GMM tests for the Katz family of distributions

Yue Fang *

*Department of Decision Sciences, Lundquist College of Business, University of Oregon, Eugene,
OR 97403-1208, USA*

Abstract

Generalized method of moments (GMM) is used to develop tests for discriminating discrete distributions among the two-parameter family of Katz distributions. Relationships involving moments are exploited to obtain identifying and over-identifying restrictions. The asymptotic relative efficiencies of tests based on GMM are analyzed using the local power approach and the approximate Bahadur efficiency. The paper also gives results of Monte Carlo experiments designed to check the validity of the theoretical findings and to shed light on the small sample properties of the proposed tests. Extensions of the results to compound Poisson alternative hypotheses are discussed. © 2001 Elsevier Science B.V. All rights reserved.

1. Introduction

One of the most prominent families of discrete distributions whose successive probabilities satisfy first-order recurrence relations has been formulated by Katz (1965). The Katz family of distributions cover a wide spectrum including binomial, negative binomial, and Poisson distributions. It has been used as a basis of developing more general families of distributions such as distributions defined by a discrete analogue to the Pearson system of continuous distributions studied in Ord (1967a, b, 1972), and the extended Katz family of Gurland and Tripathi (1975) and Tripathi and Gurland (1977, 1979). These distributions often serve as embryonic forms of data generating distributions and provide an ideal benchmark in many applications such as econometric modeling, industrial quality control, risk and insurance, and sampling theory, among others.

* Tel.: +1-541-346-3265.

E-mail address: yfang@darkwing.uoregon.edu (Y. Fang).

A major motivation of Katz's (1965) work was the problem of discriminating among binomial, negative binomial, and Poisson distributions when a given set of data is known to come from one or other of them. More specifically, Katz suggested the use of a test based on the first two sample moments as a discriminating test statistic for testing the null hypothesis of equi-dispersion (the Poisson distribution) against the alternative hypothesis of under- or over-dispersions (the binomial or negative binomial distribution).

The aim of this paper is to investigate discriminating tests for discrete distributions in the generalized method of moments (GMM) framework developed by Hansen (1982). The paper analyzes formally the asymptotic and finite-sample properties of discriminating tests which are based on a finite set of moment restrictions for the null hypothesis of a Poisson distribution. In particular, it is concerned with the potential for efficiency gains from using additional moment restrictions other than the first two moments utilized in the Katz test. The results of the GMM tests are also compared to the performance of the Poisson index of dispersion, a commonly used statistic for testing hypothesis concerning the Poisson distribution, attributed to Fisher. The final section of the paper extends the discussion to compound Poisson distributions. We show that since compound Poisson distributions exhibit over-dispersion, the results of the negative binomial hold qualitatively for compound Poisson alternatives.

The paper is organized as follows. Section 2 introduces briefly the Katz family of distributions and GMM estimators. Section 3 discusses identifying and over-identifying moment restrictions. Section 4 analyzes the asymptotic relative efficiency of the tests. Section 5 presents finite-sample power results of tests against binomial and negative binomial alternatives. Section 6 extends the results to compound Poisson distributions. Section 7 concludes. All proofs are given in the Appendix.

2. The Katz family and GMM estimators

2.1. The Katz family

For the Katz family of distributions, the recurrence relation for probabilities can be written as

$$\frac{P_{j+1}}{P_j} = \frac{\alpha + \beta j}{1 + j}, \quad j = 0, 1, 2, \ldots, \tag{1}$$

where $\alpha > 0$ and $\beta < 1$. It is understood that if $\alpha + \beta j < 0$ then $P_{j+i} = 0$ for all $i > 0$. The probability generating function for distributions defined in (1), $g(t)$, satisfies the equation (Johnson et al., 1992):

$$\frac{\mathrm{d} \log g(t)}{\mathrm{d}t} = \frac{\alpha}{1 - \beta t}$$

with $g(1) = 1$. Hence,

$$g(t) = \begin{cases} e^{\alpha(t-1)} & \text{if } \beta = 0, \\ [(1 - \beta t)/(1 - \beta)]^{-\alpha/\beta} & \text{otherwise.} \end{cases}$$

The moments about the origin are

$$\mu_{r+1} \equiv E(X^{r+1}) = \sum_{j=0}^{r} \binom{r}{j} (\alpha \mu_j + \beta \mu_{j+1}). \tag{2}$$

In particular, the mean $\mu_1 = \alpha/(1 - \beta)$ and the variance $\sigma^2 \equiv \mu_2 - (\mu_1)^2 = \alpha/(1 - \beta)^2$.

It can be shown that $\beta < 0$, $0 < \beta < 1$, and $\beta = 0$ give rise to the binomial distribution ($B(N, P)$), negative binomial distribution ($NB(k, p)$) and Poisson distribution ($P(\lambda)$), respectively, with parameters $N = -\alpha/\beta$, $P = \beta/(\beta - 1)$, $k = \alpha/\beta$, $p = \beta$, and $\lambda = \alpha$. The area of the (α, β) plane occupied by these three distributions can be found, for example, in Katz (1965). Distributions in the Katz family have the property of being equi-, under-, or over-dispersed. The Poisson distribution has $\sigma^2/\mu_1 = 1$ and exhibits equi-dispersion. The binomial and negative binomial distributions are under-dispersed ($\sigma^2/\mu_1 < 1$) and over-dispersed ($\sigma^2/\mu_1 > 1$), respectively. For many discrete distributions, in particular binomial and negative binomial distributions, violation of the variance assumption is a sufficient condition for a violation of the Poisson assumption.

2.2. GMM estimators

Let $(x_1, x_2, \ldots, x_n)$ be a sample from a discrete population with a parameter λ. Suppose that we are interested in estimating λ based on the first q moment restrictions,

$$E[f(x_i, \lambda)] = 0, \tag{3}$$

where $f(x_i, \lambda) = (x_i - \mu_1, x_i^2 - \mu_2, \ldots, x_i^q - \mu_q)'$. The jth moment about the origin of the Poisson distribution, μ_j, is given by (2) with $\alpha = \lambda$ and $\beta = 0$. We note that if one multiplies both sides of (3) by a $q \times q$ non-singular matrix Ξ, one obtains an equivalent set of moment restrictions which are based on different model parameterizations. For example, if Ξ is taken to be the lower triangular matrix with 1 as diagonal elements and Stirling numbers of the first kind as the lower triangular elements, then the moment restrictions $E[\Xi f(x_i, \lambda)] = 0$ are based on the factorial moments (see, for example, Chapter 3 of Stuart and Ord, 1987).

System (3) is overdetermined if $q > 1$. One way to reconcile the conflicting estimates that will emerge from this overdetermined system is to minimize the quadratic form

$$Q_n(\lambda) = f_n(\lambda)' W^{-1} f_n(\lambda), \tag{4}$$

where $f_n(\lambda) \equiv n^{-1} \sum_{i=1}^{n} f(x_i, \lambda)$ and W is a positive definite weighting matrix which may be a function of data. The estimators defined by choosing λ to minimize (4) are minimum distance estimators. Under certain regularity conditions on W the minimum

distance estimator is consistent. See Malinvaud (1970) and Amemiya (1985) for discussions on minimum distance estimators. See also Gerlein and Pincus (1978) for an approach based on the concept of w-divergence introduced by Kagan (1975).

The Hansen's GMM estimate of λ, $\hat{\lambda}$, is the value of λ that minimizes (4) by taking $W \equiv V_n$, where V_n is a consistent estimator of $V = \lim_{n \to \infty} \text{Var}[n^{1/2} f_n(\lambda)]$. The use of this weighted criterion in which the weights are inversely proportional to the variances of the moments is based on the logic that motivates generalized least squares (GLS).

Hansen (1982) showed that the optimal weighting matrix W in (4) turns out to be given by V_n. It can be verified that the (i,j)th element of V is $(\mu_{i+j} - \mu_i \mu_j)$. Under a model of Poisson $\hat{\lambda}$ is $\sqrt{n}$-consistent and asymptotically normally distributed. Furthermore, since $\bar{x}$ is sufficient for λ one has the following result.

Proposition 2.1. *For $q \geqslant 1$, $n^{1/2}(\hat{\lambda} - \lambda) \xrightarrow{d} \text{N}(0, \lambda)$.*

3. Testing over-identifying restrictions

When the number of moment restrictions in (3) exceeds the number of parameters ($q > 1$), the model is over-identified. In the over-identified case, Hansen (1982) suggested a test of whether all of the sample moments represented by $f_n(\hat{\lambda})$ are as close to zero as would be expected if the corresponding population moments (3) were truly zero. Applying Hansen's results, one can show that if a Poisson distribution is the true data generator and $Q_n(\lambda)$ is evaluated at the GMM estimate $\hat{\lambda}$, then

$$J_n(q) \equiv n f_n(\hat{\lambda})' V_n^{-1}(\hat{\lambda}) f_n(\hat{\lambda}) \xrightarrow{d} \chi_{q-1}^2. \tag{5}$$

Hansen's χ^2 test, $J_n(q)$, can be used to test the null hypothesis that the first q moments of the population coincide with their Poisson counterparts for some λ, that is,

$$\text{H}_0 : \text{E}[f(x_i, \lambda)] = 0 \tag{6}$$

against an alternative hypothesis

$$\text{H}_a : \text{E}[f(x_i, \lambda)] \neq 0. \tag{7}$$

To analyze the properties of $J_n(q)$ we follow Sowell (1996) to decompose the population moment restrictions in the null hypothesis (6) into

$$\text{H}_0^{\text{I}} : D V^{-1/2} \text{E}[f(x_i, \lambda)] = 0$$

and

$$\text{H}_0^{\text{O}} : (I_{q \times q} - D) V^{-1/2} \text{E}[f(x_i, \lambda)] = 0,$$

where $D \equiv M(M'M)^{-1}M'$ and $M \equiv V^{-1/2} \text{E}(F)$ with $F \equiv \text{d} f_n(\lambda)/\text{d}\lambda$. The moment restriction in H_0^{I} is the identifying restriction for λ. It represents the part of the moment restrictions which actually goes into parameter estimation. In fact, the first-order conditions associated with the minimization of (4) with $W = V_n(\hat{\lambda})$ are

$$\frac{\partial Q_n(\hat{\lambda})}{\partial \lambda} = F(\hat{\lambda})' V_n^{-1}(\hat{\lambda}) f_n(\hat{\lambda}) = [V_n^{-1/2}(\hat{\lambda}) F(\hat{\lambda})]' V_n^{-1/2}(\hat{\lambda}) f_n(\hat{\lambda}) = 0. \tag{8}$$

This characterization of the GMM estimator yields an interesting interpretation that $V_n^{-1/2}(\hat{\lambda})f_n(\hat{\lambda})$ lies in the null space of $[V_n^{-1/2}(\hat{\lambda})F(\hat{\lambda})]'$.

Sowell (1996) has shown that the information about λ in (8) is equivalent to the information in

$$\hat{D}V_n^{-1/2}(\hat{\lambda})f_n(\hat{\lambda}) = 0$$

with $\hat{D} \equiv \hat{M}(\hat{M}'\hat{M})^{-1}\hat{M}'$ and $\hat{M} \equiv V_n^{-1/2}\hat{F}$, which is the sample analog to H_0^I (see also Hall, 2001). Therefore, although one begins with (3), GMM estimation is actually based on information in H_0^I. Since the projection of $V_n^{-1/2}(\hat{\lambda})f_n(\hat{\lambda})$ on to the column space of $[V_n^{-1/2}(\hat{\lambda})F(\hat{\lambda})]'$ is zero, the rank of $\hat{D}$ is 1. Hence restrictions in (6) set only 1 linear combination of the q vector $E[f(x_i, \lambda)]$ to zero (see Proposition 3.3 for the identifying moment restriction).

On the other hand, moment restrictions in H_0^O are over-identifying restrictions. Unlike the moment restriction in H_0^I which is used to obtain the GMM estimate of λ, the moment restrictions in H_0^O will be satisfied by the Poisson distribution but not by the binomial or negative binomial distribution. In our case, the over-identifying restrictions contain $(q - 1)$ linear combinations of $E[f(x_i, \lambda)] = 0$. Since the over-identifying restrictions are not imposed it is possible to test whether they hold in the sample. Under the decomposition described above it is clear that $H_0 = H_0^I \cap H_0^O$.

Results stated in the following two propositions are useful in determining identifying and over-identifying restrictions specified in H_0^I and H_0^O.

Proposition 3.1. *Under* H_0, $(M'M)^{-1} = \lambda$.

Proposition 3.2. *Under* H_0, *the* $q \times q$ *matrix* $(FF')V^{-1}$ *has rank 1 and*

$$(FF')V^{-1} = [-(M'M)F, \mathcal{O}], \tag{9}$$

where $\mathcal{O}$ *is a* $q \times (q - 1)$ *zero matrix*.

The identifying and over-identifying moment restrictions are given in Propositions 3.3 and 3.4, respectively.

Proposition 3.3. *For any* $q \geqslant 1$, *the identifying moment restriction in* H_0^I *is* $E(x_i) = \mu_1$.

Proposition 3.4. *The over-identifying restrictions are* $(I_{q \times q} + [F, \mathcal{O}])E[f(x_i, \lambda)] = 0$, *where* $\mathcal{O}$ *is a* $q \times (q - 1)$ *zero matrix*.

From Propositions 3.3 and 3.4, both identifying and over-identifying moment restrictions are formulated by $E[f(x_i, \lambda)]$, which, by definition, is based on the first q moments about the origin. Note that there are $q - 1$ over-identifying restrictions since the first element of vector F is -1. If we assume that the identifying restriction is satisfied by the sample the over-identifying restrictions simplify to $E(x_i^j) = \mu_j$, for $j = 2, 3, \ldots, q$ (Proposition 3.4).

Before we discuss the asymptotic power of $J_n(q)$, two simple examples are appropriate to demonstrate the method. The first example, the case in which $q = 2$, yields a GMM version of the Katz's (1965) test. The second one, utilizing the first three moments ($q = 3$), serves as an alternative test, which will be used in the power comparison analysis in later sections.

Example 1 ($J_n(2)$ *and the Katz test*). When $q = 2$, $f_n(\lambda) = (\bar{x} - \lambda, m_2 - (\lambda + \lambda^2))'$, where $m_2 = n^{-1} \sum x_i^2$. The variance matrix is

$$V = \begin{bmatrix} \lambda & 2\lambda^2 + \lambda \\ 2\lambda^2 + \lambda & 4\lambda^3 + 6\lambda^2 + \lambda \end{bmatrix}.$$

It can be shown that the GMM test statistic is

$$J_n(2) = \frac{n}{2} \left(\frac{s^2 - \bar{x}}{\bar{x}} \right)^2, \tag{10}$$

where $s^2 = \sum(x_i - \bar{x})^2/n$. $J_n(2)$ follows asymptotically a χ_1^2 distribution by (5).

Note that $J_n(2)$ is equivalent to the test proposed by Katz (1965), in which the square root of $J_n(2)$, $\sqrt{n/2}((s^2 - \bar{x})/\bar{x})$, is distributed asymptotically as the standard normal distribution. Therefore, we can analyze the Katz test using the results established above.

By Propositions 3.3 and 3.4, the identifying and over-identifying restrictions in H_0^I and H_0^O are $(E(x_i) - \mu_1) = 0$ and $(E(x_i^2) - \mu_2) = 0$, respectively. The interpretation of the decomposition result is that in the Katz's test or in the $J_n(2)$ statistic, the first-moment restriction, which is automatically satisfied by the sample, is used in estimating the population mean, resulting the MME, which is also the MLE estimator. The second-moment restriction is satisfied by the Poisson distribution but not by the binomial or negative binomial distribution. In fact, the second-moment restriction for binomial and negative binomial distributions is

$$0 = E(x_i^2) - (E(x_i) + [E(x_i)]^2) = \begin{cases} -NP^2 & \text{binomial,} \\ kp^2 & \text{negative binomial.} \end{cases} \tag{11}$$

It is apparent that $J_n(2)$ has power against violation of the over-identifying restriction for any departures from the Poisson distribution. We will have formal discussion on the asymptotic power of $J_n(2)$ in Section 4 and on its finite-sample power properties in Section 5.

Example 2 ($J_n(3)$). If $q = 3$, $f_n(\lambda) = (\bar{x} - \lambda, s^2 - (\lambda + \lambda^2), m_3 - (\lambda + 3\lambda^2 + \lambda^3))'$, where $m_3 = n^{-1} \sum x_i^3$. It can be verified that

$$V = \begin{bmatrix} \lambda & \lambda + 2\lambda^2 & \lambda + 6\lambda^2 + 3\lambda^3 \\ \lambda + 2\lambda^2 & \lambda + 6\lambda^2 + 4\lambda^3 & \lambda + 14\lambda^2 + 21\lambda^3 + 6\lambda^4 \\ \lambda + 6\lambda^2 + 3\lambda^3 & \lambda + 14\lambda^2 + 21\lambda^3 + 6\lambda^4 & \lambda + 30\lambda^2 + 84\lambda^3 + 54\lambda^4 + 9\lambda^5 \end{bmatrix}.$$

The $J_n(3)$ test statistic is given by

$$J_n(3) = \hat{v}_{22}(s^2 - \bar{x})^2 + 2\hat{v}_{23}(s^2 - \bar{x})(m_3' - \bar{x}) + \hat{v}_{33}(m_3' - \bar{x})^2, \tag{12}$$

where $\hat{v}_{ij}$ is the (i,j)th element of $\hat{V}_n^{-1}$, and $m_3' = n^{-1}\sum(x_i - \bar{x})^3$. The direct but tedious calculation of V^{-1} yields that $v_{22} = 2^{-1}\lambda^{-3}(3 + 7\lambda + 3\lambda^2)$, $v_{23} = -2^{-1}\lambda^{-3}(1+\lambda)$, and $v_{33} = 6^{-1}\lambda^{-3}$. Therefore, $\hat{v}_{22} = 2^{-1}\bar{x}^{-3}(3 + 7\bar{x} + 3\bar{x}^2)$, $\hat{v}_{23} = -2^{-1}\bar{x}^{-3}(1 + \bar{x})$, and $\hat{v}_{33} = 6^{-1}\bar{x}^{-3}$. (The expression of V^{-1}, and hence V_n^{-1}, may be obtained from *Mathematica* (see, for example, Wolfram, 1999 for more information on the software), which provides a useful tool to derive the required elements of V^{-1}, especially for V with high values of q.)

Again, the identifying restriction is $E(x_i) = \mu_1$. In contrast to the case of $q = 2$, the over-identifying restrictions consist of two equations: $E(x_i^2) = \mu_2$ and $E(x_i^3) = \mu_3$ (Proposition 3.4). The two over-identifying moment restrictions for the binomial and negative binomial alternative hypotheses are given by (11) and the following equation:

$$0 = E(x_i^3) - [E(x_i) + 3[E(x_i)]^2 + [E(x_i)]^3]$$

$$= \begin{cases} -3NP^2 + 2NP^3 - 3N^2P^3 & \text{binomial,} \\ 3kp^2 + 2kp^3 - 3k^2p^3 & \text{negative binomial.} \end{cases} \tag{13}$$

4. Asymptotic power

This section first investigates the asymptotic power of $J_n(q)$ using a local power analysis. To this end, we introduce the following sequences of local alternatives to H_0^I and H_0^O:

$$H_A^I : DV^{-1/2}E[f(x_i, \lambda)] = n^{-1/2}v_I \tag{14}$$

and

$$H_A^O : (I_{q\times q} - D)V^{-1/2}E[f(x_i, \lambda)] = n^{-1/2}v_O, \tag{15}$$

in which $v_I \neq 0$ and $v_O \neq 0$. Since it is always possible to decompose $V^{-1/2}E[f(x_i, \lambda)]$ into

$$V^{-1/2}E[f(x_i, \lambda)] = DV^{-1/2}E[f(x_i, \lambda)] + (I_{q\times q} - D)V^{-1/2}E[f(x_i, \lambda)],$$

H_0^I and H_0^O translate directly into sequences of local alternatives to $V^{-1/2}E[f(x_i, \lambda)] = 0$, which is equivalent to H_0: $E[f(x_i, \lambda)] = 0$. The following proposition is adapted from Hall (1999), showing that for local alternatives given in (14) and (15), $J_n(q)$ follows asymptotically a non central χ^2 distribution.

Proposition 4.1. *If the data satisfy* $H_0^I \cap H_A^O$, $J_n(q) \sim \chi_{q-1}^2(v_O'v_O)$.

Proposition 4.1 implies that $J_n(q)$ has power against the alternative H_A^O whenever $v_O'v_O > 0$. In Proposition 4.1, we assume that the identifying restriction is satisfied. Note that $J_n(q)$ has the same asymptotic distribution χ_{q-1}^2 under H_0^O regardless of whether H_0^I or H_A^I holds (Newey, 1985; Hall, 1999). Hence, $J_n(q)$ has no power to discriminate between H_0^I or H_A^I. This implies that the rejection of H_0 indicates the invalidity of H_0^O.

Proposition 4.2. *If the data satisfy* $H_0^I \cap H_A^O$, $v_O'v_O = nE[f(x_i, \lambda)']V^{-1}E[f(x_i, \lambda)]$.

An alternative approach to analyze the power of tests is to study the approximate slope of a test, which has been proposed by Bahadur (1960) and further studied by Bahadur (1967) and Geweke (1981), among others. Although the Bahadur's efficiency is asymptotically equivalent to the local power method under certain regularity conditions (Proposition 4.3; see also Sievers, 1969; Wieand, 1976; Groeneboom and Oosterhoff, 1981, who compared the Bahadur efficiency measure with other methods), it has an intuitively appealing aspect in understanding the role of $E[f(x_i, \lambda)']V^{-1}E[f(x_i, \lambda)]$ in determining the asymptotic power of the test. According to Bahadur (1960) the approximate slope of a test is defined to be the rate at which the logarithm of the asymptotic marginal significance level of the test decreases as sample size increases. Geweke (1981) has shown that if the test statistic's limiting distribution under the null hypothesis is a χ^2 distribution, then the approximate slope of the test equals the probability limit of the statistic divided by the sample size n. Let c_q be the approximate slope of $J_n(q)$. Applying Geweke's result, we have

Proposition 4.3. $c_q = E[f(x_i, \lambda)']V^{-1}E[f(x_i, \lambda)]$.

For any given q, it is straightforward to apply Proposition 4.3 to calculate c_q under different alternative hypotheses. For example, consider once again the case $q = 2$ discussed in Example 1 of Section 3. When $q = 2$, we have that the approximate slope

$$
c_2 = (0, E(x_i^2) - (E(x_i) + [E(x_i)]^2)) \begin{bmatrix} \frac{1+3E(x_i)+4[E(x_i)]^2}{2[E(x_i)]^2} & -\frac{1+2E(x_i)}{2[E(x_i)]^2} \\ -\frac{1+2E(x_i)}{2[E(x_i)]^2} & \frac{1}{2[E(x_i)]^2} \end{bmatrix}
$$

$$
\times \begin{pmatrix} 0 \\ E(x_i^2) - (E(x_i) + [E(x_i)]^2) \end{pmatrix}
$$

$$
= \frac{[E(x_i^2) - \{E(x_i) + [E(x_i)]^2\}]^2}{2[E(x_i)]^2} = \begin{cases} \frac{1}{2}NP^3 & \text{binomial,} \\ \frac{1}{2}kp^3 & \text{negative binomial.} \end{cases} \tag{16}
$$

Note that the approximate slopes of $J_n(2)$ for binomial and negative binomial distributions are identical if $NP^3 = kp^3$.

Approximate slopes can also be used in making comparisons across different q. For example, under the binomial alternative, c_3 is given by the following equation:

$$c_3 = \frac{18P - 12P^2 + 4P^3 + 57P^2N - 24P^3N + 54P^3N^2 - 12P^4N^2 + 18P^4N^3}{6N}.$$

It can be shown that $c_3 \geqslant c_2$ on the parameter space of N and P. The difference between c_3 and c_2 is less significant when both N and P are small but becomes enormous as either N or P increases.

Theoretically speaking, any violations in either one of the two over-identifying restrictions of (11) and (13) can be tested by $J_n(3)$. One can argue that $J_n(3)$ should be more asymptotically efficient than $J_n(2)$ since an additional moment condition is used in $J_n(3)$. The inequality $c_3 \geqslant c_2$ indicates that including an additional moment restriction will not reduce asymptotic efficiency. However, one must be cautious since asymptotic results may not be valid in finite samples (see finite-sample properties of tests in Section 5).

5. Finite-sample results

5.1. The simulation setup

In this section we will compare the finite-sample properties of GMM tests for discriminating discrete distributions via Monte Carlo simulations under the Poisson null and the binomial and negative binomial alternative hypotheses. Other alternatives including compound Poisson and a zero truncated geometric distributions are examined in Section 6.

Since the use of GMM tests with $q > 3$ is computationally inconvenient and our results (not reported here for brevity) indicate that there is no major gain by using GMM tests with high values of q for hypotheses investigated in this study, we will focus only on $J_n(2)$ and $J_n(3)$. All simulations are based on 3000 replications. The nominal significance level is taken to be 1%, 5% and 10%.

For comparison, we also report the results of Monte Carlo experiments performed for the Poisson index of dispersion $\mathscr{L}$, which is defined as

$$\mathscr{L} = \sum (x_i - \bar{x})^2 / \bar{x}.$$

If the x_i's follow a Poisson distribution then $\mathscr{L}$ is approximately χ^2 with $n-1$ degrees of freedom (Fisher, 1950; Cochran, 1936, 1954).

5.2. The size

To gauge the quality of asymptotic approximations in Section 3, we perform simulation experiments for $J_n(2)$, $J_n(3)$ and $\mathscr{L}$ under the Poisson null hypothesis with various values of λ and sample sizes. Results, which are presented in Table 1, show that for 5% and 10% significance levels, the empirical sizes of $J_n(2)$ and $J_n(3)$ are close to

Table 1
Sizes of tests[a]

Sample size	λ	Size of 1% test			Size of 5% test			Size of 10% test		
		$J_n(2)$	$J_n(3)$	$\mathscr{L}$	$J_n(2)$	$J_n(3)$	$\mathscr{L}$	$J_n(2)$	$J_n(3)$	$\mathscr{L}$
10	1	0.008	0.011	0.008	0.026	0.027	0.047	0.042	0.046	0.072
	2.5	0.016	0.013	0.016	0.029	0.030	0.056	0.060	0.055	0.113
	5	0.009	0.022	0.009	0.022	0.039	0.042	0.062	0.050	0.097
	10	0.011	0.016	0.010	0.019	0.034	0.040	0.055	0.046	0.092
	20	0.007	0.024	0.007	0.029	0.044	0.047	0.063	0.062	0.110
	30	0.012	0.023	0.010	0.031	0.040	0.052	0.067	0.063	0.102
20	1	0.011	0.010	0.011	0.048	0.043	0.053	0.089	0.076	0.090
	2.5	0.014	0.018	0.014	0.030	0.036	0.051	0.071	0.062	0.098
	5	0.008	0.019	0.008	0.033	0.037	0.045	0.074	0.060	0.092
	10	0.007	0.017	0.007	0.041	0.046	0.054	0.085	0.068	0.104
	20	0.012	0.022	0.012	0.044	0.052	0.056	0.102	0.076	0.107
	30	0.006	0.025	0.007	0.035	0.042	0.037	0.068	0.077	0.081
50	1	0.005	0.013	0.007	0.041	0.040	0.051	0.085	0.075	0.099
	2.5	0.009	0.019	0.011	0.044	0.048	0.045	0.092	0.085	0.092
	5	0.008	0.021	0.009	0.037	0.054	0.047	0.081	0.083	0.096
	10	0.008	0.013	0.007	0.041	0.040	0.045	0.099	0.073	0.089
	20	0.011	0.014	0.012	0.052	0.051	0.055	0.101	0.082	0.109
	30	0.013	0.023	0.012	0.045	0.045	0.061	0.104	0.088	0.096
100	1	0.008	0.019	0.100	0.039	0.046	0.043	0.087	0.074	0.083
	2.5	0.009	0.028	0.011	0.041	0.058	0.052	0.099	0.083	0.099
	5	0.013	0.023	0.015	0.047	0.056	0.055	0.097	0.096	0.108
	10	0.010	0.022	0.011	0.039	0.045	0.044	0.088	0.078	0.100
	20	0.013	0.014	0.012	0.049	0.053	0.052	0.106	0.092	0.105
	30	0.008	0.014	0.008	0.047	0.039	0.049	0.092	0.091	0.106

[a]Empirical sizes of nominal 1%, 5% and 10% significance GMM tests of the Poisson (λ) null hypothesis. The statistic $J_n(q)$ is asymptotically χ^2 with degree of freedom $q - 1$. For comparison, the empirical sizes of the Poisson index of dispersion, $\mathscr{L}$, which is asymptotically χ^2 with degree of freedom $n - 1$, are also reported. The simulation experiment is based on 3000 replications.

their nominal values for sample sizes > 20 and are considerably below their nominal values for sample sizes < 20. For the 1% nominal significance level, the empirical size of $J_n(2)$ is close to its nominal value but the empirical size of $J_n(3)$ is substantially above its nominal value for all sample sizes. In contrast to the two statistics based on GMM, empirical sizes of $\mathscr{L}$ are close to their nominal values regardless of the sample size and the significance level.

5.3. The power

Table 2 reports the power of $J_n(2)$, $J_n(3)$ and $\mathscr{L}$ at the 1%, 5% and 10% significance levels against the binomial alternative. The values of two parameters of binomial distribution, P and N, were chosen to not only conform with other previously published studies such as Rao and Chakravarti (1956), and Darwin (1957), but also to

Table 2
Power of tests against binomial alternatives[a]

Sample size	N	P	1% test			5% test			10% test		
			Power $J_n(2)$	Power $J_n(3)$	Power $\mathscr{L}$	Power $J_n(2)$	Power $J_n(3)$	Power $\mathscr{L}$	Power $J_n(2)$	Power $J_n(3)$	Power $\mathscr{L}$
10	5	0.10	0.004	0.006	0.004	0.012	0.011	0.028	0.017	0.027	0.030
	5	0.25	0.000	0.001	0.000	0.004	0.026	0.005	0.049	0.052	0.011
	5	0.50	0.000	0.000	0.000	0.017	0.005	0.000	0.176	0.072	0.000
	5	0.75	0.000	0.003	0.000	0.153	0.006	0.000	0.667	0.283	0.000
	10	0.10	0.004	0.000	0.003	0.013	0.025	0.025	0.036	0.040	0.047
	10	0.25	0.000	0.000	0.000	0.004	0.002	0.003	0.060	0.023	0.011
	10	0.50	0.000	0.000	0.000	0.020	0.000	0.000	0.185	0.019	0.000
	10	0.75	0.000	0.004	0.000	0.155	0.000	0.000	0.686	0.081	0.000
	25	0.10	0.004	0.001	0.004	0.018	0.021	0.034	0.052	0.043	0.070
	25	0.25	0.000	0.000	0.000	0.006	0.003	0.002	0.059	0.013	0.026
	25	0.50	0.000	0.000	0.000	0.019	0.000	0.001	0.163	0.005	0.002
	25	0.75	0.000	0.007	0.000	0.172	0.000	0.000	0.682	0.027	0.000
20	5	0.10	0.003	0.004	0.003	0.014	0.010	0.012	0.030	0.029	0.028
	5	0.25	0.000	0.003	0.000	0.027	0.027	0.003	0.125	0.062	0.006
	5	0.50	0.003	0.002	0.000	0.251	0.130	0.000	0.550	0.307	0.000
	5	0.75	0.264	0.127	0.000	0.947	0.747	0.000	0.988	0.922	0.000
	10	0.10	0.002	0.006	0.002	0.017	0.017	0.022	0.072	0.052	0.034
	10	0.25	0.002	0.003	0.000	0.048	0.018	0.003	0.153	0.069	0.008
	10	0.50	0.007	0.002	0.000	0.307	0.086	0.000	0.557	0.265	0.000
	10	0.75	0.257	0.020	0.000	0.957	0.652	0.000	0.993	0.986	0.000
	25	0.10	0.003	0.006	0.003	0.018	0.020	0.016	0.062	0.037	0.036
	25	0.25	0.002	0.002	0.001	0.057	0.014	0.005	0.151	0.054	0.012
	25	0.50	0.005	0.000	0.000	0.303	0.047	0.000	0.586	0.200	0.000
	25	0.75	0.251	0.002	0.000	0.955	0.570	0.000	0.994	0.887	0.000
50	5	0.10	0.000	0.003	0.001	0.025	0.018	0.007	0.081	0.057	0.023
	5	0.25	0.020	0.021	0.000	0.163	0.092	0.000	0.331	0.174	0.000
	5	0.50	0.490	0.343	0.000	0.886	0.682	0.000	0.964	0.832	0.000
	5	0.75	0.999	0.996	0.000	1.000	1.000	0.000	1.000	1.000	0.000
	10	0.10	0.003	0.008	0.002	0.047	0.035	0.013	0.114	0.068	0.030
	10	0.25	0.021	0.010	0.000	0.191	0.082	0.000	0.348	0.167	0.000
	10	0.50	0.490	0.239	0.000	0.888	0.658	0.000	0.957	0.824	0.000
	10	0.75	0.999	0.992	0.000	1.000	0.999	0.000	1.000	1.000	0.000
	25	0.10	0.004	0.002	0.001	0.051	0.027	0.011	0.114	0.066	0.033
	25	0.25	0.026	0.009	0.000	0.188	0.071	0.000	0.340	0.159	0.004
	25	0.50	0.512	0.221	0.000	0.878	0.684	0.000	0.956	0.840	0.000
	25	0.75	1.000	0.996	0.000	1.000	1.000	0.000	1.000	1.000	0.000

Table 2 (*continued*)

Sample size	N	P	1% test			5% test			10% test		
			Power $J_n(2)$	Power $J_n(3)$	Power $\mathscr{L}$	Power $J_n(2)$	Power $J_n(3)$	Power $\mathscr{L}$	Power $J_n(2)$	Power $J_n(3)$	Power $\mathscr{L}$
100	5	0.10	0.006	0.010	0.000	0.064	0.044	0.002	0.146	0.083	0.009
	5	0.25	0.131	0.085	0.000	0.418	0.249	0.000	0.596	0.389	0.000
	5	0.50	0.976	0.898	0.000	0.998	0.986	0.000	1.000	0.998	0.000
	5	0.75	1.000	1.000	0.000	1.000	1.000	0.000	1.000	1.000	0.000
	10	0.10	0.008	0.010	0.001	0.070	0.043	0.010	0.156	0.085	0.020
	10	0.25	0.139	0.078	0.000	0.432	0.256	0.000	0.594	0.390	0.000
	10	0.50	0.973	0.884	0.000	0.999	0.985	0.000	1.000	0.995	0.000
	10	0.75	1.000	1.000	0.000	1.000	1.000	0.000	1.000	1.000	0.000
	25	0.10	0.007	0.006	0.001	0.077	0.041	0.009	0.156	0.092	0.023
	25	0.25	0.124	0.047	0.000	0.418	0.224	0.000	0.598	0.358	0.000
	25	0.50	0.967	0.863	0.000	0.998	0.981	0.000	0.999	0.998	0.000
	25	0.75	1.000	1.000	0.000	1.000	1.000	0.000	1.000	1.000	0.000

[a]Power of GMM tests ($J_n(q)$) against the binomial (N, P) alternative hypothesis. For comparison, the power of the Poisson index of dispersion ($\mathscr{L}$) are also reported. The simulation experiment is based on 3000 replications.

provide a representative set of examples of possible models commonly encountered in practice. As can be seen from Table 2, two tests based on GMM are comparable for the binomial alternative. However, $\mathscr{L}$ has no power against binomial distributions regardless of the values of binomial parameters, significance levels and sample sizes — a well-known result reported in previous studies (see, for example, Nass, 1959; Olkin et al., 1981). For two GMM test statistics $J_n(2)$ and $J_n(3)$, the power depends largely on P rather than on N. For a fixed sample size, the power increases as P increases. For small P ($P \leqslant 0.25$), both $J_n(2)$ and $J_n(3)$ are not powerful unless the sample size is relatively large ($n \geqslant 20$ for 5% and 10% significance levels, and $n \geqslant 50$ for 1% significance level). The results also suggest that although the approximate slope c_q increases with q, there appears to be no gain of using $J_n(3)$ over $J_n(2)$ in finite samples with sample sizes up to at least 100. This may be attributed to the imprecision with which the higher moments are estimated for a fixed sample size. The results indicate that the local power (and the approximate slope) comparisons discussed in Section 4 are only strictly valid asymptotically. It is, therefore, inappropriate in finite, especially in small samples (Bahadur, 1967; Geweke, 1981; Faust, 1992).

Table 3 reports the results for the negative binomial alternative with parameters p and k. For comparison, we use the same parameter values for p and k as those of P and N of binomial distributions. Although the local power (and the approximate slope) of test statistics based on GMM against the negative binomial distribution is the same as that against the binomial distribution (see Section 4), the results show that both $J_n(2)$ and $J_n(3)$ are, in general, much powerful against the negative binomial than the binomial alternative except for a few cases in which p is high and the sample size is large. As expected, similar to the binomial case, the power increases as either p or

Table 3
Power of tests against negative binomial alternatives[a]

Sample size	k	p	1% test			5% test			10% test		
			Power $J_n(2)$	Power $J_n(3)$	Power $\mathscr{L}$	Power $J_n(2)$	Power $J_n(3)$	Power $\mathscr{L}$	Power $J_n(2)$	Power $J_n(3)$	Power $\mathscr{L}$
10	5	0.10	0.010	0.009	0.010	0.033	0.033	0.075	0.044	0.048	0.087
	5	0.25	0.044	0.040	0.042	0.082	0.071	0.129	0.115	0.089	0.194
	5	0.50	0.108	0.107	0.107	0.174	0.164	0.224	0.218	0.208	0.318
	5	0.75	0.195	0.190	0.191	0.296	0.272	0.350	0.325	0.332	0.458
	10	0.10	0.019	0.017	0.018	0.031	0.032	0.071	0.054	0.050	0.110
	10	0.25	0.041	0.042	0.039	0.083	0.077	0.131	0.128	0.123	0.229
	10	0.50	0.119	0.135	0.118	0.179	0.192	0.247	0.229	0.228	0.356
	10	0.75	0.188	0.212	0.185	0.284	0.308	0.366	0.358	0.358	0.472
	25	0.10	0.016	0.024	0.015	0.046	0.046	0.082	0.083	0.075	0.148
	25	0.25	0.037	0.054	0.033	0.084	0.102	0.138	0.128	0.141	0.219
	25	0.50	0.102	0.141	0.099	0.182	0.210	0.252	0.234	0.270	0.356
	25	0.75	0.195	0.227	0.192	0.294	0.297	0.377	0.358	0.358	0.488
20	5	0.10	0.025	0.025	0.025	0.053	0.046	0.086	0.088	0.070	0.138
	5	0.25	0.078	0.086	0.080	0.139	0.141	0.194	0.191	0.183	0.293
	5	0.50	0.195	0.200	0.196	0.291	0.279	0.355	0.353	0.341	0.485
	5	0.75	0.341	0.327	0.349	0.475	0.460	0.553	0.543	0.535	0.670
	10	0.10	0.032	0.037	0.033	0.066	0.065	0.087	0.101	0.091	0.143
	10	0.25	0.069	0.080	0.069	0.134	0.134	0.185	0.187	0.176	0.259
	10	0.50	0.183	0.199	0.186	0.278	0.289	0.376	0.364	0.360	0.513
	10	0.75	0.339	0.361	0.340	0.465	0.474	0.548	0.541	0.538	0.668
	25	0.10	0.022	0.026	0.022	0.063	0.064	0.091	0.098	0.099	0.171
	25	0.25	0.070	0.096	0.073	0.140	0.153	0.193	0.197	0.201	0.301
	25	0.50	0.190	0.215	0.191	0.310	0.311	0.391	0.385	0.375	0.504
	25	0.75	0.358	0.379	0.362	0.494	0.491	0.571	0.561	0.547	0.676
50	5	0.10	0.039	0.049	0.042	0.088	0.082	0.125	0.137	0.126	0.212
	5	0.25	0.128	0.138	0.143	0.229	0.224	0.301	0.300	0.273	0.419
	5	0.50	0.403	0.390	0.423	0.550	0.515	0.626	0.625	0.604	0.735
	5	0.75	0.693	0.663	0.707	0.804	0.776	0.853	0.853	0.837	0.908
	10	0.10	0.041	0.045	0.047	0.087	0.084	0.121	0.133	0.122	0.214
	10	0.25	0.150	0.146	0.160	0.246	0.244	0.321	0.323	0.308	0.443
	10	0.50	0.432	0.432	0.449	0.567	0.566	0.669	0.671	0.642	0.768
	10	0.75	0.711	0.674	0.723	0.825	0.796	0.874	0.874	0.844	0.929
	25	0.10	0.029	0.040	0.031	0.075	0.079	0.110	0.132	0.128	0.193
	25	0.25	0.108	0.128	0.125	0.239	0.246	0.311	0.318	0.319	0.437
	25	0.50	0.374	0.383	0.387	0.533	0.513	0.610	0.609	0.599	0.750
	25	0.75	0.681	0.669	0.712	0.818	0.790	0.872	0.871	0.844	0.919

Table 3 (*continued*)

Sample size	k	p	1% test			5% test			10% test		
			Power $J_n(2)$	Power $J_n(3)$	Power $\mathscr{L}$	Power $J_n(2)$	Power $J_n(3)$	Power $\mathscr{L}$	Power $J_n(2)$	Power $J_n(3)$	Power $\mathscr{L}$
100	5	0.10	0.059	0.077	0.070	0.115	0.124	0.174	0.187	0.176	0.268
	5	0.25	0.231	0.246	0.264	0.401	0.370	0.489	0.489	0.456	0.602
	5	0.50	0.682	0.650	0.712	0.838	0.780	0.888	0.888	0.844	0.934
	5	0.75	0.926	0.914	0.935	0.967	0.956	0.981	0.981	0.972	0.992
	10	0.10	0.069	0.063	0.080	0.122	0.130	0.178	0.188	0.182	0.276
	10	0.25	0.231	0.224	0.258	0.377	0.352	0.481	0.483	0.427	0.600
	10	0.50	0.690	0.652	0.728	0.826	0.800	0.883	0.883	0.854	0.936
	10	0.75	0.945	0.926	0.953	0.978	0.966	0.987	0.987	0.979	0.995
	25	0.10	0.059	0.078	0.069	0.127	0.146	0.176	0.188	0.199	0.294
	25	0.25	0.206	0.212	0.231	0.378	0.345	0.487	0.490	0.450	0.609
	25	0.50	0.717	0.676	0.750	0.830	0.808	0.889	0.889	0.858	0.941
	25	0.75	0.940	0.932	0.952	0.976	0.970	0.987	0.987	0.979	0.996

[a]Power of GMM tests ($J_n(q)$) against the negative binomial (k, p) alternative hypothesis. For comparison, the power of the Poisson index of dispersion ($\mathscr{L}$) are also reported. The simulation experiment is based on 3000 replications.

the sample size increases. Again, it seems that $J_n(2)$ is as powerful as $J_n(3)$. Note that $\mathscr{L}$ is either comparable to or more commonly, slightly more powerful than $J_n(2)$ and $J_n(3)$.

To understand why the power of GMM tests for the binomial can be lower than that for the negative binomial in finite samples (especially in small samples), consider $J_n(2)$ or equivalently, $\mathscr{L}$ statistic. Results from Cramér (1946, p. 354) asserts that

$$E(\mathscr{L}) = \frac{\sqrt{n}}{2} \frac{(\sigma^2 - \mu_1)}{\mu_1} + O\left(\frac{1}{n^{1/2}}\right)$$

and

$$\mathrm{Var}(\mathscr{L}) = 2\rho^2 + \frac{\rho(\rho - 1)(3\rho - 1)}{\mu_1} + O\left(\frac{1}{n^{1/2}}\right),$$

where $\rho = \sigma^2/\mu_1$. From these expressions it is apparent that when $N = k$ and $P = p$, binomial and negative binomial distributions will have approximately the same $|E(\mathscr{L})|$ and hence, c_2 in (16), but different $\mathrm{Var}(\mathscr{L})$ because $\rho < 1$ for the binomial and $\rho > 1$ for the negative binomial distribution. The difference in power observed in Tables 2 and 3 is attributed mainly to the second-order dynamics of $\mathscr{L}$, namely $\mathrm{Var}(\mathscr{L})$, which behaves fundamentally different for over- and under-dispersed distributions.

6. Extensions

Since compound Poisson distributions have been frequently used as an alternative to the Poisson distribution, it is natural and practically important to extend the power analysis to compound Poisson alternatives. Asymptotic results for compound Poisson

Table 4
Power of tests against compound poisson alternatives[a]

Panel A: $Poisson(\lambda) \wedge Rectangular(0, \pi)$

	1% test			5% test			10% test		
	Power $J_n(2)$	Power $J_n(3)$	Power $\mathscr{L}$	Power $J_n(2)$	Power $J_n(3)$	Power $\mathscr{L}$	Power $J_n(2)$	Power $J_n(3)$	Power $\mathscr{L}$
π									
1	0.037	0.038	0.037	0.076	0.074	0.117	0.113	0.105	0.186
2	0.094	0.084	0.095	0.164	0.144	0.227	0.225	0.201	0.320
3	0.147	0.136	0.151	0.271	0.246	0.347	0.342	0.323	0.483
4	0.261	0.251	0.262	0.426	0.409	0.532	0.520	0.517	0.666
5	0.408	0.414	0.415	0.550	0.559	0.643	0.631	0.633	0.760
6	0.521	0.525	0.525	0.668	0.681	0.759	0.743	0.741	0.845
7	0.632	0.655	0.643	0.765	0.782	0.824	0.815	0.830	0.884
8	0.721	0.746	0.726	0.832	0.852	0.884	0.876	0.894	0.936
9	0.797	0.822	0.800	0.870	0.879	0.907	0.901	0.917	0.955
10	0.855	0.884	0.859	0.926	0.938	0.956	0.946	0.955	0.975

Panel B: $Poisson(\lambda) \wedge Lognormal(\xi, \sigma^2)$

ξ	σ									
0	0.1	0.010	0.011	0.010	0.026	0.029	0.047	0.066	0.054	0.089
	0.5	0.017	0.026	0.017	0.039	0.041	0.052	0.071	0.073	0.104
	1	0.692	0.676	0.693	0.752	0.743	0.792	0.792	0.774	0.847
1	0.1	0.013	0.020	0.014	0.042	0.036	0.058	0.086	0.065	0.101
	0.5	0.043	0.057	0.044	0.096	0.095	0.143	0.147	0.133	0.230
	1	0.971	0.971	0.971	0.984	0.980	0.989	0.988	0.989	0.992

[a] Power of GMM tests ($J_n(q)$) against compound Poisson alternative hypotheses. For comparison, the power of the Poisson index of dispersion ($\mathscr{L}$) are also reported. The simulation experiment is based on 3000 replications with sample size 20.

distributions can be developed along the similar lines as those provided for the binomial and negative binomial in Section 4. Extensive simulation experiments indicate that $J_n(q)$ for various values of q and $\mathscr{L}$ are comparable and there is no major gain in using $J_n(q)$ with $q > 2$ than $J_n(2)$ for a wide variety of compound Poisson distributions. See Table 4 in which, for illustrative purposes, we report the results of power of test statistics against two compound Poisson alternatives: Poisson-rectangular and Poisson-log-normal distributions.

Combining the results from the previous section, we conclude that, against binomial, negative binomial and many compound Poisson alternatives, $J_n(2)$ is comparable to $J_n(3)$ in power. The key intuition for these results is that for all alternative distributions studied so far, the population mean is not equal to the population variance. Hence, $J_n(2)$ is able to detect a departure from the Poisson distribution by comparing the sample mean to the sample variance (the moment restriction in H_0^O). In fact, since most commonly used mixed distributions are over-dispersed, by applying Satterthwaite's (1942) result, the corresponding compound Poisson distributions (including two compound Poisson distributions in Table 4) are also over-dispersed. The finite-sample

Table 5
Power of tests against the geometric alternative[a]

Sample size	1% test			5% test			10% test		
	Power $J_n(2)$	Power $J_n(3)$	Power $\mathscr{L}$	Power $J_n(2)$	Power $J_n(3)$	Power $\mathscr{L}$	Power $J_n(2)$	Power $J_n(3)$	Power $\mathscr{L}$
10	0.036	0.073	0.035	0.061	0.136	0.082	0.153	0.192	0.116
20	0.050	0.155	0.048	0.146	0.288	0.094	0.246	0.408	0.132
50	0.080	0.526	0.064	0.194	0.748	0.119	0.280	0.846	0.166
100	0.105	0.909	0.068	0.224	0.980	0.131	0.315	0.992	0.176

[a]Power of GMM tests ($J_n(q)$) against the geometric ($p = 0.5$) alternative hypothesis. For comparison, the power of the Poisson index of dispersion ($\mathscr{L}$) are also reported. The simulation experiment is based on 3000 replications.

power properties of $J_n(2)$, $J_n(3)$ and $\mathscr{L}$ in Table 4 are, therefore, similar to those against the negative-binomial distribution (Table 3).

To see what happens if the distribution in the alternative hypothesis has its mean equal to the variance, we consider the following geometric distribution with positive integers as the sample space

$$p(j) = pq^{j-1}, \quad j = 1, 2, 3, \ldots, \tag{17}$$

where $0 \leqslant p \leqslant 1$ and $q = 1 - p$. Assuming the observational apparatus becomes active only when at least one event occurs, this distribution has mean $1/p$ and variance q/p^2. Table 5 presents the power results for $J_n(2)$, $J_n(3)$ and $\mathscr{L}$ against this geometric alternative with $p = \frac{1}{2}$ in which the distribution has a mean equal to its variance. Since both $J_n(2)$ and $\mathscr{L}$ rely only on the first two moments, they have difficulty distinguishing between the Poisson distribution and the geometric distribution defined in (17). It can be seen from Table 5 that with the inclusion of the third moment, $J_n(3)$ is much more powerful than both $J_n(2)$ and $\mathscr{L}$ across all significance levels and sample sizes considered.

These results also emphasize the fact that as semi-parametric tests, $J_n(q)$s (as well as $\mathscr{L}$), which rely solely on moments, may not utilize efficiently all available information such as that on the sample space of the variable.

7. Concluding remarks

This paper uses the generalized method of moments framework to investigate discriminating tests for discrete distributions with a focus on the two-parameter family of Katz and with an extension to compound Poisson distributions. It shows that although it has less asymptotic power, Katz's test statistic, which is based on the first two moments is, in general, comparable in finite-sample power with other GMM test statistics based on higher-order moments against binomial, negative binomial and various compound Poisson distributions with variances noticeably not equal to the means. The results also

suggest that tests based on GMM have reasonable power against the over-dispersed alternatives. However, although they show substantially better power properties than the Poisson index of dispersion, for testing the under-dispersed distributions (which, exampled by the binomial distribution, are less commonly encountered in practice), GMM tests may not be reliable unless the sample size is relatively large.

GMM tests based on higher-moment restrictions may have advantages when low-moment restrictions are satisfied by the distribution specified in the alternative hypothesis. Therefore, if the Katz's test or the Poisson index of dispersion fails to reject the null hypothesis of a Poisson distribution, it may not be superfluous to try a GMM test based on higher-order moment restrictions to make a further investigation before fitting a Poisson distribution to the data and comparing the observed and expected cell frequencies.

Although the discussion in this paper focuses on the Poisson distribution and the Katz family, the proposed GMM approach can be extended to developing discriminating tests in many other situations such as hypotheses using various generalizations of the Poisson distribution. The approach is also applicable to other discrete and continuous distributions. In general, a sensible test criterion can be found by the method of this paper if the hypothesized distributions are well characterized by a set of moment conditions and the over-identifying moment restrictions of the test are not satisfied by alternative distributions.

Acknowledgements

Appendix A. Proofs of Propositions

Proof of Proposition 3.1. The result follows from $(M'M)^{-1} = (F'V^{-1}F)^{-1}$ and Proposition 2.1. $\square$

Proof of Proposition 3.2. It is clear that F has rank 1 and so is $(FF')V^{-1}$. By the definition of F, (FF') is a $q \times q$ matrix with (i,j)th element $(\partial\mu_i/\partial\lambda)(\partial\mu_j/\partial\lambda)$. Now, we show that the first column of $FF'V^{-1}$ is $-(M'M)F$ and all other columns are 0. Using the result that $(M'M)^{-1} = \lambda$ in Proposition 3.1 and the property of the Poisson distribution that $\lambda\partial\mu_j/\partial\lambda = \mu_{j+1} - \mu_1\mu_j$ (Haight, 1967), we can show that $v_{1,j} = \lambda\partial\mu_j/\partial\lambda$, where $v_{1,j}$ is the $(1,j)$th element of V. Then, it can be shown that $FF' = [-(M'M)F, \mathcal{O}]V$ and therefore, $FF'V^{-1} = [-(M'M)F, \mathcal{O}]$. $\square$

Proof of Proposition 3.3. Since $D = M(M'M)^{-1}M' = (M'M)^{-1}V^{-1/2}(FF')V^{-1/2}$, $\mathrm{H}_0^{\mathrm{I}}$ is equivalent to

$$(FF')V^{-1}\mathrm{E}[f(x_i, \lambda)] = 0. \qquad \square$$

The result follows from Proposition 3.2.

Proof of Proposition 3.4. By Proposition 3.2 and the fact that the over-identifying restrictions in $\mathrm{H}_0^{\mathrm{O}}$ are specified by $(I - D)V^{-1/2}\mathrm{E}[f(x_i, \lambda)] = V^{-1/2}[I - (M'M)^{-1}FF'V]\mathrm{E}[f(x_i, \lambda)] = 0$, which is equivalent to $[I - (M'M)^{-1}FF'V]\mathrm{E}[f(x_i, \lambda)] = 0$. $\quad \square$

Proof of Proposition 4.2. The result follows from (14) and (15). $\quad \square$

Proof of Proposition 4.3. From the definition of $J_n(q)$ and Theorem 1 of Geweke (1981). $\quad \square$

References

Amemiya, T., 1985. Advanced Econometrics. Harvard University Press, Cambridge.

Bahadur, R., 1960. Stochastic comparison of tests. Ann. Math. Statist. 31, 276–295.

Bahadur, R., 1967. Rates of convergence of estimates and test statistics. Ann. Math. Statist. 38, 303–324.

Cramér, H., 1946. Mathematical Methods of Statistics. Princeton University Press, Princeton.

Cochran, W., 1936. The χ^2 distribution for binomial and Poisson series, with small expectations. Ann. Eugenics 7, 207–217.

Cochran, W., 1954. Some methods for strengthening the common χ^2 tests. Biometrics 10, 417–451.

Darwin, J., 1957. The power of the Poisson index of dispersion. Biometrika 44, 286–289.

Faust, J., 1992. When are variance ratio tests for serial dependence optimal? Econometrica 60, 1215–1226.

Fisher, R., 1950. The significance of deviations from expectation in a Poisson series. Biometrics 6, 17–24.

Gerlein, O.V., Pincus, R., 1978. Goodness-of-fit tests based on a measure of discrepancy between two scalar products. Theory Probab. Appl. 23, 248–258.

Geweke, J., 1981. The approximate slope of econometric tests. Econometrica 49, 1427–1442.

Groeneboom, P., Oosterhoff, J., 1981. Bahadur efficiency and small-sample efficiency. Internat. Statist. Rev. 49, 127–141.

Gurland, J., Tripathi, R., 1975. Estimation of parameters on some extensions of the Katz family of discrete distributions involving hypergeometric functions. In: Patil, G., Kotz, S., Ord, J. (Eds.), Statistical Distributions in Scientific Work, Vol. 1: Models and Structures. Reidel, Dordrecht, pp. 59–82.

Haight, F., 1967. Handbook of the Poisson Distribution. Wiley, New York.

Hall, A., 1999. Hypothesis testing in models estimated by GMM. In: Màtyàs, L. (Ed.), Generalized Method of Moments Estimation. Cambridge University Press, Cambridge.

Hall, A., 2001. Generalized method of moments. In: Baltagi, B. (Ed.), A Companion to Theoretical Econometrics. Blackwell Publishers, Malden.

Hansen, L., 1982. Large sample properties of generalized method of moment estimators. Econometrica 50, 1029–1054.

Johnson, N., Kotz, S., Kemp, A., 1992. Univariate Discrete Distributions, 2nd Edition. Wiley, New York.

Kagan, A.M., 1975. On a measure of divergency between two scalar products and its statistical application. Sankhya Ser. A 37, 211–228.

Katz, L., 1965. Unified treatment of a broad class of discrete probability distributions. Proceedings of the International Symposium on Discrete Distributions. Montreal, pp. 175–182.

Malinvaud, E., 1970. Statistical Methods of Econometrics. North-Holland, Amsterdam.

Nass, C., 1959. The χ^2 test for small expectations in contingency tables, with special reference to accidents and absenteeism. Biometrika 46, 365–385.

Newey, W., 1985. Generalized method of moments specification testing. J. Econom. 29, 229–256.

Olkin, N., Petkau, J., Zidek, J., 1981. A comparison of *n* estimators for the binomial distribution. J. Amer. Statist. Assoc. 76, 637–642.

Ord, J., 1967a. Graphical methods for a class of discrete distributions. J. Roy. Statist. Soc. Ser. A 130, 232–238.

Ord, J., 1967b. On a system of discrete distributions. Biometrika 54, 649–656.

Ord, J., 1972. Families of Frequency Distributions. Griffin, London.

Rao, C., Chakravarti, I., 1956. Some small sample tests of significance for a Poisson distribution. Biometrics 12, 264–281.

Satterthwaite, F., 1942. Generalized Poisson distribution. Ann. Math. Statist. 13, 410–417.

Sievers, G., 1969. On the probability of large deviations and exact slopes. Ann. Math. Statist. 40, 1908–1921.

Sowell, F., 1996. Optimal tests for parameter instability in the generalized method of moments framework. Econometrica 64, 1085–1108.

Stuart, A., Ord, J., 1987. Kendall's Advanced Theory of Statistics: Vol. 1, Distribution Theory. Oxford University Press, Oxford.

Tripathi, R., Gurland, J., 1977. A general family of discrete distributions with hypergeometric probabilities. J. Roy. Statist. Soc. Ser. B 39, 349–356.

Tripathi, R., Gurland, J., 1979. Some aspects of the Kemp families of distributions. Comm. Statist. — Theory Methods 8, 855–869.

Wieand, H., 1976. A condition under which the Pitman and Bahadur approaches to efficiency coincide. Ann. Statist. 4, 1003–1011.

Wolfram, S., 1999. The Mathematica Book. Cambridge University Press, Cambridge.