







330  
B385  
1993:171 COPY 2

STX

## A Large Sample Normality Test

The Library of the  
JAN 6 1994  
University of Illinois  
of Urbana-Champaign

*Anil K. Bera*  
*Department of Economics*  
*University of Illinois*

*Pin T. Ng*  
*Department of Economics*  
*University of Houston, TX*



# BEBR

FACULTY WORKING PAPER NO. 93-0171

College of Commerce and Business Administration


University of Illinois at Urbana-Champaign

November 1993

## A Large Sample Normality Test

Anil K. Bera

Pin T. Ng



Digitized by the Internet Archive  
in 2011 with funding from  
University of Illinois Urbana-Champaign

# A LARGE SAMPLE NORMALITY TEST

ANIL K. BERA and PIN T. NG

Department of Economics, University of Illinois, Champaign, IL 61820

Department of Economics, University of Houston, TX 77204-5882

November 22, 1993

## Abstract

The score function, defined as the negative logarithmic derivative of the probability density function, plays an ubiquitous role in statistics. Since the score function of the normal distribution is linear, testing normality amounts to checking the linearity of the empirical score function. Using the score function, we present a graphical alternative to the Q-Q plot for detecting departures from normality. Even though graphical approaches are informative, they lack the objectivity of formal testing procedures. We, therefore, supplement our graphical approach with a formal large sample chi-square test. Our graphical approach is then applied to a wide range of alternative data generating processes. The finite sample size and power performances of the chi square test are investigated through a small scale Monte Carlo study.

KEY WORDS: Normality test; Score function; Graphical approach

## 1 Introduction

Since Geary's (1947) suggestion of putting the statement "Normality is a myth. There never was and will never be, a normal distribution" in front of all statistical texts, the need to test for the normality assumption in many statistical models has been widely acknowledged. As a result, a wide range of tests for normality are currently available. Most of these tests basically fall into the following categories: (1) tests based on probability or Q-Q plots, (2) moments tests, (3) distance tests based on the empirical distribution function, (4) goodness of fit tests, and (5) tests based on the empirical characteristic function.

No single test statistic can reveal as much information as a graphical display. In Section 2, we present a graphical alternative to the Q-Q plot using the score function, defined as the negative logarithmic derivative of the probability density function. Even though graphical approaches are informative, they lack the objectivity of formal testing procedures. We therefore supplement our graphical approach with a formal large sample  $\chi^2$  test based on the score function in Section 3. The performances of our graphical approach and score function based  $\chi^2$  test depend on our ability to estimate the score function accurately. We review some score function estimators in Section 4.



In Jarque and Bera (1987), a moment test was shown to possess superior powers compared to most other normality tests. Their moment test utilizes the normal distribution's skewness measure  $\sqrt{b_1} = 0$  and kurtosis measure  $b_2 = 3.0$ . As a result, under certain non normal distributions with skewness and kurtosis measures identical to the normal distribution, moment tests based on  $\sqrt{b_1}$  and  $b_2$  will have no power. Some of such distribution are Tukey's  $\lambda$  distributions when  $\lambda = 0.135$  and  $5.2$  [see Joiner and Rosenblatt (1971)]. Moment based tests also have power against only certain alternatives. Our score function based  $\chi^2$  test, on the other hand, does not have this disadvantage. The superior power of our score function based  $\chi^2$  test is demonstrated in a small scale Monte Carlo study in Section 5.

## 2 A Graphical Approach

The score function, defined as  $\psi(x) = -\log' f(x) = -\frac{f'(x)}{f(x)}$ , of a random variable having probability density function  $f(x)$  plays an ubiquitous role in statistics. It is related to the constructions of L-, M- and R-estimators for location and scale model as well as regression models in the robustness literatures. [See Joiner and Hall (1983) for an excellent overview]. It is also used in constructing various adaptive L-, M- and R-estimators which achieve the Cramer-Rao efficiency bounds asymptotically. [See Koenker (1982)]. It can also be used to estimate the Fisher information. In hypothesis testing, the score function plays a crucial role in robustifying conventional testing procedures. [See Bickel (1978) and Bera and Ng (1992)]. Its fundamental contribution to statistics, however, can best be seen in the realm of exploratory data analysis.

The plots of the density and score functions of some common distributions are presented in Figure 1 and Figure 2 respectively. While it is difficult to differentiate the tails of a Gaussian distribution from those of a Cauchy distribution through the density functions, the tails of their score functions are very distinct. In fact, we can easily distinguish among various distributions by investigating the score functions.

It is clear from Figure 1 and Figure 2 that the mode of a distribution is characterized by an upward crossing of the score function at the horizontal axis while an anti-mode is located at the point of downward crossing. An exponential distribution has a horizontal score function. A tail thicker than the exponential has a negatively sloped score while a tail thinner than the exponential corresponds to an upward sloping score.

A Gaussian distribution has a linear score function passing through the horizontal axis at its location parameter with a slope equal to the reciprocal of its variance. This suggests an alternative to the familiar and popular probability or Q-Q plot. An estimated score function with a redescending tail towards the horizontal axis indicates departure towards distributions with thicker tails than the normal distribution while a diverging tail suggests departure in the direction of thinner tailed distributions.

We can even recover the estimate of the density function through exponentiating the negative integral of the estimated score function although this may seem to be a roundabout approach.



Figure 1: Probability Density Functions

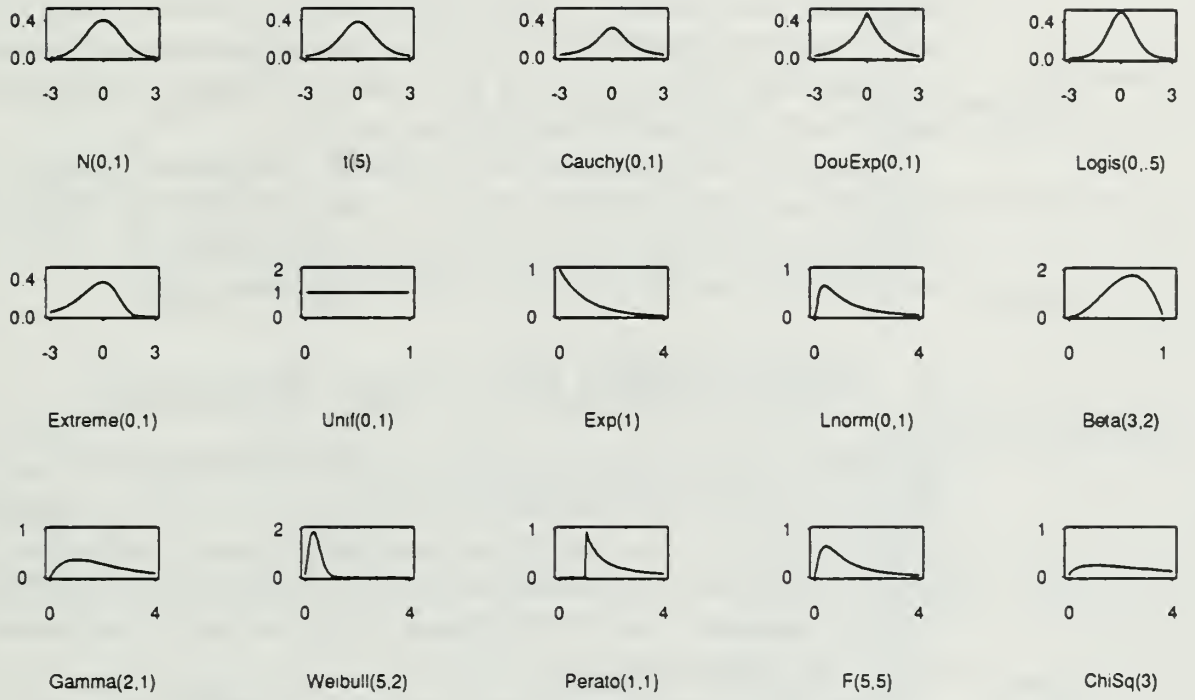
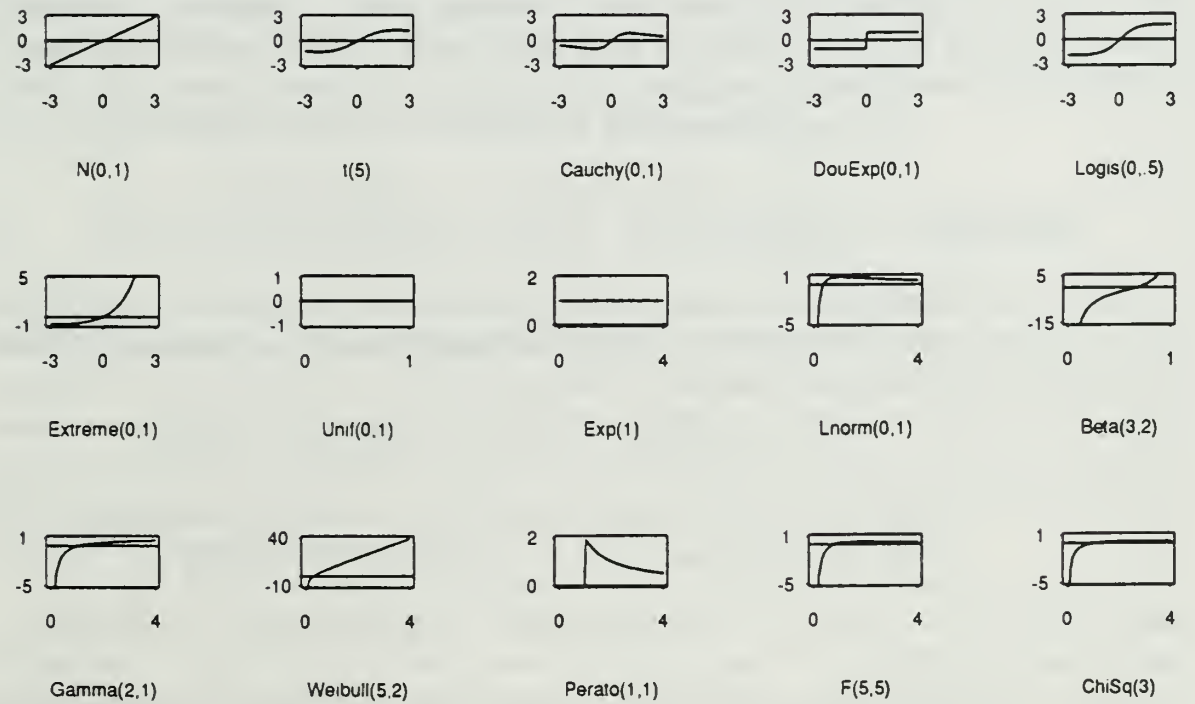


Figure 2: Score Functions



### 3 A Formal Test

A formal "objective" test on the null hypothesis of a Gaussian distribution is equivalent to testing the linearity of the score function. Since a straight line can be viewed as a first order approximation to any polynomial, the normality test can easily be carried out through the asymptotic  $\chi^2$  test of regressing the estimated score function  $\hat{\psi}(x_i)$  on some polynomial of  $x_i$ . The null hypothesis of a Gaussian distribution will correspond to the linear relationship between  $\hat{\psi}(x_i)$  and  $x_i$ .

When the null hypothesis of a Gaussian distribution cannot be rejected, we can estimate the location parameter by the point at which the ordinary least squares regression line intersects the horizontal axis and the estimate of the scale parameter will be the square root of the reciprocal of the regression slope.

### 4 Estimating the Score Function

Performances of the above graphical approach and formal  $\chi^2$  test rely on accurate estimates of the score functions. Numerous score function estimators are available, most of which are constructed from some kernel density estimators. [See Stone (1975), Manski (1984) and Cox and Martin (1988)]. Csörgö and Révész (1983) used a nearest-neighbor approach. Cox (1985) proposed a smoothing spline version, which is further refined and implemented in Ng (1994).

It has often been argued that the choice of kernel is not crucial in kernel density estimation. The correct choice of kernel, however, becomes important in the tails where density is low and few observations will help smooth things out. This sensitivity to kernel choice is further amplified in score function estimation where higher derivatives of the density function are involved [see Portnoy and Koenker (1989), and Ng (1994)].

Ng (1994) found that the smoothing spline score estimator, which finds its theoretical justification from an explicit mean squared errors minimization criterion, is more robust than the kernel estimators to distributional variations. We use this score estimator in the paper.

The smoothing spline score estimator is the solution to

$$\min_{\psi \in H_2[a,b]} \int (\psi^2 - 2\psi') dF_n + \lambda \int (\psi''(x))^2 dx \quad (1)$$

where  $H_2[a,b] = \{\psi : \psi, \psi' \text{ are absolutely continuous, and } \int_a^b [\psi''(x)]^2 dx < \infty\}$ . The objective function (1) is the (penalized) empirical analogue of minimizing the following mean-squared error:

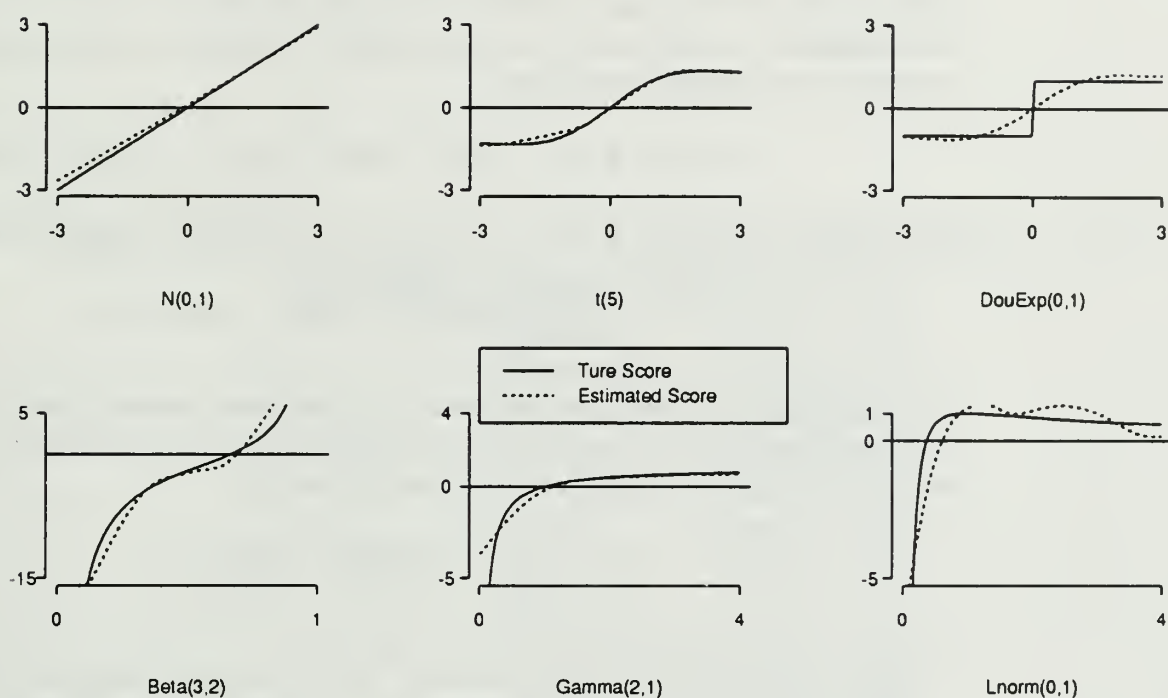
$$\int (\psi - \psi_0)^2 dF_0 = \int (\psi^2 - 2\psi') dF_0 + \int \psi_0^2 dF_0 \quad (2)$$

in which  $\psi_0$  is the unknown true score function and the equality is due to the fact that under some mild regularity conditions [see Cox (1985)]

$$\int \psi_0 \psi dF_0 = - \int f'_0(x) \psi(x) dx = \int \psi' dF_0.$$

Since the second term on the right hand side of (2) is independent of  $\psi$ , minimizing the mean-squared error may focus exclusively on the first term. Minimizing (1) yields a

Figure 3: Estimated Score Functions



balance between “fidelity-to-data” measured by the mean-squared error term and the smoothness represented by the second term. As in any nonparametric score function estimator, the smoothing spline score estimator has the penalty parameter  $\lambda$  to choose. The penalty parameter merely controls the trade-off between “fidelity-to-data” and smoothness of the estimated score function. We use the automated penalty parameter choice mechanism, the *adaptive information criteria*, suggested and implemented in Ng (1994) [see Ng (1991) for a FORTRAN source codes].

## 5 Some Examples and Simulation Results

In Figure 3, we present the smoothing spline estimated score functions for each of the 100 random observations drawn from some of the distributions in Figure 2. The random number generators were Marsaglia’s Super-Duper random number generators available in “S” [Becker, Chambers and Wilks (1988)] installed on a Sun SPARCstation 10. The smoothing spline score estimator was Ng’s (1991) Fortran version adapted for “S”. It is obvious from Figure 3 that any departure from the Gaussian distribution can be easily detected from the plots.

To study the finite sample properties of our score function based  $\chi^2$  test and the moment based  $LM$  test of Jarque and Bera (1987), we perform a small scale Monte Carlo study. The  $LM$  test was shown in Jarque and Bera (1987) to possess very good power as compared to the skewness measure test  $\sqrt{b_1}$ , the Kurtosis measure



test  $b_2$ , D'Agostino's (1971)  $D^*$  test, Pearson, D'Agostino and Bowman's (1977)  $R$  test, Shapiro and Wilk's (1965)  $W$  test, and Shapiro and Francia's (1972)  $W'$  test against the various alternative distributions investigated. As a result, we use it as our bench mark to evaluate the performance of our  $\chi^2$  test. The null distribution here is the standard normal distribution and the alternatives are Gamma (2,1), Beta (3,2), Student's  $t$  (5) and Tukey's  $\lambda$  distribution with  $\lambda = 5.2$ . All distributions are standardized to have zero mean and variance twenty-five. Our  $\chi^2$  test is obtained by running the following regression

$$\hat{\psi}(x_i) = \gamma_0 + \gamma_1 x_i + \gamma_2 x_i^2 + \gamma_3 x_i^3 + \gamma_4 x_i^4 + \gamma_5 x_i^5 + \gamma_6 x_i^6 + \epsilon_i$$

and testing for  $H_0 : \gamma_2 = \gamma_3 = \gamma_4 = \gamma_5 = \gamma_6 = 0$ . The  $\chi^2$  test statistic is then given by

$$\frac{(RSS_r - RSS)}{RSS/(N - 7)} \xrightarrow{D} \chi_5^2 \quad \text{under } H_0$$

where  $RSS_r$  is the restricted residual sum of squares obtained from regressing  $\hat{\psi}(x_i)$  on the intercept and  $x_i$  alone,  $RSS$  is the residual sum of squares of the whole regression and  $N$  is the sample size. The  $LM$  test is given by

$$LM = N \left[ \frac{\sqrt{b_1}^2}{6} + \frac{(b_2 - 3)^2}{24} \right]$$

Under the null hypothesis of normality,  $LM$  is asymptotically distributed as a  $\chi_2^2$ .

The estimated sizes and powers of the  $LM$  and  $\chi^2$  test in 1000 Monte Carlo replications are reported in Table 1. The standard errors of the estimated probabilities are no greater than  $\sqrt{.25/1000} \cong .016$ . The sample sizes considered are 25, 50, 100 and 250. The performances of the  $\chi^2$  test from regressing the  $\hat{\psi}(x_i)$  on some higher order polynomials of  $x_i$  were also investigated. The results are similar so we choose not to report those here. Under the Gaussian distribution, the estimated probabilities of Type I error are computed from the true  $\chi^2$  critical points.

From Table 1, we can see that the estimated Type I errors of the  $\chi^2$  test are much closer than the  $LM$  test to the nominal value of .10 for all sample sizes. The  $LM$  test under estimated the sizes of the test in all the sample sizes we investigated.

To make a valid power comparison, we size adjust the power under all alternative distributions. The empirical significance level we use is 10%. At the smaller sample sizes of  $N=25$  and 50, the  $LM$  test has higher powers than the  $\chi^2$  test under Gamma (2,1), Log (0,1) and  $t(5)$ . The discrepancies, however, become less prominent as the sample size increases. This is due to the fact that the score functions of both Log (0,1) and Gamma (1,2) are approximately linear in the high density regions as can be seen from Figure 2. More observations in the tails will be needed to facilitate estimation of the score functions that are distinguishable from the linear Gaussian score. The situation is similar in the Student's  $t(5)$ . However, an even bigger sample size will probably be needed for some realizations in the tails to discern the estimated score function of the Student's  $t$  from that of the Gaussian. As expected, the  $\chi^2$  test has some power for  $\lambda(5.2)$  and this increases rapidly with the sample size. The  $\chi^2$  test also performs better for Beta (3,2) alternative. The  $LM$  test has powers even lower than its sizes for the Tukey's  $\lambda$  alternative in all sample sizes.

Table 1: Estimated Powers for 1000 Replications (Empirical size = .10)

Sample Sizes	Distributions	LM	$\chi^2$
$N = 25$	Gaussian	.044	.126
	Beta (3,2)	.063	.213
	Gamma (2,1)	.700	.625
	t (5)	.352	.163
	Log (0,1)	.962	.795
	$\lambda(5.2)$	.09	.152
$N = 50$	Gaussian	.059	.111
	Beta (3,2)	.154	.370
	Gamma (2,1)	.944	.872
	t (5)	.517	.260
	Log(0,1)	1.00	.927
	$\lambda(5.2)$	.061	.334
$N = 100$	Gaussian	.067	.099
	Beta (3,2)	.433	.548
	Gamma (2,1)	.998	.990
	t (5)	.699	.347
	Log(0,1)	1.00	1.00
	$\lambda(5.2)$	.019	.641
$N = 250$	Gaussian	.085	.091
	Beta (3,2)	.980	.901
	Gamma (2,1)	1.00	1.00
	t (5)	.933	.701
	Log(0,1)	1.00	1.00
	$\lambda(5.2)$	.011	.980

Based on our examples and simulation results, we conclude that the estimated score function is informative in performing exploratory data analysis. It also allows us to formulate a formal large sample test for normality that possesses reasonable size and good power properties under finite sample situations.

### Acknowledgement

The authors would like to thank Roger Koenker and Robin Sickles for their helpful suggestions and incisive comments. The computations were performed on computing facilities support by the National Science Foundation Grants SES 89-22472.

## References

- [1] Bera, A.K. and Ng, P.T. (1992), "Robust Tests for Heteroskedasticity and Autocorrelation Using Score Function," Tilburg University, CentER for Economic Research, Discussion Paper No. 9245.
- [2] Bickel, P.J. (1978), "Using Residuals Robustly I: Tests for Heteroscedasticity, Nonlinearity," *The Annals of Statistics*, 6, 266-291.
- [3] Cox, D.D. (1985), "A Penalty Method for Nonparametric Estimation of the Logarithmic Derivative of a Density Function," *Annals of the Institute of Statistical Mathematics*, 37, 271-288.
- [4] Cox, D.D. and Martin, D.R. (1988), "Estimation of Score Functions", Technical Report, University of Washington.
- [5] Csörgö, M. and Révész, P. (1983), "An N.N-estimator for the Score Function," *Seminarbericht Nr.49, Proceedings of the First Easter Conference on Model Theory, Sektion Mathematik*.
- [6] D'Agostino, R.B. (1971), "An Omnibus Test for Normality for Moderate and Large Size Samples," *Biometrika*, 58, 341-348.
- [7] Geary, R.C. (1947), "Testing for Normality," *Biometrika*, 34, 209-242.
- [8] Jarque, C.M. and Bera, A.K. (1987), "Test for Normality of Observations and Regression Residuals," *International and Statistical Review*, 55, 163-172.
- [9] Joiner, B.L. and Hall, D.L. (1983), "The Ubiquitous Role of  $f'/f$  in Efficient Estimation of Location," *The American Statistician*, 37, 128-133.
- [10] Joiner, B.L. and Rosenblatt, J.R. (1971), "Some Properties of the Range in Samples from Tukey's Symmetric Lambda Distributions," *Journal of the American Statistical Association*, 66, 394-399.
- [11] Koenker, R. (1982), "Robust Methods in Econometrics," *Econometric Review*, 1, 214-255.
- [12] Manski, C.F. (1984), "Adaptive Estimation of Non-linear Regression Models," *Econometric Reviews*, 3, 145-194.
- [13] Ng, Pin T. (1991), "Computing Smoothing Spline Score Estimator," working paper.



- [14] Ng, Pin T. (1994), "Smoothing Spline Score Estimation," *SIAM, Journal of Scientific Computing*, forthcoming.
- [15] Pearson, E.S., D'Agostino, R.B. and Bowman, K.O. (1977), "Tests for Departure from Normality: Comparison of Powers," *Biometrika*, 64, 231-246.
- [16] Portnoy, S. and Koenker, R. (1989), "Adaptive L-Estimation of Linear Models," *The Annals of Statistics*, 17, 362-381.
- [17] Shapiro, S.S. and Wilk, M.B. (1965), "An Analysis of Variance Test for Normality (Complete Samples)," *Biometrika*, 52, 591-611.
- [18] Shapiro, S.S. and Francia, R.S. (1972), "Approximate Analysis of Variance Test for Normality," *Journal of the American Statistical Association*, 67, 215-216.
- [19] Stone, C.J. (1975), "Adaptive Maximum Likelihood Estimators of a Location Parameter," *The Annals of Statistics*, 3, 267-284.







UNIVERSITY OF ILLINOIS-URBANA



3 0112 005702854