Negative binomial estimation and testing : comparison to minimum disparity methods
by Wendy Lee Swanson

a thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Statistics
Montana State University

Abstract:
Various methods have been proposed for comparing the means of independent samples from two negative binomial distributions, but no method is recognized as the standard. The t-test, after log-transforming the data, is often used. But the Mest is unreliable, especially for small means (i.e., one of the means $\mu < 5$). In this dissertation, a new test procedure, called the Disparity Difference Test (DDT) is derived and compared to existing methods. The new method is based on an idea of Lindsay (1994 Annals of Statistics) who introduced a general approach for estimation and testing based on the Negative Exponential Disparity (NED) measure. The DDT is compared to the t-test, the generalized likelihood ratio test, and some generalized score tests. Because all the tests, except the t-test, are asymptotically equivalent, the comparison is based on a simulation study that used small means and realistic sample sizes.

Estimation is embedded in the significance testing methodology because each method requires an estimate of the common negative binomial variance parameter, as well as estimates of the means. A derivation of the NED estimator is provided. The statistical properties of the NED estimator of the variance parameter is compared to the maximum likelihood estimator and to some robust estimators, including the extended quasi-likelihood estimator, the pseudolikelihood estimator, and a conditional maximum likelihood estimator. The comparisons are based on simulation studies.

The results are that the NED estimator performs well, and the DDT not so well, compared to the other methods. There are no practical differences among the empirical average errors for the various estimators. The DDT has smaller power than the likelihood ratio and scores tests for a majority of the parameter settings. There are no practical differences among the score and likelihood ratio tests. Recommendations are provided.

NEGATIVE BINOMIAL ESTIMATION AND TESTING: COMPARISON TO MINIMUM DISPARITY METHODS

by

Wendy Lee Swanson

a thesis submitted in partial fulfillment
of the requirements for the degree

of

Doctor of Philosophy

in

Statistics

MONTANA STATE UNIVERSITY-BOZEMAN
Bozeman, Montana

May 1997

APPROVAL

of a thesis submitted by

Wendy Lee Swanson

This thesis has been read by each member of the thesis committee and has been found to be satisfactory regarding content, English usage, format, citations, bibliographic style, and consistency, and is ready for submission to the College of Graduate Studies.

Martin A. Hamilton — Martin A. Hamilton (Signature) — 5/19/97 Date

Approved for the Department of Mathematical Sciences

John R. Lund — John R. Lund (Signature) — 5/19/97 Date

Approved for the College of Graduate Studies

Robert L. Brown — R L Brown (Signature) — 6/3/97 Date

STATEMENT OF PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a doctoral degree at Montana State University-Bozeman, I agree that the Library shall make it available to borrowers under rules of the Library. I further agree that copying of this thesis is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for extensive copying or reproduction of this thesis should be referred to University Microfilms International, 300 North Zeeb Road, Ann Arbor, Michigan 48106, to whom I have granted "the exclusive right to reproduce and distribute my dissertation in and from microform along with the non-exclusive right to reproduce and distribute my abstract in any format in whole or in part."

Signature Wendy Lee Swanson

Date May 8, 1997

ACKNOWLEDGMENT

Partial support was provided by Cooperative Agreement No. CR818324 between the USEPA and Montana State University and by Cooperative Agreement No. EEC-8907039 between the National Science Foundation and Montana State University. This dissertation has not been subjected to peer or administrative review by the USEPA and therefore may not necessarily reflect the views of the Agency, and no official endorsement should be inferred.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

ABSTRACT

Various methods have been proposed for comparing the means of independent samples from two negative binomial distributions, but no method is recognized as the standard. The t-test, after log-transforming the data, is often used. But the t-test is unreliable, especially for small means (i.e., one of the means $\mu \leq 5$). In this dissertation, a new test procedure, called the Disparity Difference Test (DDT) is derived and compared to existing methods. The new method is based on an idea of Lindsay (1994 *Annals of Statistics*) who introduced a general approach for estimation and testing based on the Negative Exponential Disparity (NED) measure. The DDT is compared to the t-test, the generalized likelihood ratio test, and some generalized score tests. Because all the tests, except the t-test, are asymptotically equivalent, the comparison is based on a simulation study that used small means and realistic sample sizes.

Estimation is embedded in the significance testing methodology because each method requires an estimate of the common negative binomial variance parameter, as well as estimates of the means. A derivation of the NED estimator is provided. The statistical properties of the NED estimator of the variance parameter is compared to the maximum likelihood estimator and to some robust estimators, including the extended quasi-likelihood estimator, the pseudolikelihood estimator, and a conditional maximum likelihood estimator. The comparisons are based on simulation studies.

The results are that the NED estimator performs well, and the DDT not so well, compared to the other methods. There are no practical differences among the empirical average errors for the various estimators. The DDT has smaller power than the likelihood ratio and scores tests for a majority of the parameter settings. There are no practical differences among the score and likelihood ratio tests. Recommendations are provided.

CHAPTER 1

INTRODUCTION

Background

The negative binomial distribution has provided a representation for count data in many areas of research. As noted by Bliss and Fisher (1953), the earliest fit (empirical) of count data to a negative binomial was applied to microscopic counts of yeast cells by Student (1907). Some of the early uses of the negative binomial model in statistical analyses were on counts of insects (Anscombe, 1949), microbes (Jones, Mollison and Quenouille, 1948) and accidents (Greenwood and Yule, 1920). In more recent years, negative binomial analysis of count data has been applied by researchers in a variety of disciplines. It has been applied to market research (Chatfield, 1975), purchasing (Schmittlein et al., 1985; Ramaswamy et al,1994) and reliability (Bain and Wright, 1982). Jones et al (1991) analyzed negative binomial models for counts of parasites on a host species; Hubbard and Allan (1991) used a sequential negative binomial analysis for an insect pest management strategy, and Morton (1987) used a negative binomial model to analyze insect trap catches in a nested block design. Gold et al (1996) compare sampling protocols for weeds clustered in fields in a spatial distribution that can be described by the negative binomial. Manton et. al. (1981) applied hierarchical negative binomial models to an epidemiological study of lung cancer mortality rates; Maul, El-Shaarawi and Ferard (1991) analyzed a chronic toxicity response using a negative binomial model.

In ecology, the NB is used in the analysis of a pollution impact study on fish abundance (Ramakrishnan and Meeter, 1993). Counts of species richness-area relationship were analyzed using competing models based on the negative binomial distribution (Stein, 1988). Seber (1973) applied negative binomial models arising from several different sampling strategies to the estimation of animal abundance.

Barnwal and Paul (1988) derived two C(α) statistics (Neyman, 1959) for testing equality of means in a one-way layout for negative binomial data and applied them to field and laboratory research counts. Collings and Margolin (1985) compared tests for departure from the Poisson assumption using a negative binomial alternative for data in a one-way layout.

Microbiological applications included analyses of the counts of revertant colonies in the Ames salmonella microsome assay (Margolin et al, 1981; Breslow, 1984; Krewski et al, 1993), spatial and temporal variation of bacterial counts (Maul and El-Shaarawi, 1991) and the estimation of coliform density in drinking water (Pipes et al. 1977).

There are a large number of chance mechanisms which give rise to the negative binomial and have plausible physical applications. Some of them are described below with an outline of their derivation.

A common representation of the negative binomial random variable, Y, uses parameters p, $(0 < p < 1)$ and $k > 0$, denoted NB(p,k) and having discrete density or probability mass function (pmf) $P(Y = y) = \binom{y+k-1}{y} p^k (1-p)^y$ $y = 0,1,\ldots$. This representation has associated moment generating function

(mgf) $M_Y(t) = E\left(e^{tY}\right) = \left(\frac{p}{1-(1-p)e^t}\right)^k$ and probability generating function (pgf) $G_Y(s) = E\left(s^Y\right) = \left(\frac{p}{1-(1-p)s}\right)^k$.

In the above representation k is not limited to the integers. When k is limited to the integers, the distribution is sometimes called the Pascal distribution [Pascal (1679)]. In this spirit, Ross (1989) motivated the NB distribution as "a coin having probability p of coming up heads (being) successively flipped until the kth head appears". This example of the NB derivation as the distribution of the number of tosses of a coin required to achieve a fixed number of successes is due to Montmort (1714). The presentation of NB for integer k is common in statistics textbooks. A geometric random variable is the special case of a negative binomial random variable for $k=1$. One way to generate the NB with integer k is as the sum of n independent and identically distributed (iid) geometric random variables with parameter $p, (0<p<1)$ and mgf $M_{X_i}(t) = \frac{p}{1-qe^t}$. It follows from independence of the X_i's that the mgf of their sum, $Y = X_1 + \cdots + X_n$, is the product of their mgfs, i.e.: $M_Y(t) = M_{X_1+\ldots+X_n}(t) = \prod_{i=1}^{n} M_{X_i}(t) = \left(\frac{p}{1-qe^t}\right)^n$ which is the mgf of the $NB(p, k=n)$. This model was used by Seber (1973, p.174) for recapture data from multiple traps (acting independently) within an animal's home range.

The negative binomial can be derived as a Poisson-stopped sum of logarithmic random variables (Luders, 1934; see also Quenouille, 1949). Let $Y = X_1 + \cdots + X_N$, where the X_i's are iid logarithmic random variables with parameter $\eta > 0$, and have pmf, $P(X=x) = \frac{-\eta^x}{x\ln(1-\eta)}$, $x = 1,2,\ldots$, and probability generating function (pgf), $G_{X_i}(s) = \ln(1-\eta s)/\ln(1-\eta)$. Let N be independently

distributed Poisson with parameter $\lambda > 0$ and pgf $G_N(s) = \exp(\lambda(s-1))$. Then the pgf of the random sum $Y = X_1 + \cdots + X_N$ is $G_Y(s) = G_N\left(G_{X_i}(s)\right)$

$$= \exp\left(\lambda\left(\frac{\ln(1-\eta s)}{\ln(1-\eta)} - 1\right)\right) = \exp\left(\ln\left(\frac{1-\eta s}{1-\eta}\right)^{\lambda/\ln(1-\lambda)}\right) = \left[(1-\eta)/(1-\eta s)\right]^{-\lambda/\ln(1-\lambda)}$$ which is

the NB pgf with parameters $p = 1-\eta$ and $k = -\lambda/\ln(1-\lambda)$. Quenouille's derivation of this model stemmed from research counts of soil microbes (Jones, Mollison and Quenouille, 1948) in which the "colony counts followed a Poisson's distribution, the numbers of bacteria per colony were logarithmically distributed, and that, consequently, the bacterial counts were distributed in the negative binomial form."

Another derivation of the negative binomial (due to Greenwood and Yule, 1920) is as a gamma mixture of Poisson random variables. Suppose X, given λ, is distributed Poisson(λ) with (conditional) pmf $P(X = x|\lambda) = e^{-\lambda}\lambda^x/x!$, for $\lambda > 0, x = 0,1,2,\ldots$; and that independently, λ is distributed as gamma(ν,τ) with probability density function (pdf) $f(\lambda;\nu,\tau) = \left(\lambda^{\nu-1}e^{-\lambda/\tau}\right)/\left(\Gamma(\nu)\tau^{\nu}\right)$ for $\lambda,\nu,\tau > 0$. For the gamma distribution ν is referred to as the index or *shape parameter* and τ is the *scale parameter*. Then unconditionally, X is distributed as NB with parameters $p = 1/(\tau+1)$ and $k = \nu$. That is,

$$P(X = x) = \int_0^{\infty} P(X = x|\lambda) f(\lambda;\nu,\tau)\,d\lambda = \left\{\Gamma(\nu)\tau^{\nu}x!\right\}^{-1} \times \int_0^{\infty} e^{-\lambda}\lambda^x\lambda^{\nu-1}e^{-\lambda/\tau}\,d\lambda$$

$$= \frac{\Gamma(\nu+x)\left(\tau/(\tau+1)\right)^{\nu+x}}{\Gamma(\nu)\tau^{\nu}x!} \times \int_0^{\infty} \frac{e^{-\lambda/(\tau/(\tau+1))}\lambda^{(\nu+x)-1}\,d\lambda}{\Gamma(\nu+x)\left(\tau/(\tau+1)\right)^{\nu+x}} = \frac{\Gamma(\nu+x)}{\Gamma(\nu)x!}\left(\frac{\tau}{\tau+1}\right)^x\left(\frac{1}{\tau+1}\right)^{\nu} \times 1,$$ the NB

pmf with mean, $E(X) = \nu\tau = \mu$ and variance $Var(X) = \nu\tau + \nu\tau^2 = \mu + \frac{\mu^2}{\nu} = \mu(1+\tau)$.

Greenwood and Yule (1920; see also Arbous and Kerrich, 1951) used this derivation as a model for accident statistics in which individuals differed in accident-proneness.

Boswell and Patil (1970) give a list of processes which produce the NB distribution. Included are population growth models and other stochastic processes and their derivations. Johnson, Kotz, and Kemp (1992, chapter 5) provide an excellent review of the literature on the negative binomial distribution and a large list of references. They listed many parameterizations for the negative binomial. The forms that I find most useful are denoted NB(μ,k) [Anscombe (1950)] and NB(μ,a) [Bliss and Owen (1958)]. The Anscombe parameterization pmf is $\Pr(Y=y;\mu,k)=\frac{\Gamma(y+k)}{y!\Gamma(k)}\left(\frac{\mu}{\mu+k}\right)^{y}\left(\frac{k}{\mu+k}\right)^{k}$ $y=0,1,\ldots.$ Bliss and Owen used $a=k^{-1}$ and $\Pr(Y=y;\mu,a)=\frac{\Gamma\left(y+a^{-1}\right)}{y!\,\Gamma\left(a^{-1}\right)}\left(\frac{a\mu}{1+a\mu}\right)^{y}\left(\frac{1}{1+a\mu}\right)^{a^{-1}}$ $y=0,1,\ldots.$ Both these forms parameterize using the mean, $E(Y)=\mu$ and an additional parameter in the variance, $\mathrm{var}(Y)=V=\mu+\frac{\mu^2}{k}=\mu+a\mu^2$. Recent literature has made use of the NB(μ,a) because it yields the Poisson distribution as the limiting case when a goes to 0. [Lawless (1987), Piegorsch (1990)]. Clark and Perry (1989) suggest the use of a because it "eliminates the problems of infinite values of $\hat{k}=\frac{\bar{y}^2}{S^2-\bar{y}}$ (method-of-moments estimator) when $S^2=\bar{y}$ (where S^2 is the sample variance); also confidence intervals for a are continuous and usually more symmetric than those for k, which may be discontinuous."

The negative binomial is often used for empirical reasons. It has provided an adequate fit to count data. Martin and Katti (1965) fit various "contagious" distributions (distributions produced as mixtures of other distributions) to 35 data sets and found the negative binomial and Neyman Type A to have wide applicability. Bliss and Fisher (1953) used the NB to fit numerous biological data sets. Evans (1953) found the NB provided the only

satisfactory fit to counts of insect populations he examined. Christian and Pipes (1983) found the NB distribution compatible to the coliform frequencies in nine drinking water systems. In each case, the data could be considered as "overdispersed" relative to a Poisson model (variability larger than expected, or variance larger than the mean). In a Poisson model, the distribution of numbers of individuals per unit space or time is random, the presence of an individual not influencing the probability of the presence of another. Bliss (Bliss and Fisher, 1953) noted that "when the numbers of individuals per unit space or time cannot be assumed to have the same expected value, they may represent a mixture of several homogeneous Poisson distributions" in which the means are distributed as a positive continuous variate. This produces heterogeneity or density-dependence rather than the random spatial distribution associated with the Poisson. Animals that occur in herds display this tendency. So do microbes and insects which tend to aggregate or clump. Hubbard and Allen (1991) reasoned that insect distributions are heterogeneous because they "migrate to hospitable environments, or if the insects migrate little, (because) their eggs are laid in clumps." Taylor, Woiwod, and Perry (1978) reported on the rarity of randomness of spatial behavior in nature. Among the possible distributions that can fit heterogeneity, the negative binomial provides the simplest Poisson mixture. And because the gamma mixing variable can take on a wide variety of shapes, the resulting NB distributions fit many data sets.

Some analysts worked with data that fit only the form of the mean and variance of the NB, i.e., $E(Y)=\mu$ and $\mathrm{var}(Y)=\mu+a\mu^2$. Minkin (1991) arrives at this form using a random effects Poisson model. A multiplicative random effect was introduced to accommodate the extra-Poisson variability in data sets involving medical cancer clonogenic assays and the analysis of dose-response

curves. Chase and Hoel (1975) examine the mechanism of serial dilutions used to estimate particle concentration and show how small measurement errors of individual dilutions can produce a mean and variance with the same form as seen in the NB.

I became interested in plate counts of microbes while working in the Center for Biofilm Engineering, at Montana State University, on EPA-sponsored research concerning the efficacy of disinfectants. The counts appeared overdispersed relative to a Poisson model and the results of interest were tests of control versus treatment. My literature search lead me to believe that the NB distribution represented a reasonable model for this type of data for both empirical and methodological reasons. I wanted to compare the power of tests for data of NB form framed in a generalized linear model (GLM) setting with the specific purpose of testing control versus treatment. I was attracted to the characterization of the NB as a gamma mixture of Poissons. I followed the setting of the mixed Poisson (NB) regression model as presented in Lawless (1987). I was interested in the estimation of the variance parameter because it is an open, important problem. The two-parameter NB is not exponential family, the minimal sufficient statistic (order statistic) is not complete, and plots of the log-likelihood function illuminate the difficulty of the estimation problem for the variance parameter (Willson, Folks and Young (1986)). Many robust estimators have been proposed in recent literature. I wanted to compare the newest estimation techniques to the some of the most frequently used older techniques. My focus was two-fold: estimation of the variance parameter and testing for a difference between the means of two populations.

Notation and Terminology

Matrices shall be denoted by upper case letters (X), elements or scalars (in italics, x_{ij} and β_2) and vectors (not italic, bold for non-greek font, $\mathbf{x}_i$ and β) in lower case letters.

The extension of the NB(μ,a) to a log-linear regression model, or (generalized) linear model (GLM), denoted NB$(\mu(\mathbf{x}),a)$ is given in Lawless (1987). Given that the Y_i's are n independent negative binomial random variables with a common variance parameter, a, and the $\mathbf{x}_i$'s are px1 explanatory variables, the probability that Y_i equals y_i given $\mathbf{x}_i$ is

$$\Pr(Y_i = y_i|\mathbf{x}_i) = \frac{\Gamma(y_i + a^{-1})}{y_i!\,\Gamma(a^{-1})}\left(\frac{a\mu_i}{1+a\mu_i}\right)^{y_i}\left(\frac{1}{1+a\mu_i}\right)^{a^{-1}}, \; y_i = 0,1,\ldots \text{ where}$$

$E(Y_i|\mathbf{x}_i) = \mu(\mathbf{x}_i) = e^{\mathbf{x}_i^t\beta}$, abbreviated μ_i, and $Var(Y_i) = \mu_i + a\mu_i^2$, abbreviated V_i. The link function, or link between the mean and the explanatory variables is $\log(\mu) = \mathbf{x}^t\beta$, hence the term log-linear model. I will refer to a as the *variance parameter*. Note that the variance parameter does not depend on i. Other authors refer to a (or to k=a^{-1}) as an index parameter, shape parameter or dispersion parameter. [I do not refer to a (or to k) as a dispersion parameter because it does not qualify as such in the way "dispersion parameter" is defined and used in some of the methods described below.]

The simplest way to adapt this generalized linear model setting to a negative binomial model for testing treatment versus control is to use a design matrix consisting of columns for an intercept and a control versus treatment contrast. Then each $\mathbf{x}_i^t$ is a row of a design matrix, X, and $\mathbf{x}_i$ is the transpose of the ith row. Just as in linear models for the one-way ANOVA setting, if we have n_1 observations from the control group (population one), and n_2

observations from the treatment group (population two), and $n=n_1+n_2$ then the design matrix consists of a column of 1's of length n and a column of 1's and -1's of lengths n_1 and n_2 , i.e.: $X=\begin{bmatrix}X_1\\X_2\end{bmatrix}\begin{matrix}n_1\times 2\\n_2\times 2\end{matrix}=\begin{bmatrix}1 & 1\\1 & -1\end{bmatrix}\begin{matrix}n_1\times 2\\n_2\times 2\end{matrix}$. The means for control and treatment are $\mu_1=e^{\beta_1+\beta_2}$ and $\mu_2=e^{\beta_1-\beta_2}$, or $\log(\mu_1)=\beta_1+\beta_2$ and $\log(\mu_2)=\beta_1-\beta_2$, respectively, where $\beta=\begin{bmatrix}\beta_1\\\beta_2\end{bmatrix}$ and β_1 is the intercept and β_2 is the slope of the contrast between treatment and control.

The objective is to test whether or not the treatment is effective. The interest is in a one-sided test. In the disinfectant setting, the disinfected object (treatment) should be at least as effective as the control (no treatment) in reducing the counts of microbes. The counts for the treatment group (Y_2) should be at least as small as those of control (Y_1). Thus, I would want to test the null hypothesis (H_0) that the treatment has no effect versus the alternative (H_1) that it reduces the counts of colonies. These hypotheses can be stated in either of two equivalent ways,

$$H_0: \beta_2 = 0 \text{ vs } H_1: \beta_2 > 0 \text{ , or}$$

$$H_0: \mu_1 = \mu_2 = \mu \text{ vs } H_1: \mu_1 > \mu_2.$$

For results reported to a regulatory agency, often test results and point estimates (rather than interval estimates or confidence intervals) are all that is required. If interval estimates are desired, they are obtainable by inverting the test statistics.

I studied a variety of test statistics and a variety of methods to estimate the variance parameter a. One test statistic was based on a generalization of Rao's score test which accommodates estimating equations from methods other than maximum likelihood. One advantage of (generalized) score tests is that parameters need only be estimated under the null hypothesis. The estimation

methods used in conjunction with generalized score tests were Extended Quasi-likelihood (EQL), Pseudolikelihood (PL), Conditional Maximum Likelihood (CML), Optimal Quadratic (OQ) and Maximum Likelihood (ML). The powers of these tests was compared to the power of the most commonly used test based on the complete distribution, the (generalized) Likelihood Ratio Test. I also studied a new type of estimation and testing, minimum Disparity estimation and a Disparity Difference Test (DDT), based on the Negative Exponential Disparity (NED) measure introduced by Lindsay (1994). Because it is easily applied and conventionally used by many researchers, a t-test (t) was performed on log-transformed NB data. I used a version of the t-test that allows for unequal variances, namely Welch's modified t-test.

CHAPTER 2

ESTIMATION METHODS

I shall explain the methods of estimation in general terms, then the resultant forms of the estimating equations applied to a NB GLM setting (p. 8) where possible. The adaptations of the equations and estimators to a one-way layout will follow.

In my reading, I noted that a number of authors examine the form or properties of the estimating equations themselves rather than those of the estimators [Godambe (1960), Godambe and Heyde (1987), Godambe and Thompson (1987, 1989), Anraku and Yanagimoto (1990), McCullaugh and Nelder (1989, see sections 9.4 and 9.5)]. They develop theory based on the distributions of the estimating functions rather than distributions of the estimators. Citing work by Boos (1980) and others, Godambe and Thompson (1989, p. 171) give examples where use of the "estimating function" improves accuracy of confidence intervals over those based on the corresponding "estimate". An *estimating equation* $g(\mathbf{y},\theta)$ is any function of the data and parameters having $E_\theta[g(\mathbf{y},\theta)]=0$ for all θ (Godambe and Thompson, 1984). Provided there are as many equations as parameters, the estimates $\tilde{\theta}_i$, are obtained by solving the vector equation $g(\mathbf{y},\theta)=0$ for θ.

Maximum Likelihood (ML) Estimation

Maximum likelihood is a long-used and familiar technique for estimation. The full distribution is assumed. When that distribution is correct, maximum likelihood estimators often provide the standard for efficiency against which robust estimators are compared. Maximum likelihood estimation for the NB is discussed by Fisher (1941). Its use in estimation for the variance parameter, $a = 1/k$ is discussed in Piegorsch (1990). Its application to the negative binomial in a generalized linear regression setting is summarized in Lawless (1987). The object of the estimation procedure is to maximize the likelihood, or equivalently the log-likelihood, for the given distribution.

For a sample from the general NB log-linear regression model, the likelihood is proportional to $L(\beta,a)=\prod_{i=1}^{n}\frac{\Gamma\left(y_i+a^{-1}\right)}{\Gamma\left(a^{-1}\right)}\left(\frac{a\mu_i}{1+a\mu_i}\right)^{y_i}\left(\frac{1}{1+a\mu_i}\right)^{a^{-1}}$, where $\mu_i=\mu(\mathbf{x}_i)=e^{\mathbf{x}_i^t\beta}$. Note that if y is an integer ≥ 1, then $\forall c>0$, $\frac{\Gamma(y+c)}{\Gamma(c)}=c(c+1)(c+2)\cdots(c+y-1)$; the gamma ratio equals one for $y=0$. It follows that $\frac{\Gamma\left(y_i+a^{-1}\right)}{\Gamma\left(a^{-1}\right)}=\prod_{j=0}^{y_i-1}a^{-1}(1+aj)=\left(\frac{1}{a^{y_i}}\prod_{j=0}^{y_i-1}(1+aj)\right)$, and thus $L(\beta,a)=\prod_{i=1}^{n}\left(\prod_{j=0}^{y_i-1}(1+aj)\right)\left(\frac{\mu_i}{1+a\mu_i}\right)^{y_i}\left(\frac{1}{1+a\mu_i}\right)^{a^{-1}}$. The log-likelihood is

$$l(\beta,a)=\log\left(L(\beta,a)\right)=\sum_{i=1}^{n}\left[\left(\sum_{j=0}^{y_i-1}\log(1+aj)\right)+y_i\log\mu_i-\left(y_i-a^{-1}\right)\log(1+a\mu_i)\right].$$

The ML estimating equations for the mean and variance parameters are:

$\frac{\partial l}{\partial\beta_s}=\sum_{i=1}^{n}\frac{(y_i-\mu_i)}{(1+a\mu_i)}x_{is}=0$ for $s=1,\ldots,p$ and

$$\frac{\partial l}{\partial a}=\sum_{i=1}^{n}\left[\left(\sum_{j=0}^{y_i-1}\frac{j}{(1+aj)}\right)+a^{-2}\log(1+a\mu_i)-\frac{\left(y_i+a^{-1}\right)\mu_i}{(1+a\mu_i)}\right]=0.$$

Adapting the generalized linear model to a control versus treatment setting, we have $\mu(\mathbf{x}_i) = \mu_1 = e^{\beta_1+\beta_2}$ for $i = 1,\ldots,n_1$ and $\mu(\mathbf{x}_i) = \mu_2 = e^{\beta_1-\beta_2}$ for $i = n_1+1,\ldots,n$. The ML estimating equations for the means become $\frac{\partial l}{\partial \beta_1} = n_1\frac{(\bar{y}_1-\mu_1)}{(1+a\mu_1)} + n_2\frac{(\bar{y}_2-\mu_2)}{(1+a\mu_2)} = 0 = U_1(\beta_1,\beta_2;a)$, say, and $\frac{\partial l}{\partial \beta_2} = n_1\frac{(\bar{y}_1-\mu_1)}{(1+a\mu_1)} - n_2\frac{(\bar{y}_2-\mu_2)}{(1+a\mu_2)} = 0 = U_2(\beta_1,\beta_2;a)$. The ML estimates for the mean are $(\tilde{\beta}_1,\tilde{\beta}_2) = \left(\log\left[(\bar{y}_1\bar{y}_2)^{1/2}\right],\log\left[(\bar{y}_1/\bar{y}_2)^{1/2}\right]\right)$ or $(\tilde{\mu}_1,\tilde{\mu}_2) = (\bar{y}_1,\bar{y}_2)$ under H_1 and $\tilde{\mu} = \bar{y}$ under H_0.

The variance parameter estimate is obtained by inserting the mean estimates into $\frac{\partial l}{\partial a} = 0$ and solving for the root of the equation. A number of methods are applicable, including gradient methods, the scoring algorithm or Newton Raphson method (Walsh, 1975). For solving for the ML estimator of a, and *for all other methods and estimators requiring numerical techniques to obtain roots, I used the Newton Raphson* (NR) *method.* NR was used because it has good convergence properties when starting with good initial values (Walsh, 1975, chapter 4). Initial values were obtained using a grid search.

The above use of the GLM setting provides a model for obtaining an estimate for a common variance parameter in a one-way layout. An alternative method outside the GLM structure in given in Bliss and Owen (1958). In settings where the design matrix is not of the simplified ANOVA form, the mean estimates may not be obtainable in closed form and numerical techniques are required to calculate estimates of the $\tilde{\beta}$'s. Lawless (1987) suggests a method of profile likelihood to find $\tilde{a}$ in the general setting.

Extended Quasi-likelihood (EQL) estimation

Quasi-likelihood (QL) is a robust form of estimation used in generalized linear models (see McCullaugh and Nelder, 1989, chapter 9; first use of the term "quasi-likelihood" was by Wedderburn, 1974). The user specifies only the mean and variance structure rather than the complete form of the likelihood. The general assumptions are in agreement with the model as stated above. Namely, the components of the response vector $\mathbf{Y}$ are independent with mean vector μ and covariance matrix $\sigma^2\mathbf{V}(\mu)$. The mean vector is a known function of β and covariates $\mathbf{x}$. The covariance matrix is the product of σ^2, a scalar dispersion parameter which does not depend on β, and $\mathbf{V}(\mu)$, a diagonal matrix of known functions where the i^{th} diagonal element V_i depends on the i^{th} element of μ (rather than several components of μ). For the NB as defined above, σ^2=1 and $V_i = \mu_i + a\mu_i^2$.

The integral $Q(\mu;y) = \int_y^{\mu} \frac{y-t}{\sigma^2 V(t)} dt$, if it exists, is called the quasi-likelihood, or more correctly, the log quasi-likelihood for μ based on data y. When $Q(\mu;y)$ exists, the estimates for the mean parameters are the solutions to equations obtained by differentiating $Q(\mu;y)$ and are written in the general form $U(\tilde{\beta}) = 0$. The QL estimating equations, $U(\tilde{\beta}) = 0$, are also referred to as the quasi-score functions and have the same form whether or not $Q(\mu;y)$ exists. They can be written as $\underset{px1}{U}(\beta) = D^t V^{-1}(\mathbf{y} - \mu)/\sigma^2$, where D is the matrix of derivatives of the mean with respect to β, $D_{nxp} = \frac{\partial \mu}{\partial \beta^t} = \{D_{ij}\} = \left\{\frac{\partial \mu_i}{\partial \beta_j}\right\}$. Because the components of the response vector are independent, we can express the quasi-score function

for the complete data set as a sum of the individual contributions:

$$\underset{p\,x\,1}{U}(\beta)=\sum_{i=1}^{n}\mathbf{u}_i=\sum_{i=1}^{n}\frac{(y_i-\mu_i)}{V_i}\frac{\partial\mu_i}{\partial\beta_{p\,x\,1}}\Big/\sigma^2.$$

For the NB, $D_{nxp}=\dfrac{\partial\mu_{nx1}}{\partial\beta^{t}_{1xp}}=\left\{\dfrac{\partial\left(e^{\mathbf{x}_i^t\beta}\right)}{\partial\beta_j}\right\}=\left\{x_{ij}e^{\mathbf{x}_i^t\beta}\right\}=\left\{x_{ij}\mu_i\right\}=\underset{nxn}{\left\{diag(\mu_i)\right\}}X_{nxp}$,

where $\underset{nxp}{X}=\begin{bmatrix}\mathbf{x}_1^t\\ \vdots\\ \mathbf{x}_n^t\end{bmatrix}$ and each $\underset{px1}{\mathbf{x}_i}$, the covariates for the i^{th} response, is the transpose of a row of the design matrix, X. The resultant QL estimating equation for the mean parameters is the same as that obtained by maximum likelihood. In the QL estimating equation for the mean, as in the ML estimating equation for the mean, the value of a is treated as fixed while solving for the mean parameters. An additional equation is needed to solve for a.

This method was "extended" to enable estimation of parameters in the variance function by adding a term to the estimating equation (introduced in Nelder and Pregibon, 1987). The Extended Quasi-likelihood equation can be written as: $Q^{+}=Q-\frac{1}{2}\log\left\{2\pi\sigma^2V(y)\right\}$, where $V(y)$ is the variance expressed as a function of y instead of μ, a sort of "empirical" variance. Thus, addition of this term does not change the estimates of the mean parameters. For the NB, $V(y)=y(1+ay)$ and σ^2=1.

For many distributions, the EQL equation is related to the log-likelihood equations. When using the NB mean and variance structure, Q^{+} is directly obtainable from the NB log-likelihood. The only difference in the equation is that the factorials $z!$ in the NB likelihood are replaced by Sterling's approximation, $z!\approx(2\pi z)^{\frac{1}{2}}z^ze^{-z}$. Because the Sterling approximation fails for z=0, an amended form of the approximation, $z!\approx\left\{2\pi(z+c)\right\}^{\frac{1}{2}}z^ze^{-z}$ where c>0, is used for discrete variables with zero in their support space. Nelder and

Pregibon (1987) suggest use of $c=1/6$. (See the Q^+ log-likelihood relationship for the NB in Appendix A).

An "adjustment for degrees of freedom" used to estimate variance or dispersion parameters is recommended by McCullaugh and Nelder (1989, p. 362; see also Nelder and Lee, 1992, p. 281). They suggest multiplying the term of Q^+ containing the empirical variance by $\frac{n-p}{n}$ to account for the fact that p parameters have been fitted to the means.

Adapting EQL to the NB null hypothesis model, the estimate for the mean using EQL is, again, $\tilde{\mu}=\bar{y}$, and p=1 parameter was fit to the mean. The estimate for the NB variance parameter is the solution to $\frac{\partial Q^+}{\partial a}=0=\tilde{U}(a)_{EQL}$, using $\tilde{\mu}=\bar{y}$, which reduces to solving for $\tilde{a}=\tilde{a}_{EQL}$ in $\sum_1^n\left[\frac{1}{\tilde{a}^2}\log\left(\frac{1+\tilde{a}\bar{y}}{1+\tilde{a}y_i}\right)-\frac{(n-1)}{n}\frac{y_i}{1+\tilde{a}y_i}+\frac{(n-1)}{n}\frac{(1+6y_i)}{2(\tilde{a}+6+6\tilde{a}y_i)}\right]=\frac{n-1}{2(\tilde{a}+6)}$. This is the equation listed in Clark and Perry (1989), adjusted for degrees of freedom. (See simplifications in Appendix A).

The above version of EQL was outlined in McCullaugh and Nelder (1983, pp. 212-214), but not present in the second edition (1989; pp. 373-374, 349-350, 360-362). In the later addition for the given model they suggest estimation of a via setting the mean deviance equal to unity, a type of "method of moments" approach. Their general suggestion in the later edition is to view any overdispersion as stemming from a dispersion parameter. Thus using $Var(Y)=\sigma^2V(\mu)$ or more generally $Var(Y_i)=\sigma_i^2V(\mu_i)$ where $\sigma^2\neq1$ and $V(\mu)$ does not contain any additional parameters beyond those found in the mean. This type of analysis could accommodate the NB modeled as a Poisson mixed by a gamma distribution in which the scale parameter (τ), rather than the shape parameter ($\nu=a^{-1}$), is held constant (McCullaugh and Nelder, 1989, p. 199 and

problem 9.1, p. 352; Nelder and Lee, 1992, p. 277). This results in an NB with the variance function as a linear function of the mean and a dispersion parameter, i.e., $Var(Y) = \sigma^2 V(\mu)$ where $V(\mu) = \mu$ and $\sigma^2 = (1+\tau)$, a function of the gamma scale parameter. Nelder and Lee (1992) point out that this model does not belong to the GLM family of distributions, i.e., one that can be written in exponential family form for fixed τ, and show that the Quasi-likelihood estimating equations for the mean will be different than those based on ML for this model. The NB parameterized with a is GLM family for fixed (but unknown)a, with canonical link of $\mathbf{x}^t\beta = \eta = \log\left(\frac{a\mu}{1+a\mu}\right)$ (McCullaugh and Nelder, 1989, p. 373).

I chose to follow the more common approach and assume that the shape parameter remains constant and that the mean varies with the scale of the gamma. In this setting, the mean estimates for the NB model coincide with those obtained by using a Poisson model. This choice is advantageous if one wants to consider the Poisson as a limiting case of the NB. There are differences in the both the Gammas and resultant NBs when one varies either the shape or scale parameters of the Gammas (Figure 1, p. 126). The difference between these choices is more evident in the gammas. Holding the shape fixed while increasing the scale parameter (Figure 1C) is similar to grabbing the right hand tail of the gamma distribution and stretching it out. The relative probability of values near zero is more closely preserved. Holding the scale fixed while increasing the shape parameter (Figure 1A) results in a change of symmetry and density values near zero. In this figure, the NBs produced by the gammas are plotted directly below the gammas. The NB shapes mimic the gamma shapes to a lesser degree. In both cases the NBs have the same means as the gammas that mixed them. The difference is in the

variance of the NBs. The NBs produced by the gammas with shape held constant (Figure 1D) exhibit a wider range of variance for the same change in mean with scale held constant (Figure 1B). The NB samples in this study were generated by holding the shape fixed while varying the mean by changing the scale.

Pseudolikelihood (PL) Estimation

Pseudolikelihood, the term used by Gong and Samaniego (1981), is a robust method for parameter estimation similar to EQL estimation in that one specifies only the mean and variance structure rather than the complete form of the distribution. It is used in those settings in which the variance can be expressed as a function of the mean and additional parameter(s), θ, where θ is single (or vector) valued. Assuming that the mean (regression) parameters are known and equal to the current estimate $\tilde{\beta}$, the general PL equation $l_{PL}\left(\theta,\tilde{\beta}\right)$, is obtained by incorporating the assumed mean and variance function ($v(\theta,\beta)$) into a normal log-likelihood equation, which has the form $l_{PL}\left(\theta,\tilde{\beta}\right)=-\sum_{i=1}^{n}\log\left(2\pi v_i\left(\theta,\tilde{\beta}\right)^{1/2}\right)-\frac{1}{2}\sum_{i=1}^{n}\left(y_i-\mu_i\left(\tilde{\beta}\right)\right)^2\Big/v_i\left(\theta,\tilde{\beta}\right)$ (see Carroll and Ruppert, 1982, for examples of general variance functions.). The PL estimators are the maximizers of the "pseudo-normal" likelihood equations. The PL estimating equation for a variance function parameter is $0=\frac{\partial l_{PL}}{\partial\theta}=\sum_{i=1}^{n}\left[\left(\frac{r_i^2}{v_i}-1\right)\frac{\partial v_i}{\partial\theta}\frac{1}{v_i}\right]$, where $r_i=y_i-\mu_i\left(\tilde{\beta}\right)$ is the i^{th} residual. A modification of the PL estimating equation that incorporates leverage was introduced by Davidian and Carroll (1987). The modification to $0=\frac{\partial l_{PL}}{\partial\theta}$

replaces $\left(\frac{r_i^2}{v_i}-1\right)$ with $\left[\frac{r_i^2}{v_i}-(1-h_i)\right]$ where the h_i are diagonal elements of the projection or "hat" matrix produced from estimation of the mean parameters.

PL estimation was used in the NB setting by Breslow (1989, 1990), who estimated the mean parameters using QL and used PL for the single equation to estimate the variance parameter. The equations use $\theta=a$ and $v_i=\mu_i+a\mu_i^2=V_i$ (previous NB notation for variance function). The PL estimating equation for a is $\tilde{U}(a)_{PL}=\sum_{i=1}^{n}\left[\frac{r_i^2}{V_i}-(1-h_i)\right]\frac{\partial V_i}{\partial a}.\frac{1}{V_i}=0$, where the h_i are the diagonal elements of the projection matrix that arises at convergence of the (QL) iterated weighted least squares solution for the mean parameters for a given value of the variance parameter. The projection matrix, H, is written: $H=Q\left(Q^tQ\right)^{-1}Q^t$, where $Q_{nxp}=\left(diag\left(\frac{\mu_i}{V_i^{1/2}}\right)\right)_{nxn}X_{nxp}$.

PL estimation for the variance function is usually alternated with use of generalized least squares for estimation of β. Use of GLS for estimating β. would amount to minimizing only the second term in $l_{PL}\left(\theta,\tilde{\beta}\right)$ above and arriving at the same solutions for β as QL. Note that one could use the same PL "pseudo-normal" method to obtain estimates for the mean parameters given the current estimate of the variance parameter(s). In this case, only for the Gaussian model with constant variance, would the QL and PL estimators for mean parameters be the same. The PL estimates differ because β occurs in both terms of the PL equation, whereas the added term in EQL is an empirical variance expression which does not contain the mean parameters (Nelder, 1992). For use in variance function estimation, Davidian and Carroll (1988) note that the PL method is asymptotically equivalent to weighted regression on

squared residuals with estimated weights, both being based on the method of moments. Thus the estimating equation for variance parameter(s) is unbiased and the estimates are consistent under general conditions. They prefer PL to EQL for this reason and others based on asymptotic results.

I used PL estimation for the variance parameter only, as implemented for NB by Breslow (1989). For the NB single population model the mean estimate (using QL) and variance estimator (using PL) are $\tilde{\mu} = \bar{y}$ and $\tilde{a}_{PL} = \frac{S^2 - \bar{y}}{\bar{y}^2}$, where $S^2 = \frac{1}{(n-1)} \sum_{i=1}^{n} (y_i - \bar{y})^2$ is the usual sample variance. So, the pseudolikelihood estimator for a in this setting is just the Method of Moments (MOM) estimator. (See calculations and simplifications in Appendix B).

Optimal Quadratic (OQ) Estimation

If, in addition to knowledge about the functional relationship between the mean and variance, one had knowledge about the skewness and kurtosis of the distribution, one could follow the outline of Godambe and Thompson (1989) and use their optimal quadratic estimating equations. In earlier work, Godambe and Thompson (1987) had shown that the QL equation is optimal among linear estimating equations, i.e., for equations of the form $g = \sum_{1}^{n} (y_i - \mu)\alpha_i$ QL provides the optimal constants α_i. The optimality property is defined as equations with minimum variance or maximal information or, equivalently, the highest correlation with the score statistic among all estimating equations of the form g. The emphasis on highest correlation with the score function $[\partial \log f(y,\theta)/\partial\theta]$ stems from the fact that under regularity conditions the score statistic provides the minimal sufficient partitioning of the sample space even when the ML

estimator is not minimal sufficient. In this sense the score function contains all the information in the sample.

Godambe and Thompson extended Godambe's work on the optimal combination of "orthogonal" estimating equations (1985, 1987) to include higher moments. They introduced an "extended quasi-score function" by adding a term to the QL equation and an additional equation to estimate a dispersion parameter. Godambe and Thompson demonstrate the connection between the quasi-score function and the extended quasi-score function and claim that the former is the natural substitute for the latter when the likelihood is undefined. Requirement of knowledge about third and fourth moments raises questions about efficiency and robustness relative to maximum likelihood and to methods requiring knowledge of only the first two moments.

The extended quasi-score function, or OQ estimating equations are written (Godambe and Thompson, 1989) in terms of a parameter vector θ_{px1} and a dispersion parameter σ^2. The means (μ's) and variances (V's) can be any specified function of the parameters in θ_{px1} and the variance is not assumed to depend on θ_{px1} only through the means. The OQ equations are of the general form $\sum_{i=1}^{n} h_{1i} w_{1i} + \sum_{i=1}^{n} h_{2i} w_{2i} = 0$. Here the orthogonal estimating functions are $h_{1i} = y_i - \mu_i$ and $h_{2i} = (y_i - \mu_i)^2 - (\sigma^2 V_i) - \gamma_1 (\sigma^2 V_i)^{1/2} (y_i - \mu_i)$. The orthogonality implies that $E(h_{1i} h_{2j}) = 0 \; \forall i, j = 1 \ldots n$ and that the estimating equations are additive or provide "additive information". The optimal weights (optimality as defined in previous paragraph) are $w_{1i} = \dfrac{(\partial \mu_i / \partial \theta_r)}{\sigma^2 V_i}$ and $w_{2i} = \dfrac{\left[\gamma_1 \left(\dfrac{\partial \mu_i}{\partial \theta_r}\right) - \sigma \left(\dfrac{\partial V_i}{\partial \theta_r}\right) \Big/ V_i^{1/2}\right]}{(\sigma^2 V_i)^{3/2} (\gamma_2 + 2 - \gamma_1^2)}$ (proof in Godambe and Thompson (1989)). The first

summation ($\sum_{i=1}^{n} h_{1i} w_{1i} = 0$) is just the quasi-score function from QL. The second summation ($\sum_{i=1}^{n} h_{2i} w_{2i} = 0$) incorporates the higher moments, skewness (γ_1) and kurtosis (γ_2).

For the general GLM NB model (p. 8) with means expressed in terms of px1 explanatory variables **x** and parameter vector β and a common variance parameter (a), the OQ estimating equations reduce to QL for the mean parameters and solving the additional OQ estimating equation for the variance parameter: $\tilde{U}(a)_{OQ} = \sum_{i=1}^{n} \left\{ \frac{\left[(y_i - \tilde{\mu}_i)^2 - \tilde{\mu}_i(1+a\tilde{\mu}_i) - (1+2a\tilde{\mu}_i)(y_i - \tilde{\mu}_i) \right]}{(1+a\tilde{\mu}_i)^2} \right\} = 0$. Under H_0 for the treatment versus control NB model, the QL estimators are $\tilde{\mu} = \bar{y}$ and $\tilde{a}_{OQ} = \frac{\left(\frac{n-1}{n}\right) S^2 - \bar{y}}{\bar{y}^2}$ (See calculations in Appendix C).

Conditional Maximum Likelihood (CML) Estimation

A technique used to obtain an estimating equation for a parameter of interest in a multiple parameter distribution is conditional likelihood. The full likelihood is factored into a conditional likelihood and a residual likelihood. The conditional likelihood depends on only the parameter of interest. The residual likelihood depends on the other parameters. The estimating equation for the parameter of interest is obtained by maximizing the conditional likelihood rather than the full likelihood. This is the same rationale that underlies the popular Restricted Maximum Likelihood Estimate (REML) for variance components in analysis of variance settings (Searle, et. al., 1992, Chapter 3).

The following factorization was presented in Anraku and Yanagimoto (1990). The negative binomial likelihood for a single population can be factored into a conditional likelihood, L_C, which depends on the variance parameter but not on the mean, and a residual likelihood, L_R, as follows. Conditioning on $t=\sum_{i=1}^{n} y_i$, factor the likelihood as

$L(y_i;\beta,a) \quad = \quad L_C(y_i;a|t) \quad \times \quad L_R(t;\beta,a)$, which is

$$\prod_{i=1}^{n}\frac{\Gamma(y_i+a^{-1})}{y_i!\Gamma(a^{-1})}\left(\frac{a\mu}{1+a\mu}\right)^{y_i}\left(\frac{1}{1+a\mu}\right)^{a^{-1}} = \frac{\prod_{i=1}^{n}\frac{\Gamma(y_i+a^{-1})}{y_i!\Gamma(a^{-1})}}{\binom{n/a+t-1}{t}}\times\binom{n/a+t-1}{t}\left(\frac{a\mu}{1+a\mu}\right)^{t}\left(\frac{1}{1+a\mu}\right)^{n/a}.$$

The conditional likelihood, $L_C(y_i;a|t)$, is proportional to

$$LC=\frac{\prod_{i=1}^{n}\frac{\Gamma(y_i+a^{-1})}{\Gamma(a^{-1})}}{\frac{\Gamma(n/a+t)}{\Gamma(t)}}=\frac{\prod_{i=1}^{n}\{1/a(1/a+1)(1/a+2)\cdots(1/a+y_i-1)\}}{n/a(n/a+1)(n/a+2)\cdots(n/a+t-1)}$$

$$= \frac{\prod_{i=1}^{n}\prod_{j=0}^{y_i-1} 1/a(1+aj)}{\prod_{j=0}^{t-1} 1/a(n+aj)} = \frac{\prod_{i=1}^{n}\prod_{j=0}^{y_i-1}(1+aj)}{\prod_{j=0}^{t-1}(n+aj)}, \text{ and,}$$

$$\log(LC) = \left[\sum_{i=1}^{n}\sum_{j=0}^{y_i-1}\log(1+aj)\right] - \left[\sum_{j=0}^{t-1}\log(n+aj)\right].$$

The maximizer of the conditional likelihood is the solution $\tilde{a}_{CML}$ to $\frac{\partial \log(LC)}{\partial a} = 0 = \left[\sum_{i=1}^{n}\sum_{j=0}^{y_i-1}\frac{j}{1+aj}\right] - \left[\sum_{j=0}^{t-1}\frac{j}{n+aj}\right]$. It is obtainable using root finding techniques.

Anraku and Yanagimoto state that "$\bar{y}$ is a reasonable estimator of μ irrespective of the estimator of a." They suggest $\tilde{a}_{CML}$ as a robust alternative to the maximum likelihood estimate for the variance parameter when $\frac{n-1}{n}S^2 > \bar{y}$. Anraku and Yanagimoto claim that the uniqueness of $\tilde{a}_{CML}$ when $S^2 > \bar{y}$ follows from work by Levin and Reeds (1977) in terms of the compound multinomial distribution. When $S^2 < \bar{y}$, they define $\tilde{a}_{CML}$ as zero. I used $\bar{y}$ and $\tilde{a}_{CML}$ as estimates under the null hypothesis setting.

Although the conditioning method as described above is not applicable to the general GLM setting, it is applicable to the one-way layout structure with a common variance parameter using conditioning on the group totals. The one-way layout was studied by Anraku and Yanagimoto (1990). Results from their simulation showed that the CML estimators for a, $\frac{a}{(a+1)}$ and $\frac{1}{a}$ performed better than those based on ML and MOM in terms of bias, mean square error and Kullback-Leibler risks for multiple populations.

Negative Exponential Disparity (NED) Estimation

The Negative Exponential Disparity estimator belongs to a class of estimators known as minimum disparity estimators (Lindsay, 1994). These estimators correspond to the minimization of a "disparity" or a measure of the distance between a pair of densities, namely the data (or empirical) density and the model (or probability) density. In a simplified sense, one can think of looking for an estimator which yields the best match between the heights corresponding to the histogram of the data and the heights corresponding to a family of (discrete) probability densities. The general theory for minimum disparity estimators is easily applied to discrete distributions. Though it can be applied with modifications to continuous distributions (Basu and Lindsay, 1994; Basu and Sarkar, 1994), I will present the general theory in the context of discrete distributions only.

Let the sample space be Y={0,1,2,...,K} with K possibly infinite, and assume m_θ is a family of probability densities on Y indexed by θ, a vector of parameters. Assume, unless otherwise noted, that summations (denoted by $\sum$) are taken over the entire sample space. Assume that the data are n iid observations made from m_θ. Let $d(y)$ be the proportion of the n sample observations which has value y. Lindsay defines the Pearson residual function $\delta(y)$ as $\delta = \delta(y) = \frac{[d(y) - m_\theta(y)]}{m_\theta(y)}$ (this name was used because the model-weighted sum of the squared residuals $\sum m_\theta(y)\delta(y)^2$ is Pearson's chi-squared distance). The disparity measure, ρ, between the data proportions, $d(y)$, and the model density values, $m_\theta(y)$, is a function of the Pearson residual function and defined as $\rho_\theta(d, m_\theta) = \sum G(\delta(y))m_\theta(y)$ where G is a strictly convex thrice-differentiable function. Assuming the differentiability of the model density,

minimization of the disparity measure involves the solution of an estimating equation for θ_j of the form $-\frac{\partial}{\partial\theta_j}\rho(d,m_\theta)=\sum A(\delta(y))\frac{\partial m_\theta(y)}{\partial\theta_j}=0$,where $A(\delta)=(1+\delta)G'(\delta)-G(\delta)$ is termed the residual adjustment function (RAF). The form and properties of the estimation procedure depend on the choice of G. Lindsay presents disparity measure methods as an unifying concept and lists the choices of G (and associated RAFs, A) for many estimation techniques, including maximum likelihood (ML) and the common distance-type measures of minimum Pearson's chi-squared (PCS), minimum Neyman's chi-squared (NCS), minimum Kullback-Leibler divergence (KL) and minimum Hellinger distances (HD) (see Lindsay, 1994, pp. 1086-1089, 1101, 1103). The shape of the RAF determines the tradeoff between the estimator's robustness and efficiency. RAF construction or choice is listed as a starting point for building new disparity measures. The RAFs chosen with strict convexity and differentiability insure some desirable properties of the estimators.

Lindsay (1994, p. 1103) introduces the Negative Exponential disparity (NED) as a new disparity measure, determined by $G(\delta)=e^{-\delta}-1$. NED estimators are shown to be second-order efficient, and robust in the sense that they reduce the effects of outliers and "inliers" in the data (proofs in Lindsay). Inliers (term first used by Lindsay), or smaller than expected sample proportions, are defined as values of δ near -1. They can be due to empty cells, i.e., $d(y)$=0. The RAFs in Figure 2 (p. 127) present a combination of Lindsay's Figures 3 and 5. Curvature towards the x-axis indicates the degree of robustness to outliers (for δ>0) and inliers (for δ<0). The NED has the only RAF in the plot that is robust for both (is convex for δ>0 and concave for δ<0).

I chose to apply the NED method for estimation and testing. To my knowledge, it has not been used for the NB(μ,a) distribution. Construction for the estimators and a test follow.

NED applied to NB single population

I provide here some notation and abbreviations of disparity functions in terms of NED and the NB(μ,a) family.

Let $\theta=(\mu(\beta),a)$ be the vector of parameters which indexes the NB family of pmfs, and $m_\theta=m_\theta(y)=\frac{\Gamma\left(y+a^{-1}\right)}{y!\Gamma\left(a^{-1}\right)}\left(\frac{a\mu}{1+a\mu}\right)^y\left(\frac{1}{1+a\mu}\right)^{a^{-1}}$, $y=0,1,2,\ldots$, denote the model density or likelihood. I will also use the NB likelihood with the parameterization of $k=a^{-1}$, yielding $m_\theta=\frac{\Gamma(y+k)}{y!\Gamma(k)}\left(\frac{\mu}{\mu+k}\right)^y\left(\frac{k}{k+\mu}\right)^k$. Let $l_\theta=l_\theta(y)=\log[m_\theta(y)]$ denote the log-likelihood. Thus, $\frac{\partial m_\theta}{\partial\theta_j}=m_\theta\frac{\partial l_\theta}{\partial\theta_j}=m_\theta u_j$, where u_j is the score function or derivative of the log-likelihood with respect to θ_j (similarly, u_μ will denote derivative of the log-likelihood with respect to μ). Recall, the Pearson residual, $\delta=\delta(y)=\frac{[d(y)-m_\theta(y)]}{m_\theta(y)}$, takes on the value δ=-1 whenever $d(y)=0$. A sample proportion of zero, $d(y)=0$, results when the value y does not occur in the realized data. Use of this fact results in some simplifications in numerical calculations of several equations below.

The disparity function is $G(\delta)=e^{-\delta}-1$ and has derivatives $G'(\delta)=-e^{-\delta}$ and $G''(\delta)=e^{-\delta}$. Thus, $A^*(\delta)=(1+\delta)G'(\delta)-G(\delta)=1-(2+\delta)e^{-\delta}$ and $A(\delta)=A^*(\delta)-A^*(0)=2-(2+\delta)e^{-\delta}$ is the RAF, adjusted so that $A(0)=0$.

The disparity measure is defined as

$$\begin{aligned}\rho_\theta(d,m_\theta) &= \sum G(\delta(y))m_\theta(y) \\ &= G(-1)+\sum\left[G(\delta(y))-G(-1)\right]m_\theta(y) \\ &= (e-1)+\sum\left[e^{-\delta(y)}-e\right]m_\theta(y), \text{ for NED.}\end{aligned}$$

The equations in the second and third lines are simplifications used in numerical computations and are equivalent to the first equation because $G(-1)$ is constant and $\sum m_\theta(y)=1$. The equations are a useful representation since $\left[G(\delta)-G(-1)\right]=0$ whenever $\delta=\delta(y)=-1$, so the disparity measure need be summed only over observed data values rather than the entire support space. I shall refer to equations summed over observed data points as "working" equations. They are the forms I used in computation.

The general form for the NED estimating equations is $-\frac{\partial}{\partial\theta_j}\rho(d,m_\theta)=0=\sum A(\delta)\frac{\partial m_\theta(y)}{\partial\theta_j}=\sum\left[A(\delta)-A(-1)\right]m_\theta(y)u_j$. The last equation is the working equation and is equivalent because $A(-1)$ is a constant and $\sum m_\theta(y)u_j=E(u_j)=0$. So, the estimators are the solutions to $0=\sum\left[e-(2+\delta)e^{-\delta}\right]m_\theta(y)u_j$, where u_j is the score function for the estimated parameter.

To estimate the mean of NB(μ,a), include the mean score function, $u=u_\mu=\frac{y-\mu}{\mu(1+a\mu)}$. The NED estimating equation for the mean becomes

$$-\frac{\partial}{\partial\mu}\rho(d,m_\theta)=0=\sum\left[A(\delta)-A(-1)\right]m_\theta(y)u_\mu=\sum\left\{\frac{\left[e-(2+\delta)e^{-\delta}\right]m_\theta}{\mu(1+a\mu)}\right\}(y-\mu).$$

The estimator for the NB mean, $\tilde{\mu}$, can be iteratively reweighted. The updated estimate $\tilde{\mu}_{r+1}=\frac{\sum w_r y}{\sum w_r}$, where the weights from the rth iteration, w_r, are

$w_r = \dfrac{\left[e-(2+\delta_r)e^{-\delta_r}\right]m_{\theta r}}{\tilde{\mu}_r(1+a\tilde{\mu}_r)}$, and where a is the current estimate of the variance parameter.

The estimator for the NB variance parameter, a, must be obtained by a root finding routine in conjunction with the NED equations. This is because one cannot factor or otherwise separate out the variance parameter in the score function. The working form of the estimating equation is, again, $-\dfrac{\partial\rho(d,m_\theta)}{\partial a} = 0 = \sum\left[A(\delta)-A(-1)\right]m_\theta(y)u_a$, where u_a is the variance parameter score function, i.e., $u = u_a = \sum_{j=0}^{y-1}\left(\dfrac{j}{1+aj}\right)+a^{-2}\log(1+a\mu)-\dfrac{\left(y+a^{-1}\right)\mu}{(1+a\mu)}$. I solve for the root of this NED estimating equation using the Newton-Raphson method. In order to apply this numerical technique, I need an additional partial derivative.

For the disparity measures in general,

$$\frac{-\partial^2\rho}{\partial\theta_j^{\,2}} = \frac{\partial}{\partial\theta_j}\left(\frac{-\partial\rho}{\partial\theta_j}\right) = \frac{\partial}{\partial\theta_j}\left[\sum A(\delta)\frac{\partial m_\theta}{\partial\theta_j}\right]$$

$$= \sum\left[\frac{\partial A(\delta)}{\partial\theta_j}\frac{\partial m_\theta}{\partial\theta_j} + A(\delta)\frac{\partial^2 m_\theta}{\partial\theta_j^{\,2}}\right] = \sum\left\{\left(\frac{\partial A(\delta)}{\partial\delta}\frac{\partial\delta}{\partial m_\theta}\frac{\partial m_\theta}{\partial\theta_j}\right)\frac{\partial m_\theta}{\partial\theta_j} + A(\delta)m_\theta\left[\left(\frac{\partial l}{\partial\theta_j}\right)^2 + \frac{\partial^2 l}{\partial\theta_j^{\,2}}\right]\right\}$$

$$= \sum m_\theta\left\{-A'(\delta)(\delta+1)\left(\frac{\partial l}{\partial\theta_j}\right)^2 + A(\delta)\left[\left(\frac{\partial l}{\partial\theta_j}\right)^2 + \frac{\partial^2 l}{\partial\theta_j^{\,2}}\right]\right\}.$$

For NED and the NB variance parameter, note that $A'(\delta) = (1+\delta)e^{-\delta}$, which is zero whenever $\delta(y) = -1$. Use of $\left[A(\delta)-A(-1)\right]$ in place of $A(\delta)$ does not change the equation, since $\sum m_\theta\left[\left(\dfrac{\partial l}{\partial\theta_j}\right)^2 + \dfrac{\partial^2 l}{\partial\theta_j^{\,2}}\right] = E\left[\left(\dfrac{\partial l}{\partial\theta_j}\right)^2 + \dfrac{\partial^2 l}{\partial\theta_j^{\,2}}\right] = 0.$

The NB working equation becomes

$$\frac{-\partial^2\rho}{\partial a^2}=\frac{\partial}{\partial a}\left(\frac{-\partial\rho}{\partial a}\right)=\sum m_\theta\left\{\left[(1+\delta)^2e^{-\delta}\right]\left(\frac{\partial l}{\partial a}\right)^2+\left[e-(2+\delta)e^{-\delta}\right]\left[\left(\frac{\partial l}{\partial a}\right)^2+\frac{\partial^2 l}{\partial a^2}\right]\right\},$$

where $\frac{\partial l}{\partial a}=u_a$ is listed above and

$$-\frac{\partial^2 l}{\partial a^2}=-\frac{\partial u_a}{\partial a}=\sum_{j=0}^{y-1}\left(\frac{j}{1+aj}\right)^2+2a^{-3}\log(1+a\mu)-\frac{2a^{-2}\mu}{1+a\mu}-\frac{\left(y+a^{-1}\right)\mu^2}{(1+a\mu)^2}.$$

Thus, the estimator for a is obtained by iterative solutions to the Newton Raphson equation, updating estimates using $\tilde{a}_{r+1}=\tilde{a}_r-\frac{(-\partial\rho/\partial a)}{\left(-\partial^2\rho/\partial a^2\right)}\Big|a=\tilde{a}_r,\mu=\tilde{\mu}.$

Estimates for μ and a are obtained alternately until convergence of the overall disparity measure ρ. The convergence criteria for both the individual estimates and the overall system or set of estimates is at each stage based on the convergence in the relative size ofρ, i.e., $(\text{abs}(\rho_{new}-\rho_{old}/\rho_{new})<\text{tol}.$

Treatment versus Control setting

The NED methodology may be extended to a treatment versus control setting. Just as the estimates in a single population setting can be thought of as minimizing an expectation, $\rho=E(G(\delta))$, one can think of the two population setting as minimizing a ρ defined as the expected value of a conditional expectation. Here I will calculate ρ for the sample from each population and use the sampling technique of weights proportional to size. Thus, I will use weights $w_i=n_i/n$. and minimize $\rho=E_{\mathbf{x}}\left[E\left(G|\mathbf{x}_i\right)\right]=\sum_{i=1}^{2}\frac{n_i}{n}\sum G\left(\delta\left(y|\mathbf{x}_i\right)\right)m_\theta\left(y|\mathbf{x}_i\right),\quad (*)$

where $\mathbf{x}_i^t$ for $i=1,2$ are the unique rows of X. Here, $\mathbf{x}_1^t=[1\ \ 1]$ and $\mathbf{x}_2^t=[1\ \ -1]$ indicate the groups of control or treatment.

In the 2 population setting with parameters $\theta=(\mu_1,\mu_2,a)$, use $\delta_i=\delta\left(y|\mathbf{x}_i\right)=\frac{\left[d\left(y|\mathbf{x}_i\right)-m_\theta\left(y|\mathbf{x}_i\right)\right]}{m_\theta\left(y|\mathbf{x}_i\right)}$ as the Pearson residuals for the sample proportions and model density per group or population. Similarly,

$G_i = G\big(\delta(y|\mathbf{x}_i)\big)$ and $A_i = A\big(\delta(y|\mathbf{x}_i)\big)$ are functions are defined in terms the two populations via δ_i. Again, let $G_i' = \partial G_i/\partial\delta_i$ and $A_i' = \partial A_i/\partial\delta_i$.

Solving for the parameters in the mean one obtains

$$-\frac{\partial\rho}{\partial\beta_r} = \sum_{i=1}^{2}\frac{n_i}{n}\sum A\big(\delta(y|\mathbf{x}_i)\big)\frac{\partial m_\theta(y|\mathbf{x}_i)}{\partial\beta_r} = \sum_{i=1}^{2}\sum\frac{n_i}{n}A\big(\delta(y,\mathbf{x}_i)\big)m_\theta(y|\mathbf{x}_i)\frac{\partial l_\theta(y|\mathbf{x}_i)}{\partial\mu_i}\frac{\partial\mu_i}{\partial\beta_r}$$

$$= \sum_{i=1}^{2}\sum w_i(y|\mathbf{x}_i)(y-\mu_i)\mathbf{x}_i \text{ for } r=1,2 \text{ where } w_i = w(y|\mathbf{x}_i) = \frac{n_i}{n}A\big(\delta(y|\mathbf{x}_i)\big)\frac{m_\theta(y|\mathbf{x}_i)}{(1+a\mu_i)}.$$

So, $\dfrac{-\partial\rho}{\partial\beta_1} = \sum w_1(y-\mu_1)+\sum w_2(y-\mu_2) = 0$,

and $\dfrac{-\partial\rho}{\partial\beta_2} = \sum w_1(y-\mu_1)-\sum w_2(y-\mu_2) = 0$.

Solving these simultaneously leads to $\dfrac{-\partial\rho}{\partial\beta_1}+\dfrac{-\partial\rho}{\partial\beta_2} = 0 = 2\sum w_1(y-\mu_1)$ and $\dfrac{-\partial\rho}{\partial\beta_1}-\dfrac{-\partial\rho}{\partial\beta_2} = 0 = 2\sum w_2(y-\mu_2)$. Thus one can iteratively solve for the estimate of the mean of the population (treatment or control) by using the sample from that population and the same technique as for a single population, i.e.,

$$\tilde{\mu}_{i,r+1} = \frac{\sum w_{i,r}y}{\sum w_{i,r}}.$$

One could solve for the β parameters, because $\mu_1 = e^{\beta_1+\beta_2}$ and $\mu_2 = e^{\beta_1-\beta_2}$ imply $\mu_1\mu_2 = e^{2\beta_1}$ and $\mu_1/\mu_2 = e^{2\beta_2}$. As estimates of the β's we can use $\tilde{\beta}_1 = \log\Big[(\tilde{\mu}_1\tilde{\mu}_2)^{\frac{1}{2}}\Big]$ and $\tilde{\beta}_2 = \log\Big[(\tilde{\mu}_1/\tilde{\mu}_2)^{\frac{1}{2}}\Big]$. This choice is in the spirit of Lehmann (TPE, p. 112) using the invariance property of MLEs, $\hat{g}(\theta) = g\big(\hat{\theta}\big)$. If the invariance property holds for NED estimators, the β's would be estimated as shown. These estimates seem a reasonable starting point in any case. Note, however, that only the $\tilde{\mu}_i$'s are needed in order to construct a test.

To solve for the variance parameter, histogram information from both populations is needed. Again go back to the definition of ρ and minimize the disparity measure. Using $\rho = E_{\mathbf{x}}\Big[E\big(G|\mathbf{x}_i\big)\Big] = \sum_{i=1}^{2}\frac{n_i}{n}\sum G\big(\delta(y|\mathbf{x}_i)\big)m_\theta(y|\mathbf{x}_i)$ the NED

estimating equation for the variance parameter is $-\frac{\partial \rho}{\partial a} = \sum_{i=1}^{2} \frac{n_i}{n} \sum A\left(\delta\left(y|\mathbf{x}_i\right)\right) m_\theta\left(y|\mathbf{x}_i\right) \frac{\partial l_\theta}{\partial a} = 0$. The NED variance estimator, $\tilde{a}_{NED}$, provides the root or solution to this equation.

To apply Newton-Raphson use the derivative $\frac{\partial}{\partial a}\left(-\frac{\partial \rho}{\partial a}\right) =$ $\sum_{i=1}^{2} \frac{n_i}{n} \sum m_\theta\left(y|\mathbf{x}_i\right) \left\{ -A'\left(\delta\left(y|\mathbf{x}_i\right)\right)\left(1+\delta\left(y|\mathbf{x}_i\right)\right)\left(\frac{\partial l_\theta}{\partial a}\right)^2 + A\left(\delta\left(y|\mathbf{x}_i\right)\right)\left[\left(\frac{\partial l_\theta}{\partial a}\right)^2 + \frac{\partial {l_\theta}^2}{\partial a^2}\right] \right\}$ and iterate until convergence using updates $\tilde{a}_{r+1} = \tilde{a}_r - \frac{\left(-\frac{\partial \rho}{\partial a}\right)}{\frac{\partial}{\partial a}\left(-\frac{\partial \rho}{\partial a}\right)}$. The working equations are obtainable using the same steps as for the single population estimates. Convergence criterion are based on convergence of ρ as in the single population setting (above).

CHAPTER 3

TESTING METHODS

Likelihood Ratio Test (LRT)

The statistic for the generalized Likelihood Ratio test is the logarithm of a ratio of likelihoods, $LRT = 2n(l_{H0} - l_{H1})$. McCullagh and Nelder refer to the LRT for GLMs as a "deviance" and, analogous to (Normal error) Analysis of Variance, to partitioned LRTs as Analysis of Deviance. A discussion of the difficulties that arise is covered in McCullaugh and Nelder (1989, pp. 35-36). The LRT has an asymptotic chi-square distribution with degrees of freedom equal to the difference in the numbers of parameters estimated under the null and alternative hypotheses. (McCullaugh and Nelder,1989; see Appendix C).

Disparity Difference Test (DDT)

Disparity Difference Tests (DDT) are presented by Lindsay (1994) in the simple setting of testing the parameters equal to a given value. Using the estimators obtained in the treatment versus control setting (above) I apply DDT to the desired two-sample test. DDTs constructed from NED estimators have an asymptotic Chi-square distribution if the family of model densities satisfies the conditions in Lehmann (1991, pp 409 and 429) and additional bounded expectations (Assumption 31, p. 1109 of Lindsay,1994). The DDT has an analogous form to the Likelihood Ratio Test (LRT) and is expressed as $DDT = 2n(\rho_{H0} - \rho_{H1})$. The NB family does satisfy the required conditions (proofs in Appendix D).

The disparity measures differenced in the DDT are based on conditional ρ, a weighted average of the measures calculated for the two samples. The test amounts to determining the degree of improvement or reduction in disparity which occurs when more parameters are estimated under H_1 than estimated under H_0. The asymptotic distribution of the DDT is chi-square with degrees of freedom (df) equal to the difference in the numbers of parameters calculated in the two models (Lindsay, 1994, Theorem 6 and Appendix A).

Figure 3 (p. 128) illustrates the difference in the fit to a data set due to disparity measures minimized for estimation under H_0 (p. 27) and H_1 (p. 30). This sample was generated from two NBs with $n_1 = n_2$=15 (overall n=30), $\mu_1 > \mu_2$ and a common value of a. The top left histogram is that of the total sample or the combination of the 2 samples into a single sample. The X's show the best "fit" under H_0 (the pmf based on estimation under H_0 multiplied by the total sample size). The bottom 2 histograms are those of the 2 samples (control and treatment) considered separately. The diamonds represent the best fit under H_1 to the sample means and a common variance parameter. The X's show the fit to each sample based on H_0 estimation. The difference in the disparity under H_0 and H_1 is a function of the distances between the X's or diamonds and the tops of the histograms, respectively (as pictured in the bottom row). [The top right histogram is the sum of the 2 lower histograms and the diamonds are the sum of the diamond heights. This combined sample with combined frequency estimates (diamonds) is not used in analysis but presented to show the similarity of fit to combined samples (top row) using the estimates obtained under H_0 versus H_1 (X's versus diamonds). The best fit under H_0 (top left) appears to give us the same fit as the sum of the H_1 fits on the separate samples applied to the combined sample (top right).] Again, differences in disparity are

not based on disparity calculations applied to the combined sample. The disparity measures are calculated from the fits in the bottom row where one can see a difference between the X's and diamonds. Each disparity measure is the NED-based function ((*) p. 30) of the differences between the histogram heights and the heights based on estimation (X's and diamonds). The H_0 disparity is ρ_{H0}=0.2987 and the H_1 disparity is ρ_{H1}=0.1451, resulting in a DDT statistic of 9.216 and a p-value of less than 0.01. Thus the conclusion is that there is a difference between the means of the 2 populations.

Welch Modified Two-Sample t-test (t)

This version of the t-test is conventionally recommended when the variances are unequal (Miller,1986, section 2.3; Snedecor and Cochran, 1980). The statistic is $t = \frac{(\bar{y}_1 - \bar{y}_2) - (\mu_1 - \mu_2)}{s_d}$, where $s_d = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$. Its distribution under the null hypothesis can be approximated by a t-distribution where the degrees of freedom (df) are calculated by a formula based on the variances of the two samples, i.e., $\mathrm{df} = \left[\frac{c^2}{n_1 - 1} + \frac{(1-c)^2}{n_2 - 1}\right]^{-1}$, where $c = \frac{s_1^2}{n_1 s_d^2}$. In the simulation study, the t is applied to log(y+0.5), where y is the sample data, as this is the standard transformation used on microbial count data (Niemelä, 1983). The intent is not to provide the true variance stabilizing (or normalizing) transformation but to mimic what is commonly used in practice. The Welch t-test is used instead of the standard t-test because it is available in software packages and should handle problems remaining if the transformation is less than optimal.

Generalized Score Tests (S)

Boos (1992) presented generalizations of Rao's score test (Rao 1948; see also Cox and Hinkley, 1974) which make use of general estimating equations (rather than just derivatives of the (Normal) log-likelihood) and empirical variance estimates. Rao's (original) score statistics (notation from Boos (1992)) have the general form $S(\tilde{\theta})^t \tilde{I}_f^{-1} S(\tilde{\theta})$, where $S(\theta)$ is the vector of partial derivatives of the log-likelihood function, $\tilde{\theta}$ is the vector of restricted maximum likelihood estimates under H_0 and $\tilde{I}_f$ is the Fisher information of the sample evaluated at $\tilde{\theta}$. One advantage of the score statistics is that they only require computation of the parameter estimates under the null hypothesis. They are asymptotically equivalent to the likelihood ratio statistics under the null hypothesis, H_0. Both the likelihood ratio statistic and the score statistic are invariant under reparameterization or non-linear transformations.

Recall, a (general) estimating equation $g(\mathbf{y},\theta)$ is defined as any function of the data and parameters having zero mean for all parameter values (Godambe and Thompson, 1984). Provided there are as many equations as parameters, the estimates $\tilde{\theta}_i$ are obtained by solving the vector equation $g(\mathbf{y},\theta)=0$ for θ. Boos (1992) refers to the estimating equations as "score functions".

The generalized score test is constructed in the following manner, outlined in Boos (1992, p. 328).

> "Find the asymptotic covariance matrix of the score function $g(\mathbf{y},\theta_0)$ under H_0, say Σ_g. Then define the generalized score statistic to be $T_{GS} = g(\mathbf{y},\theta)^t \tilde{\Sigma}_g^- g(\mathbf{y},\theta)$, where $\tilde{\Sigma}_g^-$ is a generalized inverse of a consistent estimate $\tilde{\Sigma}_g$ of Σ_g. Usually a version of $\tilde{\Sigma}_g^-$ is available which is easily computed."

The score statistics have an asymptotic null chi-squared distribution. Boos shows how the various forms of generalized score tests arise from Taylor expansion of the estimating equations.

General information on Score tests in a GLM setting are found in Pregibon (1982). Breslow (1989, 1990) presents the theory and form of the score tests in the NB setting for testing a hypothesis about a subset of the mean vector. Estimation of the mean parameters is via QL and the variance parameter via an additional equation. Breslow uses PL for the variance parameter but the following discussion would hold for any estimation procedure described above (see Appendix E). The QL estimating equations for the mean are $\underset{p\times 1}{U}(\beta,a)=\sum_{i=1}^{n}\mathbf{u}_i=\sum_{i=1}^{n}\frac{(y_i-\mu_i)}{V_i}\frac{\partial\mu_i}{\partial\beta_{p\times 1}}=\sum_{i=1}^{n}\frac{(y_i-\mu_i)}{(1+a\mu_i)}\mathbf{x}_i=X^t\left[diag\left(\frac{1}{1+a\mu_i}\right)\right](\mathbf{y}-\mu)$,

where $\mathbf{x}_i$, the covariates for the i^{th} observation, is the transpose of a row of X (refer to section on EQL Estimation). Breslow assumes that the mean structure is correctly specified but allows for possible misspecification of the variance. The empirical covariance matrix for score statistics of the vector of mean parameters, β, is $G=\sum_{i=1}^{n}\mathbf{u}_i\mathbf{u}_i^t=\sum_{i=1}^{n}\frac{(y_i-\mu_i)^2\mu_i^2}{V_i^2}\mathbf{x}_i\mathbf{x}_i^t=X^t\left[diag\left(\frac{(y_i-\mu_i)^2\mu_i^2}{V_i^2}\right)\right]X$. The negative expectation of the partial derivatives of the mean score statistic is $A=-\sum_{i=1}^{n}E\frac{\partial\mathbf{u}_i}{\partial\beta^t}=\sum_{i=1}^{n}\frac{\mu_i^2}{V_i}\mathbf{x}_i\mathbf{x}_i^t=X^t\left[diag\left(\frac{\mu_i^2}{V_i}\right)\right]X$. The asymptotic variance of $\tilde{\beta}$ is estimable by $A^{-1}GA^{-1}$, even if the variance is misspecified. This is called the "empirical covariance matrix". If the variance is correctly specified then $E(G)=A$ and one may estimate Var(β) by A^{-1}, which Breslow refers to as "model based".

In the general setting, β is partitioned into sub vectors of lengths p_1 and p_2, i.e., $\beta_{px1} = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}_{p2\,x\,1}^{p1\,x\,1}$. The set of covariables and other matrices are conformably partitioned. Generalized score tests for testing $H_0 : \beta_2 = \beta_2^0$ are based on $U_2 = X_2^t \left[diag\left(\frac{\mu_i}{V_i} \right) \right] (\mathbf{y} - \mu) = \sum_{i=1}^{n} \frac{(y_i - \mu_i)\mu_i}{V_i} \mathbf{x}_{2i}$ where $\tilde{\beta}$ is the quasi-likelihood estimate under H_0, i.e., $\tilde{\beta}_2 = \beta_2^0$ and $\tilde{\beta}_1^0$ is the solution to $U_1 = \sum_{i=1}^{n} \frac{(y_i - \mu_i)\mu_i}{V_i} \mathbf{x}_{1i} = 0$. The test statistic is $T_{GS} = U_2{}^T \tilde{\Sigma}_g^{-1} U_2$, where $\tilde{\Sigma}_g = I_{2\cdot 1} = A_{22} - A_{21} A_{11}^{-1} A_{12}$ for a model based score test and $\tilde{\Sigma}_g = I_{2\cdot 1}^e = G_{22} - A_{21} A_{11}^{-1} G_{12} - G_{21} A_{11}^{-1} A_{12} + A_{21} A_{11}^{-1} G_{11} A_{11}^{-1} A_{12}$ for an empirical score test. The matrices A and G have been partitioned into appropriate sub matrices of the dimensions $p_1 \times p_1$, $p_1 \times p_2$, etc.

For the treatment versus control setting testing $H_0 : \beta_2 = 0$ with $n_1 = n_2 = n/2$, the model based score test is $T_{GS}^m = \frac{n(\bar{y}_c - \bar{y}_t)^2}{4\bar{y}(1 + \tilde{a}\bar{y})}$ where $\tilde{a}$ is the null hypothesis estimator and $\bar{y}_t$, $\bar{y}_c$ are the sample means for treatment and control. A model based score test statistic is calculated for each of the estimation methods described above. The empirical score test statistic for the same setting is $T_{GS}^e = \frac{n^2(\bar{y}_c - \bar{y}_t)^2}{4 \sum_{total} r_i^2}$ for all methods where $r_i = y_i - \bar{y}$ are the residuals. The model score test statistic with $\tilde{a}$ estimated via OQ Estimation gives the same test statistic as the empirical score test. (For more details on Score Tests, see Appendix E).

CHAPTER 4

SIMULATION STUDY

A simulation study was conducted with two objectives: (i) compare estimators of the NB variance parameter and (ii) compare two-sample test methods.

Study Methods

Data generated under the null hypothesis (H_0) are generated from a single population setting. Data generated under the alternative hypothesis (H_1) are from two populations with different means but a common variance parameter. Each sample is comprised of two subsamples with $n_1=n_2=n/2$.

For simulation under the null hypothesis, a 3x3x2 factorial design was used. The first factor was the NB mean (3 levels: μ=1,2,5), the second factor was the NB variance parameter (3 levels: a=.2,.5,1) and the third factor was sample size (2 levels: $n_1=n_2$=15, $n_1=n_2$=25). One reason for this choice of levels is that they overlap with those from previous simulation studies that compared estimation methods for the NB variance parameter (Willson et. al. (1984), van de Ven (1993), Clark and Perry(1989), Piegorsch (1990)). A second reason for the choice of levels is due to applicability to analyses of drinking water and disinfection studies in environmental microbiology. Sample estimates of means and NB variance parameters reported in the literature often fall within the ranges of the levels chosen. The range of $\tilde{\mu} \in [1,5]$ was reported in water quality studies [El-Shaarawi, Esterby and Dutka (1981); Pipes and Christian (1982); Pipes, Ward and Ahn (1977)]; and $\tilde{a} \in [.2,1]$ (or $\tilde{k} \in [1,5]$) was estimated from

microbiological count data (Pipes, Ward and Ahn (1977); Maul, El-Shaarawi and Block (1985); Maul and El-Shaarawi (1991)); $n \in$30-50 is about the maximum sample sizes in environmental microbiology. I noted that the chosen parameter ranges were more representative of coliform counts or other microbes occurring at low density. I found estimates of negative binomial parameters reported in the literature that were outside this range (eg: means in Maul, El-Shaarawi and Block (1985); variance parameter in Pipes and Christian (1982), Christian and Pipes (1983)). The estimate values outside the ranges were due to microbes that commonly occur at moderate to high concentrations and due to the greater spatial and temporal variability that often occurs in environmental sampling as opposed to a controlled laboratory setting.

For simulation under the alternative hypothesis, the mean used to generate one of the subsamples was kept at the null hypothesis setting and the mean of the other subsample was varied for the desired power comparisons. I shall refer to both the choice of parameters and (sometimes) the set of samples generated for that choice using NB($\mu_1=\mu_2=\mu,a$) or NB($\mu_1>\mu_2,a$) as the "H_0 setting" or "power setting" respectively.

The computing environment used was S-PLUS version 3.2. Random NB counts were generated as random Poisson variates with means generated as random gamma variates with the desired shape and scale. Thus, the NB counts are generated as a gamma mixture of Poissons. The random NB generator available in S-PLUS was based on the Pascal form that uses only integer k ($=a^{-1}$) and I wrote a routine for general k. The estimation and testing routines were built using S-PLUS commands. I noted that Venables (1994) has written S-PLUS code to fit and analyze NB GLMs for a nested sequence of models using analysis of deviance (LRT).

The variance parameter was estimated using the single population assumption (under H_0) for all methods. Because use of the generalized score tests does not require estimation under the alternative hypothesis, the 2-sample variance parameter estimate was not calculated for several methods and comparisons were made on the single population (H_0) estimates.

The estimation methods are listed below along with their acronymns and the page numbers on which they are described.

Estimation Method	Acronym	Description
Maximum Likelihood	ML	pg. 12
Extended Quasi-Likelihood	EQL	pg. 14
Pseudolikelihood	PL	pg. 18
Optimal Quadratic	OQ	pg. 20
Conditional Maximum Likelihood	CML	pg. 23
Negative Exponential Disparity	NED	pg. 25

The estimation methods are rated according to the size of the errors of the estimates ($\tilde{a}-a$) calculated under the null hypothesis. The criteria for comparison of the estimation methods are average bias, mean square error (MSE) and means of root squared errors (MRSE) as a robust alternative to MSE. The MRSE is discussed in the next to last section of this chapter. The differences in MRSE are tested via simultaneous confidence intervals using the "all possible contrasts" form of Hotelling T^2 (Seber 1984, section 3.4; Miller 1966, p. 197; Scheffé 1959, section 3.5).

The two-sample tests of the null hypothesis $H_0: \beta_2 = 0$ versus $H_1: \beta_2 > 0$ (or $H_0: \mu_1 = \mu_2$ versus $H_1: \mu_1 > \mu_2$) are listed below with acronyms and the page numbers on which they are described.

Testing Method	Acronym	Description
Generalized Score Tests		pg. 36
ML Score test	ML.S	
EQL Score test	EQL.S	
PL Score test	PL.S	
OQ Score test	OQ.S	
CML Score test	CML.S	
Other tests		
NED Disparity Difference test	DDT	pg. 33
Generalized Likelihood Ratio test	LRT	pg. 33
Welch Modified 2-sample t-test	t	pg. 35

The criterion for comparison of testing methods is based on the power of the test. The Welch t-test has an approximate t distribution. All tests other than the t rely on asymptotic chi-square distributions. Table 1 (p. 95) lists the rejection rates of the testing methods at the H_0 settings using the asymptotic distribution critical values. The t was compared to a critical value of $t(\alpha=0.05)$ with degrees of freedom calculated by formula. The others are compared to a critical value (cv) of $\chi^2(df=1, \alpha=0.1) = 2.7055$; H_0 was rejected if the test statistic was larger than the critical value and $\tilde{\mu}_1 > \tilde{\mu}_2$ and accepted otherwise. All tests are remarkably close to the nominal 0.05 level. In general, LRT is the most liberal and DDT the most conservative. No test is the overall winner in achieving the nominal rate using the asymptotic critical value.

Adjustments were made to the critical values in order to make the test size the same for all methods at the null settings. The critical values were determined empirically to make the test size or alpha level equal to $\alpha = 0.05$. It would be possible to follow the work of McCullagh and Nelder (1989, pp. 459-463; general derivations in McCullagh and Cox, 1986) to obtain a Bartlett

correction factor for the LRT statistics to make them more closely chi-square and use a chi-square critical value. However, I found no similar adjustments in the literature for the other methods. I followed the commonly used practice of setting empirical critical values.

I generated (N=) 10000 samples (of size n) at each H_0 setting. I used as the critical value for a testing method that value of the test statistic which produced a rejection rate of $\alpha = 0.05$ for the samples generated under the H_0 setting [less the "non-NB" samples (defined below) for that testing method]. The empirical critical values are listed in Table 2 (p. 96).

I generated (N=) 5000 samples (of size n; comprised of two subsamples of size $n_1=n_2=n/2$) at each of the power settings. Differences in power are tested using the Generalized Estimating Equation (GEE) approach (Zeger, Liang and Albert, 1988) performed on the indicator matrix representing the outcomes of testing method results. The outcomes are 0 or 1 depending on whether H_0 was accepted or rejected for the 2-way array of samples by testing methods.

An advantage of using the Hotelling T^2 for MRSE or the GEE approach for power comparisons is it capitalizes on the high degree of correlation between the different estimators or testing methods results on individual samples. A disadvantage is the complication due to "non-NB samples", a sample for which one or more of the estimators or tests was not calculable (discussed in next section).

I used tables to display the actual values of comparison criteria for the different methods. Tables containing Hotelling T^2 results on MRSE or GEE results on power are coded to show statistically significant differences. I used

plots to show general trends from the result tables and to distinguish between significant results and practically important results.

Estimators and Tests when $\tilde{a} \leq 0$

During the generation of samples, I discarded "underdispersed" samples and generated new samples until a total of 10000 samples were obtained. Underdispersed samples are those for which $S^2 < \bar{y}$ where $\bar{y}$ and S^2 are the sample mean and variance. There is no (real) root to the ML (variance parameter) estimating equation for underdispersed samples (Aragon et. al., 1992). Both PL and OQ produce negative estimates of the NB variance parameter for underdispersed samples and it is questionable whether the remaining estimates (based on CML, EQL, NED) would be positive. When a sample is underdispersed ($S^2 < \bar{y}$), one might believe that fitting the NB is not appropriate. Other simulation studies reported discarding or replacing these samples [Pieters et. al. (1977), Wilson et. al. (1984), Van de Ven (1993)] or assigned the value zero to the NB variance parameter estimate [Anraku and Yanagimoto (1990)]. Underdispersed samples can occur by chance, especially when generating samples from NB distributions with small means and variance parameters (Wilson et. al., 1984). The numbers of underdispersed samples from my simulation settings are listed in the right-most column (of Table 3 and Table 4, pp. 97-98) under the heading $S^2 < \bar{y}$. The numbers of samples discarded increase as a and μ decrease. The proportions of samples affected agree with those of Wilson et. al. (1984, p. 110-111) and Clark and Perry (1989, p. 313).

Even when $S^2 > \bar{y}$ it is possible to obtain negative estimates, $\tilde{a} \leq 0$. The variance parameter of the NB distribution is defined for positive values only. If

an estimation method failed to produce an NB solution, i.e.: estimation produces $\tilde{a} \leq 0$, I refer to the estimates as "non-NB" estimates.

Some authors did not discard any samples with negative estimates for a or k and included the negative values in estimates of bias and MSE. Clark and Perry (1989) allow negative values for the solutions to PL and EQL estimating equations which arise solely due to sampling variation. They cite work by Binet (1986) who obtained (positive binomial) estimates using the analogy between the negative and positive binomial distributions and by Ross and Preece (1985) who worked with a setting in which a negative MLE of k leads to a valid censored probability distribution. Piegorsch (1990) allowed negative MLEs for a but restricted the range, for numerical stability, to $\tilde{a}_{ML} > -1/\max(y)$ where max(y) is the largest observed value from the sample.

I chose to interpret negative variance parameter estimates as evidence supporting a Poisson distribution and assigned the value $\tilde{a} = 0$ when testing comparisons on all samples. I did not pursue theory and tests for non-NB estimates related to a (positive) binomial. Recall that the Poisson is the limiting distribution of the NB(μ,a) as $a \to 0$. The pmfs of a Poisson and NB with small a are almost indistinguishable. Figure 4 (p. 129) illustrates this point by showing a sequence of NBs pmfs with the same mean as a Poisson converging to the Poisson as a decreases. The estimating equations for ML, CML, EQL, and NED involve taking logs of quantities which may become zero or negative (numerical instability). I did not find instability to be a problem for ML, CML or EQL estimation under H_0. Negative estimates converged to values within the range used by Piegorsch. Thus I was assigning $\tilde{a} = 0$ for negative estimates with values close to zero. NED produced many more non-NB estimates under H_0 and values often further from zero, and became unstable or divergent in the

negative direction. When I investigated a subset of the cases where instabilities occurred and employed slower, but more reliable numerical analysis techniques, I found that the majority (>90%) of the estimates became very small positive estimates.

The other kind of instability was when a diverged toward infinity. Only NED estimates exhibited this type of divergence (estimation was considered divergent and stopped after $\tilde{a} \geq 1000$). The divergence ($\tilde{a} \geq 1000$) was much less frequent than non-NB estimates ($\tilde{a} \leq 0$); it amounted to less than 8% of the overall difficulties associated with NED (more frequently it was less than 2% or did not occur) at any simulation parameter setting. Again, when numerical analysis techniques were applied, the vast majority of these divergent estimates converged.

When conducting the two-sample test, I performed each test based on either the NB or the Poisson distributions depending on whether NB or non-NB variance parameter estimates were obtained. For LRT and DDT, the two-sample test required obtaining ML or NED estimates for the NB variance parameter under H_1. There were non-NB estimates produced under H_1 for samples that had produced NB estimates under H_0. Estimation under H_0 is based on a single sample of size n=30 (or 50) but estimation under H_1 is based on two samples of size n_1=n_2=15 (or 25). The difference in sample size was a factor in occurrance of non-NB estimates (under H_1). Included in the (H_1) estimation algorithms for ML and NED was a grid search for better starting values if the first estimation attempt did not produce NB estimates. I did not investigate the (H_1) ML estimates that did not produce NB estimates to determine if any were due to divergence. The incidence of divergence for NED under H_1 were at least as small as those under H_0. I did not update any NED

non-NB estimates to reflect the results of numerical analysis techniques, but recorded the numbers and types of non-NB estimates that occurred.

Tabulations of all types of non-NB estimates are found in Tables 3 and 4 (pp. 97-98) for simulations at H_0 and power settings respectively. The tabulations of non-NB estimates produced under H_0 estimation for the various estimation methods are in the columns under the heading of "$\tilde{a} \leq 0$ (H_0)". Note that after underdispersed ($S^2 < \bar{y}$) samples are discarded (right-most column of table), PL estimation cannot produce non-NB estimates. I included (H_0) NED divergence failures in the same category with its non-NB estimates. A Poisson test was performed whenever non-NB estimates were obtained under H_0 or H_1. For LRT and DDT the additional non-NB estimates produced under H_1 (for samples that produced NB estimates under H_0) are listed under the heading "$\tilde{a} \leq 0$ (H_1)". I was unable to perform either an NB or Poisson test for DDT on a few samples. For these samples the NED mean estimate (assuming Poisson distribution) for one of the subsamples converged to zero (indicating $P(Y=0)=1$). These counts are listed in Tables 3 and 4 under the heading "no tests". Some LRT non-NB estimates under H_1 estimation were identified but the samples had not been saved. These results were listed under "no tests" (Table 4 only) because I was unable to determine whether they would have produced an NB test with new starting values or would have produced a Poisson test. The (Welch) t-test always produced a result and is indicated as such by listing zeroes under "no tests".

For statistical comparisons, I distinguish between the following sets of samples (out of N samples of size n generated at a setting): the set of samples where all methods produced NB results (I shall refer to these as "NB samples"), and the set of <u>all</u> samples but using $\tilde{a} = 0$ for non-NB estimates and Poisson

tests where non-NB estimates were obtained (I shall refer to these as "all samples" or "NB+P samples"). MRSE and power comparisons were performed separately on NB samples and NB+P samples.

Choice of transformation (MRSE).

Because the Hotelling T^2 is based on the multivariate normal distribution, I used a transformation which (i) produced symmetric histograms and boxplots and (ii) the best looking normal quantile-quantile plots (qq plots) for the combined error distribution and (iii) produced error distributions of (approximately) the same shape for each estimation method. Seber(1984) reports that Hotelling T^2 is more sensitive to skewness than kurtosis. I wanted to use a single transformation, if possible, that would work for all data sets so that sizes of (mean) errors could be compared between the various parameter settings. I followed up on my choice by determining the Box-Cox transformation for the sets of samples at the various parameter settings (discussion below).

I considered the following transformations on squared errors (sq.er) or absolute errors (abs.er): sq.er, abs.er, log(sq.er +c) for c=.5, .1, .01, .001, .0001, and root(sq.er) for root=1/5, 1/7, 1/9. The constants in the log transformations were added to avoid the usual problem of infinite results for errors that are essentially zero. The vast majority of errors fall in the range of (-5.0,5.0). For values in that range the transformations have different effects on errors relative to their size. They magnify or reduce the differences between "similar errors", or errors within the same small intervals, depending on the interval's distance from the origin. These transformations in intervals close to and more distant from zero are displayed in Figure 5 (p. 130). As expected, squared errors magnify the errors that are distant from zero and the log transformations magnify the differences between errors in the interval about zero with the stretching or

magnification increasing as c approaches zero. The root transformations have a more moderate effect on errors close to zero and a stronger leveling effect on errors far from zero. They also produce more symmetric histograms for the data from the range of simulation settings that I used. In Figure 6 (p. 131) are histograms of transformed sq.er for all the transformations considered. The data are the set of errors, combined across all six estimation methods, for one of the simulation settings (NB(μ=1,a=.2)). Thus, there are 6 x 10000 (NB+P) results in the combined set. I had originally wanted to perform contrasts on MSEs but decided against it because sq.er's produced extremely skewed distributions and mean values were most influenced by the few large errors. The symmetry of the distributions for the various log transformations was effected by the choice of the constant added to the errors. The smaller the constant, the more symmetric the histogram. The root transformations produce the most symmetric histograms, but with their restricted positive range and stubby tails they fall short of appearing normal. Among these choices of transformations, I decided to go with $(sq.er)^{1/7}$ but I suspect that the test results from any of the root transformations would be essentially the same. Henceforth, I shall refer to $(sq.er)^{1/7}$ as root square error (RSE) and its mean value as MRSE. In Figure 7 (p. 132) are the histograms of RSE broken out per estimation method (example from one simulation parameter setting (NB(μ=1,a=.2))). Note their similarity in shape.

Figures 8 (n=30) and 9 (n=50) (pp. 133-134) contain box plots of RSE per estimation method for all the H_0 simulation settings. The y-axis is the same in all plots so the viewer can compare the range for RSE at the various simulation settings and samples sizes. The box plot features are standard. The box encloses the interquartile range with an inner line at the median, whiskers

(dashed line) extend to 1.5 times the inter-quartile range and data outside that range plotted as dots. The MRSE is indicated by an "x". The intent of the box plots is to confirm that the RSE transformation achieves symmetry at all settings for all estimation methods prior to testing differences using Hotelling T^2.

In order to see how the RSE compared to the "best" transformation, I applied the Box-Cox procedure. The Box-Cox transformation determines the MLE of the power transformation to produce constant normally-distributed residuals (Box and Cox, 1964; Box and Draper, 1987). The "best" (Box-Cox) transformation per parameter setting resulted in power (λ) in the range $[1/20 < \lambda < 1/5]$. For a majority of the parameter settings used, λ=1/7 (RSE) was included in the 99% confidence interval on λ. When I retested those parameter settings with "best" transformations furthest from RSE, the results of the Hotelling T^2 using the "best" λ produced the same (significance) results as those using RSE. In order to compare sizes of errors across parameter settings, all results will be presented in terms of RSE and MRSE.

GEE Analysis of Power Results

Differences in power are tested based on the generalized estimating equation (GEE) approach outlined in Zeger, Liang and Albert (1988). The power results are an indicator matrix representing the outcomes of the testing methods. The outcomes are 0 or 1 depending on whether H_0 was accepted or rejected for the 2-way array of samples by testing methods. As might be expected, there is a high degree of correlation between the different testing methods results on individual samples. The rows (samples) of the matrix are independent and the columns (test methods) are a set of correlated Bernoulli trials (see a summary of the correlation structure in Tables 21 and 22, pp. 115-

116). The GEE approach is a extension of GLMs for data displaying dependence (data with correlated outcomes per subject; the paper by Zeger et. al. (1988) was presented in terms of longitudinal data). The interest is on the power over the complete set of samples or a "population-averaged" GEE model.

For the power result matrix from a set of d testing methods applied to N independent samples $T_{N\times d}=\{t_{ij}\}$, the expected or mean response is the same for every row, $E(T_i)=E(t_{i1},\ldots,t_{id})^t=\tau_{d\times 1},\forall i$, and the variance function $g()$ is $Var(t_{ij})=g(\tau_j)=\tau_j(1-\tau_j),\forall i$. The means for the binary power outcomes are modeled with a logit link i.e., $\beta_{ij}^*=h(\tau_j)=\text{logit}(\tau_j)=\log\left(\frac{\tau_j}{1-\tau_j}\right),\forall i$. Thus, $\tau_j=\frac{e^{\beta_j^*}}{1+e^{\beta_j^*}}$. The estimates of β^* are obtained by solving the GEE $U(\beta^*)=\sum_{i=1}^{N}\frac{\partial\tau^t}{\partial\beta^*}V_i^{-1}(T_i-\tau)=0$. The covariance structure for each row is the same and estimated by $\tilde{V}_i=\tilde{V}_{d\times d}=\tilde{A}^{1/2}R\tilde{A}^{1/2}$, where $\tilde{A}=diag\{\tilde{\tau}_j(1-\tilde{\tau}_j)\}$ and R is the sample correlation matrix of the power results. Liang and Zeger (1986) show that $\tilde{\beta}^*$, the solution to the GEE is consistent and asymptotically $(N\to\infty)$ Gaussian given only correct specification of the mean and the usual regularity conditions. A consistent variance estimate is $V_{\tilde{\beta}^*}=M_0^{-1}M_1M_0^{-1}$, where

$$M_0=\sum_{i=1}^{N}\frac{\partial\tilde{\tau}^t}{\partial\beta^*}\tilde{V}_i^{-1}\frac{\partial\tilde{\tau}}{\partial\beta^*}\text{ and }M_1=\sum_{i=1}^{N}\frac{\partial\tilde{\tau}^t}{\partial\beta^*}\tilde{V}_i^{-1}(T_i-\tilde{\tau})(T_i-\tilde{\tau})^t\tilde{V}_i^{-1}\frac{\partial\tilde{\tau}}{\partial\beta^*}.$$

For the power result matrix, $\frac{\partial\tilde{\tau}}{\partial\beta^*}$ is invertible ($\frac{\partial\tilde{\tau}}{\partial\beta^*}=\tilde{A}$). The GEE simplifies to $U(\beta^*)=\sum_{i=1}^{N}A^{1/2}R^{-1}A^{-1/2}(T_i-\tau)=0$, hence $U(\beta^*)=0=\sum_{i=1}^{N}(T_i-\tau)$ or $\tilde{\tau}_j=\overline{T}_j$ and $\tilde{\beta}_j^*=\text{logit}(\overline{T}_j)$ the logit of the power (empirical rejection rate) of the j^{th} testing method. The variance estimate is $V_{\tilde{\beta}^*}=\frac{N-1}{N^2}\tilde{A}^{-1}S\tilde{A}^{-1}$ where $(N-1)S=\sum_{i=1}^{N}(T_i-\tilde{\tau})(T_i-\tilde{\tau})^t$ and $\tilde{A}=diag\{\tilde{\tau}_j(1-\tilde{\tau}_j)\}$. Contrasts between the power

of the testing methods, on the logit scale, are based on Z-statistics, $Z = C^t\tilde{\beta}^* \Big/ C^t V_{\tilde{\beta}^*} C$ where C is the vector of contrast coefficients. A simultaneous (Bonferroni) α-level test of all possible contrasts between d methods are based on whether the Z-statistics fall into the associated acceptance regions, $\left[\Phi^{-1}_{\alpha/2k}, \Phi^{-1}_{1-\alpha/2k}\right]$, where $k = \frac{d(d-1)}{2}$, and Φ is the normal cumulative distribution function.

CHAPTER 5

RESULTS

I used tables to display statistically significant differences. I used plots to show general trends and to distinguish between significant results and practically important results.

Results for Estimation Methods

Differences between methods used to obtain estimates of the NB variance parameter were tested in terms of means of root squared errors. General trends in size of MRSE for the methods are shown in Figure 10 (p. 135) for samples generated at the H_0 settings. The simulation settings for μ are listed along the x-axis and values of MRSE along the y-axis. Results are plotted using a different symbol for each of the six estimation methods. Three different line types are used to connect a method's MRSE results for a given simulation setting of a across the settings of μ. Separate plots are presented for sample sizes of n=30 and n=50. For all estimation methods at all settings, smaller MRSE is achieved for n=50 (plot B) than for n=30 (plot A). The set of lines for the different settings of a fall in separate bands with no overlap between line types on a plot. All the lines (and bands of line types) are decreasing. Thus, for any method, the estimation errors decrease for a given setting of the variance parameter as the mean setting increases and for a given setting of the mean as the variance parameter setting decreases. The relative performances of the different estimation methods is variable, with no overall winner or loser.

The Tables 5-8 (pp. 99-102) display results of Hotelling T^2 test comparing MRSE for NED versus the other estimators. Each table is partitioned into two sections by double bold horizontal lines. The results for the H_0 settings are listed in the top section and results for the H_1 power settings are in the bottom section. I chose to separate the table into two subsections because the performance of NED is different for H_0 data than for H_1 data. Recall that the MRSE comparisons are being done on the variance estimates obtained under the H_0 assumption. Thus a comparison of performance where that assumption is true (the H_0 settings, as summarized in Figure 10, p. 135) is typically all that would be reported. I observed that the performance of NED did not follow the pattern of the other estimators at the H_1 settings and included the power setting results because I thought them of interest (discussion below).

Results for different sample sizes (n=30 or 50) and sets of samples included (NB or NB+P samples) are found in different tables. The values listed in the tables are the MRSE per estimation method (columns) for each simulation setting (rows). Thus, each row represents the results from a Hotelling T^2 test for the listed simulation setting. The MRSEs are coded so that significant differences at overall α=0.05 can be seen at a glance. The MRSE for NED, and all estimators not significantly different from NED, are listed in italics and shaded in gray. Any MRSEs that are significantly smaller (better) than that for NED are listed in bold and underlined. Any MRSEs listed in plain text are significantly larger (worse) than those for the NED method.

For NB samples of size n=30 (Table 5) or n=50 (Table 6) of the H_0 settings (top section of tables), the NED method performs moderately well. Here NED is significantly better than at least one other method in 3 out of 9 settings and performs at least as well as one other method in 7 out of 9 settings.

NED is significantly worse than all other methods for 2 out of 9 settings and outperformed by at least one other method in all but 1 out of 9 settings for each sample size. When looking at MRSE for NB+P samples results are similar. There is general agreement in MRSE entries to two significant digits between the tables at the same sample size (Tables 5 and 7, Tables 6 and 8). Even so, the small changes in MRSE due to including samples with at least one non-NB estimate (NB+P) for such a large set of samples (N=10000) results in a general decline in performance for the NED. Recall that the NED method had the largest number of non-NB estimates for all simulation settings (Tables 3 and 4; pp. 97-98). Thus it received the largest number of estimates with $\tilde{a} = 0$.

ML has significantly smaller MRSE (NB+P samples) than all the other methods for 5 out of 9 of the H_0 settings for samples of size n=30 and for 8 out of 9 settings for samples of size n=50. The methods that are significantly better than ML for some H_0 settings are PL (at 4 settings for n=30; at one setting for n=50) and OQ (at 3 settings for n=30; at one setting for n=50). MRSE for CML is not significantly different from that of ML at 2 out of 9 H_0 settings for samples of size n=30 and 5 out of 9 settings for samples of size n=50.

The performance for NED is dramatically improved in the same tables under the power settings (bottom section of tables). NED has MRSE at least as small as all others for all but 3 out of 24 power settings and all but 2 out of 26 power settings for samples of size n=30 and n=50 respectively. This is an example of estimation under a misspecified model. The samples are generated from two populations with a common variance parameter but estimation takes place assuming a single population. (The means of the two populations are different, hence the population variances are different even though the variance parameter is the same.) We compare estimates of the variance parameter

when the mean structure has been misspecified. The mixing of samples from two populations does not drastically change the probabilities of small counts but increases the probability of a few of large counts. In a sense this may provide a look at robustness in estimation. The difference between NED and the other methods in terms of (MRSE) errors of estimating the variance parameter can be seen in plots which display results for both the H_0 and power settings. Figure 11 (n=30) and Figure 12 (n=50) (pp. 136-137) present the totality of the MRSE results for NB samples in a matrix of plots displaying the factorial design at the H_0 setting. The matrix is organized with the levels of a in rows and levels of μ in columns. Different line types are used to connect the results per method. The actual settings for which MRSE was calculated are only at (a subset of) the "percent increase in μ" values listed across the x-axis. The actual settings and values are listed in Tables 7 and 8. The "0% increase" mark indicates the MRSE result at the H_0 setting. The legend for the line types connecting the MRSEs for an estimation method across the settings is found in the upper left plot. The MRSE lines, with the exception of the one for NED, appear parallel across the settings and show that MRSE increases with percent increase in μ. The NED line stands out because it drops below the other lines at the higher values of percent increase in μ. The percent increase in μ indicates the departure of the simulation setting from the H_0 ($\mu_1=\mu_2$) assumption being used to calculate the sample estimates. The NED method which is purported to be robust or less sensitive to outliers and inliers provides the smallest variance parameter estimates in these settings. The tendency to be biased low seems to be the advantage in simulations with parameters set at small positive values. The plots for NB+P samples would be essentially the same due to the relatively small differences in means (same values to two significant digits).

Plots also provide a visual impression of the sizes of differences in MRSE results. There is a relatively large range of root square errors for each method compared to much smaller differences in the means (MRSEs). Refer back to the RSE histograms for NB+P samples in Figure 7 (p. 132). There is a histogram of the RSEs for each estimation method and the MRSEs plotted with an X along the x-axis. All the histograms are plotted with the same break points and y-axis so the shapes and heights are directly comparable. All the MRSEs are just to the left or right of the break point at 0.6. The similarity of the shape and spread of the histograms and the small differences between MRSEs indicate the impact of the large set of samples (N=10000) and correlation of results in detecting significant differences. Similar information is conveyed in the box plots of Figures 8 and 9 (pp. 133-134) for all the null simulation settings.

The degree of correlation of RSE across samples between the different estimators is summarized in Tables 9 (n=30) and 10 (n=50) (pp. 103-104). The average, maximum and minimum correlation values across all the settings for MRSE Hotelling T^2 test analyses on NB samples are listed. Note the high values for PL with OQ, and for EQL with CML and ML. NED has the lowest correlation with any other estimator.

Numerical values for Bias and MSE based on the same data sets as for MRSE are found in Tables 11-14 (pp. 105-108) and plots found in Figures 13-16 (pp. 138-141). This presentation is included because bias and mean square error are often the measures used to evaluate estimation methods. A horizontal line at the origin is added to the Bias plots. NED tends to underestimate the variance parameter yielding negative bias. Bias in estimates of a decrease as the setting for μ increases, except when a=1. The plots of MSE appear quite similar to the plots of MRSE; the most notable exception being that there are

larger differences between the methods based on MSE than MRSE for settings with a=1.

Results for Testing Methods

For all the H_0 settings, simulations were run to determine the power to detect a 100% increase in the mean of the treatment subsample. Other increases, in even 25% or 50% incremental settings above the null setting, were used. The choice of the settings was driven by the intent to see how small (approximately) a percent increase could be detected without dropping below 50% power and how large (approximately) a percent increase was necessary to achieve a minimum of 80% power by all testing methods. In some settings the t-test and Disparity test were performing poorly and the settings were stopped after the LRT and all the score tests achieved 80% power.

The general performance trends across all tests can be seen in plots. Figure 17 (p. 142) summarizes the power to detect a 100% increase in the mean. The null settings for μ are listed along the x-axis and power plotted along the y-axis. The power values are indicated by plotting symbols, with a separate symbol for each test. Three different line types are used to connect a test's power results for a given value of a across the settings of μ. Separate plots are presented for sample sizes of n=30 and n=50. Out of the set of score tests, only the PL.S score test is plotted. All of the other score tests produce power lines which criss-cross between or are indistinguishable from either the LRT or PL.S lines. For all tests at all settings, better power is achieved for n=50 than for n=30. The smaller the value of the variance parameter, the less difference in power between the testing methods. The power lines for the various methods are closer together for a=.2 and more spread for a=.5 and a=1.

For n=50 the range of power for a=1 is visibly greater than for a=.5 at every setting for μ. The set of lines for the different values of a for the various testing methods fall in separate bands, as there is almost no overlap between line types on a plot. Thus, the power to detect a difference increases as the variance parameter decreases from a=1 to a=.2. For each setting of a, the lines representing LRT and PL.S are above those for t and DT. The only exception is for n=30, a=.2 and μ=1. Almost all the lines are increasing. Thus, for each testing method at a given setting of variance parameter, power to detect a 100% difference in μ improves as μ gets larger. The only exception to this pattern is for DDT at a=1.

Figure 18 (n=30) and Figure 19 (n=50) (pp. 143-144) present the totality of the power results in a matrix of plots displaying the factorial design at the H_0 setting, with the levels of a in rows and levels of μ in columns. Lines connect the points at which power was calculated to aid the viewer. The power calculations were made at only at (a subset of the) settings labeled as "percent increase in μ" along the x-axis. The actual settings and values are listed in Tables 15 and 16 (pp. 109-110). The legend of line types for the various testing methods is in the upper leftmost plot but the lines of tests with only small power differences blur together. In each plot, the lines for the score tests and LRT are uppermost and close together, often indistinguishable. In most plots the lines for t and DDT run some distance below them, indicating a practical drop in power. The total spread between the lines is smallest in the upper row of plots and greatest in the bottom row. Again, this demonstrates that there is less difference in power achieved by the testing methods at smaller values of a. All the lines are increasing. Thus for all tests and all null settings used, power increases with increased difference between the means of the two subsamples. The plots in

the top row have lines beginning further to the left, at smaller percent increases in μ. Looking down a column (a given value of H_0 setting for μ) and at a fixed percent increase in μ, tests achieve greater power for smaller a. Looking across a row from left to right (at a given value of H_0 setting for a, across increased (H_0) settings of μ), greater power is achieved by the tests at a given "percent increase in μ". This is seen as a shift upwards for a given position along the x-axis by the top group of lines if not all lines.

To understand the relative difficulty of distinguishing between NBs with different means and a common variance parameter as influenced by the settings for a and μ, consider the model pmfs from which the samples were generated. The pmfs for generating the 2 subsamples, each of size n/2, used to determine the power to detect a 100% increase in μ are plotted in Figure 20 (a=.2) Figure 21 (a=.5) and Figure 22 (a=1) (pp. 145-147). The plots show the degree of similarity in the probabilities of producing sample counts. The pmf's on a plot appear quite similar and overlapping but the degree of similarity depends on the settings for a and μ. For a given value of a, the larger μ, the greater the differences in the heights of the NB pmf's especially as regards probabilities of larger counts. For a given value of μ, the larger the value of a, the smaller the differences of the heights of the NB pmf's are. For a simple measure of the differences in the pmf's, see Figure 23 (p. 148) for the cumulative sums of the absolute differences in heights. This gives some rational for the differences in power performance for the tests. The settings with larger cumulative differences yield greater power for all tests. Printed on the pmf plots in Figures 20, 21, and 22, are the empirical power values for a selection of the 2-sample tests performed on the N=5000 samples [less the non-NB samples] generated at sample sizes n=30 and n=50. The relative size of the

power to detect a difference between the subsamples at different settings follows the size pattern of cumulative differences found in Figure 23.

The results of the significance tests for differences in power between tests are found in Tables 15-20 (pp. 109-114). The results are based on the GEE "all possible contrasts" test performed on the indicator matrix representing the outcomes of test results. The outcomes are 0 or 1 depending on whether H_0 was accepted or rejected for the 2-way array of samples by tests. Thus the analysis takes advantage of the high degree of correlation between the different tests' results on individual samples. A summary of these correlation matrices across all the power settings is found in Table 21 (n=30) and Table 22 (n=50) (pp. 115-116). Note the high correlations between the score tests and LRT. DDT has the lowest minimum and average correlations with any of the other tests, but its maximum correlations are higher than those of the t-test. The t-test is most highly correlated with the DDT. DDT has a more variable performance across the settings and its highest correlations are with tests other than the t-test.

If any testing method did not produce an NB test for a particular sample, that sample was not included in the comparison of NB power results due to the missing value. Thus the NB power comparison is done over the subset of samples where all methods produced NB results (Tables 15, 16, and 19). The NB+P power comparisons includes all those samples producing a result (NB or Poisson test) for all testing methods (Tables 17, 18, and 20). (For more discussion, refer back to section titled Estimators and Tests when $\tilde{a} \leq 0$).

The power results in Tables 15-18 compare DDT to the other tests. DDT performs poorly. It has significantly less power than the LRT for all power settings and significantly less power than all the score tests for a=0.5 and a=1

and for a=0.2 when μ_2>1. It has significantly more power than the t and a subset of the score tests for a few of the settings with a=0.2.

Tables 19-20 compare LRT to the t and Score tests. The LRT is always significantly more powerful than the t. It is significantly more powerful than the score tests at most settings for n=30. However, the majority of the settings at n=50 show the score tests as not significantly different than the LRT. Comparisons of Score tests to the t-test for the samples as in Tables 19 and 20 show that all the score tests are always significantly more powerful than the t. Among the score tests, PL.S is at least as powerful as the other score tests at 10 out of 24 settings for samples of size n=30 and at 22 out of 26 settings for samples of size n=50. The methods that are significantly more powerful than PL.S include one of more of (EQL.S, ML.S, CML.S) for 14 out of 25 settings at n=30 and for 4 out of 26 settings at n=50.

A summary of the similarity of performance between the testing methods is found in Tables 23-30 (pp. 117-124). The tables list the unique rows of power result matrices for 5 tests (EQL.S, PL.S, DDT, LRT, and t) at all the power settings. The left section of each table indicates the rejection patterns (0=accept and 1=reject) listing all possible outcomes for different combinations of the testing methods. The remainder of the table tabulates the occurrences of that pattern, one column of counts for each simulation setting. Table 23 contains the power settings with variance parameter a=.2 and samples of size n=30. For these settings all the tests perform the same over 75 percent of the time (see right column; all accept is 24.54% and all reject is 56.89%). Looking at the totals (right column) for Table 23, the tests that are individually most different are DDT which rejects alone a total of 480 times and t which accepts when the other four reject 1509 times. The combinations of 2 and 3 tests rejecting together (or

accepting together) most frequently are DDT with t and LRT with the score tests (EQL.S and PL.S). DDT and t reject together when others accept 262 times and LRT, EQL.S and PL.S reject together when DDT and t accept 823 times. Some rejection patterns don't occur at these settings. LRT never accepts when all others reject and EQL.S never rejects alone. Tables 24 to 30 exhibit similar patterns. The biggest differences are within the categories of "rejected by no tests" and "rejected by all tests" depending on the setting.

CHAPTER 6

CONCLUSIONS

The NED estimation method is intuitively appealing and has proven to be competitive in contaminated Normal models (Basu and Lindsay, 1994; Basu and Sarkar, 1994). However, I can not propose it as a simple, most efficient estimator for the NB variance parameter based on my simulation study. I concentrated on comparing estimation under a correct model where the ML method is expected to perform the best in terms of efficiency. The NED performed moderately well, performing significantly better than some of the estimators in terms of MRSE for some of the H_0 settings. It may not have been the best overall, but it stayed within the range of the others and there were no great practical differences in MRSE.

The testing procedure DDT based on NED produced low relative power. The fault may have been with the 2-sample estimator of a common variance parameter. Perhaps a technique that combines the estimates obtained separately from the 2 samples would be more appropriate. Bliss and Owen (1958) present a technique for weighting separate MOM estimates to get a combined estimate of a common variance parameter.

At present, the t-test is a common approach for testing differences between two populations for any type data. For my NB simulations the power of the t-test was dominated by all but DDT. I recommend against its use when the population means are as small as those in my simulations. The LRT is attractive because it is well understood. But the LRT fails to produce an NB test (due to

non-NB estimates) more often than the score tests and is computationally more intensive. An advantage of generalized score tests over the LRT is their standard null asymptotic chi-squared distributions (Boos,1992). The LRT under misspecification has a nonstandard null asymptotic distribution. It is generally distributed as a weighted sum of chi-squared random variables with unknown weights (Foutz and Srivastava, 1977; Kent, 1982).

The score tests are easy to compute, perform well and show little practical difference in power from the LRT for the sample sizes considered. Among the score tests, there was no clear evidence of a best estimation procedure to use based on RSE practical differences. I would recommend the PL and OQ estimators based on ease of computation. I would recommend EQL and CML over OQ and ML based on fewest failures.

Further Work

I intend to make my score test S-PLUS code available on the web for use by applied statisticians.

Interval estimates for the NB parameters or difference between the NB subsample means may be obtained by profile likelihood confidence intervals (Nelder and Pregibon, 1987) or inverting the test statistics. Details for those based on NED may be of interest and will be considered at a later time.

In the future I plan to perform a simulation study to compare the estimation methods of NED to ML and other robust methods (a subset of EQL, PL, CML, OQ) by looking at estimates of both the mean and variance parameter for the NB. I intend to use contaminated NB models to investigate the robustness of the various estimators. The results of estimation under a misspecified model, that is, H_0 estimation of the variance parameter at power

settings, indicated that the NED would be most robust. The biggest obstacle was that NED failed most often to produce an NB estimator of the variance parameter. Perhaps a method other than Newton Raphson, like a gradient or profile method, would have produced fewer failures in conjunction with NED. I did not investigate the effect of a larger sample size but expect that it would reduce the numbers of failures.

I would like to look at improvements to the DDT testing procedure for NB data and investigate them in robust settings. I am interested in additional problems deserving further investigation. A modification of the Bartlett correction factor which is used for the LRT or a similar derivation to make score tests more closely chi-square would enable the use of p-values rather than using empirical critical values. It may be that the t-test performed on log-transformed counts has acceptable power for samples from populations with large means. Instead of the log-transformation, one could use the variance-stabilizing transformation suggested by the Delta method. Extension of the simulation study to include larger means would provide insight into when it is appropriate to use the t-test on log-transformed data.

APPENDICES

APPENDIX A.

DERIVATION OF Q^+ FOR THE EQL METHOD.

To prove: Q^+, the extended quasi-likelihood, (or more correctly, the extended quasi log likelihood), is obtainable from the log likelihood of the Negative Binomial where all factorials z! are replaced by Stirling's approximation, $z! \approx (2\pi z)^{\frac{1}{2}} z^z e^{-z}$. I am using the notation as defined in Nelder and Pregibon (1987). It is consistent with that found in McCullagh and Nelder (1989).

Definition: The extended quasi-likelihood, denoted by Q^+, is $Q^+ = -\frac{1}{2}\log\{2\pi\sigma^2 V(y)\} - \frac{1}{2}D(y;\mu)/\sigma^2$, where $\sigma^2$is the dispersion parameter, $D(y;\mu) = -2\int_y^\mu \frac{y-u}{V(u)}du$ is the deviance function, and $\text{var}(y) = \sigma^2 V(\mu)$ (Nelder and Pregibon, p.223; McCullaugh and Nelder, p. 360).

Proof: (i) Show: $-\frac{1}{2}D(y;\mu)/\sigma^2 = y\log\left(\frac{\mu}{y}\right) - (y+k)\log\left(\frac{\mu+k}{y+k}\right) = Q + h(y,k)$.

For y from distribution NB$\left(\mu, k = a^{-1}\right)$ with fixed k,

$\text{var}(y) = \sigma^2 V(\mu) = \mu + \frac{\mu^2}{k} = \frac{\mu(\mu+k)}{k} = V(\mu)$, so $\sigma^2 = 1$.

$Q = -\frac{1}{2}D(y;\mu)/\sigma^2 = \int_y^\mu \frac{y-u}{\sigma^2 V(u)}du$. (definitions of Q, D)

Evaluating the integral using integration by parts and the indefinite integrals $\int \frac{dx}{x(ax+b)} = \frac{1}{b}\log\left(\frac{x}{ax+b}\right)$ and $\int \log(x)dx = x\log(x) - x$ yields:

$$Q = \int_y^\mu \frac{y-u}{V(u)}du = \int_y^\mu (y-u)\frac{1}{u\left(\frac{1}{k}u+1\right)}du = (y-u)\log\left(\frac{ku}{u+k}\right)\Big|_y^\mu - \int_y^\mu \log\left(\frac{ku}{u+k}\right)(-1)du$$

$$= (y-u)\log\left(\frac{ku}{u+k}\right)\Big|_y^\mu + \left\{[u\log ku - u]\Big|_y^\mu - [(u+k)\log(u+k) - (u+k)]\Big|_y^\mu\right\}$$

$$= y\log\left(\frac{\mu}{y}\right) - (y+k)\log\left(\frac{\mu+k}{y+k}\right) \qquad (1)$$

This is the same as McCullagh and Nelder's Q for NB in Table 9.1 plus a function of y and *k*, i.e.,

$y\log\left(\frac{\mu}{y}\right) - (y+k)\log\left(\frac{\mu+k}{y+k}\right) = y\log\left(\frac{\mu}{\mu+k}\right) + k\log\left(\frac{k}{\mu+k}\right) + h(y,k)$, where

$h(y,k) = k\log\left(\frac{y+k}{k}\right) - y\log\left(\frac{y}{y+k}\right)$.

(ii) Show $Q^+ = -\frac{1}{2}\log\left\{2\pi\sigma^2 V(y)\right\} - \frac{1}{2}D(y;\mu)/\sigma^2$ is equal to the log likelihood of y when factorials are replaced by Sterling approximations.

$V(y) = y(1+ay)$ is an empirical variance, expressed in terms of y rather than μ.

The log likelihood of y is:

$$\log\left[\binom{y+k-1}{y}\left(\frac{k}{k+\mu}\right)^k\left(\frac{\mu}{k+\mu}\right)^y\right] = \log\left[\left(\frac{(y+k)!}{y!k!}\frac{k}{(y+k)}\right)\left(\frac{k}{k+\mu}\right)^k\left(\frac{\mu}{k+\mu}\right)^y\right]$$

and, when substituing in Stirling's approximation for factorials, becomes:

$$\approx \log\left\{\frac{[2\pi(y+k)]^{1/2}(y+k)^{y+k}e^{-(y+k)}}{(2\pi y)^{1/2}y^y e^{-y}(2\pi k)^{1/2}k^k e^{-k}}\frac{k}{(y+k)}\left(\frac{k}{k+\mu}\right)^k\left(\frac{\mu}{k+\mu}\right)^y\right\}$$

$$= \log\left\{\left[\frac{k}{2\pi y(y+k)}\right]^{\frac{1}{2}}\left[\frac{(y+k)^{y+k}}{y^y}\left(\frac{1}{k+\mu}\right)^k\left(\frac{\mu}{k+\mu}\right)^y\right]\right\}$$

$$= -\frac{1}{2}\log\left(2\pi\frac{y(y+k)}{k}\right) + y\log\left(\frac{\mu}{y}\right) - (y+k)\log\left(\frac{\mu+k}{y+k}\right)$$

$= -\frac{1}{2}\log\left\{2\pi\sigma^2 V(y)\right\} - \frac{1}{2}D(y;\mu)/\sigma^2$ (from (1))$= Q^+$

QED

Note: Nelder and Pregibon suggest the following modifications for NB and other distributions where y=0 is part of the support space. Rather than the standard Stirlings's approximation, they use the modified form:

$z! \approx \left\{2\pi(z+c)\right\}^{\frac{1}{2}} z^z e^{-z}$ for c>0 and suggest using $c=\frac{1}{6}$. Following their suggestion, the modified approximation produces $0! \approx 1.023$ instead of $0! \approx 0$ (for

c=0, as in the standard Stirling's approximation). This approximation is better than the standard one for all integer z. Using the modified Stirling's approximation in the NB log likelihood produces a change in the calculations above. The above result holds if one defines the empirical variance function with the following modification given in Nelder and Pregibon (1987, page 226). Rather than $V(y)=V(y;0)=\frac{y(y+k)}{k}$ use $V(y;c)=\frac{(y+k)^2(y+c)(k+c)}{k^2(y+k+c)}$. This modified empirical variance function reduces to the original form for c=0. Using c=$\frac{1}{6}$, the modified formula becomes $Q^+=-\frac{1}{2}\log\left\{2\pi V\left(y;c=\frac{1}{6}\right)\right\}-\frac{1}{2}D(y;\mu)/\sigma^2$

$$=-\frac{1}{2}\log\left\{2\pi\frac{(y+k)^2\left(y+\frac{1}{6}\right)\left(k+\frac{1}{6}\right)}{k^2\left(y+k+\frac{1}{6}\right)}\right\}+y\log\left(\frac{\mu}{y}\right)-(y+k)\log\left(\frac{\mu+k}{y+k}\right)$$

Substituting in $a=k^{-1}$ we get the form of Q^+ for NB(μ,a) used in Clark & Perry (1989, p. 311). They list the extended quasi-likelihood for a sample of counts y_i, i=1,...,n as $Q^+=\sum_1^n\left[y_i\log\left(\frac{\mu}{y_i}\right)-\left(\frac{1+ay_i}{a}\right)\log\left(\frac{1+a\mu}{1+ay_i}\right)\right.$

$$\left.-\frac{1}{2}\log(2\pi)-\log(1+ay_i)-\frac{1}{2}\log\left(y_i+\frac{1}{6}\right)-\frac{1}{2}\log\left(1+\frac{a}{6}\right)+\frac{1}{2}\log\left(ay_i+1+\frac{a}{6}\right)\right]$$

An adjustment for the degrees of freedom is suggested in McCullaugh and Nelder, 1989, p. 363 and Nelder (1992b). Nelder and Lee, 1992, p.281, refer to this as the "finite-sample-adjusted" form. This is an adjustment to the empirical variance term, i.e., $\left(\frac{n-p}{n}\right)\log\left\{2\pi\sigma^2V\left(y;c=\frac{1}{6}\right)\right\}$. Then the NB$(\mu,a)$ EQL equation is $Q^+=\sum_1^n\left[y_i\log\left(\frac{\mu}{y_i}\right)-\left(\frac{1+ay_i}{a}\right)\log\left(\frac{1+a\mu}{1+ay_i}\right)\right.$

$$\left.+\left(\frac{n-p}{n}\right)\left\{-\frac{1}{2}\log(2\pi)-\log(1+ay_i)-\frac{1}{2}\log\left(y_i+\frac{1}{6}\right)-\frac{1}{2}\log\left(1+\frac{a}{6}\right)+\frac{1}{2}\log\left(ay_i+1+\frac{a}{6}\right)\right\}\right]$$

Adapting EQL to the NB null hypothesis model, the estimate for the mean using EQL is $\tilde{\mu} = \bar{y}$, and p=1. The estimate for the NB variance parameter is the solution to $\frac{\partial Q^{+}}{\partial a} = 0 = \tilde{U}(a)_{EQL}$, using $\tilde{\mu} = \bar{y}$, which reduces to solving for $\tilde{a} = \tilde{a}_{EQL}$ in $\sum_{1}^{n}\left[\frac{1}{\tilde{a}^2}\log\left(\frac{1+\tilde{a}\bar{y}}{1+\tilde{a}y_i}\right) - \frac{(n-1)}{n}\frac{y_i}{1+\tilde{a}y_i} + \frac{(n-1)}{n}\frac{(1+6y_i)}{2(\tilde{a}+6+6\tilde{a}y_i)}\right] = \frac{n-1}{2(\tilde{a}+6)}$.

APPENDIX B

DERIVATION OF THE PL ESTIMATOR FOR THE NB VARIANCE PARAMETER.

The PL estimator is the solution to: $\tilde{U}(a)_{PL}=\sum_{i=1}^{n}\left[\frac{r_i^2}{V_i}-(1-h_i)\right]\frac{\partial V_i}{\partial a}.\frac{1}{V_i}=0$, where the h_i are the diagonal elements of the projection or "hat" matrix that arises at convergence of the (QL) IRLS solution for the mean parameters for a given value of the variance parameter. Let $\tilde{a}_{PL}$ denote the solution.

The projection matrix, H, is written: $H=Q\left(Q^tQ\right)^{-1}Q^t$, where $Q_{nxp}=V^{-1/2}D=\left(diag\left(\frac{\mu_i}{V_i^{1/2}}\right)\right)_{nxn}X_{nxp}$ is the matrix listed in Breslow (1989) and derived by Davidian and Carroll (1987). Note, Q is not the log quasi-likelihood equation. This Q is discussed in a note immediately after the derivation of $\tilde{a}_{PL}$.

For the NB in the single population or intercept-only GLM setting the following simplifications occur. In this setting $\mu_i=\mu$ and $V_i=V$ for all i and the design matrix, $X_{nx1}=\mathbb{1}$, is a column of 1's, so $Q=\mathrm{diag}\left(\frac{\mu}{V^{1/2}}\right)_{n\times n}\mathbb{1}_{n\times 1}=\frac{\mu}{V^{1/2}}\mathbb{1}_{n\times 1}$ and $H=\mathrm{ppo}(\mathbb{1})=\frac{1}{n}\mathbb{1}\mathbb{1}^t$ is an nxn matrix of $\frac{1}{n}$'s. The QL estimate for the mean is $\tilde{\mu}=\bar{y}$. The general form for the PL estimating equation for the variance parameter is $\tilde{U}(a)_{PL}=\sum_{i=1}^{n}\left[\frac{(y_i-\hat{\mu}_i)^2}{\hat{\mu}_i(1+a\hat{\mu}_i)}-(1-h_i)\right]\frac{\hat{\mu}_i}{(1+a\hat{\mu}_i)}$=0. Then, after substitution and simplification, $\tilde{U}(a)_{PL}$=0$=\sum_{i=1}^{n}\left[\frac{(y_i-\bar{y})^2}{\bar{y}(1+a\bar{y})}-\left(1-\frac{1}{n}\right)\right]\frac{\bar{y}}{(1+a\bar{y})}$ or $\frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{\bar{y}(1+\tilde{a}_{PL}\bar{y})}=(n-1)$, so $\tilde{a}_{PL}=\frac{S^2-\bar{y}}{\bar{y}^2}$, where $S^2=\frac{1}{(n-1)}\sum_{i=1}^{n}(y_i-\bar{y})^2$ is the usual

sample variance. Thus, the psuedolikelihood estimator for a in this NB single population setting is just the Method of Moments (MOM) estimator.

NOTE: General comments regarding the form of Q as defined above and its connection to IRLS and properties of least squares.

Nelder and Wedderburn (1972) propose IRLS as a fitting method for GLMs with distributions belonging to the linear exponential family and parameterized with a canonical link. The NB is not exponential family. The variance function contains the parameter a which is not a dispersion parameter. McCullaugh and Nelder (p.373) get around this by noting that NB(μ,a) is exponential family for fixed a and the canonical link for the NB is $\log\left(\frac{\mu}{k+\mu}\right)$ or $\log\left(\frac{a\mu}{1+a\mu}\right)$ (McCullaugh and Nelder (1989, p.372)). Hence the link is a function of μ and (fixed) a. They apply QL and IRLS to estimate the mean parameter(s). They suggest an additional equation to subsequently estimate a.

General insight regarding the form of $Q_{nxp} = \left(diag\left(\frac{\mu_i}{V_i^{1/2}}\right)\right)_{nxn} X_{nxp}$ follows (approximately) McCullaugh & Nelder's outline in section 2.5.

For EQL, we use IRLS to solve: $\underset{px1}{U}(\beta,\tilde{a}) = D^t V^{-1}(\mathbf{y}-\mu) = 0$ for β for a given (fixed) a, where $D_{nxp} = \frac{\partial \mu}{\partial \beta^t} = \{D_{ij}\} = \left\{\frac{\partial \mu_i}{\partial \beta_j}\right\}$ and V=diag($\mu_i + a\mu_i^2$). This amounts to using a log link function and η=log(μ)=$X\beta$ as the linear predictor and (iteratively) regressing an adjusted dependent variate $\mathbf{z} = \eta + \left(\frac{\partial \eta}{\partial \mu^t}\right)(\mathbf{y}-\mu)$ on the covariates, $\mathbf{x}_i : p \times 1$, with quadratic weight defined by $W^{-1} = \left(\frac{\partial \eta}{\partial \mu^t}\right) V(\mu) \left(\frac{\partial \eta}{\partial \mu^t}\right)^t = diag\left(\frac{V_i}{\mu_i^2}\right)$. Vector $\mathbf{z}$ is the first order Taylor expansion of $\log(\mathbf{y})$. We assume that $E(\mathbf{y}) = \mu$ and $\mathrm{var}(\mathbf{y}) = V(\mu)$. Consequently,

$E\left[\left(\frac{\partial \eta}{\partial \mu^t}\right)(\mathbf{y}-\mu)\right]=0$, $\text{var}\left[\left(\frac{\partial \eta}{\partial \mu^t}\right)(\mathbf{y}-\mu)\right]=W^{-1}$ and approximately $E(\mathbf{z}-\eta)\approx 0_{n\times 1}$ and $\text{var}(\mathbf{z}-\eta)\approx W^{-1}$. If the approximate moments are taken as actual and we let $\mathbf{z}^*=W^{1/2}\mathbf{z}$, then $E(\mathbf{z}^*)=W^{1/2}\eta=W^{1/2}X\beta=Q\beta$ and $\text{var}(\mathbf{z}^*)=\mathbf{I}$. The resulting estimate of β is $\tilde{\beta}=\left(X^tWX\right)^{-1}X^tW\mathbf{z}$ and $Q\tilde{\beta}=W^{\frac{1}{2}}X\left(X^tWX\right)^{-1}X^tW\mathbf{z}=Q\left(Q^tQ\right)^{-1}QW^{\frac{1}{2}}\mathbf{z}=ppo(Q)\mathbf{z}^*$, where ppo denotes perpendicular projection operator. Thus $H=ppo(Q)$ is the projection matrix arising at convergence of the (QL) IRLS solutions for the mean parameters.

APPENDIX C:

VERIFICATION OF OQ ESTIMATING EQUATIONS FOR NB IN GLM SETTING.

Notation and equations follow those of Godambe and Thompson (1989).

The general extended quasi-score function, or OQ estimating equations for the r[th] component of the parameter vector θ is:

$$\sum_{i=1}^{n} h_{1i} w_{1i} + \sum_{i=1}^{n} h_{2i} w_{2i} = 0$$

$$= \sum_{i=1}^{n} (y_i - \mu_i) \left(\frac{\partial \mu_i / \partial \theta_r}{\sigma^2 V_i} \right) + \sum_{i=1}^{n} \left\{ \left[(y_i - \mu_i)^2 - (\sigma^2 V_i) - \gamma_1 (\sigma^2 V_i)^{1/2} (y_i - \hat{\mu}_i) \right] \frac{\left[\gamma_1 \left(\frac{\partial \mu_i}{\partial \theta_r} \right) - \sigma \left(\frac{\partial v_i}{\partial \theta_r} \right) \Big/ V_i^{1/2} \right]}{(\sigma^2 V_i)^{3/2} (\gamma_2 + 2 - \gamma_1^2)} \right\},$$

where $E(y_i) = \mu_i(\theta)$, $\mathrm{var}(y_i) = \sigma^2 V_i$ for $V_i = V_i(\theta)$ and σ^2 = scale or dispersion parameter, and where $\gamma_1 = E\left[\frac{y_i - \mu_i}{\left(E[y_i - \mu_i]^2 \right)^{1/2}} \right]^3$ =(standardized) skewness, and

$\gamma_2 = E\left[\frac{y_i - \mu_i}{\left(E[y_i - \mu_i]^2 \right)^{1/2}} \right]^4 - 3 =$ (standardized) kurtosis.

For the NB model, θ=(β,a) with $\mu_i = \exp(\mathbf{x}_i^t \beta)$ and common a, $\sigma^2 = 1$,

$V_i = \mu_i + a\mu_i^2$, $\gamma_{1i} = \frac{(1 + 2a\mu_i)}{(\mu_i(1 + a\mu_i))^{1/2}} = \frac{(1 + 2a\mu_i)}{V_i^{1/2}} = \frac{\partial V_i / \partial \mu_i}{V_i^{1/2}}$,

$\gamma_{2i} = 6a + \{\mu_i(1 + a\mu_i)\}^{-1} = 6a + \frac{1}{V_i}$, and $(\gamma_{2i} + 2 - \gamma_{1i}^2) = 2(a+1)$ for all i.

For the mean parameters, the estimating equations are the same as for quasi-likelihood because $\sum_{i=1}^{n} h_{2i} w_{2i} = 0$ which follows because the term in the numerator of w_{2i}, namely $\left[\gamma_{1i} \left(\frac{\partial \mu_i}{\partial \beta_r} \right) - \sigma \left(\frac{\partial V_i}{\partial \beta_r} \right) \Big/ V_i^{1/2} \right]$ is zero due to the form of γ_{1i} for the NB.

This same property holds for GLM settings in general when θ is the canonical parameter. The GLM (exponential family) property was noted by Nelder in his contribution to the discussion of the Godambe and Thompson (1989) article.

For the variance parameter, $\tilde{U}(a)_{OQ}$ simplifies to $\tilde{U}(a)_{OQ}=\sum_{i=1}^{n}\left\{\frac{\left[(y_i-\tilde{\mu}_i)^2-V_i-(1+2a\tilde{\mu}_i)(y_i-\tilde{\mu}_i)\right]}{(1+a\tilde{\mu}_i)^2}\right\}$. This simplification for the NB with common a was noted by Dean and Lawless in their contribution to the discussion of the Godambe and Thompson (1989) article.

There are further simplifications for the one-way layout. Here the third term in $\tilde{U}(a)_{OQ}$ sums to zero. In the single population model the solution to $\tilde{U}(a)_{OQ}=0$ yields $\sum_{i=1}^{n}\left[(y_i-\bar{y})^2-V_i\right]=0$, so that $\sum_{i=1}^{n}(y_i-\bar{y})^2=n\bar{y}\left(1+\tilde{a}_{OQ}\bar{y}\right)$ or

$$\tilde{a}_{OQ}=\frac{\frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})^2-\bar{y}}{\bar{y}^2}=\frac{\left(\frac{n-1}{n}\right)S^2-\bar{y}}{\bar{y}^2}.$$

APPENDIX D

VERIFICATION OF ASSUMPTIONS FOR $\mathrm{NB}(\mu,k)$ FAMILY NEEDED TO CONFIRM ASYMPTOTIC DISTRIBUTION OF NED ESTIMATORS FOR THE $\mathrm{NB}(\mu,k)$ PARAMETERS.

Lindsay (1994) lists the following assumptions about the family of models (Assumption 31, Appendix A) along with conditions on the residual adjustment function which are used to establish the limiting distribution of minimum disparity estimators. Lindsay states that the conditions on the residual adjustment factor are satisfied by the NED estimator. Thus for a family of models satisfying the listed assumptions, the Disparity Difference Test (DDT) has an asymptotic chi-square distribution under H_0 (Theorem 6, Lindsay).

The NB(μ,k) is both the model and true density. Many of the following properties of the NB distribution may be listed and proven elsewhere, but I could not find them in the literature. Therefore, before using the NED method, I felt it necessary to show that the NB family of models satisfies the regularity conditions of TPE (Lehmann, *Theory of Point Estimation*, 1983) on pages 409 and 429, and an additional assumption. The conditions from TPE p. 409 are listed below as (A0)-(A3); the conditions from TPE p. 429 are listed below as (A)-(D). I used Lehmann's numbering and wording for the conditions so that the reader could refer to TPE if desired. The additional condition from Lindsay is listed below as (E).

For all proofs, let ν=counting measure, $\theta=(\mu,k)$, P_θ =NB(μ,k).

(A0) The distributions P_θ of the observations are distinct.

Proof: This follows directly from the fact that the probability generating function (pgf) uniquely determines the distribution of a discrete random variable. (proof in Fisz, Prob Thy & Math Stat, p.125) From the Fisz proof it follows that:

$$E_\theta\left(s^y\right) = E_{\theta'}\left(s^y\right) \Leftrightarrow \sum_{y=0}^{\infty} p_\theta(y)s^y = \sum_{y=0}^{\infty} p_{\theta'}(y)s^y \Leftrightarrow \frac{p_\theta(y)}{p_{\theta'}(y)} = 1 \text{ for } y = 0,1,2,\ldots$$

Thus,

for y=0, $$\frac{p_\theta(0)}{p_{\theta'}(0)} = \frac{\left(\frac{k}{\mu+k}\right)^k}{\left(\frac{k'}{\mu'+k'}\right)^{k'}} = 1 \qquad \text{(a)}$$

for y=1, $$\frac{p_\theta(1)}{p_{\theta'}(1)} = \frac{k\left(\frac{\mu}{\mu+k}\right)\left(\frac{k}{\mu+k}\right)^k}{k'\left(\frac{\mu'}{\mu'+k'}\right)\left(\frac{k'}{\mu'+k'}\right)^{k'}} = 1 \qquad \text{(b)}$$

$$= \frac{k\left(\frac{\mu}{\mu+k}\right)}{k'\left(\frac{\mu'}{\mu'+k'}\right)}, \text{ from (a)} \qquad \text{(c)}$$

for y=2, $$\frac{p_\theta(2)}{p_{\theta'}(2)} = \frac{k(k+1)\left(\frac{\mu}{\mu+k}\right)^2\left(\frac{k}{\mu+k}\right)^k}{k'(k'+1)\left(\frac{\mu'}{\mu'+k'}\right)^2\left(\frac{k'}{\mu'+k'}\right)^{k'}} = 1$$

$$= \frac{k'(k+1)}{k(k'+1)}, \text{ from (b) and (c).}$$

$$\Leftrightarrow k = k'. \qquad \text{(d)}$$

Substituting (d) into (c) $\frac{p_\theta(1)}{p_{\theta'}(1)} = 1 \Leftrightarrow \mu = \mu'$

Thus, P_θ =NB(μ,k) distns are distinct for values of θ=(μ,k).

(A1) The distributions P_θ have common support.

NB(μ,k) have the non-negative integers as common support,

(A2) The observations are $\boldsymbol{Y}=(\boldsymbol{Y_1},...,\boldsymbol{Y_n})$ where the $\boldsymbol{Y_i}$ are iid with probability density $f(y_i)$ with respect to ν, counting measure.

We assume throughout that this assumption is true.

(A3) The parameter space Ω contains an open interval ω of which the true parameter value θ_0 is an interior point.

The parameter space $\Omega=\{(\mu,k):0<\mu<\infty,0<k<\infty\}$ contains an open interval ω of which (μ_0,k_0) is an interior point. Specifically, for any given $(\mu_0,k_0)\in\Omega$ we can let $\omega=(.5\mu_0,2\mu_0)\cup(.5k_0,2k_0)$.

(A) There exists on open subset ω of Ω containing the true parameter point θ_0 such that for almost all y the density $f(y,\theta)$ admits all third derivatives $\left(\partial^3/\partial\theta_i\partial\theta_j\partial\theta_k\right)f(y,\theta)$ for all θ in ω.

Proof: Let ω be as defined in (A3). All third derivatives of the density are composed of a product of the original density and a function (sums and products) of the derivatives of the log likelihood. The first and second derivatives of the log-likelihood are listed in many sources. The derivatives of the density can be derived through use of the chain rule. I will now produce the third derivatives.

Let $f=p$, $l=\log(p)$, $V=\text{var}(y)=\mu+\frac{\mu^2}{k}=\frac{\mu(\mu+k)}{k}$.

$$p=p_{\mu,k}(y)=\Pr(Y=y;\mu,k)=\frac{\Gamma(y+k)}{y!\Gamma(k)}\left(\frac{\mu}{\mu+k}\right)^y\left(\frac{k}{\mu+k}\right)^k$$
$$=\frac{k(k+1)\cdots(k+(y-1))}{y!}\left(\frac{\mu}{\mu+k}\right)^y\left(\frac{k}{\mu+k}\right)^k$$
$$l=\log\left(p_{\mu,k}(y)\right)=\left[\sum_{j=0}^{y-1}\log(k+j)\right]-\log y!-(y+k)\log(\mu+k)+y\log(\mu)+k\log(k)$$

The required derivatives of the log likelihood are as follows:

$$\frac{\partial l}{\partial\mu}=\frac{(y-\mu)k}{\mu(\mu+k)}=\frac{(y-\mu)}{V};$$

$$\frac{\partial^2 l}{\partial\mu^2}=\frac{\partial l}{\partial\mu}A=\frac{k\left(\mu^2-ky-2\mu y\right)}{\mu^2(\mu+k)^2}\text{ where }A=-\frac{1}{(y-\mu)}-\frac{1}{\mu}-\frac{1}{(\mu+k)};$$

$$\frac{\partial^3 l}{\partial\mu^3}=\frac{\partial^2 l}{\partial\mu^2}A+\frac{\partial l}{\partial\mu}\frac{\partial A}{\partial\mu}=\frac{\partial l}{\partial\mu}A^2+\frac{\partial l}{\partial\mu}A'=\frac{k\left(6\mu ky+6\mu^2y+2k^2y-2\mu^3\right)}{\mu^3(\mu+k)^3};$$

$$\frac{\partial l}{\partial k}=\left[\sum_{j=0}^{y-1}\frac{1}{(k+j)}\right]-\frac{(y-\mu)}{(\mu+k)}-\log\left(\frac{k+\mu}{k}\right);$$

$$\frac{\partial^2 l}{\partial k^2}=\left[\sum_{j=0}^{y-1}\frac{-1}{(k+j)^2}\right]+\frac{\left(ky+\mu^2\right)}{k(\mu+k)^2}\qquad\frac{\partial^3 l}{\partial k^3}=\left[\sum_{j=0}^{y-1}\frac{2}{(k+j)^3}\right]-\frac{\left(\mu^3+3k\mu^2+2k^2y\right)}{k^2(\mu+k)^3};$$

$$\frac{\partial^2 l}{\partial k \partial \mu} = \frac{y-\mu}{(\mu+k)^2}; \qquad \frac{\partial^3 l}{\partial \mu^2 \partial k} = \frac{\mu-k-2y}{(\mu+k)^3}; \text{ and } \frac{\partial^3 l}{\partial \mu \partial k^2} = \frac{2(\mu-y)}{(\mu+k)^3}.$$

The first and second derivatives of p are:

$$\frac{\partial p}{\partial \mu} = p\frac{\partial l}{\partial \mu} \text{ and } \frac{\partial p}{\partial k} = p\frac{\partial l}{\partial k};$$

$$\frac{\partial^2 p}{\partial \mu^2} = \frac{\partial p}{\partial \mu}\frac{\partial l}{\partial \mu} + p\frac{\partial^2 l}{\partial \mu^2} = p\left[\left(\frac{\partial l}{\partial \mu}\right)^2 + \frac{\partial^2 l}{\partial \mu^2}\right];$$

$$\frac{\partial^2 p}{\partial k^2} = p\left[\left(\frac{\partial l}{\partial k}\right)^2 + \frac{\partial^2 l}{\partial k^2}\right]; \text{ and}$$

$$\frac{\partial^2 p}{\partial \mu \partial k} = \frac{\partial p}{\partial k}\frac{\partial l}{\partial \mu} + p\frac{\partial^2 l}{\partial \mu \partial k} = p\left[\frac{\partial l}{\partial k}\frac{\partial l}{\partial \mu} + \frac{\partial^2 l}{\partial \mu \partial k}\right].$$

The third derivatives of the density:

$$\frac{\partial^3 p}{\partial \mu^3} = \frac{\partial p}{\partial \mu}\left[\left(\frac{\partial l}{\partial \mu}\right)^2 + \frac{\partial^2 l}{\partial \mu^2}\right] + p\left[2\frac{\partial l}{\partial \mu}\frac{\partial^2 l}{\partial \mu^2} + \frac{\partial^3 l}{\partial \mu^3}\right] = p\left[\left(\frac{\partial l}{\partial \mu}\right)^3 + 3\frac{\partial l}{\partial \mu}\frac{\partial^2 l}{\partial \mu^2} + \frac{\partial^3 l}{\partial \mu^3}\right];$$

$$\frac{\partial^3 p}{\partial k^3} = p\left[\left(\frac{\partial l}{\partial k}\right)^3 + 3\frac{\partial l}{\partial k}\frac{\partial^2 l}{\partial k^2} + \frac{\partial^3 l}{\partial k^3}\right];$$

$$\frac{\partial^3 p}{\partial \mu^2 \partial k} = \frac{\partial p}{\partial \mu}\left[\frac{\partial l}{\partial k}\frac{\partial l}{\partial \mu} + \frac{\partial^2 p}{\partial \mu \partial k}\right] + p\left[\frac{\partial^2 l}{\partial \mu \partial k}\frac{\partial l}{\partial \mu} + \frac{\partial l}{\partial k}\frac{\partial^2 l}{\partial \mu^2} + \frac{\partial^3 l}{\partial \mu^2 \partial k}\right]$$

$$= p\left[\frac{\partial l}{\partial k}\left(\frac{\partial l}{\partial \mu}\right)^2 + 2\frac{\partial^2 l}{\partial k \partial \mu}\frac{\partial l}{\partial \mu} + \frac{\partial l}{\partial k}\frac{\partial^2 l}{\partial \mu^2} + \frac{\partial^3 l}{\partial k \partial \mu^2}\right]; \text{ and}$$

$$\frac{\partial^3 p}{\partial \mu \partial k^2} = p\left[\frac{\partial l}{\partial \mu}\left(\frac{\partial l}{\partial k}\right)^2 + 2\frac{\partial^2 l}{\partial \mu \partial k}\frac{\partial l}{\partial k} + \frac{\partial l}{\partial \mu}\frac{\partial^2 l}{\partial k^2} + \frac{\partial^3 l}{\partial \mu \partial k^2}\right].$$

QED (A)

Note: The following facts are required for Condition (B)

To show the first and second derivatives of the likelihood exist and are finite for all $y \in A$ and $\theta \in \Omega$.

Proof: The lower derivatives of the density are finite because they are dominated by the functions of the log-likelihoods (or more correctly, functions of the derivatives of the log-likelihoods). The expected values of these functions are finite for the following reason. Summation of terms over the support of Y is monotone for either the derivative of the density or the functions of the log likelihood because all terms are of like sign. Thus, the sum of the terms (or its negative) represents the maximum. Also, the sum of terms of the density derivatives equals the expected value of the function of log likelihoods. Using the Dominated Convergence Theorem, it suffices to show that the expectation of each term in the functions of the log-likelihood is finite for (μ,k) in Ω.

In the calculations below, I make use of the following:

Lemma 1: (from Durrett, p.34) If $X \geq 0$ is integer-valued, then $E(X) = \sum_{i=0}^{\infty} \Pr(X > i)$.

Pf: $E(X) = \sum_{n=0}^{\infty} n \Pr(X = n) = \sum_{n=1}^{\infty} \Pr(X = n) \left(\sum_{i=1}^{n} 1 \right) = \sum_{i=1}^{\infty} \sum_{n=i}^{\infty} \Pr(X = n) = \sum_{i=0}^{\infty} \Pr(X > i)$ where reversal of summation allowed since all terms are ≥ 0 and series on left is known convergent (Buck, Thm 14, p 174). QED

Lemma 2: (an extension of Lemma 1) If $X \geq 0$ is integer-valued, and $0 \leq s(j) < 1$ is a sequence not involving X, then $E\left(\sum_{j=0}^{x-1} s(j) \right) = \sum_{j=0}^{\infty} s(j) \Pr(X > j)$.

Pf: $E\left(\sum_{j=0}^{x-1} s(j) \right) = \sum_{n=1}^{\infty} \Pr(X = n) \sum_{j=0}^{n-1} s(j) = \sum_{j=0}^{\infty} s(j) \sum_{n=j+1}^{\infty} \Pr(X = n) = \sum_{j=0}^{\infty} s(j) \Pr(X > j)$ where reversal of summation allowed since all terms are ≥ 0 and series is bounded above by a known convergent series, ie: series in Lemma 1 where $s(j) = 1$. QED

Let $E=E_\theta$ for the remainder of the Appendix.

Lemma 3: $E\left[\sum_{j=0}^{y-1}\frac{(y-\mu)}{(k+j)}\right]=\frac{\mu}{k}$.

Proof: Let $B=E\left[\sum_{j=0}^{y-1}\frac{y}{k+j}\right]$.

Now by definition $B=\sum_{y=0}^{\infty}\left\{\frac{\Gamma(y+k)}{y!\Gamma(k)}\left(\frac{\mu}{\mu+k}\right)^{y}\left(\frac{k}{\mu+k}\right)^{k}\left[y\sum_{j=0}^{y-1}\frac{1}{k+j}\right]\right\}$

The first term is zero, so summation can begin for y=1. Rewriting terms,

$$B=\left(\frac{\mu}{\mu+k}\right)\sum_{y=1}^{\infty}\left\{\frac{\Gamma((y-1)+k)}{(y-1)!\Gamma(k)}\left(\frac{\mu}{\mu+k}\right)^{y-1}\left(\frac{k}{\mu+k}\right)^{k}\left[[(y-1)+k]\sum_{j=0}^{y-1}\frac{1}{k+j}\right]\right\}$$

Let x=y -1, and the equation for B is just

$$\left(\frac{\mu}{\mu+k}\right)E\left[\sum_{j=0}^{x}\frac{x+k}{k+j}\right]=\left(\frac{\mu}{\mu+k}\right)E\left[1+\sum_{j=0}^{x-1}\frac{x+k}{k+j}\right]=\left(\frac{\mu}{\mu+k}\right)\left\{1+kE\left[\sum_{j=0}^{x-1}\frac{1}{k+j}\right]+B\right\}.$$

Using algebra, $\frac{k}{\mu}B=1+kE\left[\sum_{j=0}^{x-1}\frac{1}{k+j}\right]$. Thus, $E\left[\sum_{j=0}^{x-1}\frac{(x-\mu)}{(k+j)}\right]=\frac{\mu}{k}$. QED

Now I shall show that the following expectations are bounded by a finite function for (μ,k) in $\Omega=\{(\mu,k):0<\mu<\infty,0<k<\infty\}$.

(1) $E\left[\frac{\partial l}{\partial\mu}\right]=E\left[\frac{(y-\mu)k}{\mu(\mu+k)}\right]=\frac{k}{\mu(\mu+k)}E[(y-\mu)]=0.$

(2) $E\left[\frac{\partial l}{\partial k}\right]=E\left[\sum_{j=0}^{y-1}\frac{1}{(k+j)}\right]-\log\left(\frac{k+\mu}{k}\right)=\sum_{j=0}^{\infty}\frac{1}{(k+j)}\Pr(Y>j)-\log\left(\frac{k+\mu}{k}\right)$

$$<\frac{1}{k}\sum_{j=0}^{\infty}\Pr(Y>j)+\log\left(\frac{k+\mu}{k}\right)=\frac{\mu}{k}+\frac{k+\mu}{k}<\infty.$$

(3) $E\left[\left(\frac{\partial l}{\partial\mu}\right)^{2}\right]=\frac{E[(y-\mu)^{2}]}{V^{2}}=\frac{1}{V}<\infty$

(4) $E\left[-\frac{\partial^{2}l}{\partial\mu^{2}}\right]=\frac{k\mu(\mu+k)}{\mu^{2}(\mu+k)^{2}}=\frac{k}{\mu(\mu+k)}=\frac{1}{V}<\infty$

$$(5)\ E\left[-\frac{\partial^2 l}{\partial k^2}\right]=E\left[\sum_{j=0}^{y-1}\frac{1}{(k+j)^2}\right]-\frac{\mu}{k(\mu+k)}<E\left[\sum_{j=0}^{y-1}\frac{1}{(k+j)^2}\right]+\frac{1}{k}$$

$$=\frac{1}{k}+\sum_{j=0}^{\infty}\frac{1}{(k+j)^2}\Pr(Y\geq j)<\frac{1}{k}+\sum_{j=0}^{\infty}\frac{1}{(k+j)^2}$$

$$<\frac{1}{k}+\left(\frac{1}{k^2}+\sum_{j=0}^{\infty}\frac{1}{j^2}\right)=\frac{k+1}{k^2}+\frac{\pi^2}{6}<\infty$$

$$(6)\ E\left[\frac{\partial l}{\partial k}\frac{\partial l}{\partial \mu}\right]=E\left\{\left[\left(\sum_{j=0}^{y-1}\frac{1}{(k+j)}\right)-\left(\frac{y-\mu}{\mu+k}\right)\right]\left[\frac{(y-\mu)k}{(\mu+k)}\right]\right\}$$

$$=\frac{1}{(\mu+k)}\left\{E\left[\sum_{j=0}^{y-1}\frac{(y-\mu)k}{(k+j)}\right]-\frac{kE\left[(y-\mu)^2\right]}{\mu+k}\right\}=\frac{1}{(\mu+k)}\left\{E\left[\sum_{j=0}^{y-1}\frac{(y-\mu)k}{(k+j)}\right]-\mu\right\}=0$$

by Lemma 3

$$(7)\ E\left[\frac{\partial^2 l}{\partial k\partial \mu}\right]=\frac{E(y-\mu)}{(\mu+k)^2}=0$$

$$(8)\ E\left[\left(\frac{\partial l}{\partial k}\right)^2\right]=E\left[\left(\sum_{j=0}^{y-1}\frac{1}{(k+j)}\right)^2\right]+E\left\{\left[\frac{(y-\mu)}{(\mu+k)}\right]^2\right\}+\log^2\left(\frac{k+\mu}{k}\right)$$

$$-2E\left\{\frac{(y-\mu)}{(\mu+k)}\sum_{j=0}^{y-1}\frac{1}{(k+j)}\right\}-2\log\left(\frac{k+\mu}{k}\right)E\left[\sum_{j=0}^{y-1}\frac{1}{(k+j)}\right]$$

$$-2\log\left(\frac{k+\mu}{k}\right)E\left[\frac{(y-\mu)}{(\mu+k)}\right]$$

$$=E\left\{\left[\sum_{j=0}^{y-1}\frac{1}{(k+j)}\right]^2\right\}+\frac{\mu}{k(\mu+k)}+\log^2\left(\frac{k+\mu}{k}\right)-\frac{2\mu}{k(\mu+k)}-2\log\left(\frac{k+\mu}{k}\right)E\left[\sum_{j=0}^{y-1}\frac{1}{(k+j)}\right]$$

4th term by Lemma 3,

$$<E\left\{\left[\sum_{j=0}^{y-1}\frac{1}{(k+j)^2}+\sum_{j=0}^{y-1}\frac{2y}{(k+j)k}\right]\right\}+\frac{\mu}{k(\mu+k)}+\log^2\left(\frac{k+\mu}{k}\right)+2\frac{\mu}{k}\log\left(\frac{k+\mu}{k}\right)$$

$$<\left\{\left(\frac{1}{k^2}+\frac{\pi^2}{6}\right)+\left[\frac{2\mu(\mu+1)}{k^2}\right]\right\}+\frac{\mu}{k(\mu+k)}+\left(\frac{k+\mu}{k}\right)^2+2\frac{\mu}{k}\left(\frac{k+\mu}{k}\right)<\infty\ \text{QED.}$$

(end of Note from (A)).

(B) The first and second logarithmic derivatives of f satisfy the equations

$$E\left[\frac{\partial}{\partial\theta_j}\log f(Y,\theta)\right]=0 \qquad \text{for } j=1,\ldots,s$$

and

$$I_{jk}(\theta)=E\left[\frac{\partial}{\partial\theta_j}\log f(Y,\theta)\cdot\frac{\partial}{\partial\theta_k}\log f(Y,\theta)\right]=E\left[-\frac{\partial^2}{\partial\theta_j\partial\theta_k}\log f(Y,\theta)\right].$$

Proof: This follows from the extension of Lemma 2.6.1 to the multiparameter case (Lehmann, TPE, pp. 125-126). The hypotheses of this extension are: Ω is an open interval (finite, infinite, or semi-infinite) (see (A3)),The distributions P_θ have common support, so that without loss of generality the set $A=\{y: p_\theta(y)>0\}$ is independent of θ (see (A1)), the first and second derivatives of the likelihood exist and are finite for all $y\in A$ and $\theta\in\Omega$ (see Note from A) and, the derivatives with respect to each θ_i on the left side of $\int p_\theta(y)dv(y)=1$ can be obtained by differentiating under the integral sign (see Ash (1972), problem 3, page 52). Because the hypotheses hold, the extension follows.

(C) Since the s xs matrix $I(\theta)$ is a covariance matrix, it is positive semidefinite. In generalization of condition (v) of Theorem 1.1 we shall assume that the $I_k(\theta)$ are finite and that the matrix $I(\theta)$ is positive definite for all θ in ω and hence that the s statistics

$$\frac{\partial}{\partial\theta_1}f(Y,\theta),\ldots,\frac{\partial}{\partial\theta_s}f(Y,\theta)$$

are affinely independent with probability 1.

Proof: According to Lehmann, this proof is covered by the same Lehmann extension and hypotheses as for (B) (see above).

(D) There exist functions M_{jkl} such that

$$\left|\frac{\partial^3}{\partial\theta_j\partial\theta_k\partial\theta_l}\log f(y,\theta)\right| \le M_{jkl}(y) \qquad \text{for all } \theta\in\omega$$

where

$$m_{jkl} = E_{\theta_0}\left[M_{jkl}(Y)\right] < \infty \qquad \text{for all } j,k,l.$$

Proof: There are four cases to consider. Let the symbol := mean "denoted by".

(i) $$\frac{\partial^3 l}{\partial\mu^3} = \frac{k\left(6\mu ky + 6\mu^2 y + 2k^2 y - 2\mu^3\right)}{\mu^3(\mu+k)^3} < \frac{k\left(6\mu ky + 6\mu^2 y + 2k^2 y + 2\mu^3\right)}{\mu^3(\mu+k)^3} := M_{jkl}(y).$$

Then $$m_{jkl} = E_{\theta_0}\left[M_{jkl}(Y)\right] = \frac{2k\left(3\mu k + 4\mu^2 + 2k^2\right)}{\mu^2(\mu+k)^3} < \frac{8k}{\mu^2(\mu+k)} < \infty,\ \forall(\mu,k)\in\omega.$$

(ii) $$\frac{\partial^3 l}{\partial\mu^2\partial k} = \frac{\mu-k-2y}{(\mu+k)^3} < \frac{\mu+k+2y}{(\mu+k)^3} := M_{jkl}(y).$$

Then $$m_{jkl} = E_{\theta_0}\left[M_{jkl}(Y)\right] = \frac{3\mu+k}{(\mu+k)^3} < \frac{3}{(\mu+k)^2} < \infty,\ \forall(\mu,k)\in\omega.$$

(iii) $$\frac{\partial^3 l}{\partial\mu\partial k^2} = \frac{2(\mu-y)}{(\mu+k)^3} < \frac{2(\mu+y)}{(\mu+k)^3} := M_{jkl}(y).$$

Then $$m_{jkl} = E_{\theta_0}\left[M_{jkl}(Y)\right] = \frac{4\mu}{(\mu+k)^3} < \infty,\ \forall(\mu,k)\in\omega.$$

(iv) $$\frac{\partial^3 l}{\partial k^3} = \left[\sum_{j=0}^{y-1}\frac{2}{(k+j)^3}\right] - \frac{\left(\mu^3+3k\mu^2+2k^2y\right)}{k^2(\mu+k)^3} < \left[\sum_{j=0}^{y-1}\frac{2}{(k+j)^3}\right] + \frac{\left(\mu^3+3k\mu^2+2k^2y\right)}{k^2(\mu+k)^3}$$

$:= M_{jkl}(y)$. Then

$$m_{jkl} = E_{\theta_0}\left[M_{jkl}(Y)\right] = \left[\sum_{j=0}^{\infty}\frac{2P(Y>j)}{(k+j)^3}\right] + \frac{\left(\mu^3+3k\mu^2+2k^2\mu\right)}{k^2(\mu+k)^3} < \left[\sum_{j=0}^{\infty}\frac{2}{(k+j)^3}\right] + \frac{1}{k^2} < \infty,$$

$\forall(\mu,k)\in\omega$.

QED.

(E) There exist $M_{ijk}(y)$, $M_{ij,k}(y)$ and $M_{i,j,k}(y)$ that dominate in absolute value $u_{ijk}(y,\theta)$, $u_{ij}(y,\theta)u_i(y,\theta)$, and $u_i(y,\theta)u_j(y,\theta)u_k(y,\theta)$respectively, for all θ in a neighborhood ω of θ_0, and that are uniformly bounded in expectation E_θ for all θ in some, possibly smaller, open neighborhood of θ_0.

$u_i(y,\theta), u_{ij}(y,\theta)$ and $u_{ijk}(y,\theta)$ are defined in Lindsay as the first, second and third derivatives of the log likelihood with respect to θ, where the subscripts denote the components of θ involved. Let ω be the neighborhood of $\theta_0=(\mu_0,k_0)$.

Proof: The choice of functions in part (D) and the calculations satisfy the conditions for $u_{ijk}(y,\theta)$.

$$\text{(i) } u_{ij}(y,\theta)u_i(y,\theta)=\frac{\partial^2 l}{\partial\beta_r\partial\beta_s}\frac{\partial l}{\partial k}=\left\{\frac{-(y+k)k\mu x_r x_s}{(\mu+k)^2}\right\}.$$

Let $M_{ij,k}(y) = \dfrac{(y+k)k\mu}{(\mu+k)^2}\left\{\left[\sum_{j=0}^{y-1}\frac{1}{(k+j)}\right]+\frac{(y-\mu)}{(\mu+k)}+\log\left(\frac{k+\mu}{k}\right)\right\}$. Then,

$$E\left[M_{ij,k}(y)\right]=\frac{k\mu}{(\mu+k)^2}\left\{E\left[\sum_{j=0}^{y-1}\frac{y}{k+j}\right]+kE\left[\sum_{j=0}^{y-1}\frac{1}{k+j}\right]+\frac{E[y(y-\mu)]}{\mu+k}+(\mu+k)\log\left(\frac{\mu+k}{k}\right)\right\}$$

$$=\frac{k\mu}{(\mu+k)^2}\left\{\left[\frac{\mu}{k}+\mu\log\left(\frac{\mu+k}{k}\right)\right]+k\log\left(\frac{\mu+k}{k}\right)+\left[\frac{\mu(\mu+k)/k}{\mu+k}\right]+(\mu+k)\log\left(\frac{\mu+k}{k}\right)\right\}$$

$$=\frac{k\mu}{(\mu+k)^2}\left[\frac{2\mu}{k}+2(\mu+k)\log\left(\frac{\mu+k}{k}\right)\right]<2(\mu+1)<\infty,\ \forall(\mu,k)\in\omega.$$

$$\text{(ii) } u_i(y,\theta)u_j(y,\theta)u_k(y,\theta)=\frac{\partial l}{\partial\beta_i}\frac{\partial l}{\partial\beta_j}\frac{\partial l}{\partial k}$$

$$=\left\{\frac{(y-\mu)kx_i}{(\mu+k)}\right\}\left\{\frac{(y-\mu)kx_j}{(\mu+k)}\right\}\left\{\left[\sum_{j=0}^{y-1}\frac{1}{(k+j)}\right]-\frac{(y-\mu)}{(\mu+k)}-\log\left(\frac{k+\mu}{k}\right)\right\}.$$

Let $M_{i,j,k}(y)=\left\{\dfrac{(y+\mu)k}{(\mu+k)}\right\}^2\left\{\left[\sum_{j=0}^{y-1}\frac{1}{(k+j)}\right]+\frac{(y+\mu)}{(\mu+k)}+\log\left(\frac{k+\mu}{k}\right)\right\}$.

Then $E(M_{i,j,k}) = \frac{k^2}{(\mu+k)^2}\left\{E\left[\sum_{j=0}^{y-1}\frac{y^2}{k+j}\right] - 2\mu E\left[\sum_{j=0}^{y-1}\frac{y}{k+j}\right] + \mu^2 E\left[\sum_{j=0}^{y-1}\frac{1}{k+j}\right]\right.$

$\left. + \left[\frac{E(y-\mu)^3}{\mu+k}\right] + \log\left(\frac{\mu+k}{k}\right)E(y-\mu)^2\right\}$

$= \frac{k^2}{(\mu+k)^2}\left\{\left[\frac{\mu(2\mu+1)}{k} + \left(\frac{k\mu(\mu+1)+\mu^2}{k}\right)\log\left(\frac{\mu+k}{k}\right)\right] - 2\mu\left[\frac{\mu}{k} + \mu\log\left(\frac{\mu+k}{k}\right)\right]\right.$

$\left. + \mu^2\log\left(\frac{\mu+k}{k}\right) + \left[\frac{\mu(\mu+k)(k+2\mu)/k^2}{\mu+k}\right] + \frac{\mu(\mu+k)}{k}\log\left(\frac{\mu+k}{k}\right)\right\}$

$= \frac{k^2}{(\mu+k)^2}\left\{\frac{\mu}{k} + \frac{2\mu(\mu+k)}{k}\log\left(\frac{\mu+k}{k}\right)\right\} < 2(\mu+1) < \infty, \ \forall(\mu,k) \in \omega.$

QED.

Additional interesting facts that result from Condition (B):

(1) $E\left[\sum_{j=0}^{y-1}\frac{1}{(k+j)}\right] = \log\left(\frac{k+\mu}{k}\right)$, from $E\left[\frac{\partial l}{\partial k}\right] = 0$

Note: $\sum_{j=0}^{y-1}\frac{1}{(k+j)} = \psi(y+k) + \gamma$, where ψ is the digamma function and γ is Euler's constant.

(2) $E\left[\left(\sum_{j=0}^{y-1}\frac{1}{k+j}\right)^2\right] - E\left[\sum_{j=0}^{y-1}\left(\frac{1}{k+j}\right)^2\right] = \log^2\left(\frac{k+\mu}{k}\right)$, from $E\left[-\frac{\partial^2 l}{\partial k^2}\right] = E\left[\left(\frac{\partial l}{\partial k}\right)^2\right]$ Note:

$\sum_{j=0}^{y-1}\left(\frac{1}{k+j}\right)^2 = \frac{\pi^2}{6} - \psi'(y+k)$, where ψ' is the tri-gamma function.

An informative section on the gamma function and related functions (log gamma, digamma, polygamma) can be found in Abramowitz and Stegun (1964).

APPENDIX E

GENERALIZED SCORE TESTS. SUMMARY OF WORK BY BOOS (1992) AND BRESLOW (1989, 1990) AND SIMPLIFICATIONS.

The setting is the generalized NB log-linear model. The mean structure has the form $E(Y_i|\mathbf{x}_i)=\mu(\mathbf{x}_i)=e^{\mathbf{x}_i^t\beta}$ and the variance is $\mathrm{var}(Y_i)=\mu_i+a\mu_i^2=V_i$ for the n observations indexed by i. The estimating equations used in the generalized score test are quasi-likelihood for the mean and an additional equation for estimating the variance parameter. The equations for the mean are $\underset{px1}{U}=\underset{px1}{U}(\beta)=\underset{px1}{U}(\beta;a)=D^tV^{-1}(\mathbf{y}-\mu)=X^t\left[diag\left(\frac{1}{1+a\mu_i}\right)\right](\mathbf{y}-\mu)$ where

$D_{nxp}=\frac{\partial\mu}{\partial\beta^t}=\{D_{ij}\}=\left\{\frac{\partial\mu_i}{\partial\beta_j}\right\}=\{diag(\mu_i)\}X$ or

$\underset{p\,x\,1}{U}(\beta)=\sum_{i=1}^{n}\underset{p\,x\,1}{\mathbf{u}_i}=\sum_{i=1}^{n}\frac{(y_i-\mu_i)}{V_i}\underset{p\,x\,1}{\frac{\partial\mu_i}{\partial\beta}}=\sum_{i=1}^{n}\frac{(y_i-\mu_i)}{(1+a\mu_i)}\mathbf{x}_i$ where $\mathbf{x}_i$ is the transpose of a row of X. The additional equation for the variance parameter is denoted $\ddot{U}=\ddot{U}(a)=\ddot{U}(a;\beta)=\sum_{i=1}^{n}\ddot{u}_i$ and the form of this equation depends on the estimation method chosen.

Let $\hat{\beta}$ denote the solution to $\underset{px1}{U}(\beta;a)=0$ and $\tilde{\beta}$ denote the restricted estimate under H_0. In the following discussion, it is assumed that there is a unique β^* in the parameter space such that $\hat{\beta}\xrightarrow{p}\beta^*$ as $n\to\infty$ and $\tilde{\beta}\xrightarrow{p}\beta^*$ under H_0 as $n\to\infty$. In addition, assume that $\hat{a}$ converges to the solution of the limiting equation $\lim_n n^{-1}\ddot{U}(a^*;\beta^*)=0$.

Under suitable regularity conditions, distributional properties of the joint solution $\hat{\theta}=(\hat{\beta},\hat{a})$ to the "score" equations, $S(\theta)=\begin{bmatrix}U\\ \ddot{U}\end{bmatrix}_{(p+1)\times 1}$, follow from the asymptotic theory of estimating equations. The theory follows from a Taylor

expansion of the score functions, $0 = S(\hat{\theta}) = S(\theta^*) + \frac{\partial S(\theta^*)}{\partial \theta^t}(\hat{\theta} - \theta^*) + R_{n1}$. Because the laws of large numbers guarantee that $\left[E\left(\frac{\partial S}{\partial \theta^t}\right)\right]^{-1} \frac{\partial S}{\partial \theta^t} \xrightarrow{p} I$, the identity matrix, we are lead to $\hat{\theta} - \theta^* = \left[E\left(\frac{\partial S}{\partial \theta}\right)\Big|_{\theta=\theta^*}\right]^{-1} S(\theta^*) + R_{n2}$. Under suitable regularity conditions such that R_{n2} is asymptotically negligible, we have that $\hat{\theta}$ is asymptotically normal $\left(\theta^*, \overset{*}{V}\right)$ where $V = \left[E\left(\frac{\partial \mathbf{S}}{\partial \theta^t}\right)\right]^{-1}\left[\sum_{i=1}^{n} \mathbf{s}_i(Y_i, \mathbf{x}_i, \theta)\mathbf{s}_i(Y_i, \mathbf{x}_i, \theta)^t\right]\left[E\left(\frac{\partial \mathbf{S}}{\partial \theta^t}\right)\right]^{-1}$. Let $\overset{*}{V}, \hat{V}, \tilde{V}$ be the values of V when $\theta = \overset{*}{\theta}, \hat{\theta}, \tilde{\theta}$, respectively. Similar notation will be used for other (partitioned) matrices and vectors.

Let summation, unless otherwise noted, be over the n observations. Adapting the above paragraph to the NB setting, denote the empirical covariance matrix of the NB scores as $\begin{bmatrix} G & H^t \\ H & J \end{bmatrix} = \left[\sum_{i=1}^{n} \mathbf{s}_i(Y_i, x_i, \theta)\mathbf{s}_i(Y_i, x_i, \theta)^t\right]$ where $G = \sum \mathbf{u}\mathbf{u}^t = \sum \frac{(y_i - \mu_i)^2 \mu_i^2}{V_i^2} \mathbf{x}_i \mathbf{x}_i^t = X^t \left[diag \frac{(y_i - \mu_i)^2}{(1 + a\mu_i)^2}\right] X$, $H = \sum i\mathbf{u}^t$ and $J = \sum iii^t$. The negative expectation of their partial derivatives is $\begin{bmatrix} A & B \\ C & D \end{bmatrix}$ or $-E\left(\frac{\partial S}{\partial \theta^t}\right)$ where $A = -\sum E\frac{\partial U}{\partial \beta^t} = X^t\left[diag\left(\frac{\mu_i^2}{V_i}\right)\right]X = \sum \frac{\mu_i^2}{V_i^2}\mathbf{x}_i\mathbf{x}_i^t$, $B = -\sum E\frac{\partial U}{\partial a} = 0$, $C = -\sum E\frac{\partial i\mathbf{u}}{\partial \beta}$ and $D = -\sum E\frac{\partial i\mathbf{u}}{\partial a}$. In the evaluation of C and D, assume that the variance function has been correctly specified.

The joint estimate $(\hat{\beta}, \hat{a})$ is asymptotically normal with mean $\left(\overset{*}{\beta}, \overset{*}{a}\right)$ and a covariance matrix of $\begin{bmatrix} A^{-1} & 0 \\ -D^{-1}CA^{-1} & D^{-1} \end{bmatrix}\begin{bmatrix} G & H^t \\ H & J \end{bmatrix}\begin{bmatrix} A^{-1} & 0 \\ -D^{-1}CA^{-1} & D^{-1} \end{bmatrix}$. Regardless of the choice of $\ddot{U}(a)$, the fact that B=0 allows us to conclude that the asymptotic

variance of $\hat{\beta}$ is estimable by $A^{-1}GA^{-1}$ even if the variance is misspecified. This is the "empirical covariance matrix" reported by Breslow (1989). If the variance is correctly specified then $E(G) = A$ and we may estimate var(β) by A^{-1}, which is referred to as "model based".

The fact that B=0 together with sufficient regularity conditions implies that for any sequence of estimates $\tilde{\beta}_n$ converging to β^* the statistics $U(\tilde{\beta};\hat{a})$ and $U(\tilde{\beta};a^*)$ are asymptotically equivalent. Hence, in the following derivations we may assume that a is fixed at its limiting value. Substitution of $\hat{a}$ for a^* does not effect the asymptotic distribution of the test statistics based on U.

For a test on a subset of the mean parameters, $H_0 : \beta_2 = \beta_2^0$, partition the the regression coefficients into the appropriate subvectors of lengths say p_1 and p_2, $\beta_{px1} = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}_{p2\,x\,1}^{p1\,x\,1}$. The matrix of covariables and vector of "score" functions are similarly partitioned as $\underset{n\,x\,p}{X} = \left[\underset{n\,x\,p1}{X_1} \middle| \underset{n\,x\,p2}{X_2} \right]$ and

$$\underset{p\,x\,1}{U}(\beta) = \begin{bmatrix} U_1(\beta) \\ U_2(\beta) \end{bmatrix}_{p2\,x\,1}^{p1\,x\,1} = \begin{bmatrix} X_1^t \left\{ diag\left(\frac{1}{1+a\mu_i} \right) \right\} (y_i - \mu_i) \\ X_2^t \left\{ diag\left(\frac{1}{1+a\mu_i} \right) \right\} (y_i - \mu_i) \end{bmatrix}.$$ The generalized score tests are based on an expansion of U_2 under H_0. Thus, the asymptotic distribution will be normal with mean and variance determined by the expansion and partitioned matrices described below.

Under $H_0: \beta_2 = \beta_2^0$, $\tilde{\beta}^0 = \begin{bmatrix} \tilde{\beta}_1^0 \\ \tilde{\beta}_2^0 \end{bmatrix}^{p1 \times 1}_{p2 \times 1}$ where $\tilde{\beta}_2^0$ is fixed at β_2^0 and $\tilde{\beta}_1^0$ is the subsequent solution to $U_1 = \sum \frac{(y_i - \mu_i)\mu_i}{V_i} \mathbf{x}_{i1} = 0$. An expansion of $\underset{p \times 1}{U}(\tilde{\beta}^0)$ about $\underset{p \times 1}{U}(\beta^*)$ is

$$\begin{bmatrix} U_1(\tilde{\beta}^0) \\ U_2(\tilde{\beta}^0) \end{bmatrix} = \begin{bmatrix} 0 \\ U_2(\tilde{\beta}^0) \end{bmatrix} = \begin{bmatrix} U_1(\beta^*) \\ U_2(\beta^*) \end{bmatrix} + \begin{bmatrix} \frac{\partial U_1(\beta)}{\partial \beta_1} & \frac{\partial U_1(\beta)}{\partial \beta_2} \\ \frac{\partial U_2(\beta)}{\partial \beta_1} & \frac{\partial U_2(\beta)}{\partial \beta_2} \end{bmatrix}_{\beta=\beta^*} \begin{pmatrix} \tilde{\beta}_1 - \beta_1^* \\ 0 \end{pmatrix} + R_n.$$

Replacing the matrix of derivatives by their asymptotic equivalents, the partitions of the A matrix (understood as evaluated at $\beta = \beta^*$) we get

$\begin{bmatrix} 0 \\ U_2(\tilde{\beta}^0) \end{bmatrix} = \begin{bmatrix} U_1(\beta^*) \\ U_2(\beta^*) \end{bmatrix} - \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{pmatrix} \tilde{\beta}_1 - \beta_1^* \\ 0 \end{pmatrix} + R_{n3}$. Thus

$0 = U_1(\beta^*) - A_{11}(\tilde{\beta}_1 - \beta_1^*) + R_{n4}$, and an approximation for $\tilde{\beta}_1 - \beta_1^*$ is $A_{11}^{-1} U_1(\beta^*)$.

Using this approximation, the corresponding expansion for U_2 is

$U_2(\tilde{\beta}^0) \approx U_2(\beta^*) - A_{21}A_{11}^{-1}U_1(\beta^*) = \left[-A_{21}A_{11}^{-1} \middle| I_{p2}\right] \begin{bmatrix} U_1(\beta^*) \\ U_2(\beta^*) \end{bmatrix}$. Since the covariance of U is G (understood as evaluated at $\beta = \beta^*$), the asymptotic covariance of U_2 is

$$\left[-A_{21}A_{11}^{-1} \middle| I_{p2}\right] \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix} \begin{bmatrix} -A_{11}^{-1}A_{12} \\ I_{p2} \end{bmatrix} = G_{22} - G_{21}A_{11}^{-1}A_{12} - A_{21}A_{11}^{-1}G_{12} + A_{21}A_{11}^{-1}G_{11}A_{11}^{-1}A_{12}.$$

The test statistic is $T_{GS} = U_2^t \tilde{\Sigma}^{-1} U_2$, where $\tilde{\Sigma} = \tilde{\Sigma}^e$ or $\tilde{\Sigma}^m$. Use $\tilde{\Sigma}^e = G_{22} - A_{21}A_{11}^{-1}G_{12} - G_{21}A_{11}^{-1}A_{12} + A_{21}A_{11}^{-1}G_{11}A_{11}^{-1}A_{12}$ for an "empirical score test" and $\tilde{\Sigma}^m = A_{22} - A_{21}A_{11}^{-1}A_{12}$ for a "model based score test". Again, estimation of the parameters is done under H_0.

For the treatment versus control setting with n_1 observations from control and n_2 observations from treatment where $n=n_1+n_2$, the matrices defined above have the following form. Let $\bar{y}_t$, $\bar{y}_c$, $\bar{y}$ denote the sample means for treatment, control and the combined (total) sample and $\mathbf{r}$ denote the vector of residuals, i.e., $r_i = y_i - \bar{y}$. Then,

$$\tilde{A} = X^t\left[diag\left(\frac{\tilde{\mu}_i^2}{\tilde{V}_i}\right)\right]X = \frac{\bar{y}}{1+\tilde{a}\bar{y}}X^tX = \frac{\bar{y}}{1+\tilde{a}\bar{y}}\begin{bmatrix} X_1^tX_1 & X_1^tX_2 \\ X_2^tX_1 & X_2^tX_2 \end{bmatrix} = \frac{\bar{y}}{1+\tilde{a}\bar{y}}\begin{bmatrix} n & n_1-n_2 \\ n_1-n_2 & n \end{bmatrix}$$

$$\tilde{G} = X^t\left[diag\left(\frac{(y_i-\tilde{\mu}_i)^2\tilde{\mu}_i^2}{\tilde{V}_i^2}\right)\right]X = \frac{1}{(1+\tilde{a}\bar{y})^2}\left[X_1^t\middle|X_2^t\right]\left[diag\left(r_i^2\right)\right]\begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$$

$$= \frac{1}{(1+\tilde{a}\bar{y})^2}\begin{bmatrix} \sum_{total} r_i^2 & \sum_{control} r_i^2 - \sum_{treatment} r_i^2 \\ \sum_{control} r_i^2 - \sum_{treatment} r_i^2 & \sum_{total} r_i^2 \end{bmatrix} \text{ and}$$

$$U_2 = X_2^t\left[diag\left(\frac{1}{1+\tilde{a}\tilde{\mu}_i}\right)\right]\mathbf{r} = \left[n_1(\bar{y}_c-\bar{y}) - n_2(\bar{y}_t-\bar{y})\right]/(1+\tilde{a}\bar{y}).$$

Thus, $\tilde{\Sigma}^m = A_{22} - A_{21}A_{11}^{-1}A_{12} = \frac{\bar{y}}{1+\tilde{a}\bar{y}}\frac{n^2-(n_1-n_2)^2}{n} = \frac{\bar{y}}{1+\tilde{a}\bar{y}}\frac{4n_1n_2}{n}$ and

$$\tilde{\Sigma}^e = G_{22} - A_{21}A_{11}^{-1}G_{12} - G_{21}A_{11}^{-1}A_{12} + A_{21}A_{11}^{-1}G_{11}A_{11}^{-1}A_{12}$$

$$= \frac{1}{(1+\tilde{a}\bar{y})^2}\left[\sum_{total} r_i^2 - 2\frac{n_1-n_2}{n}\left(\sum_{control} r_i^2 - \sum_{treatment} r_i^2\right) + \frac{(n_1-n_2)^2}{n^2}\sum_{total} r_i^2\right]$$

$$= \frac{4}{n^2(1+\tilde{a}\bar{y})^2}\left[n_2^2\sum_{control} r_i^2 + n_1^2\sum_{treatment} r_i^2\right]$$

The model based score test is $T_{GS}^m = \frac{\left[n_1(\bar{y}_c-\bar{y}) - n_2(\bar{y}_t-\bar{y})\right]^2}{4n_1n_2\bar{y}(1+\tilde{a}\bar{y})/n}$ where $\tilde{a}$ is the null hypothesis estimator for a chosen method. The empirical score test is $T_{GS}^e = \frac{n^2\left[n_1(\bar{y}_c-\bar{y}) - n_2(\bar{y}_t-\bar{y})\right]^2}{4n_2^2\sum_{control} r_i^2 + 4n_1^2\sum_{treatment} r_i^2}$ for all methods. When $n_1 = n_2 = n/2$,

$T_{GS}^m = \frac{n(\bar{y}_c-\bar{y}_t)^2}{4\bar{y}(1+\tilde{a}\bar{y})}$ and $T_{GS}^e = \frac{n^2(\bar{y}_c-\bar{y}_t)^2}{4\sum_{total} r_i^2}$.

The model based score test with $n_1 = n_2 = n/2$ and $\tilde{a}$ estimated via optimal quadratic estimation gives the same test statistic as the empirical score test. This follows directly from substitution and algebra using $\tilde{a}_{OQ} = \frac{\left(\frac{n-1}{n}\right)S^2 - \bar{y}}{\bar{y}^2} = \frac{\frac{1}{n}\sum_{total} r_i^2 - \bar{y}}{\bar{y}^2}$ in T_{GS}^m yeilds T_{GS}^e.

APPENDIX F.

TABLES

Table 1: Null Rates using asymptotic distribution critical values.

n	a	μ	EQL.S	PL.S	ML.S	OQ.S	CML.S	DDT	LRT	t
30	0.2	1	0.0490	0.0509	0.0540	0.0556	0.0489	0.0308	0.0458	0.0502
		2	0.0453	0.0473	0.0482	0.0504	0.0452	0.0295	0.0450	0.0494
		5	0.0519	0.0516	0.0550	0.0547	0.0520	0.0345	0.0583	0.0514
	0.5	1	0.0452	0.0508	0.0501	0.0544	0.0452	0.0444	0.0555	0.0502
		2	0.0472	0.0505	0.0527	0.0545	0.0473	0.0440	0.0574	0.0498
		5	0.0465	0.0505	0.0506	0.0535	0.0465	0.0473	0.0577	0.0489
	1	1	0.0421	0.0477	0.0479	0.0518	0.0423	0.0450	0.0567	0.0464
		2	0.0437	0.0477	0.0479	0.0509	0.0438	0.0516	0.0581	0.0501
		5	0.0449	0.0503	0.0500	0.0529	0.0454	0.0538	0.0585	0.0495
50	0.2	1	0.0476	0.0485	0.0496	0.0496	0.0480	0.0332	0.0465	0.0484
		2	0.0491	0.0506	0.0517	0.0528	0.0491	0.0394	0.0506	0.0471
		5	0.0493	0.0496	0.0514	0.0502	0.0493	0.0420	0.0539	0.0459
	0.5	1	0.0462	0.0484	0.0494	0.0498	0.0466	0.0403	0.0510	0.0476
		2	0.0466	0.0479	0.0492	0.0500	0.0469	0.0449	0.0530	0.0480
		5	0.0493	0.0497	0.0510	0.0518	0.0493	0.0467	0.0539	0.0489
	1	1	0.0501	0.0549	0.0536	0.0572	0.0506	0.0504	0.0582	0.0565
		2	0.0462	0.0501	0.0491	0.0512	0.0474	0.0442	0.0553	0.0483
		5	0.0429	0.0480	0.0462	0.0495	0.0436	0.0470	0.0524	0.0483

Notes: Null rates implies rejection rates (of H_0) by the testing methods for samples generated at H_0 settings. The rate for each testing method is calculated over all the samples for which it obtained NB tests. There were 10000 samples generated at each H_0 setting, less failures as listed in Table 3. The Welchs' t-test (t) has an approximate t distribution with df calculated by formula (see Testing Methods section); it (was compared to a t(α=0.05,df calculated), not an asymptotic critical value; the rate is the number of Welchs' t-statistics produced such that α<=0.05. The t rejection rate is listed for interest. The remainder of the tests have an asymptotic χ^2(df=1) distribution and are compared to a chi-square(df=1,α=.10)=2.7055 critical value wtih rejection if $\hat{\mu}_1 > \hat{\mu}_2$. and acceptance if $\hat{\mu}_1 < \hat{\mu}_2$.

Table 2: Empirical Critical Values.

n	a	μ	EQL.S	PL.S	ML.S	OQ.S	CML.S	DDT	LRT
30	0.2	1	2.6661	2.7282	2.7972	2.8337	2.6625	2.1208	2.5882
		2	2.5422	2.6364	2.6414	2.7125	2.5420	2.1240	2.6159
		5	2.7615	2.7594	2.8567	2.8560	2.7627	2.2511	2.9517
	0.5	1	2.5707	2.7224	2.7059	2.8215	2.5728	2.4893	2.8588
		2	2.6488	2.7100	2.7629	2.8056	2.6541	2.5036	2.9614
		5	2.6074	2.7103	2.7170	2.8038	2.6174	2.5679	2.9184
	1	1	2.5029	2.6567	2.6487	2.7576	2.5011	2.5369	2.9156
		2	2.5078	2.6416	2.6376	2.7347	2.5268	2.7428	2.9478
		5	2.5800	2.7113	2.7006	2.8048	2.5924	2.7746	2.9890
50	0.2	1	2.6127	2.6216	2.6725	2.6831	2.6114	2.1666	2.6100
		2	2.6835	2.7273	2.7464	2.7833	2.6841	2.4056	2.7277
		5	2.6906	2.6660	2.7445	2.7168	2.6916	2.4442	2.8438
	0.5	1	2.5950	2.6305	2.6833	2.7010	2.6031	2.4721	2.7521
		2	2.6027	2.6486	2.6804	2.7027	2.6133	2.5382	2.7990
		5	2.6526	2.6973	2.7289	2.7523	2.6662	2.5645	2.8595
	1	1	2.7062	2.8434	2.8094	2.9027	2.7258	2.7283	3.0187
		2	2.5973	2.7061	2.6890	2.7613	2.6166	2.5326	2.8699
		5	2.5257	2.6269	2.6139	2.6805	2.5554	2.6145	2.7635

Notes: These are the values of the test statistic for the given method which produces the desired test size (α=.05) for the samples generated at H_0 settings. There were 10000 samples generated at each H_0 setting, less failures as listed in Table 3. Compare these to the asymptotic critical values of χ^2(df=1 ,α=.10)=2.7055.

Table 3: Counts of non-NB estimates at H_0 settings out of 10000.

			$\tilde{a}\leq 0$ (H_0)						$\tilde{a}\leq 0$ (H_1)		no test		discard
n	a	μ	EQL	PL	ML	OQ	CML	NED	ML	NED	DDT	t	$s^2<\bar{y}$
30	0.2	1	57	0	977	978	57	1211	505	1971	9	0	4274
		2	15	0	371	374	15	1082	430	1601	3	0	1772
		5	1	0	45	45	1	473	78	1245	0	0	148
	0.5	1	28	0	399	399	28	937	287	1328	17	0	1378
		2	1	0	67	67	1	477	100	789	0	0	196
		5	0	0	0	0	0	50	2	351	0	0	0
	1	1	14	0	116	116	14	560	120	868	45	0	331
		2	0	0	8	8	0	176	16	430	7	0	15
		5	0	0	0	0	0	22	0	116	1	0	0
50	0.2	1	0	0	592	591	0	1003	349	1314	0	0	2140
		2	7	0	183	187	7	616	221	884	0	0	571
		5	0	0	5	5	0	101	7	225	0	0	18
	0.5	1	0	0	181	177	0	503	131	746	0	0	580
		2	0	0	10	11	0	118	22	207	0	0	28
		5	0	0	0	0	0	2	0	19	0	0	0
	1	1	0	0	24	24	0	192	32	353	0	0	79
		2	0	0	0	0	0	14	1	40	0	0	0
		5	0	0	0	0	0	0	0	0	0	0	0

Notes (Table 3 and Table 4): The terms "H_0 settings" (Table 3) and "Power settings" (Table 4) refer to samples (of size n) generated from parameter settings of NB($\mu_1=\mu_2=\mu,a$) and NB($\mu_1>\mu_2,a$) respectively. There were a total of 10000* samples (of size n) generated at H_0 settings and 5000* samples (of size n) generated at Power settings. *These are the totals after discarding underdispersed samples, i.e., those with $s^2<\bar{y}$. Counts of the discards are in right-most column. The section denoted " $\tilde{a}\leq 0$ (H_0)" are counts of samples producing a non-NB variance parameter estimate, ($\tilde{a}\leq 0$), during estimation under H_0 assumptions. PL has no failures because samples that would have produced failures for PL are those which were discarded. The discarded samples would have produced non-NB estimates for OQ and ML, and (probably) for EQL, CML and NED, as well. The section denoted " $\tilde{a}\leq 0$ (H_1)" are counts of samples producing a non-NB result, ($\tilde{a}\leq 0$), under H_1 assumptions; these same samples had produced a NB result $\tilde{a} > 0$ under H_0 assumptions. The category of "no test" for DDT are the counts of samples for which non-NB results were produced under H_0 **or** H_1 assumptions, followed by non-Poisson results (one of the subsamples of size n/2 produced a mean estimate of $\tilde{\mu} = 0$). The category of "no test" for LRT are samples that produced non-NB results under H_1 assumptions but samples had not been saved for further evaluation. The t test does not involve estimation of *a* and obtains results for all samples. For further information, see discussion in Methods section.

Table 4: Counts of non-NB estimates at Power settings out of 5000.

simulation settings				$\tilde{a}\leq 0$ (H_0)						$\tilde{a}\leq 0$ (H_1)		no test			discard
n	a	μ2	μ1	EQL	PL	ML	OQ	CML	NED	ML	NED	LRT	DDT	t	$s^2<\bar{y}$
30	0.2	1	1.75	12	0	232	232	12	598	744	1235	0	2	0	1001
			2	4	0	128	129	4	514	826	1268	13	0	0	669
			2.5	5	0	57	58	5	349	928	1376	0	2	0	217
		2	3.5	0	0	56	57	0	315	424	953	9	0	0	203
			4	0	0	21	21	0	252	426	1063	31	0	0	16
		5	7.5	0	0	6	6	0	137	41	669	7	0	0	10
			8.75	0	0	1	1	0	86	36	682	0	0	0	3
			10	0	0	0	0	0	45	28	668	105	0	0	0
	0.5	1	2	4	0	50	50	4	318	219	612	0	4	0	143
			2.5	2	0	25	25	2	221	217	609	0	2	0	49
			3	0	0	4	4	0	154	157	479	0	5	0	15
		2	4	0	0	2	2	0	112	35	306	0	0	0	6
			5	0	0	2	2	0	60	21	300	0	0	0	0
		5	10	0	0	0	0	0	12	0	202	0	0	0	0
			12.5	0	0	0	0	0	12	0	238	0	0	0	0
	1	1	2	0	0	15	15	0	188	43	326	0	13	0	27
			2.5	0	0	10	10	0	153	28	280	0	14	0	8
			3	0	0	3	3	0	65	10	251	0	14	0	8
			3.5	0	0	0	0	0	74	12	171	0	11	0	3
		2	4	0	0	0	0	0	45	1	124	0	1	0	0
			5	0	0	0	0	0	24	1	128	0	1	0	0
			6	0	0	0	0	0	27	0	106	0	1	0	0
		5	10	0	0	0	0	0	5	0	86	0	0	0	0
			15	0	0	0	0	0	13	0	138	0	0	0	0
50	0.2	1	1.5	1	0	143	143	1	392	434	789	0	0	0	780
			1.75	1	0	87	87	1	364	541	871	0	0	0	478
			2	0	0	49	49	0	259	642	981	2	0	0	246
		2	3	1	0	24	24	1	184	141	438	0	0	0	112
			3.5	0	0	8	8	0	97	184	493	0	0	0	39
			4	0	0	5	5	0	62	145	414	56	0	0	19
		5	7.5	0	0	0	0	0	8	5	101	0	0	0	0
			8.75	0	0	0	0	0	4	1	95	0	0	0	0
			10	0	0	0	0	0	2	2	90	0	0	0	0
	0.5	1	1.75	0	0	19	19	0	137	84	296	0	0	0	51
			2	0	0	2	2	0	112	69	240	0	0	0	28
			2.5	1	0	2	2	1	49	32	196	0	1	0	7
		2	3.5	0	0	0	0	0	20	4	54	0	0	0	0
			4	0	0	0	0	0	10	5	42	0	0	0	0
		5	7.5	0	0	0	0	0	0	0	3	0	0	0	0
			8.75	0	0	0	0	0	0	0	3	0	0	0	0
			10	0	0	0	0	0	0	0	3	0	0	0	0
	1	1	2	0	0	0	0	0	24	2	61	0	0	0	4
			2.5	0	0	0	0	0	20	2	50	0	0	0	0
			3	0	0	0	0	0	17	1	27	0	0	0	0
		2	4	0	0	0	0	0	4	0	11	0	0	0	0
			5	0	0	0	0	0	1	0	4	0	0	0	0
		5	8.75	0	0	0	0	0	0	0	2	0	0	0	0
			10	0	0	0	0	0	0	0	4	0	0	0	0
			12.5	0	0	0	0	0	1	0	3	0	0	0	0
			15	0	0	0	0	0	0	0	3	0	0	0	0

Notes: See Notes for Table 3.

Table 5: MRSE Significance Results comparing NED vs. other estimators of *a*. Comparison over samples of size n=30 which produce NB results for all methods

	μ_2	μ_1	var2	var1	runs	EQL	PL	ML	OQ	CML	NED
0.2	1	1	1.2	1.2	8062	0.6195	**0.5791**	0.5975	**0.5716**	0.6187	*0.5927*
	2	2	2.8	2.8	8686	0.5300	*0.5150*	*0.5209*	*0.5133*	0.5297	*0.5176*
	5	5	10	10	9507	**0.4569**	**0.4577**	**0.4543**	**0.4570**	**0.4565**	*0.4650*
0.5	1	1	1.5	1.5	8765	0.6854	**0.6567**	0.6721	**0.6540**	0.6840	*0.6653*
	2	2	4	4	9491	*0.6124*	**0.6065**	**0.6059**	*0.6079*	*0.6105*	*0.6126*
	5	5	17.5	17.5	9950	**0.5463**	**0.5512**	**0.5433**	**0.5528**	**0.5444**	*0.5651*
1	1	1	2	2	9349	*0.7760*	**0.7563**	**0.7670**	**0.7613**	*0.7761*	*0.7721*
	2	2	6	6	9819	**0.7030**	**0.7068**	**0.6971**	*0.7101*	**0.7011**	*0.7172*
	5	5	30	30	9978	**0.6425**	**0.6658**	**0.6361**	*0.6687*	**0.6393**	*0.6735*
0.2	1	1.75	1.2	2.36	4249	0.6146	*0.5829*	0.59312	**0.5695**	0.6142	*0.5806*
		2		2.8	4408	0.6227	*0.5915*	0.59923	*0.5775*	0.6223	*0.5845*
		2.5		3.75	4617	0.6612	0.6287	0.6361	*0.6107*	0.6605	*0.6133*
	2	3.5	2.8	5.95	4653	0.5461	0.5356	0.5303	0.5251	0.5456	*0.5175*
		4		7.2	4734	0.5613	0.5473	0.5442	0.5353	0.5607	*0.5259*
	5	7.5	10	18.8	4860	*0.4646*	*0.4661*	*0.4569*	*0.4600*	*0.4640*	*0.4600*
		8.75		24.1	4914	0.4882	0.4885	0.4750	0.4769	0.4875	*0.4600*
		10		30	4955	0.5339	0.5341	0.5173	0.5185	0.5329	*0.4747*
0.5	1	2	1.5	4	4647	0.6771	*0.6546*	0.6578	*0.6461*	0.6748	*0.6503*
		2.5		5.63	4767	0.7049	0.6796	0.6802	*0.6677*	0.7022	*0.6583*
		3		7.5	4842	0.7317	0.7000	0.7025	0.6824	0.7286	*0.6691*
	2	4	4	12	4886	0.6226	0.6186	0.6051	0.6105	0.6198	*0.5947*
		5		17.5	4939	0.6550	0.6493	0.6318	0.6354	0.6515	*0.5992*
	5	10	17.5	60	4988	0.5733	0.5878	0.5565	0.5777	0.5698	*0.5434*
		12.5		90.6	4988	0.6219	0.6353	0.5996	0.6198	0.6175	*0.5406*
1	1	2	2	6	4800	0.7650	*0.7437*	*0.7454*	*0.7393*	0.7628	*0.7457*
		2.5		8.75	4838	0.7676	0.7483	0.7433	*0.7406*	0.7644	*0.7320*
		3		12	4933	0.7921	0.7680	0.7615	0.7547	0.7880	*0.7310*
		3.5		15.8	4926	0.8049	0.7872	0.7735	0.7722	0.8002	*0.7318*
	2	4	6	20	4955	*0.6955*	0.7046	**0.6768**	0.6998	*0.6910*	*0.6860*
		5		30	4976	0.7189	0.7277	0.6969	0.7182	0.7137	*0.6821*
		6		42	4973	0.7442	0.7500	0.7156	0.7364	0.7376	*0.6822*
	5	10	30	110	4995	**0.6496**	*0.6755*	**0.6310**	*0.6692*	**0.6434**	*0.6642*
		15		240	4987	0.6989	0.7380	0.6725	0.7228	0.6899	*0.6550*

Notes: Coded MRSE results based on Hotelling's T^2 test.
MRSE values for each estimation procedure are listed in rows along with the simulation (parameter) setting for generating the samples.
"runs" = number of samples of size n=30 where all estimation methods obtain NB values ($\hat{a}>0$) during estimation under H_0 assumptions. These are the samples over which the comparison is made.
NED and all tests with MRSE results not significantly different from NED are listed in italics and shaded in gray. Tests with MRSE significantly less (better) than NED are listed as bold and underlined.
Tests with MRSE significantly greater (worse) than NED are listed in plain text.
A summary of the RSE correlation matrices between the estimators across all the simulation settings in Table 5 is found in Table 9.

Table 6: MRSE Significance Results comparing NED vs. other estimators of *a*. Comparison over samples of size n=50 which produce NB results for all methods.

a	μ_2	μ_1	var2	var1	runs	EQL	PL	ML	OQ	CML	NED
0.2	1	1	1.2	1.2	8558	0.5711	*0.5502*	0.5577	**0.5454**	0.5692	*0.5521*
	2	2	2.8	2.8	9263	*0.4946*	*0.4910*	*0.4906*	*0.4894*	*0.4937*	*0.4928*
	5	5	10	10	9899	**0.4295**	**0.4313**	**0.4285**	**0.4324**	**0.4290**	*0.4374*
0.5	1	1	1.5	1.5	9351	0.6389	**0.6285**	*0.6312*	*0.6287*	*0.6356*	*0.6344*
	2	2	4	4	9875	**0.5713**	**0.5727**	**0.5688**	**0.5738**	**0.5691**	*0.5813*
	5	5	17.5	17.5	9998	**0.5087**	*0.5194*	**0.5056**	*0.5203*	**0.5068**	*0.5215*
1	1	1	2	2	9790	*0.7225*	*0.7178*	**0.7147**	*0.7197*	*0.7194*	*0.7237*
	2	2	6	6	9986	**0.6539**	*0.6658*	**0.6486**	*0.6680*	**0.6507**	*0.6636*
	5	5	30	30	10000	**0.5933**	0.6232	**0.5892**	0.6250	**0.5907**	*0.6153*
0.2	1	1.5	1.2	1.95	4504	0.5649	*0.5474*	0.5528	*0.5427*	0.5631	*0.5454*
		1.75		2.36	4580	0.5757	0.5620	0.5617	*0.5542*	0.5739	*0.5507*
		2		2.8	4710	0.5876	0.5721	0.5714	*0.5627*	0.5858	*0.5597*
	2	3	2.8	4.8	4798	0.4961	0.4927	*0.4876*	*0.4883*	0.4950	*0.4865*
		3.5		5.95	4899	0.5162	0.5141	0.5052	0.5057	0.5150	*0.4975*
		4		7.2	4933	0.5470	0.5404	0.5331	0.5296	0.5457	*0.5131*
	5	7.5	10	18.8	4992	0.4421	0.4444	0.4341	0.4389	0.4413	*0.4281*
		8.75		24.1	4996	0.4767	0.4815	0.4648	0.4725	0.4757	*0.4381*
		10		30	4998	0.5239	0.5289	0.5107	0.5176	0.5227	*0.4718*
0.5	1	1.75	1.5	3.28	4849	0.6311	*0.6214*	*0.6180*	*0.6174*	0.6272	*0.6169*
		2		4	4888	0.6375	0.6254	0.6209	*0.6193*	0.6331	*0.6144*
		2.5		5.63	4950	0.6677	0.6556	0.6470	0.6452	0.6626	*0.6272*
	2	3.5	4	9.63	4980	0.5687	0.5701	*0.5581*	0.5652	0.5652	*0.5546*
		4		12	4990	0.5897	0.5928	0.5747	0.5855	0.5858	*0.5620*
	5	7.5	17.5	35.6	5000	*0.5103*	0.5278	**0.5032**	0.5246	*0.5076*	*0.5104*
		8.75		47	5000	0.5228	0.5404	0.5111	0.5339	0.5194	*0.5008*
		10		60	5000	0.5511	0.5694	0.5358	0.5607	0.5469	*0.5037*
1	1	2	2	6	4976	0.7089	0.7103	*0.6917*	0.7065	0.7027	*0.6919*
		2.5		8.75	4980	0.7313	0.7289	0.7076	0.7214	0.7238	*0.6930*
		3		12	4983	0.7609	0.7606	0.7337	0.7503	0.7523	*0.6932*
	2	4	6	20	4996	0.6552	0.6739	*0.6392*	0.6691	0.6485	*0.6364*
		5		30	4999	0.6839	0.7038	0.6598	0.6946	0.6756	*0.6295*
	5	8.75	30	85.3	5000	*0.6052*	0.6412	**0.5920**	0.6369	*0.5990*	*0.6065*
		10		110	5000	*0.6100*	0.6480	*0.5943*	0.6418	*0.6030*	*0.6019*
		12.5		169	4999	0.6417	0.6918	0.6205	0.6820	0.6327	*0.5930*
		15		240	5000	0.6826	0.7394	0.6588	0.7273	0.6726	*0.5899*

Notes: Coded MRSE results based on Hotelling's T^2 test.
MRSE values for each estimation procedure are listed in rows along with the simulation (parameter) setting for generating the samples.
"runs" = number of samples of size n=50 where all estimation methods obtain NB values ($\hat{a}>0$) during estimation under H_0 assumptions. These are the samples over which the comparison is made.
NED and all tests with MRSE results not significantly different from NED are listed in italics and shaded in gray. Tests with MRSE significantly less (better) than NED are listed as bold and underlined.
Tests with MRSE significantly greater (worse) than NED are listed in plain text.
A summary of the RSE correlation matrices between the estimators across all the simulation settings in Table 6 is found in Table 10.

Table 7: MRSE Significance Results comparing NED vs. other estimators of *a*. Comparison over all samples of size n=30 using $\hat{a}$=0 for negative estimates.

a	μ_2	μ_1	var2	var1	runs	EQL	PL	ML	OQ	CML	NED
0.2	1	1	1.2	1.2	10000	0.6112	**0.5842**	*0.5956*	**0.5789**	0.6105	*0.5983*
	2	2	2.8	2.8	10000	*0.5301*	**0.5184**	**0.5237**	**0.5181**	*0.5297*	*0.5323*
	5	5	10	10	10000	**0.4597**	**0.4604**	**0.4576**	**0.4601**	**0.4592**	*0.4732*
0.5	1	1	1.5	1.5	10000	*0.6862*	**0.6648**	**0.6756**	**0.6628**	*0.6848*	*0.684*
	2	2	4	4	10000	**0.6146**	**0.6097**	**0.6091**	**0.6112**	**0.6128**	*0.6232*
	5	5	17.5	17.5	10000	**0.5466**	**0.5515**	**0.5437**	**0.5531**	**0.5447**	*0.5664*
1	1	1	2	2	10000	**0.7779**	**0.7618**	**0.7702**	**0.7664**	**0.7781**	*0.7868*
	2	2	6	6	10000	**0.7039**	**0.7081**	**0.6983**	**0.7114**	**0.7020**	*0.7223*
	5	5	30	30	10000	**0.6426**	**0.6659**	**0.6363**	*0.6688*	**0.6394**	*0.6742*
0.2	1	1.75	1.2	2.36	5000	0.6047	*0.5833*	*0.587*	**0.5713**	0.6043	*0.5877*
		2		2.8	5000	0.6141	*0.5914*	*0.5929*	**0.5786**	0.6137	*0.5898*
		2.5		3.75	5000	0.6523	0.6266	0.6288	*0.6093*	0.6516	*0.6146*
	2	3.5	2.8	5.95	5000	0.5439	0.5365	*0.5294*	*0.5266*	0.5434	*0.5254*
		4		7.2	5000	0.5582	0.5480	0.5422	*0.5363*	0.5576	*0.5314*
	5	7.5	10	18.8	5000	*0.4651*	*0.4668*	**0.4579**	*0.4611*	*0.4645*	*0.4648*
		8.75		24.1	5000	0.4876	0.4884	0.4747	0.4769	0.4869	*0.463*
		10		30	5000	0.5333	0.5340	0.5168	0.5186	0.5323	*0.4762*
0.5	1	2	1.5	4	5000	0.6768	*0.6598*	*0.6591*	**0.6516**	0.6747	*0.6622*
		2.5		5.63	5000	0.7017	0.6807	0.6786	*0.6691*	0.6991	*0.6659*
		3		7.5	5000	0.7295	0.7012	0.7008	*0.6838*	0.7264	*0.6738*
	2	4	4	12	5000	0.6221	0.6200	*0.605*	0.6119	0.6194	*0.5998*
		5		17.5	5000	0.6540	0.6499	0.6310	0.6360	0.6506	*0.6019*
	5	10	17.5	60	5000	0.5731	0.5878	0.5564	0.5776	0.5696	*0.5441*
		12.5		90.6	5000	0.6220	0.6357	0.5997	0.6202	0.6175	*0.5413*
1	1	2	2	6	5000	*0.7646*	*0.7467*	**0.7461**	**0.7423**	*0.7623*	*0.7559*
		2.5		8.75	5000	0.7676	0.7528	*0.7436*	*0.7446*	0.7643	*0.7406*
		3		12	5000	0.7917	0.7694	0.7613	0.7561	0.7876	*0.7346*
		3.5		15.8	5000	0.8045	0.7885	0.7733	0.7736	0.7998	*0.7358*
	2	4	6	20	5000	*0.6958*	0.7052	**0.6772**	0.7005	*0.6913*	*0.6889*
		5		30	5000	0.7191	0.7286	0.6971	0.7190	0.7139	*0.6837*
		6		42	5000	0.7444	0.7508	0.7159	0.7371	0.7378	*0.6839*
	5	10	30	110	5000	**0.6495**	*0.6755*	**0.6309**	*0.6692*	**0.6432**	*0.6646*
		15		240	5000	0.6991	0.7387	0.6727	0.7235	0.6900	*0.6559*

Notes: Coded MRSE results based on Hotelling's T^2 test.
MRSE values for each estimation procedure are listed in rows along with the simulation (parameter) setting for generating the samples.
"runs" = number of samples of size n=30 over which the comparison is made. Non-positive estimates obtained under H_0 assumptions by any method were replaced by $\hat{a}$=0.
NED and all tests with MRSE results not significantly different from NED are listed in italics and shaded in gray. Tests with MRSE significantly less (better) than NED are listed as bold and underlined.
Tests with MRSE significantly greater (worse) than NED are listed in plain text.

Table 8: MRSE Significance Results comparing NED vs. other estimators of *a*. Comparison over all samples of size n=50 using $\hat{a}$=0 for negative estimates.

a	μ_2	μ_1	var2	var1	runs	EQL	PL	ML	OQ	CML	NED
0.2	1	1	1.2	1.2	10000	0.5676	**0.5530**	**0.5569**	**0.5496**	*0.5658*	*0.5626*
	2	2	2.8	2.8	10000	**0.4972**	**0.4940**	**0.4942**	**0.4931**	**0.4962**	*0.5029*
	5	5	10	10	10000	**0.4306**	**0.4322**	**0.4297**	**0.4333**	**0.4301**	*0.4393*
0.5	1	1	1.5	1.5	10000	0.6423	**0.6332**	**0.6356**	**0.6338**	**0.6392**	*0.6463*
	2	2	4	4	10000	**0.5726**	**0.5736**	**0.5702**	**0.5748**	**0.5704**	*0.5843*
	5	5	17.5	17.5	10000	**0.5088**	*0.5194*	**0.5056**	*0.5203*	**0.5068**	*0.5216*
1	1	1	2	2	10000	**0.7243**	**0.7198**	**0.7169**	**0.7216**	**0.7213**	*0.7295*
	2	2	6	6	10000	**0.6540**	*0.6659*	**0.6487**	*0.6680*	**0.6507**	*0.6641*
	5	5	30	30	10000	**0.5933**	0.6232	**0.5892**	0.6250	**0.5907**	*0.6153*
0.2	1	1.5	1.2	1.95	5000	0.5606	*0.5477*	*0.5507*	**0.5438**	0.5590	*0.5535*
		1.75		2.36	5000	0.5717	*0.5611*	0.5593	*0.5542*	0.5700	*0.5573*
		2		2.8	5000	0.5837	0.5710	*0.5685*	*0.5621*	0.5819	*0.5637*
	2	3	2.8	4.8	5000	*0.4967*	*0.4942*	*0.4887*	*0.4901*	*0.4957*	*0.4923*
		3.5		5.95	5000	0.5152	0.5138	*0.5046*	*0.5056*	0.5140	*0.5002*
		4		7.2	5000	0.5464	0.5401	0.5327	0.5295	0.5451	*0.5147*
	5	7.5	10	18.8	5000	0.4421	0.4445	0.4341	0.4389	0.4413	*0.4284*
		8.75		24.1	5000	0.4766	0.4815	0.4647	0.4725	0.4756	*0.4383*
		10		30	5000	0.5239	0.5289	0.5107	0.5176	0.5227	*0.4719*
0.5	1	1.75	1.5	3.28	5000	0.6321	*0.6239*	*0.6196*	*0.6201*	*0.6282*	*0.6230*
		2		4	5000	0.6374	*0.6263*	*0.6211*	*0.6202*	0.6329	*0.6190*
		2.5		5.63	5000	0.6674	0.6562	0.6467	0.6459	0.6623	*0.6291*
	2	3.5	4	9.63	5000	0.5687	0.5703	*0.5582*	0.5655	0.5652	*0.5657*
		4		12	5000	0.5898	0.5931	0.5750	0.5858	0.5859	*0.5625*
	5	7.5	17.5	35.6	5000	*0.5103*	0.5278	**0.5032**	0.5246	*0.5076*	*0.5104*
		8.75		47	5000	0.5228	0.5404	0.5111	0.5339	0.5194	*0.5008*
		10		60	5000	0.5511	0.5694	0.5358	0.5607	0.5469	*0.5037*
1	1	2	2	6	5000	0.7088	0.7111	*0.6916*	0.7072	0.7027	*0.6934*
		2.5		8.75	5000	0.7310	0.7297	0.7074	0.7222	0.7235	*0.6942*
		3		12	5000	0.7607	0.7611	0.7336	0.7508	0.7522	*0.6942*
	2	4	6	20	5000	0.6552	0.6742	*0.6393*	0.6693	0.6485	*0.6367*
		5		30	5000	0.6839	0.7038	0.6598	0.6946	0.6756	*0.6295*
	5	8.75	30	85.3	5000	*0.6052*	0.6412	**0.5920**	0.6369	*0.5990*	*0.6065*
		10		110	5000	*0.6100*	0.6480	*0.5943*	0.6418	*0.6030*	*0.6019*
		12.5		169	5000	0.6417	0.6919	0.6206	0.6822	0.6328	*0.5931*
		15		240	5000	0.6826	0.7394	0.6588	0.7273	0.6726	*0.5899*

Notes: Coded MRSE results based on Hotelling's T^2 test.
MRSE values for each estimation procedure are listed in rows along with the simulation (parameter) setting for generating the samples. "runs" = number of samples of size n=50 over which the comparison is made.
Non-positive estimates obtained under H_0 assumptions by any method were replaced by $\hat{a}$=0.
NED and all tests with MRSE results not significantly different from NED are listed in italics and shaded in gray. Tests with MRSE significantly less (better) than NED are listed as bold and underlined.
Tests with MRSE significantly greater (worse) than NED are listed in plain text.

Table 9. Summary statistics across correlation matrices (R) produced for MRSE analyses in Table 5 on samples of size n=30.

average correlations:

	EQL	PL	ML	OQ	CML	NED
EQL	1.0000	0.7072	0.9236	0.6452	0.9980	0.5513
PL	0.7072	1.0000	0.7291	0.9428	0.7072	0.3903
ML	0.9236	0.7291	1.0000	0.7005	0.9293	0.6018
OQ	0.6452	0.9428	0.7005	1.0000	0.6461	0.3630
CML	0.9980	0.7072	0.9293	0.6461	1.0000	0.5577
NED	0.5513	0.3903	0.6018	0.3630	0.5577	1.0000

minimum correlations:

	EQL	PL	ML	OQ	CML	NED
EQL	1.0000	0.4183	0.8808	0.3714	0.9904	-0.0524
PL	0.4183	1.0000	0.4587	0.8765	0.4250	0.0615
ML	0.8808	0.4587	1.0000	0.4382	0.8874	0.0133
OQ	0.3714	0.8765	0.4382	1.0000	0.3827	0.0609
CML	0.9904	0.4250	0.8874	0.3827	1.0000	-0.0351
NED	-0.0524	0.0615	0.0133	0.0609	-0.0351	1.0000

maximum correlations:

	EQL	PL	ML	OQ	CML	NED
EQL	1.0000	0.9061	0.9701	0.8482	1.0000	0.8853
PL	0.9061	1.0000	0.9190	0.9744	0.9046	0.7401
ML	0.9701	0.9190	1.0000	0.8982	0.9700	0.8610
OQ	0.8482	0.9744	0.8982	1.0000	0.8460	0.6610
CML	1.0000	0.9046	0.9700	0.8460	1.0000	0.8810
NED	0.8853	0.7401	0.8610	0.6610	0.8810	1.0000

Notes: A separate correlation matrix (R) is produced for each MRSE analysis (row) in Table 5. The summary statistics (averages, minimums and maximums) were taken over the 33 R matrices.

Table 10. Summary statistics across correlation matrices (R) produced for MRSE analyses in Table 6 on samples of size n=50.

average correlations:

	EQL	PL	ML	OQ	CML	NED
EQL	1.0000	0.7005	0.9391	0.6639	0.9952	0.5745
PL	0.7005	1.0000	0.7135	0.9655	0.7030	0.3856
ML	0.9391	0.7135	1.0000	0.6974	0.9522	0.6176
OQ	0.6639	0.9655	0.6974	1.0000	0.6683	0.3695
CML	0.9952	0.7030	0.9522	0.6683	1.0000	0.5871
NED	0.5745	0.3856	0.6176	0.3695	0.5871	1.0000

minimum correlations:

	EQL	PL	ML	OQ	CML	NED
EQL	1.0000	0.4149	0.8935	0.3869	0.9807	0.0298
PL	0.4149	1.0000	0.4536	0.9339	0.4310	0.0654
ML	0.8935	0.4536	1.0000	0.4400	0.9224	0.0830
OQ	0.3869	0.9339	0.4400	1.0000	0.4070	0.0663
CML	0.9807	0.4310	0.9224	0.4070	1.0000	0.0480
NED	0.0298	0.0654	0.0830	0.0663	0.0480	1.0000

maximum correlations:

	EQL	PL	ML	OQ	CML	NED
EQL	1.0000	0.9023	0.9830	0.8612	1.0000	0.8686
PL	0.9023	1.0000	0.9125	0.9870	0.9022	0.6999
ML	0.9830	0.9125	1.0000	0.8934	0.9838	0.8362
OQ	0.8612	0.9870	0.8934	1.0000	0.8615	0.6448
CML	1.0000	0.9022	0.9838	0.8615	1.0000	0.8685
NED	0.8686	0.6999	0.8362	0.6448	0.8685	1.0000

Notes: A separate correlation matrix (R) is produced for each MRSE analysis (row) in Table 6. The summary statistics (averages, minimums and maximums) were taken over the 35 R matrices.

Table 11: Bias and MSE for estimators of *a*. Based on samples of size n=30 where all methods obtain NB results.

			Bias						MSE					
a	μ2	μ1	EQL	PL	ML	OQ	CML	NED	EQL	PL	ML	OQ	CML	NED
0.2	1	1	0.251	0.178	0.182	0.131	0.252	0.183	0.185	0.115	0.135	0.094	0.185	0.134
	2	2	0.080	0.053	0.047	0.028	0.080	0.033	0.042	0.033	0.034	0.029	0.042	0.032
	5	5	0.012	0.005	-0.003	-0.009	0.012	-0.029	0.012	0.012	0.011	0.011	0.012	0.011
0.5	1	1	0.190	0.079	0.103	0.025	0.191	0.075	0.286	0.186	0.223	0.168	0.290	0.208
	2	2	0.053	0.000	0.005	-0.034	0.051	-0.047	0.100	0.087	0.085	0.083	0.097	0.084
	5	5	0.014	-0.015	-0.016	-0.038	0.011	-0.086	0.040	0.043	0.036	0.042	0.039	0.043
1	1	1	0.185	-0.019	0.059	-0.087	0.188	-0.035	0.643	0.437	0.527	0.415	0.679	0.509
	2	2	0.064	-0.051	-0.018	-0.100	0.056	-0.150	0.263	0.252	0.226	0.243	0.259	0.238
	5	5	0.042	-0.035	-0.017	-0.074	0.030	-0.184	0.135	0.169	0.118	0.162	0.129	0.145
0.2	1	1.75	0.220	0.168	0.168	0.131	0.219	0.151	0.136	0.098	0.104	0.083	0.135	0.096
		2	0.239	0.190	0.189	0.154	0.238	0.162	0.143	0.107	0.110	0.090	0.141	0.098
		2.5	0.294	0.242	0.246	0.208	0.292	0.210	0.174	0.131	0.137	0.111	0.171	0.119
	2	3.5	0.120	0.103	0.093	0.081	0.120	0.057	0.050	0.046	0.040	0.040	0.049	0.032
		4	0.148	0.131	0.120	0.108	0.147	0.079	0.056	0.051	0.045	0.044	0.055	0.036
	5	7.5	0.050	0.047	0.035	0.033	0.049	-0.007	0.015	0.016	0.013	0.014	0.015	0.011
		8.75	0.085	0.084	0.070	0.069	0.085	0.020	0.020	0.022	0.017	0.019	0.020	0.012
		10	0.131	0.132	0.114	0.117	0.130	0.052	0.032	0.035	0.026	0.030	0.031	0.016
0.5	1	2	0.230	0.155	0.161	0.110	0.227	0.085	0.235	0.191	0.184	0.169	0.231	0.157
		2.5	0.323	0.258	0.253	0.213	0.318	0.147	0.300	0.271	0.235	0.237	0.293	0.180
		3	0.395	0.331	0.325	0.286	0.388	0.196	0.343	0.304	0.270	0.264	0.334	0.196
	2	4	0.169	0.146	0.124	0.113	0.164	0.016	0.117	0.124	0.094	0.109	0.113	0.076
		5	0.258	0.244	0.211	0.209	0.251	0.073	0.161	0.176	0.128	0.153	0.155	0.085
	5	10	0.136	0.147	0.105	0.121	0.131	-0.036	0.062	0.088	0.051	0.077	0.060	0.037
		12.5	0.221	0.248	0.187	0.220	0.215	0.010	0.095	0.141	0.077	0.122	0.091	0.039
1	1	2	0.270	0.134	0.164	0.073	0.265	-0.012	0.514	0.474	0.412	0.432	0.514	0.349
		2.5	0.356	0.247	0.251	0.185	0.348	0.026	0.556	0.546	0.440	0.488	0.552	0.332
		3	0.473	0.387	0.366	0.324	0.462	0.093	0.654	0.690	0.513	0.610	0.644	0.353
		3.5	0.540	0.465	0.433	0.401	0.527	0.123	0.709	0.751	0.557	0.661	0.694	0.355
	2	4	0.195	0.145	0.118	0.095	0.183	-0.100	0.250	0.331	0.200	0.299	0.240	0.180
		5	0.306	0.289	0.226	0.236	0.291	-0.048	0.335	0.454	0.265	0.403	0.320	0.186
		6	0.389	0.391	0.307	0.336	0.373	-0.005	0.392	0.561	0.308	0.494	0.374	0.195
	5	10	0.159	0.147	0.099	0.104	0.144	-0.161	0.155	0.247	0.125	0.221	0.145	0.130
		15	0.319	0.400	0.254	0.350	0.300	-0.126	0.241	0.473	0.189	0.415	0.224	0.123

Notes: Bias and MSE calculated over all samples of size n=30 which produced NB results ($\hat{a}>0$) for all estimation methods at a given simulation (parameter) setting. For counts of samples per setting, see "runs" column in Table 5.

Table 12: Bias and MSE for estimators of *a*. Based on samples of size n=50 where all methods obtain NB results.

			Bias						MSE					
a	μ_2	μ_1	EQL	PL	ML	OQ	CML	NED	EQL	PL	ML	OQ	CML	NED
0.2	1	1	0.146	0.110	0.108	0.083	0.144	0.106	0.080	0.060	0.064	0.053	0.078	0.064
	2	2	0.037	0.024	0.019	0.010	0.036	0.006	0.022	0.020	0.020	0.019	0.022	0.019
	5	5	0.004	0.000	-0.005	-0.008	0.003	-0.023	0.007	0.008	0.007	0.007	0.007	0.008
0.5	1	1	0.087	0.028	0.035	-0.003	0.080	0.006	0.138	0.115	0.116	0.110	0.132	0.114
	2	2	0.021	-0.009	-0.010	-0.029	0.016	-0.048	0.058	0.057	0.052	0.055	0.056	0.056
	5	5	0.010	-0.005	-0.009	-0.019	0.007	-0.057	0.024	0.029	0.022	0.028	0.023	0.025
1	1	1	0.089	-0.031	0.006	-0.071	0.074	-0.069	0.323	0.292	0.279	0.285	0.314	0.279
	2	2	0.045	-0.035	-0.012	-0.065	0.031	-0.097	0.148	0.161	0.132	0.157	0.142	0.143
	5	5	0.025	-0.029	-0.016	-0.053	0.012	-0.125	0.074	0.100	0.067	0.098	0.070	0.080
0.2	1	1.5	0.137	0.109	0.105	0.087	0.135	0.093	0.072	0.059	0.058	0.053	0.069	0.055
		1.75	0.158	0.134	0.128	0.113	0.155	0.108	0.075	0.064	0.061	0.057	0.073	0.057
		2	0.177	0.154	0.148	0.134	0.174	0.125	0.081	0.071	0.067	0.064	0.078	0.060
	2	3	0.066	0.059	0.050	0.045	0.065	0.027	0.024	0.023	0.021	0.021	0.023	0.019
		3.5	0.099	0.093	0.083	0.080	0.098	0.054	0.030	0.030	0.025	0.027	0.029	0.022
		4	0.138	0.132	0.121	0.119	0.136	0.089	0.040	0.040	0.034	0.036	0.039	0.028
	5	7.5	0.047	0.047	0.038	0.039	0.047	0.011	0.010	0.011	0.008	0.010	0.010	0.007
		8.75	0.085	0.088	0.075	0.080	0.084	0.041	0.015	0.017	0.013	0.016	0.015	0.009
		10	0.123	0.128	0.113	0.119	0.122	0.075	0.024	0.026	0.021	0.024	0.023	0.014
0.5	1	1.75	0.139	0.102	0.094	0.075	0.132	0.039	0.127	0.122	0.105	0.113	0.121	0.097
		2	0.186	0.156	0.142	0.129	0.178	0.074	0.139	0.139	0.113	0.127	0.132	0.100
		2.5	0.275	0.251	0.230	0.224	0.266	0.147	0.178	0.182	0.146	0.164	0.170	0.116
	2	3.5	0.110	0.098	0.081	0.079	0.105	0.017	0.061	0.070	0.051	0.064	0.058	0.046
		4	0.152	0.152	0.122	0.133	0.145	0.045	0.075	0.093	0.062	0.084	0.071	0.050
	5	7.5	0.050	0.050	0.031	0.036	0.047	-0.036	0.027	0.036	0.023	0.033	0.025	0.023
		8.75	0.090	0.099	0.070	0.084	0.086	-0.010	0.032	0.047	0.027	0.043	0.031	0.021
		10	0.133	0.155	0.112	0.139	0.128	0.013	0.043	0.067	0.036	0.061	0.041	0.023
1	1	2	0.234	0.178	0.160	0.141	0.216	0.015	0.303	0.344	0.248	0.320	0.288	0.218
		2.5	0.337	0.301	0.263	0.263	0.318	0.088	0.356	0.420	0.287	0.386	0.336	0.225
		3	0.435	0.431	0.360	0.392	0.414	0.151	0.437	0.572	0.353	0.525	0.412	0.239
	2	4	0.182	0.165	0.127	0.135	0.165	-0.025	0.167	0.240	0.138	0.223	0.156	0.115
		5	0.276	0.300	0.220	0.269	0.258	0.029	0.207	0.329	0.167	0.302	0.192	0.113
	5	8.75	0.108	0.114	0.067	0.089	0.093	-0.100	0.087	0.159	0.074	0.148	0.081	0.073
		10	0.143	0.171	0.101	0.145	0.127	-0.089	0.095	0.186	0.078	0.172	0.087	0.069
		12.5	0.227	0.307	0.183	0.278	0.210	-0.059	0.129	0.297	0.104	0.272	0.118	0.066
		15	0.311	0.431	0.264	0.401	0.292	-0.032	0.182	0.408	0.149	0.374	0.167	0.066

Notes: Bias and MSE calculated over all samples of size n=50 which produced NB results ($\hat{a}>0$) for all estimation methods at a given simulation (parameter) etting. For counts of samples per setting, see "runs" column in Table 6.

Table 13: Bias and MSE for estimators of *a*. Based on all samples of size n=30 using $\hat{a}$=0 for negative estimates.

			Bias						MSE					
a	μ_2	μ_1	EQL	PL	ML	OQ	CML	NED	EQL	PL	ML	OQ	CML	NED
0.2	1	1	0.196	0.146	0.133	0.101	0.197	0.111	0.159	0.110	0.126	0.097	0.159	0.126
	2	2	0.058	0.040	0.028	0.015	0.058	0.002	0.040	0.033	0.033	0.029	0.039	0.032
	5	5	0.007	0.001	-0.008	-0.012	0.007	-0.037	0.012	0.012	0.011	0.012	0.012	0.011
0.5	1	1	0.146	0.062	0.064	0.010	0.146	0.005	0.271	0.198	0.214	0.177	0.274	0.209
	2	2	0.040	-0.004	-0.007	-0.038	0.038	-0.070	0.100	0.091	0.085	0.085	0.097	0.084
	5	5	0.013	-0.015	-0.017	-0.038	0.010	-0.088	0.040	0.043	0.036	0.042	0.039	0.043
1	1	1	0.156	-0.017	0.034	-0.085	0.160	-0.097	0.634	0.465	0.519	0.434	0.668	0.513
	2	2	0.058	-0.049	-0.023	-0.098	0.051	-0.166	0.264	0.260	0.228	0.249	0.260	0.238
	5	5	0.042	-0.034	-0.017	-0.073	0.030	-0.186	0.135	0.170	0.118	0.163	0.130	0.145
0.2	1	1.75	0.185	0.150	0.137	0.114	0.185	0.099	0.122	0.097	0.097	0.084	0.121	0.094
		2	0.213	0.178	0.165	0.143	0.212	0.120	0.131	0.105	0.103	0.090	0.129	0.096
		2.5	0.275	0.234	0.229	0.200	0.274	0.178	0.164	0.129	0.130	0.110	0.162	0.118
	2	3.5	0.109	0.097	0.083	0.075	0.109	0.039	0.048	0.046	0.039	0.039	0.047	0.032
		4	0.140	0.128	0.113	0.105	0.139	0.064	0.054	0.052	0.044	0.044	0.054	0.036
	5	7.5	0.046	0.044	0.031	0.031	0.046	-0.012	0.015	0.016	0.013	0.014	0.015	0.011
		8.75	0.083	0.082	0.068	0.068	0.083	0.016	0.020	0.022	0.017	0.019	0.020	0.012
		10	0.130	0.132	0.113	0.116	0.129	0.049	0.032	0.035	0.026	0.030	0.031	0.016
0.5	1	2	0.211	0.154	0.143	0.109	0.207	0.044	0.230	0.202	0.181	0.178	0.226	0.157
		2.5	0.305	0.252	0.237	0.207	0.300	0.117	0.290	0.270	0.228	0.236	0.284	0.180
		3	0.385	0.332	0.317	0.287	0.379	0.174	0.338	0.310	0.266	0.270	0.329	0.196
	2	4	0.164	0.147	0.120	0.114	0.159	0.004	0.116	0.126	0.093	0.110	0.112	0.076
		5	0.254	0.243	0.208	0.209	0.248	0.066	0.160	0.176	0.127	0.153	0.154	0.085
	5	10	0.136	0.147	0.104	0.121	0.131	-0.037	0.062	0.088	0.051	0.077	0.060	0.037
		12.5	0.221	0.249	0.187	0.221	0.215	0.008	0.095	0.143	0.077	0.124	0.091	0.039
1	1	2	0.253	0.136	0.148	0.074	0.247	-0.052	0.509	0.488	0.408	0.443	0.509	0.351
		2.5	0.348	0.259	0.244	0.197	0.340	-0.007	0.554	0.578	0.438	0.516	0.549	0.333
		3	0.469	0.390	0.362	0.326	0.458	0.078	0.652	0.698	0.512	0.617	0.643	0.354
		3.5	0.537	0.470	0.430	0.405	0.524	0.106	0.708	0.763	0.556	0.671	0.693	0.355
	2	4	0.193	0.147	0.116	0.097	0.181	-0.108	0.249	0.334	0.200	0.302	0.239	0.180
		5	0.305	0.291	0.226	0.238	0.291	-0.052	0.335	0.457	0.265	0.405	0.320	0.186
		6	0.389	0.393	0.307	0.338	0.372	-0.011	0.393	0.564	0.309	0.497	0.374	0.195
	5	10	0.159	0.147	0.099	0.105	0.144	-0.162	0.154	0.247	0.125	0.222	0.145	0.130
		15	0.319	0.403	0.254	0.353	0.300	-0.128	0.241	0.481	0.189	0.423	0.224	0.123

Notes: Bias and MSE of $\hat{a}$ calculated over 5000 samples of size n=30 generated at each simulation (parameter) setting. Values of $\hat{a}$=0 used for non-positive estimates. See Tables 3 and 4 for the counts of non-positive estimates produced per estimation method.

Table 14: Bias and MSE for estimators of *a*. Based on all samples of size n=50 using $\hat{a}$=0 for negative estimates.

			Bias						MSE					
a	μ2	μ1	EQL	PL	ML	OQ	CML	NED	EQL	PL	ML	OQ	CML	NED
0.2	1	1	0.115	0.089	0.079	0.063	0.113	0.063	0.073	0.058	0.060	0.053	0.070	0.062
	2	2	0.026	0.017	0.009	0.002	0.026	-0.009	0.022	0.020	0.019	0.019	0.022	0.020
	5	5	0.002	-0.001	-0.007	-0.009	0.002	-0.025	0.007	0.008	0.007	0.008	0.007	0.008
0.5	1	1	0.065	0.016	0.014	-0.015	0.058	-0.026	0.137	0.119	0.115	0.111	0.132	0.116
	2	2	0.017	-0.011	-0.013	-0.031	0.013	-0.053	0.058	0.057	0.053	0.056	0.056	0.056
	5	5	0.010	-0.005	-0.009	-0.019	0.007	-0.057	0.024	0.029	0.022	0.028	0.023	0.025
1	1	1	0.079	-0.033	-0.003	-0.073	0.064	-0.088	0.324	0.302	0.280	0.292	0.316	0.280
	2	2	0.044	-0.036	-0.012	-0.065	0.030	-0.098	0.148	0.161	0.132	0.157	0.142	0.143
	5	5	0.025	-0.029	-0.016	-0.053	0.012	-0.125	0.074	0.100	0.067	0.098	0.070	0.080
0.2	1	1.5	0.117	0.097	0.087	0.075	0.115	0.065	0.067	0.057	0.055	0.051	0.065	0.055
		1.75	0.140	0.123	0.111	0.102	0.138	0.082	0.071	0.063	0.059	0.057	0.069	0.057
		2	0.164	0.146	0.136	0.126	0.162	0.106	0.077	0.070	0.064	0.062	0.075	0.060
	2	3	0.060	0.055	0.044	0.041	0.059	0.018	0.024	0.023	0.020	0.021	0.023	0.019
		3.5	0.096	0.091	0.080	0.078	0.095	0.049	0.029	0.030	0.025	0.026	0.029	0.022
		4	0.136	0.131	0.119	0.117	0.134	0.085	0.039	0.040	0.034	0.036	0.039	0.028
	5	7.5	0.047	0.047	0.038	0.039	0.046	0.011	0.010	0.011	0.008	0.010	0.010	0.007
		8.75	0.085	0.088	0.075	0.079	0.084	0.041	0.015	0.017	0.013	0.016	0.015	0.009
		10	0.123	0.128	0.113	0.119	0.122	0.075	0.024	0.026	0.021	0.024	0.023	0.014
0.5	1	1.75	0.129	0.098	0.085	0.071	0.122	0.023	0.126	0.124	0.105	0.115	0.120	0.097
		2	0.181	0.155	0.136	0.128	0.173	0.061	0.137	0.141	0.112	0.129	0.131	0.100
		2.5	0.271	0.249	0.226	0.222	0.262	0.141	0.178	0.182	0.145	0.165	0.169	0.116
	2	3.5	0.109	0.098	0.080	0.078	0.104	0.015	0.061	0.070	0.051	0.064	0.058	0.046
		4	0.151	0.152	0.122	0.133	0.145	0.043	0.075	0.093	0.062	0.085	0.071	0.050
	5	7.5	0.050	0.050	0.031	0.036	0.047	-0.036	0.027	0.036	0.023	0.033	0.025	0.023
		8.75	0.090	0.099	0.070	0.084	0.086	-0.010	0.032	0.047	0.027	0.043	0.031	0.021
		10	0.133	0.155	0.112	0.139	0.128	0.013	0.043	0.067	0.036	0.061	0.041	0.023
1	1	2	0.233	0.180	0.158	0.142	0.215	0.010	0.303	0.348	0.248	0.324	0.288	0.218
		2.5	0.337	0.303	0.262	0.265	0.317	0.084	0.356	0.424	0.287	0.390	0.336	0.225
		3	0.435	0.432	0.359	0.393	0.414	0.147	0.437	0.574	0.353	0.526	0.412	0.239
	2	4	0.182	0.166	0.127	0.136	0.165	-0.025	0.167	0.245	0.138	0.227	0.156	0.115
		5	0.276	0.301	0.220	0.269	0.258	0.029	0.207	0.330	0.167	0.302	0.192	0.113
	5	8.75	0.108	0.114	0.067	0.089	0.093	-0.100	0.087	0.159	0.074	0.148	0.081	0.073
		10	0.143	0.171	0.101	0.145	0.127	-0.089	0.095	0.186	0.078	0.172	0.087	0.069
		12.5	0.227	0.307	0.184	0.279	0.210	-0.059	0.129	0.299	0.104	0.275	0.118	0.066
		15	0.311	0.431	0.264	0.401	0.292	-0.032	0.182	0.408	0.149	0.374	0.167	0.066

Notes: Bias and MSE of $\hat{a}$ calculated over 5000 samples of size n=50 generated at each simulation (parameter) setting. Values of $\hat{a}$=0 used for non-positive estimates. See Tables 3 and 4 for the counts of non-positive estimates produced per estimation method.

Table 15: Power Comparison of DDT vs. others for samples of size n=30; Significance Results for NB tests only.

a	μ2	μ1	runs	EQL.S	PL.S	ML.S	OQ.S	CML.S	DDT	LRT	t
0.2	1	1.75	3052	0.4037	0.4089	0.4043	PL	0.4043	*0.4364*	**0.4600**	0.3607
		2	3121	0.5694	*0.5806*	0.5678	*0.5796*	0.5694	*0.5934*	**0.6267**	0.5059
		2.5	3185	*0.8182*	*0.8264*	*0.8170*	*0.8261*	*0.8188*	*0.8223*	**0.8587**	0.7608
	2	3.5	3676	**0.6151**	**0.6058**	**0.6162**	**0.6077**	**0.6148**	*0.5762*	**0.6417**	0.5182
		4	3628	**0.7889**	**0.7875**	**0.7897**	**0.7883**	**0.7891**	*0.7552*	**0.8112**	0.6987
	5	7.5	4185	**0.5350**	**0.5429**	**0.5362**	**0.5424**	**0.5345**	*0.4915*	**0.5496**	*0.4889*
		8.75	4231	**0.7920**	**0.8003**	**0.7927**	**PL.S**	**EQL.S**	*0.7261*	**0.8005**	*0.7426*
		10	4181	**0.9316**	**0.9294**	**EQL.S**	**0.9290**	**EQL.S**	*0.8536*	**0.9366**	**0.8928**
0.5	1	2	4021	**0.4949**	**0.4947**	**0.4942**	**0.4934**	**0.4949**	*0.4457*	**0.5158**	*0.4432*
		2.5	4135	**0.7320**	**0.7289**	**0.7313**	**0.7282**	**0.7325**	*0.6665*	**0.7434**	*0.6566*
		3	4322	**0.8714**	**0.8637**	**0.8707**	**0.8635**	**0.8716**	*0.7999*	**0.8764**	**0.8144**
	2	4	4581	**0.6202**	**0.6197**	**0.6213**	**0.6195**	**0.6208**	*0.5139*	**0.6296**	**0.5394**
		5	4637	**0.8361**	**0.8342**	**0.8372**	**0.8339**	**0.8361**	*0.7194*	**0.8430**	**0.7608**
	5	10	4786	**0.7528**	**0.7386**	**0.7524**	**PL.S**	**0.7537**	*0.5389*	**0.7530**	**0.6544**
		12.5	4750	**0.9200**	**0.9124**	**0.9202**	**PL.S**	**0.9206**	*0.7229*	**0.9213**	**0.8442**
1	1	2	4458	**0.3921**	**0.3948**	**0.3910**	**0.3926**	**0.3932**	*0.3333*	**0.4069**	*0.3479*
		2.5	4547	**0.5914**	**0.5797**	**0.5909**	**0.5791**	**0.5931**	*0.4799*	**0.6048**	**0.5241**
		3	4667	**0.7461**	**0.7392**	**0.7450**	**0.7379**	**0.7469**	*0.6100*	**0.7574**	**0.6687**
		3.5	4740	**0.8414**	**0.8297**	**0.8407**	**0.8285**	**0.8439**	*0.6886*	**0.8508**	**0.7622**
	2	4	4830	**0.4708**	**0.4625**	**0.4714**	**0.4623**	**0.4706**	*0.3070*	**0.4712**	**0.3894**
		5	4847	**0.6833**	**0.6666**	**0.6825**	**PL.S**	**0.6833**	*0.4539*	**0.6841**	**0.5740**
		6	4866	**0.8126**	**0.7963**	**0.8132**	**0.7961**	**0.8132**	*0.5561*	**0.8169**	**0.7100**
	5	10	4909	**0.5117**	**0.4956**	**0.5119**	**PL.S**	**0.5138**	*0.2730*	**0.5105**	**0.4099**
		15	4849	**0.8575**	**0.8363**	**0.8579**	**PL.S**	**0.8600**	*0.4774*	**0.8612**	**0.7313**

Notes: Coded power results from tests based on the GEE approach for correlated data.
Each row contains the results of the (simultaneous α=0.05) GEE test comparing power values obtained for the listed simulation (parameter) setting.
"runs" = number of samples of size n=30 where all testing methods obtain NB test results and over which the comparison is made.
DDT and all tests with power not significantly different from DDT have power values italicized and shaded in gray.
Tests with power significantly greater (better) than DDT are listed in bold and underlined.
Tests with power significantly less (worse) than DDT are listed power values in plain text.
Test results listed as another test method abbreviation acheived the same power as, and were highly correlated with, the abbreviated test. The results of that test were removed to avoid singularity of the covariance matrix. A summary of the correlation matrices produced for these power comparisons is found in Table 21.

Table 16: Power Comparison of DDT vs. others for samples of size n=50; Significance Results for NB tests only.

a	μ2	μ1	runs	EQL.S	PL.S	ML.S	OQ.S	CML.S	DDT	LRT	t
0.2	1	1.5	3703	*0.3846*	*0.3916*	*0.3867*	*0.3897*	*0.3856*	*0.3967*	*0.4080*	0.3305
		1.75	3671	*0.6186*	*0.6233*	*0.6189*	*0.6230*	*ML.S*	*0.6268*	**0.6415**	0.5554
		2	3676	*0.7973*	*0.8088*	*0.7990*	*0.8085*	*0.7976*	*0.8055*	*0.8172*	0.7378
	2	3	4359	**0.5446**	**0.5451**	**0.5458**	**PL.S**	**0.5455**	*0.5155*	**0.5614**	0.4735
		3.5	4385	**0.8098**	**0.8080**	**0.8105**	**PL.S**	**0.8100**	*0.7742*	**0.8223**	0.7446
		4	4453	**0.9367**	**0.9362**	**EQL.S**	**PL.S**	**EQL.S**	*0.9093*	**0.9421**	0.8945
	5	7.5	4891	**0.7495**	**0.7538**	**0.7502**	**PL.S**	**0.7497**	*0.6763*	**0.7504**	*0.6847*
		8.75	4901	**0.9451**	**0.9463**	**0.9455**	**0.9465**	**0.9453**	*0.8945*	**0.9455**	*0.9053*
		10	4908	**0.9955**	**0.9953**	**EQL.S**	**PL.S**	**EQL.S**	*0.9817*	**EQL.S**	*0.9835*
0.5	1	1.75	4546	**0.5350**	**0.5405**	**0.5348**	**0.5385**	**0.5359**	*0.4798*	**0.5458**	0.4648
		2	4633	**0.7058**	**0.7108**	**0.7049**	**0.7075**	**0.7056**	*0.6454*	**0.7138**	0.6311
		2.5	4752	**0.9112**	**0.9141**	**0.9114**	**0.9125**	**0.9116**	*0.8641*	**0.9160**	*0.8559*
	2	3.5	4926	**0.6630**	**0.6626**	**0.6634**	**PL.S**	**0.6630**	*0.5684*	**0.6646**	*0.5715*
		4	4947	**0.8278**	**0.8231**	**0.8274**	**PL.S**	**0.8272**	*0.7328*	**0.8280**	*0.7425*
	5	7.5	4997	**0.5369**	**0.5361**	**0.5361**	**PL.S**	**0.5367**	*0.4040*	**0.5349**	**0.4481**
		8.75	4997	**0.7757**	**0.7641**	**0.7755**	**PL.S**	**0.7765**	*0.6196*	**0.7757**	**0.6678**
		10	4997	**0.9113**	**0.9087**	**0.9109**	**PL.S**	**0.9115**	*0.7797*	**ML.S**	**0.8279**
1	1	2	4913	**0.5455**	**0.5363**	**0.5463**	**PL.S**	**0.5463**	*0.4464*	**0.5496**	**0.4918**
		2.5	4930	**0.7769**	**0.7738**	**0.7769**	**0.7736**	**EQL.S**	*0.6590*	**0.7795**	**0.7168**
		3	4956	**0.9044**	**0.8967**	**0.9048**	**PL.S**	**0.9054**	*0.8008*	**0.9056**	**0.8577**
	2	4	4985	**0.6441**	**0.6389**	**0.6443**	**PL.S**	**0.6455**	*0.5105*	**0.6455**	**0.5551**
		5	4995	**0.8623**	**0.8553**	**0.8629**	**PL.S**	**0.8631**	*0.7101*	**0.8639**	**0.7688**
	5	8.75	4998	**0.5746**	**0.5686**	**0.5738**	**PL.S**	**0.5744**	*0.3659*	**0.5748**	**0.4526**
		10	4996	**0.7188**	**0.7102**	**0.7186**	**PL.S**	**0.7188**	*0.4692*	**0.7200**	**0.5811**
		12.5	4996	**0.9217**	**0.9129**	**0.9215**	**PL.S**	**ML.S**	*0.6703*	**0.9217**	**0.8050**
		15	4997	**0.9762**	**0.9720**	**0.9760**	**PL.S**	**0.9762**	*0.7819*	**0.9762**	**0.9077**

Notes: Coded power results from tests based on the GEE approach for correlated data.

Each row contains the results of the (simultaneous α=0.05) GEE test comparing power values obtained for the listed simulation (parameter) setting.

"runs" = number of samples of size n=50 where all testing methods obtain NB test results and over which the comparison is made.

DDT and all tests with power not significantly different from DDT have power values italicized and shaded in gray.

Tests with power significantly greater (better) than DDT are listed in bold and underlined.

Tests with power significantly less (worse) than DDT are listed power values in plain text.

Test results listed as another test method abbreviation acheived the same power as, and were highly correlated with, the abbreviated test. The results of that test were removed to avoid singularity of the covariance matrix. A summary of the correlation matrices produced for these power comparisons is found in Table 22.

Table 17: Power Comparison of DDT vs. others for samples of size n=30; Significance Results for NB and Poisson tests.

a	μ2	μ1	runs	EQL.S	PL.S	ML.S	OQ.S	CML.S	DDT	LRT	t
0.2	1	1.75	4998	0.4750	0.4718	0.4736	0.4716	0.4754	*0.5040*	**0.5242**	0.4418
		2	4987	*0.6256*	*0.6256*	0.6228	*0.6248*	*0.6256*	*0.6379*	**0.6746**	0.5771
		2.5	4998	*0.8475*	*0.8495*	*0.8457*	*0.8493*	*0.8483*	*0.8441*	**0.8800**	0.8005
	2	3.5	4991	**0.6518**	**0.6390**	**0.6522**	**0.6406**	**0.6516**	*0.6135*	**0.6748**	0.5632
		4	4969	**0.8114**	**0.8056**	**0.8116**	**0.8062**	**0.8116**	*0.7716*	**0.8308**	0.7325
	5	7.5	4993	**0.5492**	**0.5534**	**0.5504**	**0.5528**	**0.5488**	*0.5159*	**0.5628**	*0.5055*
		8.75	5000	**0.8018**	**0.8076**	**0.8024**	**PL.S**	**EQL.S**	*0.7400*	**0.8094**	*0.7538*
		10	4895	**0.9359**	**0.9332**	**EQL.S**	**0.9328**	**EQL.S**	*0.8621*	**0.9408**	**0.8970**
0.5	1	2	4996	**0.5136**	**0.5048**	**0.5126**	**0.5032**	**0.5136**	*0.4602*	**0.5310**	*0.4640*
		2.5	4998	**0.7373**	**0.7271**	**0.7361**	**0.7261**	**0.7375**	*0.6717*	**0.7471**	*0.6699*
		3	4995	**0.8747**	**0.8635**	**0.8739**	**0.8629**	**0.8749**	*0.7972*	**0.8793**	**0.8188**
	2	4	5000	**0.6274**	**0.6242**	**0.6284**	**0.6240**	**0.6280**	*0.5208*	**0.6360**	**0.5500**
		5	5000	**0.8398**	**0.8362**	**0.8408**	**0.8360**	**0.8398**	*0.7212*	**0.8462**	**0.7650**
	5	10	5000	**0.7544**	**0.7396**	**0.7540**	**PL.S**	**0.7554**	*0.5476*	**0.7546**	**0.6566**
		12.5	5000	**0.9214**	**0.9140**	**0.9216**	**PL.S**	**0.9220**	*0.7308*	**0.9226**	**0.8474**
1	1	2	4987	**0.3984**	**0.3928**	**0.3970**	**0.3908**	**0.3994**	*0.3395*	**0.4115**	**0.3563**
		2.5	4986	**0.5913**	**0.5728**	**0.5901**	**0.5722**	**0.5929**	*0.4761*	**0.6029**	**0.5247**
		3	4986	**0.7475**	**0.7365**	**0.7465**	**0.7353**	**0.7487**	*0.6105*	**0.7589**	**0.6729**
		3.5	4976	**0.8413**	**0.8276**	**0.8404**	**0.8264**	**0.8437**	*0.6825*	**0.8499**	**0.7611**
	2	4	4999	**0.4727**	**0.4629**	**0.4731**	**0.4627**	**0.4723**	*0.3119*	**0.4729**	**0.3931**
		5	4999	**0.6843**	**0.6661**	**0.6837**	**PL.S**	**0.6843**	*0.4563*	**0.6851**	**0.5763**
		6	4999	**0.8130**	**0.7956**	**0.8136**	**0.7954**	**0.8136**	*0.5563*	**0.8174**	**0.7109**
	5	10	5000	**0.5136**	**0.4980**	**0.5138**	**PL.S**	**0.5156**	*0.2790*	**0.5122**	**0.4130**
		15	5000	**0.8596**	**0.8378**	**0.8600**	**PL.S**	**0.8620**	*0.4844*	**0.8634**	**0.7340**

Notes: Coded power results from tests based on the GEE approach for correlated data.
Poisson test results are used for individual tests failing to obtain a NB result due to a non-positive estimate for the variance parameter.
Values obtained for the same simulation (parameter) settings are listed in rows.
"runs" = number of samples of size n=30 where all testing methods obtain NB or Poisson test results and over which the comparison is made.
DDT and all tests with power not significantly different from DDT are listed in italics and shaded in gray.
Tests with power significantly greater (better) than DDT are listed in bold and underlined.
Tests with power significantly less (worse) than DDT are listed in plain text.
Test results listed as another test method abbreviation acheived the same power as, and were highly correlated with, the abbreviated test. The results of that test were removed to avoid singularity of the covariance matrix.

Table 18: Power Comparison of DDT vs. others for samples of size n=50; Significance Results for NB and Poisson tests

a	μ2	μ1	runs	EQL.S	PL.S	ML.S	OQ.S	CML.S	DDT	LRT	t
0.2	1	1.5	5000	0.4192	0.4228	0.4208	0.4214	0.4202	*0.4378*	*0.4398*	0.3756
		1.75	5000	*0.6506*	*0.6524*	*0.6504*	*0.6520*	*0.6510*	*0.6578*	**0.6702**	0.5964
		2	4998	*0.8243*	*0.8305*	*0.8255*	*0.8301*	*0.8245*	*0.8251*	**0.8413**	0.7697
	2	3	5000	**0.5574**	**0.5570**	**0.5584**	**PL.S**	**0.5582**	*0.5304*	**0.5722**	0.4936
		3.5	5000	**0.8208**	**0.8164**	**0.8210**	**PL.S**	**0.8210**	*0.7842*	**0.8322**	0.7580
		4	4944	**0.9391**	**0.9387**	**EQL.S**	**PL.S**	**EQL.S**	*0.9122*	**0.9444**	0.8989
	5	7.5	5000	**0.7512**	**0.7552**	**0.7518**	**PL.S**	**0.7514**	*0.6790*	**0.7520**	*0.6884*
		8.75	5000	**0.9454**	**0.9464**	**0.9458**	**0.9466**	**0.9456**	*0.8950*	**0.9458**	*0.9052*
		10	5000	**0.9956**	**0.9954**	**EQL.S**	**PL.S**	**EQL.S**	*0.9818*	**EQL.S**	*0.9832*
0.5	1	1.75	5000	**0.5412**	**0.5446**	**0.5406**	**0.5428**	**0.5416**	*0.4908*	**0.5518**	*0.4786*
		2	5000	**0.7092**	**0.7126**	**0.7078**	**0.7092**	**0.7088**	*0.6466*	**0.7166**	*0.6366*
		2.5	4999	**0.9114**	**0.9136**	**0.9116**	**0.9120**	**0.9118**	*0.8636*	**0.9162**	*0.8590*
	2	3.5	5000	**0.6638**	**0.6624**	**0.6642**	**PL.S**	**0.6638**	*0.5692*	**0.6652**	*0.5722*
		4	5000	**0.8282**	**0.8234**	**0.8278**	**PL.S**	**0.8276**	*0.7332*	**0.8284**	*0.7432*
	5	7.5	5000	**0.5366**	**0.5358**	**0.5358**	**PL.S**	**0.5364**	*0.4040*	**0.5346**	**0.4478**
		8.75	5000	**0.7756**	**0.7640**	**0.7754**	**PL.S**	**0.7764**	*0.6194*	**0.7756**	**0.6674**
		10	5000	**0.9114**	**0.9088**	**0.9110**	**PL.S**	**0.9116**	*0.7798*	**ML.S**	**0.8280**
1	1	2	5000	**0.5444**	**0.5344**	**0.5452**	**PL.S**	**0.5454**	*0.4460*	**0.5486**	**0.4910**
		2.5	5000	**0.7774**	**0.7728**	**0.7774**	**0.7726**	**EQL.S**	*0.6578*	**0.7796**	**0.7168**
		3	5000	**0.9044**	**0.8964**	**0.9048**	**PL.S**	**0.9054**	*0.7998*	**0.9056**	**0.8580**
	2	4	5000	**0.6436**	**0.6378**	**0.6438**	**PL.S**	**0.6450**	*0.5104*	**0.6450**	**0.5552**
		5	5000	**0.8622**	**0.8552**	**0.8628**	**PL.S**	**0.8630**	*0.7100*	**0.8638**	**0.7686**
	5	8.75	5000	**0.5748**	**0.5688**	**0.5740**	**PL.S**	**0.5746**	*0.3662*	**0.5750**	**0.4528**
		10	5000	**0.7186**	**0.7102**	**0.7184**	**PL.S**	**0.7186**	*0.4690*	**0.7198**	**0.5810**
		12.5	5000	**0.9218**	**0.9130**	**0.9216**	**PL.S**	**ML.S**	*0.6706*	**0.9218**	**0.8052**
		15	5000	**0.9762**	**0.9720**	**0.9760**	**PL.S**	**0.9762**	*0.7818*	**0.9762**	**0.9074**

Notes: Coded power results from tests based on the GEE approach for correlated data.
Poisson test results are used for individual tests failing to obtain a NB result due to a non-positive estimate for the variance parameter.
Values obtained for the same simulation (parameter) settings are listed in rows.
"runs" = number of samples of size n=30 where all testing methods obtain NB or Poisson test results and over which the comparison is made.
DDT and all tests with power not significantly different from DDT are listed in italics and shaded in gray.
Tests with power significantly greater (better) than DDT are listed in bold and underlined.
Tests with power significantly less (worse) than DDT are listed in plain text.
Test results listed as another test method abbreviation acheived the same power as, and were highly correlated with, the abbreviated test. The results of that test were removed to avoid singularity of the covariance matrix.

Table 19: Power Comparison of LRT vs. others excluding DDT for samples of size n=30; Significance Results for NB tests only.

a	$\mu 2$	$\mu 1$	var2	var1	runs	EQL.S	PL.S	ML.S	OQ.S	CML.S	LRT	t
0.2	1	1.75	1.2	2.36	4024	0.4120	0.4100	0.4118	0.4098	0.4125	*0.4657*	0.3805
		2		2.8	4033	0.5713	0.5720	0.5686	0.5710	0.5713	*0.6268*	0.5185
		2.5		3.75	4015	0.8207	0.8242	0.8192	0.8239	0.8217	*0.8598*	0.7691
	2	3.5	2.8	5.95	4511	0.6294	0.6160	0.6300	0.6178	0.6291	*0.6537*	0.5356
		4		7.2	4522	0.7970	0.7910	0.7974	0.7917	0.7972	*0.8178*	0.7134
	5	7.5	10	18.8	4946	0.5469	0.5512	0.5481	0.5505	0.5465	*0.5607*	0.5028
		8.75		24.1	4963	0.8005	*0.8064*	0.8011	*PL.S*	EQL.S	*0.8082*	0.7524
		10		30	4867	0.9355	0.9328	EQL.S	0.9324	EQL.S	*0.9404*	0.8964
0.5	1	2	1.5	4	4731	0.5012	0.4925	0.5001	0.4910	0.5012	*0.5198*	0.4519
		2.5		5.63	4758	0.7299	0.7192	0.7287	0.7184	0.7301	*0.7402*	0.6608
		3		7.5	4839	0.8725	0.8609	0.8717	0.8603	0.8727	*0.8772*	0.8155
	2	4	4	12	4963	0.6252	0.6220	0.6262	0.6218	0.6258	*0.6339*	0.5479
		5		17.5	4977	0.8397	0.8360	0.8407	0.8358	0.8397	*0.8461*	0.7647
	5	10	17.5	60	5000	*0.7544*	0.7396	*0.7540*	PL.S	*0.7554*	*0.7546*	0.6566
		12.5		90.6	5000	*0.9214*	0.9140	*0.9216*	PL.S	*0.9220*	*0.9226*	0.8474
1	1	2	2	6	4942	0.3954	0.3901	0.3942	0.3881	0.3964	*0.4087*	0.3541
		2.5		8.75	4962	0.5907	0.5728	0.5901	0.5721	0.5923	*0.6030*	0.5250
		3		12	4987	0.7481	0.7371	0.7471	0.7359	0.7493	*0.7596*	0.6736
		3.5		15.8	4988	0.8414	0.8278	0.8406	0.8266	0.8438	*0.8500*	0.7612
	2	4	6	20	4999	*0.4725*	*0.4629*	*0.4729*	*0.4627*	*0.4721*	*0.4727*	0.3931
		5		30	4999	*0.6841*	0.6659	*0.6835*	PL.S	*0.6841*	*0.6849*	0.5761
		6		42	5000	0.8130	0.7956	0.8136	0.7954	0.8136	*0.8174*	0.7110
	5	10	30	110	5000	*0.5136*	0.4980	*0.5138*	PL.S	**0.5156**	*0.5122*	0.4130
		15		240	5000	0.8596	0.8378	0.8600	PL.S	*0.8620*	*0.8634*	0.7340

Notes: Coded power results from tests based on the GEE approach for correlated data.
Values obtained for the same simulation (parameter) settings are listed in rows.
"runs" = number of samples of size n=30 where all testing methods obtain NB test results and over which the comparison is made.
LRT and all tests with power not significantly different from LRT have power values italicized and shaded in gray.
Tests with power significantly greater (better) than LRT are listed in bold and underlined.
Tests with power significantly less (worse) than LRT are listed power values in plain text.
Test results listed as another test method abbreviation acheived the same power as, and were highly correlated with, the abbreviated test. The results of that test were removed to avoid singularity of the covariance matrix.

Table 20: Power Comparison of LRT vs. others excluding DDT for samples of size n=50; Significance Results for NB tests only.

a	$\mu 2$	$\mu 1$	var2	var1	runs	EQL.S	PL.S	ML.S	OQ.S	CML.S	LRT	t
0.2	1	1.5	1.2	1.95	4423	0.3832	0.3873	0.3850	0.3857	0.3844	*0.4049*	0.3405
		1.75		2.36	4372	0.6212	0.6233	0.6212	0.6231	0.6217	*0.6432*	0.5659
		2		2.8	4307	0.8078	0.8150	0.8091	0.8145	0.8080	*0.8263*	0.7504
	2	3	2.8	4.8	4835	0.5481	0.5477	0.5491	PL.S	0.5489	*0.5634*	0.4836
		3.5		5.95	4808	0.8153	0.8107	0.8155	PL.S	0.8155	*0.8272*	0.7517
		4		7.2	4794	0.9378	0.9374	EQL.S	PL.S	EQL.S	*0.9433*	0.8972
	5	7.5	10	18.8	4995	*0.7512*	*0.7552*	*0.7518*	*PL.S*	*0.7514*	*0.7520*	0.6883
		8.75		24.1	4999	*0.9454*	*0.9464*	*0.9458*	*0.9466*	*0.9456*	*0.9458*	0.9052
		10		30	4998	*LRT*	*0.9954*	*LRT*	*PL.S*	*LRT*	*0.9956*	0.9832
0.5	1	1.75	1.5	3.28	4897	0.5371	0.5405	0.5365	0.5387	0.5375	*0.5477*	0.4740
		2		4	4929	0.7056	*0.7095*	0.7046	0.7060	0.7054	*0.7135*	0.6324
		2.5		5.63	4966	0.9108	*0.9130*	0.9110	*0.9114*	0.9112	*0.9156*	0.8582
	2	3.5	4	9.63	4996	*0.6637*	*0.6623*	*0.6641*	*PL.S*	*0.6637*	*0.6651*	0.5721
		4		12	4995	*0.8280*	*0.8232*	*0.8276*	*PL.S*	*0.8274*	*0.8282*	0.7431
	5	7.5	17.5	35.6	5000	*0.5366*	*0.5358*	*0.5358*	*PL.S*	*0.5364*	*0.5346*	0.4478
		8.75		47	5000	*0.7756*	0.7640	*0.7754*	PL.S	*0.7764*	*0.7756*	0.6674
		10		60	5000	*0.9114*	*0.9088*	*EQL.S*	*PL.S*	*0.9116*	*0.9110*	0.8380
1	1	2	2	6	4998	0.5442	0.5342	0.5450	PL.S	*0.5452*	*0.5484*	0.4908
		2.5		8.75	4998	*0.7773*	*0.7727*	*0.7773*	*0.7725*	*EQL.S*	*0.7795*	0.7167
		3		12	4999	*0.9044*	0.8964	*0.9048*	PL.S	*0.9054*	*0.9056*	0.8580
	2	4	6	20	5000	*0.6436*	*0.6378*	*0.6438*	*PL.S*	*0.6450*	*0.6450*	0.5552
		5		30	5000	*0.8622*	0.8552	*0.8628*	PL.S	*0.8630*	*0.8638*	0.7686
	5	8.75	30	85.3	5000	*0.5748*	*0.5688*	*0.5740*	*PL.S*	*0.5746*	*0.5750*	0.4528
		10		110	5000	*0.7186*	*0.7102*	*0.7184*	*PL.S*	*0.7186*	*0.7198*	0.5810
		12.5		169	5000	*0.9218*	0.9130	*0.9216*	PL.S	*ML.S*	*0.9218*	0.8052
		15		240	5000	*0.9762*	*0.9720*	*0.9760*	*PL.S*	*0.9762*	*0.9762*	0.9074

Notes: Coded power results from tests based on the GEE approach for correlated data.
Values obtained for the same simulation (parameter) settings are listed in rows.
"runs" = number of samples of size n=50 where all testing methods obtain NB test results and over which the comparison is made.
LRT and all tests with power not significantly different from LRT have power values italicized and shaded in gray.
Tests with power significantly greater (better) than LRT are listed in bold and underlined.
Tests with power significantly less (worse) than LRT are listed power values in plain text.
Test results listed as another test method abbreviation acheived the same power as, and were highly correlated with, the abbreviated test. The results of that test were removed to avoid singularity of the covariance matrix.

Table 21. Summary statistics across correlation matrices (R) produced for power analyses in Table 15 on samples of size n=30.

average correlations:

	EQL.S	PL.S	ML.S	OQ.S	CML.S	DDT	LRT	t
EQL.S	1.0000	0.9007	0.9961	0.9009	0.9975	0.6316	0.9561	0.6881
PL.S	0.9007	1.0000	0.9032	0.9984	0.9008	0.6548	0.8957	0.7071
ML.S	0.9961	0.9032	1.0000	0.9035	0.9952	0.6324	0.9573	0.6887
OQ.S	0.9009	0.9984	0.9035	1.0000	0.9010	0.6553	0.8959	0.7074
CML.S	0.9975	0.9008	0.9952	0.9010	1.0000	0.6317	0.9573	0.6880
DDT	0.6316	0.6548	0.6324	0.6553	0.6317	1.0000	0.6244	0.7195
LRT	0.9561	0.8957	0.9573	0.8959	0.9573	0.6244	1.0000	0.6771
t	0.6881	0.7071	0.6887	0.7074	0.6880	0.7195	0.6771	1.0000

minimum correlations:

	EQL.S	PL.S	ML.S	OQ.S	CML.S	DDT	LRT	t
EQL.S	1.0000	0.8124	0.9928	0.8131	0.9905	0.3671	0.8605	0.5474
PL.S	0.8124	1.0000	0.8155	0.9953	0.8129	0.3840	0.8188	0.5795
ML.S	0.9928	0.8155	1.0000	0.8163	0.9882	0.3680	0.8569	0.5484
OQ.S	0.8131	0.9953	0.8163	1.0000	0.8137	0.3840	0.8195	0.5795
CML.S	0.9905	0.8129	0.9882	0.8137	1.0000	0.3664	0.8624	0.5482
DDT	0.3671	0.3840	0.3680	0.3840	0.3664	1.0000	0.3675	0.5522
LRT	0.8605	0.8188	0.8569	0.8195	0.8624	0.3675	1.0000	0.5512
t	0.5474	0.5795	0.5484	0.5795	0.5482	0.5522	0.5512	1.0000

maximum correlations:

	EQL.S	PL.S	ML.S	OQ.S	CML.S	DDT	LRT	t
EQL.S	1.0000	0.9539	1.0000	0.9539	1.0000	0.8420	0.9916	0.7670
PL.S	0.9539	1.0000	0.9593	1.0000	0.9512	0.8672	0.9430	0.7838
ML.S	1.0000	0.9593	1.0000	0.9593	1.0000	0.8460	0.9938	0.7686
OQ.S	0.9539	1.0000	0.9593	1.0000	0.9512	0.8672	0.9430	0.7838
CML.S	1.0000	0.9512	1.0000	0.9512	1.0000	0.8406	0.9957	0.7672
DDT	0.8420	0.8672	0.8460	0.8672	0.8406	1.0000	0.8460	0.8248
LRT	0.9916	0.9430	0.9938	0.9430	0.9957	0.8460	1.0000	0.7422
t	0.7670	0.7838	0.7686	0.7838	0.7672	0.8248	0.7422	1.0000

Notes: A separate correlation matrix (R) is produced for each power analysis (row) in Table 15. The summary statistics (averages, minimums and maximums) were taken over the 24 R matrices.

Table 22. Summary statistics across correlation matrices (R) produced for power analyses in Table 16 on samples of size n=50.

average correlations:

	EQL.S	PL.S	ML.S	OQ.S	CML.S	DDT	LRT	t
EQL.S	1.0000	0.9161	0.9971	0.9163	0.9980	0.6233	0.9807	0.6380
PL.S	0.9161	1.0000	0.9179	0.9988	0.9167	0.6371	0.9180	0.6512
ML.S	0.9971	0.9179	1.0000	0.9180	0.9983	0.6230	0.9830	0.6380
OQ.S	0.9163	0.9988	0.9180	1.0000	0.9168	0.6374	0.9176	0.6515
CML.S	0.9980	0.9167	0.9983	0.9168	1.0000	0.6231	0.9823	0.6379
DDT	0.6233	0.6371	0.6230	0.6374	0.6231	1.0000	0.6198	0.7098
LRT	0.9807	0.9180	0.9830	0.9176	0.9823	0.6198	1.0000	0.6341
t	0.6380	0.6512	0.6380	0.6515	0.6379	0.7098	0.6341	1.0000

minimum correlations:

	EQL.S	PL.S	ML.S	OQ.S	CML.S	DDT	LRT	t
EQL.S	1.0000	0.7529	0.9871	0.7529	0.9914	0.2353	0.9381	0.3504
PL.S	0.7529	1.0000	0.7655	0.9894	0.7609	0.2744	0.7768	0.3658
ML.S	0.9871	0.7655	1.0000	0.7655	0.9941	0.2368	0.9429	0.3504
OQ.S	0.7529	0.9894	0.7655	1.0000	0.7609	0.2744	0.7768	0.3658
CML.S	0.9914	0.7609	0.9941	0.7609	1.0000	0.2353	0.9389	0.3504
DDT	0.2353	0.2744	0.2368	0.2744	0.2353	1.0000	0.2385	0.4352
LRT	0.9381	0.7768	0.9429	0.7768	0.9389	0.2385	1.0000	0.3504
t	0.3504	0.3658	0.3504	0.3658	0.3504	0.4352	0.3504	1.0000

maximum correlations:

	EQL.S	PL.S	ML.S	OQ.S	CML.S	DDT	LRT	t
EQL.S	1.0000	0.9740	1.0000	0.9734	1.0000	0.8636	1.0000	0.7639
PL.S	0.9740	1.0000	0.9746	1.0000	0.9746	0.8694	0.9648	0.7604
ML.S	1.0000	0.9746	1.0000	0.9740	1.0000	0.8614	1.0000	0.7623
OQ.S	0.9734	1.0000	0.9740	1.0000	0.9740	0.8699	0.9648	0.7609
CML.S	1.0000	0.9746	1.0000	0.9740	1.0000	0.8625	1.0000	0.7642
DDT	0.8636	0.8694	0.8614	0.8699	0.8625	1.0000	0.8599	0.8284
LRT	1.0000	0.9648	1.0000	0.9648	1.0000	0.8599	1.0000	0.7475
t	0.7639	0.7604	0.7623	0.7609	0.7642	0.8284	0.7475	1.0000

Notes: A separate correlation matrix (R) is produced for each power analysis (row) in Table 16. The summary statistics (averages, minimums and maximums) were taken over the 26 R matrices.

Table 23. Rejection Patterns for Power Comparisons in Table 15 with a=0.2 and samples of size n=30.

Testing Methods					a=.2, n=30			Settings for Means						
						μ2=1			μ2=2			μ2=5		
E Q L S	P L S	D D T	L R T	t	μ1= 1.75	μ1= 2	μ1= 2.5	μ1= 3.5	μ1= 4		μ1= 7.5	μ1= 8.75	μ1= 10	
rejection pattern					counts of occurence in this pattern								total	percent
rejected by no tests														
0	0	0	0	0	1542	1052	400	1176	592		1566	658	194	7180	24.54
rejected by 1 of 5 tests														
1	0	0	0	0	0	0	0	0	0		0	0	0	0	0
0	1	0	0	0	0	0	0	3	3		19	12	5	42	0.144
0	0	1	0	0	49	54	23	73	36		139	74	32	480	1.641
0	0	0	1	0	68	65	47	32	33		17	11	7	280	0.957
0	0	0	0	1	25	28	14	36	30		57	36	9	235	0.803
rejected by 2 of 5 tests														
1	1	0	0	0	0	0	0	0	0		0	0	0	0	0
1	0	1	0	0	0	0	0	0	0		0	0	0	0	0
1	0	0	1	0	20	12	15	49	24		31	14	18	183	0.625
1	0	0	0	1	0	0	0	0	0		0	0	0	0	0
0	1	1	0	0	0	0	0	1	0		9	7	2	19	0.065
0	1	0	1	0	7	17	10	6	6		9	9	3	67	0.229
0	1	0	0	1	0	0	0	0	0		7	8	3	18	0.062
0	0	1	1	0	34	32	18	15	9		0	1	1	110	0.376
0	0	1	0	1	32	31	13	27	24		80	38	17	262	0.895
0	0	0	1	1	7	4	6	9	3		4	2	3	38	0.130
rejected by 3 of 5 tests														
1	1	1	0	0	0	0	0	0	0		0	0	0	0	0
1	1	0	1	0	39	67	53	135	128		160	150	91	823	2.813
1	1	0	0	1	0	0	0	0	0		0	0	0	0	0
1	0	1	1	0	5	5	0	2	3		2	2	2	21	0.072
1	0	1	0	1	0	0	0	0	0		0	0	0	0	0
1	0	0	1	1	0	2	1	14	5		10	5	4	41	0.140
0	1	1	1	0	15	17	17	12	12		4	3	1	81	0.277
0	1	1	0	1	0	0	0	1	0		8	11	3	23	0.079
0	1	0	1	1	2	1	0	5	3		8	1	1	21	0.072
0	0	1	1	1	21	22	15	12	7		2	0	2	81	0.277
rejected by 4 of 5 tests														
1	1	1	1	0	172	221	179	267	247		183	148	92	1509	5.157
1	1	1	0	1	0	0	0	0	0		0	0	0	0	0
1	1	0	1	1	10	21	20	93	61		240	253	274	972	3.322
1	0	1	1	1	1	2	1	4	5		5	4	6	28	0.096
0	1	1	1	1	18	21	16	7	8		17	9	3	99	0.338
rejected by all tests														
1	1	1	1	1	985	1447	2337	1697	2389		1608	2775	3408	16646	56.89
			total		3052	3121	3185	3676	3628		4185	4231	4181	29259	100.0

Notes: Rejection patterns (left section) indicate which combinations of the testing methods (listed in the header) rejected (=1) or accepted (=0) H_0. The remainder of the table is a tabulation of the occurrences per parameter settings (columns, as listed in header).

Table 24. Rejection Patterns for Power Comparisons in Table 15 with a=0.5 and samples of size n=30.

Testing Methods					*a*=.5, n=30			Settings for Means					
						μ2=1			μ2=2		μ2=5		
EQLS	PLS	DDT	LRT	t	μ1= 2	μ1= 2.5	μ1= 3	μ1= 4	μ1= 5	μ1= 10	μ1= 12.5		
rejection pattern					counts of occurence in this pattern							total	percent
rejected by no tests													
0	0	0	0	0	1766	927	445	1434	565	910	271	6318	20.23
rejected by 1 of 5 tests													
1	0	0	0	0	1	0	0	0	0	5	0	6	0.019
0	1	0	0	0	12	10	6	38	26	38	14	144	0.461
0	0	1	0	0	46	41	19	77	29	58	28	298	0.954
0	0	0	1	0	25	16	5	13	6	2	1	68	0.218
0	0	0	0	1	58	25	22	58	39	69	23	294	0.941
rejected by 2 of 5 tests													
1	1	0	0	0	0	0	0	0	0	1	0	1	0.003
1	0	1	0	0	0	0	0	0	0	0	0	0	0
1	0	0	1	0	53	50	44	74	53	100	42	416	1.332
1	0	0	0	1	0	0	0	1	0	0	0	1	0.003
0	1	1	0	0	0	0	0	5	2	10	1	18	0.058
0	1	0	1	0	24	11	7	16	16	4	0	78	0.25
0	1	0	0	1	3	4	8	10	12	7	6	50	0.16
0	0	1	1	0	0	1	0	0	1	0	0	2	0.006
0	0	1	0	1	58	42	29	59	47	67	26	328	1.05
0	0	0	1	1	6	1	1	3	1	1	1	14	0.045
rejected by 3 of 5 tests													
1	1	1	0	0	0	0	0	0	0	0	0	0	0
1	1	0	1	0	179	210	169	316	293	411	305	1883	6.029
1	1	0	0	1	0	0	0	0	0	0	0	0	0
1	0	1	1	0	1	1	3	2	5	9	5	26	0.083
1	0	1	0	1	0	0	0	0	0	1	0	1	0.003
1	0	0	1	1	13	10	8	13	11	17	8	80	0.256
0	1	1	1	0	3	1	0	1	1	1	0	7	0.022
0	1	1	0	1	3	12	5	15	8	16	5	64	0.205
0	1	0	1	1	4	6	2	3	2	0	1	18	0.058
0	0	1	1	1	2	2	1	1	1	0	1	8	0.026
rejected by 4 of 5 tests													
1	1	1	1	0	129	152	104	134	112	105	73	809	2.59
1	1	1	0	1	0	0	0	0	0	0	0	0	0
1	1	0	1	1	85	109	148	248	277	642	644	2153	6.894
1	0	1	1	1	3	5	12	7	11	12	10	60	0.192
0	1	1	1	1	21	9	6	7	4	0	2	49	0.157
rejected by all tests													
1	1	1	1	1	1526	2490	3278	2046	3115	2300	3283	18038	57.75
				total	4021	4135	4322	4581	4637	4786	4750	31232	100

Notes: Rejection patterns (left section) indicate which combinations of the testing methods (listed in the header) rejected (=1) or accepted (=0) H_0. The remainder of the table is a tabulation of the occurrences per parameter settings (columns, as listed in header).

Table 25. Rejection Patterns for Power Comparisons in Table 15 with a=1 and samples of size n=30.

Testing Methods					a=1, n=30				Settings for Means						
					$\mu 2$=1				$\mu 2$=2			$\mu 2$=5			
EQLS	PLS	DDT	LRT	t	$\mu 1$= 2	$\mu 1$= 2.5	$\mu 1$= 3	$\mu 1$= 3.5	$\mu 1$= 4	$\mu 1$= 5	$\mu 1$= 6	$\mu 1$= 10	$\mu 1$= 15		
rejection pattern					counts of occurence in this pattern									totals	percent
rejected by no tests															
0	0	0	0	0	2350	1572	901	568	2199	1245	676	1998	505	12014	28.13
rejected by 1 of 5 tests															
1	0	0	0	0	2	3	4	4	17	9	2	7	4	52	0.122
0	1	0	0	0	45	32	48	21	76	69	48	74	41	454	1.063
0	0	1	0	0	96	59	62	28	80	55	35	88	30	533	1.248
0	0	0	1	0	23	25	16	10	7	5	6	4	6	102	0.239
0	0	0	0	1	39	30	19	27	52	40	29	66	26	328	0.768
rejected by 2 of 5 tests															
1	1	0	0	0	0	1	1	0	4	1	1	3	0	11	0.026
1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0.000
1	0	0	1	0	74	106	97	97	127	136	126	153	126	1042	2.440
1	0	0	0	1	0	1	0	0	0	2	0	1	0	4	0.009
0	1	1	0	0	2	1	2	0	3	1	2	15	0	26	0.061
0	1	0	1	0	21	21	25	16	8	5	8	3	8	115	0.269
0	1	0	0	1	9	12	12	7	25	14	21	24	16	140	0.328
0	0	1	1	0	1	1	0	0	0	1	0	0	2	5	0.012
0	0	1	0	1	78	73	70	34	77	77	60	99	32	600	1.405
0	0	0	1	1	4	1	2	2	1	2	0	1	2	15	0.035
rejected by 3 of 5 tests															
1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0.000
1	1	0	1	0	198	270	302	316	381	489	466	518	546	3486	8.161
1	1	0	0	1	3	0	0	0	2	2	0	4	0	11	0.026
1	0	1	1	0	2	3	3	2	2	3	1	7	6	29	0.068
1	0	1	0	1	1	0	0	0	0	0	0	1	0	2	0.005
1	0	0	1	1	19	23	27	20	19	23	33	36	31	231	0.541
0	1	1	1	0	0	0	0	0	1	0	0	0	0	1	0.002
0	1	1	0	1	18	13	13	18	19	15	17	23	19	155	0.363
0	1	0	1	1	12	3	9	7	2	2	2	1	2	40	0.094
0	0	1	1	1	2	0	0	0	1	2	1	0	0	6	0.014
rejected by 4 of 5 tests															
1	1	1	1	0	93	70	85	65	44	46	40	27	29	499	1.168
1	1	1	0	1	1	0	0	0	0	1	0	0	0	2	0.005
1	1	0	1	1	173	265	357	381	427	603	742	676	1221	4845	11.34
1	0	1	1	1	7	14	16	15	14	16	22	15	24	143	0.335
0	1	1	1	1	10	15	6	14	5	2	7	1	2	62	0.145
rejected by all tests															
1	1	1	1	1	1175	1933	2590	3088	1237	1981	2521	1064	2171	17760	41.58
			total		4458	4547	4667	4740	4830	4847	4866	4909	4849	42713	100.0

Notes: Rejection patterns (left section) indicate which combinations of the testing methods (listed in the header) rejected (=1) or accepted (=0) H_0. The remainder of the table is a tabulation of the occurrences per parameter settings (columns, as listed in header).

Table 26. Rejection Patterns for Power Comparisons in Table 16 with a=0.2 and samples of size n=50.

Testing Methods					a=.2, n=50			Settings for Means							
						μ2=1			μ2=2			μ2=5			
E Q L S	P L S	D D T	L R T	t	μ1= 1.5	μ1= 1.75	μ1= 2	μ1= 3	μ1= 3.5	μ1= 4	μ1= 7.5	μ1= 8.75	μ1= 10		
rejection pattern					counts of occurence in this pattern									total	percent
rejected by no tests															
0	0	0	0	0	2046	1181	555	1711	665	204	969	192	14	7523	22.1
rejected by 1 of 5 tests															
1	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0.003
0	1	0	0	0	4	1	2	1	2	0	30	9	0	49	0.144
0	0	1	0	0	53	56	44	58	33	14	69	23	1	350	1.028
0	0	0	1	0	28	29	19	30	15	9	1	0	0	131	0.385
0	0	0	0	1	37	34	27	82	42	27	72	19	4	340	0.999
rejected by 2 of 5 tests															
1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	1	0	14	11	6	23	23	8	21	9	1	115	0.338
1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
0	1	1	0	0	1	1	0	2	0	0	6	3	0	13	0.038
0	1	0	1	0	17	8	7	13	11	5	1	2	0	64	0.188
0	1	0	0	1	1	0	0	0	1	0	9	4	0	15	0.044
0	0	1	1	0	12	9	5	4	7	2	1	1	0	41	0.12
0	0	1	0	1	49	43	36	58	35	12	59	14	2	306	0.899
0	0	0	1	1	1	5	5	4	6	2	0	0	0	23	0.068
rejected by 3 of 5 tests															
1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	0	1	0	65	81	78	163	140	76	214	99	23	916	2.691
1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0
1	0	1	1	0	0	1	0	0	2	1	3	0	0	7	0.021
1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0
1	0	0	1	1	2	2	1	2	5	3	5	2	1	22	0.065
0	1	1	1	0	12	12	15	3	0	1	0	0	0	43	0.126
0	1	1	0	1	1	0	8	0	1	1	7	2	1	20	0.059
0	1	0	1	1	2	1	1	3	5	1	0	0	0	13	0.038
0	0	1	1	1	8	12	5	7	6	2	0	0	0	40	0.118
rejected by 4 of 5 tests															
1	1	1	1	0	227	242	233	287	222	150	227	125	42	1713	5.032
1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0
1	1	0	1	1	17	17	14	80	75	69	261	180	47	713	2.095
1	0	1	1	1	3	0	0	4	3	0	4	2	0	16	0.047
0	1	1	1	1	7	8	16	9	5	2	1	0	0	48	0.141
rejected by all tests															
1	1	1	1	1	1096	1917	2599	1815	3081	3864	2931	4214	4772	21517	63.21
				total	3703	3671	3676	4359	4385	4453	4891	4901	4908	34039	100.0

Notes: Rejection patterns (left section) indicate which combinations of the testing methods (listed in the header) rejected (=1) or accepted (=0) H_0. The remainder of the table is a tabulation of the occurrences per parameter settings (columns, as listed in header).

Table 27. Rejection Patterns for Power Comparisons in Table 16 with a=0.5 and samples of size n=50.

Testing Methods					a=.5, n=50			Settings for Means						
						μ2=1			μ2=2			μ2=5		
E Q L S	P L S	D D T	L R T	t	μ1= 1.75	μ1= 2	μ1= 2.5	μ1= 3.5	μ1= 4	μ1= 7.5	μ1= 8.75	μ1= 10		
rejection pattern					counts of occurence in this pattern								total	percent
rejected by no tests														
0	0	0	0	0	1850	1161	313	1402	685	1941	891	326	8569	22.09
rejected by 1 of 5 tests														
1	0	0	0	0	0	0	0	1	0	8	1	1	11	0.028
0	1	0	0	0	19	23	14	39	22	48	25	21	211	0.544
0	0	1	0	0	53	46	21	70	30	89	61	23	393	1.013
0	0	0	1	0	15	8	3	1	2	1	1	0	31	0.08
0	0	0	0	1	66	42	23	61	46	105	66	33	442	1.139
rejected by 2 of 5 tests														
1	1	0	0	0	0	0	0	2	3	1	2	0	8	0.021
1	0	1	0	0	0	0	0	0	0	0	0	2	2	0.005
1	0	0	1	0	34	28	24	50	48	62	76	32	354	0.912
1	0	0	0	1	0	0	0	0	0	1	1	0	2	0.005
0	1	1	0	0	2	1	1	2	3	12	5	1	27	0.07
0	1	0	1	0	16	18	10	3	1	1	2	0	51	0.131
0	1	0	0	1	10	3	5	5	6	23	6	7	65	0.168
0	0	1	1	0	1	0	1	0	0	0	0	0	2	0.005
0	0	1	0	1	59	42	18	61	51	80	54	28	393	1.013
0	0	0	1	1	2	2	2	0	1	0	0	0	7	0.018
rejected by 3 of 5 tests														
1	1	1	0	0	0	0	0	0	0	1	0	0	1	0.003
1	1	0	1	0	246	254	160	344	309	405	426	321	2465	6.354
1	1	0	0	1	0	0	0	0	0	0	1	0	1	0.003
1	0	1	1	0	1	1	0	2	1	3	2	3	13	0.034
1	0	1	0	1	0	0	0	0	0	0	1	0	1	0.003
1	0	0	1	1	5	3	0	7	8	20	12	5	60	0.155
0	1	1	1	0	2	1	0	1	0	0	0	1	5	0.013
0	1	1	0	1	6	8	4	9	4	14	7	3	55	0.142
0	1	0	1	1	5	2	2	3	0	0	0	0	12	0.031
0	0	1	1	1	0	2	1	1	0	0	0	0	4	0.01
rejected by 4 of 5 tests														
1	1	1	1	0	194	168	138	194	170	186	168	129	1347	3.472
1	1	1	0	1	0	0	0	0	1	1	0	0	2	0.005
1	1	0	1	1	97	99	90	208	191	362	391	355	1793	4.622
1	0	1	1	1	3	5	2	6	3	8	13	3	43	0.111
0	1	1	1	1	8	4	4	2	1	0	3	0	22	0.057
rejected by all tests														
1	1	1	1	1	1852	2712	3916	2452	3361	1625	2782	3703	22403	57.75
			total		4546	4633	4752	4926	4947	4997	4997	4997	38795	100.0

Notes: Rejection patterns (left section) indicate which combinations of the testing methods (listed in the header) rejected (=1) or accepted (=0) H_0. The remainder of the table is a tabulation of the occurrences per parameter settings (columns, as listed in header).

Table 28. Rejection Patterns for Power Comparisons in Table 16 with a=1 and samples of size n=50.

Testing Methods					a=1, n=50			Settings for Means						
					μ2=1			μ2=2		μ2=5				
EQLS	PLS	DDT	LRT	t	μ1=2	μ1=2.5	μ1=3	μ1=4	μ1=5	μ1=8.75	μ1=10	μ1=12.5	μ1=15		
rejection pattern					counts of occurence in this pattern									total	percent
rejected by no tests															
0	0	0	0	0	1913	872	348	1436	527	1789	1108	283	78	8354	18.66
rejected by 1 of 5 tests															
1	0	0	0	0	4	2	3	5	1	0	3	1	3	22	0.049
0	1	0	0	0	36	35	17	61	29	53	54	25	11	321	0.717
0	0	1	0	0	61	27	14	70	26	88	61	21	3	371	0.829
0	0	0	1	0	5	2	2	3	3	4	2	0	0	21	0.047
0	0	0	0	1	77	57	31	62	33	60	55	19	7	401	0.896
rejected by 2 of 5 tests															
1	1	0	0	0	1	1	1	0	0	3	1	0	0	7	0.016
1	0	1	0	0	0	0	0	0	1	0	0	0	0	1	0.002
1	0	0	1	0	80	72	48	103	75	92	105	65	29	669	1.494
1	0	0	0	1	1	0	0	2	0	0	0	0	0	3	0.007
0	1	1	0	0	1	1	1	4	1	7	5	0	2	22	0.049
0	1	0	1	0	7	9	4	4	6	3	4	1	2	40	0.089
0	1	0	0	1	10	20	12	21	14	13	20	6	2	118	0.264
0	0	1	1	0	1	0	0	0	0	1	0	0	0	2	0.004
0	0	1	0	1	89	55	33	84	38	94	74	28	10	505	1.128
0	0	0	1	1	1	0	0	0	0	0	1	0	0	2	0.004
rejected by 3 of 5 tests															
1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	0	1	0	302	295	210	397	386	613	656	519	308	3686	8.234
1	1	0	0	1	0	0	0	1	0	2	0	0	0	3	0.007
1	0	1	1	0	2	1	0	9	4	2	5	3	0	26	0.058
1	0	1	0	1	1	1	2	0	1	1	0	0	0	6	0.013
1	0	0	1	1	27	15	15	13	8	15	21	6	7	127	0.284
0	1	1	1	0	0	0	0	1	0	0	0	0	0	1	0.002
0	1	1	0	1	19	16	5	21	9	14	18	8	3	113	0.252
0	1	0	1	1	6	2	3	4	1	0	1	0	0	17	0.038
0	0	1	1	1	3	0	2	0	0	0	0	0	0	5	0.011
rejected by 4 of 5 tests															
1	1	1	1	0	84	79	57	125	96	81	89	56	25	692	1.546
1	1	1	0	1	0	0	1	0	0	1	0	0	0	2	0.004
1	1	0	1	1	250	299	293	328	365	522	621	722	643	4043	9.031
1	0	1	1	1	13	11	14	13	6	10	13	9	3	92	0.206
0	1	1	1	1	4	4	2	3	1	0	2	0	1	17	0.038
rejected by all tests															
1	1	1	1	1	1915	3054	3838	2215	3364	1530	2077	3224	3860	25077	56.02
				total	4913	4930	4956	4985	4995	4998	4996	4996	4997	44766	100.0

Notes: Rejection patterns (left section) indicate which combinations of the testing methods (listed in the header) rejected (=1) or accepted (=0) H_0. The remainder of the table is a tabulation of the occurrences per parameter settings (columns, as listed in header).

Table 29. Rejection Patterns for H_0 settings for samples of size n=30.

Testing Methods: EQL. PLS DDT LRT t	n=30			H_0 Parameter settings							
		a=2			*a*=5			*a*=1			
	μ=1	μ=2	μ=5	μ=1	μ=2	μ=5	μ=1	μ=2	μ=5		
rejection pattern	counts of occurence in this pattern									total	percent
rejected by no tests											
0 0 0 0 0	6043	6688	7692	7021	8053	8790	7825	8607	8916	69635	92.36
rejected by 1 of 5 tests											
1 0 0 0 0	0	0	0	1	2	2	3	5	2	15	0.020
0 1 0 0 0	0	0	5	3	16	17	21	31	31	124	0.164
0 0 1 0 0	34	72	110	49	97	166	114	147	202	991	1.314
0 0 0 1 0	42	35	6	20	7	1	10	2	4	127	0.168
0 0 0 0 1	8	33	46	31	56	63	21	44	68	370	0.491
rejected by 2 of 5 tests											
1 1 0 0 0	0	0	0	0	0	0	1	3	2	6	0.008
1 0 1 0 0	0	0	0	0	0	0	0	0	0	0	0
1 0 0 1 0	6	6	12	13	22	25	17	32	57	190	0.252
1 0 0 0 1	0	0	0	0	0	0	1	1	0	2	0.003
0 1 1 0 0	0	0	1	0	4	6	1	3	1	16	0.021
0 1 0 1 0	4	9	4	11	8	4	9	3	2	54	0.072
0 1 0 0 1	0	0	3	2	2	12	3	9	22	53	0.070
0 0 1 1 0	20	9	4	0	0	0	0	0	0	33	0.044
0 0 1 0 1	20	34	50	33	41	67	35	68	103	451	0.598
0 0 0 1 1	2	4	1	1	2	0	0	1	0	11	0.015
rejected by 3 of 5 tests											
1 1 1 0 0	0	0	0	0	0	0	0	0	1	1	0.001
1 1 0 1 0	9	26	41	41	60	100	67	86	145	575	0.763
1 1 0 0 1	0	0	0	0	0	0	0	2	1	3	0.004
1 0 1 1 0	0	2	1	0	0	0	0	1	3	7	0.009
1 0 1 0 1	0	0	0	0	0	0	0	0	1	1	0.001
1 0 0 1 1	3	0	6	1	7	10	6	8	5	46	0.061
0 1 1 1 0	7	10	0	4	1	0	0	2	0	24	0.032
0 1 1 0 1	0	0	4	3	7	5	7	10	18	54	0.072
0 1 0 1 1	0	0	1	1	1	1	3	3	2	12	0.016
0 0 1 1 1	19	3	1	1	0	0	1	0	0	25	0.033
rejected by 4 of 5 tests											
1 1 1 1 0	53	62	47	33	42	29	22	16	13	317	0.420
1 1 1 0 1	0	0	0	0	0	1	0	0	1	2	0.003
1 1 0 1 1	2	6	42	18	31	94	37	76	111	417	0.553
1 0 1 1 1	1	1	3	4	3	4	0	2	4	22	0.029
0 1 1 1 1	10	5	4	8	7	1	6	3	2	46	0.061
rejected by all tests											
1 1 1 1 1	154	157	187	238	232	200	235	218	144	1765	2.341
total	6437	7162	8271	7537	8701	9598	8445	9383	9861	75395	100.0

Notes: Rejection patterns (left section) indicate which combinations of the testing methods (listed in the header) rejected (=1) or accepted (=0) H_0. The remainder of the table is a tabulation of the occurrences per parameter settings (columns, as listed in header).

Table 30. Rejection Patterns for H_0 settings for samples of size n=50.

Testing Methods					n=50			H_0 Parameter settings							
E Q L. S	P L. S	D D T	L R T	t		*a*=.2			*a*=.5			*a*=1			
					μ=1	μ=2	μ=5	μ=1	μ=2	μ=5	μ=1	μ=2	μ=5		
rejection pattern					counts of occurence in this pattern									total	percent
rejected by no tests															
0	0	0	0	0	6897	7816	8923	8022	8918	9152	8707	9131	9072	76638	92.28
rejected by 1 out of 5 tests															
1	0	0	0	0	0	0	0	0	0	0	1	3	3	7	0.008
0	1	0	0	0	0	0	8	10	14	21	13	18	38	122	0.147
0	0	1	0	0	51	50	121	68	100	122	79	130	182	903	1.087
0	0	0	1	0	21	17	1	8	2	1	2	3	1	56	0.067
0	0	0	0	1	21	59	81	43	80	84	76	69	75	588	0.708
rejected by 2 out of 5 tests															
1	1	0	0	0	0	0	0	0	1	1	1	2	1	6	0.007
1	0	1	0	0	0	0	0	0	0	1	0	1	0	2	0.002
1	0	0	1	0	5	8	15	12	15	23	23	31	46	178	0.214
1	0	0	0	1	0	0	1	0	0	0	1	0	0	2	0.002
0	1	1	0	0	0	0	3	4	2	5	1	7	1	23	0.028
0	1	0	1	0	1	7	1	7	0	1	4	1	1	23	0.028
0	1	0	0	1	0	0	6	2	5	7	9	7	11	47	0.057
0	0	1	1	0	9	3	0	0	0	0	0	0	1	13	0.016
0	0	1	0	1	30	38	47	38	64	80	73	70	105	545	0.656
0	0	0	1	1	4	3	0	0	1	0	2	0	1	11	0.013
rejected by 3 out of 5 tests															
1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0.000
1	1	0	1	0	29	47	96	64	101	114	68	104	149	772	0.930
1	1	0	0	1	0	0	0	0	0	0	0	1	0	1	0.001
1	0	1	1	0	0	0	2	1	2	2	0	1	3	11	0.013
1	0	1	0	1	0	0	0	0	0	0	1	1	1	3	0.004
1	0	0	1	1	0	1	2	5	6	5	5	4	6	34	0.041
0	1	1	1	0	7	0	1	1	0	0	0	2	0	11	0.013
0	1	1	0	1	0	0	3	2	7	7	5	9	11	44	0.053
0	1	0	1	1	0	2	0	0	1	0	1	1	0	5	0.006
0	0	1	1	1	5	5	0	0	1	0	0	0	1	12	0.014
rejected by 4 out of 5 tests															
1	1	1	1	0	58	86	64	60	60	48	28	32	19	455	0.548
1	1	1	0	1	0	0	0	0	1	0	0	0	0	1	0.001
1	1	0	1	1	2	11	55	12	37	71	57	72	96	413	0.497
1	0	1	1	1	0	1	1	2	1	9	1	5	3	23	0.028
0	1	1	1	1	3	5	0	7	3	0	3	1	0	22	0.026
rejected by all tests															
1	1	1	1	1	207	224	242	248	241	225	280	239	173	2079	2.503
				total	7350	8383	9673	8616	9663	9979	9441	9945	10000	83050	100.0

Notes: Rejection patterns (left section) indicate which combinations of the testing methods (listed in the header) rejected (=1) or accepted (=0) H_0. The remainder of the table is a tabulation of the occurrences per parameter settings (columns, as listed in header).

APPENDIX G.

FIGURES

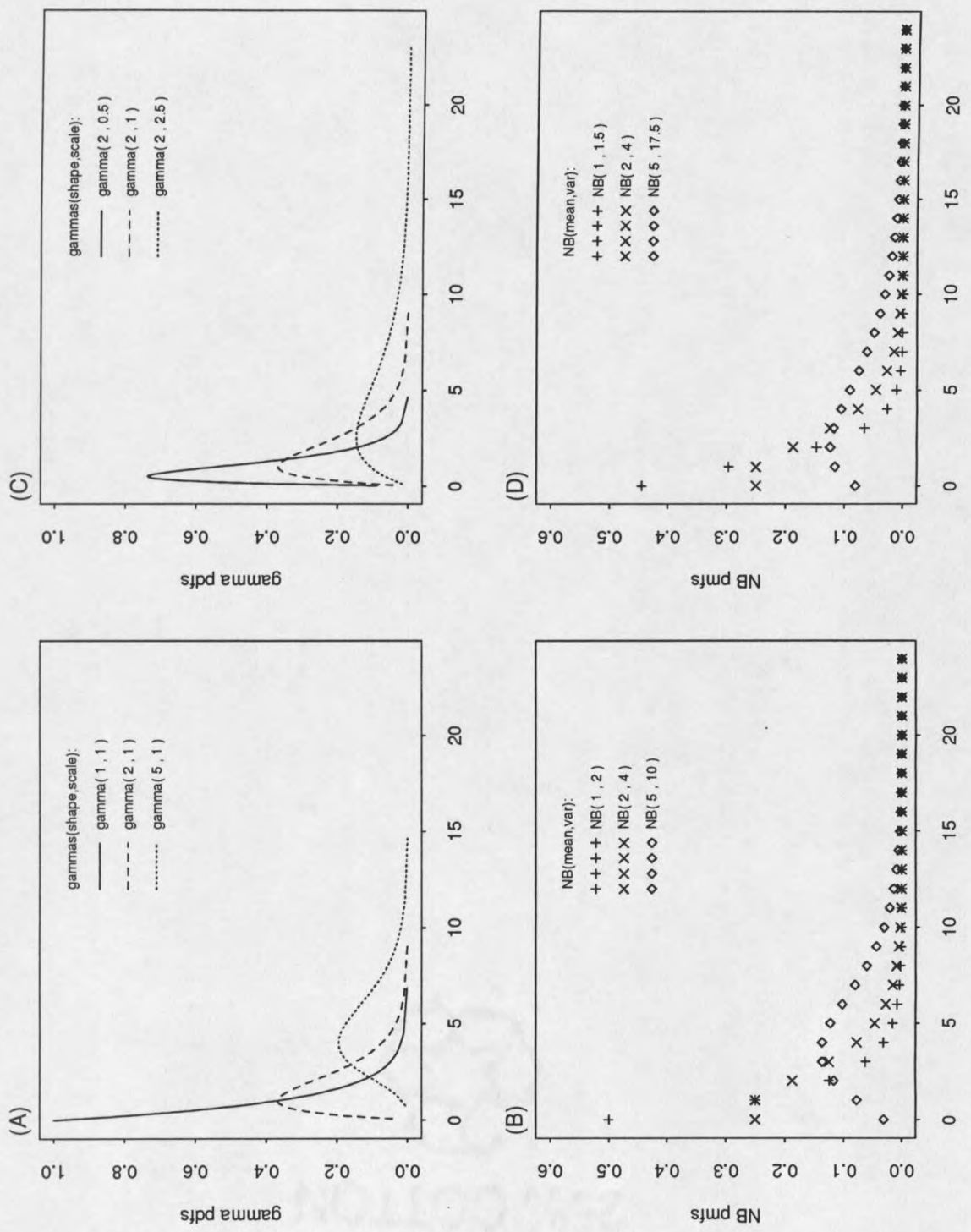

Figure 1. Fixed shape vs. fixed scale: Gammas and resultant NBs.

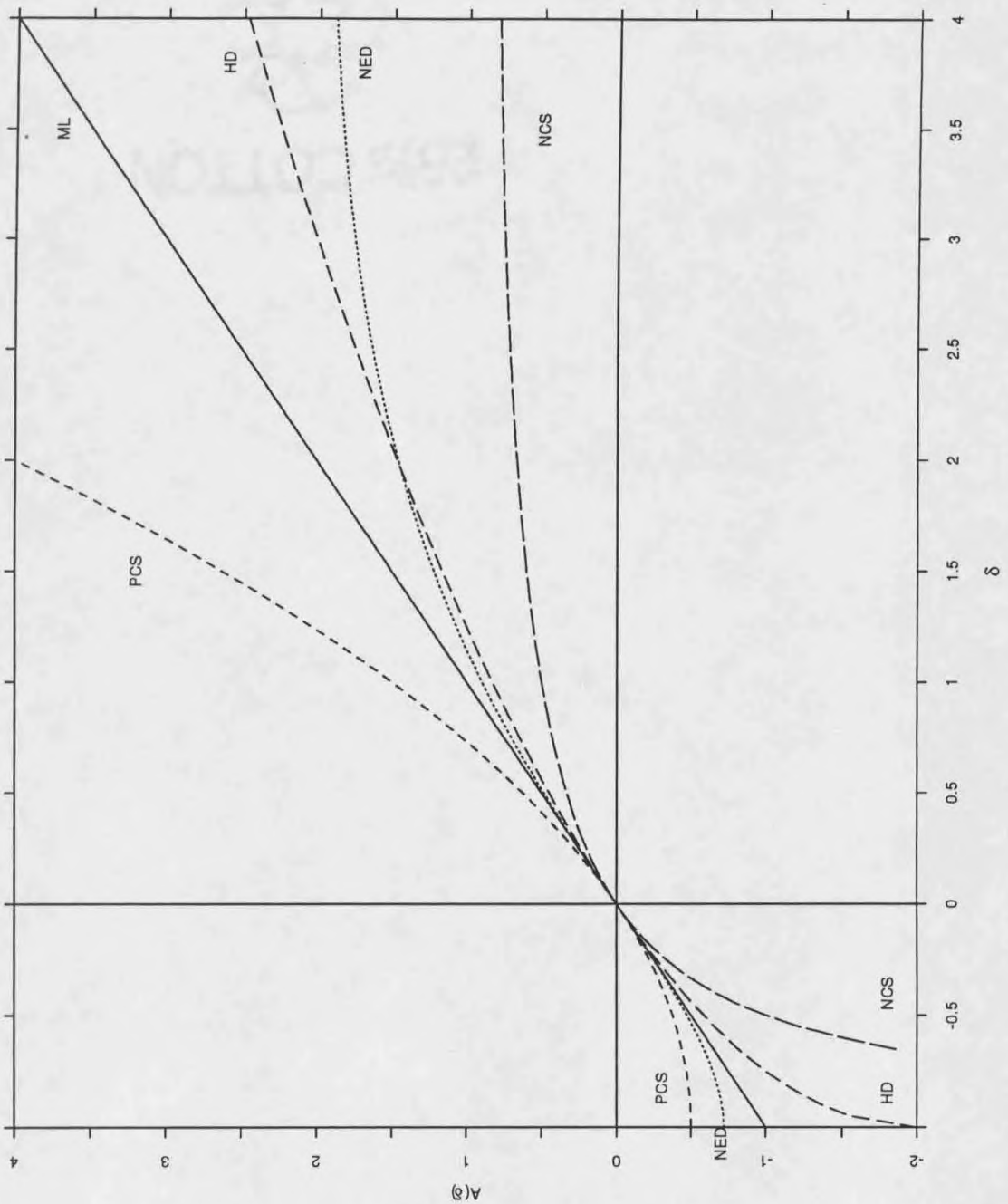

Figure 2. Residual Adjustment Functions for ML, NED, NCS, PCS, HD.

NCS=Neyman's chi-squared, PCS=Pearson's chi-squared, HD=Hellinger distance.

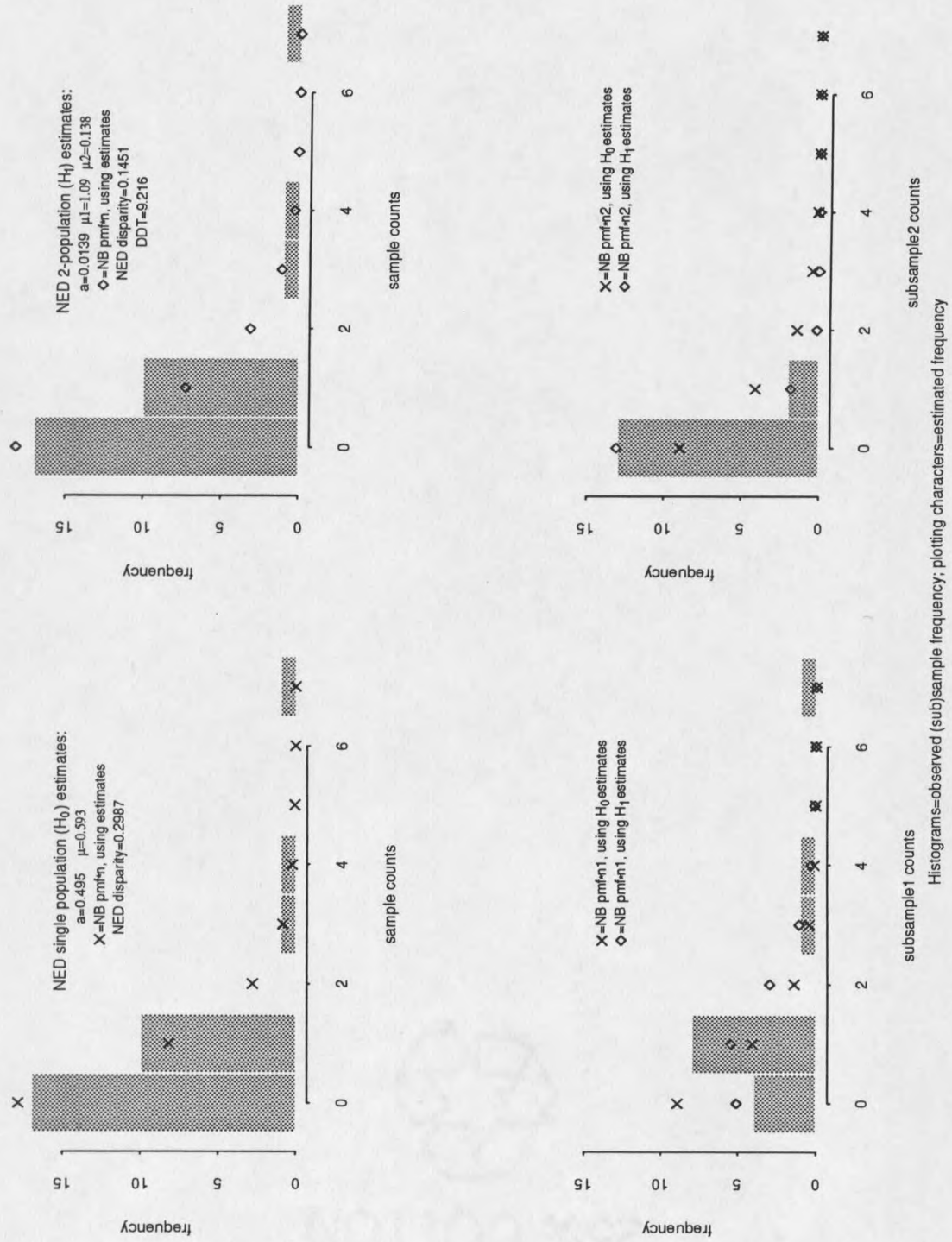

Figure 3. NED estimated frequency under H_0 and H_1 estimation and resultant DDT statistic.

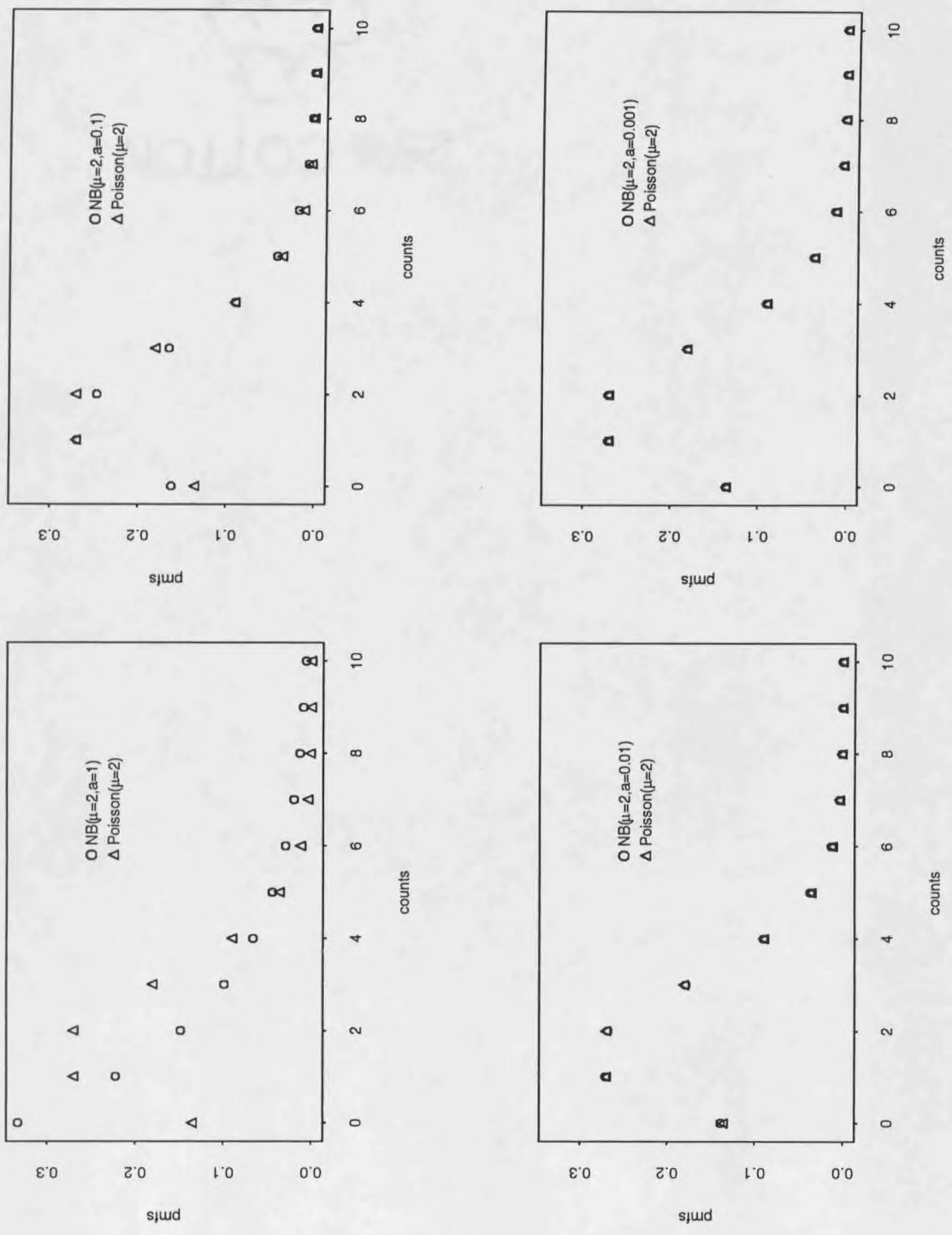

Figure 4. NB(μ,*a*) approaching Poisson(μ) with decreasing *a*.

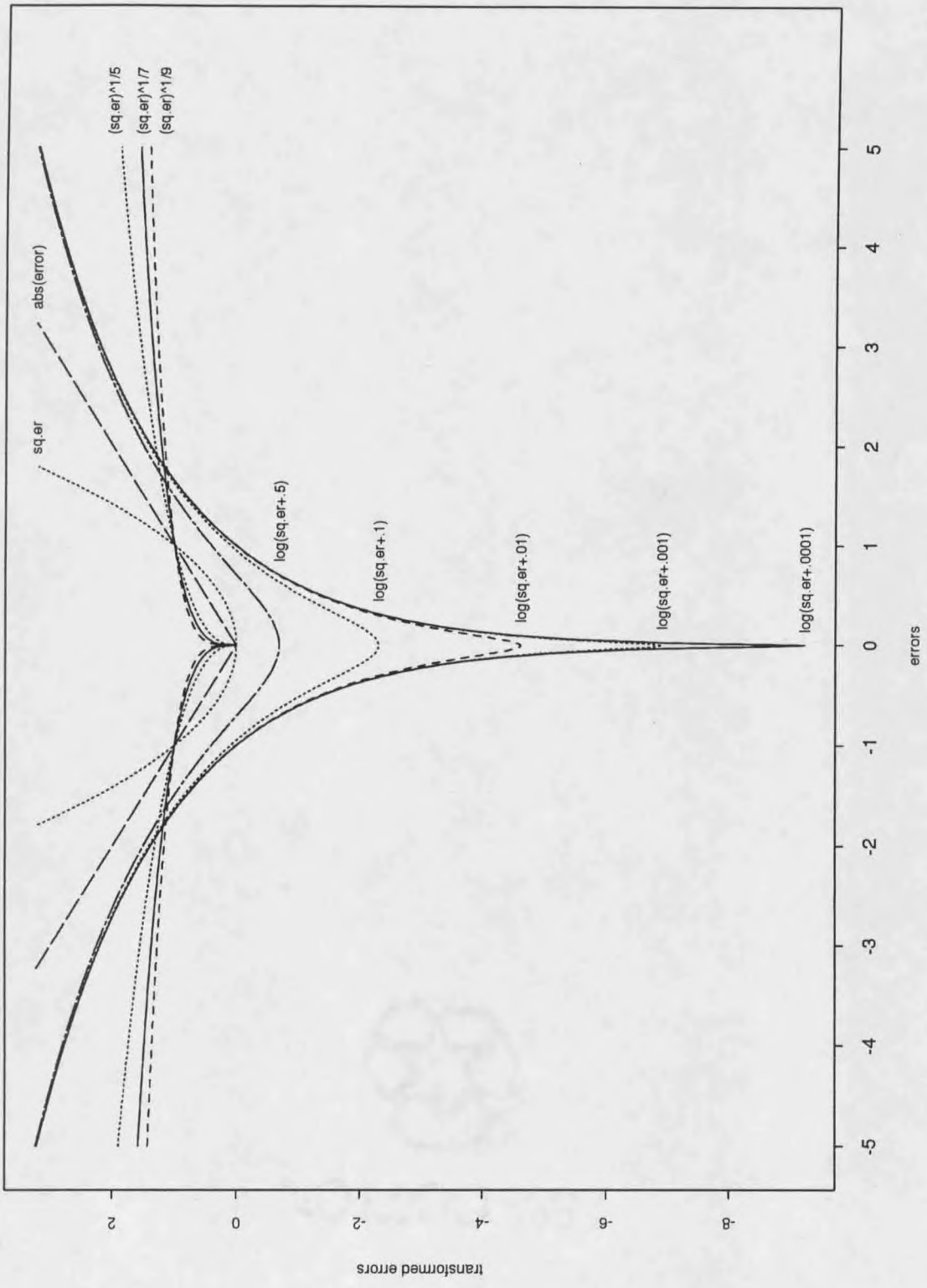

Figure 5. Comparison of transformations on errors in range about zero.

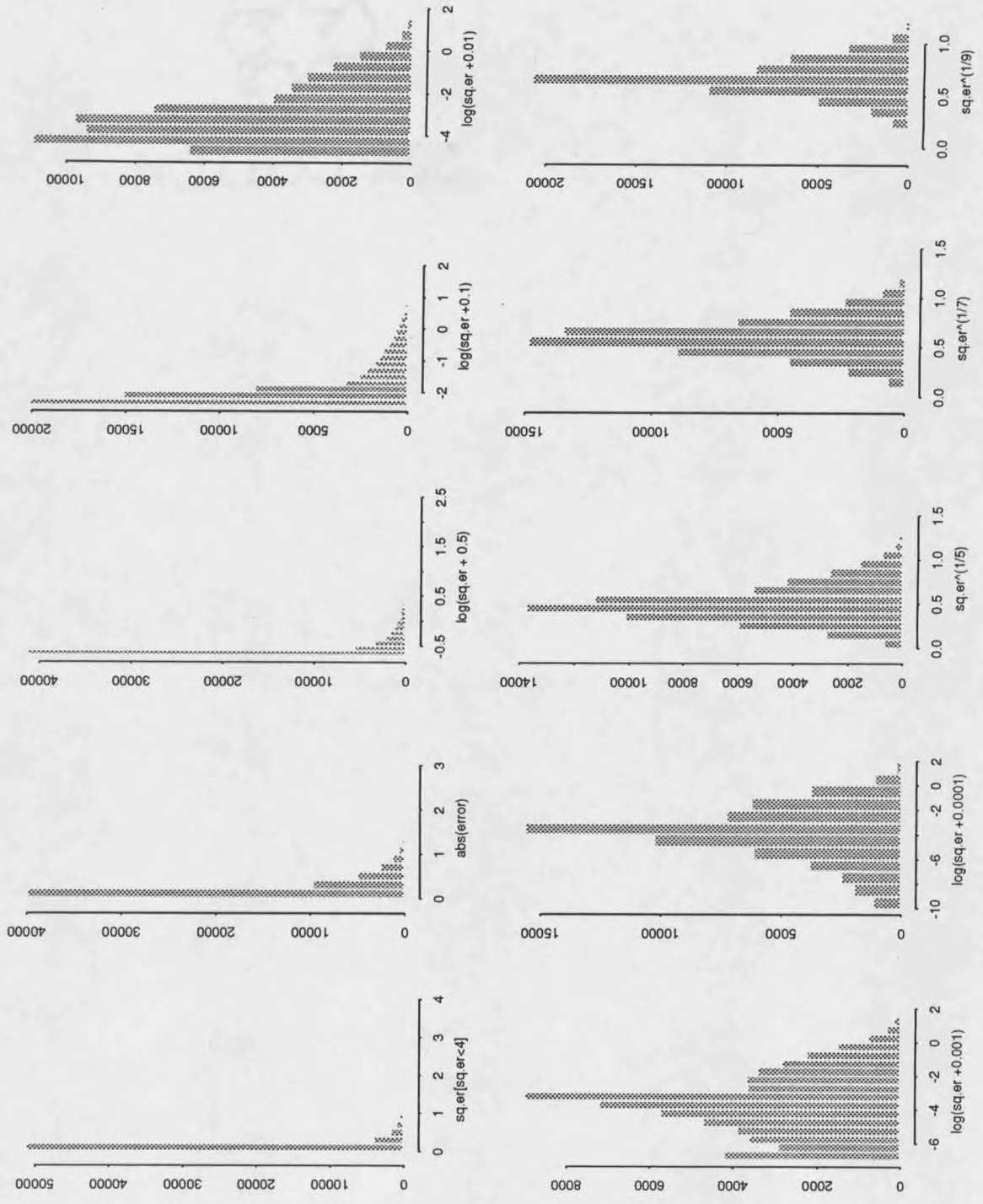

Figure 6. Distributions of transformed sq.er for compared transformations.

Data are combined for all estimation methods from 10000 samples of size n=30 generated from NB(μ=1,a=0.2).

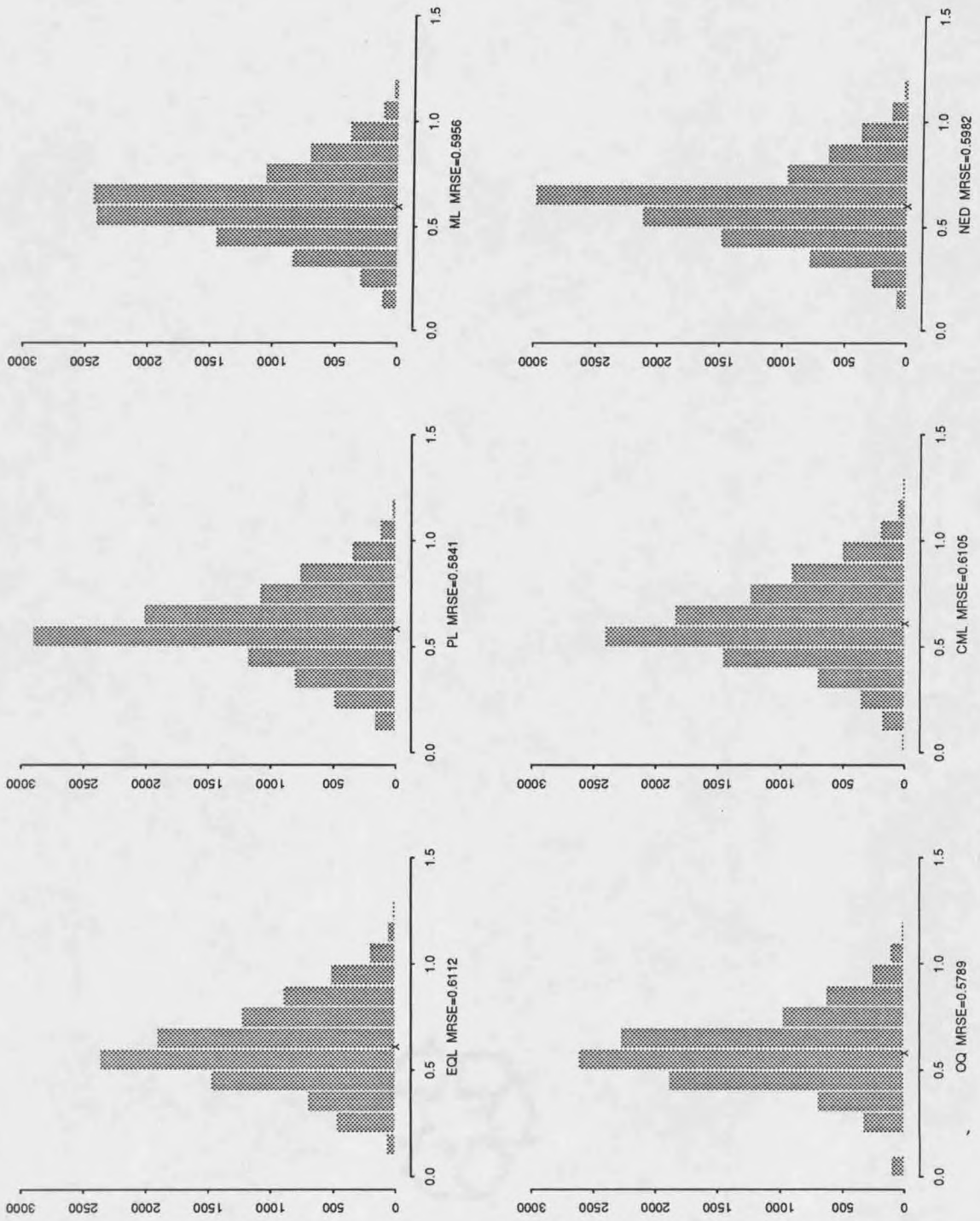

Figure 7. Distributions of RSE per estimation method, using $\hat{a}=0$ for $\hat{a}<0$. MRSE values plotted as X. Data from 10000 samples of size n=30 generated from NB(μ=1,a=.2).

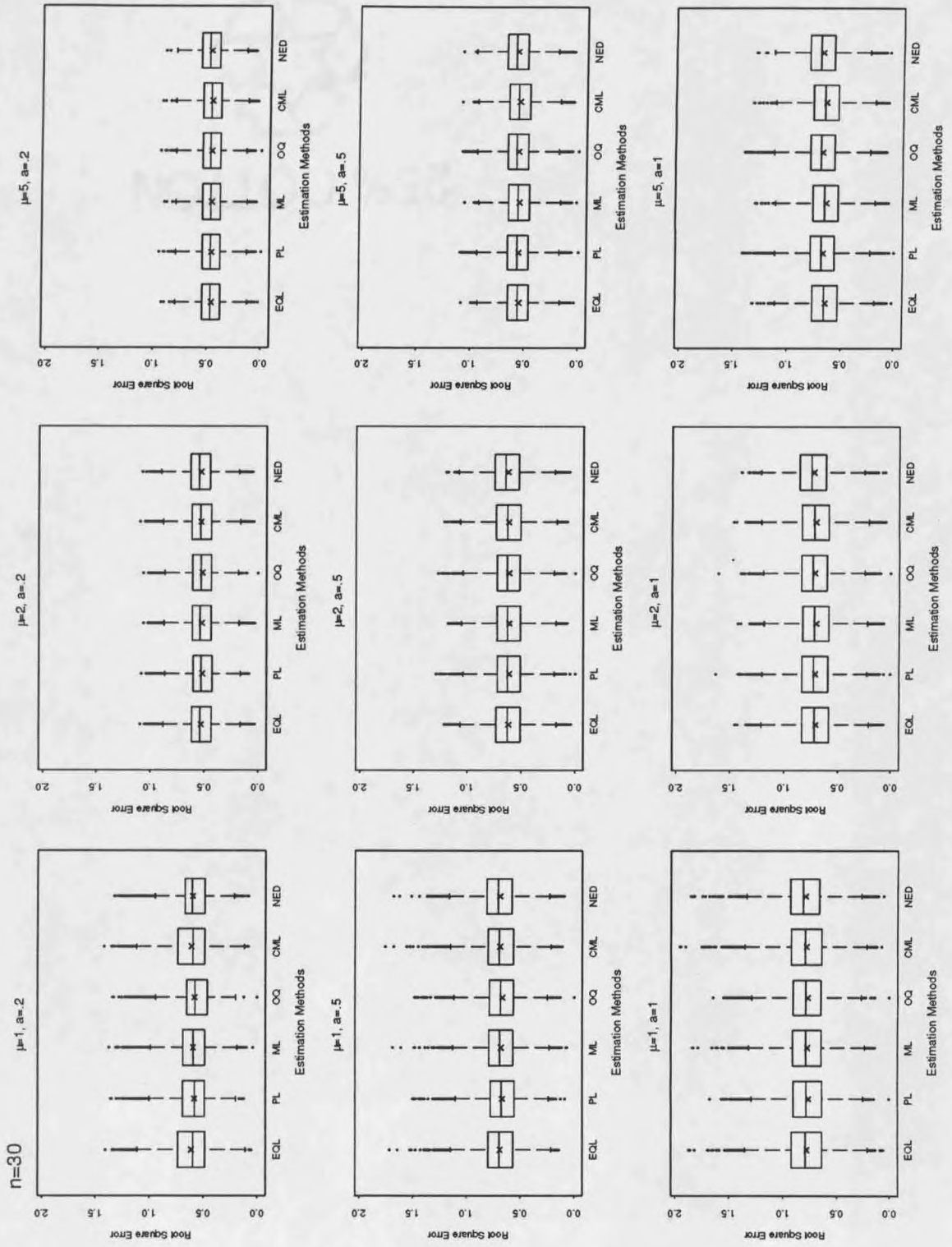

Figure 8. RSE boxplots per estimation method at H_0 settings for samples of size n=30.

MRSE values plotted as X.

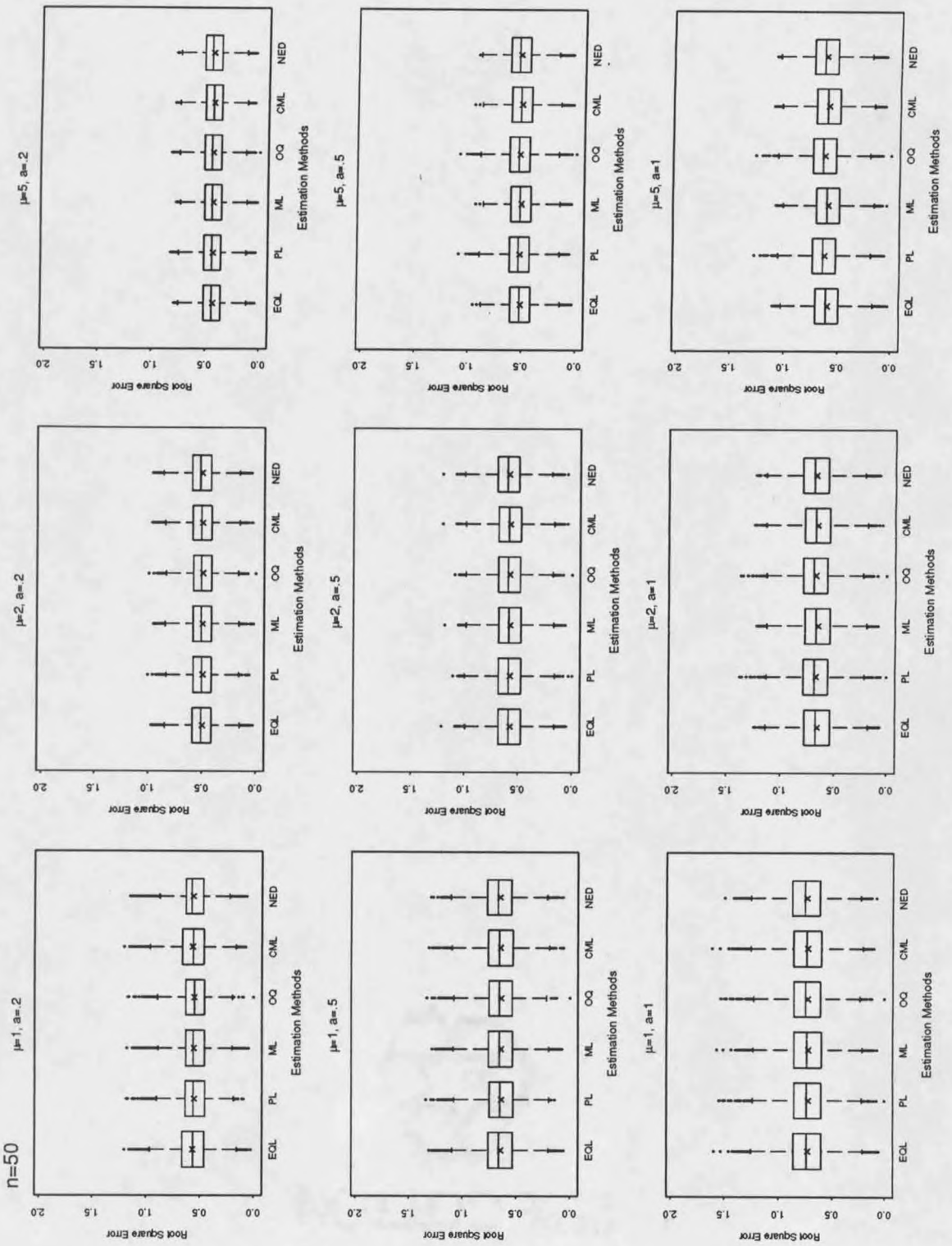

Figure 9. RSE boxplots per estimation method at H_0 setting for samples of size n=50.

MRSE values plotted as X.

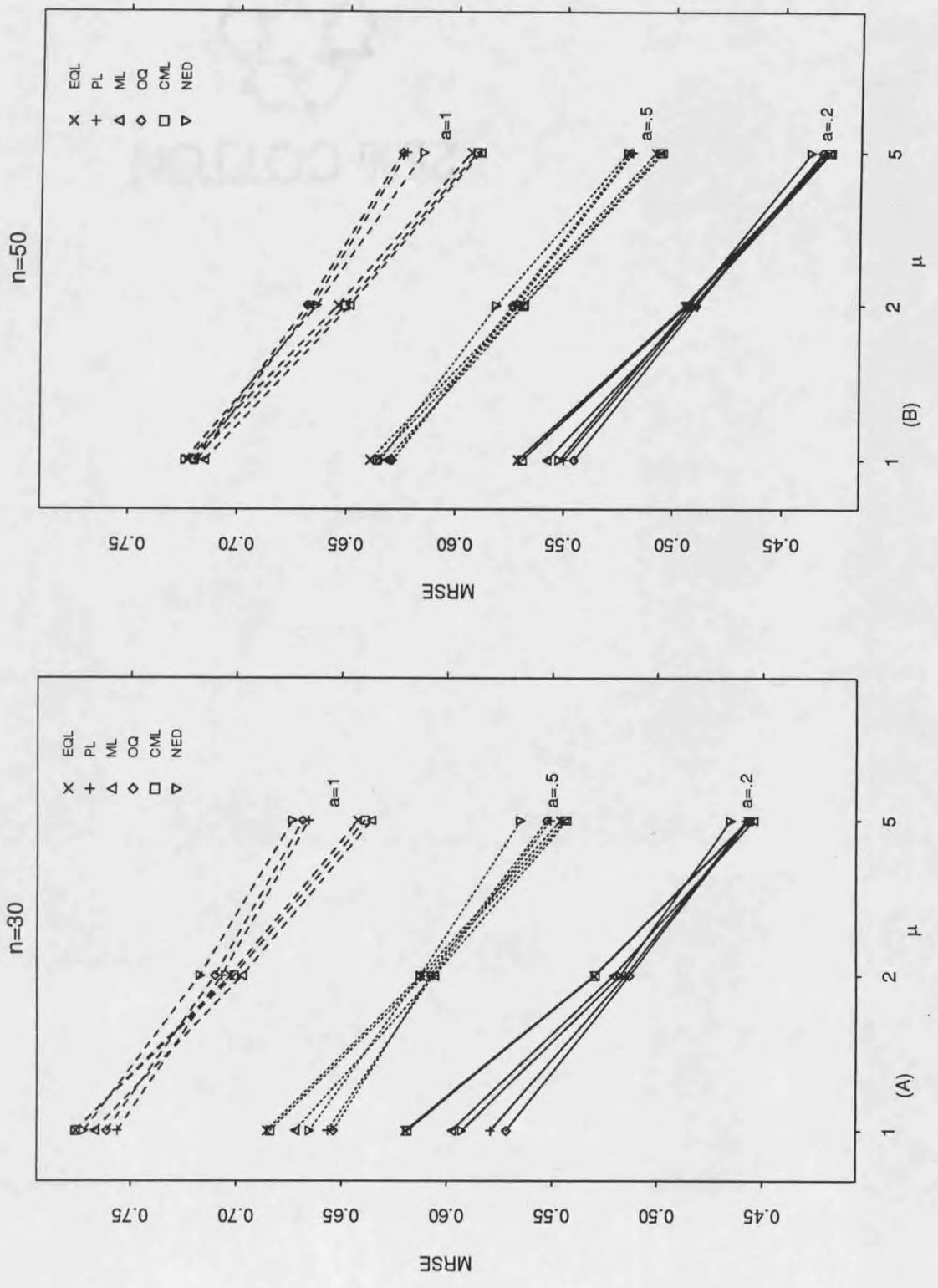

Figure 10. Comparison of MRSE across H_0 settings.

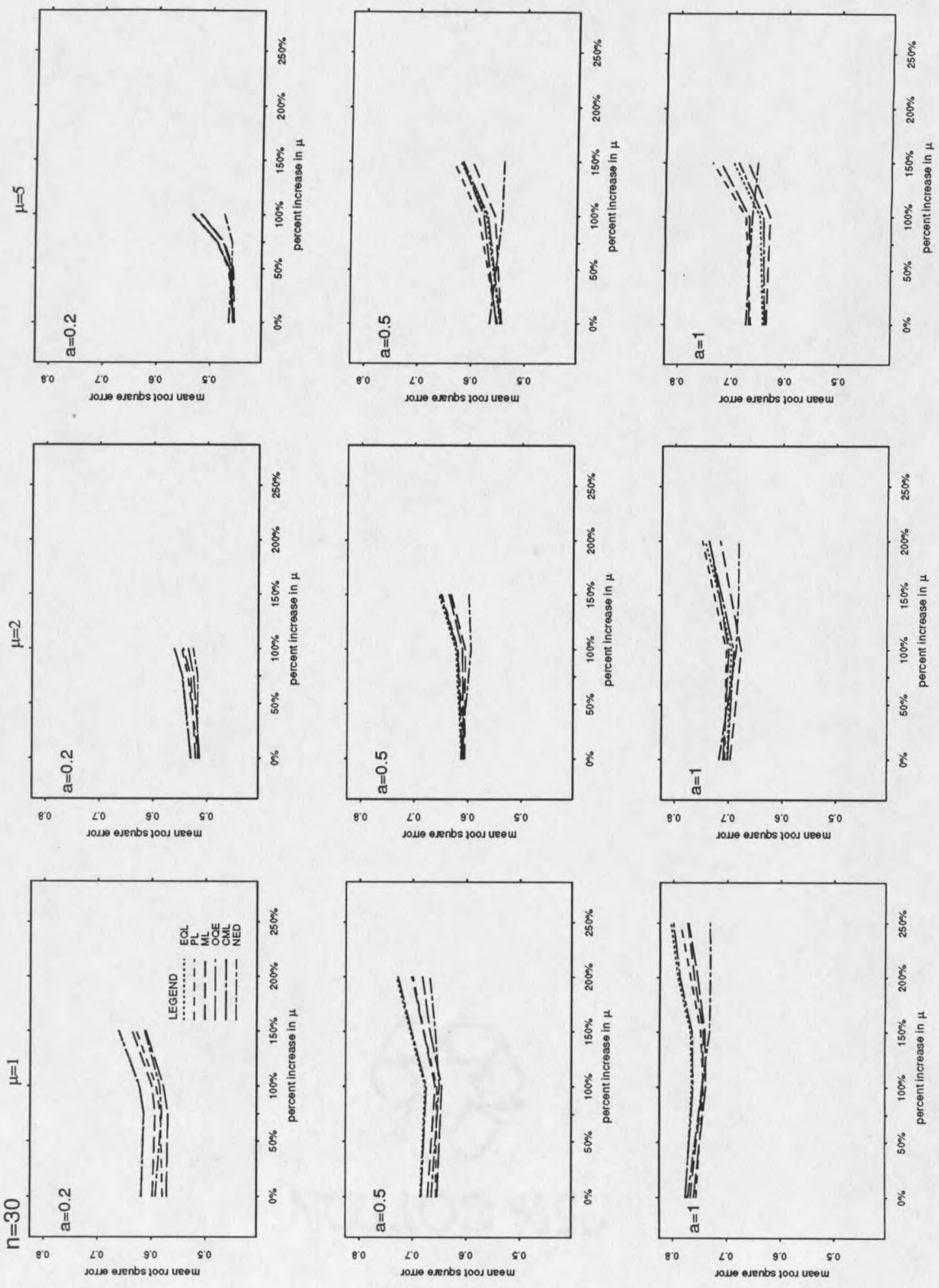

Figure 11. MRSE plots for samples of size n=30.

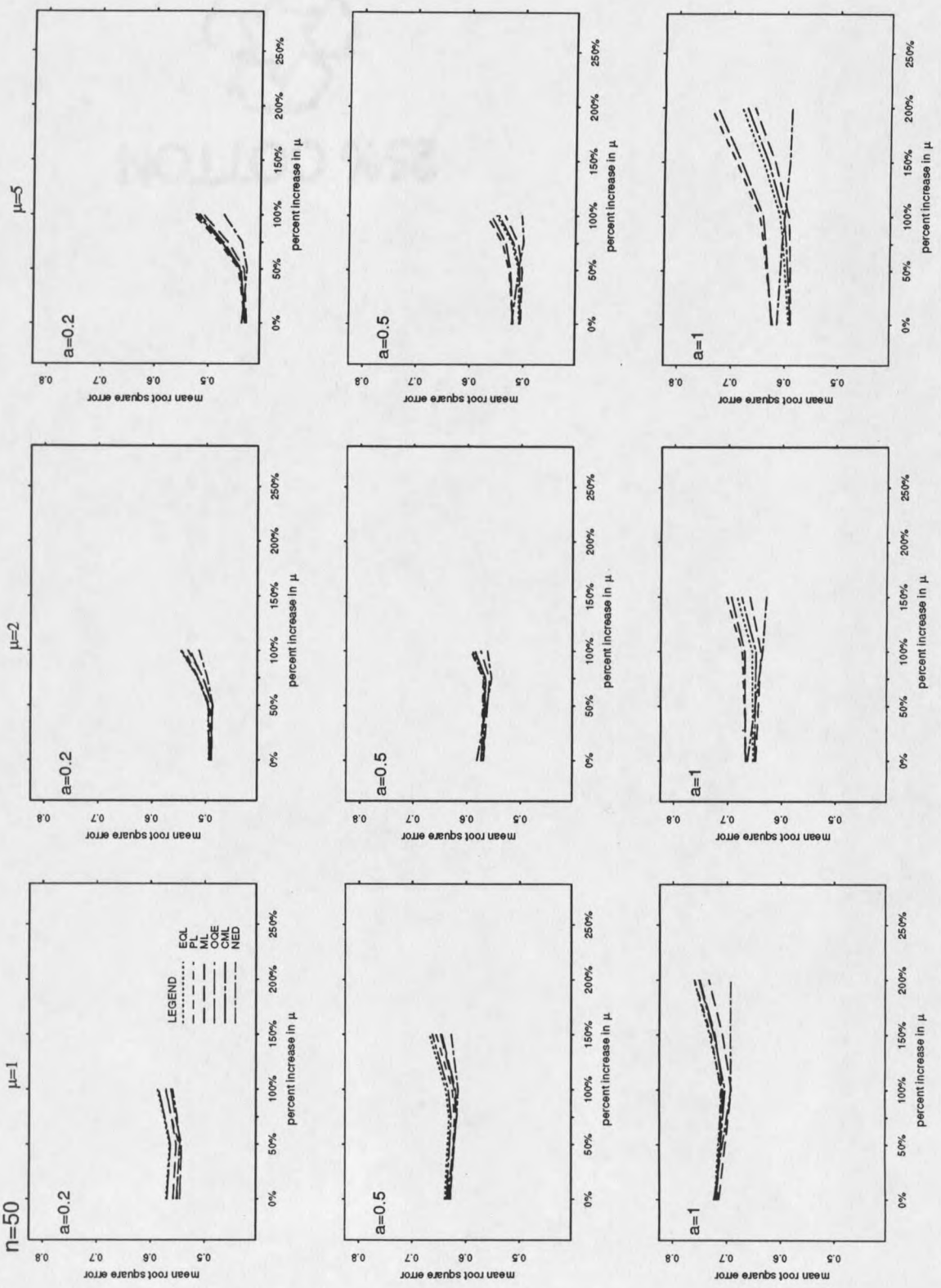

Figure 12. MRSE plots for samples of size n=50.

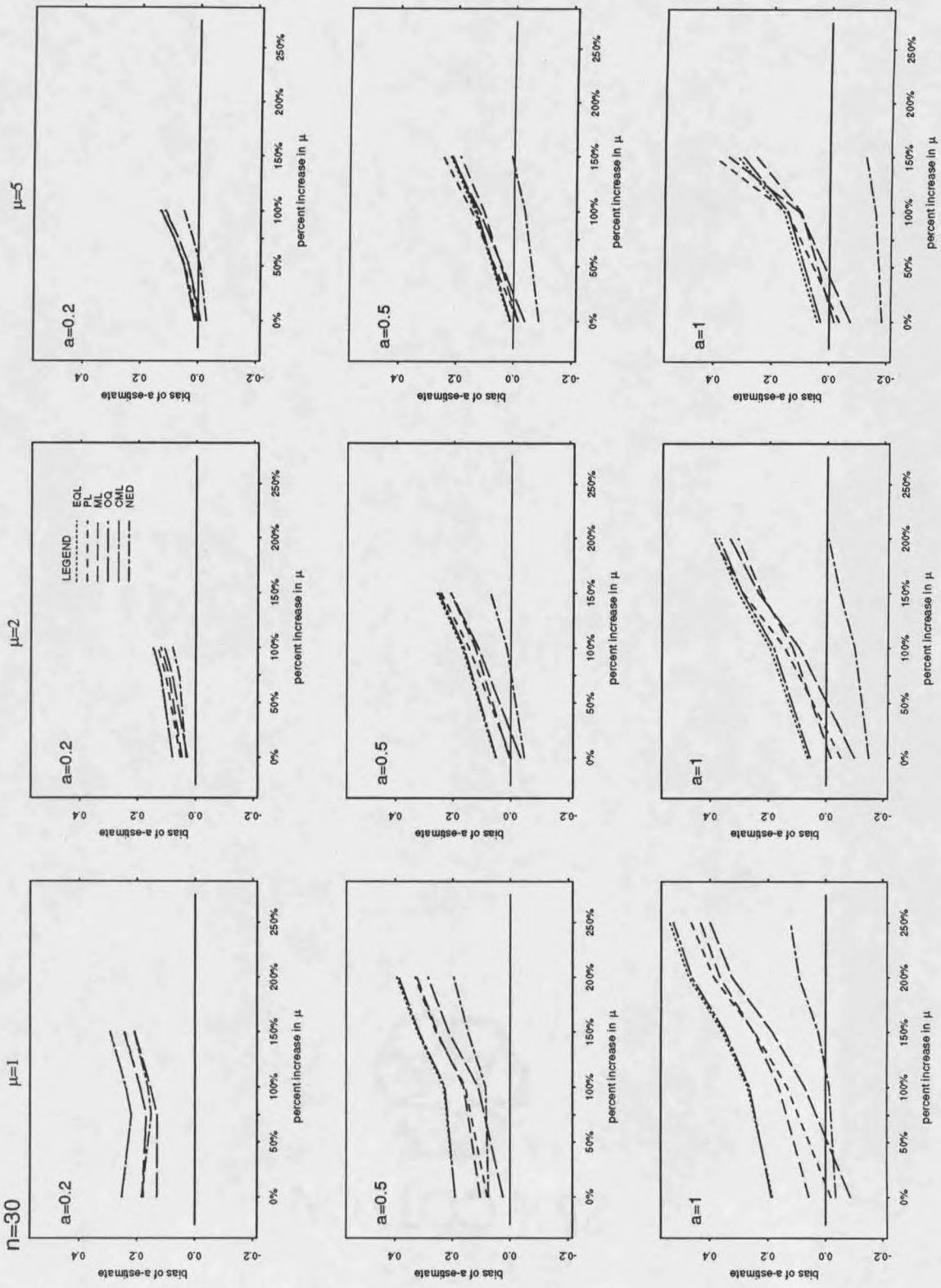

Figure 13. Bias plots for samples of size n=30.

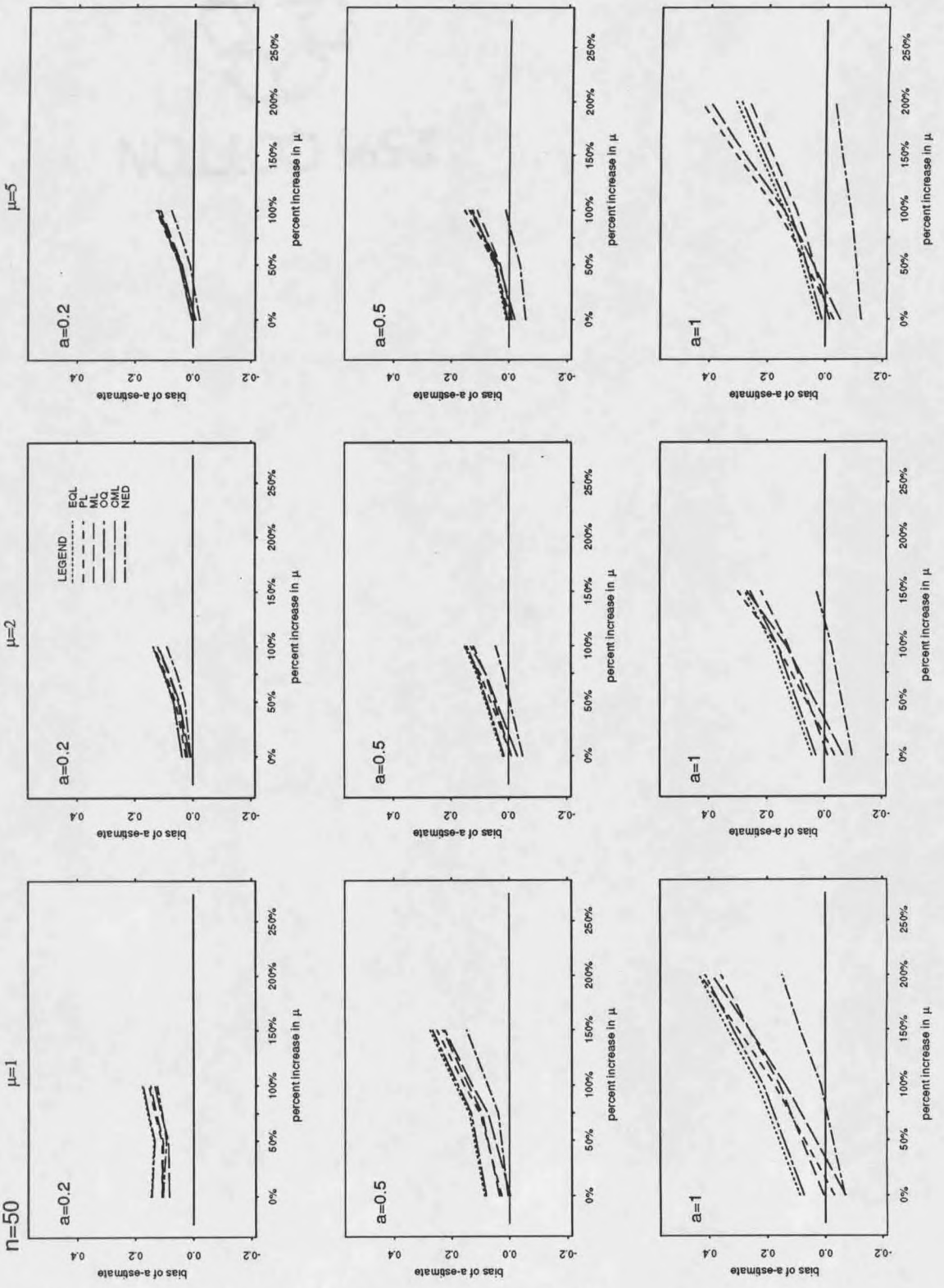

Figure 14. Bias plots for samples of size n=50.

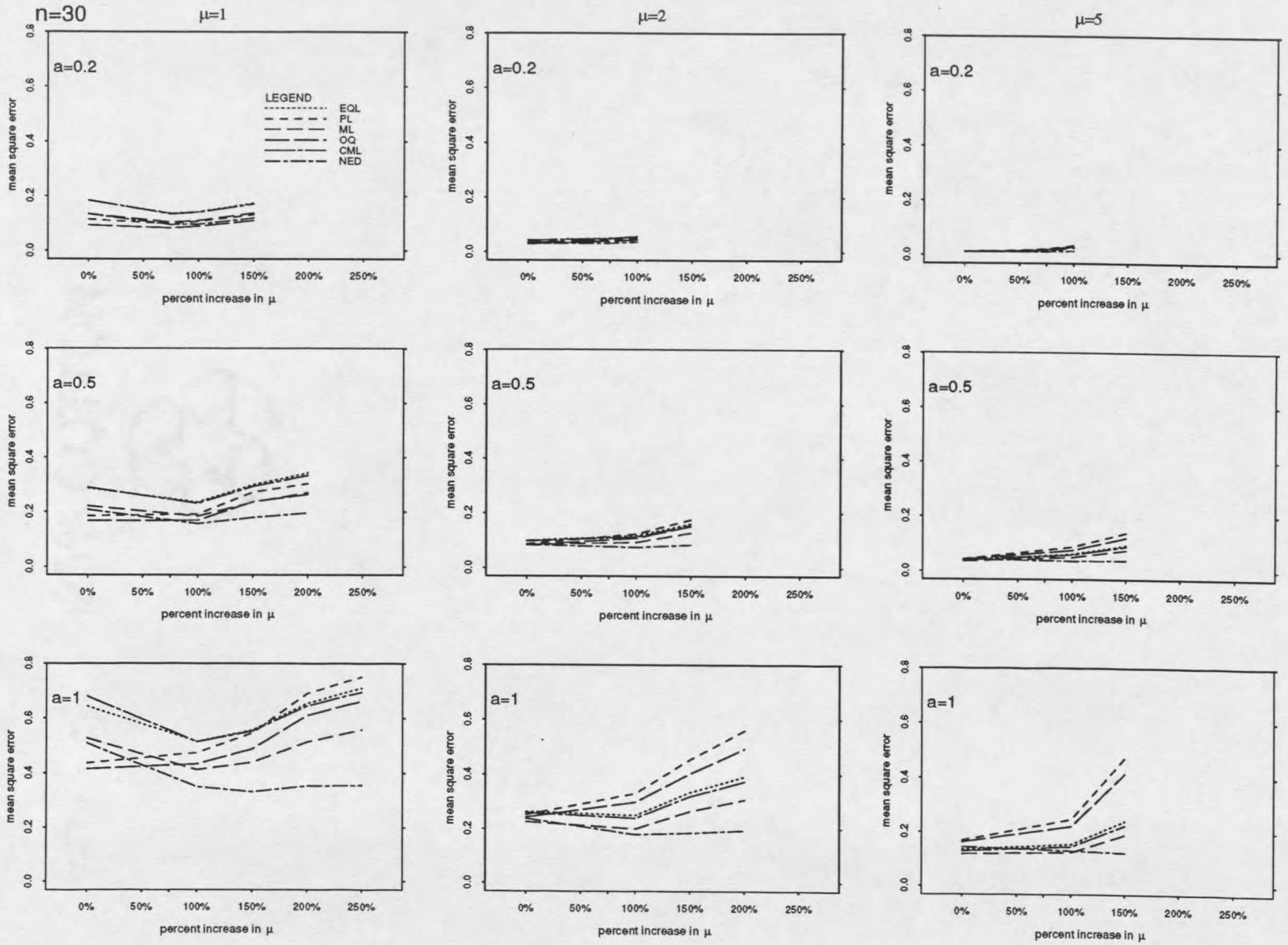

Figure 15. MSE plots for samples of size n=30.

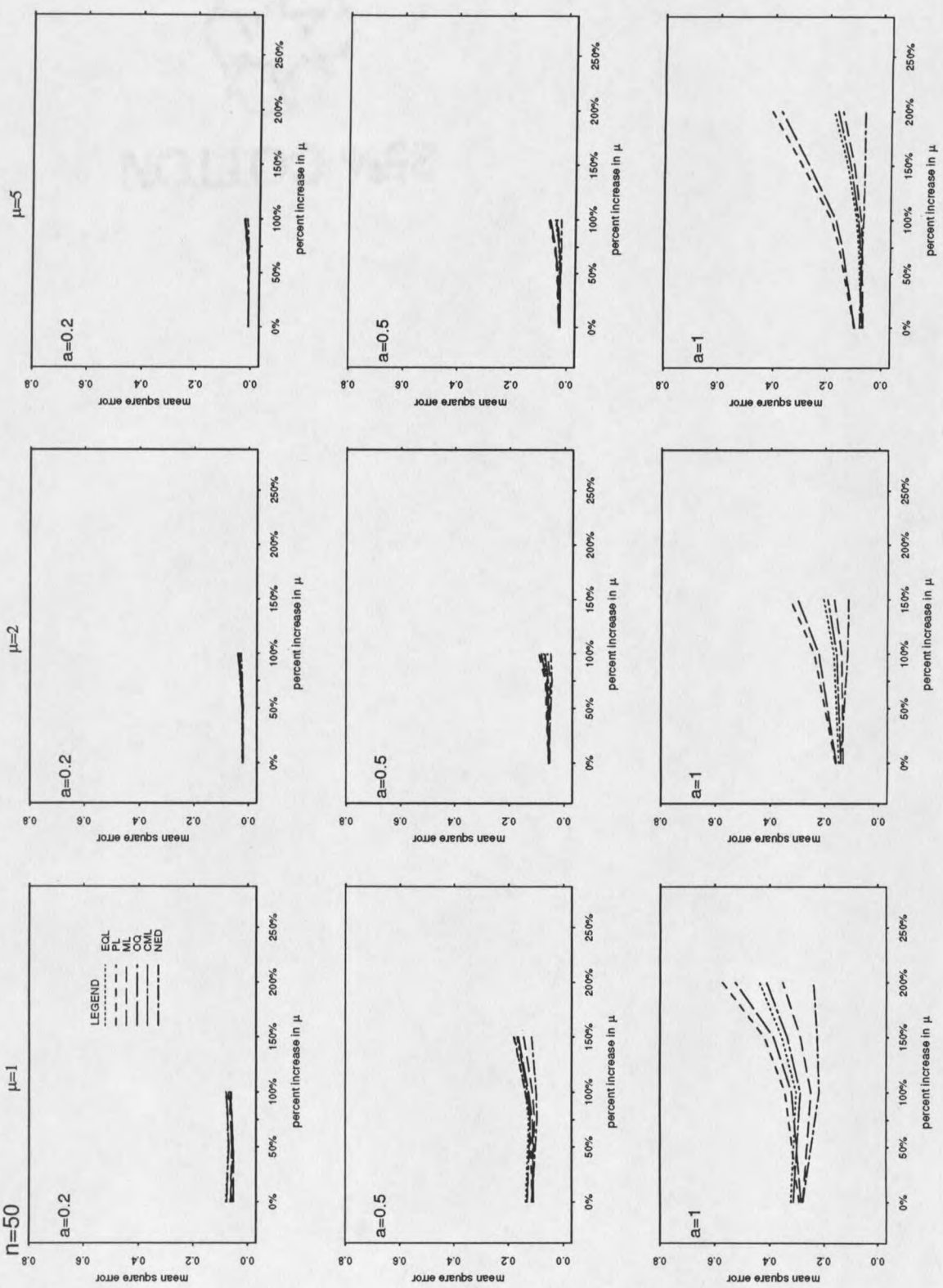

Figure 16. MSE plots for samples of size n=50.

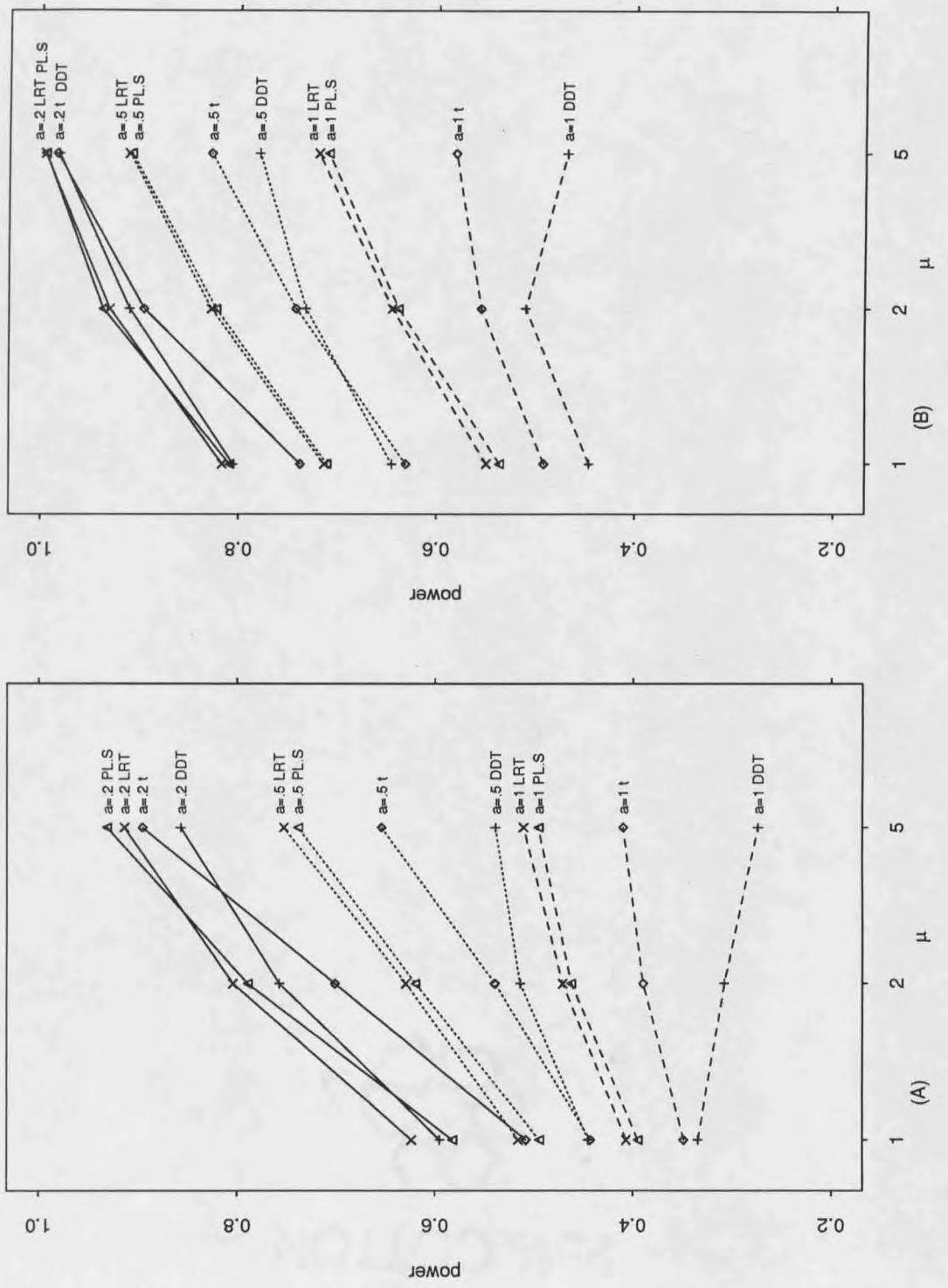

Figure 17. Comparison of Power to Detect 100% increase in μ.

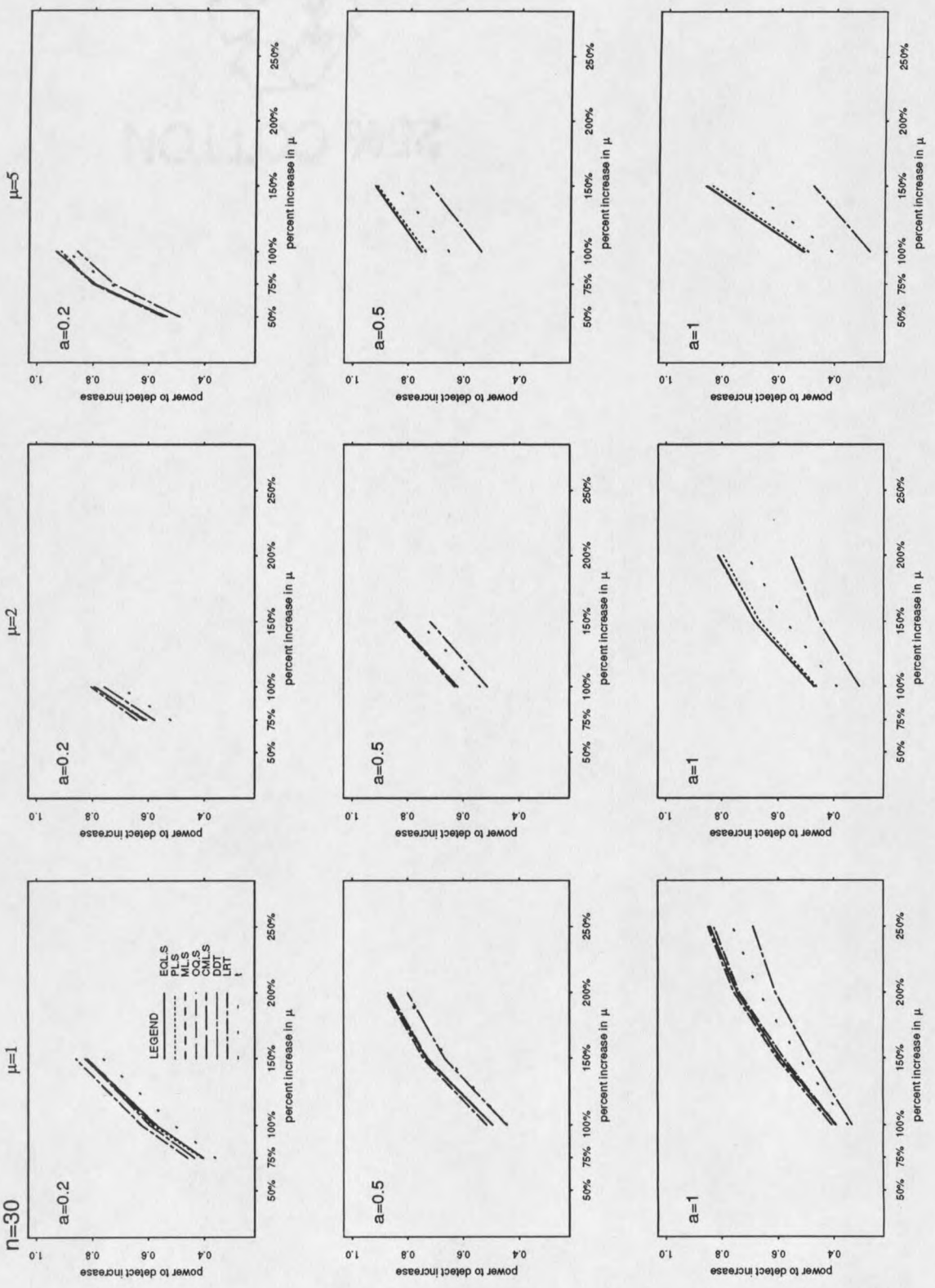

Figure 18. Power plots for all testing methods on samples of size n=30.

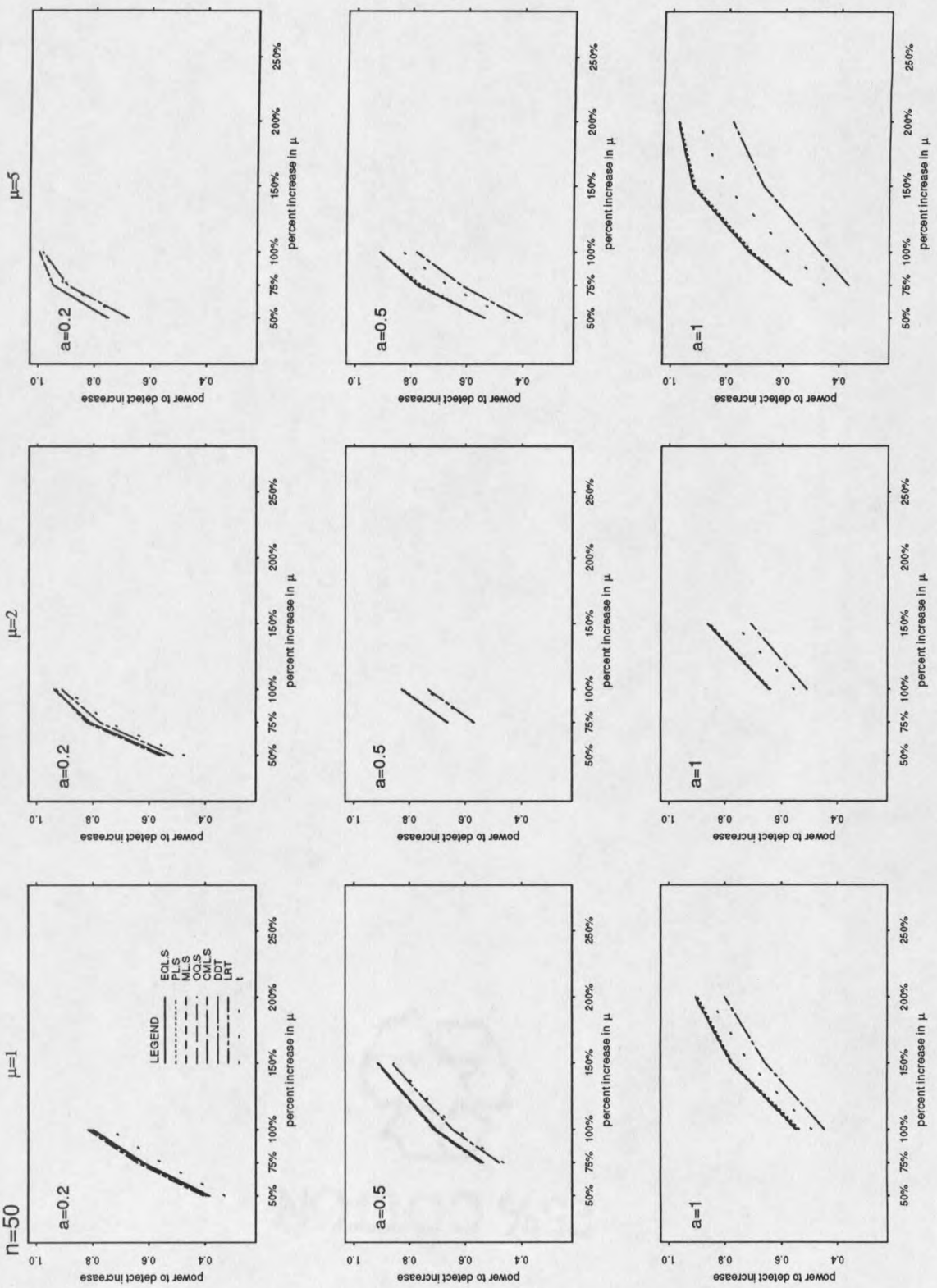

Figure 19. Power plots for all testing methods on samples of size n=50.

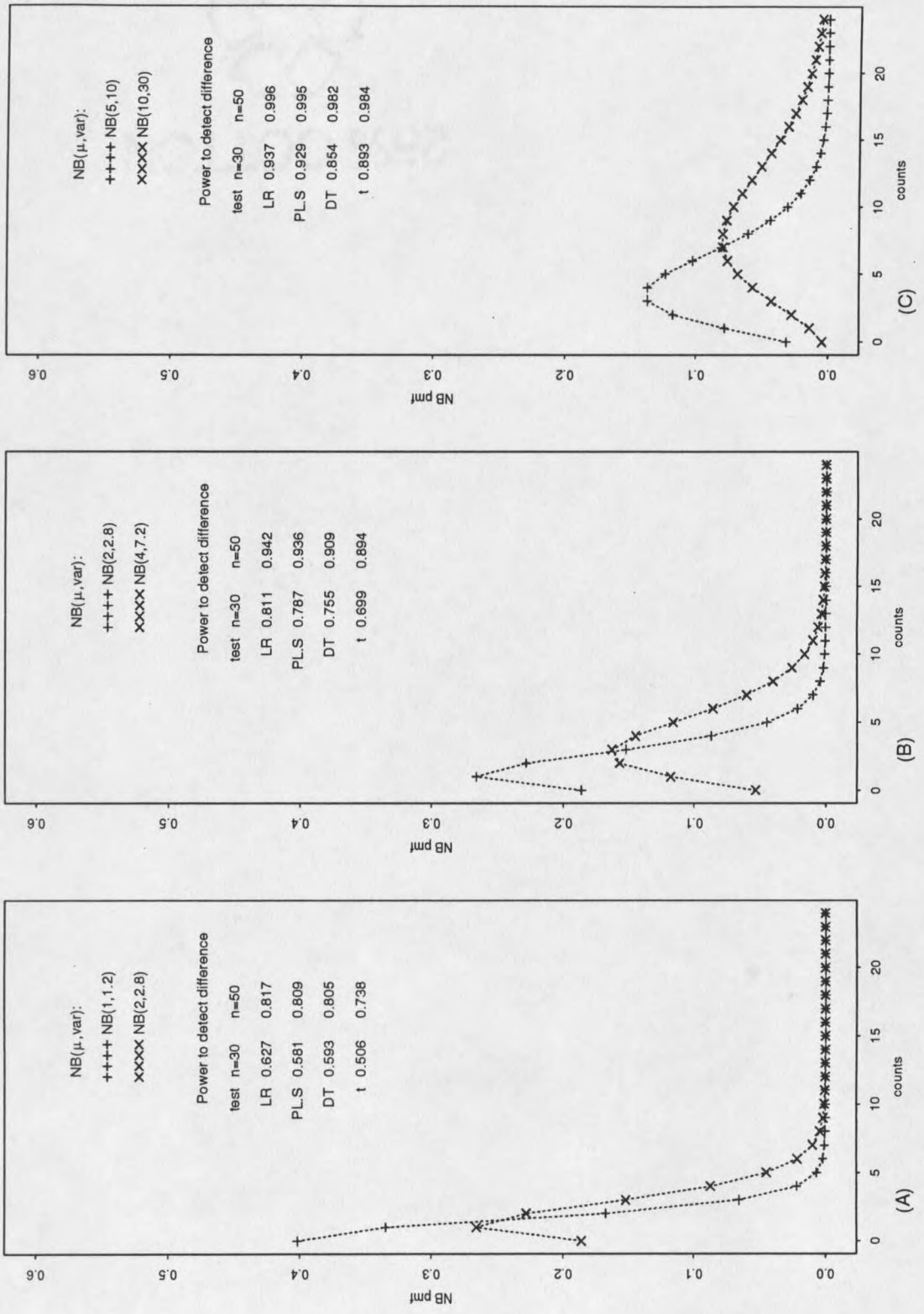

Figure 20. NB pmfs at 100% increase in μ for a=0.2.

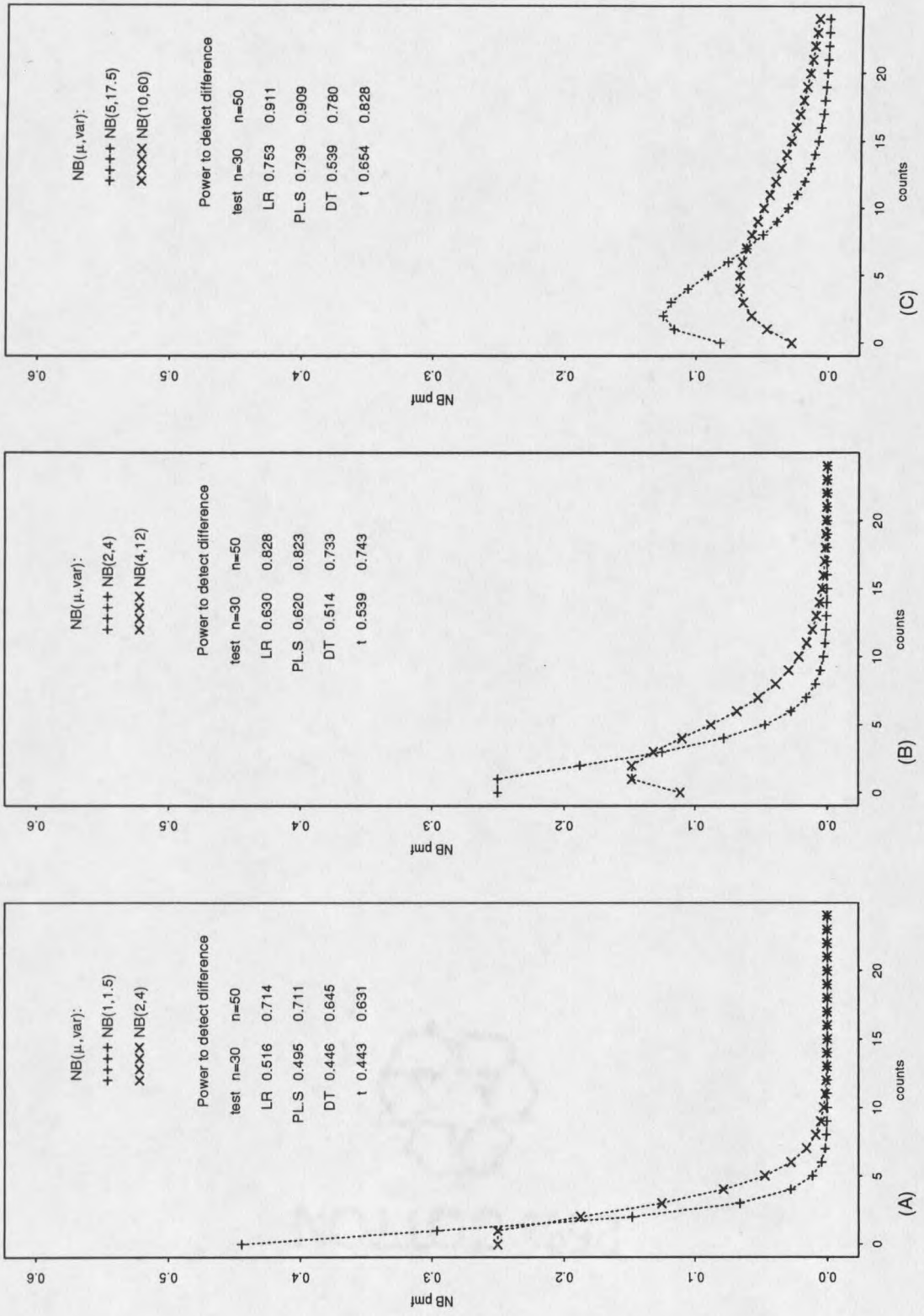

Figure 21. NB pmfs at 100% increase in μ for a=0.5.

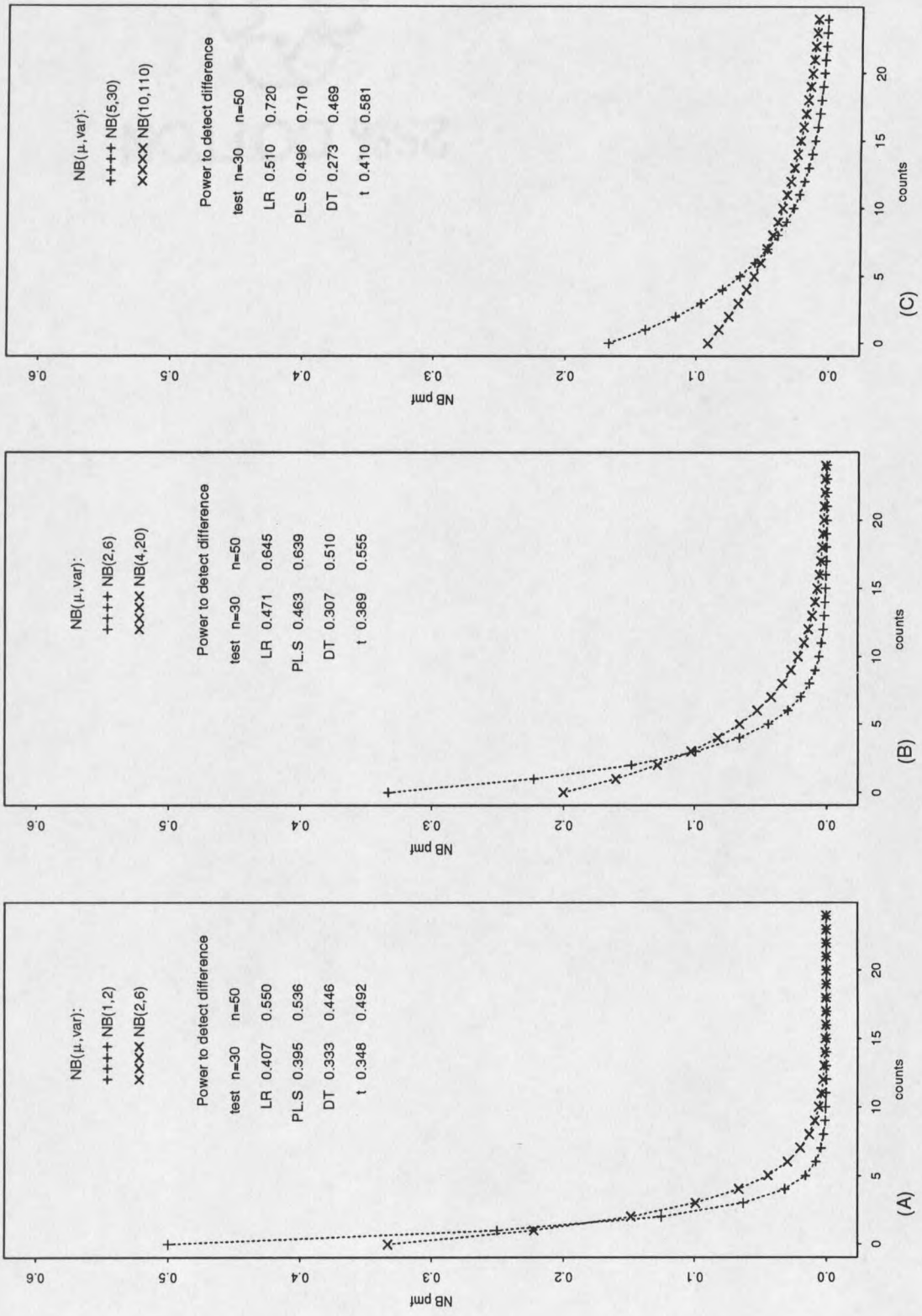

Figure 22. NB pmfs at 100% increase in μ for a=1.0.

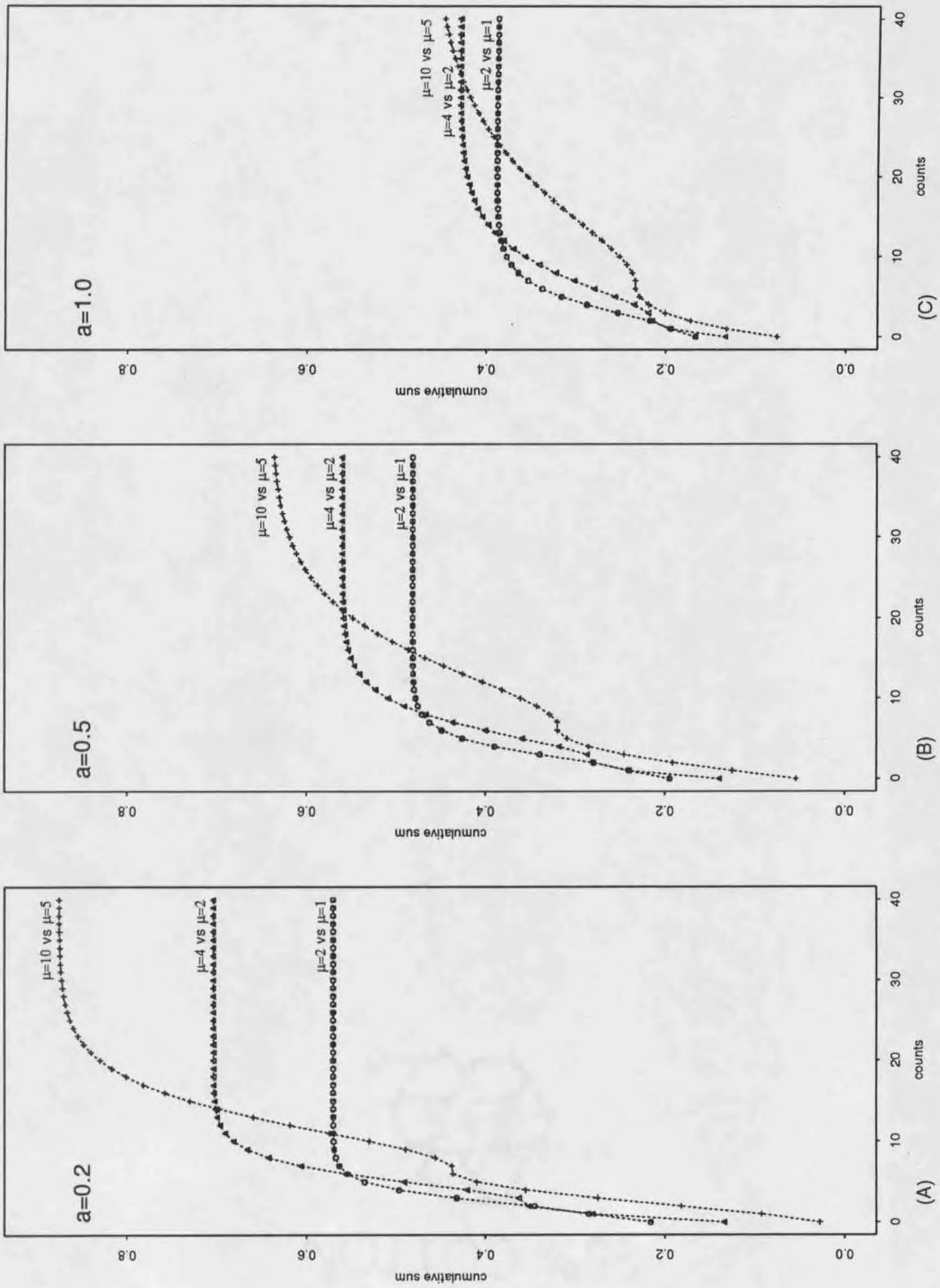

Figure 23. Cumulative sum of absolute differences between NB pmfs at 100% increase in μ.

REFERENCES CITED

Abramowitz, M. and Stegun, I.A.(eds.) (1965), *Handbook of Mathematical Functions*, New York: Dover.

Anraku, K. and Yanagimoto, T. (1990), "Estimation for the Negative Binomial Distribution Based on the Conditional Likelihood," *Communications in Statistics: Simulation*, 19, 771-786.

Anscombe, F.J. (1949), "The Statistical Analysis of Insect Counts Based on the Negative Binomial Distribution," *Biometrics*, 5, 165-173.

Anscombe, F.J. (1950), "Sampling theory of the negative binomial and logarithmic series distributions," *Biometrika*, 37, 358-382.

Aragón, J., Eberly, D. and Eberly, S. (1992), "Existence and uniqueness of the maximum likelihood estimator for the two-parameter negative binomial distribution," *Statistics & Probability Letters*, 15, 375-379.

Arbous, A.G. and Kerrich, J.E. (1951), "Accident Statistics and the Concept of Accident Proneness," *Biometrics*, 7, 340-432.

Ash, R.B. (1972), *Real Analysis and Probability*, New York: Academic Press.

Bain, L.J. and Wright, F.T. (1982), "The Negative Binomial Process with Applications to Reliability," *Journal of Quality Technology*, 14, 60-66.

Barnwal, R.K. and Paul, S.R. (1988), "Analysis of one-way layout of count data with negative binomial variation," *Biometrika*, 75, 215-222.

Basu, A. and Lindsay, B.G. (1994), "Minimum Disparity Estimation for Continuious Models: Efficiency, Distributions and Robustness," *Annals of Institute of Statistical Mathematics*, 46, 683-705.

Basu, A. and Sarkar, S. (1994), "The trade-off between robustness and efficiency and the effect of model smoothing in minimum disparity inference," *Journal of Statistical Computation and Simulation*, 50, 173-185.

Binet, F.E. (1986), "Fitting the Negative Binomial Distribution," *Biometrics*, 42, 989-992.

Bliss, C.I. and Fisher, R.A. (1953), "Fitting the Negative Binomial Distribution to Biological Data," *Biometrics*, 9, 176-200.

Bliss, C.I. and Owen, A.R.G. (1958), "Negative binomial distributions with a common *k*," *Biometrika*, 45, 37-58.

Boos, D.D. (1980), "A New Method For Constucting Approximate Confidence Intervals From M Estimates," *Journal of the American Statisical Association*, 75, 142-145.

Boos, D.D. (1992), "On Generalized Score Tests," *The American Statistician*, 46, 327-333.

Boswell, M.T. and Patil, G.P. (1970), "Chance Mechanisms Generating the Negative Binomial Distributions," in*Random Counts in Models and Structures*, ed. Patil, G.P., University Park: The Pennsylvania State University Press, 3-22.

Bowman, K.O. (1984), "Extended Moment Series and the Parameters of the Negative Binomial Distribution," *Biometrics*, 40, 249-252.

Box, G.E.P. and Cox, D.R. (1964), "An analysis of transformations," *Journal of the Royal Statistical Society, Ser.B*, 26, 211-243.

Box, G.E.P. and Draper, N.R. (1987), *Empirical Model-Building and Response Surfaces*, New York: John Wiley and Sons.

Breslow, N.E. (1984), "Extra-Poisson Variation in Log-Linear Models," *Applied Statistics*, 33, 38-44.

Breslow, N.E. (1989), "Score Tests in Overdispersed GLM's," in *Proceedings of GLIM 89 and the Fourth International Workshop on Statistical Modelling*, eds. Decarli, A., Francis, B.J., Gilchrist, R. and Seeber, G.U.H., New York: Springer-Verlag, 64-74.

Breslow, N.E. (1990), "Tests of Hypotheses in Overdispersed Poisson Regression and Other Quasi-Likelihood Models," *Journal of the American Statisical Association*, 85, 565-571.

Buck, R.C. (1965), *Advanced Calculus*, ed. 2, New York: McGraw-Hill.

Carroll, R.J. and Ruppert, D. (1982), "Robust Estimation in Heteroscedastic Linear Models," *The Annals of Statistics*, 10, 429-441.

Chase, G.R. and Hoel, D.G. (1975), "Serial dilutions: Error effects and optimal designs," *Biometrika*, 62, 329-334.

Chatfield, C. (1975), "A marketing application of a characterization theorem," *Statistical Distributions in Scientific Work*, eds. Patil, G.P., Kotz, S. and Ord, J.K., Dordrecht-Holland: D. Reidel Publishing Company, 2, 175-185

Christian, R.R. and Pipes, W.O. (1983), "Frequency Distribution of Coliforms in Water Distribution Systems," *Applied and Environmental Microbiology*, 45, 603-609.

Clark, S.J. and Perry, J.N. (1989), "Estimation of the Negative Binomial Parameter *k* by Maximum Quasi-Likelihood," *Biometrics*, 45, 309-316.

Collings, B.J. and Margolin, B.H. (1985), "Testing Goodness of Fit for the Poisson Assumption When Observations Are Not Identically Distributed," *Journal of the American Statisical Association*, 80, 411-418.

Cox, D.R. and Hinkley, D.V. (1974), *Theoretical Statistics*, London: Chapman and Hall.

Davidian, M. and Carroll, R.J. (1987), "Variance Function Estimation," *Journal of the American Statisical Association*, 82, 1079-1091.

Davidian, M. and Carroll, R.J. (1988), "A Note on Extended Quasi-likelihood," *Journal of the Royal Statistical Society,Ser.B*, 50, 74-82.

El-Shaarawi, A.H., Esterby,S.R., and Dutka, B.J. (1981), "Bacterial Density in Water Determined by Poisson or Negative Binomial Distributions," *Applied Environmental Microbiology*, 41, 107-116.

Evans, D.A. (1953), "Experimental Evidence Concerning Contagious Distributions In Ecology," *Biometrika*, 40, 186-211.

Fisher, R.A. (1941), "The negative binomial distribution," *Annals of Eugenics*, 11, 182-187.

Fisz, M. (1963), *Probability Theory and Mathematical Statistics*, New York: John Wiley and Sons.

Foutz, R.V. and Srivastava, R.C. (1977), "The Performace of the Likelihood Ratio Test when the Model in Incorrect," *The Annals of Statistics*, 5, 1183-1194.

Godambe, V.P. (1960), "An optimum property of regular maximum likelihood estimation," *Annals of Mathematical Statistics*, 31, 1208-1211.

Godambe, V.P. (1985), "The foundations of finite sample estimation in stochastic processes," *Biometrika*, 72, 419-428.

Godambe, V.P. (1987), "The foundations of finite sample estimation in stochastic processes-II," in*Proceedings First World Congress if Bernoulli Society*, Utrecht: VNU Science Press, 49-54.

Godambe, V.P. and Heyde, C.C. (1987), "Quasi-likelihood and optimal estimation," *International Statistical Review*, 55, 231-244.

Godambe, V.P. and Thompson, M.E. (1984), "Robust estimation through estimating equations," *Biometrika*, 71, 115-125.

Godambe, V.P. and Thompson, M.E. (1985), *Logic of Least Squares-Revisited*, preprint.

Godambe, V.P. and Thompson, M.E. (1989), "An Extension of Quasi-likelihood Estimation (with discussion)," *Journal of Statistical Planning and Inference*, 22, 137-152.

Gold, H.J., Bay, J and Wilkerson, G.G. (1996), "Scouting for weeds, based on the negative binomial distribution," *Weed Science*, 44: 504-510.

Gong, G. and Samaniego, F.J. (1981), "Pseudo Maximum Likelihood Estimation: Theory and Applications," *The Annals of Statistics*, 9, 861-869.

Greenwood, M. and Yule, G.U. (1920), "An Enquiry into the Nature of Frequency Distributions Representative of Multiple Happenings, with Particular Reference to the Occurrence of Multiple Attacks of Disease or of Repeated Accidents," *Journal of the Royal Statistical Society*, 83, 255-279.

Hubbard, D.J. and Allen, O.B. (1991), "Robustness of the SPRT for a Negative Binomial to Misspecification of the Dispersion Parameter," *Biometrics*, 47, 419-427.

Johnson, N.L., Kotz, S. and Kemp, A.W. (1992), "Negative Binomial Distribution," in *Univariate Discrete Distributions*, New York: John Wiley and Sons, 199-235.

Jones, K.S., Herod, D. and Huffman, D. (1991), "Fitting the Negative Binomial Distribution to Parasitological Data," *The Texas Journal of Science*, 43, 357-371.

Jones, P.C.T., Mollison, J.E. and Quenouille, M.H. (1948), "A Technique for the Quantitative Estimation of Soil Micro-Organisms. Statistical note," *Journal of General Microbiology*, 2, 54-69.

Kent, J.T. (1982), "Robust properties of likelihood ratio tests," *Biometrika*, 69, 19-27.

Krewski, D., Leroux, B.G., Bleuer, S.R. and Broekhoven, L.H. (1993), "Modeling the Ames *Salmonella*/Microsome Assay," *Biometrics*, 49, 499-510.

Lawless, J.F. (1987), "Negative binomial and mixed Poisson regression," *The Canadian Journal of Statistics*, 15, 209-225.

Lehmann, E.L. (1991), *Theory of Point Estimation*, Belmont, CA: Wadsworth. [also: Lehmann, E.L. (1983), *Theory of Point Estimation*, New York: John Wiley and Sons.]

Levin, B. and Reeds, J. (1977), "Compound multinomial likelihood functions are unimodal: Proof of a conjecture of I. J. Good," *The Annals of Statistics*, 5, 79-87.

Liang, K. and Zeger, S.L., "Longitudinal data anlysis using generalized linear models," *Biometrika*, 73, 13-22.

Lindsay, B.G. (1994), "Efficiency versus robustness: the case for minimum Hellinger distance and related methods," *The Annals of Statistics*, 22, 1081-1114.

Luders, R. (1934), "Die statistic der seltenen ereignisse," *Biometrika*, 26, 108-128.

Manton, K.G, Woodbury, M.A. and Stallard, E. (1981), "A Variance Components Approach to Categorical Data Models with Heterogeneous Cell Populations: Analysis of Spatial Gradients in Lung Cancer Mortality Rates in North Carolina Counties," *Biometrics*, 37, 259-269.

Margolin, B.H., Kaplan, N. and Zeiger, E. (1981), "Statistical analysis of the Ames *Salmonella*/microsome test," *Proceedings of the National Academy of Science USA*, 78, 3779-3783.

Martin, D.C. and Katti, S.K. (1965), "Fitting of Certain Contagious Distributions to some available data by the Maximum Likelihood Method," *Biometrics*, 21, 34-41.

Maul, A. and El-Shaarawi, A.H. (1991), "Analysis of Two-way Layout of Count Data with Negative Binomial Variation," *Environmental Monitoring and Assessment*, 17, 237-244.

Maul, A., El-Shaarawi, A.H. and Block, J.C. (1985), "Heterotrophic Bacteria in Water Distribution Systems. I. Spatial and Temporal Variation," *The Science of the Total Environment*, 44, 201-214.

Maul, A., El-Shaarawi, A.H. and Ferard, J.F. (1991), "Application of Negative Binomial Regression Models to the Analysis of Quantal Bioassays Data," *Environmetrics*, 2, 253-261.

McCullagh, P. and Cox, D.R. (1986), "Invariants and Likelihood Ratio Statistics," *The Annals of Statistics*, 14, 1419-1430.

McCullagh, P. and Nelder, J.A. (1983), *Generalized Linear Models*, London: Chapman and Hall.

McCullagh, P. and Nelder, J.A. (1989), *Generalized Linear Models*, ed. 2, London: Chapman and Hall.

Miller, R.G.,Jr. (1966), *Simultaneous Statistical Inference*, New York: McGraw-Hill.

Miller, R.G.,Jr. (1986), *Beyond Anova, Basics of Applied Statistics*, New York: John Wiley and Sons.

Minkin, S. (1991), "A Statistical Model for In-Vitro Assessment of Patient Sensitivity to Cytotoxic Drugs," *Biometrics*, 47, 1581-1591.

Montmort, P.R. (1714), "Essai d'analyse sur les jeux de hasards," *Paris*. [as cited in Tripathi, 1982]

Morton, R. (1987), "A generalized linear model with nested strata of extra-Poisson variation," *Biometrika*, 74, 247-257.

Neyman, J. (1959), "Optimal asymtotic tests for composite hypotheses," in *Probability and Statistics*, ed. Grenander,U., New York: Wiley, 213-234.

Nelder, J.A. (1992), "Pseudolikelihood and Quasi-likelihood," *Applied Statistics*, 41, 595-600.

Nelder, J.A. and Lee, Y. (1992), "Likelihood, Quasi-likelihood and Pseudolikelihood: Some Comparisons," *Journal of the Royal Statistical Society,Ser.B*, 54, 273-284.

Nelder, J.A. and Pregibon, D. (1987), "An extended quasi-likelihood function," *Biometrika*, 74, 221-232.

Niemelä, S. (1983), *Statistical Evaluation of Results from Quantitative Microbiological Examinations*, ed. 2, Uppsala: Nordic Committee on Food Analysis.

Pascal, B. (1679), *Varia Opera Mathematica*, Tolossae: D. Pettri de Fermet. [as cited in Johnson, Kotz and Kemp, 1992]

Piegorsch, W.W. (1990), "Maximum Likelihood Estimation for the Negative Binomial Dispersion Parameter," *Biometrics*, 46, 863-867.

Pieters, E.P., Gates, C.E., Matis, J.H. and Sterling, W.L. (1977), "Small Sample Comparison of Different Estimators of Negative Binomial Parameters," *Biometrics*, 33, 718-723.

Pipes, W.O. and Christian, R.R. (1982), "Sampling frequency - microbiological drinking water regulations - final report," U.S.E.P.A., EPA R-805-637/9-82-001, Washinton DC.

Pipes, W.O., Ward, P. and Ahn, S.H. (1977), "Frequency Distributions for Coliform Bacteria in Water," *Journal American Water Works Association*, 69, 664-668.

Pregibon, D. (1982), "Score Tests in GLIM with Applications," in *GLIM 82: Proceedings of the International Conference on Generalised Linear Models*, ed. Gilchrist, R., New York: Springer-Verlag, 87-97.

Quenouille, M.H. (1949), "A Relation Between the Logarithmic, Poisson, and Negative Binomial Series," *Biometrics*, 5, 162-164.

Ramakrishnan, V. and Meeter, D. (1993), "Negative Binomial Cross-Tabulations, with Applications to Abundance Data," *Biometrics*, 49, 195-207.

Ramaswamy, V., Anderson, E.W. and DeSarbo, W.S. (1994), "A Disaggregate Negative Binomial Regression Procedure for Count Data Analysis," *Management Science*, 40, 405-417.

Rao, C.R. (1948), "Large Sample Tests of Statistical Hypotheses Concerning Several Parameters With Applications to Problems of Estimation," *Proceedings of the Cambridge Philosophical Society*, 44, 50-57.

Ross, S.M. (1989), *Introduction to Probability Models*, ed. 4, San Diego: Academic Press.

Ross, G.J.S. and Preece, D.A. (1985), "The negative binomial distribution," *The Statistician*, 34, 323-336.

Scheffé, H. (1959), *The Analysis of Variance*, New York: Wiley.

Schmittlein, D.C, Bemmaor, A.C. and Morrison, D.G. (1985), "Why does the NBD Model Work? Robustness of Representing Product Purchases, Brand Purchases and Imperfectly Recorded Purchases," *Marketing Science*, 4, 255-266.

Searle, S.R., Casella, G. and McCulloch, C.E. (1992), *Variance Components*, New York: Wiley.

Seber, G.A.F. (1973), *The Estimation of Animal Abundance*, New York: Hafner Press.

Seber, G.A.F. (1984), *Multivariate Observations*, New York: Wiley.

Snedecor, G.W. and Cochran, W.G. (1980), *Statistical Methods*, ed. 7, Ames, Iowa: Iowa State University Press.

Stein, Gillian Z. (1988), "Modelling Counts in Biological Populations," *Mathematical Scientist*, 13, 56-65.

Student, (1907), "On the error of counting with a haemocytometer," *Biometrika*, 5, 351-360.

Taylor, L.R., Woiwod, I.P. and Perry, J.N. (1978), "The density-dependence of spatial behavior and the rarity of randomness," *Journal of Animal Ecology*, 47, 383-406.

Tripathi, R.C. (1982), "Negative Binomial Distribution," in *Encyclopedia of Statistical Sciences*, eds. Kotz, S., Johnson, N.L. and Read, C.B., New York: Wiley, 169-177.

Van de Ven, R. (1993), "Estimating the Shape Parameter for the Negative Binomial Distribution," *Journal of Statistical Computation and Simulation*, 46, 111-123.

Venables, B. (1994), *Functions for fitting and analysing Negative Binomial GLMs*, email: venables@stats.adelaide.edu.au.

Walsh, G.R. (1975), "Gradient Methods for Unconstrained Optimization," in *Methods of Optimization*, London: John Wiley and Sons, 105-142.

Wedderburn, R.W.M. (1974), "Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method," *Biometrika*, 61, 439-447.

Willson, L.J., Folks, J.L. and Young, J.H. (1984), "Multistage Estimation Compared with Fixed-Sample-Size Estimation of the Negative Binomial Parameter *k*," *Biometrics*, 40, 109-117.

Willson, L.J., Folks, J.L. and Young, J.H. (1986), "Complete Sufficiency and Maximum Likelihood Estimation for the Two-Parameter Negative Binomial Distribution," *Metrika*, 33, 349-362.

Zeger, S.L., Liang, K. and Albert, P.S. (1988), "Models for Longitudinal Data: A Generalized Estimating Equation Approach," *Biometrics*, 44, 1049-1060.