Simulated Minimum Quadratic Distance Methods Using Grouped Data for Some Bivariate Continuous Models

Andrew Luong

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

Abstract

Quadratic distance methods based on a special distance which make use of survival functions are developed for inferences for bivariate continuous models using selected points on the nonegative quadrant. A related version which can be viewed as a simulated version is also developed and appears to be suitable for bivariate distributions with no closed form expressions and numerically not tractable but it is easy to simulate from these distributions. The notion of an adaptive basis is introduced and the estimators can be viewed as quasilikelihood estimators using the projected score functions on an adaptive basis and they are closely related to minimum chi-square estimators with random cells which can also be viewed as quasilikeliood estimators using a projected score functions on a special adaptive basis but the elements of such a basis were linearly dependent. A rule for selecting points on the nonegative quadrant which make use of quasi Monte Carlo (QMC) numbers and two sample quantiles of the two marginal distributions is proposed if complete data is available and like minimum chi-square methods; the quadratic distance methods also offer chi-square statistics which appear to be useful in practice for model testing.

Keywords

Projected Score Functions, Adaptive Basis, Complete Basis, Chi-Square Tests Statistics, Random Cells, Contingency Table

1. Introduction

In actuarial science or biostatistics, we often encounter bivariate data which are already grouped into cells forming a contingency table and we would like to make

inferences for a continuous bivariate model used to model the complete data, see Partrat [1] (p. 225), Gibbons and Chakraborti [2] (pp. 511-512) for examples.

The bivariate distributions if they have closed form expressions then there is no difficulty in general to fit these distributions using maximum likelihood or minimum chi-square methods based on grouped data for examples but many useful distributions might only be computable numerically as they are expressible only using an integral representation and if the quadrature numerical methods often fail then it appears to be natural to develop simulated methods of inferences for these distributions. We would like to have methods which offer a unified approach to estimation and model testing as well beside they should be able to handle the situation where the lack of closed form expressions for the model survival distributions might create numerical difficulties. We shall see subsequently that new distributions created using the bivariate survival mixture operator (BSPM) introduced by Marshall and Olkin [3] (pp. 834-836) and by the trivariate reduction techniques often lead to distributions with no closed form expressions for survival functions but it is easy to draw samples from such distributions. Since we focus on nonnegative bivariate distributions in actuarial science, it is natural to use survival functions instead of just using distribution functions alone. The BSPM operator will be introduced and we shall see a few examples to illustrate the numerical difficulties we might encounter when fitting these distributions.

1.1. Bivariate Survival Power Mixture Operator

Marshall and Olkin [3] in their seminal paper have introduced the following operator to create a new bivariate survival function $S(x, y)$ from two univariate survival functions $\overline{F_1}(x), \overline{F_2}(y)$ and a mixing distribution $G(\theta)$ for a nonnegative mixing random variable $\theta \geq 0$. We shall call their operator bivariate survival power mixture operator and use the acronym BSPM and we shall see how this operator works to create new bivariate survival functions. The new survival function created can be expressed as an integral given by

$$S(x, y) = \int_0^\infty \left(\overline{F_1}(x) \right)^\theta \left(\overline{F_2}(y) \right)^\theta \, \mathrm{d}G(\theta).$$

Since there is an integral representation, the new survival function might still be computable numerically depending on the expressions for $\overline{F_1}(x), \overline{F_2}(y)$ and $G(\theta)$. An algorithm to simulate a sample from $S(x, y)$ has also been given by Marshall and Olkin [3] (p. 840).

Later on in section (1.2) we shall examine another way to create new survival functions. Unlike new distributions created using the BSPM operator, new distributions created using means of trivariate reduction techniques often do not even have an integral representation despite the functions used are simple for examples linear functions. We shall discuss further trivariate reduction techniques in section 1.2 and consider first a few examples of new distributions cretead using the BSPM operator subsequently.

1.2. Some Examples of New Bivariate Distributions Created

Example 1

We let $\overline{F_1}(x;\alpha_1,\lambda_1) = e^{-\lambda_1 x^{\alpha_1}}, x > 0, \lambda_1, \alpha_1 > 0$ which is the survival function of a Weibull distribution and similarly let $\overline{F_2}(y;\alpha_2,\lambda_2) = e^{-\lambda_2 x^{\alpha_2}}, y > 0, \lambda_2, \alpha_2 > 0$. For the mixing random variable θ let θ follows a Pareto type II distribution which is also called Lomax distribution with density function given by

$f(\theta;\tau,\delta) = \dfrac{\delta\tau^{\delta}}{(\theta+\tau)^{\delta+1}}$ with the domain given by $\theta > 0$ and the parameters α

and δ are positive. Note that θ has no closed form Laplace transform (LT).

The new bivariate distribution created using the BSPM operator is

$$S_{\boldsymbol{\beta}}(x,y) = \int_0^\infty e^{-\theta\lambda_1 x^{\alpha_1}} e^{-\theta\lambda_2 x^{\alpha_2}} \frac{\delta\tau^{\delta}}{(\theta+\tau)^{\delta+1}} \, \mathrm{d}\theta, \ \boldsymbol{\beta} = (\alpha_1,\alpha_2,\lambda_1,\lambda_2,\delta,\tau)'.$$

For most of the univariate distributions used in the paper, see Appendix A given by Klugman *et al.* [4] (pp. 459-482).

Observe that if we specify a Gamma distribution for θ instead of a Lomax distribution as discussed earlier where the density function of θ is given by

$$f(\theta;\tau,\delta) = \frac{1}{\Gamma(\delta)}\theta^{\delta-1}e^{-\theta}, \theta > 0$$

and its Laplace transform is given by $\phi(s) = (1+s)^{-\delta}$, the newly created bivariate survival distribution can be expressed as

$$S_{\boldsymbol{\beta}}(x,y) = \int_0^\infty e^{-\theta\lambda_1 x^{\alpha_1}} e^{-\theta\lambda_2 x^{\alpha_2}} \frac{1}{\Gamma(\delta)}\theta^{\delta-1}e^{-\theta}\mathrm{d}\theta$$

and since $\phi(s)$ has a closed form expression $S_{\boldsymbol{\beta}}(x,y)$ has closed form expression which is given by

$$S_{\boldsymbol{\beta}}(x,y) = \phi\big(H_1(x;\alpha_1,\lambda_1) + H_2(y;\alpha_2,\lambda_2)\big)$$

where $H_1(x;\alpha_1,\lambda_1)$ and $H_2(y;\alpha_2,\lambda_2)$ are respectively the cumulative hazard rate functions of $\overline{F_1}(x;\alpha_1,\lambda_1)$ and $\overline{F_2}(y;\alpha_2,\lambda_2)$ with

$$H_1(x,\alpha_1,\lambda_1) = -\ln\overline{F_1}(x;\alpha_1,\lambda_1) = \lambda_1 x^{\alpha_1},$$

$$H_2(y,\alpha_2,\lambda_2) = -\ln\overline{F_2}(y;\alpha_2,\lambda_2) = \lambda_2 y^{\alpha_2}.$$

By using the usual conditioning argument, by conditioning on θ often we can obtain the first two moments of the vector $Z = (X,Y)'$ and even higher positive integer moments can be obtained without having a closed form for $S_{\beta}(x,y)$; see the conditioning argument for the univariate case given by Klugman *et al.* [4] (pp. 62-65). If complete data are available then the parameters of some bivariate distribution created using the BSPM operator can be estimated using the methods of moment (MM).

In this paper we emphasize grouped data. In general, with grouped data MM estimators cannot be obtained and furthermore with four or five parameters in the bivariate model, high order moments must be used and as a result the MM esti-

mators are not robust in general. We consider the situation where the data have been grouped into a contingency table so that we must analyse data in this from or the complete data is available but we must group them to perform chi-square tests for model testing. If complete data are available then we have choices to group the data; in this situation we hope to be able to propose a way to group them so that inference methods based on such a grouping rule will have have high efficiencies, see the discussions in section (5) by Klugman and Parsa [5] (pp. 146-147) on the difficulties on grouping data to perfom goodness of fit tests. We use the notion of complete data to describe the situation where we have bivariate observations.

$Z_i = (X_i, Y_i)'$, $i = 1, \cdots, n$ which are independent and identically distributed(iid)from a bivariate distribution specified by a bivariate survival function $S_{\boldsymbol{\beta}}(x, y)$; this includes a situation where the original observations have been left truncated by d_1 and d_2 where the values d_1 and d_2 are known; d_1 and d_2 are the amount of deductibles in actuarial science for example. We can view

$$S_{\boldsymbol{\beta}}(x, y) = \frac{S_{\boldsymbol{\beta}}^o(x, y)}{S_{\boldsymbol{\beta}}^o(d_1, d_2)}$$ and we only need $S_{\boldsymbol{\beta}}(x, y)$ for fitting with $S_{\boldsymbol{\beta}}^o(x, y)$

being specified as well; see Klugman and Parsa [5] (p. 142) for these models in actuarial science. Furthermore, in our set up we emphasize survival function $S_{\boldsymbol{\beta}}(x, y)$ but clearly the bivariate distribution function $K_{\boldsymbol{\beta}}(x, y)$ can be obtained from $S_{\boldsymbol{\beta}}(x, y)$ using the relation

$$S_{\boldsymbol{\beta}}(x, y) = 1 - F_{\boldsymbol{\beta}}(x) - G_{\boldsymbol{\beta}}(y) + K_{\boldsymbol{\beta}}(x, y),$$

$F_{\boldsymbol{\beta}}(x)$ and $G_{\boldsymbol{\beta}}(y)$ are the two marginal distributions of $K_{\boldsymbol{\beta}}(x, y)$. For non-negative parametric families where the bivariate distribution functions are commonly used for specifying the families, it is not difficult to convert them to bivariate survival functions and consequently, MQD methods are still applicable; only some minor modifications are needed.

Example 2

In this example, we let $\overline{F_1}$ and $\overline{F_2}$ to be Burr survival functions, see Hogg *et al.* [6] (p. 201) for the Burr distribution with

$$\overline{F_1}(x; \alpha_1, \lambda_1, \gamma_1) = (1 + \delta_1 x^{\tau_1})^{-\gamma_1}, \ x > 0$$

and the parameters $\delta_1, \tau_1, \gamma_1$ are positive and similarly, let

$$\overline{F_2}(y; \alpha_2, \lambda_2, \gamma_2) = (1 + \delta_2 y^{\tau_2})^{-\gamma_2}.$$

For the distribution of θ, we specify a Weibull distribution with density function given by

$$f(\theta; \lambda, \alpha) = \alpha \lambda \theta^{\alpha-1} \mathrm{e}^{-\lambda \theta^\alpha}, \ \theta > 0.$$

The new distribution created using the BSPM operator will have bivariate survival function given by

$$S_{\boldsymbol{\beta}}(x, y) = \int_0^\infty (1 + \delta_1 x^{\tau_1})^{-\theta \gamma_1} (1 + \delta_2 y^{\tau_2})^{-\theta \gamma_2} \cdot \alpha \lambda \theta^{\alpha-1} \mathrm{e}^{-\lambda \theta^\alpha} \mathrm{d}\theta.$$

The bivariate survival function has no closed form expression. For bivariate

distributions without a closed form expression for their survival functions and only representable using an integral representation on an unbounded domain, $S_\beta(x, y)$ can be evaluated numerically or not depends on the integrand and the numerical quadrature method used. In the same vein, we can mention the class of bivariate contingency tables studied by Mardia [7], Mardia [8], Plackett [9] as these distributions have numerical tractable bivariate distribution functions but no closed form expression for the bivariate distributions. In this paper, we emphasize statistical aspects and do not go into details in the question of dependence of the two components of the new bivariate survival function. Marshall and Olkin [3], Marshall and Olkin [10] have discussed some of the issues of dependence and infinitely divisibility for distributions created using mixture procedures.

For $S_\beta(x, y)$ which is not numerical tractable we propose simulated minimum quadratic distance (SMQD) methods and providing that we can draw simulated samples from $S_\beta(x, y)$, estimation of β is still possible and minimum quadratic distance methods offer a unified approach for $S_\beta(x, y)$ with a closed form expression or without a closed form expression for grouped data without choice and with choices. MQD methods version D will be suitable for numerical tractable $S_\beta(x, y)$ and a corresponding simulated version, version S will be suitable if $S_\beta(x, y)$ is not numerically tractable. For version S, $S_\beta(x, y)$ will be replaced by a sample bivariate survival distribution $S_\beta^s(x, y)$ based on a simulated sample of size $U = \tau n, \tau \geq 10$ drawn from $S_\beta(x, y)$ with $S_\beta^s(x, y)$ converges in probability to $S_\beta(x, y)$ for each point $z = (x, y)'$ fixed,

$$\text{i.e., } S_\beta^s(x, y) \xrightarrow{\ p\ } S_\beta(x, y),$$

$S_\beta^s(x, y)$ is defined similarly as $S_n(x, y)$, see expression (1).

Data under the form of contingency tables are often encountered in actuarial science and biostatistics where bivariate observations are grouped into a two dimensions array or matrix and only the numbers of observations or proportions of the original sample which belong to the elements of such a matrix are recorded. The original data set is lost. Obviously, when the original complete data set is available, we can always group them into a contingency table but once grouped it is impossible to convert grouped data to complete data. We focus on the situations where the complete data are observations $Z_1, \cdots, Z_n$ which are independent and identically distributed as $Z = (X, Y)'$ which follows a bivariate absolutely continuous distribution with domain given by the nonnegative quadrant and subsequently they are grouped and we develop statistical inference techniques using grouped data.

1.3. New Bivariate Distributions Created by Trivariate Reductions

A version of bivariate gamma distribution as introduced by Mathai and Moschopolus [11] (pp. 137-138) can be introduced with the use of two linear functions $\phi_1(V_0, V_1, V_2)$ and $\phi_2(V_0, V_1, V_2)$, these functions are known and their arguments as given by V_0, V_1, V_2 are independent univariate random variables. It is simple to simulate a pair of observation $(X, Y)'$ as we have the following equalities in distributionng for the pair of observations $(X, Y)'$,

$X =^d \phi_1(V_0, V_1, V_2), Y =^d \phi_2(V_0, V_1, V_2)$ and the functional forms for $\phi_1(V_0, V_1, V_2)$ and $\phi_2(V_0, V_1, V_2)$ are given.

Let V_0, V_1, V_2 be gamma random variables with their respective density functions given by

$$f(v_i; \alpha_i, \beta_i) = \frac{1}{\Gamma(\alpha_i)\beta_i^{\alpha_i}} v_i^{\alpha_i - 1} e^{-\frac{v_i}{\beta_i}}, v_i \geq 0, \alpha_i, \beta_i \geq 0$$

and

$$X = \phi_1(V_0, V_1, V_2) = \frac{\beta_1}{\beta_0} V_0 + V_1,$$

$$Y = \phi_2(V_0, V_1, V_2) = \frac{\beta_1}{\beta_0} V_0 + V_2.$$

The bivariate density function $f(x, y; \boldsymbol{\beta})$ has no closed form expression and very complicated. It has five parameters which can be represented by the vector $\boldsymbol{\beta} = (\alpha_0, \alpha_1, \alpha_2, \beta_1, \beta_2)'$, see section (5) as given by Mathai and Moschopoulos [11] (pp. 145-148). Mathai and Moschopolus [11] also give methods of moment estimators (MM) in section (7) of their paper. We shall consider their estimators and compare with the simulated quadratic distance estimators in section (4). The bivariate gamma distribution as introduced by Furman [12] can also be obtained similarly using another pair of linear functions $\phi_1(.)$ and $\phi_2(.)$. For the use of nonlinear functions $\phi_1(.)$ and $\phi_2(.)$ to create bivariate distributions, see Chapter 15 given by Hutchinson and Lai [13] (pp. 218-224).

1.4. 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 $Z_i = (X_i, Y_i)', i = 1, \cdots, n$ which are independent and identically distributed as $Z = (X, Y)'$ which follows a non-negative continuous bivariate distribution with model survival function given by $S_{\boldsymbol{\beta}}(x, y)$.

The vector of parameters is $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_m)'$, the true vector of parameters is denoted by $\boldsymbol{\beta}_0$. 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. The grouping in two dimensional cells generalize the grouping of univariate into disjoint intervals of the nonnegative real line in one dimension. Contingency tables data are often encountered in actuarial science and biostatistics, see Partrat [1] (p. 225), Gibbons and Chakraborti [2] (pp. 511-512). We shall give a brief description below.

Let the nonnegative axis X be partitioned into disjoints interval $\bigcup_{i=1}^{J} [s_{i-1}, s_i)$ with $s_0 = 0, s_J = \infty$ and similarly, the axis Y be partitioned into disjoints interval $\bigcup_{j=0}^{K} [t_{j-1}, t_j)$ with $t_0 = 0, t_K = \infty$.

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

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

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

$$C_{ij}, i = 1, \cdots, I, j = 1, \cdots, J$$

We can define the empirical bivariate survival function as

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

and we have $S_n(x, y) \xrightarrow{p} S_{\boldsymbol{\beta}_0}(x, y)$.

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

$$p_n(C_{ij}) = S_n(s_{i-1}, t_{j-1}) - S_n(s_{i-1}, t_j) - S_n(s_i, t_{j-1}) + S_n(s_j, t_j) \tag{2}$$

and the corresponding model probability is

$$p_{\boldsymbol{\beta}}(C_{ij}) = S_{\boldsymbol{\beta}}(s_{i-1}, t_{j-1}) - S_{\boldsymbol{\beta}}(s_{i-1}, t_j) - S_{\boldsymbol{\beta}}(s_i, t_{j-1}) + S_{\boldsymbol{\beta}}(s_j, t_j). \tag{3}$$

Note that

$$S_{\boldsymbol{\beta}}(s_i, t_J) = 0, S_n(s_i, t_J) = 0, i = 1, \cdots, I \tag{4}$$

and similarly,

$$S_{\boldsymbol{\beta}}(s_I, t_j) = 0, S_n(s_I, t_j) = 0, j = 1, \cdots, J. \tag{5}$$

so they can be discarded without affecting the efficiency of inference methods and this is precisely the approach quadratic distance methods use by discarding redundant elements and create a basis with only linearly independent elements but the basis will span the same linear space. Consequently, we gain in numerical effciency and at the same time retaining the same efficiency.

1.5. Efficient Modified Minimum Chi-Square Methods

Using the contingency table data, the modified minimum chi-square estimators which are as efficient as the likelihood estimators using the grouped data is obtained by minimizing the objective function given by

$$\sum_{i,j} \frac{\left(p_n(C_{ij}) - p_{\boldsymbol{\beta}}(C_{ij})\right)^2}{p_n(C_{ij})}$$

and since $p_n(C_{ij}) \xrightarrow{p} p_{\boldsymbol{\beta}_0}(C_{ij})$, the minimum chi-square estimators given by the vector $\tilde{\boldsymbol{\beta}}$ which minimizes the expression given above have the same asymptotic efficiency as the vector $\boldsymbol{\beta}^*$ which minimize

$\sum_{i,j} \dfrac{\left(p_n(C_{ij}) - p_{\boldsymbol{\beta}}(C_{ij})\right)^2}{p_{\boldsymbol{\beta}_0}(C_{ij})}$ and under differentiability assumptions given by the

roots of the system of equation

$$\sum_{i,j} \frac{\left(p_n\left(C_{ij}\right) - p_{\boldsymbol{\beta}}\left(C_{ij}\right)\right)}{p_{\boldsymbol{\beta}_0}\left(C_{ij}\right)} \frac{\partial p_{\boldsymbol{\beta}}\left(C_{ij}\right)}{\partial \boldsymbol{\beta}} = 0.$$

Note that the quasi-score functions generated belong to the linear space spanned by

$$\left\{ p_n\left(C_{ij}\right) - p_{\boldsymbol{\beta}}\left(C_{ij}\right), i = 1, \cdots, I, \ j = 1, \cdots, J \right\}. \tag{6}$$

But since we have the property given by expressions (4-5), the same linear space is spanned by

$$\left\{ S_n\left(s_{ij}, t_{ij}\right) - S_{\boldsymbol{\beta}}\left(s_{ij}, t_{ij}\right), i = 1, \cdots, I-1, \ j = 1, \cdots, J-1 \right\}. \tag{7}$$

Therefore, an equivalent method but possibly numerically more efficient, remember we need to evaluate $p_{\beta}\left(C_{ij}\right)$ numerically or by simulation is to minimize a quadratic form using the elements of the basis given by expression (7) with an optimum matrix which is no longer diagonal as in the minimum chi-square objective function but it turns out to be quite simple and can be estimated empirically as $S_n\left(s_{ij}, t_{ij}\right)$ are relatively simple and well defined for $i = 1, \cdots, I-1, j = 1, \cdots, J-1$. Furthermore, for performing minimum chi square methods in practice if the cells do not have more than 5 elements, they need to be regrouped into larger cells, this will reduce the efficiency of the minimum chi-square methods in practice. We do not need as many regrouping operations with the proposed methods which are quadratic distance methods. For the equivalent efficiency of the new proposed methods, see the projection argument in Luong [14] (pp. 463-468) for quadratic distance methods.

Also using expression (7) is equivalent to use overlapping cells of the form

$$O_{ij} = I\left[x > s_{ij}, y > t_{ij}\right], i = 1, \cdots, I-1, \ j = 1, \cdots, J-1. \tag{8}$$

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 components being elements of the basis so that we only need one subscript by collapsing the matrix given by expression (7) 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

$$z_n = \left(S_n\left(s_1, t_1\right), \cdots, S_n\left(s_M, t_M\right) \right)', \ M = (I-1)(J-1). \tag{9}$$

and its model counterpart is

$$z_{\boldsymbol{\beta}} = \left(S_{\boldsymbol{\beta}}\left(s_1, t_1\right), \cdots, S_{\boldsymbol{\beta}}\left(s_M, t_M\right) \right)'. \tag{10}$$

The number of components of z_n is M with the assumption $M > m$.

Efficient quadratic distance methods can be constructed using the inverse of the covariance matrix of z_n as the weight matrix for the quadratic form but such an optimum matrix is only defined up to a constant, so the inverse of the matrix of the following vector

$$\boldsymbol{h}(x, y) = \left(I\left[x > s_1, y > t_1\right] - S_{\boldsymbol{\beta}}\left(s_1, t_1\right), \cdots, I\left[x > s_M, y > t_M\right] - S_{\boldsymbol{\beta}}\left(s_M, t_M\right) \right)'. \tag{11}$$

can also be considered as optimum. It can also be replaced by a consistent estimate and since the elements of the basis can be identified as the corresponding elements of the vector $\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{h}(x_i,y_i)$ with the observations being independent and identically distributed, choosing a basis is equivalent to choose the vector $\boldsymbol{h}(x,y)$.

It is not difficult to see that the infinite basis of the form

$$\left\{ I\left[x > s_l, y > t_l \right] - S_{\boldsymbol{\beta}}(s_l,t_l), l = 1,2,\cdots \right\}$$

is complete and the projected score functions will give estimators which have the same efficiency as the maximum likelihood estimators as the score functions belong to the space spanned by the infinite basis, see Carrasco and Florens [15] for a similar property for the univariate case; also see Luong [14] (pp. 461-468) for the notion of MQD estimators as quasilikelihood estimators based on the projected score functions on a finite basis. The MQD methods which make use of such a basis for the bivariate case will be introduced below and they are similar to the univariate case. Note if the data has been grouped, this means that we have no choice of points $(s_l,t_l)', i = 1,\cdots,M$ as they are already predetermined by the way data are grouped into cells.

Beside the predetermined grouped scenario, we shall also examine the question: if we have complete data and we would like to choose a finite basis with M elements and since these elements are identified as points, the same question can be phrased as how we should choose M points or equivalently, how we should group the data into cells?

The question on how to choose cells appear to be already difficult for the minimum chi square methods with univariate data, see Greenwood and Nikulin [16] (pp. 194-208), we shall propose a solution based on quasi-Monte Carlo (QMC) methods via a Halton sequences and two empirical quantiles from the marginal distributions to create an artificial sample with values on the nonnegative quadrant. The selected points used to construct quadratic distances are based on these artificial sample points. Since these points are random it is similar to the use random cells for minimum chi-square methods and naturally we would like to introduce the notion of an adaptive basis used to achieve high efficiency and it will also unify quadratic distance and minimum chi-square methods. An adaptive basis is data dependent and therefore carry informations about the true vector of parameters $\boldsymbol{\beta}_0$ despite that $\boldsymbol{\beta}_0$ is unknown and consequently the projected scores functions on such a basis will lead to inference with better efficiency than without using the informations obtainable from data concerning $\boldsymbol{\beta}_0$, in general.

We shall discuss on how to construct such an adaptive basis when complete data is available in section (3) where we would like to give some light on the question on how to form a finite basis so that inference methods using the basis will have high efficiencies for a restricted parameter space which appear to be reasonable for the applications. It appears that such an adaptive basis which is formed from a com-

plete basis is very practical for applications. Using an infinite basis such as in the case of generalized method of moments (GMM) with univariate observations based on a continuum moment conditions as introduced by Carrasco and Florens [15] appears to be complicated for practitioners. Consequently, only finite bases are considered in this paper.

Using the remarks given by Luong ([14] (p. 472), we need to work with a restricted parameter space for a basis with finite number of elements so that statistical inferences using MQD methods or GMM methods can have high efficiencies. An adaptive basis only has a finite number of elements so that numerically it is not so complicated to use such a basis to construct MQD or SMQD methods. The elements of the basis will be adapted to β_0 and efficiencies of the SMQD methods using such a basis come from the fact that the elements of the basis are chosen accordingly and adjustable depending on the value of the vector of true parameters β_0 despite that β_0 is unknown.

In general, it is not obvious on how to construct an adaptive basis from a complete basis, for example it is a difficult task to construct an adaptive basis from the complete basis of polynomials $\left\{ x^j y^k - E_\beta \left(x^j y^k \right), j = 0,1,\cdots, k = 0,1,\cdots \right\}$ so that the MQD methods using such a basis might have good efficiencies. Howewer, it is natural to construct an adaptive basis with only finite number of elements extracted from a complete basis as given by expression (11). This will be further developed in section (3) of the paper. The notion of an adaptive basis constructed from a complete basis used to project the score functions appears to be relatively new despite implicitly it has been used in the minimum chi-square methods without using explicitly this notion, see Moore and Spruill [17] and Pollard [18] (pp. 317-318).

In the literature, attentions seem to be given to complete bases. Adaptive bases will be further developed and discussed in section (3) and used to develop MQD methods with a deterministic version (version D) and a simulated version (version S) or SMQD methods.

For the deterministic version (version D) with $S_\beta(u,v)$ considered to be fixed, asymptotic properties of the methods are similar to the univariate case as given Luong and Thompson [19], Duchesne *et al.* [20]; also see related results of generalized methods of moment (GMM) in Newey and McFadden [21] (p. 2148). For the simulated version where we replace $S_\beta(u,v)$ by a sample survival function using a simulated sample of size U drawn from $S_\beta(u,v)$, *i.e.*, $S_\beta^s(u,v)$. We can make use of results given by Theorem (3.1) and Theorem (3.3) given by Pakes and Pollard [22] (pp. 1038-1043) to establish asymptotic properties for MQD and SMQD methods.

It might be worth to mention in practice without an infinite and complete basis and only using a finite basis which is a subset of an infinite basis high efficiency for the procedures can only be attained in some restricted parameter space in general unless the score functions belong to the span of the finite base. One viable strategy is to identify a restricted parameter space for the type of applications being considered; often it suffices to use a restricted parameter space then try to identify

elements to form a finite basis so that the procedures will retain high efficiency overall for β_0 which belongs the restricted parameter space, see Luong [14] (pp. 463-468). If such a strategy is not feasible then we might want to turn to using an adaptive basis with finite elements constructed from an infinite complete basis but the elements are adapted to data. Implicilty, the idead behind is to let the data points point to a restricted parameter space and the elements of the basis are adjusted accordingly. Minimum chi-square (MCS) methods using nonoverlapping random cells make use of an adaptive basis where the elements are linearly dependent but for MQD methods and the simulated version SMQD methods we make use of overlapping random cells in the form of random points on the nonnegative quadrant and unlike minimum chi-square methods where it is difficult to have a rule to choose random cells we shall have a rule to select these points.

For MQD or SMQD methods, the matrix Ω_0 which is the covariance matrix of the vector $h(x, y)$ under β_0 plays an important role as we can obtain estimators with good efficiencies for estimators using Ω_0. Despite that Ω_0 is unknown, its elements are not complicated and moreover, it can be replaced by a consistent estimate constructed empirically without affecting the asymptotic efficiency of the procedures. It is also needed for constructing chi-square tests statistics. We shall give more details about this matrix and construct an empirical estimate $\widehat{\Omega}_0$ which is data dependent for Ω_0.

Let Ω_0 the covariance matrix of the vector $h(x, y)$ under β_0, its elements are given by

$$
\begin{aligned}
\Omega_0(i, j) &= cov\left(I[x > s_i, y > t_i], I[x > s_j, y > t_j]\right) \\
&= E\left(\left(I[x > s_i, y > t_i]\right)\left(I[x > s_j, y > t_j]\right)\right) \\
&\quad - \left(E\left(I[x > s_i, y > t_i]\right)\right)\left(E\left(I[x > s_i, y > t_i]\right)\right) \\
&= S_{\beta_0}\left(\max(s_i, s_j), \max(t_i, t_j)\right) \\
&\quad - \left(S_{\beta_0}(s_i, t_i)\right)\left(S_{\beta_0}(s_j, t_j)\right), i = 1, \cdots, M, j = 1, \cdots, M
\end{aligned}
\tag{12}
$$

Clearly, these elements can be estimated empirically with the bivariate empirical survival function using grouped data provided by the contingency table; we then have the corresponding estimates given by

$$
\begin{aligned}
\widehat{\Omega}_0(i, j) &= S_n\left(\max(s_i, s_j), \max(t_i, t_j)\right) \\
&\quad - \left(S_n(s_i, t_i)\right)\left(S_n(s_j, t_j)\right), i = 1, \cdots, M, j = 1, \cdots, M
\end{aligned}
\tag{13}
$$

Therefore, we can define the matrix $\widehat{\Omega}_0$ and its inverse is denoted by $\widehat{W}_0$ and similarly let W_0 be the inverse of Ω_0. Clearly, $\widehat{W}_0 \xrightarrow{p} W_0$ as $S_n(x, y) \xrightarrow{p} S_{\beta_0}(x, y)$. Now, we can define the objective functions to be minimized for the implementations of MQD and SMQD methods.

For version D, let

$$
G_n(\beta) = \left(z_n - z_\beta\right)'
\tag{14}
$$

and let $S_{\boldsymbol{\beta}}^{s}(x,y)$ be an estimate of $S_{\boldsymbol{\beta}}(x,y)$ using a simulated sample of size $U = \tau n$ so that we can define

$$z_{\boldsymbol{\beta}}^{s} = \left(S_{\boldsymbol{\beta}}^{s}(s_1,t_1), \cdots, S_{\boldsymbol{\beta}}^{s}(s_M,t_M) \right)'$$

and let

$$G_n(\boldsymbol{\beta}) = \left(z_n - z_{\boldsymbol{\beta}}^{s} \right)' \tag{15}$$

for version S.

We can define the length of the random function $G_n(\boldsymbol{\beta})$ as $\|G_n(\boldsymbol{\beta})\|$ with the norm $\|.\|$ defined as

$$\|G_n(\boldsymbol{\beta})\|^2 = (z_n - z_{\boldsymbol{\beta}})' \widehat{W_0} (z_n - z_{\boldsymbol{\beta}}) \tag{16}$$

or equivalently

$$\|G_n(\boldsymbol{\beta})\|^2 = (z_n - z_{\boldsymbol{\beta}})' W_0 (z_n - z_{\boldsymbol{\beta}}) \tag{17}$$

as they will give asymptotic equivalent estimators and goodness-of-fit tests statistics; for finding estimators numerically, we need to minimize expression (16) and for asymptotic properties it might be slightly simpler to work with expression (17) as less notations are involved. Similarly, for version S let

$$Q_n(\boldsymbol{\beta}) = \|G_n(\boldsymbol{\beta})\|^2 = \left(z_n - z_{\boldsymbol{\beta}}^{s} \right)' \widehat{W_0} \left(z_n - z_{\boldsymbol{\beta}}^{s} \right) \tag{18}$$

or equivalently

$$Q_n(\boldsymbol{\beta}) = \|G_n(\boldsymbol{\beta})\|^2 = \left(z_n - z_{\boldsymbol{\beta}}^{s} \right)' W_0 \left(z_n - z_{\boldsymbol{\beta}}^{s} \right). \tag{19}$$

Note that the weight matrix $\widehat{W_0}$ is the same for both versions and the norm $\|.\|$ is a weighted Euclidean norm which obeys the triangle inequality so that results of Theorems given by Pakes and Pollard despite they are stated with the Euclidean norm remain valid if the Euclidean norm is replaced by a weighted Euclidean norm. Minimum QD estimators are obtained as the vector $\hat{\boldsymbol{\beta}}$ which minimizes the objective function $Q_n(\boldsymbol{\beta})$ or equivalently $\|G_n(\boldsymbol{\beta})\|$ as defined by expression (16) for version D or $\hat{\boldsymbol{\beta}}^{S}$ which minimizes $Q_n(\boldsymbol{\beta})$ or equivalently $\|G_n(\boldsymbol{\beta})\|$ as defined by expression (18) for version S.

The paper is organized as follows.

In Section 2, MQD and SMQD methods will be developed using predetermined grouped data. Asymptotic properties of the estimators are studied and asymptotic distribution for the model testing statistics are derived. The methods can be extended to the situation when comple data is available but will be grouped by defining a rule to choose points on the nonnegative quadrant to group the data. An artificial sample constructed using two sample quantiles and QMC numbers are proposed in Section 3 to select points and the methods developed with preselected points or cells in section (2) are shown to be still applicable, the methods can be seen as equivalent to minimum chi-square methods with random cells but with a rule to define these cells; using random cells is equivalent to using an adaptive

basis. Both QD estimation and minimum chi-square estimation can be unified with the approach of quasilikelihood estimation using an adaptive basis and to implement MQD or SMQD methods require less computing time than the related minimum chi-square versions due to the adaptive basis being used by MQD and SMQD methods only has linearly independent elements and there is less numerical evaluations or simulations for computing probabilities assigned to points than cells. Section 4 illustrates the implementations of SMQD methods by comparing the methods of moment estimators (MM) with the SMQD estimators. The SMQD estimators appear to be much more efficient and robust than MM estimators in a limited study for a bivariate gamma model with the range of parameters often encountered in actuarial science and chosen in the study.

2. SMQD Methods Using Grouped Data

2.1. Estimation

Consistency for both versions of quadratic distance estimators using predetermined grouped data can be treated in a unified way using the following Theorem 1 which is essentially Theorem 3.1 of Pakes and Pollard [22] (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 (3) where we have complete data and have choices to regroup the data into cells or equivalently forming the artificial sample points on the nonnegative quadrant.

Theorem 1 (Consistency)

Under the following conditions $\tilde{\boldsymbol{\beta}}$ converges in probability to $\boldsymbol{\beta}_0$:

1) $\left\| G_n \left(\tilde{\boldsymbol{\beta}} \right) \right\| \le o_p (1) + \inf_{\boldsymbol{\beta} \in \boldsymbol{\Omega}} \left(\left\| G_n \left(\boldsymbol{\beta} \right) \right\| \right)$, the parameter space space Ω is compact

2) $\left\| G_n \left(\boldsymbol{\beta}_0 \right) \right\| = o_p (1)$,

3) $\sup_{\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| > \delta} \left(\dfrac{1}{\left\| G_n \left(\boldsymbol{\beta} \right) \right\|} \right) = O_p (1)$ for each $\delta > 0$.

Theorem 3.1 states condition 2) as $G_n \left(\boldsymbol{\beta}_0 \right) = o_p (1)$ but in the proof the authors just use $\left\| G_n \left(\boldsymbol{\beta}_0 \right) \right\| = o_p (1)$ so we state condition b) as $\left\| G_n \left(\boldsymbol{\beta}_0 \right) \right\| = o_p (1)$.

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

$n^{-\frac{1}{2}} \to 0$. For version D and version S, we have $\inf_{\boldsymbol{\theta} \in \Omega} \left(\left\| G_n \left(\boldsymbol{\beta} \right) \right\| \right)$ occurs at the values of the vector values of the MQD estimators, so the conditions a) and b) are satisfied for both versions and compactness of the parameter space Ω is assumed. Also, for both versions $\left\| G_n \left(\boldsymbol{\beta} \right) \right\| \overset{p}{\longrightarrow} 0$ only at $\boldsymbol{\beta} = \boldsymbol{\beta}_0$ in general if the number of components of $G_n \left(\boldsymbol{\beta} \right)$ is greater than the number of parameters of the model, i.e., $M > m$. For $\boldsymbol{\beta} \ne \boldsymbol{\beta}_0$ we have $0 < Q_n \left(\boldsymbol{\beta} \right) \le B$ for some $B > 0$ since survival functions evaluated at points are components of $G_n \left(\boldsymbol{\beta} \right)$ and these functions are bounded.

This implies that there exist real numbers u and v with $0 < u < v < \infty$ such that

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

Therefore, for both versions of $Q_n(\boldsymbol{\beta})$ whether deterministic or simulated, the minimum quadratic distance estimators MQD and SMQD estimators are consistent using Theorem 1, *i.e.*, the vector of MQD estimators and the vector MSQD estimators converge in probability to β_0. Theorem 3.1 of Pakes and Pollard [22] (pp. 1038-1039) is an elegant theorem, its proof is also concise using the norm concept of functional analysis and it allows many results to be unified. 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 [22] (pp. 1040-1043) which we shall restate their Theorem as Theorem 2 and Corollary 1 given subsequently after the following discussion on the ideas behind their Theorem which allow to get asymptotic normality results for estimators obtained from extremum of a smooth or nonsmooth objective function.

For both versions we can express $Q_n(\boldsymbol{\beta}) = \left(\|\boldsymbol{G}_n(\boldsymbol{\beta})\| \right)^2$, $\boldsymbol{G}_n(\boldsymbol{\beta})$ is as given by expression (14) for version D and as given by expression(15) for version S and since $\boldsymbol{G}_n(\boldsymbol{\beta})$ not differentiable for version S, the traditional Taylor expansion argument cannot be used to establish asymptotic normality of estimators obtained by minimizing $\left(\|\boldsymbol{G}_n(\boldsymbol{\beta})\| \right)^2$.

For both versions, $\boldsymbol{G}_n(\boldsymbol{\beta}) \xrightarrow{p} \boldsymbol{G}(\boldsymbol{\beta})$ with

$$\boldsymbol{G}(\boldsymbol{\beta}) = \left(z_{\boldsymbol{\beta}_0} - z_{\boldsymbol{\beta}} \right)'. \tag{20}$$

Explicitly,

$$\boldsymbol{G}(\boldsymbol{\beta}) = \left(S_{\boldsymbol{\beta}_0}(s_1,t_1) - S_{\boldsymbol{\beta}}(s_1,t_1), \cdots, S_{\boldsymbol{\beta}_0}(s_M,t_M) - S_{\boldsymbol{\beta}}(s_M,t_M) \right)'. \tag{21}$$

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

Note that $\boldsymbol{G}(\boldsymbol{\beta})$ is non-random and if we assume $\boldsymbol{G}(\boldsymbol{\beta})$ is differentiable with derivative matrix $\boldsymbol{\Gamma}(\boldsymbol{\beta})$, then we can define the random function $Q_n^a(\boldsymbol{\beta})$ to approximate $Q_n(\boldsymbol{\beta})$ for both versions in a unified way with

$$Q_n^a(\boldsymbol{\beta}) = \left(\|\boldsymbol{L}_n(\boldsymbol{\beta})\| \right)^2, \ \boldsymbol{L}_n(\boldsymbol{\beta}) = \boldsymbol{G}_n(\boldsymbol{\beta}_0) + \boldsymbol{\Gamma}(\boldsymbol{\beta}_0)(\boldsymbol{\beta}-\boldsymbol{\beta}_0). \tag{22}$$

The matrix $\boldsymbol{\Gamma}(\boldsymbol{\beta})$ can be displayed explicitly as

$$\boldsymbol{\Gamma}(\boldsymbol{\beta}) = - \begin{bmatrix} \dfrac{\partial S_{\boldsymbol{\beta}}(s_1,t_1)}{\partial \beta_1} & \cdots & \dfrac{\partial S_{\boldsymbol{\beta}}(s_1,t_1)}{\partial \beta_m} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial S_{\boldsymbol{\beta}}(s_M,t_M)}{\partial \beta_1} & \cdots & \dfrac{\partial S_{\boldsymbol{\beta}}(s_M,t_M)}{\partial \beta_m} \end{bmatrix}. \tag{23}$$

Note that $Q_n^a(\boldsymbol{\beta})$ is differentiable for both versions. Since $Q_n^a(\boldsymbol{\beta})$ is a quadratic function of $\boldsymbol{\beta}$, the vector $\boldsymbol{\beta}^*$ which minimizes $Q_n^a(\boldsymbol{\beta})$ can be obtained explicitly and

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

and since $\widehat{\boldsymbol{W}}_0 \xrightarrow{p} \boldsymbol{W}_0$, $\boldsymbol{W}_0$ is assumed to be a positive matrix, we have

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

Let $\tilde{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}^*$ be the vectors which minimize $Q_n\left(\boldsymbol{\beta}\right)$ and $Q_n^a\left(\boldsymbol{\beta}\right)$ respectively. If the approximation is of the right order then $\tilde{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}^*$ are asymptotically equivalent. This set up will allow a unified approach for establishing asymptotic normality for both versions. For version D, it suffices to let $\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}$ and for version S, let $\tilde{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}^S}$.

Clearly the set up fits into the scopes of their Theorem (3.3) which we shall rearrange the results of these two theorems before applying to version D and version S for MQD methods and verify that we can satisfy the regularity conditions of these two Theorems. We shall state Theorem 2 and Corollary 1 which are essentially their Theorem (3.3) and the proofs have been given by Pakes and Pollard [22] Note that the condition 4) is slightly more stringent but simpler than the condition iii) in their Theorem.

Also, for version S, the simulated samples are assumed to have size $U = \tau n$ and the same seed is used across different values of $\boldsymbol{\beta}$ to draw samples of size U. We make these assumptions and they are standard assumptions for simulated methods of inferences, see section 9.6 for method of simulated moments (MSM) given by Davidson and McKinnon [23] (pp. 383-394). For numerical optimization to find the minimum of the objective function, we rely on direct search simplex methods and the R package already has prewritten functions to implement direct search methods

Theorem 2.

Let $\tilde{\boldsymbol{\beta}}$ be a vector of consistent estimators for $\boldsymbol{\beta}_0$, the unique vector which satisfies $\boldsymbol{G}\left(\boldsymbol{\beta}_0\right) = \boldsymbol{0}$.

Under the following conditions:

1) The parameter space Ω is compact, $\tilde{\boldsymbol{\beta}}$ is an interior point of Ω.

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

3) $\boldsymbol{G}\left(.\right)$ is differentiable at $\boldsymbol{\beta}_0$ with a derivative matrix $\boldsymbol{\Gamma} = \boldsymbol{\Gamma}\left(\boldsymbol{\beta}_0\right)$ of full rank

4) $\sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|\le\delta_n}\sqrt{n}\left\|\boldsymbol{G}_n\left(\boldsymbol{\beta}\right) - \boldsymbol{G}\left(\boldsymbol{\beta}\right) - \boldsymbol{G}_n\left(\boldsymbol{\beta}_0\right)\right\| = o_p\left(1\right)$ for every sequence $\left\{\delta_n\right\}$ of positive numbers which converge to zero.

5) $\left\|\boldsymbol{G}_n\left(\boldsymbol{\beta}\right)\right\| = o_p\left(1\right)$.

6) $\boldsymbol{\beta}_0$ is an interior point of the parameter space Ω.

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

$$\sqrt{n}\left(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) = -\left(\boldsymbol{\Gamma}'\widehat{\boldsymbol{W}}_0\boldsymbol{\Gamma}\right)^{-1}\boldsymbol{\Gamma}'\widehat{\boldsymbol{W}}_0\sqrt{n}\boldsymbol{G}_n\left(\boldsymbol{\beta}_0\right) + o_p\left(1\right), \tag{24}$$

or equivalently, using equality in distribution,

$$\sqrt{n}\left(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) =^d -\left(\boldsymbol{\Gamma}'\boldsymbol{W}_0\boldsymbol{\Gamma}\right)^{-1}\sqrt{n}\boldsymbol{\Gamma}'\boldsymbol{W}_0\boldsymbol{G}_n\left(\boldsymbol{\beta}_0\right) \tag{25}$$

or equivalently,

$$\sqrt{n}\left(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right)=^d -\left(\boldsymbol{\Gamma}'\widehat{\boldsymbol{W}}_0\boldsymbol{\Gamma}\right)^{-1}\sqrt{n}\boldsymbol{\Gamma}'\widehat{\boldsymbol{W}}_0\boldsymbol{G}_n\left(\boldsymbol{\beta}_0\right) \tag{26}$$

The proofs of these results follows from the results used to prove Theorem 3.3 given by Pakes and Pollard [22] (pp. 1040-1043). For expression (13) or expression (14) to hold, in general only condition 5) of Theorem 2 is needed and there is no need to assume that $\boldsymbol{G}_n\left(\boldsymbol{\beta}_0\right)$ 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 $\boldsymbol{Y}_n = \sqrt{n}\boldsymbol{\Gamma}'\boldsymbol{W}_0\boldsymbol{G}_n\left(\boldsymbol{\beta}_0\right)$, if $\boldsymbol{Y}_n \xrightarrow{L} N\left(\boldsymbol{0},\boldsymbol{V}\right)$ then $\sqrt{n}\left(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right)\xrightarrow{L} N\left(\boldsymbol{0},\boldsymbol{T}\right)$ with

$$\boldsymbol{T} = \left(\boldsymbol{\Gamma}'\boldsymbol{W}_0\boldsymbol{\Gamma}\right)^{-1}\boldsymbol{V}\left(\boldsymbol{\Gamma}'\boldsymbol{W}_0\boldsymbol{\Gamma}\right)^{-1}, \tag{27}$$

The matrices $\boldsymbol{T}$ and $\boldsymbol{V}$ depend on $\boldsymbol{\beta}_0$, we also adopt the notations $\boldsymbol{T}=\boldsymbol{T}\left(\boldsymbol{\beta}_0\right),\boldsymbol{V}=\boldsymbol{V}\left(\boldsymbol{\beta}_0\right)$.

We observe that condition 4) of Theorem 2 when applies to SMQD methods in general involve technicalities. The condition 4) holds for version D, we only need to verifiy for version S. Note that to verify the condition 4, it is equivalent to verify

$$\sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|\leq\delta_n} n\left(\left\|\boldsymbol{G}_n\left(\boldsymbol{\beta}\right)-\boldsymbol{G}\left(\boldsymbol{\beta}\right)-\boldsymbol{G}_n\left(\boldsymbol{\beta}_0\right)\right\|\right)^2 = o_p\left(1\right),$$

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

$$g_n\left(\boldsymbol{\beta}\right) = n\left(\left\|\boldsymbol{G}_n\left(\boldsymbol{\beta}\right)-\boldsymbol{G}\left(\boldsymbol{\beta}\right)-\boldsymbol{G}_n\left(\boldsymbol{\beta}_0\right)\right\|\right)^2 \tag{28}$$

and for version S, using

$$\begin{aligned}\boldsymbol{u}_n\left(\boldsymbol{\beta}\right) = \Big(&\left(S^s_{\boldsymbol{\beta}_0}\left(s_1,t_1\right)-S_{\boldsymbol{\beta}_0}\left(s_1,t_1\right)\right)-\left(S^s_{\boldsymbol{\beta}}\left(s_1,t_1\right)-S_{\boldsymbol{\beta}}\left(s_1,t_1\right)\right)\\ &-\left(S^s_{\boldsymbol{\beta}_0}\left(s_M,t_M\right)-S_{\boldsymbol{\beta}_0}\left(s_M,t_M\right)\right)-\left(S^s_{\boldsymbol{\beta}}\left(s_M,t_M\right)-S_{\boldsymbol{\beta}}\left(s_M,t_M\right)\right)\Big)'\end{aligned} \tag{29}$$

Consequently, $g_n\left(\boldsymbol{\beta}\right)$ can also be expressed as

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

Since the elements of $\sqrt{n}\boldsymbol{u}'_n\left(\boldsymbol{\beta}\right)$ are bounded in probability, it is not difficult to see that the sequence $\left\{g_n\left(\boldsymbol{\beta}\right)\right\}$ is bounded in probability and continuous in probability with $g_n\left(\boldsymbol{\beta}\right)\xrightarrow{p} g_n\left(\boldsymbol{\beta}\right)$ as $\boldsymbol{\beta}\to\boldsymbol{\beta}'$, using the assumption that same seed is used across different values of $\boldsymbol{\beta}$ and assuming that $S_{\boldsymbol{\beta}}\left(x,y\right)$ is differentiable with respect to $\boldsymbol{\beta}$ and note that $g_n\left(\boldsymbol{\beta}_0\right)=0$. Therefore, results given in section of Luong et al. [24] (p. 218) can be used to justify the sequence of functions $\left\{g_n\left(\boldsymbol{\beta}\right)\right\}$ attains its maximum on the compact set

$C_n = \{ \boldsymbol{\beta} \big| \| \boldsymbol{\beta} - \boldsymbol{\beta}_0 \| \leq \delta_n \}$ in probability and hence has the property $\sup_{\| \boldsymbol{\beta} - \boldsymbol{\beta}_0 \| \leq \delta_n} g_n(\boldsymbol{\beta}) \xrightarrow{p} 0$ as $n \to \infty$ and $\boldsymbol{\beta} \to \boldsymbol{\beta}_0$.

We can see for version D as $V = W_0^{-1}$,

$$\sqrt{n} G_n(\boldsymbol{\beta}_0) \xrightarrow{p} N(0, W_0^{-1}). \tag{30}$$

For version S, note that

$$\left(z_n - z_{\boldsymbol{\beta}}^s \right)' = \left(z_n - z_{\boldsymbol{\beta}_0} - \left(z_{\boldsymbol{\beta}}^S - z_{\boldsymbol{\beta}_0} \right) \right)' = \left(z_n - z_{\boldsymbol{\beta}_0} \right)' - \left(z_{\boldsymbol{\beta}}^S - z_{\boldsymbol{\beta}_0} \right)',$$

we can see that

$$\sqrt{n} G_n(\boldsymbol{\beta}_0) \xrightarrow{p} N\left(0, \left(1 + \frac{1}{\tau} \right) W_0^{-1} \right) \tag{31}$$

as the simulated samples size is $U = \tau n$ and the simulated samples are independent of the original sample given by the data. Implicitly, we assume that the same seed is used across different values of $\boldsymbol{\beta}$ to obtain simulated samples.

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

$$\sqrt{n} \left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \right) \xrightarrow{L} N\left(0, (\Gamma' W_0 \Gamma)^{-1} \right), \tag{32}$$

Γ is as given by expression (23) which can be estimated easily.

For version S, the SMQD estimators also follow an asymptotic normal distribution with

$$\sqrt{n} \left(\hat{\boldsymbol{\beta}}^S - \boldsymbol{\beta}_0 \right) \xrightarrow{L} N\left(0, \left(1 + \frac{1}{\tau} \right) (\Gamma' W_0 \Gamma)^{-1} \right), \tag{33}$$

an estimate of Γ can be obtained using the technique as given by Pakes and Pollard [22] (p. 1043).

2.2. Model Testing

2.2.1. Simple Hypothesis

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

H_0: data comes from a specified distribution with distribution $F_{\boldsymbol{\beta}_0}$, $\boldsymbol{\beta}_0$ is specified. The chi-square test statistics, their asymptotic distributions and their degree of freedoms r are given below with

$$n Q_n(\boldsymbol{\beta}_0) \xrightarrow{L} \chi^2 (r = M) \text{ for version D and} \tag{34}$$

$$n \left(\frac{\tau}{\tau + 1} \right) Q_n(\boldsymbol{\beta}_0) \xrightarrow{L} \chi^2 (r = M) \text{ for version S.} \tag{35}$$

The version S is of interest since it allows testing goodness of fit for continuous distributions without closed form bivariate survival functions as we only need to be able to simulate from these distributions. We shall justify the asymptotic chi-square distributions given by expression (34) and expression (35) below.

Note that

$$nQ_n(\boldsymbol{\theta}_0) = \sqrt{n}\boldsymbol{G}'_n(\boldsymbol{\theta}_0)\widehat{\boldsymbol{W}}_0\sqrt{n}\boldsymbol{G}_n(\boldsymbol{\theta}_0) \quad \text{and for version D}$$

$$\sqrt{n}\boldsymbol{G}_n(\boldsymbol{\theta}_0) \xrightarrow{L} N(0, \boldsymbol{W}_0^{-1}), \ \boldsymbol{W}_0^{-1} = \boldsymbol{\Omega}_0.$$

For version S,

$$\sqrt{n}\boldsymbol{G}_n(\boldsymbol{\theta}_0) \xrightarrow{L} N\left(0, \left(1 + \frac{1}{\tau}\right)\boldsymbol{W}_0^{-1}\right).$$

We have the asymtoptic chi-square distributions as given, using standard results for distribution of quadratic forms and $\widehat{\boldsymbol{W}}_0 \xrightarrow{p} \boldsymbol{W}_0$, see Luong and Thompson [19] (p. 247) for example.

2.2.2. Composite Hypothesis

The quadratic distances $Q_n(\boldsymbol{\beta})$ can also be used for construction of the test satistics for the composite hypothesis, $Q_n(\boldsymbol{\beta})$ is as defined by expression (18) for version D and as defined by expression (19) for version S. The null hypothesis can be stated as

H$_0$: data comes from a parametric model $\{F_{\boldsymbol{\beta}}\}$. The chi-square test statistics are given by

$$nQ_n(\hat{\boldsymbol{\beta}}) \xrightarrow{L} \chi^2(r = M - m), \tag{36}$$

for version D and for version S,

$$n\left(\frac{\tau}{\tau+1}\right)Q_n(\hat{\boldsymbol{\beta}}^S) \xrightarrow{L} \chi^2(r = M - m) \tag{37}$$

where $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}^S$ are the vector of MQD and SMQD estimators which minimize $Q_n(\boldsymbol{\beta})$ version D and version S respectively and assuming $M > m$. To justify these asymptotic chi-square dsitributions, note that we have for version D, $nQ_n(\hat{\boldsymbol{\beta}}) = nQ_n^a(\hat{\boldsymbol{\beta}}) + o_p(1)$. It suffices to consider the asymptotic distribution of $nQ_n^a(\hat{\boldsymbol{\beta}})$ as we have the following equalities in distribution,

$$nQ_n(\hat{\boldsymbol{\beta}}) =^d nQ_n^a(\hat{\boldsymbol{\beta}}) = n\|L_n(\hat{\boldsymbol{\beta}})\|^2 = \sqrt{n}L'_n(\hat{\boldsymbol{\beta}})\widehat{\boldsymbol{W}}_0\sqrt{n}\boldsymbol{L}_n(\boldsymbol{\beta}), \ \boldsymbol{L}_n(\boldsymbol{\beta}) \text{ as given by}$$

expression (22). Also, using expressions (24-26).

$$\sqrt{n}\boldsymbol{L}_n(\hat{\boldsymbol{\beta}}) =^d \sqrt{n}\boldsymbol{G}_n(\boldsymbol{\beta}_0) + \boldsymbol{\Gamma}\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \text{ which can be reexpressed as}$$

$$\sqrt{n}\boldsymbol{L}_n(\hat{\boldsymbol{\beta}}) =^d \sqrt{n}\boldsymbol{G}_n(\boldsymbol{\beta}_0) - \boldsymbol{\Gamma}(\boldsymbol{\Gamma}'\boldsymbol{W}_0\boldsymbol{\Gamma})^{-1}\boldsymbol{\Gamma}'\boldsymbol{W}_0\sqrt{n}\boldsymbol{G}_n(\boldsymbol{\theta}_0) \text{ using expressions (25-26).}$$

or equivalently, $\sqrt{n}\boldsymbol{L}_n(\hat{\boldsymbol{\beta}}) =^d \left(\boldsymbol{I} - \boldsymbol{\Gamma}(\boldsymbol{\Gamma}'\boldsymbol{W}_0\boldsymbol{\Gamma})^{-1}\boldsymbol{\Gamma}'\boldsymbol{W}_0\right)\sqrt{n}\boldsymbol{G}_n(\boldsymbol{\theta}_0)$ with

$$\sqrt{n}\boldsymbol{G}_n(\boldsymbol{\theta}_0) \xrightarrow{L} N(0, \boldsymbol{W}_0^{-1}).$$

We have $\sqrt{n}\boldsymbol{L}_n(\hat{\boldsymbol{\beta}}) \xrightarrow{L} N(0, \boldsymbol{\Sigma})$,

$\boldsymbol{\Sigma} = \left(\boldsymbol{I} - \boldsymbol{\Gamma}(\boldsymbol{\Gamma}'\boldsymbol{W}_0\boldsymbol{\Gamma})^{-1}\boldsymbol{\Gamma}'\boldsymbol{W}_0\right)\boldsymbol{W}_0^{-1}\left(\boldsymbol{I} - \boldsymbol{W}_0\boldsymbol{\Gamma}(\boldsymbol{\Gamma}'\boldsymbol{W}_0\boldsymbol{\Gamma})^{-1}\boldsymbol{\Gamma}'\right)$ and note that $\boldsymbol{\Sigma}\boldsymbol{W}_0 = \boldsymbol{B}$

and the trace of the matrix $B = \boldsymbol{I} - \boldsymbol{\Gamma}(\boldsymbol{\Gamma}'\boldsymbol{W}_0\boldsymbol{\Gamma})^{-1}\boldsymbol{\Gamma}'\boldsymbol{W}_0$ is $trace(\boldsymbol{B}) = M - m$; the rank of the matrix $\boldsymbol{B}$ is also equal to its trace. The argument used is very similar to the one used for the Pearson's statistics, see Luong and Thompson [19] (pp. 248-249).

Similarly, for version S, $Q_n(\hat{\boldsymbol{\beta}}^S) =^d nQ_n^a(\hat{\boldsymbol{\beta}}^S) = n\|L_n(\hat{\boldsymbol{\beta}}^S)\|^2$ and

$$\sqrt{n}\boldsymbol{L}_n(\hat{\boldsymbol{\beta}}^S) =^d \left(\boldsymbol{I} - \boldsymbol{\Gamma}(\boldsymbol{\Gamma}'\boldsymbol{W}_0\boldsymbol{\Gamma})^{-1}\boldsymbol{\Gamma}'\boldsymbol{W}_0\right)\sqrt{n}\boldsymbol{G}_n(\boldsymbol{\beta}_0) \text{ with}$$

$\sqrt{n}G_n(\boldsymbol{\beta}_0) \xrightarrow{L} N\left(0, \left(1+\frac{1}{\tau}\right)\boldsymbol{W}_0^{-1}\right)$. This justifies the asymptotic chi-square distributions as given by expression (36) and expression (37).

3. Estimation and Model Testing Using Complete Data

3.1. Preliminaries: Statistical Functional and Its Influence Function

In section (3.1) and section (3.2), we shall define a rule of selecting the points $(s_l, t_l), l = 1, \cdots, M$ if complete data are available. Equivalently, we would like to define the cells used to group the data and we shall see that random cells will be used as the points $(s_l, t_l), l = 1, \cdots, M$ 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. 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 [16] (pp. 194-208). We need a few tools to develop such a rule. We shall define sample quantiles then statistics can be viewed as functionals of the sample distribution; their influence functions are also needed and it allows us to find their asymptotic variance.

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.

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 $F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I[x_i \leq x]$ and it is defined to be $\alpha_p^{(n)} = \inf\{F_n(x) \geq p\}$ and its model counterpart is given by $\alpha_p = \inf\{F(x) \geq p\}$. 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 quantile functions $\alpha_p^{(n)}$ or $\beta_q^{(n)}$ can be viewed as statisticaf 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 δ_x be the degenerate distribution at x, *i.e.*, $\delta_x(u) = 1$ if $u \geq x$ and $\delta_x(u) = 0$, otherwise; 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 $(\delta_x - H)$ and by letting $H_\varepsilon = H + \varepsilon(\delta_x - H)$, *i.e.*, $IC_{T,H}(x) = \lim_{\varepsilon \to 0} \frac{T(H_\varepsilon) - T(H)}{\varepsilon} = T'_H(\delta_x - H)$ and T'_H is a linear functional.

Alternatively, it is easy to see that $IC_{T,H}(x) = \left.\frac{\partial H_\varepsilon}{\partial \varepsilon}\right|_{\varepsilon=0}$ and this gives a con-

venient 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 [25] (p. 56), Hogg *et al.* [6] (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 $IC_{T,H}(x)$ is based on a linear functional, the asymptotic variance of $T(H_n)$ is simply $\frac{1}{n}V(IC_{T,H}(x))$ with $V(.)$ being the variance of the expression inside the bracket since in general we have $E(IC_{T,H}(x)) = 0$ and the following representation when $IC_{T,H}(x)$ 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.* [6] (p. 593). Consequently, in general we have for bounded influence functionals with the use of means of central limit theorems (CLT) the following convergence in distribution,

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

The influence function representation of a functional which depends only on one function such as H_n is the equivalent of a Taylor expansion of a univariate function and the influence function representation of a functional which depends on many functions is the equivalent of a Taylor expansion of a multivariate function with domain in an Euclidean space and having range being the real line. We will encounter an example of functionals which depend on three functions S_n, F_n, G_n in section (3.2).

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 $(u_l, v_l) = (\varphi_{b_1}(l), \varphi_{b_2}(l)), l = 1, 2, \cdots, M$, we also use H_M to denote set of points $\{(u_l, v_l), l = 1, 2, \cdots, M\}$. 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 $b_1 = 3$, 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 functions to generate the sequences and see Glaserman [26] (pp. 293-297) for the related pseudo codes; also see the seminal paper by Halton [27]. For the general principles of QMC methods, see Glasserman [26] (pp. 281-292). The Halton sequences 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 sample quantiles which are robust statistics, 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 sequences are deterministic 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 can be useful for integrating a function of the form $A = \int_0^1 \int_0^1 \psi(x,y)\,\mathrm{d}x\mathrm{d}y$. Using the M terms of the Halton sequences it can be approximated as $A \approx \frac{1}{M}\sum_{l=1}^{M}\psi(s_l,t_l)$ which is similar to the sample mean from a random sample of size M.

From observations which are given by $Z_i = (X_i, Y_i)', i = 1,\cdots,n$ iid with common bivariate distribution function $K_{\boldsymbol{\beta}}(x,y)$ and survival function $S_{\boldsymbol{\beta}}(x,y)$. Let the two marginal distributions of $K_{\boldsymbol{\beta}}(x,y)$ be denoted by $F_{\boldsymbol{\beta}}(x)$ and $G_{\boldsymbol{\beta}}(y)$ and also let $F = F_{\boldsymbol{\beta}_0}$ and $G = G_{\boldsymbol{\beta}_0}, K = K_{\boldsymbol{\beta}_0}, S = S_{\boldsymbol{\beta}_0}$.

Define the bivariate empirical distribution function which is similar to the bivariate empirical survival function as

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

We might want to think that it admits a bivariate empirical density estimate $k_n(x,y)$ so that the following Cramer-Von Mises distances expressions are equivalent,

$$\int_0^\infty \int_0^\infty \left(S_n(x,y) - S_{\boldsymbol{\beta}}(x,y)\right)^2 \mathrm{d}K_n(x,y)$$

is equivalent to

$$\int_0^\infty \int_0^\infty \left(S_n(x,y) - S_{\boldsymbol{\beta}}(x,y)\right)^2 k_n(x,y)\mathrm{d}x\mathrm{d}y.$$

For univariate Cramér-Von Mises methods, see Luong and Blier-Wong [28].

In the next section we shall give details on how to form a type of quasi sample or artificial sample of size M from $k_n(x,y)$ using the Halton sequence of M terms and the pth-sample quantiles of the marginal distributions F and G which allow us to define the sequence $(s_l, t_l), l = 1,\cdots,M$ so that the above integrals can be approximated by the following finite sum of the type of an average of M terms,

$$\frac{1}{M}\sum_{l=1}^{M}\left(S_n(s_l,t_l) - S_{\boldsymbol{\beta}}(s_l,t_l)\right)^2. \tag{38}$$

We can see the expression (38) is an unweighted quadratic distance using the identity matrix $\boldsymbol{I}$ as weight matrix instead of $\widehat{\boldsymbol{W}}_0$. The unweighted quadratic distance still produces consistent estimators but possibly less efficient estimators than estimators using the quadratic distance with $\widehat{\boldsymbol{W}}_0$ for large samples and for finite samples the estimators obtained using $\boldsymbol{I}$ might still have reasonable performances and yet being simple to obtained.

The set of points $(s_l, t_l), l = 1, \cdots, M$ 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 overlapping 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 [18] (pp. 324-326) and Moore and Spruill [17] for the notion of random cells; quadratic distance methods will share the same properties that is to say the fact that the chosen points are random but it will be shown that they do stabilize and therefore these random points can be viewed as fixed points since they do not affect efficiencies of the estimators or asymptotic distributions of goodness-of-fit test statistics which make use of them. These properties will be 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 $(s_l, t_l), l = 1, \cdots, M$ on the nonegative quadrant which give us a guideline on how to choose points if complete data is available.

3.2. Halton Sequences and an Artificial Sample

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

Let $\eta = \dfrac{1}{\max(u_l, l = 1, \cdots, M)}$ and $\varrho = \dfrac{1}{\max(v_l, l = 1, \cdots, M)}$, we can form the artificial sample with elements given by $(s_l, t_l), l = 1, \cdots, M$ with $s_l = \eta u_l F_n^{-1}(p), t_l = \varrho v_l G_n^{-1}(p)$ with $0.90 \leq p \leq 0.99$. We can view $(s_l, t_l), l = 1, \cdots, M$ being a form of quasi random sample on the nonegative quadrant and these are the points proposed to be used in case of complete data is avalaible. In general, we might want to choose $20 \leq M \leq 30$ if $M^2 \leq n$ and if n is small we try to ensure $M \leq \sqrt{n}$. Consequently as $n \to \infty$, M remains bounded.

Since $F_n^{-1}(p) \xrightarrow{\ p\ } F^{-1}(p)$ and $G_n^{-1}(p) \xrightarrow{\ p\ } G^{-1}(p)$, $(s_l, t_l) \xrightarrow{\ p\ } (s_l^0, t_l^0)$ with $s_l^0 = \eta u_l F^{-1}(p)$ and $t_l^0 = \varrho t_l G^{-1}(p)$ for $l = 1, \cdots, M$ and the points $(s_l^0, t_l^0), l = 1, \cdots, M$ 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 [18] (pp. 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 $\|G_n(\boldsymbol{\beta})\|^2$ as defined by expression (16) for version D and expression(18) for version S. Now to emphasize z_n and $z_{\boldsymbol{\beta}}$ which depend on $\{(s,t)\}$, we also use respectively the notations $z_n(\{(s,t)\})$ and $z_{\boldsymbol{\beta}}(\{(s,t)\})$ and let $z_n^0 = z_n(\{(s^0,t^0)\})$, $z_{\boldsymbol{\beta}}^0 = z_{\boldsymbol{\beta}}(\{(s^0,t^0)\})$.

It suffices to verify that results of Theorem 1, Theorem 2 and its corollary in section (2) continue to hold.

Now observe that for both versions D and S we have

$$\left(z_n - z_{\boldsymbol{\beta}}\right) \overset{p}{\longrightarrow} \left(z_{\boldsymbol{\beta}_0}^0 - z_{\boldsymbol{\beta}}^0\right) \tag{39}$$

and

$$\left(z_n - z_{\boldsymbol{\beta}}^S\right) \overset{p}{\longrightarrow} \left(z_{\boldsymbol{\beta}_0}^0 - z_{\boldsymbol{\beta}}^0\right) \tag{40}$$

so that $G(\boldsymbol{\beta})$ remains the same for both versions since $S_{\boldsymbol{\beta}}(x,y)$ is continuous with respect to (x,y), $S_n(x,y) \overset{p}{\longrightarrow} S_{\boldsymbol{\beta}}(x,y)$ and $\{(s,t)\} \overset{p}{\longrightarrow} \{(s^0,t^0)\}$. Clearly, $\widehat{W_0}(\{(s,t)\}) \overset{p}{\longrightarrow} W_0(s^0,t^0)$. It remains to establish $\sqrt{n}(z_n - z_{\boldsymbol{\beta}}) = \sqrt{n}(z_n^0 - z_{\boldsymbol{\beta}}^0) + o_p(1)$ and note that $z_{\boldsymbol{\beta}}$ is random instead of being fixed in section (2) as we are using an adaptive basis here. Using results on the influence functions representations for functionals as discussed, it suffices to show that the vector $(z_n - z_{\boldsymbol{\beta}})$ has the same influence representation as the vector $(z_n^0 - z_{\boldsymbol{\beta}}^0)$ 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 functionals $(z_n - z_{\boldsymbol{\beta}})$ and show that it is the same for the corresponding elements of the vector of functionals $(z_n^0 - z_{\boldsymbol{\beta}}^0)$.

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

$$F_{\varepsilon_1}(u) = F(u) + \varepsilon_1\left(\delta_x(u) - F(u)\right), 0 \le \varepsilon_1 \le 1.$$

Similarly,

$$G_{\varepsilon_2}(v) = G(v) + \varepsilon_2\left(\delta_y(v) - G(v)\right), 0 \le \varepsilon_2 \le 1.$$

Now we consider $\left(z_{jn}-z_{j\boldsymbol{\beta}}\right)$ the jth element of $\left(z_n-z_{\boldsymbol{\beta}_0}\right)$,

$$\left(z_{jn}-z_{j\boldsymbol{\beta}_0}\right)=S_n\left(s_j\left(F_n\right),t_j\left(G_n\right)\right)-S\left(s_j\left(F_n\right),t_j\left(G_n\right)\right)=T_j\left(S_n,F_n,G_n\right), j=1,\cdots,M .$$

Clearly, $T_j\left(S_n,F_n,G_n\right)$ depend on S_n,F_n,G_n but we can use the influence function representation as given by Reid [29] which allows the asymptotic representation with three influence functions given by

$$\left.\frac{\partial T_j\left(S_\varepsilon,F_{\varepsilon_1},G_{\varepsilon_2}\right)}{\partial\varepsilon}\right|_{\varepsilon=\varepsilon_1=\varepsilon_2=0}=I\left[x>s_l^0, y>t_l^0\right]-S\left(s_l^0,t_l^0\right) \text{ which is bounded with}$$

respect to (x,y), $\left.\dfrac{\partial T_j\left(S_\varepsilon,F_{\varepsilon_1},G_{\varepsilon_2}\right)}{\partial\varepsilon_1}\right|_{\varepsilon=\varepsilon_1=\varepsilon_2=0}$ and $\left.\dfrac{\partial T_j\left(S_\varepsilon,F_{\varepsilon_1},G_{\varepsilon_2}\right)}{\partial\varepsilon_2}\right|_{\varepsilon=\varepsilon_1=\varepsilon_2=0}$.

It is interesting to note that $\left.\dfrac{\partial T_j\left(S_\varepsilon,F_{\varepsilon_1},G_{\varepsilon_2}\right)}{\partial\varepsilon_1}\right|_{\varepsilon=\varepsilon_1=\varepsilon_2=0}=0$ and

$$\left.\frac{\partial T_j\left(S_\varepsilon,F_{\varepsilon_1},G_{\varepsilon_2}\right)}{\partial\varepsilon_2}\right|_{\varepsilon=\varepsilon_1=\varepsilon_2=0}=0 ; \text{ so they do not contribute to the influence function}$$

representation of $T_j\left(S_n,F_n,G_n\right)$.

If we consider the jth term of $\left(z_n^0-z_{\boldsymbol{\beta}_0}^0\right)$ given by the functional $G_j\left(S_n\right)=S_n\left(s_l^0,t_l^0\right)-S\left(s_l^0,t_l^0\right)$ which does not depend on F_n and G_n, we find that

$$\left.\frac{\partial G_j\left(S_\varepsilon\right)}{\partial\varepsilon}\right|_{\varepsilon=\varepsilon_1=\varepsilon_2=0}=\left.\frac{\partial T_j\left(S_\varepsilon,F_{\varepsilon_1},G_{\varepsilon_2}\right)}{\partial\varepsilon}\right|_{\varepsilon=\varepsilon_1=\varepsilon_2=0} .$$

Therefore, all the asymptotic results of section (2) remain valid and all these influence 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 $\left\{\left(s^0,t^0\right)\right\}$ and if needed they can be replaced by $\left\{(s,t)\right\}$ without affecting the asymptotic results already established in section (2).

4. An Illustration and a Simulation Study

For iilustration, we consider a bivariate gamma model as discussed in section (1.2) introduced by Mathai and Moschopoulos [11] (pp. 137-139) which is a model with 5 parameters given by the vector $\boldsymbol{\beta}=\left(\alpha_0,\alpha_1,\alpha_2,\beta_1,\beta_2\right)'$. Mathai and Moschopoulos [11] (pp. 138-140) also give the following model moments using the bivariate Laplace transform of the model and replacing them with the corresponding empirical moments and solving a system of equations give the method of moments (MM) estimators.

The moments being considered are given by

$$E(X)=\left(\alpha_0+\alpha_1\right)\beta_1, V(X)=\left(\alpha_0+\alpha_1\right)\beta_1^2, E\left(X-E(X)\right)^3=2\left(\alpha_0+\alpha_1\right)\beta_1^3,$$

they are the first three cumulants of the marginal distribution of X; the first three cumulants of the marginal distribution of Y are given by

$$E(Y) = (\alpha_0 + \alpha_2)\beta_2, \ V(Y) = (\alpha_0 + \alpha_2)\beta_2^2, \ E(Y - E(Y))^3 = 2(\alpha_0 + \alpha_2)\beta_2^3,$$

and the covariance between X and Y, $cov(X,Y) = \alpha_0\beta_1\beta_2$.

Let $\bar{X}, s_X^2, m_3^{(1)}$ be the first three sample cumulants of the marginal distribution of X and similarly let $\bar{Y}, s_Y^2, m_3^{(2)}$ be the first three sample cumulants of the marginal distribution of Y and finally let s_{XY} be the sample covariance between X and Y; the MM estimators for $\boldsymbol{\beta}$ as given by Mathai and Moschopoulos [11] (p. 151) can be expressead as

$$\widetilde{\beta_1} = \frac{m_3^{(1)}}{2s_X^2}, \ \widetilde{\beta_2} = \frac{m_3^{(2)}}{2s_Y^2}, \ \widetilde{\alpha_0} = \frac{s_{XY}}{\widetilde{\beta_1}\widetilde{\beta_2}}$$

and

$$\widetilde{\alpha_1} = \frac{s_X^2}{\widetilde{\beta_1}^2} - \widetilde{\alpha_0}, \ \widetilde{\alpha_2} = \frac{s_Y^2}{\widetilde{\beta_2}^2} - \widetilde{\alpha_0}.$$

As the MM estimators depend on statistics of polynomials of degree 3, they will not be robust and they are very sensitive to outliers.

As the bivariate density function for the model is complicated, we consider version S for quadratic distance with $M = 25$ and we compare the performance of MM estimators versus SMQD estimators using the overall asymptotic relative criterion

$$ARE = \frac{MSE\left(\widehat{\alpha_0^S}\right) + MSE\left(\widehat{\alpha_1^S}\right) + MSE\left(\widehat{\alpha_2^S}\right) + MSE\left(\widehat{\beta_1^S}\right) + MSE\left(\widehat{\beta_2^S}\right)}{MSE\left(\widetilde{\alpha_0}\right) + MSE\left(\widetilde{\alpha_1}\right) + MSE\left(\widetilde{\alpha_2}\right) + MSE\left(\widetilde{\beta_1}\right) + MSE\left(\widetilde{\beta_2}\right)}.$$

The mean square errors (*MSE*) and *ARE* are estimated using simulated samples. For references on simulations, we find Ross [30] and Johnson [31] useful. The mean square error of an estimator $\hat{\pi}$ for π_0 is defined as

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

We fix $M = 25$, the two samples quantiles are 0.99 quantiles and we do not have problem to obtain the inverse matrix used as optimum matrix estimated from data. If we fix $M = 30$, occasionally the matrix $\hat{\boldsymbol{\Omega}}_0$ is nearly singular and the package R give us a warning sign. It takes around one to two minutes to complete one run on a laptop computer as we do not have large computers, so we fix $N = 50$ replications with each sample of size $n = 500$. We observe that in general, each individual ARE is smaller for SMQD estimators than for MM estimators and over all more efficient and robust than MM estimators confirming asymptotic theory; also note that MM estimators require that complete data is available. The unweighted simulated quadratic distance estimators using $\boldsymbol{I}$ perform almost as well as the estimators obtained using an optimum weight matrix. The range of parameters are fixed as follows, we let $\alpha_0 = 2$, $\alpha_1 = \alpha_2$, $\beta_1 = \beta_2$, $\alpha_1 = 2, 3, 4, 6, 8, 9, 10$ and $\beta_1 = 2, 3, 4, 6, 8, 9, 10$.

The chosen ranges are often encountered in actuarial science, we observe that in general when the samples are not contaminated the SMQD estimators are at least

five times better than their counterparts, the results are summarized in the upper table of Table 1.

For robustness study we contaminate the distributions used above with 5% of observations coming from a bivariate gamma with parameters outside the range being considered used to generate outliers. The bivariate gamma distribution used to generate outliers has parameters given by the vector $\beta^c = \left(\alpha_0^c = 2, \alpha_1^c = 2, \alpha_2^c = 2, \beta_1^c = 50, \beta_2^c = 50 \right)'$ .We observe that SMQD estimators are at least 1000 times more efficient than MM estimators, we just display a row for illustrations of the order of efficiency gained by using SMQD methods and it is given at the bottom of Table 1. The limited study seems to point to better efficiencies and robustness for SMQD estimators than MM estimators. MM estimators are very vulnerable to outliers as expected. Despite the study is limited but it tends to confirm theoretical asymptotic results; more numerical and simulation works need to be done to further study the performances of the estimators proposed especially the performances in finite samples.

5. Conclusion

Minimum Quadratic distance methods offer a unified and numerical efficient

Table 1. Overall Asymptotic relative efficiencies comparisons between SMQD estimators and MM estimators using noncontaminated samples of size $n = 500$.

$$\alpha_0 = 2, \beta_1 = \beta_2 = b, \alpha_1 = \alpha_2 = a$$

$a \therefore b$	2	3	4	6	8	9	10
2	0.0372	0.0264	0.0446	0.0981	0.1595	0.1450	0.1807
3	0.0311	0.0039	0.0036	0.0028	0.0027	0.0288	0.0512
4	0.0448	0.0621	0.0376	0.0444	0.0790	0.0392	0.0565
6	0.0110	0.0528	0.0064	0.0031	0.0022	0.0158	0.0948
8	0.0063	0.0084	0.0046	0.0222	0.0039	0.0152	0.0949
9	0.0027	0.0039	0.0035	0.0072	0.0028	0.0058	0.0024
10	0.0055	0.0044	0.0649	0.0105	0.0195	0.0145	0.0180

$$ARE = \frac{MSE\left(\widehat{\alpha_0^s}\right) + MSE\left(\widehat{\alpha_1^s}\right) + MSE\left(\widehat{\alpha_2^s}\right) + MSE\left(\widehat{\beta_1^s}\right) + MSE\left(\widehat{\beta_2^s}\right)}{MSE\left(\widetilde{\alpha_0}\right) + MSE\left(\widetilde{\alpha_1}\right) + MSE\left(\widetilde{\alpha_2}\right) + MSE\left(\widetilde{\beta_1}\right) + MSE\left(\widetilde{\beta_2}\right)}$$

Legend: Tabulated values are estimates of overall ARE (SMQD vs MM) based on simulated samples. Overall Asymptotic relative efficiencies comparisons between SMQD estimators and MM estimators using contaminated samples of size $n = 500$ with 5% contamination

$$\alpha_0 = 2, \beta_1 = \beta_2 = b, \alpha_1 = \alpha_2 = a$$

$a \therefore b$	2	3	4	6	8	9	10
6	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Legends: Tabulated values are estimates of overall ARE (SMQD vs MM) based on simulated contaminated samples. Samples are drawn from a contaminated model of the form pbivariate-Gamma $\left(\alpha_0 = 2, \beta_1 = \beta_2 = b, \alpha_1 = \alpha_2 = a \right)$ + qbivariate-Gamma $\left(\alpha_0 = 2, \beta_1 = \beta_2 = 50, \alpha_1 = \alpha_2 = 2 \right)$, $p = 0.95$, $q = 0.05$.

approach for estimation and model testing using grouped data and unify minimum chi-square methods with fixed or random cells based on the notion of projected score functions on finite bases and adaptive bases. The simulated version is especially suitable for handling bivariate distributions where numerical complications might arise. The rule on how to select points on the nonnegative quadrant for using minimum quadratic distance methods is also more clearly defined whereas it is not clear on how to form random cells for minimum chi-square methods. The methods appear to be relatively simple to implement and yet being more efficient and more robust than methods of moments in general especially when the model has more than three parameters.

Acknowledgements

References

[1] Partrat, C. (1995) Compound Model for Two Dependent Kinds of Claims. *Insurance: Mathematics and Economics*, **15**, 219-231. https://doi.org/10.1016/0167-6687(94)90796-X

[2] Gibbons, J.D. and Chakraborti, S. (2011) Nonparametric Statistical Inference. Fifth Edition, CRC Press, Boca Raton. https://doi.org/10.1007/978-3-642-04898-2_420

[3] Marshall, A.W. and Olkin, I. (1988) Families of Multivariate Distributions. *Journal of the American Statistical Association*, **83**, 834-841. https://doi.org/10.1080/01621459.1988.10478671

[4] Klugman, S.A., Panjer, H.H. and Willmot, G.E. (2012) Loss Models: From Data to Decisions. Fourth Edition, Wiley, New York.

[5] Klugman, S.A. and Parsa, A. (1999) Fitting Bivariate Distributions with Copulas. *Insurance: Mathematics and Economics*, **24**, 139-148. https://doi.org/10.1016/S0167-6687(98)00039-0

[6] Hogg, R.V., McKean, J.W. and Craig, A.T. (2013) Introduction to Mathematical Statistics. Pearson, New York.

[7] Mardia, K.V. (1967) Some Contributions to Contingency Type Bivariate Distributions. *Biometrika*, **56**, 449-451.

[8] Mardia, K.V. (1970) Families of Bivariate Distributions. Griffin, London.

[9] Plackett, R.L. (1965) A Class of Bivariate Distributions. *Journal of the American Statistical Association*, **60**, 516-522. https://doi.org/10.1080/01621459.1965.10480807

[10] Marshall, A.W. and Olkin, I. (1990) Multivariate Distributions Generated from Mixture of Convolution and Product Families. Lectures Notes and Monograph Series, Volume 16, Institute of Mathematical Statistics, 371-393. https://doi.org/10.1214/lnms/1215457574

[11] Mathai, A.M. and Moschopoulos, P.G. (1991) On a Multivariate Gamma. *Journal of Multivariate Analysis*, **39**, 135-153. https://doi.org/10.1016/0047-259X(91)90010-Y

[12] Furman, E. (2008) On a Multivariate Gamma Distribution. *Statistics and Probability Letters*, **78**, 2353-2360. https://doi.org/10.1016/j.spl.2008.02.012

[13] Hutchinson, T. and Lai, C. (1990) Continuous Bivariate Distributions Emphasizing Applications. Rumbsby, Adelaide.

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

[15] 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

[16] Greenwood, P. and Nikulin, M.S. (1996) A Guide to Chi-Square Testing. Wiley, New York.

[17] 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

[18] Pollard, D. (1979) General Chi-Square Goodness-of-Fit Tests with Data Dependent Cells. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, **50**, 317-331. https://doi.org/10.1007/BF00534153

[19] Luong, A. and Thompson, M.E. (1987) Minimum Quadratic Distance Methods Based on Quadratic Distances for Transforms. *Canadian Journal of Statistics*, **15**, 239-251. https://doi.org/10.2307/3314914

[20] Duchesne, T., Rioux, J. and Luong, A. (1997) Minimum Cramér-Von Mises Distance Methods for Complete and Grouped Data. *Communications in Statistics*, *Theory and Methods*, **26**, 401-420. https://doi.org/10.1080/03610929708831923

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

[22] Pakes, A. and Pollard, D. (1989) Simulation Asymptotic of Optimization Estimators. *Econometrica*, **57**, 1027-1057. https://doi.org/10.2307/1913622

[23] Davidson, R. and MacKinnon, J.G. (2004) Econometric Theory and Methods. Oxford University Press, Oxford.

[24] Luong, A., Bilodeau 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

[25] Huber, P. (1981) Robust Statistics. Wiley, New York. https://doi.org/10.1002/0471725250

[26] Glasserman, P. (2003) Monte Carlo Methods in Financial Engineering. Springer, New York. https://doi.org/10.1007/978-0-387-21617-1

[27] Halton, J. (1960) On the Efficiency of Certain Quasi Random Sequences of Points in Evaluating Multi-Dimensional Integrals. *Numerische Mathematik*, **2**, 84-90. https://doi.org/10.1007/BF01386213

[28] 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. https://doi.org/10.4236/ojs.2017.75058

[29] Reid, N. (1981) Influence Functions for Censored Data. *Annals of Statistics*, **9**, 78-92. https://doi.org/10.1214/aos/1176345334

[30] Ross, S.M. (2013) Simulation. Fifth Edition, Elsevier, New York.

[31] Johnson, M.E. (1986) Multivariate Statistical Simulation. Wiley, New York.