

# Specification Tests for Families of Discrete Distributions with Applications to Insurance Claims Data

Yue Fang

*China Europe International Business School, Shanghai, P.R.China*

*Abstract:* Families of distributions are commonly used to model insurance claims data that require flexible distributional forms in a satisfactory manner, but the specification problem to assess the goodness-of-fit of the hypothesized model can sometimes be a challenge due to the complexity of the likelihood function of the family of distributions involved. The previous work shows that these specification problems can be attacked by means of semi-parametric tests based on generalized method of moment (GMM) estimators. While the approach can be directly applied to both discrete and continuous families of distributions, the paper focuses on developing a testing strategy within a framework of discrete families of distributions. Both the local power analysis and the approximate slope method demonstrate the excellent performance of these tests. The finite-sample performance of the tests, based on both asymptotic and bootstrap critical values, are also discussed and are compared with established methods that require the complete specification of likelihood functions.

*Keywords:* Claims data; goodness-of-fit; model specification.

## 1. Introduction

A fundamental issue in the actuarial study is selecting a probability model for the insurance claim data. Both overestimation and underestimation result in a loss for an insurance company. An overestimate implies an idle fund, whereas an underestimate may lead to higher borrowing costs and large reputational risk. Thus, an accurate estimation of the claims distribution, especially in its tail area, is crucial. To model insurance claims data, one can use parametric families of distributions, which allow considerable flexibility to provide reasonable fit to the data (see, for example, recent studies by Ahn et al. 2012; Lee et al. 2012; Eling 2012; Jeon and Kim, 2013). However, distinguishing among families of distributions can be challenging because of the complexities of the distribution functions involved. In this article, we consider a GMM-based test strategy to solve these specification problems.

The use of families of distributions to represent observed data can be traced back to the work of Karl Pearson on evolution in the 1890s. Pearson came across data sets that often exhibited substantial skewness and other systematic departures from normality. These discrepancies led to the development of the Pearson system of frequency curves. Characterizing the probability density function with a four-parameter differential equation, the family permits

the flexibility critical for a wide range of empirical studies. Families of discrete distributions are also widely used in analyses of count data that may not be adequately described by simpler models. Such cases typically occur in modeling insurance claim data in actuarial literature (Klugman et al., 2004).

The topic of specification tests for distributions has been studied extensively. Various methods exist for testing whether a particular distribution adequately describes a set of observed data. For instance, given a hypothesized distribution, several classical tests such as the likelihood ratio (LR), the Wald, and score tests can be developed (Cox and Hinkley 1974). Testing procedures for discriminating between two continuous distributions (so-called separate families of hypotheses) have been studied by Cox (1961) and further considered by, for example, Dyer (1974) with principle emphasis on procedures invariant under location and scale transformations. Another widely used test is Pearson's chi-square ( $\chi^2$ ) goodness-of-fit test. As a general purpose test of goodness-of-fit, Pearson's  $\chi^2$  test is easy to understand in principle and remains a useful tool for testing the null hypothesis that the data are from a certain probability distribution, though the asymptotic distribution of the Pearson statistic depends on the estimators employed and may lead to difficulties related to computing a non-standard limit distribution (Stuart, Ord and Arnold 1999).

The previous work discusses how these specification problems can be attacked by means of semi-parametric tests based on GMM estimators. The approach provides a general framework for dealing with a wide range of specification problems, with particular attention on choosing among discrete families of distributions. Unlike the widely applied testing procedures that rely primarily on the assumption of the density or likelihood function of the hypothesized distribution, this test does not require this sort of full knowledge; rather, it only demands the specification of a set of moment conditions the model should satisfy. Consequently, it is relatively easy to implement, particularly when the plausible hypothesized model has a complex distributional form for which likelihood-based methods are not easy to apply. Furthermore, testing results from GMM tests can often provide information regarding which models may be appropriate for data that cannot be satisfactorily modeled by alternative distributions (see, for example, Fang 2003, which used GMM-based tests to discriminate between the Poisson and (negative) binomial distributions).

To motivate and demonstrate the proposed specification tests for families of distributions, we use the Katz family of distributions as the null hypothesis in this study. The main reason for choosing the Katz family is that this two-parameter family includes the most basic counting distributions as special cases: the Poisson, negative binomial, and binomial distributions. This family usually serves as the embryonic model in the initial data analysis stage. Since the chosen model should provide a balance between simplicity and conformity to the available data, it is often of interest to know whether one of these simple distributions can adequately describe the observed data before considering more complex families of distributions. As illustrated by the empirical examples discussed in Section 5, the Katz family is relatively simple, but is also flexible enough to provide a good fit, especially to the tail area of the data.

The article is organized as follows. Section 2 provides some background on the Katz family of distributions and presents the GMM-based test strategy. Sections 3 and 4 discuss the

asymptotic and finite-sample properties of the proposed test, respectively. Two empirical examples are given to demonstrate the usefulness of the test strategy in Section 5. In Section 6, we draw conclusions.

## 2. Specification Tests for Families of Discrete Distributions

### 2.1 Characteristics for Families of Discrete Distributions

We first examine the characteristics of the Katz family of distributions and related distributions that all belong to the more general Ord's family. Let  $X$  be a discrete random variable and  $P_j = \Pr(X = j)$  be the probability that  $X$  takes the value  $j$ . The Ord's family of distributions is defined by the following relationship:

$$\frac{P_{j+1} - P_j}{P_j} = \frac{a_0 - (j+1)}{b_0 + b_1(j+1) + b_2j(j+1)}, \quad j = 0, 1, 2, \dots \quad (1)$$

where  $a_0$ ,  $b_0$ ,  $b_1$  and  $b_2$  are distribution parameters. The difference equation (1) is a discrete analogue of the Pearson system of continuous distributions, and has been thoroughly explored and studied (see, for example, Johnson, Kemp and Kotz, 2005). The special case  $b_0 = b_2 = 0$  yields the Katz family of distributions, which is specified by the following difference equation:

$$\frac{P_{j+1}}{P_j} = \frac{\alpha + \beta j}{j+1}, \quad j = 0, 1, 2, \dots \quad (2)$$

where  $\alpha = (a_0 + b_1 - 1)/b_1 > 0$  and  $\beta = (b_1 - 1)/b_1 < 1$ . The Katz family (2) is one of the most prominent families of discrete distributions whose successive probabilities satisfy first-order recurrence relations. The Katz family regards three commonly used discrete distributions as its special cases: (i) negative binomial, for  $0 < \beta < 1$ ; (ii) Poisson, for  $\beta = 0$ ; and (iii) binomial, for  $\beta < 0$  and a positive integer  $-\alpha/\beta$  (or generalized binomial, for  $\beta < 0$ ,  $\alpha > 0$ , and a non-integer  $\alpha/\beta$ ). When  $b_2 \neq 0$ ,  $\kappa = (b_1 - b_2 - 1)^2/4(a + b_0)b_2$  is defined with restriction  $\kappa \geq 0$  and  $I = \mu_2/\mu_1$  ( $\mu_r$  is the  $r$ -th central moment) is considered to yield other well-known distributions, including (iv) hypergeometric, for  $I < 1$  and  $\kappa > 1$ ; (v) beta-binomial (negative hypergeometric), for  $\kappa < 0$ ; and (vi) beta-Pascal, for  $I > 1$  and  $\kappa = 1$ . Note that distributions (iv), (v), and (vi) are not members of the Katz family.

Define  $\mu_r$  as the  $r$ -th central moment (moment about the mean) of (1) and  $\mu'_s$  as the  $s$ -th moment about the origin of (1). It is of interest to identify the characteristics of the Katz family in terms of moments to achieve the goal of developing specification tests. Although distributions in the Katz family have the property of being equi-, under-, or over-dispersed, all distributions in this family can be characterized by the following relation:

$$S = 2I - 1, \quad (3)$$

where  $I = \mu_2/\mu_1$ ,  $S = \mu_3/\mu_2$ , and  $\mu_1 = \mu'_1$  is the mean. Note that the beta-binomial distribution arises as  $S < 2I - 1$ , and  $S > 2I - 1$  gives the hypergeometric or the beta-Pascal distributions

depending on whether  $S < 1$  or  $S > 1$ . Such differences in moment conditions between the Katz family and the non-Katz family distributions have fundamental implications for GMM-based statistics which, as shown in later sections, play a central role in developing specification test statistics.

To gain further insight into the Katz family, it is helpful to rewrite (1) as

$$u_j \equiv \frac{(j+1)P_{j+1}}{P_j} = \frac{(j+1)[a_0 + b_0 + (b_1 - 1)(j+1) + b_2j(j+1)]}{b_0 + b_1(j+1) + b_2j(j+1)}. \quad (4)$$

When  $b_0 = b_2 = 0$ , the right-hand side of (4) is a linear function of  $j$ . Hence, for the Katz family of distributions, the plot of  $u_j$  against  $j$  gives a straight line with a slope less than, equal to, or greater than zero for the binomial, Poisson, and negative binomial, respectively. When  $b_0 \neq 0$  or  $b_2 \neq 0$ , the  $u_j$  curve is no longer a straight line, indicating that the underlying data-generating distribution is not a member of the Katz family.

In fact, several authors have suggested the use of (4) or its varieties in selecting a distribution from a given family. For instance, Katz (1965) pointed out that plotting  $u_j$  can help discriminate among the three distributions in the Katz family. Ord (1967) is an example of using graphical methods based on (4) for identifying distributions among the Ord's family. Note that different shapes of  $u_j$  curves for different distributions appear to be useful for developing strategies to determine whether the data are generated from the Katz family vis-à-vis other families and distributions. However, as demonstrated in Jinkinson and Slater (1981) and Hoaglin (1985), results based only on the  $u_j$  curve can be quite illusive. Therefore, these graphical methods should only be used as a starting-point for more rigorous and sophisticated inference procedures, such as the GMM-based specification test that we developed in this paper.

## 2.2 Test Statistics

Consider the null hypothesis  $H_0$  specified by the following moment restrictions:

$$E[f(x_i, \theta)] = 0, \quad (5)$$

where

$$f(x_i, \theta) = \begin{pmatrix} x_i - \mu'_1 \\ (x_i - \mu'_1)^2 - \mu_2 \\ (x_i - \mu'_1)^3 - \mu_3 \\ \dots \\ (x_i - \mu'_1)^q - \mu_q \end{pmatrix}. \quad (6)$$

As long as the focus of our attention is on the moment restrictions defined by (6), it may be convenient to define  $\theta = (\mu'_1, \mu_2)$ , rather than with the parameters  $\alpha$  and  $\beta$ . Under the hypothesis of the Katz family,  $\mu'_1 = \alpha/(1 - \beta)$  and  $\mu_2 = \alpha/(1 - \beta)^2$ . All other higher order moments can be expressed in terms of  $\mu'_1$  and  $\mu_2$ :

$$\mu_r = \sum_{j=0}^r (-1)^j \binom{r}{j} \mu'^j_1 \mu'_{r-j}, \quad (7)$$

where

$$\mu'_{s+1} \equiv E(X^{s+1}) = \sum_{j=0}^s \binom{s}{j} (\alpha \mu'_j + \beta \mu'_{j+1}).$$

When the number of orthogonality moment restrictions in (5) exceeds the number of parameters (denoted by  $p$ ;  $p = 2$  in our case), the model is over-identified. In the over-identified case, (5) implies substantive restrictions. If the hypothesis of the model that leads to (5) is incorrect in the first place, some of the sample moment restrictions will be systematically violated, providing a basis for developing a specification test. Here, we consider the following statistic to test the validity of the moment restrictions in (5):

$$J_n(q) \equiv n f_n(\hat{\theta})' \hat{V}_n^{-1} f_n(\hat{\theta}), \quad (8)$$

where  $f_n(\theta) \equiv n^{-1} \sum_{i=1}^n f(x_i, \theta)$ ;  $V_n$  is a consistent estimator of  $V = \lim_{n \rightarrow \infty} \text{var}[n^{1/2} f_n(\theta)]$ ; and  $\theta$  is evaluated at the GMM estimate  $\hat{\theta}$ , the value of  $\theta$  that minimizes the quadratic form

$$Q_n(\theta) = f_n(\theta)' V_n^{-1} f_n(\theta).$$

The test statistic (8) has an asymptotic  $\chi^2$  distribution with  $(q - 2)$  degrees of freedom under the null hypothesis of the model (Hansen, 1982). To obtain the  $q \times q$  matrix  $V_n$ , we first compute the weighting matrix  $V$ . It can be shown that the  $(i, j)$ -th element of  $V$  is  $(\mu_{i+j} - \mu_i \mu_j)$ , where the  $\mu_r$ s are given by (7). Then,  $\hat{V}_n$  is obtained when the two parameters  $\mu'_1$  and  $\mu_2$  in  $V$  are evaluated at the GMM estimates  $\hat{\mu}'_1 = \bar{x} \equiv n^{-1} \sum_{i=1}^n x_i$  and  $\hat{\mu}_2 = s^2 \equiv n^{-1} \sum_{i=1}^n (x_i - \bar{x})^2$ , respectively. With  $f(x_i, \theta)$  defined in (6) and  $V$  given above, the test statistic  $J_n(q)$  in (8) can now be calculated.

Routine matrix calculations provide little insight into the GMM test statistic. To gain a proper understanding of the test statistic, it is helpful to have the closed-form expression of  $J_n(q)$ . As a demonstration, we provide the closed-form expression of  $J_n(3)$  here. When  $q = 3$ , there are three moment restrictions in  $E[f(x_i, \theta)] = 0$ . By the construction of  $f(x_i, \theta)$  and the choice of the model parameterization, the first two equations in  $E[f(x_i, \theta)] = 0$  are identifying moment restrictions. They are satisfied by families of distributions under both the null and alternative hypotheses. The third equation is the over-identifying restriction, which is satisfied

by the Katz family but not by others such as the non-Katz family distributions. As such,  $f_n(\theta)$  becomes  $(0, 0, m_3 - \hat{\mu}_3)'$ , where  $m_3 = n^{-1} \sum_{i=1}^n (x_i - \bar{x})^3$  and  $\hat{\mu}_3 = \hat{\mu}_2(2\hat{\mu}_2/\hat{\mu}_1 - 1)$ . Hence,

$$J_n(3) = \hat{w}_{33}(m_3 - \hat{\mu}_3)^2, \quad (9)$$

where  $\hat{w}_{33}$  is the (3, 3)-th element of  $\widehat{V}_n^{-1}$  given by

$$\hat{w}_{33} = \frac{-\hat{\mu}_2^3 - \hat{\mu}_3^2 + \hat{\mu}_2\hat{\mu}_4}{\hat{\mu}_3^4 - 3\hat{\mu}_2\hat{\mu}_3^2\hat{\mu}_4 + \hat{\mu}_2^2\hat{\mu}_4^2 - \hat{\mu}_4^3 + 2\hat{\mu}_2^2\hat{\mu}_3\hat{\mu}_5 + 2\hat{\mu}_3\hat{\mu}_3\hat{\mu}_5 - \hat{\mu}_2\hat{\mu}_5^2 - \hat{\mu}_2^3\hat{\mu}_6 - \hat{\mu}_3^2\hat{\mu}_6 + \hat{\mu}_2\hat{\mu}_4\hat{\mu}_6}, \quad (10)$$

and  $\widehat{\mu}_r$  ( $r = 3, \dots, 6$ ) are functions of  $\widehat{\mu}_1^I$  and  $\widehat{\mu}_2^I$ . Observe that the condition  $(m_3 - \widehat{\mu}_3) = 0$  is the sample analogue of (3) under the null hypothesis of the Katz family (albeit with a factor  $\mu_2$ ). Hence, the equation (9) suggests that  $J_n(3)$  has power against violation of (3) for any departures from the Katz family. As long as (3) is not satisfied,  $J_n(3)$  will not follow the central  $\chi^2$  distribution.

### 3. Asymptotic Efficiency of Specification Tests

In this section, we use two alternative approaches to investigate the asymptotic properties of  $J_n(q)$ : the local power analysis and the approximate slope method. The first approach examines local or contiguous alternatives, while the second considers a non-local analysis by employing large deviation theory. These two approaches provide different insights into the asymptotic properties of the tests.

#### 3.1 Local Power Analysis

Define

$$H_0^I : DV^{-1/2}E[f(x_i, \theta)] = 0 \quad (11)$$

and

$$H_0^O : (I_q - D)V^{-1/2}E[f(x_i, \theta)] = 0, \quad (12)$$

where  $D = M(M' M)^{-1}M'$  and  $M = V^{-1/2}E[f(x_i, \theta)]$ . We consider the local power analysis by introducing the following sequences of local alternatives to  $H_0^I$  and  $H_0^O$ :

$$H_A^I : DV^{-1/2}E[f(x_i, \theta)] = n^{-1/2}\nu_I \quad (13)$$

and

$$H_A^O : (I_q - D)V^{-1/2}E[f(x_i, \theta)] = n^{-1/2}\nu_O, \quad (14)$$

where  $\nu_I \neq 0$  and  $\nu_O \neq 0$ . Since it is always possible to decompose  $V^{-1/2}E[f(x_i, \theta)]$  into

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

$H_0^I$  and  $H_0^O$  translate directly into sequences of local alternatives to  $V^{-1/2}E[f(x_i, \theta)] = 0$ , which is equivalent to  $E[f(x_i, \theta)] = 0$ . Hall (1999) showed that for local alternatives given in (11) and (12),  $J_n(q)$  follows asymptotically a non-central  $\chi^2$  distribution. Furthermore, if the data satisfy

$H_0^I \cap H_A^O$ ,  $J_n(q) \sim \chi_{q-p}^2(\nu_O' \nu_O)$ . This result suggests that when  $V_O' V_O > 0$ ,  $J_n(q)$  has power against the alternative  $H_A^O$ , which is  $\mu_3 \neq \mu_2(2\mu_2/\mu_1' - 1)$  or, equivalently,  $S \neq (2I - 1)$  when  $q = 3$  in our previous example. The result also implies that the local power of  $J_n(q)$  depends on both  $V_O' V_O$  and  $(q - 2)$ . In particular, although additional moment conditions are used in  $J_n(q)$  with higher value  $q$ , this does not necessarily imply that large values of  $q$  are more desirable, even asymptotically: increasing  $q$  increases both  $V_O' V_O$  (causing an increase in power) and the variance of  $\chi_{q-p}^2(\nu_O' \nu_O)$  (causing a decrease in power). More rigorously, as  $n \rightarrow \infty$  and  $q \rightarrow \infty$ , with  $q = q(n) = o(n)$ ,

$$\lim_{n \rightarrow \infty} \beta_\alpha(q, \nu_O) = 1 \quad \text{iff} \quad \lim_{q \rightarrow \infty} \nu_O' \nu_O / \sqrt{q} = \infty \quad (15)$$

and

$$\lim_{n \rightarrow \infty} \beta_\alpha(q, \nu_O) = \alpha \quad \text{if} \quad \lim_{q \rightarrow \infty} \nu_O' \nu_O / \sqrt{q} = 0, \quad (16)$$

where  $\beta_\alpha(q, \nu_0)$  is the power of the size- $\alpha$  test. The theoretical results (15) and (16) directly indicate the following, as  $q \rightarrow \infty$ : the asymptotic local power of the test statistic  $J_n(q)$  increases if  $\lim_{q \rightarrow \infty} \nu'_0 \nu_0 / \sqrt{q} = \infty$ , and decreases to  $\alpha$  if  $\lim_{q \rightarrow \infty} \nu'_0 \nu_0 / \sqrt{q} = 0$ . Moreover, the results emphasize that the local power of the test may differ substantially for different values of  $q$  and across alternatives. The proof of these results is based on the asymptotic normality of the properly normalized statistic  $J_n(q)$  and can be derived along the lines of Kallenberg, Oosterhoff, and Schriever (1985).

### 3.2 Approximate Slope Method

An alternative approach to analyzing the power of tests is to study its approximate slope, which has been proposed by Bahadur (1960) and further studied by Bahadur (1967) and Geweke (1981), among others. The approximate slope of a test is defined to be the rate at which the logarithm of its asymptotic marginal significance level decreases as the sample size increases. Geweke (1981) has shown that if the test statistic's limiting distribution under the null hypothesis is a  $\chi^2$  distribution, 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 the results in Geweke (1981), we have

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

For any given  $q$ , one may apply the above results to calculate  $c_q$  under a specified alternative hypothesis. Take the example in Section 2.2 and consider the case  $q = 3$ , for which the approximate slope is

$$c_3 = (0, 0, \mu_3 - \mu_2(2\frac{\mu_2}{\mu'_1} - 1)) V^{-1} \begin{pmatrix} 0 \\ 0 \\ \mu_3 - \mu_2(2\frac{\mu_2}{\mu'_1} - 1) \end{pmatrix} = w_{33} [\mu_3 - \mu_2(2\frac{\mu_2}{\mu'_1} - 1)]^2,$$

where  $w_{33}$  is the (3,3)-th element of  $V^{-1}$  and is the population moment analogue to (10) under the alternative hypothesis. For the non-Katz family distributions, one can derive the required  $\mu_s$ s (Johnson, Kemp and Kotz, 2005). Figure 1 also provides a comparison of approximate slopes of  $c_q$  across different  $q$  for various hypergeometric, beta-binomial and beta-Pascal distributions. Two observations emerge from Figure 1. First, given  $q$ ,  $c_q$  depends on both the probability structures (hypergeometric, beta-binomial or beta-Pascal) and levels of model parameters. In general,  $c_q$  increases with  $\mu_1$  for each of three distribution types. Additionally,  $c_q$  increases with  $q$ , where the increase can be considerable. This implies that including additional moment restrictions will not reduce the asymptotic efficiency of the test. For example, consider the hypergeometric distribution with parameterization  $N = 20$ ,  $M = 15$  and  $l = 5$ . The ratio of the approximate slopes for  $q = 3$  and  $q = 4$  is approximately 8 percent. This implies that about 12 times as many observations are needed to reject the alternative hypothesis when  $q$  is chosen to be 3.



With respect to the GMM approach discussed in this paper, it is possible to calculate the approximate slopes of  $J_n(q)$  for any  $q$ . This allows us to quantify the potential for efficiency gain in power from additional moment restrictions. Furthermore, the ratio of these slopes between any two test statistics, denoted A and B, will be  $T_*^B/T_*^A$ , where  $T_*$  is the minimum number of observations needed to reject the alternative hypothesis for a given power.

#### 4. Monte Carlo Simulation Studies

In this section, the finite sample properties of  $J_n(q)$  are investigated via Monte Carlo simulations. Since various previous studies suggest that GMM tests based on their asymptotic critical values often have true sizes that differ greatly from the nominal sizes (Brown and Newey, 1995; Hall and Horowitz, 1996; see also Table 1 of the present paper), tests based on both asymptotic and bootstrap critical values are examined. More technical details on GMM tests based on bootstrap critical values are provided in the Appendix. For comparison, we generated simulated data from both the Katz family of distributions (negative binomial, Poisson, and binomial distributions; formats are listed underneath Table 1) and other non-Katz family distributions (hypergeometric, beta-binomial, and beta-Pascal distributions; formats are listed underneath Table 2). For the simulations, the sample size is taken to be 50, 100, and 200. Then, simulation experiments are performed for the tests of  $J_n(q)$  for three values of  $q$ : 3, 4, and 5. The nominal significance level is chosen to be 5 percent. To compare the finite sample performance of the GMM tests with those of commonly used likelihoodbased test procedures, we perform simulation experiments for the LR tests based on the constrained (the Katz family) and non-constrained likelihood functions. For brevity, we will report results of the LR tests only for  $n = 100$ .

Table 1 presents the results of sizes of tests for the Poisson, binomial and negative binomial distributions with various parameterizations that yield means ranging from approximately 1 to 20. As shown, the empirical sizes of the tests based on the asymptotic critical values are generally below their nominal values (5%), except for the  $J_n(3)$  and  $J_n(4)$  statistics for the binomial distribution. It appears that the quality of the asymptotic approximations of the distributions of the tests deteriorates as  $q$ , the number of moment conditions, increases. One reason for this is that in the Monte Carlo study,  $q$  affects the size via two routes. First, it controls the degree of overidentification. The quality of the asymptotic approximation tends to deteriorate as the degree of overidentification increases. Second,  $q$  controls the order of moments included, and it is known that higher sample moments converge more slowly than lower moments do (Hall, 2005).

The results in Table 1 show that with the bootstrap critical values, the empirical sizes of the tests are much close to the nominal levels. This is especially true for the  $J_n(5)$  test. However, the bootstrap does not completely remove the size distortion; as the case of 50 observations demonstrates, the size of  $J_n(3)$  is still only 1.8 percent for the negative-binomial with  $\alpha = 1.25$  and  $\beta = 0.25$ . The behavior of the LR tests appears to be comparable to that of the GMM tests in the case of the Poisson distribution. However, the sizes of the LR tests are quite a bit too large for the binomial and negative-binomial distributions, a phenomenon that has been noticed before

(Vu and Maller, 1996). It may be worth noting that the algorithm used for computing the maximum likelihood estimates cannot always converge; hence, the sizes of the LR tests reported in Table 1 (and also the power in Table 2) are based on converged replications, which may be less than 3,000. Table 2 contains the results of the empirical power of tests against the hypergeometric, beta-binomial, and beta-Pascal distributions. We chose parameters that yield means similar to those of distributions investigated in the null hypothesis. As can be seen from Table 2, the tests have satisfactory performance for moderate sample sizes, but may not perform particularly well in small samples under certain situations. For both beta-binomial and beta-Pascal distributions, tests based on the bootstrap critical values appear to have greater power than corresponding tests based on the asymptotic critical values, regardless of the sample sizes. However, tests based on the asymptotic critical values seem to be more powerful than those based on the bootstrap critical values, especially when the sample size is 50 or 100. We note that by selecting the value  $q$  appropriately, the power of the GMM tests is comparable to that of the LR tests.

Finally, we remark that, in general, the actual finite-sample power performance agrees well with the approximate slopes of the tests in Figure 1, especially for relatively large sample sizes. However, the results also suggest that, although additional moment conditions are used in  $J_n(q)$  as  $q$  increases, large values of  $q$  (i.e.  $q > 4$ ) may not necessarily be more desirable, even with a relatively large sample size ( $n = 200$ ). This is an anticipated outcome following the local asymptotic results in Section 3.

## 5. Two Empirical Examples

Two empirical examples will serve to illustrate the types of situation in which the proposed test is useful.

### **Example 1. The number of claims on automobile insurance policies in Australia**

The data in the second column of Table 3 are taken from Hossack et al. (1983) and describes the distribution of the number of claims on automobile insurance policies in Australia; the average number of claims per policy is 0.125 and the variance is 0.130. Although the MME of  $\beta$  is 0.0385, which suggests that the negative binomial may be a candidate model for the data, Klugman et al. (2004) have shown that the negative binomial is a poor fit because of its short tail (see the third column of Table 3). We apply our GMM test  $J_n(q)$  to the data set for  $q = 3, 4, \dots, 9$ . As can be seen from Table 3, the test statistics with  $q = 3$  and 4 are not significant at any commonly used significance levels. The test statistic with  $q = 5$  is significant at the 10 percent level, but not at 1 or 5 percent. The test statistics based on  $q > 5$  are significant at the 1 percent level. The results provide strong evidence against the Katz family of distributions, which includes the negative binomial as a special case. Note that the data set is very large and, hence, requires a close-fitting model. The test statistic based on higher moments (i.e.  $q \geq 5$ ) provides considerable power, making a convincing case that one must look for alternative models beyond the three basic distributions in the Katz family. See Klugman et al. (2004) for more discussions on fitting the data to the Poisson-inverse Gaussian and the Poisson-ETNB distributions.

**Example 2. The number of claims on automobile insurance policies in Switzerland**

The second example considers data describing the number of claims on automobile insurance policies in Switzerland taken from Bühlmann (1970). The data are given in the second column in Table 4. Gathy and Lefèvre (2010) have shown that both the Poisson and negative binomial distributions are bad fits (see the third and fourth columns of Table 4). The Poisson is clearly a poor fit, especially for large values; while the negative binomial distribution is a slight improvement, the distribution still underestimates the heavy tail of the data. Here, we fit the data to the Katz family of distributions. As can be seen from the results reported in Table 4, the  $p$ -values of  $J_n(q)$  for all considered  $q$  are much higher than the commonly used significance levels, suggesting that the Katz family is an acceptable model to the data. Although both the Poisson and negative binomial distributions are rejected as two separate null hypotheses, the Katz family provides enough flexibility to capture the heavy tail of the data.

**6. Conclusion**

This paper develops a semi-parametric specification test for discrete families of distributions from recent developments in GMM. The testing procedure provides flexibility while maintaining simplicity. The test can be constructed systematically, making it easy to implement using matrix languages in statistical software. By studying the Katz family null vis-à-vis the non-Katz family alternative, we demonstrate that the asymptotic properties of the test can be analyzed using standard asymptotic theory. Additionally, the finite sample properties of the test were investigated via Monte Carlo simulations. The simulation results suggest that the distribution of the test is well approximated by the asymptotic theory, and that the test has satisfactory performance for moderate size samples. Two empirical examples illustrate the practical relevance of the proposed tests. In conclusion, the flexibility, simplicity and reliability of this semi-parametric GMM-based test make it a valuable tool for modeling insurance claim data.

**Appendix: Procedure of GMM Specification Tests Based on Bootstrap Critical****Values**

For each experiment, we specify a data-generation process from the Katz family of distributions, and then carry out the following steps: (1) sample the data-generation process and compute  $J_n(q)$ ; (2) obtain bootstrap samples from equ. (2) based on the estimated parameters  $\alpha$  and  $\beta$ , and compute the bootstrap critical values; (3) estimate the empirical sizes of the tests using the bootstrap critical values by repeating steps (1)–(3) 3,000 times. The bootstrap critical values are based on 100 replications of the bootstrap sampling processes. Note that greater number of replications does not affect the results substantially (Hall, 1986; Hall and Horowitz, 1996).

## References

- [1] Ahn, S., Kim, J., and Ramaswami, V. (2012). A new class of models for heavy tailed distributions in finance and insurance risk. *Insurance: Mathematics and Economics* **51** 43-52.
- [2] Bahadur, R. (1960). Stochastic comparison of tests. *Annals of Mathematical Statistics* **31** 276- 295.
- [3] Bahadur, R. (1967). Rates of convergence of estimates and test statistics. *Annals of Mathematical Statistics* **38** 303-324.
- [4] Brown, B. W., and Newey, W. K. (1995). Bootstrapping for GMM. Unpublished manuscript, Rice University, Dept. of Economics.
- [5] Bühlmann, H. (1970). *Mathematical Methods in Risk Theory*. New York: Springer-Verlag.
- [6] Cox, D. (1961). Tests of separate families of hypotheses. *Proceedings of the fourth Berkeley Symposium*, Vol.1. Berkeley: University of California Press.
- [7] Cox, D. and Hinkley, D. (1974). *Theoretical Statistics*. London: Chapman and Hall.
- [8] Dyer, A. (1974). Hypothesis testing procedures for separate families of hypotheses. *Journal of the American Statistical Association* **69** 140-145.
- [9] Eling, M. (2012). Fitting insurance claims to skewed distributions: Are the skew-normal and skew-student good models? *Insurance: Mathematics and Economics* **51** 239-248.
- [10] Fang, Y. (2003) GMM tests for the Katz family of distributions. *Journal of Statistical Planning and Inference* **110** 55-73.
- [11] Gathy, M. and Lefevre, C. (2010). On the Lagrangian Katz family of distributions as a claim frequency model. *Insurance: Mathematics and Economics* **47** 76-83.
- [12] Geweke, J. (1981). The approximate slope of econometric tests. *Econometrica* **49** 1427-1442.
- [13] Hall, P. (1986). On the number of bootstrap simulations required to construct a confidence interval. *Annals of Statistics* **14** 1453-1462.
- [14] Hall, P., and Horowitz, J. (1996). Bootstrap critical values for tests based on generalized-method-of-moments estimators. *Econometrica* **64** 891-916.
- [15] Hall, A. (1999). Hypothesis testing in models estimated by GMM. In *Generalized Method of Moments Estimation*, eds. L. Matyas, Cambridge: Cambridge University Press.

- 
- [16] Hall, A. (2005). *Generalized Method of Moments*. Oxford: Oxford University Press.
- [17] Hansen, L. (1982). Large sample properties of generalized method of moment estimators. *Econometrica* **50** 1029-1054.
- [18] Hoaglin, D. C. (1985). A Poissonness Plot. Chapter 9 in David C. Hoaglin, Frederick Mosteller, and John W. Tukey, *Exploring Data Tables, Trends, and Shapes*. New York: John Wiley & Sons.
- [19] Jeon, Y. and Kim, J. (2013). A gamma kernel density estimation for insurance loss data. *Insurance: Mathematics and Economics* **53** 569-579.
- [20] Jinkinson, R. and Slater, M. (1981). Critical discussion of a graphical method for identifying discrete distributions. *The Statistician* **30** 239-248.
- [21] Johnson, N., Kemp, A. and Kotz, S. (2005). *Univariate Discrete Distributions*, 3rd. New York: John Wiley & Sons.
- [22] Kallenberg, W., Oosterhoff, J. and Schriever, B. (1985). The number of classes in Chi-squared goodness-of-fit test. *Journal of American Statistical Association* **80** 959-968.
- [23] Katz, L. (1965). Unified treatment of a broad class of discrete probability distributions. *Proceedings of the International Symposium on Discrete Distributions*, Montreal, 175-182.
- [24] Klugman, S., Panjer, H. and Willmot, G. (2004). *Loss Models: From Data to Decisions*. New York: John Wiley & Sons.
- [25] Hossack, I., Pollard, J., and Zehnwirth, B. (1983). *Introduction Statistics with Applications in General Insurance*. Cambridge: Cambridge University Press.
- [26] Lee, D., Li, W. K., and Wong, T.S.T. (2012). Modeling insurance claims via a mixture exponential model combined with peaks-over-threshold approach. *Insurance: Mathematics and Economics* **51** 538-550.
- [27] Ord, J. (1967). On a system of discrete distributions. *Biometrika* **54** 649-656.
- [28] Stuart, A., Ord, J. and Arnold, S. (1999). *Classical Inference and the Linear Model- Kendall's Advanced Theory of Statistics*, Volume 2A, 6th. New York: Oxford University Press.
- [29] Vu, H. T. V., and Maller, R. (1996). The likelihood ratio test for Poisson versus binomial distributions. *Journal of American Statistical Association* **91** 818-824.

Table 1: Sizes of Tests: Empirical sizes of nominal 5 percent significance GMM, bootstrap GMM, and LR tests of the Katz family null hypothesis. The bootstrap critical values for the GMM tests are in parentheses.

	$\alpha$	$\beta$	$n = 50$			$n = 100$			LR	$n = 200$		
			$J_n(3)$	$J_n(4)$	$J_n(5)$	$J_n(3)$	$J_n(4)$	$J_n(5)$		$J_n(3)$	$J_n(4)$	$J_n(5)$
Poisson <sup>a</sup>	1	0	0.026 (0.052)	0.024 (0.053)	0.024 (0.059)	0.026 (0.049)	0.037 (0.056)	0.03 (0.056)	0.028	0.035 (0.048)	0.029 (0.039)	0.036 (0.037)
	2	0	0.028 (0.038)	0.033 (0.041)	0.021 (0.041)	0.036 (0.055)	0.04 (0.051)	0.02 (0.05)	0.032	0.045 (0.05)	0.037 (0.048)	0.02 (0.043)
	5	0	0.033 (0.046)	0.03 (0.046)	0.022 (0.056)	0.035 (0.047)	0.044 (0.065)	0.018 (0.065)	0.046	0.038 (0.038)	0.039 (0.048)	0.018 (0.052)
	10	0	0.034 (0.034)	0.035 (0.045)	0.019 (0.058)	0.048 (0.058)	0.037 (0.048)	0.017 (0.049)	0.052	0.05 (0.05)	0.034 (0.052)	0.026 (0.059)
	20	0	0.038 (0.044)	0.028 (0.038)	0.024 (0.053)	0.045 (0.036)	0.043 (0.044)	0.032 (0.047)	0.054	0.047 (0.044)	0.036 (0.05)	0.033 (0.047)
	Binomial <sup>b</sup>	5	-1/3	0.047 (0.049)	0.044 (0.035)	0.026 (0.039)	0.048 (0.044)	0.04 (0.046)	0.021 (0.046)	0.061	0.044 (0.045)	0.046 (0.044)
15		-1	0.058 (0.063)	0.056 (0.043)	0.042 (0.043)	0.059 (0.055)	0.062 (0.054)	0.026 (0.058)	0.096	0.057 (0.037)	0.061 (0.052)	0.034 (0.052)
45		-3	0.068 (0.032)	0.06 (0.043)	0.041 (0.044)	0.078 (0.056)	0.053 (0.047)	0.046 (0.05)	0.091	0.074 (0.05)	0.072 (0.036)	0.049 (0.045)
10		-1/3	0.046 (0.047)	0.053 (0.052)	0.036 (0.061)	0.042 (0.048)	0.049 (0.051)	0.032 (0.055)	0.078	0.041 (0.04)	0.048 (0.05)	0.029 (0.05)
30		-1	0.049 (0.046)	0.053 (0.051)	0.047 (0.064)	0.042 (0.046)	0.056 (0.045)	0.037 (0.046)	0.113	0.074 (0.07)	0.062 (0.047)	0.038 (0.044)
90		-3	0.065 (0.044)	0.034 (0.042)	0.029 (0.053)	0.077 (0.062)	0.036 (0.047)	0.031 (0.045)	0.095	0.072 (0.045)	0.042 (0.047)	0.026 (0.048)
Negative Binomial <sup>c</sup>	1.25	0.25	0.008 (0.018)	0.019 (0.042)	0.01 (0.048)	0.019 (0.033)	0.018 (0.036)	0.019 (0.036)	0.051	0.033 (0.051)	0.013 (0.043)	0.022 (0.057)
	2.5	0.5	0.009 (0.029)	0.024 (0.044)	0.011 (0.058)	0.022 (0.061)	0.03 (0.054)	0.02 (0.059)	0.053	0.035 (0.053)	0.034 (0.05)	0.021 (0.05)
	3.75	0.75	0.02 (0.048)	0.027 (0.053)	0.014 (0.054)	0.028 (0.05)	0.037 (0.045)	0.018 (0.043)	0.069	0.035 (0.036)	0.033 (0.041)	0.023 (0.059)
	3.75	0.25	0.022 (0.038)	0.018 (0.04)	0.013 (0.042)	0.038 (0.059)	0.032 (0.055)	0.013 (0.054)	0.105	0.033 (0.045)	0.041 (0.057)	0.018 (0.051)
	7.5	0.5	0.027 (0.06)	0.023 (0.044)	0.011 (0.043)	0.04 (0.064)	0.032 (0.044)	0.021 (0.041)	0.067	0.045 (0.053)	0.04 (0.053)	0.016 (0.056)
	11.25	0.75	0.03 (0.042)	0.037 (0.055)	0.014 (0.046)	0.045 (0.038)	0.034 (0.045)	0.019 (0.046)	0.072	0.044 (0.047)	0.028 (0.043)	0.022 (0.061)

a. Poisson:  $P_j = \alpha^j e^{-\alpha} / j!$ ;  $I = 1$  and  $S = 1$ .

b. Binomial:  $P_j = \binom{-\alpha/\beta}{j} \left(\frac{-\beta}{1-\beta}\right)^j \left(1 - \frac{-\beta}{1-\beta}\right)^{-\alpha/\beta-j}$ ;  $I < 1$  and  $S = 2I - 1$ .

(Generalized Binomial:  $P_j = P_j' / \sum_{i=0}^{[-\alpha/\beta]+1} P_i'$ , where  $P_j' = \frac{\Gamma(-\alpha/\beta+1)}{j! \Gamma(-\alpha/\beta+1-j)} \left(\frac{-\beta}{1-\beta}\right)^j \left(1 - \frac{-\beta}{1-\beta}\right)^{-\alpha/\beta-j}$ ).

c. Negative Binomial:  $P_j = \binom{\alpha/\beta + j - 1}{j} \beta^j (1 - \beta)^{\alpha/\beta}$ ;  $I > 1$  and  $S = 2I - 1$ .

Table 2: Power of Tests: Empirical power of nominal 5 percent significance GMM, bootstrap GMM, and LR tests of the alternative hypothesis of hypergeometric series distributions. The bootstrap critical values for the GMM tests are in in parentheses.

				n = 50			n = 100				n = 200		
				$J_n(3)$	$J_n(4)$	$J_n(5)$	$J_n(3)$	$J_n(4)$	$J_n(5)$	LR	$J_n(3)$	$J_n(4)$	$J_n(5)$
hyper-geometric <sup>a</sup>	N=20	M=10	l=5	0.322	0.501	0.442	0.383	0.629	0.553	0.525	0.508	0.754	0.708
				(0.212)	(0.344)	(0.324)	(0.342)	(0.508)	(0.429)		(0.475)	(0.684)	(0.658)
	20	10	10	0.403	0.551	0.498	0.621	0.745	0.685	0.609	0.811	0.902	0.891
				(0.305)	(0.396)	(0.324)	(0.498)	(0.586)	(0.546)		(0.77)	(0.853)	(0.818)
	20	15	5	0.511	0.692	0.63	0.626	0.806	0.809	0.839	0.806	0.948	0.94
				(0.369)	(0.484)	(0.458)	(0.505)	(0.667)	(0.681)		(0.773)	(0.879)	(0.867)
	20	15	10	0.789	0.466	0.403	0.934	0.639	0.6	0.93	0.993	0.647	0.633
				(0.537)	(0.319)	(0.458)	(0.788)	(0.51)	(0.48)		(0.919)	(0.504)	(0.576)
	40	25	20	0.301	0.151	0.12	0.443	0.277	0.221	0.541	0.723	0.494	0.403
				(0.192)	(0.13)	(0.09)	(0.376)	(0.227)	(0.193)		(0.661)	(0.671)	(0.356)
	40	25	30	0.23	0.092	0.063	0.317	0.136	0.121	0.345	0.561	0.32	0.261
				(0.211)	(0.057)	(0.09)	(0.446)	(0.129)	(0.099)		(0.797)	(0.694)	(0.227)
beta-binomial <sup>b</sup>	L=15	a=2	b=5	0.01	0.069	0.037	0.081	0.175	0.107	0.153	0.346	0.409	0.283
				(0.014)	(0.152)	(0.171)	(0.206)	(0.226)	(0.295)		(0.472)	(0.488)	(0.528)
	15	5	5	0.011	0.145	0.087	0.399	0.313	0.213	0.376	0.766	0.561	0.464
				(0.014)	(0.205)	(0.199)	(0.438)	(0.369)	(0.325)		(0.759)	(0.607)	(0.634)
	15	10	5	0.197	0.151	0.121	0.49	0.298	0.211	0.506	0.767	0.649	0.513
				(0.21)	(0.168)	(0.178)	(0.463)	(0.307)	(0.338)		(0.749)	(0.623)	(0.633)
	25	2	5	0.012	0.093	0.05	0.125	0.243	0.138	0.32	0.419	0.537	0.363
				(0.124)	(0.174)	(0.205)	(0.301)	(0.344)	(0.298)		(0.53)	(0.585)	(0.575)
	25	5	5	0.163	0.202	0.101	0.457	0.359	0.269	0.5	0.839	0.713	0.6
				(0.187)	(0.257)	(0.229)	(0.478)	(0.401)	(0.44)		(0.834)	(0.743)	(0.756)
	25	10	5	0.24	0.201	0.133	0.554	0.378	0.306	0.512	0.877	0.759	0.638
				(0.264)	(0.217)	(0.214)	(0.562)	(0.359)	(0.447)		(0.868)	(0.743)	(0.736)
beta-Pascal <sup>c</sup>	k=5	a=2	b=5	0.002	0.016	0.045	0.066	0.106	0.117	0.143	0.19	0.311	0.29
				(0.011)	(0.038)	(0.128)	(0.079)	(0.13)	(0.217)		(0.209)	(0.331)	(0.518)
	5	5	5	0.014	0.065	0.112	0.16	0.268	0.283	0.271	0.35	0.502	0.449
				(0.047)	(0.104)	(0.256)	(0.221)	(0.3)	(0.387)		(0.406)	(0.514)	(0.609)
	5	10	5	0.042	0.134	0.145	0.201	0.328	0.372	0.289	0.481	0.632	0.495
				(0.112)	(0.192)	(0.283)	(0.304)	(0.38)	(0.515)		(0.538)	(0.645)	(0.613)
	10	2	5	0.001	0.032	0.07	0.071	0.179	0.206	0.236	0.245	0.392	0.363
				(0.019)	(0.056)	(0.16)	(0.142)	(0.202)	(0.317)		(0.266)	(0.408)	(0.575)
	10	5	5	0.047	0.111	0.165	0.215	0.291	0.42	0.293	0.465	0.646	0.6
				(0.112)	(0.161)	(0.286)	(0.287)	(0.39)	(0.476)		(0.527)	(0.654)	(0.756)
	10	10	5	0.077	0.172	0.2	0.316	0.511	0.504	0.346	0.611	0.77	0.638
				(0.174)	(0.221)	(0.337)	(0.348)	(0.479)	(0.63)		(0.641)	(0.772)	(0.736)

a. Hypergeometric:  $P_j = \binom{M}{j} \binom{N-M}{l-j} / \binom{N}{l}$ ;  $l < 1$  and  $2l - 1 < S < 1$ .

b. Beta-Binomial:  $P_j = \binom{L}{j} B(j+a, L+b-j) / B(a, b)$ ;  $S < 2l - 1$

c. Beta-Pascal:  $P_j = \binom{k+j-1}{j} B(k+a, j+b) / B(a, b)$ ;  $S > \min(2l - 1, 1)$ .

Table 3: The number of claims on automobile insurance policies in Australia

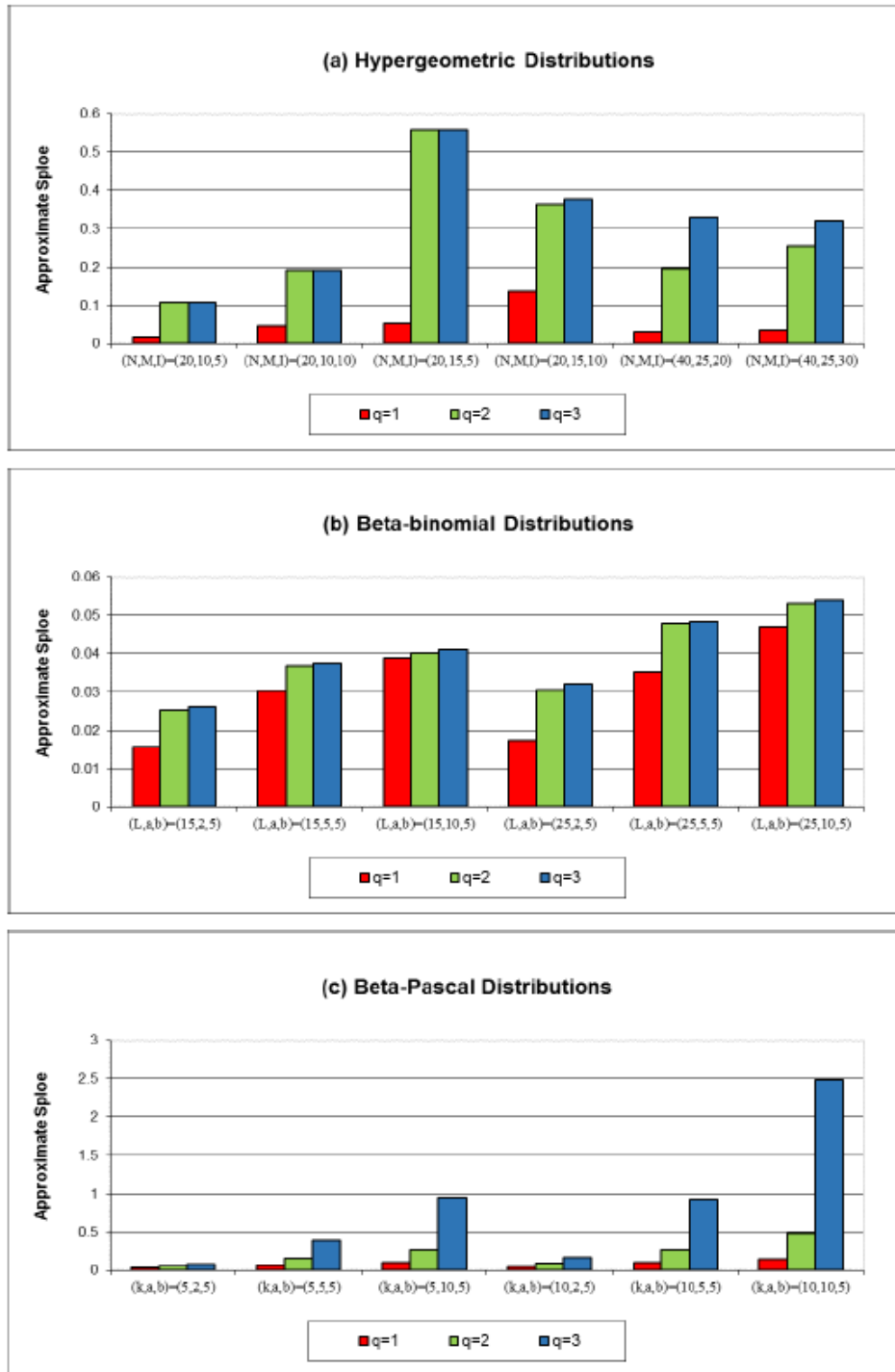
	Observed	NB	GMM Estimated Frequency						
			$q = 3$	$q = 4$	$q = 5$	$q = 6$	$q = 7$	$q = 8$	$q = 9$
0	565664	565708	565824	565941	566064	566174	566257	566312	566345
1	68714	68570	68384	68219	68053	67910	67803	67732	67692
2	5177	5317	5374	5413	5447	5473	5492	5504	5510
3	365	335	347	355	363	369	374	377	378
4	24	19	20	21	22	22	23	23	24
5	6	1	1	1	1	1	1	1	1
6	0	0	0	0	0	0	0	0	0
$\geq 7$	0	0	0	0	0	0	0	0	0
$p$ -value		< 1%	50.3%	19.9%	2.20%	< 1%	< 1%	< 1%	< 1%

Table 4: The number of claims on automobile insurance policies in Switzerland

	Observed	Poisson	NB	GMM Estimated Frequency						
				$q = 3$	$q = 4$	$q = 5$	$q = 6$	$q = 7$	$q = 8$	$q = 9$
0	103704	102630	103724	103808	103851	103882	103889	103907	103910	103911
1	14075	15922	13990	13858	13799	13759	13737	13727	13723	13722
2	1766	1235	1857	1887	1896	1902	1904	1906	1906	1906
3	255	64	245	259	263	266	268	269	269	269
4	45	3	32	36	37	38	38	38	38	38
5	6	0	4	5	5	5	5	5	5	5
6	2	0	0	1	1	1	1	1	1	1
$\geq 7$	0	0	0	0	0	0	0	0	0	0
$p$ -value		< 1%	< 1%	46.7%	35.1%	24.6%	22.5%	26.4%	34.3%	44.4%



Figure 1. The approximate slopes of  $J_n(q)$



Yue Fang  
China Europe International Business School  
Shanghai, P.R.China