



ISSN Online: 2161-7198 ISSN Print: 2161-718X

# Minimum Quadratic Distance Methods Using Grouped Data for Parametric Families of Copulas

# **Andrew Luong**

École d'actuariat, Université Laval, Sainte-Foy, Canada Email: Andrew.Luong@act.ulaval.ca

How to cite this paper: Luong, A. (2018) Minimum Quadratic Distance Methods Using Grouped Data for Parametric Families of Copulas. *Open Journal of Statistics*, **8**, 427-456.

https://doi.org/10.4236/ojs.2018.83028

Received: April 8, 2018 Accepted: June 4, 2018 Published: June 7, 2018

Copyright © 2018 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/





## **Abstract**

Minimum quadratic distance (MQD) methods are used to construct chi-square test statistics for simple and composite hypothesis for parametric families of copulas. The methods aim at grouped data which form a contingency table but by defining a rule to group the data using Quasi-Monte Carlo numbers and two marginal empirical quantiles, the methods can be extended to handle complete data. The rule implicitly defines points on the nonnegative quadrant to form quadratic distances and the similarities of the rule with the use of random cells for classical minimum chi-square methods are indicated. The methods are relatively simple to implement and might be useful for applied works in various fields such as actuarial science.

## **Keywords**

Influence Functions, Quasi-Monte Carlo Numbers, Chi-Square Tests Statistics, Random Cells, Contingency Tables

## 1. Introduction

In actuarial science or biostatistics we often encounter bivariate data which are already grouped into cells forming a contingency table, see Partrat [1] (p 225), Gibbons and Chakraborti [2] (p 511-512) for examples, and the primary focus is on dependency study and we only want like to make inference on association parameters of the parametric survival copula used to model the dependency of the two components of the bivariate observations.

For the complete data, in actuarial science or biostatistics usually we assume to have a sample of nonnegative bivariate observations  $Z_i = (X_i, Y_i)'$  which are independent and identically distributed (iid) as Z = (X, Y)' with the bivariate

survival function expressible as

$$S_{\theta}(x,y) = C_{\theta}(\overline{F}(x), \overline{G}(y)) \tag{1}$$

where  $C_{\theta}(u,v), 0 \le u \le 1, 0 \le v \le 1$  is the survival copula function,  $\overline{F}(x) = P(X > x)$  and  $\overline{G}(y) = P(Y > y)$  are the marginal survival functions. The bivariate model with deductibles in actuarial science as given by Klugman and Parsa [3] can be considered as having complete data within this framework as we still have a sample of bivariate observations which are iid.

In this paper, we emphasize nonnegative distributions. So in general we use survival functions and survival copula functions but it is not difficult to see that the statistical procedures developed can be adjusted to handle the situation where we use distribution functions and distribution copula instead of survival function and survival copula. If we use distributions functions then the bivariate distribution function

$$H(x,y) = C_{\theta}(F(x),G(y))$$

where the marginal distribution functions are given respectively by F(x) and G(y). In the paper by Dobric and Schmid [4], distributions functions are used as the authors emphasize financial applications instead of actuarial science applications. It is not difficult to see that statistical procedures are similar.

For illustrations, we shall discuss of few examples of parametric model for survival copulas. In general, a survival copula can be viewed as a bivariate survival function but the bivariate sample of observations which is given by the complete data is not drawn directly from this bivariate survival function. This should be taken into account when developing inferences methods even when the data is complete. It is natural to have procedures which provide a unified approach for grouped data and for complete data but must be grouped so a rule for grouping the complete data needs to be specified. We shall see that a rule for grouping the data is equivalent to a rule for choosing points on the nonnegative quadrant. We propose inference procedures which are based on quadratic distance and which lead to chi-square tests statistics for the composite hypothesis.

$$H_0: C(u,v) \in \left\{ C_{\theta}(u,v) \right\} \tag{2}$$

with the vector of parameters given by  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_m)'$  and in most of the applications, we just need one or two parameters and the true vector of parameters for the copula model is denoted by  $\boldsymbol{\theta}_0$ . Also, by copula in general we mean survival copula.

In actuarial science we often encounter grouped data, see Klugman *et al.* [5] for the univariate case. Inferences procedures for bivariate censored data have been developed by Shih and Louis [6], see the review paper by Genest *et al.* [7] but inference procedures for grouped data do not seem to receive attention and furthermore, despite that the chi-square tests statistics that Dobric and Schmid [4] propose to make use of a contingency table, complete data must be available first, and then transformed by the marginal empirical distribution functions,

subsequently put into cells of a contingency table. By making use of multinomial distributions which are induced by a contingency table, chi-square tests can be proposed. In practice, if data are gouped into a contingency table without being transformed, then the tests procedures are no longer applicable. They also note that chi-square tests statistics can have good power along some direction of the alternatives yet being simple to apply and might be of interest for practitioners.

We also know that chi-square tests statistics in one dimension might not be consistent in all direction of the alternatives yet due to its simplicity to apply as there is a unique asymptotic chi-square distribution across the composite hypothesis, one can control the size of the test. Depending on the alternatives and by carefully choosing the intervals to partition the real line, chi-square tests can still have good power against some directions of the alternatives and in practice. Often we are primarily concerned about some type of alternatives instead of all alternatives. For these advantages, chi-square tests are still used despite there are more powerful tests such as the Cramer-Von Mises tests, see Greenwood and Nikulin [8] (p 124-126) for power under contaminated mixture distributions alternatives and Lehmann [9] (p 326-329) for discussions on power of chi-square tests which are related to the way to create intervals to group the data in one dimension.

Therefore, if we can retain the advantages of the chi-square tests in two dimensions of having a unique chi-square distribution across the null composite hypothesis and improve on the issue of arbitrariness of a grouping rule, the inference procedures might still be attractive for practitioners as implementing other tests procedures might need extensive simulations to approximate a null distribution which depends on  $\theta_0 \in \Omega$ .

In this paper, we would like to develop minimum quadratic distance (MQD) procedures for grouped data and the procedures can be extended to the situation of having complete data and they must be grouped by specifying a rule which make use of the Halton sequence of Quasi-Monte Carlo (QMC) numbers and two empirical quantiles from the two marginal distributions or marginal survival functions. Tests for copula models can be performed using chi-square tests statistics with data already grouped and if complete data is available they can be grouped according a more clearly defined rule. As mentioned earlier, the rule to select cells to group the data is a rule to select points on the nonnegative quadrant to construct quadratic distances. If complete data is available then it is established using QMC methods and based on the idea of selecting points in the nonnegative quadrant so that Cramer-Von Mises distances can be approximated by quadratic distances. The methods can also be applied to Copula models with a singular component when u = v provided that the Copula function is differentiable with respect to the parameters given by  $\theta$ . An example of such a copula is the one parameter Marshall Olkin(MO) copula, for discussions on MO copulas, see Dobrowolski and Kumar [10] and Marshall and Olkin [11].

We briefly list some copula models often encountered in practice. Most of

them just have one or two parameters. A subclass of Archimedean copulas has the representation using a generator which is the Laplace transform (LT) of a nonnegative random variable  $T \ge 0$  denoted by  $\psi(s;\theta) = \psi_{\theta}(s)$ . The class can be represented as

$$C_{\theta}(u,v) = \psi_{\theta} \left[ \psi_{\theta}^{-1}(u) + \psi_{\theta}^{-1}(v) \right], \ 0 \le u \le 1, 0 \le v \le 1.$$

If we specify a gamma LT with  $\psi_{\theta}(s) = (1+s)^{-\theta}$ ,  $\theta > 0$ , then we have the Clayton or Cook-Johnson copula model

$$C_{\theta}(u,v) = (u^{-\theta} + v^{-\theta} - 1)^{-\frac{1}{\theta}}, \theta > 0.$$

If we specify a positive stable LT  $\psi_{\theta}(s) = e^{-s^{\theta}}$  then we have the positive stable copula model which is also called positive stable frailties model with

$$C_{\theta}(u,v) = \exp\left[-\left\{\left(-\log u\right)^{\frac{1}{\theta}} + \left(-\log v\right)^{\frac{1}{\theta}}\right\}^{\theta}\right], 0 < \theta < 1,$$

see Shih and Louis [6] for these families and for simulations from these copulas, and see the algorithms given by Mai and Scherer [12] (p 98-99).

Beside this subclass the one and two parameters Marshall Olkin copula models are also frequently used. The two parameters MO model can be expressed as

$$C_{\theta}(u,v) = u^{1-\theta_1}v$$
 if  $u^{\theta_1} \ge v^{\theta_2}$  and  $C_{\theta}(u,v) = uv^{1-\theta_2}$  if  $u^{\theta_1} \le v^{\theta_2}$ , 
$$0 < \theta_1 < 1, 0 < \theta_2 < 1, \boldsymbol{\theta} = (\theta_1, \theta_2)'.$$

The model has a singular component and if  $\theta_1 = \theta_2$ , the MO copula model just has one parameter and

$$C_{\theta}(u,v) = u^{1-\theta}v$$
 if  $u \ge v$  and  $C_{\theta}(u,v) = uv^{1-\theta}$  if  $u \le v$ ,

note that  $C_{\theta}(u,v)$  is singular for u=v but a function of  $\theta$ ,  $C_{\theta}(u,v)$  is differentiable. For further discussions on MO copulas see Dobrowolski and Kumar [10] and see Ross [13] (p 103-108) for simulations from MO copulas and Gaussian copulas. The Gaussian Copula model can be represented by

$$C_{\rho}(u,v) = \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}(x^2+y^2-2xy)\right\} dxdy,$$

with the standard normal univariate quantile function denoted by  $\phi^{-1}(.)$  and the integrand of the above integral is a bivariate normal density function with standard normal marginals and parameter  $\rho$ .

Copulas are often used to create bivariate distributions and for inference procedures for these distributions for actuarial science, see Klugman and Parsa [3], Klugman *et al.* [14], Frees and Valdez [15] for examples.

Before giving further details and properties of MQD methods, we shall give the logic behind the MQD procedures.

Let the bivariate empirical survival function be defined as

$$S_n(x,y) = \frac{1}{n} \sum_{i=1}^n I[x_i > x, y_i > y]$$

with I[.] being the usual indicator function, let  $C(u,v) = C_{\mathbf{e}_0}(u,v)$  and define the two univariate empirical marginal survival functions as

$$\overline{F}_n(x) = \frac{1}{n} \sum_{i=1}^n I[x_i > x] \text{ and } \overline{G}_n(x) = \frac{1}{n} \sum_{i=1}^n I[y_i > x],$$

we then have the following convergence in probability properties,

$$S_n(x,y) \xrightarrow{p} S(x,y), \overline{F}_n(x) \xrightarrow{p} \overline{F}(x), \overline{G}_n(x) \xrightarrow{p} \overline{G}(x)$$

with the true survival function and the marginal survival functions are given respectively by S(x,y),  $\overline{F}(x)$  and  $\overline{G}(y)$ . We shall assume that  $\overline{F}(x)$  and  $\overline{G}(y)$  are absolutely continuous, S(x,y) is either absolutely continuous or S(x,y) is absolutely continuous everywhere except when x=y where the survival distribution can be singular as in the case of the bivariate exponential model introduced by Marshall and Olkin [11].

Now if the parametric survival copula model is valid,

$$S(x,y) = C(\overline{F}(x), \overline{G}(y)), C(u,v) = C_{\theta_0}(u,v)$$

so that

$$S_n(x,y) - C_{\theta}(\overline{F}_n(x), \overline{G}_n(y)) \xrightarrow{p} 0 \text{ for } \theta = \theta_0.$$

For the time being assume that the M points given by  $(x_l, y_l)'$ ,  $l = 1, \dots, M$  are already chosen, then we can define the vector of empirical components,

$$\hat{\mathbf{z}}_n = \left(S_n(x_1, y_1), \dots, S_n(x_M, y_M)\right)'$$

with the counterpart vector which makes use of the copula model,

$$\hat{\boldsymbol{z}}_{\boldsymbol{\theta}} = \left( C_{\boldsymbol{\theta}} \left( \overline{F}_{n} \left( \boldsymbol{x}_{1} \right), \overline{G}_{n} \left( \boldsymbol{y}_{1} \right) \right), \cdots, C_{\boldsymbol{\theta}} \left( \overline{F}_{n} \left( \boldsymbol{x}_{M} \right), \overline{G}_{n} \left( \boldsymbol{y}_{M} \right) \right) \right)',$$

and form the vector of differences  $G_n(\theta) = (\hat{z}_n - \hat{z}_\theta)'$ , by choosing a symmetric positive definite matrix W we can form a class of quadratic distances (QD) given by

$$Q_{n}(\boldsymbol{\theta}) = (\hat{z}_{n} - \hat{z}_{\theta})' W (\hat{z}_{n} - \hat{z}_{\theta}).$$

A positive definite matrix can be used to create a weighted Euclidean norm, so we can also let

$$Q_n(\boldsymbol{\theta}) = \|G_n(\boldsymbol{\theta})\|^2$$
,

 $\|.\|$  is the weighted Euclidean norm induced by W and if we let W = I then we obtain the classical Euclidean norm. QD inferences procedures developed subsequently are based on  $Q_n(\theta)$  which are similar to the univariate case. For MQD procedures with univariate observations, see Luong and Thompson [16].

The paper is organized as follows.

In Section 3, MQD methods will be developed using predetermined grouped data such as data presented using a contingency table. The efficient quadratic distances is derived and can be used for estimation and model testing. Asymp-

totic theory is established for MQD estimators and chi-square tests using quadratic distances can be constructed for testing copula models. In Section 4, by viewing grouped data as defining a set of points on the nonnegative quadrant, a rule to select points is proposed based on Quasi-Monte-Carlo numbers and two sample quantiles if complete data is available and the methods can be extended to the situation where complete data is available. The methods can be seen as similar to minimum chi-square methods with random cells but with a rule to define these cells. The choice of random cells for minimum chi-square methods is less well defined. Section 5 illustrates the implementations of MQD methods using a limited simulation study by comparing the methods of moment estimator (MM) estimators based on sample Spearman rho which requires the availability of complete data versus the MQD estimator which uses grouped data for the one parameter Marshall-Olkin model and it appears that the chi-square tests have some power to detect alternatives which can be represented as mixture or contaminated copula model such as the mixture of one parameter Marshall-Olkin copula model and Gaussian copula model from the study. The findings appear to be in line with chi-square tests in one dimension which also display similar properties if intervals are chosen properly.

# 2. MQD Methods Using Grouped Data

# 2.1. Contingency Tables

Contingency table data can be viewed as a special form of two-dimensional grouped data. We will give some more details about this form of grouped data.

Assume that we have a sample  $Z_i = (X_i, Y_i)', i = 1, \dots, n$  which are independent and identically distributed as Z = (X, Y)' which follows a non-negative continuous bivariate distribution with model survival function given by  $C_{\theta}(\bar{F}(x), \bar{G}(y))$ . The marginal survival functions are given respectively by  $\bar{F}(x)$  and  $\bar{G}(y)$  assumed to be absolutely continuous but there is no parametric model assumed for the marginals.

The vector of parameters is  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_m)'$ , the true vector of parameters is denoted by  $\boldsymbol{\theta}_0$ . We do not observe the original sample but observations are grouped and put into a contingency table and only the number which fall into each cells of the contingency table are recorded or equivalently the sample proportions which fall into these cells are recorded. Contingency tables are often encountered in actuarial science and biostatistics, see Partrat [1] (p 225), Gibbons and Chakraborti [2] (p 511-512) and we shall give a brief description below.

Let the nonnegative axis X be partitioned into disjoints interval  $\bigcup_{i=1}^{I} \left[ S_{i-1}, S_i \right]$  with  $s_0 = 0, s_I = \infty$  and similarly, the axis Y be partitioned into disjoints interval  $\bigcup_{i=0}^{J} \left[ t_{j-1}, t_j \right]$  with  $t_0 = 0, t_J = \infty$ .

The nonnegative quadrant can be partitioned into nonoverlapping cells of the form.

$$C_{ij} = [s_{i-1}, s_i) \times [t_{j-1}, t_j], i = 1, \dots, I, j = 1, \dots, J.$$

The contingency table  $T = (C_{ij})$  is formed which can be viewed as a matrix with elements given by

$$C_{ii}$$
,  $i = 1, \dots, I$ ,  $j = 1, \dots, J$ .

The empirical bivariate survival function is as defined earlier with  $S_n(x,y) \xrightarrow{p} S(x,y)$ , the underlying bivariate survival distribution. We assume that S(x,y) is either absolutely continuous or it can have a singular component when X=Y as in the case of the bivariate exponential distribution of Marshall Olkin [9] but absolutely continuous elsewhere. Implicitly, the marginal survival functions  $\overline{F}(x)$  and  $\overline{G}(y)$  are assumed to be absolutely continuous.

The sample proportion or empirical probability for one observation which falls into cell  $C_{ii}$  can be obtained using  $S_n(x, y)$ 

$$\widehat{p_n}(C_{ij}) = S_n(s_{i-1}, t_{j-1}) - S_n(s_{i-1}, t_j) - S_n(s_i, t_{j-1}) + S_n(s_j, t_j)$$
(3)

and the corresponding probability  $\widehat{p}_{\theta}(C_{ij})$  using the copula model coupled with the empirical survival distributions  $\overline{F}_n$  and  $\overline{G}_n$  with

$$\widehat{S_{\theta}}(s,t) = C_{\theta}(\overline{F_n}(s), \overline{G_n}(t))$$
 is given by

$$\widehat{p_{\boldsymbol{\theta}}}\left(C_{ij}\right) = \widehat{S_{\boldsymbol{\theta}}}\left(s_{i-1}, t_{j-1}\right) - \widehat{S_{\boldsymbol{\theta}}}\left(s_{i-1}, t_{j}\right) - \widehat{S_{\boldsymbol{\theta}}}\left(s_{i}, t_{j-1}\right) + \widehat{S_{\boldsymbol{\theta}}}\left(s_{j}, t_{j}\right).$$

It is not difficult to see that there is redundant information displayed by a contingency table, one way to see that there is duplication is to note

$$\widehat{S_{\boldsymbol{\theta}}}(s_i, t_J) = 0, \ S_n(s_i, t_J) = 0, \ i = 1, \dots, I$$
(4)

and similarly,  $\widehat{S}_{\theta}(s_I, t_j) = 0$ ,  $S_n(s_I, t_j) = 0$ ,  $j = 1, \dots, J$ .

Therefore, the set points given by  $\{(s_I,t_j),(s_i,t_J),i=1,\cdots,I,j=1,\cdots,J\}$  can be discarded without affecting the information provided by the contingency table. Consequently, we can view a contingency table implicitly define a grid on the nonnegative quadrant with only M=(I-1)(J-1) points. It is also clear that if we want a rule to choose cells, the same rule will allow us to choose points on the nonnegative quadrant.

The objective function of the proposed quadratic form will be given below. It is a natural extension of the objective function used in the univariate case. Define a vector with empirical components so that we only need one subscript by collapsing the points of the contingency table given by

 $\{(s_i, t_j), i = 1, \dots, I - 1, j = 1, \dots, J - 1\}$  into a vector by putting the first row of the matrix as the first batch of elements of the vector and the second row being the second batch of elements so forth so on, *i.e.*, let

$$\widehat{\boldsymbol{z}_n} = \left(S_n(s_1, t_1), \dots, S_n(s_M, t_M)\right)', M = (I - 1)(J - 1). \tag{5}$$

and its counterpart which makes use of the copula model is

$$\widehat{\boldsymbol{z}_{\boldsymbol{\theta}}} = \left(\widehat{\boldsymbol{S}_{\boldsymbol{\theta}}}(\boldsymbol{s}_{1}, t_{1}), \cdots, \widehat{\boldsymbol{S}_{\boldsymbol{\theta}}}(\boldsymbol{s}_{M}, t_{M})\right)'. \tag{6}$$

The number of components of  $z_n$  is M with the assumption M > m. A class of quadratic distances can be defined as

$$Q_{n}(\boldsymbol{\theta}) = (\hat{z}_{n} - \hat{z}_{\theta})' W (\hat{z}_{n} - \hat{z}_{\theta})$$
(7)

with W being a symmetric and positive definite matrix. In this class, we focus on two choices of W.

Letting W=I, we obtain the unweighted quadratic distance, this choice is not optimum but it produces consistent estimators and can be used as preliminary estimates for  $\theta$  to start the numerical procedures for finding more efficient estimators. The matrix W is defined up to a positive constant as minimizing the objective function multiplied by a positive constant still gives the same estimators and  $\hat{W}$  a consistent estimate of W can be used to replace W without affecting the asymptotic theory for estimation and asymptotic distribution for test statistics. Using quadratic distance theory or generalized methods of moment (GMM) theory, it is not difficult to see that an optimum choice for W is to let  $W=W_0$  where  $\Omega_0=W_0^{-1}$  and  $\Omega_0$  is an asymptotic covariance matrix which is given by

$$\sqrt{n}\left(\hat{z}_n - \hat{z}_{\theta_0}\right) \xrightarrow{L} N\left(\mathbf{0}, \mathbf{\Omega}_0\right),$$

see Remark 2.4.3 given by Luong and Thompson [16] (p 245).

Clearly,  $\Omega_0$  depends on  $\theta_0$ . We shall obtain the expression for  $\Omega_0$  and show that  $\Omega_0$  can be estimated by  $\hat{\Omega}_0$  in the next section as we can obtain a preliminary consistent estimate for  $\theta_0$  by using the unweighted quadratic distance or other quick methods; see the methods of moment using Spearman-rho in Section 5.2 for example. Consequently, by quadratic distance we mean the following efficient version with the objective function defined as

$$Q_n(\boldsymbol{\theta}) = (\hat{z}_n - \hat{z}_{\theta})' \widehat{\boldsymbol{W}}_0(\hat{z}_n - \hat{z}_{\theta}) \text{ with } \widehat{\boldsymbol{W}}_0 = \widehat{\boldsymbol{\Omega}}_0^{-1}.$$
 (8)

The version with W=I will be called unweighted quadratic distance. In the next section we shall use the influence function representation for  $\sqrt{n}\left(\hat{z}_n-\hat{z}_{\theta_0}\right)$  to derive  $\Omega_0$  and we shall also propose  $\hat{\Omega}_0$  a consistent estimate for  $\Omega_0$ .

#### 2.2. Optimum Matrix $W_0$

The matrix  $\Omega_0$  which is the asymptotic covariance matrix of the vector  $\sqrt{n}\left(\hat{z}_n-\hat{z}_{\theta_0}\right)$  plays an important role for MQD methods as we can obtain estimators with good efficiencies for estimators using  $\Omega_0$  or a consistent estimate of  $\Omega_0$  and we also have chi-square tests statistics. Despite that  $\Omega_0$  is unknown, its elements are not complicated and moreover, it can be replaced by a consistent estimate without affecting the asymptotic properties of the procedures. We shall give more details about this matrix and construct  $\widehat{\Omega}_0$ , a consistent estimate of  $\Omega_0$ .

Using influence representation for the vector of functions of  $\sqrt{n}(\hat{z}_n - \hat{z}_{\theta_0})$  which depend on three functions  $S_n, \overline{F_n}, \overline{G_n}$  as discussed by Reid [17], see tech-

nical appendix (TA1) in the Appendices for more details, it can be seen that  $\Omega_0$  is the covariance matrix of the vector  $\boldsymbol{h}(x,y)$  under  $\boldsymbol{\theta}_0$  with

$$\boldsymbol{h}(x,y) = \begin{bmatrix} \boldsymbol{a}_{1}'Y_{1} \\ \vdots \\ \boldsymbol{a}_{M}'Y_{m} \end{bmatrix},$$

$$\boldsymbol{a}_{i}' = \left(1, -\frac{\partial C(\overline{F}(s_{i}), \overline{G}(t_{i}))}{\partial u}, -\frac{\partial C(\overline{F}(s_{i}), \overline{G}(t_{i}))}{\partial v}\right), i = 1, \dots, M,$$

$$\boldsymbol{Y}_{i} = \begin{bmatrix} I[x > s_{i}, y > t_{i}] - S(s_{i}, t_{i}) \\ I[x > s_{i}] - \overline{F}(s_{i}) \\ I[y > t_{i}] > \overline{G}(t_{i}) \end{bmatrix}, i = 1, \dots, M$$

and I[.] is the usual indicator function,

$$S(s_i,t_i) = C(\overline{F}(s_i),\overline{G}(t_i)), C = C_{\theta_0}.$$

 $\frac{\partial C(u,v)}{\partial u}$ ,  $\frac{\partial C(u,v)}{\partial v}$  are respectively the partial derivatives of C(u,v) with respect to u and v.

It is not difficult to see that the elements of  $\Omega_0$  are

$$\Omega_0(i,j) = \mathbf{a}_i' Cov(\mathbf{Y}_i, \mathbf{Y}_j) \mathbf{a}_j$$

with  $Cov(Y_i, Y_j) = E(Y_i Y_j')$  and since  $Y_i$  and  $Y_j$  are not identically distributed  $Cov(Y_i, Y_j)$  is not symmetric, the matrix has 9 elements, see technical Appendix (TA2) in the Appendices for more details. The elements can be expressed as

$$c_{11} = S\left(\max\left(s_{i}, s_{j}\right), \max\left(t_{i}, t_{j}\right)\right) - S\left(s_{i}, t_{i}\right) S\left(s_{j}, t_{j}\right)$$

$$c_{12} = S\left(\max\left(s_{i}, s_{j}\right), t_{i}\right) - S\left(s_{i}, t_{i}\right) \overline{F}\left(s_{j}\right)$$

$$c_{13} = S\left(s_{i}, \max\left(t_{i}, t_{j}\right)\right) - S\left(s_{i}, t_{i}\right) \overline{G}\left(t_{j}\right)$$

$$c_{21} = S\left(\max\left(s_{i}, s_{j}\right), t_{j}\right) - S\left(s_{j}, t_{j}\right) \overline{F}\left(s_{i}\right)$$

$$c_{22} = \overline{F}\left(\max\left(s_{i}, s_{j}\right)\right) - \overline{F}\left(s_{i}\right) \overline{F}\left(s_{j}\right)$$

$$c_{23} = S\left(s_{i}, t_{j}\right) - \overline{F}\left(s_{i}\right) \overline{G}\left(t_{j}\right)$$

$$c_{31} = S\left(s_{j}, \max\left(t_{i}, t_{j}\right)\right) - S\left(s_{j}, t_{j}\right) \overline{G}\left(t_{i}\right)$$

$$c_{32} = S\left(s_{i}, t_{j}\right) - \overline{F}\left(s_{j}\right) \overline{G}\left(t_{i}\right)$$

$$c_{33} = \overline{G}\left(\max\left(t_{i}, t_{j}\right)\right) - \overline{G}\left(t_{i}\right) \overline{G}\left(t_{j}\right)$$

$$(9)$$

The elements  $c_{ij}$  can be estimated empirically by replacing  $S, \overline{F}, \overline{G}$  in the expressions of  $c_{ij}$  by  $S_n, \overline{F_n}, \overline{G_n}$  for i=1,2,3, j=1,2,3. The estimates  $\widehat{c_{ij}}, i=1,2,3, j=1,2,3$  can be formed.

Therefore, we can form  $\widehat{Cov}(Y_i, Y_j)$  which estimates  $Cov(Y_i, Y_j)$ . Similarly, by replacing  $\theta_0$  by a consistent preliminary estimate  $\theta_0^{(0)}$  which can be obtained using the unweighted quadratic distance for example and replacing  $\overline{F}, \overline{G}$ 

by  $\overline{F_n}, \overline{G_n}$  we can estimate  $a_i$  by  $\widehat{a_i}, i = 1, \dots, M$ .  $\widehat{\Omega}_n$  an estimate for  $\Omega_n$  will have the elements given by

$$\widehat{\Omega_0}(i,j) = \widehat{a}_i' \widehat{Cov}(Y_i, Y_j) \widehat{a}_j, i = 1, \dots, 3, j = 1, \dots, 3$$
(10)

and define  $\widehat{W}_0 = \widehat{\Omega}_0^{-1}$ .  $\widehat{W}_0$  will be used as an optimum matrix for constructing quadratic distance as the asymptotic property remain unchanged. We can replace the unknown matrix  $W_0 = \Omega_0^{-1}$  by its consitent estimate which is  $\widehat{W}_0$ without affecting asymptotic theory for estimation and tests.

# 3. MQD Methods Using Grouped Data

#### 3.1. Estimation

The MQD estimators can be seen as given by the vector  $\hat{\boldsymbol{\theta}}$  which minimizes

$$Q_n(\boldsymbol{\theta}) = (\hat{z}_n - \hat{z}_{\theta})' \widehat{\boldsymbol{W}}_0(\hat{z}_n - \hat{z}_{\theta}). \tag{11}$$

and since

$$\widehat{\boldsymbol{z}}_{n} = \left(S_{n}\left(\boldsymbol{s}_{1}, t_{1}\right), \cdots, S_{n}\left(\boldsymbol{s}_{M}, t_{M}\right)\right)',\tag{12}$$

$$\widehat{\boldsymbol{z}_{\boldsymbol{\theta}}} = \left(\widehat{S_{\boldsymbol{\theta}}}\left(\boldsymbol{s}_{1}, t_{1}\right), \cdots, \widehat{S_{\boldsymbol{\theta}}}\left(\boldsymbol{s}_{M}, t_{M}\right)\right)', \tag{13}$$

$$\widehat{S_{\theta}}(s_{i},t_{i}) = C_{\theta}(\overline{F_{n}}(s_{i}), \overline{G_{n}}(t_{i})), i = 1, \dots, M,$$

$$G_{n}(\theta) = (\hat{z}_{n} - \hat{z}_{\theta}),$$
(14)

we can also used the weighted Euclidean norm  $\|\cdot\|$  with the use of  $\widehat{W}_0$  and let

$$Q_n(\boldsymbol{\theta}) = \left\| \boldsymbol{G}_n(\boldsymbol{\theta}) \right\|^2. \tag{15}$$

Consistency for quadratic distance estimators using predetermined grouped data or if complete data is available but must be grouped according a rule can be treated in a unified way using the following Theorem 1 which is essentially Theorem 3.1 of Pakes and Pollard [18] (p 1038) and the proof has been given by the authors. In fact, their Theorems 3.1 and 3.3 are also useful for Section 4 where we have complete data and we have choices to group the data into cells or equivalently forming the artificial sample points on the nonnegative quadrant to form the quadratic distances.

## Theorem 1 (Consistency)

Under the following conditions  $\hat{\theta}$  converges in probability to  $\theta_0$ :

- 1)  $\|G_n(\hat{\theta})\| \le o_p(1) + \inf_{\theta \in \Omega} (\|G_n(\theta)\|)$ , the parameter space  $\Omega$  is compact
- 2)  $\|\mathbf{G}_{n}(\mathbf{\theta}_{0})\| = o_{n}(1)$ ,
- 3)  $\sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\| > \delta} \left( \frac{1}{\|\boldsymbol{G}_n(\boldsymbol{\theta})\|} \right) = O_p(1)$  for each  $\delta > 0$ .

Theorem 3.1 states condition b) as  $G_n(\theta_0) = o_n(1)$  but in the proof the authors just use  $\|\mathbf{G}_n(\mathbf{\theta}_0)\| = o_p(1)$  so we state condition b) as  $\|\mathbf{G}_n(\mathbf{\theta}_0)\| = o_p(1)$ .

An expression is  $o_p(1)$  if it converges to 0 in probability,  $O_p(1)$  if it is

bounded in probability and  $o_p\left(n^{\frac{-1}{2}}\right)$  if it converges to 0 in probability faster

than  $n^{\frac{1}{2}} \to 0$ . We have  $\inf_{\theta \in \Omega} \left( \| G_n(\theta) \| \right)$  occurs at the values of the vector values of the MQD estimators, so the conditions 1) and 2) are satisfied for both versions. Implicitly, we make the assumption that the parameter space  $\Omega$  is compact. Also, for both versions  $\| G_n(\theta) \|^{-p} \to 0$  only at  $\theta = \theta_0$  in general if the number of components of  $G_n(\theta)$  is greater than the number of parameters of the model, *i.e.*, M > m.

For  $\theta \neq \theta_0$  we have  $0 < Q_n(\theta) \le B$  for some B > 0 since survival functions evaluated at points are components of  $G_n(\theta)$  and these functions are bounded. This implies that there exist real numbers u and v with  $0 < u < v < \infty$  such that

$$P\left(u \leq \sup_{\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| > \delta} \left(\frac{1}{\|\boldsymbol{G}_n(\boldsymbol{\theta})\|}\right) \leq v\right) \to 1 \quad \text{as} \quad n \to \infty.$$

Therefore, the minimum quadratic distance (MQD) estimators are consistent, *i.e.*,  $\hat{\theta} \xrightarrow{p} \theta_0$ . The Theorem 3.1 given by Pakes and Pollard [18] (p 1038-1039) is an elegant theorem using the norm concept of functional analysis. Now we turn our attention to the question of asymptotic normality for the quadratic distance estimators and it is possible to have unified approach using their Theorem 3.3, see Pakes and Pollard [18] (p 1040-1043) where we shall restate their Theorem as Theorem 2 and Corollary 1 given subsequently after the following discussions on the ideas behind their theorem, allowing us to get asymptotic normality results for estimators obtained from extremum of a smooth or nonsmooth objective function.

Note that 
$$G_n(\theta) \xrightarrow{p} G(\theta)$$
 (16)

with

$$\boldsymbol{G}(\boldsymbol{\theta}) = \left(S(s_1, t_1) - C_{\boldsymbol{\theta}}(\overline{F}(s_1), \overline{G}(t_1)), \dots, S(s_M, t_M) - C_{\boldsymbol{\theta}}(\overline{F}(s_M), \overline{G}(t_M))\right)'. (17)$$

The points  $(s_1, t_1)', \dots, (s_M, t_M)'$  are predetermined by a contingency table we give and we have no choice but to analyze the grouped data as they are presented.

Note that  $G(\theta)$  is non-random and if we assume  $G(\theta)$  is differentiable with repect to with derivative matrix  $\Gamma(\theta)$ , then we can define the random function  $Q_n^a(\theta)$  to approximate  $Q_n(\theta)$  with

$$Q_n^a(\boldsymbol{\theta}) = (\|\boldsymbol{L}_n(\boldsymbol{\theta})\|)^2, \, \boldsymbol{L}_n(\boldsymbol{\theta}) = \boldsymbol{G}_n(\boldsymbol{\theta}_0) + \boldsymbol{\Gamma}(\boldsymbol{\theta})(\boldsymbol{\theta} - \boldsymbol{\theta}_0).$$
 (18)

By using  $\frac{\partial C_{\theta}(u,v)}{\partial \theta_i}$  which is the partial derivative of  $C_{\theta}(u,v)$  with repect

to  $\theta_j$ ,  $j = 1, \dots, m$ , the matrix  $\Gamma(\theta)$  can be displayed explicitly as

$$\Gamma(\boldsymbol{\theta}) = -\begin{bmatrix} \frac{\partial C_{\theta}(\overline{F}(s_{1}), \overline{G}(t_{1}))}{\partial \theta_{1}} & \cdots & \frac{\partial C_{\theta}(\overline{F}(s_{1}), \overline{G}(t_{1}))}{\partial \theta_{m}} \\ \vdots & \ddots & \vdots \\ \frac{\partial C_{\theta}(\overline{F}(s_{M}), \overline{G}(t_{M}))}{\partial \theta_{1}} & \cdots & \frac{\partial C_{\theta}(\overline{F}(s_{M}), \overline{G}(t_{M}))}{\partial \theta_{m}} \end{bmatrix}.$$
(19)

Note that  $Q_n^a(\boldsymbol{\theta})$  is differentiable and a quadratic function of  $\boldsymbol{\theta}$ , the vector  $\boldsymbol{\theta}^*$  which minimizes  $Q_n^a(\boldsymbol{\theta})$  can be obtained explicitly with

$$\boldsymbol{\theta}^* - \boldsymbol{\theta}_0 = -\left(\boldsymbol{\Gamma}'\widehat{\boldsymbol{W}}_0\boldsymbol{\Gamma}\right)^{-1}\boldsymbol{\Gamma}'\widehat{\boldsymbol{W}}_0\boldsymbol{G}_n\left(\boldsymbol{\theta}_0\right) \tag{20}$$

and since  $\widehat{W_0} \xrightarrow{p} W_0$ .  $W_0$  is assumed to be a positive define matrix; we have

$$\sqrt{n} \left( \boldsymbol{\theta}^* - \boldsymbol{\theta}_0 \right) = - \left( \Gamma' \widehat{\boldsymbol{W}}_0 \Gamma \right)^{-1} \Gamma' \widehat{\boldsymbol{W}}_0 \sqrt{n} \boldsymbol{G}_n \left( \boldsymbol{\theta}_0 \right) 
= - \left( \Gamma' \boldsymbol{W}_0 \Gamma \right)^{-1} \Gamma' \boldsymbol{W}_0 \sqrt{n} \boldsymbol{G}_n \left( \boldsymbol{\theta}_0 \right) + o_n (1).$$
(21)

Clearly set up fits into the scopes of their Theorem 3.3 where we shall rearrange the results to make them more suitable for MQD methods and verify that we can satisfy the regularity conditions of Theorem 3.3. We shall state Theorem 2 and Corollary 1 below which are essentially their Theorem (3.3) and the proofs have been given by Pakes and Pollard [18]. Note that the condition 4) is slightly more stringent but simpler to check than the condition 3) in their Theorem.

#### Theorem 2

Let  $\hat{\theta}$  be a vector of consistent estimators for  $\theta_0$ , the unique vector which satisfies  $G(\theta_0) = 0$ .

Under the following conditions:

1) The parameter space  $\Omega$  is compact,  $\hat{\boldsymbol{\theta}}$  is an interior point of  $\Omega$ .

2) 
$$\|\boldsymbol{G}_n(\hat{\boldsymbol{\theta}})\| \le o_p\left(n^{-\frac{1}{2}}\right) + \inf_{\boldsymbol{\theta} \in \boldsymbol{\Omega}} \|\boldsymbol{G}_n(\boldsymbol{\theta})\|$$

- 3) G(.) is differentiable at  $\theta_0$  with a derivative matrix  $\Gamma = \Gamma(\theta_0)$  of full rank.
- 4)  $\sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\| \leq \delta_n} \sqrt{n} \|\boldsymbol{G}_n(\boldsymbol{\theta}) \boldsymbol{G}(\boldsymbol{\theta}) \boldsymbol{G}_n(\boldsymbol{\theta}_0)\| = o_p(1)$  for every sequence  $\{\delta_n\}$  of positive numbers which converge to zero.
  - 5)  $\|\boldsymbol{G}_n(\boldsymbol{\theta}_0)\| = o_p(1)$ .
  - 6)  $\theta_0$  is an interior point of Ω.

Then, we have the following representation which will give the asymptotic distribution of  $\hat{\theta}$  in Corollary 1, *i.e.*,

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_{0}\right) = -\left(\mathbf{\Gamma}'\widehat{\boldsymbol{W}}_{0}\mathbf{\Gamma}\right)^{-1}\mathbf{\Gamma}'\widehat{\boldsymbol{W}}_{0}\sqrt{n}\boldsymbol{G}_{n}\left(\boldsymbol{\theta}_{0}\right) + o_{p}\left(1\right),\tag{21}$$

or equivalently, using equality in distribution,

$$\sqrt{n} \left( \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 \right) = {}^{d} - \left( \boldsymbol{\Gamma}' \boldsymbol{W}_0 \boldsymbol{\Gamma} \right)^{-1} \sqrt{n} \boldsymbol{\Gamma}' \boldsymbol{W}_0 \boldsymbol{G}_n \left( \boldsymbol{\theta}_0 \right)$$
 (22)

or equivalently,

$$\sqrt{n} \left( \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 \right) = {}^{d} - \left( \Gamma' \widehat{\boldsymbol{W}}_0 \Gamma \right)^{-1} \sqrt{n} \Gamma' \widehat{\boldsymbol{W}}_0 \boldsymbol{G}_n \left( \boldsymbol{\theta}_0 \right)$$
 (23)

The proofs of these results follow the results used to prove Theorem 3.3 given by Pakes and Pollard [18] (p 1040-1043). For expression (22) or expression (23) to hold, in general only condition 5) of Theorem 2 is needed and there is no need to assume that  $G_n(\theta_0)$  has an asymptotic distribution. From the results of Theorem 2, it is easy to see that we can obtain the main result of the following Corollary 1 which gives the asymptotic covariance matrix for the quadratic distance estimators for both versions.

# Corollary 1

Let 
$$Y_n = \sqrt{n} \Gamma' W_0 G_n(\theta_0)$$
, if  $Y_n \xrightarrow{L} N(\mathbf{0}, V)$  then
$$\sqrt{n} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{L} N(\mathbf{0}, T) \text{ with}$$

$$T = (\Gamma' W_0 \Gamma)^{-1} V (\Gamma' W_0 \Gamma)^{-1}, \qquad (24)$$

The matrices T and V depend on  $\theta_0$ , we also adopt the notations  $T = T(\theta_0), V = V(\theta_0)$ .

We observe that when applying condition 4) of Theorem 2 to MQD methods in general involves technicalities. Note that to verify the condition 4, it is equivalent to verify

$$\sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_{0}\| \leq \delta_{n}} n(\|\boldsymbol{G}_{n}(\boldsymbol{\theta}) - \boldsymbol{G}(\boldsymbol{\theta}) - \boldsymbol{G}_{n}(\boldsymbol{\theta}_{0})\|)^{2} = o_{p}(1), \qquad (25)$$

a regularity condition for the approximation is of the right order which implies the condition 3 given by their Theorem 3.3, which might be the most difficult to check. The rest of the conditions for Theorem 2 are satisfied in general.

Let

$$g_n(\boldsymbol{\theta}) = n(\|\boldsymbol{G}_n(\boldsymbol{\theta}) - \boldsymbol{G}(\boldsymbol{\theta}) - \boldsymbol{G}_n(\boldsymbol{\theta}_0)\|)^2$$
(26)

and define  $u_n(\theta) = G_n(\theta) - G(\theta) - G_n(\theta_0)$  which can be expressed as

$$\boldsymbol{u}_{n}(\boldsymbol{\theta}) = \left(C_{\theta_{0}}\left(\overline{F_{n}}(s_{1}), \overline{G_{n}}(t_{1})\right) - S(s_{1}, t_{1}), \cdots, C_{\theta_{0}}\left(\overline{F_{n}}(s_{M}), \overline{G_{n}}(t_{M})\right) - S(s_{M}, t_{M})\right)$$

$$, \cdots, C_{\theta}\left(\overline{F_{n}}(s_{1}), \overline{G_{n}}(t_{1})\right) - C_{\theta}\left(\overline{F}(s_{1}), \overline{G}(t_{1})\right), \cdots,$$

$$C_{\theta}\left(\overline{F_{n}}(s_{M}), \overline{G_{n}}(t_{M})\right) - C_{\theta}\left(\overline{F}(s_{M}), \overline{G}(t_{M})\right)\right).$$

$$(27)$$

Consequently,  $g_n(\theta)$  can also be expressed as

$$g_n(\boldsymbol{\theta}) = n\boldsymbol{u}_n'(\boldsymbol{\theta})\widehat{\boldsymbol{W}_0}\boldsymbol{u}_n'(\boldsymbol{\theta}).$$

Since the elements of  $\sqrt{n}u_n'(\boldsymbol{\theta})$  are bounded in probability, it is not difficult to see that the sequence  $\left\{g_n(\boldsymbol{\theta})\right\}$  is bounded in probability and continuous in probability with  $g_n(\boldsymbol{\theta}) \stackrel{p}{\longrightarrow} g_n(\boldsymbol{\theta}')$  as  $\boldsymbol{\theta} \rightarrow \boldsymbol{\theta}'$ . Also note that  $g_n(\boldsymbol{\theta}_0) = 0$ . Therefore, results given in section of Luong *et al.* [19] (p 218) can be used to justify the sequence of functions.  $\left\{g_n(\boldsymbol{\theta})\right\}$  attains its maximum on the compact set  $C_n = \left\{\boldsymbol{\theta} \left\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\right\| \leq \delta_n\right\}$  in probability and hence has the property

$$\sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\| \leq \delta_n} g_n(\boldsymbol{\theta}) \xrightarrow{p} 0 \text{ as } n \to \infty \text{ and } \boldsymbol{\theta} \to \boldsymbol{\theta}_0.$$
Since  $\sqrt{n} G_n(\boldsymbol{\beta}_0) \xrightarrow{p} N(0, \boldsymbol{W}_0^{-1}).$ 

Using results of Corollary 1, we have asymptotic normality for the MQD estimators which is given by

$$\sqrt{n} \left( \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 \right) \xrightarrow{L} N \left( 0, \left( \boldsymbol{\Gamma}' \boldsymbol{W}_0 \boldsymbol{\Gamma} \right)^{-1} \right),$$
 (28)

 $\Gamma$  as given by expression (19) can be estimated once the parameters are estimated.

# 3.2. Model Testing

## 3.2.1. Simple Hypothesis

In this section, the quadratic distance  $Q_n(\theta)$  will be used to construct goodness of fit test statistics for the simple hypothesis

 $H_0$ : data coming from a specified distribution with distribution  $F_{\theta_0}$ ,  $\theta_0$  is specified. The chi-square test statistic with its chi-square asymptotic distribution and its degree of freedom r are given below, *i.e.*,

$$nQ_n(\boldsymbol{\theta}_0) \xrightarrow{L} \chi^2(r=M)$$
. (29)

It is not difficult to see that indeed we have the above asymptotic chi-square distribution as  $nQ_n(\boldsymbol{\theta}_0) = \sqrt{n}G_n'(\boldsymbol{\theta}_0)\widehat{W_0}\sqrt{n}G_n(\boldsymbol{\theta}_0)$  and  $\sqrt{n}G_n(\boldsymbol{\theta}_0) \stackrel{L}{\longrightarrow} N(0,\boldsymbol{W}_0^{-1})$ ,  $\boldsymbol{W}_0^{-1} = \boldsymbol{\Omega}_0$ ,  $\widehat{\boldsymbol{W}}_0 \stackrel{p}{\longrightarrow} \boldsymbol{W}_0$ , using standard results for distribution of quadratic forms, see Luong and Thompson [16] (p 247) for example.

# 3.2.2. Composite Hypothesis

The quadratic distances  $Q_n(\theta)$  can also be used for construction of the test satisfies for the composite hypothesis

 $H_0$ : data comes from a parametric model  $\{S_{\theta}\}$ . The chi-square test statistic and its asymptotic distribution are given similarly in this case by

$$nQ_n(\hat{\boldsymbol{\beta}}) \xrightarrow{L} \chi^2(r = M - m),$$
 (30)

with M > m. To justify the asymptotic chi-square distribution given above, note that we have the equality in probability,  $nQ_n(\boldsymbol{\theta}) = nQ_n^a(\hat{\boldsymbol{\theta}}) + o_p(1)$ . It suffices to consider the asymptotic distribution of  $nQ_n^a(\hat{\boldsymbol{\theta}})$  as we also have the following equalities in distribution,

$$nQ_n(\hat{\boldsymbol{\theta}}) = {}^{d} nQ_n^a(\hat{\boldsymbol{\theta}}) = n \left\| L_n(\hat{\boldsymbol{\theta}}) \right\|^2 = \sqrt{n}L'_n(\hat{\boldsymbol{\theta}})\widehat{W}_0\sqrt{n}\boldsymbol{L}_n(\hat{\boldsymbol{\theta}}),$$

 $L_n(\boldsymbol{\theta})$  as given by expression. Therefore we also have the following equalities in distribution,  $\sqrt{n}L_n(\hat{\boldsymbol{\theta}}) = \sqrt{n}G_n(\boldsymbol{\theta}_0) + \Gamma\sqrt{n}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)$  which can be reexpressed as

$$\sqrt{n}\boldsymbol{L}_{n}\left(\hat{\boldsymbol{\theta}}\right) = {}^{d}\sqrt{n}\boldsymbol{G}_{n}\left(\boldsymbol{\theta}_{0}\right) - \boldsymbol{\Gamma}\left(\boldsymbol{\Gamma}'\boldsymbol{W}_{0}\boldsymbol{\Gamma}\right)^{-1}\boldsymbol{\Gamma}'\boldsymbol{W}_{0}\sqrt{n}\boldsymbol{G}_{n}\left(\boldsymbol{\theta}_{0}\right)$$

or equivalently,  $\sqrt{n} \boldsymbol{L}_{n}(\hat{\boldsymbol{\theta}}) = {}^{d} \left( \boldsymbol{I} - \boldsymbol{\Gamma} (\boldsymbol{\Gamma}' \boldsymbol{W}_{0} \boldsymbol{\Gamma})^{-1} \boldsymbol{\Gamma}' \boldsymbol{W}_{0} \right) \sqrt{n} \boldsymbol{G}_{n}(\boldsymbol{\theta}_{0})$  with  $\sqrt{n} \boldsymbol{G}_{n}(\boldsymbol{\theta}_{0}) \stackrel{L}{\longrightarrow} N(0, \boldsymbol{W}_{0}^{-1})$ .

We have

$$\sqrt{n} \boldsymbol{L}_{n} \left( \hat{\boldsymbol{\theta}} \right) \xrightarrow{L} N \left( 0, \boldsymbol{\Sigma} \right),$$

$$\boldsymbol{\Sigma} = \left( \boldsymbol{I} - \boldsymbol{\Gamma} \left( \boldsymbol{\Gamma}' \boldsymbol{W}_{0} \boldsymbol{\Gamma} \right)^{-1} \boldsymbol{\Gamma}' \boldsymbol{W}_{0} \right) \boldsymbol{W}_{0}^{-1} \left( \boldsymbol{I} - \boldsymbol{W}_{0} \boldsymbol{\Gamma} \left( \boldsymbol{\Gamma}' \boldsymbol{W}_{0} \boldsymbol{\Gamma} \right)^{-1} \boldsymbol{\Gamma}' \right) \tag{31}$$

and note that  $\Sigma W_0 = B$  and the trace of the matrix  $B = I - \Gamma (\Gamma' W_0 \Gamma)^{-1} \Gamma' W_0$  is trace(B) = M - m; the rank of the matrix B is also equal to its trace using the techniques as given by Luong and Thompson [16] (p 248-249).

# 4. Estimation and Model Testing Using Complete Data

#### 4.1. Preliminaries

In Section 4.1 and Section 4.2, we shall define a rule of selecting the points  $(s_l, t_l), l = 1, \dots, M$  if complete data are available. Selecting points is equivalent to define the cells used to group the data and we shall see that random cells will be used as the points  $(s_l, t_l), l = 1, \dots, M$  constructed using Quasi-Monte Carlo (QMC) numbers on the unit square multiplied by two chosen sample quantiles from the two marginal distributions will be used. They are random and can be viewed as sample points on the nonnegative quadrant forming an artificial sample. For minimum chi-square methods it appears to be difficult to have a rule to choose cells to group the data, see discussions by Greenwood and Nikulin [8] (p 194-208). We need a few preliminary notions tools and define sample quantiles then statistics can be viewed as functionals of the sample distribution; the notion of influence function is also introduced and this useful tool will be used to find their asymptotic variance of the functional.

We shall define the pth sample quantile of a distribution as we shall need two sample quantiles from the marginal distributions together with QMC numbers to construct an approximation of an integral. Our quadratic distance based on selected points can be viewed as an approximation of a continuous version given by an integral as given by expression (33).

From a bivariate distribution we have two marginal distributions F(x) and G(y). The univariate sample pth quantile of the distribution F(x) assumed to be continuous is based the sample distribution function

 $F_n(x) = \frac{1}{n} \sum_{i=1}^n I[x_i \le x]$  and it is defined to be  $\alpha_p^{(n)} = \inf\{F_n(x) \ge p\}$  and its model counterpart is given by  $\alpha_p = \inf\{F(x) \ge p\}$ . We also use the notation  $\alpha_p^{(n)} = F_n^{-1}(p)$  and  $\alpha_p = F^{-1}(p)$ . We define similarly the qth sample quantile for the distribution G(y) as  $\beta_q^{(n)} = G_n^{-1}(q)$  and its model counterpart  $\beta_q = G^{-1}(q)$  with 0 < p, q < 1.

The sample survival function is defined as

$$\overline{F_n}(x) = \frac{1}{n} \sum_{i=1}^n I[x_i > x] = 1 - F_n(x)$$

The sample quantile functions  $\alpha_p^{(n)}$  or  $\beta_q^{(n)}$  can be viewed as statistical functionals of the form  $T(H_n)$  with  $H_n = F_n$  or  $H_n = G_n$ . The influence function of  $T(H_n)$  is a valuable tool to study the asymptotic properties of the statistical functional and will be introduced below. Let H be the true distribution and  $H_n$  is the usual empirical distribution which estimates H; also let  $\delta_x$  be the degenerate distribution at x, i.e.,  $\delta_x(u) = 1$  if  $u \ge x$  and  $\delta_x(u) = 0$ , oth-

erwise; the influence function of T viewed as a function of x,  $IC_{T,H}(x)$  is defined as a functional directional derivative at H in the direction of  $(\delta_x - H)$ . Letting  $H_{\varepsilon} = H + \varepsilon (\delta_x - H)$ ,  $IC_{T,H}(x)$  is defined as

$$IC_{T,H}(x) = \lim_{\varepsilon \to 0} \frac{T(H_{\varepsilon}) - T(H)}{\varepsilon} = T'_{H}(\delta_{x} - H)$$
 and  $T'_{H}$  is a linear functional.

Alternatively, it is easy to see that  $IC_{T,H}(x) = \frac{\partial H_{\varepsilon}}{\partial \varepsilon}\Big|_{\varepsilon=0}$  and this gives a convenient way to compute the influence function. It can be shown that the influence function of the pth sample quantile  $T(H_n)$  is given by

$$IC_{T,H}(x) = \frac{p-1}{h(H^{-1}(p))}, x < H^{-1}(p) \text{ and } IC_{T,H}(x) = \frac{p}{h(H^{-1}(p))}, x > H^{-1}(p)$$

with h being the density function of the distribution H which is assumed to be absolutely continuous, see Huber [20] (p 56), Hogg  $et\ al.$  [21] (p 593). A statistical functional with bounded influence function is considered to be robust, B-robust and consequently the pth sample quantile is robust. The sample quantiles are robust statistics.

Furthermore, as  $IC_{T,H}\left(x\right)$  is based on a linear functional, the asymptotic variance of  $T\left(H_n\right)$  is simply  $\frac{1}{n}V\left(IC_{T,H}\left(x\right)\right)$  with  $V\left(.\right)$  being the variance of the expression inside the bracket since in general we have  $E\left(IC_{T,H}\left(x\right)\right)=0$  and we have following representation when  $IC_{T,H}\left(x\right)$  is bounded as a function of x,

$$T(H_n) = T(H) + T'_F(H_n - H) + o_p\left(\frac{1}{\sqrt{n}}\right)$$

and 
$$T'_F(H_n-H)=\frac{1}{n}\sum_{i=1}^n T'_F(\delta_{x_i}-H)$$
,

 $T_F'(\delta_{x_i} - H) = IC_{T,H}(x_i)$ , see Hogg *et al.* [21] (p 593). Consequently, in general we have for bounded influence functional with the use of means of central limit theorems (CLT) the following convergence in distribution

$$\sqrt{n}\left(T\left(H_{n}\right)-T\left(H\right)\right) \xrightarrow{L} N\left(0,\sigma_{IC}^{2}\right), \quad \sigma_{IC}^{2}=V\left(IC_{T,H}\left(x\right)\right).$$

The influence function representation of a functional which depends only on one function such as  $H_n$  is the equivalence of a Taylor expansion of a univariate function and the influence function representation of a functional which depends on many functions is the equivalence of a Taylor expansion of a multivariate function with domain in an Euclidean space and having range being the real line. Since we work with marginal survival functions, we define the pth sample quantiles of the marginals survival functions as

$$\overline{F}_n^{-1}(p) = F_n^{-1}(1-p), \overline{G}_n^{-1}(p) = G_n^{-1}(1-p).$$

The influence functions for  $\overline{F}_n^{-1}(p)$  and  $\overline{G}_n^{-1}(p)$  can be derived using the definitions of influence functions or obtained from the influence functions of  $F_n^{-1}(1-p)$  and  $G_n^{-1}(1-p)$ .

Subsequently, we shall introduce the Halton sequences with the bases  $b_1 = 2$  and  $b_2 = 3$  and the first M terms are denoted by

$$(u_l, v_l) = (\varphi_{b_1}(l), \varphi_{b_2}(l)), l = 1, 2, \dots, M.$$

We also use  $H_M$  to denote set of points  $\{(u_l, v_l), l = 1, 2, \dots, M\}$ . The sequence of points belong to the unit square  $(0,1) \times (0,1)$  can be obtained as follows.

For  $b_1 = 2$ , we divide the interval (0,1) into half  $(b_1 = 2)$  then in fourth  $(b_1^2 = 2^2)$  so forth so on to obtain the sequence  $\frac{1}{2}, \frac{1}{4}, \frac{3}{4}, \cdots$ .

For  $b_2 = 3$ , we divide the interval (0,1) into third ( $b_2 = 3$ ) then in ninth ( $b_2^2 = 3^2$ ) so forth so on to obtain the sequence  $\frac{1}{3}, \frac{2}{3}, \frac{1}{9}, \cdots$ . Now pairing them

up we obtain the Halton sequence  $\left(\frac{1}{2},\frac{1}{3}\right), \left(\frac{1}{4},\frac{2}{3}\right), \left(\frac{3}{4},\frac{1}{9}\right), \cdots$ . Matlab and R

have packages to generate the sequences and see Glaserman [22] (p 293-297) for the related pseudo codes; also see the seminal paper by Halton [23]; for the general principles of QMC methods, see Glasserman [22] (p 281-292). The Halton sequence together with two chosen sample quantiles from the two marginal distributions will allow us to choose points to match the bivariate empirical survival function with its model counterpart as we shall have an artificial sample with values on the nonnegative quadrant with the use of two empirical quantiles from the marginal distributions. These points can be viewed as sample points from an artificial sample and since they depend on quantiles which are robust, the artificial sample can be viewed as free of outliers and the methods which make use of them will be robust.

Note that the Halton sequence of numbers are deterministic and useful for approximating an integral, if we would like to compute numerically an integral of the form

 $A = \int_0^1 \int_0^1 \psi(x, y) dxdy$  with  $\psi(x, y)$  being a bivariate function. Using the M terms of the Halton sequence and QMC principles, it can be approximated as

$$A \approx \frac{1}{M} \sum_{l=1}^{M} \psi(s_l, t_l). \tag{32}$$

but if we are used to integration by simulation we might want to think the M terms represent a quasi random sample of size M from a bivariate uniform distribution which is useful for approximating A.

From observations which are given by  $Z_i = (X_i, Y_i)'$ ,  $i = 1, \dots, n$  iid with common bivariate survival distribution S(x, y). Let the two marginal survival functions be denoted by  $\overline{F}(x)$  and  $\overline{G}(y)$  and they are absolutely continuous by assumption; also define the bivariate empirical distribution function which is similar to the bivariate empirical survival function as

$$K_n(x,y) = \frac{1}{n} \sum_{i=1}^n I[x_i \le x, y_i \le y].$$

The two empirical marginal survival functions are defined respectively by

$$\overline{F}_n(x) = \frac{1}{n} \sum_{i=1}^n I[x_i > x]$$
 and  $\overline{G}_n(y) = \frac{1}{n} \sum_{i=1}^n I[y_i > y].$ 

We might want to think that we would like to approximate the following Cramer-Von Mises distance expressed as an integral given by

$$\int_{0}^{\infty} \int_{0}^{\infty} \left( S_{n}(x, y) - \hat{S}_{\theta}(x, y) \right)^{2} dK_{n}(x, y) \tag{33}$$

which is similar to univariate Cramér-Von Mises (CVM) distance and minimizing the distance with respect to  $\theta$  will give the CVM estimator for  $\theta$ , see Luong and Blier-Wong [24] for CVM estimation for example.

In the next section we shall give details on how to form a type of quasi sample or artificial sample of size M using the M terms of the Halton sequence of M terms and the two sample quantiles of the marginal distributions F and G or equivalently using the corresponding empirical function quantiles as discuss earlier and this will allow us to define the sequence  $(s_l,t_l), l=1,\cdots,M$  so that the above integral can be approximated by the following finite sum of the type of an average of M terms

$$\frac{1}{M} \sum_{l=1}^{M} \left( S_n \left( s_l, t_l \right) - \hat{S}_{\boldsymbol{\theta}} \left( s_l, t_l \right) \right)^2. \tag{34}$$

We can see the expression (34) is an unweighted quadratic distance using the identity matrix I as weight matrix instead of  $\widehat{W}_0$ . The unweighted quadratic distance still produces consistent estimators but possibly less efficient estimators than estimators using the quadratic distance with  $\widehat{W}_0$  for large samples and for finite samples the estimators obtained using I might still have reasonable performances and yet being simple to obtained.

The set of points  $(s_l, t_l), l = 1, \dots, M$  is a set of points proposed to be used to form optimum quadratic distances in case that complete data is available. We shall see the set of points depend on two quantiles chosen from the two marginal distributions and they are random consequently. We might want to think that we end up working with random cells.

As for the minimum chi-square methods if random cells stabilize into fixed cells minimum chi-square methods in general have the same efficiency as based on stabilized fixed cells, see Pollard [25] (p 324-326) and Moore and Spruill [26] for the notion of random cells; quadratic distance methods will share the same properties. The chosen points are random but it will be shown that they do stabilize and therefore these random points can be viewed as fixed at stabilized points and despite that they are random, it does not affect efficiencies of the estimators or asymptotic distributions of goodness-of-fit test statistics which make use of them. These properties will be further discussed and studied in more details in the next section along with the introduction of an artificial sample of size M given by the points  $(s_l, t_l), l = 1, \dots, M$  on the nonegative quadrant which give us a guideline on how to choose points if complete data is available.

# 4.2. Halton Sequences and an Artificial Sample

From the *M* terms of the Halton sequences, we have  $(u_l, v_l), l = 1, \dots, M$ .

Let 
$$\eta = \frac{1}{\max(u_l, l = 1, \dots, M)}$$
 and  $\varrho = \frac{1}{\max(v_l, l = 1, \dots, M)}$ , we can form the

artificial sample with elements given by  $(s_l, t_l), l = 1, \dots, M$  with  $s_l = \eta u_l \overline{F}_n^{-1}(p), t_l = \varrho v_l \overline{G}_n^{-1}(p)$  with  $0.01 \le p \le 0.05$ . Note that we have the following relationships between empirical quantile based on distributions and survival functions with  $\overline{F}_n^{-1}(0.01) = F_n^{-1}(0.99)$  and  $\overline{G}_n^{-1}(0.01) = G_n^{-1}(0.99)$ .

We can view  $(s_l,t_l), l=1,\cdots,M$  being a form of quasi random sample on the nonnegative quadrant and these are the points proposed to be used in case of complete data is available. In general, we might want to choose  $20 \le M \le 35$  if  $M^2 \le n$  and if n is small we try to ensure  $M \le \sqrt{n}$ . Consequently as  $n \to \infty$ , M remains bounded. If M > 35, there might be difficulty to obtain the matrix  $\hat{W}_0$  as  $\hat{\Omega}_0$  might be nearly singular. In practice we tend to replace  $\hat{W}_0$  by a near optimum matrix  $\hat{W}_0$  obtained from  $\hat{\Omega}_0$  by regularizing the eigenvalues of  $\hat{\Omega}_0$  which might not be stable which causes the matrix to be nearly singular hence  $\hat{W}_0$  will not be available; see Section 5.1 for more discussions on these issues.

Since 
$$\overline{F}_n^{-1}(p) \xrightarrow{p} \overline{F}^{-1}(p)$$
 and  $\overline{G}_n^{-1}(p) \xrightarrow{p} \overline{G}^{-1}(p)$ ,  $(s_l, t_l) \xrightarrow{p} (s_l^0, t_l^0)$  with  $s_l^0 = \eta u_l \overline{F}^{-1}(p)$  and  $t_l^0 = \varrho v_l \overline{G}^{-1}(p)$  for  $l = 1, \dots, M$  and the points  $(s_l^0, t_l^0), l = 1, \dots, M$  are non-random or fixed.

It turns out that quadratic distances for both versions constructed with the points  $(s_l,t_l), l=1,\cdots,M$  are asymptotic equivalent to quadratic distances using the points  $(s_l^0,t_l^0), l=1,\cdots,M$  so that asymptotic theory developed using the points  $(s_l,t_l), l=1,\cdots,M$  considered to be fixed continue to be valid; we shall show indeed this is the case. Similar conclusions have been established for the minimum chi-square methods with the use of random cells provide that these cells stabilize to fixed cells, see Theorem 2 given by Pollard [25] (p 324-326). We shall define a few notations to make the arguments easier to follow.

Define 
$$\{(s,t)\}=\{(s_l,t_l),l=1,\cdots,M\}$$
 and similarly let 
$$\{(s^0,t^0)\}=\{(s_l^0,t_l^0),l=1,\cdots,M\}.$$

We work with the quadratic distance defined using  $\{(s,t)\}$  which leads to consider quadratic of the form  $\|G_n(\boldsymbol{\beta})\|^2$ . Now to emphasize  $\widehat{z_n}$  and  $\widehat{z_{\boldsymbol{\theta}}}$  which depend on  $\{(s,t)\}$ , we also use respectively the notations  $\widehat{z_n} = \widehat{z_n} \left(\{(s,t)\}\right)$  and  $\widehat{z_{\boldsymbol{\theta}}} = \widehat{z_{\boldsymbol{\theta}}} \left(\{(s,t)\}\right)$  and define

$$\widehat{\boldsymbol{z}}_{n}^{0} = \widehat{\boldsymbol{z}}_{n} \left( \left\{ \left( \boldsymbol{s}^{0}, \boldsymbol{t}^{0} \right) \right\} \right), \, \widehat{\boldsymbol{z}}_{\boldsymbol{\theta}}^{0} = \widehat{\boldsymbol{z}_{\boldsymbol{\theta}}} \left( \left\{ \left( \boldsymbol{s}^{0}, \boldsymbol{t}^{0} \right) \right\} \right).$$

It suffices to verify that results of Theorem 1, Theorem 2 and its corollary in Section 3 continue to hold.

Observe that we have

$$\left(\widehat{z_n} - \widehat{z_{\theta}}\right) \xrightarrow{p} G(\theta) \tag{35}$$

and

$$\left(\hat{\boldsymbol{z}}_{n}^{0} - \hat{\boldsymbol{z}}_{\boldsymbol{\theta}}^{0}\right) \xrightarrow{p} \boldsymbol{G}(\boldsymbol{\theta}) \tag{36}$$

$$\boldsymbol{G}(\boldsymbol{\theta}) = \left(S\left(s_{1}^{0}, t_{1}^{0}\right) - C_{\boldsymbol{\theta}}\left(\overline{F}\left(s_{1}^{0}\right), \overline{G}\left(t_{1}^{0}\right)\right), \cdots, S\left(s_{M}^{0}, t_{M}^{0}\right) - C_{\boldsymbol{\theta}}\left(\overline{F}\left(s_{M}^{0}\right), \overline{G}\left(t_{M}^{0}\right)\right)\right)'.$$

This also means that we have the same limit in probability for  $(\widehat{z_n} - \widehat{z_{\theta}})$  and  $(\widehat{z_n} - \widehat{z_{\theta}})$  as we have  $\{(s,t)\} \xrightarrow{p} \{(s^0,t^0)\}$  and  $S_n(x,y) \xrightarrow{p} S(x,y)$ .

Clearly, 
$$\widehat{W_0}(\{(s,t)\}) \xrightarrow{p} W_0(s^0,t^0)$$
.

It remains to establish  $\sqrt{n}\left(\widehat{z_n} - \widehat{z_{\theta}}\right) = \sqrt{n}\left(\widehat{z}_n^0 - \widehat{z_{\theta}}\right) + o_p\left(1\right)$ .

Using results on the influence functions representations for functionals as discussed, it suffices to show that the vector  $(\widehat{z_n} - \widehat{z_\theta})$  has the same influence representation as the vector  $(\widehat{z_n}^0 - \widehat{z_\theta}^0)$  to conclude that all the asymptotic results are valid even  $\{(s,t)\}$  are random.

We shall derive the influence functions for elements of the vector of functional  $(\widehat{z_n} - \widehat{z_\theta})$  and show that they are the same for the corresponding elements of the vector of functional  $(\widehat{z}_n^0 - \widehat{z_\theta}^0)$ . Let S(x, y) be the true bivariate survival function and under the parametric model being considered,

 $S(x,y) = C_{\theta_0}(\overline{F}(x), \overline{G}(y))$  and we also use the notation  $C(u,v) = C_{\theta_0}(u,v)$ . Let  $\delta_{x,y}^S(u,v)$  be the degenerate bivariate survival function at the point (x,y), i.e.,  $\delta_{x,y}^S(u,v) = 1$  if u < x and v < y and  $\delta_{x,y}^S(u,v) = 0$ , otherwise.

Let the degenerate survival function at x be defined as  $\delta_x^S(u)=1$  if x>u and  $\delta_x^S(u)=0$ , otherwise. Similarly, let the degenerate survival function at y be defined as  $\delta_y^S(v)=1$  if y>v and  $\delta_y(v)=0$ , otherwise. Now we can define the following contaminated bivariate survival and marginal survival functions,

$$S_{\varepsilon}(u,v) = S(u,v) + \varepsilon \left(\delta_{x,y}^{S}(u,v) - S(u,v)\right), 0 \le \varepsilon \le 1$$

which is a contaminated bivariate survival function and

$$\overline{F}_{\varepsilon_{1}}\left(u\right) = \overline{F}\left(u\right) + \varepsilon_{1}\left(\delta_{x}^{S}\left(u\right) - \overline{F}\left(u\right)\right), 0 \le \varepsilon_{1} \le 1.$$

Similarly for the marginals,

$$\overline{G}_{\varepsilon_2}(v) = \overline{G}(v) + \varepsilon_2(\delta_y^S(v) - \overline{G}(v)), 0 \le \varepsilon_2 \le 1.$$

Now, we consider  $(\hat{z}_{jn} - \hat{z}_{j\theta_0})$  the jth element of  $(\hat{z}_n - \hat{z}_{\theta_0})$ ,  $(\hat{z}_{jn} - \hat{z}_{j\theta_0}) = T_j(S_n, \overline{F}_n, \overline{G}_n)$ ,  $j = 1, \dots, M$  with each

$$T_{j}\left(S_{n},\overline{F}_{n},\overline{G}_{n}\right)=S_{n}\left(S_{j}\left(\overline{F}_{n}\right),t_{j}\left(\overline{G}_{n}\right)\right)-C_{\theta_{0}}\left(\overline{F}_{n}\left(S_{j}\left(\overline{F}_{n}\right)\right),\overline{G}_{n}\left(t_{j}\left(G_{n}\right)\right)\right),j=1,\cdots,M.$$

Clearly,  $T_j(S_n, \overline{F}_n, \overline{G}_n)$  depend on  $S_n, \overline{F}_n, \overline{G}_n$  and  $T_j(S_\varepsilon, \overline{F}_\varepsilon, \overline{G}_\varepsilon)$ 

$$=S_{\varepsilon}\left(s_{j}\left(\overline{F}_{\varepsilon_{1}}\right),t_{j}\left(\overline{G}_{\varepsilon_{2}}\right)\right)-C_{\theta_{0}}\left(\overline{F}_{\varepsilon_{1}}\left(s_{j}\left(\overline{F}_{\varepsilon_{1}}\right)\right),\overline{G}_{n}\left(t_{j}\left(\overline{G}_{\varepsilon_{2}}\right)\right)\right),j=1,\cdots,M,$$

but we can use the influence function representation of  $T_j(S_n, \overline{F}_n, \overline{G}_n)$ , a technique proposed by Reid [29] (p 80-81) but in this case it will need three influence functions which are given by

$$\left. \frac{\partial T_{j}\left(S_{\varepsilon}, \overline{F}_{\varepsilon_{1}} \overline{G_{\varepsilon_{2}}}\right)}{\partial \varepsilon} \right|_{\varepsilon=\varepsilon_{1}=\varepsilon_{2}=0} = I\left[x > s_{j}^{0}, y > t_{j}^{0}\right] - S\left(s_{j}^{0}, t_{j}^{0}\right)$$

which is bounded with respect to (x, y),

$$\begin{split} & \frac{\partial T_{j}\left(S_{\varepsilon}, \overline{F}_{\varepsilon_{1}}, \overline{G_{\varepsilon_{2}}}\right)}{\partial \varepsilon_{1}} \Bigg|_{\varepsilon=\varepsilon_{1}=\varepsilon_{2}=0} \\ & = \frac{\partial S\left(s_{j}^{0}, t_{j}^{0}\right)}{\partial s} \frac{\partial s_{j}\left(\overline{F}_{\varepsilon_{1}}\right)}{\partial \varepsilon_{1}} \Bigg|_{\varepsilon_{1}=0} - \frac{\partial C_{\boldsymbol{\theta}_{0}}\left(\overline{F}\left(s_{j}^{0}\right), \overline{G}\left(t_{j}^{0}\right)\right)}{\partial u} \left(\frac{\partial \overline{F}\left(s_{j}^{0}\right)}{\partial s} \frac{\partial s_{j}\left(\overline{F}_{\varepsilon_{1}}\right)}{\partial \varepsilon_{1}} \Bigg|_{\varepsilon_{1}=0}\right) \\ & - \frac{\partial C_{\boldsymbol{\theta}_{0}}\left(\overline{F}\left(s_{j}^{0}\right), \overline{G}\left(t_{j}^{0}\right)\right)}{\partial u} \left(\delta_{x}^{S}\left(s_{j}^{0}\right) - \overline{F}\left(s_{j}^{0}\right)\right) \end{split}$$

and the expression is reduced to

$$\left.\frac{\partial T_{j}\left(S_{\varepsilon},\overline{F}_{\varepsilon_{1}},\overline{G_{\varepsilon_{2}}}\right)}{\partial\varepsilon_{1}}\right|_{\varepsilon=\varepsilon_{1}=\varepsilon_{2}=0}=-\frac{\partial C_{\boldsymbol{\theta}_{0}}\left(\overline{F}\left(s_{j}^{0}\right),\overline{G}\left(t_{j}^{0}\right)\right)}{\partial u}\left(\delta_{x}^{S}\left(s_{j}^{0}\right)-\overline{F}\left(s_{j}^{0}\right)\right)$$

by noting the first two terms of the RHS of the above expression cancel each other since we have  $S(s,t) = C_{\theta_0}(\overline{F}(s), \overline{G}(t))$  which implies

$$\frac{\partial S(s,t)}{\partial s} = \frac{\partial C_{\theta_0}(\bar{F}(s),\bar{G}(t))}{\partial u} \frac{\partial \bar{F}(s)}{\partial s}.$$

Similarly,

$$\left.\frac{\partial T_{j}\left(S_{\varepsilon},\overline{F}_{\varepsilon_{1}},\overline{G_{\varepsilon_{2}}}\right)}{\partial\varepsilon_{2}}\right|_{\varepsilon=\varepsilon_{1}=\varepsilon_{2}=0}=-\frac{\partial C_{\boldsymbol{\theta}_{0}}\left(\overline{F}\left(s_{j}^{0}\right),\overline{G}\left(t_{j}^{0}\right)\right)}{\partial\boldsymbol{v}}\left(\boldsymbol{\delta}_{\boldsymbol{y}}^{S}\left(t_{j}^{0}\right)-\overline{G}\left(t_{j}^{0}\right)\right)$$

If we compare with the corresponding jth term of  $(\hat{z}_n^0 - \hat{z}_{\theta_0}^0)$  given by the functional  $G_j(S_n, \overline{F}_n, \overline{G}_n) = S_n(s_l^0, t_l^0) - C_{\theta_0}(\overline{F}_n(s_l^0), \overline{G}_n(t_l^0))$ , we can verify the functional  $G_j(S_n, \overline{F}_n, \overline{G}_n)$  has the same influence functions as the functional  $T_j(S_\varepsilon, \overline{F}_{\varepsilon_1}, \overline{G_{\varepsilon_2}})$ . It is not difficult to see that we have the equalities

$$\begin{split} \frac{\partial G_{j}\left(S_{\varepsilon},\overline{F}_{\varepsilon_{1}}\overline{G_{\varepsilon_{2}}}\right)}{\partial \varepsilon} \Bigg|_{\varepsilon=\varepsilon_{1}=\varepsilon_{2}=0} &= \frac{\partial T_{j}\left(S_{\varepsilon},\overline{F}_{\varepsilon_{1}}\overline{G_{\varepsilon_{2}}}\right)}{\partial \varepsilon} \Bigg|_{\varepsilon=\varepsilon_{1}=\varepsilon_{2}=0}, \\ \frac{\partial G_{j}\left(S_{\varepsilon},\overline{F}_{\varepsilon_{1}}\overline{G_{\varepsilon_{2}}}\right)}{\partial \varepsilon_{1}} \Bigg|_{\varepsilon=\varepsilon_{1}=\varepsilon_{2}=0} &= \frac{\partial T_{j}\left(S_{\varepsilon},\overline{F}_{\varepsilon_{1}}\overline{G_{\varepsilon_{2}}}\right)}{\partial \varepsilon_{1}} \Bigg|_{\varepsilon=\varepsilon_{1}=\varepsilon_{2}=0}, \\ \frac{\partial G_{j}\left(S_{\varepsilon},\overline{F}_{\varepsilon_{1}}\overline{G_{\varepsilon_{2}}}\right)}{\partial \varepsilon_{2}} \Bigg|_{\varepsilon=\varepsilon_{1}=\varepsilon_{2}=0} &= \frac{\partial T_{j}\left(S_{\varepsilon},\overline{F}_{\varepsilon_{1}}\overline{G_{\varepsilon_{2}}}\right)}{\partial \varepsilon_{2}} \Bigg|_{\varepsilon=\varepsilon_{1}=\varepsilon_{2}=0}, j=1,\cdots,M. \end{split}$$

Therefore, all the asymptotic results of Section 3 remain valid and all these in-

fluence functions are bounded so that inference methods making use of these functionals are robust in general. Furthermore, we can consider the inference procedures based on quadratic distances as we have non-random points  $\{(s^0,t^0)\}$  if they can be replaced by  $\{(s,t)\}$  without affecting the asymptotic results already established in Section 3. For more discussions on random cells and influence function techniques for minimum chi-square methods and related quadratic distance methods, see Luong [27].

# 5. Numerical Issues and a Limited Study

#### 5.1. Numerical Issues

In this section we shall consider the numerical problem of not being able to obtain the matrix  $\hat{W}_0$  as  $\hat{\Omega}_0$  might be nearly singular and we need to replace  $\hat{W}_0$  by a near optimum matrix  $\tilde{W}_0$  obtained from  $\hat{\Omega}_0$ . The techniques of regularizing a matrix have been introduced by Carrasco and Florens [28] (p 809-810) for GMM estimation with continuum moment conditions, MQD methods can be viewed as similar to GMM with a finite number of moment conditions and clearly the techniques can also be applied for MQD methods. We use the spectral decomposition of  $\hat{\Omega}_0$  to obtain its eigenvalues and eigenvectors, see Hogg *et al.* [21] (p 179) for the spectral decomposition of a symmetric positive definite matrix which allows us to express

$$\hat{\mathbf{\Omega}}_0 = \sum_{i=1}^M \lambda_i \mathbf{v}_i \mathbf{v}_i'$$

where the  $\lambda_i's$  are positive eigenvalues with corresponding eigenvectors given by the  $v_i's$  of the matrix  $\hat{\Omega}_0$ . Now, observe that

$$\hat{\boldsymbol{W}}_0 = \hat{\boldsymbol{\Omega}}_0^{-1} = \sum_{i=1}^M \frac{1}{\lambda_i} \boldsymbol{v}_i \boldsymbol{v}_i'$$

is not obtainable numerically. It is due to the eigenvalues which are not stable, the regularization of  $\hat{\Omega}_0$  will lead to the following matrix which hopefully is obtainable and approximate  $\hat{W}_0$ . It consists of perturbing the  $\lambda_i's$  by a small positive number a and define the approximate optimum matrix as

$$\tilde{W}_0 = \sum_{i=1}^M \frac{|\lambda_i|}{\lambda_i^2 + a} v_i v_i', a > 0.$$

Carrasco and Florens [28] (p 809-810) for GMM estimation with continuum moment conditions have shown that asymptotic theory remains unchanged if  $a \to 0$  at a suitable rate as  $n \to \infty$ . This condition is difficult to verify in practice. However, we might want to continue to use the asymptotic theory in an approximate sense, *i.e.*, we can replace  $\hat{W}_0$  by  $\tilde{W}_0$  and view such a replacement does not modify the asymptotic theory in practice.

A more rigorous approach to justify the chi-square distribution for goodness of fit tests is to divide into 2 steps, first using  $\tilde{W}_0$  to construct the distance for estimation and letting  $\hat{\theta}$  be the vector which minimizes

$$(\hat{z}_n - \hat{z}_{\theta})' \tilde{W}_0 (\hat{z}_n - \hat{z}_{\theta}).$$

Using Equation (31) we have

$$\sqrt{n} \boldsymbol{L}_{n} \left( \hat{\boldsymbol{\theta}} \right) \xrightarrow{L} N(\boldsymbol{0}, \boldsymbol{\Sigma}),$$

$$\hat{\boldsymbol{\Sigma}} = \left( \boldsymbol{I} - \hat{\boldsymbol{\Gamma}} \left( \hat{\boldsymbol{\Gamma}}' \tilde{\boldsymbol{W}}_{0} \hat{\boldsymbol{\Gamma}} \right)^{-1} \hat{\boldsymbol{\Gamma}}' \tilde{\boldsymbol{W}}_{0} \right) \hat{\boldsymbol{\Omega}}_{0} \left( \boldsymbol{I} - \hat{\boldsymbol{\Gamma}} \left( \hat{\boldsymbol{\Gamma}}' \tilde{\boldsymbol{W}}_{0} \hat{\boldsymbol{\Gamma}} \right)^{-1} \hat{\boldsymbol{\Gamma}}' \tilde{\boldsymbol{W}}_{0} \right)^{\prime},$$

also see expression (3.4.2) given by Luong and Thompson [16] (p 248). The matrices  $\hat{\Sigma}$  and  $\hat{\Gamma}$  are respectively consistent estimates of  $\Sigma$  and  $\Gamma$ .

It suffices to find the Moore-Penrose  $\hat{\Sigma}^-$  generalized inverse of  $\hat{\Sigma}$  and construct the test statistics as

$$n(\hat{z}_n - \hat{z}_{\hat{\theta}})' \hat{\Sigma}^-(\hat{z}_n - \hat{z}_{\hat{\theta}}).$$

The asymptotic distribution of the test statistics will be again chi-square with M-m degree of freedom using distribution theory for quadratic forms, see Luong and Thompson [16] (p 247) for example and for generalized inverses, see Harville [29] (p 493-514).

Note that if  $\hat{W}_0$  can be used for estimation then we can let  $\hat{\Sigma}^- = \hat{W}_0$ , *i.e.* there is no need to use two quadratic distances separately.

#### 5.2. A Limited Simulation Study

For the study, we fix the number of points M=25. The two samples quantiles are 0.99 quantiles or 0.01 survival functions quantiles if marginal empirical survival functions are used instead of distribution functions for estimation without construction of goodness-of-fit tests. The points used are constructed using the procedures given in Section 4.2. We consider the one parameter MO copula model with

$$C_{\theta}(u,v) = u^{1-\theta}v$$
 if  $u \ge v$  and  $C_{\theta}(u,v) = uv^{1-\theta}$  if  $u \le v$ . (37)

 $C_{\theta}(u,v)$  is differentiable with respect to  $\theta$  and  $C_{\theta}(u,v)$  is singular if u=v and  $\theta \in (0,1)$  see Dobrowolski and Kumar [10] (p 2). For this model, the model

Spearman rho 
$$\rho_{SP} = \frac{3\theta^2}{4\theta - \theta^2} = \frac{3\theta}{4 - \theta}$$
, see Dobrowolski and Kumar [10] (p 5).

The sample Spearman rho  $\rho_{SP}$  is simply the Pearson correlation coefficient but computed using ranks of the observations from the two empirical marginal distributions, see Conover [30] (p 314-318).

If complete data are available, equating  $\widehat{\rho_{SP}} = \rho_{SP}$  gives the moment estimator  $\widetilde{\theta} = \frac{4\widehat{\rho_{SP}}}{3+\widehat{\rho_{SP}}}$  and one might expect that the moment estimator has reasonable

efficiency as we only has one parameter in this model and the estimate is based on ranks.

The moment estimate can be used to compute  $\hat{\Omega}_0^{-1} = \widehat{W}_0$  which is needed for chi-square tests and for estimation using quadratic distances. We use M=25 and there is no problem on inverting the matrix  $\hat{\Omega}_0$ . Clearly if data is already

grouped we can use the unweighted quadratic distance to provide a consistent preliminary estimate for  $\theta_0$ . The efficient MQD estimator is denoted by  $\hat{\theta}$ . In the simulation study since we have so many marginal survival functions which can be used so we decide to draw observations directly from the Copula Models. This is not what happens in real life situation but we want to test the procedures. We do not have the computing resources for a large scale study and try various marginal survival functions. More works need to be done but we want to illustrate the procedures.

We use sample size n=2000 and the number of samples used is N=100. For comparison of MQD estimator  $\hat{\theta}$  versus Methods of moment (MM) estimator  $\tilde{\theta}$  we use the ratio of relative efficiency

$$ARE = \frac{MSE(\hat{\theta})}{MSE(\tilde{\theta})}$$

where the mean square error of an estimator  $\hat{\pi}$  for  $\pi_0$  is defined as

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

which can be estimated using M samples each of size n.

The unweighted QD estimator is denoted by  $\widehat{\theta}_I$  as the identity matrix I is used for the unweighted quadratic distance. The corresponding

$$ARE = \frac{MSE\left(\widehat{\theta_I}\right)}{MSE\left(\widetilde{\theta}\right)}$$

can similarly be used for comparison and it can be estimated using simulated samples.

The range of parameter being considered is  $\theta = 0.1, 0.2, \cdots, 0.9$ , the results are summarized using the first table of **Table 1** where we find that the MM estimator and the two quadratic distance estimators have practically equal efficiency up to 4 or 5 decimal precisions.

To study the size of the chi-square tests and the power of the tests let  $H_0$ : The MO copula model  $C^{MO}$  with  $C_{\theta}(u,v)$  as given by expression (37) and  $\theta = \frac{1}{2}$ . With  $\theta = \frac{1}{2}$ ,  $\rho_{SP} = \frac{3}{7}$ . Observations are drawn from the model specified by by  $H_a$  which specifies the model is a contaminated one given by

$$(1-\lambda)C^{MO}(u,v) + \lambda C^{Gaussian}(u,v), 0 < \lambda < 1.$$

 $C^{MO}(u,v)$  is as defined earlier,  $C^{Gaussian}(u,v)$  is the Gaussian copula defined as

$$C_{\rho}(u,v) = \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}(x^2+y^2-2xy)\right\} dxdy$$

with  $\rho = 0.5$ .

Procedures to simulate from Gaussian and MO copulas are given in chapter 6 by Ross [13] (p 97-108). We use M = 25 and M = 35, the sample size

**Table 1.** Asymptotic relative efficiencies comparisons for MQD estimators versus MM estimator using N = 1000 samples of size n = 1000 for the one parameter MO copula Model.

(a)										
$ARE$ . $\theta$	$\frac{1}{6}$	1/5	$\frac{1}{4}$	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{2}{3}$	$\frac{4}{5}$			
$\frac{\mathit{MSE}\left(\hat{\theta}\right)}{\mathit{MSE}\left(\hat{\theta}\right)}$	0.9999	0.9999	0.9999	0.999	0.9999	0.9999	0.9999			
$\frac{MSE\left(\hat{\theta}_{_{I}}\right)}{MSE\left(\hat{\theta}\right)}$	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999			

Power study using M = 25 points, n = 3000 and the alternative hypothesis specified as the contaminated model  $(1 - \lambda)C^{MO}(u, v) + \lambda C^{Gaussian}(u, v)$ ,  $0 < \lambda < 1$ .

	(b)									
· <i>\lambda</i>	0	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
Mean of the chi-square statistics	13.96	20.89	21.01	22.19	19.67	16.43	22.35	55.72	102.79	338.95
SD of the chi-square statistics	2.48	5.04	9.65	10.89	9.63	10.11	9.37	39.47	45.49	184.10

Critical point for the test using the 95th percentile of a chi-square distribution,  $\chi^2_{0.95}(24) = 36.41$ . Power study using M = 35 points, n = 3000 and the alternative hypothesis specified as the contaminated model  $(1-\lambda)C^{MO}(u,v) + \lambda C^{Giaussian}(u,v)$ ,  $0 < \lambda < 1$ .

	(C)									
·λ	0	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
Mean of the										
chi-square	32	761	2332	5045	9289	15284	23711	34981	48675	66843
statistics										
SD of the										
chi-square	5.48	36.94	66.10	82.25	99.29	112.74	219.00	88.84	85.02	26.72
statistics										

Critical point for the test using the 95th percentile of a chi-square distribution,  $\chi^2_{0.95}(34) = 48.60$ .

n=3000 and we use N=30. Dobric and Schmid [6] (p 1060-1061) in their study have used n=2500 and their chi-square tests have around 70 degrees of freedom. With M=35 only occasionally that  $\hat{\Omega}_0$  is nearly singular, if this happens we discard the sample. We do not have resources for larger scale study; each run takes around three minutes to complete. As most of the time we are drawing observations using an alternative model but for testing we must estimate the parameter  $\theta$  of the MO model, the algorithm tends to take time to converge. The study is very limited as the number of simulated samples is small with N=30 and only a few copula models are considered but it seems to point to the potential uses of MQD chi-square tests. The tests especially with M=35 seem to have power especially along some directions which can be represented as a mixture type of models as shown by the means and standard deviations of

the chi-square statistics as displayed in the second and third table of **Table 1**. More simulation works are needed to assess the power of the MQD tests using various copula models. There are not many statistical procedures for copula models using data that have been already grouped. MQD methods might be useful for this type of situation.

#### 6. Conclusions

Minimum Quadratic Distance Methods (MQD) offer a unified for estimation and model testing using grouped data under the form of a contingency table for parametric copula models without having to assume parametric models for the marginal distributions. The methods share with minimum chi-square methods by having a unique asymptotic distribution across the composite hypothesis for testing which make the implementations relatively simple without requiring extensive simulations for approximating the null asymptotic distribution. It is shown in this paper that if complete data are available, a rule to define points based on QMC numbers can be proposed to alleviate the arbitrariness on the choice of points to construct quadratic distances. The rule will also make quadratic distances close to Cramer-Von Mises distances. It is well known that in one dimension, chi-square tests cannot be consistent against all alternatives but if the intervals are chosen properly the tests still can have good power against some form of alternatives considered to be useful for applications.

MQD tests statistics with the rule of choosing points might preserve the same properties and by being relative simple to implement, they can be useful for applied works. More numerical and simulation works are needed for further study the power of the MQD tests.

# Acknowledgements

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

#### References

- Partrat, C. (1995) Compound Model for Two Dependent Kinds of Claims. *Insurance: Mathematics and Economics*, 15, 219-231.
   <a href="https://doi.org/10.1016/0167-6687(94)90796-X">https://doi.org/10.1016/0167-6687(94)90796-X</a>
- [2] Gibbons, J.D. and Chakraborti, S. (2011) Nonparametric Statistical Inference. 5th Edition, CRC Press, Boca Raton. https://doi.org/10.1007/978-3-642-04898-2\_420
- [3] Klugman, S.A. and Parsa, A. (1999) Fitting Bivariate Distributions with Copulas. *Insurance. Mathematics and Economics*, 24, 139-148. <a href="https://doi.org/10.1016/S0167-6687(98)00039-0">https://doi.org/10.1016/S0167-6687(98)00039-0</a>
- [4] Dobric, J. and Schmid, F. (2005) Testing Goodness of Fit for Parametric Families of Copulas—Applications to Financial Data. *Communications in Statistics: Simulation and Computation*, **34**, 1053-1068. <a href="https://doi.org/10.1080/03610910500308685">https://doi.org/10.1080/03610910500308685</a>
- [5] Klugman, S.A., Panjer, H.H. and Willmot, G.E. (2012) Loss Models: From Data to

- Decisions. Wiley, New York.
- [6] Shih, J.H. and Louis, T.A. (1995) Inferences on the Association Parameter in Copula Models for Bivariate Survival Data. *Biometrics*, 51, 1384-1399. https://doi.org/10.2307/2533269
- [7] Genest, C., Rémillard, B. and Beaudoin, D. (2009) Goodness-of-Fit Tests for Copulas: A Review and Power Study. *Insurance: Mathematics and Economics*, 44, 199-213. https://doi.org/10.1016/j.insmatheco.2007.10.005
- [8] Greenwood, P. and Nikulin, M.S. (1996) A Guide to Chi-Square Testing. Wiley, New York.
- [9] Lehmann, E.L. (1999) Elements of Large Sample Theory. Springer, New York. https://doi.org/10.1007/b98855
- [10] Dobrowolski, E. and Kumar, P. (2014) Some Properties of the Marshall-Olkin and Generalized Cuadras-Augé Families of Copulas. The Australian Journal of Mathematical Analysis and Applications, 11, 1-13.
- [11] Marshall, A.W. and Olkin, I. (1967) A Multivariate Exponential Distribution. *Journal of the American Statistical Association*, 62, 30-44. <a href="https://doi.org/10.1080/01621459.1967.10482885">https://doi.org/10.1080/01621459.1967.10482885</a>
- [12] Mai, J.-F. and Scherer, M. (2012) Simulating Copulas: Stochastic Models, Sampling Algorithms and Applications. Imperial College Press, London. https://doi.org/10.1142/p842
- [13] Ross, S.M. (2012) Simulations. 5th Edition, Elsevier, New York.
- [14] Klugman, S.A., Panjer, H.H. and Willmot, G.E. (2012) Loss Models: Further Topics. Wiley, New York.
- [15] Frees, E.W. and Valdez, E.A. (1998) Understanding Relationships Using Copulas. *North American Actuarial Journal*, **2**, 1-25.
- [16] Luong, A. and Thompson, M.E. (1987) Minimum-Distance Methods Based on Quadratic Distances for Transforms. *Canadian Journal of Statistics*, 15, 239-251. https://doi.org/10.2307/3314914
- [17] Reid, N. (1981) Influence Functions for Censored Data. *Annals of Statistics*, **9**, 78-92. https://doi.org/10.1214/aos/1176345334
- [18] Pakes, A. and Pollard, D. (1989) Simulation Asymptotic of Optimization Estimators. *Econometrica*, **57**, 1027-1057. https://doi.org/10.2307/1913622
- [19] Luong, A., Bilodeau, C. and Blier-Wong, C. (2018) Simulated Minimum Hellinger Distance Inference Methods for Count Data. *Open Journal of Statistics*, 8, 187-219. https://doi.org/10.4236/ojs.2018.81012
- [20] Huber, P. (1981) Robust Statistics. Wiley, New York. https://doi.org/10.1002/0471725250
- [21] Hogg, R.V, McKean, J.W. and Craig, A.T. (2013) Introduction to Mathematical Statistics. Pearson, New York.
- [22] Glasserman, P. (2003) Monte Carlo Methods in Financial Engineering. Springer, New York. <a href="https://doi.org/10.1007/978-0-387-21617-1">https://doi.org/10.1007/978-0-387-21617-1</a>
- [23] Halton, J. (1960) On the Efficiency of Certain Quasi Random Sequences of Points in evaluating Multi-Dimensional Integrals. *Numerische Mathematik*, 2, 84-90. <a href="https://doi.org/10.1007/BF01386213">https://doi.org/10.1007/BF01386213</a>
- [24] Luong, A. and Blier-Wong, C. (2017) Simulated Minimum Cramér-Von Mises Distance Estimation for Some Actuarial and Financial Models. *Open Journal of Statistics*, 7, 815-833. <a href="https://doi.org/10.4236/ojs.2017.75058">https://doi.org/10.4236/ojs.2017.75058</a>

- [25] Pollard, D. (1979) General Chi-Square Goodness-of-Fit Tests with Data Dependent Cells. *Probability and Related Fields*, 50, 317-331. <a href="https://doi.org/10.1007/BF00534153">https://doi.org/10.1007/BF00534153</a>
- [26] Moore, D.S. and Spruill, M.C. (1975) Unified Large-Sample Theory of General Chi-Squared Statistics for Tests of Fit. *Annals of Statistics*, 3, 599-616. https://doi.org/10.1214/aos/1176343125
- [27] Luong, A. (2018) Simulated Quadratic Distance Methods Using Grouped Data for Some Bivariate Continuous Models. *Open Journal of Statistics*, **8**, Article ID: 84193. https://doi.org/10.4236/ojs.2018.82024
- [28] Carrasco, M. and Florens, J.-P. (2000) Generalization of GMM to a Continuum of Moment Conditions. *Econometric Theory*, 16, 797-834. https://doi.org/10.1017/S0266466600166010
- [29] Harville, D.A. (1997) Matrix Algebra from a Statistician Perspective. Springer, New York. https://doi.org/10.1007/b98818
- [30] Conover, W.J. (1999) Practical Nonparametric Statistics. Wiley, New York.

# **Appendices**

## Technical Appendix 1 (TA1)

In this technical appendix, we shall consider influence function representation for the vector of functionals  $(\hat{z}_n - \hat{z}_{\theta_0})$  to justify the expression.  $\Omega_0(i,j)$  is as given by expression (9) in Section 3.2.

Let 
$$(\hat{z}_n - \hat{z}_{\theta_0}) = (T^{(1)}(S_n, \overline{F_n}, \overline{G_n}), \dots, T^{(M)}(S_n, \overline{F_n}, \overline{G_n}))'$$
,

consider the *I*-th element of  $(\hat{z}_n - \hat{z}_{\theta_0})$ , it is given by

$$T^{(l)}\left(S_{n},\overline{F_{n}},\overline{G_{n}}\right)=S_{n}\left(s_{l},t_{l}\right)-C_{\theta_{0}}\left(\overline{F}_{n}\left(s_{l}\right),\overline{G}_{n}\left(t_{l}\right)\right)$$

which is a functional which depend on three functions  $S_n, \overline{F_n}, \overline{G_n}$  but we still can applied the techniques given by Reid [17] (p 80) to have an influence representation of the functional. Since it depends on three functions we shall have three coresponding influence functions. Let  $S_\varepsilon = S + \varepsilon \left( \delta_{x,y}^S - S \right)$  with  $\delta_{x,y}^S \left( u,v \right) = 1$  if u < x and v < y and  $\delta_{x,y}^S \left( u,v \right) = 0$ , elsewhere; also, similarly let  $\overline{F}_{\varepsilon 1} = \overline{F} + \varepsilon_1 \left( \delta_x^S - \overline{F} \right)$  with  $\delta_x^S \left( u \right) = 1$  if u < x and  $\delta_x^S \left( u \right) = 0$ , elsewhere and let  $\overline{F}_{\varepsilon_2} = \overline{F} + \varepsilon_2 \left( \delta_x^S - \overline{F} \right)$  with  $\delta_y^S \left( v \right) = 1$  if v < y and  $\delta_y^S \left( v \right) = 0$ , elsewhere, with  $0 \le \varepsilon, \varepsilon_1, \varepsilon_2 \le 1$ . Consequently,

$$T^{(l)}\left(S_{\varepsilon},\overline{F_{\varepsilon_{l}}},\overline{G_{\varepsilon_{2}}}\right)=S_{\varepsilon}\left(s_{l},t_{l}\right)-C_{\theta_{0}}\left(\overline{F_{\varepsilon_{1}}}\left(s_{l}\right),\overline{G_{\varepsilon_{2}}}\left(t_{l}\right)\right) \text{ and } T^{(l)}\left(S,\overline{F},\overline{G}\right)=0 \ .$$

The three influence functions are given respectively by

$$\begin{split} \frac{\partial T^{(l)}\left(S_{\varepsilon},\overline{F_{\varepsilon_{l}}},\overline{G_{\varepsilon_{2}}}\right)}{\partial \varepsilon} &= IC_{1}^{(l)}\left(x,y\right) = I\left[x > s_{l},y > t_{l}\right] - S\left(s_{l},t_{l}\right), \\ \frac{\partial T^{(l)}\left(S_{\varepsilon},\overline{F_{\varepsilon_{l}}},\overline{G_{\varepsilon_{2}}}\right)}{\partial \varepsilon_{1}} &= IC_{1}^{(l)}\left(x,y\right) = I\left[x > s_{l},y > t_{l}\right] - S\left(s_{l},t_{l}\right), \\ &= IC_{2}^{(l)}\left(S_{\varepsilon},\overline{F_{\varepsilon_{l}}},\overline{G_{\varepsilon_{2}}}\right) \\ &= IC_{2}^{(l)}\left(x\right) = -\frac{\partial C_{\theta_{0}}\left(\overline{F}\left(s_{l}\right),\overline{G}\left(t_{l}\right)\right)}{\partial u}\left(I\left[x > s_{l}\right] - \overline{F}\left(s_{l}\right)\right), \\ &\frac{\partial T^{(l)}\left(S_{\varepsilon},\overline{F_{\varepsilon_{l}}},\overline{G_{\varepsilon_{2}}}\right)}{\partial \varepsilon_{2}} \\ &= IC_{3}^{(l)}\left(y\right) = -\frac{\partial C_{\theta_{0}}\left(\overline{F}\left(s_{l}\right),\overline{G}\left(t_{l}\right)\right)}{\partial v}\left(I\left[y > t_{l}\right] - \overline{G}\left(t_{l}\right)\right). \end{split}$$

Consequently, we have the influence representation for the *I*-th element of  $\sqrt{n}(\hat{z}_n - \hat{z}_{g_0})$  with

$$\begin{split} \sqrt{n} \left( \hat{z}_{n,l} - \hat{z}_{\theta_{0,l}} \right) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} IC_{1}^{(l)} \left( x_{i}, y_{i} \right) + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} IC_{2}^{(l)} \left( x_{i} \right) \\ &+ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} IC_{3}^{(l)} \left( y_{i} \right) + o_{p} \left( 1 \right) \end{split}$$

and since  $(x_i, y_i)'$  are iid we have the equality in distribution asymptotically,

$$\sqrt{n}\left(\hat{z}_{n,l}-\hat{z}_{\theta_{0,l}}\right)=^{d}IC_{1}^{(l)}(x,y)+IC_{2}^{(l)}(x)+IC_{3}^{(l)}(y), l=1,\cdots,M.$$

Equivalently, using vector notations we have the following equality in distribution asymptotically by letting

$$\boldsymbol{a}_{l}' = \left(1, -\frac{\partial C_{\theta_{0}}\left(\overline{F}\left(s_{l}\right), \overline{G}\left(t_{l}\right)\right)}{\partial u}, -\frac{\partial C_{\theta_{0}}\left(\overline{F}\left(s_{l}\right), \overline{G}\left(t_{l}\right)\right)}{\partial v}\right),$$

$$\boldsymbol{Y}_{l} = \begin{bmatrix} I\left[x > s_{l}, y > t_{l}\right] - S\left(s_{l}, t_{l}\right)\\ I\left[x > s_{l}\right] - \overline{F}\left(s_{l}\right)\\ I\left[y > t_{l}\right] > \overline{G}\left(t_{l}\right) \end{bmatrix}.$$

 $\sqrt{n}(\hat{z}_{n,l} - \hat{z}_{\theta_{0,l}}) = d' \mathbf{a}_{l}' \mathbf{Y}_{l}$ , a result which is needed in Section 3.2.

## Technical Appendix 2 (TA2)

In this technical appendix, we shall justify the validity of expression (9) of Section 3.2.

The covariance matrix  $Cov(Y_i, Y_j)$  is defined as  $E(Y_i Y_j)$ , the vector

$$Y_{i} = \begin{pmatrix} I[x > s_{i}, y > t_{i}] - S(s_{i}, t_{i}) \\ I[x > s_{i}] - \overline{F}(s_{i}) \\ I[y > t_{i}] - \overline{G}(t_{i}) \end{pmatrix} \text{ and }$$

$$Y'_{i} = (I[x > s_{i}, y > t_{i}] - S(s_{i}, t_{i}), I[x > s_{i}] - \overline{F}(s_{i}), I[y > t_{i}] - \overline{G}(t_{i})).$$

Therefore the elements of the matrix  $Cov(Y_i, Y_i)$  are given by

$$c_{11} = E\left(\left(I\left[x > s_{i}, y > t_{i}\right] - S\left(s_{i}, t_{i}\right)\right)\left(I\left[x > s_{j}, y > t_{j}\right] - S\left(s_{j}, t_{j}\right)\right)\right)$$

$$c_{12} = E\left(\left(I\left[x > s_{i}, y > t_{i}\right] - S\left(s_{i}, t_{i}\right)\right)\left(I\left[x > s_{j}\right] - \overline{F}\left(s_{j}\right)\right)\right)$$

$$c_{13} = E\left(\left(I\left[x > s_{i}, y > t_{i}\right] - S\left(s_{i}, t_{i}\right)\right)\left(I\left[y > t_{j}\right] - \overline{G}\left(t_{j}\right)\right)\right)$$

$$c_{21} = E\left(\left(I\left[x > s_{i}\right] - \overline{F}\left(s_{i}\right)\right)\left(I\left[x > s_{j}, y > t_{j}\right] - S\left(s_{j}, t_{j}\right)\right)\right)$$

$$c_{22} = E\left(\left(I\left[x > s_{i}\right] - \overline{F}\left(s_{i}\right)\right)\left(I\left[x > s_{j}\right] - \overline{F}\left(s_{j}\right)\right)\right)$$

$$c_{23} = E\left(I\left[x > s_{i}\right] - \overline{F}\left(s_{i}\right)\left(I\left[y > t_{j}\right] - \overline{G}\left(t_{j}\right)\right)\right)$$

$$c_{31} = E\left(\left(I\left[y > t_{i}\right] - \overline{G}\left(t_{i}\right)\right)\left(I\left[x > s_{j}, y > t_{j}\right] - S\left(s_{j}, t_{j}\right)\right)\right)$$

$$c_{32} = E\left(\left(I\left[y > t_{i}\right] - \overline{G}\left(t_{i}\right)\right)\left(I\left[x > s_{j}\right] - \overline{F}\left(s_{j}\right)\right)\right)$$

$$c_{33} = E\left(\left(I\left[y > t_{i}\right] - \overline{G}\left(t_{i}\right)\right)\left(I\left[y > t_{j}\right] - \overline{G}\left(t_{j}\right)\right)\right)$$

Now, note that the above equalities which give the elements of the matrix  $Cov(Y_i, Y_j)$  can be reexpressed as the equalities as given by expression (9) in Section 3.2.