REVSTAT – Statistical Journal
Volume 0, Number 0, Month 0000, 000–000
https://doi.org/00.00000/revstat.v00i0.000

Inferences on the Association Parameter in FGM Copula Based Bivariate Distribution: An Application to Water Quality Data

Authors: Rahul Chatterjee
– Department of Mathematics,
University of Louisiana at Lafayette, USA
chatterjee.rahul@epa.gov; rxc9305@gmail.com

Nabendu Pal ✉
– Faculty of Mathematics and Statistics
Ton Duc Thang University, Vietnam
nabendu.pal@tdtu.edu.vn
and
Department of Mathematics,
University of Louisiana at Lafayette, USA
nabendu.pal@louisiana.edu

Received: May 2023 Revised: September 2024 Accepted: November 2024

Abstract:

- This paper examines the inferential aspects of the association parameter in a bivariate distribution using the Farlie-Gumbel-Morgernstern Copula ($FGMC$). While most studies focus on the association parameter's relationship with correlation measures, this work highlights its inferential properties. Through extensive analysis, we show that the maximum likelihood estimator (MLE) of the parameter performs poorly, whereas Bayes estimators with noninformative priors offer superior results. We also demonstrate how the $FGMC$-based bivariate distribution effectively models arsenic contamination in groundwater, offering a valuable alternative to the traditional yet sometimes questionable use of multivariate normal models in such applications.

Keywords:

- *copula; maximum likelihood estimation; Bayes estimation; noninformative prior; parametric bootstrap; asymptotic tests.*

AMS Subject Classification:

- 62F10, 62F15, 62C05.

✉ Corresponding author

1. INTRODUCTION

1.1. Preliminaries

The normal distribution, though widely popular and heavily used in modeling datasets, has its own limitations. In the univariate case, when the variable of interest is positively skewed, one can use a host of other distributions such as Gamma, Weibull, Lognormal etc., just to name a few. However, in a multivariate set-up, the multivariate normal distribution appears to be the default choice, either by omission or by commission. The multivariate normal model has a host of advantages as its inferential problems are well studied, and the sampling distributions of its key statistics are relatively convenient to deal with. To be precise, let $\mathbf{X} = (X_1, X_2, ..., X_p)'$ be a p-variate random vector whose distribution is assumed to be $N_p(\boldsymbol{\mu}, \Sigma)$ where $\boldsymbol{\mu} \in \mathbb{R}^p$ and $\Sigma = ((\sigma_{ij})) > 0$ (p.d.). Based on a random sample $\mathbf{X}_i$, $1 \leq i \leq n$, (i.e., n copies of $\mathbf{X}$), assuming $n > p$, the maximum likelihood estimators (MLEs) of $\boldsymbol{\mu}$ and Σ are respectively $\hat{\boldsymbol{\mu}} = \overline{\mathbf{X}} = \sum_{i=1}^n \mathbf{X}_i/n$, and $\hat{\Sigma} = S/n$, where $S = \sum_{i=1}^n (\mathbf{X}_i - \overline{\mathbf{X}})(\mathbf{X}_i - \overline{\mathbf{X}})'$. Further, under the above normality of $\boldsymbol{X} = (X_1, ..., X_p)'$, it is well known that $E(X_1|X_2, ..., X_p) = \beta_1 + \sum_{k=2}^p \beta_k X_k$, for a suitable value of $\boldsymbol{\beta} = (\beta_1, \beta_2, ..., \beta_p)'$ which depends on $\boldsymbol{\mu}$ and Σ, and this is the motivation behind the usual multiple linear regression where X_1 is intended to be explained as a linear function of $(X_2, ..., X_p)$ subject to some variation. But what happens if $\mathbf{X}$ does **not** follow $N_p(\mu, \Sigma)$?
(a): Can we have the aforementioned $\overline{\mathbf{X}}$ and S/n as the MLE of $\boldsymbol{\mu}$ and Σ? - Possibly not.
(b): Can we have the independence of $\overline{\mathbf{X}}$ and S (which is the foundation of most of the normality based inferential results)? - Definitely not. (Kagan et al. (1973))
(c): Does regressing X_1 on $(X_2, ..., X_p)$ through a linear function make sense? - Certainly not, since $E(X_1|X_2, ..., X_p)$ would not be linear at all if the distribution of $\mathbf{X}$ is non-elliptically symmetric (assuming that the conditional expectation exists) Ley et al. (2021).

Yet, in multivariate modelling, ranging from psychology to anthropology, from agriculture to environmental science, especially in a 'Big Data' setting, multivariate normal distribution is being used hastily without paying closer attention to whether such model fitting is appropriate or not.

If the multivariate normal is found to be inappropriate for the data $\boldsymbol{X}_i, 1 \leq i \leq n$, then one may transform the variable(s) suitably hoping that the transformed data would follow normal. But there are two major issues with such transformation. There is no magic formula to tell us what transformation would be suitable for normality. Secondly, often such transformed variables are hard to interpret, and they lose significance to the original problem which gave rise to the data set to begin with.

This study has been motivated by several data sets where component-wise histograms indicate that marginals are heavily skewed, and therefore the joint distribution of the marginals ought to be something other than a multivariate normal distribution (not even elliptically symmetric one). In such a situation, it makes sense to follow a 'ground-up' approach to build a multivariate model starting with marginals, rather than the 'top-down' approach of starting with a (questionable) multivariate model and then live with its consequences at the marginal level.

Copula theory is a convenient 'ground-up' approach where one theorizes a multivariate distribution for the random vector $\mathbf{X} = (X_1, X_2, ..., X_p)'$ based on the marginal of each $X_k, 1 \leq k \leq p$. This is based on the understanding that the desired joint distribution ought to obey a particular structure involving the marginals which we have much control over. The following subsection gives a brief introduction of the copula theory. Though the focus of this work is the bivariate set-up, occasionally we present some general multivariate results.

1.2. A Brief Introduction of the Copula Theory and motivation

The theorem by Sklar (1959) plays the most important role in the copula theory. In the simplest case of a bivariate distribution, it tells us that given a random vector (X_1, X_2) with absolutely continuous marginal cumulative distribution functions (*cdfs*), F_1 and F_2, with corresponding probability density functions (*pdfs*) f_1 and f_2 respectively, and its joint *cdf* denoted by F, with joint *pdf* f, there exists a unique copula C (a functional), such that

$$\begin{aligned} F(x_1, x_2) &= C(F_1(x_1), F_2(x_2)), \\ i.e., f(x_1, x_2) &= \partial^2 F(x_1, x_2)/\partial x_1 \partial x_2 \\ &= c(F_1(x_1), F_2(x_2)) f_1(x_1) f_2(x_2), \end{aligned} \tag{1.1}$$

where $c(u, v) := \partial^2 C(u, v)/\partial u \partial v$.

In general, given a continuous random vector in p-dimension, i.e., $\mathbf{X} = (X_1, X_2, ..., X_p)'$, with marginal *cdfs* $F_k, k = 1, 2, ..., p$, if we use the transformations such that $U_k := F_k(X_k)$, $k = 1, 2, .., p$, then we have $U_k \sim \mathit{Uniform}(0, 1), k = 1, 2, ..., p$. The copula function $C : [0, 1]^p \to [0, 1]$ is a joint multivariate *cdf* of $\boldsymbol{U} := (U_1, U_2, ..., U_p)'$, i.e.,

$$C(u_1, u_2, ..., u_p) = P(U_1 \leq u_1, ..., U_p \leq u_p). \tag{1.2}$$

The joint *cdf* of $\boldsymbol{X}$, denoted by $F(x_1, ..., x_p)$, can be given in terms of $C(u_1, ..., u_p)$. By Sklar's theorem there exists a unique copula C such that

$$F(x_1, x_2, ..., x_p) = C(F_1(x_1), F_2(x_2), ..., F_p(x_p)). \tag{1.3}$$

Simply put, the copula C is viewed as a dependence structure among the marginal *cdfs*.

Since the inception of the copula idea, one can find several copulas in the literature such as Gaussian copula, Exponential copula, Clayton copula, Frank Copula etc., just to name a few. Out of the many available copula structures we focus on the Farlie-Gumbel-Morgenstern copula ($FGMC$) (see Morgenstern (1956)). The following subsection gives a brief introduction about the joint distribution based on $FGMC$, henceforth referred to as Farlie - Gumbel - Morgernstern Distribution ($FGMD$). The main reason behind our choice of $FGMC$ (and subsequently that of $FGMD$) is its simplicity. Moreover, the nature of our investigation is completely new, and to the best of our knowledge the type of our investigation has not been carried out not only for $FGMC$, but also for any other copula. Therefore, this work of ours can be used as a template of future research for all other copulas.

1.3. A Brief History of $FGMD$

As mentioned in the earlier section, a host of Copula structures have been discussed in the existing literature and one can find an overview of the available copula structures in Nelsen (2007). Several bivariate and multivariate non-normal probability distributions based on copula structures can be found in Kotz et al. (2004).

Morgenstern (1956) first introduced the following bivariate probability distribution on the square $[-1,1]\times[-1,1]$ of the form:

$$f(x_1,x_2)=0.25*(1+\lambda x_1 x_2), \tag{1.4}$$

where $|\lambda|\leq 1$ and $-1\leq x_1,x_2\leq 1$. Note that the above joint *pdf* in (1.4) is the result of $c(u,v)=1+\lambda(u-1)(v-1)$, which is equivalently due to the copula $C(u,v)=uv(1+\lambda((u/2)-1)((v/2)-1))$. Farlie (1960) further studied various standard correlation coefficients between X_1 and X_2 for the bivariate distribution in (1.4). The limitations that a bivariate normal distribution brings to a dataset were first pointed out by Gumbel (1960) while he constructed a bivariate distribution with exponential marginals using Morgenstern's underlying copula.

The *pdf* of the bivariate Farlie-Gumbel-Morgenstern distribution ($FGMD$) with general marginals based on the $FGMC$ is given by:

$$f(x_1,x_2)=f_1(x_1)f_2(x_2)[1+\lambda(2F_1(x_1)-1)(2F_2(x_2)-1)], \tag{1.5}$$

where $|\lambda|\leq 1$ is the *association* parameter, f_1,f_2 are the marginal *pdf*s of the components X_1 and X_2, with corresponding marginal *cdf*s F_1,F_2 respectively.

As a special case of (1.5), D'este (1981) considered a special biavriate Gamma distribution with gamma marginals and studied the structures of the covariance, conditional expectations as well as other distributional properties. The following Fig **??** shows the two - dimensional contour plots of the bivariate $FGMD$ for various λ values, using uniform marginals. This is to give an idea about how different the bivariate $FGMD$ is from a bivariate normal distribution.

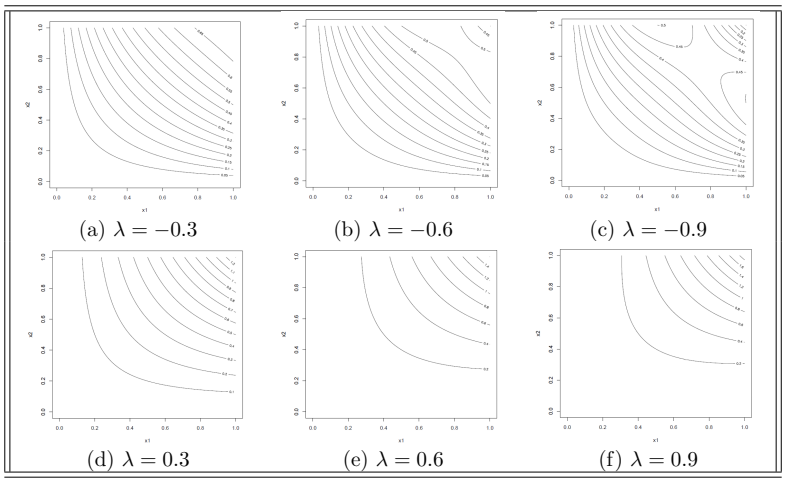
Since the inception of the $FGMC$, it has undergone several modifications over the years leading to some wider family of $FGMC$ by different researchers. All these modifications were done with the goal of capturing a wider range of dependence among the components through common dependence measures such as Pearson's Correlation Coefficient (ρ), Spearman's Correlation coefficient (ρ_s), Kendall's Tau (τ), etc. In their modified $FGMD$ Huang and Kotz (1999) showed that with a polynomial type single parameter extension of the $FGMC$ with uniform marginals the maximal attainable range of ρ is $[-0.39,0.333...]$. Bairamov and Kotz (2002) proposed a new generalization of $FGMD$ by introducing new association parameters and were able to attain a maximal positive (Pearson's) correlation of $\rho=0.5021$ for some specific values of the model parameters. All these generalizations were made to accommodate a larger spectrum of the Pearson's correlation coefficient values. However the Pearson's correlation coefficient measures the strength of linear relationship between the components; therefore, paying attention only to this aspect of dependency, at the cost of adding more parameters to the model, is a rather narrow approach. Amblard and Girard (2009) gave a new family of copulas by generalizing the $FGMC$ and highlighted the main

feature of the proposed family as to permit modelling of data with high positive dependency, in particular over the range of $\rho_s \in [-0.75, 1]$. Another new generalization of the $FGMC$ was put forward by Bekrizadeh et al. (2012) and they were able to show the usefulness of the proposed generalized model in data with high negative dependence value by showing the (Spearman's rank correlation) values of $\rho_s \in [-0.5, 0.43]$. All these generalizations were made by introducing new parameters which only adds to the complexity of the statistical inferences of the $FGMD$ model.

1.4. Scope of this work with $FGMD$

The flexibility of the copula structure lies in allowing the freedom of choice of the desired marginal distributions. Hence, the association parameter λ of the copula structure (1.5) becomes a pivotal parameter in conserving the dependency between the components. As a result, inferences on the association parameter λ in (1.5) is of paramount interest. The basis of this current work has been the $FGMD$ given in (1.5) with the goal of studying the inferential aspects of the association parameter λ comprehensively, with known marginals.

If one looks at the existing applications of the copula theory with real-life data sets then it becomes abundantly clear that the preferred estimator of the association parameter has always been the maximum likelihood estimator (MLE). But how good is the MLE? From an asymptotic point of view the MLE has nice tractable limiting distributional properties. But, for small to moderate sample sizes the performance of the MLE of the $FGMD$ association parameter λ is totally unknown. Worse, the existing literature is completely silent on other possible estimators, especially the Bayes ones under noninformative priors. In a parametric set-up, one should study various estimators of all the model parameters simultaneously which include the association parameter λ as well as other parameters of the marginal distributions. (For example, if one assumes a two parameter gamma model for each of the two marginals, then one ends up with a total five parameters.) It has been noted that estimating just the association parameter with known marginals itself is a research problem as it entails several point estimators with corresponding sampling distributions, followed by hypothesis testing which allows us to verify, under the $FGMD$ assumption, whether the components are independent or not. The computational challenges that one faces with Bayes estimators in this simplistic scenario (i.e., just for the association parameter) can be quite overwhelming. However, the simplistic model that we are using in this work can be applied in a totally nonparametric marginal set-up where one can use the empirical marginal *cdf* of each component to replace the aforementioned known marginal, and then can proceed with the subsequent inferences. The results developed here will be useful in dealing with the unknown parametric marginals where the total number of parameters would be at least three (one for each marginal plus the association parameter). This will be reported in our forthcoming work where parametric marginals with unknown parameters are to be considered under the $FGMD$. With that above objective in mind, the following subsection presents a brief review of parametric estimation of the association parameter λ as available in the existing literature.



(a) $\lambda = -0.3$ (b) $\lambda = -0.6$ (c) $\lambda = -0.9$

(d) $\lambda = 0.3$ (e) $\lambda = 0.6$ (f) $\lambda = 0.9$

Figure 1: *FGMD* Contour Plots with Uniform marginals and selected λ.

1.5. A Brief survey of the Parameter Estimation under *FGMD*

Most of the available literature focus only on point estimation of the association parameter λ in a parametric set-up. Suppose we have *iid* observations $\boldsymbol{X}_i$, $1 \leq i \leq n$, from the above *pdf* (1.5) where the marginal $f_k(x_{ik}) = f(x_{ik}|\boldsymbol{\theta}_k)$ for the k^{th} component involves the parameter $\boldsymbol{\theta}_k$ with dimension $d_k \geq 1$, $k = 1, 2$. Note that x_{ik} is the realized value of the k^{th} component of $\boldsymbol{X}_i$, i.e., X_{ik}, $1 \leq i \leq n$, $k = 1, 2$. The entire parameter vector $\boldsymbol{\omega} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \lambda)$ has the combined dimension of $d = d_1 + d_2 + 1$. The combined likelihood function of the data is

$$\begin{aligned} L &= L(\boldsymbol{\omega}|\boldsymbol{X}_i, 1 \leq i \leq n) \\ &= \prod_{i=1}^{n}[(\prod_{k=1}^{2} f_k(x_{ik}|\boldsymbol{\theta}_k))(1 + \lambda \prod_{k=1}^{2}(2F_k(x_{ik}|\boldsymbol{\theta}_k) - 1))], \end{aligned} \tag{1.6}$$

where $F_k(.|\boldsymbol{\theta}_k)$ is the *cdf* associated with the *pdf* $f_k(.|\boldsymbol{\theta}_k)$. It would make sense to find the MLE of $\boldsymbol{\omega}$ by maximizing L w.r.t. $\boldsymbol{\omega}$. But finding the global maxima of L w.r.t to $\boldsymbol{\omega}$ is a computational challenge. Even for the above bivariate case (1.6) with $d_1 = d_2 = 2$ (which happens for typical gamma or Weibull marginals), $d = 5$, and hence the standard maximization techniques often get bogged down to local maxima. However, the MLE of $\boldsymbol{\omega}$, say $\hat{\boldsymbol{\omega}}_{ML}$ remains as the preferred estimator as it is consistent, and asymptotically $\hat{\boldsymbol{\omega}}_{ML} \sim N_p(\boldsymbol{\omega}, I^{-1})$, where I^{-1} is the inverse of the Fisher information matrix I given as $I = ((I_{ij}))$, where $I_{ij} = -E(\partial^2 L/\partial\omega_i\partial\omega_j)$ with $\boldsymbol{\omega} = (\omega_1, ..., \omega_{d_1}, \omega_{d_1+1}, ..., \omega_{d_1+d_2}, \omega_d) = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \lambda)$. Also, because of the *iid* observations, $I = nI^0$, where $I^0 =$ Fisher Information Matrix per observation $(FIPO)$.

To overcome the computational challenge of maximizing L w.r.t $\boldsymbol{\omega}$ in the $(d+1)$ dimension, Joe (2005) proposed a two-stage estimation process which has a relative computational simplicity, but it may come at a cost of efficiency. This simplistic approach is known as 'Inference Function for Margins'(IFM) method where the parameter $\boldsymbol{\theta_k}$ for the k^{th} marginal is estimated based on the observations $X_{1k}, ..., X_{nk}$ following $f_k(.|\boldsymbol{\theta}_k)$. To be precise, in the

first stage, the IFM calls for estimating $\boldsymbol{\theta}_k$ using the marginal likelihood function

$$L_k = L_k(\boldsymbol{\theta}_k | X_{ik}, 1 \le i \le n) = \prod_{i=1}^{n} f_k(x_{ik}|\boldsymbol{\theta}_k), \tag{1.7}$$

for each k, $1 \le k \le 2$. The MLE of $\boldsymbol{\theta}_k$ thus obtained is called the 'marginal MLE', and denoted by $\hat{\boldsymbol{\theta}}_k = \hat{\boldsymbol{\theta}}_{k(MML)}$. Then, in the second stage, replace $\boldsymbol{\theta}_k$ by $\hat{\boldsymbol{\theta}}_{k(MML)}$ in (1.6) to obtain the profile likelihood given as

$$\begin{aligned} \widetilde{L} &= L(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2, \lambda | \boldsymbol{X}_i, 1 \le i \le n) \\ &= \prod_{i=1}^{n} [(\prod_{k=1}^{2} f_k(x_{ik}|\hat{\boldsymbol{\theta}}_{k(MML)}))(1 + \lambda \prod_{k=1}^{2} (2F_k(x_{ik}|\hat{\boldsymbol{\theta}}_{k(MML)}) - 1))]. \end{aligned} \tag{1.8}$$

Then obtain the 'profile MLE' of λ, say $\hat{\lambda}_{PML}$, by maximizing $\widetilde{L}$ w.r.t. λ, i.e.,

$$\hat{\lambda}_{PML} = \arg\max_{\lambda \in [-1,1]} \widetilde{L}(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2, \lambda | \boldsymbol{X}_i, 1 \le i \le n).$$

Note that the estimator $\hat{\boldsymbol{\theta}}_1 = \hat{\boldsymbol{\theta}}_{1(MML)}$ and $\hat{\boldsymbol{\theta}}_2 = \hat{\boldsymbol{\theta}}_{2(MML)}$ are mutually independent, but $\hat{\lambda}_{PML}$ is dependent on $(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2)$. As a result, given $(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2)$ fixed, the approximate distribution of $\hat{\lambda}_{(PML)}$ is $N(\lambda, \hat{\sigma}_*^2)$, where $\hat{\sigma}_*^2 = \sigma_*^2(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2)$ is the conditional approximate variance of $\hat{\lambda}_{(PML)}$ obtained from $(\sigma_*^2)^{-1} = -E_*(\partial^2 \widetilde{L}/\partial \lambda^2)$, where E_* represents conditional expectation based on $\boldsymbol{\theta}$. Under regularity conditions $\hat{\theta}_i \xrightarrow{p} \theta_i$ when $n \to \infty$, $i = 1, 2$, therefore, asymptotically $\hat{\lambda}_{PML} \xrightarrow{d} N(\lambda, \sigma_*^2)$, where $\sigma_*^2 = \sigma_*^2(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$.

Taheri et al. (2018) considered a related, but different model where it was assumed that a fixed number of the original observations could be potential outliers, and hence these outliers follow the $FGMD$ in (1.5) with the association parameter $\lambda^* = \lambda\beta$, while the rest of the dataset followed (1.5) with the association parameter λ. But the set-up of the aforementioned outliers will not be considered here as it assumes that certain observations are known to be outliers which is not feasible to implement in practice. It is not easy to decide what constitutes an outlier under the proposed $FGMD$, and once detected, based on whatever subjectivity it might involve, how much information does it carry for the underlying model parameters is not clear.

Application of $FGMC$ to construct a joint survival function based on Weibull marginals was taken up by Suzuki et al. (2013). These authors considered a generalized beta-type prior for the association parameter λ, and then implemented Metropolis - Hastings algorithm for the Bayesian estimation of λ. However, to keep the theory simple, straightforward, and implementable, we use only noninformative priors without any extra super (or hyper) parameters.

1.6. A Brief Outline of this work

As mentioned earlier, the primary focus of this work is inferences on the association parameter λ, especially point estimation and hypothesis testing with known marginals. In Section 2 we address point estimation of λ assuming that the marginals f_k's $(k = 1, 2)$ are completely known. We have proposed several estimators including Bayes' estimators under

suitable non informative priors such as the flat prior (FP), Jeffrey's prior (JP) and approximate Jeffrey's prior (AJP). We have investigated the performance of these estimators in terms of bias and mean squared error (MSE) in section 3. Detailed results from our simulation study to assess the performances of the competing estimators have been provided for different sample sizes. This includes the simulated sampling distribution of the estimators in subsection 3.1 and the corresponding bias and MSE of the estimators in subsection 3.2. In Section 4 we have taken up the problem of testing the null hypothesis $H_0 : \lambda = \lambda_0$ against a two sided alternative. In particular, our interest lies in $\lambda_0 = 0$, since, if this null is retained, then under the $FGMD$, it implies that the marginals are independent of each other. In this section, we have proposed a new family of hypothesis tests developed on the basis of the likelihood ratio test principle that incorporates the parametric bootstrap (PB) approach in determining the critical regions rather than the asymptotic Chi-square cut-off points. The size and power of these tests have been studied extensively through simulation for $\lambda_0 = 0$. A real-life dataset has been used in Section 5 to demonstrate how the FGMD can be used to investigate the pairwise association among different variables under study. This dataset, from the Mekong Delta Region of Vietnam, pertains to Arsenic contamination in groundwater in the presence of other apparently benign elements. Finally, the paper ends with some useful observations and a concluding remark.

1.7. A useful result for known marginals

The following Fisher information for $FGMD$ will be useful in section 2 when the marginals are (assumed to be) fully known. To the best of our knowledge this has not been reported before in the literature.

Lemma 1.1. *Based on the iid observations $\boldsymbol{X}_1, \boldsymbol{X}_2, ..., \boldsymbol{X}_n$ from (1.5) with marginals f_1 and f_2 completely known, the Fisher information $I(\lambda)$ is given as $I(\lambda) = nI^0(\lambda)$, where $I^0(\lambda)$ is the Fisher information per observation ($FIPO$), and*

$$I^0(\lambda) = (1/4)\int_{-1}^{+1}\int_{-1}^{+1} u_1^2 u_2^2 (1+\lambda u_1 u_2)^{-1} du_1 du_2. \tag{1.9}$$

Note that the $FIPO$ expression is free from f_1 and f_2. A further simplification yields

$$I^0(\lambda) = \sum_{m=0}^{\infty} \lambda^{2m}/(2m+3). \tag{1.10}$$

It is worth noting here that the infinite sum in the above expression is convergent and has a finite expression. Using that expression of the infinite sum, the final form of the FIPO, is given by

$$I^0(\lambda) = \{-\lambda + tanh^{-1}(\lambda)\}/\lambda^3, \tag{1.11}$$

where $tanh^{-1}(\lambda) = (0.5)log((1+\lambda)/(1-\lambda))$. The proof is available in Section 1 of the supplementary information (SI).

Remark 1.1. It is not at all surprising to see that the expression of I^0 in (1.10) or (1.11) is free from f_k's $(k = 1, 2)$. Since the marginals are assumed to be completely known, without any loss of information one can look at $Y_{ik} = F_k(X_{ik})$, $k = 1, 2$, $1 \leq i \leq n$. Note that Y_{ik}'s are *iid* $Uniform(0, 1)$. Each $\boldsymbol{Y}_i = (Y_{i1}, Y_{i2})'$ then follows the $FGMD(\lambda)$ with joint *pdf*, say $g(\boldsymbol{y}) = [1 + \lambda(2y_1 - 1)(2y_2 - 2)]$ on the unit square $[0, 1] \times [0, 1]$. The transformation $\boldsymbol{X}_i \to \boldsymbol{Y}_i$ does not change the problem as far as inference on λ is concerned, and yields the $FIPO$ expression as stated above.

2. Point Estimation of λ

In this section, we propose a wide variety of estimators of the association parameter λ based on n *iid* observations from (1.5) with known marginals f_1 and f_2. As stated in Remark 1.1, we refer to (1.5) as $FGMD(\lambda)$. Let $\mathbb{R}_1$ and $\mathbb{R}_2$ denote the support of X_1 and X_2 respectively.

2.1. Method of Moment Estimator $(\hat{\lambda}_{MM})$

Method of moment estimator is attained essentially by equating the sample raw moment with the population moment. For the joint population moment, using the form of (1.5) and some further simplification lead us to the following form

$$E(X_1X_2) = E(X_1)E(X_2) + \lambda I_1 I_2, \tag{2.1}$$

where $I_k = \int_{-1}^{1}(u/2)F_k^{-1}((1 + u)/2)\partial u$, $k = 1, 2$. For convenience define $\mu_k = E(X_k)$, $k = 1, 2$, i.e., the means of the known marginals. Therefore from (2.1) it can be easily established that $Cov(X_1, X_2) = \lambda I_1 I_2$. For the method of moment estimator $\hat{\lambda}_{MM}$ we equate $\lambda I_1 I_2$ with the sample equivalent of $Cov(X_1, X_2)$ which is $(1/n)\sum_{i=1}^{n}(X_{1i} - \overline{X}_1)(X_{2i} - \overline{X}_2)$ where $\overline{X}_k = (1/n)\sum_{i=1}^{n} X_{ki}$, $k = 1, 2$. Therefore,

$$\hat{\lambda}_{MM} = (nI_1I_2)^{-1}\sum_{i=1}^{n}(X_{1i} - \overline{X}_1)(X_{2i} - \overline{X}_2). \tag{2.2}$$

The above estimator may take values outside the parameter space with a positive probability and hence it ought to be truncated at the boundary -1 or +1.

2.2. Maximum Likelihood Estimator (MLE or $\hat{\lambda}_{ML}$)

For the brevity of derivation let us denote $2F_k(x_{ik}) - 1 = G_k(x_{ik}), k = 1, 2$. The log-likelihood function of the data denoted by $l(\lambda)$ is as follows

$$l(\lambda) = C + \sum_{i=1}^{n} ln(1 + \lambda G_1(x_{i1})G_2(x_{i2})), \tag{2.3}$$

where C is a constant, free of λ. It is tempting to take derivative of $l(\lambda)$ and equating it with zero, i.e.,

$$\sum_{i=1}^{n} G_1(x_{i1})G_2(x_{i2})/(1+\lambda G_1(x_{i1})G_2(x_{i2})) = 0, \tag{2.4}$$

to find the MLE of λ. But this can lead to a computational error as the solution may lie outside the parameter space which may go unnoticed in simulation studies. (We suspect that this issue may arise for other copula - based joint distributions as well, and may have gone unnoticed in applications.)

Theorem 2.1. *The MLE of λ i.e., $\hat{\lambda}_{ML}$ as it is called here, which maximizes $l(\lambda)$ in (2.3), exists, and it is unique. (See SI: Section 2 for proof)*

Remark 2.1. Define $a_i = G_1(x_{i1})G_2(x_{i2})$, $1 \le i \le n$, and $G_k(x_{ik}) = 2F_k(x_{ik}) - 1$, $k = 1, 2$. Let $h(\lambda) = \sum_{i=1}^{n} a_i/(1+\lambda a_i)$, $\lambda \in [-1, +1]$. As seen from the details of the proof of the above theorem, $\hat{\lambda}_{ML}$ takes the following form:

$$\hat{\lambda}_{ML} = \begin{cases} -1 & \text{if } h(-1) < 0 \\ \text{solution of (2.4)} & \text{if } h(-1) > 0 \text{ and } h(+1) < 0 \\ +1 & \text{if } h(+1) > 0, \end{cases} \tag{2.5}$$

Remark 2.2. It is further seen that if all the a_i's are > 0, which happens with probability $(0.5)^n$, then $l(\lambda)$ is monotonically increasing in λ. Hence $\hat{\lambda}_{ML}$ is $+1$. Thus, $\{(a_1, ..., a_n) |\ a_i > 0\ \forall\ i\} \subseteq \{(a_1, ..., a_n) |\ h(+1) = \sum_{i=1}^{n} a_i/(1+a_i) > 0\}$. Similarly, if all the a_i's are < 0, which again happens with probability $(0.5)^n$, then $l(\lambda)$ is monotonically decreasing in λ. Hence $\hat{\lambda}_{ML}$ is -1. Thus, $\{(a_1, ..., a_n) |\ a_i < 0\ \forall\ i\} \subseteq \{(a_1, ..., a_n) |\ h(-1) = \sum_{i=1}^{n} a_i/(1-a_i) < 0\}$. We will see later in our simulation study that $\hat{\lambda}_{ML}$ can take ± 1 with substantially high probabilities depending on the sample size as well as λ.

2.3. Regression Estimators based on the distributions of $X_2|X_1$ and $X_1|X_2$

To estimate any parameter in a bivariate distribution, especially the association parameter, a natural direction is to consider a suitable regression model and then estimating the parameter of interest based on the regression model. So, two structurally similar yet different estimators of the association parameter emerge from the $FGMD(\lambda)$ as follows.

First, we consider the conditional expectation $E(X_2|X_1 = x_1)$. The conditional density function of $X_2|X_1 = x_1$ is denoted by $f(x_2|x_1)$ and can be easily derived from the $FGMD(\lambda)$. The form is as follows:

$$f_{2|1}(x_2|x_1) = f_2(x_2)(1 + \lambda(2F_1(x_1) - 1)(2F_2(x_2) - 1)). \tag{2.6}$$

The regression model is as follows:

$$\begin{aligned} E(X_2|X_1 = x_1) &= \int_{\mathbb{R}_2} x_2 f_{2|1}(x_2|x_1)\partial x_2 \\ &= E(X_2) + \lambda G_1(x_1) I_2, \end{aligned} \tag{2.7}$$

where $G_1(x_1) = (2F_1(x_1) - 1)$ and I_2 is given after (2.1) respectively. Let us denote $E(X_2) = \mu_2$, which assumed to be a known constant value. Also, for a given $X_1 = x_1$, we can treat $G_1(x_1)$ and I_2 as known constants. Let us denote $G_1(x_1)I_2 = B_1(x_1)$. The conditional expectation in equation (2.7) would then look like $E(X_2|X_1 = x_1) = \mu_2 + B_1(x_1)\lambda$, which is a simple linear regression equation. Given a bivariate data set, a least squares estimator of λ from the above regression equation (2.7) can be attained by minimizing the expression $\sum_{i=1}^{n}(x_{2i} - \mu_2 - \lambda B_1(x_{1i}))^2$ with respect to λ. Hence, by solving the normal equation, the regression estimator of type-1 , i.e., $\hat{\lambda}_{RE1}$ is given by

$$\hat{\lambda}_{RE1} = [\sum_{i=1}^{n} X_{2i}B_1(X_{i1}) - \mu_2 \sum_{i=1}^{n} B_1(X_{i1})]/\sum_{i=1}^{n}[B_1(X_{i1}]^2. \tag{2.8}$$

Since the density function of $FGMD(\lambda)$ is structurally symmetric, by reversing the roles of X_1 and X_2 in equation (2.7) we get another (type-2) regression based estimator i.e., $\hat{\lambda}_{RE2}$ of the association parameter and is given by -

$$\hat{\lambda}_{RE2} = [\sum_{i=1}^{n} X_{1i}B_2(X_{i2}) - \mu_1 \sum_{i=1}^{n} B_2(X_{i2})]/\sum_{i=1}^{n}[B_2(X_{i2}]^2, \tag{2.9}$$

where $\mu_1 = E(X_1)$, $G_2(x_2)I_1 = B_2(x_2)$ and $G_2(.), I_2$ are defined in Subsections 2.2 and 2.1 respectively. The above two regression estimators ought to be truncated at the boundary of -1 or +1 because they can take values outside the parameter space with positive probabilities.

If the marginals are completely known or they are unknown but identically distributed then both the above regression estimators will have the same sampling distribution. Therefore as far as the sampling distribution is concerned it is enough to look at either one of them.

2.4. Bayes' Estimation of λ

For any suitable prior $\pi(\lambda)$ over the parameter space $[-1, +1]$, the posterior distribution of $(\lambda|data)$, denoted by $g(\lambda|data)$, is

$$g(\lambda|data) = \frac{\prod_{i=1}^{n}[1 + \lambda G_1(X_{i1})G_2(X_{i2})]\pi(\lambda)}{\int_{-1}^{1}\prod_{i=1}^{n}[1 + \lambda G_1(X_{i1})G_2(X_{i2})]\pi(\lambda)\partial\lambda}. \tag{2.10}$$

A natural choice of the prior for the association parameter is a modification of the beta distribution which is originally defined over the space $(0, 1)$. The beta-type prior density function defined over the parameter space $[-1, +1]$ is

$$\pi(\lambda) = 1/(2B(a, b))((1 + \lambda)/2)^{a-1}((1 - \lambda)/2)^{b-1}, \tag{2.11}$$

where a, b are the hyper-parameters.

The most common loss function for estimating a parameter is the usual squared error loss. However, when a parameter is restricted to a finite range, as we have here for the

association parameter λ, a weighted quadratic loss is more meaningful which can assign a heavy penalty near the boundary. Hence, we consider a general structure of the loss function of the form

$$L(\hat{\lambda}, \lambda) = w(\lambda)(\hat{\lambda} - \lambda)^2, \tag{2.12}$$

where $w(\lambda)$ is a suitable weight function. In this work we are going to consider weight function $w(\lambda)$ of the form

$$w_\delta(\lambda) = (1 - \lambda^2)^{-\delta},\ \delta \geq 0. \tag{2.13}$$

Note that $\delta = 0$ leads to the usual squared error loss. For any $\delta > 0$, the loss (2.12) goes to ∞ as λ approaches ± 1 and $|\hat{\lambda} - \lambda| > 0$. In other words, a small deviation of $\hat{\lambda}$ from λ near the boundary can be very costly.

Under the general weighted quadratic loss (2.12), the general structure of the Bayes' rule is given as

$$\begin{aligned} \hat{\lambda}_B &= \frac{E(\lambda w(\lambda)|\lambda \sim g(\lambda|data))}{E(w(\lambda)|\lambda \sim g(\lambda|data))} \\ &= \frac{\int_{-1}^{1} \lambda w(\lambda) \prod_{i=1}^{n} [1 + \lambda G_1(X_{i1}) G_2(X_{i2})] \pi(\lambda) d\lambda}{\int_{-1}^{1} w(\lambda) \prod_{i=1}^{n} [1 + \lambda G_1(X_{i1}) G_2(X_{i2})] \pi(\lambda) d\lambda}. \end{aligned} \tag{2.14}$$

With the special structure of $w(\lambda) = w_\delta(\lambda) = (1 - \lambda^2)^{-\delta}$, we are now ready to derive the Bayes' rule, denoted by $\hat{\lambda}_{B\delta}$ as follows.

In order to attain a tractable structure of the Bayes' rule, we resort to a simple algebraic manipulation within the equation (2.14). Let us focus on the term $\prod_{i=1}^{n}(1+\lambda G_1(X_{i1})G_2(X_{i2}))$ in the equation (2.14). Recalling from Remark 2.2 that $a_i = G_1(X_{i1})G_2(X_{i2})$, the following product term can be rewritten as

$$\begin{aligned} \prod_{i=1}^{n}(1 + \lambda G_1(X_{i1})G_2(X_{i2})) &= (1 + \lambda a_1)(1 + \lambda a_2)....(1 + \lambda a_n) \\ &= 1 + \lambda \sum_{i_1=1}^{n} a_{i_1} + \lambda^2 \sum_{(1 \leq i_1 < i_2 \leq n)} a_{i_1} a_{i_2} + ... \\ &\quad ... + \lambda^k \sum_{(1 \leq i_1 < i_2 < ... < i_k \leq n)} a_{i_1} a_{i_2} ... a_{i_k} + ... + \lambda^n a_{i_1} a_{i_2} ... a_{i_n}. \end{aligned} \tag{2.15}$$

Call $\sum_{(1 \leq i_1 < i_2 < ... < i_k \leq n)} a_{i_1} a_{i_2} ... a_{i_k} = D_k$, $1 \leq k \leq n$, and define $D_0 = 1$. Therefore $\prod_{i=1}^{n}(1 + \lambda G_1(X_{i1})G_2(X_{i2})) = \sum_{k=0}^{n} \lambda^k D_k$. Hence, the Bayes' rule in (2.14) can be simplified as -

$$\hat{\lambda}_B = \frac{\sum_{k=0}^{n} D_k \int_{-1}^{1} \lambda^{k+1}(1 - \lambda^2)^{-\delta} \pi(\lambda) d\lambda}{\sum_{k=0}^{n} D_k \int_{-1}^{1} \lambda^k (1 - \lambda^2)^{-\delta} \pi(\lambda) d\lambda}. \tag{2.16}$$

Further, we will consider the special case of $a = b = d$, which implies a symmetric prior about 0. We are going to introduce the notation β as $\beta = d - \delta$ and the estimator (2.16) with the prior in (2.11) will be denoted as $\hat{\lambda}_{B\beta}$, i.e.,

$$\hat{\lambda}_{B\beta} = \frac{\sum_{k=0}^{n} D_k \int_{-1}^{1} \lambda^{k+1}(1 - \lambda^2)^{\beta - 1} d\lambda}{\sum_{k=0}^{n} D_k \int_{-1}^{1} \lambda^k (1 - \lambda^2)^{\beta - 1} d\lambda}. \tag{2.17}$$

2.4.1. Special Case of $\beta = 1$ (Bayes estimator under flat prior or BFP)

A particular case of interest is $\beta = 1$ which can happen if $\delta = 0$ and $d = 1$ or $\delta = 1$ and $d = 2$ etc. Sine $\beta = 1$ (due to $\delta = 0$ and $d = 1$) also implies the Bayes' estimator under the flat prior (FP) using the ordinary squared error loss function, we denote $\hat{\lambda}_{B1}$ as $\hat{\lambda}_{BFP}$ and is given by

$$\begin{aligned}\hat{\lambda}_{BFP} &= \frac{\sum_{k=0}^{n} D_k \int_{-1}^{1} \lambda^{k+1} d\lambda}{\sum_{k=0}^{n} D_k \int_{-1}^{1} \lambda^{k} d\lambda} \\ &= \frac{\sum_{k=0}^{n} (D_k/(k+2))\{1-(-1)^k\}}{\sum_{k=0}^{n} (D_k/(k+1))\{1-(-1)^{k+1}\}}.\end{aligned} \tag{2.18}$$

2.4.2. Bayes' Estimator under Jeffrey's Prior (BJP)

Let us step back to the initial form of the Bayes' estimator as mentioned in equation (2.14). A natural non-informative prior is the Jeffrey's prior, denoted by $\pi_{JP}(\lambda)$, which is

$$\pi_{JP}(\lambda) \propto (I(\lambda))^{1/2},$$

where $I(\lambda)$ = Fisher Information of λ from a sample of size n. Hence, from (1.10), we have

$$\pi_{JP}(\lambda) \propto \sum_{m=0}^{\infty} \lambda^{2m}/(2m+3)$$

Hence, the Bayes' estimator under Jeffrey's prior, denoted by $\hat{\lambda}_{BJP}$, is given by

$$\hat{\lambda}_{BJP} = \frac{\sum_{k=0}^{n} D_k \int_{-1}^{1} \lambda^{k+1}(1-\lambda^2)^{-\delta}(\sum_{m=0}^{\infty} \lambda^{2m}/(2m+3))^{1/2} d\lambda}{\sum_{k=0}^{n} D_k \int_{-1}^{1} \lambda^{k}(1-\lambda^2)^{-\delta}(\sum_{m=0}^{\infty} \lambda^{2m}/(2m+3))^{1/2} d\lambda}. \tag{2.19}$$

Using $\delta = 0$ we have a special case, which will be of interest, that is

$$\hat{\lambda}_{BJP} = \frac{\sum_{k=0}^{n} D_k \int_{-1}^{1} \lambda^{k+1}(\sum_{m=0}^{\infty} \lambda^{2m}/(2m+3))^{1/2} d\lambda}{\sum_{k=0}^{n} D_k \int_{-1}^{1} \lambda^{k}(\sum_{m=0}^{\infty} \lambda^{2m}/(2m+3))^{1/2} d\lambda}. \tag{2.20}$$

2.4.3. Bayes' Estimator under an approximate Jeffrey's Prior $(BAJP)$

Note that in either of (2.19) or (2.20) the Bayes' estimator involves an infinite series. For the ease of simplification and being able to study the performance of a suitable Bayes'

estimator analytically, we propose a simplistic approximation of the Jeffrey's prior which is given by $\sum_{m=0}^{\infty} |\lambda|^m/(2m+3)^{1/2}$. Also, note that this infinite series is convergent and has a finite value. In fact, the above series converges to $\sqrt{2}\Phi(|\lambda|, 1/2, 3/2)/2$, where $\Phi(x, y, z)$ is called the confluent hypergeometric function of the first kind (Abramowitz and Stegun (1964)), which is a function of x when y, z are held constants. Due to this fact, we can use this approximation as a new prior distribution. We call this as the approximate Jeffrey's prior and is given by

$$\pi_{AJP}(\lambda) \propto \sum_{m=0}^{\infty} |\lambda|^m/(2m+3)^{1/2}. \tag{2.21}$$

Hence the Bayes' estimator with respect to (2.21), denoted by $\hat{\lambda}_{BAJP}$, is

$$\hat{\lambda}_{BAJP} = \frac{\sum_{k=0}^{n}\sum_{m=0}^{\infty} D_k(2m+3)^{-1/2}(1+(-1)^{k+1})(m+k+2)^{-1}}{\sum_{k=0}^{n}\sum_{m=0}^{\infty} D_k(2m+3)^{-1/2}(1+(-1)^{k})(m+k+1)^{-1}}. \tag{2.22}$$

The derivation of (2.22) is given in SI: Section 3. Figure 2 shows the plots of $\pi_{JP}(\lambda)$ and $\pi_{AJP}(\lambda)$.

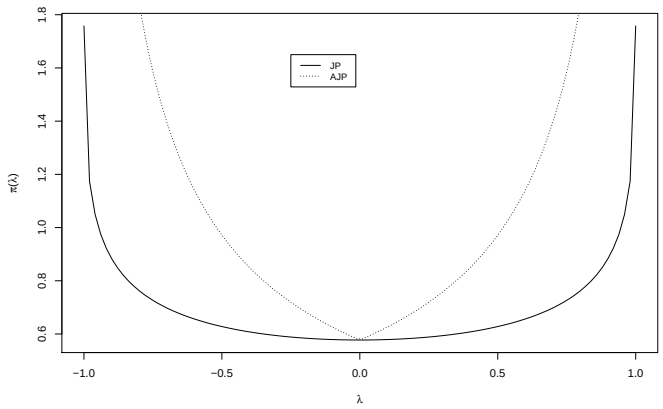


Figure 2: Plots of the Jeffrey's prior and the approximate Jeffrey's prior

Remark 2.3. Obtaining the D_k terms in the Bayesian estimators as defined in equation (2.15), is a time consuming computation, not for a single use, but for the simulations, because each D_k term takes into account $\binom{n}{k}$ terms, and this can be very large for $n > 25$ and $k \approx (n/2)$. For example, when the sample size is $n = 30$ and $k = 15$, D_k involves a sum of the products of $\binom{30}{15} = 155,117,520$ terms. One can easily imagine how computation of the D_k terms becomes quite heavy very quickly as the sample size grows. For the exact computation for a single dataset the D_k terms are still manageable. However for the purpose of simulation with replicated samples, dealing with D_k - terms become prohibitively time consuming, and hence we use numerical integration, specifically the adaptive gaussian quadrature in the numerator and the denominator (of each Bayes estimator) to circumvent the computational complexity associated with the D_k - terms. This has been discussed elaborately in Remark 3.2.

3. Performance of the Six Estimators

We study the sampling distribution of the six estimators in a comprehensive way through simulation. For the simulation purpose we have generated 10^5 iid samples for each n. Since $\lambda \in [-1, 1]$, a symmetric parameter space, we presented the results for the positive part of the parameter space only. From the generated data set in each replication we calculate all the six estimators and then obtain the relative frequency histogram of the estimators for the different λ, such as 0, 0.3, 0.6, 0.9, and various n, such as 10, 20, 30, 40 and 50. For the simulation purpose we have used f_1 and f_2 as $Uniform(0, 1)$, and hence the conditional distributions of $X_1|X_2$ and $X_2|X_1$ are the same. As a result, the estimators $\hat{\lambda}_{RE1}$ and $\hat{\lambda}_{RE2}$ are the same, and denoted by $\hat{\lambda}_{RE}$ for convenience. Further note that with known marginals, the performance of all the estimators is independent of f_k, $k = 1, 2$ (refer to Remark 1.1).

3.1. Relative Frequency Histograms of the Six Estimators

The Figures 1 - 5 in SI: Section 5 provide the simulated sampling distributions of the six estimators as discussed earlier, i.e., $\hat{\lambda}_{MM}$, $\hat{\lambda}_{RE}$, $\hat{\lambda}_{ML}$, $\hat{\lambda}_{BFP}$, $\hat{\lambda}_{BJP}$ and $\hat{\lambda}_{BAJP}$.

Remark 3.1. From the relative frequency distributions of the six estimators achieved through simulation, we can clearly see that the estimators, more specifically the non-Bayesian ones such as $\hat{\lambda}_{MM}$, $\hat{\lambda}_{RE}$ and $\hat{\lambda}_{ML}$ demonstrate somewhat similar distributional properties. One obvious criticism for the above-mentioned non-Bayesian estimators is that as the value of λ approaches towards the boundary, these estimators tend to take values outside the parameter space resulting in huge probability concentration at the boundary since the analytical expressions of the estimators are truncated at the boundary values of +1 and -1. There is a natural correction of this behavior observed in the Bayesian estimators because the prior distributions are defined strictly on the parameter space [-1,+1]. One can expect, with the increase of sample size $\hat{\lambda}_{ML}$ would avoid the heavy probability concentrations at the boundaries, but even for moderately large sample sizes such as 40 and 50 we don't see this happening; however, this probability concentration near the boundary values tend to diminish as the sample size increases, as it is evident in our SI: Figures 1 - 5. Even for $\lambda = 0$, the non-Bayesian estimators tend to exhibit somewhat probability concentrations near the boundary values which is gradually diminishing as the sample size increases. This is happening perhaps due to the fact that these estimators are converging to a limiting normal distribution asymptotically at a very slow rate of convergence. In other words, the non-Bayesian estimators are exhibiting large dispersions even when λ is near the center of the parameter space. On the other hand, the relative frequency distributions of the Bayesian estimators exhibit "within boundary" behavior for all sample sizes. This observation leads us to suspect that MLE might not be the best choice even for moderately large sample sizes. This fact is re-iterated in the following subsection where we study the bias and the MSE of these six competing estimators to evaluate their performance over the entire parameter space.

Remark 3.2. The computation of the Bayesian estimators defined in (2.18), (2.19) and (2.22) poses a major challenge in terms of CPU time. The challenge in the computation

is due to the D_k - terms present in all the three Bayesian estimator expressions. Remark 2.4 discusses the computational challenge elaborately. To navigate around this challenge, we have implemented numerical integration, specifically Simpson's adaptive quadrature, on the expression (2.14) with the specific forms of the prior distributions to get the three different Bayesian estimators. This implementation makes the computations heavily time efficient. We were able to achieve the same results as that of the exact expressions at least up to 7 decimal places as verified in sample sizes up to 25.

3.2. Bias and MSE Study

Derivation of multiple estimators comes with an immediate follow up query about the comparison among the different estimators. We study the bias and the MSE of the six estimators simultaneously to get a better idea about the overall performance. We have looked at the performances of the estimators for several sample sizes such as $n = 10, 20, 30$, 40 and 50. Two representative plots, one each for bias and MSE of the 6 estimators had been presented in Figures 6 and 7 respectively for $n = 30$ in SI: Section 5. All other sample sizes exhibit identical trends. The performance of the six estimators as seen have been summarized in Remark 3.3.

Remark 3.3. (a) For the sake of a better understanding the performances of the proposed estimators over the parameter space, we have identified the following two distinct regions in the parameter space of λ. The following distinction has been made based on the contrast in performances over these two regions as portrayed by each of the estimators: (i) 'mid - region' or approximately $\lambda \leq 0.75$; and (ii) 'boundary - region' or approximately $|\lambda| > 0.75$. The estimators show contrasting performance patterns both in terms of bias and MSE over the two regions (symmetric about 0) in the parameter space as mentioned above.
(b) The bias of each estimator shows a monotonically decreasing behavior as we move from left to right along the parameter space. All the estimators tend to overestimate in the negative part of the parameter space and underestimate in the positive part. At $\lambda = 0$ all the estimators appear to be perfectly unbiased. This is not surprising because at $\lambda = 0$ all the 6 estimators are exhibiting perfectly symmetric distributions (see SI, Figures 1 - 5). It is worth noticing that the Bayesian estimators tend to have higher magnitude of bias than the non-Bayesian estimators. For $n \geq 30$, $\hat{\lambda}_{BAJP}$ and $\hat{\lambda}_{ML}$ seem to have very close bias curves. Also, $\hat{\lambda}_{BFP}$ and $\hat{\lambda}_{BJP}$ seem to have much higher bias, in terms of magnitude, compared to the other estimators.
(c) Note that the MSE plots of the Bayesian and the non-Bayesian estimators show a very contrasting patterns among themselves. The non-Bayesian estimators $\hat{\lambda}_{ML}$, $\hat{\lambda}_{MM}$ and $\hat{\lambda}_{RE}$ perform in a similar fashion over the entire parameter space and have concave down MSE curves. The estimator $\hat{\lambda}_{ML}$ has the highest MSE with $\hat{\lambda}_{RE}$ closely following and with a slightly higher MSE than $\hat{\lambda}_{RE}$. The overall trend of these estimator stays the same over the entire parameter space. The curves gradually flatten out as the sample size increases, but even when the sample size is as large as 50 the MSE of non-Bayesian estimators are still concave down with $\hat{\lambda}_{ML}$ having the highest MSE in the mid - region of the parameter space. On the other hand, the Bayesian estimators have concave up MSE curves. Over the mid - region $\hat{\lambda}_{BFP}$ has the lowest MSE, closely followed by $\hat{\lambda}_{BJP}$ and $\hat{\lambda}_{BAJP}$. For $n < 30$, $\hat{\lambda}_{BFP}$ seems to have an overall robust performance over the entire parameter space. However,

for $n \geq 30$, $\hat{\lambda}_{BJP}$ and $\hat{\lambda}_{BAJP}$ appear to have a much more appealing performance over the mid-range which is the majority (about three-fourth) of the entire parameter space.

4. Testing a hypothesis on the association parameter

In this section we study the performance of different types of hypothesis testing procedures to test the hypothesis $H_0 : \lambda = \lambda_0$ vs $H_A : \lambda \neq \lambda_0$. One particular value of λ_0 is of interest, that is $\lambda_0 = 0$, as in most of the applied cases one would be interested in knowing whether the components are associated or not. The following tests have been proposed and studied through size and power for the aforementioned hypotheses.

1. Asymptotic Tests:
   (a) Asymptotic Normal Test. (ANT)
   (b) Asymptotic Likelihood Ratio Test ($ALRT$)
2. Parametric Bootstrap Tests based on the LRT statistic ($PBLRT$)

The following subsections provide the details of each of the above test procedures.

4.1. Asymptotic Tests

4.1.1. Asymptotic Normal Test (ANT)

While testing the values of λ, this is probably the simplest approach of developing a hypothesis test utilizing the asymptotic property of the MLE of λ. In Theorem 2.1, we have already seen that for an iid sample of size n from $FGMD(\lambda)$ the MLE exists and it is unique. It is a well known result that as $n \to \infty$, $\hat{\lambda}_{ML} \xrightarrow{d} N(\lambda,\ AV(\lambda))$, where $AV(\lambda)$ is the asymptotic variance of the MLE and is given by the inverse of the fisher information of λ, i.e. $I^{-1}(\lambda)$, assuming that marginals are fully known. Therefore, if we assume that the null hypothesis is true, then $\hat{\lambda}_{ML} \xrightarrow{d} N(\lambda_0, 1/I(\lambda_0))$ as $n \to \infty$. Therefore we reject the null hypothesis if

$$|\sqrt{nI^0(\lambda_0)}(\hat{\lambda}_{ML} - \lambda_0)| > z_{(1-\alpha/2)},$$

where $z_{(1-\alpha/2)}$ is the right tail $(\alpha/2)$ - probability cutoff point of the standard normal distribution and $I^0(\lambda_0)$ is the $FIPO$ (see (1.10)).

4.1.2. Asymptotic Likelihood Ratio test ($ALRT$)

Based on the iid observations derive the likelihood ratio statistic Λ as

$$\Lambda = \frac{\underset{H_0}{Sup}\, L(\lambda|data)}{\underset{H_0 U H_A}{Sup}\, L(\lambda|data)} = \frac{L(\lambda_0|data)}{L(\hat{\lambda}_{ML}|data)}.$$

Define $\Lambda_* = -2ln(\Lambda)$. Asymptotically, as $n \to \infty$, $\Lambda_* \xrightarrow{d} \chi^2_1$ under H_0. So we reject the null hypothesis at level α if

$$\Lambda_* > \chi^2_{1;(1-\alpha)},$$

where $\chi^2_{1;(1-\alpha)}$ is the right tail (α) - probability cut off point of Chi squared distribution with 1 degree of freedom. The following Table 1 shows the simulated size values of the two asymptotic tests based on the MLE $\hat{\lambda}_{ML}$.

Table 1: Simulated size of the two asymptotic tests for $\lambda_0 = 0, \alpha = 0.05$

Test	$n = 10$	$n = 20$	$n = 30$	$n = 40$	$n = 50$	$n = 75$	$n = 100$
ANT	0.000	0.297	0.284	0.274	0.270	0.265	0.262
$ALRT$	0.002	0.017	0.026	0.028	0.028	0.027	0.026

Remark 4.1. Both the asymptotic tests are far from satisfactory as far as size is concerned. For $n = 10$, both of them are hopelessly conservative. For $n \geq 20$, ANT is overall a very liberal test and $ALRT$ on the other hand is a very conservative test. It is seen that ANT has a monotonically decreasing (albeit very slowly) size property with the increase in n, whereas $ALRT$'s size values indicate a conservative behavior. Even for $n = 100$, which is generally considered to be large, these tests are still unable to achieve α satisfactorily. The poor performance of these two asymptotic tests are expected at the boundary values because the typical regularity conditions (the ones needed for the Cramer-Rao Inequality) do not hold at these boundary parameter values. However, in Table 1 our H_0 value of λ was taken as 0 which is well within the parameter space where the regularity conditions hold without any difficulty. Yet the poor performance of these two asymptotic tests which are based on the MLE indicate only one thing, that is - the MLE converges to its asymptotic normal distribution very slowly.

4.2. Parametric Bootstrap (PB) tests based on the LRT Concept

As seen in the earlier sections, the asymptotic tests do not perform well for small to moderately large sample sizes. Therefore, in this section we propose a class of four tests based on the idea of LRT with the added parametric bootstrap (PB) concept.

The traditional LRT calls for using $\Lambda_* = -2\ ln(\Lambda)$ which, under H_0, follows χ^2_1 asymptotically. However, in this present $FGMD$ case, the null distribution of Λ_* has been found

to be way off from the asymptotic distribution χ^2_1. Therefore, the cut-off point $\chi^2_{1;(1-\alpha)}$ is not applicable for the statistic Λ_* in order to test H_0. A situation like this calls for coming up with different cut-off points for Λ_* depending on sample size n as well as the data $\mathbf{X}$ through a PB method. Note that the expression Λ has $\hat{\lambda}_{ML}$ in the denominator as an estimator of λ while the numerator uses the null value λ_0 of λ. As a result, the value of Λ is always between 0 and 1, and a value of Λ closer to 1 implies a probable validity of H_0.

We extend the above traditional LRT concept a bit further by incorporating the other three estimates of λ which have shown considerable improvement over $\hat{\lambda}_{ML}$, especially in the mid region of the parameter space. In this regard we are going to consider $\hat{\lambda}_{BFP}$, $\hat{\lambda}_{BJP}$ and $\hat{\lambda}_{BAJP}$ (along with $\hat{\lambda}_{ML}$) in the LRT structure. In its generic form, the structure of Λ_* is going to be redefined as $\Lambda_*(\hat{\lambda}) = -2\ ln(\Lambda(\hat{\lambda}))$, where $\Lambda(\hat{\lambda}) = [\ L(\lambda_0|data)/L(\hat{\lambda}|data)\ ]$, where $\hat{\lambda}$ can be any one of the four aforementioned estimators of λ.

One difficulty with the above $\Lambda(\hat{\lambda})$ is that the denominator is not guaranteed to be greater or equal to the numerator unless $\hat{\lambda} = \hat{\lambda}_{ML}$. In other words, $\Lambda_*(\hat{\lambda}) = -2\ ln\Lambda(\hat{\lambda})$ is not guaranteed to be non-negative unless $\hat{\lambda} = \hat{\lambda}_{ML}$. However, a value of $\Lambda_*(\hat{\lambda})$ closer to 0 still conforms the validity of H_0. Therefore, to find suitable cut-off points for the statistic Λ_*, we consider

$$\Lambda_{**}(\hat{\lambda}) = |\Lambda_*(\hat{\lambda})|, \tag{4.1}$$

which is always nonnegative. The four versions of Λ_{**} using four aforementioned estimators will be referred to as

$$\begin{aligned} \Lambda_{**1}\ (or PBLRT\ 1) &= \Lambda_{**}(\hat{\lambda}_{ML}) \\ \Lambda_{**2}\ (or PBLRT\ 2) &= \Lambda_{**}(\hat{\lambda}_{BFP}) \\ \Lambda_{**3}\ (or PBLRT\ 3) &= \Lambda_{**}(\hat{\lambda}_{BJP}) \\ \Lambda_{**4}\ (or PBLRT\ 4) &= \Lambda_{**}(\hat{\lambda}_{BAJP}) \end{aligned} \tag{4.2}$$

Algorithmic Steps to Implement $\Lambda_{**}(\hat{\lambda})$ as a Test:

Step - 1: For the given data $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_n)$ from $FGMD$, compute $\hat{\lambda}$ (which is one of the above 4 estimators as mentioned above). Obtain the corresponding $\Lambda_{**}(\hat{\lambda})$.

Step - 2: Assume that $H_0 : \lambda = \lambda_0$ is true. Generate a bootstrap sample of size n (say, $\mathbf{X}^*_1, \mathbf{X}^*_2, ..., \mathbf{X}^*_n$) from $FGMD(\lambda_0)$. Once this bootstrap data is generated, pretend that λ is unknown, estimate λ using the bootstrap data by $\hat{\lambda}$, and call it $\hat{\lambda}^*$, which in turn produces the value of $\Lambda_{**}(\hat{\lambda}^*)$.

Step - 3: Repeat the above Step - 2 a large number of times (say, B times). This produces B copies of $\Lambda_{**}(\hat{\lambda}^*)$, and call them as $\Lambda^{(b)}_{**}(\hat{\lambda}) = \Lambda_{**}(\hat{\lambda}^{*(b)})$, $1 \le b \le B$, where $\hat{\lambda}^{*(b)}$ is the b^{th} copy of $\hat{\lambda}^*$ as mentioned in Step - 2. These $\Lambda^{(b)}_{**}(\hat{\lambda})$ values are supposed to approximate the null distribution of $\Lambda_{**}(\hat{\lambda})$.

Step - 4: Order $\Lambda^{(b)}_{**}(\hat{\lambda})$, $1 \le b \le B$, and let $\Lambda_{**}(\hat{\lambda}|\alpha)$ be the $100(1-\alpha)^{th}$ percentile value of $\Lambda^{(b)}_{**}(\hat{\lambda})$, $1 \le b \le B$. This $\Lambda_*(\hat{\lambda}|\alpha)$ is the critical value for $\Lambda_*(\hat{\lambda})$ in Step - 1.

Step - 5: Reject H_0 if $\Lambda_{**}(\hat{\lambda})$ (from Step -1) $> \Lambda_{**}(\hat{\lambda}|\alpha)$, and retain H_0 if otherwise.

4.3. Computational Results Comparing $PBLRTi$, $1 \leq i \leq 4$

4.3.1. Comparison of Four $PBLRT$s in Terms of Size

In this subsection we provide the results of our comprehensive simulation study to compare the above four tests in terms of size and power to test $H_0 : \lambda = \lambda_0$ vs $H_A : \lambda \neq \lambda_0$. Since in applications one is mostly interested in knowing whether the components are associated or not under the assumed $FGMD$, we are going to use $\lambda_0 = 0$. The simulated sizes of the four $PBLRT$s have been presented in Table 2 for various sample sizes. This simulation requires two loops of replications. The outer loop has $M = 10^4$ replications, and the inner loop which produces the critical value also uses $B = 10^4$ replications.

Table 2: Size values of the four $PBLRT$s for $\lambda = \lambda_0 = 0$

Test	$n = 10$	$n = 20$	$n = 30$	$n = 40$	$n = 50$
$PBLRT1$	0.054	0.052	0.054	0.052	0.050
$PBLRT2$	0.052	0.051	0.055	0.052	0.049
$PBLRT3$	0.052	0.051	0.055	0.052	0.049
$PBLRT4$	0.053	0.051	0.054	0.052	0.050

Remark 4.2. (a) First of all, in terms of simulated size, all the four $PBLRT$s appear to be adhering to the level condition very closely. The differences between the simulated size and α appear to be within one standard error (SE) of the simulated size; and hence, all the four $PBLRT$s appear to be α-reliable for all n. (The SE of each simulated size is about 0.002.)
(b) Notice the stark difference between the $ALRT$ and $PBLRT1$ (which uses the classical LRT statistic under a PB framework). While $PBLRT1$ looks highly α-reliable, its asymptotic version doesn't appear to be so for those selected values of n ($10 \leq n \leq 100$) (see Table 1). In other words, the asymptotic cut-off point $\chi^2_{1;(1-\alpha)}$ is not a reliable cut-off point for the classical LRT statistic for small to moderately large sample sizes. The $PBLRT1$ method uses a more accurate cut-off point for the statistic $\Lambda_{**}(\hat{\lambda}_{ML})$ than the asymptotic one (i.e., $\chi^2_{1;(1-\alpha)}$).

To see for what sample size the Chi-square cut-off point is "close" to the true one, we have computed $\Lambda_{**}(\hat{\lambda}_{ML}|\alpha)$ (which is a function of n and λ_0 for various values of n ranging from 0 to 300, and $\lambda_0 = 0$. These cut-off points have been plotted against n in Figure 3 for $\alpha = 0.05$. It appears that the asymptotic Chi-square cut-off point can be used for $PBLRT1$ only when $n \geq 300$, since $\Lambda_{**}(\hat{\lambda}|\alpha) \approx \chi^2_{1;(1-\alpha)} = 3.84$ for $\alpha = 0.05$.

4.3.2. Comparisons of Four $PBLRT$s in terms of Power

From the simulated size study, it is established that the $PBLRT$s are complying with the level condition whereas the asymptotic tests are far from satisfactory. Since all the

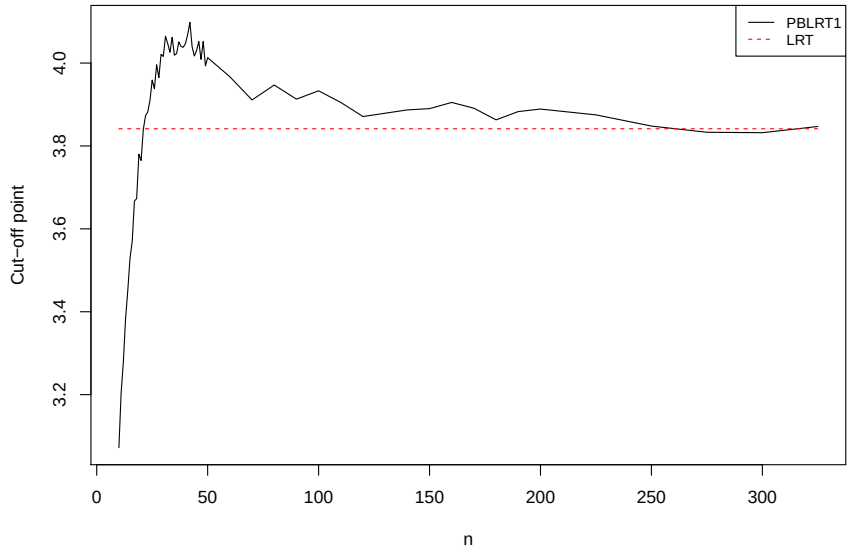


Figure 3: Plot of $PBLRT1$ cut-off points against n and comparison with fixed Chi-square $ALRT$ cut-off point (dashed line)

Table 3: Power of four $PBLRT$s at $\lambda \neq \lambda_0$ when $\lambda_0 = 0$ and $\alpha = 0.05$

n	Test	$\lambda = 0.1$	$\lambda = 0.2$	$\lambda = 0.3$	$\lambda = 0.4$	$\lambda = 0.5$	$\lambda = 0.6$	$\lambda = 0.7$	$\lambda = 0.8$	$\lambda = 0.9$
10	$PBLRT1$	0.052	0.055	0.061	0.067	0.078	0.093	0.110	0.125	0.149
	$PBLRT2$	0.054	0.055	0.059	0.074	0.082	0.095	0.107	0.127	0.150
	$PBLRT3$	0.051	0.055	0.059	0.064	0.078	0.091	0.106	0.127	0.142
	$PBLRT4$	0.051	0.055	0.059	0.064	0.078	0.091	0.107	0.126	0.142
20	$PBLRT1$	0.052	0.060	0.074	0.092	0.113	0.142	0.179	0.220	0.271
	$PBLRT2$	0.051	0.054	0.071	0.087	0.113	0.145	0.177	0.212	0.270
	$PBLRT3$	0.051	0.059	0.070	0.086	0.108	0.131	0.177	0.209	0.259
	$PBLRT4$	0.051	0.058	0.068	0.085	0.106	0.130	0.175	0.208	0.259
30	$PBLRT1$	0.054	0.064	0.084	0.110	0.146	0.192	0.245	0.315	0.388
	$PBLRT2$	0.051	0.065	0.080	0.107	0.141	0.192	0.246	0.309	0.382
	$PBLRT3$	0.058	0.064	0.083	0.107	0.140	0.197	0.240	0.307	0.387
	$PBLRT4$	0.059	0.065	0.082	0.108	0.139	0.197	0.242	0.308	0.389
40	$PBLRT1$	0.055	0.070	0.095	0.130	0.180	0.240	0.315	0.400	0.494
	$PBLRT2$	0.052	0.069	0.097	0.125	0.178	0.232	0.305	0.391	0.487
	$PBLRT3$	0.056	0.066	0.099	0.135	0.175	0.241	0.324	0.398	0.498
	$PBLRT4$	0.056	0.066	0.100	0.134	0.175	0.242	0.324	0.399	0.500
50	$PBLRT1$	0.056	0.075	0.109	0.154	0.217	0.292	0.384	0.488	0.591
	$PBLRT2$	0.055	0.077	0.114	0.162	0.217	0.300	0.388	0.500	0.595
	$PBLRT3$	0.053	0.075	0.110	0.156	0.208	0.292	0.380	0.471	0.588
	$PBLRT4$	0.052	0.074	0.108	0.156	0.207	0.289	0.379	0.470	0.588

$PBLRT$-s are α-reliable, in a bid to choose the most powerful test among the four proposed tests, we have studied the power of the $PBLRT$s for testing $H_0 : \lambda = 0$ vs $H_A : \lambda \neq 0$. Due to the symmetric parameter space, we have presented the power results in Table 3 on the positive part of the parameter space only.

Remark 4.3. The four $PBLRT$s, as it appears from the simulated power study in Table 3, are equivalent in nature. The difference in the power values are negligible. Each simulated value has an SE bounded above by 0.004. In other words, all the differences are within 2SE of each other. Hence to test the association parameter in $FGMD$ any one of the four tests is applicable and recommended.

5. An application with real-life datasets

5.1. Background of the dataset

Severe arsenic concentration in groundwater has been a major problem in densely populated glacial river downstream across the globe. One of the hotspots that has been hit by arsenic contamination in groundwater is the Mekong Delta Region (MDR) in southern part of Vietnam. Researchers have been carrying out surveys in the MDR to assess arsenic prevalence and to have a better understanding on how this carcinogenic element is associated with other elements present in groundwater. Along with the concentration of Arsenic (As) in the groundwater which is well known to have adverse effects on the human health when consumed beyond the tolerance limit, concentration of Chlorine (Cl), Redox level (Eh) and the pH level of the groundwater have been collected from several water-wells with locations scattered across the Dong Thap Province of the MDR as reported by Merola et al. (2015). The study area covers two sub regions, namely - North and South, and the complete dataset pertaining to the aforementioned elements have been reproduced in SI: Section 4.

The fact that the two subregions (North and South) are vastly different can be seen from the following Table 4 which gives the p-values of the Kolmogorov - Smirnov test (KST) which compares the empirical distributions of each element for the two subregions. From the

Table 4: KST P-values to test the equality of distributions in two subregions.

	As	Cl	Eh	pH
P-value	< 0.0001	0.0004	< 0.0001	< 0.0001

P - values it is evident that the distribution of each variable is significantly different for the two subregions i.e. Groundwater samples from the North (near Tan Hong) with a sample of size $n_T = 23$; and Groundwater samples from the South (near Thanh Binh) with a sample of size $n_S = 44$. SI: Figure 8 provides the sample relative frequency histograms of all the four elements for the two subregions mentioned above.

The relative frequency histograms of the four elements (As, Cl, Eh and pH) in two subregions (North and South) as shown in SI: Figure 8 clearly demonstrate mostly non-normality, and formal normality tests like Anderson - Darling test (ADT) and Shapiro - Wilks Test (SWT) do confirm this.

Table 5: ADT and SWT P-values to test the normality in two subregions

Test	Subregion	As	Cl	Eh	pH
ADT	North	< 0.0001	0.0414	0.1404	0.0029
	South	0.0001	< 0.0001	< 0.0001	0.4363
SWT	North	< 0.0001	0.0286	0.0914	0.0012
	South	0.0002	< 0.0001	< 0.0001	0.6933

Our main objective is to study the relationship between As and other three elements

through bivariate distributions using $FGMC$. The bivariate plots as shown in SI: Figure 9 show the highly non-linear association in the bivariate distributions.

Remark 5.1. The initial exploratory analysis points towards the fact that the MDR dataset consists of components which, firstly, have distinct distributions over the two subregions; and secondly, have mostly skewed marginal distributions. We further delve into the exploration of the nature of pairwise association present among the variables in this dataset. We employ the $FGMD$ model for modeling pairwise datasets as shown in Figure 9 in SI. However, one can raise a question about the goodness of fit of $FGMD$ to these datasets. It is worth noting that there is no "one stop solution" for the goodness of fit problem for the host of available copula in the literature and it remains an open problem. Several goodness of fit tests are available across the literature but to the best of our knowledge there doesn't exist one for $FGMC$. We have proposed and developed a novel data driven goodness of fit test for $FGMD$, which does not assume any known asymptotic distribution of the test statistic under the null hypothesis. A detailed study of this goodness of fit test including the test procedures, as well as the size and the power of the tests is given in Chatterjee and Pal (2024). The goodness of fit tests accept $FGMD$ for the aforementioned bivariate datasets.

5.2. Analysis of the groundwater data in MDR using $FGMD$

From the scatter plots in SI: Figure 9, all but (e) and (f) show possible nonlinear trends. To explore the association among the pairwise variables, a common approach is to resort to the standard correlation coefficients such as Pearson's (ρ_P), Spearman's (ρ_S) and Kendall's (τ). Pearson's correlation is inappropriate here because of the obvious non-linear association between the two components. On the other hand Spearman's and Kendall's correlation measures do go beyond linearity but is not of much help when it comes to regressing one variable on the other. In this regard our proposed $FGMD$ not only helps us in measuring the association between the two components under study but also helps us in regressing one variable from the other one thereby predicting Arsenic based on the other components in a pairwise set-up. Table 6 reports the standard correlation coefficients along with the corresponding (asymptotic) P-values in parentheses for testing H_0 : Standard Correlation $=$ 0 vs H_A : Standard Correlation $\neq 0$. These tests are easily available in any standard statistical package. The three standard correlation measures portray some interesting scenarios as noted in Remark 5.2.

Remark 5.2. (a) In the North, all three correlations indicate that there is no association between As and Cl. However, in the South, they all indicate a significant negative association.
(b) Between As and Eh, there appears to be a significant negative association in both the subregions. A conflicting picture provided by ρ_P in the North which shows a significant positive association. Between As and pH, there appears to have no significant association as all but one p-values are quite high (more than 20%). However, ρ_P shows a strong positive association.

Now we turn our attention to our proposed $FGMD$ model to the MDR dataset. Using

Table 6: Standard correlation coefficients along with P-Values

	North ($n_T = 23$)			South ($n_S = 44$)		
	As vs Cl	*As vs Eh*	*As vs pH*	*As vs Cl*	*As vs Eh*	*As vs pH*
ρ_P	0.182	0.525	0.754	-0.325	-0.668	0.119
	(0.405)	(0.010*)	(< 0.001***)	(0.031**)	(< 0.001***)	(0.442)
ρ_S	0.018	-0.414	0.260	-0.320	-0.753	0.156
	(0.936)	(0.050**)	(0.231)	(0.035**)	(< 0.001***)	(0.314)
τ	0.012	-0.323	0.188	-0.230	-0.577	0.101
	(0.937)	(0.032**)	(0.213)	(0.028**)	(< 0.001***)	(0.336)

Table 7: Estimates of the $FGMD$ association parameter in the MDR sub-regions

Pair of Elements	North				South			
	$\hat{\lambda}_{ML}$	$\hat{\lambda}_{BFP}$	$\hat{\lambda}_{BJP}$	$\hat{\lambda}_{BAJP}$	$\hat{\lambda}_{ML}$	$\hat{\lambda}_{BFP}$	$\hat{\lambda}_{BJP}$	$\hat{\lambda}_{BAJP}$
As vs Cl	0.085	0.053	0.064	0.108	-0.982	-0.621	-0.674	-0.81
As vs Eh	-1	-0.587	-0.646	-0.803	-1	-0.872	-0.892	-0.941
As vs pH	0.746	0.423	0.475	0.666	0.611	0.431	0.469	0.648

Table 8: $PBLRT$ test statistic values along with the simulated P-Values

Test	North			South		
	As vs Cl	*As vs Eh*	*As vs pH*	*As vs Cl*	*As vs Eh*	*As vs pH*
$PBLRT1$	0.017	3.959*	1.366	4.447*	17.099*	1.364
	(0.899)	(0.047**)	(0.263)	(0.039**)	(< 0.001***)	(0.251)
$PBLRT2$	0.013	2.792	1.108	3.841	15.266*	1.262
	(0.208)	(0.101*)	(0.485)	(0.076*)	(0.002***)	(0.503)
$PBLRT3$	0.015	2.986	1.186	4.006	15.577*	1.305
	(0.216)	(0.101*)	(0.491)	(0.076*)	(< 0.001***)	(0.506)
$PBLRT4$	0.016	3.475	1.350	4.309	16.280*	1.353
	(0.239)	(0.098*)	(0.503)	(0.076*)	(< 0.001***)	(0.495)

the estimates from the MDR dataset in Table 7, we proceed to perform $PBLRT1$, $PBLRT2$, $PBLRT3$ and $PBLRT4$ to test the hypothesis $H_0 : \lambda = 0$ vs $H_A : \lambda \neq 0$. Table 8 gives the $PBLRT$ test statistic values along with the simulated P-values.

Remark 5.3. (a) With the application of $FGMD$ we were able to estimate the underlying association among the pairwise variables using the four estimators. According to the estimates in Table 7 there exists a strong negative association between Eh and As in the northern region. $\hat{\lambda}_{ML}$ estimates the strongest negative association among the variables, followed by $\hat{\lambda}_{BAJP}$, $\hat{\lambda}_{BJP}$ and $\hat{\lambda}_{BFP}$. Only through $PBLRT1$ we see the presence of strong association between As and Eh in the northern subregion. The highly negative association between As and Eh in the northern subregion, which was partially captured by the Spearman's and Kendall's, is ratified by the estimates of the association parameter of $FGMD$ and $PBLRT1$ at a 5% level of significance.
(b) In the case of As vs pH it is crucial to note that in the northern sub-region, Spearman's rho and Kendall's Tau contradicted Pearson's correlation coefficient which showed a strong linear association.
(c) In the southern subregion, the standard correlation coefficients estimate a considerable

negative linear association between As and Eh. Although there is evidence of association between As and Eh, and to some extent in between As and Cl, but labeling it as a linear association will be an over simplification and inaccurate. The estimates in Table 7 of the association parameter λ shows a strong negative association between As and Eh, and in between As and Cl. The MLE registers the strongest association among the variables As and Eh followed by $\hat{\lambda}_{BAJP}$, $\hat{\lambda}_{BJP}$ and $\hat{\lambda}_{BFP}$. The same case holds for As and Cl as well. There is a positive association among the variables As and pH as estimated by all the estimators, reiterating the same phenomenon by our $FGMD$ model as that of Spearman's rho and Kendall's tau.

6. Observations and Concluding Remarks

(a) In our study we have provided a systematic approach of estimating the association parameter in $FGMD$ with the help of a few competing estimators. With the help of extensive simulation we have been able to demonstrate that the MLE may not be the best choice as an estimator for the association parameter. A family of Bayesian estimators shows a better performance in terms of MSE.

(b) In order to test $H_0 : \lambda = \lambda_0$ vs $H_A : \lambda \neq \lambda_0$, we have proposed four parametric bootstrap likelihood ratio type tests ($PBLRTs$) and through extensive simulation we have been able to show that they have better size and power than the usual asymptotic normal or the asymptotic likelihood ratio tests. We were able to demonstrate that a requirement of a large sample size (approximately $n = 300$ or higher) is necessary to attain the asymptotic property of the $ALRT$.

(c) The $FGMD$ is found to be appropriate in modeling Arsenic contamination in MDR in the presence of other benign elements. Through our proposed tests Arsenic can be highly associated some other benign elements which can be a beneficial information for the applied researchers.

(d) The authors will be happy to share the R codes used to run all computations with any interested reader/researcher.

ACKNOWLEDGMENTS

We would like to thank the Executive Editor and a Co-Editor of REVSTAT as well as two anonymous reviewers for their time and patience in handling our manuscript. We are particularly indebted to one anonymous reviewer for many constructive suggestions which have helped significantly in the presentation of our work.

REFERENCES

Abramowitz, M. and Stegun, I. A. (1964). *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, volume 55. US Government printing office.

Amblard, C. and Girard, S. (2009). A new extension of bivariate fgm copula. *Metrika*, 70(1):1–17.

Bairamov, I. and Kotz, S. (2002). Dependence structure and symmetry of huang-kotz fgm distributions and their extensions. *Metrika*, 56(1):55–72.

Bekrizadeh, H., Parham, G. A., and Zadkarmi, M. R. (2012). The new generalization of farlie - gumbel - morgenstern copulas. *Applied Mathematical Sciences*, 6(71):3527–3533.

Chatterjee, R. and Pal, N. (2024). A distribution-free goodness of fit test for copula model: An application to farlie–gumbel–morgernstern copula. *Journal of Statistical Theory and Practice*, 18(2):27.

D'este, G. (1981). A morgenstern-type bivariate gamma distribution. *Biometrika*, 68(1):339–340.

Farlie, D. J. (1960). The performance of some correlation coefficients for a general bivariate distribution. *Biometrika*, 47(3/4):307–323.

Gumbel, E. J. (1960). Bivariate exponential distributions. *Journal of the American Statistical Association*, 55(292):698–707.

Huang, J. S. and Kotz, S. (1999). Modifications of the farlie-gumbel-morgenstern distributions. a tough hill to climb. *Metrika*, 49(2):135–145.

Joe, H. (2005). Asymptotic efficiency of the two-stage estimation method for copula-based models. *Journal of Multivariate Analysis*, 94(2):401–419.

Kagan, A. M., Linnik, I., and Rao, C. R. (1973). Characterization problems in mathematical statistics. *Wiley*.

Kotz, S., Balakrishnan, N., and Johnson, N. L. (2004). *Continuous multivariate distributions, Volume 1: Models and applications*. John Wiley & Sons.

Ley, C., Babic, S., and Palangetic, M. (2021). Elliptical symmetry tests in r. *R Journal*, 13(2).

Merola, R., Hien, T., Quyen, D., and Vengosh, A. (2015). Arsenic exposure to drinking water in the mekong delta. *Science of the Total Environment*, 511:544–552.

Morgenstern, D. (1956). Einfache beispiele zweidimensionaler verteilungen. *Mitteilingsblatt fur Mathematische Statistik*, 8:234–235.

Nelsen, R. B. (2007). *An introduction to copulas*. Springer Science & Business Media.

Sklar, M. (1959). Fonctions de repartition an dimensions et leurs marges. *Publ. inst. statist. univ. Paris*, 8:229–231.

Suzuki, A. K., Louzada, F., and Cancho, V. G. (2013). On estimation and influence diagnostics for a bivariate promotion lifetime model based on the fgm copula: A fully bayesian computation. *TEMA (São Carlos)*, 14(3):441–461.

Taheri, B., Jabbari, H., and Amini, M. (2018). Parameter estimation of bivariate distributions in presence of outliers: An application to fgm copula. *Journal of Computational and Applied Mathematics*, 343:155–173.