SMALL SAMPLE EMPIRICAL CRITICAL VALUES AS A TOOL FOR THE COMPARISON OF MULTIVARIATE NORMALITY GOF TESTS.

Jorge Luis Romeu
Mathematics
SUNY-Cortland, NY 13045
jromeu@suvm.bitnet

Abstract.

Existing Multivariate Normality (MVN) Goodness of Fit (GOF) Tests either follow a known asymptotic distribution (e.g. Mardia's) or are empirical (e.g. Malkovich and Affifi). When samples are small the asymptotic theory cannot be safely invoked. Hence, their asymptotic distribution percentiles are no longer accurate. In such cases empirical critical values (ecv) are derived via Monte Carlo.

We have thus obtained ecv's for eight well known MVN GOF tests for n=25(25)200, p=2(1)6(2)10 and medium and high p-variate correlations. Using the ecv's we statistically study several characteristics of the unknown small sample distributions of these MVN tests. Then we present criteria as to when and where the asymptotic critical values can be safely used.

1.0 Introduction and Background.

This research stems from the work to demonstrate our newly developed MVN GOF test (Ozturk and Romeu, 1992). We conducted an extensive Monte Carlo power study (under a Cornell Theory Center award) to compare it with eight carefully selected and well established MVN GOF tests: Mardia's Skewness and Kurtosis, Royston's W, Cox and Small, Malkovich-Affifi, Hawkins and Koziol's Angles and Chi Square tests. For space and brevity we refer the reader to Romeu (1990), for a complete list of references.

We soon realized how several of these MVN tests lacked any asymptotic distributions. And those who had one, converged to it very slowly, rendering them impractical when samples were "small".
We also observed how some of these tests (i) were prone to detect certain types of departure from multivariate normality (say skewness) but not others (say kurtosis). Or how (ii) correlation among p-variates affected some tests very seriously. Or how the ecv's were (iii) severely affected by sample size or (iv) by number of p-variates. Or how (v) certain algorithm results varied from one computer to the other. Or a combination of any of the above mentioned problems.

We wanted to further investigate such problems and to study how the different tests fared under them. But, lacking the theoretical tool to undertake such comparative study (i.e. small sample distributions) we took an indirect approach. We thus, used the ecv's obtained by simulating five or ten thousand test results, to characterize the unknown statistical distributions.

2.0 Research Methods.

The validation of our Monte Carlo study is discussed elsewhere (Romeu, 1992). In this paper we present the statistical results regarding the mentioned problems of the eight MVN GOF tests.

First, the effect of sample size was investigated by regressing ecv's on inverse sample size 1/n, for fixed number of p-variates and percentile (PCT). Some results, for bivariate normals, are shown in Table 1. Each test (MTD), their asymptotic critical values (CV), the regression independent term (Bo) and Index of Fit (IF) are given. A 95% confidence interval (CI) for each regression's Bo, for Mardia's Skewness (MSK) and Cox-Small (Cox) test, cover their asymptotic CV. However, Mardia's Kurtosis (MKT), converges much slower and its 95% CI for Bo doesn't cover its asymptotic CV.

Table 1: Regression ecv = f(n).

MTD	PCT	Bo	CV	IF
MSK	0.90	7.79	7.78	0.99
MSK	0.95	9.67	9.49	0.98
MSK	0.99	13.77	13.28	0.98
MKT	0.90	1.63	1.65	0.99
MKT	0.95	2.08	1.95	0.99
MKT	0.99	2.90	2.58	0.92
COX	0.90	4.58	4.61	0.98
COX	0.95	5.99	5.99	0.94
COX	0.99	9.10	9.21	0.94

Next we investigated the "smallest" sample size n, required for the safe use of asymptotic critical values. Table 2 show examples of coverage (COV) of the 95th asymptotic percentile (CV) by a 95% non parametric CI, derived for that sample size (n) and no. of p-variates (p). Verify how, for different n and p, methods varied widely. Again, Mardia's Skewness ecv's cover the true asymptotic percentile for n=200 while Kurtosis, which converges at a slower rate, doesn't.

Table 2: Approximate CI for ecv's.

MTD	n	p	COV	n	p	COV
MSK	50	2	NO	50	8	NO
MSK	100	2	NO	100	8	NO
MSK	200	2	YES	200	8	YES
MKT	50	2	NO	50	8	NO
MKT	100	2	NO	100	8	NO
MKT	200	2	NO	200	8	NO

Next we investigated the effect of p-variate correlation. In practice, the covariance matrix is seldom known. Instead, it is estimated from the data and used in the GOF tests. In our power study we had also observed wide differences for low and high correlation.

We investigated this problem by taking, at prefixed and regular intervals of the order statistics (at 0.9(0.05)0.995) differences of ecv's obtained for rho=0.5 and 0.9. There were two alternatives: (i) this difference was statistically zero (no correlation effect) or (ii) there was an effect of correlation. If so, this effect produced a shift problem or a scale one. Hence, the differences in ecv's would (or would not) be independent of how far out they were obtained. In the first case we used (i) paired (Wilcoxon/Sign) tests. In the second, (ii) regression of these differences on its percentiles. If there was a significance difference in (i) above, we obtained its 95% non parametric CI.

In Table 3 we report some of these non parametric 95% confidence intervals for ecv differences, for those methods that showed none or very small effect of p-variate correlation.

Table 3: CI for ecv Differences

MTD	p	Lower	Upper
Skewness	2	-0.22	-0.16
Skewness	4	0.12	0.26
Hawkins	2	0.018	0.025
Hawkins	4	-0.008	0.015
Cox-Small	2	-0.077	-0.012
Kurtosis	4	0.047	0.062

We also investigated the effect of the number of p-variates on ecv's. Several MVN methods (e.g. Koziol) required large n and small p for its use, which is not always met in practice. We again used regression of ecv's on p-variates (for p=2,3,4,5,6,8,10) for fixed sample size and asymptotic percentile. We verified how, with the exception of Koziol's Angles and Hawkins' tests, all others heavily depend on p-variates too.

We then reanalyzed ecv differences for low/high correlation, now for fixed asymptotic percentile (PC) and n, but increasing no. of p-variates. We verified how, with the exception of Royston's and Koziol's Angles tests, all other power results could be pooled. An example is shown in Table 4.

Table 4: 95% CI for ecv diff(p).

Test	PC	n	LB.	UB.
Skewness	0.90	25	-0.48	0.13
Skewness	0.95	25	-0.47	0.36
Skewness	0.99	25	-1.02	1.40

Of practical consideration is the effect of hardware on the calculation of ecv's. We observed some variation for two tests (Royston's and Koziol's Angles). We performed non parametric paired comparisons between ecv's obtained in Syracuse University's IBM 3090 and Cornell's Supercomputer. Differences (fixed p=2, n=50) for correlations of 0.5 and 0.9 were obtained for successive ecv percentiles 0.9(0.005)0.995. We followed the same approach above described for the analysis of correlation effect. Descriptive statistics of some of our analyses are presented in Table 5. For relative comparison of the effect of these differences on ecv's, the mean ecv value, by method, is also given.

Table 5 :Hardware effect comparison.

MTD	Q.25	Q.5	Q.75	Mean
MSK	-0.048	-0.020	0.075	8.75
ROY	0.883	0.975	1.208	5.26
HAW	0.008	0.014	0.039	1.30
CHI	0.002	0.005	0.006	0.22

Notice how ecv's for Royston's test vary in the order of 15%, while those for Skewness (MSK), Hawkins and Koziol's Chi Square tests are, for practical purposes, negligible. We conjecture that calculation of sensitive quantities (e.g. eigenvalues/eigenvectors) in the denominators of Koziol's Angles and Royston's algorithms account for such large differences, when processed in two significantly different machines as those used.

A complete set of tables of small sample ecv's for the MVN GOF tests discussed in this paper can be found in Romeu and Ozturk (1991).

3.0 Research Results.

The main result of this paper concerns the determination of the appropriate sample size for asymptotic values. Mardia's skewness test requires more than

100 observations before the use of asymptotic critical values is appropriate. The same holds for Cox and Small and Koziol. For n=200, our ecv's 95% CI results cover the asymptotic values. Mardia's Kurtosis test converges much slower and 95% CI obtained for n=200 do not cover the asymptotic percentiles. Hence, a sample of size 200 is still inadequate for using asymptotic critical values. Since, in practice, samples may be much smaller than that, our empirical critical values provide a useful tool for the practitioner.

Our ecv's regression results also indicate that test that don't have a known asymptotic distribution (e.g. Malkovich and Affifi) also converge as a function of 1/n. Hence, a function may be found that approximates this test's unknown asymptotic distribution for large samples.

The next result of interest pertains to the effect of p-variate correlation on the power of the tests (or equivalently on their ecv's). This is of importance, since the covariance matrix is generally unknown and estimated from the samples. We concluded that only two procedures, Royston's W and Koziol's Angles test (as well as the Sigma Inverse implementation of our own multivariate Qn test) are seriously affected by p-variate correlation. Separate ecv's have been provided for medium (0.5) and high (0.9) rho, for those two tests. All other methods analyzed may be considered, for practical purpose, as approximately correlation free and single tables of ecv's are provided.

Two MVN GOF tests have been found quite sensitive to hardware effect: Koziol's Angles and Royston's W. We caution the practitioner to calibrate our results with those of his own machine before using our ecv's.

Finally, we have also shown how empirical critical values, obtained from simulation, can become effective tools. The statistical study and comparison of the small sample (unknown) distributions of these MVN GOF tests was, both, required but infeasible with the conventional research tools (closed form distribution). We had few other alternatives, since the true distributions were either unknown or available only when the sample size was very large, which rendered them useless for our needs.

The use of the ecv's as a characterization of these unknown small sample distributions allowed us to investigate this problem.

<u>4.0 Bibliography</u>.

Ozturk A. and J. L. Romeu (1992). A New Method for Assessing Multivariate Normality With Graphical Applications. Comm. in Stat. -Simula. (21)1, 15-34.

Romeu, J. L. (1990). Development and Evaluation of a General Procedure for Assessing Multivariate Normality. CASE Center Tech. Report 9022. Syracuse University, Syracuse NY 13244.

Romeu, J. L. (1992) Small Sample Empirical Critical Values For Multivariate Normality Tests. ASA Winter Conference. Louisville, KY.

Romeu, J. L. and A. Ozturk (1991). A Comparative Study of Goodness of Fit Tests for Multivariate Normality. (Submitted for publication).

Table 6: Example of ecv's power comparison results (n=25; p=2).

PERCENT REJECTIONS FOR N= 20000 TOTAL CASES.

METHOD:	ALPHA=0.10	ALPHA=0.05	ALPHA=0.01
CHOLESKI	0.09710	0.04675	0.00920
SIGMA	0.09755	0.04845	0.01025
M-SKEW	0.09860	0.04645	0.00910
M-KURT	0.09960	0.04975	0.01060
COX-SMAL	0.09560	0.04860	0.00895
ROYSTONW	0.10585	0.05415	0.01065
MALKOV	0.09960	0.04860	0.00910
KOZ-CHI	0.10155	0.05135	0.00985
KOZANGLE	0.10230	0.05140	0.00975
HAWKINS	0.10150	0.05100	0.01005

TABLE NO. 7 CRITICAL VALUES FOR THE CASE P = 2 VARIATES.

RHO=0.5 N	%	SKEWNESS TEST	KURTOSIS LOWER	KURTOSIS UPPER	ROYSTON W	MALKOVICH AFIFI	KOZIOL CHI-SQR.	COX-SMAL REG	HAWKINS TEST	KOZIOL ANGLES
25	90	5.87	-1.22	0.88	4.56	0.913	0.167	5.13	1.036	4.53
25	95	7.48	-1.33	1.24	5.98	0.895	0.271	6.90	1.290	5.87
25	99	11.24	-1.52	2.03	9.28	0.855	0.317	11.62	1.937	9.07
50	90	6.67	-1.37	1.19	4.50	0.953	0.173	4.73	1.074	4.45
50	95	8.27	-1.53	1.58	5.93	0.945	0.219	6.29	1.346	5.94
50	99	12.18	-1.76	2.49	9.29	0.926	0.337	9.97	1.997	9.15
75	90	7.03	-1.42	1.39	4.07	0.965	0.172	4.79	1.067	4.59
75	95	9.06	-1.58	1.84	5.43	0.959	0.218	6.28	1.321	5.93
75	99	13.07	-1.90	3.12	8.61	0.946	0.338	9.75	1.977	8.76
100	90	7.33	-1.44	1.45	4.63	0.971	0.170	4.71	1.051	4.62
100	95	9.14	-1.61	1.85	5.98	0.967	0.217	6.23	1.308	6.06
100	99	13.26	-1.95	2.89	9.43	0.957	0.326	9.68	1.931	9.48
125	90	7.35	-1.52	1.45	4.31	0.975	0.180	4.72	1.077	4.52
125	95	9.35	-1.70	1.88	5.60	0.971	0.227	5.99	1.343	5.82
125	99	13.35	-2.05	2.93	8.93	0.963	0.337	9.34	1.988	8.74
150	90	7.44	-1.47	1.47	4.25	0.977	0.171	4.79	1.035	4.64
150	95	9.48	-1.70	1.87	5.80	0.973	0.218	6.14	1.293	5.85
150	99	13.61	-2.11	2.80	9.05	0.967	0.324	9.65	1.942	9.08
175	90	7.62	-1.48	1.55	4.41	0.978	0.170	4.71	1.044	4.67
175	95	9.40	-1.69	2.02	5.88	0.976	0.214	6.35	1.303	6.12
175	99	13.07	-2.06	2.78	9.25	0.970	0.347	9.33	2.025	9.37
200	90	7.64	-1.51	1.57	4.73	0.979	0.174	4.62	1.063	4.61
200	95	9.43	-1.75	1.99	6.31	0.977	0.220	6.10	1.319	5.99
200	99	13.37	-2.12	3.02	9.94	0.972	0.350	9.68	1.986	9.07

Small Sample Empirical Critical Values as a Tool for the Comparison Of Multivariate Normality GOF Tests.

Jorge Luis Romeu
Mathematics
SUNY-Cortland
JROMEU@SUVM

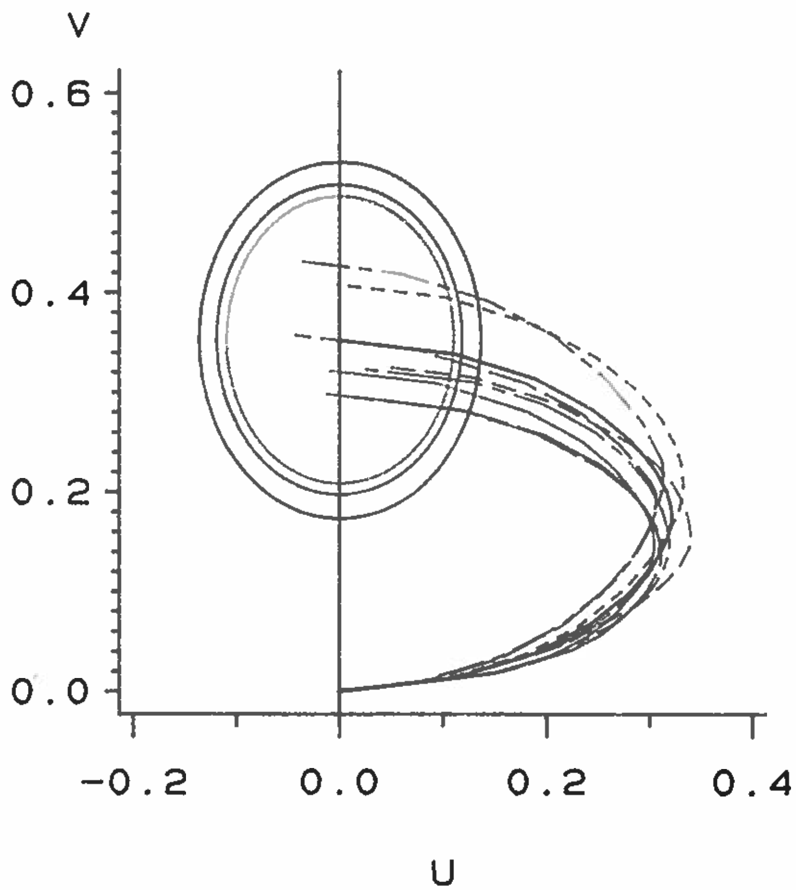
Presented at the
Interface '92 Simposium

Texas A & M University
College Station, TX

March of 1992.

MULTIVARIATE STANDARIZED GOF TEST.

ORSA DATA: MULTIVAR.

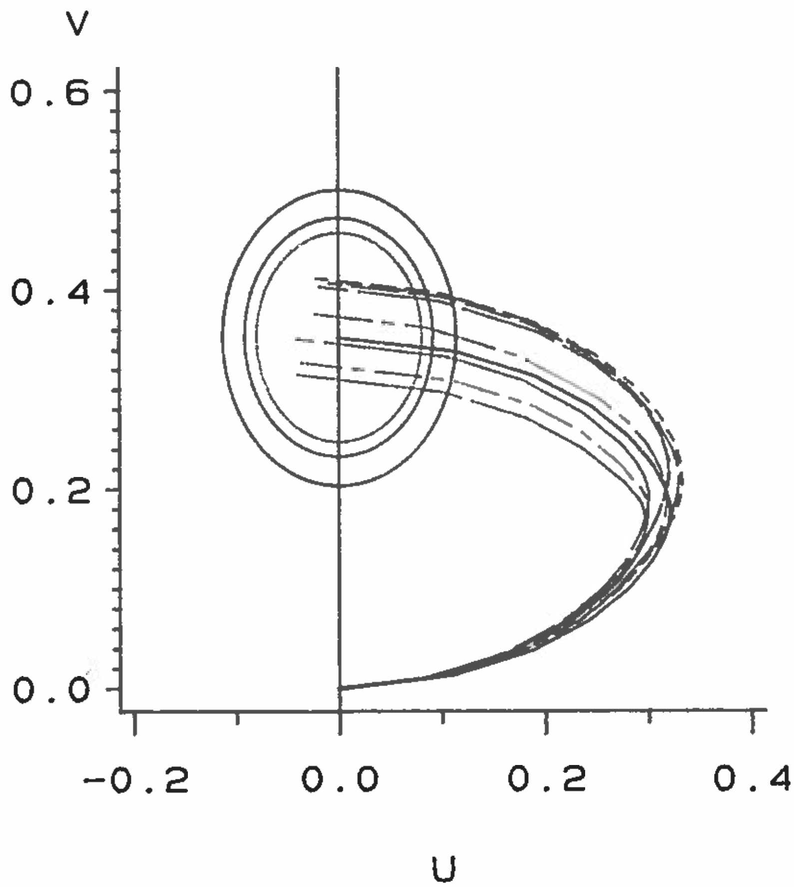


NULL: SOLID BLACK.

```
   ORSA DATA: MULTIVARIATE.
 UU= 0.00000000000000000E+00 UV= 0.352405015379190445     RHO=
  0.00000000000000000E+00
 VARU= 0.139582037809304893E-02 VARV= 0.239539076574146748E-02
 U(          1 )=-0.133451058921983667E-02
 V(          1 )= 0.407880592788740157
 U(          2 )= 0.395858632066939942E-01
 V(          2 )= 0.326028326352823500
 U(          3 )=-0.421448828398449937E-01
 V(          3 )= 0.357090115939490235
 U(          4 )=-0.117652969875390451E-01
 V(          4 )= 0.298031182791999805
 U(          5 )= 0.140294618081499151E-01
 V(          5 )= 0.295249213806604696
 U(          6 )=-0.925887306348702199E-02
 V(          6 )= 0.321402206577264038
 U(          7 )=-0.434878810762080598E-01
 V(          7 )= 0.431905416609520462
 U(          8 )= 0.171338962014607465E-01
 V(          8 )= 0.324031796822023391
 U(          9 )= 0.467694689032516209E-06
 V(          9 )= 0.352499614142144896
 PROB OF ERROR OF            1  IS:  0.525699373074244747
 PROB OF ERROR OF            2  IS:  0.493340453424635075
 PROB OF ERROR OF            3  IS:  0.526852196354867491
 PROB OF ERROR OF            4  IS:  0.513394828844348602
 PROB OF ERROR OF            5  IS:  0.471236427420124920
 PROB OF ERROR OF            6  IS:  0.793470992940267461
 PROB OF ERROR OF            7  IS:  0.135780554148641944
 PROB OF ERROR OF            8  IS:  0.760940726713653426
 PROB OF ERROR OF            9  IS:  0.999998131976397697
```

MULTIVARIATE STANDARIZED GOF TEST.

ORSA DATA: UNIVARIATE.



NULL: SOLID BLACK.

```
 ORSA ANALYSIS: UNIVARIATE.
UU= 0.000000000000000000E+00 UV= 0.352405015379190445     RHO=
 0.000000000000000000E+00
VARU= 0.139582037809304893E-02 VARV= 0.239539076574146748E-02
U(           1 )=-0.133451056005173818E-02
V(           1 )= 0.407880592764176458
U(           2 )=-0.208448254037331276E-01
V(           2 )= 0.411457020439948626
U(           3 )=-0.892307703477879244E-02
V(           3 )= 0.406485598382310867
U(           4 )=-0.182316038099993027E-01
V(           4 )= 0.403102811103540723
U(           5 )=-0.399306685207473228E-01
V(           5 )= 0.314848399837278622
U(           6 )=-0.406319976815460726E-01
V(           6 )= 0.350861596067087209
U(           7 )=-0.362134233556734991E-01
V(           7 )= 0.326848602034276112
U(           8 )=-0.223762058402355857E-01
V(           8 )= 0.375435488879668156
U(           9 )= 0.467694689032516209E-06
V(           9 )= 0.352499614142144896
PROB OF ERROR OF          1  IS:  0.525699373387963728
PROB OF ERROR OF          2  IS:  0.413320928690073350
PROB OF ERROR OF          3  IS:  0.527815626757881143
PROB OF ERROR OF          4  IS:  0.519146470097831700
PROB OF ERROR OF          5  IS:  0.420809933678129675
PROB OF ERROR OF          6  IS:  0.553279983330174738
PROB OF ERROR OF          7  IS:  0.545477062421547032
PROB OF ERROR OF          8  IS:  0.748212582174009572
PROB OF ERROR OF          9  IS:  0.999998131976397697
```