Scientific
Research
Publishing

Unified Asymptotic Results for Maximum Spacing and Generalized Spacing Methods for Continuous Models

Andrew Luong

école d'Actuariat, Université Laval, Québec, Canada
Email: Andrew.Luong@act.ulaval.ca

Abstract

Asymptotic results are obtained using an approach based on limit theorem results obtained for α-mixing sequences for the class of general spacings (GSP) methods which include the maximum spacings (MSP) method. The MSP method has been shown to be very useful for estimating parameters for univariate continuous models with a shift at the origin which are often encountered in loss models of actuarial science and extreme models. The MSP estimators have also been shown to be as efficient as maximum likelihood estimators in general and can be used as an alternative method when ML method might have numerical difficulties for some parametric models. Asymptotic properties are presented in a unified way. Robustness results for estimation and parameter testing results which facilitate the applications of the GSP methods are also included and related to quasi-likelihood results.

Keywords

Maximum Product of Spacings, M-Estimators, Quasi-Likelihood Ratio Test Statistic, α-Mixing Sequences

1. Introduction

Let $X_1, \cdots, X_n$ be a random sample from a continuous parametric family with a distribution function which belongs $\{F_{\boldsymbol{\theta}}\}$, $\boldsymbol{\theta}$ is the vector of parameters and instead of fitting distribution using maximum likelihood (ML) method, Cheng and Amin [1], Ranneby [2] proposed the maximum product of spacings method which is also called maximum spacing (MSP) method which makes use of spacings which are gaps between order statistics of the sample instead of using directly the observations of the sample. The method consists of maximizing or equivalently

minimizing the following objective function to obtain the MSP estimators,

$$\sum_{i=1}^{n+1} -\log\left((n+1)\left(D_i(\boldsymbol{\theta})\right)\right)$$

where $D_i(\boldsymbol{\theta}) = F_{\boldsymbol{\theta}}\left(x_{(i)}\right) - F_{\boldsymbol{\theta}}\left(x_{(i-1)}\right), i = 1, \cdots, n+1$ are the spacings and we define $F_{\boldsymbol{\theta}}\left(x_{(n+1)}\right) = 1, F_{\boldsymbol{\theta}}\left(x_{(0)}\right) = 0$ with the order statistics of the sample given by

$$X_{(1)} < X_{(2)} < \cdots < X_{(n)}.$$

It is quite obvious that it is not more difficult to obtain the GSP estimators than the ML estimators and it has been proven that the MSP estimators are as efficient as the ML estimators in general and can be consistent when ML estimators might fail to be consistent. MSP method can be used as an alternative to ML method as ML method might encounter numerical difficulties when used for fitting some models with shifted origin which are often encountered in loss models and extreme value models. We shall examine a few examples for illustrations. Anatolyev and Kosenok [3] have discussed the model of example 1 below where they find the MSP estimators have better finite sample properties than ML estimators and notice that the method has not received much attention in econometrics; it has received even less attention in actuarial science.

Example 1 (Pareto)

The Pareto model considered by Anatolyev and Kosenok [3] has density function given by

$$f(x;\alpha) = \frac{\alpha}{x^{\alpha+1}}$$

and distribution function given by

$$F(x;\alpha) = 1 - \left(\frac{1}{x}\right)^{\alpha}, x > 1, \alpha > -1.$$

The model is a sub-model of the larger model with two parameters α and θ, density function

$$f(x;\alpha) = \frac{\alpha\theta^{\alpha}}{x^{\alpha+1}}, x > \theta, \theta \geq 0,$$

where θ is a shift parameter, $\alpha > -1$. The distribution function is given by

$$F(x;\alpha,\theta) = 1 - \left(\frac{\theta}{x}\right)^{\alpha}, x > \theta, \alpha > -1,$$

see Klugman *et al.* [4] (p. 465) for the larger model which is a shifted origin model and used in actuarial science.

The following example gives the Fréchet model which is an extreme value model and it is also a shifted origin model, see more properties and details in the book by Castillo *et al.* [5] (p. 63-64) for the Fréchet model.

Example 2(Fréchet)

The Fréchet model has three parameters β, δ, λ with density function and distribution function given by

$$f(x;\beta,\delta,\lambda) = \frac{\beta\delta}{(x-\lambda)^2}\left(\frac{\delta}{x-\lambda}\right)^{\beta-1}\exp\left[-\left(\frac{\delta}{x-\lambda}\right)^{\beta}\right], x > \lambda \geq 0, \beta > 1, \delta > 0$$

and

$$F(x;\beta,\delta,\lambda) = \exp\left[-\left(\frac{\delta}{x-\lambda}\right)^{\beta}\right], x > \lambda \geq 0,$$

where λ is as shift parameter, $\beta > 1, \delta > 0$.

Ghosh and Jammalamadaka [6] have generalized the MSP method by considering objective functions of the form $\sum_{i=1}^{n+1} h\left((n+1)\left(D_i(\theta)\right)\right)$ with $h(x)$ being a convex function and twice differentiable, we shall call methods based on this class generalized spacings methods (GSP) and give more details in the next section where we restrict $h(x)$ being the commonly used convex functions and introduce the GJ class named after Ghosh and Jammalamadaka [6] and by GSP methods we refer to GSP methods but with $h(x)$ belongs to this GJ class.

As we have seen that despite the GSP methods are powerful methods for univariate continuous models but they are not used as often as they should be. It might be due to the asymptotic results are scattered in the literature and in particular previous approaches for asymptotic normality have been based on distribution of spacings and order statistics which make further results such as the distributions of counterparts of Wald test statistic, Score statistic and likelihood ratio test statistics of likelihood theory difficult to establish which prevent the use of these methods for applications. In this paper, a different approach is taken for establishing asymptotic normality. The approach is a based on using uniform weak law of large numbers (UWLLN) for establishing consistency of the GSP estimators and central limit theorem for α-mixing sequences as given by White and Domowitz [7] to establish asymptotic normality for the GSP estimators, asymptotic distributions for the trinity test for hypothesis testing for parameters for GSP methods are also obtained by relating the GSP methods to quasi-likelihood methods so that robustness of the GSP methods can also be studied. With a unified and simpler presentation, we hope to put the GSP methods parallel to likelihood methods and by doing so we hope to encourage more use of these methods by practitioners for their applied works in various fields.

The paper is organized as follows. Section 2 gives the preliminary results already established by Ranneby [2] but needed for further results. The notions of mixing sequences are also introduced to facilitate the developments in subsequent sections. Section 3 gives the asymptotic properties of the GSP methods using UWWLN and CLT; robustness properties are established. Parameter hypothesis testing is treated with results obtained for the trinity test for the GSP methods which are related to quasi-likelihood methods and parallel to ML methods. Results which are available for M-estimation theory are also used to investigate the asymptotic distributions for the trinity test for the GSP methods.

2. Preliminaries

For further study the class of generalized spacing (GSP) methods, we shall present the GJ class being considered by Ghosh and Jammalamadaka [6] and define some notations. We shall make use of some results already established by Ranneby [2] so that subsequently we shall use a different approach for studying asymptotic properties of the GSP methods; the approach is based on uniform law of large numbers (ULLN) and central limit theorem as given by White and Domowitz [7] for α-mixing sequences. The approach seems to give results in a more unified way than approaches using results on spacings and order statistics; see Pyke [8], Shorack and Wellner [9] for results on spacings and David and Nagaraja [10] for results on order statistics.

The GSP methods can be seen to be closely related to quasi-likelihood methods and M-estimation theory can be used to study estimation, robustness and parameter testing via Wald test, Lagrange multiplier test or score test and quasi-likelihood ratio test but with a GSP version for each of these tests forming the classical trinity. The results appear to be natural and parallel to maximum likelihood methods (ML) given that the vector of MSP estimators which belong to the class of GSP is as efficient as the vector of ML estimator and therefore it is natural to establish inference methods based on this class which parallel ML methods. It is also worth to note that this class can be used for robust estimation which parallel Hellinger distance methods given by Beran [11] and like the class of pseudo-distance studied by Broniatowski [12] there is no need of density estimate to implement GSP methods. We try to reduce technicalities for the methods introduced subsequently so that practitioners might find that it is not so difficult to follow and make use of the results and it is quite clear that it is relatively simple to implement GSP methods just as in the case for ML methods.

Now we shall use the set up as given by Ghosh and Jammalamadaka [6] and Ranneby [2] by assuming a random sample of size n where we have observations $X_1, \cdots, X_n$ which are independent and identically distributed (iid) as X with density function and distribution function given respectively by $f_{\boldsymbol{\theta}}(x)$ and $F_{\boldsymbol{\theta}}(x)$. The vector of parameters is denoted by $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_m)'$. The vector of the true parameters is $\boldsymbol{\theta}_0$.

The order statistics are denoted by $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$ and the spacings $D_i(\boldsymbol{\theta})$ which can be viewed as transforms of the order statistics and they are given by

$$D_i(\boldsymbol{\theta}) = F_{\boldsymbol{\theta}}(X_{(i)}) - F_{\boldsymbol{\theta}}(X_{(i-1)}), i = 1, \cdots, n+1$$

with $F_{\boldsymbol{\theta}}(X_{(0)}) = 0$ and $F_{\boldsymbol{\theta}}(X_{(n+1)}) = 1$ by definitions.

Ghosh and Jammalamadaka [6] studied the class of estimators namely GSP estimators obtained by minimizing a criterion function of the following form which make use of a class of convex function $h(x)$, i.e.,

$$T(\boldsymbol{\theta}) = \sum_{i=1}^{n+1} h((n+1)D_i(\boldsymbol{\theta})) \tag{1}$$

or equivalently,

$$Q_n(\boldsymbol{\theta}) = \frac{1}{n+1}\sum_{i=1}^{n+1} h\big((n+1)D_i(\boldsymbol{\theta})\big). \tag{2}$$

The class considered include the following functional form for $h(x)$ which is a convex function with domain $(0,\infty)$ an range being the real line. With $h(x) = -\log(x)$ we have the MSP method and the function $h(x) = -\log(x)$ is optimum in term of efficiency of statistical methods generated but for robustness other choices for $h(x)$ might include $h(x) = x^\alpha$ for $\alpha > 1$, $h(x) = -x^\alpha$ for $0 < \alpha < 1$,

$$h(x) = x\log(x), \quad h(x) = x^\alpha \quad \text{for} \quad -\frac{1}{2} < \alpha < 0. \tag{3}$$

Note that for all these choices the first and second derivatives $h'(x)$, $h''(x)$ exist and since $h(x)$ is a convex function $h'(x) \geq 0$, see Lehmann and Casella [13] (p. 45-47) for properties of convex function and Jensen's inequality which is based on the property of convexity of a function.

We shall call the class defined by using functions given by expression (3) and including $h(x) = -\log(x)$, the GJ class as it was introduced by Ghosh and Jammalamadaka [6]. The GJ class includes the commonly used $h(x)$, see expression (6) given by Ghosh and Jammalamadaka [6] (p73).

For this class, we can see that the sub-class with $h(x) = -x^\alpha$ for $0 < \alpha < 1$ and up to an additive and multiplicative constant, it can be expressed equivalently as

$$h(x) = -\frac{x^\alpha - 1}{\alpha} \quad \text{and as} \quad \alpha \to 0_+$$

$$h(x) = -\frac{x^\alpha - 1}{\alpha} \to h^*(x) = -\log(x)$$

which is the optimum $h(x)$ as using $h^*(x)$ will generate the MSP estimators which are the most efficient within this class and asymptotically equivalent to maximum likelihood(ML) estimators.

Ghosh and Jammalamadaka [6] establish consistency and asymptotic normality for this class of GSP estimators and make use of limit theorems for spacings which are based on order statistics. We shall use another approach which bypass limit theorems for spacings and order statistics and obtain consistency and asymptotic normality for the multi-parameter case for this class of GSP methods and based on these results, robustness properties for this class can be studied and parameter testing can be developed in Section 4 and we shall be able to consider GSP methods as quasi-likelihood methods and hence unify the asymptotic theory with M-estimation theory.

We shall also make use of the following results and notions introduced by Ranneby [2] to overcome the need of using spacings or order statistics in expression (1) and expression (2). With a sample of size n; for each observation X_i, it can be associated to a random variable $Y_i = (n+1)d_i$ where d_i is random

variable which represents the distance from X_i to the nearest observation to the right of X_i and this distance is defined to be infinity if $X_i = X_{(n)}$.

It is clear that if we use (X_i, Y_i) to re-express $Q_n(\boldsymbol{\theta})$ or an asymptotic equivalent expression for $Q_n(\boldsymbol{\theta})$ can be given with the use of $(X_i, Y_i), i = 1, \cdots, n$ instead of $D_i(\boldsymbol{\theta}), i = 1, \cdots, n$ in expression (2).

Now if we define

$$z_i(\boldsymbol{\theta}, n) = (n+1)\left(F_{\boldsymbol{\theta}}\left(x_i + \frac{y_i}{n+1}\right) - F_{\boldsymbol{\theta}}(x_i)\right) \tag{4}$$

and

$$Q_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n} h(z_i(\boldsymbol{\theta}, n)) \tag{5}$$

then $Q_n(\boldsymbol{\theta})$ defined as above is asymptotically equivalent to $Q_n(\boldsymbol{\theta})$ defined using expression (2) as only the first term in the summation of expression (2) is left out for the summation of expression (5) and we focus on asymptotic theory here. Also, expression (5) is asymptotically equivalent to the expression denoted by $S_n(\boldsymbol{\theta})$ given by Ranneby [2] (p. 98).

We shall see that most of limit theorems such as the uniform weak law of large numbers(UWLLN) or Central limit theorem (CLT) are stated using the form given by expression (5) and we need to subject $Q_n(\boldsymbol{\theta})$ to these limit theorems; it is more convenient to use $Q_n(\boldsymbol{\theta})$ as given by expression (5) with a factor n as it simplifies the notations.

Now we can note that if $\{z_i(\boldsymbol{\theta}, n)\}$ is a sequence of independent identically distributed (iid) terms, then there is no problem to apply UWLN and CLT but we will see that $\{z_i(\boldsymbol{\theta}, n)\}$ is a dependent sequence but with a weak form of dependency so that we can apply a dependent version of UWLN then we can draw the same conclusion just as assuming $\{z_i(\boldsymbol{\theta}, n)\}$ are iid.

Clearly, we need to make use of the distribution of $V_i = (X_i, Y_i)'$ and the dependence of the sequence $\{V_i\}$ to study the sequence $\{z_i(\boldsymbol{\theta}, n)\}$. Ranneby [2] (p. 99-100) has shown that V_i for $i = 1, 2, \cdots, n$ has a common bivariate distribution which depends on n and given by

$$P_n(x, y) = \int_{-\infty}^{x}\left[1 - \left[1 - \left(F_{\boldsymbol{\theta}_0}\left(u + \frac{y}{n+1}\right)\right)^{n-1}\right]\right]f_{\boldsymbol{\theta}_0}(u)\,\mathrm{d}u. \tag{6}$$

As $n \to \infty$,

$$P_n(x, y) \to P_{\boldsymbol{\theta}_0}(x, y) = \int_{-\infty}^{x}\left(1 - \mathrm{e}^{-yf_{\boldsymbol{\theta}_0}(u)}\right)f_{\boldsymbol{\theta}_0}(u)\,\mathrm{d}u$$

with the bivariate density function given by

$$p_{\boldsymbol{\theta}_0}(x, y) = f_{\boldsymbol{\theta}_0}^2(x)\mathrm{e}^{-yf_{\boldsymbol{\theta}_0}(x)}, y > 0, \tag{7}$$

see Ranneby [2] (p. 98) for the expression for the bivariate density function $p_{\boldsymbol{\theta}_0}(x, y)$. The density $p_{\boldsymbol{\theta}_0}(x, y)$, with respect to x, has $f_{\boldsymbol{\theta}_0}(x)$ as marginal density function.

Let $V^0 = \left(X^0, Y^0 \right)$ and its bivariate density is as given by expression (7) and to derive asymptotic results subsequently we let $n \to \infty$, then we have the equalities in distribution for $i = 1, 2, \cdots$.

$$V_i = \left(X_i, Y_i \right)' =^d V^0 = \left(X^0, Y^0 \right)'. \tag{8}$$

Furthermore, we have pairwise asymptotically independent of $\{V_i, V_{i'}\}$ for $i \neq i'$ in the sense that the joint distribution of V_i and $V_{i'}$ which is denoted by

$$Q_n \left(x, y, x', y' \right) \to P_{\theta_0} \left(x, y \right) P_{\theta_0} \left(x', y' \right) \quad \text{as} \quad n \to \infty, \tag{9}$$

see Ranneby [2] (p. 100). Intuitively, it makes sense that we can assume a similar form of asymptotically independent for the pair $\{z_i \left(\theta, n \right), z_{i'} \left(\theta, n \right)\}$ based on expression (4) and as UWLLN is often stated using sequences with a weak form of dependency of a sequence defined as α-mixing sequence as given by White and Domowitz [7] in the econometric literature, we shall examine the notion of α-mixing sequence and show that we can assume the sequence $\{z_i \left(\theta, n \right)\}$ is α-mixing as we can assume that $\{z_i \left(\theta, n \right)\}$ is ρ-mixing using property given by expression (9). We shall discuss first the notion of ρ-mixing which makes use of the correlation of a pair of random variables as it is more intuitively appealing than the notion α-mixing which makes use of two sets of two separate sigma algebras.

We shall define some more notations.

Let $V_i^0 = \left(X_i^0, Y_i^0 \right)', i = 1, 2, \cdots$. As they have a common distribution let V^0 be one of them, its bivariate density function is given by

$$p_{\theta_0} \left(x, y \right) = f_{\theta_0}^2 \left(x \right) e^{-y f_{\theta_0}(x)}. \tag{10}$$

From the mean value theorem, we have

$$z_i \left(\theta, n \right) = \left(n + 1 \right) \left(F_\theta \left(x_i + \frac{y_i}{n+1} \right) - F_\theta \left(x_i \right) \right) \to y_i f_\theta \left(x_i \right) \tag{11}$$

for each i as $n \to \infty$ which is also given by Ranneby [2] (p. 99); therefore, we also have the equality in distribution $y_i f_\theta \left(x_i \right) =^d y_i^0 f_\theta \left(x_i^0 \right)$ as $n \to \infty$ and we define $z_i^0 \left(\theta \right) = z_i^0$, let z^0 be one of the z_i^0's as they have a common distribution. Consequently, for establishing asymptotic properties as we let $n \to \infty$, we can make use of the distribution of $z_i \left(\theta, n \right)$ which converges in distribution to the distribution of $y_i^0 f_\theta \left(x_i^0 \right)$ and to simplify the notations, let

$$z_i = z_i \left(\theta, n \right) \quad \text{and} \quad z_i \xrightarrow{L} z_i^0 =^d z^0 \tag{12}$$

for all i and the property given by expression (12) can be used to establish asymptotic results subsequently.

It is not difficult to see that the following covariance relationships hold,

$$\lim_{n \to \infty} Cov \left(z_i, z_{i'} \right) = Cov \left(z_i^0, z_{i'}^0 \right) = 0 \tag{13}$$

as V_i^0 and $V_{i'}^0$ are independent for $i \neq i'$. Equivalently, if we use correlation, $\lim_{n \to \infty} \rho_{z_i, z_{i'}} = 0$. The notion of ρ-mixing makes use of the correlation of two

random variables and appears to be easier to grasp, so we shall define this notion first and verify that indeed we can assume the sequence $\{z_i\}$ is ρ-mixing which in turn implies α-mixing which allows us to use available limit theorems results which are stated either using α or ϕ mixing.

Define two sets of random variables which are apart of a distance m as follows $\Omega_a = \{\cdots, z_{a-1}, z_a\}$ and $\Omega_{a+m} = \{z_{a+m}, z_{a+m+1}, \cdots\}$. Pick $z' \in \Omega_a$ and $z'' \in \Omega_{a+m}$ and form the covariance $Cov(z', z'')$ and the correlation coefficient

$$\rho_{z',z''} = \frac{Cov(z', z'')}{\sqrt{Var(z')Var(z'')}}$$

and since z' and z'' have the same distribution $\rho_{z',z''} = \frac{Cov(z', z'')}{Var(z')}$.

Now we can define the ρ-mixing coefficient for the sequence $\{z_i\}$ as $\rho(m) = \sup_a (\rho_{z',z''})$ and the sequence $\{z_i, i = 1, 2, \cdots\}$ will be ρ-mixing if $\rho(m) \to 0$ as $m \to \infty$. Note that Ranneby [2] (p. 99) has pointed out that for finite n, the random variable $z_i's, i = 1, \cdots, n$ are exchangeable in the proof for Lemma (2.1) and from the existing results already established, it appears that $\{z_i, i = 1, 2, \cdots\}$ forms a covariance stationary sequence.

Now $m \to \infty$ will imply $n \to \infty$ and using expression (13), we might conclude that the sequence $\{z_i\}$ is ρ-mixing using the inequality $\rho(m) \leq b\alpha(m)$ with $b > 0$ as given by expression (5.39) in Hall and Heyde [14] (p. 147).

The coefficient $\alpha(m)$ for α-mixing sequence will be defined subsequently. Note that from the above inequality, $\rho(m) \to 0$ implies $\alpha(m) \to 0$ so that $\{z_i, i = 1, 2, \cdots\}$ is also α-mixing. The coefficient $\alpha(m)$ is defined using two sets from two sigma algebras which are of a distance m apart is more abstract but nice results follow. If we can establish a sequence is α-mixing then we can form a new mixing sequence using a transformation applied on a finite number of elements of the original mixing sequence; for example if $\{z_i\}$ is α mixing and define $w_i = g_i(z_i, z_{i-1}, z_{i-2})$ then $\{w_i\}$ is again an α-mixing sequence and furthermore if $\alpha(m)$ for $\{z_i\}$ is $O(m^{-\lambda})$, i.e., $\alpha(m) \to 0$ at the rate as $m^{-\lambda} \to 0$, the same rate is preserved for $\alpha(m)$ of the new sequence $\{w_i\}$, see Lemma 2.1 given by White and Domowitz [7] (p. 146) and discussions as given by Martin *et al.* [15] (p. 38). This property facilitates the use of limit theorems for establishing consistency and asymptotic normality for estimators obtained by minimizing a nonlinear objective function.

Now we shall define the coefficient $\alpha(m)$ for the sequence $\{z_i, i = 1, 2, \cdots\}$. Let S_1 and S_2 be two sigma algebras. Choose a set $F \in S_1$ and another set $G \in S_2$ and define $\alpha(S_1, S_2)$ using the following probability,

$$\alpha(S_1, S_2) = \sup_{F \in S_1, G \in S_2} |P(F \cap G) - P(F)P(G)|$$

The coefficient $\alpha(m)$ for the sequence $\{z_i, i = 1, 2, \cdots\}$ can be defined as

$$\alpha(m) = \sup_a \alpha(\sigma(\Omega_a), \sigma(\Omega_{a+m}))$$

using $\sigma(\Omega_a)$ and $\sigma(\Omega_{a+m})$ which denote respectively the sigma algebras

generated by $\Omega_a = \{\cdots, z_{a-1}, z_a\}$ and $\Omega_{a+m} = \{z_{a+m}, z_{a+m+1}, \cdots\}$.

3. Asymptotic Properties

For establishing asymptotic results for the GSP methods which include the MSP method we do not aim to obtain the results with a minimum amount of regularity conditions as by doing so the technicalities will be increased and might discourage practitioners to use the methods. The regularity conditions used are comparable to regularity conditions for maximum likelihood methods under usual circumstances in order to put GSP methods parallel to ML methods. The aims are to put the GSP methods as equally practical as ML methods for univariate continuous models and to show that it is not more difficult to use these methods than ML methods. Furthermore, by related this class of estimators with to the class of M-estimators, it will be shown that this class can offer more flexible choices for robust estimators should the MSP estimators which are equivalent to ML estimators are not robust and they share similarities with the class of estimators considered by Broniatowski *et al.* [12]. We will treat consistency and asymptotic normality for the GSP estimators in the next two sections.

3.1. Consistency

The objective function to be minimized to obtain the GSP estimators which is denoted by the vector $\hat{\boldsymbol{\theta}}$ is

$$Q_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n} h\left(z_i(\boldsymbol{\theta}, n)\right).$$

The following Theorems can be used to establish consistency for $\hat{\boldsymbol{\theta}}$ and they are listed below as Theorem 1 and Theorem 2. Theorem 1 is the basic consistency Theorem (2.1) given by Newey and McFadden [16] (p. 2121-2122) for estimators obtained by minimizing an objective function $Q_n(\boldsymbol{\theta})$ in general and clearly applicable for $Q_n(\boldsymbol{\theta})$ as defined by expression (5). Theorem 2 is essentially Theorem 2.3 given by White and Domowitz [7] (p. 147), Their Theorem 2.3 is a Theorem on uniform weak law of large numbers (UWLLN) but we shall restate it so that it is more suitable for our purposes. The proofs have been given by the authors and for related Theorems on UWLLN, see Davidson [17] (p. 340-344).

Theorem 1(Consistency)

Assume that:

1) The parameter space $\boldsymbol{\theta}$ is compact, the true vector of parameters is denoted by $\boldsymbol{\theta}_0$,

2) $Q_n(\boldsymbol{\theta}) \xrightarrow{p} Q(\boldsymbol{\theta})$ uniformly and $Q(\boldsymbol{\theta})$ is a non-random and continuous function of $\boldsymbol{\theta}_0$,

3) $Q(\boldsymbol{\theta})$ is uniquely minimized at $\boldsymbol{\theta}_0$,

Then $\hat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}_0$, *i.e.*, we have the convergence in probability which implies consistency for $\hat{\boldsymbol{\theta}}$, where $\hat{\boldsymbol{\theta}}$ is the vector which minimizes $Q_n(\boldsymbol{\theta})$.

To apply this Theorem condition 2) is a condition on uniform convergence,

conditions which ensure UWWLN can be applied will imply condition 2 for $Q_n(\boldsymbol{\theta})$ as given by expression (5) and $Q_n(\boldsymbol{\theta})$ is an average over n terms, this makes it easier to follow the notations for applying the following Theorem 2 on UWLLN. For a detail proof of UWLLN, see Bierens [18] (p. 187).

Implicitly, we assume that $\{z_i, i = 1, 2, \cdots\}$ is a α-mixing sequence with mixing coefficient $\alpha(m)$ which imply the new sequence created using the transformation $h(.)$ is again a α-mixing sequence, *i.e.*, $\{h(z_i), i = 1, 2, \cdots\}$ is also a α-mixing sequence with the same order for $\alpha(m)$.

Theorem 2 (UWLLN)

Assume that:

1) $h(z_i, \boldsymbol{\theta})$ is measurable and continuous for each $\boldsymbol{\theta} \in \boldsymbol{\theta}$, $\boldsymbol{\theta}$ is compact.
2) There exists a function $g(z_i)$ such that $|h(z_i, \boldsymbol{\theta})| \le g(z_i)$ for all $\boldsymbol{\theta} \in \boldsymbol{\theta}$.
3) For $r \ge 1$ and any $\delta > 0$, the expectation $E|g(z_i)|^{r+\delta} \le \Delta < \infty$ for all *i*.
4) $\alpha(m) = O(m^{-\lambda}), \lambda > \dfrac{r}{r-1}, r > 1$, $O(m^{-\lambda})$ is as defined earlier.

Then we have:

1) $E_{\boldsymbol{\theta}_0}(h(z_i, \boldsymbol{\theta}))$ is continuous as a function of $\boldsymbol{\theta}$ for all *i*.
2) $\left| \dfrac{1}{n} \sum_{i=1}^{n} \left(h(z_i, \boldsymbol{\theta}) - E_{\boldsymbol{\theta}_0}(h(z_i, \boldsymbol{\theta})) \right) \right| \xrightarrow{\ p\ } 0$ uniformly as $n \to \infty$ on $\boldsymbol{\theta}$.

Applying Theorem 1 and Theorem 2 will show consistency for the GSP estimators given by the vector $\hat{\boldsymbol{\theta}}$ which minimizes expression (5). We can say that with $\{z_i, i = 1, 2, \cdots\}$ being an α-mixing sequence, the regularity conditions of Theorem 1 and Theorem 2 can be assumed to be satisfied in general with $Q_n(\boldsymbol{\theta}) \xrightarrow{\ p\ } Q(\boldsymbol{\theta})$ uniformly, $Q(\boldsymbol{\theta})$ is continuous and given by

$$Q(\boldsymbol{\theta}) = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} E_{\boldsymbol{\theta}_0}(h(z_i, \boldsymbol{\theta}))$$

and since the $z_i's$ are identically distributed as *z, z* being z_1 for example. We then have $Q(\boldsymbol{\theta}) = \lim_{n \to \infty} E_{\boldsymbol{\theta}_0}(h(z, \boldsymbol{\theta})) = E_{\boldsymbol{\theta}_0}\left(h(z^0, \boldsymbol{\theta})\right)$ with $z^0 = y^0 f_{\boldsymbol{\theta}}(x^0)$ using expression (11) and Dominated Convergence Theorem (DCT). For expressions expressible as a sequence of integrals, the Dominated Convergence Theorem (DCT) might be useful for finding their limits; details and proof of the DCT can be found in standard real analysis books. The joint density function for $V^0 = (X^0, Y^0)'$ is as given by expression (8). Consequently, we can express $Q(\boldsymbol{\theta})$ as

$$Q(\boldsymbol{\theta}) = \int_{-\infty}^{\infty} \int_{0}^{\infty} h(y f_{\boldsymbol{\theta}}(x)) f_{\boldsymbol{\theta}_0}^2(x) \mathrm{e}^{-y f_{\boldsymbol{\theta}_0}(x)} \mathrm{d}y \mathrm{d}x \qquad (14)$$

which is more general but similar to the expression given by Ranne by [2] (p. 97) for the MSP case. We can also express $Q(\boldsymbol{\theta})$ as

$$Q(\boldsymbol{\theta}) = \int_{-\infty}^{\infty} \int_{0}^{\infty} h\left(y f_{\boldsymbol{\theta}_0}(x) \frac{f_{\boldsymbol{\theta}}(x)}{f_{\boldsymbol{\theta}_0}(x)} \right) f_{\boldsymbol{\theta}_0}^2(x) \mathrm{e}^{-y f_{\boldsymbol{\theta}_0}(x)} \mathrm{d}y \mathrm{d}x \,.$$

If we consider the inner integral and make a change of variables with $w = y f_{\boldsymbol{\theta}_0}(x)$ which implies $\mathrm{d}w = f(x) \mathrm{d}y$, this allows to re-express

$$Q(\boldsymbol{\theta}) = \int_{-\infty}^{\infty} \left(\int_0^{\infty} h\left(w \frac{f_{\boldsymbol{\theta}}(x)}{f_{\boldsymbol{\theta}_0}(x)} \right) \mathrm{e}^{-w} \mathrm{d}w \right) f_{\boldsymbol{\theta}_0}(x) \mathrm{d}x \,. \tag{15}$$

Therefore,

$$Q(\boldsymbol{\theta}_0) = \int_{-\infty}^{\infty} \left(\int_0^{\infty} h(w) \mathrm{e}^{-w} \mathrm{d}w \right) f_{\boldsymbol{\theta}_0}(x) \mathrm{d}x \,. \tag{16}$$

For consistency based on Theorem 1, we need to show that $Q(\boldsymbol{\theta}) \geq Q(\boldsymbol{\theta}_0)$ or $Q(\boldsymbol{\theta}) - Q(\boldsymbol{\theta}_0) \geq 0$. We shall make similar assumptions as in the case of ML methods as given by Theorem 2.5 of Newey and McFadden [16] (p. 2131) and use Jensen's inequality to show $Q(\boldsymbol{\theta}) - Q(\boldsymbol{\theta}_0) \geq 0$. The assumptions are:

1) $f_{\boldsymbol{\theta}} \neq f_{\boldsymbol{\theta}'}$, *i.e.*, identification assumption for the parametric family.

2) The vector $\boldsymbol{\theta}_0$ is an interior point of the compact parameter space $\boldsymbol{\theta}$.

3) $F_{\boldsymbol{\theta}}(x)$ is continuous with respect to $\boldsymbol{\theta}$.

The conditions (1 - 3) as given above hold in general, it has been shown that conditions for MSP estimators to be consistent are more relaxed than the condition

$$E_{\boldsymbol{\theta}_0} \left(\sup_{\boldsymbol{\theta}} |\log f_{\boldsymbol{\theta}}(x) \right) < \infty \tag{17}$$

as given by Theorem 2.5 of Newey and McFadden [16] for ML estimators to be consistent but the proofs are very technical and might discourage practitioners to use MSP method or GSP methods in general, see Shao and Hahn [19] and Ekström [20]. We can also compare expression (17) with condition 2 of Theorem 2. In fact, the MSP method originally proposed by Cheng and Amin [1] is for circumstances where ML estimators fail to be consistent and the method was called maximum product of spacings method. Anatolyev and Kosenok [3] also found in many parametric families MSP estimators perform better than ML estimators in finite samples yet being as efficient as ML estimators in large samples. These findings make this class of GSP methods interesting and it is not more complicated to implement GSP methods than ML methods. Now having the entire class defined using a convex function $h(x)$, it also allows the flexibility to choose a robust method within this class.

Since $Q(\boldsymbol{\theta})$ is obtained from a limit operation as $n \to \infty$, we work with the limit density as given by expression (10) with $V^0 = (X^0, Y^0)'$ for the expression $Q(\boldsymbol{\theta})$. By making the change of variables using $u = x^0$, $w = y^0 f_{\boldsymbol{\theta}_0}(x^0)$, the joint density using expression (10) for u and v can be seen to be

$$f_{\boldsymbol{\theta}_0}(u) \mathrm{e}^{-w}, w > 0 \quad \text{and} \quad -\infty < u < \infty \,. \tag{18}$$

It is not difficult to see that:

1) $w = y^0 f_{\boldsymbol{\theta}_0}(x^0)$ and x^0 are independent,

2) the marginal density of W is standard exponential, *i.e.*, the density for W is e^{-w},

3) the marginal density for x^0 is simply $f_{\boldsymbol{\theta}_0}(x^0)$.

We shall see subsequently in the next sections that these properties allow many asymptotic results to be obtained in a unified way and simplify proofs for some results which already appeared in the literature and allow asymptotic nor-

mality to be established for the GSP estimators for multi-parameters estimation which have been established in the paper by Ghosh and Jammalamadaka [6] but the approach is different in this paper. This also facilitates the establishment of asymptotic distribution theory for the parametric tests such as the Wald, score and quasi-likelihood ratio tests for the GSP methods. These tests have their counterparts in likelihood theory or quasi-likelihood theory.

In fact, for asymptotic properties we essentially work with $\{z_i^0\}$ and $\{V_i^0\}$; $\{z_i^0\}$ have a common distribution and are pairwise independent as discussed earlier.

We shall see in the next sections by considering $\{z_i^0, i = 1, 2, \cdots\}$ as an iid sequence, we still have the same asymptotic statistical results for estimators and parametric tests as considering the terms of the sequence $\{z_i^0\}$ has a common distribution and they are pairwise independent. This shows that the assumption of an iid sequence provides a close enough approximation to the true sequence so that asymptotic results for efficiency on estimation and asymptotic distributions for parameter test statistics are unaffected.

For establishing $Q(\boldsymbol{\theta}) \geq Q(\boldsymbol{\theta}_0)$, it suffices to show that $Q(\boldsymbol{\theta}) - Q(\boldsymbol{\theta}_0) \geq 0$. Using expression (16) we can see that

$$Q(\boldsymbol{\theta}_0) = E_W(h(w)) \tag{19}$$

and using expression (16) with a change of order of integration,

$$Q(\boldsymbol{\theta}) = \int_{-\infty}^{\infty} e^{-w} \int_0^{\infty} h\left(w \frac{f_{\boldsymbol{\theta}}(x)}{f_{\boldsymbol{\theta}_0}(x)}\right) f_{\boldsymbol{\theta}_0}(x) \, dx \, dw \,.$$

The inner integral can be expressed as

$$E_{\boldsymbol{\theta}_0}\left(h\left(w \frac{f_{\boldsymbol{\theta}}(x)}{f_{\boldsymbol{\theta}_0}(x)}\right)\right)$$

and since $h(x)$ is convex, we can use Jensen's inequality, see Lehmann and Casella [19] (p. 46-47) for example to conclude

$$E_{\boldsymbol{\theta}_0}\left(h\left(w \frac{f_{\boldsymbol{\theta}}(x)}{f_{\boldsymbol{\theta}_0}(x)}\right)\right) \geq h\left(E_{\boldsymbol{\theta}_0}\left(w \frac{f_{\boldsymbol{\theta}}(x)}{f_{\boldsymbol{\theta}_0}(x)}\right)\right) = h\left(w \cdot E_{\boldsymbol{\theta}_0}\left(\frac{f_{\boldsymbol{\theta}}(x)}{f_{\boldsymbol{\theta}_0}(x)}\right)\right) = h(w)$$

since $E_{\boldsymbol{\theta}_0}\left(\frac{f_{\boldsymbol{\theta}}(x)}{f_{\boldsymbol{\theta}_0}(x)}\right) = 1$.

Therefore,

$$Q(\boldsymbol{\theta}) \geq \int_{-\infty}^{\infty} h(w) e^{-w} \, dw = E_W(h(w)) = Q(\boldsymbol{\theta}_0) \,.$$

This completes the proof for the inequality.

Furthermore, by making a change of variable we can put

$$Q(\boldsymbol{\theta}) - Q(\boldsymbol{\theta}_0) = \int_0^1 \left[E_W\left(h\left(w \frac{f_{\boldsymbol{\theta}}\left(F_{\boldsymbol{\theta}_0}^{-1}(u)\right)}{f_{\boldsymbol{\theta}_0}\left(F_{\boldsymbol{\theta}_0}^{-1}(u)\right)}\right)\right) - E_W(h(w)) \right] du \tag{20}$$

which is the expression used by Ghosh and Jammalamadaka [6] (p. 80) to justify consistency for GSP estimators. In the next section, we turn our attention to asymptotic normality and we shall see that GSP methods can be viewed as quasi-likelihood methods and M-estimation theory can be used to establish asymptotic results.

3.2. Asymptotic Normality and Robustness

For asymptotic normality, often we work with an expression with n being finite then passing to the limit to get the asymptotic results by letting $n \to \infty$ and to make the presentation of the proof easier to follow, we define some notations. For finite n, we have seen $z_i = z_i(\boldsymbol{\theta}, n)$ and to alleviate the notations we also use $z_i = z_i(n)$ and since they have the same distribution we let $z = z(n)$ to denote one of them and to emphasize the dependence on n if necessary.

When passing to the limit by letting $n \to \infty$ we end up working with the sequence $\{z_i^0, i = 1, 2, \cdots\}$ and again since the terms of this sequence have the same distribution we let z^0 to denote one of them. For establishing asymptotic normality under standard conditions which are similar to the ones for ML estimators, we make some assumptions on differentiability on the term $h(z_i, \boldsymbol{\theta})$ which plays a similar role for $\log(f_{\boldsymbol{\theta}}(x_i))$ for likelihood theory.

The asymptotic normality results might continue to hold with less stringent conditions for some parametric families but the proofs would be technical and similar to proofs for maximum likelihood estimators under the nonstandard conditions as in M-estimation theory which are given by Huber [21] (p. 43-51). Fist we state a Theorem for CLT for an α-mixing sequence which is Theorem2.4 by White and Domowitz [6] (p. 147-148) but restated as Theorem 3 below.

Theorem 3 (CLT)

Let $\{M_t\}$ be an α-mixing sequence and define the partial sums $S_n = \sum_{i=1}^{n} M_t$ and

$$T_n^{(a)} = \frac{1}{n} \sum_{i=a+1}^{a+n} M_t \,,$$

Assume that:

1) $E(M_t) = 0$.

2) There exists K finite and nonzero such that $E\left(T_n^{(a)}\right)^2 - K \to 0$ as $n \to \infty$ for all a, a is a positive integer.

3) $E\left(|M_t|^{2r}\right)$ for all t and some $r > 1$.

4) The mixing coefficient $\alpha(m) = O(m^{-\lambda})$ with $\lambda > \dfrac{r}{r-1}$.

Then $\dfrac{S_n}{\sqrt{nK}} \xrightarrow{L} N(0,1)$.

Often, we need to apply Theorem 3 in a multivariate context, Cramer-Wold devices can be used together with Theorem 3, see Davidson [17] (p. 405-407) for these devices. We also define the following notations.

Let

$$\frac{\partial h}{\partial \theta_i}, \frac{\partial h^2}{\partial \theta_j \partial \theta_i}, \frac{\partial h^3}{\partial \theta_k \partial \theta_j \partial \theta_i}, i = 1, \cdots, m, j = 1, \cdots, m, k = 1, \cdots, m$$

be the first, second and third derivatives of the function $h(.)$ with respect to the elements of the vector $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_m)'$ and we shall assume that:

1) The above partial derivatives are continuous with respect to elements of $\boldsymbol{\theta}$.

2) The expectations $E_{\boldsymbol{\theta}_0}\left(\left\|\frac{\partial h(z, \boldsymbol{\theta}_0)}{\partial \theta_i}\right\|\right) < \infty$ and $E_{\boldsymbol{\theta}_0}\left(\left\|\frac{\partial h(z^0, \boldsymbol{\theta}_0)}{\partial \theta_i}\right\|\right) < \infty$.

3) $E_{\boldsymbol{\theta}_0}\left(\frac{\partial h(z, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial h(z, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\right) < \infty$ and $E_{\boldsymbol{\theta}_0}\left(\frac{\partial h(z^0, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial h(z^0, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\right) < \infty$.

4) Interchanging order of integration and differentiation is allowed as in likelihood theory so that

$$E_{\boldsymbol{\theta}}\left(\frac{\partial \log(f_{\boldsymbol{\theta}}(x))}{\partial \boldsymbol{\theta}}\right) = \mathbf{0} ,$$

and the Fisher information $\boldsymbol{I}(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}}\left(\frac{\partial \log(f_{\boldsymbol{\theta}}(x))}{\partial \boldsymbol{\theta}} \frac{\partial \log(f_{\boldsymbol{\theta}}(x))}{\partial \boldsymbol{\theta}'}\right)$ can also be

expressed as

$$\boldsymbol{I}(\boldsymbol{\theta}) = -E_{\boldsymbol{\theta}}\left(\frac{\partial^2 \log(f_{\boldsymbol{\theta}}(x))}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}\right) .$$

5) The convergence of $z_i(\boldsymbol{\theta}, n) \to y_i f_{\boldsymbol{\theta}}(x_i)$ as $n \to \infty$ needs to be strengthened to uniform convergence so that

$$\frac{\partial z_i(\boldsymbol{\theta}, n)}{\partial \boldsymbol{\theta}} \to y_i \frac{\partial f_{\boldsymbol{\theta}}(x_i)}{\partial \boldsymbol{\theta}} .$$

These functions are with respect to $\boldsymbol{\theta}$.

6) The vector $\boldsymbol{\theta}_0$ is an interior point of the parameter space $\boldsymbol{\theta}$, $\boldsymbol{\theta}$ assumed to be compact.

For condition (5), a sufficient condition to have uniform convergence is the sequence of functions with respect to $\boldsymbol{\theta}$, $\{z_i(\boldsymbol{\theta}, n)\}$ is equicontinuous, see Rudin [22] (p. 152-158, p. 168) for these related properties and Davidson and Donsig [23](p. 54) for the Lipchitz property as it can be used to show equicontinuity for a sequence of functions and the Lipchitz property is related to the partial derivatives of the sequence of functions.

Now we can state the following Theorem which is Theorem 4 which give the asymptotic normality results for the GSP estimators in general, *i.e.*, for the multi parameters case and we also verify the result given by expression (9) obtained by Ghosh and Jammalamadaka [6] (p. 76).

Theorem 4 (Asymptotic Normality)

Under Assumptions (165) as given above, then we have the following convergence in distribution for the vector of GSP estimators $\hat{\theta}$ to a multivariate normal distribution,

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)\overset{L}{\longrightarrow}N\left(\mathbf{0},\boldsymbol{\Sigma}_0\right),\quad \boldsymbol{\Sigma}_0=\boldsymbol{A}_0^{-1}\boldsymbol{B}_0\boldsymbol{A}_0^{-1},$$

1) $A_0=E_{\boldsymbol{\theta}_0}\left(\dfrac{\partial^2\left(z^0,\boldsymbol{\theta}_0\right)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right)=E_{\theta_0}\left(\dfrac{\partial^2 h\left(yf_{\boldsymbol{\theta}_0}(x)\right)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right),$

2) $\boldsymbol{B}_0$ being the covariance matrix of the vector of $\dfrac{\partial h\left(z^0,\boldsymbol{\theta}_0\right)}{\partial\theta}$ under $\boldsymbol{\theta}_0$.

3) $\dfrac{\partial h\left(z^0,\boldsymbol{\theta}_0\right)}{\partial\theta}$ is the vector of partial derivatives of $h(.)$ and $\dfrac{\partial^2 h\left(z^0,\boldsymbol{\theta}_0\right)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}$ is the second derivative matrix of $h(.)$ with respect to the elements of the vector $\boldsymbol{\theta}=\left(\theta_1,\cdots,\theta_m\right)'$, $\boldsymbol{B}_0=E_{\boldsymbol{\theta}_0}\left(\dfrac{\partial h\left(z^0,\boldsymbol{\theta}_0\right)}{\partial\boldsymbol{\theta}}\dfrac{\partial h\left(z^0,\boldsymbol{\theta}_0\right)}{\partial\boldsymbol{\theta}'}\right)$, $E_{\boldsymbol{\theta}_0}\left(\dfrac{\partial h\left(z^0,\boldsymbol{\theta}_0\right)}{\partial\boldsymbol{\theta}}\right)=0$ as given by expression (23).

Proof.

Under differentiability assumptions made, the vector of GSP estimators $\hat{\boldsymbol{\theta}}$ is given as roots of the following system of equation as $\hat{\boldsymbol{\theta}}$ minimizes $Q_n(\boldsymbol{\theta})$, *i.e.*,

$$\frac{\partial Q_n(\hat{\boldsymbol{\theta}})}{\partial\theta}=0.$$

Using a Taylor expansion around the true vector of parameters $\boldsymbol{\theta}_0$ of the above system allows us to express

$$0=\frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}+\frac{\partial^2 Q_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)+o_p\left(n^{-\frac{1}{2}}\right) \tag{21}$$

with $o_p\left(n^{-\frac{1}{2}}\right)$ is an expression which converges to 0 in probability faster than

$$\frac{1}{n^{1/2}}\to 0.$$

From expression (21), we have the following representation using equality in distribution

$$-\sqrt{n}\frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}=_d\frac{\partial^2 Q_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\sqrt{n}\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right).$$

We can proceed by using $\{z_i\}$ first but for asymptotic property we let $n\to\infty$, so the same conclusion is reached by assuming we have the model as $\{z_i^0,i=1,2,\cdots\}$ or alternatively, if we derive asymptotic results we can consider $\{(X_i,Y_i)',i=1,2,\cdots\}$ and the bivariate observations which form the sequence follow the common bivariate density as given by expression (10); also they are pairwise independent or it turns out the same asymptotic results can be obtained as the sequence is an iid sequence as it does not affect the asymptotic results that we aim to have. By doing so, we do not need to carry the notation $\left(X_i^0,Y_i^0\right)'$. For asymptotic properties with $n\to\infty$, simply we can let $\left(X_i^0,Y_i^0\right)'=\left(X_i,Y_i\right)'$, and use the properties which states that $y_i f_{\boldsymbol{\theta}_0}(x_i)$ and x_i are independent, $W=y_i f_{\boldsymbol{\theta}_0}(x_i)$ follows a standard exponential distribution.

It is not difficult to see that:

1) $\dfrac{\partial^2 Q_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}$ being the average of α-mixing random variables with a common distribution so that the law of large numbers can be applied and therefore, $\dfrac{\partial^2 Q_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\xrightarrow{p} \boldsymbol{A}_0 = E_{\boldsymbol{\theta}_0}\left(\dfrac{\partial^2\left(z^0,\boldsymbol{\theta}_0\right)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right)$.

2) Similarly, $\sqrt{n}\,\dfrac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\xrightarrow{L} N\left(\boldsymbol{0},\boldsymbol{B}_0\right)$ with $\boldsymbol{B}_0$ being the covariance matrix of the vector of $\dfrac{\partial h\left(z^0,\boldsymbol{\theta}_0\right)}{\partial\boldsymbol{\theta}}$ under $\boldsymbol{\theta}_0$.

We shall show that $E_{\boldsymbol{\theta}_0}\left(\dfrac{\partial h\left(z^0,\boldsymbol{\theta}_0\right)}{\partial\boldsymbol{\theta}}\right)=0$, see expression (23) given below so that

$$\boldsymbol{B}_0 = E_{\boldsymbol{\theta}_0}\left(\frac{\partial h\left(z^0,\boldsymbol{\theta}_0\right)}{\partial\boldsymbol{\theta}}\frac{\partial h\left(z^0,\boldsymbol{\theta}_0\right)}{\partial\boldsymbol{\theta}'}\right). \tag{20}$$

Now applying Slutzky's Theorem if needed as in likelihood theory, we have $\sqrt{n}\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)\xrightarrow{L} N\left(\boldsymbol{0},\boldsymbol{\Sigma}_0\right)$, $\boldsymbol{\Sigma}_0 = \boldsymbol{A}_0^{-1}\boldsymbol{B}_0\boldsymbol{A}_0^{-1}$ is similar to the result obtained for quasi-likelihood estimators or M-estimators, see expression (12.18) given by Woolridge [24] (p. 407) for the asymptotic covariance matrix of M-estimators. Consequently, using the similarities with M-estimators, M-estimation theory can be used if needed to investigate the GSP estimators. This ends the proof.

Subsequently, we shall display the matrices $\boldsymbol{A}_0,\boldsymbol{B}_0,\boldsymbol{\Sigma}_0$ and after simplifications made. First, we consider the matrix

$$\frac{\partial^2 h\left(yf_{\boldsymbol{\theta}_0}(x)\right)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}$$

and with expectation taken will give

$$\boldsymbol{A}_0 = E_{\boldsymbol{\theta}_0}\left(\frac{\partial^2\left(z^0,\boldsymbol{\theta}_0\right)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right)=E_{\boldsymbol{\theta}_0}\left(\frac{\partial^2 h\left(yf_{\boldsymbol{\theta}_0}(x)\right)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right).$$

Note that

$$
\begin{aligned}
\frac{\partial\left(z^0,\boldsymbol{\theta}_0\right)}{\partial\boldsymbol{\theta}} &= \frac{\partial h\left(yf_{\boldsymbol{\theta}_0}(x)\right)}{\partial\boldsymbol{\theta}} = h'\left(yf_{\boldsymbol{\theta}_0}(x)\right)y\,\frac{\partial f_{\boldsymbol{\theta}_0}(x)}{\partial\boldsymbol{\theta}} \\
&= h'\left(yf_{\boldsymbol{\theta}_0}(x)\right)yf_{\boldsymbol{\theta}_0}(x)\frac{\partial\log\left(f_{\boldsymbol{\theta}_0}(x)\right)}{\partial\boldsymbol{\theta}}
\end{aligned}
\tag{22}
$$

by letting

$$\frac{\partial\log\left(f_{\boldsymbol{\theta}_0}(x)\right)}{\partial\boldsymbol{\theta}} = \frac{\dfrac{\partial f_{\boldsymbol{\theta}_0}(x)}{\partial\boldsymbol{\theta}}}{f_{\boldsymbol{\theta}_0}(x)},$$

$h'(x)$ and $h''(x)$ to denote the first and second derivatives of $h(x)$. Clearly,

$$E_{\boldsymbol{\theta}_0}\left(\frac{\partial h\left(yf_{\boldsymbol{\theta}_0}(x)\right)}{\partial \boldsymbol{\theta}}\right) = E_{\boldsymbol{\theta}_0}\left(h'\left(yf_{\boldsymbol{\theta}_0}(x)\right)yf_{\boldsymbol{\theta}_0}(x)\right)E_{\boldsymbol{\theta}_0}\left(\frac{\partial \log\left(f_{\boldsymbol{\theta}_0}(x)\right)}{\partial \boldsymbol{\theta}}\right) = 0 \quad (23)$$

using the independence of $yf_{\boldsymbol{\theta}_0}(x)$ and x. This property is similar to the unbiasedness of the quasi-scores of quasi-likelihood methods or in M-estimation theory. This property also justifies the equality given by expression (20).

The elements of the matrix A_0 are given by taking expectations of the following elements,

$$\ddot{h}_{ij} = h''\left(yf_{\boldsymbol{\theta}_0}(x)\right)y\frac{\partial f_{\boldsymbol{\theta}_0}(x)}{\partial \theta_j}y\frac{\partial f_{\boldsymbol{\theta}_0}(x)}{\partial \theta_i}$$
$$+ h'\left(yf_{\boldsymbol{\theta}_0}(x)\right)y\frac{\partial^2 f_{\boldsymbol{\theta}_0}(x)}{\partial \theta_j \partial \theta_i}, \quad i=1,\cdots,m, \; j=1,\cdots,m$$

Note that the second term of the RHS of the above equality can be expressed as

$$b_{ij} = h'\left(yf_{\boldsymbol{\theta}_0}(x)\right)yf_{\boldsymbol{\theta}_0}(x)\frac{\dfrac{\partial^2 f_{\boldsymbol{\theta}_0}(x)}{\partial \theta_j \partial \theta_i}}{f_{\boldsymbol{\theta}_0}(x)}$$

and upon taking expectation, it is reduced to 0 as

$$E_{\boldsymbol{\theta}_0}(b_{ij}) = E_{\boldsymbol{\theta}_0}\left[h'\left(yf_{\boldsymbol{\theta}_0}(x)\right)yf_{\boldsymbol{\theta}_0}(x)\right]E_{\boldsymbol{\theta}_0}\left[\frac{\dfrac{\partial^2 f_{\boldsymbol{\theta}_0}(x)}{\partial \theta_j \partial \theta_i}}{f_{\boldsymbol{\theta}_0}(x)}\right] \quad \text{and} \quad E_{\boldsymbol{\theta}_0}\left[\frac{\dfrac{\partial^2 f_{\boldsymbol{\theta}_0}(x)}{\partial \theta_j \partial \theta_i}}{f_{\boldsymbol{\theta}_0}(x)}\right] = 0. \quad (24)$$

Let

$$c_{ij} = h''\left(yf_{\boldsymbol{\theta}_0}(x)\right)y\frac{\partial f_{\boldsymbol{\theta}_0}(x)}{\partial \theta_j}y\frac{\partial f_{\boldsymbol{\theta}_0}(x)}{\partial \theta_i}$$
$$= h''\left(yf_{\boldsymbol{\theta}_0}(x)\right)\left(yf_{\boldsymbol{\theta}_0}(x)\right)^2\frac{\partial \log\left(f_{\boldsymbol{\theta}_0}(x)\right)}{\partial \theta_j}\frac{\partial \log\left(f_{\boldsymbol{\theta}_0}(x)\right)}{\partial \theta_i}$$

which implies

$$E_{\boldsymbol{\theta}_0}(c_{ij}) = E_{\boldsymbol{\theta}_0}\left(W^2 h''(W)\right)E_{\boldsymbol{\theta}_0}\left(\frac{\partial \log\left(f_{\boldsymbol{\theta}_0}(x)\right)}{\partial \theta_j}\frac{\partial \log\left(f_{\boldsymbol{\theta}_0}(x)\right)}{\partial \theta_i}\right)$$

using $W = yf_{\boldsymbol{\theta}_0}(x)$ which follows a standard exponential distribution; W and x are independent and properties from expression (18). Using these properties if necessary, the matrix $A_0 = E_{\boldsymbol{\theta}_0}\left(W^2 h''(W)\right)I(\boldsymbol{\theta}_0)$, $I(\boldsymbol{\theta}_0)$ is the Fisher information matrix as commonly defined in likelihood theory.

For comparison with results given by Ghosh and Jammalamadaka [6], A_0 coincides with the corresponding result given as the limit in probability of the expression $\frac{1}{n}T''(\theta_0)$ by Ghosh and Jammalamadaka [6] (p. 81) for the one parameter set-up.

It is not difficult to see that the elements of $\boldsymbol{B}_0$ with

$$\boldsymbol{B}_0 = E_{\boldsymbol{\theta}_0}\left(\frac{\partial h\left(z^0, \boldsymbol{\theta}_0\right)}{\partial \theta} \frac{\partial h\left(z^0, \boldsymbol{\theta}_0\right)}{\partial \boldsymbol{\theta}'}\right)$$

are

$$\left(E_{\boldsymbol{\theta}_0}\left[\left(h'\left(yf_{\boldsymbol{\theta}_0}(x)\right)\right)^2 \left(yf_{\boldsymbol{\theta}_0}(x)\right)^2\right]\right) E_{\boldsymbol{\theta}_0}\left(\frac{\partial \log\left(f_{\boldsymbol{\theta}_0}(x)\right)}{\partial \theta_j} \frac{\partial \log\left(f_{\boldsymbol{\theta}_0}(x)\right)}{\partial \theta_i}\right),$$

using the independence of $yf_{\theta_0}(x)$ and x which implies the independence of $yf_{\theta_0}(x)$ and

$$\frac{\partial \log\left(f_{\boldsymbol{\theta}_0}(x)\right)}{\partial \theta_j} \frac{\partial \log\left(f_{\boldsymbol{\theta}_0}(x)\right)}{\partial \theta_i}$$

for $i = 1, \cdots, m, j = 1, \cdots, m$ and expression (23).

Therefore, $\boldsymbol{B}_0 = \left(E_W\left[\left(h'(W)\right)^2 (W)^2\right]\right) \boldsymbol{I}(\boldsymbol{\theta}_0)$ and $\boldsymbol{\Sigma}_0 = \boldsymbol{A}_0^{-1} \boldsymbol{B}_0 \boldsymbol{A}_0^{-1}$ which implies $\boldsymbol{\Sigma}_0 = \sigma_h^2 \boldsymbol{I}(\boldsymbol{\theta}_0)^{-1}$,

$$\sigma_h^2 = \frac{\left(E_W\left[\left(h'(W)\right)^2 (W)^2\right]\right)}{\left(E_W\left(W^2 h''(W)\right)\right)^2}. \tag{25}$$

The expression for $\boldsymbol{\Sigma}_0$ is very similar to the one in M-estimation theory, see expression (12.18) given by Woolridge [24] (p. 407).

The asymptotic covariance for $\hat{\boldsymbol{\theta}}$ is given as

$$\frac{\sigma_h^2}{n} \boldsymbol{I}(\boldsymbol{\theta}_0)^{-1}. \tag{26}$$

At this point, we would like to make some remarks which are given below.

Remark 1

It appears that a minor adjustment is needed for expression (9) given by Ghosh and Jammalamadaka [6] (p. 76) which gives

$$\sigma_h^2 = \frac{\left(E_W\left[\left(h'(W)\right)^2 (W)^2\right]\right) - \left(2E_W\left[W h'(W)\right]\right) Cov\left(W h'(W), W\right)}{\left(E_W\left(W^2 h''(W)\right)\right)^2} \tag{27}$$

It appears that the term $\left(2E\left[W h(W)\right]\right) Cov\left(W h'(W), W\right)$ which appears in the numerator of the above expression for σ_h^2, $E_W() = E(.)$ is not needed and can be removed, using the properties based on expression (18), also see expression (23) and the derivations of the elements of $\boldsymbol{B}_0$ for the proof of Theorem 4.

An interpretation of asymptotic relative efficiency of the GSP method versus the MSP method can be given to σ_h^2. Also, using the gamma function $\Gamma(.)$, we can obtain for k being a real number and $k > -1$, $E\left(W^k\right) = \Gamma(1 + k)$ from the moment generating function of the log-gamma distribution which gives

$$E\left(W^k\right) = E\left(e^{k\log(W)}\right).$$

For the moment generating function of the log-gamma distribution, see Chan [25].

Remark 2

By relating with M-estimators, we can study the efficiency and robustness for GSP estimators based on a function $h(x)$ used to generate the estimators. It can be seen from the asymptotic results established earlier, the GSP estimators are asymptotically equivalent to M-estimators defined be the following estimating equations

$$\frac{1}{n}\sum_{i=1}^{n}\psi\left(x_i, y_i\right) = 0\,,\quad \psi\left(x_i, y_i\right) = h'\left(y_i f_{\boldsymbol{\theta}}\left(x_i\right)\right) y_i f_{\boldsymbol{\theta}}\left(x_i\right) \frac{\partial \log\left(f_{\boldsymbol{\theta}}\left(x_i\right)\right)}{\partial \boldsymbol{\theta}}$$

is the vector of quasi-score functions.

From M-estimation theory, we already know that for efficiency $\psi\left(x_i, y_i\right)$ must be proportional to the score functions

$$\frac{\partial \log\left(f_{\theta}\left(x_i\right)\right)}{\partial \boldsymbol{\theta}}\quad \text{clearly}\quad h^*(x) = -\log(x)\quad \text{with}\quad h^{*\prime}(x) = -\frac{1}{x}$$

is optimal as in this case,

$$\psi\left(x_i, y_i\right) = \frac{\partial \log\left(f_{\theta}\left(x_i\right)\right)}{\partial \boldsymbol{\theta}}$$

which shows that MSP method is efficient as ML method. This finding has been reported by Ghosh and Jammalamadaka [6] (p. 76) and within the class considered, this is the only optimum choice but the question of robustness has not been discussed. If

$$\frac{\partial \log\left(f_{\boldsymbol{\theta}}(x)\right)}{\partial \boldsymbol{\theta}}$$

is not bounded as a function of x, the MSP estimators might not be robust despite they are efficient.

For robustness we might want to choose $h(x) = -x^{\alpha}$ will $\alpha > 0$ but near 0, such choices of $h(x)$ are suboptimal within the GJ class but they can balance efficiency and robustness. With these choices, the corresponding quasi-score functions are given by

$$\psi\left(x_i, y_i\right) = \left(y_i f_{\boldsymbol{\theta}}\left(x_i\right)\right)^{\alpha} \frac{\partial \log\left(f_{\boldsymbol{\theta}}\left(x_i\right)\right)}{\partial \boldsymbol{\theta}} \tag{27}$$

and clearly $\psi\left(x_i, y_i\right)$ will be bounded as in general having the component $\left(y_i f_{\boldsymbol{\theta}}\left(x_i\right)\right)^{\alpha}$ will be able to keep $\psi\left(x_i, y_i\right)$ bounded as $f_{\boldsymbol{\theta}}(x)$ is a density function and assuming $|y| \to \infty$ only when $|x| \to \infty$ and we have $f_{\theta}(x) \to 0$ as $|x| \to \infty$. Clearly, with $\alpha = 0$, we reobtain the true score functions. The class of pseudo-distance methods as introduced by Broniatowski et al. [12] also share the same type of properties as members of the class can generate quasi-score functions which can approximate the true score functions. One interesting fea-

ture of the GJ class is the class also includes the member that generates the true score functions.

Remark 3

Only for the MSP case that we have $A_0 = B_0$, A_0 and B_0 are as given in Theorem 4 with $A_0 = B_0 = I(\theta_0)$ and only in this case all three tests with their asymptotic distributions as given in section 4 can be used, for other GSP methods we no longer have $A_0 = B_0$ only the Wald test and score test can be used for hypothesis testing with the asymptotic distributions are established and given in section 4, this situation is similar to the one encountered for quasi-likelihood estimation. These three classical tests form the trinity test and the versions for GSP methods will be presented in the next section.

4. Parameter Hypothesis Testing

Now having the asymptotic results for $\hat{\theta}$, we can turn to the question of testing hypothesis and construction of the classical tests such as the Wald test, Lagrange multipliers or Rao's score test and test based on the change of the objective function for the GSP methods. These tests do not seem to have been discussed in the literature for the GSP methods and they are parallel to the tests used for likelihood or quasi-likelihood methods. For these tests using maximum likelihood methods, see Gallant [26] (p. 178-182); also see Woolridge [24] (p. 420-429). For these tests, implicitly we assume that θ_0 is an interior point of the restricted parameter space. The original parameter space is restricted by the conditions imposed by the null hypothesis H_0.

4.1. Wald Test

Often, we are interested to test the null hypothesis which specifies that θ_0 belongs to a subset of the parameter space θ, this is phrased as restrictions imposed on θ_0 via a vector functions $c(\theta) = (c_1(\theta), \cdots, c_q(\theta))'$ which satisfies $c(\theta_0) = 0$ under the null hypothesis H_0 and we assume that $q < m$. With $c(\theta)$, we then have the Jacobian matrix

$$C(\theta) = \begin{bmatrix} \dfrac{\partial c_1(\theta)}{\partial \theta_1} & \cdots & \dfrac{\partial c_1(\theta)}{\partial \theta_m} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial c_q(\theta)}{\partial \theta_1} & \cdots & \dfrac{\partial c_q(\theta)}{\partial \theta_m} \end{bmatrix} \quad \text{and let} \quad C_0 = C(\theta_0).$$

These matrices are assumed to have rank q. With an application of the delta method, we can say that the asymptotic covariance matrix of $\sqrt{n}P_n = \sqrt{n}(c(\hat{\theta}))$, as $c(\theta_0) = 0$ is simply $C_0 \Sigma_0 C_0'$.

Applying Wald's method to construct chi-square statistic using $\sqrt{n}P_n$ will lead to the following quadratic form with a chi-square asymptotic distribution with q degree of freedom using standard results for distribution of quadratic forms.

Therefore, we have an asymptotic chi-square distribution with q degree of

freedom as given below, $\xrightarrow{L}$ is used to denote convergence in distribution and $o_p(1)$ is used to denote an expression which converges to 0 in probability,

$$n(P_n)'(C_0\Sigma_0C_0')^{-1}(P_n)\xrightarrow{L}\chi_q^2.$$

Replacing $\boldsymbol{\theta}_0$ by $\tilde{\boldsymbol{\theta}}$, we can estimate $C_0\Sigma_0C_0'$ by $\tilde{C}\tilde{\Sigma}\tilde{C}'$ and the Wald test statistic is given by

$$W = n(P_n)'\left(\tilde{C}\tilde{\Sigma}\tilde{C}'\right)^{-1}P_n\xrightarrow{L}\chi_q^2 \tag{28}$$

4.2. Score Test or LM Test

The score test is also called Lagrange multiplier (LM) test, it can be derived using the Lagrange multipliers but they do not need to be calculated explicitly as they can be expressed using the quasi-score function of the GSP methods. We only need to fit the restricted model which is specified by the null hypothesis. The vector for restricted estimators is denoted by $\tilde{\boldsymbol{\theta}}$ obtained by minimizing under the constraints $c(\tilde{\boldsymbol{\theta}}) = 0$. The vector for the unrestricted estimators is denoted by $\hat{\boldsymbol{\theta}}$.

For minimizing under q constraints, we introduce the vector of Lagrange multipliers $\boldsymbol{\lambda} = (\lambda_1, \cdots, \lambda_q)'$ and form the Lagrangian function $L(\boldsymbol{\theta}, \boldsymbol{\lambda}) = Q_n(\boldsymbol{\theta}) + \boldsymbol{\lambda}'c(\boldsymbol{\theta})$, also let

$$s_i(\boldsymbol{\theta}) = \frac{\partial h(z_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$$

which can be viewed as the quasi-score of the GSP methods, this will parallel GSP methods with quasi-likelihood methods or likelihood methods. The first order conditions using $L(\boldsymbol{\theta}, \boldsymbol{\lambda})$ will give us the following two systems of equations so that $\tilde{\boldsymbol{\theta}}$ and $\tilde{\boldsymbol{\lambda}}$ should satisfy

$$0 = \frac{\partial L}{\partial \boldsymbol{\theta}} = \frac{\partial Q_n(\tilde{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}} + C'(\tilde{\boldsymbol{\theta}})\tilde{\boldsymbol{\lambda}}, \tag{29}$$

$$0 = \frac{\partial L}{\partial \boldsymbol{\lambda}} = c(\tilde{\boldsymbol{\theta}}). \tag{30}$$

A Taylor series expansion on the system (29) and (30) around $\boldsymbol{\theta}_0$, with $c(\boldsymbol{\theta}_0) = 0$ and assuming $\boldsymbol{\lambda}_0 = \mathbf{0}$ under H_0 coupled with multiplying by a factor $\sqrt{n}$ yields the following two systems of equations which are given by

$$0 =^d \sqrt{n}\frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} + A_0\sqrt{n}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + C'(\tilde{\boldsymbol{\theta}})\tilde{\boldsymbol{\lambda}}$$

as $\boldsymbol{\lambda}_0 = \mathbf{0}$, $0 = C_0\sqrt{n}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + o_p(1)$, $o_p(1)$ is an expression which converges to 0 in probability.

Multiply the first system by $C_0A_0^{-1}$ and using the second system gives the following system of equation expressed using equality in distribution,

$$C_0A_0^{-1}C'(\tilde{\boldsymbol{\theta}})\sqrt{n}\tilde{\boldsymbol{\lambda}} =^d -C_0A_0^{-1}\sqrt{n}\frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}$$

with

$$\sqrt{n}\frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \xrightarrow{L} N(\boldsymbol{0}, \boldsymbol{B}_0)$$

as in the proof of Theorem 4. Therefore, using expression (29), we then have

$$\boldsymbol{C}_0\boldsymbol{A}_0^{-1}\sqrt{n}\frac{\partial Q_n(\tilde{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}} =^d -\boldsymbol{C}_0\boldsymbol{A}_0^{-1}\sqrt{n}\frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}.$$

Let

$$\sqrt{n}\boldsymbol{P}_n = \boldsymbol{C}_0\boldsymbol{A}_0^{-1}\sqrt{n}\frac{\partial Q_n(\tilde{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}}$$

then $\sqrt{n}\boldsymbol{P}_n \xrightarrow{L} N(\boldsymbol{0}, \boldsymbol{M}_0)$ with $\boldsymbol{M}_0 = \boldsymbol{C}_0\boldsymbol{A}_0^{-1}\boldsymbol{B}_0\boldsymbol{A}_0^{-1}\boldsymbol{C}_0'$, Wald's method applying on $\sqrt{n}\boldsymbol{P}_n$ leads to

$$n\left(\frac{\partial Q_n(\tilde{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}}\right)'\boldsymbol{A}_0^{-1}\boldsymbol{C}_0'(\boldsymbol{M}_0)^{-1}\boldsymbol{C}_0\boldsymbol{A}_0^{-1}\left(\frac{\partial Q_n(\tilde{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}}\right)$$

and replacing $\boldsymbol{\theta}_0$ by $\tilde{\boldsymbol{\theta}}$ leads to the Rao's score test or LM test statistic defined as

$$R = n\left(\frac{\partial Q_n(\tilde{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}}\right)'\tilde{\boldsymbol{A}}^{-1}\tilde{\boldsymbol{C}}'\tilde{\boldsymbol{M}}^{-1}\tilde{\boldsymbol{C}}\tilde{\boldsymbol{A}}^{-1}\left(\frac{\partial Q_n(\tilde{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}}\right) \to \chi_q^2, \tag{31}$$

R is with an asymptotic chi-square distribution with q degree of freedom.

Equivalently, if we can assume that $\boldsymbol{A}_0 = \boldsymbol{B}_0 = \boldsymbol{I}(\boldsymbol{\theta}_0)$ as indicated by Wooldridge [24] (p. 425), we can let

$$R = n\left(\frac{\partial Q_n(\tilde{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}}\right)'\tilde{\boldsymbol{A}}^{-1}\left(\frac{\partial Q_n(\tilde{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}}\right), \tilde{\boldsymbol{A}}^{-1} = \tilde{\boldsymbol{B}}^{-1}, \tag{32}$$

using a result established by Wooldrige [24] (p. 424) and expression (12.69) which is based on optimization theory and linear algebra which states that for this type of minimization under constraints, we have with the assumption $\boldsymbol{A}_0 = \boldsymbol{B}_0$ i.e., only for the MSP case that we have: $\tilde{\boldsymbol{A}} = \tilde{\boldsymbol{B}}$ and a matrix $\tilde{\boldsymbol{D}}$ with $\tilde{\boldsymbol{C}}\tilde{\boldsymbol{D}} = 0$ and

$$\tilde{\boldsymbol{D}}'\left(\sqrt{n}\sum_{i=1}^n \frac{\partial Q_n(\tilde{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}}\right) = \boldsymbol{0}$$

such that $\tilde{\boldsymbol{A}}^{-1}\tilde{\boldsymbol{C}}'(\tilde{\boldsymbol{C}}\tilde{\boldsymbol{A}}^{-1}\tilde{\boldsymbol{C}}')^{-1}\tilde{\boldsymbol{C}}\tilde{\boldsymbol{A}}^{-1} = \tilde{\boldsymbol{A}}^{-1} - \tilde{\boldsymbol{D}}(\tilde{\boldsymbol{D}}'\tilde{\boldsymbol{A}}\tilde{\boldsymbol{D}})^{-1}\tilde{\boldsymbol{D}}'$.

The proof is involved and requires preliminary results for linear algebra, we shall not reproduce here, see Wooldrige [24] (p. 424) instead. Now pre and post multiply the RHS of the above expression by

$$\left(\sqrt{n}\sum_{i=1}^n \frac{\partial Q_n(\tilde{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}}\right)$$

will give the expression (32) for R. Note that expression (31) holds in general

without the additional assumption $A_0 = B_0$ and suitable for GSP methods other than MSP method.

Note that for the use of the score test, only the reduced model under H_0 needs to be fitted to obtain $\tilde{\theta}$. This test is of interest under the circumstance when fitting the reduced model is less problematic than fitting the full model without the restrictions. The first test which is the Wald test, one needs to work with the full model. For Wald test and score test, the asymptotic covariances of some expressions need to be obtained.

The following test which is the quasi-likelihood ratio test, we do not need the expressions for asymptotic covariance matrices where partial derivatives are involved but we need to fit the reduced model and the full model to obtain both $\tilde{\theta}$ and $\hat{\theta}$ and the quasi-likelihood likelihood ratio test is only applicable for the MSP method.

4.3. Quasi-Likelihood Ratio Test

The quasi-likelihood ratio test makes use of a statistic which is based on the change of the objective function obtained by fitting the full model and the reduced model, it can be expressed as

$$QLR = 2n\left(Q_n\left(\tilde{\theta}\right) - Q_n\left(\hat{\theta}\right)\right)$$

and we shall see that again we have a chi-square asymptotic distribution for the QLR statistic with $QLR \xrightarrow{L} \chi_q^2$ assuming $A_0 = B_0 = I\left(\theta_0\right)$. Without this condition, the score test as given by expression (31) and Wald test can be used but not the QLR test with an asymptotic chi-square distribution with q degree of freedom. These two tests are asymptotic equivalent under H_0.

We justify the asymptotic distribution as given above for the QLR statistic by expanding $Q_n\left(\tilde{\theta}\right)$ around the vector of unrestricted estimators $\hat{\theta}$ and using A_0, $\dfrac{\partial Q_n\left(\hat{\theta}\right)}{\partial \theta} = 0$ to express

$$Q_n\left(\tilde{\theta}\right) = Q_n\left(\hat{\theta}\right) + \frac{1}{2}\left(\tilde{\theta} - \hat{\theta}\right)' A_0 \left(\tilde{\theta} - \hat{\theta}\right) + o_p\left(n^{-\frac{1}{2}}\right)$$

with $o_p\left(n^{-\frac{1}{2}}\right)$ converges to 0 in probability faster than $n^{-\frac{1}{2}} \to 0$ which gives

$$2n\left(Q_n\left(\tilde{\theta}\right) - Q_n\left(\hat{\theta}\right)\right) =^d \sqrt{n}\left(\tilde{\theta} - \hat{\theta}\right)' A_0 \sqrt{n}\left(\tilde{\theta} - \hat{\theta}\right).$$

But with a Taylor expansion again around $\hat{\theta}$,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} s_i\left(\tilde{\theta}\right) =^d \frac{1}{\sqrt{n}}\sum_{i=1}^{n} s_i\left(\hat{\theta}\right) + A_0 \sqrt{n}\left(\tilde{\theta} - \hat{\theta}\right)$$

with

$$\sqrt{n}\sum_{i=1}^{n}\frac{\partial Q_n\left(\tilde{\theta}\right)}{\partial \theta} = \left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} s_i\left(\tilde{\theta}\right)\right), \quad \frac{1}{\sqrt{n}}\sum_{i=1}^{n} s_i\left(\tilde{\theta}\right) = 0.$$

This implies

$$\sqrt{n}\left(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}\right) \stackrel{d}{=} A_0^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} s_i\left(\tilde{\boldsymbol{\theta}}\right)$$

and using the quasi-score functions,

$$QLR \stackrel{d}{=} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} s_i\left(\tilde{\boldsymbol{\theta}}\right)\right)' A_0^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} s_i\left(\tilde{\boldsymbol{\theta}}\right)\right)$$

which is equivalent to the score statistic,

$$QLR \xrightarrow{L} \chi_q^2. \tag{33}$$

We end this section by noting that GSP methods for multivariate models have been introduced by Kuljus and Ranneby [27] with consistency properties established for the GSP estimators. The approach used in this paper might also be used for a multivariate set up for asymptotic normality results and might lead to similar results as the ones obtained for the univariate case.

5. Conclusion

Asymptotic results for the GSP methods are obtained and presented in a unified way with fewer technicalities which parallel likelihood methods. The implementation of the methods is not more complicated than the implementation of likelihood or quasi-likelihood methods, and the GJ class is large enough to allow more choices for robustness if needed for some parametric models, and at the same time the MSP method within this class is as efficient as likelihood method for continuous univariate models. With all these properties of the GSP methods and simple presentation, we hope to show that these methods are indeed very powerful and useful for continuous univariate models but appear to be under used. Practitioners might want to implement these methods in various fields which include actuarial science for their applied works as they are not more complicated than quasi-likelihood methods.

Acknowledgements

References

[1] Cheng, R.C.H. and Amin, N.A.K. (1989) Estimating Parameters in Continuous Univariate Distributions with a Shifted Origin. *Journal of the Royal Statistical Society Series B*, **45**, 394-403.

[2] Ranneby, B. (1984) The Maximum Spacing Method. An Estimation Method Related to the Maximum Likelihood Method. *Scandinavian Journal of Statistics*, **11**, 93-112.

[3] Anatolyev, S. and Kosenok, G. (2005) An Alternative to Maximum Likelihood Based on Spacings. *Econometric Theory*, **21**, 472-476.

[4] Klugman, S.A., Panjer, H.H. and Willmott, G.E. (2012) Loss Models: From Data to

Decision. Wiley, New York.

[5] Castillo, E., Hadi, A.C., Balakrishnan, N. and Sarabia, J.M. (2005) Extreme Value and Related Models with Applications in Engineering and Science. Wiley, New York.

[6] Ghosh, K. and Jammalamadaka, S.R. (2001) A General Estimation Method Using Spacings. *Journal of Statistical Planning and Inference*, **93**, 71-82. https://doi.org/10.1016/S0378-3758(00)00160-9

[7] White, H. and Domowitz, I. (1984) Nonlinear Regression with Dependent Observations. *Econometrica*, **52**, 143-161. https://doi.org/10.2307/1911465

[8] Pyke, R. (1965) Spacings. *Journal of the Royal Statistical Society*, *Series B*, **27**, 395-449.

[9] Shorack, G.R. and Wellner, J.A. (1986) Empirical Processes with Applications to Statistics. Wiley, New York.

[10] David, H.A. and Nagaraja, H.N. (2003) Order Statistics. Wiley, New York. https://doi.org/10.1002/0471722162

[11] Beran, R. (1977) Minimum Hellinger Distance Estimates for Parametric Models. *Annals of Statistics*, **5**, 445-463. https://doi.org/10.1214/aos/1176343842

[12] Broniatowski, M., Toma, A. and Vajda, I. (2012) Decomposable Pseudodistances and Applications in Statistical Estimation. *Journal of Statistical Planning and Inference*, **142**, 2574-2585. https://doi.org/10.1016/j.jspi.2012.03.019

[13] Lehmann, E.L. and Casella, G. (1998) Theory of Point Estimation. 2nd Edition, Wiley, New York.

[14] Hall, P. and Heyde, C.C. (1980) Martingale Limit Theory and Its Applications. Academic Press, New York.

[15] Martin, V., Hurn, N. and Harris, D. (2012) Econometric Modelling with Time Series: Specification, Estimation and Testing. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9781139043205

[16] Newey, W.K. and McFadden, D. (1994) Large Sample Estimation and Hypothesis Testing. In: Engle, R.F. and McFadden, D., Eds., *Handbook of Econometrics*, Volume 4, Amsterdam, 2111-2245.

[17] Davidson, J. (2002) Stochastic Limit Theory. Oxford University Press, Oxford.

[18] Bierens, H.J. (2005) Introduction to the Mathematical and Statistical Foundations of Econometrics. Cambridge University Press, Cambridge.

[19] Shao, Y. and Hahn, M.G. (1999) Strong Consistency of the Maximum Product of Spacings Estimates with Applications in Nonparametrics and in Estimation of Unimodal Densities. *Annals of the Institute of Statistical Mathematics*, **51**, 31-49. https://doi.org/10.1023/A:1003827017345

[20] Ekström, M. (2001) Consistency of Generalized Maximum Spacing Estimates. *Scandinavian Journal of Statistics*, **28**, 343-354. https://doi.org/10.1111/1467-9469.00241

[21] Huber, P.J. (1982) Robust Statistics. Wiley, New York.

[22] Rudin, W. (1976) Principles of Mathematical Analysis. 3rd Edition, McGraw-Hill, New York.

[23] Davidson, K.R. and Donsig, A.P. (2009) Real Analysis and Applications. Springer, New York.

[24] Woolridge, J.M. (2010) Econometric Analysis of Cross Section and Panel Data. 2nd Edition, MIT Press, Cambridge.

[25] Chan, P.S. (1993) A Statistical Study of Log-Gamma Distribution. Unpublished PhD Thesis, Department of Mathematics, McMaster University, Hamilton.

[26] Gallant, R.A. (1997) An Introduction to Econometric Theory. Princeton University Press, Princeton.

[27] Kuljus, K. and Ranneby, B. (2015) Generalized Maximum Spacing Estimation for Multivariate Observations. *Scandinavian Journal of Statistics*, **42**, 1092-1108.
https://doi.org/10.1111/sjos.12153