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

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

September 14, 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. 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 is investigated. We study the behavior of the Kendall's τ dependence coefficient under the same framework for comparison purposes. For the Mallows distance, we consider an estimator based on the empirical marginal distribution function. Based on our simulation results, we propose both a semiparametric estimator for the fractional differencing parameter and a testing procedure to assess the presence of strong long range dependence in the components of VARFIMA processes of any (finite) dimension.

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

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

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 central limit theorem-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 Sena Junior, Reisen and Lopes, 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

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

of stock prices' volatility, modeling and forecasting high frequency data, among others. See, for instance, Chiriac and Voev (2011), 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 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. In the Gaussian innovation case, the degree of dependence is measured by the correlation coefficient. In the copula innovations case, the degree of dependence is measured 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 introduced in Section 3. It is based on the empirical quantile function of the process' marginals. 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.

Our simulation results are applied to construct a semiparametric estimator for the fractional differencing parameter of Gaussian VARFIMA$(0, \boldsymbol{d}, 0)$ process of any (finite) dimension, including the univariate case. The main advantage of the estimator is to be fast while providing good estimates. The estimator is especially useful in very high dimensional estimation problems. In those cases, most estimators of the fractional differencing parameter depend on maximization procedures which require an initial guess for the value of $\boldsymbol{d}$. Usually a poor choice of the initial point will lead to slow convergence of the procedure. Given that the estimator we propose is easy to implement, fast to calculate and usually provide reasonable estimates, it is a good candidate for the initial guess necessary in the mentioned maximization procedures. In high dimension, this simple procedure may provide a significant reduction in computational time. We assess the estimator's performance through simple Monte Carlo experiments. The obtained simulations results are further applied to construct an approximate test for the presence of strong long range dependence in the components of Gaussian VARFIMA$(0, \boldsymbol{d}, 0)$ process of any (finite) dimension, including the univariate case. The performance of the test is also assessed by simple Monte Carlo experiments.

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 Mallows distance estimator considered in the simulation studies. 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 considering 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. As an application of the estimation results, we present in Section 6 the construction of a semiparametric estimator for the fractional differencing parameter as well as a simple test to assess the presence of strong long range dependence in the components of VARFIMA processes. Conclusions and final remarks are reserved to Section 7.

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 $\alpha \geq 1$, $\mathscr{D}_\alpha(\cdot,\cdot)$ given by (2.1) can be shown to be equivalent to a much simpler expression (cf. Major, 1978) 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 (see Rachev and Rüschendorf, 1998 and Major, 1978):

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

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 last expression in (2.3) will be useful in defining the estimator introduced on Section 3.

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).

Definition 2.2. 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 class of the so-called VARFIMA$(p, \boldsymbol{d}, q)$ processes.

Definition 2.3. 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

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

assumed to have no common roots, where $\boldsymbol{\phi}_1, \cdots, \boldsymbol{\phi}_p, \boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_q$ 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 2.3 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

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

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 $\det\big(\boldsymbol{\Theta}(z)\big) \neq 0$ and $\det\big(\boldsymbol{\Phi}(z)\big) \neq 0$, for all $z \in \mathbb{C}$ such that $|z| \leq 1$, and also $\boldsymbol{d} \in (-0.5, 0.5)^m$, so that $(-0.5, 0.5)^m$ will be the range we shall assume for the fractional differencing 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 Sena Junior, Reisen and Lopes, 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 $k \in \{1, \cdots, m\}$. In this work, the case where $d > 0.3$ is given special attention.

Definition 2.4. Let $\{X_t\}_{t\in\mathbb{N}}$ be an ARFIMA(p, d, q) process. We say that X_t is *strong long range dependent*, or present *strong long range dependence* (SLRD for short), if $d > 0.3$.

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 (2011), 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,*

$$H(x_1, \cdots, x_n) = C\big(F_1(x_1), \cdots, F_n(x_n)\big), \quad \textit{for all } (x_1, \cdots, x_n) \in \mathbb{R}^n. \tag{2.5}$$

If F_i's are continuous, then C is unique. Otherwise, C is uniquely determined on $\mathrm{Ran}(F_1) \times \cdots \times \mathrm{Ran}(F_n)$*, where, for a function f,* $\mathrm{Ran}(f)$ *denotes the range of f. 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 in the bidimensional case can be found in Nelsen (2006).

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. Letting $C_{\boldsymbol{X}}$ denote the copula associated to $\boldsymbol{X}$ (that is, a function satisfying (2.5)), if the copula $C_{\boldsymbol{X}}$ is known, 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 (determined 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).

3 An Estimator for the Mallows Distance

We proceed to define the Mallows distance estimator considered in the simulations, prove its strong consistency and derive its limiting distribution. Given two i.i.d. samples $X_1, \cdots, X_n$ and $Y_1, \cdots, Y_m$ from distributions F and G, respectively, let $\widehat{F}_n$ and $\widehat{G}_m$ denote the empirical distribution functions based on these samples. For any $\alpha \geq 1$, the Mallows α-distance estimator is given by

$$\widehat{\mathscr{D}}_\alpha(F, G) := \mathscr{D}_\alpha(\widehat{F}_n, \widehat{G}_m) = \left(\int_0^1 \left| \widehat{F}_n^{(-1)}(u) - \widehat{G}_m^{(-1)}(u) \right|^\alpha \mathrm{d}u \right)^{1/\alpha}. \tag{3.1}$$

If F and G are absolutely continuous distributions and $m = n$, the expression (3.1) takes a very simple form. Let $X_{(1)}, \cdots, X_{(n)}$ and $Y_{(1)}, \cdots, Y_{(n)}$ be the ordered samples. Since repetition in the sample occurs with probability 0, $\widehat{F}_n^{-1}(t) = X_{(i)}$, for all $t \in \left[\frac{i-1}{n}, \frac{i}{n}\right)$ and similarly for $\widehat{G}_n^{-1}(t)$. Therefore, in this case,

$$\widehat{\mathscr{D}}_\alpha(F, G) = \left(\frac{1}{n} \sum_{i=1}^{n} |X_{(i)} - Y_{(i)}|^\alpha \right)^{1/\alpha}. \tag{3.2}$$

In the case where F and/or G is not absolutely continuous or the samples do not have the same length, expression (3.1) will become a summation depending on the jump points of both functions. In this case the estimated Mallows α-distance can be obtained by calculating (3.1) directly upon implementation of the functions $\widehat{F}_n^{(-1)}$ and $\widehat{G}_n^{(-1)}$ to derive the step function $s(u) = \left| \widehat{F}_n^{(-1)}(u) - \widehat{G}_n^{(-1)}(u) \right|^\alpha$. We then calculate the area under $s(u)$ in $[0, 1]$ and take it to the $1/\alpha$ power to obtain the estimate. Most statistical softwares already have a built in empirical quantile function (such as the function `quantile` on the software R), which facilitates the implementation of the estimator.

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 (3.1).

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.

4.1 Gaussian innovations with equal variances

Figure A.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 A.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 A.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 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 whenever the latter is greater or equal than -0.2. We shall explore this fact in Subsection 6.1 to construct a simple and fast semiparametric estimator for the fractional differencing parameter $\boldsymbol{d}$.

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 when one coordinate of $\boldsymbol{d}$ is greater or equal than 0.1 and behaves close to 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 0.1.

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

[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.

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 A.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 A.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 A.1 and A.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 A.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 A.3 and A.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 A.1 (equal variances case) and A.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 considerably 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 A.2, A.3 and A.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 A.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 (see Nelsen, 2006 for its definition and properties) 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 estimate the 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 A.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 A.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, especially between $\rho = 0$ and $\rho \in \{3.3, 18\}$. 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 A.5 shows the simulation results for the Frank-Normal couple. Figure A.6, the Frank-t_3 couple results are shown. In Figure A.7, the case Frank-t_7 is presented. Comparing Figures A.5 and A.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 $[0.003, 0.051]$, 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 A.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 basically 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 marginals' 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.

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 are two processes, 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 A.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 A.1 in the Mallows distance case. Notice that the Kendall's τ is 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.

Although it is difficult to see in Figure A.8 given the scale, 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 exactly the values of d_i for which $d_1 + d_2 \geq 0.6$, that is, when there is SLRD. This result is interesting since it reinforces the findings of Tsay and Chung (2000), where the authors study the presence of spurious regression in stationary and ergodic processes with long range dependence. In the aforementioned paper, the authors detected spurious effects in long range dependent time series whenever $d_1 + d_2 > 0.5$. In the present case, we observed that whenever $d_1 + d_2 > 0.6$, even though the components of the process are independent of each other by construction, the null hypothesis of independence is rejected, which indicates the presence of spurious effects. For more details on the matter, see Tsay and Chung (2000).

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 A.8 to A.11 present the graphs of d_2 by the Kendall's

[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.

[3]At 95% confidence level, the critical point of the two sided 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 $n = 2,000$.

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 for 3 (out of 48) combinations the hypothesis of independence is rejected when at least one parameter in $\boldsymbol{d}$ is negative. This suggests that differences in the process' variances amplify the spurious effect present in the equal variances case and give evidence of spurious regression in a broader context than reported in Tsay and Chung (2000). 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 A.12 to A.14, which are the analogous of Figures A.5 to A.7, respectively. As expected, given the Kendall's τ nature, there is no relevant difference among the estimated values of the Kendall's τ between the three cases studied (see Section 4.4 for details on the experiment).

6 Applications

In this section we shall present two simple applications of the results obtained in the previous sections. In the first one, we use the empirical results obtained in Section 4 to propose a semiparametric estimator for the fractional differencing parameter $\boldsymbol{d}$ for VARFIMA processes of any finite dimension. We present a simple Monte Carlo experiment to assess the estimator's performance. In the second application we propose a Mallows distance based test to assess the presence of SLRD components in VARFIMA process of any finite dimension. We also perform a Monte Carlo experiment to analyze its performance.

6.1 Estimation in Gaussian ARFIMA$(0, \boldsymbol{d}, 0)$ processes

In this section we shall present a simple application of the results obtained in Section 4. The idea is based on the observation that the results of Subsection 4.1 (see Figure A.1) show that in the context of Gaussian VARFIMA$(0, \boldsymbol{d}, 0)$ process with independent components, when one coordinate of the fractional differencing vector $\boldsymbol{d} = (d_1, d_2)$ is negative, say $d_1 < 0$, the Mallows distance behaves like an increasing function of d_2, whenever $d_2 > 0$. We shall explore this result to develop a simple and fast way to estimate the fractional differencing parameter in Gaussian (univariate) ARFIMA$(0, d, 0)$ processes.

Let us illustrate the idea by considering Figure A.1. Suppose that $\{X_t\}_{t=1}^n$ is a Gaussian ARFIMA$(0, d, 0)$ process with $d \in (0, 1/2)$. Our goal is to estimate the fractional differencing parameter d. The method starts by choosing d_1, among the possibilities in Figure A.1, and fitting some one-to-one curve, say $h(\cdot)$, to the portion of the graph for which $d_2 > 0$ and $\rho = 0$. For example, consider $d_1 = -0.3$ corresponding to Figure A.1(b). In that graph, the Mallows distance behaves closely to an exponential function for $d_2 > 0$ and $\rho = 0$. To illustrate the procedure we fit three different curves to the points in the positive portion of Figure A.1(c), for $\rho = 0$, as shown in Figure 6.1. The parametric form of the curves are the following

$$h_1(t; \boldsymbol{a}) := a_1 t \exp\big(a_2 t^2 + a_3 t + a_4\big), \quad h_2(t, \boldsymbol{b}) := b_1(t - 0.1)^3 + b_2(t - 0.1) + b_3,$$

for $\boldsymbol{a} := (a_1, a_2, a_3, a_4)'$ and $\boldsymbol{b} := (b_1, b_2, b_3)'$. The third curve h_3 is defined as the piecewise linear function joining the points in the graphs and extrapolated below and above by the nearest fitted line. The parameters were chosen to minimize the squared error between the observed and fitted values. In Figure 6.1 we present the fitted curves for the data from Figure A.1(b), where the parameters are

$$\boldsymbol{a} = (0.0862, 22.1222, -9.0890, 2.8109)' \quad \text{and} \quad \boldsymbol{b} = (15.8494, 0.0417, 0.0789)'.$$

The next step is to simulate an auxiliary Gaussian ARFIMA$(0, d_1, 0)$ process, say $\{Y_t^{(1)}\}_{t=1}^n$, independent of X_t, and calculate the estimated Mallows distance between the two processes, say $\widehat{\mathscr{D}}_{(1)}$. Notice that it is advantageous to simulate the auxiliary process with the same length as the one fixed in order to apply the fast computable expression (3.2). The process is repeated m times to obtain a sequence $\{\widehat{\mathscr{D}}_{(k)}\}_{k=1}^m$ from which we define the estimator $\widehat{d}$ of d by

$$\widehat{d} := h^{-1}\big(\overline{\mathscr{D}}\big), \quad \text{where} \quad \overline{\mathscr{D}} := \frac{1}{m}\sum_{k=1}^{m} \widehat{\mathscr{D}}_{(k)}. \tag{6.1}$$

To measure the performance of the method, we carry out a simple experiment whose results are presented in Table 6.1. For each $d_2 \in \{0.10, 0.15, \cdots, 0.45\}$ in Table 6.1, we simulate a single Gaussian ARFIMA$(0, d, 0)$, say $\{X_t\}_{t=1}^n$, for $n = 2,000$, with the objective of estimating d. Next, we simulate $m = 1,000$ auxiliary Gaussian ARFIMA$(0, -0.3, 0)$ processes with the same length n as X_t. We then calculate the Mallows distance between X_t and each auxiliary series to obtain the sequence $\{\widehat{\mathscr{D}}_{(k)}\}_{k=1}^{1000}$. From $\{\widehat{\mathscr{D}}_{(k)}\}_{k=1}^{1000}$ the estimator is calculated by (6.1) with h_1, h_2 and h_3 in the role of h. The estimated values of d, along with its bias (in parenthesis) and the standard deviation of the sequence $\{\widehat{\mathscr{D}}_{(k)}\}_{k=1}^{1000}$

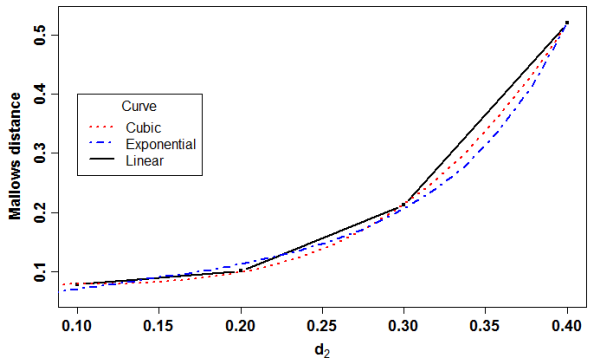


Figure 6.1: Fitted exponential, cubic and piecewise linear functions to the positive part of Figure A.1(b) for $\rho = 0$.

(at the bottom), are presented in Table 6.1. Notice that the inverse of $h_1(\cdot)$ has no closed form, so the value $h_1^{-1}(x)$ has to be numerically calculated. The inverse of h_2 can be explicitly derived after some tedious calculations, and is given by

$$h_2^{-1}(y) := \frac{\ell_0(y)^{1/3}}{6a_1} - \frac{2a_2}{\left(a_1^2\ell_0(y)\right)^{1/3}} + 0.1,$$

where

$$\ell_0(y) := 108(y - a_3) + 12\left(\frac{12a_2^3}{a_1} + \left[9(a_3 - y)\right]^2\right)^{\frac{1}{2}}.$$

Given the rather small sample to which the curves were fitted, the estimated values

Table 6.1: Estimation results for $\widehat{d}$, based on the approximations shown in Figure 6.1. Presented are the estimated values along with the bias (in parenthesis).

Method	d_2							
	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45
Exponential	0.0762 (-0.0238)	0.1453 (-0.0047)	0.1765 (-0.0235)	0.2793 (0.0293)	0.3198 (0.0198)	0.3512 (0.0012)	0.4020 (0.0020)	0.4486 (-0.014)
Cubic	0.0067 (-0.0933)	0.1749 (0.0249)	0.2044 (0.0044)	0.2792 (0.0292)	0.3128 (0.0128)	0.3430 (-0.0070)	0.4030 (0.0030)	0.4756 (0.0256)
Linear	0.0655 (-0.0345)	0.1353 (-0.0147)	0.1684 (-0.0316)	0.2518 (0.0018)	0.3011 (0.0011)	0.3256 (-0.0244)	0.3953 (-0.0047)	0.5252 (0.0752)
Std. Dev.	0.0234	0.0398	0.0598	0.1124	0.1638	0.2158	0.5755	0.6471

presented in Table 6.1 are relatively good, although some high bias can be seen for $d_2 = 0.1$, especially for the cubic fit (due to a root very close to 0). The estimators tend to perform better for d_2 corresponding to values for which data is available, as expected. In extrapolating to $d_2 = 0.45$, the estimates obtained from h_1 and h_2 are close to the true value, while for h_3, as one could expect given the exponential trend on the values, the estimative is off. Overall, the exponential curve (h_1) yield better results, even though,

visually, it appears that h_2 provides a better fit for the data. At first sight, the standard deviation seems high compared to the respective estimated value, especially when d is high. However, since the estimates are skewed to the right, higher values of standard deviation are expected, especially compared to normally distributed estimators.

On one hand, for m small the method provide reasonable approximation and is extremely fast. Basically, the time spent for computing the estimator is dominated by the time spent in simulating the auxiliary processes. On the other hand, the procedure yield more precise estimative for m large. The limitation of the method, as seen in Table 6.1, are basically due to three sources: firstly we rely on approximated curves from few empirically calculated points to obtain the estimates, which may not be close enough to the true values to allow a good fitting; secondly, the results are based on independent pairs of process, while in the procedure, one of the coordinates is keep fixed and the other varies. Therefore, in doing so, some level of bias is expected, which is reflected on the estimated Mallows distance values and ultimately repassed to the final estimative; thirdly, the shape of the data near 0.1 is almost flat, causing high bias, but small variability in the estimates observed in Table 6.1. In contrast, near 0.4 the data increase rapidly, causing high variability but smaller bias on the estimates for large d. Generally, the method would certainly benefit from more empirically calculated values for the Mallows distance.

The method can be useful to calculate a first step for a more complicated estimation method, or to determine a "good" initial guess for estimators based on maximization procedures, especially if the objective function is non-smooth with several local maximums. The method is even more useful in the multivariate case, especially for high dimensional process. Estimators like the one proposed by Shimotsu (2007), are based on multidimensional optimization problems dependent on an initial guess for the parameter of differentiation. In that case, one could apply the method componentwise to obtain a reasonable initial guess for the optimization procedure. Furthermore, each auxiliary process can be used to calculate the estimated Mallows distance with respect to all components. For high dimensional problems, this could considerably reduce the usually time consuming optimization procedure, especially if no initial guess is available and a random one has to be used.

6.2 Hypothesis test

In this section we shall develop a test based on the results of Section 4.1 to test the presence of a certain level of dependence in multidimensional Gaussian VARFIMA$(0, \boldsymbol{d}, 0)$ processes. We start with the bidimensional case $\boldsymbol{d} = (d_1, d_2)$. The goal is to construct a test to detect the presence of a SLRD component in bivariate Gaussian VARFIMA$(0, \boldsymbol{d}, 0)$ process.

The idea of the test is as follows. Let $\boldsymbol{\theta} := (d_1, d_2, \rho)$ and let $\mathscr{D}_2(\boldsymbol{\theta})$ denote the theoretical Mallows distance between the components of a Gaussian VARFIMA$(0, \boldsymbol{d}, 0)$ process whose innovations are independent and identically distributed $\mathcal{N}(\mathbf{0}, \Omega)$ with variance-covariance matrix given by $\Omega := \left(\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\right)$. Let VARFIMA$(0, \boldsymbol{d}, 0; \rho)$ denote such a process. Observe that in calculating each estimate presented on Figure A.1, we obtained, as a by-product of the replications, independent sequences of estimates from the theoretical unknown Mallows distance associated to each given Gaussian VARFIMA$(0, \boldsymbol{d}, 0; \rho)$. Each

of these sequences, denoted by $\left\{\widehat{\mathscr{D}}_2^k(\boldsymbol{\theta})\right\}_{k=1}^n$, can be regarded as an independent and identically distributed approximated sample from $\mathscr{D}_2(\boldsymbol{\theta})$. We shall explore this fact to construct an approximated test of level α for assessing the presence of a SLRD component in the process.

Suppose we have a VARFIMA$(0, \boldsymbol{d}, 0; \rho)$ process $\{\boldsymbol{X}_t\}_{t=0}^n$, with unknown $d_1 \neq d_2$ and $\rho \geq 0$, and we would like to test the presence of a SLRD component in the process. From $\boldsymbol{X}_t$ we can calculate an estimate for the Mallows distance between its components, say $\widehat{\mathscr{D}}_2(\boldsymbol{\theta})$, by using either (3.1) or (3.2). Our goal is to obtain a suitable test statistic and critical values for the test

$$H_0\text{: } \boldsymbol{X}_t \text{ has no SLRD component} \quad \text{vs} \quad H_1\text{: at least one component of } \boldsymbol{X}_t \text{ is SLRD.} \tag{6.2}$$

The problem of testing (6.2) is important in contexts where estimation near the non-stationarity zone $d \geq 0.5$ cannot be performed with precision, such as in spectral density based estimators (see Lopes, 2008). We start by analyzing the following related test

$$\begin{aligned} &H_0' : \mathscr{D}_2(\boldsymbol{\theta}) < \mathscr{D}_2(\boldsymbol{\theta}_0) \quad \text{vs} \quad H_1' : \mathscr{D}_2(\boldsymbol{\theta}) > \mathscr{D}_2(\boldsymbol{\theta}_0). \\ &\text{Rule: reject } H_0' \text{ for large values of } \widehat{\mathscr{D}}_2(\boldsymbol{\theta}). \end{aligned} \tag{6.3}$$

Testing (6.2) is the same as simultaneously test (6.3) for all combinations of $\boldsymbol{\theta}_0 := (d_1^0, d_2^0, \rho^0)$ for which at least one $d_i^0 > 0.3$. To avoid multiple testing procedures, we observe that from Figure A.1, the estimated Mallows distance is strictly increasing and consistently assumes high values whenever $\boldsymbol{d}$ contains at least one component greater or equal than 0.3 (except when $d_1 = d_2$). Consider the particular case of (6.3), obtained by taking

$$\mathscr{D}_2(\boldsymbol{\theta}_0) = \mathscr{D}_2(\boldsymbol{\theta}^*),$$

where $\boldsymbol{\theta}^*$ is the triple obtained from the combinations of $\boldsymbol{d}$ and ρ presented in Figure A.1 (and from which the sequences $\left\{\widehat{\mathscr{D}}_2^k(\boldsymbol{\theta})\right\}_{k=1}^n$ are available) such that at least one $d_i^* > 0.3$ and the value $\mathscr{D}_2(\boldsymbol{\theta}^*)$ is minimum among all possible candidates in Figure A.1. The reasoning behind such a choice is as follows: if there is no sufficient statistical evidence to reject H_0' with this choice, this means that $\widehat{\mathscr{D}}_2(\boldsymbol{\theta})$ is too small to be considered a member of the alternative class, relatively to $\mathscr{D}_2(\boldsymbol{\theta}^*)$. But if that is the case, clearly it won't be large enough for any other possible candidate in Figure A.1, hence $\mathscr{D}_2(\boldsymbol{\theta}^*)$ minimizes the type I error uniformly among all possible choices, provided $d_1 \neq d_2$. Of course, this protectiveness against type I error will eventually lead to an increase on the type II error. In the present case, considering all possible choices in Figure A.1, the best one in the sense just explained is $\boldsymbol{\theta}^* = (0.3, 0.2, 0.95)$, corresponding to Figure A.1(g).

Returning to the initial problem of testing (6.2), we notice that the uniformity of the choice $\mathscr{D}_2(\boldsymbol{\theta}^*)$ allows one to consider the test statistic $T(\boldsymbol{X}) := \widehat{\mathscr{D}}_2(\boldsymbol{\theta})$ and the rule

$$\text{Reject } H_0 \text{ if } \widehat{\mathscr{D}}_2(\boldsymbol{\theta}) > \widehat{q}_\alpha^{\,0}(\boldsymbol{\theta}^*), \tag{6.4}$$

where $\widehat{q}_\alpha^{\,0}(\boldsymbol{\theta}^*)$ is the $100 \times \alpha\%$-percentile based on the empirical distribution function of $\left\{\widehat{\mathscr{D}}_2^k(\boldsymbol{\theta}^*)\right\}_{k=1}^{1000}$. Taking $\alpha = 0.05$ in (6.4), it is routine to show that the test based on $\widehat{q}_{0.05}^{\,0}(\boldsymbol{\theta}^*) = \widehat{\mathscr{D}}_2^{(951)}(\boldsymbol{\theta}^*)$, where $\widehat{\mathscr{D}}_2^{(951)}(\boldsymbol{\theta}^*)$ is the 951-th order statistic of $\left\{\widehat{\mathscr{D}}_2^k(\boldsymbol{\theta}^*)\right\}_{k=1}^{1000}$, is of level 0.05 (see, for instance, Bickel and Doksum, 2007, p.271). In our situation[4], $\widehat{q}_{0.05}^{\,0}(\boldsymbol{\theta}^*) \approx 0.2751$.

[4]the simulated data used throughout this paper is available upon request.

To assess the performance of the test (6.2) with rule (6.4) for $\alpha = 0.05$, we generate 500 replications of a VARFIMA$(0, \boldsymbol{d}, 0; \rho)$ process of sample size 2,000 for eight combinations of (d_1, d_2) and $\rho \in \{0, 0.5, 0.95\}$. In Table 6.2 we present the rejection rate (in percentage) of the test (6.2) with rule (6.4). The first four columns of Table 6.2 correspond to values of $\boldsymbol{d}$ for which H_0 holds, so that we expect the rejection rate to be small, while in the last four columns the null hypothesis does not hold and a large rejection rate is expected. Half of the values of $\boldsymbol{d}$ in the table are borderline (identified by *) and were included to assess the test performance under very weak evidence of either hypothesis, which usually induce type I and type II errors.

Table 6.2: Hypothesis testing results based on 500 replications of a VARFIMA$(0, \boldsymbol{d}, 0; \rho)$ process. Presented is the rejection ratio in testing (6.2) by using rule (6.4) with $\boldsymbol{\theta}^* = (0.3, 0.2, 0.95)$.

ρ	$(-0.2, 0.2)$	$(-0.1, 0.2)$	$(-0.4, 0.25)^*$	$(-0.3, 0.25)^*$	$(0.1, 0.45)$	$(0.4, -0.3)$	$(0.2, 0.35)^*$	$(0.3, -0.1)^*$
0	0.4%	0.8%	7.2%	6.4%	99.8%	80.0%	53.4%	28.2%
0.5	1.0%	0.8%	9.8%	9.2%	99.6%	82.2%	47.4%	32.0%
0.95	0.4%	0.8%	5.8%	8.2%	100.0%	81.4%	41.4%	28.2%

* Borderline cases

From the results in Table 6.2, we notice that the test's overall performance is good when the values of $\boldsymbol{d}$ are not the borderline ones. The test performs well in terms of type I error even in borderline cases such as $(-0.4, 0.25)$ and $(-0.3, 0.25)$. As expected, the performance of the test in terms of type II error deteriorate in the borderline case $(0.2, 0.35)$ and is especially poor in the extreme case $\boldsymbol{d} = (0.3, -0.1)$. The test results clearly reflect the particular choice of $\mathscr{D}_2(\boldsymbol{\theta}_0)$, which was chosen to be conservative in terms of type I error, but permissive in terms of type II error.

Now, suppose we have $\{X_t\}_{t=1}^n$ a Gaussian ARFIMA$(0, d, 0)$ process for which we would like to test $H_0' : d < 0.3$ versus $H_1' : d > 0.3$. In order to do that, we reduce the problem to the previous case by considering the following construction. Choose $d_0 \in (-0.5, 0.1)$ and simulate a Gaussian ARFIMA$(0, d_0, 0)$ process $\{Y_t\}_{t=1}^n$ of same length as X_t. By considering $\boldsymbol{d} = (d, d_0)$, $\{(\boldsymbol{X}_t, \boldsymbol{Y}_t)\}_{t=1}^n$ is a VARFIMA$(0, \boldsymbol{d}, 0; 0)$, for which the Mallows distance estimate $\widehat{\mathscr{D}}_2(\boldsymbol{\theta})$, $\boldsymbol{\theta} = (d, d_0, 0)$, can be calculated. Upon applying the test (6.2) with rule (6.4), we obtain the conclusion. Obviously, if H_0 in (6.2) is refuted, we conclude $H_1' : d > 0.3$. One can replicate Y_t to obtain more information on the rejection rate and improve the conclusion. Also, as long as the sample size is not too small, even for $n \neq 1,000$ the test should yield valid conclusions. The same reasoning covers the case of Gaussian VARFIMA$(0, \boldsymbol{d}, 0)$ processes of dimension higher than two by pairwise testing or by applying the previous idea componentwise.

We end up this section by observing that the results in this section were based on Subsection 4.1, but, naturally, the ideas could be easily adapted to any case presented in Section 4.

7 Conclusions and Final Remarks

In this work we present an extensive empirical analysis on the dependence among the components of VARFIMA$(0, \boldsymbol{d}, 0)$ processes 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, in Section 3 we propose an estimator based on the marginals' empirical quantiles. 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 the 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. Also worth noticing is the evidence of spurious regression when the process' components present long range dependence. This finding reinforces (and even extend) the conclusions on the paper by Tsay and Chung (2000), where the authors study the existence of spurious effects in stationary ergodic processes presenting strong long range dependence.

In Section 6 we present some applications of our simulation results. We propose a

Mallows distance based semiparametric estimator for the fractional differencing parameter on Gaussian VARFIMA$(0, \boldsymbol{d}, 0)$ process of any finite dimension. Its usefulness lies especially in its simplicity and fast computability, which makes it an attractive first step estimator for high dimensional problems. Its performance is assessed by a simple Monte Carlo experiment. We also propose a test to identify the presence of strong long range dependence in the components of Gaussian VARFIMA$(0, \boldsymbol{d}, 0)$ process of any finite dimension based solely on our simulation results. Its properties are assessed by simple Monte Carlo experiments which shows that the test performance is reasonably good, especially in terms of type I error.

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, to the editor and an anonymous referee whose comments lead to considerable improvement of this paper.

References

[1] Bickel, P.J. and Doksum, K.A. (2007). *Mathematical Statistics: Basic Ideas and Selected Topics*. New Jersey: Prentice Hall. 2nd edition.

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

[3] Chiriac, R. and Voev, V. (2011), "Modeling and Forecasting Multivariate Realized Volatility". *Journal of Applied Econometrics*. Forthcoming.

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

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

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

[8] Major, P. (1978). "On the Invariance Principle for Sums of Independent Identically Distributed Random Variables". *Journal of Multivariate Analisys*. **8**(4), 487-517

[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] Rachev, S.T. and Rüschendorf, L. (1998). *Mass transportation problems. Vol. I. Theory*. New York: Springer-Verlag.

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

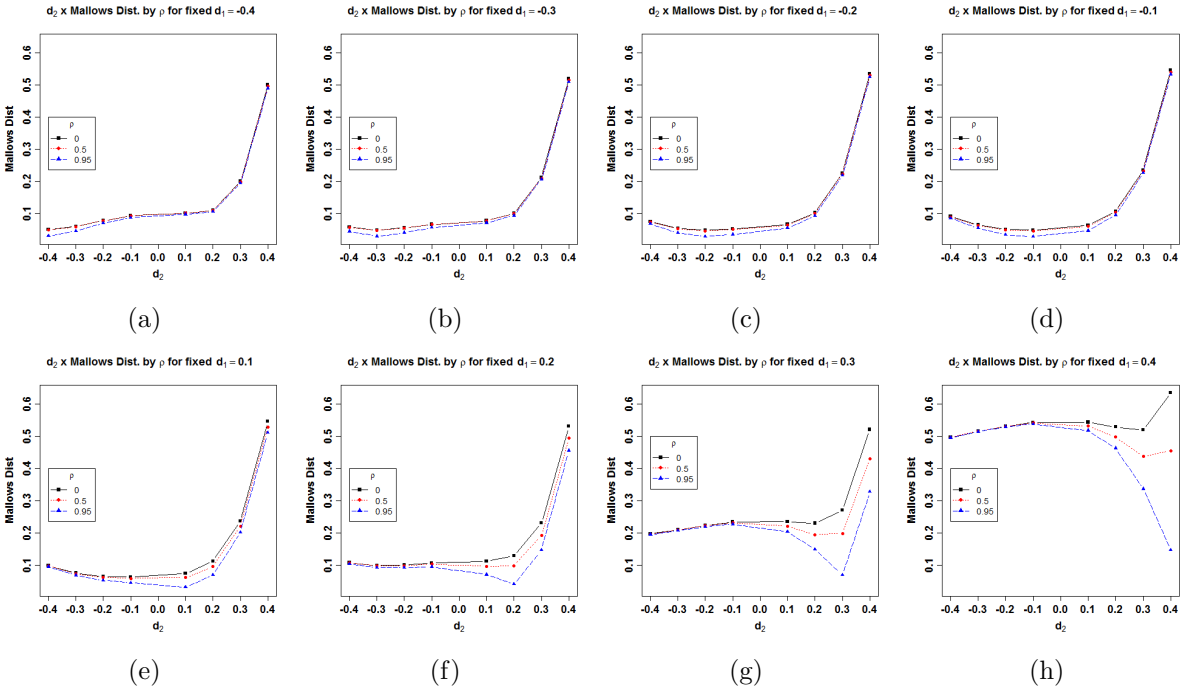
[13] Shimotsu, K. (2007). "Gaussian Semiparametric Estimation of Multivariate Fractionally Integrated Processes". *Journal of Econometrics*, **137**, 277-310.

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

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

[16] Tsay, W.-J. and Chung, C.-F. (2000). "The spurious regression of fractionally integrated processes". *Journal of Econometrics*, **96**(1), 155-182.

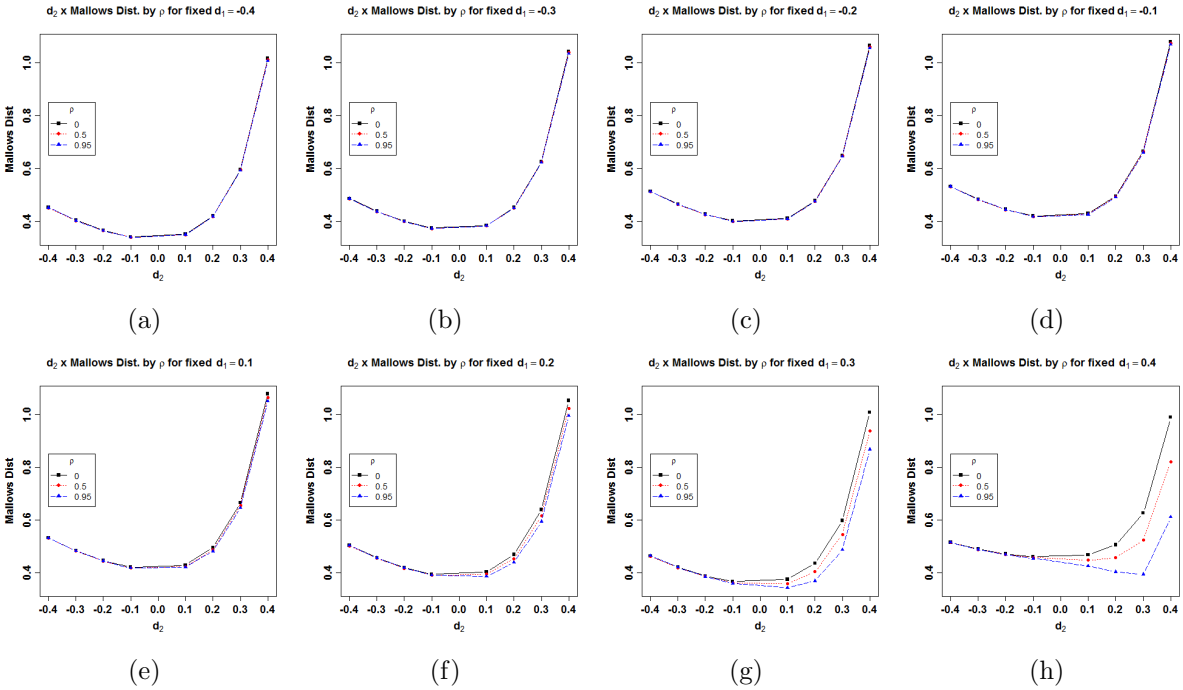
Appendix: Figures



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

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

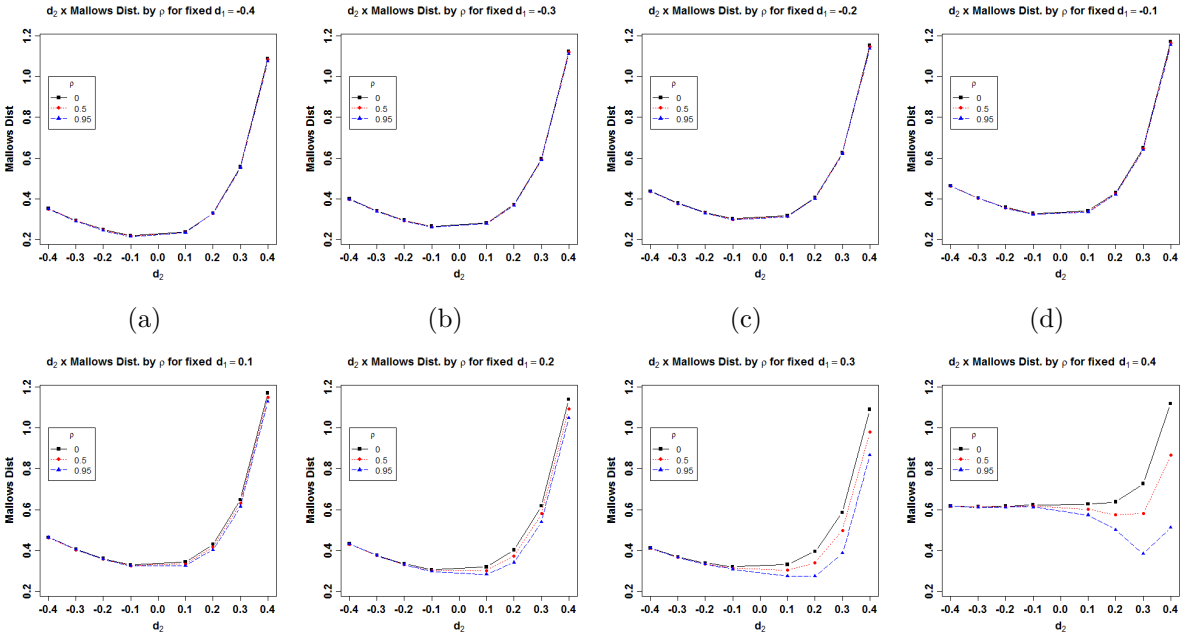
Figure A.1: Plots of d_2 by Mallows distance for fixed d_1.



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

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

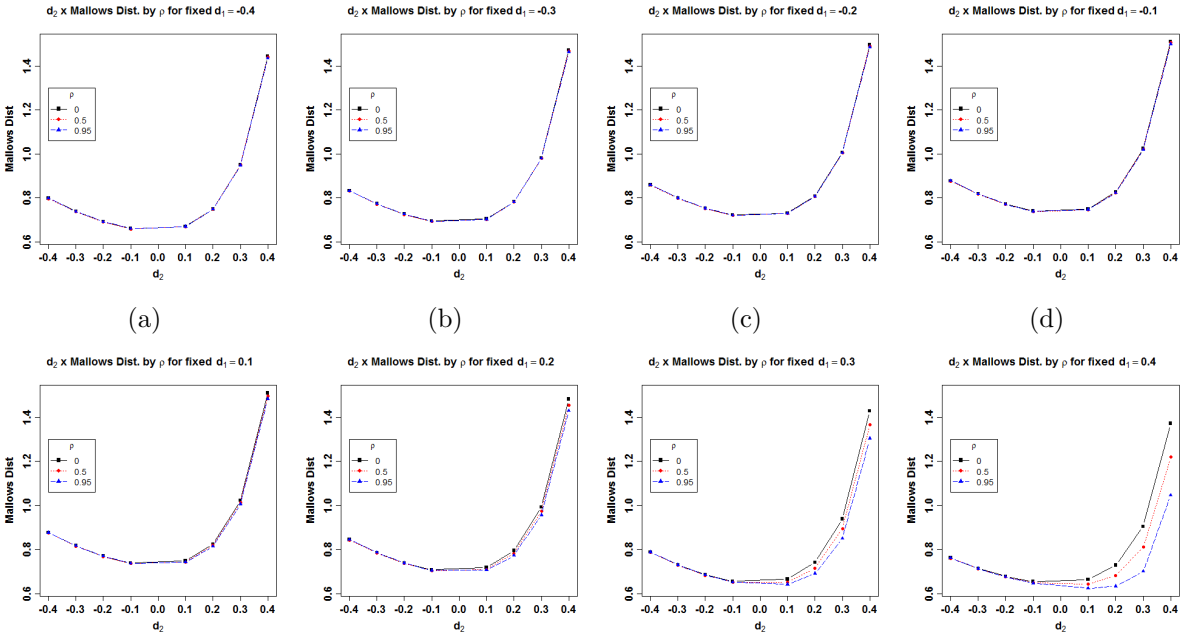
Figure A.2: Plots of d_2 by Mallows distance for fixed d_1 and $\boldsymbol{\sigma}^2 = (1, 2)$.



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

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

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

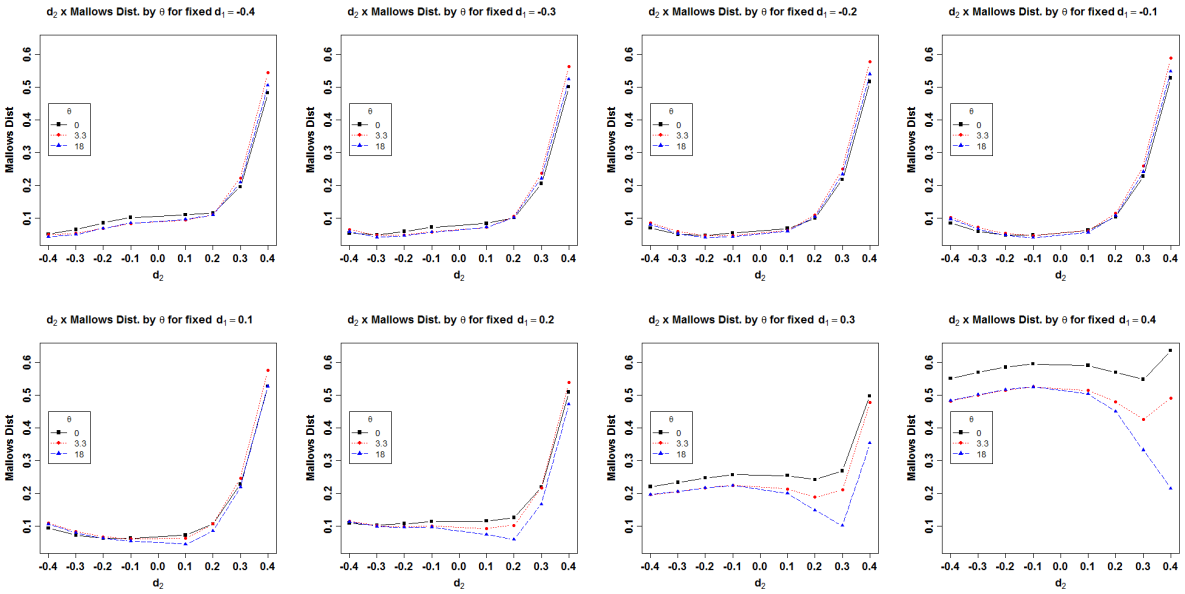


Figure A.5: Plots of d_2 by Mallows distance for fixed d_1 and $\theta \in \{0, 3.3, 18\}$. Marginals were taken to be standard normal.

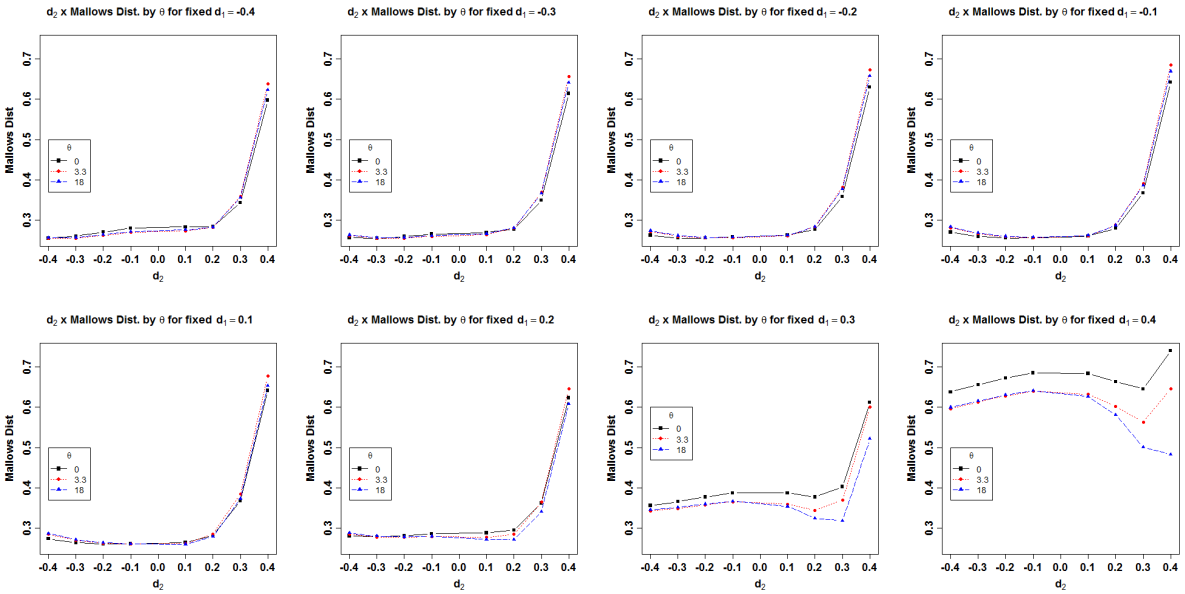


Figure A.6: Plots of d_2 by Mallows distance for fixed d_1 and $\theta \in \{0, 3.3, 18\}$. Marginals were taken to be (standardized) t_3.

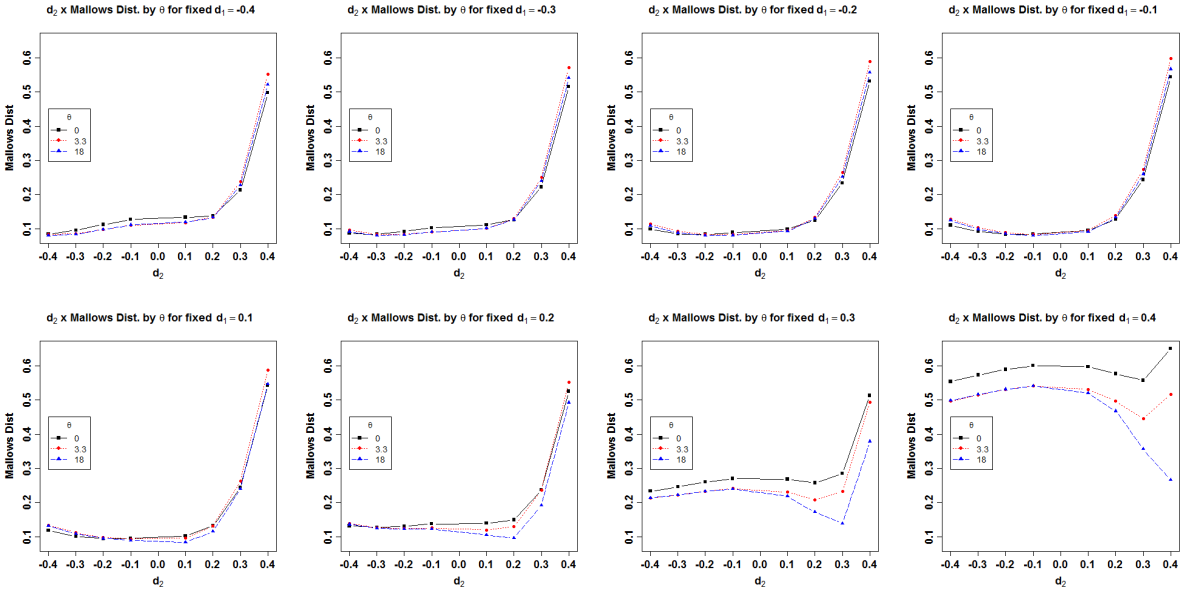


Figure A.7: Plots of d_2 by Mallows distance for fixed d_1 and $\theta \in \{0, 3.3, 18\}$. Marginals were taken to be (standardized) t_7.

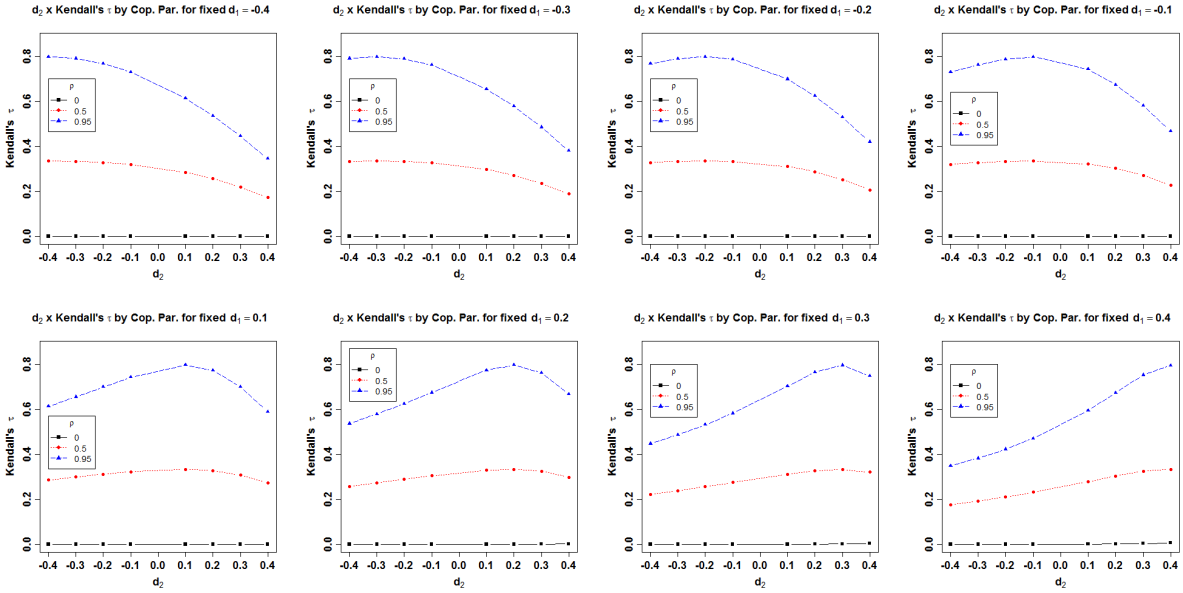


Figure A.8: Plots of d_2 by Kendall's τ for fixed d_1.

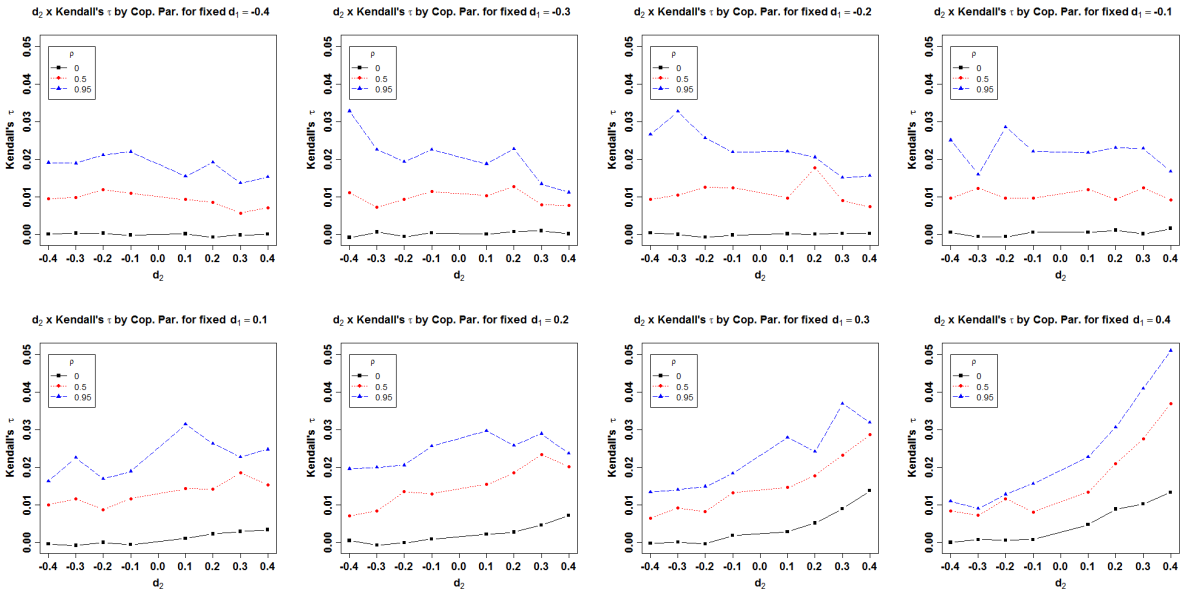


Figure A.9: Plots of d_2 by Kendall's τ for fixed d_1 and $\boldsymbol{\sigma}^2 = (1, 2)$.

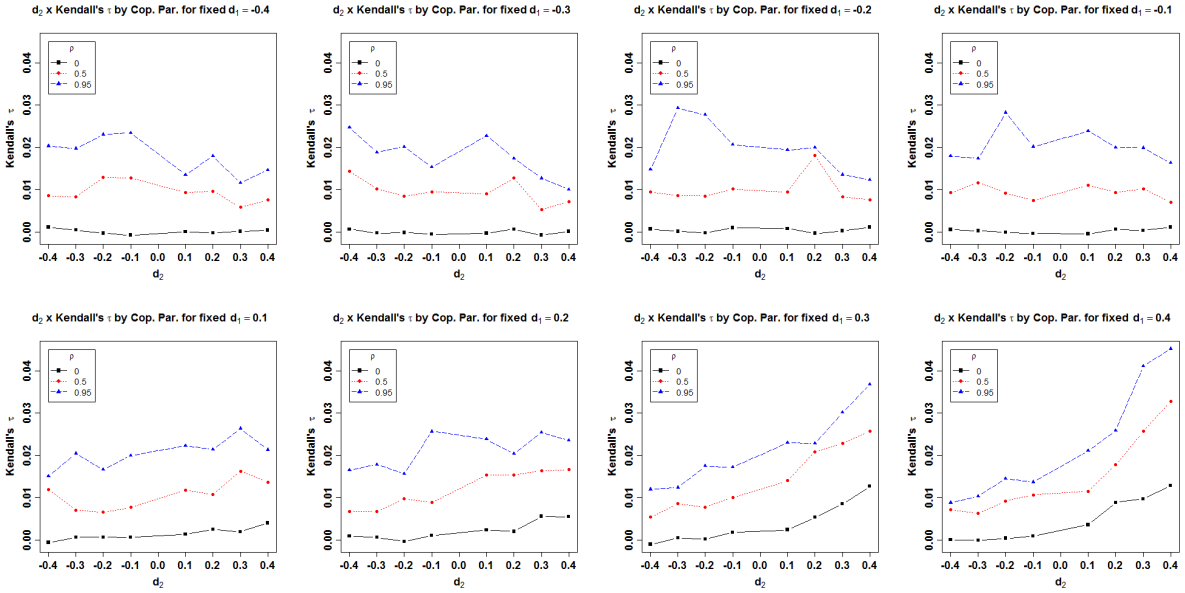


Figure A.10: Plots of d_2 by Kendall's τ for fixed d_1 and $\boldsymbol{\sigma}^2 = (1, 3)$.

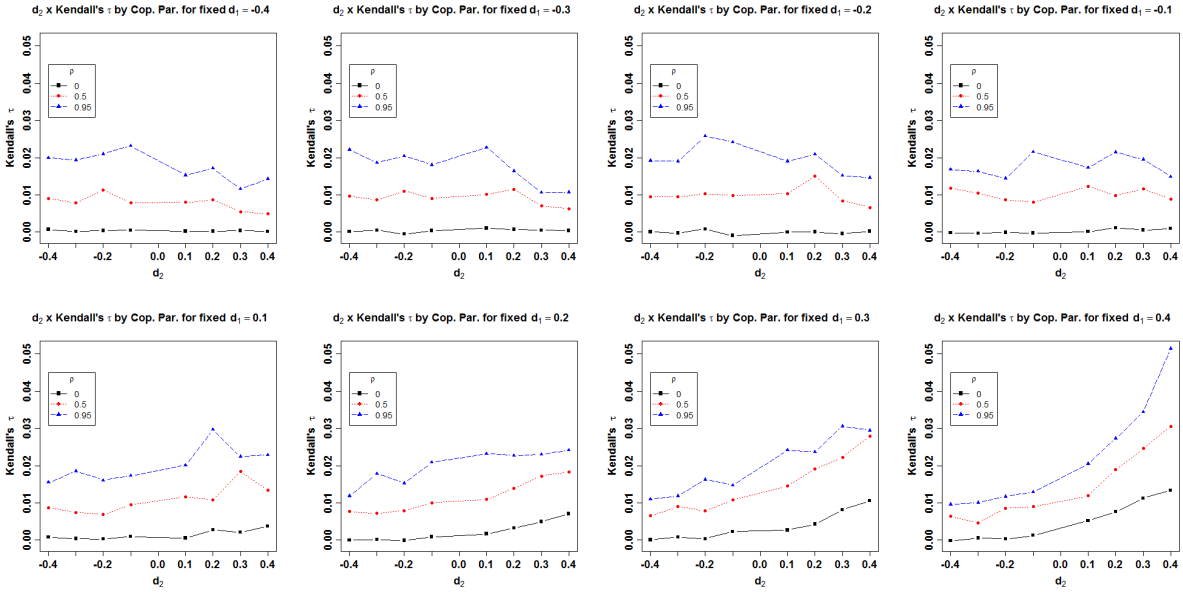


Figure A.11: Plots of d_2 by Kendall's τ for fixed d_1 and $\boldsymbol{\sigma}^2 = (2, 3)$.

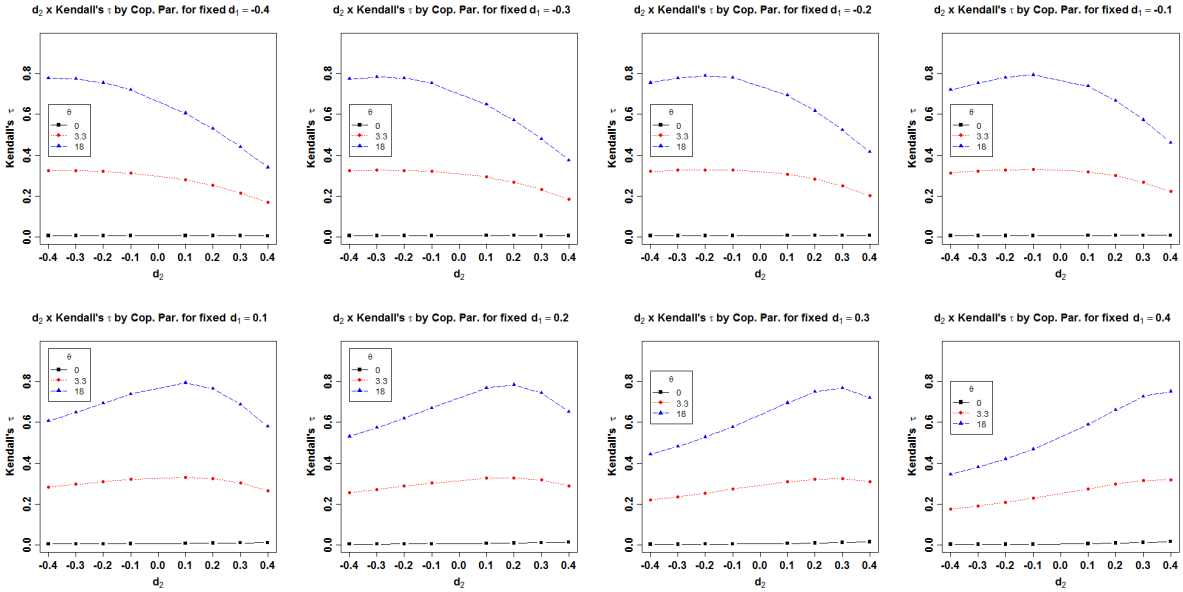


Figure A.12: Plots of d_2 by Kendall's τ for fixed d_1 and $\theta \in \{0, 3.3, 18\}$. Marginals were taken to be standard normal.

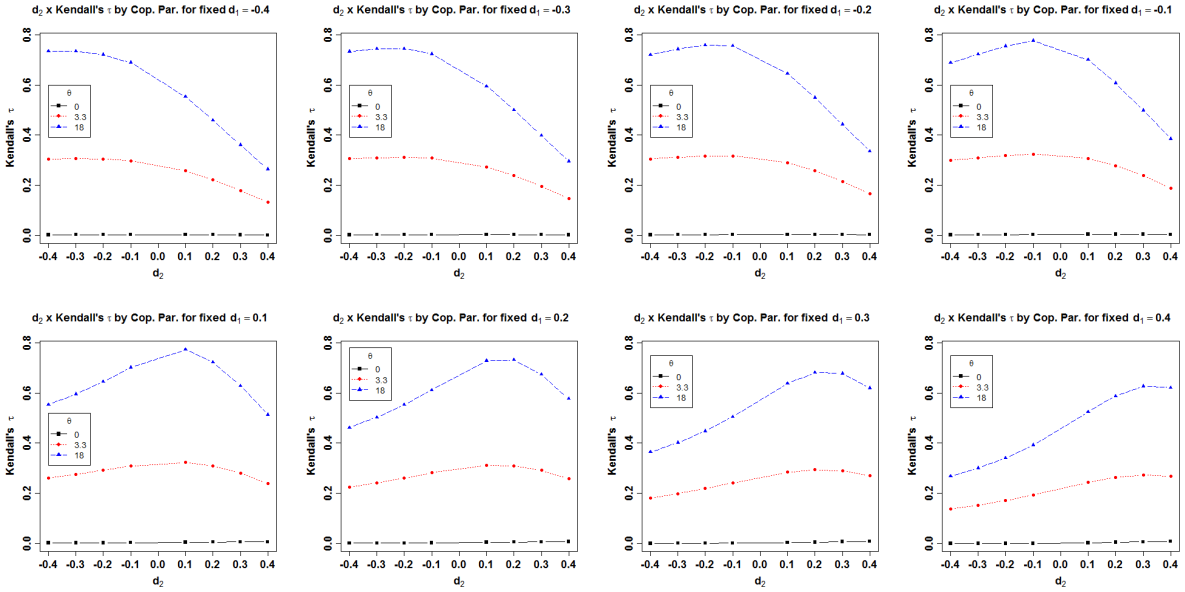


Figure A.13: Plots of d_2 by Kendall's τ for fixed d_1 and $\theta \in \{0, 3.3, 18\}$. Marginals were taken to be (standardized) t_3.

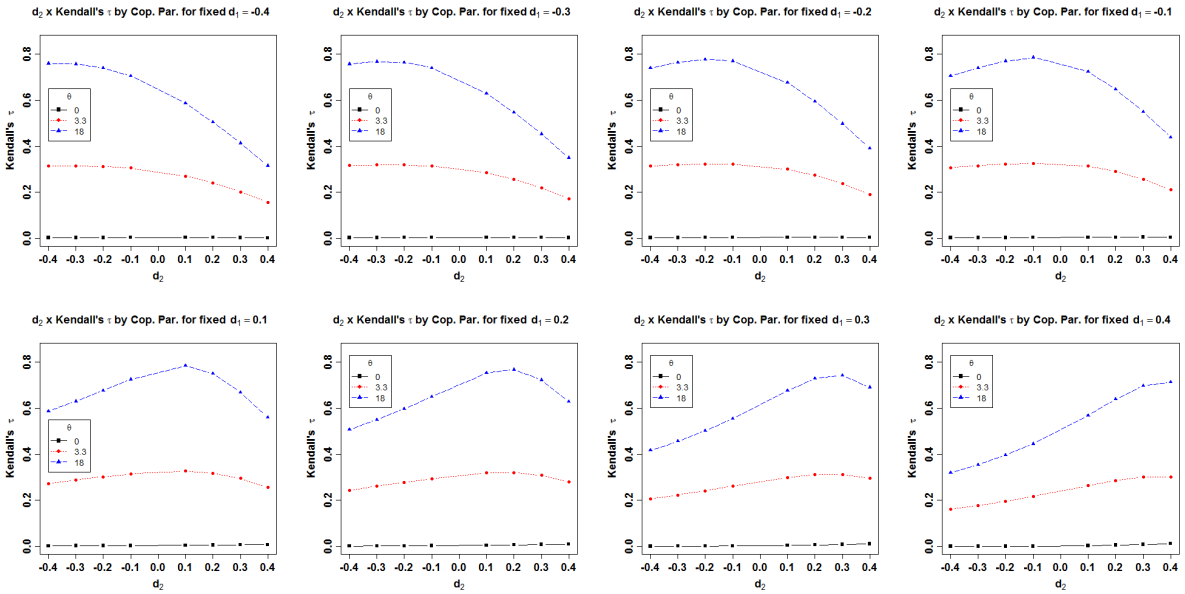


Figure A.14: Plots of d_2 by Kendall's τ for fixed d_1 and $\theta \in \{0, 3.3, 18\}$. Marginals were taken to be (standardized) t_7.