Mallows Distance in VARFIMA$(0, \boldsymbol{d}, 0)$ Processes

Sílvia R.C. Lopes[‡], Guilherme Pumi and Karine Zaniol
Federal University of Rio Grande do Sul
Mathematics Institute

April, 30th, 2011

Abstract

In this work we present an extensive simulation study on Mallows distance in the context of Gaussian and non-Gaussian VARFIMA processes. Our main goal is to analyze the dependence among the components of VARFIMA processes through the Mallows distance point of view. We also investigate a possible relationship between the Mallows distance and the fractional differencing parameter $\boldsymbol{d}$, the type and level of dependence in the innovation process as well as its marginal behavior. For the Mallows distance, we consider an estimator based on the empirical marginal distribution function, which is shown to converge to the theoretical Mallows distance.

Keywords: Mallows Distance; VARFIMA Processes; Copulas; Empirical Estimation; Long Range Dependence, Kendall's τ Coefficient.

Mathematical Subject Classification (2010). Primary 62H20, 60G10, 62M10, 62E10; Secondary 54E99.

1 Introduction

The Mallows distance was introduced by Mallows (1972) as a tool to prove the asymptotic normality for sums of independent random variables. After this work, several other applications for the Mallows distance were found, especially in proving convergence of random variables and certain CLT-type results. In Bickel and Freedman (1981) the Mallows distance is used as a tool to provide asymptotic results for the bootstrap technique. An account of the theory and history of the Mallows distance can also be found there and references therein.

By its turn, the class of VARFIMA processes was introduced by Sowell (1989). It can be seen as a natural multidimensional extension of the classical ARFIMA processes, on which each component follows an ARFIMA process (see, for instance, Lopes, 2008 and Lopes, Sena Junior and Reisen, 2006), but the components in the innovation process can be correlated to each other. VARFIMA processes have been applied in a variety of fields such as hydrology, econometrics, statistics among others. Applications include modeling of stock prices' volatility, modeling and forecasting high frequency data, among others. See, for instance, Chiriac and Voev (2010), Diongue (2010) and references therein.

Our goal in this work is to investigate through extensive Monte Carlo simulations the Mallows distance behavior among the components of VARFIMA processes in several different settings. More specifically, we are interested in a possible relationship between

[‡]Corresponding author. E-mail: silvia.lopes@ufrgs.br

the Mallows distance and the fractional differencing parameter $\boldsymbol{d}$, the type and level of dependence induced in the innovation process as well as its marginal behavior.

The study is based on Monte Carlo simulations of VARFIMA$(0, \boldsymbol{d}, 0)$ processes with Gaussian innovations as well as parametric copula-type innovations. Different dependence parameters and different types of marginal behavior are also considered. The degree of dependence is measured, in the case of Gaussian innovations, by the correlation coefficient, and, in the copula innovations case, by its parameter. In the simulations we also consider the Kendall's τ dependence coefficient, which is used as a benchmark to compare with the results obtained by using the Mallows distance.

The Mallows distance estimator considered in the simulations is presented in Section 3 and is based on the empirical quantile function of the process' marginals. The estimator is shown to converge to the theoretical Mallows distance. Time series are generated by using the infinite moving average representation of the individual components in the process, truncated at a certain cut-off point. In the simulations, several different features are studied in the Mallows distance point of view. These features are usually introduced directly into the innovation process. Some copula tools are also explored in order to separate the joint dependence in the process from its marginals. The advantage of splitting the process' joint behavior from its marginal structure is the possibility to compare time series with the exactly same joint behavior but completely different marginal configurations.

The paper is organized as follows. In the next section we present some preliminary concepts and results necessary for this work. In Section 3 we introduce the empirical version of the Mallows distance we shall use to perform the simulation studies. We also prove its convergence to the theoretical Mallows distance. In Section 4 we present the simulation results on the Mallows distance among the components of VARFIMA processes in several different settings. In Section 5 we present the simulation results, in the same configuration as in Section 4, by using the Kendall's τ coefficient as dependence measure, instead of the Mallows distance. We also compare the results with the ones obtained in Section 4. Conclusions and final remarks are reserved to Section 6.

2 Preliminaries Concepts and Results

In this section we present some basic definitions and results necessary for this work. For $\alpha > 0$, let $\mathscr{F}_\alpha$ denote the space of all distribution functions satisfying $\int_{\mathbb{R}} |x|^\alpha \mathrm{d}F < \infty$.

Definition 2.1. (Mallows α-distance). Let $\alpha > 0$ and let F and G be two distribution functions in $\mathscr{F}_\alpha$. The *Mallows α-distance* of F and G is given by

$$\mathscr{D}_\alpha(F, G) := \inf_{A(F,G)} \Big\{ \mathbb{E}\big(|X - Y|^\alpha\big)^{\frac{1}{\alpha}} \Big\}, \tag{2.1}$$

where $A(F, G) := \big\{(X, Y) : X \sim F, Y \sim G\big\}$, that is, $A(F, G)$ is the set of all pairs (X, Y) of random variables with marginals given by F and G, respectively.

It can be shown that, for $\alpha \geq 1$, $\mathscr{D}_\alpha(\cdot, \cdot)$ is a metric in $\mathscr{F}_\alpha$, while for $\alpha < 1$, $\mathscr{D}_\alpha^\alpha(\cdot, \cdot)$ is a metric in $\mathscr{F}_\alpha$ (see Bickel and Freedman, 1981). If $\alpha \geq 1$, (2.1) can be shown to be equivalent to a much simpler expression as follows: let $U \sim U(0, 1)$ be a uniformly

distributed random variable, let $F, G \in \mathscr{F}_\alpha$ and set $X^* := F^{-1}(U)$ and $Y^* := G^{-1}(U)$. Then, it can be shown that

$$\mathscr{D}_\alpha^\alpha(F,G) = \mathbb{E}\big(|X^* - Y^*|^\alpha\big). \tag{2.2}$$

Alternative expressions for (2.2) are the following:

$$\mathscr{D}_\alpha^\alpha(F,G) = \int_0^1 \big|F^{-1}(u) - G^{-1}(u)\big|^\alpha \mathrm{d}u = \iint_{\mathbb{R}^2} |x-y|^\alpha \mathrm{d}\mu, \tag{2.3}$$

where μ is the probability measure in $\mathbb{R}^2$ with joint distribution function given by $H(x,y) = \min\{F(x), G(y)\}$, that is, μ is defined in the semi-ring of rectangles $R = [x_1, x_2]\times[y_1, y_2] \subseteq \mathbb{R}^2$ by

$$\begin{aligned}\mu(R) \;=\; & \min\{F(x_1), G(y_1)\} + \min\{F(x_2), G(y_2)\} - \\ & - \min\{F(x_1), G(y_2)\} - \min\{F(x_2), G(y_1)\}.\end{aligned}$$

The representations in (2.3) are useful for estimation purposes.

Now let X and Y be two continuous random variables defined in a common probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Let (X_1, Y_1) and (X_2, Y_2) be two independent copies of (X, Y). Given $\omega \in \Omega$, the pairs $\big(X_1(\omega), Y_1(\omega)\big)$ and $\big(X_2(\omega), Y_2(\omega)\big)$ are called *concordant* if $\big(X_1(\omega) - X_2(\omega)\big)\big(Y_1(\omega) - Y_2(\omega)\big) > 0$ and *discordant* if $\big(X_1(\omega) - X_2(\omega)\big)\big(Y_1(\omega) - Y_2(\omega)\big) < 0$ (equality happens with probability 0). The Kendall's τ coefficient between X and Y, denoted by $\tau_{X,Y}$ (or simply by τ if no confusion is possible), is defined as the probability of concordance minus the probability of discordance, that is,

$$\tau = \tau_{X,Y} := \mathbb{P}\big((X_1 - X_2)(Y_1 - Y_2) > 0\big) - \mathbb{P}\big((X_1 - X_2)(Y_1 - Y_2) < 0\big).$$

Next we define the so-called VARFIMA$(p, \boldsymbol{d}, q)$ processes.

Definition 2.2. Let $\{\boldsymbol{X}_t\}_{t\in\mathbb{Z}}$ be an m-dimensional process with mean $\boldsymbol{\mu}$. The process $\{\boldsymbol{X}_t\}_{t\in\mathbb{Z}}$ is called a VARFIMA$(p, \boldsymbol{d}, q)$ process if it is a stationary solution of the difference equations

$$\boldsymbol{\Phi}(\mathcal{B})\,\mathrm{diag}\big\{(1-\mathcal{B})^{\boldsymbol{d}}\big\}(\boldsymbol{X}_t - \boldsymbol{\mu}) = \boldsymbol{\Theta}(\mathcal{B})\boldsymbol{\varepsilon}_t, \tag{2.4}$$

where $\mathcal{B}$ is the backward shift operator, $\{\boldsymbol{\varepsilon}_t\}_{t\in\mathbb{Z}}$ is an m-dimensional stationary process (the innovation process), $\boldsymbol{\Phi}(\mathcal{B})$ and $\boldsymbol{\Theta}(\mathcal{B})$ are $m\times m$ matrices in $\mathcal{B}$, given by the equations

$$\boldsymbol{\Phi}(\mathcal{B}) = \sum_{\ell=0}^{p} \boldsymbol{\phi}_\ell \mathcal{B}^\ell \;\text{ and }\; \boldsymbol{\Theta}(\mathcal{B}) = \sum_{\ell=0}^{q} \boldsymbol{\theta}_\ell \mathcal{B}^\ell,$$

with $\boldsymbol{\phi}_1, \cdots, \boldsymbol{\phi}_p, \boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_q$ real $m\times m$ matrices and $\boldsymbol{\phi}_0 = \boldsymbol{\theta}_0 = \boldsymbol{I}_{m\times m}$, the $m\times m$ identity matrix.

Notice that Definition 2.2 is more general than the classical definition of VARFIMA processes, introduced by Sowell (1989), in which $\boldsymbol{\varepsilon}_t$ is assumed to be Gaussian. In this work we analyze only the simpler case $p = 0 = q$, for which equation (2.4) simplifies to

$$\mathrm{diag}\big\{(1-\mathcal{B})^{\boldsymbol{d}}\big\}(\boldsymbol{X}_t - \boldsymbol{\mu}) = \boldsymbol{\varepsilon}_t, \quad \text{for all } t \in \mathbb{Z}.$$

It can be shown that, if $\boldsymbol{\varepsilon}_t$ has finite variance for all $t \in \mathbb{Z}$, a necessary and sufficient condition for the existence of stationary solutions for (2.4) is that $\boldsymbol{d} \in (-\infty, 0.5)^m$. On the other hand, to guarantee that the stationary solution is also causal and invertible, it can be shown that we must have $\boldsymbol{d} \in (-0.5, 0.5)^m$ and this will be the range we shall assume for the fractional parameter $\boldsymbol{d}$.

VARFIMA processes can be seen as a natural extension of the classical ARFIMA processes. For the ARFIMA case, the so-called *long range dependence* occurs whenever the fractional differencing parameter $d \in (0, 0.5)$ while if d is in the $(-0.5, 0)$ zone, some authors refer to it as the *intermediate dependence* (see, for instance Lopes, 2008 and Lopes, Sena Junior and Reisen, 2006). In the case of VARFIMA process, we shall say that the k-th coordinate process $\{X_t^{(k)}\}_{t\in\mathbb{Z}}$ present *long range dependence* (*intermediate dependence*) whenever $d_k \in (0, 0.5)$ ($d_k \in (-0.5, 0)$), for $k \in \{1, \cdots, m\}$.

Several applications and extensions for VARFIMA processes (and general fractionally differentiated multivariate models) have been studied in recent years, see for instance, Chiriac and Voev (2010), Diongue (2010) and references therein. For estimation in VARFIMA processes, see Lobato (1999), Shimotsu (2007), Tsay (2010) and references therein. More details can also be found in Sowell (1989) and Luceño (1996).

A few results on copulas will also be necessary. The literature on the subject has grown very rapidly in the last decade especially in finance, statistics and econometrics where copulas have been widely used as tools for analyzing and modeling financial time series. An m-dimensional copula is a distribution function whose marginals are uniformly distributed on $[0, 1]$ and whose support is the $[0, 1]^m$ hypercube. The main theorem in the theory is the celebrated Sklar's theorem, which elucidates the usefulness of copulas.

Theorem 2.1 (Sklar's Theorem)**.** *Let $X_1, \cdots, X_n$ be random variables with marginals $F_1, \cdots, F_n$, respectively, and joint distribution function H. Then, there exists a copula C such that,*

$$H(x_1, \cdots, x_n) = C\big(F_1(x_1), \cdots, F_n(x_n)\big), \quad \textit{for all } \ (x_1, \cdots, x_n) \in \mathbb{R}^n.$$

If F_i's are continuous, then C is unique. Otherwise, C is uniquely determined on $\mathrm{Ran}(F_1) \times \cdots \times \mathrm{Ran}(F_n)$*. The converse also holds. Furthermore,*

$$C(u_1, \cdots, u_n) = H\big(F_1^{(-1)}(u_1), \cdots, F_n^{(-1)}(u_n)\big), \quad \textit{for all } \ (u_1, \cdots, u_n) \in I^n,$$

where for a function F, $F^{(-1)}$ denotes its pseudo-inverse given by $F^{(-1)}(x) = \inf\big\{u \in \mathrm{Ran}(F) : F(u) \geq x\big\}$.

A proof of Sklar's Theorem can be found, for instance, in Schweizer and Sklar (2005).

Remark 2.1. Among many applications of Sklar's Theorem, one will be particularly useful in the simulations. Suppose we have an m-dimensional continuous random vector $\boldsymbol{X}$, with marginal distributions $F_1, \cdots, F_m$, for $m > 1$. Suppose that we want to estimate some quantity and investigate what happens when the marginal behavior of $\boldsymbol{X}$ is changed to, say, $G_1, \cdots, G_m$, but the joint dependence structure is kept as intact as possible. If one knows the copula $C_{\boldsymbol{X}}$ associated to $\boldsymbol{X}$, this problem can be easily solved. Let $\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n$ be a sample from the copula $C_{\boldsymbol{X}}$, where $\boldsymbol{u}_k = (u_k^{(1)}, \cdots, u_k^{(m)})$, $k = 1, \cdots, n$. Consider the following samples based on $\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n$:

a) $\{\boldsymbol{x}_k\}_{k=1}^n$, where $x_k^{(j)} = F_j^{-1}(u_k^{(j)})$, for $k = 1, \cdots, n$ and $j = 1, \cdots, m$.

b) $\{\boldsymbol{y}_k\}_{k=1}^n$, where $y_k^{(j)} = G_j^{-1}(u_k^{(j)})$, for $k = 1, \cdots, n$ and $j = 1, \cdots, m$.

By Sklar's Theorem, $\{\boldsymbol{x}_k\}_{k=1}^n$ and $\{\boldsymbol{y}_k\}_{k=1}^n$ are samples with the same joint dependence as $\boldsymbol{X}$, but the former has marginals $F_1, \cdots, F_m$, while the latter has marginals $G_1, \cdots, G_m$. One can now calculate and compare the quantity of interest by using $\{\boldsymbol{x}_k\}_{k=1}^n$ and $\{\boldsymbol{y}_k\}_{k=1}^n$. This method allows one to study how the marginal behavior affects some quantity of interest by keeping the joint behavior of the sample (determine by $C_{\boldsymbol{X}}$) fixed and introducing the features of interest directly into the marginals.

For more details on copulas, we refer the reader to Nelsen (2006). For connections with probabilistic metric spaces, see Schweizer and Sklar (2005).

3 Empirical Version of the Mallows Distance

In this section we shall define the Mallows distance estimator considered in the simulations and prove its convergence to the theoretical Mallows distance. Given two i.i.d. samples $X_1, \cdots, X_n$ and $Y_1, \cdots, Y_n$ from distributions F and G, respectively, let $\widehat{F}_n$ and $\widehat{G}_n$ denote the empirical distribution functions based on these samples. Then, for any $\alpha \geq 1$ the Mallows α-distance estimator is given by

$$\widehat{\mathscr{D}}_\alpha(F, G) := \mathscr{D}_\alpha(\widehat{F}_n, \widehat{G}_n) = \left(\int_0^1 \left| \widehat{F}_n^{(-1)}(u) - \widehat{G}_n^{(-1)}(u) \right|^\alpha \mathrm{d}u \right)^{1/\alpha}. \tag{3.5}$$

The Glivenko-Cantelli theorem assures that $\widehat{F}_n(x) \to F(x)$ uniformly over x except, perhaps, in a set of measure zero, which also implies $\widehat{F}_n^{(-1)}(x) \to F^{(-1)}(x)$ almost everywhere uniformly over x. Also, it can be shown that $\mathscr{D}_\alpha(f_n, f) \to 0$ for a sequence $\{f_n\}_n \in \mathscr{F}_\alpha$ and $f \in \mathscr{F}_\alpha$, implies $f_n \to f$ (see, for instance, Shao, 2003). Our aim is to show that

$$\mathscr{D}_\alpha(\widehat{F}_n, \widehat{G}_n) \longrightarrow \mathscr{D}_\alpha(F, G).$$

Proposition 3.1. *Let $\alpha \geq 1$ and let $\widehat{F}_n$ and $\widehat{G}_n$ denote the empirical distribution functions based on i.i.d. samples $X_1, \cdots, X_n$ and $Y_1, \cdots, Y_n$ from distributions F and G in $\mathscr{F}_\alpha$, respectively. Then,*

$$\mathscr{D}_\alpha(\widehat{F}_n, \widehat{G}_n) \underset{n\to\infty}{\longrightarrow} \mathscr{D}_\alpha(F, G).$$

Proof. First, it is clear that $\widehat{F}_n$ and $\widehat{G}_n$ are in $\mathscr{F}_\alpha$. Since, for $\alpha \geq 1$, the α-Mallows distance is a metric in $\mathscr{F}_\alpha$, on one hand we have

$$\mathscr{D}_\alpha(F, G) \leq \mathscr{D}_\alpha(F, \widehat{F}_n) + \mathscr{D}_\alpha(\widehat{F}_n, \widehat{G}_n) + \mathscr{D}_\alpha(\widehat{G}_n, G),$$

and on the other hand

$$\mathscr{D}_\alpha(\widehat{F}_n, \widehat{G}_n) \leq \mathscr{D}_\alpha(\widehat{F}_n, F) + \mathscr{D}_\alpha(F, G) + \mathscr{D}_\alpha(G, \widehat{G}_n),$$

from where it follows that

$$\left| \mathscr{D}_\alpha(\widehat{F}_n, \widehat{G}_n) - \mathscr{D}_\alpha(F, G) \right| \leq \mathscr{D}_\alpha(\widehat{F}_n, F) + \mathscr{D}_\alpha(\widehat{G}_n, G) \longrightarrow 0,$$

since $\mathscr{D}_\alpha(\widehat{F}_n, F) \to 0$ and $\mathscr{D}_\alpha(\widehat{G}_n, G) \to 0$, as n goes to infinity. This shows that $\mathscr{D}_\alpha(\widehat{F}_n, \widehat{G}_n) \to \mathscr{D}_\alpha(F, G)$ and completes the proof. □

Asymptotic results for $\mathscr{D}_\alpha(\widehat{F}_n, F)$ are known in some special cases. For instance, some CLT-type results, convergence rates and limiting distributions for $\mathscr{D}_\alpha(\widehat{F}_n, F)$ can be found in Samworth and Johnson (2004) and references therein and also in Johnson and Samworth (2005). At our best knowledge, there are no asymptotic results available for $\mathscr{D}_\alpha(\widehat{F}_n, \widehat{G}_n)$.

4 Simulation Results: Mallows Distance

In this section we present the Monte Carlo simulation results regarding the Mallows distance between the components of a bidimensional VARFIMA$(0, \boldsymbol{d}, 0)$. In the simulations, the fractional differencing parameter $\boldsymbol{d} := (d_1, d_2)$ is taken to range over all combinations of $d_i \in \{-0.4, -0.3, -0.2, -0.1, 0.1, 0.2, 0.3, 0.4\}$, $i = 1, 2$. In this work, we always calculate the Mallows α-distance for $\alpha = 2$, and refer to it simply by Mallows distance. The estimator used is the one presented in (3.5).

All Monte Carlo simulations are based on time series of fixed sample size 2,000 obtained from bidimensional VARFIMA$(0, \boldsymbol{d}, 0)$ processes. We perform 1,000 replications of each experiment. To generate the time series, we apply the traditional method of truncating the multidimensional infinite moving average representation of the process. The truncation point is fixed in 50,000 for all $\boldsymbol{d}$.

All simulations are performed using the computational resources from the (Brazilian) National Center of Super Computing (CESUP-UFRGS). The routines are all implemented in FORTRAN 95 language optimized by using OpenMP directives for parallel computing.

We start by presenting the results for the Gaussian innovation case.

4.1 Gaussian innovations with equal variances

Figure 1 shows the graph of d_2 by Mallows distance for different correlations. The results are based on Gaussian innovations with fixed mean $\boldsymbol{\mu} = (0, 0)$ and variance $\boldsymbol{\sigma}^2 = (1, 1)$, for correlations $\rho \in \{0, 0.5, 0.95\}$.

An interesting feature shown in Figure 1[1] is that, for small values of $\boldsymbol{d}$ (both coordinates smaller than 0.1), the Mallows distance behaves homogeneously across different correlation values. This behavior suggests that the Mallows distance is not significatively sensitive to the correlation for small values of d_i, $i = 1, 2$.

There is, however, difference when both coordinates start to increase. A clear differentiation across the correlation appears when the parameters d_1 and d_2 are both greater or equal than 0.2. From the graphs on Figure 1, we can infer that the greater the long range dependence in each coordinate is (or, equivalently, the greater the values of d_1 and d_2

[1]Tables containing the results from which the graphs are draw from are not presented here due to the restriction on the number of pages. They can be found, along with additional graphs and information, as an addendum at http://mat.ufrgs.br/∼slopes/selected_publications.htm.

are) the greater the difference among the estimated Mallows distance values for different correlations. Also notice that, when at least one coordinate of $\boldsymbol{d}$ is small (less or equal than 0.1), the Mallows distance values behave like an increasing function of the other coordinate.

From the graphs, it is clear that the correlation starts to influence and differentiate the Mallows distance values only when both coordinates are greater than 0.1. Furthermore, it appears that the magnitude of the values in the fractional differencing parameter $\boldsymbol{d}$ has more influence in differentiating the Mallows distance than the magnitude of the correlation itself. However, as expected, the Mallows distance values decrease as the correlation increases for almost all cases.

We can summarize our findings as follows:

1. The Mallows distance generally decreases as the correlation increases for almost all $\boldsymbol{d}$ and behaves like an exponential when the coordinates of $\boldsymbol{d}$ assume values over 0.1;

2. The Mallows distance appears not be affected by the correlation when at least one coordinate in $\boldsymbol{d}$ is smaller than 0.2.

3. The higher the parameter $\boldsymbol{d}$, the greater the difference among the Mallows distance estimates across different correlations. This suggests that the Mallows distance is less sensitive to the correlation than to the parameter $\boldsymbol{d}$.

4.2 Gaussian innovations with unequal variances

The results from the previous subsection bring some light into the dynamics between the components of Gaussian VARFIMA processes from the Mallows distance point of view. All simulations involving Gaussian innovations are performed by using a variance-covariance matrix whose values in the main diagonal are identical (equal to one, to be precise).

In this section we investigate the following question: how (if at all) the Mallows distance behavior change when the innovation's second moments are altered? In other words, we study what happens if the innovation process' marginals have different variances. A simple approach is to compare the case where the marginals have equal variances to the case where they are different in the spirit of Remark 2.1.

Figure 2 shows the graphs of d_2 by Mallows distance for fixed d_1 and correlation $\rho \in \{0, 0.5, 0.95\}$. The variance was taken to be $\boldsymbol{\sigma}^2 = (1, 2)$. From the graphs, it is clear that increasing the variance of one innovation component also increases the Mallows distance in comparison to the equal variances case (Figure 1). Increasing the correlation in the unequal variances case produces little to no difference in the global behavior of the Mallows distance, except when both innovation components present strong long range dependence, particularly when $\boldsymbol{d} = (0.4, 0.4)$. Notice that the magnitude of the Mallows distance is larger in the unequal variances case, as can be seen by the scale on the graphs. Figures 1 and 2 show that the equal variances case present a more erratic Mallows distance behavior across correlation when compared to the smooth curves on the unequal variances one. As a function of d_2, high values of d_1 (> 0.2) produce a more

erratic behavior for the Mallows distance, which becomes sensitive to the correlation for $d_2 > 0.2$. We also observe that the Mallows distance behaviors in Figure 2(a)-(f) are all very similar to each other. Surprisingly, the results suggest that the different variances in the marginals stabilize the Mallows distance behavior, weakening the influence of the fractional differencing parameter $\boldsymbol{d}$.

We can summarize our findings as follows:

1. Compared to the equal variances case, the unequal variances situation is more stable with respect to the Mallows distance and to the parameter $\boldsymbol{d}$, so that for $d_i \leq 0.2$, there is little to no difference in the Mallows distance values within d_2.

2. The magnitude of the Mallows distance is generally larger than the equal variances case.

3. The difference in the marginal variances appear to reduce the influence of the fractional differencing parameter $\boldsymbol{d}$ causing the Mallows distance to be less sensitive to the correlation in the innovation.

4.3 More on the unequal variances

After studying the differences when the innovations process have unequal variances ($\boldsymbol{\sigma}^2 = (1,1)$ and $\boldsymbol{\sigma}^2 = (1,2)$ cases), two questions naturally arise:

1. How (if at all) does the magnitude of the components of $\boldsymbol{\sigma}^2$ influence the Mallows distance? That is, compared to the case $\boldsymbol{\sigma}^2 = (1,2)$, will the Mallows distance significantly change if we take $\boldsymbol{\sigma}^2 = (2,3)$?

2. How (if at all) does the magnitude of the difference between the components of $\boldsymbol{\sigma}^2$ influence the Mallows distance? That is, compared to the case $\boldsymbol{\sigma}^2 = (1,2)$, will the Mallows distance change much if we take $\boldsymbol{\sigma}^2 = (1,3)$?

In order to answer the questions above, we apply the ideas explained in Remark 2.1 to simulate Gaussian VARFIMA processes with innovation marginal variances equal to $\boldsymbol{\sigma}^2 = (1,3)$ and $\boldsymbol{\sigma}^2 = (2,3)$. We compare the results with the case $\boldsymbol{\sigma}^2 = (1,2)$ from last subsection (the simulations were actually performed all together using the ideas in Remark 2.1). We apply the Gaussian copula with parameter ρ and the marginals are taken to be normally distributed with zero mean and the desired marginal variances. Also, since the marginals are normally distributed, the parameter ρ still represents the correlation between the components.

Figure 3 and 4 show the graphs of d_2 by Mallows distance for fixed d_1 for variances $\boldsymbol{\sigma}^2 = (1,3)$ and $\boldsymbol{\sigma}^2 = (2,3)$, respectively. These are the analogous of Figure 1 (equal variances case) and 2 ($\boldsymbol{\sigma}^2 = (1,2)$). Upon analyzing the graphs, one notice that cases $\boldsymbol{\sigma}^2 = (1,2)$ and $\boldsymbol{\sigma}^2 = (2,3)$ present very similar magnitudes for the Mallows distance values, while the case $\boldsymbol{\sigma}^2 = (1,3)$ present much larger values. The equal variances case present the smallest values among all. This indicates that the difference between the components in $\boldsymbol{\sigma}^2$ has a strong influence on the magnitude of the Mallows distance, stronger than the magnitude of the components in $\boldsymbol{\sigma}^2$.

For $d_1 < 0.2$ (Figures 2, 3 and 4 (a)-(e)), we notice that the overall behavior of the Mallows distance is similar among the unequal variances case and basically no differentiation across correlation can be observed in any of the graphs. For $d_1 > 0.2$, we observe that the cases $\boldsymbol{\sigma}^2 = (1, 2)$ and $\boldsymbol{\sigma}^2 = (2, 3)$ are closer to each other than to any other cases. For $d_1 = 0.3$, in the unequal variances case, a small differentiation across correlation start to surface for $d_2 > 0$, which become stronger when $d_1 = 0.4$. We notice that the differentiation for $d_1 > 0.2$ is stronger when $\boldsymbol{\sigma}^2 = (2, 3)$ which may indicate that the magnitude of the $\boldsymbol{\sigma}^2$ components somehow affects the Mallows distance differentiation across correlation. Also the Mallows distance behavior for $d_1 = 0.4$ is similar for $\boldsymbol{\sigma}^2 \in \{(1, 1), (1, 2), (2, 3)\}$, but clearly different for $\boldsymbol{\sigma}^2 = (1, 3)$. In comparison to the equal variances case, it seems that the components of $\boldsymbol{\sigma}^2$ strongly influences the Mallows distance behavior, making it more stable across the correlation. Notice the resemblance of the graphs for $d_1 = 0.3$ (frame (g) in the respective figures) in the unequal variance case to the one for $d_1 = 0.2$ (in Figure 1(f)) in the case of equal variances. This also happens for the case $d_1 = 0.4$ which is close to the case $d_1 = 0.3$ in the equal variances case.

We conclude that the magnitude of the Mallows distance responds positively to both, the magnitude and the difference between the components of $\boldsymbol{\sigma}^2$, but clearly the response is stronger to the latter. This is no surprise, since the higher the difference between the variances of two normally distributed random variables with same mean, the more distant the values assumed by a sample of each are, which is directly reflected into the Mallows distance values.

We can summarize our findings as follows:

1. The Mallows distance respond positively to the difference between the components of $\boldsymbol{\sigma}^2$. The higher the difference, the higher the magnitude of the Mallows distance.

2. The magnitude of $\boldsymbol{\sigma}^2$ seems to influence positively the Mallows distance sensitivity regarding the correlation, especially for high $\boldsymbol{d}$. That is, the higher the magnitude of $\boldsymbol{\sigma}^2$, the more sensitive the Mallows distance become with respect to the correlation in the presence of strong long range dependence.

3. The overall Mallows distance behavior seem to be unaffected by the unequal variances in the innovation process, except for the magnitude of the values.

4. For $d_1 < 0.2$, the correlation has no affect in the Mallows distance in the unequal variances case. Also, it appears that the influence of the fractional differencing parameter $\boldsymbol{d}$ is attenuated in the unequal variance case.

4.4 Non-Gaussian innovation and heavy-tailed marginals

So far, all results presented are based on the bivariate Gaussian distribution. A question that naturally arises is does the Mallows distance behave in the same way if the innovations generating the VARFIMA process have a distribution other than the bivariate Gaussian? In order to partially answer this question, we simulate innovations from the Frank copula with different parameters and added to it, via Sklar's theorem, standard normal marginals. In this way we obtain a sample with dependence following a Frank copula, but standard

normal marginals, from which we generate the VARFIMA$(0, \boldsymbol{d}, 0)$ process and calculate the empirical Mallows distance between the components of the process. Our aim is to compare the results obtained this way with the bivariate Gaussian with standard normal marginals ones presented in Subsection 4.1.

In order to make a fair comparison, we would like to somehow match the dependence strength in the Frank innovation case with the Gaussian innovation case. We use the Kendall's τ theoretical values as a measure of the dependence strength in the innovation. The Kendall's τ for the Gaussian case with correlation ρ and for the Frank copula with parameter θ are given, respectively, by

$$\tau_\rho = \frac{2}{\pi}\arcsin(\rho) \quad \text{and} \quad \tau_\theta = 1 - \frac{4}{\theta^2}\left(\theta - \int_0^\theta \frac{t}{e^t - 1}\,\mathrm{d}t\right).$$

To match the Kendall's τ in the Gaussian case for $\rho \in \{0, 0.5, 0.95\}$, a good approximation is $\theta \in \{0, 3.3, 18\}$. These are the values chosen for the simulations.

Figure 5 presents the results for the Mallows distance between the components of VARFIMA$(0, \boldsymbol{d}, 0)$ process with Frank innovations and standard normal marginals. This is the analogous of Figure 1 in the Gaussian case. From the figures, we observe in both the same shape and almost the same magnitude of the Mallows distance. We observe that when both components of $\boldsymbol{d}$ have high values (≥ 0.2), there is a clear differentiation across the parameters. The higher the dependence in the innovation, the higher the Mallows distance sensitivity to this dependence in both cases. For high values of d_1, we observe that the Mallows distance values are slightly higher in the Frank-Normal case.

Another question not discussed so far is whether or not the marginals distribution tails influence the Mallows distance in any way. In other words, is there any difference in the Mallows distance behavior if we only change the type of the marginal from, say, Gaussian marginals to heavy-tailed ones?

To answer this question we apply the ideas of Remark 2.1 to simulate Frank copula innovations with parameter $\theta \in \{0, 3.3, 18\}$ coupled with standard normal, t_3 and t_7 marginals, where, as usual, t_ν stands for the Student's t distribution with ν degrees of freedom. We recall that our previous experiments indicate that the innovation variances interfere in the Mallows distance. Since t_3, t_7 and normal distributions have different variances, we use a standardized version of the t distribution with unitary variance to avoid any differences in the Mallows distance from sources other than the marginals' tail behavior. For simplicity, whenever t_3 and t_7 marginal cases are mentioned, we mean the respective standardized version.

Figures 5 shows the simulation results for the Frank-Normal couple. Figure 6, the Frank-t_3 couple results are shown. In Figure 7, the case Frank-t_7 is presented. Comparing Figures 5 and 7, we notice that they all look very similar. In fact, the absolute difference on the estimated Mallows distance between the case t_7 and standard normal marginals ranges on $[0.2311, 1.5897]$, for all $\boldsymbol{d}$. Also in these cases (Frank-Normal and Frank-t_7), the Mallows distance behavior follows a similar pattern to the Gaussian case with equal variances.

In Figure 6, we observe that the Mallows distance values in the Frank-t_3 case are higher than the respective ones in the other cases, but the overall curve pattern is the

same. A differentiation across the parameter occurs only when at least one coordinate of $\boldsymbol{d}$ is greater or equal than 0.2. Otherwise, little differentiation appears.

The similarities between the cases t_7 and standard normal marginals are not really surprising because the difference in the tail of these distribution is small. However, the t_3 marginals case shows that the Mallows distance is sensitive to tail fatness in the innovation, which is reflected mainly in the magnitude of the Mallows distance. The pattern followed by the Mallows distance however, does not seem to be significatively affected neither by the innovation's non-Gaussianity, nor by the marginals' tails. We can summarize our findings as follows:

1. The Mallows distance behavior does not seem to be affected by the type of innovations, within the same dependence strength (here measured by the innovation theoretical Kendall's τ).
2. The marginal's tail seem to have little influence in the Mallows distance. Nevertheless, when present, this effect seem basically to be reflected in the magnitude of the Mallows distance values, which are slightly higher in the presence of heavy-tailed marginals.

Remark 4.1. The tables from which the graphs presented in this section were drawn from, along with the estimates' standard deviation and extra graphs, can be found as an addendum at http://mat.ufrgs.br/∼slopes/selected_publications.htm. The standard deviation for the estimated Mallows distance may look, at first glance, high compared to the magnitude of the respective estimate. This is because the sample distribution of the estimated Mallows distance (which is always non-negative) are generally skewed to the right, but concentrated at the mean. This suggest that the limiting distribution for the empirical Mallows distance, as defined in (3.5), is skewed to the right.

5 Simulation Results: Kendall's τ comparison

In the previous section we investigated the behavior of the Mallows distance between the components of VARFIMA$(0, \boldsymbol{d}, 0)$ processes in several contexts. So, as a measure of how close two process are our findings natural or surprising? Are they shared for all types of dependence measure or are they unique to the Mallows distance? In order to provide a comparison, we perform the same experiment presented in Section 4 applying the Kendall's τ as a dependence measure instead of the Mallows distance. To insure fidelity, we performed the calculations using the same methodology as in the previous section. We start by presenting the Gaussian noise with equal variance case.

5.1 Gaussian innovations with equal variances

Figure 8[2] presents the plots of d_2 by Kendall's τ for fixed d_1 and $\rho \in \{0, 0.5, 0.95\}$. This is the analogous of Figure 1 in the Mallows distance case. Notice that the Kendall's τ is

[2]Tables containing the results from which the graphs are draw can also be found in the addendum at http://mat.ufrgs.br/∼slopes/selected_publications.htm. The tables also contain the standard deviation for the estimated Kendall's τ presented in this section. They are generally small, as one could expect given its asymptotic distribution.

much more sensitive to the correlation in the innovation than the Mallows distance. As the difference between the parameters d_1 and d_2 increases, the difference on the Kendall's τ between the components become higher and the range of the Kendall's τ values also increases as $|d_1|$ increases.

Also notice that, for $\rho = 0$, a simple hypothesis test show that in most cases (58 out of 64), the components can be regarded as statistically independent[3]. The exceptions are the values of d_i for which $d_1 + d_2 \geq 0.6$, that is, when there is strong long range dependence.

5.2 Gaussian innovations with unequal variances

In this subsection we present simulation results analogous to those in Subsection 4.3 in the context of the Kendall's τ. Figures 8 to 11 present the graphs of d_2 by the Kendall's τ for fixed d_1 for the cases $\boldsymbol{\sigma}^2$ equal to $(1,1), (1,2), (1,3)$ and $(2,3)$, respectively. From all graphs, it is clear that the variance strongly influences the Kendall's τ between the components. We notice the differences in the scale between unequal and equal variances cases. The much smaller values in the unequal variances case indicates that the different variances in the innovation somehow balances the number of concordant and discordant pairs, so that, at the Kendall's τ point of view, the components are more distinct compared to the equal variances case. For $\rho \neq 0$, even though the magnitude of the Kendall's τ is small, the components cannot be considered statistically independent regardless $\boldsymbol{d}$. For $\rho = 0$ and $\boldsymbol{\sigma}^2 = (1,2)$, the independence hypothesis is always rejected when $d_i > 0$, for $i = 1, 2$, and only in 3 (out of 48) combinations the hypothesis of independence is rejected when at least one parameter in $\boldsymbol{d}$ is negative. Similar results hold for $\boldsymbol{\sigma}^2 \in \{(1,3), (2,3)\}$.

The Kendall's τ behavior as the correlation increases is erratic, especially compared to the smooth behavior in the equal variances case. In all cases, a more or less similar pattern is followed and, arguably, the cases $\boldsymbol{\sigma}^2 = (1,2)$ and $\boldsymbol{\sigma}^2 = (2,3)$ are more alike then any other combination. Based on these results, it is clear that the variance heavily influences the Kendall's τ behavior. However, it is not clear what influences the most the value of τ, if the magnitude of the difference between the components on the variance, or the magnitude of the components themselves. Also notice that there are no overlaps/crossings within the correlation, which means that in all cases, the Kendall's τ is sensitive to the correlation in the innovation process.

5.3 Non-Gaussian innovation and heavy-tailed marginals

We also repeated the experiment of Section 4.4 and the simulation results are shown in Figures 12 to 14, which are the analogous of Figures 5 to 7, respectively. As expected, given the Kendall's τ nature, there is no relevant difference among the estimated values

[3] At 95% confidence level, the critical point of the two tailed test H_0 : *the components are independent* is 0.0009356. The test is obtained by the normal approximation to the Kendall's τ, which rejects H_0 if

$$|\hat{\tau}| > u_{\alpha/2}\sqrt{\frac{2(2n+5)}{9n(n-1)}},$$

where u_k stands for the $100 \times k$-percentile of the standard normal distribution. In this work, all hypothesis tests are performed at 95% significance level and $n = 2,000$.

of the Kendall's τ between the three cases studied (see Section 4.4 for details on the experiment).

6 Conclusions and Final Remarks

In this work we present an extensive empirical analysis on the dependence among the components of VARFIMA$(0, \boldsymbol{d}, 0)$ process through the Mallows distance point of view. We examine several cases, including Gaussian and non-Gaussian innovation processes, heavy-tailed marginals, equal and unequal marginal variances, among others. These marginal features are introduced in the process at the innovation's level, mostly by using copula tools. The goal is to investigate a possible relationship between the Mallows distance, the fractional differencing parameter $\boldsymbol{d}$, the type and dependence in the innovation process as well as its marginal behavior.

To estimate the Mallows distance we use an estimator based on the marginals' empirical quantiles. In Section 3, the estimator is shown to converge to the theoretical Mallows distance. Section 4 is dedicated to present the simulation results. It is divided in 4 subsections. In Subsection 4.1 we present the results for the case where the innovation process is Gaussian, for several combinations of $\boldsymbol{d}$ and ρ. In this standard case, our findings suggest that the Mallows distance is not generally sensitive to the correlation on the innovation, except in the presence of strong long range dependence. As expected, the Mallows distance decreases as the correlation increases.

In Subsections 4.2 and 4.3 we study the case where the innovation process is still Gaussian, but the marginal variances are different. We conclude that the different variances do not affect the Mallows distance behavior, but do affect its magnitude. We discover that the higher the difference between the variance components, the higher the magnitude of the Mallows distance. We find that equal variances produce smaller Mallows distance compared to unequal variances. The Mallows distance also responds to the magnitude of the variance components. The higher the magnitude, the more sensitive the Mallows distance becomes with respect to correlation in the innovation process in the presence of long range dependence. The overall Mallows distance behavior seems to be indifferent with respect to the marginal variances, except, as mentioned, for its magnitude. We also conclude that the fractional differencing parameter influence is attenuated by differences in the variances.

In Subsection 4.4 we attack the problem of non-Gaussianity in the innovation process by considering innovations with Frank copula distribution, but Gaussian marginals. We compared innovations with the same dependence strength, measured here by the Kendall's τ. We find that, as long as the dependence strength is kept at the same level, non-Gaussian innovations produce no change in the Mallows distance. We also investigate whether heavy-tailed marginals influence the Mallows distance at all. We discover that, except for a small change in the magnitude of the estimates, the Mallows distance behavior do not change under heavy-tailed innovations.

Are our results unique to the Mallows distance or are they shared by other dependence measures? To partially answer this question, we repeat all the simulations presented in Section 4 calculating the Kendall's τ instead of the Mallows distance. We find that the Kendall's τ, in clear opposition to the Mallows distance, is highly sensitive to the

innovations dependence in all experiments, but indifferent for most marginal changes we have applied. A bold exception is the unequal variances case, for which the Kendall's τ changes from its usual smooth behavior to an erratic one.

Overall we conclude that the Mallows distance is usually indifferent to changes in the innovation dependence, except in the presence of strong long range dependence and that the fractional differencing parameter plays an important role in determining the dependence structure of VARFIMA$(0, \boldsymbol{d}, 0)$ processes whatever the marginal or joint dependence considered is.

Acknowledgements

S.R.C. Lopes research was partially supported by CNPq-Brazil, by CAPES-Brazil, by INCT *em Matemática* and also by Pronex *Probabilidade e Processos Estocásticos* - E-26/170.008/2008 -APQ1. G. Pumi was partially supported by CAPES/Fulbright Grant BEX 2910/06-3 and by CNPq-Brazil. K. Zaniol was partially supported by CNPq-Brazil and by UFRGS. The authors are also grateful to the (Brazilian) National Center of Super Computing (CESUP-UFRGS) for the computational resources.

References

[1] Bickel, P. J. and Freedman, D. A. (1981). "Some Asymptotic Theory for the Bootstrap". *The Annals of Statistics*, **9**, 1196-1217.

[2] Chiriac, R. and Voev, V. (2010), "Modeling and Forecasting Multivariate Realized Volatility". *Journal of Applied Econometrics*. Forthcoming.

[3] Dionge, A.K. (2010). "A Multivariate Generalized Long Memory Model". *Comptes Rendus de l'Académie des Sciences - Série I.* **348**, 327-330.

[4] Johnson, O. and Samworth, R. (2005). "Central Limit Theorem and Convergence to Stable Laws in Mallows Distance". *Bernoulli*, **11(5)**, 829-845.

[5] Lobato, I.N. (1999). "A Semiparametric two Steps Estimator in a Multivariate Long Memory Model". *Journal of Econometrics*, **90**, 129-153.

[6] Lopes, S.R.C.; Sena Jr, M.R. and Reisen, V.A. (2006). "Correlated Errors in the Parameters Estimation of ARFIMA Model: A Simulated Study". *Communications in Statistics. Simulation and Computation*, **35(2)**, 789-802.

[7] Lopes, S.R.C. (2008). "Long-Range Dependence in Mean and Volatility: Models, Estimation and Forecasting". *Progress in Probability*, **60**, 497-525.

[8] Luceño, A. (1996). "A Fast Likelihood Approximation for Vector General Linear Processes with Long Series: Application to Fractional Differencing". *Biometrika*, **83**, 603-14.

[9] Mallows, C. L. (1972). "A Note on Asymptotic Joint Normality". *The Annals of Mathematical Statistics*, **43**, 508-515.

[10] Nelsen, R.B. (2006). *An Introduction to Copulas*. New York: Springer Verlag. 2nd edition.

[11] Samworth, R. and Johnson, O. (2004). "Convergence of the Empirical Process in Mallows Distance, with an Application to Bootstrap Performance". Unpublished Manuscript. Source: arXiv:math/0406603v1

[12] Schweizer, B. and Sklar, A. (2005). *Probabilistic Metric Spaces*. Mineola: Dover Publications.

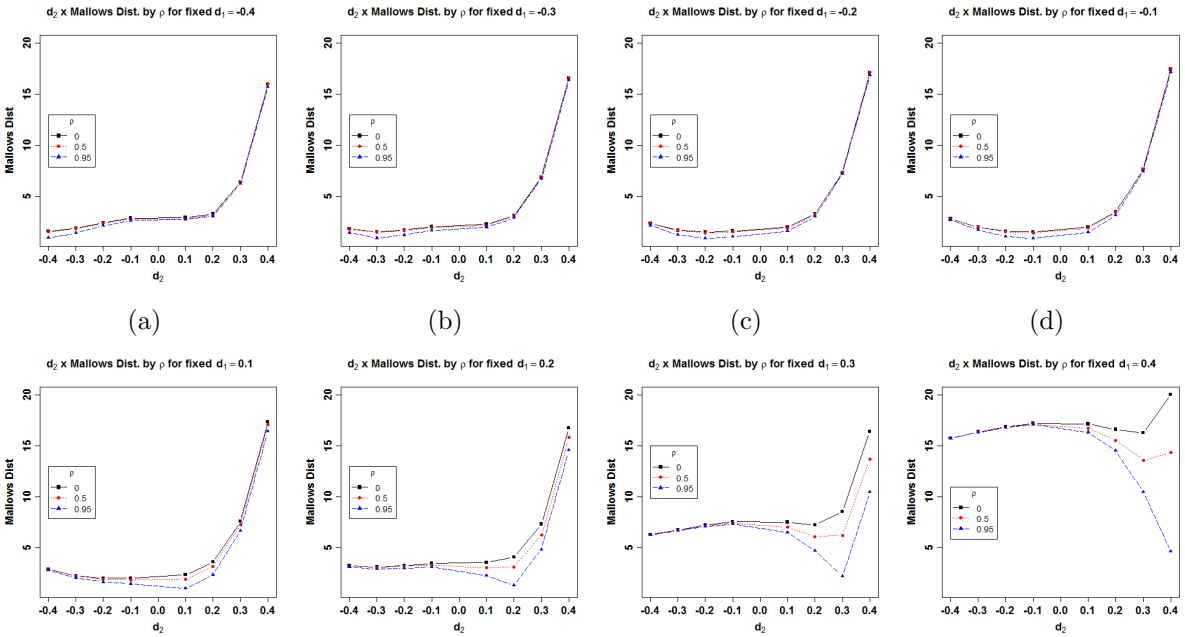
[13] Shimotsu, K. (2007). "Gaussian Semiparametric Estimation of Multivariate Fractionally Integrated Processes". *Journal of Econometrics*, **137**, 277-310.

[14] Shao, J. (2003). *Mathematical Statistics*. New York: Springer.

[15] Sowell, F. (1989). "Maximum Likelihood Estimation of Fractionally Integrated Time Series Models", Working Paper, Carnegie-Mellon University.

[16] Tsay, W.-J. (2010), "Maximum Likelihood Estimation of Stationary Multivariate ARFIMA Processes", *Journal of Statistical Computation and Simulation*, **80**, 729 - 745.

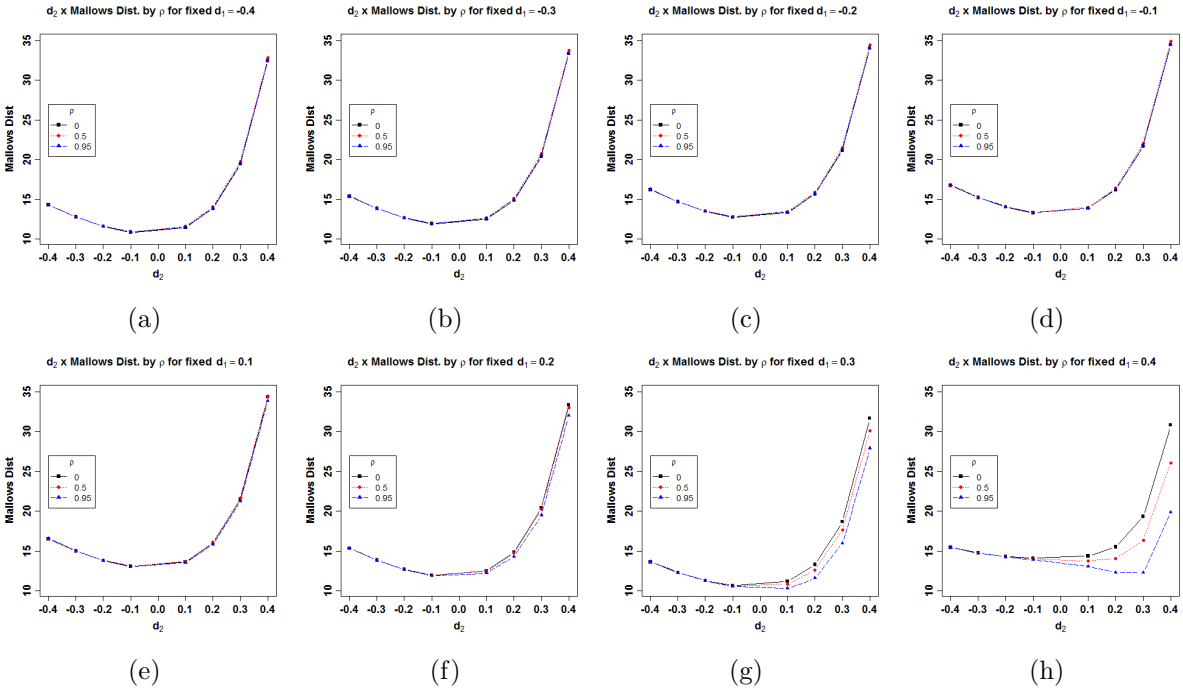
Appendix: Figures



(a) (b) (c) (d)

(e) (f) (g) (h)

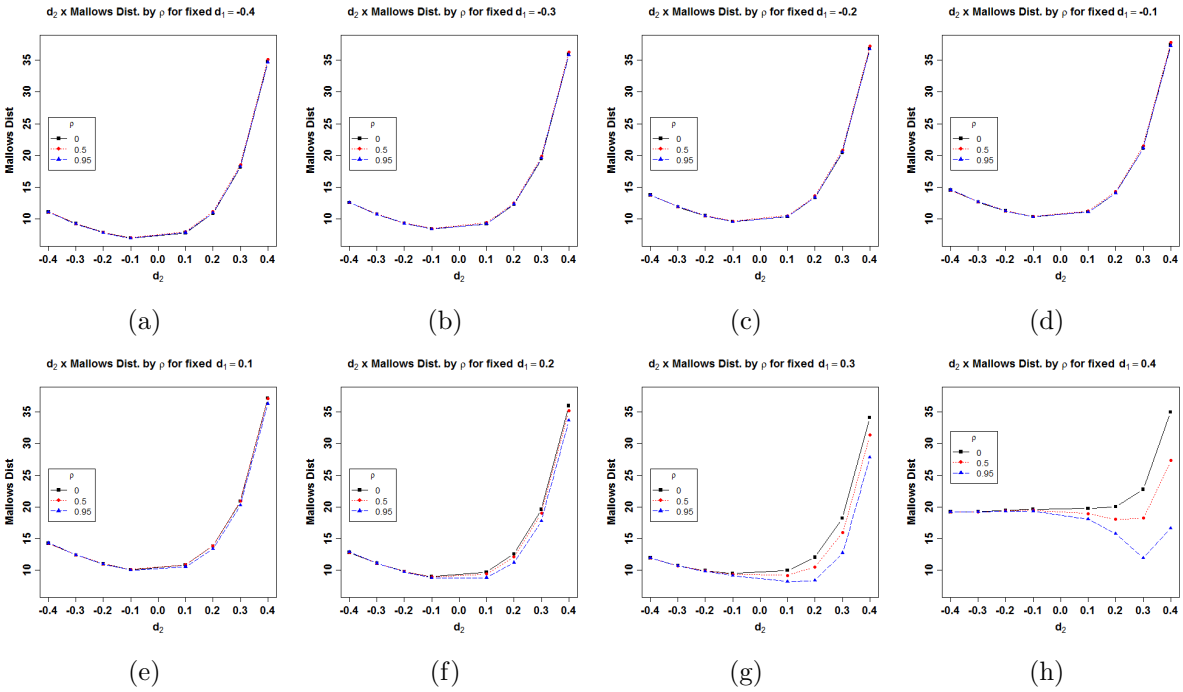
Figure 1: Plots of d_2 by Mallows distance for fixed d_1.



(a) (b) (c) (d)

(e) (f) (g) (h)

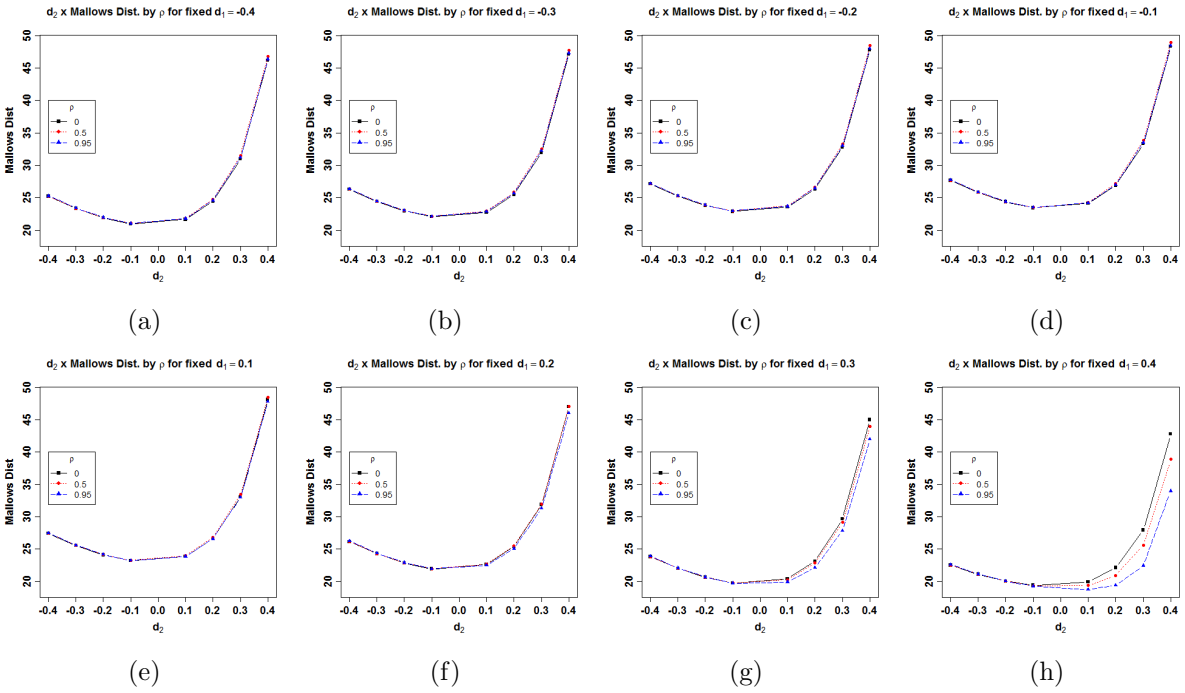
Figure 2: Plots of d_2 by Mallows distance for fixed d_1 and $\boldsymbol{\sigma}^2 = (1, 2)$.



(a) (b) (c) (d)

(e) (f) (g) (h)

Figure 3: Plots of d_2 by Mallows distance for fixed d_1 and $\boldsymbol{\sigma}^2 = (2, 3)$.



(a) (b) (c) (d)

(e) (f) (g) (h)

Figure 4: Plots of d_2 by Mallows distance for fixed d_1 and $\boldsymbol{\sigma}^2 = (1, 3)$.

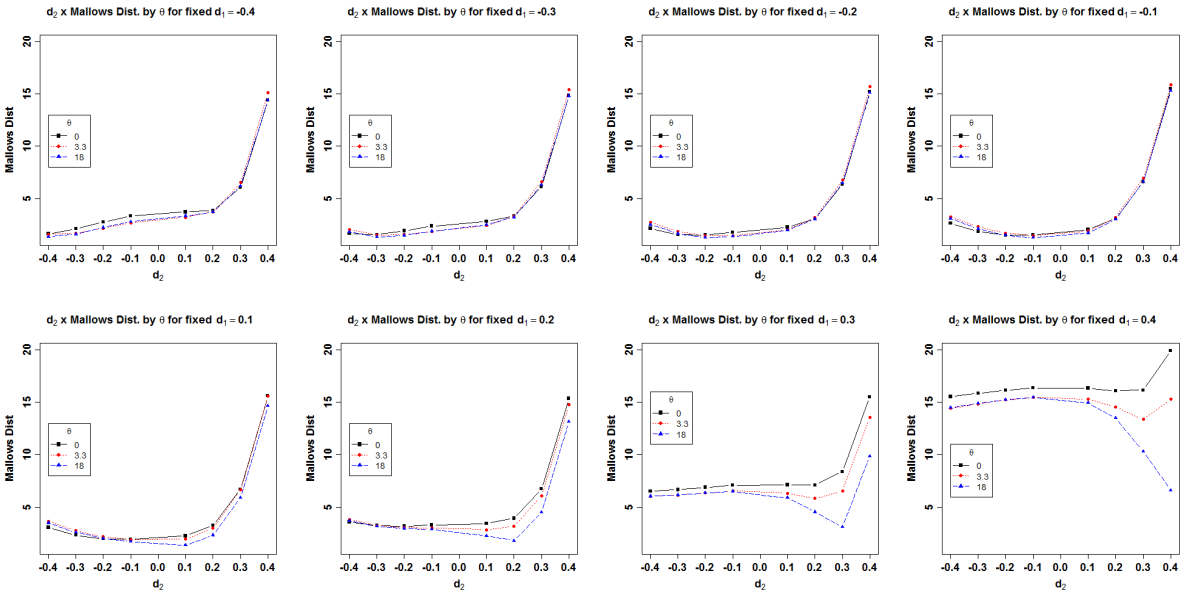


Figure 5: Plots of d_2 by Mallows distance for fixed d_1 and $\theta \in \{0, 3.3, 18\}$. Marginals were taken to be standard normal.

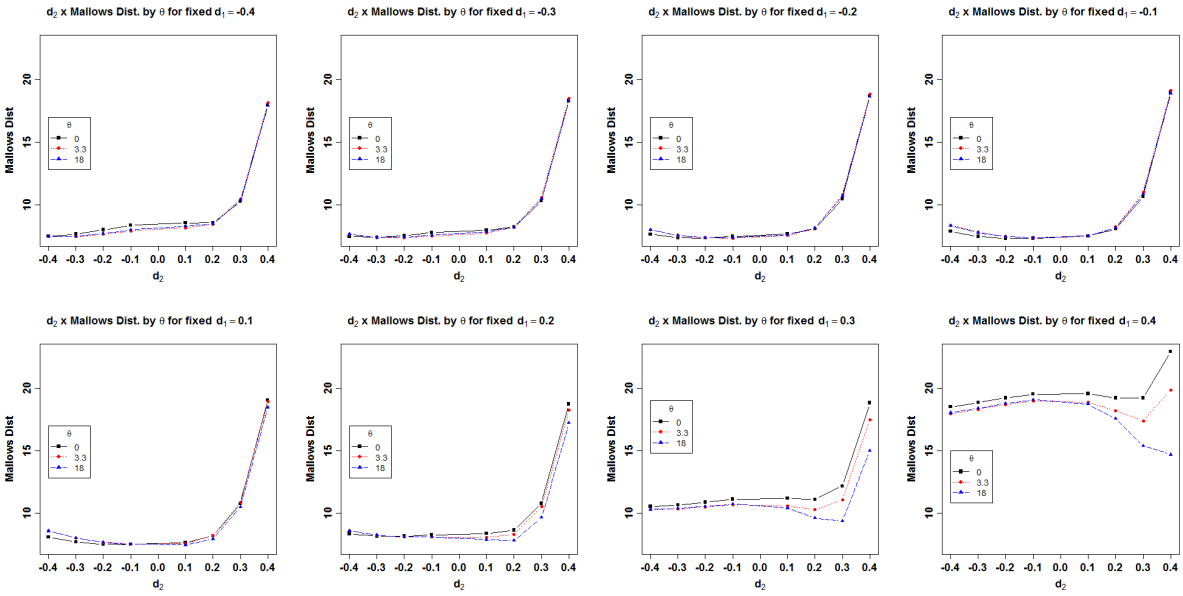


Figure 6: Plots of d_2 by Mallows distance for fixed d_1 and $\theta \in \{0, 3.3, 18\}$. Marginals were taken to be (standardized) t_3.

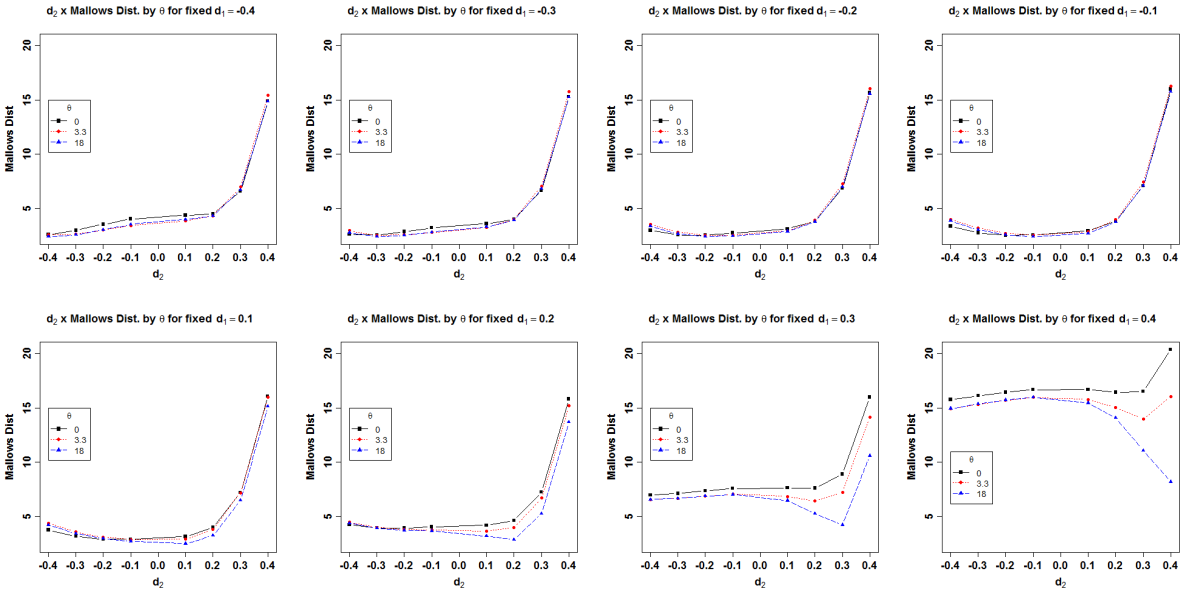


Figure 7: Plots of d_2 by Mallows distance for fixed d_1 and $\theta \in \{0, 3.3, 18\}$. Marginals were taken to be (standardized) t_7.

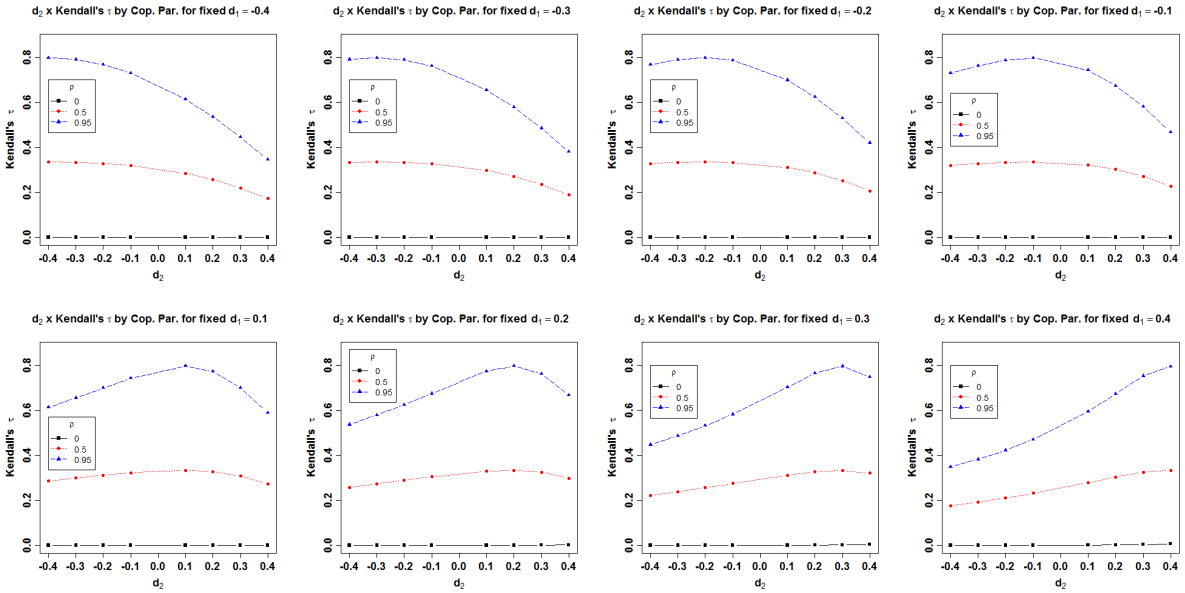


Figure 8: Plots of d_2 by Kendall's τ for fixed d_1.

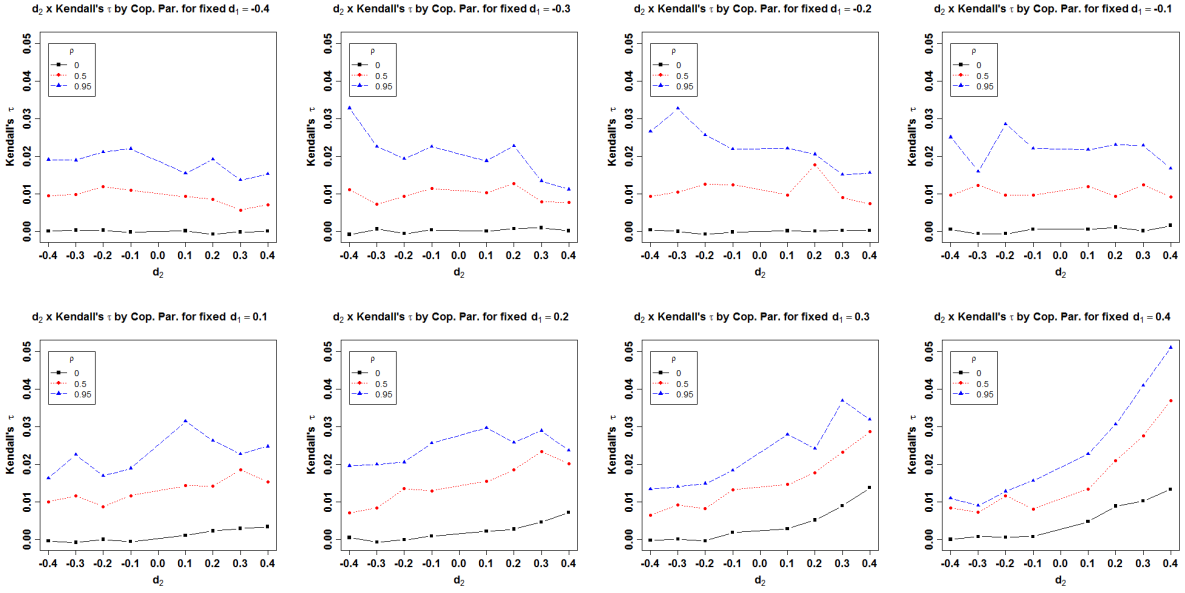


Figure 9: Plots of d_2 by Kendall's τ for fixed d_1 and $\boldsymbol{\sigma}^2 = (1, 2)$.

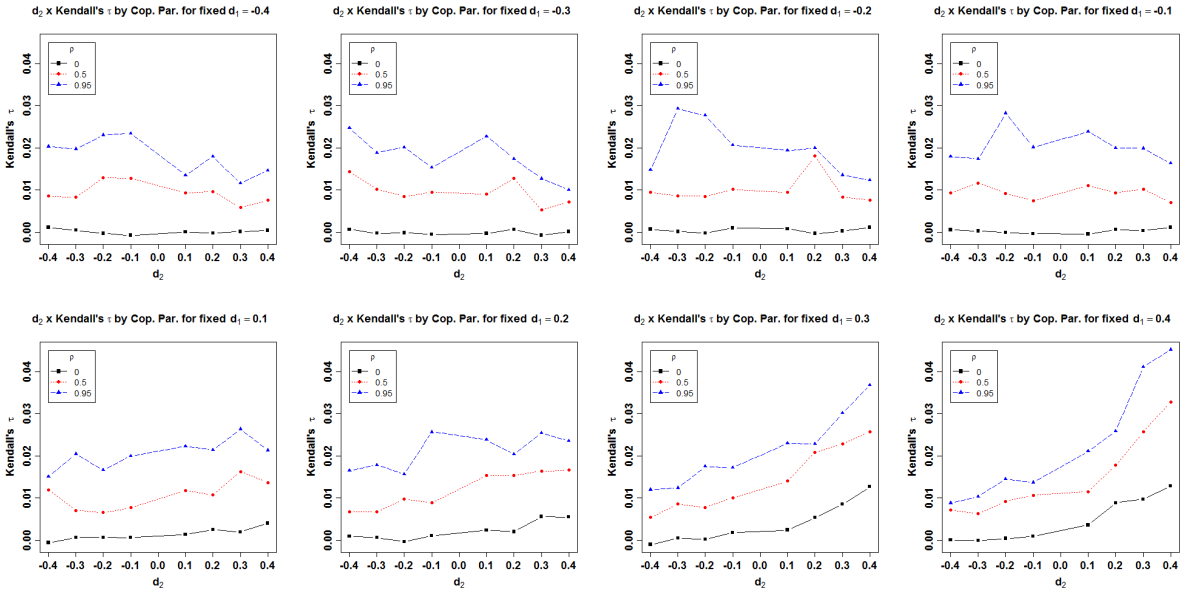


Figure 10: Plots of d_2 by Kendall's τ for fixed d_1 and $\boldsymbol{\sigma}^2 = (1, 3)$.

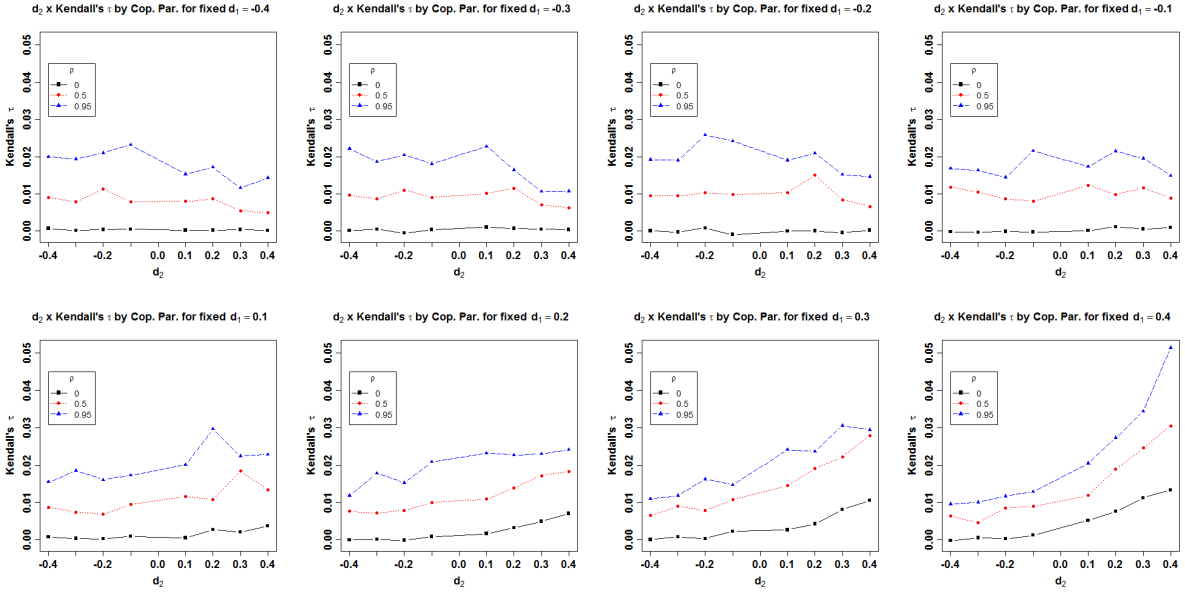


Figure 11: Plots of d_2 by Kendall's τ for fixed d_1 and $\boldsymbol{\sigma}^2 = (2, 3)$.

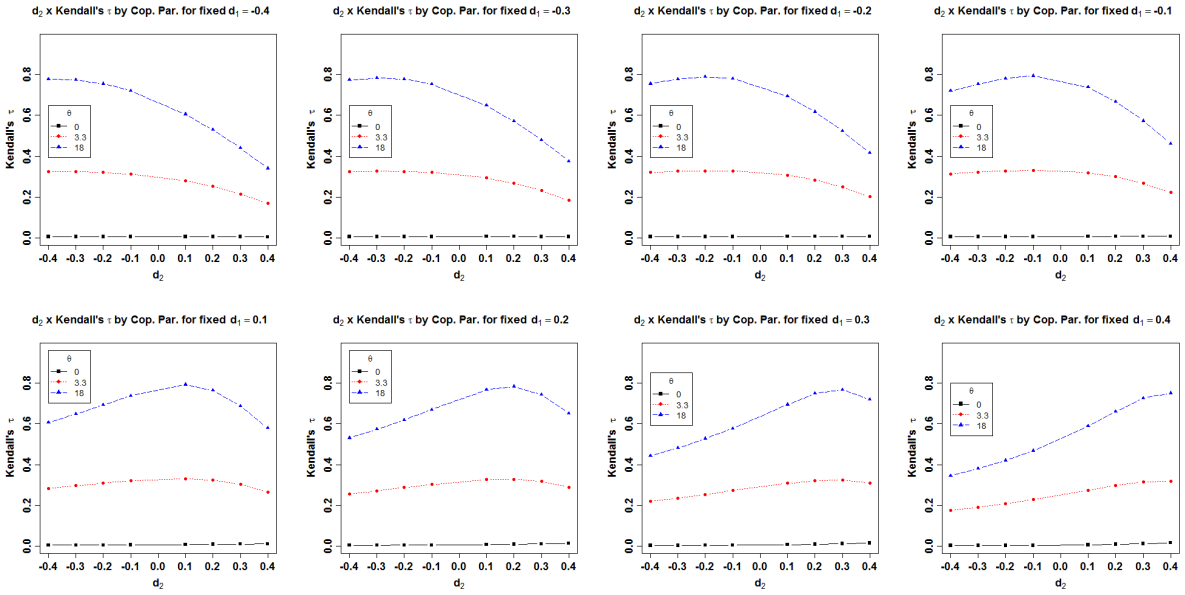


Figure 12: Plots of d_2 by Kendall's τ for fixed d_1 and $\theta \in \{0, 3.3, 18\}$. Marginals were taken to be standard normal.

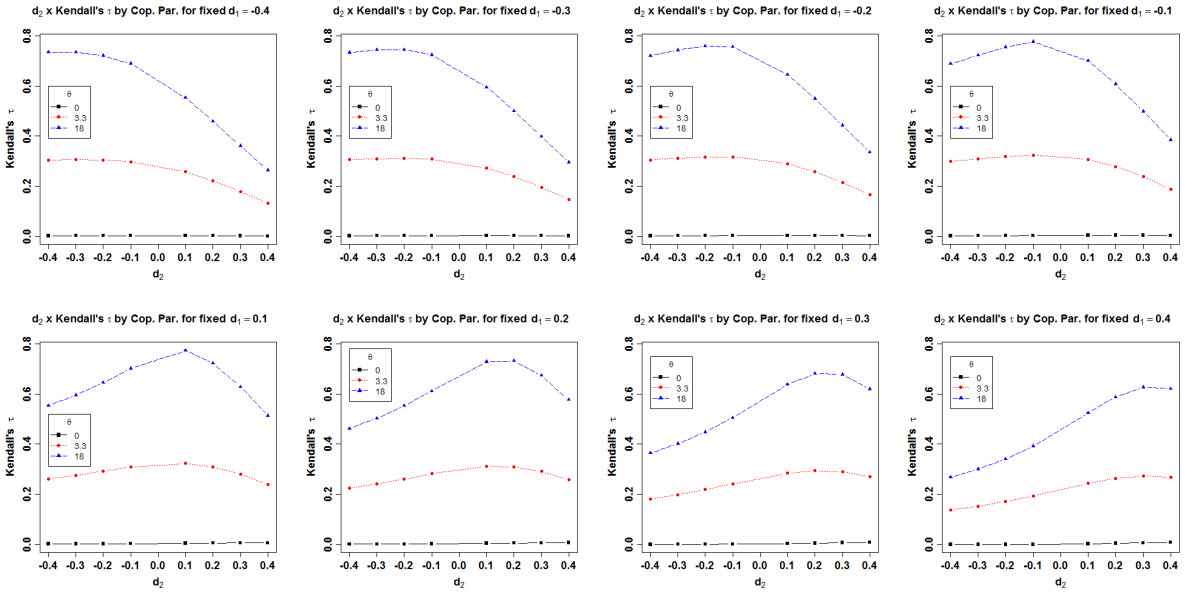


Figure 13: Plots of d_2 by Kendall's τ for fixed d_1 and $\theta \in \{0, 3.3, 18\}$. Marginals were taken to be (standardized) t_3.

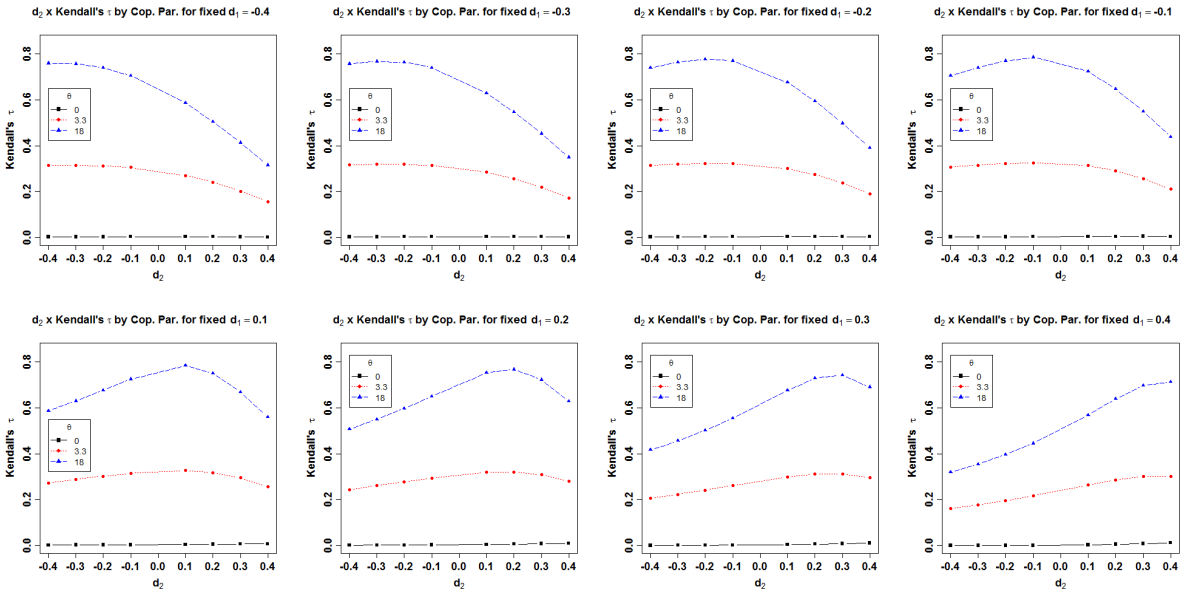


Figure 14: Plots of d_2 by Kendall's τ for fixed d_1 and $\theta \in \{0, 3.3, 18\}$. Marginals were taken to be (standardized) t_7.