Available online at www.sciencedirect.com

ScienceDirect

Computational Statistics & Data Analysis 52 (2008) 2829–2839

COMPUTATIONAL STATISTICS & DATA ANALYSIS

www.elsevier.com/locate/csda

Semi-parametric specification tests for mixing distributions

Yue Fang[a, b, *]

[a] Lundquist College of Business, University of Oregon, Eugene, OR 97403, USA
[b] China Europe International Business School, Shanghai 201206, PR China

Received 15 September 2006; received in revised form 16 October 2007; accepted 18 October 2007
Available online 26 October 2007

Abstract

We present a semi-parametric method for testing mixing distributions in the mixed Poisson model. The proposed method, which is based on the generalized method of moments, does not demand the complete specification of the probability function but only requires a specification of a set of moment conditions which the model should satisfy. We demonstrate that an explicit expression for moment relations between the mixing and the mixed distributions provides a natural way in selecting moment restrictions and model parameterization. The Monte Carlo evidence suggests that the test has satisfactory performance for moderate size samples.

Keywords: Mixed Poisson distribution; Model specification; Monte Carlo simulation; Semi-parametric method

1. Introduction

Mixed Poisson distributions are obtained from the homogeneous Poisson distribution by assuming that the Poisson parameter is itself a random variable with positive support. Special types of mixed Poisson distributions are obtained by considering specific distributional types of the mixing distribution. In the class of continuous distributions, the gamma distribution is one well-known example. Other members of the Pearson family of distributions may also serve as the mixing distribution. Commonly used discrete mixing distributions include the logarithmic series or a more general class of discrete distributions—the class of generalized power series distributions (Haight, 1967; Johnson et al., 1993).

In this article we consider the problem of testing for the mixing distribution when the mixed distribution is assumed to be a mixed Poisson. More specifically, let $\mathscr{F}$ denote a family of distributions with distribution function $U(x)$, where $U(0) = 0$. Suppose that we have an observed sample $\{x_i : i = 1, 2, \ldots, n\}$ of the variable X defined by

$$P_k \equiv P(X = k) = \int_0^\infty \mathrm{e}^{-\lambda} \lambda^k \Gamma^{-1}(k+1)\, \mathrm{d}U(\lambda). \tag{1}$$

We would like to test the null hypothesis H_0 that the mixing distribution $\in \mathscr{F}$. Initially we develop the arguments in a general setting. Then, we focus our attention on the Pearson family of distributions or its sub-families as $\mathscr{F}$ to further investigate the properties of the test and to demonstrate its use by applying it to a claim frequency example.

* Corresponding author at: Lundquist College of Business, University of Oregon, Eugene, OR 97403, USA. Tel.: +1 541 346 3265.
E-mail address: yfang@darkwing.uoregon.edu.

0167-9473/$ - see front matter
doi:10.1016/j.csda.2007.10.015

It should be pointed out that the mixing distribution is unobservable, since the data are drawn from the mixed distribution. The identifiability result from Douglas (1980) allows us to identify the mixing distribution based on the mixed distribution.

Note that by allowing $\mathscr{F}$ to be a general family, the probability function of the mixed Poisson distribution can become unwieldy. For example, consider the Pearson family as $\mathscr{F}$. Then, there is no general explicit expression of the probability function for the mixed Poisson since the integral in (1) can be carried out only for some special cases. As a result, many widely used methods such as the Pearson chi-square (χ^2) and likelihood-based tests may not be easy to apply. In contrast, as will be shown, our testing procedure provides a flexible methodology in dealing with such a complicated $\mathscr{F}$ and has a computational advantage since it does not require this sort of full knowledge but only demands the specification of a set of moment conditions, which are usually easy to establish.

We examine approximate slopes of the tests and show that asymptotic properties of the tests depend intimately on the over-identifying moment restrictions. The Monte Carlo simulations suggest that in general, the test has satisfactory performance for moderate size samples, in particular the proposed test performs well in comparison with the Pearson χ^2 test, as demonstrated by a study of the data of Johnson and Hey (1971).

A common feature of mixed Poisson models is the presence of heterogeneity. The heterogeneity property has spurred broad applications of mixed Poisson distributions. For example, mixed Poisson distributions have been widely used in modeling insurance claims (Bühlmann, 1970; Panjer, 1986; Fang, 2003a). The number of claims occurred in a given time period is often approximated by a Poisson distribution with a risk parameter λ representing the claim intensity (i.e., the intensity of claim-causing events). The claim intensity is rarely a constant because it is subject to variations of many factors including external background factors, such as weather and economic conditions. A mixed Poisson distribution provides a more realistic model for the number of claims than the homogeneous Poisson distribution with a constant λ. One commonly used techniques to introduce the mixing effect is the use of method of moments. As a semi-parametric specification approach, the method of moments focuses only on the key characteristics, in particular lower-order moment such as the standard deviation and skewness. It is a natural approach in modeling insurance claims because the lower-order moments are the most important in specifying the mixing distribution. The differences in the structures of the tail characterized by higher-order moments, even though maybe significant, concern only a very small number of accidents and do not necessarily have any major effect on the finance of the business (Daykin et al., 1994).

The structure of the paper is as follows. Section 2 describes our method and asymptotic properties of the test. Section 3 presents finite-sample properties of the test via Monte Carlo simulations. An empirical example is given in Section 4. Section 5 contains a discussion.

2. Method and tests statistics

2.1. Preliminaries

As aspect of particular importance when dealing with the inference of mixing distributions is the concept of identifiability. Douglas (1980) has shown that for any mixed Poisson distribution, the mixing distribution is unique: if probability generating functions of mixed Poisson distributions equal, then the mixing distributions have the same distribution function. Thus, in dealing with mixed Poisson distributions, the mixing distribution is identifiable.

Since our approach depends upon the moments of mixing distributions and those of their mixtures with Poisson, we start with moment generating functions. Let Λ be a mixing distribution with distribution function $U(x) \in \mathscr{F}$. It can be shown that the factorial moment generating function of X equals the moment generating function of Λ:

$$\sum (1+s)^j P_j = \int_0^\infty \mathrm{e}^{sx}\, \mathrm{d}U(x). \tag{2}$$

In fact, a necessary and sufficient condition that a distribution be a mixed Poisson is that its factorial moment generating function be equal to the moment generating function of a random variable with positive support (Haight, 1967). From (2) one can obtain moments of mixed Poisson in terms of moments of the mixing distribution.

Let μ_k, μ_k' and $\mu_{(k)}$ denote respectively the kth central moment, kth moment about zero and kth factorial moment of X. Similarly, let ν_k denote the kth central moment and ν_k' the kth moment about zero of Λ. Supposing all moments to exist we have the mean and variance of X

$$\mu_1' = v_1' \quad \text{and} \quad \mu_2 = v_2 + v_1' \tag{3}$$

and in general, for the kth factorial moments

$$\mu_{(k)} = v_k'. \tag{4}$$

The property (3) characterizes the "over-dispersed" aspect of mixed Poisson distributions. The moment relationships between X and Λ in (4) have fundamental implications for statistics based on the generalized method of moments (GMM) which, as shown in later sections, play a central role in developing specification tests for mixing distributions.

The following results, which we will use later, are adapted from Johnson et al. (1993, p. 42), showing the connections between different types of moments:

$$\mu_k = \sum_{j=0}^{k} (-1)^j \binom{k}{j} \mu_1'^j \mu_{k-j}' \tag{5}$$

and

$$\mu_k' = \sum_{j=0}^{k} (\Delta^j 0^k / j!) \mu_{(k)}, \tag{6}$$

where $\Delta^j 0^k / j!$ is the Stirling number of the second kind. In particular, we may write (6) in the matrix form: $(\mu_1', \mu_2', \dots, \mu_j')' = \Omega(\mu_{(1)}, \mu_{(2)}, \dots, \mu_{(j)})'$, where the transformation matrix Ω is lower triangular with 1 as diagonal elements. For the inverse calculation of (6), Ω^{-1}, which is also lower triangular, has Stirling numbers of the first kind as its elements. Note that both (5) and (6) are valid for any random variables, assuming all required moments exist. Hence, they hold for the mixing variable Λ as well.

2.2. Test statistics

Suppose that $\mathscr{F}$ involves p unknown parameters Θ. Let $f_n(x_i, \Theta)$ be a continuous q-dimensional vector function of Θ with $q \geqslant p$. Consider the moment restrictions:

$$E[f_n(x_i, \Theta)] = 0. \tag{7}$$

When the number of orthogonality moment restrictions in (7) exceeds the number of parameters, the model is over-identified. In the over-identified case, (7) implies substantive restrictions. If the hypothesis of the model that leads to (7) in the first place is incorrect, some of the sample moment restrictions will be systematically violated, providing a basis for developing a specification test (Hall, 2005).

Let $\hat{\Theta}$ be the GMM estimate of Θ, which is the value of Θ that minimizes the quadratic form

$$Q_n(\Theta) \equiv f_n(\Theta)' V_n^{-1} f_n(\Theta), \tag{8}$$

where $f_n(\Theta) \equiv n^{-1} \sum_{i=1}^{n} f_n(x_i, \Theta)$ and V_n is a consistent estimator of $V = \lim_{n\to\infty} Var[n^{1/2} f_n(\Theta)]$. By construction, the following statistic

$$J_n(q) \equiv n f_n(\hat{\Theta})' \hat{V}_n^{-1} f_n(\hat{\Theta}) \tag{9}$$

can be used to test the validity of the hypothesis $H_0 : E[f_n(x_i, \Theta)] = 0$. It can be shown that under the null hypothesis of the model, $J_n(q)$ has an asymptotic χ^2 distribution with $(q - p)$ degrees of freedom under some regularity conditions (Hansen, 1982).

Consider the moment restrictions (7), where $f_n(x_i, \Theta) = \Xi g_n(x_i, \Theta)$, Ξ is a given $q \times q$ non-singular weighting matrix, and $g_n(x_i, \Theta) = (x_i - \mu_1', x_i^2 - \mu_2', \dots, x_i^q - \mu_q')'$. It is important to note that the test statistic $J_n(q)$ is invariant to the weighting matrix Ξ, suggesting that the choice of Ξ has no impact on the properties of the tests statistic. However, by taking different values in Ξ, one obtains different moment restrictions in (7). Although these moment restrictions are equivalent, we suggest that Ξ is taken to be Ω^{-1}. In this case, applying (4), (7) becomes

$$E[f_n(x_i, \Theta)] = E[(x_i - v_1', x_i^{(2)} - v_2', \ldots, x_i^{(q)} - v_q')'] = 0, \tag{10}$$

where $x_i^{(k)} = x_i(x_i - 1)\ldots(x_i - k + 1)$. Having (7) being represented as moment restrictions of the mixing distribution is of great interest. It will simplify the analysis of moment restrictions and as demonstrated later, will provide insight into how the test works.

An advantage of having $f_n(x_i, \Theta)$ be written in terms of $g_n(x_i, \Theta)$ is that it offers a natural way to carry out some required numerical computations. For example, it can be verified that the (i, j)-th element of W is $(\mu'_{i+j} - \mu'_i\mu'_j)$, where W is the matrix of $\lim_{n\to\infty} Var[n^{1/2} g_n(\Theta)]$ with $g_n(\Theta) \equiv n^{-1}\sum_{i=1}^{n} g_n(x_i, \Theta)$. Consequently, V^{-1} can be easily obtained as $V^{-1} = \Omega' W^{-1} \Omega$ and the close form of the test statistic in (9) may be derived similarly as in Fang (2003b).

2.3. Asymptotic properties

To examine the asymptotic properties of the proposed testing strategy, we study the approximate slope 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 sample size n. Geweke's result provides a simple way to obtain the approximate slope of test statistics in many situations including our cases. Let c_q be the approximate slope of $J_n(q)$. Applying Geweke's result, we have

$$c_q = E[f_n(x_i, \Theta)']V^{-1}E[f_n(x_i, \Theta)] = E[g_n(x_i, \Theta)']W^{-1}E[g_n(x_i, \Theta)]. \tag{11}$$

With respect to the GMM approach discussed in this article, it is possible to calculate the approximate slopes of $J_n(q)$ for any q given the alternative hypothesis. This is of interest because we can quantify the potential for efficiency gain in power from additional moment restrictions. We note however that despite its simplicity, the approximate slope may be a poor measure of the exact slope in many cases (Bahadur, 1967) and has to be used with caution. Especially it cannot be used by itself to compare the merits of tests in finite samples.

As an illustrative example, we consider the standard gamma distribution with one-parameter distribution function $U(x) = \int_0^x f(x)\,dx$, where

$$f(x) = \Gamma(\alpha)^{-1} x^{\alpha - 1} e^{-x}; \quad \alpha > 0; x > 0. \tag{12}$$

In this simple case $\mathscr{F}$ consists of a one-parameter family of distributions and $p = 1$. Let $\Theta = \alpha$. The only one identifying restriction is $E(x_i) - v_1' = 0$ and the GMM estimator for α, $\hat{\alpha}$, is $\bar{x}$. For $q > 1$, consider the null hypothesis (7) with $f_n(x_i, \Theta) = (x_i - v_1', x_i^{(2)} - v_2', \ldots, x_i^{(q)} - v_q')'$. With the moment $v_k' = \Gamma(\alpha + k)/\Gamma(\alpha)$ (Stuart and Ord, 1987) and moment relationships (4) through (6), the test statistic $J_n(q)$ is ready to be obtained. Under the null hypothesis that $E[f_n(x_i, \alpha)] = 0$, $J_n(q)$ is asymptotically distributed as a χ^2_{q-1} random variable.

Noting that the gamma distribution belongs to the Pearson family, we consider two other members of the Pearson family as alternative mixing distributions: the chi-square and the beta distributions. To provide some perspectives for non-Pearson distributions, we also include the rectangular and log-normal distributions in our study. Our choice of parameters in the alternative mixing distributions was determined by several considerations. We would like to have parameters of the mixing distribution such that the mean of the mixed Poisson has the span ranging from about 1 to 20. We also avoid confronting null hypotheses with extreme alternatives, since such alternatives hold little interest and have no bearing on asymptotics, the power being near 1 for very small sample sizes.

Table 1 provides a comparison of approximate slopes of $J_n(q)$ ($q = 2$, 3 and 4) for the four alternative mixing distributions. As can be seen, c_q appears to depend not only on the probability structures but also on the parameterizations of the alternative hypotheses. In general, approximate slopes increase with q. This implies that including additional moment restrictions will not reduce asymptotic efficiency of the test. For the chi-square and the beta distributions, the change in c_q is limited for most parameterizations as q goes from 2 to 4. However, the gain in c_q with higher value of q

Table 1
Approximate slopes c_q

Mixing Distribution	v	Approximate slopes c_2	c_3	c_4	Mixing Distribution	α	ρ	Approximate slopes c_2	c_3	c_4
Chi-square[a]	1	0.06250	0.06250	0.07129	Beta[b]	1	1	0.02170	0.04209	0.05331
	2	0.08333	0.08333	0.09583		2	1	0.02375	0.02563	0.02576
	5	0.10417	0.10417	0.12126		5	1	0.01843	0.02465	0.02861
	10	0.11364	0.11364	0.13333		1	5	0.02170	0.04209	0.05331
	15	0.11719	0.11719	0.13795		2	5	0.02232	0.02563	0.02576
	20	0.11905	0.11905	0.14045		5	5	0.02389	0.02465	0.02861
	π	c_2	c_3	c_4		ξ	σ	c_2	c_3	c_4
Rectangular[c]	1	0.02894	0.05679	0.07436	Lognormal[d]	0.0	0.5	0.03054	0.05104	0.05829
	2	0.02778	0.05247	0.07843		0.0	1.0	0.26142	2.45964	121.0823
	5	0.00248	0.03564	0.53536		1.0	0.5	0.00148	0.00643	0.00686
	10	0.04627	0.32848	10.06836		1.0	1.0	4.58872	152.9395	32134.3
	20	0.61869	4.67032	173.4152		−1.0	0.5	0.02858	0.05680	0.07317
	40	3.82275	51.80903	2751.452		−1.0	1.0	0.00008	0.02497	0.12618

[a]The mixing variable Λ follows $\chi^2_v(0)$.
[b]The mixing variable $\Lambda = \rho\Lambda'$, where Λ' is $Beta(\alpha, 2.0)$.
[c]The mixing variable Λ follows $Rectangular(0, \pi)$.
[d]The mixing variable Λ is $Lognormal(\xi, \sigma^2)$.

Table 2
Sizes of tests of the standard gamma null hypothesis

Sample size	α	1% test Size $J_n(2)$	Size $J_n(3)$	Size $J_n(4)$	5% test Size $J_n(2)$	Size $J_n(3)$	Size $J_n(4)$	10% test Size $J_n(2)$	Size $J_n(3)$	Size $J_n(4)$
50	1	0.022	0.020	0.021	0.040	0.039	0.040	0.069	0.065	0.058
	2.5	0.016	0.022	0.020	0.043	0.044	0.039	0.086	0.072	0.060
	5	0.014	0.023	0.019	0.048	0.046	0.038	0.091	0.075	0.061
	10	0.013	0.018	0.019	0.047	0.046	0.042	0.091	0.079	0.067
	20	0.010	0.015	0.021	0.045	0.048	0.044	0.090	0.083	0.072
100	1	0.018	0.020	0.018	0.042	0.046	0.042	0.078	0.069	0.062
	2.5	0.014	0.021	0.018	0.048	0.049	0.039	0.095	0.079	0.063
	5	0.012	0.021	0.020	0.048	0.045	0.040	0.099	0.078	0.065
	10	0.010	0.017	0.018	0.051	0.051	0.048	0.097	0.088	0.076
	20	0.009	0.020	0.023	0.046	0.048	0.047	0.087	0.080	0.078
200	1	0.013	0.021	0.017	0.051	0.047	0.039	0.098	0.075	0.062
	2.5	0.016	0.019	0.017	0.056	0.052	0.040	0.110	0.091	0.071
	5	0.014	0.023	0.022	0.055	0.058	0.050	0.110	0.092	0.080
	10	0.013	0.019	0.022	0.055	0.053	0.051	0.104	0.090	0.080
	20	0.012	0.020	0.022	0.058	0.055	0.056	0.112	0.095	0.090
500	1	0.017	0.024	0.021	0.058	0.055	0.047	0.112	0.093	0.073
	2.5	0.012	0.020	0.023	0.054	0.058	0.050	0.109	0.096	0.082
	5	0.012	0.021	0.023	0.053	0.056	0.054	0.102	0.099	0.087
	10	0.008	0.014	0.020	0.051	0.052	0.050	0.105	0.095	0.082
	20	0.011	0.012	0.018	0.052	0.047	0.048	0.098	0.096	0.087

can be considerable for the rectangular and log-normal mixing distributions, especially when the mixed distribution has a relatively large mean value. For example, the approximate slope for the rectangular with $\pi = 10$ starts at about 0.05 when $q = 2$ and jumps to 0.3 at $q = 3$. Comparing the approximate slopes for $q = 2$ and 3, the ratio is approximately 14%. This implies that about 7 times as many observations are needed to reject the alternative hypothesis when q is chosen

Table 3
Power of tests against chi-square as mixing distributions

Sample size	v	1% test			5% test			10% test		
		Power $J_n(2)$	Power $J_n(3)$	Power $J_n(4)$	Power $J_n(2)$	Power $J_n(3)$	Power $J_n(4)$	Power $J_n(2)$	Power $J_n(3)$	Power $J_n(4)$
50	1	0.218	0.216	0.194	0.318	0.317	0.278	0.380	0.379	0.341
	2	0.281	0.278	0.244	0.398	0.400	0.358	0.473	0.477	0.430
	5	0.343	0.338	0.305	0.482	0.480	0.433	0.556	0.565	0.509
	10	0.359	0.359	0.333	0.508	0.499	0.458	0.600	0.583	0.538
	15	0.384	0.391	0.365	0.543	0.538	0.501	0.630	0.620	0.581
	20	0.382	0.386	0.364	0.540	0.527	0.496	0.620	0.612	0.577
100	1	0.376	0.380	0.339	0.509	0.512	0.465	0.585	0.592	0.547
	2	0.487	0.489	0.440	0.633	0.643	0.590	0.704	0.714	0.665
	5	0.609	0.598	0.553	0.749	0.740	0.693	0.810	0.805	0.764
	10	0.651	0.633	0.601	0.793	0.776	0.734	0.851	0.840	0.802
	15	0.660	0.649	0.615	0.801	0.786	0.750	0.861	0.845	0.809
	20	0.676	0.650	0.622	0.818	0.790	0.758	0.877	0.852	0.822
200	1	0.625	0.643	0.599	0.754	0.772	0.732	0.811	0.831	0.794
	2	0.766	0.780	0.731	0.879	0.889	0.853	0.918	0.929	0.899
	5	0.881	0.877	0.850	0.951	0.947	0.925	0.969	0.969	0.956
	10	0.912	0.903	0.879	0.967	0.956	0.941	0.979	0.977	0.963
	15	0.929	0.919	0.900	0.974	0.968	0.957	0.986	0.983	0.974
	20	0.932	0.912	0.895	0.975	0.968	0.952	0.986	0.981	0.971
500	1	0.946	0.965	0.949	0.978	0.986	0.978	0.988	0.992	0.987
	2	0.988	0.991	0.989	0.997	0.998	0.997	0.997	0.999	0.998
	5	0.998	0.998	0.998	1.000	1.000	0.999	1.000	1.000	1.000
	10	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	15	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	20	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

to equal 2. Similar astonishing increases in c_q as q varies from 2 to 4 can be observed in the cases of the rectangular distributions with $\pi = 20$ and 40, and the log-normal with $(\xi, \sigma) = (0.0, 1.0)$ and $(1.0, 1.0)$.

3. Monte Carlo evidence

Although the above results characterize the asymptotic power of $J_n(q)$ for large samples, further insight into the operating characteristics of the test in finite samples can be gained by means of Monte Carlo simulation. In this section, as an illustration, we conduct Monte Carlo simulations to examine the finite-sample properties of $J_n(q)$ of the standard gamma null hypothesis via simulation experiments under the four alternative hypotheses considered in Section 2.3. All simulations are based on 10,000 replications and performed in single-precision FORTRAN using various random number generators of the IMSL subroutine library. The nominal significance level is chosen to be 1%, 5% and 10%, while the sample size is taken to be 50, 100, 200, and 500.

3.1. The size

Table 2 reports the empirical sizes of tests with $q = 2$, 3, and 4 under the null hypothesis of the standard gamma distribution. The results show that the empirical sizes of 1% tests are above the nominal value for small samples and decrease, in general, with the sample size. For 5% and 10% tests, the empirical sizes start low, increase as the sample size goes from 50 to 200, and settle at their nominal values when $n > 200$ in most cases. For sample sizes less than 100, the empirical sizes of $J_n(q)$ appear to be closer to the nominal values for mixing distributions with relatively large mean values. Not surprisingly, for a fixed sample size, the empirical size of the test with smaller value q is, in general, closer to its nominal value than that with the higher value q for the same size.

Table 4
Power of tests against beta as mixing distributions

Sample size	α	ρ	1% test			5% test			10% test		
			Power $J_n(2)$	Power $J_n(3)$	Power $J_n(4)$	Power $J_n(2)$	Power $J_n(3)$	Power $J_n(4)$	Power $J_n(2)$	Power $J_n(3)$	Power $J_n(4)$
50	1	1	0.000	0.001	0.005	0.000	0.025	0.048	0.014	0.113	0.115
	2	1	0.000	0.011	0.040	0.012	0.146	0.190	0.153	0.330	0.295
	5	1	0.000	0.074	0.122	0.116	0.368	0.352	0.460	0.571	0.507
	1	5	0.001	0.001	0.002	0.008	0.010	0.008	0.032	0.028	0.020
	2	5	0.002	0.007	0.004	0.074	0.061	0.026	0.205	0.131	0.064
	5	5	0.066	0.109	0.049	0.457	0.374	0.218	0.686	0.543	0.371
100	1	1	0.000	0.033	0.068	0.057	0.263	0.269	0.330	0.481	0.427
	2	1	0.006	0.234	0.297	0.407	0.624	0.585	0.749	0.788	0.728
	5	1	0.145	0.557	0.557	0.776	0.862	0.814	0.942	0.935	0.894
	1	5	0.001	0.001	0.001	0.013	0.014	0.009	0.052	0.031	0.022
	2	5	0.038	0.043	0.018	0.253	0.167	0.094	0.453	0.285	0.171
	5	5	0.542	0.520	0.356	0.891	0.808	0.671	0.962	0.903	0.806
200	1	1	0.061	0.433	0.463	0.674	0.800	0.755	0.904	0.911	0.868
	2	1	0.565	0.856	0.841	0.965	0.976	0.960	0.994	0.992	0.984
	5	1	0.942	0.984	0.978	0.999	0.999	0.997	1.000	1.000	0.999
	1	5	0.003	0.002	0.001	0.036	0.016	0.016	0.096	0.043	0.041
	2	5	0.251	0.193	0.109	0.639	0.453	0.307	0.793	0.607	0.464
	5	5	0.976	0.957	0.907	0.999	0.994	0.982	1.000	0.999	0.994
500	1	1	0.975	0.994	0.990	0.999	1.000	0.999	1.000	1.000	1.000
	2	1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	5	1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1	5	0.018	0.004	0.006	0.119	0.038	0.075	0.234	0.090	0.161
	2	5	0.894	0.788	0.675	0.983	0.938	0.891	0.994	0.975	0.946
	5	5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Since the sampling theory for $J_n(q)$ is obtained only asymptotically, the actual size of those tests will of course differ from their nominal values in finite samples. Although Table 2 indicates that such differences may not be large for reasonable sample sizes, it may nevertheless seem more desirable to base tests upon corrected χ^2 critical values. It can be done, for example, using standard regression techniques to related the percentage points of $J_n(q)$ derived from Monte Carlo experiments, against the parameter α, as in Rayner and Best (1989). The finite-sample bias correction depends however on the null hypothesis and must be resolved with the particular model and data at hand. Since our main objective for using the gamma distribution is to demonstrate the proposed approach, we have not made any effect to correct for finite-sample biases in this article.

3.2. The power

Tables 3 through 6 report the power results of $J_n(q)$ for the four alternatives. In examining the results from Tables 3 through 6 four general conclusions emerge.

First, for a fixed number of observations, it becomes apparent that choosing an appropriate value q for the tests depends intimately on the alternative hypothesis of interest. For example, the power of $J_n(q)$ against the chi-square alternative declines slightly as q increases (Table 3). On the other hand, when the rectangular distribution is considered, the power of the tests increases with q and the increase can be significant; as the case of 100 observations demonstrates, the power against the rectangular distribution with $\pi = 10$ is 29.8% when $q = 2$ but jumps to 70.9% when $q = 3$ (Table 5).

Second, the power of the test depends on both the structure and the parameterization of the mixing distribution considered in the alternative hypothesis. For example, except for several cases (which will be discussed next), the power of the test seems to increase with the mean value of the mixing distribution for a given value of q, regardless of the size of the test.

Table 5
Power of tests against rectangular as mixing distributions

Sample size	π	1% test			5% test			10% test		
		Power $J_n(2)$	Power $J_n(3)$	Power $J_n(4)$	Power $J_n(2)$	Power $J_n(3)$	Power $J_n(4)$	Power $J_n(2)$	Power $J_n(3)$	Power $J_n(4)$
50	1	0.000	0.040	0.072	0.007	0.269	0.253	0.097	0.470	0.398
	2	0.000	0.044	0.040	0.038	0.209	0.148	0.193	0.357	0.249
	5	0.000	0.002	0.004	0.008	0.013	0.022	0.034	0.033	0.047
	10	0.134	0.556	0.603	0.274	0.742	0.772	0.382	0.824	0.840
	20	0.963	0.999	0.999	0.988	1.000	1.000	0.993	1.000	1.000
	40	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
100	1	0.002	0.134	0.174	0.272	0.471	0.422	0.624	0.656	0.567
	2	0.022	0.092	0.071	0.306	0.335	0.238	0.570	0.495	0.367
	5	0.002	0.004	0.009	0.018	0.018	0.029	0.060	0.040	0.065
	10	0.298	0.709	0.762	0.534	0.859	0.881	0.651	0.908	0.922
	20	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	40	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
200	1	0.346	0.700	0.674	0.897	0.918	0.875	0.974	0.967	0.939
	2	0.361	0.448	0.339	0.815	0.753	0.621	0.930	0.866	0.755
	5	0.003	0.005	0.019	0.037	0.027	0.094	0.102	0.055	0.175
	10	0.648	0.975	0.983	0.835	0.992	0.996	0.902	0.996	0.997
	20	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	40	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
500	1	0.999	1.000	0.999	1.000	1.000	1.000	1.000	1.000	1.000
	2	0.988	0.984	0.959	0.999	0.998	0.993	1.000	0.999	0.997
	5	0.018	0.011	0.191	0.122	0.108	0.492	0.251	0.298	0.661
	10	0.982	1.000	1.000	0.997	1.000	1.000	0.998	1.000	1.000
	20	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	40	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Third, the results show that in general, the tests have satisfactory performance for moderate sample sizes but may not perform well in certain situations in which some or all over-identifying moment restrictions are approximately satisfied by the alternative mixing distribution. For example, one over-identifying restriction $E[x_i^{(2)}] - v_2' = 0$ is satisfied when the rectangular distribution with $\pi = 6$ is assumed. In the case that $\pi = 5$, the above over-identifying restriction is approximately satisfied and therefore, $J_n(2)$ has difficulty distinguishing between the standard gamma and the rectangular distributions even with sample sizes as large as 500, as demonstrated by the fact that the inclusion of the third moment, $J_n(3)$ is comparable to $J_n(2)$ (Table 5). In contrast, the use of the fourth moment can enhance the power (of $J_n(4)$) considerably especially when the sample size is large. Other two similar examples are the beta distribution with $\alpha = 1.0$, $\beta = 2.0$ and $\rho = 5$ (Table 4), and the log-normal distribution with $\xi = 1.0$ and $\sigma = 0.5$ (Table 6).

Fourth, the approximate slope in many cases does not pick up the test that has maximum power in finite samples. Consequently, as mentioned in Section 2.3, the criterion that q is chosen to maximize the approximate slope of the test against the alternative of interest may not be reliable.

4. An empirical example

Such specification problems arise naturally in many applications. In an insurance context, for instance, if the number of claims follows a Poisson distribution with a risk parameter λ which is assumed to itself be a random variable, the mixed model is used to describe the heterogeneity of risks that are in a single classification of the insurer.

As an empirical example, we examine the data from Johnson and Hey (1971), republished by Beard et al. (1984). Johnson's data contain 421,240 observations. They are the United Kingdom comprehensive motor policies in 1968 which are classified according to the number of claims ranging from 0 to 5, with the average number of claims per policy being 0.1317 and the sample variance 0.1385.

Table 6
Power of tests against lognormal as mixing distributions

Sample size	ξ	σ	1% test			5% test			10% test		
			Power $J_n(2)$	Power $J_n(3)$	Power $J_n(4)$	Power $J_n(2)$	Power $J_n(3)$	Power $J_n(4)$	Power $J_n(2)$	Power $J_n(3)$	Power $J_n(4)$
50	0.0	0.5	0.001	0.032	0.036	0.092	0.174	0.142	0.285	0.297	0.237
		1.0	0.363	0.349	0.336	0.444	0.420	0.403	0.494	0.468	0.447
	1.0	0.5	0.016	0.034	0.030	0.064	0.071	0.061	0.129	0.110	0.093
		1.0	0.945	0.938	0.931	0.969	0.964	0.956	0.975	0.974	0.966
	−1.0	0.5	0.000	0.006	0.246	0.003	0.095	0.140	0.079	0.255	0.241
		1.0	0.042	0.042	0.476	0.063	0.064	0.754	0.087	0.102	0.103
100	0.0	0.5	0.059	0.198	0.174	0.430	0.472	0.395	0.662	0.621	0.528
		1.0	0.579	0.566	0.549	0.645	0.651	0.629	0.723	0.704	0.679
	1.0	0.5	0.014	0.041	0.037	0.075	0.095	0.079	0.155	0.150	0.122
		1.0	0.998	0.997	0.996	1.000	0.999	0.998	1.000	1.000	0.999
	−1.0	0.5	0.000	0.142	0.208	0.235	0.498	0.484	0.598	0.689	0.633
		1.0	0.063	0.074	0.834	0.099	0.121	0.129	0.158	0.165	0.169
200	0.0	0.5	0.485	0.616	0.562	0.840	0.851	0.791	0.924	0.922	0.882
		1.0	0.818	0.804	0.788	0.881	0.860	0.843	0.905	0.891	0.876
	1.0	0.5	0.023	0.072	0.068	0.108	0.158	0.137	0.195	0.233	0.212
		1.0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	−1.0	0.5	0.332	0.744	0.751	0.985	0.932	0.917	0.969	0.976	0.957
		1.0	0.080	0.121	0.135	0.152	0.192	0.201	0.234	0.255	0.268
500	0.0	0.5	0.984	0.995	0.991	0.997	0.999	0.993	0.999	1.000	1.000
		1.0	0.991	0.990	0.988	0.996	0.995	0.993	0.998	0.996	0.995
	1.0	0.5	0.058	0.176	0.173	0.181	0.343	0.327	0.274	0.453	0.434
		1.0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	−1.0	0.5	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		1.0	0.109	0.244	0.278	0.213	0.365	0.414	0.300	0.453	0.500

Table 7
UK comprehensive motor policies in 1968

Number of claims	Observed frequencies	Fitted Poisson[a]	Fitted negative binomial[b]
0	370 412	369 246	370 460
1	46 545	48 644	46 411
2	3935	3204	4045
3	317	141	301
4	28	5	21
5	3	.	1
Parameters		$\lambda = 0.1317$	$\alpha = 2.5598$ $\beta = 0.0514$
Chi-square		543.0	6.9
Degrees of freedom		2	4
p-value		$< 1\%$	14%

[a] $Pr(N=k) = \mathrm{e}^{-\lambda}(\lambda^k/k)$, $k = 1, 2, 3, \ldots, \lambda > 0$.
[b] $Pr(N=k) = \binom{k+\alpha-1}{k}(1/1+\beta)^{\alpha}(\beta/1+\beta)^{k}$, $k = 1, 2, 3, \ldots, \alpha > 0, \beta > 0$.

Beard et al. (1984) have shown that the Poisson distribution is a poor fit because of its short tail. They, therefore, used the "over-dispersed" negative binomial distribution and anticipated that the negative binomial provides a much better fit in the tail region, an observation confirmed by the Pearson χ^2 test using the parameters based on ML estimates. According to Beard et al., the value of the Pearson χ^2 test statistic is 6.9 which gives a significance level 14% for 4 degrees of freedom. They concluded that although there is a slight indication that the negative binomial may be under-representing the tail, for most applications the model may be safely used (Table 7).

It is well known the negative binomial is a mixed Poisson with a two-parameter gamma as the mixing distribution. Here we apply the test statistic $J_n(q)$ to test the negative binomial model or equivalently the mixing distribution specified by

$$f(x) = \frac{x^{\alpha-1}\,\mathrm{e}^{-x/\beta}}{\beta^{\alpha}\Gamma(\alpha)}, \quad (\alpha > 0,\, \beta > 0;\, x > 0). \tag{13}$$

Since $\mathscr{F}$ is the set of gamma distributions with two-parameters (i.e., $p = 2$), we calculate $J_n(q)$ for $q = 3, 4$, and 5. The values of $J_n(q)$ for three values of q are 3.7130, 11.3074 and 10.2359, respectively. The test statistic based on $q = 4$ is significant at the 1%, although the test statistic for $q = 3$ is significant at the 10% but not at 1 or 5% level, and that with $q = 5$ is significant at the 5% but not at the 1% level.

Over all, the evidence is fairly strong against the negative binomial, especially in view of the result of $q = 4$ that the hypothesis is rejected at all three significance levels. The result is a complete contrast to that of Beard et al. that the Pearson χ^2 test cannot reject the negative binomial at significance levels less than 14%. We interpret this result as evidence that $J_n(q)$ is more powerful than the Pearson χ^2 test.

Several models can be considered as alternatives to the negative binomial distribution. Klugman et al. (1998) proposed the zero-modified Poisson and zero-modified geometric distributions for the Johnson's data. Here we consider the three-parameter gamma with the density function given in (14) as the mixing distribution:

$$f(x) = \frac{(x-\gamma)^{\alpha-1}\,\mathrm{e}^{-(x-\gamma)/\beta}}{\beta^{\alpha}\Gamma(\alpha)}, \quad (\alpha > 0,\, \beta > 0;\, x > \gamma). \tag{14}$$

The mixed Poisson with three-parameter gamma as the mixing distribution is clearly superior according to the results from $J_n(q)$ tests. For example, $J_n(5)$ is only 0.021 with the significance level 99.41%. The fit is almost perfect, implying that the inclusion of one more parameter is justified. Same conclusion is obtained by having other values of q.

Finally, we note that unlike the two-parameter case, if (14) is used, handling (1) with three-parameter gamma as the mixing distribution is difficult and hence, the Pearson χ^2 test and methods based on the likelihood function may be inconvenient.

5. Discussion

Although we restrict our attention to the mixed Poisson distributions, the approach can be easily modified to apply to many other situations. For example, consider the claim frequency distribution that is the sum of two or more independent components, which represent common and particular risks. This kind of model has been studied by Ruohonen (1988). Ruohonen made remarks on the difficulty of testing the model by using the Pearson χ^2 test since the mixture nature of the model, but did not make any effort to resolve the problem. For this mixture model GMM tests can be developed along similar lines as those provided in this article, though relatively large size samples may be desirable to obtain reliable results due to large values of p in mixture models, especially if one allows correlations among components. Nevertheless, the flexibility, simplicity and reliability of the GMM-based test make it a valuable tool for inference.

References

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.
Beard, R., Pentikäinen, T., Pesonen, E., 1984. Risk Theory. Chapman & Hall, London.
Bühlmann, H., 1970. Mathematical Methods in Risk Theory. Springer, Berlin.
Daykin, C.D., Pentikäinen, T., Pesonen, E., 1994. Practical Risk Theory for Actuaries. Chapman & Hall, London.
Douglas, J., 1980. Analysis with Standard Contagious Distributions. International Co-operative Publishing House, Fairland.
Fang, Y., 2003a. Semi-parametric specification tests for discrete probability models. Journal of Risk and Insurance 70, 75–86.
Fang, Y., 2003b. GMM tests for the Katz family of distributions. Journal of Statistical Planning and Inference 110, 55–73.
Geweke, J., 1981. The approximate slope of econometric tests. Econometrica 49, 1427–1442.
Haight, F., 1967. Handbook of the Poisson Distribution. Wiley, New York.
Hall, A., 2005. Generalized Method of Moments. Oxford University Press, Oxford.
Hansen, L., 1982. Large sample properties of generalized method of moment estimators. Econometrica 50, 1029–1054.
Johnson, P.D., Hey, G.B., 1971. A review of the scope for using claim histories of individual policies in risk assessment. Bull. Assoc. Roy. Actuaires Belges 66, 159–195.

Johnson, N., Kotz, S., Kemp, A., 1993. Univariate Discrete Distributions. Wiley, New York.
Klugman, S., Panjer, H., Willmot, G., 1998. Loss Models: From Data to Decisions. Wiley, New York.
Panjer, H., 1986. Actuarial Mathematics. Proceedings of Symposia in Applied Mathematics, vol. 35. American Mathematical Society, Providence, RI.
Rayner, J., Best, D., 1989. Smooth Tests of Goodness of Fit. Oxford University Press, Oxford.
Ruohonen, M., 1988. On a model for the claim number process. ASTIN Bull. 18, 57–68.
Stuart, A., Ord, J., 1987. Kendall's Advanced Theory of Statistics: Volume 1 Distribution Theory. Oxford University Press, Oxford.