Metrika manuscript No.
(will be inserted by the editor)

Efficient parameter estimation for independent and INAR(1) negative binomial samples

V. SAVANI e-mail: SavaniV@Cardiff.ac.uk,
A. A. ZHIGLJAVSKY e-mail: ZhigljavskyAA@Cardiff.ac.uk

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

Received: date / Revised version: date

Abstract We consider moment based estimation methods for estimating parameters of the negative binomial distribution that are almost as efficient as maximum likelihood estimation and far superior to the celebrated zero term method and the standard method of moments estimator. Maximum likelihood estimators are difficult to compute for dependent samples such as samples generated from the negative binomial first-order autoregressive integer-valued processes. The power method of estimation is suggested as an alternative to maximum likelihood estimation for such samples and a comparison is made of the asymptotic normalized variance between the power method, method of moments and zero term method estimators.

Key words Negative binomial distribution, maximum likelihood, method of moments, efficiency of estimators, INAR(1) process.

1 Introduction

Parameter estimation for independent samples from the negative binomial distribution (NBD) has been thoroughly investigated in [5] and [6] using the maximum likelihood approach. The maximum likelihood estimator for the shape parameter of the NBD cannot be presented in an explicit form and must be found as the solution to an equation. Anscombe in [2] noted that the maximum likelihood equation for the shape parameter can be tedious to compute and therefore considered moment based estimators for the parameters of an independent and identically distributed (i.i.d.) NBD sample.

In this paper we study a family of moment based estimation methods, called the power method estimators, for estimating parameters of the NBD. These estimators are almost as efficient as maximum likelihood estimators and are far superior to the celebrated zero term method and method of moments. We further investigate the power method estimators when used to estimate NBD parameters for dependent NBD samples generated from the NBD first-order autoregressive integer-valued process, or INAR(1) process.

In Section 2 the maximum likelihood estimators and generalized moment based estimators for the mean m and shape parameter k of an i.i.d. NBD sample are presented. We discuss reasons why maximum likelihood estimators are unsuitable for use in practice when analyzing, say, market research data and show that, in such cases, it is preferable to use moment based estimators. In Section 3 we prove that the asymptotic normalized variance of the power method estimator can be smaller than the asymptotic normalized variance of both the method of moments and zero term method estimator discussed in [2]. Simple approaches of implementing the power method in practice are offered. Finally, in Section 4 the power method of estimation for dependent NBD samples generated from an INAR(1) process is studied.

2 Background

2.1 The negative binomial distribution

The NBD is a two parameter distribution usually defined by the mean m and shape parameter k. The probabilities of the NBD are given by

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

The distribution has a number of different parameterizations. The parameter pair (k, m) is statistically preferred as the maximum likelihood estimators $(\hat{k}_{ML}, \hat{m}_{ML})$ and all natural moment based estimators $(\hat{k}, \hat{m})$ are asymptotically uncorrelated given an i.i.d. NBD sample (see [2]).

Market research is an example of one of the areas in which the distribution has been successfully applied. Ehrenberg in [4] used the NBD to model the purchase frequencies of brands and categories by a population of individuals. Ehrenberg showed, using empirical evidence, that the number of purchases of a product by a population could be modeled using two NBD parameters called the penetration and purchase frequency. The penetration $b = 1-p_0$ represents the probability for a random individual to make at least one purchase and $w = m/b \ (w \geqslant 1)$ denotes the mean purchase frequency per buyer.

In this paper we use 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. The NBD is only defined for the parameter pairs $(b, w') \in (0, 1) \times (0, 1)$ such that $w' < -b/\log(1-b)$. Note that there is a one-to-one correspondence between the parameter pairs (b, w') and (k, m).

2.2 Parameter estimation

Maximum likelihood Parameter estimation using maximum likelihood (ML) for independent observations from a NBD has been investigated by Fisher in [5] and Haldane in [6]. The ML estimator for m is simply given by the sample mean $\bar{x}$, however there is no closed form solution for $\hat{k}_{ML}$ and the estimator is obtained as the solution in z to the equation

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

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

Difficulties with maximum likelihood Anscombe noted in [2] that the ML estimator is impractical even in the case of i.i.d. samples. In market research, for example, analyzing consumer purchase data often requires investigating data over different time periods of varying lengths. The n_i therefore differ according to the analysis period taken into consideration. Fitting the NBD to such data using the ML approach will require calculation of the n_i from raw transaction data for each analysis period. This data must record the date and the frequency of the product bought per household over a continuous time period. It is, however, very uneconomical to store and very difficult to obtain such raw transaction data and ML estimation is therefore hardly ever used in the practice of market research.

When modeling the number of purchase occasions by the NBD it may be the case that the data are not independent. In this paper we additionally consider estimation for dependent NBD samples generated from the NBD INAR(1) process. The dependency in the observations makes it extremely difficult to both formulate and solve the ML equations in order to obtain ML estimators. Since the NBD INAR(1) process is a stationary and ergodic process, moment based estimators provide a simple way of estimating parameters of the NBD in the case where ML estimators cannot be formulated.

Table 1 Three moment based estimation methods for the NBD

Method		$f(x)$	$Ef(X)$	Sample moment $\bar{f}$
Method of moments	(MOM)	x^2	$m(m+1)+\frac{m^2}{k}$	$\overline{x^2} = \frac{1}{N}\sum_{i=1}^{N} x_i^2$
Zero term method	(ZTM)	$I_{[x=0]}$	$\left(1+\frac{m}{k}\right)^{-k}$	$\widehat{p_0} = \frac{n_0}{N}$
Power method	(PM)	c^x	$\left(1+\frac{m(1-c)}{k}\right)^{-k}$	$\widehat{c^X} = \frac{1}{N}\sum_{i=1}^{N} c^{x_i}$

Table 2 Estimators for k

Method		Notation	Estimator or equation for $\hat{k}$
Maximum Likelihood	(ML)	$\hat{k}_{ML}$	Equation (2)
Method of Moments	(MOM)	$\hat{k}_{MOM}$	$\frac{\bar{x}^2}{s^2-\bar{x}}$
Zero Term Method	(ZTM)	$\hat{k}_{ZTM}$	$\widehat{p_0} = \left(1+\frac{\bar{x}}{z}\right)^{-z}$
Power Method	(PM)	$\hat{k}_{PM}$	$\widehat{c^X} = \left(1+\frac{\bar{x}(1-c)}{z}\right)^{-z}$

Generalized moment based estimators As an alternative to ML estimation, the use of moment based estimation methods were considered by Anscombe in [2]. The first moment used in estimation is the sample mean $\bar{x}$ as this is both a natural estimator and ML estimator for m. An additional sample moment, given by $\bar{f} = \frac{1}{N}\sum_{i=1}^{N} f(x_i)$, is required to estimate k; this is done by equating the sample moment to $Ef(X)$, with m replaced by $\hat{m} = \bar{x}$, and solving the equation with respect to k. Here N denotes the sample size and $f(x)$ can be any convex or concave function on the non-negative integers.

In this paper we shall consider three different functions as shown in Table 1. (In Table 1, $I_{[x=0]}$ denotes the indicator function of the event $x = 0$, n_0 denotes the number of zeros in the sample and $c \neq 1$ is a positive constant.) Table 2 presents the estimators of k for the ML method and the three moment based methods of Table 1. (In Table 2, n_i denotes the frequency of $i = 0, 1, 2, \ldots$ in the sample and s^2 denotes the sample variance.)

The MOM estimator for k is the only closed form estimator for k. The ML, PM, ZTM methods all require a numerical solution with respect to z for estimating k. Anscombe in [2] noted that all the equations in Table 2 for estimating k provide solutions which are uniquely defined with probability one, note however that there is a small probability that the estimator for k may be negative. The estimators for k considered above all have finite means and variances.

Note that the PM at $c = 0$ is equivalent to the ZTM. The PM at $c = 1$ yields no direct solution for estimating k. As $c \to 1$, however, the PM estimator tends to the MOM estimator. The PM therefore generalizes the MOM and ZTM methods of estimation. In literature and in practice the ZTM has been the most preferred method of estimation in the field of market research. This is related to the fact that the parameter $b = 1 - p_0$ has a natural interpretation and that the estimator n_0/N for b is often readily available. In consumer research, for example, the penetration may be estimated either from questionnaires, retail data or even shipment data.

2.3 Asymptotic efficiency of estimators

In this section we compare the relative efficiencies of moment based estimators of k with respect to the ML estimator. The estimator $\hat{m} = \bar{x}$ for m remains the same for all methods and since the parameters $\hat{m}$ and $\hat{k}$ are asymptotically uncorrelated for an i.i.d. sample (see e.g. [2]), only a comparison of the variance of $\hat{k}$ is required. The variances of the ML estimators are in fact 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. In separate papers both Fisher [5] and Haldane [6] derived expressions for the asymptotic variances of the NBD parameter estimators. The asymptotic normalized variance of $\hat{k}_{ML}$ obtained from the Fisher information matrix is given by

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

where $p = 1 + m/k$. (We use the notation p rather than m as it simplifies the formulae.)

The asymptotic normalized variances for the moment based estimators of k for a given function $f(\cdot)$ were derived by Anscombe (1950) to be

$$\lim_{N\to\infty} N\ Var\left(\hat{k}\right) = \frac{Ef^2(X) - [Ef(X)]^2 - \left(m + \frac{m^2}{k}\right)\left[\frac{\partial}{\partial m}Ef(X)\right]^2}{\left[\frac{\partial}{\partial k}Ef(X)\right]^2}\ . \tag{4}$$

The asymptotic normalized variances of $\hat{k}$ for MOM, ZTM and PM estimation methods are respectively given by

$$v_{MOM} = \lim_{N\to\infty} N\,Var(\hat{k}_{MOM}) = \frac{2k\,(k+1)\,p^2}{(p-1)^2}, \tag{5}$$

$$v_{ZTM} = \lim_{N\to\infty} N\,Var(\hat{k}_{ZTM}) = \frac{p^{k+2} - p^2 - kp(p-1)}{[p\log(p) - p + 1]^2}, \tag{6}$$

$$v_{PM}(c) = \lim_{N\to\infty} N\,Var(\hat{k}_{PM}) = \frac{\left(p-pc^2+c^2\right)^{-k}g^{2k+2} - g^2 - kp(p-1)(1-c)^2}{[g\log(g) - g + 1]^2} \tag{7}$$

where $p = 1 + m/k$ (as above) and $g = p - pc + c$.

3 The power method for i.i.d. samples

3.1 Inadmissability of MOM/ZTM

Let us define the more efficient of the MOM and ZTM estimator as the 'combined MOM/ZTM' estimator. Anscombe stated that the PM estimator is nowhere, within the parameter space, more efficient than the combined MOM/ZTM estimator. This statement has been assumed true in a number of textbooks and surveys (for example [7]) and is indeed true when comparing against the PM with fixed $c \in (0,1)$ for all parameter values of the NBD simultaneously. The following theorem states that this statement is always untrue if we allow $c \in (0,1)$ to change according to the parameter values.

Theorem 1 *The combined MOM/ZTM estimator is inadmissible in the class of PM estimators in the following sense: for any fixed k and m there exists c_*, with $0 < c_* < 1$, such that $v_{PM}(c_*) < \min\{v_{ZTM}, v_{MOM}\}$, where v_{MOM}, v_{ZTM} and $v_{PM}(\cdot)$ are the normalized asymptotic variances of $\hat{k}$ as defined in* (5), (6) *and* (7) *for the MOM, ZTM and PM respectively.*

Proof See Appendix B.

Fig. 1 shows the ratio $v_{PM}(c)/v_{ML}$ versus $c \in (0,1)$ for different parameter values of (k,m). Fig. 1 shows that there is, in fact, a range of values of c_*, such that $v_{PM}(c_*) < \min\{v_{ZTM}, v_{MOM}\}$. Moreover, there is a single optimum c, which we denote by c_o, that will give a PM estimator for k that is almost as efficient as the ML estimator for k. The value of c_o can be obtained numerically by minimizing $v_{PM}(c)$ for every (k,m).

3.2 Power method with sub-optimum c

In this section we study the efficiency of power method estimators with non-optimum c. The first type of estimators provide a simple and practical alternative to ML estimators. These estimators require the collection of statistics $\hat{c} = \frac{1}{N}\sum_{i=1}^{N} c^{x_i}$ for different values of c. The second estimator achieves efficiency very close to that of the ML estimator. The implementation of these methods, however, requires knowledge of the parameter values. Use of these methods in practice is considered in [10].

Simple and efficient estimators. Fig. 2 shows the efficiency of two simple PM estimators of k relative to the PM estimator computed at optimum c. The simple estimators are obtained by computing the PM($\tilde{c}$) estimator, where $\tilde{c} = \mathrm{argmin}_{c \in \mathcal{C}} v_{PM}(c)$ with $\mathcal{C} = \{0,1\}$ and $\mathcal{C} = \{1/4, 1/2, 3/4\}$. In the case of market research, this PM estimator offers a simple and viable alternative to ML since only the statistics $\hat{c} = \frac{1}{N}\sum_{i=1}^{N} c^{x_i}$ are required for each $c \in \mathcal{C}$ to estimate k. In market research such statistics are easily requested in comparison to actual frequency counts which are required for ML estimation.

The PM estimator computed at $\tilde{c} = \mathrm{argmin}_{c \in \mathcal{C}} v_{PM}(c)$ in the case $\mathcal{C} = \{0,1\}$ is the combined MOM/ZTM estimator and is commonly used in practice. One can verify that the efficiencies of the combined MOM/ZTM estimator with respect to the PM(c_o) estimator cover the whole interval $(0,1)$. The graph of the efficiency of the combined MOM/ZTM estimator relative to the ML estimator is almost identical to Fig. 2(a).

The efficiency, with respect to the PM(c_o) estimator, of the PM estimator computed at $\tilde{c} = \mathrm{argmin}_{c \in \mathcal{C}} v_{PM}(c)$ in the case $\mathcal{C} = \{1/4, 1/2, 3/4\}$ is shown in

Fig. 2(b). Comparing Fig. 2(a) against Fig. 2(b) we can conclude that the efficiency of PM($\tilde{c}$) estimator with $\mathcal{C} = \{1/4, 1/2, 3/4\}$ is much higher than the efficiency of the combined MOM/ZTM estimator.

Highly efficient estimators. To obtain more efficient parameter estimates than the simple estimators proposed above, the value of c must vary according to the true parameter values. It is possible to obtain values of c_o numerically. We can then use regression techniques to approximate c_o. As a reasonably good approximation to c_o we suggest

$$\widehat{c_o} = (4.5w'^2 - 1.9w' + 0.5)b^2 + (3.1w'^3 - 2.4w'^2 + 0.7w' + 0.4)b\,. \tag{8}$$

Fig. 3(a) shows the efficiency of the PM estimator computed at c_o relative to the ML estimator. Fig. 3(b) shows the efficiency of the PM estimator at $\widehat{c_o}$ relative to the PM estimator at c_o; this efficiency is practically equal to 1 for all NBD parameter values. The implementation of the PM(c_o) and PM($\widehat{c_o}$) estimators in practice require knowledge of the unknown parameter values. An adaptive procedure for updating values of c may be used. For example we may update values of b and w' in (8) as the sample increases.

4 The power method for the NBD INAR(1) process

In this section we consider the power method for estimating parameters of the NBD INAR(1) process. For the rest of this section, unless otherwise stated, we assume that X_t is a NBD INAR(1) process and we let $x_t,\ t = 1, 2, \dots N$, denote a realization from the process X_t. A non-negative integer-valued process X_t is an INAR(1) process if

$$X_t \stackrel{d}{=} \alpha \circ X_{t-1} + \varepsilon_t, \tag{9}$$

where $\alpha \circ X_{t-1}$ and ε_t are independent, the ε_t form a sequence of uncorrelated random variables and $X_t \stackrel{d}{=} X_{t-1}$ for all t, with $\stackrel{d}{=}$ denoting equivalence in distribution, see e.g. [1,9]. Here '$\circ$' is called the thinning operator and $\alpha \circ X$ is defined as $\alpha \circ X \stackrel{d}{=} \sum_{j=1}^{X} Z_j$ where the Z_j are i.i.d. Bernoulli random variables with $P(Z_j = 1) = \alpha$ and $P(Z_j = 0) = 1 - \alpha$. It is assumed that the X_t and ε_t have finite means and variances. Note that if $\alpha = 0$ then then INAR(1) sample is i.i.d.

The INAR(1) process X_t with marginal distribution π will only be an INAR(1) process if the random variable X_π, with distribution π, is discrete self-decomposable so that the probability generating function $G_{X_\pi}(c) = Ec^{X_\pi}$ satisfies

$$G_{X_\pi}(c) = G_{X_\pi}(1 - \alpha + \alpha c) G_\varepsilon(c;\ \alpha), \qquad \alpha \in (0, 1). \tag{10}$$

McKenzie in [8] derived the NBD INAR(1) process. Note that if the process X_t has a NBD marginal distribution with parameters (k, m) then X_t is discrete self-decomposable since the probability generating function of X_t can be written in the form of equation (10) with

$$\underbrace{\left(1+\frac{m(1-c)}{k}\right)^{-k}}_{G_{X_\pi}(c)} = \underbrace{\left(1+\frac{m\alpha(1-c)}{k}\right)^{-k}}_{G_{X_\pi}(1-\alpha+\alpha c)} \underbrace{\left(\frac{k+m(1-c)}{k+m\alpha(1-c)}\right)^{-k}}_{G_\varepsilon(c;\ \alpha)}. \quad (11)$$

If X_t is NBD with parameters (k, m) then the errors ε_t must follow the NBD-Geometric distribution with

$$P\{\varepsilon_t = x\} = \sum_{j=0}^{\infty} \binom{j+x-1}{x} \left(\frac{k}{k+m\alpha}\right)^j \left(\frac{m\alpha}{k+m\alpha}\right)^x \binom{k+j-1}{j} \alpha^k (1-\alpha)^j,$$

where $x = 0, 1, 2 \ldots$.

4.1 Estimating parameters of the NBD INAR(1) process

Al-Osh and Alzaid in [1] addressed the problem of estimating the parameter α and the mean parameter λ of a Poisson INAR(1) process using three different types of estimators. The first two methods of estimation use moment based methods and are asymptotically equivalent. The final method uses a maximum likelihood approach. These estimation methods and their suitability for estimating parameters of the NBD INAR(1) process will now be reviewed.

For the moment based estimators, the thinning parameter α is estimated from the autocorrelation function of the INAR(1) process. Since the autocorrelation function of the INAR(1) process is identical to that of the AR(1) process, the problem of estimating α is well documented in many textbooks (see e.g. [3]). Once α is estimated by $\hat{\alpha}$, Al-Osh and Alzaid obtain a sequence of estimators $\hat{\varepsilon_t}$ through the equation $\hat{\varepsilon_t} = x_t - \hat{\alpha} x_{t-1}$. Standard moment based estimation methods are then used to estimate the parameters of the marginal distribution of the error process. The distribution of the errors for the NBD INAR(1) process is not simple and this makes inference about the estimators of (k, m) difficult.

The conditional maximum likelihood method discussed in [1] maximizes the likelihood function given knowledge of the process at time zero (i.e. given x_0). In the case of the NBD INAR(1) model, the maximum likelihood equations are already complex in the i.i.d. case (i.e. when $\alpha = 0$) and the maximum likelihood equations become much harder even to formulate when $\alpha > 0$. It is also very difficult to compute the Fisher information matrix and therefore difficult to obtain the asymptotic distribution of ML estimators.

In this paper we consider estimating the distributional parameters of the NBD INAR(1) process by using the moments of X_t as opposed to the moments of ε_t. Since the INAR(1) process is a stationary and ergodic process, the expected values of the sample moments for an observed realization and the stationary distribution are equal. An unbiased estimator for m is therefore $\hat{m} = \bar{x} = \frac{1}{N}\sum_{t=1}^{N} x_t$. The power method estimator $\hat{k}_{PM}$ for the shape parameter of the NBD distribution is computed by solving, in z, the equation $\widehat{c^X} = \frac{1}{N}\sum_{t=1}^{N} c^{x_t} = (1 + \bar{x}(1-c)/z)^{-z}$.

4.2 Efficiency of the power method estimator

Although computing moment based estimators for a NBD INAR(1) process and an i.i.d. NBD sample are identical, the fact that the values of x_t are correlated for INAR(1) samples implies that the covariance matrices of the estimators of (k, m) are different. In this section we derive the normalized asymptotic covariance matrix for the statistics $\left(\bar{x}, \widehat{c^X}\right)$, from which the normalized covariance matrix of $\left(\hat{k}_{PM}, \hat{m}\right)$ is consequently derived.

Theorem 2 *Let $\{x_t;\ t = 1, 2, \ldots, N\}$ be a sample realization from an INAR(1) process X_t with stationary distribution π. Let $\hat{\beta}_s = \left(\bar{x}, \widehat{c^X}\right)^T$ with $\bar{x} = \frac{1}{N}\sum_{t=1}^{N} x_t$ and $\widehat{c^X} = \frac{1}{N}\sum_{t=1}^{N} c^{x_t}$, with $c > 0$ and $c \neq 1$. Then $\hat{\beta}_s$ has an asymptotic normal distribution given by $\lim_{N\to\infty}\sqrt{N}\left(\hat{\beta}_s - E\hat{\beta}_s\right) \sim \mathcal{N}(0, D_c)$ with*

$$D_c = \begin{pmatrix} V_{\bar{X}} & C_{\bar{X},\widehat{c^X}} \\ C_{\bar{X},\widehat{c^X}} & V_{\widehat{c^X}} \end{pmatrix}. \tag{12}$$

Here

$$\begin{aligned}
V_{\bar{X}} &= \lim_{N\to\infty} N\ Var\left(\bar{X}\right) = \left(\frac{1+\alpha}{1-\alpha}\right) Var[X_\pi], \\
V_{\widehat{c^X}} &= \lim_{N\to\infty} N\ Var\left(\widehat{c^X}\right) = Var\left(c^{X_\pi}\right) \\
&\quad + 2\lim_{N\to\infty}\sum_{r=1}^{N-1}\left(1-\frac{r}{N}\right)\left\{G_{X_\pi}\left(c\left[1-\alpha^r+\alpha^r c\right]\right)G_\varepsilon(c;\alpha^r) - G^2_{X_\pi}(c)\right\}, \\
C_{\bar{X},\widehat{c^X}} &= \lim_{N\to\infty} N\ Cov\left(\bar{X}, \widehat{c^X}\right) = Cov\left(X_\pi, c^{X_\pi}\right) \\
&\quad + \lim_{N\to\infty}\sum_{r=1}^{N-1}\left(1-\frac{r}{N}\right)\left\{E\left[X_\pi\left(1-\alpha^r+\alpha^r c\right)^{X_\pi}\right]G_\varepsilon(c;\alpha^r) - E[X_\pi]G_{X_\pi}(c)\right\} \\
&\quad + \lim_{N\to\infty}\sum_{r=1}^{N-1}\left(1-\frac{r}{N}\right)\left\{G_{X_\pi}\left(c\left[1-\alpha^r+\alpha^r c\right]\right) - \alpha^r E[X_\pi]G_{X_\pi}(c)\right\}.
\end{aligned}$$

Proof See Appendix C.

Note that the asymptotic distribution of $\hat{\beta}_s = \left(\bar{x}, \widehat{c^X}\right)^T$ derived in Theorem 2 holds for any INAR(1) process and not just the NBD INAR(1) process. In the case of a NBD INAR(1) sample we have $Var[X_\pi] = m + m^2/k$, $Var\left(c^{X_\pi}\right) = G_{X_\pi}(c^2) - G^2_{X_\pi}(c)$ and $E[X_\pi c^{X_\pi}] = mc\left(1 + m(1-c)/k\right)^{-k-1}$. The generating functions $G_{X_\pi}(c)$ and $G_\varepsilon(c;\alpha)$ are given in equation (11). Fig. 4 shows 95% asymptotic bivariate normal confidence ellipses for $E\hat{\beta}_s$, centered at zero, given by the equation

$$\left(\hat{\beta}_s - E\hat{\beta}_s\right)^T D_c^{-1}\left(\hat{\beta}_s - E\hat{\beta}_s\right) \leqslant \chi_{0.95}(2) \simeq 5.99\,.$$

Fig. 5 shows estimates $\hat{\beta}_s - E\hat{\beta}_s$ obtained from 1000 simulations from a NBD INAR(1) process together with corresponding 95% asymptotic bivariate normal confidence ellipses. The parameters used for the NBD INAR(1) process were $m = 1$, $k = 2$, $N = 1000$, $\alpha \in \{0, 0.5\}$ and the PM estimator $\widehat{c^X}$ was computed using the value $c = 0.5$.

Corollary 1 *Let $\{x_t;\ t = 1, 2, \ldots, N\}$ be a sample realization from a NBD INAR(1) process X_t with NBD parameters (k, m). Let $\hat{\beta} = (\hat{k}_{PM}, \hat{m})^T$ be the power method estimators, with fixed c, $0 < c < 1$, obtained from the NBD INAR(1) sample, then $\hat{\beta}$ has an asymptotic normal distribution with $\lim_{N\to\infty} \sqrt{N}\left(\hat{\beta} - E\hat{\beta}\right) \sim \mathcal{N}(0, \Sigma_\alpha(c))$ where*

$$\Sigma_\alpha(c) = \begin{pmatrix} D_{\hat{k},\bar{x}} & D_{\hat{k},\widehat{c^X}} \\ D_{\hat{m},\bar{x}} & D_{\hat{m},\widehat{c^X}} \end{pmatrix} \begin{pmatrix} V_{\bar{X}} & C_{\bar{X},\widehat{c^X}} \\ C_{\bar{X},\widehat{c^X}} & V_{\widehat{c^X}} \end{pmatrix} \begin{pmatrix} D_{\hat{k},\bar{x}} & D_{\hat{k},\widehat{c^X}} \\ D_{\hat{m},\bar{x}} & D_{\hat{m},\widehat{c^X}} \end{pmatrix}^T . \tag{13}$$

Here $D_{f(\nu),\,\nu}$ is the derivative of $f(\nu)$ with respect to ν and $D_{f(\nu),\,\nu}$ is evaluated at the point $(\hat{k}, \hat{m}) = (k, m)$. The matrix of partial derivatives is

$$\begin{pmatrix} D_{\hat{k},\bar{x}} & D_{\hat{k},\widehat{c^X}} \\ D_{\hat{m},\bar{x}} & D_{\hat{m},\widehat{c^X}} \end{pmatrix} = \begin{pmatrix} \dfrac{c-1}{g\log(g) - g + 1} & -\dfrac{g^{k+1}}{g\log(g) - g + 1} \\ 1 & 0 \end{pmatrix}$$

where $g = p - pc + c = 1 + m(1-c)/k$.

The asymptotic distribution of the estimators $(\hat{k}_{PM}, \hat{m})$ has been derived by using a multivariate version of the so-called δ-method (see e.g. [11]).

Fig. 6 shows 95% asymptotic bivariate normal confidence ellipses for $E\hat{\beta}$, centered at zero. For $\alpha = 0$ the estimators $\hat{m}$ and $\hat{k}_{PM}$ are clearly uncorrelated. For $\alpha \in (0, 1)$, however, there is a positive correlation between the estimators $\hat{m}$ and $\hat{k}_{PM}$. A comparison of the efficiency of estimation methods may, therefore, no longer be made by comparing just the variance of $\hat{k}_{PM}$. A traditional method for comparing the efficiency of correlated estimators is by minimizing the determinant of the covariance matrix.

Fig. 7(a) shows $\det(\Sigma_\alpha(c))$, the determinant of the asymptotic normalized covariance matrix as defined by equation (13), plotted against c for $k = 0.5$ and $m = 1$ with $\alpha \in \{0, 0.25, 0.5, 0.75\}$. Note that the optimal value of c is never equal to 0 or 1. The figure shows that increasing the value of α can substantially increase the determinant of the covariance matrix. Fig. 7(b) shows values of efficiency defined by $\det(\Sigma_{0.25}(c_{0.25})) / \det(\Sigma_0(c_o))$ for all NBD values. Here c_α is the value of c that minimizes $\det(\Sigma_\alpha(c))$. Fig. 7(b) indicates that increasing the value of α from 0 to 0.25 means that we can only hope to be at most 55% as efficient when estimating from INAR(1) samples with $\alpha = 0.25$ in comparison to estimating from i.i.d. NBD samples for the majority of NBD parameter values. For $\alpha = 0.75$ the value of $\det(\Sigma_{0.75}(c_{0.75})) / \det(\Sigma_0(c_o))$ is at most 4% for the majority of NBD parameter values.

4.3 Estimating parameters of NBD INAR(1) samples under the false assumption of independence

In practice it may happen that the dependence within samples is ignored. In the case of dependent NBD INAR(1) samples, when implementing the power method using the optimum value of c, the incorrect optimum value c_o would then be used to compute the power method estimators rather than the true optimum value c_α. Fig. 8 shows contour levels of c_o and $c_{0.75}$. For fixed k and m, as α increases, the value of optimum c also increases.

Fig. 9 shows properties of the power method estimator in the case $\alpha = 0.75$. Fig. 9(a) shows the correlation between the estimators $\hat{k}_{PM}(c_{0.75})$ and $\hat{m}$. Fig. 9(b) shows the efficiency, defined by $\det(\Sigma_{0.75}(c_{0.75})) / \det(\Sigma_{0.75}(c_o))$, when estimating using the $\hat{k}_{PM}(c_o)$ relative to the $\hat{k}_{PM}(c_{0.75})$ estimator. Fig. 9(b) therefore shows the efficiency retained when estimating under the assumption of independent samples when in fact the sample is a NBD INAR(1) sample with $\alpha = 0.75$. Note that the Fisher information matrix is difficult to formulate and compute for NBD INAR(1) samples and therefore it is difficult to make a comparison between the covariance matrix of moment based estimators and maximum likelihood estimators for independent and dependent samples.

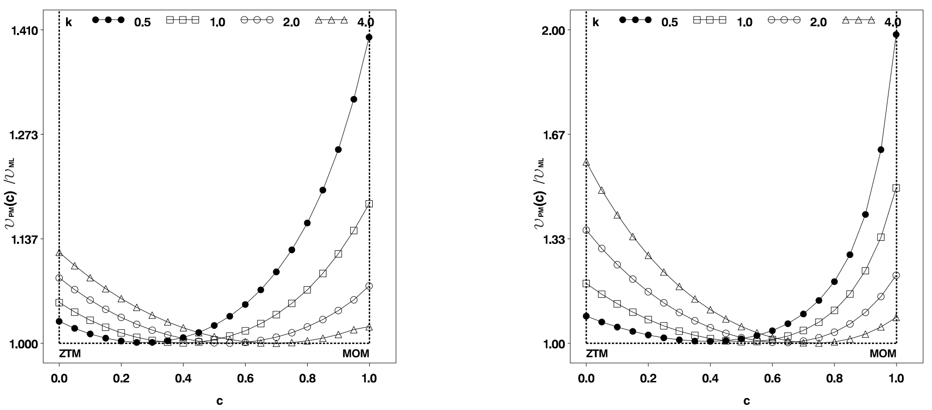
Conclusion

Standard method of moments and zero term method estimators are moment based estimators, popular in the field of market research, for estimating the parameters of the NBD. These moment based estimators are inefficient in certain regions of the NBD parameter space. The proposed power method estimator is almost as efficient as the maximum likelihood estimator and can be simple to implement in practice. The power method estimator, when correctly implemented, is always more efficient than the combined method of moments and zero term method estimator.

In the case of NBD INAR(1) samples, the dependency between observations within samples makes it difficult to formulate maximum likelihood equations and especially to obtain asymptotic normalized variances for the estimators. The power method estimators, however, can still be easily implemented. It is straightforward to compute the asymptotic normalized variances for the power method estimators as presented in this paper. Unlike in the case of an i.i.d. sample, however, parameter estimates for (k, m) are correlated for an INAR(1) sample when the value of the thinning parameter $\alpha > 0$. In implementing the power method, the optimum value of the power method parameter c depends on α. The loss of efficiency in assuming an i.i.d. NBD sample when in fact the sample is a NBD INAR(1) sample with $\alpha = 0.75$ is presented graphically to show the sensitivity of using the power method estimator to changes in c. The loss of efficiency depends on values of the NBD parameters.

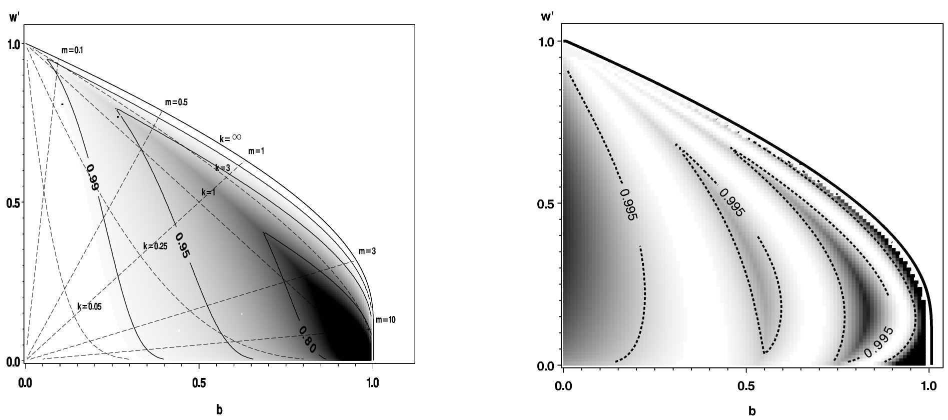
Appendicies

Appendix A: Figures



$m = 0.5$ $m = 2$

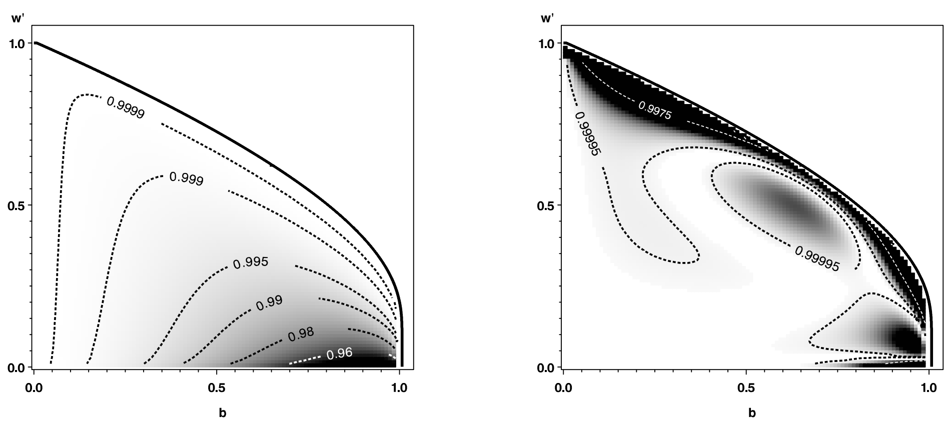
Fig. 1 $v_{PM}(c)/v_{ML}$ vs. c and asymptotic normalized variances for ML, MOM and ZTM.



(a) $v_{PM}(c_o)/v_{PM}(\tilde{c})$ with $\mathcal{C} = \{0, 1\}$
(Black shows efficiencies $\leqslant 0.7$)

(b) $v_{PM}(c_o)/v_{PM}(\tilde{c})$ with
$\mathcal{C} = \{1/4, 1/2, 3/4\}$
(Black shows efficiencies $\leqslant 0.975$)

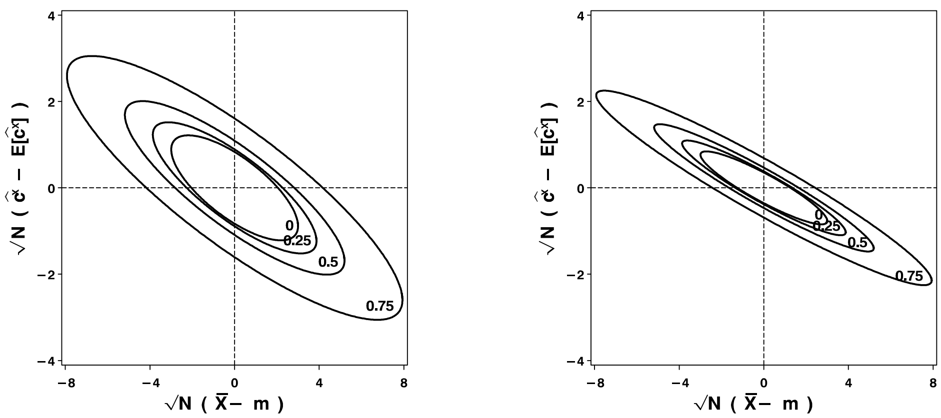
Fig. 2 Efficiency of the power method using $\tilde{c} = \text{argmin}_{c \in \mathcal{C}} v_{PM}(c)$ for some set $\mathcal{C}$.



(a) $v_{ML}/v_{PM}(c_o)$
(Black shows efficiencies $\leqslant 0.95$)

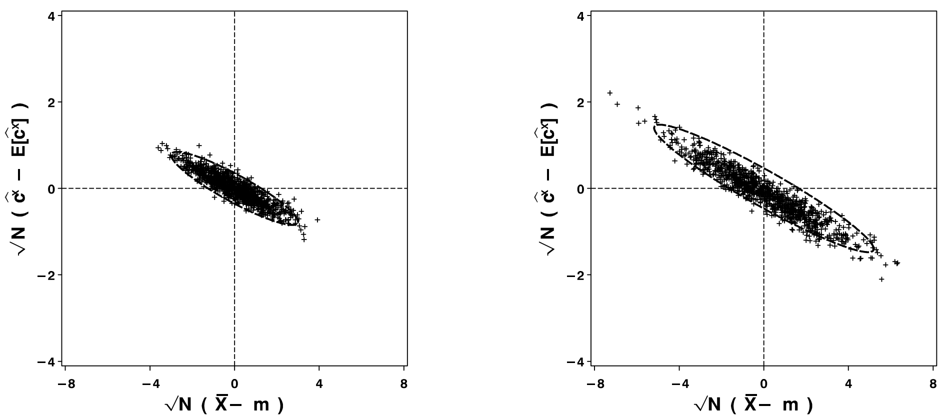
(b) $v_{PM}(c_o)/v_{PM}(\widehat{c_o})$
(Black shows efficiencies $\leqslant 0.999$)

Fig. 3 Efficiency of the power method using (a) $c = c_o$ and (b) $c = \widehat{c_o}$.



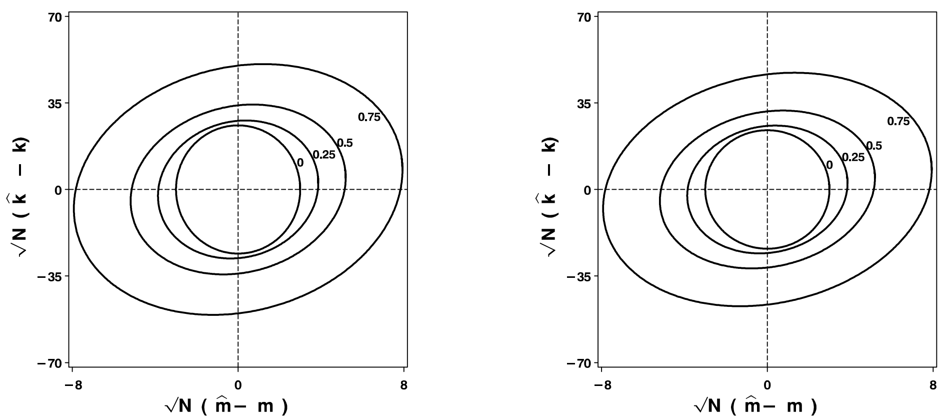
$c = 0$ $\qquad$ $c = 0.5$

Fig. 4 95% asymptotic bivariate normal confidence ellipses for $E\hat{\beta}_s$ ($\hat{\beta}_s = (\bar{x}, \widehat{c^X})^T$), centered at zero, for a NBD INAR(1) sample with $k = 2$, $m = 1$, $\alpha \in \{0, 0.25, 0.5, 0.75\}$ and $c \in \{0, 0.5\}$.



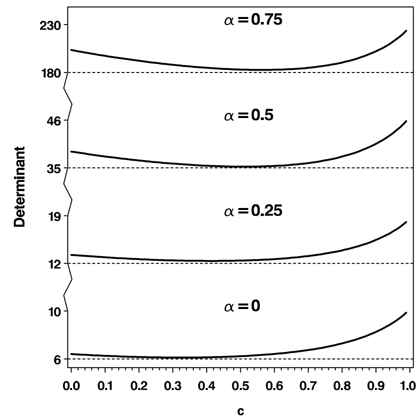
$\alpha = 0$ $\qquad$ $\alpha = 0.5$

Fig. 5 1000 simulated $\hat{\beta}_s - E\hat{\beta}_s$ ($\hat{\beta}_s = (\bar{x}, \widehat{c^X})^T$) with 95% asymptotic bivariate normal confidence ellipses for a NBD INAR(1) sample with $k = 2$, $m = 1$, $N = 1000$, $c = 0.5$ and $\alpha \in \{0, 0.5\}$.

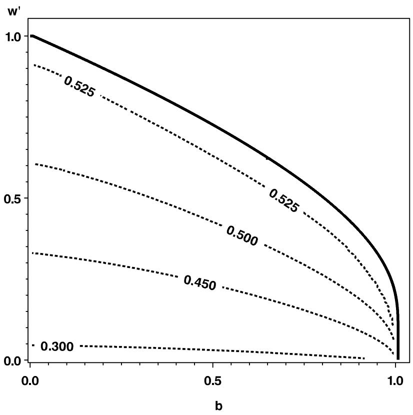


$c = 0$ $\qquad$ $c = 0.5$

Fig. 6 95% asymptotic bivariate normal confidence ellipses for $E\hat{\beta}$ ($\hat{\beta} = (\hat{k}_{PM}, \hat{m})$), centered at zero, for a NBD INAR(1) sample with $k = 2$, $m = 1$, $\alpha \in \{0, 0.25, 0.5, 0.75\}$.

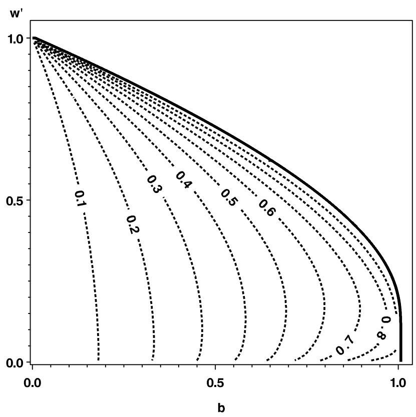


(a) $\det(\Sigma_\alpha(c))$ for $k = 0.5$, $m = 1$

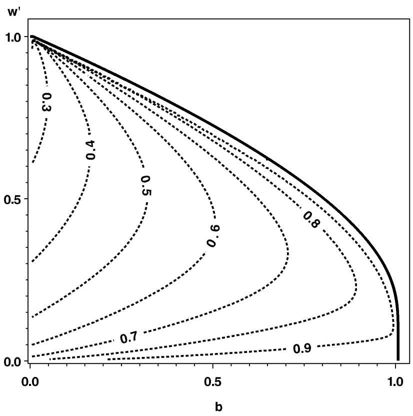


(b) $\det(\Sigma_{0.25}(c_{0.25})) \,/\, \det(\Sigma_0(c_o))$

Fig. 7 (a) $\det(\Sigma_\alpha(c))$ for $k = 0.5$, $m = 1$ and (b) $\det(\Sigma_{0.25}(c_{0.25})) \,/\, \det(\Sigma_0(c_o))$ where $c_{0.25}$ and c_o are values of $c \in (0, 1)$ that minimize $\det(\Sigma_{0.25}(c_{0.25}))$ and $\det(\Sigma_0(c_o))$ respectively.

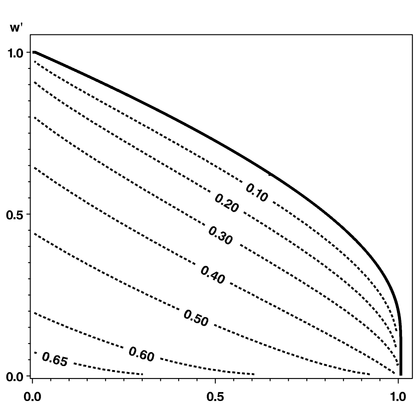


Values of c_o

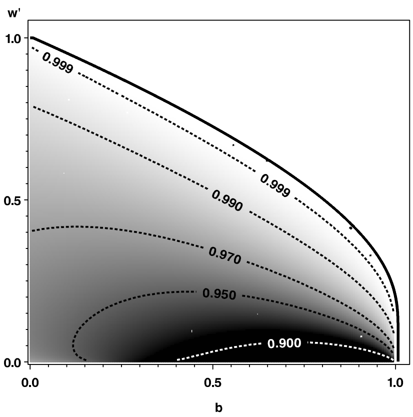


Values of $c_{0.75}$

Fig. 8 Values of c optimum, c_o and $c_{0.75}$, for power method estimators when estimating from a NBD INAR(1) sample with $\alpha = 0$ and $\alpha = 0.75$ respectively.



(a) $\mathrm{Corr}(\hat{k}_{PM}(c_{0.75}), \hat{m})$



(b) $\det(\Sigma_{0.75}(c_{0.75})) \,/\, \det(\Sigma_{0.75}(c_o))$
(Black shows efficiencies $\leqslant 0.92$)

Fig. 9 Estimating from NBD INAR(1) samples with $\alpha = 0.75$. (a) Correlation between $(\hat{k}_{PM}(c_{0.75}), \hat{m})$ where the power method estimator is computed at optimum c and (b) relative efficiency $\det(\Sigma_{0.75}(c_{0.75})) \,/\, \det(\Sigma_{0.75}(c_o))$.

Appendix B: Inadmissability of the MOM/ZTM

Proof of Theorem 1

Let m and k be fixed and set $p = 1 + m/k$. Note that $0 < k, m < \infty$ and $1 < p < \infty$.

i) Inadmissability of MOM. A Taylor expansion of $v_{PM}(c)$ in the neighbourhood of $c = 1$ gives

$$v_{PM}(c) = \frac{2k(k+1)p^2}{(p-1)^2} - \frac{8k(k+1)p^2}{3(p-1)}(1-c) + O\left((1-c)^2\right), \quad c \to 1. \tag{14}$$

In view of (5) this implies $v_{PM}(1) = v_{MOM}$. Additionally, the derivative of $v_{PM}(c)$ at $c = 1$ is

$$\left.\frac{\partial v_{PM}(c)}{\partial c}\right|_{c=1} = \frac{8k(k+1)p^2}{3(p-1)},$$

which is strictly positive for all k, m. Hence, there always exists c' such that $0 < c' < 1$ and $v_{PM}(c') < v_{PM}(1) = v_{MOM}$.

ii) Inadmissability of ZTM. A Taylor expansion of $v_{PM}(c)$ in the neighborhood of $c = 0$ gives

$$v_{PM}(c) = v_{PM}(0) + c \left.\frac{\partial v_{PM}(c)}{\partial c}\right|_{c=0} + O\left(c^2\right), \quad c \to 0\,. \tag{15}$$

Equation (6) and (7) directly imply that $v_{PM}(0) = v_{ZTM}$. The derivative of $v_{PM}(c)$ at $c = 0$ can be written as

$$\begin{aligned}\left.\frac{\partial v_{PM}(c)}{\partial c}\right|_{c=0} &= -\,2p(p-1)\frac{p^k\left[kp\log(p) - (p-1)(k+1)\right] + (p-1)(k+1) - k\log(p)}{(p\log(p) - p + 1)^3} \\ &= -\frac{2p(p-1)}{[h(p)]^3\log(p)}\sum_{j=2}^{\infty}\frac{[k\log(p)]^j}{j!}h_j(p),\end{aligned} \tag{16}$$

where $h(p) = p\log(p) - p + 1$ and $h_j(p) = [(j-1)p + 1]\log(p) - jp + j$. The infinite series in (16) is derived by a Taylor expansion of p^k (at $k = 0$) in the numerator. Lemma 1 implies that $h(p) > 0$ and $h_j(p) > 0$ for all $p > 1$ and all $j \geqslant 2$. All the terms in the infinite series in (16) are therefore positive for all k and p. This implies first, that the series is absolutely convergent for all k and p and second, that the derivative (16) is negative for all k and p. Hence, there always exists c'' such that $0 < c'' < 1$ and $v_{PM}(c'') < v_{PM}(0) = v_{ZTM}$.

iii) Inadmissability of MOM/ZTM. Let c' and c'' be particular values as above. Define

$$c_* = \begin{cases} c' & \text{if} \quad v_{ZTM} \geq v_{MOM} \\ c'' & \text{if} \quad v_{ZTM} < v_{MOM} \end{cases} \tag{17}$$

then we obviously have $v_{PM}(c_*) < \min\{v_{ZTM}, v_{MOM}\}$. $\square$

Lemma 1 *The functions*

$$h(t) = t\log(t) - t + 1 \;\; \text{and} \;\; h_j(t) = [(j-1)t + 1]\log(t) - jt + j$$

are positive for all $t > 1$ *and* $j \geqslant 2$.

Proof We have $h(1)=0$ and $h'(t)=\log(t)>0$ for all $t>1$, implying that $h(t)>0$ for all $t>1$. Similarly, for all $j\geqslant 2$ we have $h_j(1)=0$ and $h_j'(t)=(j-2)\log(t)+h(t)/t > 0$ for all $t>1$, implying that $h_j(t)>0$ for all $t>1$ and $j\geqslant 2$. □

Appendix C: Asymptotic variances of the PM estimators for the NBD INAR(1) process

The proof of Theorem 2 uses the statistical properties of the thinning operator and the form of the INAR(1) process. Recall that the thinning operation $\alpha \circ X$ is defined as

$$\alpha \circ X \stackrel{d}{=} \sum_{j=1}^{X} Z_j \quad \alpha \in (0,1),$$

where the Z_j are i.i.d. Bernoulli random variables with $P(Z_j=1) = \alpha$ and $P(Z_j = 0) = 1-\alpha$. From the definition of the thinning operator above it follows that

$$E[\alpha \circ X] = \alpha E[X] \quad \text{and} \quad E[f(X)(\alpha \circ X)] = \alpha E[Xf(X)],$$

where all expectations are assumed to be finite.

Since the INAR(1) process is a stationary process we have for any $s \neq t$

$$E[f(X_s)] = E[f(X_t)] = E[f(X_\pi)].$$

From the definition of the INAR(1) process we note that the dependence between any two random variables X_t and X_s from the same INAR(1) process with $s > t$ can be written as

$$X_s = \alpha^{s-t} \circ X_t + \sum_{j=0}^{s-t-1} \alpha^j \circ \varepsilon_{s-j}\,.$$

Finally we note that since $X_t \stackrel{d}{=} \alpha \circ X_{t-1} + \varepsilon_t$ the expected value of the errors are

$$E[\varepsilon_t] = E[X_t] - E[\alpha \circ X_{t-1}] = (1-\alpha)E[X_\pi].$$

This result and many more relationships between the moments of the ε_π and the moments of X_π can be obtained using the relationship

$$G_{X_\pi}(c) = G_{X_\pi}(1-\alpha+\alpha c)G_\varepsilon(c;\ \alpha).$$

Proof of Theorem 2.

Proof for $\mathbf{Var(\bar{X})}$

$$\begin{aligned}
Var(\bar{X}) &= Var\left(\frac{1}{N}\sum_{t=1}^{N} X_t\right) = \frac{1}{N^2}\left\{\sum_{t=1}^{N} Var(X_t) + \sum_{t\neq s}^{N} Cov(X_t, X_s)\right\} \\
&= \frac{1}{N^2}\left\{NVar[X_\pi] + 2\sum_{r=1}^{N-1}(N-r)\alpha^r Var[X_\pi]\right\} \\
&= \frac{1}{N} Var[X_\pi]\left\{1 + 2\sum_{r=1}^{N-1}\left(1-\frac{r}{N}\right)\alpha^r\right\} \\
\lim_{N\to\infty} Var(\bar{X}) &= \left(\frac{1+\alpha}{1-\alpha}\right) Var(X_\pi). \quad \square
\end{aligned}$$

Proof for $\mathbf{Var(\widehat{c^X})}$
Note that

$$\begin{aligned}
E\left[c^{X_t}c^{X_s}\right] &= E\left[c^{X_t+X_s}\right] = E\left[c^{X_t+\alpha^{s-t}\circ X_t}\right] E\left[c^{\sum_{j=0}^{s-t-1}\alpha^j\circ\varepsilon_{s-j}}\right] \\
&= G_{X_\pi}\left(c(1-\alpha^{s-t}+\alpha^{s-t}c)\right)\prod_{j=0}^{s-t-1} G_\varepsilon(1-\alpha^j+\alpha^j c;\ \alpha) \\
&= G_{X_\pi}\left(c(1-\alpha^{s-t}+\alpha^{s-t}c)\right)\prod_{j=0}^{s-t-1} \frac{G_{X_\pi}\left(1-\alpha^j+\alpha^j c\right)}{G_{X_\pi}\left(1-\alpha+\alpha(1-\alpha^j+\alpha^j c)\right)} \\
&= G_{X_\pi}\left(c(1-\alpha^{s-t}+\alpha^{s-t}c)\right)\frac{G_{X_\pi}(c)}{G_{X_\pi}\left(1-\alpha^{s-t}+\alpha^{s-t}c\right)} \\
&= G_{X_\pi}\left(c(1-\alpha^{s-t}+\alpha^{s-t}c)\right) G_\varepsilon(c;\ \alpha^{s-t})
\end{aligned}$$

therefore

$$\begin{aligned}
Var(\widehat{c^X}) &= Var\left(\frac{1}{N}\sum_{t=1}^{N} c^{X_t}\right) = \frac{1}{N^2}\left\{\sum_{t=1}^{N} Var(c^{X_t}) + \sum_{t\neq s}^{N} Cov(c^{X_t}, c^{X_s})\right\} \\
&= \frac{1}{N^2}\left\{NVar\left[c^{X_\pi}\right] + \sum_{t\neq s}^{N}\left(E\left[c^{X_t}c^{X_s}\right] - E\left[c^{X_t}\right]E\left[c^{X_s}\right]\right)\right\} \\
&= \frac{1}{N^2}\left\{NVar\left[c^{X_\pi}\right] + 2\sum_{r=1}^{N-1}(N-r)\left(G_{X_\pi}\left(c(1-\alpha^r+\alpha^r c)\right)G_\varepsilon(c;\ \alpha^r) - G^2_{X_\pi}(c)\right)\right\}
\end{aligned}$$

$$\lim_{N\to\infty} NVar(\widehat{c^X}) = Var\left(c^{X_\pi}\right) + 2\lim_{N\to\infty}\sum_{r=1}^{N-1}\left(1-\frac{r}{N}\right)\left\{G_{X_\pi}\left(c\left[1-\alpha^r+\alpha^r c\right]\right)G_\varepsilon(c;\alpha^r) - G^2_{X_\pi}(c)\right\}. \quad \square$$

Proof for $\mathbf{Cov}(\bar{\mathbf{X}}, \widehat{\mathbf{c}^{\mathbf{X}}})$
Note that for $t < s$

$$\begin{aligned}
E\left[X_s c^{X_t}\right] &= E\left[\left(\alpha^{s-t}\circ X_t + \sum_{j=0}^{s-t-1}\alpha^j\circ\varepsilon_{s-j}\right)c^{X_t}\right] \\
&= E\left[c^{X_t}\left(\alpha^{s-t}\circ X_t\right)\right] + E\left[c^{X_t}\right]\sum_{j=0}^{s-t-1}E\left[\alpha^j\circ\varepsilon_{s-j}\right] \\
&= \alpha^{s-t}E\left[c^{X_t}X_t\right] + (1-\alpha^{s-t})E\left[c^{X_t}\right]E\left[X_t\right] \\
&= \alpha^{s-t}E\left[c^{X_\pi}X_\pi\right] + (1-\alpha^{s-t})E\left[c^{X_\pi}\right]E\left[X_\pi\right], \\
Cov\left(X_s, c^{X_t}\right) &= \alpha^{s-t}Cov\left(X_\pi, c^{X_\pi}\right)
\end{aligned}$$

and

$$\begin{aligned}
E\left[X_t c^{X_s}\right] &= E\left[X_t c^{\alpha^{s-t}\circ X_t}\right]E\left[c^{\sum_{j=0}^{s-t-1}\alpha^j\circ\varepsilon_{s-j}}\right] \\
&= E\left[X_t\left(1-\alpha^{s-t}+\alpha^{s-t}c\right)^{X_t}\right]\prod_{j=0}^{s-t-1}G_{\varepsilon_{s-j}}(1-\alpha^j+\alpha^j c) \\
&= E\left[X_\pi\left(1-\alpha^{s-t}+\alpha^{s-t}c\right)^{X_\pi}\right]G_\varepsilon(c;\ \alpha^{s-t}) \\
Cov\left(X_t, c^{X_s}\right) &= E\left[X_\pi\left(1-\alpha^{s-t}+\alpha^{s-t}c\right)^{X_\pi}\right]G_\varepsilon(c;\ \alpha^{s-t}) - E\left[X_\pi\right]E\left[c^{X_\pi}\right],
\end{aligned}$$

therefore

$$\begin{aligned}
Cov\left(\bar{X}, \widehat{c^X}\right) &= \frac{1}{N^2}\left\{\sum_{t=1}^{N}Cov\left(X_t, c^{X_t}\right) + \sum_{t<s}Cov(X_t, c^{X_s}) + \sum_{t>s}Cov(X_t, c^{X_s})\right\} \\
&= \frac{1}{N^2}\left\{NCov\left(X_\pi, c^{X_\pi}\right) + \sum_{r=1}^{N-1}(N-r)\alpha^r Cov\left(X_\pi, c^{X_\pi}\right)\right. \\
&\left. + \sum_{r=1}^{N-1}(N-r)\left(E\left[X_\pi\left(1-\alpha^{s-t}+\alpha^{s-t}c\right)^{X_\pi}\right]G_\varepsilon(c;\ \alpha^{s-t}) - E\left[X_\pi\right]E\left[c^{X_\pi}\right]\right)\right\}
\end{aligned}$$

$$\begin{aligned}
\lim_{N\to\infty}NCov\left(\bar{X}, \widehat{c^X}\right) &= \frac{1}{1-\alpha}Cov\left(X_\pi, c^{X_\pi}\right) \\
&+ \lim_{N\to\infty}\sum_{r=1}^{N-1}\left(1-\frac{r}{N}\right)\left(E\left[X_\pi\left(1-\alpha^r+\alpha^r c\right)^{X_\pi}\right]G_\varepsilon(c;\ \alpha^r) - E\left[X_\pi\right]E\left[c^{X_\pi}\right]\right). \quad \square
\end{aligned}$$

References

1. M. A. Al-Osh and A. A. Alzaid. First-order integer-valued autoregressive (INAR(1)) process. *Journal of Time Series Analysis*, 8(3):261–275, 1987.
2. F. J. Anscombe. Sampling theory of the negative binomial and logarithmic series distributions. *Biometrika*, 37:358–382, 1950.
3. P. J. Brockwell and R. A. Davis. *Introduction to time series and forecasting*. Springer-Verlag, New York, second edition, 2002.
4. A.S.C. Ehrenberg. *Repeat-Buying: Facts, Theory and Applications*. Charles Griffin & Company Ltd., London, 1988.
5. R. A. Fisher. The negative binomial distribution. *Ann. Eugenics*, 11:182–187, 1941.
6. J. B. S. Haldane. The fitting of binomial distributions. *Ann. Eugenics*, 11:179–181, 1941.
7. N. L. Johnson, S. Kotz, and A. W. Kemp. *Univariate discrete distributions*. John Wiley & Sons Inc., New York, second edition, 1992.
8. E. McKenzie. Autoregressive moving-average processes with negative-binomial and geometric marginal distributions. *Advances in Applied Probability*, 18(3):679–705, 1986.
9. E. McKenzie. Some ARMA models for dependent sequences of Poisson counts. *Advances in Applied Probability*, 20(4):822–835, 1988.
10. V. Savani and A. Zhigljavsky. Efficient estimation of parameters of the negative binomial distribution. *Communications in Statistics: Theory and Methods*, 35(5):1–17, 2006.
11. R. J. Serfling. *Approximation theorems of mathematical statistics*. John Wiley & Sons Inc., New York, 1980.