Scientific
Research
Publishing

Robust Continuous Quadratic Distance Estimation Using Quantiles for Fitting Continuous Distributions

Andrew Luong

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

Abstract

Quadratic distance estimation making use of the sample quantile function over a continuous range is introduced. It extends previous methods which are based only on a few sample quantiles and it parallels the continuous GMM method. Asymptotic properties are established for the continuous quadratic distance estimators (CQDE) and the implementation of the methods are discussed. The methods appear to be useful for balancing robustness and efficiency and useful for fitting distribution with model quantile function being simpler than its density function or distribution function.

Keywords

Covariance Kernel, Influence Function, Hilbert Space, Linear Operator, GMM Estimation, Spectral Decomposition

1. Introduction

For estimation in a classical setup, we often assume to have n independent, identically distributed observations $X_1, \cdots, X_n$ from a continuous density $f_{\theta_0}(x)$ which belongs to a parametric family $\{f_\theta\}$, *i.e.*, $f_{\theta_0}(x) \in \{f_\theta\}$ where $\theta = (\theta_1, \cdots, \theta_m)'$, $\theta \in \Omega$ and θ_0 is the true vector of parameters, Ω is assumed to be compact. One of the main objectives of inferences is to be able to estimate θ_0. In an actuarial context, the sample observations might represent losses of a certain type of contracts and an estimate of θ_0 is necessary if we want to make rates or premiums for the type of contract where we have observations.

Maximum likelihood (ML) estimation are density based and often the domain of the density function must not depend on the parameters is one of the regular-

ity conditions so that ML estimators attain the lower bound as given by the information matrix. In many applications, this condition is not met. We can consider the following example which gives the Generalized Pareto distribution (GPD) and draw the attention on the properties of the model quantile function which appears to have nicer properties than the density function and hence motivate us to develop continuous quadratic distance (CQD) estimation using quantiles on a continuum range which generalizes the quadratic distance (QD) methods based on few quantiles as proposed by LaRiccia and Wehrly [1] which can be viewed as based on a discrete range and hence CQD estimation might overcome the arbitrary choice of quantiles of QD as CQD will essentially make use of all the quantiles over the range with $0 < p < 1$.

Example (GPD).

The GP family is a two parameters family with the vector of parameter $\boldsymbol{\theta} = (\lambda, \kappa)'$.

The density, distribution function and quantile function are given respectively by

$$f(x; \lambda, \kappa) = \frac{1}{\lambda}\left(1 - \frac{\kappa x}{\lambda}\right)^{\frac{1}{\kappa}-1}, 1 - \frac{\kappa x}{\lambda} \geq 0, \kappa \neq 0, \lambda > 0 \quad \text{and}$$

$$f(x; \lambda) = \frac{1}{\lambda}e^{-x/\lambda}, x \geq 0, \kappa = 0, \lambda > 0,$$

the distribution function is given by

$$F(x; \lambda, \kappa) = 1 - \left(1 - \frac{\kappa x}{\lambda}\right)^{\frac{1}{\kappa}}, 1 - \frac{\kappa x}{\lambda} \geq 0, \kappa \neq 0, \lambda > 0 \quad \text{and}$$

$$F(x; \lambda) = 1 - e^{-x/\lambda}, x \geq 0, \kappa = 0, \lambda > 0,$$

the quantile function is given by

$$F^{-1}(t; \lambda, \kappa) = \lambda\left(1 - (1-t)^{k}\right)\Big/k, 0 < t < 1, \kappa \neq 0, \lambda > 0 \quad \text{and}$$

$$F^{-1}(t; \lambda) = -\lambda \log(1-t), \kappa = 0, \lambda > 0, 0 < t < 1$$

These functions can be found in Castillo *et al.* [2] (pages 65-66). Among these functions only the domain of the quantile function $F^{-1}(t; \lambda, \kappa)$ does not depend on the parameters and naturally if the model quantile function satisfies some additional conditions such as differentiability, it is natural to develop statistical inference methods using the sample quantile function $F_n^{-1}(t)$ instead of the sample distribution function $F_n(x)$ which are defined respectively as

$$F_n^{-1}(t) = \inf\{x \mid F_n(x) \geq t\} \quad \text{and}$$

$F_n(x) = \frac{1}{n}\sum_{i=1}^{n}\delta_{x_i}$ with δ_{x_i} being the degenerate distribution at x_i is the commonly used sample distribution. The counterpart of $F_n^{-1}(t)$ is the model quantile function $F_\theta^{-1}(t)$, see Serfling [3] (pages 74-80).

Due to the complexity of the density function for the GP model, alternative methods to ML have been developed in the literature for example with the probability weighted moments (PWM) method proposed by Hosking and Wallis [4] which leads to solve moment type of equations to obtain estimators by matching selected empirical moments with their model counterpart. The drawback of the PWM method is the range of the parameters must be restricted for the selected moments to exist, see Hosking and Wallis [4], Kotz and Nadarajah [5] (p 36). The PWM method might not be robust and some robust methods have been proposed by Dupuis [6], Juarez and Schucany [7] for estimation for the GP model.

For estimating parameters of the GPD, the percentiles matching (PM) method for fitting loss distributions as described by Klugman *et al.* [8] (pages 256-257) can also be used. It consists of first selecting two points t_1, t_2 , with $0 < t_1 < t_2 < 1$ as we only have two parameters and solve the following moment type of estimating equations to obtain the estimators, *i.e.*, $\hat{\theta}_{PM}$ is the vector of solutions of

$$F_n^{-1}(t_1) = F_{\theta}^{-1}(t_1) \quad \text{or equivalently,} \quad F_{\theta}\left(F_n^{-1}(t_1)\right) = t_1$$

and

$$F_n^{-1}(t_2) = F_{\theta}^{-1}(t_2) \quad \text{or equivalently,} \quad F_{\theta}\left(F_n^{-1}(t_2)\right) = t_2 .$$

The method is robust but not very efficient as only two points are used here to obtain moment type of equations and there is also arbitrariness on the choice of these two points. Castillo and Hadi [9] have improved this method by first selecting a set of two points, $S = \left\{(t_i, t_j)\right\}$ and obtain a set of corresponding PM estimators and finally define the final estimators according to a rule to select from the set of PM estimators generated by the set $S = \left\{(t_i, t_j)\right\}$. The question on arbitrariness on selecting the set $S = \left\{(t_i, t_j)\right\}$ is still not resolved with this method.

Instead of solving moment type of equations, for parametric estimation in general not necessary for the GPD with the vector of parameters $\theta = (\theta_1, \cdots, \theta_m)'$, LaRiccia and Wehrly [1] proposed to construct quadratic distance based on the discrepancy of $F_n^{-1}(t_i) - F_{\theta}^{-1}(t_i)$ using $k > m$ selected points t_i's with $0 < t_1 < t_2 < \cdots < t_n < 1$, so that we can define the following two vectors z_n and z_{θ} with

$$z_n = \left(F_n^{-1}(t_1), \cdots, F_n^{-1}(t_k)\right)'$$

which is based on the sample and its model counterpart defined as

$$z_{\theta} = \left(F_{\theta}^{-1}(t_1), \cdots, F_{\theta}^{-1}(t_k)\right)'.$$

This leads to a class of quadratic distance of the form

$$(z_n - z_{\theta})' H(\theta)(z_n - z_{\theta}) \tag{1}$$

and the quadratic distance (QD) estimators are found by minimizing the objec-

tive function given by expression (1), $\boldsymbol{H}(\boldsymbol{\theta})$ is a class of symmetric positive definite matrix which might depend on $\boldsymbol{\theta}$. Goodness-of-fit test statistics can also be constructed using expression (1), see Luong and Thompson [10].

By quadratic distance estimation without further specializing it is continuous we mean that it is based on quadratic form as given by expression (1), it also fits into classical minimum distance (CMD) estimation and closely related to Generalized Methods of moment (GMM) and by GMM without further specializing that it is continuous GMM, we mean GMM based on a finite number of moment conditions, see Newey and McFadden [11] (p 212-2128).

Using the asymptotic theory of QD estimation or CMD estimation, it is well known that by letting $\boldsymbol{H}(\boldsymbol{\theta})$ to be the inverse of the asymptotic covariance matrix of $\sqrt{n}\left(z_n - z_{\boldsymbol{\theta}_0}\right)'$ under $\boldsymbol{\theta}_0$, we can obtain estimators which are the most efficient within the class being considered as given by expression (1), so we can let

$$\boldsymbol{H}^*\left(\boldsymbol{\theta}_0\right) = \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}_0\right) \text{ and}$$

$\boldsymbol{\Sigma}\left(\boldsymbol{\theta}_0\right)$ is the asymptotic covariance matrix of $\sqrt{n}\left(z_n - z_{\boldsymbol{\theta}_0}\right)'$.

In fact, it has been shown that it suffices to use a consistent estimate for $\boldsymbol{\Sigma}\left(\boldsymbol{\theta}_0\right)$ to obtain asymptotic equivalent estimators. For example, first we obtain a preliminary consistent estimate $\boldsymbol{\theta}^{(0)}$ and if we can construct a consistent estimate

$$\boldsymbol{\Sigma}_n\left(\boldsymbol{\theta}^{(0)}\right) \text{ for } \boldsymbol{\Sigma}\left(\boldsymbol{\theta}_0\right), \text{ i.e., } \boldsymbol{\Sigma}_n\left(\boldsymbol{\theta}^{(0)}\right) \xrightarrow{p} \boldsymbol{\Sigma}\left(\boldsymbol{\theta}_0\right)$$

then we can construct a consistent estimate which is given $\boldsymbol{\Sigma}_n^{-1}\left(\boldsymbol{\theta}^{(0)}\right)$ for $\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}_0\right)$ as in general,

$$\boldsymbol{\Sigma}_n^{-1}\left(\boldsymbol{\theta}^{(0)}\right) \xrightarrow{p} \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}_0\right).$$

In practice, for QD estimation we let $\boldsymbol{H} = \boldsymbol{\Sigma}_n^{-1}\left(\boldsymbol{\theta}^{(0)}\right)$ to obtain QD estimators and the asymptotic efficiency is identical as QD estimators based on $\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}_0\right)$ and it is simpler to obtain them numerically.

For GMM estimation, it is quite straightforward to construct $\boldsymbol{\Sigma}_n^{-1}\left(\boldsymbol{\theta}^{(0)}\right)$, see expression (4.2) given by Newey and McFadden [11] (p2155). The authors also pointed out that this might not be the case for CMD estimation or QD estimation. This is a point that we shall address when generalizing the quadratic distance methods using a finite number of quantiles to method using quantile function over a continuous range which we shall refer to as continuous quadratic distances (CQD); we shall use an approach based on the influence functions of the sample quantiles to estimate the optimum kernel which is the analogous of the use of $\boldsymbol{\Sigma}_n^{-1}\left(\boldsymbol{\theta}^{(0)}\right)$ to estimate $\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}_0\right)$ for the continuous set-up.

Continuous GMM theory makes use of Hilbert space linear operator theory and have been developed in details by Carrasco and Florens [12] and as mentioned it is closely related to the theory for continuous QD, we shall make use of

their results to establish consistency and asymptotic normality of continuous quadratic distance estimators and since the paper aims at providing results for practitioners for their applied works, the presentation will emphasize methodologies with less technicalities so that it might be more suitable for applied researchers for their works. First, we shall briefly outline how to form the quadratic distance to obtain the CQD estimators and postpone the details for later sections of the paper.

CQD estimators can be viewed as estimators based on minimizing a continuous quadratic form as given by

$$Q_n(\boldsymbol{\theta}) = \int_a^b \int_a^b \left(F_n^{-1}(s) - F_{\boldsymbol{\theta}}^{-1}(s) \right) \left(F_n^{-1}(t) - F_{\boldsymbol{\theta}}^{-1}(t) \right) k^o(s,t) \, ds \, dt \tag{2}$$

with:

1) $k^o(s,t)$ is an optimum symmetric positive definite kernel assumed to be fully specified.

2) a and b are chosen values with a being close to 0 and b close to 1 and $0 < a < b < 1$.

In practice, we work with an asymptotic equivalent objective function $Q_n^a(\boldsymbol{\theta})$ instead of $Q_n(\boldsymbol{\theta})$ where $k^o(s,t)$ is estimated by a degenerate kernel $k_n^o(s,t)$, *i.e.*,

$$Q_n^a(\boldsymbol{\theta}) = \int_a^b \int_a^b \left(F_n^{-1}(s) - F_{\boldsymbol{\theta}}^{-1}(s) \right) \left(F_n^{-1}(t) - F_{\boldsymbol{\theta}}^{-1}(t) \right) k_n^o(s,t) \, ds \, dt . \tag{3}$$

Since the kernel $k_n(s,t)$ is degenerate and in our case, we can find explicitly n eigenvalues $\lambda_j^{(n)}, j = 1, \cdots, n$ with corresponding closed form eigenfunctions $\phi_j^{(n)}(t), j = 1, \cdots, n$. These eigenfunctions can be computed explicitly.

As in the spectral decomposition of a symmetric positive defined matrix for the Euclidean space, spectral decomposition in Hilbert space allows the kernel $k_n^o(s,t)$ to be represented as $k_n^o(s,t) = \sum_{j=1}^n \lambda_j^{(n)} \phi_j^{(n)}(s) \phi_j^{(n)}(t)$, and using this representation, we can express $Q_n^a(\boldsymbol{\theta})$ as a sum of n components, *i.e.*,

$$Q_n^a(\boldsymbol{\theta}) = \sum_{i=1}^n \lambda_j^{(n)} \left(\int_a^b \left(\phi_j^{(n)}(t) \right) \left(F_n^{-1}(t) - F_{\boldsymbol{\theta}}^{-1}(t) \right) dt \right)^2 \tag{4}$$

which is similar to the expression used to obtain continuous GMM estimators as given by Carrasco and Florens [12] (page 799).

Spectral decompositions in functional space have been used in the literature, see Feuerverger and McDunnough [13] (page 312), Durbin [14] (page 292-294). Furthermore, if $\lambda_j^{(n)}, j = 1, \cdots, n$ are not stable, they can be replaced by suitable defined $\lambda_j^{(\alpha_n)}, j = 1, \cdots, n$ without affecting the asymptotic theory of the CQD estimators. In practice, we work with

$$Q_n^{\alpha_n}(\boldsymbol{\theta}) = \sum_{i=1}^n \lambda_j^{(\alpha_n)} \left(\int_a^b \left(\phi_j^{(n)}(t) \right) \left(F_n^{-1}(t) - F_{\boldsymbol{\theta}}^{-1}(t) \right) dt \right)^2 \tag{5}$$

to obtain CQD estimators. Unless otherwise stated, by CQD estimators we mean estimators using the objective function of the form as defined by expression (5).

Carrasco and Florens [6] (page 799) developed perturbation technique, a

technique to obtain $\lambda_j^{(\alpha_n)}, j = 1, \cdots, n$ from the eigenvalues $\lambda_j^{(n)}, j = 1, \cdots, n$. The perturbation technique will also be used for constructing a degenerate optimum kernel for CQD estimation.

The objectives of the paper are to develop CQD estimation based on quantiles with the aims to have estimators which are robust in the sense of bounded influence functions and have good efficiencies. For technicalities, we refer to the paper by Carrasco and Florens [12] who have introduced continuous GMM estimation.

The paper is organized as follows. Section 2 gives some preliminary results such as statistical functional and its influence function from which the sample quantiles can be viewed as robust statistics with bounded influence functions. CQD estimation using quantiles will inherit the same robustness property. Some of the standard notions for the study of kernel functions will also be reviewed. By linking a kernel to a linear operator in the Hilbert space of functions which are square integrable over the range (a, b) with an inner product, it allows a norm $\| . \|$ to be introduced. Also, the notion of self adjoint linear operator which can be viewed as analogous to a symmetric matrix in Euclidean space is also introduced in Section 2. Section 3 gives asymptotic properties of the CQD estimators based on an estimate optimum kernel. An estimate of the covariance matrix is also given in Section 3.

Finally, we shall mention that simulation studies are not discussed in this paper as numerical quadrature methods are involved for evaluating the integrals over the range $[a, b]$ for computing the objective function, we prefer to gather numerical aspects and simulation aspects together for further works and include these type of results in a separate paper leaving this paper focusing only on the methodologies.

2. Some Preliminaries

In this section we shall review the notion of statistical functional and its influence function and view a sample quantile as a statistical functional. Using its influence function, we can see that the sample quantile is a robust statistic and using the influence functions of two sample quantiles, we can also obtain the asymptotic covariance of the two sample quantiles.

2.1. Statistical Functional and Its Influence Function

Often, a statistic can be represented as a functional of the sample distribution F_n which we can denote by $T(F_n)$. For example, the sth-sample quantile is defined as $F_n^{-1}(s)$. Associated with $T(F_n)$, there is its influence function which is a weak functional directional derivative at $F = F_{\theta_0}$ in the direction of $(\delta_x - F)$, δ_x is the degenerate distribution at x. More specifically, the influence function of $T(F_n)$ as a function of x is defined as

$$IC_T(x) = T_F'(\delta_x - F) = \lim_{\epsilon \to 0+} \frac{T(F + \epsilon(\delta_x - F)) - T(F)}{\epsilon},$$

T_F' is a linear function in the functional space. It is not difficult to see that we can also compute $IC_T(x)$ using the usual derivative

$$IC_T(x) = \frac{dT(F + \epsilon(\delta_x - F))}{d\epsilon}\Bigg|_{\epsilon=0}.$$

Furthermore, since a Taylor type of approximation in a functional space can be used, we then have the following approximation expressed with a remainder term r_n

$$T(F_n) = T(F) + T_F'(F_n - F) + r_n$$

or equivalently using $F_n = \frac{1}{n}\sum_{i=1}^{n}\delta_{x_i}$,

$$T(F_n) = T(F) + T_F'\left(\frac{1}{n}\sum_{i=1}^{n}(\delta_{x_i} - F)\right) + r_n \quad \text{and using} \quad T_F' \text{ is linear,}$$

$$T(F_n) = T(F) + \frac{1}{n}\sum_{i=1}^{n}T_F'(\delta_{x_i} - F) + r_n,$$

$$T(F_n) - T(F) = \frac{1}{n} + \sum_{i=1}^{n}IC_T(x_i) + r_n.$$

If $IC_T(x)$ as a function of x is bounded, the statistics is robust and the remainder is $o_p(n^{-1/2})$ with $o_p(n^{-1/2})$ being a term which converges to 0 in probability faster than $n^{-1/2} \to 0$ as $n \to \infty$.

Therefore, if we want to find the asymptotic variance of $Var(\sqrt{n}(T(F_n) - T(F)))$, it is given by the variance of $IC_T(x)$ as

$$\sqrt{n}(T(F_n) - T(F)) \xrightarrow{L} N(0, \sigma_T^2), \sigma_T^2 = Var(IC_T(x))$$

The influence function of the sth-sample quantile $F_n^{-1}(s)$ can be obtained and it is given by

$$IC_{Q_s}(x) = -\frac{I[x \le F^{-1}(s)] - s}{f(F^{-1}(s))}, f = f_{\theta_0}, F = F_{\theta_0}, 0 < s < 1 \tag{6}$$

from which we can obtain the asymptotic variance of

$$\sqrt{n}(F_n^{-1}(s) - F^{-1}(s)) = Var(IC_{Q_s}(x)) = \frac{s(1-s)}{(f(F^{-1}(s)))^2},$$

See Serfing [3] (page236), Hogg *et al.* [15] (page 593). Also, using the influence function representation for the sth-sample quantile $F_n^{-1}(s)$ and the corresponding one for the tth-sample quantile $F_n^{-1}(t)$, it can be shown that the asymptotic covariance of the following sample quantiles $F_n^{-1}(s)$ and $F_n^{-1}(t)$ is given by

$$Cov\left(\sqrt{n}(F_n^{-1}(s) - F^{-1}(s)), \sqrt{n}(F_n^{-1}(t) - F^{-1}(t))\right)$$
$$= \frac{\min(s,t) - st}{f(F^{-1}(s))f(F^{-1}(t))}, 0 < s < 1, 0 < t < 1$$

see LaRiccia and Wehrly [1] (page 743).

If we define the covariance kernel as

$$k(s,t) = \frac{\min(s,t) - st}{f(F^{-1}(s))f(F^{-1}(t))}, f = f_{\theta_0}, F = F_{\theta_0}, 0 < s < 1 \qquad (7)$$

then associated to this kernel there is a linear operator $\mathbf{K}$ in a functional space which can be defined as follows, let a function $f(t)$ which belongs to the functional space being considered, K is defined as

$$\mathbf{K}f = \mathbf{K}(f(t)) = \int_a^b k(s,t) f(s) \mathrm{d}s = g(t).$$

We can see that for a suitable functional space, it is natural to consider the Hilbert space of functions which are square integrable so that a norm and linear operators can be defined in this space. This will facilitate the studies of kernels which are function of (s,t). The necessary notions are introduced in the following section.

2.2. Linear Operators Associated with Kernels in a Hilbert Space

The functional space that we are interested is the space of integrable function with the range $[a,b], 0 < a < b < 1$ and it is natural to introduce an inner product $(f,g) = \int_a^b f(t)g(t)\mathrm{d}t$ and therefore, a norm $\|.\|$ can be defined as

$$\|f\| = \sqrt{(f,f)} = \left(\int_a^b f(t)g(t)\mathrm{d}t\right)^{1/2}.$$

For a Euclidean space, the composition of two linear operators $\mathbf{A}$ and $\mathbf{B}$ where $\mathbf{A}$ and $\mathbf{B}$ are matrices produces a matrix $\mathbf{C}$ with $\mathbf{C} = \mathbf{A} \cdot \mathbf{B}$. For linear operators in the Hilbert space the composition of the linear operators $\mathbf{A}$ and $\mathbf{B}$ is a linear operator $\mathbf{C} = \mathbf{A} \cdot \mathbf{B}$ with its kernel $k_C(s,t)$ and

$$\mathbf{C}f = \mathbf{C}(f(t)) = \int_a^b k_C(s,t) f(s) \mathrm{d}s = A(B(f(t))).$$

Just as a matrix $\mathbf{A}$ has its transpose $\mathbf{A}^*$ matrix and if $\mathbf{A}$ is symmetric then $\mathbf{A} = \mathbf{A}^*$, these notions can be extended to a functional space as a linear operator $\mathbf{A}$ has its adjoint $\mathbf{A}^*$ and if the kernel defining $\mathbf{A}$ is symmetric then $\mathbf{A} = \mathbf{A}^*$, $\mathbf{A}$ is called self adjoint.

More precisely, given A, A^* is found using the following equality, see Definition 6 given by Carrasco and Florens [12] (page 823),

$$(Af, g) = (f, A^* g).$$

Furthermore,

$$\text{if } A = A^* \text{ then } (Af, g) = (f, Ag).$$

In this paper we focus on positive definite symmetric kernel $k(s,t)$ which can be viewed as the covariance of (Y_s, Y_t) for some stochastic process $\{Y_t\}$; therefore, the objective function is of the type $\int_a^b \int_a^b f(s) f(t) k(s,t) \mathrm{d}s \mathrm{d}t$ is always positive unless $f(s) = 0$ then $\int_a^b \int_a^b f(s) f(t) k(s,t) \mathrm{d}s \mathrm{d}t = 0$, see Luenberger [16] (page 152) for this notion.

Unless otherwise stated, we work with linear operators associated with positive definite symmetric kernels. For the Euclidean space if the covariance matrix K is invertible with the inverse given by K^{-1} assumed to exist after regularizations then there are symmetric positive definite symmetric matrices denoted by $K^{1/2}$ and $K^{-1/2}$ so that

$$K = K^{1/2}K^{1/2}, K^{-1} = K^{-1/2}K^{-1/2} \tag{8}$$

see Hogg *et al.* [15] (pages 179-180) for square root of a symmetric positive definite matrices and they can be computed using the technique of spectral decomposition of matrices.

If K is linear operator with covariance kernel $k(s,t)$, the analogous properties given by expression (8) continues to hold but unlike matrices where closed forms for the matrices can be found, one might not be able to display the kernel of K^{-1} or $K^{-1/2}$ explicitly as no closed form expressions are available despite that both K^{-1} and $K^{-1/2}$ exist subject to some technical regularizations as discussed in section 4 by Carrasco and Florens [12] (pages 506-510).

For our purpose, we shall focus on a linear operator K with its kernel defined by Equation (7) for the rest of the paper. Since K and K^{-1} are related and if we can construct an estimator for K, we can construct an estimator for K^{-1} and the construction of these estimators will be discussed in the next sub-section.

2.3. Estimation of *K* and *K*⁻¹

The methods used to construct an estimator for K follows from the techniques proposed by Carrasco and Florens [12]. The steps are given below:

1) We need a preliminary consistent estimate $\theta^{(0)}$ for θ_0, for our case we can minimize the following simple objective function to obtain $\theta^{(0)}$,

$$\int_a^b \left(F_n^{-1}(t) - F^{-1}(t)\right)^2 dt.$$

2) Use $\theta^{(0)}$ and the sample of observations to construct a degenerate kernel $k_n(s,t)$ which has the form

$$k_n(s,t) = \frac{1}{n}\sum_{i=1}^n h_s(x_i)h_t(x_i), \ h_s(x_i) \ \text{and} \ h_t(x_i) \ \text{depends on} \ \theta^{(0)}.$$

For our set-up, *i.e.*, CQD estimation, we should use the influence function of the sample quantiles as given by expression (6) to specify $h_s(x) = IC_{Q_s}(x)$, $h_t(x) = IC_{Q_t}(x)$.

The notion of influence function was not mentioned in Carrasco and Florens [12].

3) Since $k_n(s,t)$ is a degenerate kernel it only has n eigenvalues $\mu_j^{(n)}, j = 1, \cdots, n$ with the corresponding eigenvectors $\phi_j^{(n)}(t), j = 1, \cdots, n$, these eigenvectors have closed forms. The procedures to find these eigenvalues and eigenvectors have been given Carrasco and Florens [12] (page 805) and will be summarized in the next paragraphs. Let $\mu^{(n)}$ be one of these n eigenvalues with

its corresponding eigenvector $\phi^{(n)}(t)$, $\phi^{(n)}(t)$ needs to satisfy

$$\boldsymbol{K}\phi^{(n)}(t) = \mu^{(n)}\phi^{(n)}(t), \text{ i.e., } \int_a^b k_n(s,t)\phi^{(n)}(s) = \phi^{(n)}(t).$$

4) Use the spectral decomposition to express $k_n(s,t)$ using its eigenvalues and eigenfunctions, *i.e.*,

$$k_n(s,t) = \sum_{j=1}^n \mu_j^{(n)}\phi_j^{(n)}(s)\phi_j^{(n)}(t).$$

The above expression is similar to the representation of a positive definite matrix $\boldsymbol{A}$ using the spectral decomposition and from which we only need to adjust the eigenvalues if we want to find A^{-1}, the inverse of the matrix $\boldsymbol{A}$ or the matrices $A^{1/2}$ and $A^{-1/2}$.

We can proceed as follows in order to find $\mu_j^{(n)}, j = 1, \cdots, n$ and $\phi_j^{(n)}(t), j = 1, \cdots, n$, following Carraco and Florens [12] (page 805). First we form a matrix $\boldsymbol{C} = \left(\dfrac{1}{n}c_{ij}\right), i = 1, \cdots, n; j = 1, \cdots, n$ with

$$c_{ij} = \int_a^b h_s(x_i)h_s(x_j)\mathrm{d}s.$$

It turns out that $\mu_j^{(n)}$ for each *j* is also an eigenvalue of the matrix $\boldsymbol{C}$ and its eigenvectors is $\boldsymbol{\beta}^j$ with respect to the matrix $\boldsymbol{C}$ with

$$\boldsymbol{\beta}^j = \left(\beta_1^j, \cdots, \beta_n^j\right)' \text{ and } \boldsymbol{C}\boldsymbol{\beta}^j = \mu_j^{(n)}\boldsymbol{\beta}^j, j = 1, \cdots, n.$$

The eigenfunction can be expressed as $\phi_j^{(n)}(t) = \dfrac{1}{n}\sum_{i=1}^n \beta_i^j h_t(x_i)$ and they can be computed as statistical packages often offer routines to compute eigenvalues and eigenvectors for a given matrix.

For numerical evaluations of $c_{ij}, i = 1, \cdots, n; j = 1, \cdots, n$ a numerical quadrature procedure is needed to compute the integrals over a range $[a,b]$.

Now we turn into attention of constructing $\boldsymbol{K}_n^{-1}$ and $\boldsymbol{K}_n^{-1/2}$ to estimate $\boldsymbol{K}^{-1}$ and $\boldsymbol{K}^{-1/2}$.

It appears then the kernel of $\boldsymbol{K}_n^{-1}$ can be defined as

$k_n^o(s,t) = \sum_{j=1}^n \dfrac{1}{\mu_j^{(n)}}\phi_j^{(n)}(s)\phi_j^{(n)}(t)$, see Definition 3 given by Carrasco and Florens [12] (page 807). Howewer, Carrasco and Florens [12] (page 799) have shown that $\dfrac{1}{\mu_j^{(n)}}$ will create numerical instabilities and need to be regularized and instead of $\dfrac{1}{\mu_j^{(n)}}$, we need to replace it by $\lambda_j^{(\alpha_n)} = \dfrac{\mu_j^{(n)}}{\left(\mu_j^{(n)}\right)^2 + \alpha_n}, j = 1, \cdots, n$,

and since $\mu_j^{(n)}, j = 1, \cdots, n$ are positive in probability, we can also let

$\lambda_j^{(\alpha_n)} = \dfrac{\left|\mu_j^{(n)}\right|}{\left(\mu_j^{(n)}\right)^2 + \alpha_n}, j = 1, \cdots, n$ and these expressions might be easier to handle numerically.

Now we can define define $k_n^o(s,t) = \sum_{j=1}^n \lambda_j^{(\alpha_n)}\phi_j^{(n)}(s)\phi_j^{(n)}(t)$ to be the kernel

of $K_{\alpha_n}^{-1}$, $K_{\alpha_n}^{-1}$ will be a valid estimator for K^{-1} providing that the sequence $\alpha_n \to 0$ and $n\alpha_n^{3/2} \to \infty$ using their Theorem 7 on page 810.

For example, if we let $\alpha_n = \dfrac{d}{n^{0.65}}$ for some d chosen to be positive then the requirements for the sequence α_n are met.

This also means that the kernel for $K_{\alpha_n}^{-1/2}$ can be defined as

$$\sum_{j=1}^{n} \sqrt{\lambda_j^{(\alpha_n)}} \phi_j^{(n)}(s) \phi_j^{(n)}(t) \tag{9}$$

and again $K_{\alpha_n}^{-1/2}$ is a valid estimator for $K^{-1/2}$.

This also means that K^{-1} and $K^{-1/2}$ can be replaced by $K_{\alpha_n}^{-1}$ and $K_{\alpha_n}^{-1/2}$ whenever they appear in expressions or equations used to derive asymptotic properties for the CQD estimators based on their Theorem 7.

In Section 3 we shall turn our attention to asymptotic properties of CQD estimators using the objective function $Q_n^{\alpha_n}(\boldsymbol{\theta})$ an using the norm $\|.\|$, it can also be expressed neatly as

$$Q_n^{\alpha_n}(\boldsymbol{\theta}) = \left\| K_{\alpha_n}^{-1/2} u_t(\boldsymbol{\theta}) \right\|^2 \quad \text{with} \quad u_t(\boldsymbol{\theta}) = F_n^{-1}(t) - F_{\boldsymbol{\theta}}^{-1}(t) \quad \text{and}$$

$K_{\alpha_n}^{-1/2}$ is the linear operator as defined by expression (9).

For consistency, we shall make use of the basic consistency Theorem, *i.e.*, Theorem 2.1 as given by Newey and McFadden [12] (page 2121). For establishing asymptotic normality for the CQD estimators, the procedures are similar to those used for establishing asymptotic normality of continuous GMM estimators as given by Theorem 8 given by Carrasco and Florens [12] (page 811, page 825).

3. Asymptotic Properties

3.1. Consistency

Assuming $\boldsymbol{\theta} \in \Omega$ and Ω is compact and observe that

$$Q_n^{\alpha_n}(\boldsymbol{\theta}) = \int_a^b \int_a^b \left(F_n^{-1}(s) - F_{\boldsymbol{\theta}}^{-1}(s) \right)\left(F_n^{-1}(t) - F_{\boldsymbol{\theta}}^{-1}(t) \right) k_{\alpha_n}^o(s,t)\, ds\, dt . \tag{10}$$

Now if we assume that the integrand can be dominated by a function $g(s,t)$ which does not depend $\boldsymbol{\theta}$ and furthermore $\int_a^b \int_a^b g(s,t)\, ds\, dt < \infty$ then we have uniform convergence in probability, *i.e.*, $Q_n^{\alpha_n}(\boldsymbol{\theta}) \xrightarrow{p} Q_0(\boldsymbol{\theta})$ uniformly with

$$Q_0(\boldsymbol{\theta}) = \int_a^b \int_a^b \left(F_{\boldsymbol{\theta}_0}^{-1}(s) - F_{\boldsymbol{\theta}}^{-1}(s) \right)\left(F_{\boldsymbol{\theta}_0}^{-1}(t) - F_{\boldsymbol{\theta}}^{-1}(t) \right) k^o(s,t)\, ds\, dt ,$$

$k^o(s,t)$ is the optimum symmetric positive definite kernel of K^{-1}. Therefore, $Q_0(\boldsymbol{\theta})$ is uniquely minimized at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, this implies consistency of the CQD estimators given by the vector $\hat{\boldsymbol{\theta}}$ using the basic consistency Theorem. Therefore, $\hat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}_0$, the symbol $\xrightarrow{p}$ denotes convergence in probability. We implicitly assume that the conditions $\alpha_n \to 0$ and $n\alpha_n^{3/2} \to \infty$ are met.

3.2. Asymptotic Normality

The basic assumption used to establish asymptotic normality for the CQD esti-

mators is the model quantile function is twice differentiable which allows a standard Taylor expansion the estimating equations.

Assuming the first derivative vector $\dfrac{\partial F_{\boldsymbol{\theta}}^{-1}(t)}{\partial \boldsymbol{\theta}}$ and the second derivative matrix

$$\frac{\partial^2 F_{\boldsymbol{\theta}}^{-1}(t)}{\partial \boldsymbol{\theta}'\partial \boldsymbol{\theta}} = \left(\frac{\partial^2 F_{\boldsymbol{\theta}}^{-1}(t)}{\partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}_i}\right), i = 1,\cdots,m; j = 1,\cdots,m \quad \text{exist.}$$

Before considering the Taylor expansion, we also need the following notation and the notion of a random element with zero mean and covariance given by the kernel of the associated linear operator K, *i.e.*, $Y_t \sim N(0, K)$, see Remark 2 as given by Carrasco and Florens [12] (page 803). Note that if we let $u_t(\boldsymbol{\theta}) = F_n^{-1}(t) - F_{\boldsymbol{\theta}}^{-1}(t)$, using the Mean value Theorem, we then have

$$u_t(\boldsymbol{\theta}) = u_t(\boldsymbol{\theta}_0) - \left.\frac{\partial F_{\boldsymbol{\theta}}^{-1}(t)}{\partial \boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}=\bar{\boldsymbol{\theta}}} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0),$$

$\bar{\boldsymbol{\theta}}$ lies in the segment joining $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_0$. Now we have $\hat{\boldsymbol{\theta}}$ which satisfies $\dfrac{\partial Q_n^{\alpha_n}(\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}} = 0$ which is also given by

$$\int_a^b \int_a^b \frac{\partial u_s(\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}} u_t(\hat{\boldsymbol{\theta}}) k_{\alpha_n}^o(s,t)\,\mathrm{d}s\mathrm{d}t = 0 \tag{11}$$

as $k_{\alpha_n}^o(s,t)$ is symmetric. Using inner product and Hilbert space as in the proofs of Theorem 2 by Carrasco and Florens [12] (page 825), expression (11) can be expressed as

$$\left(\boldsymbol{K}_{\alpha_n}^{-1/2} \frac{\partial u_s(\hat{\boldsymbol{\theta}})}{\partial \theta}, \boldsymbol{K}_{\alpha_n}^{-1/2} u_t(\hat{\boldsymbol{\theta}})\right) = 0 . \tag{12}$$

Using expression (12), we then have

$$\left(\boldsymbol{K}_{\alpha_n}^{-1/2} \frac{\partial u_s(\hat{\boldsymbol{\theta}})}{\partial \theta}, \boldsymbol{K}_{\alpha_n}^{-1/2} \left(u_t(\boldsymbol{\theta}_0) - \left.\frac{\partial F_{\boldsymbol{\theta}}^{-1}(t)}{\partial \boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}=\bar{\boldsymbol{\theta}}} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)\right)\right) = 0 .$$

Now using $\boldsymbol{K}_{\alpha_n}^{-1/2}$ is a linear operator, $\dfrac{\partial u_s(\boldsymbol{\theta})}{\partial \theta} = -\dfrac{\partial F_{\boldsymbol{\theta}}^{-1}(s)}{\partial \theta}$, rearranging the terms gives the following equality in distribution

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) =^d \left(\frac{\partial F_{\boldsymbol{\theta}_0}^{-1}(s)}{\partial \boldsymbol{\theta}} \boldsymbol{K}_{\alpha_n}^{-1} \frac{\partial F_{\boldsymbol{\theta}_0}^{-1}(t)}{\partial \boldsymbol{\theta}'}\right)^{-1} \left(\frac{\partial F_{\boldsymbol{\theta}_0}^{-1}(s)}{\partial \boldsymbol{\theta}} \boldsymbol{K}_{\alpha_n}^{-1} \sqrt{n} u_t(\boldsymbol{\theta}_0)\right).$$

Note that $\sqrt{n} u_t(\boldsymbol{\theta}_0) =^d Y_t$ and the symbol $=^d$ denotes equality in distribution.

Let $\mathcal{J}(\boldsymbol{\theta}_0) = \left(\dfrac{\partial F_{\boldsymbol{\theta}_0}^{-1}(s)}{\partial \boldsymbol{\theta}} K^{-1} \dfrac{\partial F_{\boldsymbol{\theta}_0}^{-1}(t)}{\partial \boldsymbol{\theta}'}\right)$

And $Z = \left(\dfrac{\partial F_{\boldsymbol{\theta}_0}^{-1}(s)}{\partial \boldsymbol{\theta}} \boldsymbol{K}^{-1} \sqrt{n} u_t(\boldsymbol{\theta}_0) \right)$ then it is easy to see that

$$\boldsymbol{Z} =^d N\left(0, \mathcal{J}(\boldsymbol{\theta}_0)\right),$$

so that

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) \xrightarrow{L} N\left(0, \left[\mathcal{J}(\boldsymbol{\theta}_0)\right]^{-1}\right) \tag{13}$$

with the symbol $\xrightarrow{L}$ denotes convergence in law or in distribution.

The matrix $\mathcal{J}(\boldsymbol{\theta}_0)$ plays the same role as the information matrix for maximum likelihood (ML) estimation. Clearly, $\mathcal{J}(\theta_0)$ needs to be estimated, an estimate is given as

$$\mathcal{J}_n\left(\hat{\boldsymbol{\theta}}\right) = \left(\dfrac{\partial F_{\hat{\boldsymbol{\theta}}}^{-1}(s)}{\partial \boldsymbol{\theta}} \boldsymbol{K}_{\alpha_n}^{-1} \dfrac{\partial F_{\hat{\boldsymbol{\theta}}}^{-1}(t)}{\partial \boldsymbol{\theta}'} \right), \tag{14}$$

using the spectral decomposition technique, the (i, j) element of $\mathcal{J}_n\left(\hat{\boldsymbol{\theta}}\right)$ can be expressed as

$$\sum_{k=1}^{n} \lambda_k^{(\alpha_n)} \left(\int_a^b \phi_k^{(n)}(s) \dfrac{\partial F_{\hat{\boldsymbol{\theta}}}^{-1}(s)}{\partial \boldsymbol{\theta}_i} \mathrm{d}s \right)\left(\int_a^b \phi_k^{(n)}(t) \dfrac{\partial F_{\hat{\boldsymbol{\theta}}}^{-1}(t)}{\partial \boldsymbol{\theta}_j} \mathrm{d}t \right), i = 1, \cdots, m; j = 1, \cdots, m \tag{15}$$

4. Summary and Conclusion

The proposed method is similar to the continuous GMM method with the estimators obtained using sample distribution function obtained by minimizing

$$\int_0^T \int_0^T \left(F_n(s) - F_{\boldsymbol{\theta}}(s)\right)\left(F_n(t) - F_{\boldsymbol{\theta}}(t)\right) w_{\alpha_n}^0(s,t) \mathrm{d}s \mathrm{d}t \quad \text{with} \quad w_{\alpha_n}^0(s,t)$$

being an optimum kernel but using a sample distribution function F_n instead of the sample quantile function as studied by Carrasco and Florens [12] (page 816) for nonnegative continuous distributions. The kernel $w_{\alpha_n}^0(s,t)$ is constructed with the use of $h_t(x) = I[x \le t] - F_{\theta^{(o)}}(t)$, $I[.]$ being the usual indicator function.

The authors also showed that by letting $T \to \infty$, the continuous GMM estimators are as efficient as ML estimators.

For robustness sake for continuous GMM estimation we might want to let T be finite and the lower bound be $a > 0$ so that the optimum kernel $w_{\alpha_n}^0(s,t)$ remains bounded for the regions of the double integrals used to define the continuous GMM objective function. This can be viewed as equivalent to choose $a > 0$ and $b < 1$ for the integrals of the objective function for CQD estimation. For robustness sake, it appears simpler to work with the domain (a, b) instead of (a, T) as numerical quadrature methods applied over the range (a, b) might be simpler to implement. We conjecture that CQD estimators can also be fully efficient just as the continuous GMM estimators as defined above despite a proof is still lacking for the time being by letting $a \to 0$, $b \to 1$. More numerical and more simulation studies are needed but we hope that based on the presentation of this paper the proposed method is implementable and its asymptotic proper-

ties useful so that applied researchers might want to consider to use them for their works especially for fitting models where the model quantile function is simpler to handle than its model distribution or density function and especially when there is a need for robust estimation with the data.

Acknowledgements

Conflicts of Interest

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

References

[1] LaRiccia, V.N. and Wehrly, T.E. (1985) Asymptotic Properties of a Family of Minimum Quantile Distance Estimators. *Journal of the American Statistical Association*, **80**, 742-747. https://doi.org/10.1080/01621459.1985.10478178

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

[3] Serfling, R.J. (1980) Approximation Theorems of Mathematical Statistics. Wiley, New York. https://doi.org/10.1002/9780470316481

[4] Hosking, J.R.M. and Wallis, J.R. (1987) Parameter and Quantile Estimation for the Generalized Pareto Distribution. *Technometrics*, **29**, 339-349. https://doi.org/10.1080/00401706.1987.10488243

[5] Kotz, S. and Nadarajah, S. (2000) Extreme Value Distributions. Imperial College Press, London. https://doi.org/10.1142/p191

[6] Dupuis, D.J. (1988) Exceedances over High Thresholds: A Guide to Threshold Selection. *Extremes*, **1**, 251-261.

[7] Juarez, S.F. and Schucany, W.R. (2004) Robust and Efficient Estimation for the Generalized Pareto Distribution. *Extremes*, **7**, 237-251. https://doi.org/10.1007/s10687-005-6475-6

[8] Klugman, S.A., Panjer, H.H. and Willmot, G.E. (2012) Loss Models: From Data to Decisions. Fourth Edition, Wiley, New York. https://doi.org/10.1002/9781118787106

[9] Castillo, E. and Hadi, A.S. (1997) Fitting the Generalized Pareto Distribution to Data. *Journal of the American Statistical Association*, **92**, 1619-1620. https://doi.org/10.1080/01621459.1997.10473683

[10] 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

[11] Newey, W.K. and McFadden, D. (1994) Large Sample Estimation and Hypothesis Testing. In: Engle, R. and McFadden, D., Eds., *Handbook of Econometrics*, Volume 4, Elsevier, Amsterdam, 419-554.

[12] Carrasco, M. and Florens, J.-P. (2000) Generalization of GMM to a Continuum of

Moment Condition. *Econometric Theory*, **16**, 797-834.
https://doi.org/10.1017/S0266466600166010

[13] Feuerverger, A. and McDunnough, P. (1984) On Statistical Transform Methods and Their Efficiency. *Canadian Journal of Statistics*, **12**, 303-317.
https://doi.org/10.2307/3314814

[14] Durbin, J. and Knott, M. (1972) Components of the Cramer-von Mises Statistics. *Journal of the Royal Statistical Society*, *Series B*, **34**, 290-307.
https://doi.org/10.1111/j.2517-6161.1972.tb00908.x

[15] Hogg, R.V., McKean, J.W. and Craig, A.T. (2013) Introduction to Mathematical Statistics. Seventh Edition, Pearson, Hoboken.

[16] Luenberger, D.G. (1968) Optimization by Vector Space Methods. Wiley, New York.