Efficient Estimation of Parameters of the Negative Binomial Distribution

V. SAVANI AND A. A. ZHIGLJAVSKY

Department of Mathematics, Cardiff University, Cardiff, CF24 4AG, U.K.

e-mail: SavaniV@cardiff.ac.uk, ZhigljavskyAA@cardiff.ac.uk (Corresponding author)

Abstract

In this paper we investigate a class of moment based estimators, called power method estimators, which can be almost as efficient as maximum likelihood estimators and achieve a lower asymptotic variance than the standard zero term method and method of moments estimators. We investigate different methods of implementing the power method in practice and examine the robustness and efficiency of the power method estimators.

Key Words: Negative binomial distribution; estimating parameters; maximum likelihood method; efficiency of estimators; method of moments.

1. The Negative Binomial Distribution

1.1. Introduction

The negative binomial distribution (NBD) has appeal in the modelling of many practical applications. A large amount of literature exists, for example, on using the NBD to model: animal populations (see e.g. Anscombe (1949), Kendall (1948a)); accident proneness (see e.g. Greenwood and Yule (1920), Arbous and Kerrich (1951)) and consumer buying behaviour (see e.g. Ehrenberg (1988)). The appeal of the NBD lies in the fact that it is a simple two parameter distribution that arises in various different ways (see e.g. Anscombe (1950), Johnson, Kotz, and Kemp (1992), Chapter 5) often allowing the parameters to have a natural interpretation (see Section 1.2). Furthermore, the NBD can be implemented as a distribution within stationary processes (see e.g. Anscombe (1950), Kendall (1948b)) thereby increasing the modelling potential of the distribution. The NBD model has been extended to the process setting in Lundberg (1964) and Grandell (1997) with applications to accident proneness, sickness and insurance in mind. NBD processes have also been considered in Barndorff-Nielsen and Yeo (1969) and Bondesson (1992), Chapter 2.

This paper concentrates on fitting the NBD model to consumer purchase occasions, which is usually a primary variable for analysis in market research. The results of this paper, however, are general and can be applied to any source of data. The mixed Poisson process with Gamma structure distribution, which we call the Gamma-Poisson process, is a stationary process that assumes that events in any time interval of any length are NBD. Ehrenberg (1988) has shown that the Gamma-Poisson process provides a simple model for describing consumer purchase occasions within a market and can be used to obtain market forecasting measures. Detecting changes in stationarity and obtaining accurate forecasting measures require efficient estimation of parameters and this is the primary aim of this paper.

1.2. Parameterizations

The NBD is a two parameter distribution which can be defined by its mean m ($m > 0$) and shape parameter k ($k > 0$). The probabilities of the $\text{NBD}(m, k)$ are given by

$$p_x = \mathbb{P}(X = x) = \frac{\Gamma(k + x)}{x!\Gamma(k)}\left(1 + \frac{m}{k}\right)^{-k}\left(\frac{m}{m + k}\right)^x \quad x = 0, 1, 2, \ldots \quad . \tag{1}$$

The parameter pair (m, k) is statistically preferred to many other parameterizations as the maximum likelihood estimators $(\hat{m}_{ML}, \hat{k}_{ML})$ and all natural moment based estimators $(\hat{m}, \hat{k})$ are asymptotically uncorrelated, so that $\lim_{N \to \infty} \text{Cov}(\hat{m}, \hat{k}) = 0$, given an independent and identically distributed (i.i.d.) NBD sample (see Anscombe (1950)).

The simplest derivation of the NBD is obtained by considering a sequence of i.i.d. Bernoulli random variables, each with a probability of success p. The probability of requiring x ($x \in \{0, 1, 2, \ldots\}$) failures before observing k ($k \in \{0, 1, 2, \ldots\}$) successes is then given by (1), with $p = k/(m + k)$. The NBD with parameterization (p, k) and k restricted to the positive integers is sometimes called the Pòlya distribution.

Alternatively, the NBD may be derived by 'heterogenous Poisson sampling' (see e.g. Anscombe (1950), Grandell (1997)) and it is this derivation that is commonly used in market research. Suppose that the purchase of an item in a fixed time interval is Poisson distributed with unknown mean λ_j for consumer j. Assume that these means λ_j, within the population of individuals, follow the $\text{Gamma}(a, k)$ distribution with probability density function given by

$$p(y) = \frac{1}{a^k \Gamma(k)} y^{k-1} \exp(-y/a), \quad a > 0, \ k > 0, \quad y \geqslant 0.$$

Then it is straightforward to show that the number of purchases made by a random individual has a NBD distribution with probability given by (1) and $a = m/k$. It is this derivation of the NBD that is the basis of the Gamma-Poisson process.

A very common parametrization of the NBD in market research is denoted by (b, w). Here $b = 1 - p_0$ represents the probability for a random individual to make at least one purchase (b is often termed the penetration of the brand) and $w = m/b$ ($w \geqslant 1$) denotes the mean 'purchase frequency' per buyer.

3

In this paper we include a slightly different parametrization, namely (b, w') with $w' = 1/w$. Its appeal lies in the fact that the corresponding parameter space is within the unit square $(b, w') \in [0, 1]^2$, which makes it easier to make a visual comparison of different estimators for all parameter values of the NBD. Fig. 1 shows the contour levels of m, p and k within the (b, w')-parameter space. Note that the relationship $p = 1/(1 + a)$ is independent of b and w' so that the contour levels for a have the same shape as the contour levels of p.

The NBD is only defined for the parameter pairs $(b, w') \in (0, 1) \times (0, 1)$ such that $w' < -b/\log(1 - b)$ (shaded region in Fig. 1). The relationship $w' = -b/\log(1 - b)$ represents the limiting case of the distribution as $k \to \infty$, when the NBD converges to the Poisson distribution with mean m. The NBD is not defined on the axis $w' = 0$ (where $m = \infty$) and is degenerate on the axis $b = 0$ (as $p_0 = 1$).
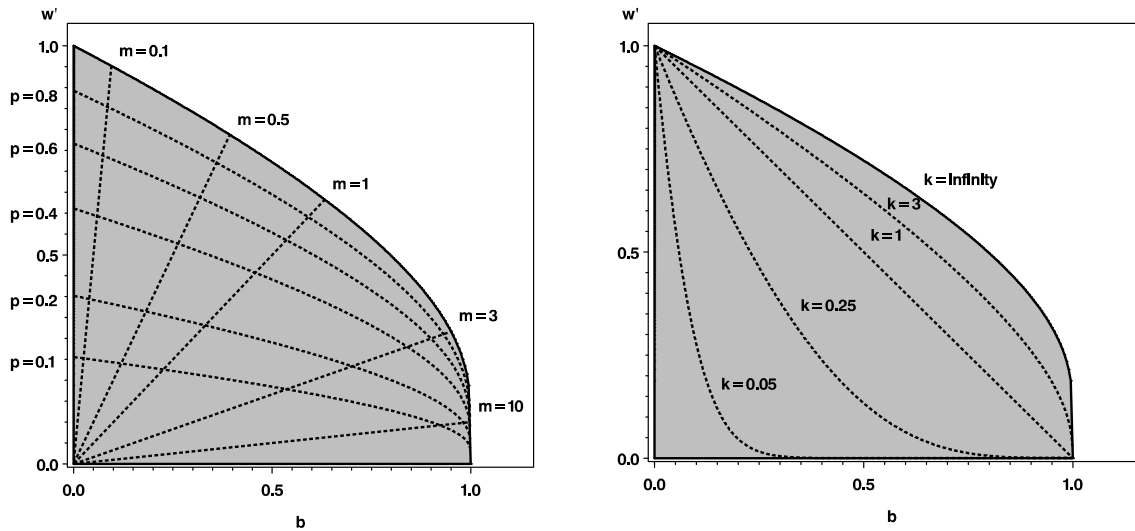

Figure 1: m, p and k versus (b, w').

2. Maximum Likelihood Estimation

Fisher (1941) and Haldane (1941) independently investigated parameter estimation for the NBD parameter pair (m, k) using maximum likelihood (ML). The ML estimator for m is simply given by the sample mean $\bar{x}$, however there is no closed form solution for $\hat{k}_{ML}$ (the

ML estimator of k) and the estimator $\hat{k}_{ML}$ is defined as the solution, in z, to the equation

$$\log\left(1 + \frac{\bar{x}}{z}\right) = \sum_{i=1}^{\infty} \frac{n_i}{N} \sum_{j=0}^{i-1} \frac{1}{z+j} \ . \tag{2}$$

Here N denotes the sample size and n_i denotes the observed frequency of $i = 0, 1, 2, \ldots$ within the sample.

The variances of the ML estimators are the minimum possible asymptotic variances attainable in the class of all asymptotically normal estimators and therefore provide a lower bound for the moment based estimators. Fisher (1941) and Haldane (1941) independently derived expressions for the asymptotic variance of $\hat{k}_{ML}$, which is given by

$$v_{ML} = \lim_{N \to \infty} N \, Var\left(\hat{k}_{ML}\right) = \frac{2k(k+1)(a+1)^2}{a^2 \left(1 + 2 \sum_{j=2}^{\infty} \left(\frac{a}{a+1}\right)^{j-1} \frac{j!\Gamma(k+2)}{(j+1)\Gamma(k+j+1)}\right)} \ . \tag{3}$$

Here $a = 1 + m/k$ is the scale parameter of the gamma and NBD distributions as described in Section 1.2. Fig. 2 shows the contour levels of v_{ML} and the asymptotic coefficient of variation $\sqrt{v_{ML}}/k$. The relative precision of $\hat{k}_{ML}$ decreases rapidly as one approaches the boundary $w' = -b/\log(1-b)$ or the axis $b = 0$.
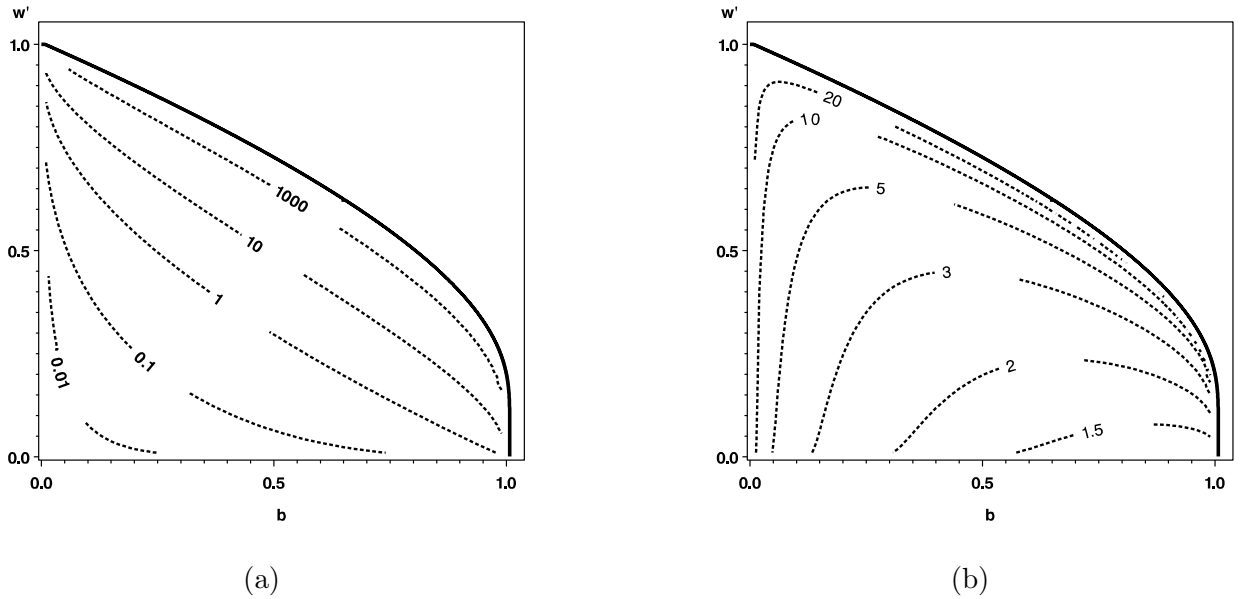

Figure 2: Contour levels of (a) v_{ML} and (b) $\sqrt{v_{ML}}/k$.

Computing the ML estimator for k may be difficult; additionally, ML estimators are not robust with respect to violations of say the i.i.d. assumption. As an alternative, it is conventional to use moment based estimators for estimating k. These estimators were suggested and thoroughly studied in Anscombe (1950).

3. GENERALIZED MOMENT BASED ESTIMATORS

In this section we describe a class of moment based estimation methods that can be used to estimate the parameter pair (m, k) from an i.i.d. NBD sample. Within this class of moment based estimators we show that it is theoretically possible to obtain estimators that achieve efficiency very close to that of ML estimators. In Section 4 we show how the parameter pair (m, k) can be efficiently estimated in practice.

Note that if the NBD model is assumed to be true, then in practice it suffices to estimate just one of the parameter pairs (m, k), (a, k), (p, k) or (b, w) since they all have a one-to-one relationship. Since all moment based estimators $(\hat{m}, \hat{k})$, with $\hat{m} = \bar{x}$, are asymptotically uncorrelated given an i.i.d. sample, estimation in literature has mainly focused on estimating the shape parameter k.

In our forthcoming paper we shall consider the problem of estimating parameters from a stationary NBD autoregressive (namely, INAR (1)) process. For such samples, it is very difficult to compute the ML estimator for k. Moment based estimators for $(\hat{m}, \hat{k})$ are much simpler than the ML estimators, although the estimators are no longer asymptotically uncorrelated. This provides additional motivation for considering moment based estimators.

3.1. METHODS OF ESTIMATION

The estimation of the parameters (m, k) requires the choice of two sample moments. A natural choice for the first moment is the sample mean $\bar{x}$ which is both an efficient and an unbiased estimator for m. An additional moment is then required to estimate the shape parameter k. Let us denote this moment by $\bar{f} = \frac{1}{N} \sum_{i=1}^{N} f(x_i)$, where N denotes the sample

6

size. The estimator for k is obtained by equating the sample moment $\bar{f}$ to its expected value, $Ef(X)$, and solving the corresponding equation $\bar{f} = Ef(X)$, with m replaced by $\hat{m} = \bar{x}$, for k. Anscombe (1950) has proved that if $f(x)$ is any integrable convex or concave function on the non-negative integers then this equation for k will have at most one solution.

In the present paper we consider three functions

$$\overline{f_j} = \frac{1}{N} \sum_{i=1}^{N} f_j(x_i)$$

for the estimation of k, with $f_1(x) = x^2$, $f_2(x) = I_{[x=0]}$ (where $f_2(x) = 1$ if $x = 0$ and $f_2(x) = 0$ if $x \neq 0$) and $f_3(x) = c^x$ with $c > 0$ ($c \neq 1$). The sample moments are

$$\overline{f_1} = \overline{x^2} = \frac{1}{N} \sum_{i=1}^{N} x_i^2, \qquad \overline{f_2} = \widehat{p_0} = \frac{n_0}{N}, \qquad \text{and} \qquad \overline{f_3} = \widehat{c^X} = \frac{1}{N} \sum_{i=1}^{N} c^{x_i}, \tag{4}$$

with $Ef_j(X)$ given by

$$Ef_1(X) = ak(a+1+ka), \qquad Ef_2(X) = (a+1)^{-k} \qquad \text{and} \qquad Ef_3(X) = (1+a-ac)^{-k}. \tag{5}$$

Note that $\overline{f_3}$ is the empirical form of the factorial (or probability) generating function $E[c^X]$. Solving $\overline{f_j} = Ef_j(X)$ for $j = 1, 2, 3$ we obtain the method of moments (MOM), zero term method (ZTM) and power method (PM) estimators of k respectively with estimators denoted by $\hat{k}_{MOM}$, $\hat{k}_{ZTM}$ and $\hat{k}_{PM(c)}$. No analytical solution exists for the PM and ZTM estimators and the estimator of k must be obtained numerically. Defining $s^2 = \overline{x^2} - \bar{x}^2$, the MOM estimator is

$$\hat{k}_{MOM} = \frac{\bar{x}^2}{s^2 - \bar{x}}. \tag{6}$$

The PM estimator $\hat{k}_{PM(c)}$ with $c = 0$ is equivalent to the ZTM estimator and as $c \to 1$ the PM estimator becomes the MOM estimator. When comparing the asymptotic distribution and the asymptotic efficiency of different estimators of k it is therefore sufficient to consider only the PM method with $c \in [0, 1]$, where the PM at $c = 0$ and $c = 1$ is defined to be the ZTM and MOM methods of estimation respectively.

3.2. Asymptotic Properties of Estimators

The asymptotic distribution of the estimators can be derived by using a multivariate version of the so-called δ-method (see e.g. Serfling (1980), Chapter 3). It is well known that the joint distribution of $\bar{x}$ and the sample moments $\overline{f_j} = \frac{1}{N}\sum_{i=1}^{N} f(x_i)$, for $j = 1, 2, 3$, follows a bivariate normal distribution as $N \to \infty$. Therefore, the estimators of (m, k), (a, k), (p, k) and (b, w) also follow a bivariate normal distribution as $N \to \infty$ (see Appendix). Note that when k is relatively large compared to m, the finite-sample distributions of the estimators are skewed and convergence to the asymptotic normal distribution is slow.

The asymptotic variances of the sample moments are given by $\mathrm{Var}(\bar{x}) = E\bar{x}^2 - (E\bar{x})^2$ and $\mathrm{Var}(\overline{f_j}) = E\overline{f_j}^2 - (E\overline{f_j})^2$, for $j = 1, 2, 3$. Each of the estimators $(\hat{m}, \hat{k}_{MOM})$, $(\hat{m}, \hat{k}_{ZTM})$ and $(\hat{m}, \hat{k}_{PM})$, with $\hat{m} = \bar{x}$, asymptotically follow a bivariate normal distribution with mean vector (m, k), $\lim_{N\to\infty} N\mathrm{Var}(\hat{m}) = ka(a+1)$ and $\lim_{N\to\infty} N\mathrm{Cov}(\hat{m}, \hat{k}) = 0$. For the PM we have

$$\lim_{N\to\infty} N\mathrm{Var}(\hat{k}_{PM(c)}) \;=\; v_{PM}(c) = \frac{(1+a-ac^2)^{-k}r^{2k+2} - r^2 - ka(a+1)(1-c)^2}{[r\log(r) - r + 1]^2} \qquad (7)$$

where $r = 1+a-ac$. The normalized variance of $\hat{k}$ for the MOM and ZTM estimators are obtained by taking $\lim_{c\to 1} v_{PM}(c)$ and $\lim_{c\to 0} v_{PM}(c)$ respectively, to give

$$\begin{aligned}
v_{MOM} &= \lim_{N\to\infty} N\mathrm{Var}\left(\hat{k}_{MOM}\right) = \frac{2k(k+1)(a+1)^2}{a^2} \\
v_{ZTM} &= \lim_{N\to\infty} N\mathrm{Var}\left(\hat{k}_{ZTM}\right) = \frac{(a+1)^{k+2} - (a+1)^2 - ka(a+1)}{[(a+1)\log(a+1) - a]^2}.
\end{aligned} \qquad (8)$$

The MOM and ZTM estimators are asymptotically efficient in the following sense

$$v_{MOM}/v_{ML} \to 1 \text{ as } k \to \infty, \qquad v_{ZTM}/v_{ML} \to 1 \text{ as } k \to 0. \qquad (9)$$

In Savani and Zhigljavsky (2005), however, we prove that for any fixed $m > 0$ and k ($0 < k < \infty$) the PM can be more efficient than the MOM or ZTM. In fact, there always exists $\tilde{c}$, with $0 < \tilde{c} < 1$, such that $v_{PM}(\tilde{c}) < \min\{v_{ZTM}, v_{MOM}\}$. We let c_* denote the value of c that minimizes $v_{PM}(c)$.

Fig. 3 (a) shows the asymptotic normalized variances of the estimators for k when $m = 1$ and $k = 1$ using the ML, MOM, ZTM and PM with $c \in [0, 1]$. Note that $v_{PM}(c)$ is a
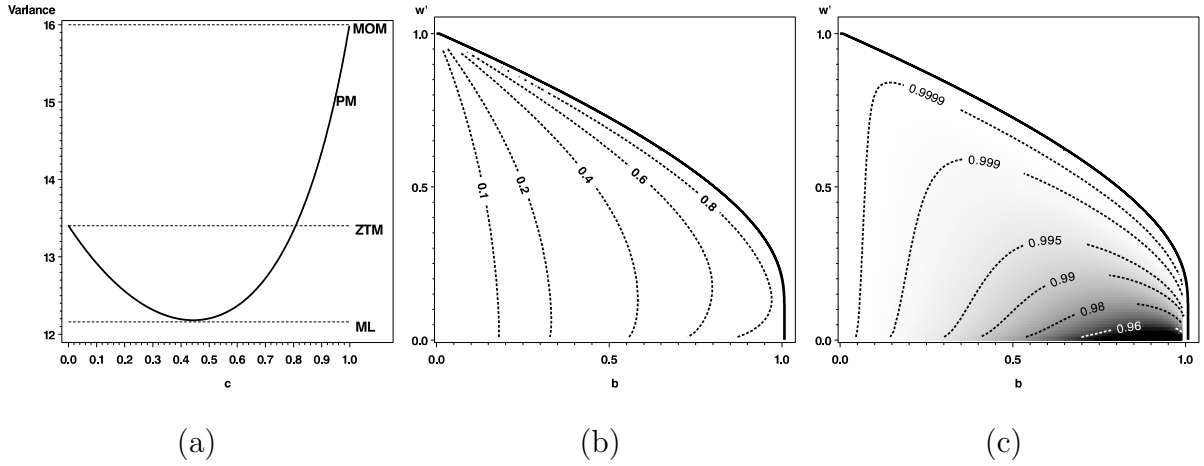
Figure 3: (a) $v_{PM}(c)$ versus c for $m = 1$ and $k = 1$ (ML, MOM, ZTM and PM). (b) Contour levels of c_*. (c) Efficiency of PM at c_*: $v_{ML}/v_{PM}(c_*)$.

continuous function in c and only has one minimum for $c \in [0, 1]$. This is true in general for all parameter values (b, w').

It is difficult to express c_* analytically since the solution (with respect to c) of the equation $\partial v_{PM}(c)/\partial c = 0$ is intractable. The problem can, however, be solved numerically and Fig. 3 (b) shows the contour levels of c_* within the NBD parameter space. The value of c_* is actually a continuous function of the parameters (b, w').

Fig. 3 (c) shows the asymptotic normalized efficiency of the PM at c_* with respect to ML. The PM at c_* is almost as efficient as ML although an efficiency of 1 is only achieved in the limit as $c \to 1$, $k \to \infty$ and as $c \to 0$, $k \to 0$, see (9). A comparison with Fig. 2 (b) notably reveals that areas of high efficiency with respect to ML relate to areas of a high coefficient of variation ($\sqrt{v_{ML}}/k$) for k and vice versa.

4. Implementing the Power Method

In this section we consider the problems of economically implementing the PM in practice to obtain robust and efficient estimators for the NBD parameters. The first problem we investigate is that of obtaining simple approximations for the optimum value c_* that minimizes the asymptotic normalized variance $v_{PM}(c)$ with respect to c, $0 \le c \le 1$.

In practice the computation or approximation of c_* by minimizing $v_{PM}(c; m, k)$ requires knowledge of the NBD parameters (m, k) and these parameters are unknown in practice. The value of c_* must therefore be estimated from data at hand by using preliminary, possibly inefficient, estimators $(\tilde{m}, \tilde{k})$ and minimizing $v_{PM}(c; \tilde{m}, \tilde{k})$ with respect to c. The estimated value for c_* may then be used to obtain 'pseudo' efficient estimators for the NBD parameters (m, k).

### 4.1. Approximating Optimum c - Method 1

The simplest and most economical method of implementing the PM is to apply a natural extension of the idea implied by Anscombe (1950), where the most efficient estimator amongst the MOM and ZTM estimators of k is chosen. In essence, the value of c_* is approximated by $c \in \{0, 1\}$ such that $v_{PM}(c) = \min \{v_{PM}(0), v_{PM}(1)\}$.

Alternatively, one can choose between the more efficient among a set of estimators $(\hat{m}, \hat{k}_{PM(c)})$, obtained by using the PM at c for a given set of values of $c \in [0, 1]$. If we denote this set of values of c by A, then the value of c_* is approximated by c_A such that $c_A = \mathrm{argmin}_{c \in A} v_{PM}(c; m, k)$ and an efficient estimator for (m, k) is obtained by using the PM at c_A.

Fig. 4 shows the asymptotic efficiency of the PM at $c_A = \mathrm{argmin}_{c \in A} v_{PM}(c; m, k)$, relative to the PM at c_* for two different sets of A. Fig. 4 (a) confirms that in all regions of the parameter space the MOM/ZTM estimator is less efficient than the PM estimator at c_*, except in the limits as $k \to 0$ and $k \to \infty$ where the efficiency is the same. Fig. 4 (b) illustrates the effect of extending the set A; it is clear that the asymptotic efficiency of the PM using c_A with $A = \{0, \frac{1}{5}, \frac{2}{5}, \frac{3}{5}, \frac{4}{5}, 1\}$ is more efficient than using just the MOM and ZTM estimators.

4.2. Approximating Optimum c - Method 2

We note that the value of c_* can be obtained by numerical minimization of $v_{PM}(c; b, w')$. Using knowledge of these numerical values over the whole parameter space, regression tech-

10
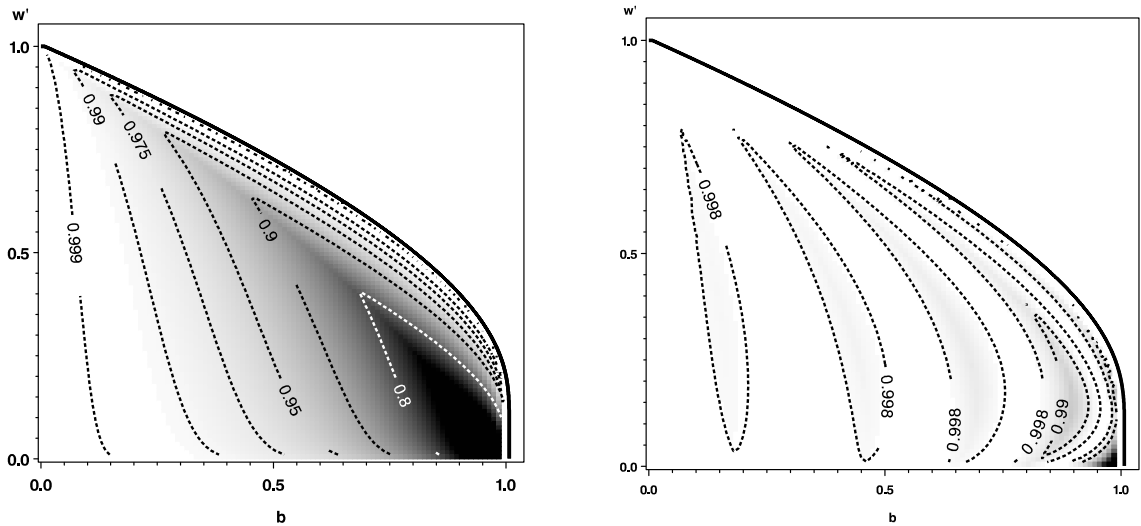
(a) MOM/ZTM ($A = \{0, 1\}$) (b) $A = \{0, \frac{1}{5}, \frac{2}{5}, \frac{3}{5}, \frac{4}{5}, 1\}$

Figure 4: Asymptotic efficiency $v_{PM}(c_*)/v_{PM}(c_A)$ with $c_A = \operatorname{argmin}_{c \in A} v_{PM}(c; m, k)$.

niques can be used to obtain an approximation for c_*. We use $v_{PM}(c)$ as a function of the parameter pair (b, w') as these parameters provide the simplest approximations.

Fig. 5 shows the asymptotic efficiency of two such regression methods relative to the PM at c_*. The first method (a) is a simple approximation independent of w' and the second method (b) optimizes the regression for values of $w' < 0.65$. The two methods are highly efficient with respect to the PM at c_*.

4.3. Estimating Optimum c

Both methods of approximating c_* described above require the NBD parameters (m, k) or (b, w'). Since these parameters are unknown in practice, the approximated value for c_* must be estimated from the data. The estimator for this approximated value may be obtained by plugging in preliminary, possibly inefficient, estimators of the NBD parameters for the values of (m, k) or (b, w'). The simplest preliminary estimates for the NBD parameters are either the MOM or ZTM estimates, although any estimator obtained using the PM at any given $c \in [0, 1]$ may be used. Let us denote the preliminary estimators for the NBD parameters by $(\tilde{m}, \tilde{k}_{PM(c)})$.
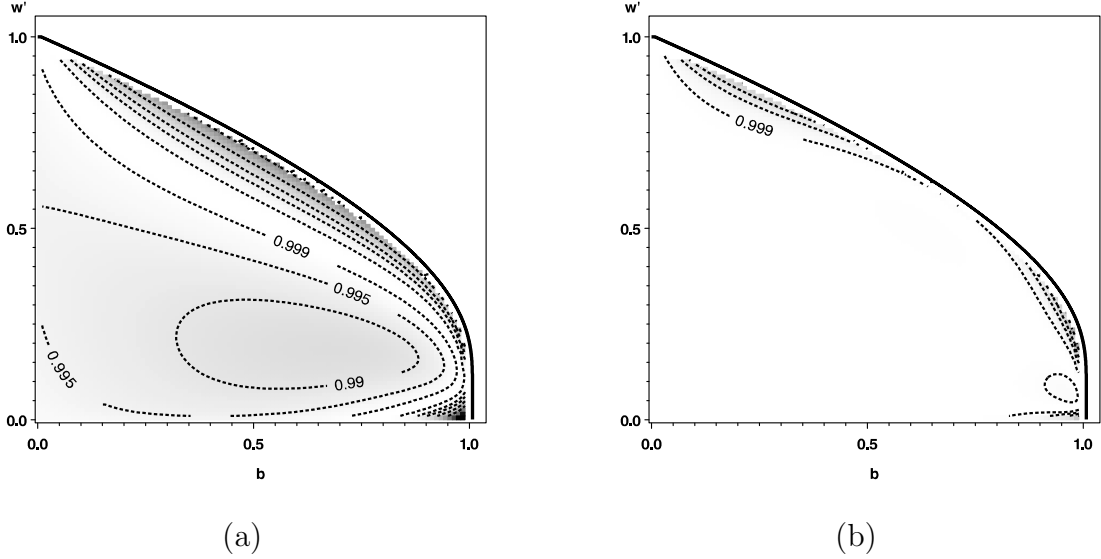
11

| (a) | (b) |

Figure 5: Efficiency of the PM $(v_{PM}(c_*)/v_{PM}(c_a))$ with (a) $c_a = 0.125 + 0.7125b$, (b) $c_a = (4.5253w'^2 - 1.8543w' + 0.523)b^2 + (3.0743w'^3 - 2.43w'^2 + 0.7115w' + 0.4255)b$.

For Method 1, the estimator $\widehat{c_A}$ for c_A is computed by minimizing $v_{PM}(c; \tilde{m}, \tilde{k}_{PM(c)})$ over the set A. Note that the preliminary estimators $(\tilde{m}, \tilde{k}_{PM(c)})$ used in the minimization of $v_{PM}(c; \tilde{m}, \tilde{k}_{PM(c)})$ over the set A are fixed. It would be incorrect to calculate $(\hat{m}, \hat{k}_{PM(c)})$ for each $c \in A$ and to choose the value c_A that minimizes $v_{PM}(c; \hat{m}, \hat{k}_{PM(c)})$; this is because the different values of $(\hat{m}, \hat{k}_{PM(c)})$ make the variances incomparable.

For Method 2, the preliminary estimates for $(\tilde{m}, \tilde{k}_{PM(c)})$ can be transformed to $(\tilde{b}, \tilde{w}')$ using the equations $\tilde{b} = 1 - (1 + \tilde{m}/\tilde{k})^{-\tilde{k}}$ and $\tilde{w}' = \tilde{b}/\tilde{m}$, under the assumption that the data are NBD. The values of $(\tilde{b}, \tilde{w}')$ can then be plugged directly into the regression equations to obtain an estimator $\widehat{c_a}$ for c_a.

4.4. Typical NBD Parameter Estimates in Market Research

Fig. 6 shows typical estimates of the NBD parameters, derived by using the MOM/ZTM, when modelling consumer purchase occasions for 50 different categories and the top 50 brands within each category. The contour levels in Fig. 6(b) clearly show that NBD parameters for a significant number of categories and some large brands are inefficiently estimated by using the MOM/ZTM, thereby justifying the use of the PM at suitable c.
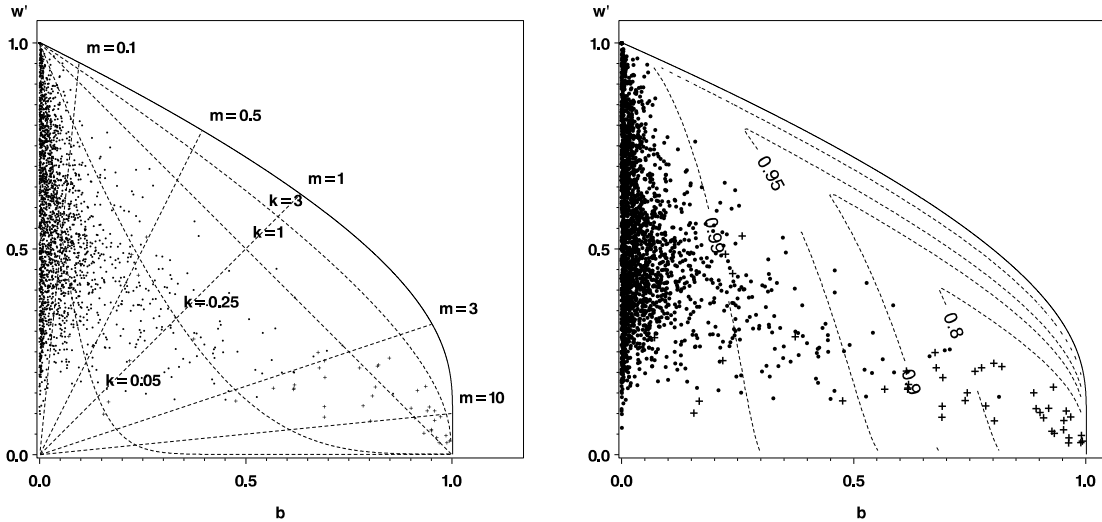
12

Figure 6: Typical NBD values for purchase occasions for 50 categories ($+$) and their top 50 brands ($\bullet$) derived using the MOM/ZTM. The contour levels in (b) show the asymptotic efficiency of the MOM/ZTM method with respect to ML. Data courtesy of ACNielsen BASES.

4.5. Robustness of Implementing the Power Method

In order to obtain robust and efficient NBD estimators by implementing the PM, it is important that approximating the value of c_* leads to a negligible loss of efficiency in the estimation of k. The value of c_* may be approximated, for example, by the approximations c_A or c_a discussed earlier.

Fig. 7 shows the variation in the approximated value of c_* between three different approximation methods. Fig. 7(a) shows regions of $c_A \in \{0, 0.2, 0.4, 0.6, 0.8, 1\}$ where the value of c_* is approximated by $c_A = \mathrm{argmin}_{c \in A} v_{PM}(c; m, k)$. The solid lines represent the boundaries for each c_A that is depicted and the dotted lines show where $c_A = c_*$. Fig. 7(b) and Fig. 7(c) both show contour levels of c_a for the two regression methods described in Section 4.2.

We have already shown that estimating k using the PM at each of the three approximations for c_* described above can achieve an efficiency very close to one with respect to using the PM at c_* (see Fig. 4 (b), Fig. 5(a) and Fig. 5(b)). Even with the difference in the values of approximated c_*, all three methods are highly efficient with respect to the PM at

13

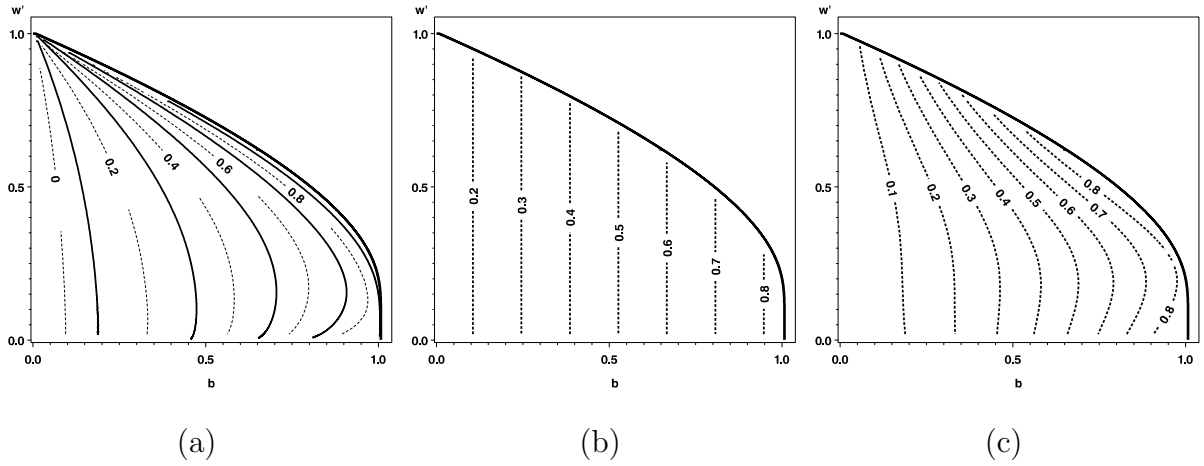(a)   (b)   (c)

Figure 7: (a) c_A with $A = \{0, \frac{1}{5}, \frac{2}{5}, \frac{3}{5}, \frac{4}{5}, 1\}$, (b) $c_a = 0.125 + 0.7125b$ and (c) $c_a = (4.5253w'^2 - 1.8543w' + 0.523)b^2 + (3.0743w'^3 - 2.43w'^2 + 0.7115w' + 0.4255)b$.

c_* and this highlights the insensitive nature of estimating k efficiently to gradual changes in the value of c.

Since both ZTM estimators for (m, k) and (b, w') are asymptotically uncorrelated, the ZTM estimators are the most convenient for estimating the approximations to c_*. Note that the PM estimators for (b, w') in general are correlated. To investigate the robustness of using the ZTM estimators as preliminary estimators we consider the worst (lowest) efficiency attainable in the estimation of $\hat{k}$, when considering all preliminary estimators $(\tilde{b}, \tilde{w}')$ within an asymptotic 95% confidence ellipse centered at the true values (b, w').

Let us denote the worst value of c that gives this lowest efficiency by $c_\times$ then $c_\times = \operatorname{argmax}_{\hat{c} \in \hat{C}}\{v_{PM}(\hat{c}; b, w')\}$. If $A = \{0, \frac{1}{5}, \frac{2}{5}, \frac{3}{5}, \frac{4}{5}, 1\}$ and we estimate c_A by $\widehat{c_A}$ then $\hat{C}$ is the set of all $\hat{c} = \operatorname{argmin}_{c \in A}\{v_{PM}(c; \tilde{b}, \tilde{w}')\}$ with $\mathbb{P}(b = \tilde{b}, w' = \tilde{w}') \geq 0.05$. If c_a is estimated by $\widehat{c_a}$ then $\hat{C}$ is the set of all values of $\hat{c}$ such that $\mathbb{P}(b = \tilde{b}, w' = \tilde{w}') \geq 0.05$ and (a) $\hat{c} = 0.125 + 0.7125\tilde{b}$ or (b) $\hat{c} = (4.5253\tilde{w}'^2 - 1.8543\tilde{w}' + 0.523)\tilde{b}^2 + (3.0743\tilde{w}'^3 - 2.43\tilde{w}'^2 + 0.7115\tilde{w}' + 0.4255)\tilde{b}$.

Fig. 8 shows the minimum efficiency attainable, given by $v_{PM}(c_*, b, w')/v_{PM}(c_\times, b, w')$, when estimating c_A or c_a by using preliminary ZTM estimators for a NBD sample of sizes $N = 1,000$ and $N = 10,000$. The black regions show regions of efficiency< 0.97, note however that this is a minimum possible efficiency and that this minimum efficiency occurs
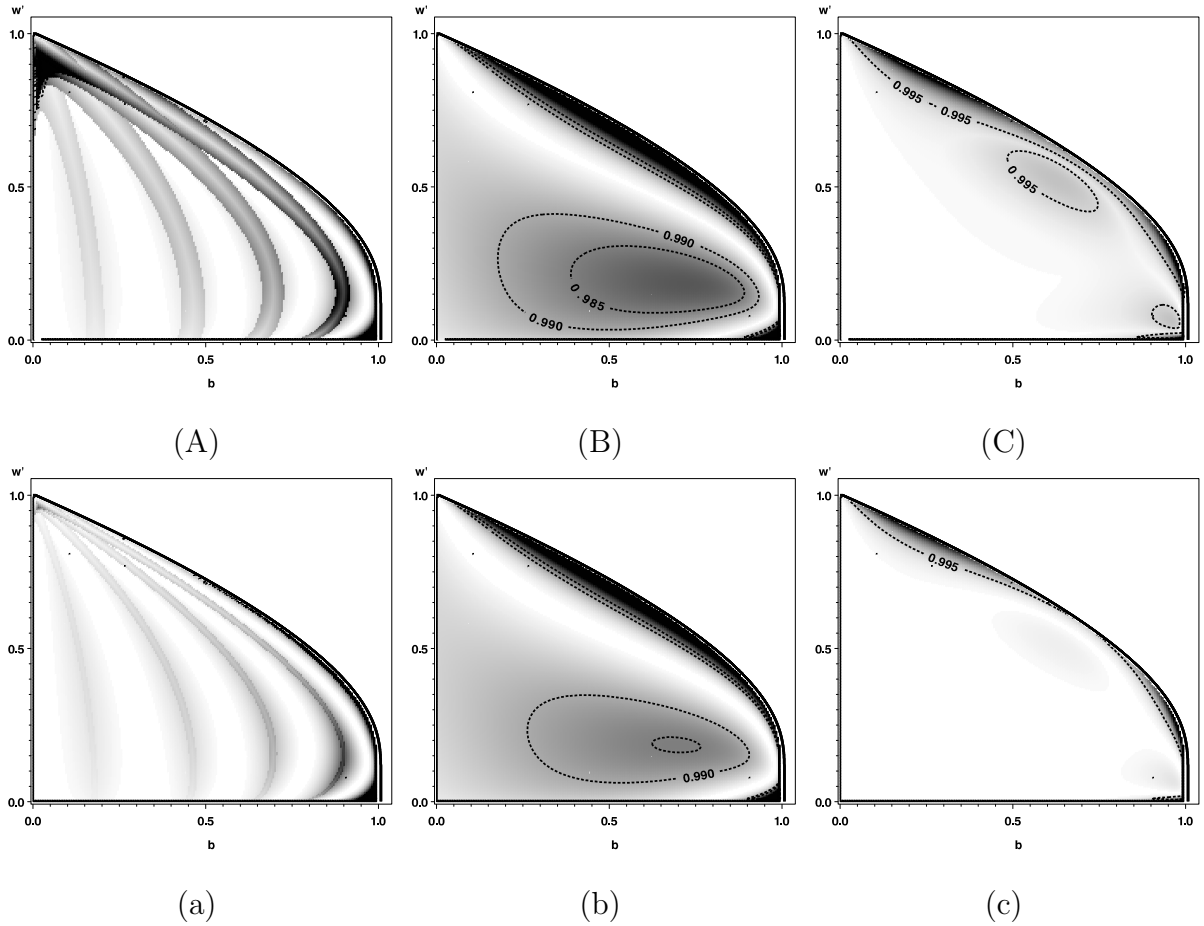
14

Figure 8: $v_{PM}(c_*, b, w')/v_{PM}(c_\times, b, w')$ for sample sizes of $N = 1,000$ (figures A, B and C) and $N = 10,000$ (figures a, b and c) and c_* approximated by (a) c_A with $A = \{0, \frac{1}{5}, \frac{2}{5}, \frac{3}{5}, \frac{4}{5}, 1\}$, (b) $c_a = 0.125 + 0.7125b$ and (c) $c_a = (4.5253w'^2 - 1.8543w' + 0.523)b^2 + (3.0743w'^3 - 2.43w'^2 + 0.7115w' + 0.4255)b$. Black regions show efficiency < 0.97.

for a given significance level and sample size. The methods become more robust as the sample size increases and also as the significance level decreases.

Fig. 8(a) shows the robustness when estimating c_A by $\widehat{c_A}$ with $A = \{0, \frac{1}{5}, \frac{2}{5}, \frac{3}{5}, \frac{4}{5}, 1\}$. This method is clearly most robust when $c_A = c_*$. Note, that for small b close to zero and large w' close to 0.8, where the ZTM should be efficient, the PM method when using preliminary ZTM estimators can become less robust. This is not surprising, however, since the coefficient

of variation for $\hat{k}$ even for ML in this area is exceptionally high (see Fig. 2(b)).

Fig. 8(b) and Fig. 8(c) show that the corresponding regression methods become less robust when the true parameters are close to the boundary $w' = -b/\log(1-b)$. In this region of the parameter space the coefficient of variation of $\hat{k}$ is exceptionally high. The possible loss of efficiency for both regression methods is clearly negligible, except for at the boundary, and both methods may therefore be implemented in practice to obtain robust and efficient estimators of the NBD distribution.

4.6. DEGENERATE SAMPLES

For any $c \in [0, 1]$, in the implementation of the power method there is a positive probability that the estimator for k has no valid positive solution when equating the sample moment to the corresponding theoretical moment, even though the sampling distribution is NBD.
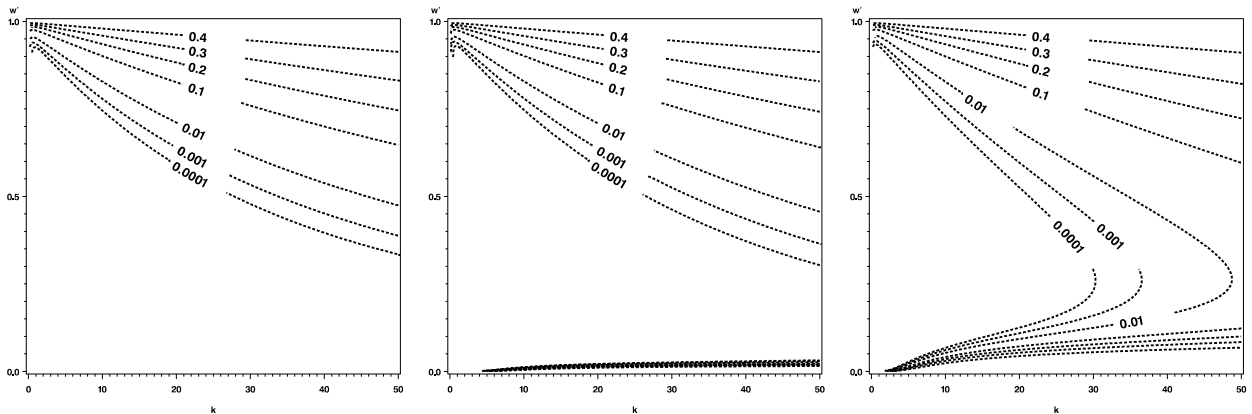

Figure 9: From left to right: Probability of obtaining a degenerate sample with sample size $N = 10,000$ using (a) MOM, (b) PM ($c = 0.5$) and (c) ZTM.

Fig. 9 shows the probability of obtaining a sample that generates an invalid estimator for k using a sample size of $N = 10,000$ within the (k, w') parameter space for the MOM, PM ($c = 0.5$) and ZTM estimators. The probability is calculated using the fact that asymptotically ($N \to \infty$) the estimators follow a bivariate normal distribution (see Section 3.2) and that an invalid estimator is obtained when $\bar{x} > s^2$, $\widehat{c^X} < \exp\left(-\bar{x}(1-c)\right)$ and $\hat{p}_0 < \exp\left(-\bar{x}\right)$ for the MOM, PM and ZTM respectively.

16

In literature (see e.g. Anscombe (1950)) it is often assumed that, when an invalid estimator for k is obtained, the Poisson distribution may be fitted and $\hat{k} = \infty$. The probability of obtaining an invalid estimator $\hat{k}$ increases as the NBD converges to the Poisson distribution and the Poisson approximation to the NBD therefore seems reasonable. If forecasting measures are sensitive to changes in the value of k then care must be taken, since an invalid estimator $\hat{k}$ may be obtained even for small values of k, albeit with a small probability.

5. Simulation Study

In this section we consider the results of a simulation study comprising $R = 1,000$ sample runs of the NBD distribution with sample size $N = 10,000$ for various parameters (m, k) with $m \in \{0.1, 0.5, 1, 5, 10\}$ and $k \in \{0.01, 0.25, 0.5, 1, 3, 5\}$. The efficiency of estimators of k are investigated by comparing the coefficient of variation for the ML, MOM, PM at c_* and ZTM estimators. The robustness of implementing the PM in practice when using preliminary ZTM estimators is investigated by showing the variation of the preliminary estimates within the (b, w') parameter space.

Table 1 shows the coefficient of variation $\hat{\kappa} = \sqrt{N}\sqrt{\frac{1}{R}\sum_{i=1}^{R}(\hat{k}_i - k)^2}/k = \sqrt{N \, \text{MSE}}/k$ for the ML, MOM, ZTM and PM at c_* against the theoretical coefficient of variation $\kappa = \sqrt{v_{ML}}/k$ (see Fig. 2 (b)). A value of $\hat{\kappa} = \infty$ indicates that $\hat{k}_i \leq 0$ or $\hat{k}_i = \infty$ for at least one sample. For all samples with $\hat{\kappa} < \infty$ the PM at c_* has a consistently lower $\hat{\kappa}$ than both the MOM and ZTM estimators. The largest percentage difference between the PM at c_* and the combined MOM/ZTM method occurs when $k = 1$ and $m = 10$ when the value of $\hat{\kappa}$ is increased by a factor of 26% by using the MOM/ZTM method.

Fig. 10 shows preliminary ZTM estimates, $(\tilde{b}, \tilde{w}')$, for different NBD parameters within the (b, w') parameter space. For each parameter pair, ZTM estimates for $1,000$ different NBD samples of size $N = 10,000$ are shown. When comparing the ZTM estimates in Fig. 10 (a) to values of c_* in Fig. 3(b) it is clear that, even with the variation in the estimates $(\tilde{b}, \tilde{w}')$ for each parameter pair (b, w'), the variation in the corresponding estimated values of c_* will be small in most regions of the (b, w')-space. The regions where c_* is sensitive to small changes

17

(a) (b) $m = 0.1\ k = 0.25$ (b) $m = 0.1\ k = 1$

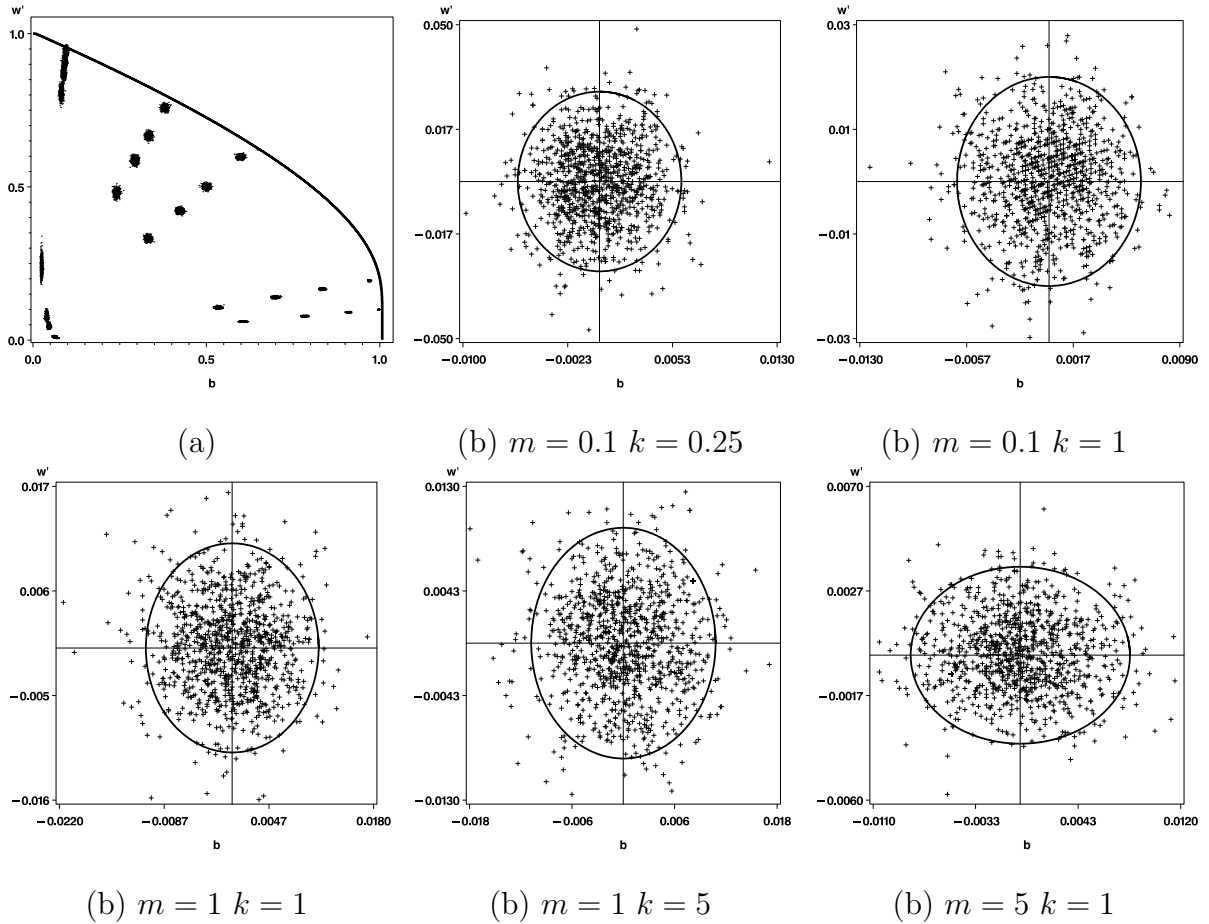(b) $m = 1\ k = 1$ (b) $m = 1\ k = 5$ (b) $m = 5\ k = 1$

Figure 10: (a) ZTM preliminary estimators for $m \in \{0.1, 0.5, 1, 5, 10\}$ and $k \in \{0.01, 0.25, 0.5, 1, 5\}$ in (b, w')-space. (b) ZTM preliminary estimators with corresponding 95% confidence ellipse.

in (b, w') and the corresponding maximum possible loss of efficiency in these regions was shown in Fig. 8. The maximum possible loss of efficiency was based on a 95% confidence ellipse. Fig. 10 (b) shows examples of preliminary ZTM estimates within the corresponding theoretical 95% confidence ellipses for (b, w'). These pictures are typical for each of the parameter pairs considered in Fig. 10 (a).

We would like to thank the referee for the careful reading of our manuscript and valuable comments.

18

m	$k=0.5$ $\sqrt{v_{ML}}/k$	MLE	ZTM	PM	MOM	$k=0.25$ $\sqrt{v_{ML}}/k$	MLE	ZTM	PM	MOM	$k=0.01$ $\sqrt{v_{ML}}/k$	MLE	ZTM	PM	MOM
0.1	**14.01**	15.34	15.34	15.34	16.25	**10.05**	10.38	10.38	10.38	11.50	**8.27**	8.78	8.78	8.78	15.21
0.5	**4.14**	4.06	4.11	4.07	4.86	**3.56**	3.62	3.63	3.62	4.76	**5.90**	5.91	5.91	5.91	13.70
1	**2.85**	2.85	2.92	2.86	3.60	**2.66**	2.65	2.67	2.65	4.00	**5.28**	5.30	5.30	5.30	13.86
5	**1.70**	1.77	1.89	1.78	2.66	**1.77**	1.80	1.83	1.81	3.34	**4.41**	4.48	4.48	4.48	13.75
10	**1.51**	1.51	1.70	1.53	2.56	**1.59**	1.61	1.66	1.61	3.30	**4.15**	4.19	4.20	4.19	13.74

m	$k=5$ $\sqrt{v_{ML}}/k$	MLE	ZTM	PM	MOM	$k=3$ $\sqrt{v_{ML}}/k$	MLE	ZTM	PM	MOM	$k=1$ $\sqrt{v_{ML}}/k$	MLE	ZTM	PM	MOM
0.1	**78.86**	∞	∞	∞	∞	**50.40**	∞	∞	∞	∞	**21.55**	∞	∞	∞	∞
0.5	**16.89**	19.92	22.26	20.00	19.90	**11.21**	11.88	12.55	11.88	12.07	**5.51**	5.52	5.68	5.53	6.00
1	**9.14**	9.28	10.73	9.29	9.38	**6.30**	6.63	7.45	6.63	6.82	**3.49**	3.51	3.65	3.52	4.07
5	**2.94**	2.96	5.63	2.96	3.09	**2.36**	2.41	3.78	2.41	2.67	**1.79**	1.81	2.07	1.83	2.38
10	**2.16**	2.14	7.41	2.15	2.34	**1.86**	1.80	4.25	1.81	2.07	**1.55**	1.58	2.02	1.60	2.22

Table 1: Comparison of $\sqrt{N}\sqrt{\frac{1}{R}\sum_{i=1}^{R}(\hat{k}_i - k)^2}/k$ against $\sqrt{v_{ML}}/k$ for the ML, ZTM, PM at c_* and MOM using $R = 1,000$ samples of the NBD distribution with sample size $N = 10,000$. A value of ∞ indicates that $\hat{k}_i \leq 0$ or $\hat{k}_i = \infty$ for at least one sample.

A. Asymptotic Distribution of Estimators

In this section we provide details on obtaining the asymptotic normal distributions of the estimators for the NBD parameters (m, k), (a, k), (b, w) and (b, w') when using the MOM, ZTM and PM methods of estimation.

The asymptotic distribution of the estimators can be derived by using a multivariate version of the so-called δ-method (see e.g. Serfling (1980), Chapter 3). Define $\overline{\mathbf{g}} = (\overline{g}_1, \overline{g}_2)$, $\mu = E\overline{\mathbf{g}}$ and $\mathbf{G}(\overline{\mathbf{g}}) = (G_1(\overline{\mathbf{g}}), G_2(\overline{\mathbf{g}}))$ and let $N_2(\mu_*, \boldsymbol{\Sigma}_*)$ denote the bivariate normal distribution with mean vector μ_* and variance-covariance matrix $\boldsymbol{\Sigma}_*$. According to the multivariate version of the δ-method, if $\lim_{N \to \infty} \sqrt{N}\, (\overline{\mathbf{g}} - \mu)' \sim N_2\left(\mathbf{0}', \boldsymbol{\Sigma}\right)$ then $\lim_{N \to \infty} \sqrt{N}\, (\mathbf{G}(\overline{\mathbf{g}}) - \mathbf{G}(\mu))' \sim N_2(\mathbf{0}', \mathbf{D}\boldsymbol{\Sigma}\mathbf{D}')$. Here $\mathbf{D} = [\partial G_u / \partial \overline{g}_v]_{u=1,2,v=1,2}$ is the matrix of partial derivatives of $\mathbf{G}$ evaluated at μ.

A.1. Asymptotic Distribution of Sample Moments

In the estimation of the NBD parameters the estimators for the MOM, ZTM and PM are functions of $\overline{\mathbf{f}}_{MOM} = (\overline{x}, \overline{x^2})'$, $\overline{\mathbf{f}}_{ZTM} = (\overline{x}, \widehat{p_0})'$ and $\overline{\mathbf{f}}_{PM} = (\overline{x}, \widehat{c^x})'$ respectively. It is well known that such functions $\overline{\mathbf{f}}$ of sample moments follow an asymptotic normal distribution (see Serfling (1980)). It is straightforward to show that each of the sample moments

$$f_0 = \overline{x} = \frac{1}{N} \sum_{i=1}^{N} x_i, \quad f_1 = \overline{x^2} = \frac{1}{N} \sum_{i=1}^{N} x_i^2, \quad f_2 = \widehat{p_0} = \frac{n_0}{N}, \quad \text{and} \quad f_3 = \widehat{c^X} = \frac{1}{N} \sum_{i=1}^{N} c^{x_i}$$

have means

$$E f_0 = m, \quad E f_1(X) = ak(a+1+ka), \quad E f_2(X) = (a+1)^{-k} \quad \text{and} \quad E f_3(X) = (1+a-ac)^{-k}.$$

The asymptotic normalized covariance matrix for $\bar{\mathbf{f}}_{MOM}$, $\bar{\mathbf{f}}_{ZTM}$ and $\bar{\mathbf{f}}_{PM}$ are

$$V_{MOM} = \lim_{N\to\infty} N\,\Sigma_{(\bar{x},\overline{x^2})} = ka(a+1)\begin{bmatrix} 1 & 2ka+2a+1 \\ 2ka+2a+1 & 4k^2a^2+6ka+10ka^2+6a^2+6a+1 \end{bmatrix}$$

$$V_{ZTM} = \lim_{N\to\infty} N\,\Sigma_{(\bar{x},\widehat{p_0})} = \begin{bmatrix} ka(a+1) & -ka(a+1)^{-k} \\ -ka(a+1)^{-k} & (a+1)^{-k}\left[1-(a+1)^{-k}\right] \end{bmatrix}$$

$$V_{PM} = \lim_{N\to\infty} N\,\Sigma_{(\bar{x},\widehat{c^X})} = \begin{bmatrix} ka(a+1) & ka(a+1)(c-1)(1+a-ac)^{-k-1} \\ ka(a+1)(c-1)(1+a-ac)^{-k-1} & (1+a-ac^2)^{-k}-(1+a-ac)^{-2k} \end{bmatrix}$$

A.2. Asymptotic Distribution of Estimators

Since the sample moments asymptotically follow the normal distribution the MOM, ZTM and PM estimators $(\hat{m},\hat{k}_{MOM})'$, $(\hat{m},\hat{k}_{ZTM})'$, $(\hat{m},\hat{k}_{PM})'$ are asymptotically normally distributed with mean vector $(m,k)'$ and asymptotically normalized covariance matrix

$$\lim_{N\to\infty} N\,\Sigma_{(\hat{m},\hat{k}_{MOM})} = D_{MOM}V_{MOM}D'_{MOM}, \quad \lim_{N\to\infty} N\,\Sigma_{(\hat{m},\hat{k}_{ZTM})} = D_{ZTM}V_{ZTM}D'_{ZTM} \quad \text{and}$$

$$\lim_{N\to\infty} N\,\Sigma_{(\hat{m},\hat{k}_{PM})} = D_{PM}V_{PM}D'_{PM}$$

for the MOM, ZTM and PM respectively. The matrix of partial derivatives are

$$D_{MOM} = \begin{bmatrix} 1 & 0 \\ \frac{2ka+2a+1}{a^2} & -\frac{1}{a^2} \end{bmatrix}, \quad D_{ZTM} = \begin{bmatrix} 1 & 0 \\ -\frac{1}{(a+1)\log(a+1)-a} & -\frac{(a+1)^{k+1}}{(a+1)\log(a+1)-a} \end{bmatrix} \quad \text{and}$$

$$D_{PM} = \begin{bmatrix} 1 & 0 \\ \frac{c-1}{r\log(r)-r+1} & -\frac{r^{k+1}}{r\log(r)-r+1} \end{bmatrix},$$

where $r = 1+a-ac$.

To find the asymptotic normal distribution of the estimators for (a,k), (b,w) and (b,w') we apply the same approach. We have the relationships

$$\hat{a} = 1+\frac{\hat{m}}{\hat{k}}, \quad \hat{b} = 1-\left(1+\frac{\hat{m}}{\hat{k}}\right)^{-\hat{k}}, \quad \hat{w} = \frac{\hat{m}}{1-\left(1+\frac{\hat{m}}{\hat{k}}\right)^{-\hat{k}}} \quad \text{and} \quad \hat{w}' = \frac{1}{\hat{m}}\left(1-\left(1+\frac{\hat{m}}{\hat{k}}\right)^{-\hat{k}}\right)$$

and since the asymptotic normal distributions of the estimators $(\hat{m}, \hat{k})$ are known for the MOM, ZTM and PM it is straightforward to compute the values of the asymptotic normal distribution of the estimators (a, k), (b, w) and (b, w') for each of these methods.

References

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.

Arbous, A., & Kerrich, J. (1951). Accident statistics and the concept of accident proneness. *Biometrics*, *7*, 340–432.

Barndorff-Nielsen, O., & Yeo, G. (1969). Negative binomial processes. *J. Appl. Probability*, *6*, 633–647.

Bondesson, L. (1992). *Generalized gamma convolutions and related classes of distributions and densities.* New York: Springer-Verlag.

Ehrenberg, A. (1988). *Repeat-buying: Facts, theory and applications.* London: Charles Griffin & Company Ltd.; New York: Oxford University Press.

Fisher, R. A. (1941). The negative binomial distribution. *Ann. Eugenics*, *11*, 182–187.

Grandell, J. (1997). *Mixed poisson processes.* Chapman & Hall.

Greenwood, M., & Yule, G. (1920). An inquiry 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. *J. Royal Statistical Society*, *93*, 255–279.

Haldane, J. B. S. (1941). The fitting of binomial distributions. *Ann. Eugenics*, *11*, 179–181.

Johnson, N. L., Kotz, S., & Kemp, A. W. (1992). *Univariate discrete distributions.* New York: John Wiley & Sons Inc.

Kendall, D. G. (1948a). On some modes of population growth leading to R. A. Fisher's logarithmic series distribution. *Biometrika*, *35*, 6–15.

Kendall, D. G. (1948b). On the generalized "birth-and-death" process. *Ann. Math. Statistics*, *19*, 1–15.

Lundberg, O. (1964). *On Random Processes and their Application to Sickness and Accident Statistics*. Uppsala: Almquist and Wiksells.

Savani, V., & Zhigljavsky, A. A. (2005). Power method for estimating the parameters of the negative binomial distribution. (Submitted to Statistics and Prob. Letters)

Serfling, R. J. (1980). *Approximation theorems of mathematical statistics*. New York: John Wiley & Sons Inc.