faculteit der economische wetenschappen

RESEARCH MEMORANDUM

ILBURG UNIVERSITY
EPARTMENT OF ECONOMICS
ostbus 90153 - 5000 LE Tilburg
etherlands

[redacted]

TESTING THE MEAN OF AN ASYMMETRIC POPULATION:
JOHNSON'S MODIFIED T TEST REVISITED

J.P.C. Kleijnen, G.L.J. Kloppenburg
and F.L. Meeuwsen

August 1985

TESTING THE MEAN OF AN ASYMMETRIC POPULATION: JOHNSON'S MODIFIED T TEST REVISITED

J.P.C. Kleijnen, G.L.J. Kloppenburg
and F.L. Meeuwsen

Catholic University Tilburg
Department of Business and Economics
5000 LE Tilburg, Netherlands

ABSTRACT

Confidence intervals for asymmetric distributions can be based on Student's t statistic or on Johnson's modified t statistic which has two variants, namely a linear and a quadratic approximation. Confidence intervals based on the quadratic approximation, are complicated and are first investigated geometrically, which results in new insight. Next Monte Carlo experiments yield estimates of the coverage and power of several variations of Johnson's test. These experiments show that the quadratic approximation is superior.

1. INTRODUCTION

We investigate several confidence intervals for the mean of an asymmetric distribution. Norman Johnson (1978, p. 537) modified Student's t statistic, explicitly accounting for skewness μ_3 :

$$t_1 = [(\bar{x}-\mu) + \frac{\mu_3}{6\sigma^2 N} + \frac{\mu_3}{3\sigma^4}(\bar{x}-\mu)^2]\ [s^2/N]^{-\frac{1}{2}} \tag{1.1}$$

where all symbols are standard (mean μ, variance σ^2, third central moment μ_3, sample mean $\bar{x}$, unbiased sample variance s^2, sample size N). Johnson (1978, p. 538) further stated "... the effect of the term involving $(\bar{x}-\mu)^2$ in small order ...neglecting the term involving $(\bar{x}-\mu)^2$ .... reduces $[t_1]$ to the variable t_1'...". Our preliminary experiments, however, showed that neglecting $(\bar{x}-\mu)^2$ definitely affects the coverage and power of the test. Those experiments also demonstrated that distribution-free alternatives like the sign and Wilcoxon's signed rank test do not work for an asymmetric distribution (and as the sample size N increases these statistics perform worse). Therefore we shall compare several versions of Johnson's statistic to the classical t statistic. First we shall present analytical results; next we shall discuss Monte Carlo estimates of coverage and power.

2. CONFIDENCE INTERVALS REANALYZED

To analyze confidence intervals based on the quadratic term $(\bar{x}-\mu)^2$, we develop a graphical representation that seems new. For didactic reasons we first discuss Student's t test; see Fig. 1(a). The t statistic with v degrees of freedom satisfies the following equation:

$$1 - \alpha = P\ [t_v^2 < (t_v^{\alpha/2})^2]$$

$$= P\ [-t_{N-1}^{\alpha/2}.s/\sqrt{N} < \bar{x} - \mu < t_{N-1}^{\alpha/2}.s/\sqrt{N}] \tag{2.1}$$

If we define

$$f_1(\mu) = \bar{x} - \mu \tag{2.2}$$

$$c = t_{N-1}^{\alpha/2}.s/\sqrt{N} \tag{2.3}$$

$$\mu_{(1)} = \bar{x} - c \tag{2.4}$$

$$\mu_{(2)} = \bar{x} + c \tag{2.5}$$

then eq. (2.1) implies

$$f_1(\mu) \leq c \wedge f_1(\mu) \geq -c \tag{2.6}$$

or

$$\mu_{(1)} \leq \mu \leq \mu_{(2)} \tag{2.7}$$

which is equivalent to the well-known result

$$\bar{x} - t_{N-1}^{\alpha/2} . s/\sqrt{N} \leq \mu \leq \bar{x} + t_{N-1}^{\alpha/2} . s/\sqrt{N}. \tag{2.8}$$

Now we consider Johnson's linear approximation t_1'. We have pictured a positive μ_3 in Fig. 1(b) so that the confidence interval moves to the right, when compared to Student's statistic; the interval is not centered around the sample mean $\bar{x}$. For negative skewness analogous results hold.

Johnson (1975, p. 538) stated: "Use of the variable t_1 does not lead to a simple expression for confidence intervals for μ since the numerator of t_1 in nonlinear in μ;" he does not elaborate. We analyze possible complications, using the graphical representation of Fig. 1(c). The function $f_3(\mu)$ equals the first factor in eq. (1.1), i.e., $f_3(\mu)$ in a second degree polynomial in μ. The mathematical analysis yields

$$f_3(\mu) < c \rightarrow \mu_{(1)} \leq \mu \leq \mu_{(4)} \tag{2.9}$$

and ($\wedge$)

$$f_3(\mu) > -c \rightarrow \mu \leq \mu_{(2)} \vee \mu \geq \mu_{(3)} \tag{2.10}$$

so that the mathematical solution (to be distinguished from the statistical solution; see below) is

$$\mu_{(1)} \leq \mu \leq \mu_{(2)} \vee \mu_{(3)} \leq \mu \leq \mu_{(4)} \tag{2.11}$$

Mathematical analysis of second-degree polynomials proves that the interval $[\mu_{(1)}, \mu_{(2)}]$ cover the sample mean $\bar{x}$, just as the first-order interval of Fig. 1(b) does. The interval $[\mu_{(3)}, \mu_{(4)}]$ does not overlap the interval of Fig. 1(b). The disjunct interval $[\mu_{(3)}, \mu_{(4)}]$ does not make sense statistically (also see the Monte Carlo results in the next section), so that we eliminate this interval. If the skewness is negative, then again two disjunct intervals result; one interval does not cover $\bar{x}$ and is rejected.

There is an more complication, namely $f_3(\mu)$ may not intersect the lower horizontal line, i.e.,

$$\min_{\mu} [f_3(\mu)] > -c \tag{2.12}$$

We investigate two heuristic solutions; see Fig. 2. The first solution takes $\mu_{(2)}$ equal to μ^*, the value where $f_3(\mu)$ reaches its minimum. The second solution uses Fig. 1(a) and 1(b), where $f(\mu)$ is a straight line with tangent minus one i.e., in Fig. 2 we replace $f_3(\mu)$ by $-\mu + c_2$ where c_2 is a constant and $\mu > \mu^*$. Obviously $\mu_{(2)}$ of the second solution exceeds μ^* of the first solution. Consequently the coverage (probability that μ lies within the confidence interval) is higher for the second solution, and the power function is smaller. The (absolute, not relative) values of the coverages and power functions of the two solutions are unknown. Therefore we shall resort to Monte Carlo experimentation in the next section. But first we note one final complication.

We estimate μ_3 in eq. (1.1) through the unbiased estimator

$$\hat{\mu}_3 = \frac{N}{(N-1)(N-2)} \sum_1^N (x_i - \bar{x})^3 \qquad (2.13)$$

see Kenney and Keeping (1954, p.100). Especially if N is small, $\hat{\mu}_3$ may have the wrong sign. Consequently, if $f_3(\mu)$ is a polynomial with a minimum or "valley", then $\hat{f}_3(\mu)$ has a maximum or "hill"; see Fig. 3. The confidence interval based on $\hat{f}_3(\mu)$ still covers the sample mean (at $\mu = \bar{x}$ the slopes of $f_3(\mu)$ and $\hat{f}_3(\mu)$ equal -1). Also see the Appendix for computational details.

3. DESIGNING THE MONTE CARLO EXPERIMENT

(i) Random number generator: We use the multiplicative congruential generator which is standard on our ICL 2960 computer. This generator was developed by NAG (Numerical Alghorithms Group) in England. It uses the multiplier 13^{13} and the modulus 2^{59}. The seeds are generated randomly by the computer itself, using the internal clock. All Monte Carlo results are independent (different seeds), except for the fact that each sample is analyzed though different statistics (t, t_1, etc.) which yields dependent results.

(ii) Sample size N: Johnson (1978) used N equal to 13 and 25. We pick N equal to 10, 16, 25 and 50 in the coverage study, and N equal to 10 and 25 in the more expensive power study.

(iii) α level: We select α is 0.10, 0.05 and 0.01 in the coverage study, and α is 0.10 and 0.05 in the power study.

(iv) Type of distribution: Many asymmetric distributions could have been selected. We choose the exponential and the lognormal distributions. Given the random numbers r, we sample from the exponential and the lognormal distributions, using standard procedures available on our computer. So we sample from the exponential distribution using the logarithmic transformation

$-(\ell n\ r)/\lambda$ where $\mu = \sigma = 1/\lambda$. And if y has a normal distribution with mean μ_y and variance σ_y^2 then we know that $x = \exp(y)$ has a lognormal distribution with mean

$$\mu_x = \exp(\mu_y + \sigma_y^2/2) \tag{3.1}$$

and variance

$$\sigma_x^2 = [\exp(2\mu_y + \sigma_y^2)][\exp(\sigma_y^2) - 1] \tag{3.2}$$

We sample y from the normal distribution using the standard Box-Muller transformation. Obviously changes in the exponential parameter λ do not affect the results of the various statistics. Therefore we fix λ at the value 1. For the lognormal distribution all combinations of μ_x and σ_x with a fixed ratio μ_x/σ_x yield identical results so that we study only three combinations: $\sigma_x = \mu_x/3$, $\sigma_x = \mu_x$ and $\sigma_x = 3\mu_x$. The exponential might seem to be the skewest distribution since its mode occurs as the extreme left; actually its "excess" μ_3/σ^3 equals 2 whereas the lognormal has excess 1, 4 and 36 for $\sigma_x = \mu_x/3$, $\sigma_x = \mu_x$ and $\sigma_x = 3\mu_x$. Johnson (1978, p. 538) selected the χ_2^2 and χ_{10}^2 distributions. The χ_2^2 is identical to the exponential; obviously the χ_{10}^2 is more symmetric. So we cover more extreme forms of asymmetry.

We also apply the different statistics to the <u>normal</u> distribution. In that situation the assumptions of Student's statistic are satisfied, i.e., the expected coverages should equal the prespecified nominal values $1-\alpha$. We use the normal distribution not only to verify our computer program, but also to examine whether the modified t statistics with or without neglection of $(\bar{x}-\mu)^2$ work when the distribution is actually symmetric (namely Gaussian).

(v) <u>Number of Monte Carlo replications R</u>: The more often we repeat the Monte Carlo experiment, the more accurate our results become. Unfortunately, Monte Carlo experimentation requires much computer time. Obviously the number of replications R needed to estimate the actual α-error within 10% with 90% probability, is

$$R = (1.6449)^2 \ (1-\alpha)/\alpha. \tag{3.3}$$

Hence if α is 0.10, 0.05, 0.01 (see iii) then R is 2435, 5140, 26786 respectively. Such high R values are prohibitive, given our computer budget, so that we use R equal to 2500 to estimate coverage and R equal to 400 to estimate power functions. Fortunately the experimental noise turns out to be small relative to the systematic effects, so that we can detect certain patterns (see next section).

(vi) Statistical procedures: We use the following statistics to derive a confidence interval for the mean:

(1) Student statistic t; see Fig. 1(a).

(2) Johnson's linear statistic t_1'; see Fig. 1(b).

(3) Johnson's quadratic statistic t_1 which yields either two disjunct intervals $[\mu_{(1)},\mu_{(2)}]$ and $[\mu_{(3)},\mu_{(4)}]$ or one long interval $[\mu_{(1)},\mu_{(4)}]$; see Fig. 1(c) and Fig. 2.

(4) Johnson's quadratic statistic t_1 with elimination of the interval that does not cover the sample mean $\bar{x}$ and, if no intersection occurs, with $\mu_{(2)}$ equal to μ^*; see Fig. 2.

(5) Like (4) but $\mu_{(2)}$ follows from the linearization of $f_3(\mu)$; see the line with slope − 1 in Fig. 2.

4. MONTE CARLO RESULTS

We do not bother the reader with the raw data of the Monte Carlo experiment. (These data were made available to the referees, and interested readers may write the authors for these details). Instead we present the information we derived from these data (see Fig. 4 to which we shall return).

4.1 Coverage: Exponential Distribution

We derive confidence intervals, using five different statistical procedures; see Section 3, sub (vi). Student's statistic t and Johnson's linear statistic t_1' give significantly low coverage, for all twelve combinations of α and N except one (N = 50; α = 0.10); we test this significance through the binomial distribution with parameters α and R and significance level 0.10; no normal approximation. For Johnson's quadratic statistic we study three procedures. Procedure (3) of Section 3 has the highest coverage, but it may have two disjunct intervals so that it is statistically unacceptable. Obviously the coverage of procedure (5) exceeds that of procedure (4). Procedure (5) gives significantly low coverage in 3 out of 12 cases; these significant values are not dramatically low (0.979 for α = 0.01; 0.941 for α = 0.05 and 0.889 for α = 0.10) The Monte Carlo experiment also shows how often no intersection of $\hat{f}_3(\mu)$ and $-c$ $(= -t_{N-1}^{\alpha/2}\, s/\sqrt{N})$ occurs. Obviously intersection occurs more often, as N and α increase. "No intersection" occurs with an estimated probability of 70% if α is 0.01 and N is 10, and 3% if α is 0.10 and N is 50. The estimate $\hat{\mu}_3$ has the wrong sign, with estimated probability of 5% if N is 10 and 0% if N is 50 (obviously α has no effect).

4.2 Coverage: Lognormal Distributions

For μ is 1 and σ is 1/3 the lognormal has smaller excess than the exponential has. Consequently the coverage of Student's statistic improves, especially as N and α increase. Johnson's linear statistic hardly improves the coverage. Johnson's quadratic statistic with linear approximation if needed (procedure 5) gives the highest coverage; only 3 out of 12 situations show significantly low coverage.

For μ is 1 and σ is 1 the lognormal has excess 4. Student's and Johnson's linear statistics give significantly low coverages in all twelve situations. The quadratic statistic always improves the coverage. Sometimes (3 out of 12) the coverage is not too low; never the coverage is dramatically low; for example, estimated coverage is 0.969 instead of 0.99 (= $1-\alpha$) if N is 10.

For μ is 1 and σ is 3 the excess is extreme, namely 36. Student's and Johnson's linear statistics (t and t_1) perform very poorly. The quadratic statistics do relatively better; their coverages are significantly low, for example, if α is 0.01 and N is 10 then estimated coverage is 0.88 and if α becomes 0.10 then coverage becomes 0.77 (t and t_1 give 0.66 and 0.67).

4.3 Power

We estimate the power function at 21 values of μ, using 400 replications per value (and 2500 replications at $\mu = \mu_0 = 1$). We investigate 13 combinations of different distributions, N values and α values. We display only three representative examples in Fig. 4. The estimated power functions of Student's and Johnson's linear statistics are so close that they cannot be distinguished in the resulting pictures. The quadratic statistic with two disjunct intervals shows curious behavior, i.e., the power shows a dip in a certain area left of μ_0. We have already explained that two disjunct intervals are statistically unacceptable; the dip of the power function emphasizes the statistical misbehavior of this procedure. The quadratic statistic with linearization (procedure 5 of Section 3) has an estimated power function the reaches its minimum at μ equal to μ_0 whereas the linear statistics reach their minimum when μ exceeds μ_0. The values of these minima are roughly equal. An extremely skew lognormal distribution ($\sigma = 3\mu$) gives an estimated power function with a shape that differs from the other distributions. If the distributions is actually normal, then all five statistics give nearly identical power functions. So as Johnson (1978, p. 539) has already noted, the modified t test also works if the distribution happens to be normal.

5. CONCLUSION

Although many studies claim that Student's statistic t is robust, Johnson (1978) proposed a modified t statistic. We find that his statistic t_1', which neglects $(\bar{x}-\mu)^2$, does not improve the t statistic. The statistic t_1, which includes the $(\bar{x}-\mu)^2$ term, requires the solution of second-degree polynomials. This t_1 gives excellent results in the exponential case while the linear statistics (t and t_1') then fail. In the lognormal cases the quadratic statistic t_1 does not give perfect results; nevertheless its coverage is quite close to the prespecified nominal value $1-\alpha$, and its results are better than the linear statistics (t and t_1'). If the distribution is actually normal, then all statistics give the desired coverage. So it is good practice to modify the classical t statistic, as proposed by Johnson (1978), provided we do include the $(\bar{x}-\mu)^2$ term; the price we pay is a slight increase in computation which, however, is negligible when using a (micro) computer.

APPENDIX. COMPUTATIONAL DETAILS OF CONFIDENCE INTERVAL USING t_1

In the main text we have already defined the following symbols: $\bar{x}$, s^2, $\hat{\mu}_3$, c. We now define $c_1 = \hat{\mu}_3 /(6\ s^2\ N)$, $c_2 = \hat{\mu}_3 / (3\ s^4)$, $d_1 = 1 - 4\ c_2\ (c_1 + c)$, $d_2 = 1 - 4\ c_2\ (c_1 - c)$. Next we introduce $K = \bar{x} + 1/(2\ c_2) - (d_1)^{\frac{1}{2}} /(2\ c_2)$, $L = \bar{x} + 1/(2\ c_2) - (d_2)^{\frac{1}{2}} /(2\ c_2)$, $M = \bar{x} + 1/(4\ c_2) + c_1 + c$, $N = \bar{x} + 1/(4\ c_2) + c_1 - c$. If $\hat{\mu}_3 > 0$ and $d_1 \geq 0$, $d_2 \geq 0$ then procedure (5) yields the confidence interval [L,K]; if $d_1 < 0$ and $d_2 \geq 0$ then [L,M]. If $\hat{\mu}_3 < 0$ and $d_1 \geq 0$, $d_2 \geq 0$ then the interval is [L,K]; if $d_1 \geq 0$ and $d_2 < 0$ then [N,K].

ACKNOWLEDGEMENT

The authors wish to thank H. Moors, statistician in the Department of Econometrics, for bringing Johnson's article to their attention, C. Cuypers and J. van Rooijen, graduate students, for preliminary research on Johnson's test, and P. Nederstigt, Dutch State Mines, for suggesting the elimination of one of the two disjunct intervals.

REFERENCES

JOHNSON, N.J. (1978), "Modified t Tests and Confidence Intervals for Asymmetric Populations", Journal of the American Statistical Association, 73, 536 - 544.

KENNEY, J.F. AND KEEPING, E.S. (1954), Mathematics of Statistics, Part One, New York, D. Van Nostrand.

Key Words and Phrases: t variable; modified t statistic; skewness; power; geometric analysis.

FIG. 1(a). Student statistic t.

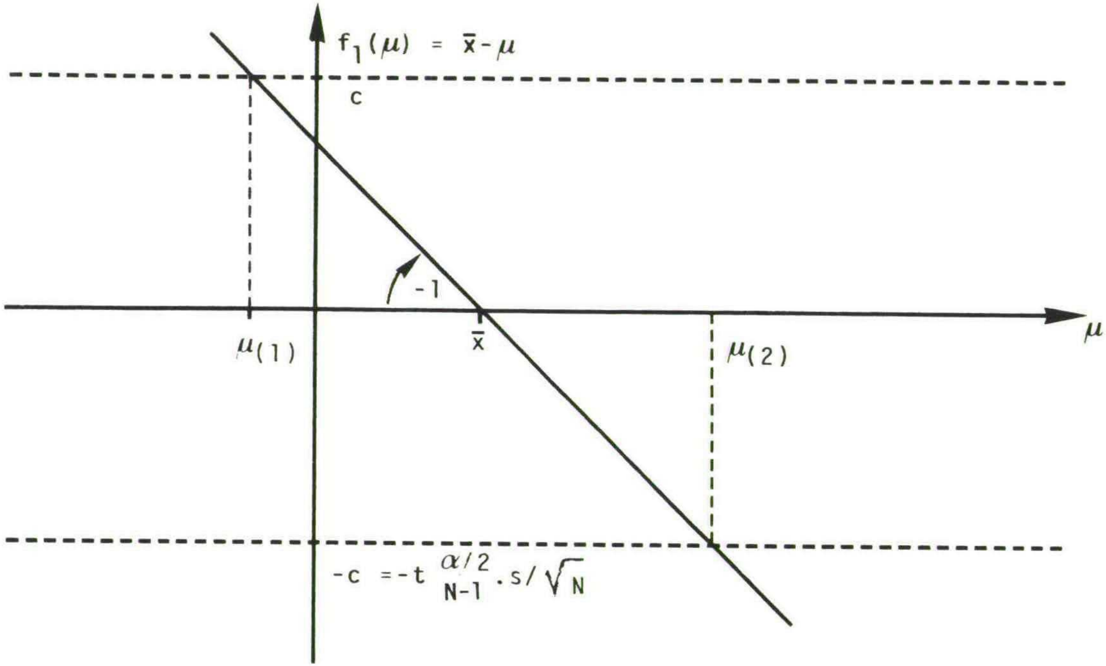


FIG.1(b). Johnson's first-order modification t_1'

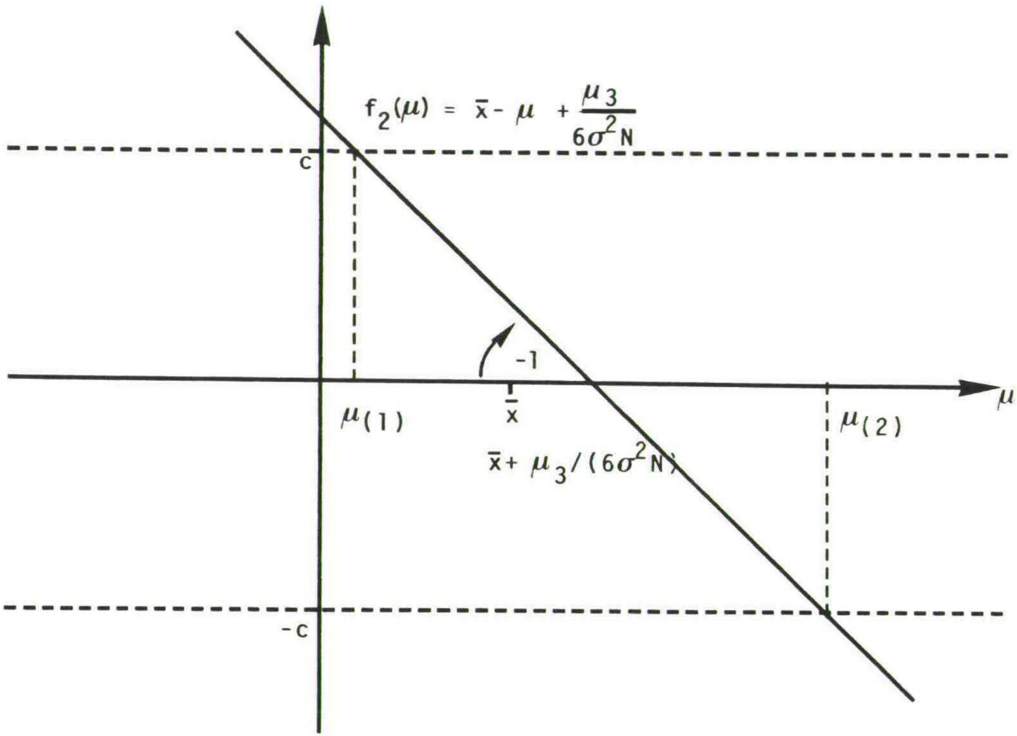


FIG. 1(c). Johnson's second-order approximation t_1.

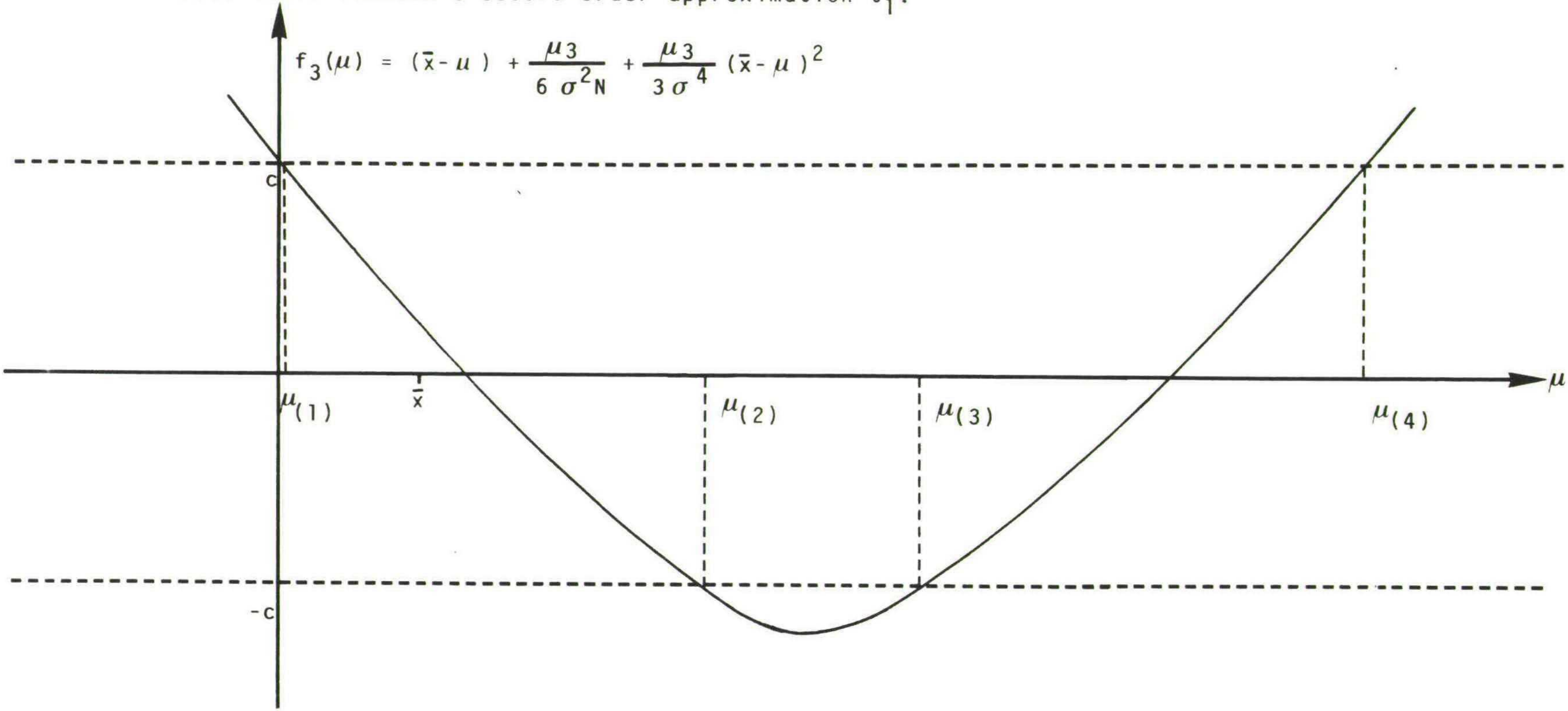


FIG. 2. No intersection with -c.

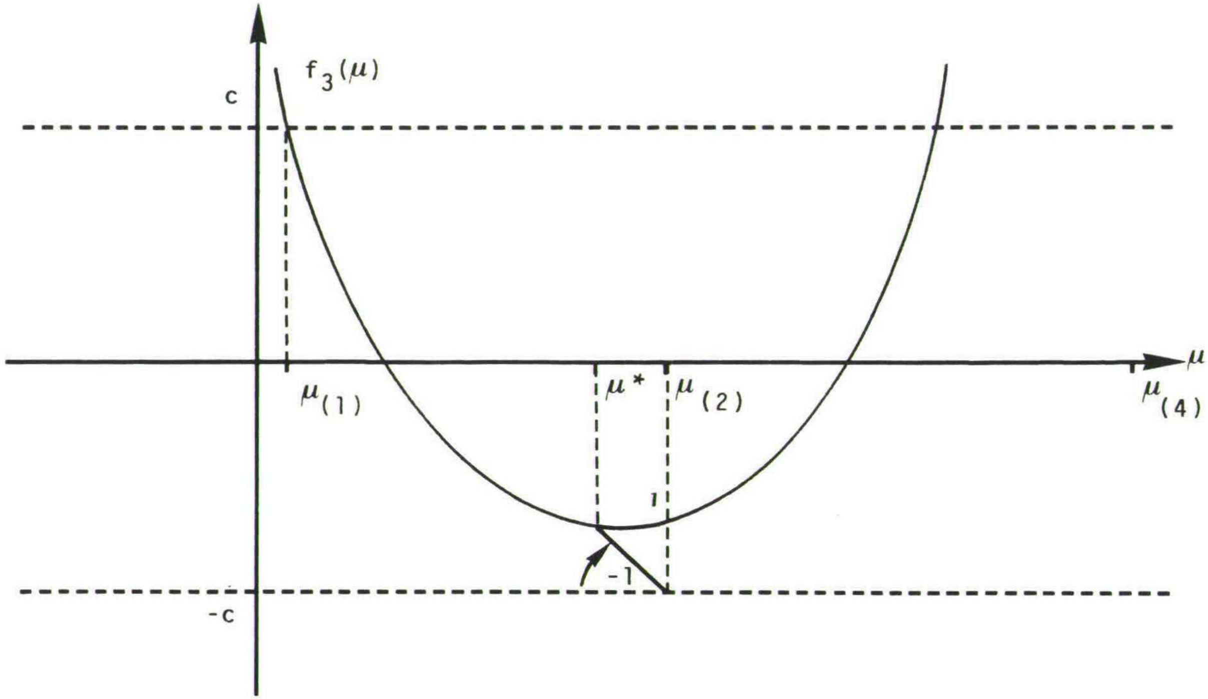


FIG. 3. Wrong sign of μ_3

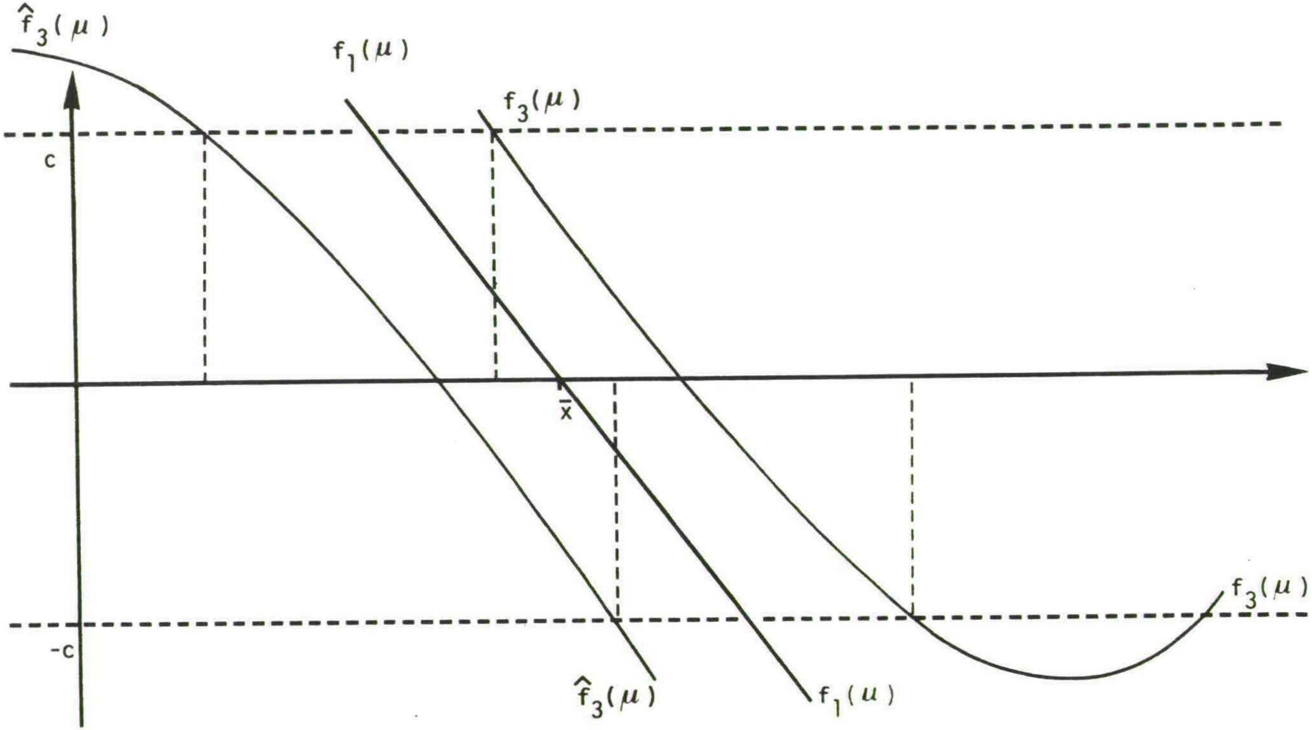
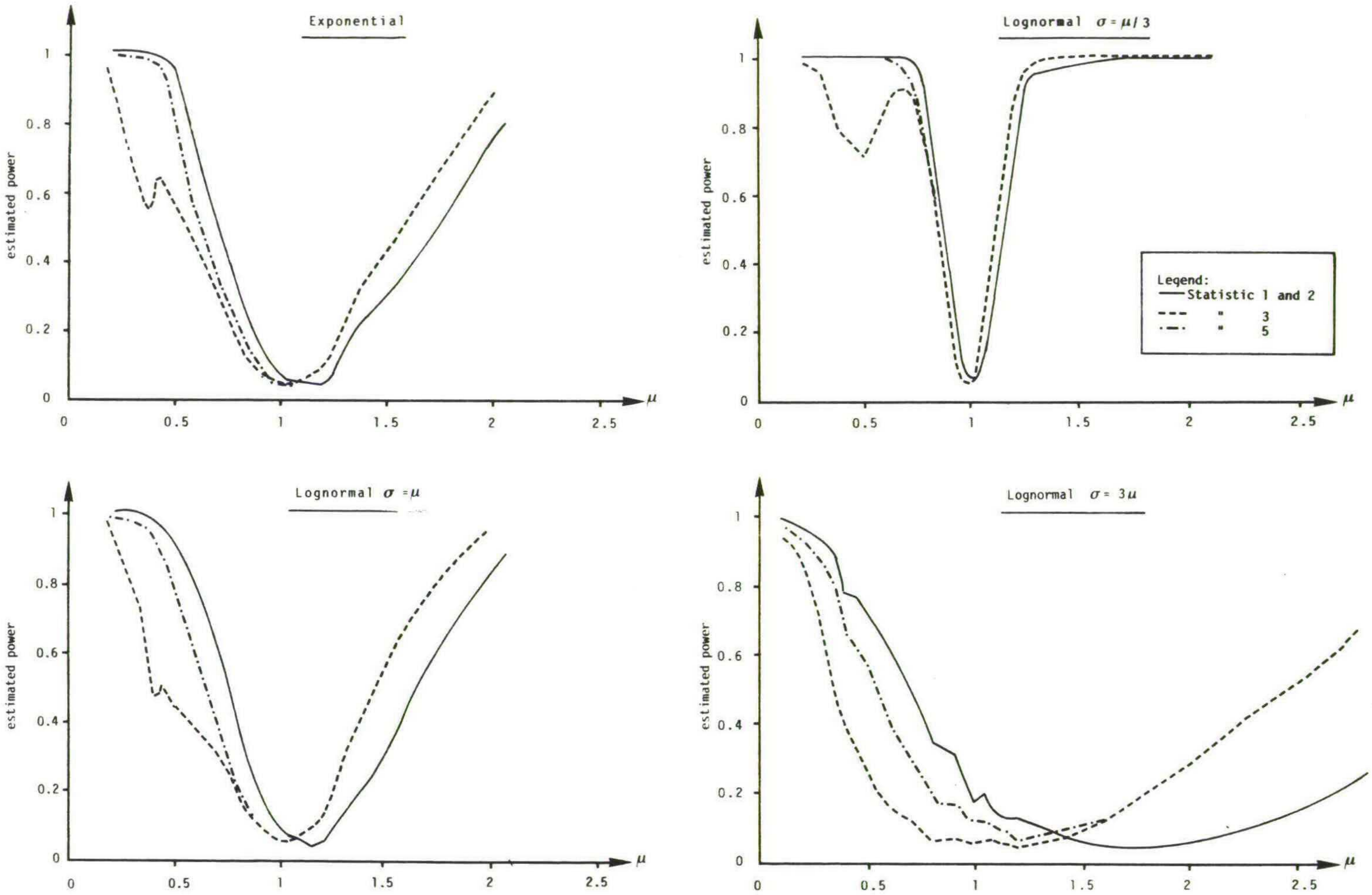


FIG. 4. Estimated power functions (N = 25; α = 0.05)



IN 1984 REEDS VERSCHENEN

138 G.J. Cuypers, J.P.C. Kleijnen en J.W.M. van Rooyen
Testing the Mean of an Asymetric Population:
Four Procedures Evaluated

139 T. Wansbeek en A. Kapteyn
Estimation in a linear model with serially correlated errors when observations are missing

140 A. Kapteyn, S. van de Geer, H. van de Stadt, T. Wansbeek
Interdependent preferences: an econometric analysis

141 W.J.H. van Groenendaal
Discrete and continuous univariate modelling

142 J.P.C. Kleijnen, P. Cremers, F. van Belle
The power of weighted and ordinary least squares with estimated unequal variances in experimental design

143 J.P.C. Kleijnen
Superefficient estimation of power functions in simulation experiments

144 P.A. Bekker, D.S.G. Pollock
Identification of linear stochastic models with covariance restrictions.

145 Max D. Merbis, Aart J. de Zeeuw
From structural form to state-space form

146 T.M. Doup and A.J.J. Talman
A new variable dimension simplicial algorithm to find equilibria on the product space of unit simplices.

147 G. van der Laan, A.J.J. Talman and L. Van der Heyden
Variable dimension algorithms for unproper labellings.

148 G.J.C.Th. van Schijndel
Dynamic firm behaviour and financial leverage clienteles

149 M. Plattel, J. Peil
The ethico-political and theoretical reconstruction of contemporary economic doctrines

150 F.J.A.M. Hoes, C.W. Vroom
Japanese Business Policy: The Cash Flow Triangle
an exercise in sociological demystification

151 T.M. Doup, G. van der Laan and A.J.J. Talman
The $(2^{n+1}-2)$-ray algorithm: a new simplicial algorithm to compute economic equilibria

IN 1984 REEDS VERSCHENEN (vervolg)

152 A.L. Hempenius, P.G.H. Mulder
Total Mortality Analysis of the Rotterdam Sample of the Kaunas-Rotterdam Intervention Study (KRIS)

153 A. Kapteyn, P. Kooreman
A disaggregated analysis of the allocation of time within the household.

154 T. Wansbeek, A. Kapteyn
Statistically and Computationally Efficient Estimation of the Gravity Model.

155 P.F.P.M. Nederstigt
Over de kosten per ziekenhuisopname en levensduurmodellen

156 B.R. Meijboom
An input-output like corporate model including multiple technologies and make-or-buy decisions

157 P. Kooreman, A. Kapteyn
Estimation of Rationed and Unrationed Household Labor Supply Functions Using Flexible Functional Forms

158 R. Heuts, J. van Lieshout
An implementation of an inventory model with stochastic lead time

159 P.A. Bekker
Comment on: Identification in the Linear Errors in Variables Model

160 P. Meys
Functies en vormen van de burgerlijke staat
Over parlementarisme, corporatisme en autoritair etatisme

161 J.P.C. Kleijnen, H.M.M.T. Denis, R.M.G. Kerckhoffs
Efficient estimation of power functions

162 H.L. Theuns
The emergence of research on third world tourism: 1945 to 1970;
An introductory essay cum bibliography

163 F. Boekema, L. Verhoef
De "Grijze" sector zwart op wit
Werklozenprojecten en ondersteunende instanties in Nederland in kaart gebracht

164 G. van der Laan, A.J.J. Talman, L. Van der Heyden
Shortest paths for simplicial algorithms

165 J.H.F. Schilderinck
Interregional structure of the European Community
Part II: Interregional input-output tables of the European Community 1959, 1965, 1970 and 1975.

IN (1984) REEDS VERSCHENEN (vervolg)

166 P.J.F.G. Meulendijks
An exercise in welfare economics (I)

167 L. Elsner, M.H.C. Paardekooper
On measures of nonnormality of matrices.

IN 1985 REEDS VERSCHENEN

168 T.M. Doup, A.J.J. Talman
A continuous deformation algorithm on the product space of unit simplices

169 P.A. Bekker
A note on the identification of restricted factor loading matrices

170 J.H.M. Donders, A.M. van Nunen
Economische politiek in een twee-sectoren-model

171 L.H.M. Bosch, W.A.M. de Lange
Shift work in health care

172 B.B. van der Genugten
Asymptotic Normality of Least Squares Estimators in Autoregressive Linear Regression Models

173 R.J. de Groof
Geïsoleerde versus gecoördineerde economische politiek in een twee-regiomodel

174 G. van der Laan, A.J.J. Talman
Adjustment processes for finding economic equilibria

175 B.R. Meijboom
Horizontal mixed decomposition

176 F. van der Ploeg, A.J. de Zeeuw
Non-cooperative strategies for dynamic policy games and the problem of time inconsistency: a comment

177 B.R. Meijboom
A two-level planning procedure with respect to make-or-buy decisions, including cost allocations

178 N.J. de Beer
Voorspelprestaties van het Centraal Planbureau in de periode 1953 t/m 1980

178a N.J. de Beer
BIJLAGEN bij Voorspelprestaties van het Centraal Planbureau in de periode 1953 t/m 1980

179 R.J.M. Alessie, A. Kapteyn, W.H.J. de Freytas
De invloed van demografische factoren en inkomen op consumptieve uitgaven

180 P. Kooreman, A. Kapteyn
Estimation of a game theoretic model of household labor supply

181 A.J. de Zeeuw, A.C. Meijdam
On Expectations, Information and Dynamic Game Equilibria

182 Cristina Pennavaja
Periodization approaches of capitalist development.
A critical survey