



ELSEVIER



## Goodness of fit test statistics for the zeta family

Andrew Luong<sup>a</sup>, Louis G. Doray<sup>b,\*</sup>

<sup>a</sup> *École d'actuariat, Université Laval, Sainte-Foy, Qué., Canada G1K7P4*

<sup>b</sup> *Département de mathématiques et de statistiques, Université de Montréal, C.P. 6128, succ. Centre-ville, Montréal, Qué., Canada H3C3J7*

Received March 1995; revised June 1996

### Abstract

Goodness of fit test procedures for the zeta parametric family based on quadratic distances and the Box–Cox transform are developed. Test statistics based on quadratic distances are shown to follow a chi-square distribution asymptotically. Test procedures based on the Box–Cox transform make use of the estimator of the parameter introduced by the Box–Cox transform, and numerical computations are based on the nonlinear weighted least squares algorithms.

*Keywords:* Zeta parametric family; Chi-square test statistics; Goodness of fit tests; Box–Cox transform; Nonlinear weighted least squares procedure; Quadratic distance estimator

### 1. Introduction

The probability mass function (p.m.f) of the zeta parametric family is given by

$$p_i = P\{x = i\} = \frac{i^{-(\rho+1)}}{\zeta(\rho+1)}, \quad i = 1, 2, \dots$$

where

$$\zeta(\rho+1) = \sum_{i=1}^{\infty} \frac{1}{i^{(\rho+1)}}, \quad \rho > 0$$

(see Olkin et al. (1980)).

In actuarial studies, the zeta distribution has been used to model the number of policies per person insured in an insurance portfolio, see Seal (1947, 1952).

The maximum likelihood estimator (MLE) is efficient. However, it is not hard to see that the MLE is nonrobust, sensitive to outliers. Its score function depends on the sample mean, consequently its influence function is unbounded, Huber (1981). The zeta family is discrete and asymmetric, robustifying the MLE is not easy.

\* Corresponding author.

An alternative procedure of estimating the parameter based on quadratic distance has been developed by Doray and Luong (1995). The quadratic distance estimator (QDE) is shown in this paper to have some desirable properties, in particular the flexibility of being able to trade efficiency for robustness, see Section 2. Among the desirable properties, we can mention:

- (1) The QDE can be made very efficient by choosing the value for  $k$  to be large,  $f_k$  denotes the observed frequency which takes on value  $k$  in the sample, making maximum use of informations in the sample.
- (2) The QDE can be made robust at the cost of being less efficient by discarding sensitive observed frequencies at the tail by choosing  $k$  at some fixed value  $k_0$ . Observed frequencies classes  $f_k$  with  $k > k_0$  are treated as outliers and rejected, the QDE is still consistent but less efficient.
- (3) The QDE can be used for testing goodness of fit for the zeta parametric family and the test statistics follow a unique chi-square distribution across the composite null hypothesis similar to the minimum chi-square methods, see Moore (1978). Consequently, the QDE methods offer a unified treatment of estimation and model testing.

In this paper, test statistics will be developed for testing goodness of fit of the zeta parametric family against various alternatives. More precisely, omnibus tests and tests for the zeta parametric family against more restricted alternatives such as against the left truncated at zero Poisson family are studied. For the sake of completeness, we review some of the results of the QDE in Section 2. We develop test statistics for the omnibus tests based on quadratic distance in Section 3. They are shown to follow an asymptotic chi-square distribution under the null hypotheses. Test statistics for more restricted alternatives are given in Section 4. The procedure involves estimating a parameter introduced by the Box–Cox transform by nonlinear weighted least squares methods and testing for a specific value of the parameter against another one. Our estimator can easily be seen to be consistent, while the Box–Cox classical maximum likelihood type estimator might not be consistent. The power of the test at a specific alternative can also be computed. Section 5 gives some discussion on the numerical implementation of the procedure introduced in Section 4.

## 2. The zeta parametric family and the QDE

In this section, the main properties of the QDE will be given. It allows the paper to be more self-contained. For more details concerning the QDE, we refer to Doray and Luong (1995).

From the p.m.f. of the zeta family, using the log transform, we have

$$\ln \frac{p_{i+1}}{p_i} = (\rho + 1) \ln \frac{i}{i+1},$$

which suggests the following representation:

$$\ln \frac{f_{i+1}}{f_i} = (\rho + 1) \ln \frac{i}{i+1} + \varepsilon_i$$

or equivalently

$$\ln \frac{\hat{p}_{i+1}}{\hat{p}_i} = (\rho + 1) \ln \frac{i}{i+1} + \varepsilon_i,$$

where  $\hat{p}_i = f_i/n$ ,  $f_i$  denotes the frequency of the observations in the sample taking on value  $i$ ,  $\varepsilon_i$  is the random error term, and  $n$  is the sample size.

Let

$$\theta = \rho + 1, \quad Y = \left[ \ln \frac{f_2}{f_1}, \ln \frac{f_3}{f_2}, \dots, \ln \frac{f_{k+1}}{f_k} \right]',$$

assuming  $f_1, f_2, \dots, f_{k+1}$  are nonzero where  $k$  is a chosen value. They are nonzero with probability one as  $n \rightarrow \infty$ . Also, let

$$X = \left[ \ln \frac{1}{2}, \ln \frac{2}{3}, \dots, \ln \frac{k}{k+1} \right]', \quad \varepsilon = [\varepsilon_1, \dots, \varepsilon_k]'$$

we have the following representation:

$$Y = X\theta + \varepsilon.$$

Asymptotically,  $E(\varepsilon) = 0$  and the variance–covariance matrix of  $\varepsilon$  is the tridiagonal matrix

$$\Sigma = \frac{1}{n} \begin{bmatrix} (p_1 + p_2)/p_1 p_2 & -1/p_2 & 0 & 0 & \dots & 0 \\ -1/p_2 & (p_2 + p_3)/p_2 p_3 & -1/p_3 & 0 & \dots & 0 \\ 0 & & & & & \\ \vdots & & & & & \\ 0 & \dots & & & -1/p_k & \\ & & & & -1/p_k & (p_k + p_{k+1})/p_k p_{k+1} \end{bmatrix}$$

or equivalently,  $\Sigma = (1/n)\Sigma^*$ ,  $\Sigma^*$  differs from  $\Sigma$  only by the constant factor term  $1/n$  as shown above.

The QDE  $\hat{\theta}$  is defined to be the value minimizing  $[Y - X\theta]' \Sigma^{-1} [Y - X\theta]$ . Explicitly,

$$\hat{\theta} = \left( X' \Sigma^{-1} X \right)^{-1} X' \Sigma^{-1} Y = (X' \Sigma^{*-1} X)^{-1} X' \Sigma^{*-1} Y,$$

$\Sigma^{*-1}$  can be replaced by a consistent estimate. The variance of  $\hat{\theta}$  is given by  $V(\hat{\theta}) = (X' \Sigma^{-1} X)^{-1}$ . The QDE has been shown to have very high efficiency for values of  $\rho \geq 1$ , see Doray and Luong (1995).

Admittedly, there is some arbitrariness in choosing a value for  $k$  for the QDE. The QDE remains consistent for all choice of values for  $k$ . For efficiency sake, we should make maximum use of the observed frequencies permissible in the sample, by letting  $k \rightarrow \infty$  as  $n \rightarrow \infty$  or fixing  $k$  at a large value. For robustness sake, we might decide to fix  $k$  at a certain value  $k_0$ , thus rejecting sensitive observed frequencies  $f_k$  at the tail with  $k > k_0$ . Small sensitive observed frequencies often appear at the tail where  $p_i \rightarrow 0$  as  $i \rightarrow \infty$ . If there is no evidence that all  $f_k$ 's are from the same population and some  $f_k$ 's with large  $k$  are quite high, one has to go to check the validity of the zeta distribution assumption.

These facts should be taken into account in constructing chi-square goodness of fit test statistic in Section 3.

### 3. Chi-square omnibus tests for the zeta parametric family

From the relationship  $\ln(f_{i+1}/f_i) = \theta \ln(i/i + 1) + \varepsilon_i$ , a visual plot of  $\ln(f_{i+1}/f_i)$  vs.  $\ln(i/i + 1)$  might give us some idea of goodness of fit of the zeta parametric family. A formal test can be performed based on test statistics constructed using the expression  $[Y - X\theta]' \Sigma^{-1} [Y - X\theta]$ , which represents a measure of discrepancy between the empirical c.d.f.  $F_n$  and  $F_\theta$  is the c.d.f. of the zeta family as given below.

Let

$$Y = \left[ \ln \frac{f_2}{f_1}, \ln \frac{f_3}{f_2}, \dots, \ln \frac{f_{k+1}}{f_k} \right], \quad I_k = (k - 1, k),$$

we can write

$$Y = \left[ \ln \frac{\int_{I_2} dF_n}{\int_{I_1} dF_n}, \dots, \ln \frac{\int_{I_{k+1}} dF_n}{\int_{I_k} dF_n} \right]'$$

and asymptotically

$$E(Y) = X\theta = \left[ \ln \frac{\int_{I_2} dF_\theta}{\int_{I_1} dF_\theta}, \dots, \ln \frac{\int_{I_{k+1}} dF_\theta}{\int_{I_k} dF_\theta} \right]',$$

we can see that a measure of distance between  $F_n$  and  $F_\theta$ ,

$$d(F_n, F_\theta) = [Y - X\theta]' \Sigma^{-1}(\theta)[Y - X\theta] = n[Y - X\theta]' \Sigma^{*-1}(\theta)[Y - X\theta],$$

is defined.

Naturally, we can make use of this expression to construct goodness of fit test statistics for testing the simple null hypothesis  $H_0 : \theta = \theta_0$  against all alternatives and for testing the composite null hypothesis  $H_0 : \theta \in \Theta$ ,  $\Theta$  is the full parameter space against all alternatives (omnibus tests). Our test statistic makes use of  $F_n$  just as others goodness of fit test statistics such as the Cramér–Von Mises statistics, see Boos (1981).

We are aiming to construct test statistics which follow a chi-square distribution under the null hypotheses. The following theorem also used in Luong and Thompson (1987) is needed; its proof can be found in Moore (1977, 1978).

**Theorem 1.** *Suppose that the random vector  $Y$  of dimension  $p$  is  $N_p(0, \Sigma)$  and  $C$  is any  $p \times p$  symmetric positive semidefinite matrix, then the quadratic form  $Y'CY$  is chi-square distributed with  $k$  degrees of freedom if  $\Sigma C$  is idempotent and trace  $(\Sigma C) = k$ . (The same result holds asymptotically if  $C$  is replaced by a consistent estimate  $\hat{C}$ .)*

For conciseness, we adopt the notation  $\Sigma = \Sigma(\theta_0)$ ,  $\Sigma^* = \Sigma^*(\theta_0)$ ,  $\Sigma_1 = \Sigma_1(\theta_0)$ ,  $\Sigma_1$  is given in (3.1). To test the simple hypothesis  $H_0 : \theta = \theta_0$ , the test statistic

$$\chi^2 = d(F_n, F_{\theta_0}) = n[Y - X\theta_0]' \Sigma^{*-1}[Y - X\theta_0]$$

can be used. Since  $\sqrt{n}[Y - X\theta_0] \xrightarrow{L} N(0, \Sigma^*)$  under conditions given by Moore (1978), Billingsley (1968) and  $\Sigma^* \Sigma^{*-1} = I_k$  is clearly idempotent, using the theorem above,  $\chi^2 \xrightarrow{L} \chi^2(k)$ .

To test the composite hypothesis  $H_0 : \theta \in \Theta$ ,  $\Theta$  is the full parameter space, the following test statistic

$$\begin{aligned} \chi^2 &= d(F_n, F_{\hat{\theta}}) = n[Y - X\hat{\theta}]' \Sigma^{*-1}(\hat{\theta})[Y - X\hat{\theta}] \\ &= n[Y - X(X' \Sigma^{*-1} X)^{-1} X' \Sigma^{*-1} Y]' \Sigma^{*-1}(\hat{\theta})[Y - X(X' \Sigma^{*-1} X)^{-1} X' \Sigma^{*-1} Y], \end{aligned}$$

where  $\hat{\theta} = (X' \Sigma^{*-1} X)^{-1} X' \Sigma^{*-1} Y$ , is the QDE,  $\hat{\theta}$  being consistent thus it belongs to the parameter space with probability 1.  $\Sigma^{*-1}$  can be replaced by a consistent estimate  $\hat{\Sigma}^{*-1}$ . Since  $\sqrt{n}[Y - X\hat{\theta}] \xrightarrow{L} N(0, \Sigma_1)$  as given by Moore (1978) where  $\Sigma_1$  the asymptotic variance–covariance matrix given by

$$\begin{aligned} \Sigma_1 &= \text{Var} \left\{ \sqrt{n} \left[ Y - X(X' \Sigma^{*-1} X)^{-1} X' \Sigma^{*-1} Y \right] \right\} = \text{Var} \left\{ \sqrt{n} \left[ 1 - X(X' \Sigma^{*-1} X)^{-1} X' \Sigma^{*-1} \right] Y \right\} \\ &= \left[ 1 - X(X' \Sigma^{*-1} X)^{-1} X' \Sigma^{*-1} \right] \Sigma^* \left[ 1 - \Sigma^{*-1} X(X' \Sigma^{*-1} X)^{-1} \right], \end{aligned} \quad (3.1)$$

we then have

$$\Sigma_1 \Sigma^{*-1} = \left[ I - X(X' \Sigma^{*-1} X)^{-1} X' \Sigma^{*-1} \right] \left[ I - X(X' \Sigma^{*-1} X)^{-1} X' \Sigma^{*-1} \right].$$

It is easy to verify that  $\Sigma_1 \Sigma^{*-1}$  is idempotent and

$$\begin{aligned} \text{rank} (\Sigma_1 \Sigma^{*-1}) &= \text{trace} \left[ 1 - X(X' \Sigma^{*-1} X)^{-1} X' \Sigma^{*-1} \right] \\ &= k - \text{trace} X(X' \Sigma^{*-1} X)^{-1} X' \Sigma^{*-1} = k - 1. \end{aligned}$$

Also,  $\Sigma^{*-1}(\hat{\theta}) \xrightarrow{P} \Sigma^{*-1} = \Sigma^{*-1}(\theta_0)$ ; consequently  $\chi^2 \xrightarrow{L} \chi^2(k - 1)$ .

These chi-square test statistics can be calculated easily and they do not depend explicitly on the expansion of the zeta function. Their null asymptotic distribution is chi-square across the composite  $H_0$  which makes it easy to use. This characteristic is not shared by test statistics such as the Cramér-von Mises type statistics, see Boos (1981). They are expected to have good power as well since they are based on the quadratic distance which generates an estimator which has high efficiency. The question of power of the tests under alternatives often requires extensive numerical and simulation studies and is not addressed in this paper.

**4. Goodness of fit tests based on the Box–Cox transform and nonlinear least squares estimators**

The Box–Cox transform is often used for normalizing the error term and for testing goodness of fit of the model in the continuous linear regression model context, see Box and Cox (1964), Amemiya (1985), Carroll and Ruppert (1988). The idea is to introduce an extra parameter  $\alpha$  in the model, which allows testing the distributional assumption, once an estimate for  $\alpha$  is obtained and its asymptotic distribution is derived.

In the same spirit, but in a discrete model context, we shall make use of the Box–Cox transform. Note that, for the zeta parametric family, we have the following relationship:

$$\ln \left( \frac{p_{i+1}}{p_i} \right) = \theta \ln \frac{i}{i + 1}, \quad i = 1, 2, \dots,$$

and for the left truncated at zero Poisson family

$$\frac{P_{i+1}}{P_i} = \frac{\lambda}{i + 1} = \theta \left( \frac{i}{i + 1} - 1 \right), \quad \theta = -\lambda, \quad i = 1, 2, \dots$$

The Box–Cox transform is defined as  $g(u) = (u^\alpha - 1)/\alpha$ ; we can see that for  $\alpha = 0$ , we get  $g(u) = \ln u$  and for  $\alpha = 1$ , we get  $g(u) = u - 1$ .

Consequently, we have

$$g \left( \frac{p_{i+1}}{p_i} \right) = -\alpha + \theta g \left( \frac{i}{i + 1} \right). \tag{4.1}$$

Note that for  $\alpha = 0$  or equivalently  $g(u) = \ln u$ , the relationship is satisfied for the zeta parametric family and for  $\alpha = 1$  or equivalently  $g(u) = u - 1$ , the relationship is satisfied for the left truncated at zero Poisson family where

$$p_i = \left( \sum_{i=1}^{\infty} \frac{\lambda^i e^{-\lambda}}{i!} \right)^{-1} \frac{\lambda^i e^{-\lambda}}{i!} \quad \text{for } i \geq 1, \quad p_0 = 0.$$

Relationship (4.1) implicitly defines a larger family where the zeta family and the left truncated at zero Poisson family are subfamilies within this larger family.

Using the above relationship, we obtain

$$\left( \frac{p_{i+1}}{p_i} \right)^\alpha = 1 - \alpha^2 + \alpha \theta g \left( \frac{i}{i + 1} \right)$$

or equivalently

$$\ln \left( \frac{p_{i+1}}{p_i} \right) = \frac{1}{\alpha} \ln \left[ 1 - \alpha^2 + \alpha \theta g \left( \frac{i}{i+1} \right) \right],$$

which suggests the following nonlinear regression model representation:

$$\ln \frac{\hat{p}_{i+1}}{\hat{p}_i} = \ln \frac{f_{i+1}}{f_i} = \frac{1}{\alpha} \ln \left[ 1 - \alpha^2 + \alpha \theta g \left( \frac{i}{i+1} \right) \right] + \varepsilon_i = \frac{1}{\alpha} \ln \left\{ 1 - \alpha^2 + \theta \left[ \left( \frac{i}{i+1} \right)^\alpha - 1 \right] \right\} + \varepsilon_i.$$

Asymptotically, the random vector  $\varepsilon = [\varepsilon_1, \dots, \varepsilon_k]'$  has  $\Sigma$  as variance–covariance matrix which depends on  $\beta = (\alpha, \theta)'$  and  $E(\varepsilon) = 0$  asymptotically provided that the relationship (4.1) holds. Letting

$$Y_i = \ln \frac{f_{i+1}}{f_i}, \quad \phi_i(\alpha, \theta) = \frac{1}{\alpha} \ln \left[ 1 - \alpha^2 + \alpha \theta g \left( \frac{i}{i+1} \right) \right],$$

we can write  $Y_i = \phi_i(\alpha, \theta) + \varepsilon_i$ ,  $i = 1, 2, \dots, k$ , which takes a nonlinear regression model form. Using matrix notation, let  $Y = [Y_1, \dots, Y_k]'$ ,  $\beta = (\alpha, \theta)'$ ,  $\varepsilon = (\varepsilon_1, \dots, \varepsilon_k)'$ ,  $\phi(\beta) = [\phi_1(\beta), \phi_2(\beta), \dots, \phi_k(\beta)]'$ , we then have

$$Y = \phi(\beta) + \varepsilon.$$

Letting  $\Sigma(\beta)$  to denote the variance–covariance matrix of  $[y - \phi(\beta)]$  under  $\beta$   $\Sigma = \Sigma(\beta_0)$ ,  $\beta_0$  is the true value of the parameters. This suggests estimating  $\beta$  by the nonlinear least squares estimator  $\hat{\beta}$ , which minimizes

$$[Y - \phi(\beta)]' \Sigma^{-1} [Y - \phi(\beta)], \quad (4.2)$$

and the nonlinear least squares estimator  $\hat{\beta}$  is consistent. This follows from the asymptotic theory of nonlinear least squares estimator as given by Jennrich (1969) and Amemiya (1985), for example. Our procedure yields consistent estimators essentially due to the fact that our model is based on proportions and consequently the skewness introduced in the model is mitigated asymptotically, while the classical Box–Cox maximum likelihood type estimator which is based directly on raw data does not have this feature, see Amemiya (1985).

Under mild conditions of differentiability on  $\phi$ , we shall show that  $\hat{\beta}$  is asymptotically multivariate normal,

$$\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{L} N(0, C),$$

where

$$C = \lim_{n \rightarrow \infty} \left[ \frac{1}{n} \left( \frac{\partial \phi'}{\partial \beta} \Sigma^{-1} \frac{\partial \phi}{\partial \beta} \right) \right]^{-1},$$

which shows that the asymptotic variance–covariance matrix of  $\hat{\beta}$  is given by

$$V(\hat{\beta}) = \left( \frac{\partial \phi'}{\partial \beta} \Sigma^{-1} \frac{\partial \phi}{\partial \beta} \right)^{-1}, \quad \Sigma^{-1} = \Sigma^{-1}(\beta_0), \quad \phi = \phi(\beta_0),$$

$\beta_0$  is the true value of the parameters,

$$\frac{\partial \phi}{\partial \beta} = \begin{bmatrix} \partial \phi_1 / \partial \alpha & \partial \phi_1 / \partial \theta \\ \partial \phi_2 / \partial \alpha & \partial \phi_2 / \partial \theta \\ \vdots & \vdots \\ \partial \phi_k / \partial \alpha & \partial \phi_k / \partial \theta \end{bmatrix},$$

evaluated at  $\beta = \beta_0$ ,

$$\frac{\partial \phi_i}{\partial \alpha} = \frac{1}{\alpha} \left\{ \frac{\theta (i/(i+1))^\alpha \ln[i/(i+1)] - 2\alpha}{1 - \alpha^2 + \theta [(i/(i+1))^\alpha - 1]} - \frac{1}{\alpha} \ln \left\{ 1 - \alpha^2 + \theta \left[ \left( \frac{i}{i+1} \right)^\alpha - 1 \right] \right\} \right\}$$

and

$$\frac{\partial \phi_i}{\partial \theta} = \frac{g(i/(i+1))}{1 - \alpha^2 + \alpha \theta g(i/(i+1))} = \frac{(1/\alpha) [(i/(i+1))^\alpha - 1]}{1 - \alpha^2 + \theta [(i/(i+1))^\alpha - 1]}$$

evaluated at  $\beta = \beta_0$ .

The above results can be obtained using nonlinear least squares theory as given by Jennrich (1969), Amemiya (1985), Seber and Wild (1989), and Bates and Watt (1988). A simple version of the proof is given below for the sake of completeness.

Note that if  $\bar{\beta}$  is consistent,  $\Sigma^{-1}(\bar{\beta}) \xrightarrow{P} \Sigma^{-1}$ . Also, an estimator  $\tilde{\beta}$  asymptotically equivalent to  $\hat{\beta}$  is obtained by minimizing

$$[Y - \phi(\beta)]' \hat{\Sigma}^{-1} [Y - \phi(\beta)], \tag{4.3}$$

where  $\hat{\Sigma}$  is a consistent estimate of  $\Sigma$ . For numerical procedures concerning how to obtain  $\tilde{\beta}$ , see Section 5.

We shall prove asymptotic normality using expression (4.2). It is not difficult to see from the proof that expressions (4.2) and (4.3) yield asymptotically equivalent estimators. Since  $\Sigma$  is not known, expression (4.2) is not used to generate estimators in practice, however the proof based on this expression is less cumbersome. In practice (4.3) is used to generate estimators numerically as indicated in Section 5.

The estimator  $\hat{\beta}$  minimizing  $[Y - \phi(\beta)]' \Sigma^{-1} [Y - \phi(\beta)]$  satisfies the following system of equations under differentiability assumptions:

$$\frac{\partial \phi'}{\partial \beta}(\hat{\beta}) \Sigma^{-1} [Y - \phi(\hat{\beta})] = 0. \tag{4.4}$$

Using a Taylor's expansion around  $\beta_0$ , (4.4) can be re-expressed as

$$\begin{aligned} & \frac{1}{\sqrt{n}} \frac{\partial \phi'}{\partial \beta}(\beta_0) \Sigma^{-1} [Y - \phi(\beta_0)] \\ & + \frac{1}{n} \left\{ \frac{\partial^2 \phi}{\partial \beta \partial \beta'}(\beta_0) \Sigma^{-1} [Y - \phi(\beta_0)] - \frac{\partial \phi'}{\partial \beta}(\beta_0) \Sigma^{-1} \frac{\partial \phi}{\partial \beta}(\beta_0) \right\} \sqrt{n}(\hat{\beta} - \beta_0) + O_p(1) = 0, \end{aligned} \tag{4.5}$$

where  $\partial^2 \phi / \partial \beta \partial \beta' = (\partial^2 \phi / \partial \beta \partial \beta')(\beta_0)$  is the second derivative Hessian matrix, evaluated at  $\beta = \beta_0$  and  $O_p(1)$  is an expression converging to 0 in probability.

Under regularity conditions,

$$\frac{1}{n} \left[ \frac{\partial^2 \phi}{\partial \beta \partial \beta'} \Sigma^{-1} (Y - \phi) \right]$$

is  $O_p(1)$  and

$$\lim_{n \rightarrow \infty} \frac{1}{n} \left( \frac{\partial \phi'}{\partial \beta} \Sigma^{-1} \frac{\partial \phi}{\partial \beta} \right) = C^{-1},$$

replacing them in (4.5) yields

$$\frac{1}{\sqrt{n}} \frac{\partial \phi'}{\partial \beta} \Sigma^{-1} (Y - \phi) = [C^{-1} + O_p(1)] \sqrt{n}(\hat{\beta} - \beta_0) + O_p(1). \tag{4.6}$$

Since the LHS converges in law to a multivariate normal distribution with  $C^{-1}$  as variance–covariance matrix, using the RHS, we can conclude that

$$\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{L} N(0, C).$$

Therefore, the asymptotic variance of  $\hat{\beta}$  which is also the asymptotic variance of  $\tilde{\beta}$  is given by

$$V(\hat{\beta}) = V(\tilde{\beta}) = \left( \frac{\partial \phi'}{\partial \beta} \Sigma^{-1} \frac{\partial \phi}{\partial \beta} \right)^{-1}.$$

This result allows us to test goodness of fit of the zeta parametric family and compute the power of the test when the alternative is the left truncated at zero Poisson family and vice versa by testing  $H_0 : \alpha = 0$  vs.  $H_a : \alpha = 1$ , for example. Letting  $\hat{v}_{ij}$  to denote a consistent estimate of the  $i - j$ th element of  $V(\hat{\beta})$ , the test statistic  $\hat{\alpha}/\sqrt{\hat{v}_{11}}$  can then be used and it follows a standard normal distribution asymptotically. By looking at the values of the objective function produced by the estimation iteration process, one can already have an idea what would be the conclusion of the test.

## 5. Discussion

For computation of the nonlinear least squares estimator, one can use the algorithm of Gauss–Newton which utilizes a series of iterated reweighted least squares procedures, see Seber and Wild (1984), for example. Statistical packages such as SAS, SYSTAT or S-plus provide subroutines for finding nonlinear least squares estimators.

A nonlinear least squares unweighted estimator  $\tilde{\beta}$  which is consistent but less efficient than  $\hat{\beta}$  can be obtained by minimizing

$$[Y - \phi(\beta)]'[Y - \phi(\beta)].$$

Using a Taylor's expansion as in the previous proof, we can see that the asymptotic variance of  $\tilde{\beta}$  is

$$V(\tilde{\beta}) = \left( \frac{\partial \beta'}{\partial \beta} \cdot \frac{\partial \phi}{\partial \beta} \right)^{-1} \left( \frac{\partial \phi'}{\partial \beta} \Sigma^{-1} \frac{\partial \phi}{\partial \beta} \right) \left( \frac{\partial \phi'}{\partial \beta} \cdot \frac{\partial \phi}{\partial \beta} \right)^{-1}.$$

Since  $\tilde{\beta}$  is consistent, we can use it as a starting point. We can estimate  $\Sigma^{-1}$  by  $\Sigma^{-1}(\tilde{\beta})$  and obtain the nonlinear weighted least squares estimator  $\tilde{\beta}$  by minimizing

$$[Y - \phi(\beta)]' \Sigma^{-1}(\tilde{\beta}) [Y - \phi(\beta)]$$

and the procedure can be repeated by re-estimating  $\Sigma^{-1}$  at each step until some criterion of convergence is met.

## References

- [1] Amemiya, T.(1985). *Advanced Econometrics*. Harvard University Press, Cambridge, MA.
- [2] Bates, D.M. and D.G. Watts (1988). *Nonlinear Regression Analysis and its Application*. Wiley, New York.
- [3] Billingsley, P. (1968). *Convergence of Probability Measure*. Wiley, New York.
- [4] Boos, D.D. (1981). Minimum distance estimators for location and goodness of fit. *Journal of the American Statistical Association*. 76, 663–670.
- [5] Carroll, R.J. and D. Ruppert (1988). *Transformation and weighting in Regression*. Chapman and Hall, London.
- [6] Doray, L. and A. Luong (1995). Quadratic distance estimators for the zeta family. *Insurance: Mathematics and Economics* 16, 225–260.
- [7] Huber, P.(1981). *Robust Statistics*. Wiley, New York.
- [8] Jennrich, R. (1969). Asymptotic properties of nonlinear least squares estimator. *Annals of Mathematical Statistics* 40, 633–648.



- [9] Luong, A. and M.E. Thompson, (1987). Minimum distance methods based on quadratic distance for transforms. *Canadian Journal of Statistics*. 15, 239–251.
- [10] Moore, D.S. (1977). Generalized inverses, Wald's method and the construction of chi-square tests of fit. *Journal of the American Statistical Association* 72, 131–137.
- [11] Moore, D.S. (1978). *Chi-Square Tests*. Studies in statistics, Vol. 19. Mathematical Association of America.
- [12] Olkin, I., Gleson, L.J. and C. Derman (1980). *Probability Models and Applications*. MacMillan, New York.
- [13] Seal, H.L. (1947). A probability distribution of deaths at age  $x$  when policies are counted instead of lives. *Skand. Aktuar. Tidskr.* 30, 18–43.
- [14] Seal, H.L. (1952). The maximum likelihood fitting of the discrete Pareto law. *Journal of the Institute of Actuaries*. 78, 115–121.
- [15] Seber, G.A.F. and C.J. Wild, (1989). *Nonlinear Regression*. Wiley, New York.