed curve; based on 1,000 simulations, $3 \le n \le 30$.

Remark 6. The jackknife covariance estimate obtained with $(n+1)$-augmented samples is $(n+1)^{-1}\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})$.

5.3 Calculations for Smooth Functionals

Let $\theta = t(F)$, and let $t_n = t(\hat{F}_n)$ be a smooth functional with second-order von Mises expansion

$$t_n = \theta + \frac{1}{n}\sum_{j=1}^n a_1(X_j) + \frac{1}{2n^2}\sum_{k=1}^n\sum_{l=1}^n a_2(X_k, X_l) + Rem_2, \quad (30)$$

and normalized kernels $a_1(x)$ and $a_2(x, y)$ such that

$$\begin{aligned} E_F a_1(X) &= 0, \\ a_2(X, Y) &= a_2(Y, X), \qquad \text{and} \\ E_F a_2(x, Y) &= 0. \end{aligned} \quad (31)$$

Under regularity conditions implying that $Rem_2 = o_P(n^{-1})$ (see Serfling 1980), it holds that

$$\begin{aligned} t_{n-1,i} &= t(\hat{F}_{n-1,i}) \\ &= \theta + \frac{1}{n-1}\left(\sum_{j=1}^n a_1(X_j) - a_1(X_i)\right) \\ &\quad + \frac{1}{2(n-1)^2}\left(\sum_{k=1}^n\sum_{l=1}^n a_2(X_k, X_l) \right. \\ &\qquad \left. - \sum_{l=1}^n a_2(X_i, X_l) - \sum_{k\neq i} a_2(X_k, X_i)\right) \\ &\quad + o_P(n^{-1}), \end{aligned}$$

thus obtaining from (23) and (30) that

$$\begin{aligned} &(n-1)\left(t_n - \frac{\sum_{i=1}^n t_{n-1,i}}{n}\right) \\ &\quad = \frac{1}{2n^2(n-1)}\sum_{k=1}^n\sum_{l=1}^n a_2(X_k, X_l) \\ &\qquad - \frac{1}{2n(n-1)}\sum_{i=1}^n a_2(X_i, X_i) + o_P(1). \end{aligned} \quad (32)$$

It also holds that

$$\begin{aligned} t_{n+1,i} &= t(\hat{F}_{n+1,i}) \\ &= \theta + \frac{1}{n+1}\left(\sum_{j=1}^n a_1(X_j) + a_1(X_i)\right) \\ &\quad + \frac{1}{2(n+1)^2}\left(\sum_{k=1}^n\sum_{l=1}^n a_2(X_k, X_l) + \sum_{l=1}^n a_2(X_i, X_l) \right. \\ &\qquad \left. + a_2(X_i, X_i) + \sum_{k=1}^n a_2(X_k, X_i)\right) \\ &\quad + o_P(n^{-1}), \end{aligned}$$

thus obtaining from (25) and (30) that

$$(n+1)\left(\frac{\sum_{i=1}^{n} t_{n+1,i}}{n} - t_n\right) = -\frac{1}{2n^2(n+1)}\sum_{k=1}^{n}\sum_{l=1}^{n} a_2(X_k, X_l) + \frac{1}{2n(n+1)}\sum_{i=1}^{n} a_2(X_i, X_i) + o_P(1). \quad (33)$$

For n large, (32) and (33) confirm the opposite signs of the t_n corrections in (23) and (25), and when in addition $ERem_2$ is negligible, they further confirm the opposite sign of their expectations.

For n large, the next proposition determines when $t_n^J, \tilde{t}_n^J$, and $\tilde{t}_n^1$ [see (2)] improve t_n.

Proposition 4. Assume that $ERem_2$ is negligible.

a. For t_n^J and $\tilde{t}_n^J$, it holds that

$$E(t_n^J - t_n)^2 \sim E(\tilde{t}_n^J - t_n)^2 \sim \frac{[Ea_2(X_1, X_1)]^2}{4n^2},$$

$$E(t_n^J - t_n)(t_n - \theta) \sim -\frac{Ea_1(X_1)a_2(X_1, X_1)}{2n^2} - \frac{[Ea_2(X_1, X_1)]^2}{4n^2},$$

and

$$E(\tilde{t}_n^J - t_n)(t_n - \theta) \sim \frac{Ea_1(X_1)a_2(X_1, X_1)}{2n^2} + \frac{[Ea_2(X_1, X_1)]^2}{4n^2}.$$

For n large, the MSE of $\tilde{t}_n^J$ (resp. t_n^J) is smaller than the MSE of t_n iff

$$Ea_1(X_1)a_2(X_1, X_1) < -.75[Ea_2(X_1, X_1)]^2$$

$$\big[\text{resp.} \quad Ea_1(X_1)a_2(X_1, X_1) > -.25[Ea_2(X_1, X_1)]^2\big].$$

b. For $\tilde{t}_n^1$, it holds that

$$E(\tilde{t}_n^1 - t_n)^2 \sim (4n^4)^{-1}[Ea_2(X_1, X_1)]^2$$

and

$$E(\tilde{t}_n^1 - t_n)(t_n - \theta) \sim (2n^3)^{-1}Ea_1(X_1)a_2(X_1, X_1) + (4n^3)^{-1}[Ea_2(X_1, X_1)]^2.$$

For n large, the MSE of $\tilde{t}_n^1$ is smaller than the MSE of t_n iff

$$Ea_1(X_1)a_2(X_1, X_1) < -.5[Ea_2(X_1, X_1)]^2.$$

For the proof see the Appendix.

Remark 7. For the normalized kernels a_1 and a_2 of the variance functional $\sigma^2 = t(F)$, it holds that $a_1(x_1) = (x_1 - \mu)^2 - \sigma^2$ and $a_2(x_1, x_1) = -2(x_1 - \mu)^2$. Thus $Ea_1(X_1)a_2(X_1, X_1) = -2(\mu_4 - \sigma^4)$ and $Ea_2(X_1, X_1) = -2\sigma^2$, and it follows from Proposition 4 that $\tilde{t}_n^J$ improves t_n and t_n^J if $\mu_4 > 2.5\sigma^4$. This holds for the T_m random variable with $\sigma^2 = \frac{m}{m-2}$ and $\mu_4 = \frac{3m^2}{(m-2)(m-4)}$, $m > 4$.

APPENDIX: PROOFS

Lemma A.1. For the kernels a_1 and a_2 in (30) that satisfy (31), it holds that

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} Ea_2(X_i, X_j)a_2(X_k, X_l) = \sum_{i=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} Ea_2(X_i, X_i)a_2(X_k, X_l) \quad (A.1)$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} Ea_2(X_i, X_i)a_2(X_j, X_j) = n(n-1)[Ea_2(X_1, X_1)]^2 + nEa_2^2(X_1, X_1), \quad (A.2)$$

$$\sum_{k=1}^{n}\sum_{l=1}^{n}\sum_{j=1}^{n} Ea_2(X_k, X_l)a_1(X_j) = \sum_{i=1}^{n}\sum_{j=1}^{n} Ea_2(X_i, X_i)a_1(X_j) = nEa_1(X_1)a_2(X_1, X_1). \quad (A.3)$$

Proof. It is sufficient to observe that (31) implies that

$$Ea_2(X_i, X_j)a_2(X_k, X_l) = 0, \qquad i \neq j \neq k \neq l;$$
$$Ea_2(X_i, X_j)a_2(X_i, X_l) = 0, \qquad i \neq j \neq l;$$
$$Ea_2(X_i, X_i)a_2(X_k, X_l) = 0, \qquad i \neq k \neq l;$$
$$Ea_2(X_i, X_i)a_2(X_i, X_l) = 0, \qquad i \neq l;$$
$$Ea_2(X_k, X_l)a_1(X_j) = 0, \qquad k \neq l \neq j;$$
$$Ea_2(X_k, X_k)a_1(X_j) = 0, \qquad k \neq j;$$
$$Ea_2(X_j, X_l)a_1(X_j) = 0, \qquad j \neq l.$$

Proof of Proposition 4. a. Equations (23), (32), and (33); negligibility of $ERem_2$, and (A.1)–(A.3) imply that

$$E(t_n^J - t_n)^2 \sim E(\tilde{t}_n^J - t_n)^2 \sim \frac{1}{4n^6}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} Ea_2(X_i, X_j)a_2(X_k, X_l) + \frac{1}{4n^4}\sum_{i=1}^{n}\sum_{j=1}^{n} Ea_2(X_i, X_i)a_2(X_j, X_j) - \frac{1}{2n^5}\sum_{i=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} Ea_2(X_i, X_i)a_2(X_k, X_l) \sim \frac{[Ea_2(X_1, X_1)]^2}{4n^2}, \quad (A.4)$$

$$E(t_n^J - t_n)(t_n - \theta) \sim \frac{1}{2n^4}\sum_{k=1}^{n}\sum_{l=1}^{n}\sum_{j=1}^{n} Ea_2(X_k, X_l)a_1(X_j) + \frac{1}{4n^5}\sum_{k=1}^{n}\sum_{l=1}^{n}\sum_{i=1}^{n}\sum_{j=1}^{n} Ea_2(X_k, X_l)a_2(X_i, X_j)$$

$$- \frac{1}{2n^3} \sum_{i=1}^{n} \sum_{j=1}^{n} E a_2(X_i, X_i) a_1(X_j)$$

$$- \frac{1}{4n^4} \sum_{i=1}^{n} \sum_{k=1}^{n} \sum_{l=1}^{n} E a_2(X_i, X_i) a_2(X_k, X_l)$$

$$\sim - \frac{E a_1(X_1) a_2(X_1, X_1)}{2n^2} - \frac{[E a_2(X_1, X_1)]^2}{4n^2}, \qquad \text{(A.5)}$$

and

$$E(\tilde{t}_n^J - t_n)(t_n - \theta) \sim -E(t_n^J - t_n)(t_n - \theta).$$

The conditions for t_n^J and $\tilde{t}_n^J$ to improve t_n for n large follow from (26) and (27).

b. From (33), it holds that

$$\tilde{t}_n^1 - t_n = -(2n^2(n+1)^2)^{-1} \sum_{k=1}^{n} \sum_{l=1}^{n} a_2(X_k, X_l) + (2n(n+1)^2)^{-1} \sum_{i=1}^{n} a_2(X_i, X_i) + Rem_2^*.$$

Negligibility of $ERem_2$ and (A.1)–(A.3) imply that

$$E(\tilde{t}_n^1 - t_n)^2 \sim (4n^8)^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} \sum_{l=1}^{n} E a_2(X_i, X_j) a_2(X_k, X_l)$$

$$+ (4n^6)^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} E a_2(X_i, X_i) a_2(X_j, X_j)$$

$$- (2n^7)^{-1} \sum_{i=1}^{n} \sum_{k=1}^{n} \sum_{l=1}^{n} E a_2(X_i, X_i) a_2(X_k, X_l)$$

$$\sim (4n^4)^{-1} [E a_2(X_1, X_1)]^2$$

and

$$E(\tilde{t}_n^1 - t_n)(t_n - \theta)$$

$$\sim -(2n^5)^{-1} \sum_{k=1}^{n} \sum_{l=1}^{n} \sum_{j=1}^{n} E a_2(X_k, X_l) a_1(X_j)$$

$$- (4n^6)^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} \sum_{l=1}^{n} E a_2(X_i, X_j) a_2(X_k, X_l)$$

$$+ (2n^4)^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} E a_1(X_j) a_2(X_i, X_i)$$

$$+ (4n^5)^{-1} \sum_{k=1}^{n} \sum_{l=1}^{n} \sum_{i=1}^{n} E a_2(X_k, X_l) a_2(X_i, X_i)$$

$$\sim (2n^3)^{-1} E a_1(X_1) a_2(X_1, X_1) + (4n^3)^{-1} [E a_2(X_1, X_1)]^2.$$

The result follows because $\tilde{t}_n^1$ improves t_n iff $E(\tilde{t}_n^1 - t_n)^2 + 2E(\tilde{t}_n^1 - t_n)(t_n - \theta) < 0$.

[Received January 2004. Revised February 2005.]

REFERENCES

Arnold, B. (1970), “Inadmissibility of the Usual Scale Estimate for a Shifted Exponential Distribution,” *Journal of the American Statistical Association*, 65, 1260–1264.

Beran, R. (1984), “Jackknife Approximations to Bootstrap Estimates,” *The Annals of Statistics*, 12, 101–118.

Brown, L. (1968), “Inadmissibility of the Usual Estimators of Scale Parameters in Problems With Unknown Location and Scale Parameters,” *The Annals of Mathematical Statistics*, 39, 29–48.

Cabrera, J., and Fernholz, L. T. (1999), “Target Estimation for Bias and MSE Reduction,” *The Annals of Statistics*, 27, 1080–1104.

Efron, B. (1979), “Bootstrap Methods: Another Look at the Jackknife,” *The Annals of Statistics*, 7, 1–26.

Hinkley, D. V. (1978), “Improving the Jackknife With Special Reference to Correlation Estimation,” *Biometrika*, 65, 13–21.

Lee, A. J. (1990), *U-Statistics: Theory and Practice*, New York: Marcel Dekker.

Lehmann, E. L. (1983), *Theory of Point Estimation*, New York: Wiley.

Quenouille, M. H. (1956), “Notes on Bias in Estimation,” *Biometrika*, 43, 353–360.

Rubin, D. B. (1987), *Multiple Imputation for Nonresponse in Surveys*, New York: Wiley.

Serfling, R. (1980), *Approximation Theorems of Mathematical Statistics*, New York: Wiley.

Shao, J., and Tu, D. (1995), *The Jackknife and Bootstrap*, New York: Springer-Verlag.

Stein, C. (1964), “Inadmissibility of the Usual Estimator for the Variance of a Normal Distribution With Unknown Mean,” *Annals of the Institute of Statistical Mathematics*, 16, 155–160.

Stuart, A., and Ord, K. (1994), *Kendall’s Advanced Theory of Statistics*, Vol. 1, London: Arnold.

Yatracos, Y. G. (2002), “Assessing the Quality of Bootstrap Samples and of the Bootstrap Estimates Obtained With Finite Resampling,” *Statistics and Probability Letters*, 59, 281–292.