

Generalized Method of Moments and Generalized Estimating Functions Based on Probability Generating Function for Count Models

Andrew Luong

École d'actuariat, Université Laval, Ste Foy, Québec, Canada

Email: andrew.luong@act.ulaval.ca

How to cite this paper: Luong, A. (2020) Generalized Method of Moments and Generalized Estimating Functions Based on Probability Generating Function for Count Models. *Open Journal of Statistics*, 10, 516-539.

<https://doi.org/10.4236/ojs.2020.103031>

Received: May 6, 2020

Accepted: June 8, 2020

Published: June 11, 2020

Copyright © 2020 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Generalized method of moments based on probability generating function is considered. Estimation and model testing are unified using this approach which also leads to distribution free chi-square tests. The estimation methods developed are also related to estimation methods based on generalized estimating equations but with the advantage of having statistics for model testing. The methods proposed overcome numerical problems often encountered when the probability mass functions have no closed forms which prevent the use of maximum likelihood (ML) procedures and in general, ML procedures do not lead to distribution free model testing statistics.

Keywords

Mixture Distributions, Consistent Chi-Square Tests, Infinitely Divisible Distributions, Mixture Distributions, Distribution Free Test Statistics, Model Testing

1. Introduction

Count data are often encountered in many fields of applications which include actuarial sciences and fitting discrete count models are of interests. Classical methods such as maximum likelihood (ML) procedures often require the probability function of the model to have closed-form and furthermore the inferences techniques do not lead to distribution free statistics when using the Pearson statistics. In fact, if a model does not fit the data, better models can be created using compound procedure, stop sum procedure or mixing procedure and the new

models might provide a better fit as they can take into account modeling processes which were omitted earlier.

For discussions on these procedures see the books by Johnson *et al.* [1], Klugman *et al.* [2] but for these better models often they do not have closed-form probability mass functions but their probability generating functions often remain simple and have closed-form expressions.

For example, if count data display long tailed behavior so that the Poisson model with probability generating function $P_\theta(s) = e^{\theta(s-1)}$, $\theta > 0$ does not provide a good fit, the positive discrete stable (DPS) distribution can be created and be used as an alternative to the Poisson distribution. The discrete positive stable distribution (DPS) does not have closed-form or simple form for probability mass function but its probability generating function is simple and given by

$$P_{\boldsymbol{\delta}}(s) = e^{\theta(s-1)^\alpha}, \boldsymbol{\delta} = (\theta, \alpha)', \alpha \in (0, 1], \theta > 0$$

see Christoph and Schreiber [3] for this distribution. In their paper, expression (6) gives the representation of the probability mass function of the DPS distribution using series,

$$p(x = k; \boldsymbol{\delta}) = (-1)^k \sum_{j=0}^{\infty} \binom{j\alpha}{k} \frac{(-\theta)^j}{j!}, k = 0, 1, \dots$$

and expression (8) gives the recursive formula to compute $p(x = k; \boldsymbol{\delta})$ using the previous terms

$$p(x = 0; \boldsymbol{\delta}), \dots, p(x = k-1; \boldsymbol{\delta})$$

with

$$(k+1)p(x = k+1; \boldsymbol{\delta}) = \theta \sum_{m=0}^k p(x = k-m; \boldsymbol{\delta}) (m+1) (-1)^m \binom{\alpha}{m+1}, k = 0, 1, \dots$$

The probability mass function appears to be complicated and for model validation there is a need for a statistic for model testing. By having these issues, it will make maximum likelihood (ML) procedures difficult to implement.

GMM procedures based on probability generating function appear to be a natural way to introduce alternatives to ML procedures, bypassing the use of the probability mass function explicitly and focus uniquely on the probability generating function. In this vein, the procedures proposed in this paper make use of GMM and generalized estimating equation theory and they are less simulation intensive oriented than inference techniques given by the paper by Luong *et al.* [4].

We shall use general GMM methodology but adapted it to situations where moment conditions are based on probability generating function so that estimation and model testing can be carried out in a unified way for discrete count models. The choice of moments of the developed GMM procedures makes use of estimating function theory which allows us the use a number of points based on probability generating which tends to infinity as the sample size $n \rightarrow \infty$. Furthermore, we also related GMM estimation with the approach using generalized

estimating equations (GEE) based on a set of elementary or basic unbiased estimating function but unlike GEE procedures, GMM procedures also provide distribution free chi-square statistics for model testing but the theory of estimating function is useful as it provides insight on the choice of sample moments for GMM estimation. In another word, the proposed methods blend classical GMM procedures and inference techniques based on estimating equation which in general will allow flexibility, efficiency and model testing yet being relatively simple to implement and might be of interests for practitioners. Consequently, the new methods differ from proposed GMM procedures in the literature from the following points:

1) GMM procedures as proposed by Doray *et al.* [5] only make use of a finite number of points of the probability generating function, our methods aim at achieving higher efficiency yet remain simple to implement and it is done by linking to the theory of estimating function, it can accommodate the use of a number of points from the probability generating function instead of being fixed, it goes to infinity as $n \rightarrow \infty$.

2) The new GMM procedures remain simpler to implement than GMM procedures using a continuum moment conditions in general as proposed by Carrasco and Florens [6] or methods on adapting GMM procedures using a continuum of moment conditions for characteristic function proposed by Carrasco and Kotchoni [7] to probability generating function. Practitioners might find the sophisticated methods based on a continuum moment conditions difficult to implement.

The paper is organized as follows. In Section 2, we review available results from general GMM theory, despite the results are not new once the moment conditions are defined but they make the paper more self-contained as these results will be adapted subsequently with moments conditions extracted from the probability generating function when count models are considered. In Section 3, GMM estimation and related GEE estimation for count models are considered. The chi-square statistics are also given in Section 3.2.2. In Section 3.2.3, we consider GMM procedures based on optimum orthogonal estimating functions. In Section 4 we illustrate the implementation of the GMM methodology and preliminary results show that the methods are simple to implement and have the potentials of being very efficient. The new methods display flexibility as they can accommodate the changes to the sample moments for better efficiencies if needed and it can be done within the framework of the inference methods developed.

2. Generalized Method of Moments (GMM) Methodology

The inferences techniques based on probability generating functions developed in this paper make use of results of Generalized Method of Moments (GMM) theory which are well established once the moment conditions are specified, see Martin *et al.* [8] (p 352-384, also see Hamilton [9]). In this section, we shall briefly review GMM methodology for estimation and moment restrictions testing to

make the paper easier to follow and connect to the problem on how to select moment conditions for discrete distributions based on probability generating functions for applying GMM methods.

The estimating equations of GMM methods will also be linked to the theory of estimating equations and generalized estimating equations (GEE) as developed by Godambe and Thompson [10], Morton [11], Liang and Zeger [12].

2.1. Generalized Estimating Equations (GEE) and GMM Estimation

For data, we shall assume that we have n independent observations y_1, \dots, y_n , these observations need not be identically distributed but each y_i will follow a distribution which depends on the same vector of parameters $\theta = (\theta_1, \dots, \theta_p)'$, $\theta \in \Omega$, Ω is compact and $\theta \in R^p$. The true vector of parameters is denoted by θ_0 .

For the time being, assume that we have identified n unbiased basic estimating functions or elementary estimating functions denoted by $h_i = h_i(y_i; \theta)$, $i = 1, \dots, n$ with the property

$$E_{\theta}(h_i(y_i; \theta)) = 0 \quad \text{for } i = 1, \dots, n. \quad (1)$$

The optimum estimating functions based on linear combinations of $\{h_i(y_i; \theta), i = 1, \dots, n\}$ for estimating θ_0 is given by

$$g^{(r)}(\theta) = \sum_{i=1}^n h_i(y_i; \theta) \frac{E_{\theta}\left(\frac{\partial h_i}{\partial \theta_r}\right)}{E_{\theta}\left((h_i)^2\right)}, r = 1, \dots, p \quad (2)$$

and $E_{\theta}\left((h_i)^2\right)$ is the variance of $h_i(y_i; \theta)$.

The vector of estimators $\hat{\theta}_{op}$ based on the optimum estimating equations are given as solutions of the system of equations $g^{(r)}(\theta) = 0, r = 1, \dots, p$. This result is given by Godambe and Thompson [10] (page 4) and Morton [11] (page 229-230).

In applications, often we restrict our attention to $h_i(y_i; \theta)$ with some common functional form so that we also use the notations $h_i(y_i; \theta) = h(y_i; \theta)$, $i = 1, \dots, n$ and more precisely $h(y_i; \theta) = h(y_i; s_i, \theta)$, s_i is a constant.

With this notation which is commonly used in the literature, notice that the random variables given by $h(y_i; \theta), i = 1, \dots, n$ need not be identically distributed.

Also, since estimating equations are defined up to a constant which does not depend on θ , the related estimating functions used can be re-expressed equivalently as

$$g^{(r)}(\theta) = \frac{1}{n} \sum_{i=1}^n h(y_i; \theta) \frac{E_{\theta}\left(\frac{\partial h(y_i; \theta)}{\partial \theta_r}\right)}{E_{\theta}\left((h(y_i; \theta))^2\right)}, r = 1, \dots, p \quad (3)$$

and the vector of estimators based on the optimum equations are given as solu-

tions of the system of equations $g^{(r)}(\boldsymbol{\theta}) = 0, r = 1, \dots, p$, using expression (3). Using vector notations, the vector of optimum estimating functions based on

$$\text{expression (3) can be expressed as } g(\boldsymbol{\theta}) = \begin{pmatrix} g_1(\boldsymbol{\theta}) \\ \vdots \\ g_r(\boldsymbol{\theta}) \end{pmatrix},$$

$$g(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^n h(y_i; \boldsymbol{\theta}) \frac{E_{\theta} \left(\frac{\partial h(y_i; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)}{E_{\theta} \left((h(y_i; \boldsymbol{\theta}))^2 \right)} \text{ and the vector of estimators } \hat{\boldsymbol{\theta}}_{op} \text{ based}$$

on $g(\boldsymbol{\theta})$ are solutions of $g(\boldsymbol{\theta}) = \mathbf{0}$ and from this observation it is clear that the factor $\frac{1}{n}$ can be omitted when defining estimating functions or equations.

$$\text{Now suppose that we have vector } \mathbf{m}(y_i; \boldsymbol{\theta}) = \begin{pmatrix} m_1(y_i; \boldsymbol{\theta}) \\ \vdots \\ m_k(y_i; \boldsymbol{\theta}) \end{pmatrix} \text{ with the property}$$

$$E_{\theta}(\mathbf{m}(y_i; \boldsymbol{\theta})) = \begin{pmatrix} E_{\theta}(m_1(y_i; \boldsymbol{\theta})) \\ \vdots \\ E_{\theta}(m_k(y_i; \boldsymbol{\theta})) \end{pmatrix} = \mathbf{0} \text{ for } i = 1, 2, \dots,$$

the optimum estimating functions for estimating $\boldsymbol{\theta}$ based on linear combinations of the elements of the set $\{m(y_i; \boldsymbol{\theta}), i = 1, \dots, n\}$ are also called generalized optimum estimating functions, see Morton [11] (p 229-230), also see expression (6) as given by Liang and Zeger [12] (page 15) are given by

$$\sum_{i=1}^n \mathbf{C}'_i(\boldsymbol{\theta}) \mathbf{V}_i^{-1}(\boldsymbol{\theta}) \mathbf{m}(y_i; \boldsymbol{\theta}) \quad (4)$$

and the estimators are given by the vector $\hat{\boldsymbol{\theta}}_{op}$ obtained by solving

$$\sum_{i=1}^n \mathbf{C}'_i(\boldsymbol{\theta}) \mathbf{V}_i^{-1}(\boldsymbol{\theta}) \mathbf{m}(y_i; \boldsymbol{\theta}) = \mathbf{0} \quad (5)$$

where $\mathbf{V}_i(\boldsymbol{\theta})$ is the covariance matrix of $\mathbf{m}(y_i; \boldsymbol{\theta})$ under $\boldsymbol{\theta}$ and its inverse $\mathbf{V}_i^{-1}(\boldsymbol{\theta})$, $\mathbf{V}_i(\boldsymbol{\theta})$ is also referred to as a working matrix in the literature of esti-

$$\text{mating equation theory and } \mathbf{C}'_i(\boldsymbol{\theta}) = \begin{pmatrix} E_{\theta} \left(\frac{\partial m_1(y_i, \boldsymbol{\theta})}{\partial \theta_1} \right) & \dots & E_{\theta} \left(\frac{\partial m_k(y_i, \boldsymbol{\theta})}{\partial \theta_1} \right) \\ \vdots & \ddots & \vdots \\ E_{\theta} \left(\frac{\partial m_1(y_i, \boldsymbol{\theta})}{\partial \theta_p} \right) & \dots & E_{\theta} \left(\frac{\partial m_k(y_i, \boldsymbol{\theta})}{\partial \theta_p} \right) \end{pmatrix}$$

which is a p by k matrix.

Clearly expression (4) is more general than expression (3) and is reduced to expression (3) when $\mathbf{m}(y_i; \boldsymbol{\theta})$ is a scalar instead of a vector.

For the studies of estimating functions, Godambe and Thompson [10] emphasized efficiency of estimating equations rather efficiency of the vector of optimum estimators $\hat{\boldsymbol{\theta}}_{op}$ obtained by solving estimating equations.

For applications, often we need the asymptotic covariance matrix of $\hat{\boldsymbol{\theta}}_{op}$. For this purpose, we use the set up for the study of generalized estimating equations

(GEE) as considered by Liang and Zeger [12] (p 15-16). Using a Taylor's expansion and results of their Theorem 2 (p 16) we can obtain the asymptotic covariance matrix of $\hat{\theta}_{op}$.

$$\sqrt{n}(\hat{\theta}_{op} - \theta_0) \xrightarrow{L} N(0, \Sigma)$$

with

$$\Sigma = \lim_{n \rightarrow \infty} n \left(\sum_{i=1}^n C_i'(\theta_0) V_i^{-1}(\theta_0) C_i(\theta_0) \right)^{-1},$$

with convergence in probability denoted by \xrightarrow{P} and convergence in distribution denoted by \xrightarrow{L} .

Therefore, the asymptotic covariance for $\hat{\theta}_{op}$ is simply $\lim_{n \rightarrow \infty} \left(\sum_{i=1}^n C_i'(\theta_0) V_i^{-1}(\theta_0) C_i(\theta_0) \right)^{-1}$ which can be estimated. A Fisher scoring algorithm as given by expression (6) as described by Liang and Zeger [12] (p 16) can be used to obtain the estimators numerically $\hat{\theta}_{op}$. The algorithm gives the j + 1-th iteration based on the previous j -th iteration as

$$\hat{\theta}_{op}^{(j+1)} = \hat{\theta}_{op}^{(j)} - \left(\sum_{i=1}^n C_i'(\hat{\theta}_{op}^{(j)}) V_i^{-1}(\hat{\theta}_{op}^{(j)}) C_i(\hat{\theta}_{op}^{(j)}) \right)^{-1} m(y_i; \hat{\theta}_{op}^{(j)}).$$

Other numerical techniques to obtain $\hat{\theta}_{op}$ can be used. For example, we can consider solving the system of equations given by expression (4) and expression (5) as $g^{(r)}(\theta) = 0, r = 1, \dots, p$ and $\hat{\theta}_{op}$ can be obtained by minimizing $\sum_{r=1}^p \left(g^{(r)}(\theta) \right)^2$, techniques for minimization can be used to obtain $\hat{\theta}_{op}$ numerically.

Now we turn our attention to GMM estimation methodology and we observe that the set of estimating equations using expression (2) can be reobtained using a GMM estimation set up. GMM estimation is based on the use of a k moment conditions specified by a vector function

$$m(y_i; \theta) = \begin{pmatrix} m_1(y_i; \theta) \\ \vdots \\ m_k(y_i; \theta) \end{pmatrix}$$

with its expectation with the property

$$E_{\theta}(m(y_i; \theta)) = \begin{pmatrix} E_{\theta}(m_1(y_i; \theta)) \\ \vdots \\ E_{\theta}(m_k(y_i; \theta)) \end{pmatrix} = 0 \quad \text{for } i = 1, 2, \dots \quad (6)$$

The sample moments being the counterparts of

$$\begin{pmatrix} E_{\theta}(m_1(y_i; \theta)) \\ \vdots \\ E_{\theta}(m_k(y_i; \theta)) \end{pmatrix}$$

are defined as $g^{(r)}(\theta) = \frac{1}{n} \sum_{i=1}^n m_r(y_i; \theta), r = 1, \dots, k$ and define the vector of sample moments as

$$\mathbf{g}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{g}^{(1)}(\boldsymbol{\theta}) \\ \vdots \\ \mathbf{g}^{(k)}(\boldsymbol{\theta}) \end{pmatrix} = \frac{1}{n} \sum_{i=1}^n \mathbf{m}(y_i; \boldsymbol{\theta}). \quad (7)$$

Now we need a positive definite symmetric matrix or a positive definite matrix symmetric with probability one which is denoted by $\hat{\mathbf{S}}^{-1}$ to define a quadratic form using $\mathbf{g}^{(r)}(\boldsymbol{\theta}), r=1, \dots, k$, $\hat{\mathbf{S}}^{-1}$ will be defined subsequently and this allows the objective function

$$Q(\boldsymbol{\theta}) = \mathbf{g}'(\boldsymbol{\theta}) \hat{\mathbf{S}}^{-1} \mathbf{g}(\boldsymbol{\theta})$$

to be formed for GMM estimation and the GMM estimators are given by the vector $\hat{\boldsymbol{\theta}}$ which minimizes $Q(\boldsymbol{\theta})$.

We shall define the matrix \mathbf{S} first then its estimate is $\hat{\mathbf{S}}$ from which we can obtain its inverse $\hat{\mathbf{S}}^{-1}$. In fact \mathbf{S} can be viewed as the limit as $n \rightarrow \infty$ of the covariance matrix of the vector $\frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{m}(y_i; \boldsymbol{\theta}_0) = \sqrt{n} \mathbf{g}(\boldsymbol{\theta}_0)$ and the co-

variance matrix of $\frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{m}(y_i; \boldsymbol{\theta}_0)$ can be seen as given by

$\frac{1}{n} \sum_{i=1}^n E_{\boldsymbol{\theta}_0} \left(\left[\mathbf{m}(y_i; \boldsymbol{\theta}_0) \right] \left[\mathbf{m}(y_i; \boldsymbol{\theta}_0) \right]' \right)$, then \mathbf{S} and its estimate $\hat{\mathbf{S}}$ can be defined respectively as

$$\mathbf{S} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n E_{\boldsymbol{\theta}_0} \left(\left[\mathbf{m}(y_i; \boldsymbol{\theta}_0) \right] \left[\mathbf{m}(y_i; \boldsymbol{\theta}_0) \right]' \right)$$

and with a preliminary consistent estimate $\hat{\boldsymbol{\theta}}^{(0)}$ for $\boldsymbol{\theta}_0$ then we can define

$$\hat{\mathbf{S}} = \frac{1}{n} \sum_{i=1}^n E_{\hat{\boldsymbol{\theta}}^{(0)}} \left(\left[\mathbf{m}(y_i; \hat{\boldsymbol{\theta}}^{(0)}) \right] \left[\mathbf{m}(y_i; \hat{\boldsymbol{\theta}}^{(0)}) \right]' \right) \text{ or}$$

$$\hat{\mathbf{S}} = \frac{1}{n} \sum_{i=1}^n \left[\mathbf{m}(y_i; \hat{\boldsymbol{\theta}}^{(0)}) \right] \left[\mathbf{m}(y_i; \hat{\boldsymbol{\theta}}^{(0)}) \right]'$$

and $\hat{\mathbf{S}}$ is positive definite with probability one and clearly $\hat{\mathbf{S}}$ is symmetric, its inverse is $\hat{\mathbf{S}}^{-1}$ which exists with probability one. Despite that these two expressions for $\hat{\mathbf{S}}$ are asymptotically equivalent but for numerical implementations of the methods in finite samples, the matrix

$\hat{\mathbf{S}} = \frac{1}{n} \sum_{i=1}^n E_{\hat{\boldsymbol{\theta}}^{(0)}} \left(\left[\mathbf{m}(y_i; \hat{\boldsymbol{\theta}}^{(0)}) \right] \left[\mathbf{m}(y_i; \hat{\boldsymbol{\theta}}^{(0)}) \right]' \right)$ has more chance to be invertible.

Under suitable differentiability assumptions imposed on the vector function $\mathbf{g}(\boldsymbol{\theta})$, the GMM estimators given by $\hat{\boldsymbol{\theta}}$ is consistent and has an asymptotic multivariate normal distribution, *i.e.*,

$$\hat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}_0$$

and

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{L} N(\mathbf{0}, \mathbf{V}).$$

The asymptotic covariance of $\hat{\boldsymbol{\theta}}$ is simply $\text{Acov}(\hat{\boldsymbol{\theta}}) = \frac{1}{n} \mathbf{V}$ and \mathbf{V} depends on

θ_0 so we also use the notation, $V = V(\theta_0)$ and $V = (D'(\theta_0)S^{-1}D(\theta_0))^{-1}$ with

$$D'(\theta_0) = \lim_{n \rightarrow \infty} \begin{bmatrix} E_{\theta_0} \left(\frac{\partial g^{(1)}(\theta_0)}{\partial \theta_1} \right) & \cdots & E_{\theta_0} \left(\frac{\partial g^{(k)}(\theta_0)}{\partial \theta_1} \right) \\ \vdots & \ddots & \vdots \\ E_{\theta_0} \left(\frac{\partial g^{(1)}(\theta_0)}{\partial \theta_p} \right) & \cdots & E_{\theta_0} \left(\frac{\partial g^{(k)}(\theta_0)}{\partial \theta_p} \right) \end{bmatrix} \text{ and } D'(\theta_0) \text{ is a } p \text{ by } k$$

matrix, its transpose is $D(\theta_0)$. Since $V = V(\theta_0)$, an estimate of $V(\theta_0)$ is

$$\hat{V} = (D'(\hat{\theta})\hat{S}^{-1}D(\hat{\theta}))^{-1}, D(\hat{\theta}) = \frac{\partial g(\hat{\theta})}{\partial \theta'}.$$

Using \hat{V} , the asymptotic covariance matrix of $\hat{\theta}$ can be estimated.

We also notice that we can recover optimum estimating equations estimators using the following GMM estimation set-up by letting $k = p$, i.e., the number of sample moments is equal to the number of parameters to be estimated and

$$g^{(r)}(\theta) = \frac{1}{n} \sum_{i=1}^n m_r(y_i; \theta), r = 1, \dots, p$$

with

$$m_r(y_i; \theta) = h(y_i; \theta) \frac{E_{\theta} \left(\frac{\partial h(y_i; \theta)}{\partial \theta_r} \right)}{E_{\theta} \left((h(y_i; \theta))^2 \right)}.$$

Minimizing the corresponding GMM objective function yields the vector of GMM estimators which are given by the following system of equations since \hat{S}^{-1} is positive definite with probability one,

$$g^{(r)}(\theta) = \frac{1}{n} \sum_{i=1}^n h(y_i; \theta) \frac{E_{\theta} \left(\frac{\partial h(y_i; \theta)}{\partial \theta_r} \right)}{E_{\theta} \left((h(y_i; \theta))^2 \right)} = 0, r = 1, \dots, p \quad (8)$$

which is the same system of equations for obtaining the optimum estimating equations estimators as discussed. Using vector notations, the vector of optimum

estimating functions is simply $g(\theta) = \frac{1}{n} \sum_{i=1}^n h(y_i; \theta) \frac{E_{\theta} \left(\frac{\partial h(y_i; \theta)}{\partial \theta} \right)}{E_{\theta} \left((h(y_i; \theta))^2 \right)}$ and the

related estimators are obtained by solving $g(\theta) = 0$.

The estimating equations based on of GMM procedures are based on partial derivatives of $Q(\theta)$ and can be seen as equivalent to

$$\sum_{i=1}^n D'(\theta) \hat{S}^{-1} m(y_i; \theta) = 0. \quad (9)$$

Observe that the vector of estimating functions is also formed based on linear combinations of elements of $\{m(y_i; \theta), i = 1, \dots, n\}$ which is similar to the vector of optimum estimating functions but it might not be optimum as the matrix $D(\theta)$

and the matrix \mathbf{S} , $\hat{\mathbf{S}}^{-1}$ no longer depends on i . With $\mathbf{m}(y_i; \boldsymbol{\theta}), i = 1, \dots, n$ not only being independent but they are also identically distributed then we have the equivalence of the two methods. We also notice that $\hat{\mathbf{S}}$ used for GMM estimation plays a similar role as the working matrix $\mathbf{V}_i(\boldsymbol{\theta})$ for GEE estimation but it is often simpler to obtain $\hat{\mathbf{S}}$ than $\mathbf{V}_i(\boldsymbol{\theta})$. Often, more derivations are needed to obtain $\mathbf{V}_i(\boldsymbol{\theta})$.

Based on expression (7) and the observation just made concerning expression (8), we shall define $\mathbf{m}(y_i; \boldsymbol{\theta})$ for GMM slightly different than $\mathbf{m}(y_i; \boldsymbol{\theta})$ used for generalized estimating functions (GEE) by letting for GMM estimation

$$m_r(y_i; \boldsymbol{\theta}) = g^{(r)}(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^n h(y_i; \boldsymbol{\theta}) \frac{E_{\boldsymbol{\theta}} \left(\frac{\partial h(y_i; \boldsymbol{\theta})}{\partial \theta_r} \right)}{E_{\boldsymbol{\theta}} \left((h(y_i; \boldsymbol{\theta}))^2 \right)}, r = 1, \dots, p$$

for the first p components of the vector $\mathbf{m}(y_i; \boldsymbol{\theta})$ depending on the models being studied.

We might also want to consider including other $m_r(y_i; \boldsymbol{\theta})$ for $r > p$ depending on the model being studied for the sake of efficiency, *i.e.*, this leads to define

$$\mathbf{g}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{g}_1(\boldsymbol{\theta}) \\ \mathbf{g}_2(\boldsymbol{\theta}) \end{pmatrix}$$

$$\text{with } \mathbf{g}_1(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^n h(y_i; \boldsymbol{\theta}) \frac{E_{\boldsymbol{\theta}} \left(\frac{\partial h(y_i; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)}{E_{\boldsymbol{\theta}} \left((h(y_i; \boldsymbol{\theta}))^2 \right)} \text{ which is the vector of optimum}$$

estimating function based on elements of the set $\{h(y_i; \boldsymbol{\theta}), i = 1, \dots, n\}$ and $\mathbf{g}_2(\boldsymbol{\theta})$ with its components depend upon $m_r(y_i; \boldsymbol{\theta})$ for $r > p$ to be defined based on the model under investigation, and define the GMM objective function as

$$Q(\boldsymbol{\theta}) = \mathbf{g}'(\boldsymbol{\theta}) \hat{\mathbf{S}}^{-1} \mathbf{g}(\boldsymbol{\theta}),$$

see Section 3 for more details for the choice of $\mathbf{g}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{g}_1(\boldsymbol{\theta}) \\ \mathbf{g}_2(\boldsymbol{\theta}) \end{pmatrix}$ for GMM methods with models based on probability generating functions.

One advantage of the GMM approach over generalized estimating equations (GEE) approach is with GMM approach, we have an objective function to be minimized and it leads to construction of chi-square tests for moment restrictions meanwhile there is no such equivalent test statistic if we use the generalized estimating equations approach. Furthermore, we shall see in Section 3 when applied to discrete distributions with moment conditions extracted from probability generating function, testing for moment restrictions can be viewed as testing goodness-of-fit for the count model being used. Consequently, estimation and model testing can be treated in a unified way using this approach.

As mentioned earlier, the GMM objective function evaluated at $\hat{\boldsymbol{\theta}}$ can be used to construct a test statistic which follows an asymptotic chi-square distribu-

tion for testing the null hypothesis which specify the validity of the vector moment conditions, *i.e.*,

$$H_0 : E_{\theta} \left(m(y_i; \theta) \right) = \begin{pmatrix} E_{\theta} \left(m_1(y_i; \theta) \right) \\ \vdots \\ E_{\theta} \left(m_k(y_i; \theta) \right) \end{pmatrix} = 0 \quad \text{for } i = 1, \dots \quad (10)$$

but we need $k > p$, *i.e.*, the number of sample moments must exceed the number of parameters to be estimated.

2.2. Testing the Validity of Moment Restrictions

We notice that since $g(\theta_0) \xrightarrow{p} 0$ and the vector of GMM estimators is consistent with $\hat{\theta} \xrightarrow{p} \theta_0$ and in general $g(\hat{\theta}) \xrightarrow{p} 0$, the following statistics can be constructed and will have an asymptotic chi-square distributions. These statistics are also known as Hansen's statistics after Hansen's seminal works, see Hansen [13] and they can be used for testing the validity of moment restrictions.

$$\text{For testing the simple hypothesis } H_0 : E_{\theta_0} \left(m(y_i; \theta_0) \right) = \begin{pmatrix} E_{\theta_0} \left(m_1(y_i; \theta_0) \right) \\ \vdots \\ E_{\theta_0} \left(m_k(y_i; \theta_0) \right) \end{pmatrix} = 0$$

for $i = 1, \dots, k$; θ_0 is specified, the Hansen's statistic is given as

$$nQ(\theta_0) = ng'(\theta_0) \hat{S}^{-1} g(\theta_0)$$

and the asymptotic distribution of the statistic is chi-square with k degree of freedom, *i.e.*, $nQ(\theta_0) \xrightarrow{L} \chi_k^2$ under H_0 .

For testing the composite hypothesis

$$H_0 : E_{\theta} \left(m(y_i; \theta) \right) = \begin{pmatrix} E_{\theta} \left(m_1(y_i; \theta) \right) \\ \vdots \\ E_{\theta} \left(m_k(y_i; \theta) \right) \end{pmatrix} = 0 \quad \text{for } i = 1, \dots, k; \quad \theta \in \Omega$$

We need to obtain $\hat{\theta}$ first by minimizing $Q(\theta)$ then the Hansen's statistic is given as

$$nQ(\hat{\theta}) = ng'(\hat{\theta}) \hat{S}^{-1} g(\hat{\theta})$$

and the asymptotic distribution of the statistic is chi-square with $k - p$ degree of freedom, *i.e.*, $nQ(\hat{\theta}) \xrightarrow{L} \chi_{k-p}^2$ under H_0 , assuming $k > p$.

These statistics will be used subsequently with moment conditions extracted from the model probability generating function in Section 3. We shall show in the next sections that these statistics are consistent test statistics in general for model testing with the discrete model specified by its probability generating function. These statistics are also distribution free. The distribution free property is not enjoyed by goodness-of-fit test statistics for model testing based on the empirical probability function which is defined as

$$P_n(s) = \frac{1}{n} \sum_{i=1}^n s^{X_i} \quad (11)$$

with X_1, \dots, X_n being independent and identically random variables from a discrete model specified by the model probability generating function

$P_{\theta}(s) = E_{\theta}(s^X)$ which are given by Rueda and O'Reilly [14], Marcheselli *et al.* [15] as the null distributions of the statistics depend on the unknown parameters. In addition, the procedures as proposed by Doray *et al.* [5] only make use of k fixed points s_1, \dots, s_k to generate moment conditions regardless of the sample size n .

The procedures proposed in this paper are different as the number of points selected from the probability generating function goes to infinity as $n \rightarrow \infty$.

3. GEE and GMM Methods with Moment Conditions from Probability Generating Function

In this section, we shall give attention to count models and we shall assume that we have a random sample of n independent and identically distributed observations X_1, \dots, X_n which follow the same distribution as X and X follows a non-negative integer discrete distribution with probability mass function $p(x; \theta)$ with no closed form but with model probability generating function $P_{\theta}(s) = E_{\theta}(s^X)$ with closed form and relatively simple to handle, $P_{\theta}(s)$ is well defined on the domain of $s \in [-1, 1]$.

It is well known that in general, the probability mass function is uniquely characterized by its corresponding probability generating function. Subsequently, two versions of GMM objective functions will be introduced based on estimating function theory. The first version is based on using points of $s \in [0, 1]$ to form moment conditions which are commonly used in the literature and it is given in Section 3.2.1 and Section 3.2.2, the second version is based on $s \in [-1, 1]$ and it is given in Section 3.2.3.

Optimum estimating functions can be used to obtain estimators but we emphasize here the GMM approach as tests for moment restrictions with asymptotic chi-square distribution free can also be obtained which can be interpreted as goodness-of-fit tests for the parametric family used. However, optimum estimating functions theory is very useful for identifying sample moments for efficiency of GMM procedures.

3.1. Generalized Estimating Functions (GEE)

First, we shall define the basic unbiased estimating functions $\{h(x_i; \theta)\}$, *i.e.*, with the property $E_{\theta}(h(x_i; s_i, \theta)) = 0$, then we shall form the optimum estimating functions based on linear combination of these elementary estimating functions. Since the basic elementary estimating functions are unbiased estimating function, the optimum estimating functions will be unbiased.

For each observation X_i , we shall associate the value

$$s_i = \frac{i-1/2}{n} \text{ for } i = 1, \dots, n.$$

As $n \rightarrow \infty$, the set $\{s_i; i = 1, \dots, n\}$ will become dense in $[0, 1]$ and define the

elementary estimating functions as

$$h(x_i; \theta) = h(x_i; s_i, \theta) = s_i^{X_i} - P_\theta(s_i); i = 1, \dots, n$$

and clearly $E_\theta(h(x_i; s_i, \theta)) = 0$.

Since $h(x_i; s_i, \theta)$ is independent of $h(x_j; s_j, \theta)$ for $i \neq j$, we have the property

$$E_\theta(h(x_i; s_i, \theta)h(x_j; s_j, \theta)) = 0, \text{ for } i \neq j. \quad (12)$$

The elements of the set $\{h(x_i; \theta), i = 1, \dots, n\}$ are said to be mutually orthogonal if elements of the set have the property as defined by expression (12), see Godambe and Thompson [10] (page 139). Therefore, using Godambe and Thompson [10] (page 139) optimality criteria the optimum estimating functions for estimating θ_0 based on linear combination of the basic estimating functions which are orthogonal are given by

$$g^{(r)}(\theta) = \frac{1}{n} \sum_{i=1}^n h(x_i; s_i, \theta) \frac{E_\theta\left(\frac{\partial h(x_i; s_i, \theta)}{\partial \theta_r}\right)}{E_\theta\left(\left(\{h(x_i; s_i, \theta)\}\right)^2\right)}, \quad (13)$$

and clearly $E_\theta(g^{(r)}(\theta)) = 0, r = 1, \dots, p$, the optimum estimating functions are also unbiased.

We define the vector

$$\beta_i(\theta) = \frac{E_\theta\left(\frac{\partial h(x_i; s_i, \theta)}{\partial \theta}\right)}{E_\theta\left(\left(\{h(x_i; s_i, \theta)\}\right)^2\right)}.$$

Since $E_\theta\left(\frac{\partial h(x_i; s_i, \theta)}{\partial \theta_r}\right) = -\frac{\partial P_\theta(s_i)}{\partial \theta_r}$ and letting $v_\theta(h(x_i; s_i, \theta))$ be the variance of $h(x_i; s_i, \theta)$, so

$$E_\theta(h(x_i; s_i, \theta)h(x_i; s_i, \theta)) = v_\theta(h(x_i; s_i, \theta))$$

and

$$v_\theta(h(x_i; s_i, \theta)) = E_\theta(s_i^{X_i} s_i^{X_i}) - (E_\theta(s_i^{X_i}))^2 = P_\theta(s_i^2) - (P_\theta(s_i))^2.$$

This implies

$$\beta_i(\theta) = \frac{E_\theta\left(\frac{\partial h(x_i; s_i, \theta)}{\partial \theta}\right)}{E_\theta\left(\left(\{h(x_i; s_i, \theta)\}\right)^2\right)} = \frac{-\frac{\partial P_\theta(s_i)}{\partial \theta}}{P_\theta(s_i^2) - (P_\theta(s_i))^2}. \quad (15)$$

Therefore, equivalently the vector of optimum estimating function is given by

$$\frac{1}{n} \sum_{i=1}^n h(x_i; s_i, \theta) \beta_i(\theta). \quad (16)$$

For GEE estimation as given by expression (4) and expression (5), we need to

specify the vector $\mathbf{m}(x_i; \boldsymbol{\theta})$. Let us partition $\mathbf{m}(x_i; \boldsymbol{\theta})$ into two components with

$$\mathbf{m}(x_i; \boldsymbol{\theta}) = \begin{pmatrix} m_1(x_i; \boldsymbol{\theta}) \\ m_2(x_i; \boldsymbol{\theta}) \end{pmatrix}, \quad m_1(x_i; \boldsymbol{\theta}) = s_i^{X_i} - P_{\boldsymbol{\theta}}(s_i), i = 1, \dots, n. \quad (17)$$

We select two points t_1 and t_2 , for example by letting $t_1 = 0.50$ and $t_2 = 0.75$ and therefore we can form two sets of elementary basic unbiased estimating functions using these two points which are given by

$$\{t_1^{X_i} - P_{\boldsymbol{\theta}}(t_1), i = 1, \dots, n\} \quad \text{and} \quad \{t_2^{X_i} - P_{\boldsymbol{\theta}}(t_2), i = 1, \dots, n\}. \quad (18)$$

These two sets of elementary unbiased estimating function are selected because as we shall see when used to form moment conditions for the GMM objective function, they allow the construction of consistent chi-square tests.

Furthermore, with the probability generating function we can derive the expectation of X which is denoted by $\mu(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}}(X)$ and another set of elementary unbiased estimating function can be created which is given by $\{X_i - \mu(\boldsymbol{\theta}), i = 1, \dots, n\}$ and since the sample mean if incorporated into estimating equations in general might help to improve the efficiency of the estimators, this set of estimating functions are also being considered and used for forming the vector of generalized estimating functions. By making use of these three sets of elementary unbiased estimating functions lead us to define

$$\mathbf{m}_2(x_i; \boldsymbol{\theta}) = \begin{pmatrix} t_1^{X_i} - P_{\boldsymbol{\theta}}(t_1) \\ t_2^{X_i} - P_{\boldsymbol{\theta}}(t_2) \\ X_i - \mu(\boldsymbol{\theta}) \end{pmatrix}, i = 1, \dots, n.$$

provided for the model $\mu(\boldsymbol{\theta})$ exists and note that $\mu(\boldsymbol{\theta})$ can be obtained from the derivative of the probability generating function, in fact $\mu(\boldsymbol{\theta}) = P'_{\boldsymbol{\theta}}(1)$,

$$P'_{\boldsymbol{\theta}}(s) = \frac{dP_{\boldsymbol{\theta}}(t)}{dt}.$$

If $\mu(\boldsymbol{\theta})$ does not exist, the last component of $\mathbf{m}_2(x_i; \boldsymbol{\theta})$ is replaced by $X_i t_3^{X_i-1} - P'_{\boldsymbol{\theta}}(t_3)$ with t_3 close to 1, say $t_3 = 0.95$ for example and see section 4 for an illustration and for finding the working matrix $V_i(\boldsymbol{\theta})$. For estimators to have a multivariate asymptotic normal distribution, we also need the existence of the common variance of $X_i, i = 1, \dots, n$ under the model.

Having specified the vector

$$\mathbf{m}(x_i; \boldsymbol{\theta}) = \begin{pmatrix} m_1(x_i; \boldsymbol{\theta}) \\ m_2(x_i; \boldsymbol{\theta}) \end{pmatrix},$$

GEE estimation can be performed using results and procedures of Section 2.1, the vector of GEE estimators $\hat{\boldsymbol{\theta}}_{op}$ is obtained by solving the system of equations as given by expression (5). Observe that with the notations being introduced, $m_1(x_i; \boldsymbol{\theta})$ denote a function which also depends on s_i , i.e., $m_1(x_i; \boldsymbol{\theta}) = m_1(x_i; s_i, \boldsymbol{\theta})$ and clearly $m_1(x_i; \boldsymbol{\theta}), i = 1, \dots, n$ are not identically distributed but $m_2(x_i; \boldsymbol{\theta}), i = 1, \dots, n$ are identically distributed vectors of ran-

dom variables. Therefore, GEE estimators are no longer asymptotically equivalent to GMM estimators using the same vectors $\mathbf{m}(x_i; \boldsymbol{\theta})$. With the notations being used, GEE estimators and GMM estimators are asymptotically equivalent only if $\mathbf{m}(x_i; \boldsymbol{\theta}), i = 1, \dots, n$ have a common multivariate distribution.

3.2. GMM Methodology

Before defining the sample moment vector $\mathbf{m}(y_i; \boldsymbol{\theta})$ for GMM methods, let us for the time being turn our attention on how to obtain a preliminary consistent estimate $\hat{\boldsymbol{\theta}}^{(0)}$ in general. Such a preliminary estimate $\hat{\boldsymbol{\theta}}^{(0)}$ is needed for numerical algorithms to implement GMM procedures and to define the matrix $\hat{\mathbf{S}}^{-1}$ which is used to define the GMM objective function. The nonlinear least-squares (NLS) estimators can be used to obtain a preliminary consistent estimate $\hat{\boldsymbol{\theta}}^{(0)}$ with $\hat{\boldsymbol{\theta}}^{(0)}$ being the vector which minimizes

$$\frac{1}{n} \sum_{i=1}^n \left(h(x_i; s_i, \boldsymbol{\theta}) \right)^2.$$

Note that the estimating functions of the nonlinear least-squares methods are

$$\frac{1}{n} \sum_{i=1}^n h(x_i; s_i, \boldsymbol{\theta}) \frac{\partial h(x_i; s_i, \boldsymbol{\theta})}{\partial \theta_r}, r = 1, \dots, p,$$

and they have some resemblance to the optimum ones as they are also based on linear combinations of $h(x_j; s_j, \boldsymbol{\theta}), j = 1, \dots, n$ but they are not optimum.

3.2.1. GMM Objective Function

Now we turn our attention to defining the vector

$$\mathbf{m}(x_i; \boldsymbol{\theta}) = \begin{pmatrix} \mathbf{m}_1(x_i; \boldsymbol{\theta}) \\ \mathbf{m}_2(x_i; \boldsymbol{\theta}) \end{pmatrix}.$$

we have seen GMM estimators are no longer equivalent to GEE estimators if we define $\mathbf{m}(x_i; \boldsymbol{\theta})$ as for GEE methods, some modifications appear to be necessary and to ensure that GMM estimators have comparable efficiencies to the ones obtained by using optimum estimating functions based on $\{h(x_i; \boldsymbol{\theta}), i = 1, \dots, n\}$, we shall let

$$\mathbf{m}_1(x_i; \boldsymbol{\theta}) = h(x_i; s_i, \boldsymbol{\theta}) \boldsymbol{\beta}_i(\boldsymbol{\theta}), i = 1, \dots, n$$

with the corresponding sample moment, $\mathbf{g}_1(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^n h(x_i; s_i, \boldsymbol{\theta}) \boldsymbol{\beta}_i(\boldsymbol{\theta})$,

$\mathbf{g}_1(\boldsymbol{\theta})$ is the vector of optimum function based on

$$\{h(x_i; s_i, \boldsymbol{\theta}), i = 1, \dots, n\}$$

and keeping $\mathbf{m}_2(x_i; \boldsymbol{\theta}), i = 1, \dots, n$ as for GEE estimation, so the corresponding sample moment vector for GMM estimation is

$$\mathbf{g}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{g}_1(\boldsymbol{\theta}) \\ \mathbf{g}_2(\boldsymbol{\theta}) \end{pmatrix} \quad (19)$$

with $\mathbf{g}_1(\boldsymbol{\theta})$ being just defined and

$$\mathbf{g}_2(\boldsymbol{\theta}) = \frac{1}{n} \begin{pmatrix} \sum_{i=1}^n (t_1^{X_i} - P_{\boldsymbol{\theta}}(t_1)) \\ \sum_{i=1}^n (t_2^{X_i} - P_{\boldsymbol{\theta}}(t_2)) \\ \sum_{i=1}^n (X_i - \mu(\boldsymbol{\theta})) \end{pmatrix}$$

if $\mu(\boldsymbol{\theta})$ exists, otherwise let $\mathbf{g}_2(\boldsymbol{\theta}) = \frac{1}{n} \begin{pmatrix} \sum_{i=1}^n (t_1^{X_i} - P_{\boldsymbol{\theta}}(t_1)) \\ \sum_{i=1}^n (t_2^{X_i} - P_{\boldsymbol{\theta}}(t_2)) \\ X_i t_3^{X_i-1} - P'_{\boldsymbol{\theta}}(t_3) \end{pmatrix}$, t_3 is chosen to

close to 1 but $t_3 < 1$. The GMM objective function can be constructed and given by

$$Q(\boldsymbol{\theta}) = \mathbf{g}'(\boldsymbol{\theta}) \hat{\mathbf{S}}^{-1} \mathbf{g}(\boldsymbol{\theta}).$$

3.2.2. Model Testing Using GMM Objective Function

Now we shall turn our attention to the problem of testing a model specified by its probability generating function. Let X_1, \dots, X_n be the random sample drawn from the nonnegative integer discrete distribution with probability generating function $P_0(t)$ and we want to test the following simple null hypothesis which specifies $P_0(t) = P_{\boldsymbol{\theta}_0}(t)$, $\boldsymbol{\theta}_0$ is specified, *i.e.*

$$H_0 : P_0(t) = P_{\boldsymbol{\theta}_0}(t)$$

and clearly if $H_0 : P_0(t) = P_{\boldsymbol{\theta}_0}(t)$ is true we have $E_{\boldsymbol{\theta}_0}(g(\boldsymbol{\theta}_0)) = 0$.

The following chi-square statistics

$$nQ(\boldsymbol{\theta}_0) \xrightarrow{L} \chi_r^2 \text{ with } r = k$$

For practical applications, the chi-square tests are in general consistent to detect common departure that we are interested as we shall see that if

$P_0(t) \neq P_{\boldsymbol{\theta}_0}(t)$, the test will allow us to reject $H_0 : P_0(t) = P_{\boldsymbol{\theta}_0}(t)$ in general as $n \rightarrow \infty$. Indeed, we have this property via the chi-square statistic, because if $P_0(t) \neq P_{\boldsymbol{\theta}_0}(t)$ the chi-square statistic will converge to infinity.

In order not to have this property, we must have

$$P_0(t) \neq P_{\boldsymbol{\theta}_0}(t) \text{ but } \mathbf{g}(\boldsymbol{\theta}_0) \xrightarrow{p} \mathbf{0}.$$

If $\mathbf{g}(\boldsymbol{\theta}_0) \xrightarrow{p} \mathbf{0}$ then two of its components given by

$\frac{1}{n} \sum_{i=1}^n (t_1^{X_i} - P_{\boldsymbol{\theta}_0}(t_1)), \frac{1}{n} \sum_{i=1}^n (t_2^{X_i} - P_{\boldsymbol{\theta}_0}(t_2))$ must simultaneously converge to 0 in probability, *i.e.*,

$$\frac{1}{n} \sum_{i=1}^n (t_1^{X_i} - P_{\boldsymbol{\theta}_0}(t_1)) \xrightarrow{p} 0 \text{ and } \frac{1}{n} \sum_{i=1}^n (t_2^{X_i} - P_{\boldsymbol{\theta}_0}(t_2)) \xrightarrow{p} 0. \quad (20)$$

We shall show that in general for $P_0(t)$ encountered for applications it cannot happen.

Suppose that

$$\frac{1}{n} \sum_{i=1}^n (t_1^{X_i} - P_{\boldsymbol{\theta}_0}(t_1)) \xrightarrow{p} 0, \text{ this implies } P_0(t_1) = P_{\boldsymbol{\theta}_0}(t_1)$$

and similarly

$$\frac{1}{n} \sum_{i=1}^n (t_2^{X_i} - P_{\theta_0}(t_2)) \xrightarrow{p} 0, \text{ this implies } P_0(t_2) = P_{\theta_0}(t_2).$$

Observe that in general for probability generating function $P(s)$ used for applications, the function $P(s)$ is convex for $0 < s < 1$ and $P(t) = 1$ when $t = 1$, i.e., $P(1) = 1$, see Resnick [16] (p 22-23).

Furthermore, for $P_0(t) \neq P_{\theta_0}(t)$ encountered in applications, we also have in general $P_0(t) \neq P_{\theta_0}(t)$ for some $t \in (0, 1)$. This also means in general, there is only one point a with $0 < a < 1$ at most where $P_0(t)$ crosses $P_{\theta_0}(t)$ since $P_0(t)$ and $P_{\theta_0}(t)$ are both strictly convex functions and $P_0(1) = P_{\theta_0}(1) = 1$. Therefore, we cannot have simultaneously convergence as given by expression (19) and the chi-square test is consistent in general as it can detect common departure from $H_0 : P_0(t) = P_{\theta_0}(t)$ as $n \rightarrow \infty$.

For testing the composite $H_0 : P_0(t) \in \{P_{\theta}(t)\}$, we need to estimate θ_0 by $\hat{\theta}$ by minimizing $Q(\theta) = \mathbf{g}'(\theta) \hat{\mathbf{S}}^{-1} \mathbf{g}(\theta)$ first and subsequently use $\hat{\theta}$ to compute the following chi-square statistic $nQ(\hat{\theta})$ and $nQ(\hat{\theta}) \xrightarrow{L} \chi_r^2$ with $r = 3$.

These chi-square statistics are distribution free as there is no unknown parameter in these chi-square distributions for the statistics used. These goodness-of-fit tests are simpler to implement than the ones based on matching sample probability generating function with its model counterpart using a continuum of moment conditions as given by Theorem 10 of Carrasco and Florens [6] (p 812-813). Note that maximum likelihood estimators if used concomitantly with the common classical Pearson statistics often have complicated distributions and the statistics are no longer distribution free, see Chernoff and Lehmann [17], Luong and Thompson [18] and these classical Pearson's test statistics are not consistent in general.

3.2.3. Further Extensions: The Use of Orthogonal Estimating Functions

Notice that beside the set of basic estimating functions

$$\{h(x_i; s_i) = s^{X_i} - P_{\theta}(s_i), i = 1, \dots, n\}$$

as defined earlier we also have another set of basic estimating functions given by $\{l(x_i; s_i, \theta), i = 1, \dots, n\}$ with $l(x_i; s_i, \theta) = (-s_i)^{X_i} - P_{\theta}(-s_i)$.

Consequently, if in addition of the first set of estimating functions, we also want to incorporate the second set of basic estimating functions for building $\mathbf{g}(\theta)$ then we can use optimum orthogonal estimating functions and instead of the first p components of the vector $\mathbf{g}(\theta)$ are given by the vector

$$\frac{1}{n} \sum_{i=1}^n h(x_i; s_i, \theta) \beta_i(\theta)$$

which is the vector optimum estimating functions based on the set of basic estimating functions $\{h(x_i; s_i, \theta), i = 1, \dots, n\}$, we shall use a more general vector of optimum estimating functions which can incorporate a larger set of basic esti-

mating functions as described below.

Observe that we also have another set of estimating functions given by $\{l(x_i; s_i, \theta), i = 1, \dots, n\}$ with $l(x_i; s_i, \theta) = (-s_i)^{x_i} - P_\theta(s_i)$ and clearly $\{l(x_i; s_i, \theta), i = 1, \dots, n\}$ form a mutually orthogonal basic estimating function but together combining the two sets of basic estimating functions to form the set

$$\{h(x_i; s_i, \theta), l(x_i; s_i, \theta), i = 1, \dots, n\},$$

the basic estimating functions of the combined set are not mutually orthogonal because $E_\theta(h(x_i; s_i, \theta)l(x_i; s_i, \theta))$ is not equal to 0. Using Gram-Schmidt orthogonalizing procedure we can replace $l(x_i; s_i, \theta)$ by

$$l^0(x_i; s_i, \theta) = l(x_i; s_i, \theta) - \alpha_i(\theta)h(x_i; s_i, \theta), i = 1, \dots, n,$$

$$\alpha_i(\theta) = \frac{E_\theta(h(x_i; s_i, \theta)l(x_i; s_i, \theta))}{E_\theta(h(x_i; s_i, \theta)h(x_i; s_i, \theta))}$$

which can also be represented as

$$\alpha_i(\theta) = \frac{P_\theta(-s_i^2) - P_\theta(s_i)P_\theta(-s_i)}{P_\theta(s_i^2) - (P_\theta(s_i))^2} \quad (21)$$

Since

$$\begin{aligned} E_\theta(h(x_i; s_i, \theta)l(x_i; s_i, \theta)) &= E_\theta(s_i^{x_i}(-s_i)^{x_i}) - E_\theta(s_i^{x_i})E_\theta((-s_i)^{x_i}) \\ &= P_\theta(-s_i^2) - P_\theta(s_i)P_\theta(-s_i) \end{aligned}$$

and $E_\theta(h(x_i; s_i, \theta)h(x_i; s_i, \theta))$ is simply the variance $v_\theta(h(x_i; s_i, \theta))$ of $h(x_i; s_i, \theta)$ since the basic estimating functions are unbiased,

$$v_\theta(h(x_i; s_i, \theta)) = P_\theta(s_i^2) - (P_\theta(s_i))^2.$$

Now, it is easy to see that that set

$$\{h(x_i; s_i, \theta), l^0(x_i; s_i, \theta), i = 1, \dots, n\}$$

is a set of mutually orthogonal of basic or elementary estimating functions, see Definition 2.2 and Theorem 2.1 as given by Godambe and Thompson [10] (p 139-140).

Li and Turtle [19] (p 177) also use a similar orthogonalization procedure and Theorem 2.1 for creating optimum estimating functions for ARCH model.

The first p components of the vector of sample moment functions $\mathbf{g}(\theta)$ are simply the optimum estimating functions based on linear combinations of basic estimating functions of the set $\{h(x_i; s_i, \theta), l^0(x_i; s_i, \theta), i = 1, \dots, n\}$ and again using Theorem 2.1 by Godambe and Thompson [10], the vector of optimum estimating functions is given by

$$\frac{1}{n} \sum_{i=1}^n (h(x_i; s_i, \theta)\beta_i(\theta) + l^0(x_i; s_i, \theta)\gamma_i(\theta)) \quad (22)$$

with

$$\gamma_i(\theta) = \frac{E_{\theta} \left(\frac{\partial l^0(x_i; s_i, \theta)}{\partial \theta} \right)}{E_{\theta} \left(\left(l^0(x_i; s_i, \theta) \right)^2 \right)}$$

and $\beta_i(\theta)$ is as defined by expression (15).

Now we shall display the expression for $\gamma_i(\theta)$, first note that

$$E_{\theta} \left(\frac{\partial l^0(x_i; s_i, \theta)}{\partial \theta} \right) = -\frac{\partial P_{\theta}(-s_i)}{\partial \theta} + \alpha_i(\theta) \frac{\partial P_{\theta}(s_i)}{\partial \theta}$$

and $E_{\theta} \left(\left(l^0(x_i; s_i, \theta) \right)^2 \right) = v_{\theta}(l^0(x_i; s_i, \theta))$, so that

$$\begin{aligned} v_{\theta}(l^0(x_i; s_i, \theta)) &= v_{\theta}(l(x_i; s_i, \theta)) + \alpha_i^2(\theta) v_{\theta}(h(x_i; s_i, \theta)) \\ &\quad - 2\alpha_i(\theta) \text{cov}_{\theta}(h(x_i; s_i, \theta), l(x_i; s_i, \theta)) \end{aligned}$$

with the variance

$$\begin{aligned} v_{\theta}(l(x_i; s_i, \theta)) &= E_{\theta} \left((-s_i)^{x_i} (-s_i)^{x_i} \right) - E_{\theta} \left((-s_i)^{x_i} \right) E_{\theta} \left((-s_i)^{x_i} \right) \\ &= P_{\theta}(s_i^2) - (P_{\theta}(-s_i))^2 \end{aligned}$$

$\alpha_i(\theta)$ is as given by expression (21) and the covariance

$$\text{cov}_{\theta}(h(x_i; s_i, \theta), l(x_i; s_i, \theta)) = P_{\theta}(-s_i^2) - P_{\theta}(s_i) P_{\theta}(-s_i)$$

The expression for $\gamma_i(\theta)$ can be displayed fully and it is given by

$$\frac{-\frac{\partial P_{\theta}(-s_i)}{\partial \theta} + \alpha_i(\theta) \frac{\partial P_{\theta}(s_i)}{\partial \theta}}{P_{\theta}(s_i^2) - (P_{\theta}(-s_i))^2 + \alpha_i^2(\theta) (P_{\theta}(s_i^2) - (P_{\theta}(s_i))^2) - 2\alpha_i(\theta) (P_{\theta}(-s_i^2) - P_{\theta}(s_i) P_{\theta}(-s_i))}.$$

with this vector of optimum estimating functions, the sample moments function $\mathbf{g}(\theta)$ for forming the corresponding GMM objective function can be defined and given below.

Let

$$\mathbf{g}_1(\theta) = \left(\frac{1}{n} \sum_{i=1}^n (h(x_i; s_i, \theta) \beta_i(\theta) + l^0(x_i; s_i, \theta) \gamma_i(\theta)) \right)$$

and keeping $\mathbf{g}_2(\theta)$ as the component vector of $\mathbf{g}(\theta)$ as specified by expression (19), so $\mathbf{g}(\theta) = \begin{pmatrix} \mathbf{g}_1(\theta) \\ \mathbf{g}_2(\theta) \end{pmatrix}$ and the choice of $\mathbf{g}(\theta) = \begin{pmatrix} \mathbf{g}_1(\theta) \\ \mathbf{g}_2(\theta) \end{pmatrix}$ with the use

of optimum orthogonal estimating functions constructed using two set of basic estimating functions for $\mathbf{g}_1(\theta)$ is to be preferred for improving the efficiency for estimation for some models if $\mathbf{g}(\theta) = \begin{pmatrix} \mathbf{g}_1(\theta) \\ \mathbf{g}_2(\theta) \end{pmatrix}$ as defined by expression

(19) in Section 3.1.1 does not give satisfactory results for efficiency for GMM estimation. Model testing procedures using this GMM objective function are identical to procedures for GMM objective function used earlier.

We might also want to enlarge the vector of $\mathbf{g}_2(\theta)$ by adding more compo-

nents but more components also tend to create numerical difficulties because the matrix \hat{S} will be nearly singular and the numerical inversion of such a matrix is often problematic.

Finally, we note that the GMM methods developed although are primarily for discrete distributions, the methods can also accommodate nonnegative continuous defined using Laplace transforms as discussed in Luong [20] as Laplace transforms are related to probability generating functions.

4. An Example and Numerical Illustrations

We shall use an example to illustrate the procedures, let us consider a random sample of observations X_1, \dots, X_n is drawn from the Poisson distribution with probability generating function $P_\theta(s) = e^{\theta(s-1)}$, $\theta > 0$. For this model θ is scalar. We would like to use GMM methods here as despite that maximum likelihood estimator for θ is available and given by $\hat{\theta}_{ML} = \bar{X}$, using $\hat{\theta}_{ML}$ does not lead to tractable distribution free goodness of fit test statistics with the use of Pearson type statistics as mentioned earlier.

For this model, the coefficient

$$\beta_i(\theta) = \frac{E_\theta \left(\frac{\partial h(x_i; s_i, \theta)}{\partial \theta} \right)}{E_\theta \left(\left(\{h(x_i; s_i, \theta)\} \right)^2 \right)} = \frac{(1-s_i)e^{\theta(s_i-1)}}{e^{\theta(s_i^2-1)} - e^{2\theta(s_i-1)}},$$

$$h(x_i; s_i) = s^{X_i} - P_\theta(s_i), i = 1, \dots, n.$$

We consider the case with the sample moment vector given by

$$g(\theta) = \begin{pmatrix} g_1(\theta) \\ g_2(\theta) \end{pmatrix}, \quad g_1(\theta) = \frac{1}{n} \sum_{i=1}^n (s_i - e^{\theta(s_i-1)}) \beta_i(\theta), \quad s_i = \frac{i-1/2}{n}$$

$$g_2(\theta) = \begin{pmatrix} \frac{1}{n} \sum_{i=1}^n (t_1^{X_i} - P_\theta(t_1)) \\ \frac{1}{n} \sum_{i=1}^n (t_2^{X_i} - P_\theta(t_2)) \\ \frac{1}{n} \sum_{i=1}^n (X_i - \mu(\theta)) \end{pmatrix}, \quad \mu(\theta) = \theta, t_1 = 0.5, t_2 = 0.75$$

The vector

$$\mathbf{m}(x_i; s_i, \theta) = \begin{pmatrix} m_1(x_i; s_i, \theta) \\ \vdots \\ m_4(x_i; s_i, \theta) \end{pmatrix}$$

will have four components with the components given respectively by

$$m_1(x_i; s_i, \theta) = (s_i^{X_i} - e^{\theta(s_i-1)}) \beta_i(\theta),$$

$$m_2(x_i; s_i, \theta) = (t_1^{X_i} - P_\theta(t_1)),$$

$$m_3(x_i; s_i, \theta) = (t_2^{X_i} - P_\theta(t_2)),$$

$$m_4(x_i; s_i, \theta) = (X_i - \mu(\theta)).$$

We can use $\hat{\theta}^{(0)} = \hat{\theta}_{ML}$ as $\hat{\theta}_{ML}$ is simple to obtain here and can be used as a preliminary consistent estimate. Now we can let

$$\hat{\mathbf{S}} = \frac{1}{n} \sum_{i=1}^n \left(\mathbf{m}(x_i; s_i, \hat{\theta}^{(0)}) \right) \left(\mathbf{m}(x_i; s_i, \hat{\theta}^{(0)}) \right)' \quad (23)$$

or

$$\hat{\mathbf{S}} = \frac{1}{n} \sum_{i=1}^n E_{\hat{\theta}^{(0)}} \left(\left[\mathbf{m}(y_i; \hat{\theta}^{(0)}) \right] \left[\mathbf{m}(y_i; \hat{\theta}^{(0)}) \right]' \right). \quad (24)$$

The elements of $\hat{\mathbf{S}}$ as given by expression (24) can be computed using only the probability generating function of the model since we have

$$E_{\theta}(Xt^X) = tP'_{\theta}(t) \quad \text{and} \quad E_{\theta}(t_1^X t_2^X) = P_{\theta}(t_1 t_2),$$

$$E_{\theta}(Xt^X s^X) = (t+s)P'_{\theta}(t+s),$$

the variance of X is

$$v_{\theta}(X) = P''_{\theta}(1) + P'_{\theta}(1) - (P'_{\theta}(1))^2, \quad P''_{\theta}(t) = \frac{d^2 P_{\theta}(t)}{dt^2}.$$

For the Poisson model,

$$v_{\theta}(X) = \theta.$$

$\hat{\mathbf{S}}$ as given by expression (24) tends to be invertible with less numerical difficulties.

The GMM objective function is given by

$$Q(\theta) = \mathbf{g}'(\theta) \hat{\mathbf{S}}^{-1} \mathbf{g}(\theta),$$

minimizing it allows us to obtain the corresponding GMM estimators $\hat{\theta}$. In order to obtain an estimated asymptotic variance for $\hat{\theta}$, we can define

$$\hat{\mathbf{D}}' = \left(\frac{1}{n} \sum_{i=1}^n E_{\theta} \left(\frac{\partial m_1(x_i; s_i, \theta)}{\partial \theta} \right), \dots, \frac{1}{n} \sum_{i=1}^n E_{\theta} \left(\frac{\partial m_4(x_i; s_i, \theta)}{\partial \theta} \right) \right)$$

evaluated at $\theta = \hat{\theta}$.

The asymptotic variance for $\hat{\theta}$ can be estimated as $\frac{1}{n} \hat{\mathbf{D}}' \hat{\mathbf{S}}^{-1} \hat{\mathbf{D}}$ and the chi-square statistic for testing the composite hypothesis $H_0: P_0(s) \in \{P_{\theta}(s)\}$ is given by $nQ(\hat{\theta})$ and $nQ(\hat{\theta}) \xrightarrow{L} \chi_3^2$.

For testing the feasibility of GMM methods with this example, limited simulation studies are conducted. GMM methods can be implemented without numerical difficulties for $\theta \leq 10$.

For values of θ with $10 < \theta \leq 100$, if expression (23) is used for $\hat{\mathbf{S}}$, the matrix $\hat{\mathbf{S}}$ tends to be nearly singular and the elements of $\hat{\mathbf{S}}$ need to be computed with higher accuracies in order to be able to invert $\hat{\mathbf{S}}$. We found that software like Maple or Mathematica is more able to compute with higher accuracy than R .

Often by using a spectral decomposition of \hat{S} , we can obtain \hat{S}^{-1} numerically although directly asking for the inverse using R , it might just give the message, matrix is nearly singular and does not return the inverse. As can be seen by using the spectral representation of \hat{S} ,

$$\hat{S} = P' \Lambda P$$

with P' being an orthonormal matrix, $P'P = I$, $P' = P^{-1}$ and Λ is a diagonal matrix with diagonal elements consist of eigenvalues of \hat{S} and these eigenvalues need to be computed with high accuracies and they also must be positive numerically, so by keeping more digits to compute the eigenvalues of \hat{S} then in general, Λ^{-1} can be obtained and computed as

$$\hat{S}^{-1} = P' \Lambda^{-1} P.$$

if expression (24) is used instead of expression (23) for \hat{S} with software which keeps more accuracy on computing numbers then we encounter less numerical problems to invert \hat{S} . For models which \hat{S}^{-1} is difficult to obtain, an empirical likelihood (EL) approach based on the same sample moments can be used and have the same efficiency as GMM methods but the numerical computations for implementing EL methods are also more involved, see Luong [20] on the use of penalty function for obtaining EL estimators.

We simulate $M = 100$ samples of size $n = 100$ from the Poisson distribution with $\theta = 1, 2, 3, 4, 5, 10, 100$ and obtain respectively the GMM estimate, the NLS estimate and the ML estimate. The NLS estimate is the non linear least-squares estimate as mentioned in the beginning of Section 3.2.

For comparison of relative efficiencies of these methods we estimate the ratios $\frac{\text{MSE}(\text{GMM})}{\text{MSE}(\text{ML})}$ and $\frac{\text{MSE}(\text{NLS})}{\text{MSE}(\text{ML})}$ where $\text{MSE}(\text{GMM})$, $\text{MSE}(\text{NLS})$, $\text{MSE}(\text{ML})$ are respectively the estimates of mean square error of GMM estimator, NLS estimator and ML estimator using simulated samples. The efficiency of GMM estimator is practically identical to the efficiency of ML estimator but the efficiency on NLS estimator is much lower and getting worse as θ increases in comparison with ML estimator. The results are displayed in Table A1.

In order to test whether the chi-square test has power to detect departure from the model used here we use the negative binomial with mean equals to θ and variance equals to $\theta + \frac{\theta^2}{\alpha}$ as departure from the Poisson model with

$\alpha = 1, 2, 3, 4, 5, 10, 100$ and simulate $M = 100$ samples of size $n = 100$ and the model used is Poisson with mean θ . We can estimate the power of the tests at these alternative and results are displayed in Table A2. The level used for the chi-square tests is $\alpha = 0.05$ with the critical point being the 0.95th percentile of a chi-square distribution with 3 degree of freedom, $\chi_{0.95}^2(3) = 7.814$. The results obtained are also encouraging and show that the chi-square tests have considerable power to detect departures. As n becomes large the estimate power also decreases as expected since as $n \rightarrow \infty$, the negative binomial distribution also

tends to the Poisson distribution. Larger scale simulation studies with more parametric families are needed to confirm the efficiencies of the proposed methods.

5. Conclusion

At this point, we can conclude that the methods appear to be relatively simple to implement and have the potentials to be efficient for some count models and have the advantage of only using of probability generating function instead of probability mass function, allowing inferences to be made for a much larger class of parametric families without relying on extensive use of simulations. The proposed GMM methodology also combines traditional GMM methodology with generalized estimating function methodology and both of these methodologies are well-known alternatives to ML methodology. There is a lack of statistics for model testing when using generalized estimating function methodology and it is overcome by the proposed procedures.

Acknowledgements

The helpful and constructive comments of a referee which lead to an improvement of the presentation of the paper and support from the editorial staff of Open Journal of Statistics to process the paper are all gratefully acknowledged.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

- [1] Johnson, N.L., Kotz, S. and Kemp, A.W. (1992) *Univariate Discrete Distributions*. Second Edition, Wiley, New York.
- [2] Klugman, S.A., Panjer, H.H. and Willmot, G.E. (2019) *Loss Models: From Data to Decisions*. Fifth Edition, Wiley, New York.
- [3] Christoph, G. and Schreiber, K. (1998) Discrete Stable Random Variables. *Statistics and Probability Letters*, **37**, 243-247.
[https://doi.org/10.1016/S0167-7152\(97\)00123-5](https://doi.org/10.1016/S0167-7152(97)00123-5)
- [4] Luong, A., Bilodeau, C. and Blier-Wong, C. (2018) Simulated Minimum Hellinger Distance Inference Methods for Count Data. *Open Journal of Statistics*, **8**, 187-219.
<https://doi.org/10.4236/ojs.2018.81012>
- [5] Doray, L.G., Jiang, S.M. and Luong, A. (2009) Some Simple Method of Estimation for the Parameters of the Discrete Stable Distribution with the Probability Generating Function. *Communications in Statistics—Simulation and Computation*, **38**, 2004-2017. <https://doi.org/10.1080/03610910903202089>
- [6] Carrasco, M. and Florens, J.-P. (2000) Generalization of GMM to a Continuum of Moment Conditions. *Econometric Theory*, **16**, 797-834.
<https://doi.org/10.1017/S0266466600166010>
- [7] Carrasco, M. and Kotchoni, R. (2017) Efficient Estimation Using Characteristic Function. *Econometric Theory*, **33**, 479-526.

- <https://doi.org/10.1017/S0266466616000025>
- [8] Martin, V., Hurn, S. and Harris, D. (2013) *Econometric Modelling with Time Series: Specification, Estimation and Testing*. Cambridge University Press, Cambridge.
 - [9] Hamilton, J.D. (1994) *Time Series Analysis*. Princeton University Press, Princeton.
 - [10] Godambe, V.P. and Thompson, M.E. (1989) An Extension of Quasi-Likelihood Estimation. *Journal of Statistical Planning and Inference*, **22**, 137-152.
[https://doi.org/10.1016/0378-3758\(89\)90106-7](https://doi.org/10.1016/0378-3758(89)90106-7)
 - [11] Morton, R. (1981) Efficiency of Estimating Equations and the Use of Pivots. *Biometrika*, **68**, 227-233. <https://doi.org/10.1093/biomet/68.1.227>
 - [12] Liang, K.Y. and Zeger, S.L. (1986) Longitudinal Data Analysis Using Generalized Linear Models. *Biometrika*, **73**, 13-22. <https://doi.org/10.1093/biomet/73.1.13>
 - [13] Hansen, L. (1982) Large Sample Properties of Generalized Method of Moment. *Econometrica*, **50**, 1029-1054. <https://doi.org/10.2307/1912775>
 - [14] Rueda, R. and O'Reilly, F. (1999) Tests of Fit for Discrete Distributions Based on the Probability Generating Function. *Communication in Statistics—Simulation and Computation*, **28**, 259-274. <https://doi.org/10.1080/03610919908813547>
 - [15] Marcheselli, M., Baccini, A. and Barabes, L. (2008) Parameter Estimation for the Discrete Stable Family. *Communications in Statistics*, **37**, 815-830.
<https://doi.org/10.1080/03610920701570298>
 - [16] Resnick, S. (1992) *Adventures in Stochastic Processes*. Birkhauser, Boston.
 - [17] Chernoff, H. and Lehmann, E.L. (1954) The Use of Maximum Likelihood Estimates in Chi-Square Tests for Goodness of Fit. *Annals of Mathematical Statistics*, **25**, 579-586. <https://doi.org/10.1214/aoms/1177728726>
 - [18] Luong, A. and Thompson, M.E. (1987) Minimum Distance Methods Based on Quadratic Distance for Transforms. *Canadian Journal of Statistics*, **15**, 239-251.
<https://doi.org/10.2307/3314914>
 - [19] Li, D.X. and Turtle, H.J. (2000) Semi-Parametric ARCH Models: An Estimating Function Approach. *Journal of Business and Economic Statistics*, **18**, 174-186.
<https://doi.org/10.1080/07350015.2000.10524860>
 - [20] Luong, A. (2017) Maximum Entropy Empirical Likelihood Methods Based on Laplace Transforms for Nonnegative Continuous Distributions with Actuarial Applications. *Open Journal of Statistics*, **7**, 459-482.
<https://doi.org/10.4236/ojs.2017.73033>

Appendix

Table A1. Estimate relative efficiency comparisons between GMM, NLS and ML estimators.

θ	1	2	3	4	5	10	100
Estimate $\frac{\text{MSE}(\text{NLS})}{\text{MSE}(\text{ML})}$	0.989	0.999	1.000	1.000	1.000	1.000	1.000
Estimate $\frac{\text{MSE}(\text{NLS})}{\text{MSE}(\text{ML})}$	1.477	2.193	1.803	1.913	2.481	3.975	454.200

$M = 100$ simulated samples are used and each with sample size $n = 100$.

Table A2. Estimate power of the chi-square tests using the Poisson model with parameter θ .

α	1	2	3	4	5	10	100
Estimate power	0.770	0.720	0.740	0.620	0.580	0.210	0.100

$M = 100$ simulated samples of size $n = 100$ for each sample are drawn from a negative binomial distribution with mean = 5 and variance = $\frac{5^2}{\alpha}$.