

Maximum Entropy Empirical Likelihood Methods Based on Laplace Transforms for Nonnegative Continuous Distribution with Actuarial Applications

Andrew Luong

École d'actuariat, Université Laval, Ste Foy, Québec, Canada
Email: andrew.luong@act.ulaval.ca

How to cite this paper: Luong, A. (2017) Maximum Entropy Empirical Likelihood Methods Based on Laplace Transforms for Nonnegative Continuous Distribution with Actuarial Applications. *Open Journal of Statistics*, 7, 459-482.
<https://doi.org/10.4236/ojs.2017.73033>

Received: April 6, 2017
Accepted: June 11, 2017
Published: June 14, 2017

Copyright © 2017 by author and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Maximum entropy likelihood (MEEL) methods also known as exponential tilted empirical likelihood methods using constraints from model Laplace transforms (LT) are introduced in this paper. An estimate of overall loss of efficiency based on Fourier cosine series expansion of the density function is proposed to quantify the loss of efficiency when using MEEL methods. Penalty function methods are suggested for numerical implementation of the MEEL methods. The methods can easily be adapted to estimate continuous distribution with support on the real line encountered in finance by using constraints based on the model generating function instead of LT.

Keywords

Quasi-Likelihood, Projection, Power Mixture Operator, Quadratic Distance Methods, Insurance Premium, Stop-Loss Premium

1. Introduction

1.1. New Distributions Created Using Laplace Transforms

Nonnegative continuous parametric families of distributions are useful for modeling loss data or lifetime data in actuarial sciences. Many of these families do not have closed form densities. The densities can only be expressed using means of infinite series representations but their corresponding Laplace transforms (LT) have closed form expressions and they are relatively simple to handle. An illustration is given below.

Example

Hougaard [1] introduced the positive tempered stable (PTS); the PTS distribution is obtained by tilting the positive stable (PS) distribution. The random variable $X \geq 0$ follows a positive stable law if the Laplace transform is given as

$$\varphi_{\beta}(s) = E(e^{-sX}) = \int_0^{\infty} e^{-sx} f_{\beta}(x) dx = e^{-\frac{\delta}{\alpha} s^{\alpha}}, 0 < \alpha < 1, \delta > 0, \beta = (\delta, \alpha)'$$

The density function $f_{\beta}(x)$ has no closed form but can be represented as an infinite series.

Now if we create a new distribution using the Esscher transform technique, the corresponding new density can be expressed using $f_{\beta}(x)$ and is given by

$$\frac{e^{-\theta x} f_{\beta}(x)}{\varphi_{\beta}(\theta)},$$

and its LT is

$$L(s) = \frac{\varphi_{\beta}(s + \theta)}{\varphi_{\beta}(\theta)}.$$

This operation adds an extra tilted parameter θ to the vector of parameter β of the original distribution and a new distribution is created. This new distribution is the positive tempered stable (PTS) distribution with Laplace transform given by

$$L(s) = E(e^{-sX}) = \exp\left(-\frac{\delta}{\alpha} [(\theta + s)^{\alpha} - \theta^{\alpha}]\right), \tag{1}$$

$s > 0, \theta > 0, 0 < \alpha < 1, \delta > 0.$

The first four cumulants are given by Hougaard [1] with

$$c_1 = \delta \theta^{\alpha-1}, c_2 = \delta(1-\alpha)\theta^{\alpha-2}, c_3 = \delta(1-\alpha)(2-\alpha)\theta^{\alpha-3},$$

$$c_4 = \delta(1-\alpha)(2-\alpha)(3-\alpha)\theta^{\alpha-4}.$$

For the limiting case $\alpha \rightarrow 0_+$ we have the gamma distribution. In general, the density function has no closed form except for $\alpha = \frac{1}{2}$. For $\alpha = \frac{1}{2}$ we obtain the inverse Gaussian (IG) distribution with density function given by Hougaard ([1], p.392) as

$$f(x; \theta, \alpha) = \frac{\delta}{x^{\frac{3}{2}} \sqrt{\pi}} \exp\left(2\delta\theta^{\frac{1}{2}}\right) \exp\left(-\theta x - \frac{\delta^2}{x}\right), x \geq 0.$$

For other parameterisation for the IG distribution, see Panjer and Wilmott ([2], p.114).

Hougaard [1] has given the name power variance for the PTS distribution and developed moment estimators. There are many names given to this distribution. In the financial literature the name PTS is commonly used, see Schoutens ([3], p.56), Kuchler and Tappe [4]. In actuarial sciences, it is also called generalized gamma distribution, see Gerber [5].

Many new infinitely distributions (ID) can be created using operations on LT based on existing distributions. One of them is the power mixture (PM) operator,

see Abate and Whitt ([6], p.92). It can be summarized as follows. Assume that X_t is an infinitely divisible random variable with LT given by $(\kappa(s))^t, t \geq 0$. The LT of X_t is formed using $\kappa(s)$ which is the LT of a continuous nonnegative, ID random variable X . With the introduction of a new random variable T which is also positive continuous and ID with distribution $H(t)$, a new nonnegative continuous distribution with LT $\kappa_H(s)$ can then be created with LT

$$\eta(s) = EP(\kappa, H) = \int_0^\infty (\kappa(s))^t dH(t) = \kappa_H(-\log(\kappa(s))). \tag{2}$$

The new distribution is created using the power mixture (PM) operator. The PM operator was introduced by Abate and Whitt ([6], p.92). The random variable Y is also called mixing random variable.

The new distribution obtained will have more parameters than the distribution of X . For other methods, such as compounding methods for creating new distributions, see Klugman *et al.* ([7]), p.141-1430. For other ID nonnegative distributions with closed form LT's, see Section (1.2) of Luong [8]. ID nonnegative distributions also appear in risk theory as they arise naturally from Lévy processes often used to model aggregate claim processes, see Gerber [9], Dufresne and Gerber [10] for examples. We often work with ID distribution and for completeness, a definition is given below.

Definition (Lévy-Khintchine Representation). A characteristic function (CF) $\omega(s)$ of a random variable X is infinitely divisible if and only if it can be represented as

$$\omega(v) = \exp\left(ibv + \int_{-\infty}^\infty \left(e^{ivx} - 1 - \frac{ivx}{1+x^2} \right) \left(\frac{1+x^2}{x^2} \right) dG(x) \right),$$

$G(x)$ is a bounded and non-decreasing function with $G(-\infty) = 0$, see Rao [11]. An equivalent expression known as the canonical Lévy-Khintchine representation is also used in the literature, see Sato [12]. A similar representation using LT for nonnegative distribution instead of CF can be found in Feller ([13], p.450).

1.2. Quasi-Likelihood Estimation

For statistical inferences, we assume that we have a random sample with n observations, X_1, \dots, X_n . These observations are independent and identically distributed as X which has a model distribution with closed form LT $L_\beta(s)$, $\beta = (\beta_1, \dots, \beta_p)'$ is the vector of parameters. The true parameter vector is denoted by β_0 . The density function $f_\beta(x)$ has no closed form which makes likelihood estimation difficult to implement. Consequently, we would like to estimate β based on $L_\beta(s)$. Quasi-likelihood among other methods which do not rely on the true density can be considered. A brief review of QL estimation is given below.

Godambe and Thompson [14] developed estimating equations theory and extended quasi-likelihood theory in their seminal paper. They also proposed esti-

mators using quadratic estimating equations which can be viewed as based on a quasi-score functions obtained by projecting the vector of the scores functions denoted by

$$\frac{\partial \log f_{\beta}(x)}{\partial \beta}$$

on the linear spanned space by the basis

$$B_Q = \left\{ g_1(x) = h_1(x) - E(x), g_2(x) = h_2(x) - E(x^2) \right\},$$

$$h_1(x) = x, h_2(x) = x^2.$$

Note that $E(x), E(x^2)$ can be obtained based on the model Laplace transform denoted by $L_{\beta}(s)$. See Chan and Ghosh [15] for a geometry point of view of estimating functions. Godambe and Thompson [14] obtained the most efficient estimators based on quasi-score functions obtained by projecting the true score functions on the linear space spanned by B_Q . They are the most efficient quadratic estimating function (EQEF) estimators within the class of quadratic estimating function introduced by Crowder [16]. The class includes Gaussian quasi-likelihood estimating equations. Consequently, The EQEF estimators are more efficient than normal quasi-likelihood (NQL) estimators in general.

The EQEF estimators are simple to obtain since the basis B_Q has only two elements. The fact they are based on best approximations of the score functions allow them to outperform moment estimators in many circumstances.

For example, we can consider a parametric model with 3 parameters which leads to solve the moment equations, *i.e.*,

$$\frac{1}{n} \sum_{i=1}^n (x_i - E(x)) = 0, \frac{1}{n} \sum_{i=1}^n (x_i^2 - E(x^2)) = 0, \frac{1}{n} \sum_{i=1}^n (x_i^3 - E(x^3)) = 0.$$

It is easy to see that the quasi-score functions of moment methods belong to the linear space spanned by the basis

$$B_{mom} = \left\{ x - E(x), x^2 - E(x^2), x^3 - E(x^3) \right\}.$$

Even that B_{mom} includes all the elements of B_Q , the EQEF methods can outperform the method of moments due to the quasi-score functions of the methods of moments are not the best approximations based on B_{mom} .

Therefore, in this paper we shall emphasize quasi-score functions which make use of best linear combinations of elements of a basis to produce quasi score functions and propose some bases that can provide better efficiencies than the basis formed by linear and quadratic polynomials. The basis should only make use of the model LT. Note that moment estimators based on selected points of the model LT have been discussed in the literature, see Read ([17], p.151-153). The methods appear to be useful for fields of applications which make use extensively of LT of the distributions such as actuarial sciences or engineering.

We shall approach in a unified way so that both QL methods and the MEEL methods are related to the notion of projection of the true score functions on a linear space spanned by a finite base. Within this framework, MEEL estimators

are shown to be asymptotically first order equivalent to QL estimators using the same base. For the first and higher order properties of empirical likelihood estimators, see Newey and Smith [18], Smith [19].

MEEL methods use of informations from the parametric model via constraints and there is one to one correspondence between the constraints, moment conditions and the elements of the basis. Despite that general theory of MEEL or QL methods is well established, the question which bases should we choose to achieve good efficiencies appears to be a relevant one for applications. There is a need to quantify the loss of efficiency as well and consequently in this paper, we also propose a measure of loss of efficiency to evaluate whether MEEL are appropriate methods for analyzing a data set from a specific field of applications.

We hope that the answers will give ideas on how to choose moment conditions or constraints for MEEL estimators. It will also give ideas how to construct semi-parametric bounds as defined by Chamberlain ([20], p.311) which can approximate the parametric bound which is the inverse of the Fisher information matrix. We emphasize MEEL methods but offer a unified view for both MEEL and QL methods as they are related. Numerical implementations of the MEEL methods are also discussed to facilitate practical implementations of these methods for applications in actuarial sciences. We shall discuss the quasi-score functions of the QD methods in the next section.

1.3. Quadratic Distance (QD) Estimation

Let $l_i = l_i(\boldsymbol{\beta}) = \frac{\partial}{\partial \beta_i} \log f(x; \boldsymbol{\beta}), i = 1, \dots, p$ be the score functions and note that

$E(l_i) = 0$ in general. If we try to approximate $l_i(\boldsymbol{\beta})$ using a quasi-score function formed by linear combinations of the functions $h_1(x), \dots, h_k(x)$, this leads to consider quasi-score functions of the form

$$l_i^a = a_0 + \sum_{j=1}^k a_{ij} h_j(x), i = 1, \dots, p.$$

We shall also impose the condition of unbiasedness of estimating function by requiring $E(l_i^a) = 0, i = 1, \dots, p$. With these restrictions, it is equivalent to consider

$$l_i^a = \sum_{j=1}^k a_{ij} (h_j(x) - E_{\boldsymbol{\beta}}(h_j(x))), i = 1, \dots, p.$$

Using vector notations,

$$l_i^a = \mathbf{a}'_i \mathbf{g},$$

$$\mathbf{g}(x) = (h_1(x) - E_{\boldsymbol{\beta}}(h_1(x)), \dots, h_k(x) - E_{\boldsymbol{\beta}}(h_k(x)))'$$

and define

$$\mathbf{h}(x) = (h_1(x), \dots, h_k(x))'.$$

For the best approximation by projecting $l_i(\boldsymbol{\beta})$ on the linear space spanned by the basis $B = \{h_1(x) - E_{\boldsymbol{\beta}}(h_1(x)), \dots, h_k(x) - E_{\boldsymbol{\beta}}(h_k(x))\}$, we look for the

vector of coefficients \mathbf{a}_i^{*} which minimizes

$$E(l_i - \mathbf{a}_i' \mathbf{g})^2,$$

$E(\cdot) = E_{\boldsymbol{\beta}}(\cdot)$, the expectation is taken under $f_{\boldsymbol{\beta}}(x)$.

Using results of the proof of Theorem 4.1 given by Luong and Doray ([21], p 150), the optimum vector is \mathbf{a}_i^* with

$$\mathbf{a}_i^{*} = E_{\boldsymbol{\beta}} \left(\frac{\partial \mathbf{g}'}{\partial \beta_i} \right) \boldsymbol{\Sigma}^{-1}, i = 1, \dots, p \tag{3}$$

$\boldsymbol{\Sigma}$ is the covariance matrix of $\mathbf{h}(x)$ under $f_{\boldsymbol{\beta}}(x)$. We also use the notation $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\beta})$ if an emphasis on the dependence on $\boldsymbol{\beta}$ is needed. It is easy to see that the best approximation is given by

$$l_i = l_i^*(\boldsymbol{\beta}) = \mathbf{a}_i^{*'}(\boldsymbol{\beta}) \mathbf{g}(\boldsymbol{\beta}), i = 1, \dots, k..$$

Note that the elements of \mathbf{a}_i^* need to be spelled out explicitly which means that the covariance matrix $\boldsymbol{\Sigma}$ needs to be known or estimated for applying quasi-likelihood estimation. MEEL estimation does not need this feature and yet produces asymptotic equivalent estimators. This is one of the main advantages of MEEL estimation over QL estimation.

Quadratic distance (QD) estimation as given by Luong and Thompson [22] can be viewed as a form of quasi-likelihood estimation. Numerically, it might be easier to implement QD methods than QL methods defined using estimating functions as there is an objective function to minimize for QD estimation rather than solving for roots of the QL estimating equations. QD estimation will be briefly discussed below.

Let \mathbf{u}_n be a vector defined based on observations,

$$\mathbf{u}'_n = \left(\int_0^\infty h_1 dF_n, \dots, \int_0^\infty h_k dF_n \right).$$

Its model counterpart is given by

$$\mathbf{u}'_{\boldsymbol{\beta}} = \left(\int_0^\infty h_1 dF_{\boldsymbol{\beta}}, \dots, \int_0^\infty h_k dF_{\boldsymbol{\beta}} \right)$$

where F_n is the sample distribution function and its model counterpart is denoted by $F_{\boldsymbol{\beta}}$.

The QD estimators $\widehat{\boldsymbol{\beta}}_Q$ are obtained by minimizing the quadratic form defined as

$$Q(\boldsymbol{\beta}) = (\mathbf{u}_n - \mathbf{u}_{\boldsymbol{\beta}})' \boldsymbol{\Sigma}^{-1} (\mathbf{u}_n - \mathbf{u}_{\boldsymbol{\beta}}).$$

The equivalent quadratic form is

$$Q(\boldsymbol{\beta}) = (\mathbf{u}_n - \mathbf{u}_{\boldsymbol{\beta}})' \widehat{\boldsymbol{\Sigma}}^{-1} (\mathbf{u}_n - \mathbf{u}_{\boldsymbol{\beta}}), \tag{4}$$

$\widehat{\boldsymbol{\Sigma}}$ is a consistent estimate of $\boldsymbol{\Sigma}(\boldsymbol{\beta})$ under the true vector of parameters $\boldsymbol{\beta}_0$. One can see that this procedure is equivalent to use quasi-score functions obtained by projecting the true score functions on the linear space spanned by B since minimizing Expression (3) leads to solve for $\boldsymbol{\beta}$ the system of equations

$$\sum_{j=1}^n E_{\beta} \left(\frac{\partial g'}{\partial \beta_i} \right) \hat{\Sigma}^{-1} g(x_i, \beta) = 0, i = 1, \dots, p,$$

$$E_{\beta} \left(\frac{\partial g'}{\partial \beta_i} \right) = - \frac{\partial E_{\beta}(h')}{\partial \beta_i}, i = 1, \dots, p.$$

Observe that the vector $\hat{a}_i^{*'} = E_{\beta} \left(\frac{\partial g'}{\partial \beta_i} \right) \hat{\Sigma}^{-1}$ is equivalent to $a_i^{*'}$.

From results in Luong and Thompson ([22], p 245) the asymptotic distribution of $\widehat{\beta}_Q$ is given as

$$\sqrt{n}(\widehat{\beta}_Q - \beta_0) \xrightarrow{L} N(0, V).$$

$$V = (S' \Sigma^{-1} S)^{-1} \tag{5}$$

The matrix S' can be expressed as

$$S' = \begin{bmatrix} \frac{\partial E_{\beta}(h_1)}{\partial \beta_1} & \dots & \frac{\partial E_{\beta}(h_k)}{\partial \beta_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial E_{\beta}(h_1)}{\partial \beta_p} & \dots & \frac{\partial E_{\beta}(h_k)}{\partial \beta_p} \end{bmatrix}.$$

The elements of the matrix V are evaluated under $\beta = \beta_0$, S' is the transpose of S .

Following Morton ([23], p.228), the matrix

$$I_B(\beta) = S'(\beta) \Sigma^{-1}(\beta) S(\beta) \tag{6}$$

can be defined as the information matrix of the vector of optimum quasi-score functions and it is related to the semiparametric bounds using the moment conditions as given by Chamberlain ([20], p.311). The moments conditions can be identified with elements of the basis B and so are the constraints used for MEEL methods.

Despite QL and MEEL methods generate asymptotic equivalent estimators, there are reasons to consider MEEL methods rather than quasi-likelihood methods.

With MEEL methods, we have the following main advantages:

1) The matrix Σ which depends on β in general needs to be specified explicitly which might restrict elements to be included in the basis. We can only include elements with relative simple form for their covariances, otherwise Σ will be complicated.

2) If Σ is replaced by a consistent estimate $\hat{\Sigma}$ under β_0 , the estimate is often not accurate enough especially when the sample size n is not large enough and therefore, $\hat{\Sigma}$ tends to be nearly singular even with a few elements in the basis, this creates numerical instability when applying QD methods or quasi-likelihood methods.

3) Goodness of fit test statistics with limiting chi-square distributions for testing the model can be constructed in a unified way with MEEL methods. This

feature is not shared by QL methods.

Within the class of empirical likelihood methods, the MEEL methods are numerically more stable than the original empirical likelihood methods (EL) which were first introduced by Owen [24]. For asymptotic properties of the empirical likelihood methods, see Qin and Lawless [25], Schennach [26], Imbens *et al.* [27]. Also, see the monograph by Owen [28], the book by Mittelhammer *et al.* ([29], p.281-325) and the book by Anatolyev and Gospodinov ([30], p.45-61). It is also worthwhile to note that the MEEL methods are less simulation oriented than indirect inference methods as proposed by Garcia *et al.* [31]. Numerical implementations using penalty function methods are relative simple and will be discussed in Section 4. We hope that with the exposition of the methods in details without too many technicalities, it will encourage people to use these methods in practice. There are many fields beside actuarial sciences where LT for the distribution is widely used. With some modifications, such as using constraints from the model moment generating function instead of the model LT, the methods can be applied to estimate distribution with support on the real line which are often used in finance, see these distributions in Fang and Osterlee [42].

The paper is organized as follows. The choice of bases for generating constraints for the MEEL methods is examined in Section 2. Two families of bases using LT are presented in this section. These two families of bases appear to be useful for actuarial applications. In Section 3, we review asymptotic properties of MEEL methods. An estimate for overall relative efficiency using Fourier cosine series expansion is proposed to quantify the loss of overall efficiency when MEEL methods are used. In Section 4 we examine numerical issues and penalty function methods are advocated to locate the global minimizer which gives the MEEL estimators. Simulations are discussed in Section 5. The simulation study from the positive tempered stable distribution shows that the MEEL estimators are much more efficient than moment estimators originally proposed by the seminal paper of Hougaard [1]. Based on the fields of application, often the full parameter space is not needed and can be restricted to a subspace by having the parameters subject to inequality bounds, the MEEL estimators have the potential to attain high efficiency when comparing to the maximum likelihood (ML) by using a reasonable number of elements in the basis. Actuarial applications are discussed in Section 6.

2. Choice of Bases

Using results in Section 1.2, we consider the basis B which can be used for non-negative continuous distribution or nonnegative distribution with a discontinuity point at the origin with mass assigned to it. The basis B will have the form

$$B = \left\{ x - E(x), x^2 - E(x^2), e^{-\tau x} - L_\beta(\tau), \dots, e^{-m\tau x} - L_\beta(m\tau) \right\}. \quad (7)$$

We observe that the number of elements in the basis is $k = m + 2$ and the elements can be obtained using the LT of the model and therefore suitable for estimation for parametric continuous distribution with density without a closed

form expression.

The number of elements in basis B is finite. It is formed based on the completeness property of the following basis with an infinite number of elements,

$$\{e^{-\tau x} - L_{\beta}(\delta), e^{-2\tau x} - L_{\beta}(2\tau), \dots\}. \tag{8}$$

This infinite basis can be traced back to the work of Zakian and Littlewood [32] who show that a density function can be expressed as an infinite series using elements of the infinite basis given by Expression (8) and develop methods to recover the density function using selected points of its LT. This might explain the potential of high efficiencies of the MEEL estimators constructed using only a finite number of elements of B on some restricted parameter spaces.

The following example will make it clear the notion of restricted parameter spaces. For example, we have a model with two parameters given by $\theta_1 \geq 0, \theta_2 \geq 0$. On a restricted parameter space, the parameters are subject to stricter inequality bounds. For example $0 < a \leq \theta_1 \leq b$ and $0 < c \leq \theta_2 \leq d$ with a, b, c, d are finite positive real numbers.

Therefore, in practice we might want to fix $m = 10$ and $\tau = 0.01$, i.e., let

$$B = \{x - E(x), x^2 - E(x^2), e^{-0.01x} - L_{\beta}(0.01), \dots, e^{-0.1x} - L_{\beta}(0.1)\}. \tag{9}$$

The basis B as indicated above often gives a good balance between numerical simplicity and efficiencies of the estimators.

If the model density has no discontinuity at all then the following basis with negative power moment elements can be considered and we shall see negative power moments can be recovered using the LT. Using the result given in lemma 1 given by Brockwell and Brown ([33], p.630), the following infinite basis with negative power moment element is again complete in general if τ belongs to some interval with $\tau > 0$

$$\{x^{-\tau} - E(x^{-\tau}), x^{-2\tau} - E(x^{-2\tau}), \dots\}.$$

Therefore, the following finite basis

$$C = \{x - E(x), x^2 - E(x^2), x^{-\tau} - E(x^{-\tau}), x^{-2\tau} - E(x^{-2\tau}), \dots, x^{-m\tau} - E(x^{-m\tau})\} \tag{10}$$

can also be considered.

The elements of a basis should respect the regularity conditions of Assumption of Section (3.2) for the estimators to be consistent and have an asymptotic normal distribution. The following example will illustrate this point. In practice for example if $E(x^{-1})$ exists and lower negative power moments do not exist, we might want to choose C to be

$$C = \{x - E(x), x^2 - E(x^2), x^{-\tau} - E(x^{-\tau}), x^{-2\tau} - E(x^{-2\tau}), \dots, x^{-m\tau+h} - E(x^{-m\tau+h})\}, \tag{11}$$

$m = 5, \tau = 0.1$.

The last element is special as it involves h which can be set equal to some

small positive value, for example let $h = 0.01$ for the regularity condition 3) of Assumption 1 to be met. Obviously, if $E(x^{-2})$ exists then we can let $h = 0$.

Now we shall state a proposition which relates negative power moments of a distribution to its LT. The results given by the following proposition are more general than results given by Cressie *et al.* [34] who only give results for negative integer moments. The general results can be traced back to Theorem 2.1 given by Brockwell and Brown ([35], p.215) but it can be difficult to find this reference, so we reproduce the results below.

Proposition

Suppose that X is a nonnegative continuous random variable with density function and Laplace transform given respectively by $f(x)$ and $L(s)$ then if $E(x^{-u})$ exists, it is given by $E(x^{-u}) = \frac{1}{\Gamma(u)} \int_0^\infty s^{u-1} L(s) ds, u > 0$, $\Gamma(u)$ is the commonly used gamma function, assuming the integral exists.

Proof.

Observe that $\int_0^\infty s^{u-1} L(s) ds = \int_0^\infty \left(\int_0^\infty s^{u-1} e^{-sx} ds \right) f(x) dx$ by switching the order of integration and note that the inner integral can be expressed as

$$\int_0^\infty s^{u-1} e^{-sx} ds = x^{-u} \Gamma(u), u > 0, ,$$

using properties of a gamma distribution. The integral $\int_0^\infty s^{u-1} L(s) ds, u > 0$ if it exists can be evaluated numerically. Most of computer packages provide built-in functions to evaluate these integrals numerically. For the positive stable distribution or gamma distribution negative power moments have closed form expressions, see Luong and Doray ([21], p.149). The bases B and C only provide guidelines to form a good basis based on LT. We can also combine or select elements from these bases to form new bases.

3. MEEL Methods

3.1. Two Stages Distance Methods

MEEL methods as discussed in chapter 13 by Mittelhammer *et al.* ([29], p.313-326) belong to the class of empirical likelihood methods. The name MEEL is given by the authors and MEEL methods are based on the Kullback-Leibler distance which belongs to the class of distance for discrete distribution introduced by Cressie-Read [36]. MEEL methods are also called exponential tilted empirical likelihood methods in the literature, see Imbens [37]. The methods are asymptotically as efficient as other EL methods. The main reason which leads us to emphasize MEEL methods over EL methods is the advantage of the numerical stability of MEEL methods, see Schennach [26] for this property. We shall discuss how to implement MEEL methods with the specifications of constraints extracted from LT of the original model. These constraints are associated with moment conditions or elements of a finite base. MEEL methods can also be viewed conceptually as a two stages distance methods based on the Kullback-Leibler distance for discrete distributions. The first stage consists of choosing the

best proxy discrete model to replace the original parametric model and the second stage consists of using the best discrete proxy model to estimate the parameters of the original model.

Assume that we have a random sample as in Section 1.2. The vector β has p components, *i.e.*,

$$\beta = (\beta_1, \dots, \beta_p)'$$

We are in the situation where the density $f_\beta(x)$ has no closed form expression but using the LT, we can extract k moments of the original parametric model, $(E_\beta(h_1(x)), \dots, E_\beta(h_k(x)))'$ assuming $k > p$.

Clearly, the sample distribution function corresponds to a discrete distribution which assigns the mass $p_{in} = \frac{1}{n}$ at the realized point of the observation $i = 1, \dots, n$ and $\sum_{i=1}^n p_i = 1$. Now instead of using the original model for inferences we shall consider proxy discrete models with mass function $\pi_i(\beta)$ assigning mass at the realized point of observations. Let

$$p_n = (p_{1n}, \dots, p_{mn})' = \left(\frac{1}{n}, \dots, \frac{1}{n}\right)'$$

$$\pi = \pi(\beta) = (\pi_1, \dots, \pi_n)'$$

The Kullback-Leibler distance between the two discrete distributions p_n and π is defined by the following measure of discrepancy,

$$KL(\pi, p_n) = \sum_{i=1}^n \pi_i \left(\ln \pi_i - \ln \left(\frac{1}{n}\right) \right).$$

We also require the proxy model beside satisfying the basic requirement, *i.e.*, $\sum_{i=1}^n \pi_i = 1, \pi_i \geq 0$, it also satisfies the same moment conditions of the original parametric model, *i.e.*,

$$E_\pi(h_j(x)) = E_\beta(h_j(x)), j = 1, \dots, k,$$

$$E_\pi(h_j(x)) = \sum_{i=1}^n \pi_i h_j(x).$$

Parametric estimation will be carried out in two stages. The first stage is to choose the best proxy model by minimizing $KL(\pi, p_n)$ which is equivalent to maximize the entropy measure with the above constraints. It leads to maximize $-\sum_{i=1}^n \pi_i \ln \pi_i$ or equivalently minimize

$$\sum_{i=1}^n \pi_i \ln \pi_i \tag{12}$$

subject to the constraints given by

$$\sum_{i=1}^n \pi_i = 1, \tag{13}$$

$$\sum_{i=1}^n \pi_i (g_j(x_i; \beta)) = 0, j = 1, \dots, k \tag{14}$$

with

$$g_j(x_i; \beta) = h_j(x_i) - E_\beta(h_j(x)), j = 1, \dots, k. \tag{15}$$

Mittelhammer *et al.* ([29], p.321) have shown that the Lagrangian of the optimization problem is

$$L(\boldsymbol{\pi}, \boldsymbol{\lambda}, \mu) = \sum_{i=1}^n \pi_i \ln \pi_i + \sum_{j=1}^k \lambda_j \left(\sum_{i=1}^n \pi_i g_j(x_i; \boldsymbol{\beta}) \right) + \mu \left(\sum_{i=1}^n \pi_i - 1 \right)$$

with $\lambda_j, j = 1, \dots, k$ and μ are Lagrange multipliers. Taking partial derivatives with respect to π_i leads to the system of equation

$$\frac{\partial L}{\partial \pi_i} = \ln \pi_i + 1 + \sum_{j=1}^k \lambda_j (g_j(x_i; \boldsymbol{\beta})) + \mu = 0, i = 1, \dots, n. \tag{16}$$

The solutions of the equation yield the best discrete proxy model with mass function given by

$$\pi_i^* = \frac{\exp\left(-\sum_{j=1}^k \lambda_j(\boldsymbol{\beta})(g_j(x_i; \boldsymbol{\beta}))\right)}{\sum_{i=1}^n \exp\left(-\sum_{j=1}^k \lambda_j(\boldsymbol{\beta})(g_j(x_i; \boldsymbol{\beta}))\right)}, i = 1, \dots, n \tag{17}$$

which is Expression (13. 2.6) given by Mittelhammer *et al.* ([29], p.321). Note that $\lambda_j = \lambda_j(\boldsymbol{\beta})$ and the λ_j 's are defined implicitly by Expression (16).

Note that since the π_i 's are defined implicitly, they depend on $\boldsymbol{\beta}$ but do not depend on the Lagrange multiplier μ as it is easy to see that we already have $\sum_{i=1}^n \pi_i^* = 1$ and $\pi_i^* \geq 0, i = 1, \dots, n$.

$$\text{Let } \boldsymbol{\pi}^* = (\pi_1^*, \dots, \pi_n^*)', \boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_k)'$$

The second stage is to use the KL distance for parametric inferences. At this stage, we minimize with respect to $\boldsymbol{\beta}$ the expression

$$\sum_{i=1}^n \pi_i^*(\boldsymbol{\beta}) \ln \pi_i^*(\boldsymbol{\beta}) \tag{18}$$

to obtain the MEEL estimators $\hat{\boldsymbol{\beta}}$.

The numerical procedures to implement MEEL methods appear to be complicated as the π_i 's are defined implicitly. Numerical procedures are simplified by using penalty function methods and will be discussed in Section 4. With this approach, it suffices to perform unconstrained minimization with respect to $k + p$ variables given by $\lambda_j, j = 1, \dots, k, \beta_i, i = 1, \dots, p$ with $k > p$ using a suitably defined objective function. The relationships between the vector $\boldsymbol{\lambda}$ and the vector $\boldsymbol{\beta}$ are given by

$$\sum_{i=1}^n \pi_i^*(\boldsymbol{\lambda}, \boldsymbol{\beta})(g_j(x_i; \boldsymbol{\beta})) = 0, j = 1, \dots, k \tag{19}$$

will be used to build the penalty function part of the new objective function.

Imbens ([37], p.501-502) also advocated the use of a specific version of penalty function approach to obtain the MEEL estimators. Chong and Zak ([38], p.564-571) give details on how to construct penalty function to handle optimization under equality and inequality constraints and can be a good reference for using penalty methods. We choose to follow more closely penalty methods of nonlinear optimization used in the literature as given by Chong and Zak ([38]) and suggest strategy to identify the global minimizer vector which is the vector

of the estimators.

The identification of the global minimizer in nonlinear estimation which gives the estimates is an important one as most of the algorithms only give local minimizers and therefore are vulnerable to starting points used to initialize the algorithm, see Davidson and McKinnon ([39], p.232-233) for a strategy using different starting points. Andrews [40] proposes to use the criterion functions of goodness of fit test statistics to limit the search for the global minimizer in a suitable restricted parameter space, this can be handled easily with penalty function methods with inequality constraints. For performing a global random search based on simulated annealing, see Robert and Casella ([41], p.140-146).

3.2. Asymptotic Properties

3.2.1. Asymptotic Covariance

The regularity conditions for the MEEL estimators $\hat{\beta}$ to be consistent and to follow an asymptotic normal distribution have been given by Assumption 1 in Schennach ([26], p.645) who also provides proofs for consistency and asymptotic normality of the MEEL estimators. The regularity conditions are reproduced below. Also, see Expressions (13.2.10), (13.2.11) of the book by Mittelhammer *et al.* ([29], p.323).

Assumption

Assume that:

1) The true parameter given by the vector β_0 is an interior point of the parametric space θ which is assumed to be compact.

2) β_0 is the unique vector which satisfies $E_{\beta_0}(\mathbf{g}(x, \beta_0)) = 0$.

3) $\mathbf{g}(x, \beta)$ is differentiable with respect to β and $E_{\beta_0}(\sup_{\beta} |g_i|^{2+h}) < \infty$ for some $h > 0, i = 1, \dots, k$.

4) The derivatives of $\mathbf{g}(x, \beta)$, $\frac{\partial g_i}{\partial \beta_j}, i = 1, \dots, k$, also satisfy the local boundedness condition $E_{\beta_0} \left(\sup_{\beta} \left| \frac{\partial g_i}{\partial \beta_j} \right|^{2+\delta} \right) < \infty$ for some $\delta > 0$ when β is restricted to some neighborhood of β_0 .

5) The covariance matrix Σ of $\mathbf{g}(x, \beta)$ has rank k .

Under Assumption 1, then the MEEL estimators given by the vector $\hat{\beta}$ is consistent and have a multinormal asymptotic distribution, $\hat{\beta} \xrightarrow{p} \beta_0$, β_0 is the vector of the true parameters, $\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{p} N(0, \Omega)$,

$$\Omega = \left[E \left[\frac{\partial \mathbf{g}(x, \beta)}{\partial \beta} \Big|_{\beta = \beta_0} \right] \left(E \left[\mathbf{g}(x, \beta) \mathbf{g}(x, \beta)' \right]_{\beta = \beta_0} \right)^{-1} E \left[\frac{\partial \mathbf{g}(x, \beta)}{\partial \beta'} \Big|_{\beta = \beta_0} \right] \right]^{-1}, \quad (20)$$

$$\mathbf{g}(x, \beta) = (g_1(x; \beta), \dots, g_k(x; \beta))', \Sigma(\beta_0) = E \left[\mathbf{g}(x, \beta) \mathbf{g}(x, \beta)' \right]_{\beta = \beta_0}.$$

An estimator $\hat{\Omega}$ for Ω can be defined,

$$\hat{\Omega} = \left[\left[\sum_{i=1}^n \hat{\pi}_i \frac{\partial \mathbf{g}(x_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} \right] \left(\sum_{i=1}^n \hat{\pi}_i \mathbf{g}(x_i, \boldsymbol{\beta}) \mathbf{g}(x_i, \boldsymbol{\beta})' \Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} \right)^{-1} \right. \\ \left. \left[\sum_{i=1}^n \hat{\pi}_i \frac{\partial \mathbf{g}(x_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}'} \Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} \right] \right]^{-1},$$

$\hat{\pi}_i = \pi_i^*(\hat{\boldsymbol{\beta}}), i = 1, \dots, n$. If we let $\hat{\pi}_i = \frac{1}{n}, i = 1, \dots, n$, we have another consistent estimator for Ω .

Note that Ω is identical to Expression (5) which shows the asymptotic equivalence between optimum quasi-likelihood estimation and MEEL estimation. Both methods do not need full specifications of the model but only require moment conditions of the true model.

3.2.2. Goodness-of-Fit Test Statistics

The use of the KL distance also allows construction of a goodness-of-fit test statistics which follows an asymptotic chi-square distribution. The validity of the original model is reduced to the validity of moment conditions, we might want to test the null hypothesis specified as $H_0 : E_{\boldsymbol{\beta}}(h_j(x)), j = 1, \dots, k$, the expectations are under the true parametric model.

The following test statistics given below is a chi-square test statistics with $r = k - p$ degree of freedom, *i.e.*,

$$2nKL(\boldsymbol{\pi}^*(\hat{\boldsymbol{\beta}}), \mathbf{p}_n) \\ = 2n \left(\sum_{i=1}^n \pi_i^*(\hat{\boldsymbol{\beta}}) \left[\ln \pi_i^*(\hat{\boldsymbol{\beta}}) - \ln \left(\frac{1}{n} \right) \right] \right) \xrightarrow{L} \chi^2(k - p). \tag{21}$$

3.3. An Estimate for the Overall Relative Efficiency

It is clear that only under special circumstances that MEEL methods are as efficient as ML methods due to the use of a finite basis. This can only happen when the true score functions belong to the linear space spanned by a finite basis. Therefore, it appears to be useful to be able to quantify the loss of efficiency when using MEEL methods despite the model density has no closed form expression to check whether MEEL methods are appropriate for a specific field of applications. Fourier series expansion can be useful to approximate the density function and will be introduced below.

The density function can be expanded using Fourier cosine series in the range $0 < x < b$, see Expressions (7-11) given by Fang and Osterlee ([42], p.6), Powers ([43], p.62), *i.e.*,

$$f_{\boldsymbol{\beta}}(x) \sim f_{\boldsymbol{\beta}}^a(x), f_{\boldsymbol{\beta}}^a(x) = F_0(\boldsymbol{\beta}) + \sum_{j=1}^{\infty} F_j(\boldsymbol{\beta}) \cos\left(\frac{j\pi}{b}x\right).$$

The coefficients $F_j(\boldsymbol{\beta}), j = 0, 1, 2, \dots$ are Fourier coefficients,

$$F_0(\boldsymbol{\beta}) = \frac{1}{b} \int_0^b \cos\left(\frac{j\pi}{b}x\right) f_{\boldsymbol{\beta}}(x) dx, j = 0,$$

$$F_j(\boldsymbol{\beta}) = \frac{2}{b} \int_0^b \cos\left(\frac{j\pi}{b}x\right) f_{\boldsymbol{\beta}}(x) dx, \quad j = 1, 2, \dots$$

Regularity conditions for uniform convergence of Fourier series are also given by Powers ([43], p.72-73). The derivatives of these coefficients with respect to $\beta_l, l = 1, \dots, p$ are given by

$$F'_{0,\beta_l}(\boldsymbol{\beta}) = \frac{1}{b} \int_0^b \cos\left(\frac{j\pi}{b}x\right) \frac{\partial f_{\boldsymbol{\beta}}(x)}{\partial \beta_l} dx$$

$$F'_{j,\beta_l}(\boldsymbol{\beta}) = \frac{2}{b} \frac{\partial}{\partial \beta_l} \left(\int_0^b \cos\left(\frac{j\pi}{b}x\right) f_{\boldsymbol{\beta}}(x) dx \right), \quad l = 1, \dots, p, \quad j = 1, 2, \dots$$

If b is chosen sufficiently large, we have the following approximations of the coefficients using either the characteristic function (CF) or LT,

$$F_j(\boldsymbol{\beta}) \approx \bar{F}_j(\boldsymbol{\beta}) = \frac{2}{b} \int_0^\infty \cos\left(\frac{j\pi}{b}x\right) f_{\boldsymbol{\beta}}(x) dx = \frac{2}{b} \operatorname{Re} \left(L_{\boldsymbol{\beta}} \left(-i \frac{j\pi}{b} \right) \right), \quad j = 1, 2, \dots$$

and $\bar{F}_0(\boldsymbol{\beta}) = \frac{1}{b}$. Similarly,

$$F'_{j,\beta_l}(\boldsymbol{\beta}) \approx \bar{F}'_{j,\beta_l} = \frac{2}{b} \frac{\partial}{\partial \beta_l} \operatorname{Re} \left(L_{\boldsymbol{\beta}} \left(-i \frac{j\pi}{b} \right) \right),$$

$$\bar{F}'_{0,\beta_l}(\boldsymbol{\beta}) = 0, \quad l = 1, \dots, p, \quad j = 1, 2, \dots, M,$$

$\operatorname{Re}(\dots)$ is the real part of the complex number inside the parenthesis and most of the computer packages can handle complex numbers computations. In practice, we can only use a finite cosine series expansion with M terms. The formulas for the coefficients given by Fang and Osterlee ([42], p.6) make use of the characteristic function but they can be converted easily to expressions using LT. Using these truncated series, it leads to approximate the score functions by

$$\frac{\partial \log \bar{f}^a(x)}{\partial \beta_l} = \frac{\frac{\partial \bar{f}^a(x)}{\partial \beta_l}}{\bar{f}^a(x)}, \quad l = 1, \dots, p$$

with

$$\bar{f}^a(x) = \frac{1}{b} + \sum_{j=1}^M \bar{F}_j(\boldsymbol{\beta}) \cos\left(\frac{j\pi}{b}x\right),$$

$$\frac{\partial \bar{f}^a(x)}{\partial \beta_l} = \sum_{j=1}^M \bar{F}'_{j,\beta_l} \cos\left(\frac{j\pi}{b}x\right), \quad l = 1, \dots, p.$$

Therefore, if $\boldsymbol{\beta}$ is estimated by $\hat{\boldsymbol{\beta}}$, the Fisher information matrix $\mathbf{I}(\hat{\boldsymbol{\beta}})$ can be estimated by $\hat{\mathbf{I}}(\hat{\boldsymbol{\beta}})$ using the original sample or simulated samples from the distribution with $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$. If the original sample is used,

$$\hat{\mathbf{I}}(\hat{\boldsymbol{\beta}}) = \frac{1}{n} \sum_{i=1}^n \left(\frac{\partial \ln \bar{f}^a(x_i)}{\partial \boldsymbol{\beta}} \right) \left(\frac{\partial \ln \bar{f}^a(x_i)}{\partial \boldsymbol{\beta}} \right)' \Bigg|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}}.$$

The estimate overall relative efficiency can be defined based on Expression (20) as

$$ARE(\hat{\beta}) = \frac{\det(\hat{U}(\hat{\beta}))}{\det(\hat{I}(\hat{\beta}))}, \quad (22)$$

$\hat{U}(\hat{\beta}) = S'(\hat{\beta})\hat{\Sigma}^{-1}S(\hat{\beta})$, $\det(\cdot)$ is the determinant of the matrix inside the paranthesis, see Expression (3.7) given by Bhapkar ([44], p.471) for overall relative efficiency using determinants of matrices. Instead of determinants of matrices, the traces of matrices can also be used, this leads to alternative measure of overall relative efficiency. Fang and Osterterlee ([42]) show that finite cosine Fourier series converge at an exponential rate which suggest that with $M \geq 500$, the approximation should be quite accurate if the model density is continuous using examples given by their paper. The value M can be increased for more accuracy if needed.

For the value of b , we can let $b = \bar{X} + Ls$, $10 \leq L \leq 15$, \bar{X} and s are respectively the sample mean and sample standard deviation. Note that $ARE(\hat{\beta})$ despite its simplicity can give an idea whether MEEL methods are appropriate for the data set and the parametric model being considered.

4. Numerical Implementations

We shall use penalty function approaches to convert the problem of minimization with constraints to a problem of minimization without constraints by introducing a surrogate objective function which is defined suitably. The techniques of penalty function are well described in Chong and Zak ([38], p.560-567). They can handle both equalities and inequalities constraints. The new objective function can be minimized using a precise direct search based on Nelder-Mead simplex methods for example. The simplex methods are derivative free and converge to local optimizers, see Chong and Zak ([38], p.274-278). The package R has built-in functions to perform simplex algorithm with constraints.

For illustration, we start with a simple example and extend it to the problem for finding MEEL estimators.

Suppose that we wish to minimize a function $f(x_1, x_2)$ with two variables x_1 and x_2 subject to a constraint $c = c(x_1, x_2) = 0$. The numerical solutions of this problem can be found by minimizing the following unconstrained objective function given by $f(x_1, x_2) + \frac{K}{2}([c(x_1, x_2)]^2)$, $K \rightarrow \infty$. In practice setting a value for K being very large gives solutions with numerical accuracy. The penalty function which makes use of the square function is the second component of the objective function. The minimization procedures can give exact solutions with the use of a more complicated nondifferentiable penalty function, see Chong and Zak ([38], p.570-571).

For the MEEL minimization problem, we have $\lambda_j, j = 1, \dots, k$ depend on β and π_i^* are given by

$$\pi_i^*(\lambda, \beta) = \frac{\exp\left(-\sum_{j=1}^k \lambda_j \left(\sum_{i=1}^n g_j(x_i; \beta)\right)\right)}{\sum_{i=1}^n \exp\left(-\sum_{j=1}^k \lambda_j \left(\sum_{i=1}^n g_j(x_i; \beta)\right)\right)}, i = 1, \dots, n.$$

The vectors λ and β are related by the equality constraints given by

$$c_1 = \sum_{i=1}^n \pi_i^*(\lambda, \beta) [g_1(x_i, \beta)] = 0, \dots, c_k = \sum_{i=1}^n \pi_i^*(\lambda, \beta) [g_k(x_i, \beta)] = 0. \tag{23}$$

Therefore, we can perform unconstrained minimization using the following objective function with respect to $\lambda_1, \dots, \lambda_k$ and β_1, \dots, β_p ,

$$\sum_{i=1}^n \pi_i^*(\lambda, \beta) \ln \pi_i^*(\lambda, \beta) + \frac{K}{2} \left[\left(\sum_{i=1}^n \pi_i^*(\lambda, \beta) [g_1(x_i, \beta)] \right)^2 + \dots + \left(\sum_{i=1}^n \pi_i^*(\lambda, \beta) [g_k(x_i, \beta)] \right)^2 \right]. \tag{24}$$

The penalty constant K is a large positive value, setting $K = 500000$ for example. If the absolute value function is used to construct the penalty function then we can only use direct search algorithms which are derivative free.

It is worth to note that only a local minimizer is found each time using these algorithms, some strategies are needed to identify the global minimizer. The following procedures can be used:

1) We might need a starting vector being close to the estimators to initialize the algorithm, this is important when working with real data. For example, we might want to consider starting the algorithm with simple but consistent estimators given by $\hat{\beta}_s$ obtained by minimizing $\sum_{j=1}^k \left(\sum_{i=1}^n g_j(x_i; \beta) \right)^2$.

If the number of parameters are not large, global random search can be performed. Simulated annealing (SA) or particle swarm optimization (PSO) are commonly used global random search technique, see Chong and Zak ([38], p. 279-285) to supplement local search algorithm. This problem is less severe for local search algorithm using simulated data since the true vector β_0 is known. The simple estimators can be considered as quasi-likelihood estimators which can make use of a larger basis than the one used to generate MEEL estimators since there are less numerical difficulties to compute the simple estimators, there is no need to estimate Σ^{-1} . However, these quasi score functions are no longer orthogonal projections on the larger basis used. Based on remark 2.4.3 given by Luong and Thompson ([22], p.245) which gives the asymptotic covariance matrix of $\hat{\beta}_s$, the overall relative efficiency can be defined as

$$ARE(\hat{\beta}_s) = \frac{\det(V_s(\hat{\beta}_s))}{\det(\hat{I}^{-1}(\hat{\beta}_s))},$$

$$V_s(\hat{\beta}_s) = (S'S)^{-1} S' \Sigma^{-1} S (S'S)^{-1},$$

evaluated at $\beta = \hat{\beta}_s$.

2) For finding the global minimizer Andrews ([40], p.919-921) has suggested the use of the criterion function of a goodness of fit test statistics to identify good starting vectors by requesting a good starting vector $\beta^{(0)}$ must satisfy the inequality

$$2nKL(\pi^*(\lambda^{(0)}, \beta^{(0)}), p_n) \leq \chi_{0.95}^2(k-p), \tag{25}$$

$\chi_{0.95}^2(k-p)$ is the 0.95 percentile of the chi-square distribution with $k-p$

degree of freedom, $k > p$. We might want to minimize not only with the equality constraints given by Expression (23) but also with the inequality constraint given by Expression (25)

$$2nKL(\boldsymbol{\pi}^*(\boldsymbol{\lambda}, \boldsymbol{\beta}), \mathbf{p}_n) \leq \chi_{0.95}^2(k-p).$$

With penalty function methods, we can define a penalty function to handle the inequality constraint as

$$\frac{H}{2}(c^+)^2, \quad c_+ = \max\left(2KL(\boldsymbol{\pi}^*(\boldsymbol{\lambda}, \boldsymbol{\beta}), \mathbf{p}_n) - \frac{\chi_q^2(k-p)}{n}, 0\right),$$

H is again a penalty constant.

This leads to find the global minimizer of a new objective function given by

$$\begin{aligned} & \sum_{i=1}^n \pi_i^*(\boldsymbol{\lambda}, \boldsymbol{\beta}) \ln \pi_i^*(\boldsymbol{\lambda}, \boldsymbol{\beta}) + \frac{K}{2} \left[\left(\sum_{i=1}^n \pi_i^*(\boldsymbol{\lambda}, \boldsymbol{\beta}) [g_1(x_i, \boldsymbol{\beta})] \right)^2 + \dots \right. \\ & \left. + \left(\sum_{i=1}^n \pi_i^*(\boldsymbol{\lambda}, \boldsymbol{\beta}) [g_k(x_i, \boldsymbol{\beta})] \right)^2 \right] + \frac{H}{2} (c^+)^2. \end{aligned} \quad (26)$$

We might also want to repeat the procedures with different starting vectors and identify the global minimizer as the value of the vector which yields the overall smallest value of

$$\sum_{i=1}^n \pi_i^*(\boldsymbol{\lambda}, \boldsymbol{\beta}) \ln \pi_i^*(\boldsymbol{\lambda}, \boldsymbol{\beta}). \quad (27)$$

5. Simulations

5.1. Simulations from the PTS Distribution

The representation of a new distribution created by performing operation on LT of the original distribution often suggests how to simulate from the new distribution if we can simulate from the original distribution. For example, to simulate from the tilted density $f'(x)$ obtained by applying the Esscher operation on $f(x)$, it suffices to simulate from the original density $f(x)$. Since we have $f'(x) = \frac{e^{-\theta x} f(x)}{\kappa(\theta)}$, $\kappa(s)$ is the LT of the density $f(x)$, we have the following

$$\text{inequality } f'(x) \leq c f(x), \quad c = \frac{1}{\kappa(\theta)}.$$

Therefore, if we know how to simulate an observation from the density $f(x)$, we can apply the acceptance and rejection method to obtain simulated observations from $f'(x)$. This is known as the acceptance and rejection method, see Robert and Casella ([41], p.51-57) for example. The constant $\frac{1}{c} = \kappa(\theta)$ is the acceptance probability which is useful for planning the sample size which is obtainable from the simulations. Note that this probability decreases as θ increase making it difficult to obtain a large sample from $f'(x)$ for large values of θ .

The acceptance and rejection method allows a simple way to simulate observations from a positive tempered stable (PTS) as it is easy to simulate from

the positive stable distribution, see Devroye ([45], p.350). Consequently, with a simple algorithm to simulate from the PTS distribution, it allows us to test the performance of the MEEL estimators versus the moment estimators originally proposed by Hougaard ([1], p.392) for the PTS distribution. The moment estimators were proposed since it is difficult to obtain the density function for the PTS which prevents the use of likelihood estimation.

5.2. A Limited Simulation Study

In this section, we illustrate the implementation of the inferences techniques by considering the MEEL estimators versus the moment estimators for the PTS family using simulated samples. The PTS distribution was introduced by Hougaard [1] with Laplace transform given by Expression (6) as

$$L_{\boldsymbol{\beta}}(s) = \exp\left[-\frac{\delta}{\alpha}(\theta + s)^{\alpha} - \theta^{\alpha}\right], \delta, \theta > 0, 0 < \alpha < 1, \boldsymbol{\beta} = (\delta, \alpha, \theta)'$$

Hougaard ([1], p.392) suggested the following moment methods to estimate the parameters. Let c_1, c_2, c_3 be respectively the first, second and third empirical cumulants, *i.e.*, $c_1 = \bar{X}$ is the sample mean and

$$c_j = \frac{\sum_{i=1}^n (X_i - \bar{X})^j}{n}, j = 2, 3.$$

Define $R = \frac{c_2^2}{c_3 c_1}$, if $c_2 > 0$ and define $R = 0$, if $c_2 = 0$. The moment estimators obtained by matching cumulants for the parameters α, θ and δ are given respectively by

$$\tilde{\alpha} = 2 - \frac{1}{1-R}, \tilde{\theta} = \frac{(1-\alpha)c_1}{c_2}, \tilde{\delta} = c_1(\tilde{\theta})^{1-\tilde{\alpha}}.$$

We compare the performance of the moment estimators with the MEEL estimators using the base

$$B = \left\{x - E(x), x^2 - E(x^2), e^{-\tau x} - L_{\boldsymbol{\beta}}(\tau), \dots, e^{-m\tau x} - L_{\boldsymbol{\beta}}(m\tau)\right\},$$

$m = 5, \delta = 0.01$.

We can only have access to laptop computer so the study is limited. The sample size used is approximately with $n = 5000$ and we draw $M = 100$ samples in our simulation. The focus is on the following ranges for the parameters, we fix $\theta = 1, 0.2 < \alpha < 0.8, 1 \leq \delta \leq 10$. Overall the MEEL estimators are much more efficient than the moment estimators for the range of the parameters considered. The moment estimators do not seem to perform well either for all the parameters values selected outside the range. The overall relative efficiency is defined as

$$ARE = \frac{MSE(\hat{\delta}) + MSE(\hat{\alpha}) + MSE(\hat{\theta})}{MSE(\tilde{\delta}) + MSE(\tilde{\alpha}) + MSE(\tilde{\theta})}.$$

The mean square errors (*MSE*) are estimated using simulated samples. The mean square error of an estimator $\hat{\pi}$ for π_0 is defined as

$$MSE(\hat{\pi}) = E(\hat{\pi} - \pi_0)^2.$$

The simulation study is not extensive and more should be done but it does suggest the potential of the MEEL methods.

Some results are summarized using **Table 1** to keep the paper within a reasonable length and they are displayed below to give an idea on the gains on using the MEEL method instead of moment methods.

Based on the theory the MEEL estimators cannot be as efficient as the ML estimators over the entire parameter space since only finite number of elements in the base is used. However, the theory suggest that the methods might still have high efficiencies on subspaces where parameters are subject to inequality bounds. The estimate of overall relative efficiency given by Expression (22) might give some ideas whether the methods are recommended. The following considerations might be useful to assess whether the use of MEEL methods are appropriate for a parametric model and data sets which come from a specific field of applications:

Table 1. Asymptotic relative efficiencies comparisons between MEEL estimators and moment (MM) estimators.

$\alpha = 0.4$							
$\delta \cdot \theta$	0.01	0.02	0.03	0.04	0.06	0.08	0.10
1	0.000	0.006	0.006	0.014	0.030	0.038	0.033
2	0.002	0.001	0.001	0.002	0.027	0.050	0.038
3	0.002	0.040	0.038	0.069	0.069	0.103	0.048
4	0.001	0.067	0.029	0.027	0.091	0.128	0.113
6	0.014	0.072	0.117	0.112	0.084	0.160	0.119
8	0.016	0.004	0.000	0.001	0.005	0.001	0.004
10	0.000	0.000	0.004	0.003	0.000	0.002	0.0018

$\alpha = 0.6$							
$\delta \cdot \theta$	0.01	0.02	0.03	0.04	0.06	0.08	0.10
1	0.008	0.017	0.030	0.028	0.087	0.145	0.195
2	0.009	0.009	0.018	0.011	0.040	0.056	0.081
3	0.010	0.024	0.041	0.040	0.045	0.073	0.087
4	0.028	0.094	0.139	0.053	0.118	0.176	0.187
6	0.068	0.112	0.157	0.182	0.186	0.202	0.399
8	0.065	0.116	0.111	0.164	0.192	0.094	0.075
10	0.007	0.006	0.004	0.006	0.005	0.005	0.003

$$ARE(\text{MEEL vs MM}) = \frac{MSE(\hat{\theta}) + MSE(\hat{\delta}) + MSE(\hat{\alpha})}{MSE(\tilde{\theta}) + MSE(\tilde{\delta}) + MSE(\tilde{\alpha})}.$$

Legend: Tabulated values are estimates of ARE (MEEL vs MM) based on simulated samples from the chosen parameters δ, θ with $\alpha = 0.6$.

1) Define a restricted space based on the fields of applications, also obtain $\widehat{\beta}_s$ and use the estimate for overall relative efficiency to evaluate the loss of efficiency of MEEL methods in a neighborhood of $\widehat{\beta}_s$ which in general should be nested inside the restricted space of interest.

2) For efficiencies of MEEL methods, we shall try to include as many elements in a finite base as possible subject to numerical limitations and try different value for τ which control the spacing of the functions in the basis to see whether there is any improvement on efficiency in a neighborhood of $\widehat{\beta}_s$.

6. Actuarial Applications

Pricing of insurance contracts is one of the main objectives in actuarial sciences. A contract defines a random loss function $g(x)$, X is the individual loss random variable for one unit of time often assumed to be nonnegative and follow a parametric model with distribution function $F_{\beta}(x)$ and LT $L_{\beta}(s)$. The pure premium is the following expectation under the true vector β_0 , i.e.,

$$P = P(\beta_0) = E_{\beta_0}(g(x)).$$

P must be estimated using data and therefore, β_0 needs to be estimated first then subsequently analytical methods or simulation methods can be used to approximate the premium. If MEEL methods are used, the parametric families with closed form LT can be validated by means of goodness-of-fit tests.

For insurance, the stop loss premium is defined as $P = E_{\beta_0}\{(X-d)_+\}$, $(X-d)_+ = \max(X-d, 0)$. The stop loss premium can be expressed using means of distribution functions instead of expectations, see Expression (8) given by Luong ([8], p.543) for analytical methods to evaluate the stop loss premium.

If sampling from the distribution is possible then the pricing of the contracts can also be approximated using simulations based on an estimate of β_0 , it involves drawing sample based on the estimated parameters. For example, it is not difficult to simulate from a compound Poisson distribution despite its complicated density function which can only be expressed in series. Clearly, once the parameters for the compound Poisson distribution are estimated pricing of insurance contracts can be done via simulations.

7. Conclusion

We conclude here that MEEL methods appear to be useful for inferences and have been considered to be active fields of research for the last twenty years in econometrics yet they do not seem to receive much attention in actuarial sciences. When the methods are oriented toward actuarial applications and since LT is widely used in actuarial sciences, it is natural to consider extracting moment conditions from LT. It is shown that MEEL estimation is equivalent to QL estimation based on the best quasi score functions obtained by projecting the true score functions on the linear space spanned by a basis specified by the moment conditions. Based on these considerations, two families of bases are proposed in this paper to generate MEEL methods with the objective to achieve high

efficiencies for actuarial applications. In general the MEEL methods using these bases are more efficient than QL methods based on quadratic estimating functions and methods of moments. With finite bases, in general the MEEL methods can attain near full efficiency on restricted parameter spaces only. MEEL methods can still be very attractive if depending on the fields of applications; we essentially work with these restricted spaces and it is important to measure the loss of efficiency to verify the appropriateness of the methods for the field of applications. The methods can easily be adapted for estimation of continuous distributions with support on the real line encountered in finance by using constraints extracted from model moment generating function instead of LT.

Acknowledgements

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

References

- [1] Hougaard, P. (1986) Survival Models for Heterogeneous Populations Derived from Stable Distributions. *Biometrika*, **73**, 387-396.
<https://doi.org/10.1093/biomet/73.2.387>
- [2] Panjer, H. and Willmot, G.E. (1992) Insurance Risk Models. Society of Actuaries, Chicago.
- [3] Schoutens, W. (2003) Lévy Processes in Finance: Pricing Financial Derivatives. Wiley, New York. <https://doi.org/10.1002/0470870230>
- [4] Kuchler, U. and Tappe, S. (2013) Tempered Stable Distributions and Processes. *Stochastic Processes and Their Applications*, **123**, 4256-4293.
- [5] Gerber, H.U. (1992) From the Generalized Gamma to the Generalized Negative Binomial Distribution. *Insurance, Mathematics and Economics*, **10**, 303-309.
- [6] Abate, J. and Whitt, W. (1996) An Operational Calculus via Laplace Transforms. *Advances in Applied Probability*, **28**, 75-113.
<https://doi.org/10.1017/S0001867800027294>
- [7] Klugman, H., Panjer, H.H. and Willmot, G.E. (2012) Loss Models: From Data to Decisions. Wiley, New York.
- [8] Luong, A. (2016) Cramer-Von Mises Distance Estimation for Some Positive Infinitely Divisible Distributions with Actuarial Applications. *Scandinavian Actuarial*, **2016**, 530-549. <https://doi.org/10.1080/03461238.2014.977817>
- [9] Gerber, H.U. (1992) On the Probability of Ruin for Infinitely Divisible Claim Amount Distributions. *Insurance: Mathematics and Economics*, **11**, 163-166.
- [10] Dufresne, F. and Gerber, H.U. (1993) The Probability of Ruin for the Inverse Gaussian and Related Processes. *Insurance: Mathematics and Economics*, **12**, 9-22.
- [11] Rao, M.M. (1984) Probability Theory with Applications. Academic Press, New York.
- [12] Sato, K.I. (1999) Lévy Processes and Infinitely Divisible Distributions. Cambridge University Press, Cambridge.
- [13] Feller, W. (1971) An Introduction to Probability and Applications. Vol. 2, Wiley, New York.

- [14] Godambe, V.P. and Thompson, M.E. (1989) An Extension of Quasi-Likelihood Estimation. *Journal of Statistical Planning and Inference*, **22**, 137-152.
- [15] Chan, S. and Gosh, M. (1998) Orthogonal Projection and the Geometry of Estimating Functions. *Journal of Statistical Planning and Inference*, **67**, 227-245.
- [16] Crowder, M. (1987) On Linear and Quadratic Estimating Functions. *Biometrika*, **74**, 591-597. <https://doi.org/10.1093/biomet/74.3.591>
- [17] Read, R.R. (1981) Representation of a Certain Covariance Matrices with Application to Asymptotic Efficiency. *Journal of the American Statistical Association*, **76**, 148-154. <https://doi.org/10.1080/01621459.1981.10477621>
- [18] Newey, W.K. and Smith, R.J. (2004) Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators. *Econometrica*, **72**, 219-255. <https://doi.org/10.1111/j.1468-0262.2004.00482.x>
- [19] Smith, R.J. (2007) Efficient Information Theoretical Inference for Conditional Moment Restrictions. *Journal of Econometrics*, **138**, 430-460.
- [20] Chamberlain, G. (1987) Asymptotic Efficiency in Estimation with Conditional Moment Restrictions. *Journal of Econometrics*, **34**, 305-334.
- [21] Luong, A. and Doray, L.G. (2009) Inference for the Stable Laws Based on a Special Quadratic Distance. *Statistical Methodology*, **6**, 147-156.
- [22] Luong, A. and Thompson, M.E. (1987) Minimum Distance Methods Based on Quadratic Distance for Transforms. *Canadian Journal of Statistics*, **15**, 239-251. <https://doi.org/10.2307/3314914>
- [23] Morton, R. (1981) Efficiency of Estimating Equations and the Use of Pivots. *Biometrika*, **68**, 227-233. <https://doi.org/10.1093/biomet/68.1.227>
- [24] Owen, A.B. (1988) Empirical Likelihood Ratio Confidence Intervals for a Single Functional. *Biometrika*, **75**, 237-249. <https://doi.org/10.1093/biomet/75.2.237>
- [25] Qin, J. and Lawless, J. (1994) Empirical Likelihood and General Estimating Equations. *Annals of Statistics*, **22**, 300-325. <https://doi.org/10.1214/aos/1176325370>
- [26] Schennach, S. (2007) Point Estimation with Exponentially Tilted Empirical Likelihood. *Annals of Statistics*, **35**, 634-672. <https://doi.org/10.1214/009053606000001208>
- [27] Imbens, G.W., Spady, R.H. and Johnson, P. (1998) Information Theoretic Approaches to Inference in Moment Condition Models. *Econometrica*, **66**, 333-357. <https://doi.org/10.2307/2998561>
- [28] Owen, A.B. (2001) Empirical Likelihood. Chapman and Hall, New York. <https://doi.org/10.1201/9781420036152>
- [29] Mittelhammer, R.C., Judge, G.G. and Miller, D.J. (2000) Econometrics Foundations. Cambridge University Press, Cambridge.
- [30] Anatolyev, S. and Gospodinov, N. (2011) Methods for Estimation and Inference in Modern Econometrics. CRC Press, New York.
- [31] Garcia, R., Renault, É. and Veredas, D. (2011) Estimation of Stable Laws by Indirect Inference. *Journal of Econometrics*, **161**, 325-337.
- [32] Zakian, V. and Littlewood, R.K. (1973) Numerical Inversion of Laplace Transforms by Weighted Least-Squares Approximations. *The Computer Journal*, **16**, 66-68. <https://doi.org/10.1093/comjnl/16.1.66>
- [33] Brockwell, P.J. and Brown, B.M. (1981) High Efficiency Estimation for the Positive Stable Laws. *Journal of the American Statistical Association*, **75**, 626-631. <https://doi.org/10.1080/01621459.1981.10477695>

- [34] Cressie, N., Davis, A.S, Folks, J.L. and Policello II, G.E. (1981) The Moment Generating Function and Negative Integer Moments. *The American Statistician*, **35**, 148-150.
- [35] Brockwell, P.J. and Brown, B.M. (1978) Expansions for the Positive Stable Laws. *Z. Wahrscheinlichkeitstheorie. Verw. Gebiete*, **45**, 213-224.
<https://doi.org/10.1007/BF00535303>
- [36] Cressie, N. and Read, T. (1984) Multinomial Goodness-of-Fit Test. *Journal of the Royal Statistical Society, Series B*, **46**, 440-464.
- [37] Imbens, G.W. (2002) Generalized Method of Moments and Empirical Likelihood. *Journal of Business and Economic Statistics*, **20**, 493-506.
<https://doi.org/10.1198/073500102288618630>
- [38] Chong, E.K.P. and Zak, S.H. (2013) An Introduction to Optimization. 4th Edition, Wiley, New York.
- [39] Davidson, R. and MacKinnon, J.G. (2004) Econometric Theory and Methods. Oxford University Press, Oxford.
- [40] Andrews, D.W.K. (1997) A Stopping Rule for the Computation of the Generalized Method of Moment Estimators. *Econometrica*, **65**, 913-931.
<https://doi.org/10.2307/2171944>
- [41] Robert, C.P. and Casella, G. (2010) Introducing Monte Carlo Methods with R. Springer, New York. <https://doi.org/10.1007/978-1-4419-1576-4>
- [42] Fang, F. and Osterlee, C.W. (2009) A Novel Pricing Method for European Options Based on Fourier Cosine Series Expansions. *SIAM Journal on Scientific Computing*, **31**, 826-848. <https://doi.org/10.1137/080718061>
- [43] Powers, D.L. (2010) Boundary Value Problems and Partial Differential Equations. Academic Press, New York.
- [44] Bhapkar, V.P. (1972) On a Measure of Efficiency in an Estimating Equation. *Sankhya, Series A*, **34**, 467-472.
- [45] Devroye, L. (1993) A Triptych of Discrete Distributions Related to the Stable Laws. *Statistics and Probability Letters*, **18**, 349-351.



Submit or recommend next manuscript to SCIRP and we will provide best service for you:

Accepting pre-submission inquiries through Email, Facebook, LinkedIn, Twitter, etc.
A wide selection of journals (inclusive of 9 subjects, more than 200 journals)
Providing 24-hour high-quality service
User-friendly online submission system
Fair and swift peer-review system
Efficient typesetting and proofreading procedure
Display of the result of downloads and visits, as well as the number of cited articles
Maximum dissemination of your research work

Submit your manuscript at: <http://papersubmission.scirp.org/>

Or contact ojs@scirp.org