Variance-Stabilizing and Confidence-Stabilizing Transformations for the Normal Correlation Coefficient with Known Variances

Bailey K. Fosdick and Michael D. Perlman

Technical Report no. 610

Department of Statistics

University of Washington

Seattle, WA 98195-4322

March 6, 2013

Abstract

Fosdick and Raftery (2012) revisited the classical problem of inference for a bivariate normal correlation coefficient ρ when the variances are known. They considered several frequentist and Bayesian estimators, the former including the maximum likelihood estimator (MLE), but did not obtain the standard errors of these estimators or confidence intervals for ρ. Here we present a new variance-stabilizing transformation y for the MLE in the known-variance case. Adjusting y appropriately according to the sample size n produces a "confidence-stabilizing" transformation y_n that provides more accurate interval estimates for ρ than the MLE, as does Fisher's classical z transformation for the MLE in the unknown-variance case. Interestingly, the z transform applied to the MLE for the unknown-but-equal-variance case performs well in the known-variance case for smaller values of ρ. Both these methods are also useful for comparing two or more correlation coefficients in the known-variance case; hypothesis testing in this case is also discussed.

1. Introduction

Let (x_i, y_i), $i = 1, \ldots, n$, be i.i.d. observations from a bivariate normal distribution with correlation coefficient ρ and variances σ_x^2 and σ_y^2. We consider the following three increasingly restrictive covariance models:

Model 1: σ_x^2 and σ_y^2 both unknown;

Model 2: $\sigma_x^2/\sigma_y^2 \equiv a\sigma^2$, a known, σ^2 unknown;

Model 3: σ_x^2 and σ_y^2 both known.

Without essential loss of generality we assume throughout that the means are both 0, that $a = 1$ (so $\sigma_x^2 = \sigma_y^2 = \sigma^2$) in Model 2, and that $\sigma_x^2 = \sigma_y^2 = 1$ in Model 3.

These well-known models form the basis for many textbook examples and exercises. Models 1 and 2 are full exponential families with complete sufficient statistics $W_1 \equiv (s_x^2,\, s_y^2,\, s_{xy})$ and $W_2 \equiv (s_x^2 + s_y^2,\, s_{xy})$ respectively, where

$$s_x^2 = \frac{1}{n}\sum x_i^2, \quad s_y^2 = \frac{1}{n}\sum y_i^2, \quad s_{xy} = \frac{1}{n}\sum x_i y_i. \tag{1}$$

Both Models 1 and 2 admit explicit maximum likelihood estimators (MLE) for their unknown parameters:

Model 1: $\hat\rho = r_1, \quad \hat\sigma_x^2 = s_x^2, \quad \hat\sigma_y^2 = s_y^2$;

Model 2: $\hat\rho = r_2, \quad \hat\sigma^2 = \frac{1}{2}(s_x^2 + s_y^2)$;

where

$$r_1 = \frac{s_{xy}}{s_x s_y}, \qquad r_2 = \frac{2s_{xy}}{s_x^2 + s_y^2}. \tag{2}$$

The geometric mean of s_x^2 and s_y^2 in r_1 is replaced by the arithmetic mean in r_2, so $|r_2| < |r_1|$.

Model 3 is a curved exponential family with log likelihood function

$$l_n(\rho) = c - \frac{n}{2}\log(1-\rho^2) - \frac{n(s_x^2 + s_y^2)}{2(1-\rho^2)} + \frac{n\rho s_{xy}}{1-\rho^2}, \tag{3}$$

cf. Stuart and Ord (1991, Ch. 18). Here W_2 is minimal sufficient but not complete; note that

W_2 is two-dimensional while the unknown parameter ρ is one-dimensional. The MLE $\hat{\rho}$ is r_3, the maximizing root of the cubic equation

$$\rho(1-\rho^2) - \rho(s_x^2 + s_y^2) + (1+\rho^2)s_{xy} = 0. \tag{4}$$

In this paper we focus on Model 3, which is less amenable to exact analysis but was recently encountered by Fosdick and Raftery (2012) when comparing fertility rate forecast errors between developed and undeveloped countries. They considered several variants of three frequentist estimators for ρ, as well as several Bayesian estimators. The frequentist estimators were r_1, r_3, and the "empirical estimator"

$$r_4 = s_{xy} \tag{5}$$

s truncated to lie in $[-1, 1]$, where the estimated standard deviation estimates s_x and s_y in r_1 and r_2 are replaced by the known values 1. However, Fosdick and Raftery did not discuss the standard errors of their estimators.

It is well known[1] that as $n \to \infty$,

$$\begin{aligned}
\sqrt{n}(r_1 - \rho) &\to N\left(0, (1-\rho^2)^2\right) && \text{under Model 1,} && (6)\\
\sqrt{n}(r_2 - \rho) &\to N\left(0, (1-\rho^2)^2\right) && \text{under Model 2,} && (7)\\
\sqrt{n}(r_3 - \rho) &\to N\left(0, \frac{(1-\rho^2)^2}{1+\rho^2}\right) && \text{under Model 3,} && (8)\\
\sqrt{n}(r_4 - \rho) &\to N\left(0, 1+\rho^2\right) && \text{under Model 3.} && (9)
\end{aligned}$$

Because Model 3 is regular (sufficiently smooth), the MLE r_3 is asymptotically optimal for this model, that is, has smallest asymptotic variance among all asymptotically normal estimators.

Whereas the MLE r_3 is optimal for large samples sizes under Model 3, Fosdick and Raftery's interest was in the case of small and moderate sample sizes. In Table 1 the mean-squared errors

[1]E.g. Lehmann (1983, (19) p.441, Problems 6.2.20 and 6.5.9), Stuart and Ord (1991, Ch. 18)). The result (7) follows from (6) because $r_2 - r_1 = O_p(n^{-1})$ under Model 2 by a standard Taylor expansion argument.

of r_1, r_2, r_3, and r_4 have been obtained[2,3] by simulation. Here r_3 performs best unless ρ^2 is small, where r_2 may be somewhat better, while r_4 performs poorly in most cases. These effects become more marked as the sample size n increases, which is in agreement with the ordering of the asymptotic variances in (6)–(9).

Table 1: Root MSE ($\times$1000); r_4 truncated to [-1,1]. (largest standard error of estimates is 0.4).

n	Estimator	ρ^2					
		0	0.1	0.3	0.5	0.7	0.9
10	r_1	316	292	239	182	117	43
	r_2	301	281	236	184	122	46
	r_3	339	311	239	156	79	23
	r_4	314	314	303	286	273	272
20	r_1	224	204	163	120	74	26
	r_2	218	200	162	121	76	27
	r_3	238	208	147	94	52	16
	r_4	224	232	237	226	210	202
40	r_1	158	144	113	82	50	17
	r_2	156	142	113	83	51	17
	r_3	165	141	99	65	36	11
	r_4	158	166	177	176	162	149

In addition to its large MSE the estimator r_4 sometimes falls outside the admissible range $[-1, 1]$, while the Bayesian estimators are not amenable to frequentist interval estimation; these will not be discussed further here.

As is the case for r_1, convergence to normality is also slow for r_2 and r_3, so for small or moderate sample sizes the asymptotic variances in (6)–(8) do not provide good approximations for their standard errors (Table 2). Fisher's celebrated z transformation greatly improves the normal approximation to the distribution of r_1 under Model 1 so is invaluable for inference about ρ in this case – see §2.1. Interestingly, the *exact* distribution of $z(r_2)$ is easy to specify under Model 2 (see §2.2) and therefore under Model 3, where it provides estimates and tests that perform well for smaller values of ρ (see §2.4 and Section 3).

[2]The MSE is shown for a uniform range of ρ^2 values rather than ρ values, because it is ρ^2 rather than $|\rho|$ that indicates the strength of the relationship between x and y.

[3]Using the formulas for r_3 given in F&R, 0.0003% of estimates did not satisfy the cubic equation in (4) within $\pm$1e$-$8. For these simulations, the default 'optimize' procedure in `R` was used to obtain the estimate r_3 that maximizes the likelihood.

Table 2: Ratio of empirical variance to asymptotic variance.

n	Estimator	ρ^2					
		0	0.1	0.3	0.5	0.7	0.9
10	r_1	1.00	1.05	1.17	1.32	1.51	1.81
	r_2	0.91	0.98	1.14	1.35	1.65	2.11
	r_3	1.15	1.31	1.52	1.47	1.18	1.02
20	r_1	1.00	1.03	1.09	1.15	1.23	1.32
	r_2	0.95	0.99	1.08	1.18	1.30	1.45
	r_3	1.13	1.18	1.15	1.05	1.01	1.00
40	r_1	1.00	1.02	1.04	1.08	1.11	1.15
	r_2	0.98	1.00	1.04	1.09	1.14	1.20
	r_3	1.09	1.08	1.04	1.01	1.00	1.00

For Model 3, a new variance-stabilizing transformation y for the MLE r_3 is presented in §2.3. Unlike the z-transform, the y-transform must be adjusted for the sample size n to stabilize the confidence coverage of r_3. When this is done the resulting "confidence-stabilized" transformation $y_n(r_3)$ provides more precise confidence intervals under Model 3 than intervals based on r_1 and r_2 for moderate and large values of ρ. For small values, intervals based on $z(r_2)$ are preferable. This is demonstrated via simulation in §2.4.

The use of $z(r_2)$, $y(r_3)$, and $y_n(r_3)$ for comparing two or more correlation coefficients under Model 3 is outlined in Section 3.

F&R also considered the problem of testing $H_0 : \rho = 0$ (independence) vs. the one-sided alternative $H_1 : \rho > 0$ under Model 3. They considered the tests that reject H_0 for large values of r_1, r_3, and r_4 and their variants, as well as several Bayes tests, and approximated the significance levels of these tests by Monte Carlo simulation. It is well known that the r_1 test is exact for this problem. In Section 4 we note that the r_2 test is also exact and has an interesting although limited optimality property. Also we show that the r_4 test is locally most powerful for alternatives $\rho \downarrow 0$ and derive the asymptotically most powerful test (also exact) for alternatives $\rho \uparrow 1$. On the basis of numerical power comparisons, the r_2 (resp., r_3) test is recommended if small (resp., large) alternative values of $|\rho|$ are expected.

2. Confidence intervals for ρ under Model 3.

2.1. Confidence intervals based on the Model 1 MLE $\boldsymbol{r_1}$. Let g_γ denote the upper γ-quantile of the standard normal $\equiv$ Gaussian distribution. From (6), for sufficiently large n

$$r_1 \pm \left(\frac{1-r_1^2}{\sqrt{n-2}}\right) g_{\alpha/2} \tag{10}$$

is an approximate $1-\alpha$ confidence interval[4] for ρ under the unrestricted Model 1, hence for Models 2 and 3.

Unfortunately the sample size required for the accuracy of the normal approximation (6) depends on the unknown ρ, but this is remedied by Fisher's celebrated z-transformation for r_1 (cf. Anderson (1984, §4.2.3)), given by the indefinite integral

$$z(\rho) = \int \frac{d\rho}{1-\rho^2} = \frac{1}{2}\log\left(\frac{1+\rho}{1-\rho}\right). \tag{11}$$

This is a variance-stabilizing transformation that satisfies

$$\sqrt{n-2}\,[z(r_1) - z(\rho)] \to N(0,1) \tag{12}$$

with faster convergence to normality than (6) (see Table 3). This provides the approximate $1-\alpha$ confidence interval for ρ given by

$$z^{-1}\left(z(r_1) \pm \frac{g_{\alpha/2}}{\sqrt{n-2}}\right), \tag{13}$$

valid for Model 1 hence for Models 2 and 3. (Note that $z^{-1}(r) = \tanh(r)$.)

An *exact* confidence interval for ρ under Model 2 and therefore Model 3 is readily obtained from the classical Student t-distribution of the sample regression coefficient of y_i given x_i (e.g. Stuart and Ord (1987, eqn. 16.92)):

$$\sqrt{n-1}\left(\frac{r_1 - (\frac{s_x}{s_y})\rho}{\sqrt{1-r_1^2}}\right) \sim t_{n-1}. \tag{14}$$

[4]As is well known for (13), using $n-2$ rather than n results in a coverage probability somewhat closer to the nominal value $1-\alpha$.

This provides the exact $1-\alpha$ confidence interval[5]

$$\left(\frac{s_y}{s_x}\right)\left(r_1 \pm \sqrt{\frac{1-r_1^2}{n-1}}\,t_{n-1;\alpha/2}\right) \tag{15}$$

for ρ under Models 2 and 3, where $t_{n-1;\gamma}$ denotes the upper γ-quantile of the t_{n-1} distribution.

Although the interval (15) is exact, it is not a function of the minimal sufficient statistic $(s_x^2 + s_y^2, s_{xy})$ for Model 3 so may be inefficient; this is confirmed in §2.4.

2.2. Confidence intervals based on the Model 2 MLE r_2. Because r_2 has the same asymptotic normal distribution[6] as r_1,

$$r_2 \pm \left(\frac{1-r_2^2}{\sqrt{n-1}}\right) g_{\alpha/2} \tag{16}$$

is an approximate $1-\alpha$ confidence interval for ρ under Model 2, hence under Model 3. Furthermore, the z-transformation also applies to r_2 in this case, yielding the same normal approximation:

$$\sqrt{n-1}\,[z(r_2) - z(\rho)] \to N(0,1); \tag{17}$$

this gives another approximate $1-\alpha$ confidence interval for ρ under Model 2, hence Model 3:

$$z^{-1}\left(z(r_2) \pm \frac{g_{\alpha/2}}{\sqrt{n-1}}\right). \tag{18}$$

It is perhaps less well known that Fisher's z-transformation in fact applies *exactly* to r_2 under Model 2. The orthogonally-transformed random vectors

$$\begin{pmatrix} u_i \\ v_i \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}}(x_i + y_i) \\ \frac{1}{\sqrt{2}}(x_i - y_i) \end{pmatrix}, \quad i = 1, \ldots, n, \tag{19}$$

[5]By regressing x_i on y_i, a second exact $1-\alpha$ confidence interval is obtained by interchanging s_x and s_y in (15).

[6]See Footnote 1. Also, the entries for $z(r_2)$ in Table 3 show that its variance is approximated much better by $1/(n-1)$ than by $1/(n-2)$, hence its use in (17).

have the zero-mean bivariate normal distribution with covariance matrix

$$\begin{pmatrix} (1+\rho)\sigma^2 & 0 \\ 0 & (1-\rho)\sigma^2 \end{pmatrix} \equiv \begin{pmatrix} \sigma_u^2 & 0 \\ 0 & \sigma_v^2 \end{pmatrix}. \tag{20}$$

Thus $u_i \perp\!\!\!\perp v_i$, $i = 1, \ldots, n$, and

$$u^2 \equiv \sum u_i^2 \sim \sigma_u^2 \chi_n^2, \qquad v^2 \equiv \sum v_i^2 \sim \sigma_v^2 \chi_n^2, \tag{21}$$

$$\frac{u^2}{v^2} = \frac{1+r_2}{1-r_2} \sim \left(\frac{\sigma_u^2}{\sigma_v^2}\right) F_{n,n} \equiv \left(\frac{1+\rho}{1-\rho}\right) F_{n,n}. \tag{22}$$

Take logarithms to obtain the *exact* relation

$$z(r_2) \sim z(\rho) + \frac{1}{2} \log F_{n,n} \equiv z(\rho) + Z_n, \tag{23}$$

where Z_n denotes Fisher's Z distribution with n and n degrees of freedom (cf. Stuart and Ord (1987, §16.16)). This yields the following *exact* $1-\alpha$ confidence interval for ρ:

$$z^{-1}\left(z(r_2) \pm Z_{n;\alpha/2}\right), \tag{24}$$

valid under Model 2 hence Model 3. Here $Z_{n;\gamma}$ denotes the upper γ-quantile of Z_n, which can be expressed in terms of the γ-quantile of $F_{n,n}$.

Unlike r_1, r_2 is a function of the minimal sufficient statistic $W_2 \equiv (s_x^2 + s_y^2, s_{xy})$ for Model 3 so it may be expected to produce more efficient estimates than r_1 in this case. This should be most noticeable when ρ is small, since $|r_2| < |r_1|$; see Tables 1 and 2.

2.3. Confidence intervals based on the Model 3 MLE r_3. From (8), for sufficiently large n

$$r_3 \pm \left(\frac{1-r_3^2}{\sqrt{n(1+r_3^2)}}\right) g_{\alpha/2} \tag{25}$$

is an approximate $1-\alpha$ confidence interval for ρ under Model 3. As for r_1, however, the normal approximation (8) for r_3 is inaccurate unless the sample size is large (see Table 2). This suggests seeking a variance-stabilizing tranformation y for r_3 under Model 3.

Starting from (8) and applying *Mathematica* we obtain

$$\begin{aligned} y(\rho) &= \int \frac{\sqrt{1+\rho^2}}{1-\rho^2} d\rho \\ &= \frac{1}{\sqrt{2}} \log \left[\left(\frac{1+\rho}{1-\rho} \right) \left(\frac{\sqrt{2}\sqrt{1+\rho^2}+1+\rho}{\sqrt{2}\sqrt{1+\rho^2}+1-\rho} \right) \right] - \log\left(\sqrt{1+\rho^2}+\rho\right), \end{aligned} \tag{26}$$

so that

$$\sqrt{n-2}\,[y(r_3) - y(\rho)] \to N(0,1) \quad \text{as } n \to \infty. \tag{27}$$

Thus an approximate $1-\alpha$ confidence interval for ρ valid under Model 3 is given by[7]

$$y^{-1}\left(y(r_3) \pm \frac{g_{\alpha/2}}{\sqrt{n-2}} \right). \tag{28}$$

It follows from (26) that $y(0) = 0$, $y(\rho) = -y(-\rho)$ (antisymmetry), and $y'(\rho) \geq 1$ so y is strictly increasing on $(-1, 1)$. In fact $y'(\rho) \geq z'(\rho)$ by comparing (26) and (11), so y increases faster than z. This is most marked in the tails, as seen from Figure 1: $y(\rho) \approx z(\rho)$ for $0 \leq |\rho| \leq .5$, while $1 < y(\rho)/z(\rho) \uparrow$ in $|\rho|$ for $|\rho| > .5$. This form of y is needed to stabilize the variance of r_3 because its asymptotic variance is smaller than that of r_1 for larger values of $|\rho|$, cf. (6) and (8).

Table 3 shows, however, that when the sample size n is small or moderate, y is not entirely successful at stabilizing the variance of r_3 for ρ near 0. For example, when $n = 10$ and $\rho = 0$, the actual variance of $y(r_3)$ is 33% greater than the asymptotic approximation $1/(n-2)$ given by (27). Furthermore, in Table 5 it is seen that the coverage probability of the confidence interval (28) based on $y(r_3)$ may deviate noticeably from the nominal value $1-\alpha$ for $\alpha = .05$ when $|\rho|$ is small. This suggests making a multiplicative adjustment

$$y_n(\rho) \equiv m_n(\rho)\, y(\rho) \tag{29}$$

of the y-transformation such that $m_n(\rho) < 1$ for $|\rho|$ near 0 so that $y_n(\rho)$ will increase slower than $y(\rho)$ for ρ in that region, while $m_n(\rho) \approx 1$ for larger values of $|\rho|$.

[7]Like $\mathrm{Var}(z(r_1))$, $\mathrm{Var}(y(r_3))$ is better approximated by $\frac{1}{n-2}$ than by $\frac{1}{n}$; see Table 3.

Table 3: Empirical variance × $(n-2)$.

n	Estimator	ρ^2 0	0.1	0.3	0.5	0.7	0.9
10	$z(r_1)$	0.99	0.99	0.99	0.98	0.98	0.97
	$z(r_2)$	0.88	0.89	0.89	0.89	0.89	0.89
	$y(r_3)$	1.33	1.30	1.19	1.05	0.95	0.93
	$y_n(r_3)$	0.89	0.95	1.00	0.99	0.97	0.99
20	$z(r_1)$	1.00	1.00	1.00	0.99	0.99	0.99
	$z(r_2)$	0.95	0.95	0.95	0.95	0.95	0.95
	$y(r_3)$	1.22	1.17	1.06	0.99	0.98	0.97
	$y_n(r_3)$	0.96	0.98	0.98	0.97	0.99	1.00
40	$z(r_1)$	1.00	1.00	1.00	1.00	1.00	0.99
	$z(r_2)$	0.98	0.98	0.97	0.98	0.98	0.97
	$y(r_3)$	1.13	1.08	1.02	1.00	0.99	0.99
	$y_n(r_3)$	0.98	0.99	0.98	0.99	0.99	1.00

This can be accomplished by an ad hoc choice for $m_n(\rho)$ of the form

$$m_n(\rho) = \left(1 - \frac{a}{(n/10)} e^{-b|y(\rho)|^c}\right), \tag{30}$$

where a, b, and c are positive constants chosen as described below. Like y, $y_n(0) = 0$ and $y_n(\rho)$ is antisymmetric and strictly increasing on $(-1, 1)$. It is seen in Figure 1 that as desired, $y_n(\rho)$ increases slower than $y(\rho)$ for ρ near 0 and $y_n(\rho) \approx y(\rho)$ outside that region.

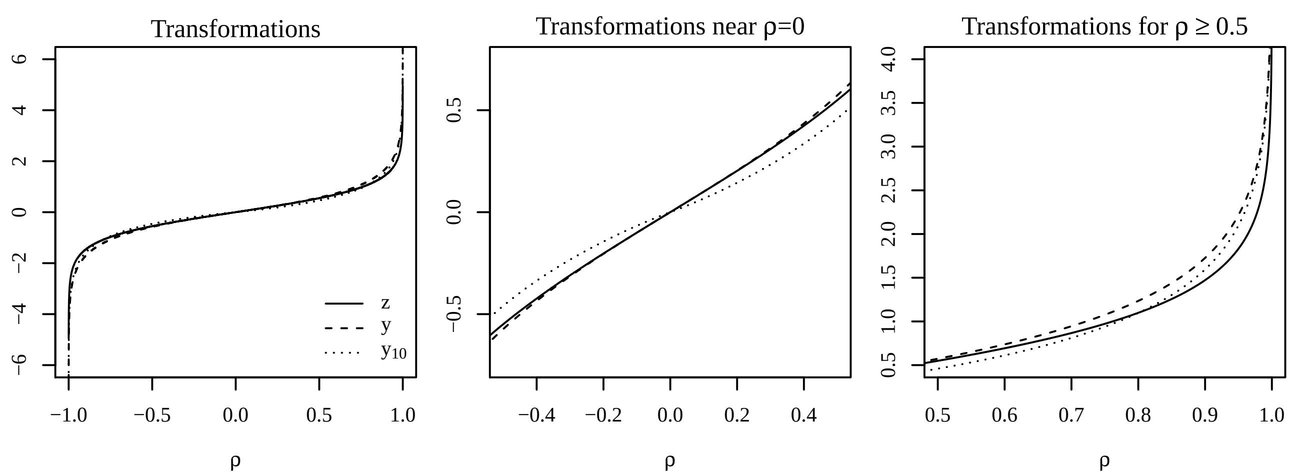

Figure 1: Transformations z, y, and y_n.

Because $y_n(r_3) - y(r_3) = O_p(n^{-1})$, $y_n(r_3)$ has the same asymptotic distribution as $y(r_3)$, namely

$$\sqrt{n-2}\,[y_n(r_3) - y_n(\rho)] \to N(0,1) \quad \text{as } n \to \infty, \tag{31}$$

so

$$y_n^{-1}\left(y_n(r_3) \pm \frac{g_{\alpha/2}}{\sqrt{n-2}}\right) \tag{32}$$

is also an approximate $1-\alpha$ confidence interval for ρ under Model 3. We choose a, b, and c to minimize the maximum difference between the empirical coverage probabilities of (32) and the nominal values $1-\alpha$ across the ranges of n, α, and ρ^2 considered in Table 4.

Specifically, we search for the triple (a,b,c) that satisfies

$$(a,b,c) = \underset{a,b,c}{\operatorname{argmin}} \max_{n\in N,\, \alpha\in L,\, \rho^2\in R} \left|(1-\alpha) - \Pr\left[y_n(\rho) \in y_n(r_3) \pm \frac{g_{\alpha/2}}{\sqrt{n-2}}\right]\right|,$$

where $N = \{10, 20, 40\}$, $L = \{.01, .05, .1\}$, and $R = \{0, .1, .3, .5, .7, .9\}$. Due to the non-convexity of the optimization problem, first we chose the best three triples (a,b,c) on the grid $A \times B \times C$, where $A = B = C = \{.1, .2, \ldots, 2.9, 3.0\}$. Next, three improved triples were obtained using the Nelder-Mead (1965) optimization procedure initialized at each of the first three chosen triples. The best of the three improved triples was selected for y_n: $a = 0.403$, $b = 1.091$, and $c = 0.775$. (See the Appendix for further details).

With this choice of (a,b,c) for y_n, Table 4 shows that the coverage probabilities for the $1-\alpha$ confidence interval (32) using $y_n(r_3)$ are acceptably close to the nominal value $1-\alpha$ for $\alpha \in A$, even though its variance remains somewhat below the nominal value $1/(n-2)$ when $\rho = 0$ (Table 3). For this reason y_n might better be called a "confidence-stabilizing" transformation, rather than "variance-stabilizing".

2.4. Comparison of the interval estimators for ρ. Table 5 shows the coverage probabilities for the nine $1-\alpha$ confidence intervals for ρ presented in §2.1, §2.2, and §2.3 when $\alpha = 0.05$, i.e., 95% confidence. Only five of these, marked by $\star$, attain or adequately approximate the nominal 95% level: the exact intervals (15) based on r_1 and (24) based on $z(r_2)$, and the approximate

Table 4: Coverage probabilities for the $(1-\alpha)$-confidence intervals based on $y_n(r_3)$.

α	n	ρ^2					
		0	0.1	0.3	0.5	0.7	0.9
.10	10	.912	.906	.908	.912	.911	.908
	20	.906	.906	.907	.908	.906	.903
	40	.904	.903	.905	.904	.903	.901
.05	10	.949	.944	.943	.949	.951	.949
	20	.947	.946	.950	.952	.951	.949
	40	.948	.949	.951	.951	.950	.950
.01	10	.983	.981	.978	.982	.984	.984
	20	.984	.983	.986	.987	.987	.987
	40	.986	.987	.988	.988	.988	.988

intervals (13) based on $z(r_1)$, (18) based on $z(r_2)$, and (31) based on the modified y-transform $y_n(r_3)$ (but not the approximate interval (28) based on $y(r_3)$ itself).

The average half-widths of these five interval estimators are shown in Table 6 for $\alpha = .10, .05, .01$. Of these five, (18) and (24), both based on $z(r_2)$, are most precise for small values of ρ while (31) based on $y_n(r_3)$ is most precise for intermediate and large values of ρ. The range of ρ values for which (31) is preferable to (18) and (24) expands as the sample size n increases, in accordance with the smaller asymptotic variance of r_3 except at $\rho = 0$, as seen in (7) and (8).

3. Comparing two or more correlations when the variances are known.

Suppose that samples of sizes $n^{(k)}$, $k = 1, \ldots, q$, are drawn from q bivariate normal distributions with unknown population correlations $\rho^{(k)}$ and known variances. Just as the z-transformation $z(r_1)$ is useful for combining or comparing two or more sample correlation coefficients under Models 1 and 2 (cf. Snedecor and Cochran (1967, §7.7)), the y-transform $y(r_3)$ or its modification $y_n(r_3)$ can be used for these purposes under Model 3, the known-variance case. The z-transform $z(r_2)$ can also be used, especially if small values of the correlations are expected.

(i) Suppose that $\rho^{(1)} = \cdots = \rho^{(q)} \equiv \rho$ and that it is desired to estimate this common ρ. If the sample sizes $n^{(k)}$ are large, we can weight the y-transforms of the Model 3 MLEs $r_3^{(1)}, \ldots, r_3^{(q)}$ according to their asymptotic inverse variances $n^{(1)}-2, \ldots, n^{(q)}-2$ to obtain the following weighted

Table 5: Coverage probabilities of 95% confidence intervals for ρ (largest standard error 0.0004). The intervals marked $\star$ have exact or approximate 95% coverage for all three samples sizes.

n		Confidence Interval	ρ^2 0	0.1	0.3	0.5	0.7	0.9
10		$r_1 \pm \left(\frac{1-r_1^2}{\sqrt{n-2}}\right) g_{\alpha/2}$	.893	.891	.889	.885	.879	.871
	$\star$	$z^{-1}\left(z(r_1) \pm \frac{g_{\alpha/2}}{\sqrt{n-2}}\right)$	.949	.949	.950	.950	.951	.952
	$\star$	$\left(\frac{s_y}{s_x}\right)\left(r_1 \pm \sqrt{\frac{1-r_1^2}{n-1}} t_{n-1;\alpha/2}\right)$	.950	.950	.950	.950	.950	.950
		$r_2 \pm \sqrt{\frac{1-r_2^2}{n-2}} g_{\alpha/2}$	.911	.912	.914	.914	.912	.909
	$\star$	$z^{-1}\left(z(r_2) \pm \frac{g_{\alpha/2}}{\sqrt{n-1}}\right)$	.949	.949	.949	.949	.949	.949
	$\star$	$z^{-1}\left(z(r_2) \pm Z_{n;\alpha/2}\right)$	.950	.950	.950	.950	.950	.950
		$r_3 \pm \left(\frac{1-r_3^2}{\sqrt{(n-2)(1+r_3^2)}}\right) g_{\alpha/2}$	.828	.842	.869	.884	.888	.887
		$y^{-1}\left(y(r_3) \pm \frac{g_{\alpha/2}}{\sqrt{n-2}}\right)$	.907	.911	.929	.946	.953	.954
	$\star$	$y_n^{-1}\left(y_n(r_3) \pm \frac{g_{\alpha/2}}{\sqrt{n-2}}\right)$	.949	.944	.943	.949	.951	.949
20		$r_1 \pm \left(\frac{1-r_1^2}{\sqrt{n-2}}\right) g_{\alpha/2}$	.921	.920	.917	.914	.910	.905
	$\star$	$z^{-1}\left(z(r_1) \pm \frac{g_{\alpha/2}}{\sqrt{n-2}}\right)$	.949	.949	.949	.950	.950	.950
	$\star$	$\left(\frac{s_y}{s_x}\right)\left(r_1 \pm \sqrt{\frac{1-r_1^2}{n-1}} t_{n-1;\alpha/2}\right)$	.950	.950	.950	.951	.950	.950
		$r_2 \pm \sqrt{\frac{1-r_2^2}{n-2}} g_{\alpha/2}$	.929	.929	.929	.929	.928	.926
	$\star$	$z^{-1}\left(z(r_2) \pm \frac{g_{\alpha/2}}{\sqrt{n-1}}\right)$	.949	.949	.949	.950	.949	.949
	$\star$	$z^{-1}\left(z(r_2) \pm Z_{n;\alpha/2}\right)$	.950	.950	.950	.950	.949	.950
		$r_3 \pm \left(\frac{1-r_3^2}{\sqrt{(n-2)(1+r_3^2)}}\right) g_{\alpha/2}$	.879	.892	.908	.913	.913	.912
		$y^{-1}\left(y(r_3) \pm \frac{g_{\alpha/2}}{\sqrt{n-2}}\right)$	.921	.929	.943	.950	.952	.952
	$\star$	$y_n^{-1}\left(y_n(r_3) \pm \frac{g_{\alpha/2}}{\sqrt{n-2}}\right)$	.947	.946	.950	.952	.951	.949
40		$r_1 \pm \left(\frac{1-r_1^2}{\sqrt{n-2}}\right) g_{\alpha/2}$	.935	.934	.933	.931	.929	.927
	$\star$	$z^{-1}\left(z(r_1) \pm \frac{g_{\alpha/2}}{\sqrt{n-2}}\right)$	.949	.949	.950	.950	.950	.950
	$\star$	$\left(\frac{s_y}{s_x}\right)\left(r_1 \pm \sqrt{\frac{1-r_1^2}{n-1}} t_{n-1;\alpha/2}\right)$	.950	.950	.950	.950	.950	.950
		$r_2 \pm \sqrt{\frac{1-r_2^2}{n-2}} g_{\alpha/2}$	.938	.939	.939	.939	.939	.938
	$\star$	$z^{-1}\left(z(r_2) \pm \frac{g_{\alpha/2}}{\sqrt{n-1}}\right)$	.949	.949	.950	.949	.949	.950
	$\star$	$z^{-1}\left(z(r_2) \pm Z_{n;\alpha/2}\right)$	.950	.950	.950	.950	.950	.950
		$r_3 \pm \left(\frac{1-r_3^2}{\sqrt{(n-2)(1+r_3^2)}}\right) g_{\alpha/2}$	.912	.920	.928	.930	.930	.931
		$y^{-1}\left(y(r_3) \pm \frac{g_{\alpha/2}}{\sqrt{n-2}}\right)$	.933	.939	.947	.950	.951	.951
	$\star$	$y_n^{-1}\left(y_n(r_3) \pm \frac{g_{\alpha/2}}{\sqrt{n-2}}\right)$	.948	.949	.951	.951	.950	.950

Table 6: Average half-width of the five $(1-\alpha)$-confidence intervals marked $\star$ in Table 5 (largest standard error 0.0002). The smallest half-width for each combination of n, ρ and α is in bold.

α	n	Confidence Interval	ρ^2					
			0	0.1	0.3	0.5	0.7	0.9
.10	10	$z^{-1}\left(z(r_1) \pm \frac{g_{\alpha/2}}{\sqrt{n-2}}\right)$	.483	.450	.377	.294	.195	.074
		$\left(\frac{s_y}{s_x}\right)\left(r_1 \pm \sqrt{\frac{1-r_1^2}{n-1}}t_{n-1;\alpha/2}\right)$	.565	.520	.429	.336	.237	.119
		$z^{-1}\left(z(r_2) \pm \frac{g_{\alpha/2}}{\sqrt{n-1}}\right)$	.463	.433	.366	.288	.194	.075
		$z^{-1}\left(z(r_2) \pm Z_{n;\alpha/2}\right)$	**.461**	**.431**	**.365**	.287	.193	.075
		$y_n^{-1}\left(y_n(r_3) \pm \frac{g_{\alpha/2}}{\sqrt{n-2}}\right)$	.532	.494	.397	**.276**	**.151**	**.046**
	20	$z^{-1}\left(z(r_1) \pm \frac{g_{\alpha/2}}{\sqrt{n-2}}\right)$	.353	.324	.263	.196	.124	.044
		$\left(\frac{s_y}{s_x}\right)\left(r_1 \pm \sqrt{\frac{1-r_1^2}{n-1}}t_{n-1;\alpha/2}\right)$	.395	.369	.313	.248	.175	.085
		$z^{-1}\left(z(r_2) \pm \frac{g_{\alpha/2}}{\sqrt{n-1}}\right)$	**.345**	.318	.259	.195	.124	.044
		$z^{-1}\left(z(r_2) \pm Z_{n;\alpha/2}\right)$	**.345**	**.317**	.259	.195	.124	.044
		$y_n^{-1}\left(y_n(r_3) \pm \frac{g_{\alpha/2}}{\sqrt{n-2}}\right)$	.377	.337	**.251**	**.168**	**.093**	**.029**
	40	$z^{-1}\left(z(r_1) \pm \frac{g_{\alpha/2}}{\sqrt{n-2}}\right)$	.255	.232	.184	.135	.083	.028
		$\left(\frac{s_y}{s_x}\right)\left(r_1 \pm \sqrt{\frac{1-r_1^2}{n-1}}t_{n-1;\alpha/2}\right)$	.270	.256	.224	.184	.133	.064
		$z^{-1}\left(z(r_2) \pm \frac{g_{\alpha/2}}{\sqrt{n-1}}\right)$	.252	**.229**	.183	.134	.083	.029
		$z^{-1}\left(z(r_2) \pm Z_{n;\alpha/2}\right)$	**.251**	**.229**	.183	.134	.083	.029
		$y_n^{-1}\left(y_n(r_3) \pm \frac{g_{\alpha/2}}{\sqrt{n-2}}\right)$	.265	.231	**.168**	**.112**	**.063**	**.020**
.05	10	$z^{-1}\left(z(r_1) \pm \frac{g_{\alpha/2}}{\sqrt{n-2}}\right)$	.558	.522	.444	.351	.239	.094
		$\left(\frac{s_y}{s_x}\right)\left(r_1 \pm \sqrt{\frac{1-r_1^2}{n-1}}t_{n-1;\alpha/2}\right)$	.670	.615	.507	.398	.282	.143
		$z^{-1}\left(z(r_2) \pm \frac{g_{\alpha/2}}{\sqrt{n-1}}\right)$	**.536**	**.504**	**.431**	.345	.237	.094
		$z^{-1}\left(z(r_2) \pm Z_{n;\alpha/2}\right)$	.538	.506	.433	.346	.238	.095
		$y_n^{-1}\left(y_n(r_3) \pm \frac{g_{\alpha/2}}{\sqrt{n-2}}\right)$	.600	.565	.471	**.340**	**.190**	**.057**
	20	$z^{-1}\left(z(r_1) \pm \frac{g_{\alpha/2}}{\sqrt{n-2}}\right)$	.414	.381	.312	.235	.150	.054
		$\left(\frac{s_y}{s_x}\right)\left(r_1 \pm \sqrt{\frac{1-r_1^2}{n-1}}t_{n-1;\alpha/2}\right)$	.476	.442	.370	.291	.204	.100
		$z^{-1}\left(z(r_2) \pm \frac{g_{\alpha/2}}{\sqrt{n-1}}\right)$	**.405**	**.374**	.308	.234	.150	.054
		$z^{-1}\left(z(r_2) \pm Z_{n;\alpha/2}\right)$	.406	.375	.309	.235	.151	.055

		$y_n^{-1}\left(y_n(r_3) \pm \frac{g_{\alpha/2}}{\sqrt{n-2}}\right)$	.436	.395	**.301**	**.203**	**.114**	**.035**
	40	$z^{-1}\left(z(r_1) \pm \frac{g_{\alpha/2}}{\sqrt{n-2}}\right)$	.301	.274	.219	.161	.100	.034
		$\left(\frac{s_y}{s_x}\right)\left(r_1 \pm \sqrt{\frac{1-r_1^2}{n-1}}t_{n-1;\alpha/2}\right)$	.324	.307	.267	.217	.155	.075
		$z^{-1}\left(z(r_2) \pm \frac{g_{\alpha/2}}{\sqrt{n-1}}\right)$	**.297**	**.271**	.218	.161	.100	.035
		$z^{-1}\left(z(r_2) \pm Z_{n;\alpha/2}\right)$	.298	.272	.218	.161	.100	.035
		$y_n^{-1}\left(y_n(r_3) \pm \frac{g_{\alpha/2}}{\sqrt{n-2}}\right)$	.311	.274	**.200**	**.134**	**.075**	**.024**
.01	10	$z^{-1}\left(z(r_1) \pm \frac{g_{\alpha/2}}{\sqrt{n-2}}\right)$	.681	.646	.565	.464	.332	.141
		$\left(\frac{s_y}{s_x}\right)\left(r_1 \pm \sqrt{\frac{1-r_1^2}{n-1}}t_{n-1;\alpha/2}\right)$	.843	.783	.657	.523	.377	.197
		$z^{-1}\left(z(r_2) \pm \frac{g_{\alpha/2}}{\sqrt{n-1}}\right)$	**.659**	**.626**	**.550**	**.454**	.326	.139
		$z^{-1}\left(z(r_2) \pm Z_{n;\alpha/2}\right)$	.671	.639	.563	.466	.337	.146
		$y_n^{-1}\left(y_n(r_3) \pm \frac{g_{\alpha/2}}{\sqrt{n-2}}\right)$	.707	.679	.599	.468	**.280**	**.084**
	20	$z^{-1}\left(z(r_1) \pm \frac{g_{\alpha/2}}{\sqrt{n-2}}\right)$	.522	.485	.405	.313	.205	.076
		$\left(\frac{s_y}{s_x}\right)\left(r_1 \pm \sqrt{\frac{1-r_1^2}{n-1}}t_{n-1;\alpha/2}\right)$	.637	.583	.476	.369	.259	.129
		$z^{-1}\left(z(r_2) \pm \frac{g_{\alpha/2}}{\sqrt{n-1}}\right)$	**.512**	**.477**	.400	.310	.205	.077
		$z^{-1}\left(z(r_2) \pm Z_{n;\alpha/2}\right)$	.518	.483	.405	.315	.208	.078
		$y_n^{-1}\left(y_n(r_3) \pm \frac{g_{\alpha/2}}{\sqrt{n-2}}\right)$	.538	.499	**.398**	**.278**	**.157**	**.049**
	40	$z^{-1}\left(z(r_1) \pm \frac{g_{\alpha/2}}{\sqrt{n-2}}\right)$	.387	.354	.287	.213	.134	.047
		$\left(\frac{s_y}{s_x}\right)\left(r_1 \pm \sqrt{\frac{1-r_1^2}{n-1}}t_{n-1;\alpha/2}\right)$	.433	.409	.350	.278	.195	.094
		$z^{-1}\left(z(r_2) \pm \frac{g_{\alpha/2}}{\sqrt{n-1}}\right)$	**.382**	**.351**	.285	.213	.134	.047
		$z^{-1}\left(z(r_2) \pm Z_{n;\alpha/2}\right)$	.385	.354	.287	.215	.135	.048
		$y_n^{-1}\left(y_n(r_3) \pm \frac{g_{\alpha/2}}{\sqrt{n-2}}\right)$	.395	.354	**.266**	**.180**	**.101**	**.032**

estimator for $y(\rho)$:

$$\bar{y}_w(r_3) = \frac{\sum_{k=1}^{q}(n^{(k)}-2)y(r_3^{(k)})}{\sum_{k=1}^{q}(n^{(k)}-2)} \approx N\left(y(\rho), \frac{1}{\sum_{k=1}^{q}(n^{(k)}-2)}\right). \tag{33}$$

This provides an approximate $1-\alpha$ confidence interval for $y(\rho)$, which is then inverted to obtain an approximate $1-\alpha$ confidence interval for ρ:

$$y^{-1}\left(\bar{y}_w(r_3) \pm \frac{g_{\alpha/2}}{\sqrt{\sum_{k=1}^q (n^{(k)}-2)}}\right). \tag{34}$$

If the sample sizes are small or moderate, however, then $\bar{y}_w(r_3)$ should be replaced by

$$\bar{y}_{n,w}(r_3) = \frac{\sum_{k=1}^q (n^{(k)}-2) y_{n^{(k)}}(r_3^{(k)})}{\sum_{k=1}^q (n^{(k)}-2)} \approx N\left(\bar{y}_{n,w}(\rho),\ \frac{1}{\sum_{k=1}^q (n^{(k)}-2)}\right), \tag{35}$$

where

$$\bar{y}_{n,w}(\rho) = \bar{m}_{n,w}(\rho)\, y(\rho) \equiv \left[\frac{\sum_{k=1}^q (n^{(k)}-2) m_{n^{(k)}}(\rho)}{\sum_{k=1}^q (n^{(k)}-2)}\right] y(\rho). \tag{36}$$

Then (35) provides an approximate $1-\alpha$ confidence interval for $\bar{y}_{n,w}(\rho)$, which is strictly increasing in ρ hence can be inverted to provide an approximate $1-\alpha$ confidence interval for ρ:

$$\bar{y}_{n,w}^{-1}\left(\bar{y}_{n,w}(r_3) \pm \frac{g_{\alpha/2}}{\sqrt{\sum_{k=1}^q (n^{(k)}-2)}}\right). \tag{37}$$

If the sample sizes are equal, i.e., $n^{(1)} = \cdots = n^{(q)} \equiv n$, then (35) and (37) simplify to

$$\bar{y}_n(r_3) = \frac{1}{q}\sum_{k=1}^q y_n(r_3^{(k)}) \approx N\left(y_n(\rho),\ \frac{1}{q(n-2)}\right), \tag{38}$$

$$\bar{y}_n^{-1}\left(\bar{y}_n(r_3) \pm \frac{g_{\alpha/2}}{\sqrt{q(n-2)}}\right). \tag{39}$$

Lastly, if it is expected that the common ρ is small, then the findings in §2.4 suggest that $y(r_3^{(k)})$ be replaced by the z-transforms $z(r_2^{(k)})$ of the Model 2 MLEs to obtain a more precise estimate of ρ. Thus (33) would be replaced by (recall (17))

$$\bar{z}_w(r_2) = \frac{\sum_{k=1}^q (n^{(k)}-1) z(r_2^{(k)})}{\sum_{k=1}^q (n^{(k)}-1)} \approx N\left(z(\rho),\ \frac{1}{\sum_{k=1}^q (n^{(k)}-1)}\right), \tag{40}$$

which provides an approximate $1-\alpha$ confidence interval for $z(\rho)$ that is then inverted to obtain an approximate $1-\alpha$ confidence interval for ρ:

$$z^{-1}\left(\bar{z}_w(r_2) \pm \frac{g_{\alpha/2}}{\sqrt{\sum_{k=1}^{q}(n^{(k)}-1)}}\right). \tag{41}$$

Of course, if the underlying bivariate normal data is available from the q populations, then these data should be combined into a single sample of size $n^{(1)}+\cdots+n^{(q)}$ from which a more efficient estimate of the common ρ can be obtained by the methods of Section 2. However, the weighted estimates obtained here would still be useful for testing homogeneity of the $\rho^{(k)}$ as now described.

(ii) The homogeneity hypothesis $H_0 : \rho^{(1)} = \cdots = \rho^{(q)}$ is equivalent to homogeneity of the y-transforms: $y(\rho^{(1)}) = \cdots = y(\rho^{(q)})$. To test this against the general alternative, if the sample sizes are large we may reject H_0 for large values of the weighted chi-square statistic

$$T_{y,w} = \sum_{k=1}^{q}(n^{(k)}-2)\left(y(r_3^{(k)}) - \bar{y}_w(r_3)\right)^2, \tag{42}$$

distributed approximately as χ^2_{q-1} under H_0.

If the samples sizes are not large but equal, i.e., $n^{(1)} = \cdots = n^{(q)} \equiv n$, then H_0 is equivalent to homogeneity of the modified y transforms, i.e., $y_n(\rho^{(1)}) = \cdots = y_n(\rho^{(q)})$, so $T_{y,w}$ can be replaced by the statistic

$$T_{y_n} = (n-2)\sum_{k=1}^{q}\left(y_n(r_3^{(k)}) - \bar{y}_n(r_3)\right)^2, \tag{43}$$

distributed approximately as χ^2_{q-1} under H_0. However, if the sample sizes are not equal then H_0 is *not* equivalent to homogeneity of $y_{n^{(1)}}(\rho^{(1)}), \ldots, y_{n^{(q)}}(\rho^{(q)})$, so the weighted test statistic

$$T_{y_n,w} = \sum_{k=1}^{q}(n^{(k)}-2)\left(y_{n^{(k)}}(r_3^{(k)}) - \bar{y}_{n,w}(r_3)\right)^2 \tag{44}$$

is not necessarily appropriate for testing H_0.

Finally, if it is expected that $\rho^{(1)}, \ldots, \rho^{(q)}$ are small, then we would reject H_0 for large values of the weighted chi-square statistic

$$T_{z,w} = \sum\nolimits_{k=1}^{q} (n^{(k)} - 1) \left(z(r_2^{(k)}) - \bar{z}_w(r_2) \right)^2, \tag{45}$$

also distributed approximately as χ^2_{q-1} under H_0.

4. Testing $\rho = 0$ (bivariate independence) under Model 3.

Fosdick and Raftery (2012) also discussed the problem of testing $\rho = 0$ vs. $\rho > 0$ in a single bivariate normal population with known variances. They considered tests that reject H_0 for large values of r_1, r_3, and r_4 and their variants, as well as several Bayes tests, determining the critical of these tests by Monte Carlo simulation. Here we add a few observations about these tests and some others for this testing problem.

Exact tests for independence are well known for Models 1 and 2. Under Model 1 the test that rejects $\rho = 0$ for large values of r_1 is the uniformly most powerful unbiased (UMPU) test for $\rho = 0$ vs. $\rho > 0$ (Lehmann (1986, §5.15)). When $\rho = 0$ it follows from (14) that

$$\sqrt{n-1} \left(\frac{r_1}{\sqrt{1 - r_1^2}} \right) \sim t_{n-1}, \tag{46}$$

from which the exact null distribution of this test is readily obtained.

Under Model 2 the problem of testing $\rho = 0$ vs. $\rho > 0$ is equivalent to the problem of comparing two normal variances, i.e., testing $\sigma_u^2 = \sigma_v^2$ vs. $\sigma_u^2 > \sigma_v^2$ (recall (22)). The F-test that rejects $\rho = 0$ if $\frac{1+r_2}{1-r_2} \equiv \frac{u^2}{v^2} > F_{n,n;\alpha}$ is UMPU level α under Model 2 (Lehmann (1986, §5.3)). Therefore this test is exact for testing $\rho = 0$ vs. $\rho > 0$ under Model 3 as well, and should perform reasonably well there. (Note that $\sigma_u^2 + \sigma_v^2 = 2$ under Model 3.)

Under Model 3, it follows from (3) that the pdf of (x_i, y_i) does not have monotone likelihood ratio (MLR) and that no uniformly most powerful (UMP) test exists for $\rho = 0$ vs. $\rho > 0$. In fact, for a fixed alternative $\rho_1 > 0$ the most powerful level α test rejects $\rho = 0$ iff

$$s_{xy} > \frac{\rho_1 (s_x^2 + s_y^2)}{2} + c_\alpha \tag{47}$$

where c_α is chosen to attain size α. Thus the locally most powerful (LMP) level α test for alternatives $\rho_1 \downarrow 0$ rejects H_0 iff $r_4 = s_{xy} > c_\alpha$, whereas the asymptotically most powerful (AMP) level α test for alternatives $\rho_1 \uparrow 1$ rejects H_0 iff $2s_{xy} - s_x^2 - s_y^2 > c_\alpha$, equivalently, iff $v^2 < \chi^2_{n;1-\alpha}$ (recall (19) and (21)). Since these two tests are different, no UMP test exists under Model 3.

The exact test based on r_2 has an interesting albeit limited optimality property under Model 3. If we set $c_\alpha = 0$ in (47), it follows that the test that rejects $\rho = 0$ if $r_2 > \rho_1$ is the MP test of its size for the fixed alternative $\rho_1 > 0$. By (23) this size is given by

$$\begin{aligned} \alpha(\rho_1) &\equiv \Pr[r_2 > \rho_1 \mid \rho = 0] \\ &= 1 - F_{n,n}\left(e^{2z(\rho_1)}\right), \end{aligned} \tag{48}$$

where $F_{n,n}(\cdot)$ denotes the cdf of the $F_{n,n}$ distribution. Values of $\alpha(\rho_1)$ are shown in Table 7.

Table 7: Size $\alpha(\rho_1)$ of the MP test of $\rho = 0$ vs $\rho = \rho_1$ that rejects when $r_2 > \rho_1$.

n	ρ_1^2					
	0^+	0.1	0.3	0.5	0.7	0.9
10	0.5	0.1583	0.0326	0.0051	3.5e-4	1.3e-6
20	0.5	0.0758	0.0042	0.0001	6.1e-7	<1e-10
40	0.5	0.0207	8.7e-5	8.3e-8	<1e-10	<1e-10

Under Model 3 the powers of the size .05 tests for $\rho = 0$ vs. $\rho > 0$ based on r_1, r_2, r_3, r_4 (the LMP test) and v^2 (the AMP test) are compared via simulation[8] in Table 8. The LMP (AMP) test is dominated in power by the other three tests for all except very small (very large) values of the alternative ρ_1, so is not recommended. The r_1 test is dominated by the r_2 test but only slightly, which suggests that for testing purposes not much power is gained from the knowledge that the variances are equal.

For sample size $n = 10$ the r_2 test dominates the r_3 test for $\rho_1^2 \le .3$ while the reverse is true when $\rho_1^2 \ge .5$. For sample sizes 20 and 40 this crossover occurs for $\rho_1^2 \in (.1, .3)$. Thus we recommend either the r_2 or r_3 test under Model 3, depending on whether small values or large

[8] The size .05 critical values for the tests based on r_3 and r_4 also were obtained by simulation, whereas the exact critical values were used for the other three tests.

Table 8: Power when testing $\rho = 0$ vs $\rho > 0$ for various alternatives ρ_1 at the $\alpha = 0.05$ significance level.

n	Test Rejection Criterion	ρ_1^2 0.01	0.05	0.1	0.3	0.5	0.7	0.9
10	$\sqrt{n-1}\left(\frac{r_1}{\sqrt{1-r_1^2}}\right) > t_{n-1;\alpha}$	0.088	0.164	0.250	0.580	0.843	0.975	1.000
	$\frac{1+r_2}{1-r_2} > F_{n,n;\alpha}$	0.088	0.166	0.252	0.585	0.847	0.976	1.000
	$r_3 > c_{\alpha,n}$	0.076	0.130	0.195	0.518	0.860	0.996	1.000
	$r_4 > c_{\alpha,n}$	0.093	0.171	0.247	0.482	0.651	0.769	0.850
	$v^2 < \chi^2_{n;1-\alpha}$	0.072	0.114	0.166	0.440	0.800	0.993	1.000
20	$\sqrt{n-1}\left(\frac{r_1}{\sqrt{1-r_1^2}}\right) > t_{n-1;\alpha}$	0.113	0.255	0.413	0.851	0.985	1.000	1.000
	$\frac{1+r_2}{1-r_2} > F_{n,n;\alpha}$	0.113	0.255	0.414	0.853	0.986	1.000	1.000
	$r_3 > c_{\alpha,n}$	0.100	0.218	0.365	0.859	0.996	1.000	1.000
	$r_4 > c_{\alpha,n}$	0.115	0.252	0.390	0.751	0.911	0.971	0.991
	$v^2 < \chi^2_{n;1-\alpha}$	0.086	0.168	0.275	0.757	0.989	1.000	1.000
40	$\sqrt{n-1}\left(\frac{r_1}{\sqrt{1-r_1^2}}\right) > t_{n-1;\alpha}$	0.154	0.410	0.656	0.986	1.000	1.000	1.000
	$\frac{1+r_2}{1-r_2} > F_{n,n;\alpha}$	0.154	0.411	0.656	0.986	1.000	1.000	1.000
	$r_3 > c_{\alpha,n}$	0.143	0.386	0.641	0.992	1.000	1.000	1.000
	$r_4 > c_{\alpha,n}$	0.155	0.398	0.619	0.959	0.997	1.000	1.000
	$v^2 < \chi^2_{n;1-\alpha}$	0.110	0.269	0.474	0.971	1.000	1.000	1.000

values of the alternative ρ_1 are of most interest. Because the r_2 test is exact, it is easier to apply than the r_3 test, so the r_2 test might be recommended for Model 3 on this basis.

Remark. Under Model 3, $\mathrm{E}_\rho(s_x^2 + s_y^2) = 2$ for all ρ, but $s_x^2 + s_y^2$ is not an ancillary statistic. Nonetheless it might be of interest to consider the conditional test based on the conditional distribution of s_{xy} given $s_x^2 + s_y^2$. □

References

Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis, 2nd edition.* Wiley, New York.

Fosdick, B. K. and Raftery, A. E. (2012). Estimating the correlation in bivariate normal data with known variances and small sample sizes. *The American Statistician*, **66**(1): 34-41.

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

Lehmann, E. L. (1986). *Testing Statistical Hypotheses, 2nd edition.* Wiley, New York.

Nelder, J. A. and Mead, R. (1965). A simplex method for function minimization. *Computer Journal*, **7**:308-313.

Snedecor, G. W. and Cochran, W. G. (1967). *Statistical Methods.* Iowa State University Press, Ames, Iowa.

Stuart, A. and Ord, J. K. (1987, 1991). *Kendall's Advanced Theory of Statistics, Volumes 1 and 2.* Oxford University Press, New York.

Appendix: Choosing a, b, and c for y_n

The $y_n(\rho)$ transformation has the form

$$y_n(\rho) = \left(1 - \frac{a}{(n/10)} e^{-b|y(\rho)|^c}\right) y(\rho).$$

The constants a, b, and c are defined as

$$(a,b,c) = \operatorname*{argmin}_{a,b,c} f(a,b,c), \text{ where}$$
$$f(a,b,c) = \max_{n\in N,\ \alpha\in L,\ \rho^2\in R} \left|(1-\alpha) - \Pr\left[y_n(\rho) \in y_n(r_3) \pm \frac{g_{\alpha/2}}{\sqrt{n-2}}\right]\right|,$$

$N = \{10, 20, 40\}$, $L = \{0.01, 0.05, 0.1\}$, and $R = \{0, 0.1, 0.3, 0.5, 0.7, 0.9\}$. The procedure used to search for (a,b,c) is outlined below; the results associated with each step are also included.

1. Find the three triples $(\tilde{a}, \tilde{b}, \tilde{c})$ on the grid $A \times B \times C$ that result in the smallest values of $f(\tilde{a}, \tilde{b}, \tilde{c})$, where $A = B = C = \{0.1, 0.2, ..., 2.9, 3.0\}$.

$\tilde{a}$	$\tilde{b}$	$\tilde{c}$	$f(\tilde{a}, \tilde{b}, \tilde{c})$
0.8	1.8	0.5	0.012626
0.4	1.1	0.8	0.012646
1.3	2.3	0.4	0.012835

2. Initialize the Nelder-Mead optimization algorithm in `R` at each of the $(\tilde{a}, \tilde{b}, \tilde{c})$ triples found in Step 1 and obtain three improved triples (a, b, c).

Initial Values			Improved Values			$f(a,b,c)$
$\tilde{a}$	$\tilde{b}$	$\tilde{c}$	a	b	c	
0.8	1.8	0.5	0.797	1.800	0.500	0.012461
0.4	1.1	0.8	0.403	1.091	0.775	0.012348
1.3	2.3	0.4	1.304	2.304	0.403	0.012567

3. From the three improved triples obtained in Step 2, select that which results in the smallest value of $f(a, b, c)$.

$$(a, b, c) = (0.403, 1.091, 0.775)$$