Concomitants of Multivariate Order Statistics With Application to Judgment Poststratification

Xinlei WANG, Lynne STOKES, Johan LIM, and Min CHEN

We generalize the definition of a concomitant of an order statistic in the multivariate case, develop general expressions for its density, and establish related properties. We study the concomitant of a normal random vector in detail and discuss methods for calculating its moments. Furthermore, we apply the theory to develop new estimators of the mean from a judgment poststratified sample, where poststrata are formed by rank classes of auxiliary variables. Our estimators are shown to be more efficient than existing ones and robust against violations of the normality assumption. They are also well suited to applications requiring cost efficiency.

KEY WORDS: Best linear unbiased estimator; Gaussian quadrature; Least squares estimator; Ranked set sampling; Selection differential.

1. INTRODUCTION

Let $(X_h, Y_h)_{h=1}^{H}$ be H independent random vectors from a common bivariate distribution. Denote by $X_{(r:H)}$ the rth-ordered X-variate, $1 \le r \le H$. The concomitant of the rth-order statistic of X is defined to be the Y-variate paired with $X_{(r:H)}$ and is denoted by $Y_{[r:H]}$. Properties of concomitants have been studied by many authors (e.g., Bhattacharya 1974; Sen 1976, 1981; David, O'Connell, and Yang 1977; Yang 1977; Goel and Hall 1994; Nagaraja and David 1994); David and Nagaraja (2003, sec. 6.8) have provided an overview. Applications of concomitants include their use in estimating correlation (Barnett, Green, and Robinson 1976), in ranking and selection (Yeo and David 1984; David 1993), and in ranked set sampling (RSS) (Stokes 1977).

In this article we extend the definition of concomitants to the multivariate case, develop general expressions for their distributions, and establish related properties. That is, we study the distribution of an Y-variate associated with ordered components of an absolutely continuous $\mathbf{X}$-vector. For example, suppose that $\mathbf{X}_h$ contains the scores of the hth employee on two pre-employment screening measures and that Y_h contains his or her score on a later job performance measure, for a sample of H employees. Our theory would allow evaluation of the distribution of the job performance measure for an employee ranked, say, best on both screening tests. It would also allow comparison of that distribution to the concomitant job performance measure for an unscreened employee or to one scoring best on a single screening measure, to evaluate our selection procedure.

Our theory was motivated by an application of concomitants to judgment poststratification (JP–S) (MacEachern, Stasny, and Wolfe 2004), a method closely related to RSS. Both JP–S and RSS, are useful when the variable of interest, Y, is expensive to measure but can be ranked, at least approximately, much more cheaply. The ranking is referred to as judgment ranking. Both RSS and JP–S allow better estimation of the mean of Y, where the reduction in variance is provided by stratification. A ranked set sample can be thought of as a stratified sample, in which judgment ranks define the strata. A judgment poststratified sample can be thought of as a simple random sample (SRS), in which judgment ranks define the poststrata. This makes JP–S more practical than RSS for some applications, where the researcher may be amenable to beginning with an SRS with the option of using auxiliary data later but reluctant to beginning with a nonstandard design, such as a RSS (MacEachern et al. 2004).

A common method of judgment ranking in RSS is through an accessible auxiliary variable X, making Y a concomitant. We introduce a similar idea for JP–S in Section 5. As in conventional poststratification, we can use multiple auxiliary variables for forming poststrata. When the ranks of these auxiliary variables jointly define poststrata, we need the theory and properties of the concomitant of multivariate order statistics to develop and compute JP–S estimators of the mean and investigate their properties.

The article is organized as follows. In Section 2 we introduce concomitants of bivariate $\mathbf{X}$-vectors and present analytical results. In Section 3 we apply these results to the normal case and show how to compute means and variances of the concomitant. In Section 4 we extend our methods with straightforward modifications to the higher-dimensional case. In Section 5 we first review methods of mean estimation that have been suggested for JP–S samples using ranking information from more than one auxiliary variable, then propose new estimators with attractive properties that are available when certain distributional assumptions about the data can be made. We report results of simulation and empirical studies comparing the estimators. We conclude with a brief discussion in Section 6.

2. CONCOMITANT OF BIVARIATE ORDER STATISTIC

2.1 The General Theory

Let $(X_{h1}, X_{h2}, Y_h)_{h=1}^{H}$ be an iid random sample from a trivariate distribution, where the random variables X_1 and X_2 are absolutely continuous. Denote the order of X_{h1} among $X_{11}, \ldots, X_{H1}$ by $R_{h:H}$ and denote the order of X_{h2} among $X_{12}, \ldots, X_{H2}$ by $S_{h:H}$. We consider the random variable Y_h given the ranks $R_{h:H} = r$ and $S_{h:H} = s$, called the *concomitant of the rth-order statistic of* X_1 *and the sth-order statistic of* X_2, and denoted by $Y_{h[r,s:H]}$. For simplicity, we ignore the subscripts H and h and denote the concomitant as $Y_{[r,s]}$, its

Xinlei Wang is Assistant Professor (E-mail: *swang@mail.smu.edu*) and Lynne Stokes is Professor (E-mail: *slstokes@mail.smu.edu*), Department of Statistical Science, Southern Methodist University, Dallas, TX 75275. Johan Lim is Assistant Professor, Department of Applied Statistics, Yonsei University, Seoul, 120-749, Korea, on leave from the Department of Statistics, Texas A&M University, College Station, TX **???** (E-mail: *johanlim@stat.tamu.edu*). Min Chen is a Doctoral Candidate, McCombs School of Business, University of Texas, Austin, TX **???** (E-mail: *minchen@mail.utexas.edu*). The authors thank Shin-Jae Lee at Seoul National University for providing data on human teeth size.

Journal of the American Statistical Association
???? 0, Vol. 0, No. 00, Theory and Methods
DOI 10.1198/00

probability distribution function (pdf) as $f_{[r,s]}(y)$, the rank random variables as R and S, and the bivariate rank distribution $\Pr[R_{h:H}=r, S_{h:H}=s]$ as π_{rs}, whenever no ambiguity exists.

Theorem 1. Suppose that (X_1, X_2, Y) follows a trivariate distribution with a joint pdf $f(x_1, x_2, y)$. Let $m(X_1, X_2)$ and $v(X_1, X_2)$ denote the conditional mean and variance of Y, $E[Y|X_1, X_2]$ and $\mathrm{var}[Y|X_1, X_2]$. Then the distribution of the concomitant $Y_{[r,s]}$ among the H iid random vectors is given by

$$f_{[r,s]}(y) = \left\{ \sum_{k=\mathcal{L}}^{\mathcal{U}} C_k \int\int_{\mathcal{X}} \theta_1^k \theta_2^{r-1-k} \theta_3^{s-1-k} \theta_4^{H-r-s+1+k} \times f(x_1, x_2, y)\, dx_1\, dx_2 \right\} \times \left\{ \sum_{k=\mathcal{L}}^{\mathcal{U}} C_k \int\int_{\mathcal{X}} \theta_1^k \theta_2^{r-1-k} \theta_3^{s-1-k} \theta_4^{H-r-s+1+k} \times f(x_1, x_2)\, dx_1\, dx_2 \right\}^{-1}, \quad (1)$$

where $\mathcal{U} = \min(r-1, s-1)$ and $\mathcal{L} = \max(0, r+s-H-1)$, $\mathcal{X}$ is the support of the distribution of the $\mathbf{X}$-vector,

$$C_k = \frac{(H-1)!}{k!(r-1-k)!(s-1-k)!(H-r-s+1+k)!},$$

$$\theta_1(x_1, x_2) = \Pr(X_1 < x_1, X_2 < x_2),$$

$$\theta_2(x_1, x_2) = \Pr(X_1 < x_1, X_2 > x_2),$$

$$\theta_3(x_1, x_2) = \Pr(X_1 > x_1, X_2 < x_2),$$

and

$$\theta_4(x_1, x_2) = \Pr(X_1 > x_1, X_2 > x_2).$$

The mean and variance of $Y_{[r,s]}$ can be expressed by

$$\mu_{[r,s]} = E\big[m\big(X_{1(r,s)}, X_{2(r,s)}\big)\big] \quad (2)$$

and

$$\sigma^2_{[r,s]} = E\big[v\big(X_{1(r,s)}, X_{2(r,s)}\big)\big] + \mathrm{var}\big[m\big(X_{1(r,s)}, X_{2(r,s)}\big)\big], \quad (3)$$

where $(X_{1(r,s)}, X_{2(r,s)})$ are bivariate order statistics of (X_1, X_2).

Proof. First, we can write

$$f_{[r,s]}(y) = \frac{\int\int_{\mathcal{X}} f(y|x_1, x_2, r, s) p(r, s|x_1, x_2) f(x_1, x_2)\, dx_1\, dx_2}{\pi_{rs}} = \frac{\int\int_{\mathcal{X}} f(x_1, x_2, y) p(r, s|x_1, x_2)\, dx_1\, dx_2}{\pi_{rs}}, \quad (4)$$

because $f(y|x_1, x_2, r, s) = f(y|x_1, x_2)$. In the spirit of the work of David et al. (1977), it can be shown that

$$p(r, s|x_1, x_2) = \sum_{k=\mathcal{L}}^{\mathcal{U}} C_k \theta_1^k \theta_2^{r-1-k} \theta_3^{s-1-k} \theta_4^{H-r-s+1+k}, \quad (5)$$

yielding the numerator of (1). Similarly, we can show that its denominator, the bivariate rank distribution, is

$$\pi_{rs} = \int\int_{\mathcal{X}} p(r, s|x_1, x_2) f(x_1, x_2)\, dx_1\, dx_2 = \sum_{k=\mathcal{L}}^{\mathcal{U}} C_k \int\int_{\mathcal{X}} \theta_1^k \theta_2^{r-1-k} \theta_3^{s-1-k} \theta_4^{H-r-s+1+k} \times f(x_1, x_2)\, dx_1\, dx_2. \quad (6)$$

Further, for the mean of $Y_{[r,s]}$, we have

$$\mu_{[r,s]} = \int_{\mathcal{Y}} y f_{[r,s]}(y)\, dy = \int\int_{\mathcal{X}} m(x_1, x_2) f_{(r,s)}(x_1, x_2)\, dx_1\, dx_2 = E\big[m\big(X_{1(r,s)}, X_{2(r,s)}\big)\big],$$

where $\mathcal{Y}$ is the support of the distribution of Y, and $f_{(r,s)}(x_1, x_2)$ is the joint pdf of $(X_{1(r,s)}, X_{2(r,s)})$, that is, $f(x_1, x_2|R=r, S=s)$. Similarly, the variance of $Y_{[r,s]}$ can be written as (3).

Remark 1. If (X_1, X_2) and Y are independent [i.e., $f(x_1, x_2, y) = f(x_1, x_2) f(y)$], then it immediately follows from (1) that $f_{[r,s]}(y) = f(y)$.

Remark 2. Suppose that there exists a monotonic function $\psi(\cdot)$ such that $X_2 = \psi(X_1)$. In this case $f(x_1, x_2) = f(x_1) I(x_2 = \psi(x_1))$ and $f(x_1, x_2, y) = f(x_1, y) I(x_2 = \psi(x_1))$, where $I(\cdot)$ is the indicator function. Based on (1), it is easy to verify that both π_{rs} and $f_{[r,s]}(y)$ degenerate to the univariate case. When $\psi(\cdot)$ is increasing (or decreasing), if $r = s$ (or $r = H+1-s$), then $\pi_{rs} = 1/H$ and

$$f_{[r,s]}(y) = \int_{\mathcal{X}_1} f(y|x_1) f_{(r)}(x_1)\, dx_1 = f_{[r,\cdot]}(y),$$

where $f_{[r,\cdot]}(y)$ is the distribution for $Y_{[r,\cdot]}$, the concomitant of the rth-order statistic of X_1, and $f_{(r)}(x_1)$ is the distribution for $X_{1(r)}$, the rth-order statistic of X_1; otherwise, $\pi_{rs} = 0$, and $f_{[r,s]}(y)$ does not exist.

Consider the application of concomitants to ranking and selection of employees. Remarks 1 and 2 formalize the intuitive notion that if the screening tests are unrelated to the performance measure, then using them for selection is of no benefit, and if the screening tests are identical, then the second one is of no marginal benefit.

Example 1. Suppose that $U_1, U_2, Y \sim \mathrm{uniform}(0, 1)$ and iid. Let $X_i = (Y + U_i)/2$ for $i = 1, 2$. We illustrate Theorem 1 by deriving $f_{[1,1]}(y)$ and $f_{[1,2]}(y)$ for $H = 2$, where we condition on the ranks of $\mathbf{X} = (X_1, X_2)$. The theorem requires both joint $f(x_1, x_2, y)$ and marginal $f(x_1, x_2)$ densities. The former can be determined to be uniform over the region $J: 0 \le y \le 1$ and $y/2 \le x_1, x_2 \le (y+1)/2$, that is,

$$f(x_1, x_2, y) = 4 I_J(x_1, x_2, y). \quad (7)$$

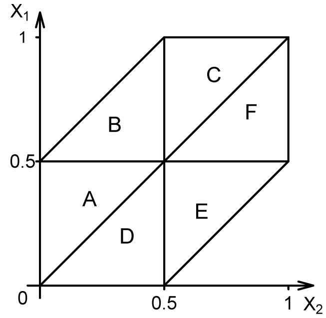

Figure 1. The Sampling Space of f(x1, x2).

The marginal density is found by integrating the joint density over the appropriate region to obtain

$$f(x_1,x_2)=\begin{cases} 8x_2 & (\text{area } A)\\ 8x_1 & (\text{area } D)\\ 4(2x_2-2x_1+1) & (\text{area } B)\\ 4(2x_1-2x_2+1) & (\text{area } E)\\ 8(1-x_1) & (\text{area } C)\\ 8(1-x_2) & (\text{area } F).\end{cases} \tag{8}$$

Areas A–F are shown in Figure 1. To find $f_{[1,1]}(y)$, first compute from (6) $\pi_{11}=\int\int_{\mathcal{X}}\theta_4 f(x_1,x_2)\,dx_1\,dx_2$, where θ_4 must be determined separately for each area of Figure 1 (e.g., in area A, $\theta_4=1-2x_1^2-2x_2^2+4x_1x_2^2-4x_2^3/3$). The result is that $\pi_{11}=1/3$. The numerator of (1) becomes

$$f(y,R=1,S=1)=\int_{y/2}^{(y+1)/2}\int_{y/2}^{(y+1)/2}4\theta_4\,dx_1\,dx_2,$$

which, after some calculation, can be shown to be

$$f_{[1,1]}(y)=\frac{1}{20}(43-45y-30y^2+50y^3-15y^4) \tag{9}$$

for $0\le y\le 1$. Noting that $\pi_{12}=1/2-\pi_{11}=1/6$, a similar calculation yields

$$f_{[1,2]}(y)=\frac{1}{10}(7+15y-30y^3+15y^4). \tag{10}$$

Suppose that (X_1,X_2,Y), with joint distribution (7), denote scores on two screening tests and a performance measure for an employee. The advantage in performance expected from an employee who performs best (in this case the lowest value, as for speed tests) on one screening test can be measured by the selection differential, which Nagaraja (1982) defined as

$$\eta_{[1]}=\frac{\mu_{[1,\cdot]}-\mu_y}{\sigma_y}, \tag{11}$$

where $\mu_y=E(Y)$, $\sigma_y^2=\operatorname{var}(Y)$, and $\mu_{[1,\cdot]}=E(Y_{[1,\cdot]})$. From (9) and (10), we have

$$f_{[1,\cdot]}(y)=\frac{\pi_{11}f_{[1,1]}(y)+\pi_{12}f_{[1,2]}(y)}{\pi_{11}+\pi_{12}}=\frac{1}{3}(5-3y-3y^2+2y^3)$$

for $0\le y\le 1$, yielding $\mu_{[1,\cdot]}=23/60$ and

$$\eta_{[1]}=\left(\frac{23}{60}-\frac{1}{2}\right)\Big/\sqrt{\frac{1}{12}}\approx -.40.$$

Generalizing the selection differential (11) to the bivariate concomitant, we compute

$$\eta_{[1,1]}=\left(\frac{13}{40}-\frac{1}{2}\right)\Big/\sqrt{\frac{1}{12}}\approx -.60$$

from (9). Comparing these two shows that the additional screening test improves selectivity by about 50%.

It is sometimes straightforward to calculate moments of the concomitant directly from the density (1), as in Example 1. In other cases, calculation is easier using (2) and (3); an example of this is given in Section 3.

2.2 Simplifying Properties

Computing densities of concomitants using Theorem 1 is tedious. However, symmetry in the distribution of (X_1,X_2,Y) can be exploited to reduce the number of calculations required to obtain densities and moments for the set of all concomitants. In this section we present some useful results for that purpose.

First, we make some observations about the rank distribution π_{rs}. For convenience, let $\bar r\equiv H+1-r$ and $\bar s\equiv H+1-s$.

Property 1. A monotonically increasing transformation on X_1 or X_2 does not change π_{rs}. A monotonically decreasing transformation on X_1 leads to $\pi'_{rs}=\pi_{\bar r s}$, and a monotonically decreasing transformation on X_2 leads to $\pi'_{rs}=\pi_{r\bar s}$, where π'_{rs} is the bivariate rank distribution based on the transformed variables.

Property 2. If the joint pdf $f(x_1,x_2)$ of X_1 and X_2 is symmetric [i.e., $f(x_1,x_2)=f(x_2,x_1)$], then $\pi_{rs}=\pi_{sr}$.

Property 3. If $f(x_1,x_2)=f(-x_1,-x_2)$, then $\pi_{rs}=\pi_{\bar r\bar s}$.

Property 1 is obvious from observing that the rank of any observation is invariant to a monotonically increasing transformation. The other two properties have been proven by David et al. (1977).

Example 2 (Example 1 continued). Properties 1–3 can be used to calculate π_{21} and π_{22}. Observe from (8) that $f(x_1,x_2)=f(x_2,x_1)$; thus $\pi_{21}=\pi_{12}=1/6$ from Property 2. Define $Z_i=X_i-1/2$ for $i=1,2$. According to Property 1, (Z_1,Z_2) has the same rank distribution as (X_1,X_2). Because the joint density of Z_1 and Z_2 satisfies $g(z_1,z_2)=g(-z_1,-z_2)$, Property 3 yields $\pi_{22}=\pi_{11}=1/3$.

Next, we observe some properties of the concomitant distribution that follow directly from Theorem 1.

Corollary 1. Suppose that there exist monotonic functions $\psi_1(\cdot)$ and $\psi_2(\cdot)$ such that $Z_1=\psi_1(X_1)$, $Z_2=\psi_2(X_2)$, and the joint pdf of Z_1, Z_2, and Y satisfies $g(z_1,z_2,y)=g(z_2,z_1,y)$. Then (a) if both $\psi_1\uparrow$ (increasing) and $\psi_2\uparrow$ or both $\psi_1\downarrow$ (decreasing) and $\psi_2\downarrow$, then $f_{[r,s]}(y)=f_{[s,r]}(y)$, and (b) if $\psi_1\uparrow$ and $\psi_2\downarrow$ or $\psi_1\downarrow$ and $\psi_2\uparrow$, then $f_{[r,s]}(y)=f_{[\bar s,\bar r]}(y)$.

Corollary 2. Suppose that there exist monotonic functions $\psi_1(\cdot)$ and $\psi_2(\cdot)$ such that $Z_1=\psi_1(X_1)$, $Z_2=\psi_2(X_2)$. Then (a) if the joint pdf of Z_1, Z_2, and Y satisfies $g(z_1,z_2,y)=g(-z_1,-z_2,y)$, then $f_{[r,s]}(y)=f_{[\bar r,\bar s]}(y)$, and (b) if $g(z_1,z_2,\mu_y+d)=g(-z_1,-z_2,\mu_y-d)$, then $f_{[r,s]}(\mu_y+d)=f_{[\bar r,\bar s]}(\mu_y-d)$.

Example 3 (Example 1 continued). From (7), $f(x_1, x_2, y) = f(x_2, x_1, y)$. Thus Corollary 1 yields $f_{[2,1]}(y) = f_{[1,2]}(y)$. Because the joint density of Z_1, Z_2, and Y satisfies $g(z_1, z_2, 1/2 + d) = g(-z_1, -z_2, 1/2 - d)$, Corollary 2 yields $f_{[2,2]}(1/2 + d) = f_{[1,1]}(1/2 - d)$. Let $y = 1/2 + d$; then $f_{[2,2]}(y) = f_{[1,1]}(1 - y) = (3 + 15y + 30y^2 + 10y^3 - 15y^4)/20$.

In the following theorem, we establish properties of the mean $\mu_{[r,s]}$ and variance $\sigma^2_{[r,s]}$ of a concomitant, where the distribution of Y is not required to be symmetric.

Theorem 2. Suppose that there exist monotonic functions $\psi_1(\cdot)$ and $\psi_2(\cdot)$ such that (a) $Z_1 = \psi_1(X_1)$, $Z_2 = \psi_2(X_2)$ and their joint pdf is symmetric about 0, that is, $g(z_1, z_2) = g(-z_1, -z_2)$, and (b) $E(Y|Z_1 = z_1, Z_2 = z_2)$ is a linear function of z_1 and z_2. Then the mean of the concomitant of bivariate order statistics of (X_1, X_2) satisfies

$$\mu_{[r,s]} + \mu_{[\bar{r},\bar{s}]} = 2\mu_y \tag{12}$$

for $r \in \{1, \ldots, H\}$ and $s \in \{1, \ldots, H\}$. Furthermore, if $\operatorname{var}(Y|z_1, z_2) = \operatorname{var}(Y|-z_1, -z_2)$, then the variance of the concomitant satisfies

$$\sigma^2_{[r,s]} = \sigma^2_{[\bar{r},\bar{s}]}. \tag{13}$$

For the proof see Appendix A.

Note that in Theorem 2, if H is odd, then $\mu_{[(H+1)/2,(H+1)/2]} = \mu_y$.

Example 4. Consider a type of regression setup: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$ where ϵ is independent of X_1 and X_2 and follows a distribution with mean 0. Also, assume that X_1 and X_2 can be linearly transformed so that their joint pdf after transformation is symmetric about 0. Then (12) and (13) hold.

More results about $\mu_{[r,s]}$ and $\sigma^2_{[r,s]}$ can be obtained easily from Corollaries 1 and 2. For example, (12) and (13) follow directly from part (b) of Corollary 2.

3. THE NORMAL CASE

Here we discuss the special case of the concomitant of the order statistics of a bivariate normal random vector. Let (X_1, X_2, Y) be trivariate normal with means μ_1, μ_2, and μ_y; variances σ_1^2, σ_2^2, and σ_y^2; and correlations ρ_{12}, ρ_{1y}, and ρ_{2y}. Properties of the normal distribution allow the conditional mean and variance of Y given $\mathbf{X} = (x_1, x_2)$ to be written as

$$m(x_1, x_2) = \mu_y + (\tau_1 z_1 + \tau_2 z_2)\sigma_y$$

and

$$v(x_1, x_2) = (1 - \tau_1\rho_{1y} - \tau_2\rho_{2y})\sigma_y^2,$$

where $\tau_1 = (\rho_{1y} - \rho_{2y}\rho_{12})/(1 - \rho_{12}^2)$, $\tau_2 = (\rho_{2y} - \rho_{1y}\rho_{12})/(1 - \rho_{12}^2)$, and $z_j = (x_j - \mu_j)/\sigma_j$, $j = 1, 2$. From (2),

$$\begin{aligned}\mu_{[r,s]} &= \mu_y + \sigma_y\big[\tau_1 E\big(Z_{1(r,s)}\big) + \tau_2 E\big(Z_{2(r,s)}\big)\big] \\ &= \mu_y + \sigma_y\big[\tau_1 E\big(Z_{1(r,s)}\big) + \tau_2 E\big(Z_{1(s,r)}\big)\big],\end{aligned} \tag{14}$$

where $(Z_1, Z_2)^T$ has the standard bivariate normal distribution with correlation ρ_{12} and $(Z_{1(r,s)}, Z_{2(r,s)})$ are the bivariate order statistics of (Z_1, Z_2) with joint density $g_{(r,s)}(z_1, z_2)$. The second line follows because $E(Z_{2(r,s)}) = E(Z_{1(s,r)})$. From (3),

$$\begin{aligned}\sigma^2_{[r,s]} = \big[&\tau_1^2 \operatorname{var}\big(Z_{1(r,s)}\big) + \tau_2^2 \operatorname{var}\big(Z_{1(s,r)}\big) \\ &+ 2\tau_1\tau_2 \operatorname{cov}\big(Z_{1(r,s)}, Z_{2(r,s)}\big) + 1 \\ &- \tau_1\rho_{1y} - \tau_2\rho_{2y}\big]\sigma_y^2.\end{aligned} \tag{15}$$

Noting that τ_1 and τ_2 are the standardized regression coefficients, (14) and (15) suggest that the concomitant and related order statistics retain the same linearity as in multiple regression.

To calculate $\mu_{[r,s]}$ and $\sigma^2_{[r,s]}$ using (14) and (15), values of the means, variances, and covariances of the bivariate standard normal order statistics are needed. Tables of these moments for $H = 2$, 3, and 4 are available in an earlier work (Wang and Stokes 2005). The method that we used to obtain these tables is briefly outlined here. To calculate the mean, one must evaluate

$$\begin{aligned}&E\big(Z_{1(r,s)}\big) \\ &\quad= \int\int_{\mathcal{R}^2} z_1 g_{(r,s)}(z_1, z_2)\, dz_1\, dz_2 \\ &\quad= \frac{\int\int_{\mathcal{R}^2} z_1 p(r, s|z_1, z_2)\phi(z_1, z_2)\, dz_1\, dz_2}{\pi_{rs}} \\ &\quad= \left(\frac{\sqrt{2}}{\pi}\sum_{k=\mathcal{L}}^{\mathcal{U}} C_k \int\int_{\mathcal{R}^2} u_1\theta_1^k\theta_2^{r-1-k}\theta_3^{s-1-k}\theta_4^{H-r-s+1+k}\right. \\ &\qquad\qquad \left.\times \exp[-(u_1^2 + u_2^2)]\, du_1\, du_2\right)\Big/\pi_{rs},\end{aligned} \tag{16}$$

where $\phi(\cdot,\cdot)$ is the joint pdf of the standard bivariate normal distribution and $\theta_i \equiv \theta_i(\sqrt{2}\mu_1, \sqrt{2}\rho_{12}\mu_1 + \sqrt{2(1-\rho_{12}^2)}\mu_2)$, for $i = 1, \ldots, 4$. The second line follows from (5) after the change of variables $z_1 = \sqrt{2}u_1$ and $z_2 = \sqrt{2}\rho_{12}u_1 + \sqrt{2(1-\rho_{12}^2)}u_2$. Similar expressions can be written for π_{rs}, $E(Z^2_{1(r,s)})$ and $E(Z_{1(r,s)}Z_{2(r,s)})$. These were all evaluated numerically using Gaussian quadrature. For example, the numerator of (16) was approximated by

$$\frac{\sqrt{2}}{\pi}\sum_{k=\mathcal{L}}^{\mathcal{U}} C_k \sum_{j=1}^{M}\sum_{i=1}^{M}\{\omega_i\omega_j t_i\theta_1^k\theta_2^{r-1-k}\theta_3^{s-1-k}\theta_4^{H-r-s+1+k}\},$$

where $\theta_l \equiv \theta_l(\sqrt{2}t_i, \sqrt{2}\rho_{12}t_i + \sqrt{2(1-\rho_{12}^2)}t_j)$ for $l = 1, \ldots, 4$, t_i is the ith zero of the Hermite polynomial $H_M(t)$, and ω_i is the ith weight factor. (Tables of t_i and ω_i are available for $M = 1$–20 in Salzer, Zucker, and Capuano 1952.)

Due to the symmetry of the normal density, the properties in Section 2.2 can be used to reduce the number of numerical evaluations needed to obtain $\mu_{[r,s]}$ and $\sigma^2_{[r,s]}$. Property 1 shows (by standardizing X_1 and X_2) that π_{rs} is related only to ρ_{12} and so can be calculated based on Z_1 and Z_2. From Properties 2 and 3, $\pi_{rs} = \pi_{sr} = \pi_{\bar{r}\bar{s}}$. Theorem 2 shows that $E(Z_{1(r,s)}) + E(Z_{1(\bar{r},\bar{s})}) = 0$, and so one need only calculate elements in the upper triangular matrix of $[E(Z_{1(r,s)})]_{H\times H}$. The number of covariance (variance) calculations needed is reduced by observing that $\operatorname{cov}(Z_{1(r,s)}, Z_{2(r,s)}) = \operatorname{cov}(Z_{1(s,r)}, Z_{2(s,r)})$ and

$\operatorname{cov}(Z_{k(r,s)}, Z_{l(r,s)}) = \operatorname{cov}(Z_{k(\bar{r},\bar{s})}, Z_{l(\bar{r},\bar{s})})$ for any k, $l = 1, 2$, the latter of which is an intermediate result in the proof of Theorem 2.

Example 5. Suppose that (X_1, X_2, Y), the joint distribution of which is trivariate normal, denote scores on two screening tests and a performance measure for an employee. The advantage in performance expected from an employee who performs best (in this case the lowest value, as for speed tests) on both screeners among $H = 2, 3$, or 4 competitors can be measured by the selection differential. It can be written, using (14), as

$$\eta_{[1,1]} = \frac{\mu_{[1,1]} - \mu_y}{\sigma_y} = \frac{\rho_{1y} + \rho_{2y}}{1 + \rho_{12}} E\big(Z_{1(1,1)}\big). \tag{17}$$

This expression shows that the selection differential increases in magnitude as screening tests grow more effective (i.e., larger values of ρ_{1y} and ρ_{2y}) and as the number of competitors increases [because $E(Z_{1(1,1)})$ is an increasing function of H]. One would also expect less advantage as screening tests become more similar (i.e., ρ_{12} increases), but this is not clear from (17) because $|E(Z_{1(1,1)})|$ increases in ρ_{12}. Figure 2(a) displays $|\eta_{[1,1]}|$ for $H = 2, 3$, and 4 as functions of ρ_{12} for two moderately effective screening tests ($\rho_{1y} = \rho_{2y} = .5$). It confirms that the second test is less useful for selection as it becomes more similar to the first.

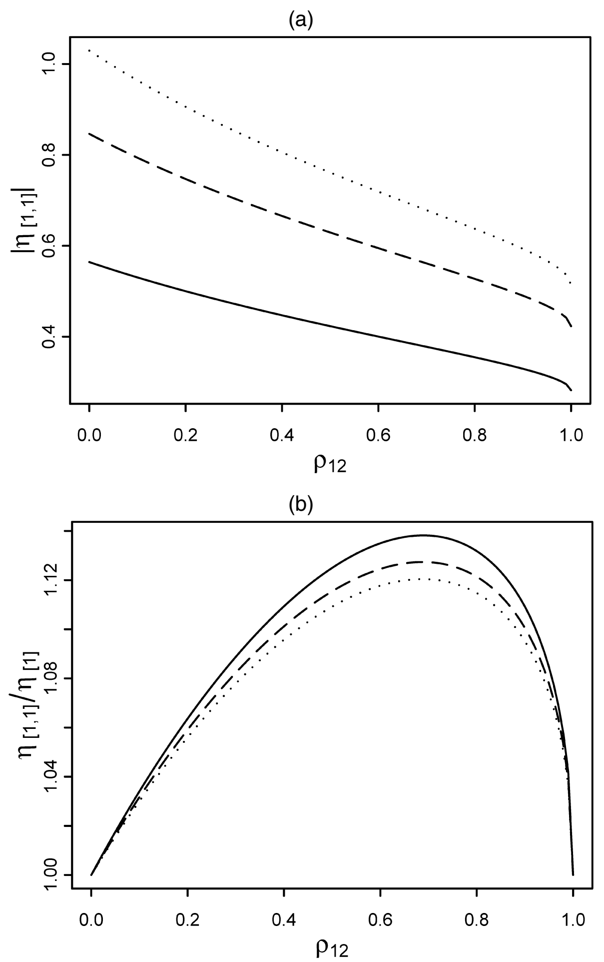

Figure 2. An Example for the Normal Case. (a) Selection differential for pairs of moderately efficient ($\rho_{1y} = \rho_{2y} = .5$) screening tests. (b) Ratio of selection differential for one and two screeners when one test is perfect. (—— $H = 2$; — — $H = 3$; · · · · $H = 4$.)

One might expect that if the first screening test were perfect ($\rho_{1y} = 1$), then the second one would provide no advantage in the selection process. This is incorrect. To see why, note first that $\rho_{2y} = \rho_{12}$ in this case. Then, from (17), $\eta_{[1,1]} = E(Z_{1(1,1)})$ and $\eta_{[1]} = E(Z_{(1)})$, where $Z_{(1)}$ is the first-order statistic of a standard normal random variable. Figure 2(b) shows the ratio $\eta_{[1,1]}/\eta_{[1]}$ as a function of ρ_{12} for $H = 2, 3, 4$. The ratio is >1, but the advantage is greatest when the second screener has a correlation of around .70 with the performance measure. Its advantage diminishes to 0 as ρ_{12} $(= \rho_{2y})$ increases to 1. An intuitive explanation for this is that even perfect ranking information does not provide complete information about the mean. The second ranking variable, to the extent that its information differs from that of the first, can still improve estimation. Note also that the finer are the perfect ranker's poststrata (larger H), the less additional information remains for the second ranker to provide.

4. EXTENSION TO THE MULTIVARIATE CASE

In the previous sections we investigated the concomitant of bivariate order statistics. We now seek analytic expressions for the general case, the concomitant of multivariate order statistics where the number of X variables ≥ 2.

Let $(\mathbf{X}_h, Y_h)_{h=1}^H$ be an iid random sample from a multivariate distribution with a joint pdf $f(\mathbf{x}, y)$, where $\mathbf{X}_h^T$ is an absolutely continuous vector of length m. Denote the order of X_{hi} among $X_{1i}, \dots, X_{Hi}$ by R_{hi} and the rank vector associated with $\mathbf{X}_h$ by $\mathbf{R}_h^T = (R_{hi})_{i=1}^m$. Given a fixed H, we consider the concomitant of multivariate order statistics of $\mathbf{X}_h$, that is, the random variable Y_h given $\mathbf{R}_h$. To obtain its density, (4) can be generalized as

$$f_{[\mathbf{r}]}(y) = \frac{\int_{\mathcal{X}} f(\mathbf{x}, y) p(\mathbf{r}|\mathbf{x})\, d\mathbf{x}}{\pi_{\mathbf{r}}}, \tag{18}$$

where $\pi_{\mathbf{r}} = \int_{\mathcal{X}} p(\mathbf{r}|\mathbf{x}) f(\mathbf{x})\, d\mathbf{x}$. Only $p(\mathbf{r}|\mathbf{x})$ is needed, which can be computed by recursion. To illustrate the idea, we describe the method for deriving $p(\mathbf{r}|\mathbf{x})$ for $m = 3$ from that for $m = 2$.

Following in David et al. (1977), we represent the ways in which the compound event $R_{h1} = r_1$ and $R_{h2} = r_2$ given $X_{h1} = x_1$ and $X_{h2} = x_2$ can occur in the following 2×2 table:

	$X_{h'2} < x_2$	$X_{h'2} > x_2$	
$X_{h'1} < x_1$	k	$r_1 - 1 - k$	$r_1 - 1$
$X_{h'1} > x_1$	$r_2 - 1 - k$	$H - r_1 - r_2 + 1 + k$	$H - r_1$
	$r_2 - 1$	$H - r_2$	$H - 1$

Then $p(r_1, r_2|x_1, x_2)$, can be obtained from the multinomial distribution, with the four outcomes defined by the cells of the table. We split each of the four cells further by a third variable:

	$X_{h'2} < x_2$		$X_{h'2} > x_2$		
	$X_{h'3} < x_3$	$X_{h'3} > x_3$	$X_{h'3} < x_3$	$X_{h'3} > x_3$	
$X_{h'1} < x_1$	l_0	$k - l_0$	l_1	$r_1 - 1 - k - l_1$	$r_1 - 1$
$X_{h'1} > x_1$	l_2	$r_2 - 1 - k - l_2$	$r_3 - 1 - (l_0 + l_1 + l_2)$	$H - r_1 - r_2 - r_3 + 2 + (k + l_0 + l_1 + l_2)$	$H - r_1$
	$r_2 - 1$		$H - r_2$		$H - 1$

After splitting, label the eight cells $1, \dots, 8$ and denote the number of observations in the jth cell by t_j, $1 \le j \le 8$ (i.e., $t_1 = l_0$, $t_2 = k - l_0$, $t_3 = l_1, \dots, t_8 = H - r_1 - r_2 - r_3 + 2 + k + l_0 + l_1 + l_2$). Thus

$$p(r_1, r_2, r_3 | x_1, x_2, x_3) = \sum_{k, l_0, l_1, l_2 \in \mathcal{A}} \left\{ C_{k, l_0, l_1, l_2} \prod_{j=1}^{8} [\theta_j(x_1, x_2, x_3)]^{t_j} \right\},$$

where $\mathcal{A}$ is an integer set $\{k, l_0, l_1, l_2 | k \ge 0; l_0 \ge 0; l_1 \ge 0; l_2 \ge 0; t_j \ge 0, \text{ for } 1 \le j \le 8\}$ and $C_{k,l_0,l_1,l_2} \equiv (H-1)!/\{\prod_{j=1}^{8} t_j!\}$. Define $\theta_j(x_1, x_2, x_3)$ as the corresponding probability in the jth cell, that is, $\theta_1(x_1, x_2, x_3) \equiv \Pr(X_1 < x_1, X_2 < x_2, X_3 < x_3)$, $\theta_2(x_1, x_2, x_3) \equiv \Pr(X_1 < x_1, X_2 < x_2, X_3 > x_3)$, and so on.

Similarly, $p(\mathbf{r}|\mathbf{x}, x_{m+1})$ can be derived from $p(\mathbf{r}|\mathbf{x})$ by partitioning each of the 2^m cells into two subcells based on the value of x_{m+1} and then applying the multinomial distribution with 2^{m+1} possible outcomes. From (18), we can obtain an analytic expression for $f_{[\mathbf{r}]}(y)$ when the length of $\mathbf{x} > 2$ by recursion. With trivial modifications, the properties discussed in Section 2 can be generalized to $m > 2$.

Consider the normal case, where

$$(\mathbf{X}, Y)^T \sim \mathrm{N}\left(\begin{pmatrix} \boldsymbol{\mu}_{\mathbf{x}} \\ \mu_y \end{pmatrix}, \operatorname{diag}(\boldsymbol{\sigma}_{\mathbf{x}}, \sigma_y) \begin{pmatrix} \boldsymbol{\rho}_{\mathbf{x}} & \boldsymbol{\rho}_{\mathbf{x}y} \\ \boldsymbol{\rho}_{\mathbf{x}y}^T & 1 \end{pmatrix} \operatorname{diag}(\boldsymbol{\sigma}_{\mathbf{x}}, \sigma_y) \right). \tag{19}$$

The mean and variance of the concomitants can be expressed as generalizations of (14) and (15) as

$$\mu_{[\mathbf{r}]} = \mu_y + \sigma_y \boldsymbol{\rho}_{\mathbf{x}y}^T \boldsymbol{\rho}_{\mathbf{x}}^{-1} E(\mathbf{Z}_{(\mathbf{r})}) \tag{20}$$

and

$$\sigma_{[\mathbf{r}]}^2 = \sigma_y^2 + \sigma_y^2 \boldsymbol{\rho}_{\mathbf{x}y}^T \boldsymbol{\rho}_{\mathbf{x}}^{-1} \left[\operatorname{cov}(\mathbf{Z}_{(\mathbf{r})}) \boldsymbol{\rho}_{\mathbf{x}}^{-1} - \mathbf{I} \right] \boldsymbol{\rho}_{\mathbf{x}y}, \tag{21}$$

where $\mathbf{Z}_{(\mathbf{r})}$ is a vector of multivariate order statistics of $\mathbf{Z}$ that follows the standard multivariate normal distribution with the correlation matrix $\boldsymbol{\rho}_{\mathbf{x}}$. Calculation of the moments of the standard normal multivariate order statistic involves $p(\mathbf{r}|\mathbf{z})$, for example,

$$E(Z_{1(\mathbf{r})}) = \frac{\int_{\mathcal{R}^m} z_1 p(\mathbf{r}|\mathbf{z}) \phi(\mathbf{z})\, d\mathbf{z}}{\pi_{\mathbf{r}}},$$

which can be evaluated numerically using Gaussian quadrature, similar to the method used in the bivariate case.

5. APPLICATION TO JUDGMENT POSTSTRATIFICATION

Here we apply the theory of the previous sections to suggest new estimators of the mean from J–PS samples and to examine their properties.

5.1 Background

MacEachern et al. (2004) introduced JP–S sampling as an alternative to RSS for estimating the mean of Y, which is expensive to quantify but relatively cheap to rank by judgment. To obtain a JP–S sample, one first draws an SRS of n units from a population and records the value of Y for each, denoted by y_i, $1 \le i \le n$. For each measured unit i, an additional sample of size $H - 1$ is chosen at random, and the order of y_i among the H units is assessed by some inexpensive (and likely imperfect) ranking method not requiring measurement of the $H - 1$ units. This rank information is used to classify the n measured units into H poststrata. MacEachern et al. (2004) proposed as an estimator of μ_y,

$$\hat{\mu}_y = \frac{1}{H} \sum_{h=1}^{H} \frac{\sum_{i=1}^{n} y_i I_{ih}}{\sum_{i=1}^{n} I_{ih}}, \tag{22}$$

where $I_{ih} = 1$ if y_i is assigned into the stratum of rank h and $I_{ih} = 0$ otherwise.

RSS differs from JP–S sampling in that in the former, judgment ranking of the group of H sample units occurs first, and then a specified rank is designated for measurement from the group. Ranking of groups of size H continues until some specified number of units of each judgment rank are quantified. The unweighted mean of such a sample can be shown to be unbiased for μ_y and to have smaller variance than an SRS of an equal number of measured observations (Dell and Clutter 1972). MacEachern et al. (2004) showed that the asymptotic relative efficiency of $\hat{\mu}_y$ to this RSS estimator is 1.

Although RSS is described as using subjective judgment in ranking, applications have often used the rank of an easily observed auxiliary variable as a proxy for the rank of Y. But what if information about the rank of more than one auxiliary variable is available? It is difficult to use this information in RSS, because any particular vector of ranks cannot be guaranteed to occur, and so prespecifying sample sizes from strata defined by joint ranks is infeasible unless a multiple-layer design of RSS is used (Chen and Shen 2003). In contrast, JP–S uses rank information only for estimation, not for sample design. Our goal is to use such rank information along with the measured y_i to estimate μ_y. An example of such data was discussed by Chen (2002) for estimating mean age of a population of fish. Aging a fish is expensive, but the rank of a fish's length and width, which are correlated with age, among a group of H fish can be easily obtained.

MacEachern et al. (2004) also cited the ability to use more than one ranker as an advantage for JP–S over RSS. In the case where assessments of ranks are available from m rankers (or auxiliary variables), they proposed as an estimator of μ_y,

$$\hat{\mu}_M^{(m)} = \frac{1}{H} \sum_{h=1}^{H} \frac{\sum_{i=1}^{n} y_i \hat{p}_{ih}}{\sum_{i=1}^{n} \hat{p}_{ih}}, \tag{23}$$

where $\hat{p}_{ih}$ is the proportion of rankers classifying y_i as having rank h; that is, they prorate the measured value among the poststrata receiving any "votes" from a ranker. This estimator requires no distributional assumptions for its justification. When $m = 1$, $\hat{\mu}_M^{(m)}$ degenerates to (22).

5.2 New Estimators of Mean

In this section we propose several new estimators of μ_y based on data from a JP–S sample, where poststrata are defined on ranks of m auxiliary variables. Our proposed estimators are designed to take advantage of knowledge of the distribution of the concomitant. Here we restrict attention to the most tractable yet important case, the multivariate normal distribution. We first assume that σ_y, $\boldsymbol{\rho}_{\mathbf{xy}}$, and $\boldsymbol{\rho}_{\mathbf{x}}$ in (19) are known, and then examine the performance of the estimators in the practical case in which they are estimated. Methods for extension to the nonnormal case (with some mild distributional assumptions) are discussed in Section 6.

JP–S again begins with selection of a random sample of n units on which Y is measured. In addition, m related and easily ranked characteristics $\mathbf{X}$ are available on each unit. For each i, an additional $H-1$ units are randomly selected, and the ranks of the components of $\mathbf{X}_i$ among its H comparison units are determined. The vector of ranks is denoted by $\mathbf{R}_i = (R_{i1}, \dots, R_{im})$. Thus there are H^m poststrata jointly grouped by the ranks $\mathbf{R} = (\mathbf{R}_i)_{i=1}^n$. Let $\mathrm{PS}_{\mathbf{r}}$ denote the poststratum in which $\mathbf{R}_i = \mathbf{r}$, and $\pi_{\mathbf{r}}$, $n_{\mathbf{r}}$, and $\bar{Y}_{[\mathbf{r}]}$ denote the probability, number, and sample mean of observations falling in $\mathrm{PS}_{\mathbf{r}}$. Let $\mu_{[\mathbf{r}]}$ and $\sigma^2_{[\mathbf{r}]}$ denote the mean and variance within the poststratum, that is, $\mu_{[\mathbf{r}]} = E(Y_i|\mathbf{R}_i = \mathbf{r})$, $\sigma^2_{[\mathbf{r}]} = \mathrm{var}(Y_i|\mathbf{R}_i = \mathbf{r})$. Define $\delta_{[\mathbf{r}]}$ as the difference between $\mu_{[\mathbf{r}]}$ and μ_y, that is, $\delta_{[\mathbf{r}]} \equiv \mu_{[\mathbf{r}]} - \mu_y$. Finally, let $I_{i\mathbf{r}}$ be the indicator variable such that $I_{i\mathbf{r}} = 1$ if $\mathbf{R}_i = \mathbf{r}$ and $I_{i\mathbf{r}} = 0$ otherwise.

We consider a class of linear JP–S estimators of the form

$$\hat{\mu}^{(m)} = \sum_{\mathbf{r}} w_{\mathbf{r}}(\mathbf{n})\big(\bar{Y}_{[\mathbf{r}]} - c_{\mathbf{r}}\big), \tag{24}$$

where the summation is over all H^m realizations of the rank vector $\mathbf{R}_i$, $\mathbf{n}$ is the random vector containing the counts of Y in the H^m poststrata, $w_{\mathbf{r}}(\cdot)$ is a weight associated with $\mathrm{PS}_{\mathbf{r}}$ that can be a function of $\mathbf{n}$, and $c_{\mathbf{r}}$ is a constant associated with $\mathrm{PS}_{\mathbf{r}}$ that can be used for bias correction. This class, denoted by $\mathcal{E}$, contains familiar members as well as useful new ones. The SRS estimator $\bar{Y}$ is in $\mathcal{E}$, with $w_{\mathbf{r}} = n_r/n$ and $c_{\mathbf{r}} = 0$. It obviously makes use of neither auxiliary nor distributional information. A version of the classical poststratified estimator that does use distributional knowledge is $\hat{\mu}_S^{(m)} = \sum_{\mathbf{r}} \pi_{\mathbf{r}} \bar{Y}_{[\mathbf{r}]} \in \mathcal{E}$. The $\pi_{\mathbf{r}}$'s can be calculated for normal data. A nonparametric variant of $\hat{\mu}_S^{(m)}$ that is also a member of the class is $\hat{\mu}_{vS}^{(m)} = \sum_{\mathbf{r}} \hat{\pi}_{\mathbf{r}}(\mathbf{n}) \bar{Y}_{[\mathbf{r}]}$, where $\hat{\pi}_{\mathbf{r}}(\cdot)$ is an estimate of the cell probability $\pi_{\mathbf{r}}$ based on $\mathbf{n}$. The cell probabilities can be estimated by the ranking procedure (Deming and Stephan 1940), because the marginal probability for each auxiliary variable rank is known to be $1/H$ because of characteristics of order statistics. The estimator of MacEachern et al. (2004) is also a member of $\mathcal{E}$, because (23) can be rewritten as

$$\hat{\mu}_M^{(m)} = \sum_{\mathbf{r}} a_{\mathbf{r}}(\mathbf{n}) \bar{Y}_{[\mathbf{r}]},$$

$$a_{\mathbf{r}}(\mathbf{n}) = \frac{1}{H} \sum_{i=1}^{H} \frac{b_{\mathbf{r}i} n_{\mathbf{r}}}{\sum_{\mathbf{r}'} b_{\mathbf{r}'i} n_{\mathbf{r}'}},$$

where $b_{\mathbf{r}i}$ is the count of rank i in the row vector $\mathbf{r}$. Now we examine several new estimators in this class suggested by commonly used estimation methods, each of which makes use of the distributional knowledge through the structure of the moments of the concomitant of multivariate order statistics.

We first consider the ordinary least squares estimator of μ_y, denoted by $\hat{\mu}_{\mathrm{oLS}}^{(m)}$ and defined as the estimator minimizing the sum of squared distances from each y_i to the mean of its poststratum, namely

$$\min_{\mu_y} \sum_{i=1}^{n} \sum_{\mathbf{r}} I_{i\mathbf{r}} \big[y_i - \big(\mu_y + \delta_{[\mathbf{r}]}\big)\big]^2. \tag{25}$$

Under the normality assumption, we have from (20) that $\delta_{[\mathbf{r}]} = \sigma_y \boldsymbol{\rho}_{\mathbf{xy}}^T \boldsymbol{\rho}_{\mathbf{x}}^{-1} E(\mathbf{Z}_{(\mathbf{r})})$. Solving (25) yields

$$\hat{\mu}_{\mathrm{oLS}}^{(m)} = \sum_{\mathbf{r}} \frac{n_{\mathbf{r}}}{n} \big(\bar{Y}_{[\mathbf{r}]} - \delta_{[\mathbf{r}]}\big).$$

Because $E(\hat{\mu}_{\mathrm{oLS}}^{(m)}|\mathbf{n}) = \mu_y$ and $\mathrm{var}(\hat{\mu}_{\mathrm{oLS}}^{(m)}|\mathbf{n}) = \sum_{\mathbf{r}} n_{\mathbf{r}} \sigma^2_{[\mathbf{r}]}/n^2$, we have $E(\hat{\mu}_{\mathrm{oLS}}^{(m)}) = \mu_y$ and

$$\begin{aligned} \mathrm{var}\big(\hat{\mu}_{\mathrm{oLS}}^{(m)}\big) &= \mathrm{var}\big[E\big(\hat{\mu}_{\mathrm{oLS}}^{(m)}|\mathbf{n}\big)\big] + E\big[\mathrm{var}\big(\hat{\mu}_{\mathrm{oLS}}^{(m)}|\mathbf{n}\big)\big] \\ &= \frac{1}{n} \sum_{\mathbf{r}} \pi_{\mathbf{r}} \sigma^2_{[\mathbf{r}]}. \end{aligned} \tag{26}$$

Next, consider the weighed least squares estimator, denoted by $\hat{\mu}_{\mathrm{wLS}}^{(m)}$, which minimizes the sum of the weighted squared distances to the poststrata means, namely

$$\min_{\mu_y} \sum_{i=1}^{n} \sum_{\mathbf{r}} I_{i\mathbf{r}} \left[\frac{y_i - (\mu_y + \delta_{[\mathbf{r}]})}{\sigma_{[\mathbf{r}]}}\right]^2. \tag{27}$$

Solving (27) yields

$$\hat{\mu}_{\mathrm{wLS}}^{(m)} = \sum_{\mathbf{r}} \frac{n_{\mathbf{r}}/\sigma^2_{[\mathbf{r}]}}{\sum_{\mathbf{r}} n_{\mathbf{r}}/\sigma^2_{[\mathbf{r}]}} \big(\bar{Y}_{[\mathbf{r}]} - \delta_{[\mathbf{r}]}\big), \tag{28}$$

where in the normal case $\sigma^2_{[\mathbf{r}]}$ is given by (21). It is easy to show that $\hat{\mu}_{\mathrm{wLS}}^{(m)}$ is unbiased and that

$$\mathrm{var}\big(\hat{\mu}_{\mathrm{wLS}}^{(m)}\big) = E\left[\left(\sum_{\mathbf{r}} \frac{n_{\mathbf{r}}}{\sigma^2_{[\mathbf{r}]}}\right)^{-1}\right], \tag{29}$$

where the $n_{\mathbf{r}}$ is multinomial with parameters n and $\pi_{\mathbf{r}}$ for all H^m possible $\mathbf{r}$.

In addition, we might naturally think of the best linear unbiased estimator, whose weights are *constant* (i.e., not functions of random variables). In our JP–S setting, this estimator, denoted by $\hat{\mu}_{\mathrm{BLU}}^{(m)}$, minimizes the variance of a subclass of $\mathcal{E}$, the unbiased estimators of the form $\sum_{\mathbf{r}} w_{\mathbf{r}}(\bar{Y}_{[\mathbf{r}]} - \delta_{[\mathbf{r}]})$, where the weights $w_{\mathbf{r}}$'s are restricted to be *constant* and sum to 1. $\hat{\mu}_{\mathrm{BLU}}^{(m)}$ has the form

$$\hat{\mu}_{\mathrm{BLU}}^{(m)} = \sum_{\mathbf{r}} \frac{[\sigma^2_{[\mathbf{r}]} E(1/n_{\mathbf{r}})]^{-1}}{\sum_{\mathbf{r}} [\sigma^2_{[\mathbf{r}]} E(1/n_{\mathbf{r}})]^{-1}} \big(\bar{Y}_{[\mathbf{r}]} - \delta_{[\mathbf{r}]}\big),$$

with

$$\mathrm{var}\big(\hat{\mu}_{\mathrm{BLU}}^{(m)}\big) = \left[\sum_{\mathbf{r}} \frac{1}{\sigma^2_{[\mathbf{r}]} E(1/n_{\mathbf{r}})}\right]^{-1}.$$

Now we proceed to compare the three unbiased estimators, $\hat{\mu}_{\mathrm{oLS}}^{(m)}$, $\hat{\mu}_{\mathrm{wLS}}^{(m)}$, and $\hat{\mu}_{\mathrm{BLU}}^{(m)}$, which are all in $\mathcal{E}$. First, $\hat{\mu}_{\mathrm{wLS}}^{(m)}$ has the smallest variance among the three, as we state in

Theorem 3. Second, $\hat{\mu}_{\text{oLS}}^{(m)}$ is easier to compute, especially when the number of poststrata H^m is large, because it does not require the variances $\sigma_{[\mathbf{r}]}^2$. Finally, $\hat{\mu}_{\text{wLS}}^{(m)}$ and $\hat{\mu}_{\text{BLU}}^{(m)}$ have similar expressions, because $\hat{\mu}_{\text{BLU}}^{(m)}$ can be obtained by replacing $1/n_{\mathbf{r}}$ by $E(1/n_{\mathbf{r}})$ in (28); they also have the same asymptotic variance, $(n\sum_{\mathbf{r}}\pi_{\mathbf{r}}/\sigma_{[\mathbf{r}]}^2)^{-1}$. However, $\hat{\mu}_{\text{BLU}}^{(m)}$ is not quite satisfactory. It is not applicable when the sample size n is small compared with the number of strata H^m. In this case there are many empty cells with inestimable means, which will cause trouble, because $\hat{\mu}_{\text{BLU}}^{(m)}$ assigns a prespecified nonzero weight to each cell. In contrast, $\hat{\mu}_{\text{oLS}}^{(m)}$ and $\hat{\mu}_{\text{wLS}}^{(m)}$ are both data-adaptive by assigning nonzero weights to nonempty cells only. Even if no empty cell occurs, the performance of $\hat{\mu}_{\text{BLU}}^{(m)}$ is very sensitive to n and is much worse than that of $\hat{\mu}_{\text{oLS}}^{(m)}$ and $\hat{\mu}_{\text{wLS}}^{(m)}$, as we show in our simulation.

The following theorem establishes an optimal property for $\hat{\mu}_{\text{wLS}}^{(m)}$.

Theorem 3. $\hat{\mu}_{\text{wLS}}^{(m)}$ has the lowest mean squared error (MSE) among the class of estimators of the form (24).

Proof. It is obvious that $\hat{\mu}_{\text{wLS}}^{(m)}$'s weights $w_{\mathbf{r}}^*(\mathbf{n}) = (n_{\mathbf{r}}/\sigma_{[\mathbf{r}]}^2)/\sum_{\mathbf{r}} n_{\mathbf{r}}/\sigma_{[\mathbf{r}]}^2$ minimize $\text{var}(\hat{\mu}^{(m)}|\mathbf{n}) = \sum_{\mathbf{r}} w_{\mathbf{r}}^2(\mathbf{n})\sigma_{[\mathbf{r}]}^2/n_{\mathbf{r}}$. Because $E(\hat{\mu}_{\text{wLS}}^{(m)}|\mathbf{n}) = \mu_y$, we have $\text{var}[E(\hat{\mu}_{\text{wLS}}^{(m)}|\mathbf{n})] = 0$. Thus $\text{var}(\hat{\mu}^{(m)}) \ge E[\text{var}(\hat{\mu}^{(m)}|\mathbf{n})] \ge E[\text{var}(\hat{\mu}_{\text{wLS}}^{(m)}|\mathbf{n})] = \text{var}(\hat{\mu}_{\text{wLS}}^{(m)})$ and $MSE(\hat{\mu}^{(m)}) \ge MSE(\hat{\mu}_{\text{wLS}}^{(m)})$, because $\hat{\mu}_{\text{wLS}}^{(m)}$ is unbiased.

This theorem assures us that $\hat{\mu}_{\text{wLS}}^{(m)}$ is the most efficient among the estimators discussed, not limited to $\hat{\mu}_{\text{oLS}}^{(m)}$, $\hat{\mu}_{\text{wLS}}^{(m)}$, and $\hat{\mu}_{\text{BLU}}^{(m)}$. However, its variance (29) is not expressed in a closed form and so is hard to compute. In the following corollary, we provide an upper bound by comparing it with $\hat{\mu}_{\text{oLS}}^{(m)}$ and also a lower bound by considering its asymptotic variance.

Corollary 3. The lower and upper bounds for the variance of $\hat{\mu}_{\text{wLS}}^{(m)}$ are

$$\frac{1}{n}\left(\sum_{\mathbf{r}}\pi_{\mathbf{r}}/\sigma_{[\mathbf{r}]}^2\right)^{-1} \le \text{var}\big(\hat{\mu}_{\text{wLS}}^{(m)}\big) \le \frac{1}{n}\sum_{\mathbf{r}}\pi_{\mathbf{r}}\sigma_{[\mathbf{r}]}^2. \tag{30}$$

The lower bound (i.e., the asymptotic variance) provides a good approximation to the variance of $\hat{\mu}_{\text{wLS}}^{(m)}$ when n is reasonably large. It also works well for small n if the difference between the upper and lower bound is small, which occurs for normal data in many cases.

Finally, the following theorem justifies our intuition that for both $\hat{\mu}_{\text{wLS}}^{(m)}$ and $\hat{\mu}_{\text{oLS}}^{(m)}$, adding an extra ranking variable improves estimation efficiency, at least in an asymptotic sense.

Theorem 4. Suppose that $\hat{\mu}_{\text{wLS}}^{(m+1)}$ ($\hat{\mu}_{\text{oLS}}^{(m+1)}$) is the weighted (ordinary) least squares estimator with ranking variables $(\mathbf{X}, X_{m+1})$, and $\hat{\mu}_{\text{wLS}}^{(m)}$ ($\hat{\mu}_{\text{oLS}}^{(m)}$) is the corresponding estimator using the first m ranking variables $\mathbf{X}$ only. Then $\hat{\mu}_{\text{wLS}}^{(m+1)}$ is more efficient than $\hat{\mu}_{\text{wLS}}^{(m)}$ for large n, and $\hat{\mu}_{\text{oLS}}^{(m+1)}$ is more efficient than $\hat{\mu}_{\text{oLS}}^{(m)}$ for any n (see App. B for the proof).

Although Theorem 4 establishes that the addition of ranking variables is helpful, a practical question is just how much gain can be expected. We investigated this for the special case of adding a second auxiliary variable to the first. The asymptotic relative efficiency (ARE), $ARE = \lim_{n\to+\infty}[\text{var}(\hat{\mu}_{\text{wLS}}^{(1)})/\text{var}(\hat{\mu}_{\text{wLS}}^{(2)})]$, was computed from the lower bound in (30), (15), and eq. (6.8.3b) of David and Nagaraja (2003, chap. 6.8) for $\rho_{12} = .25, .5, .75$; $H = 2, 3, 4$; and a range of values of ρ_{1y} and ρ_{2y}. Figure 3 shows the results for two equally effective rankers (i.e., $\rho_{1y} = \rho_{2y}$) and the three values each of ρ_{12} and H. We see that the gain from the second ranker can be substantial; it increases as either the ranking quality or the number of ranking classes increases and decreases as the two rankers become more similar. We also computed the relative efficiency for $\hat{\mu}_{\text{oLS}}^{(2)}$ over $\hat{\mu}_{\text{oLS}}^{(1)}$, in which we observed the tightness of the two bounds in (30). As a result, the values of the relative efficiency were virtually identical to those of the ARE for the weighted one, so are not shown in Figure 3.

5.3 Simulation

We have demonstrated that under the normality assumption with known σ_y, $\boldsymbol{\rho}_{\mathbf{xy}}$, and $\boldsymbol{\rho}_{\mathbf{x}}$, the weighted least squares estimator $\hat{\mu}_{\text{wLS}}^{(m)}$ is the most efficient among the class $\mathcal{E}$ including members $\bar{Y}$, $\hat{\mu}_M^{(m)}$, $\hat{\mu}_S^{(m)}$, $\hat{\mu}_{vS}^{(m)}$, $\hat{\mu}_{\text{wLS}}^{(m)}$, $\hat{\mu}_{\text{oLS}}^{(m)}$, and $\hat{\mu}_{\text{BLU}}^{(m)}$. In practice, however, these parameters will not be known, and the distributional assumptions may not hold exactly. Thus we designed

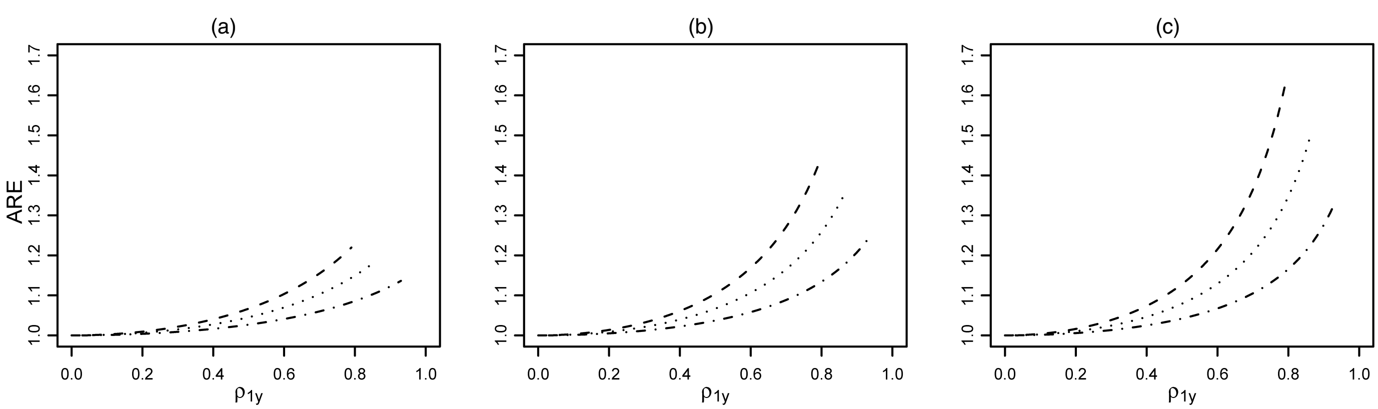

Figure 3. Asymptotic Efficiency of $\hat{\mu}_{wLS}^{(2)}$ Over $\hat{\mu}_{wLS}^{(1)}$ for Pairs of Equally Effective Rankers: (a) $H = 2$, $\rho_{1y} = \rho_{2y}$; (b) $H = 3$, $\rho_{1y} = \rho_{2y}$; (c) $H = 4$, $\rho_{1y} = \rho_{2y}$ (— —, $\rho_{12} = .25$; $\cdots$, $\rho_{12} = .5$; $\cdot - \cdot$, $\rho_{12} = .75$).

Table 1. Comparing Efficiency of the JP–S Estimators With Estimated Parameters

		H = 2; sample size				H = 4; sample size				H = 10; sample size			
$(\rho_{1y}, \rho_{2y}, \rho_{12})$	Mean estimator	10	20	50	100	10	20	50	100	10	20	50	100
(.9, .9, .65)	$\hat{\mu}_{\mathrm{wLS}}$	1.55	1.55	1.56	1.56	2.51	2.55	2.63	2.67	4.82	5.32	5.53	5.57
	$\hat{\mu}_{\mathrm{oLS}}$	1.54	1.54	1.55	1.56	2.48	2.51	2.57	2.61	4.62	5.04	5.28	5.24
	$\hat{\mu}_M$	1.38	1.44	1.47	1.48	1.80	2.11	2.29	2.35	1.82	2.74	3.94	4.21
	$\hat{\mu}_{\mathrm{BLU}}$	1.28	1.35	1.44	1.51			2.09	2.24				
(.8, .8, .5)	$\hat{\mu}_{\mathrm{wLS}}$	1.42	1.41	1.43	1.44	2.01	2.06	2.09	2.14	2.90	3.22	3.35	3.39
	$\hat{\mu}_{\mathrm{oLS}}$	1.41	1.41	1.43	1.44	2.00	2.04	2.07	2.12	2.87	3.17	3.29	3.33
	$\hat{\mu}_M$	1.29	1.32	1.35	1.36	1.56	1.74	1.84	1.90	1.55	2.00	2.56	2.70
	$\hat{\mu}_{\mathrm{BLU}}$	1.15	1.19	1.34	1.39			1.57	1.80				
(.8, .5, .5)	$\hat{\mu}_{\mathrm{wLS}}$	1.24	1.28	1.29	1.29	1.50	1.57	1.63	1.63	1.79	1.93	2.07	2.08
	$\hat{\mu}_{\mathrm{oLS}}$	1.24	1.28	1.29	1.29	1.50	1.56	1.62	1.62	1.78	1.92	2.06	2.07
	$\hat{\mu}_M$	1.15	1.20	1.22	1.22	1.23	1.34	1.42	1.43	1.19	1.34	1.61	1.67
	$\hat{\mu}_{\mathrm{BLU}}$	1.03	1.08	1.21	1.25			1.29	1.35				
(.5, .5, .5)	$\hat{\mu}_{\mathrm{wLS}}$	1.10	1.10	1.13	1.16	1.12	1.21	1.24	1.24	1.13	1.28	1.34	1.37
	$\hat{\mu}_{\mathrm{oLS}}$	1.10	1.10	1.13	1.16	1.12	1.21	1.24	1.24	1.14	1.28	1.34	1.37
	$\hat{\mu}_M$	1.08	1.08	1.11	1.14	1.08	1.15	1.20	1.21	1.03	1.12	1.25	1.30
	$\hat{\mu}_{\mathrm{BLU}}$	.90	.94	1.07	1.13			.94	1.04				

NOTE: Note that due to the empty-cell problems, $\hat{\mu}_{\mathrm{BLU}}$ is not applicable when n is small compared with H^2.

a simulation study for two purposes: (1) to compare the performance of the estimators when σ_y, $\boldsymbol{\rho}_{\mathbf{x}y}$, and $\boldsymbol{\rho}_{\mathbf{x}}$ are unknown and must be estimated from the data and (2) to test their robustness when the normality assumption is violated. In our preliminary simulations, we found that the estimator of MacEachern et al. (2004) performed consistently best among the three "sampling" estimators $\hat{\mu}_M^{(m)}$, $\hat{\mu}_S^{(m)}$, and $\hat{\mu}_{vS}^{(m)}$. Hence we included only $\hat{\mu}_M^{(m)}$, $\hat{\mu}_{\mathrm{wLS}}^{(m)}$, $\hat{\mu}_{\mathrm{oLS}}^{(m)}$, and $\hat{\mu}_{\mathrm{BLU}}^{(m)}$ in the full study presented here, along with $\bar{Y}$ as a benchmark estimator.

In our first experiment, we simulated JP–S samples from the standard multivariate normal distribution for Y and two auxiliary variables (X_1, X_2) for four sets of $(\rho_{1y}, \rho_{2y}, \rho_{12})$: (.9, .9, .65), (.8, .8, .5), (.8, .5, .5), and (.5, .5, .5). We set H to be 2, 4, or 10 and n to be 10, 20, 50, or 100. Because m is fixed at 2, we omit the superscripts in the discussion that follows. When calculating $\hat{\mu}_{\mathrm{wLS}}$, $\hat{\mu}_{\mathrm{oLS}}$, and $\hat{\mu}_{\mathrm{BLU}}$ from each sample, we substituted estimates for σ_y and the ρ's, computed using standard methods. Table 1 reports the simulated relative efficiency of the four JP–S estimators to the SRS estimator $\bar{Y}$ for each setting. Here efficiency is defined as the ratio of the variance of $\bar{Y}$ to MSE of each JP–S estimator, where the MSE is estimated from 20,000 replicates.

The results in Table 1 show that the two least squares estimators outperform the other two estimators in all cases, even though they use estimates of σ_y and the ρ's. The performance of $\hat{\mu}_{\mathrm{wLS}}$ is at most only slightly better than that of $\hat{\mu}_{\mathrm{oLS}}$. Both estimators perform well for even small n. The improvement from the two parametric estimators over the nonparametric estimator, $\hat{\mu}_M$, is considerable, especially when n is small and H is large, as long as the ranking variables are effective. In contrast, $\hat{\mu}_{\mathrm{BLU}}$ performs poorly overall; it is sensitive to sample size and is not applicable when empty cells occur. Hence we do not consider $\hat{\mu}_{\mathrm{BLU}}$ further here.

To more closely examine the effect of estimation of the unknown correlations and variance, we compute the asymptotic efficiency for $\hat{\mu}_{\mathrm{wLS}}$ over $\bar{Y}$ using the lower bound in (30) and efficiency for $\hat{\mu}_{\mathrm{oLS}}$ over $\bar{Y}$ using (26). Their theoretical values are reported in Table 2. Comparing the simulated values in Table 1 to these, we observe that estimating these parameters has a negligible effect when $n \ge 50$. For smaller sample sizes, both $\hat{\mu}_{\mathrm{wLS}}$ and $\hat{\mu}_{\mathrm{oLS}}$ lose some efficiency by doing so, although they still perform better than $\hat{\mu}_M$.

In the second experiment, we examine the performance of the JP–S estimators when the normality assumption is violated. We consider three cases: (1) $(\log X_1, \log X_2, Y)$ are generated from the standard normal distributions with the four sets of correlations as before, (2) $(\log X_1, \log X_2, \log Y)$ are generated from the same distributions as in case (1), and (3) (X_1, X_2, Y) follows the multivariate uniform distribution described in Example 1. Here we set $H = 4$, generate 20,000 JP–S samples for each setting, and calculate $\hat{\mu}_{\mathrm{wLS}}$, $\hat{\mu}_{\mathrm{oLS}}$, and $\hat{\mu}_M$ from each. The former two estimators are computed as if (X_1, X_2, Y) were multivariate normal, but using estimated values for σ_y and ρ's. Table 3 reports the simulated efficiency.

Several observations can be made from Table 3. When the ranking variables violate the normality assumption but Y is still normal, $\hat{\mu}_{\mathrm{wLS}}$ and $\hat{\mu}_{\mathrm{oLS}}$ perform comparably and have efficiencies similar to those in the normal case. When both $\mathbf{X}$ and Y are lognormal, $\hat{\mu}_{\mathrm{wLS}}$ outperforms $\hat{\mu}_{\mathrm{oLS}}$, whereas the situation is reversed when (X_1, X_2, Y) have a multivariate uniform distribution. As expected, the least squares estimators are less efficient than in the normal case when Y is no longer normally distributed. Surprisingly, $\hat{\mu}_M$ does not perform as well as $\hat{\mu}_{\mathrm{wLS}}$

Table 2. Theoretical Values of (asymptotic) Relative Efficiency of $\hat{\mu}_{wLS}$ and $\hat{\mu}_{oLS}$

Theoretical	(.9, .9, .65)			(.8, .8, .5)			(.8, .5, .5)			(.5, .5, .5)		
value	H = 2	H = 4	H = 10	H = 2	H = 4	H = 10	H = 2	H = 4	H = 10	H = 2	H = 4	H = 10
$ARE(\hat{\mu}_{\mathrm{wLS}}, \bar{Y})$	1.56	2.63	5.56	1.44	2.14	3.41	1.29	1.65	2.13	1.14	1.26	1.38
$RE(\hat{\mu}_{\mathrm{oLS}}, \bar{Y})$	1.55	2.58	5.27	1.44	2.12	3.35	1.29	1.64	2.12	1.14	1.26	1.38

Table 3. Comparing Efficiency of the JP–S Estimators With Estimated Parameters (H = 4 only) When the Normality Assumption Is Violated

Mean estimator	$(\rho'_{1y}, \rho'_{2y}, \rho'_{12})$	$(\log X_1, \log X_2, Y) \sim$ MVN; sample size 10	20	50	100	$(\log X_1, \log X_2, \log Y) \sim$ MVN; sample size 10	20	50	100	Multivariate uniform; sample size 10	20	50	100
$\hat{\mu}_{\mathrm{wLS}}$	(.9, .9, .65)	2.30	2.49	2.50	2.51	1.73	1.61	1.55	1.45	1.49	1.57	1.62	1.64
$\hat{\mu}_{\mathrm{oLS}}$		2.30	2.45	2.45	2.47	1.47	1.32	1.33	1.33	1.53	1.62	1.67	1.68
$\hat{\mu}_M$		1.81	2.11	2.26	2.30	1.33	1.36	1.49	1.44	1.34	1.49	1.57	1.59
$\hat{\mu}_{\mathrm{wLS}}$	(.8, .8, .5)	1.91	2.02	2.06	2.11	1.52	1.46	1.43	1.42				
$\hat{\mu}_{\mathrm{oLS}}$		1.91	2.01	2.05	2.09	1.33	1.25	1.24	1.26				
$\hat{\mu}_M$		1.56	1.76	1.85	1.92	1.25	1.31	1.33	1.33				
$\hat{\mu}_{\mathrm{wLS}}$	(.8, .5, .5)	1.45	1.54	1.60	1.59	1.26	1.30	1.28	1.32				
$\hat{\mu}_{\mathrm{oLS}}$		1.45	1.54	1.60	1.58	1.14	1.16	1.16	1.21				
$\hat{\mu}_M$		1.24	1.35	1.44	1.42	1.12	1.16	1.17	1.23				
$\hat{\mu}_{\mathrm{wLS}}$	(.5, .5, .5)	1.07	1.19	1.23	1.23	1.06	1.07	1.14	1.12				
$\hat{\mu}_{\mathrm{oLS}}$		1.08	1.19	1.23	1.23	1.01	1.01	1.11	1.09				
$\hat{\mu}_M$		1.04	1.16	1.20	1.20	1.00	1.07	1.11	1.11				

in any of the cases considered or as well as $\hat{\mu}_{\mathrm{oLS}}$ except when Y is heavy-tailed. This leads to our belief that even with moderate deviation from normality, $\hat{\mu}_{\mathrm{wLS}}$ and $\hat{\mu}_{\mathrm{oLS}}$ still may achieve better performance than $\hat{\mu}_M$, especially for small n.

5.4 An Empirical Study: Human Teeth Width

This section uses a real dataset to compare the JP–S estimators of the mean. To examine their performance in both infinite and finite population settings, samples were selected with and without replacement from a small population containing measurements on teeth widths for 295 subjects in a health survey conducted in Seoul, Korea (Lee, Lee, Lim, Ahn, and Kim 2006). All teeth were measured by digital Vernier calipers, a process that requires a 3-week training period to master. Here our goal is to estimate the mean width of teeth in the back of the mouth, using the middle teeth as ranking variables. A practical justification for doing this is that a tooth close to the center is much easier to order than a tooth farther back in the month.

The widths of the first upper incisors, U_1 (the first tooth from the center), and the first upper canines, U_3 (the third tooth from the center), were used as ranking variables. Selection of JP–S samples was simulated, and estimates of the mean width of the first lower molars, L_6 (the sixth tooth from the center), were calculated. The 295 subjects were treated as the "true" population, and parameter estimates computed from their data were taken as the true population parameters. These included $\rho_{16} = .503$, $\rho_{36} = .540$, $\rho_{13} = .576$, $\mu_{L6} = 10.95$, and $\sigma^2_{L6} = .319$. Standard diagnostic checking, performed on (U_1, U_3, L_6) through the macro %MULTNORM in SAS, did not reveal any gross violation of normality.

In this simulation, we set $H = 5$ and sample sizes $n = 10, 15, \ldots, 55$. To obtain a JP–S sample of size n with replacement, we repeated the following procedure n times: A set of 5 subjects was randomly selected from all 295 subjects, and bivariate ranking was done based on U_1 and U_3 within the set; then 1 of the 5 subjects was randomly selected to enter the sample. In contrast, to obtain a JP–S sample of size n without replacement, the set of five subjects selected on each draw were excluded from the dataset, so they were not available for the next selection. For each JP–S sample of size n, we calculated $\hat{\mu}^{(2)}_{\mathrm{wLS}}$, $\hat{\mu}^{(2)}_{\mathrm{oLS}}$, and $\hat{\mu}^{(2)}_M$, with poststrata determined by ranks of U_1 and U_3, and $\hat{\mu}^{(1)}_{\mathrm{wLS}}$, $\hat{\mu}^{(1)}_{\mathrm{oLS}}$, and $\hat{\mu}^{(1)}_M$, with poststrata determined by U_3 only. All least squares estimators were computed using σ_y and the ρ's estimated from the sample.

Figure 4 shows values of the simulated relative efficiency of the six JP–S estimators to $\bar{Y}$ for each sample size n. Here the MSE is estimated from 100,000 replicates. The figure shows that whether sampling is with or without replacement, the least

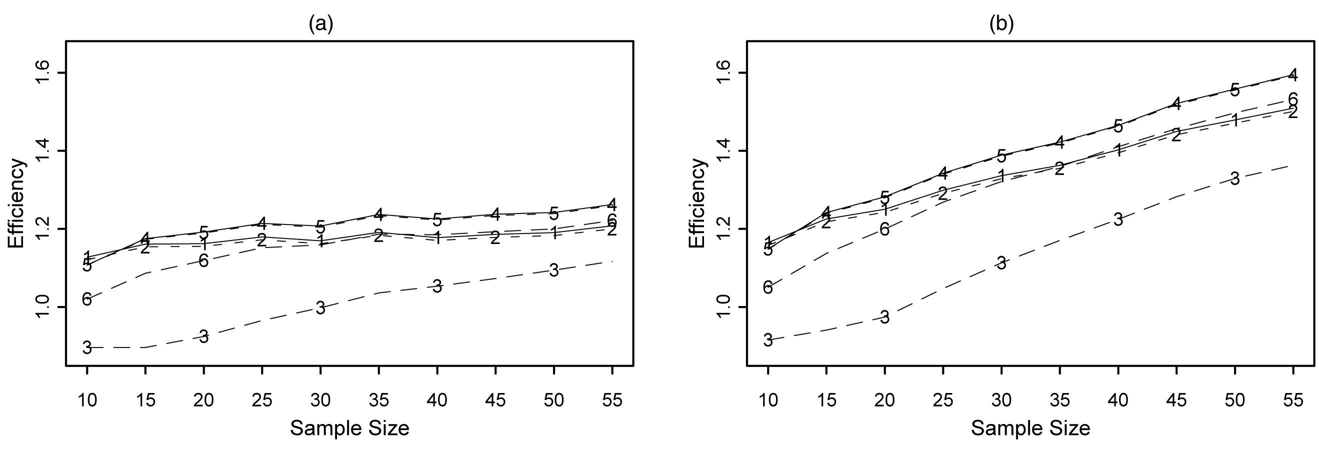

Figure 4. An Empirical Study of Human Teeth Width. (a) Sampling with replacement. (b) Sampling without replacement. (–1–, wLS1; -2-, oLS1; –3–, M1; –4–, wLS2; -5-, oLS2; –6–, M2.)

squares estimators have almost identical performance and with the two ranking variables they have the best performance of all. In addition, the results demonstrate that the benefit from using a second ranking variable when sampling without replacement is larger than that from sampling with replacement samples. Thus it may be safe to use the least squares estimators for small populations.

6. DISCUSSION

We have defined concomitants of multivariate order statistics and provided analytical expressions for their densities and moments. We have also illustrated the use of the theory by providing new estimators that use ranking information from more than one auxiliary variable for improving estimation of the mean.

We note that the proposed least squares estimators do not require normality. They are available when certain distributional assumptions about the data can be made: Development of $\hat{\mu}_{\mathrm{oLS}}$ requires that $\delta_{[\mathbf{r}]}$ not be a function of μ_y (say, $\delta_{[\mathbf{r}]} \perp \mu_y$) for each poststratum, and $\hat{\mu}_{\mathrm{wLS}}$ requires that both $\delta_{[\mathbf{r}]} \perp \mu_y$ and $\sigma^2_{[\mathbf{r}]} \perp \mu_y$. Suppose that for each ranking variable X_i ($1 \le i \le m$), there exists a monotonic function $g_i(\cdot)$ such that $Z_i = g_i(X_i)$ and $\mathbf{Z} = (Z_1, \ldots, Z_m)$ has a joint distribution $f(\mathbf{z}; \Theta)$ with the parameter set $\Theta \perp \mu_y$ (e.g., a special case is that each X_i is from a location-scale distribution family). Let $m(\mathbf{z}; \Theta_m) \equiv E(Y - \mu_y | \mathbf{Z} = \mathbf{z})$, which is a function of $\mathbf{z}$ and a set of distributional parameters Θ_m; let $v(\mathbf{z}; \Theta_v) \equiv \mathrm{var}(Y | \mathbf{Z} = \mathbf{z})$, which is a function of $\mathbf{z}$ and a set of distributional parameters Θ_v. Then a sufficient condition for $\delta_{[\mathbf{r}]} \perp \mu_y$ is $\Theta_m \perp \mu_y$, and a sufficient condition for $\sigma^2_{[\mathbf{r}]} \perp \mu_y$ is $\Theta_v \perp \mu_y$, which follow directly from Theorem 1 and its higher-dimensional generalization. These sufficient conditions may be milder than those assumptions in most regression setups, because they do not require linearity or any other functional form for the conditional expectations.

In fact, the result in Theorem 3 (i.e., the optimality of $\hat{\mu}_{\mathrm{wLS}}$ among linear estimators) is rather general. Under the assumptions discussed earlier, this result can be directly extended to situations in which the sampling space can be partitioned to strata through either sampling design or poststratification. However, obtaining the bias-correction term $\delta_{[\mathbf{r}]}$ and the variance $\sigma^2_{[\mathbf{r}]}$ for each stratum is nontrivial. In our JP–S applications, these can be derived through the theoretical developments in Sections 2–4, which greatly facilitate computation of the most efficient linear estimator.

There are other examples in the literature in which a single ranked variable is used to improve estimation of some parameter. The methods that we have developed here can be used in those applications as well. For example, the approach of Barnett et al. (1976) for obtaining linear estimates of correlation coefficients can be directly adapted when information is available from two or more ranking variables, using the moment expressions (20) and (21).

Other useful applications will require additional theoretical development. For example, the properties of concomitants of extreme order statistics have been a topic of study for their use in ranking and selection (Yeo and David 1984; Arnold and Beyer 2005). The notion of "extreme" order statistics of a vector of ranking variables can be defined in various ways, with the best way undoubtedly depending on its purpose.

Finally, we note that we have assumed that the number of ranking classes is the same for all ranking variables. There are applications in which a generalization to the case in which one ranking variable allows H classes, whereas another that allows H' classes may be needed. For example, consider the employee selection problem in which not every candidate had the complete battery of screening tests. In that case, it would be useful to have a way to examine properties of the concomitant of multivariate order statistics, some of which are partially ranked.

APPENDIX A: PROOF OF THEOREM 2

For notational clarity, let $Y_{[r,s];z}$ explicitly denote the concomitant of the rth-order statistics of Z_1 and the sth-order statistics of Z_2 with mean $\mu_{[r,s];z}$ and variance $\sigma^2_{[r,s];z}$, let q_{rs} denote the bivariate rank distribution of Z_1 and Z_2, and let $(Z_{1(r,s)}, Z_{2(r,s)})$ be the bivariate order statistics of (Z_1, Z_2) with joint density $g_{(r,s)}(z_1, z_2)$. Because ψ_1 and ψ_2 are monotonic, to show (12) and (13), it is equivalent to show for $r \in \{1, \ldots, H\}$ and $s \in \{1, \ldots, H\}$,

$$\mu_{[r,s];z} + \mu_{[\bar{r},\bar{s}];z} = 2\mu_y \tag{A.1}$$

and

$$\sigma^2_{[r,s];z} = \sigma^2_{[\bar{r},\bar{s}];z}. \tag{A.2}$$

Under the conditions that $E(Y|z_1, z_2)$ is a linear function of z_1 and z_2, and $g(z_1, z_2) = g(-z_1, -z_2)$, we can obtain

$$\mu_{[r,s];z} = \mu_y + \beta_1 E\big(Z_{1(r,s)}\big) + \beta_2 E\big(Z_{2(r,s)}\big) \tag{A.3}$$

and

$$\begin{aligned}\sigma^2_{[r,s];z} = {} & \int\int_{\mathcal{Z}} \mathrm{var}(Y|z_1, z_2) g_{(r,s)}(z_1, z_2)\, dz_1\, dz_2 \\ & + \beta_1^2 \,\mathrm{var}\big(Z_{1(r,s)}\big) + 2\beta_1\beta_2 \,\mathrm{cov}\big(Z_{1(r,s)}, Z_{2(r,s)}\big) \\ & + \beta_2^2 \,\mathrm{var}\big(Z_{2(r,s)}\big),\end{aligned} \tag{A.4}$$

where β_1 and β_2 are constants and $\mathcal{Z}$ is the support of the distribution of (Z_1, Z_2).

Now consider $E(Z_{1(r,s)})$, which can be expressed by

$$E\big(Z_{1(r,s)}\big) = \int\int_{\mathcal{Z}} z_1 g_{(r,s)}(z_1, z_2)\, dz_1\, dz_2.$$

Define $z_1^* = -z_1$ and $z_2^* = -z_2$. Then

$$E\big(Z_{1(r,s)}\big) = -\int\int_{\mathcal{Z}} z_1^* g_{(r,s)}(z_1, z_2)\, dz_1^*\, dz_2^*. \tag{A.5}$$

Because $g(z_1, z_2) = g(z_1^*, z_2^*)$, $q_{rs} = q_{\bar{r}\bar{s}}$, so that

$$g_{(r,s)}(z_1, z_2) = \frac{q(r, s|z_1, z_2) g(z_1^*, z_2^*)}{q_{\bar{r}\bar{s}}} z_2^*. \tag{A.6}$$

From (5),

$$q(r, s|z_1, z_2) = q(\bar{r}, \bar{s}|z_1^*, z_2^*), \tag{A.7}$$

by noting that $\theta_1(z_1, z_2) = \theta_4(z_1^*, z_2^*)$, $\theta_2(z_1, z_2) = \theta_3(z_1^*, z_2^*)$, $\theta_3(z_1, z_2) = \theta_2(z_1^*, z_2^*)$, and $\theta_4(z_1, z_2) = \theta_1(z_1^*, z_2^*)$. Inserting (A.7) in (A.6) and (A.8) in (A.5) yields

$$g_{(r,s)}(z_1, z_2) = g_{(\bar{r},\bar{s})}(z_1^*, z_2^*) \tag{A.8}$$

and

$$E\big(Z_{1(r,s)}\big) + E\big(Z_{1(\bar{r},\bar{s})}\big) = 0. \tag{A.9}$$

Similarly, from (A.8), the following are obtained:

$$\begin{aligned}
&E(Z_{2(r,s)}) + E(Z_{2(\bar{r},\bar{s})}) = 0,\\
&\int\int_{\mathcal{Z}} \operatorname{var}(Y|z_1, z_2) g_{(r,s)}(z_1, z_2)\, dz_1\, dz_2\\
&\quad = \int\int_{\mathcal{Z}} \operatorname{var}(Y|z_1^*, z_2^*) g_{(\bar{r},\bar{s})}(z_1^*, z_2^*)\, dz_1^*\, dz_2^*,\\
&\operatorname{var}(Z_{1(r,s)}) = \operatorname{var}(Z_{1(\bar{r},\bar{s})}),\\
&\operatorname{var}(Z_{2(r,s)}) = \operatorname{var}(Z_{2(\bar{r},\bar{s})}),\\
&\operatorname{cov}(Z_{1(r,s)}, Z_{2(r,s)}) = \operatorname{cov}(Z_{1(\bar{r},\bar{s})}, Z_{2(\bar{r},\bar{s})}).
\end{aligned} \tag{A.10}$$

Finally, combining (A.9) and (A.10) with (A.3) and (A.4) yields (A.1) and (A.2), completing the proof of (12) and (13).

APPENDIX B: PROOF OF THEOREM 4

Because the weighted (ordinary) least squares estimators are unbiased, we need only compare their variances. We want to show that for $\hat{\mu}_{\mathrm{oLS}}^{(m+1)}$ and $\hat{\mu}_{\mathrm{oLS}}^{(m)}$,

$$\sum_{\mathbf{r}} \pi_{\mathbf{r}} \sigma_{[\mathbf{r}]}^2 \ge \sum_{\mathbf{r}} \sum_{s=1}^{H} \pi_{\mathbf{r}s} \sigma_{[\mathbf{r}s]}^2, \tag{B.1}$$

and for $\hat{\mu}_{\mathrm{wLS}}^{(m+1)}$ and $\hat{\mu}_{\mathrm{wLS}}^{(m)}$,

$$\left(\sum_{\mathbf{r}} \frac{\pi_{\mathbf{r}}}{\sigma_{[\mathbf{r}]}^2}\right)^{-1} \ge \left(\sum_{\mathbf{r}} \sum_{s=1}^{H} \frac{\pi_{\mathbf{r}s}}{\sigma_{[\mathbf{r}s]}^2}\right)^{-1}, \tag{B.2}$$

where s denotes the rank of the extra ranking variable X_{m+1}. Noting that

$$\begin{aligned}
\sigma_{[\mathbf{r}]}^2 &= \sum_{s=1}^{H} \frac{\pi_{\mathbf{r}s}}{\pi_{\mathbf{r}}} \sigma_{[\mathbf{r}s]}^2 + \left[\sum_{s=1}^{H} \frac{\pi_{\mathbf{r}s}}{\pi_{\mathbf{r}}} \mu_{[\mathbf{r}s]}^2 - \left(\sum_{s=1}^{H} \frac{\pi_{\mathbf{r}s}}{\pi_{\mathbf{r}}} \mu_{[\mathbf{r}s]}\right)^2\right]\\
&\ge \sum_{s=1}^{H} \frac{\pi_{\mathbf{r}s}}{\pi_{\mathbf{r}}} \sigma_{[\mathbf{r}s]}^2
\end{aligned} \tag{B.3}$$

yields (B.1). Now noting that $\{\sum_{s=1}^{H} \pi_{\mathbf{r}s} \sigma_{[\mathbf{r}s]}^2 / \pi_{\mathbf{r}}\}\{\sum_{s=1}^{H} \pi_{\mathbf{r}s}/(\pi_{\mathbf{r}} \times \sigma_{[\mathbf{r}s]}^2)\} \ge 1$ combined with (B.3), we have $\sum_{s=1}^{H} \pi_{\mathbf{r}s}/\sigma_{[\mathbf{r}s]}^2 \ge \pi_{\mathbf{r}}/\sigma_{[\mathbf{r}]}^2$, which leads to (B.2).

[Received October 2005. Revised April 2006.]

REFERENCES

Arnold, D., and Beyer, H. (2005), "Expected Sample Moments of Concomitants of Selected Order Statistics," *Statistics and Computing*, 14, 241–250.

Barnett, V., Green, P. J., and Robinson, A. (1976), "Concomitants and Correlation Estimates," *Biometrika*, 63, 323–328.

Bhattacharya, P. K. (1974), "Covergence of Sample Paths of Normalized Sums of Induced Order Statistics," *The Annals of Statistics*, 5, 1034–1039.

Chen, Z. (2002), "Adaptive Ranked-Set Sampling With Multiple Concomitant Variables: An Effective Way to Observational Economy," *Bernoulli*, 8, 313–322.

Chen, Z., and Shen, L. (2003), "Two-Layer Ranked Set Sampling With Concomitant Variables," *Journal of Statistical Planning and Inference*, 115, 45–57.

David, H. A. (1993), "Concomitants of Order Statistics: Review and Recent Developments," in *Multiple Comparisons, Selection and Applications in Biometry*, ed. F. M. Hoppe, New York: Marcel Dekker, pp. 507–518.

David, H. A., and Nagaraja, H. N. (2003), *Order Statistics* (3rd ed.), New York: Wiley-Interscience.

David, H. A., O'Connell, M. J., and Yang, S. S. (1977), "Distribution and Expected Value of the Rank of a Concomitant of an Order Statistic," *The Annals of Statistics*, 5, 216–223.

Dell, T., and Clutter, J. L. (1972), "Ranked Set Sampling Theory With Order Statistics Background," *Biometrics*, 28, 545–555.

Deming, W. E., and Stephan, F. F. (1940), "On a Least Squares Adjustment of a Sampled Frequency Table When the Expected Marginal Totals Are Known," *The Annals of Mathematical Statistics*, 11, 427–444.

Goel, P. K., and Hall, P. (1994), "On the Average Difference Between Concomitants and Order Statistics," *The Annals of Probability*, 22, 126–144.

Lee, S.-J., Lee, S., Lim, J., Ahn, S.-J., and Kim, T.-W. (2006), "Cluster Analysis of Human Tooth Size in Subjects With Normal Occlusion," *American Journal of Orthodontics & Dentofacial Orthopedics*, **??**, **??–??**.

MacEachern, S. N., Stasny, E. A., and Wolfe, D. A. (2004), "Judgement Post-Stratification With Imprecise Rankings," *Biometrics*, 60, 207–215.

Nagaraja, H. N. (1982), "Some Nondegenerate Limit Laws for the Selection Differential," *The Annals of Statistics*, 10, 1306–1310.

Nagaraja, H. N., and David, H. A. (1994), "Distribution of the Maximum of Concomitants of Selected Order Statistics," *The Annals of Statistics*, 22, 478–494.

Salzer, H., Zucker, R., and Capuano, R. (1952), "Table of the Zeros and Weight Factors of the First Twenty Hermite Polynomials," *Journal of Research of the National Bureau of Standards*, 48, 111–116.

Sen, P. K. (1976), "A Note on Invariance Principles for Induced Order Statistics," *The Annals of Probability*, 4, 474–479.

——— (1981), "The Cox Regression Model, Invariance Principles for Some Induced Quantile Processes and Some Repeated Significance Tests," *The Annals of Statistics*, 9, 109–121.

Stokes, S. L. (1977), "Ranked Set Sampling With Concomitant Variables," *Communication in Statistics, Part A—Theory and Methods*, 6, 1207–1211.

Wang, X., and Stokes, S. L. (2005), "Moments of Bivariate Order Statistics for the Normal Distribution," technical report, Southern Methodist University, Dept. of Statistical Science.

Yang, S. S. (1977), "General Distribution Theory of the Concomitants of Order Statistics," *The Annals of Statistics*, 5, 996–1002.

Yeo, W. B., and David, H. A. (1984), "Selection Through an Associated Characteristic With Application to the Random-Effects Model," *Journal of the American Statistical Association*, 79, 399–405.