Pair copula constructions of point-optimal sign-based tests for predictive linear and nonlinear regressions

Kaveh Salehzadeh Nobari*

Lancaster University

November 9, 2021

[LATEST VERSION HERE]

Abstract

We propose pair copula constructed point-optimal sign tests in the context of linear and nonlinear predictive regressions with endogenous, persistent regressors, and disturbances exhibiting serial (nonlinear) dependence. The proposed approach entails considering the entire dependence structure of the signs to capture the serial dependence, and building feasible test statistics based on pair copula constructions of the sign process. The tests are exact and valid in the presence of heavy tailed and nonstandard errors, as well as heterogeneous and persistent volatility. Furthermore, they may be inverted to build confidence regions for the parameters of the regression function. Finally, we adopt an adaptive approach based on the split-sample technique to maximize the power of the test by finding an appropriate alternative hypothesis. In a Monte Carlo study, we compare the performance of the proposed "quasi"-point-optimal sign tests based on pair copula constructions by comparing its size and power to those of certain existing tests that are intended to be robust against heteroskedasticity. The simulation results maintain the superiority of our procedures to existing popular tests.

Keywords: Predictive regressions, D-vine, truncation, sequential estimation, persistent regressors, sign test, exact inference, adaptive method, power envelope.

JEL Codes: C12, C13, C15, C22, C52

*Department of Mathematics and Statistics, Lancaster University, Lancaster, LA1 4YR, UK (k.salehzadehnobari@lancs.ac.uk),

1 Introduction

Predictive regressions are frequently encountered in the economics and finance literature. Within this framework, the regressors are often highly persistent (and potentially nonstationary) with errors that exhibit contemporaneous correlation with the disturbances in the predictive regression. This leads to endogeneity, as a result of which the least-squared estimator of the regression parameters is biased. In such settings, t-type tests possess a non-standard distribution and inference using asymptotic critical values is no longer valid [see Mankiw and Shapiro (1986) and Stambaugh (1999) among others]. The econometric analysis of predictive regressions has been addressed extensively and numerous papers have suggested solutions to overcome this problem. These include reducing the bias in finite samples [see Nelson and Kim (1993) and Stambaugh (1999) among others] or self generated instrumental variables (IVs) that eliminate endogeneity [see Magdalinos and Phillips (2009), Phillips and Lee (2013) among others]. However, many of these studies impose strict assumptions on the degree of persistency of the regressors [see Phillips (2014) for an overview]. In this paper, we propose pair copula constructed point-optimal sign-based tests (PCC-POS-based tests hereafter) in the context of linear and nonlinear predictive regressions. The proposed tests are robust in the presence of persistent/endogenous regressors and errors, heterogeneous and persistent volatility, and disturbances that exhibit serial (nonlinear) dependence. Moreover, they are exact, consider the entire dependence structure of the signs and may be inverted to build confidence regions for the parameters of the regression function.

Sign-based tests, such as those proposed by Campbell and Dufour (1995, 1997), Luger (2003), and Dufour and Taamouti (2010) are randomized tests with a randomized distribution under the null hypothesis of unpredictability [see Pratt and Gibbons (2012) for a review of randomized tests]. Hence, under mild assumptions these procedures are distribution-free and do not suffer from the issues encountered by t-type statistics in finite samples. These class of tests are valid in the presence of non-normal distributions and heteroskedasiticty of unknown form [see Boldin et al. (1997) and Taamouti (2015) for a review of sign-based tests]. Furthermore, Dufour and Taamouti (2010) show that the heteroskedasticity and autocorrelation corrected tests developed by White (1980) (more commonly referred to as "HAC" procedures) are plagued with low power when the errors follow GARCH-type structures or there is a break in variance. To address these issues, Dufour and Taamouti (2010) propose point-optimal sign-based (POS-based) inference to test whether the conditional median of a response variable is zero against a linear regression alternative, where these procedures are further extended to nonlinear models.

In an earlier paper, we proposed an extension of the POS-based tests within a predictive regression framework. However, in order to obtain *feasible* test statistics, we imposed a Markovian assumption on the sign process. This paper relaxes the Markovian assumption on the signs, by decomposing the joint

distribution of the signs using models proposed by Panagiotelis et al. (2012) for multivariate discrete data based on pair copula constructions. The latter allow us to build feasible test statistics that are robust against heavy-tailed and asymmetric distributions, provided that the errors have zero median conditional on their own past and the explanatory variables. The PCC-POS-based tests are shown to be robust against non-standard distributions and heteroskedasticity of unknown form, and have the highest power among certain parametric and nonparametric tests that are intended to be robust against heteroskedasticity. Moreover, as in Dufour and Taamouti (2010), they can be inverted to produce a confidence region for the vector (sub-vector) of the parameters.

Although, the literature surrounding sign-based and sign-ranked inference is vast, the focus of the POS-based tests constructed by Dufour and Taamouti (2010) is to maximize power at a nominated point in the alternative parameter space, such that the power of the POS-based test is close to that of the power envelope - i.e. maximum attainable power for a given testing problem [see King (1987)]. Similarly, the PCC-POS-based tests are Neyman-Pearson type tests based on the signs, and as in Dufour and Taamouti (2010) a practical problem concerns finding an alternative at which the power of the PCC-POS-based tests is close to that of the power envelope. By conducting simulations exercises, Dufour and Taamouti (2010) find that the power of the POS-based tests is shifted close to the power envelope, when 10% of the sample is used to estimate the alternative and the remaining portion to calculate the test-statistic. Our simulations using a variety of split-sample PCC-POS-based tests confirm these findings.

Due to the nonlinear nature of the signs, there is inherent uncertainty regarding the structure of sign dependence. Therefore, it is important to consider the entire dependence structure of the signs. One approach for computing the joint distribution of the signs $s(y_1), \cdots, s(y_T)$, where $s(y_i) = \mathbb{1}_{\mathbb{R}^+ \cup \{0\}}\{y_i\}$, entails taking advantage of copula functions [see Sklar (1959)], which express the joint distribution of the signs in terms of i) the marginal distributions of the individual signs; and ii) the copula models capturing the dependence of the T signs. As the signs are discrete, the likelihood function of the POS-based tests under the alternative hypothesis can then be calculated using rectangle probabilities and in turn estimated using copulas with closed analytical form. However, this approach would not yield feasible test statistics, as the number of *multivariate* copulas that need to be evaluated increase at an exponential rate with growing sample sizes T. As a result of this curse of dimensionality, the literature concerning calculating probability mass functions (p.m.f hereafter) using discrete data is limited to low-dimensional data and copulas that are fast to calculate [see Nikoloulopoulos and Karlis (2008, 2009) and Li and Wong (2010)].

To propose feasible POS-based test statistics within a predictive regression framework, we use a discrete analogue of the vine PCCs introduced by Panagiotelis et al. (2012). The likelihood function of the signs under the alternative hypothesis can then be decomposed as a vine PCC under certain set of

conditions that are later outlined in the paper. The most important advantage of this method is that for a sample of size T, only $2T(T-1)$ *bivariate* copula evaluations are required, as opposed to 2^T *multivariate* copula evaluations using rectangle probabilities. Another advantage of the vine PCC methodology is that model selection techniques can be used to identify the conditional independence in the process of signs in order to create more parsimonious PCC models.

The rest of the paper is organized as follows: in Section 2, we motivate the use of the discrete analogue of the vine PCC for building POS-based tests. In Section 3, we outline the conditions under which vine PCCs can be implemented and we also discuss the choice of the PCC model. We then propose PCC-POS-based test statistics for linear and nonlinear models. Secion 4, discusses the estimation approach implemented for the vine PCC models. In Section 5, we derive the power envelope and highlight the choice of the alternative hypothesis for computing the PCC-POS-based test statistic. In Section 6, we discuss the problem of finding a confidence set for a vector (subvector) of parameters using the projection techniques. In Section 7, we assess the performance of the proposed tests in terms of size and power. Finally, in Section 8, we conclude the findings of the paper.

2 Framework

Consider a stochastic process $Z = \{Z_t = (y_t, \boldsymbol{x}'_{t-1}) : \Omega \to \mathbb{R}^{(k+1)}, t = 1, 2, \cdots\}$ defined on a probability space $(\Omega, \mathcal{F}, P)$. Within a framework of a predictive regression y_t can linearly be explained by a vector variable $\boldsymbol{x}_{t-1}$

$$y_t = \boldsymbol{\beta}'\boldsymbol{x}_{t-1} + \varepsilon_t, \quad t = 1, \cdots, T, \tag{1}$$

where y_t is the dependent variable and $\boldsymbol{x}_{t-1}$ is an $(k+1)\times 1$ vector of stochastic explanatory variables, say $\boldsymbol{x}_{t-1} = [1, x_{1,t-1}, \cdots, x_{k,t-1}]'$, $\boldsymbol{\beta} \in \mathbb{R}^{(k+1)}$ is an unknown vector of parameters with $\boldsymbol{\beta} = [\beta_0, \beta_1, \cdots, \beta_k]'$ and

$$\varepsilon_t \mid X \sim F_t(. \mid X)$$

where $F_t(. \mid X)$ is an unknown conditional distribution function and X=$[\boldsymbol{x}'_0, \cdots, \boldsymbol{x}'_{T-1}]'$ is an $T \times (k+1)$ information matrix.

We follow Coudin and Dufour (2009) by considering the median as an alternative measure of central tendency, which differs from the Martingale Difference Sequence (MDS hererafter) assumption. In the latter, it is generally assumed that for an adapted stochastic sequence $\{Z_t, \mathcal{F}_t, t = 1, 2, \cdots\}$, where $\mathcal{F}_t$ is a σ-field in Ω, $\mathcal{F}_s \subseteq \mathcal{F}_t$ for $s < t$, $\mathbb{E}\{\varepsilon_t \mid \mathcal{F}_{t-1}\} = 0$, $\forall t \geq 1$. Thus the alternative implies imposing a median-based analogue of the MDS on the error process - namely we suppose that ε_t is a strict conditional

mediangale

$$P[\varepsilon_t > 0 \mid \boldsymbol{\varepsilon}_{t-1}, X] = P[\varepsilon_t < 0 \mid \boldsymbol{\varepsilon}_{t-1}, X] = \frac{1}{2}, \tag{2}$$

with

$$\boldsymbol{\varepsilon}_0 = \{\emptyset\}, \quad \boldsymbol{\varepsilon}_{t-1} = \{\varepsilon_1, \cdots, \varepsilon_{t-1}\}, \quad \text{for} \quad t \geq 2.$$

Note (2) entails that $\varepsilon_t \mid X$ has no mass at zero for all t, which is only true if $\varepsilon_t \mid X$ is a continuous variable. Model (1) in conjunction with assumption (2) allows the error terms to possess asymmetric, heteroskedastic and serially dependent distributions, so long as the conditional medians are zero. Assumption (2) allows for many dependent schemes, such as those of the form $\varepsilon_1 = \sigma_1(x_1, \cdots, x_{t-2})\epsilon_1$, $\varepsilon_t = \sigma_1(x_1, \cdots, x_{t-2}, \varepsilon_1, \cdots, \varepsilon_{t-1})\epsilon_t$,$t = 2, \cdots, n$, where $\epsilon_1, \cdots, \epsilon_n$ are independent with a zero median. In time-series context this includes models such as ARCH, GARCH or stochastic volatility with non-Gaussian errors. Furthermore, in the mediangale framework the disturbances need not be second order stationary.

Suppose, we wish to test the null hypothesis of unpredictability

$$H_0 : \boldsymbol{\beta} = \mathbf{0}, \tag{3}$$

against the alternative

$$H_1 : \boldsymbol{\beta} = \boldsymbol{\beta}_1. \tag{4}$$

where $\mathbf{0}$ is a $(k+1) \times 1$ zero vector. Define the vector of signs as follows

$$U(T) = (s(y_1), \cdots, s(y_T))',$$

where for $t = 1, \cdots, T$

$$s(y_t) = \begin{cases} 1, \, if \;\; y_t \geq 0 \\ 0, \, if \;\; y_t < 0 \end{cases}.$$

We consider Neyman-Pearson type test based on the signs. Thus, to build POS-based tests for testing the null hypothesis (3) against the alternative (4), we first define the likelihood function of the sample in terms of signs $s(y_1), \cdots, s(y_T)$ conditional on X

$$L(U(T), \boldsymbol{\beta}, X) = P[s(y_1) = s_1, \cdots, s(y_T) = s_T \mid X] = \prod_{t=1}^{T} P\left[s(y_t) = s_t \mid \mathbb{S}_{t-1} = \underline{s}_{t-1}, X\right], \tag{5}$$

with

$$\mathbb{S}_0 = \{\emptyset\}, \quad \mathbb{S}_{t-1} = \{s(y_1), \cdots, s(y_{t-1})\}, \quad \text{for} \quad t \geq 2,$$

and

$$P[s(y_1) = s_1 \mid \underline{S}_0 = \underline{s}_0, X] = P[s(y_1) = s_1 \mid X],$$

where each s_t for $1 \leq t \leq T$ takes two possible values of 0 and 1. Under model (1) and assumption (2), the variables $s(\varepsilon_1), \cdots, s(\varepsilon_T)$ and in turn $s(y_1), \cdots, s(y_T)$ are i.i.d conditional on X, according to the distribution

$$P[s(\varepsilon_1) = 1 \mid X] = P[s(\varepsilon_1) = 0 \mid X] = \frac{1}{2}, \quad t = 1, \cdots, T.$$

This results holds true iff for any combination of $t = 1, \cdots, T$ there is a permutation $\pi : i \to j$ such that the mediangale assumption holds for j. Then the signs $s(\varepsilon_1), \cdots, s(\varepsilon_T)$ are i.i.d [see Theorem 1 of Coudin and Dufour (2009)]. Therefore, under the null hypothesis of unpredictability we have

$$P[s(y_t) = 1 \mid X] = P[s(y_t) = 0 \mid X] = \frac{1}{2}, \quad t = 1, \cdots, T. \tag{6}$$

Consequently, under the null hypothesis of orthogonality, the log-likelihood function conditional on X is given by

$$L_0(U(T), \mathbf{0}, X) = \prod_{t=1}^{T} P[s(y_t) = s_t \mid X] = \left(\frac{1}{2}\right)^T.$$

On the other hand, under the alternative we have

$$L_1(U(T), \boldsymbol{\beta}_1, X) = \prod_{t=1}^{T} P[s(y_t) = s_t \mid \underline{S}_{t-1} = \underline{s}_{t-1}, X],$$

where now for $t = 1, \cdots, T$

$$y_t = \boldsymbol{\beta}_1' \boldsymbol{x}_{t-1} + \varepsilon_t.$$

In an earlier paper, we considered optimal sign-based tests (in the Neyman-Pearson sense), which maximize power under the constraint $P[\text{Reject } H_0 \mid H_0] \leq \alpha$, where α is an arbitrary significance level [see Lehmann and Romano (2006)]. Let H_0 and H_1 be defined by (3) and (4) respectively. Then under the assumptions (1) and (2), the log-likelihood ratio

$$SL_T(\boldsymbol{\beta_1}) = \ln\left\{\frac{L_1(U(T), \boldsymbol{\beta}_1, X)}{L_0(U(T), \mathbf{0}, X)}\right\} > c, \tag{7}$$

is most powerful for testing H_0 against H_1 among level α tests based on the signs $(s(y_1), \cdots, s(y_T))'$, where c is the smallest constant such that

$$P[SL_T(\boldsymbol{\beta}_1) > c \mid H_0] \leq \alpha,$$

and α is an arbitrary significance level.

For POS-based tests within a predictive regression framework, the test statistic requires the calculation of $P[y_t \geq 0 \mid \underset{\sim}{S}_{t-1} = \underset{\sim}{s}_{t-1}, X]$ and $P[y_t < 0 \mid \underset{\sim}{S}_{t-1} = \underset{\sim}{s}_{t-1}, X]$. The latter is not easy to compute, as it involves the distribution of the joint process of signs $s(y_1), \cdots, s(y_T)$, conditional on X, which is unknown. Therefore, to obtain feasible test statistics, we may impose a Markovian assumption on the sign process. However, it may be important to capture the dependence structure of the entire process.

In this paper, we consider the entire dependence structure of the vector of signs by taking advantage of copulas. The Theorem of Sklar (1959) states that there exists a copula C such that

$$F(s_1, \cdots, s_T \mid X) = C(F_1(s_1 \mid X), \cdots, F_T(s_T \mid X)), \tag{8}$$

where F is a conditional joint cumulative distribution function (CDF hereafter) of the signs vector $\underset{\sim}{S} = (s(y_1), \cdots, s(y_T))'$ with conditional marginal distribution functions F_j for $j = 1, 2, \cdots, T$. Copula $C(.)$ is unique for continuous variables, but for discrete variables, it is unique only on the set

$$\text{Range}(F_1) \times \cdots \times \text{Range}(F_T),$$

which is the Cartesian product of the ranges of the conditional marginal distribution functions. To illustrate an example of non-uniqueness in the discrete case, let us consider a sample of two discrete binary variables, say $s(y_1)$ and $s(y_2)$, with corresponding conditional marginal distribution functions F_1 and F_2. We know that $F_j \sim \text{Bernoulli}(p_j)$ for $j = 1, 2$, such that

$$F_j(s_j \mid X) = \begin{cases} 0, & \text{for} \quad s_j < 0 \\ 1 - p_j, & \text{for} \quad 0 \leq s_j < 1 \\ 1, & \text{for} \quad s_j \geq 1 \end{cases} \tag{9}$$

Thus, $\text{Range}(F_1) = \{0, 1 - p_1, 1\}$ and $\text{Range}(F_2) = \{0, 1 - p_2, 1\}$, with the copula only being unique for $C(1 - p_1, 1 - p_2)$, noting that $C(0, 1 - p_j) = 0$ and $C(1, 1 - p_j) = 1 - p_j$ for $j = 1, 2$. However, this non-uniqueness does not preclude the use of parametric copulas for modelling discrete data [see. Joe (1997), Song et al. (2009)].

By considering this bivariate example, the conditional p.m.f can be expressed in terms of rectangle probabilities,

$$\begin{aligned} P[s(y_1) = s_1, s(y_2) = s_2 \mid X] &= P[s_1 - 1 < s(y_1) \leq s_1, s_2 - 1 < s(y_2) \leq s_2 \mid X] \\ &= F(s_1, s_2 \mid X) - F(s_1 - 1, s_2 \mid X) \end{aligned}$$

$$- F(s_1, s_2 - 1 \mid X) + F(s_1 - 1, s_2 - 1 \mid X),$$

and in turn in terms of copulas as follows

$$\begin{aligned} P[s(y_1) = s_1, s(y_2) = s_2 \mid X] &= F(s_1, s_2 \mid X) - F(s_1 - 1, s_2 \mid X) \\ &\quad - F(s_1, s_2 - 1 \mid X) + F(s_1 - 1, s_2 - 1 \mid X) \\ &= C(F_1(s_1 \mid X), F_2(s_2 \mid X)) - C(F_1(s_1 - 1 \mid X), F_2(s_2 \mid X)) \\ &\quad - C(F_1(s_1 \mid X), F_2(s_2 - 1 \mid X)) + C(F_1(s_1 - 1 \mid X), F_2(s_2 - 1 \mid X)), \end{aligned}$$

which implies that the T-variate conditional likelihood function (5) can be expressed in terms of 2^T finite differences

$$\begin{aligned} P[s(y_1) = s_1, \cdots, s(y_T) = s_T \mid X] &= \sum_{i_1=0,1} \cdots \sum_{i_T=0,1} (-1)^{i_1+\cdots+i_T} P[s(y_1) \leq s_1 - i_1, \cdots, s(y_T) \leq s_T - i_T \mid X] \\ &= \sum_{i_1=0,1} \cdots \sum_{i_T=0,1} (-1)^{i_1+\cdots+i_T} C(F_1(s_1 - i_1 \mid X), \cdots, F_T(s_T - i_T \mid X)). \end{aligned}$$

Evidently, the calculation of the conditional likelihood function (5) using this approach would require 2^T *multivariate* copula evaluations, which is not computationally feasible. However, by employing the vine PCC introduced later in the paper, we will show that this number can be reduced to only $2T(T-1)$ *bivariate* copula evaluations. The latter method provides us with flexibility, since any multivariate discrete distribution can be decomposed as a vine PCC under a set of conditions that are discussed in the following Section.

3 Pair copula constructions of point-optimal sign tests for predictive regressions

In this Section, we derive POS-based tests in the context of linear and nonlinear regression models based on vine PCC decomposition. Following a structure similar to Dufour and Taamouti (2010), we first consider the problem of testing whether the conditional median of a vector of observations is zero against a linear regression alternative. We further consider the conditions under which the conditional likelihood function under the alternative hypothesis can be decomposed as a vine PCC, and as such, choose an appropriate vine model. These results are later generalized to test whether the coefficients of a possibly nonlinear median regression function have a given value against an alternative nonlinear median regression.

3.1 Testing orthogonality hypothesis in linear predictive regressions

Consider the problem of testing the null hypothesis of unpredictability (3) against the alternative (4), using the test statistic (7) and given the assumptions (1) and (2). As it was shown in Section 2, under the alternative hypothesis the conditional likelihood function can be expressed as

$$L_1(U(T), \boldsymbol{\beta}_1, X) = \prod_{t=1}^{T} P\left[s(y_t) = s_t \mid \underline{\mathrm{S}}_{t-1} = \underline{\mathrm{s}}_{t-1}, X\right]. \tag{10}$$

Let $s(y_j)$ be a scalar element of $\underline{\mathrm{S}}_{t-1}$, with $\underline{\mathrm{S}}_{t-1}^{\setminus j} = \underline{\mathrm{S}}_{t-1} \setminus s(y_j)$ such that

$$\underline{\mathrm{S}}_{t-1}^{\setminus j} = \{s(y_1), s(y_2), \cdots, s(y_{j-1}), s(y_{j+1}), \cdots, s(y_{t-1})\}$$

and $s(y_t) \notin \underline{\mathrm{S}}_{t-1}$. By choosing a single element of $\underline{\mathrm{S}}_{t-1}$, say $s(y_j)$, we would have

$$\begin{aligned} P\left[s(y_t) = s_t \mid \underline{\mathrm{S}}_{t-1} = \underline{\mathrm{s}}_{t-1}, X\right] &= \frac{P[s(y_t) = s_t, s(y_j) = s_j \mid \underline{\mathrm{S}}_{t-1}^{\setminus j} = \underline{\mathrm{s}}_{t-1}^{\setminus j}, X]}{P[s(y_j) = s_j \mid \underline{\mathrm{S}}_{t-1}^{\setminus j} = \underline{\mathrm{s}}_{t-1}^{\setminus j}, X]} \\ &= \sum_{k_t=0,1} \sum_{k_j=0,1} (-1)^{k_t+k_j} \times \\ &\quad \left\{P[s(y_t) \leq s_t - k_t, s(y_j) \leq s_j - k_j \mid \underline{\mathrm{S}}_{t-1}^{\setminus j} = \underline{\mathrm{s}}_{t-1}^{\setminus j}, X]\right\} \\ &\quad / P[s(y_j) = s_j \mid \underline{\mathrm{S}}_{t-1}^{\setminus j} = \underline{\mathrm{s}}_{t-1}^{\setminus j}, X], \end{aligned} \tag{11}$$

where the bivariate conditional probability in (11) can be expressed in terms of copulas as follows

$$\begin{aligned} P[s(y_t) = s_t \mid \underline{\mathrm{S}}_{t-1} = \underline{\mathrm{s}}_{t-1}, X] = \sum_{k_t=0,1} \sum_{k_j=0,1} (-1)^{k_t+k_j} \Bigg\{ & \\ C_{s(y_t),s(y_j)|\underline{\mathrm{S}}_{t-1}^{\setminus j}} & \left(F_{s(y_t)|\underline{\mathrm{S}}_{t-1}^{\setminus j}}(s_t - k_t \mid \underline{\mathrm{s}}_{t-1}^{\setminus j}, X), F_{s(y_j)|\underline{\mathrm{S}}_{t-1}^{\setminus j}}(s_j - k_j \mid \underline{\mathrm{s}}_{t-1}^{\setminus j}, X)\right) \Bigg\} \\ & / P[s(y_j) = s_j \mid \underline{\mathrm{S}}_{t-1}^{\setminus j} = \underline{\mathrm{s}}_{t-1}^{\setminus j}, X]. \end{aligned} \tag{12}$$

Further, let $\underline{\mathrm{S}}_{t-1}^{\setminus i,j} = \underline{\mathrm{S}}_{t-1}^{\setminus j} \setminus s(y_i)$, such that $s(y_i)$ is a scalar element of $\underline{\mathrm{S}}_{t-1}^{\setminus j}$. Then the arguments $F_{s(y_t)|\underline{\mathrm{S}}_{t-1}^{\setminus j}}$

and $F_{s(y_j)|\underline{\mathrm{S}}_{t-1}^{\setminus j}}$ in copula expression (12) can be presented by the general form

$$
\begin{aligned}
&F_{s(y_t)|s(y_i),\underline{\mathrm{S}}_{t-1}^{\setminus i,j}}(s_t - k_t \mid s_i, \underline{\mathrm{s}}_{t-1}^{\setminus i,j}, X) = \\
&\left\{ C_{s(y_t),s(y_i)|\underline{\mathrm{S}}_{t-1}^{\setminus i,j}} \left(F_{s(y_t)|\underline{\mathrm{S}}_{t-1}^{\setminus i,j}}(s_t - k_t \mid \underline{\mathrm{s}}_{t-1}^{\setminus i,j}, X), F_{s(y_i)|\underline{\mathrm{S}}_{t-1}^{\setminus i,j}}(s(y_i) \mid \underline{\mathrm{s}}_{t-1}^{\setminus i,j}, X) \right) - \right. \\
&\left. \quad C_{s(y_t),s(y_i)|\underline{\mathrm{S}}_{t-1}^{\setminus i,j}} \left(F_{s(y_t)|\underline{\mathrm{S}}_{t-1}^{\setminus i,j}}(s_t - k_t \mid \underline{\mathrm{s}}_{t-1}^{\setminus i,j}, X), F_{s(y_i)|\underline{\mathrm{S}}_{t-1}^{\setminus i,j}}(s(y_i) - 1 \mid \underline{\mathrm{s}}_{t-1}^{\setminus i,j}, X) \right) \right\} \\
&/ P[s(y_i) = s_i \mid \underline{\mathrm{S}}_{t-1}^{\setminus i,j} = \underline{\mathrm{s}}_{t-1}^{\setminus i,j}, X].
\end{aligned}
\tag{13}
$$

Thus, decomposition (12), and in turn (13) can be applied recursively to the elements of the conditional likelihood function (5), such that it is expressed in terms of bivariate copulas. Let $\underline{\mathrm{S}}_{t-1} = \{s(y_1), \cdots, s(y_{t-1})\}$ be the variables that $s(y_t)$ for $t = 2, \cdots, T$ is conditioned on. We follow Joe (2014), by letting $\underline{\sigma}_{t-1} = \{\sigma(1,t), \cdots, \sigma(t-1,t)\}$ be a permutation of $\underline{\mathrm{S}}_{t-1}$, such that $s(y_t)$ is paired sequentially first with $\sigma(1,t)$, then $\sigma(2,t)$ and finally $\sigma(t-1,t)$, where in the r^{th} step $(2 \leq r \leq t-1)$, $\sigma(r,t)$ is paired to t conditional on $\sigma(1,t), \cdots, \sigma(r-1,t)$. For $n \leq 3$ (i.e. $t = 2, 3$) there are only three possible permutations with $\underline{\sigma}_1 = \{s(y_1)\}$ for $t = 2$, and $\underline{\sigma}_2 = \{s(y_1), s(y_2)\}$, as well as $\underline{\sigma}_2 = (s(y_2), s(y_1))$ for $t = 3$ respectively. Therefore, under assumptions (1) and (2), and with $T \leq 3$, let H_0 and H_1 be defined by (3) - (4), then the Neyman-Pearson type test-statistic based on the signs $(s(y_1), \cdots, s(y_T))'$ can be expressed as

$$
\begin{aligned}
SL_T(\boldsymbol{\beta}_1) = \ln P\left[s(y_1) = s_1 \mid X\right] &+ \sum_{t=2}^{T} \ln \Delta_{s_t^-}^{s_t^+} \Delta_{s_{t-1}^-}^{s_{t-1}^+} C_{t,t-1|t-2} \\
&- \sum_{t=2}^{T} \ln P[s(y_{t-1}) = s_{t-1} \mid \underline{\mathrm{S}}_{t-2} = \underline{\mathrm{s}}_{t-2}, X] - T \ln \left\{ \frac{1}{2} \right\},
\end{aligned}
$$

for $t = 2, \cdots, T$, where

$$
\begin{aligned}
\Delta_{s_t^-}^{s_t^+} \Delta_{s_{t-1}^-}^{s_{t-1}^+} C_{t,t-1|t-2} = &\sum_{k_t=0,1} \sum_{k_{t-1}=0,1} (-1)^{k_t + k_{t-1}} \\
&\times \left(C_{s(y_t),s(y_{t-1})|\underline{\mathrm{S}}_{t-2}} \left(F_{s(y_t)|\underline{\mathrm{S}}_{t-2}}(s_t - k_t \mid \underline{\mathrm{s}}_{t-2}, X), F_{s(y_{t-1})|\underline{\mathrm{S}}_{t-2}}(s_{t-1} - k_{t-1} \mid \underline{\mathrm{s}}_{t-2}, X) \right) \right)
\end{aligned}
$$

and such that

$$
\ln P\left[s(y_1) = s_1 \mid \underline{\mathrm{S}}_0 = \underline{\mathrm{s}}_0, X\right] = s(y_1) \ln \left\{ \frac{P[y_1 \geq 0 \mid X]}{P[y_1 < 0 \mid X]} \right\} + \ln P[y_1 < 0 \mid X].
$$

For $T > 3$, the permutations $\underline{\sigma}_{t-1}$ are dependent on the choice of the permutations at stages $3, \cdots, t-1$. Therefore, an issue that requires considerable attention is whether there *exists* a decomposition such as the one considered in the earlier example for $T > 3$. Furthermore, the conditional likelihood function

expressed in terms of bivariate copulas by recursively using (12) and (13), assumes that a single copula is specified for each conditional bivariate distribution $F_{s(y_t),s(y_j)|\mathbb{S}_{t-1}^{\setminus j}}$ in decomposition (10) over all possible values of $\mathbb{S}_{t-1}^{\setminus j}$. This implies that the copula is unique for the Cartesian product of the ranges of conditional CDFs $F_{s(y_t)|\mathbb{S}_{t-1}^{\setminus j}}$ and $F_{s(y_j)|\mathbb{S}_{t-1}^{\setminus j}}$. Therefore, the decomposition must be such that each conditional bivariate distribution in said vine has a constant conditional copula [see Panagiotelis et al. (2012)]. For a constant conditional copula to exist, the following conditions outlined by Panagiotelis et al. (2012) must be satisfied.

Definition 1 (Existence of constant conditional copula) *Consider the conditional bivariate distribution function* $F_{s(y_t),s(y_j)|\underline{S}_{t-1}^{\setminus j}}$. *We say that a copula* $C = C_{s(y_t),s(y_j)|\underline{S}_{t-1}^{\setminus j}}$ *is constant over all possible values of* $\underline{S}_{t-1}^{\setminus j}$ *if*

$$\sum_{m=0,1}\sum_{n=0,1}(-1)^{m+n}C(a_{k-m},b_{l-n})\geq 0,\quad \forall k,l\in\{1,2\}\times\{1,2\},$$

where $a_0<a_1<a_2$ *and* $b_0<b_1<b_2$, *are the distinct points corresponding to the ranges of the conditional Bernoulli CDFs* $F_{s(y_t)|\underline{S}_{t-1}^{\setminus j}}$ *and* $F_{s(y_j)|\underline{S}_{t-1}^{\setminus j}}$ *respectively, such that* $a_0=b_0=0$ *and* $a_2=b_2=1$, *and where further, the following constraints are satisfied:*

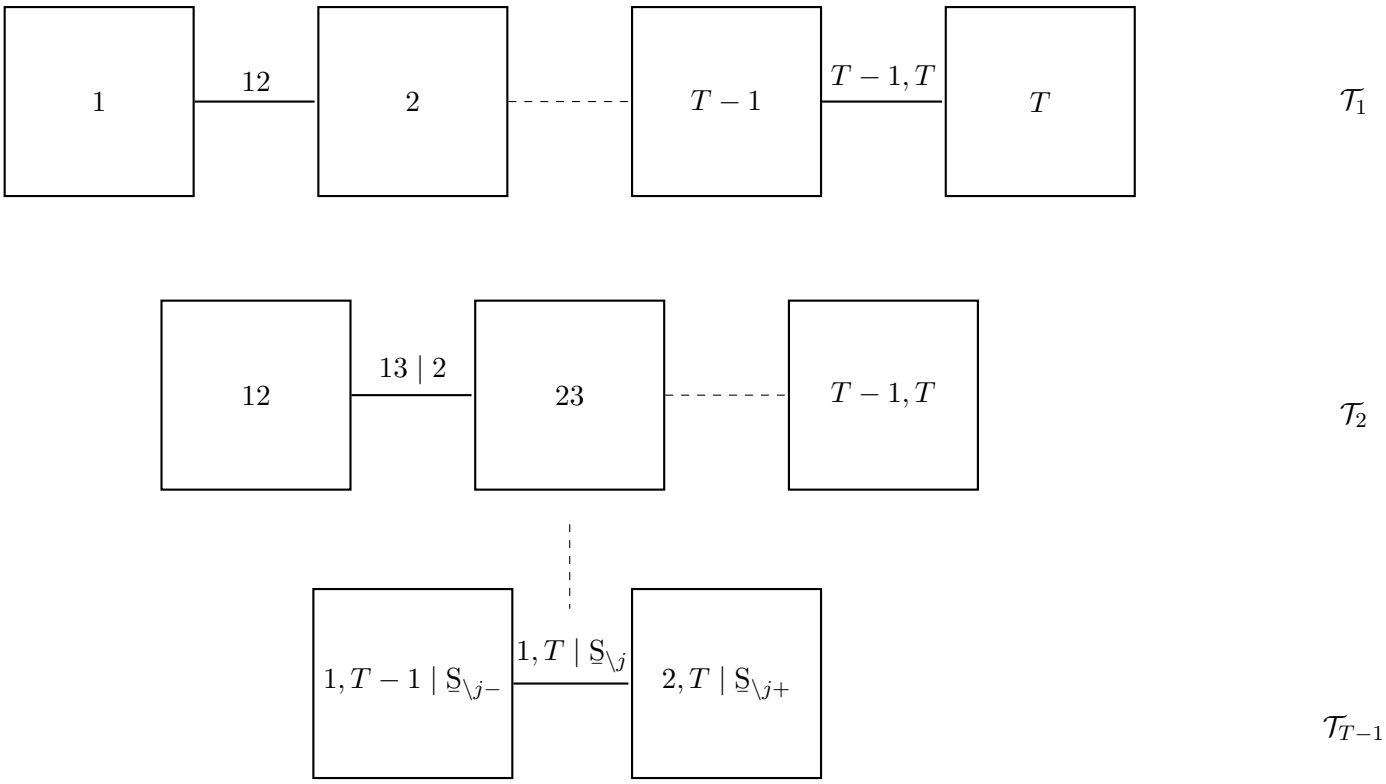
$$C_{s(y_t),s(y_j)|\underline{S}_{t-1}^{\setminus j}}\left(a_{s(y_t)|\underline{S}_{t-1}^{\setminus j}},b_{s(y_j)|\underline{S}_{t-1}^{\setminus j}}\right)=P[s(y_t)\leq s_t,s(y_j)\leq s_j\mid \underline{S}_{t-1}^{\setminus j}=\underline{s}_{t-1}^{\setminus j},X],$$
$$C_{s(y_t),s(y_j)|\underline{S}_{t-1}^{\setminus j}}\left(1,b_{s(y_j)|\underline{S}_{t-1}^{\setminus j}}\right)=b_{s(y_j)|\underline{S}_{t-1}^{\setminus j}},\quad C_{s(y_t),s(y_j)|\underline{S}_{t-1}^{\setminus j}}\left(a_{s(y_t)|\underline{S}_{t-1}^{\setminus j}},1\right)=a_{s(y_t)|\underline{S}_{t-1}^{\setminus j}},$$

with $a_{s(y_t)|\underline{S}_{t-1}^{\setminus j}}:=P[s(y_t)\leq s_t\mid \underline{S}_{t-1}^{\setminus j}=\underline{s}_{t-1}^{\setminus j},X]$ *and* $b_{s(y_j)|\underline{S}_{t-1}^{\setminus j}}:=P[s(y_j)\leq s_j\mid \underline{S}_{t-1}^{\setminus j}=\underline{S}_{t-1}^{\setminus j},X]$.

To satisfy the above conditions, the vine decomposition must be such that the strength of the dependence of the conditional bivariate distribution does not vary much across different values of the conditioning set [see Panagiotelis et al. (2012)]. As we are dealing with time-series data, the D-vine decomposition yields a constant dependence structure over different values of $\mathbb{S}_{t-1}^{\setminus j}$, and is thus, the most appropriate and intuitive choice for the decomposition of the conditional likelihood function (10).

The D-vine PCC (figure 1) is constructed by $T-1$ trees, say $D=\{\mathcal{T}_1,\cdots,\mathcal{T}_{T-1}\}$, comprised of the edges $\xi(D)=\xi(\mathcal{T}_1)\cup\cdots\cup\xi(\mathcal{T}_{T-1})$, where $\xi(\mathcal{T}_l)$ refers to the edges of the tree $\mathcal{T}_l$. In the first tree T_1, the marginals $F(s_1\mid X),F(s_2\mid X),\cdots,F(s_T\mid X)$, are arranged as nodes in consecutive order, say $N(\mathcal{T}_1):=\{1,2,\cdots,T-1,T\}$, where the nodes are of degree two, meaning that no more than two edges is connected to each node. The corresponding edges join the adjacent nodes, such that $\xi(\mathcal{T}_1):=\{12,23,\cdots,(T-1,T)\}$. Next, the edges of the first tree $\xi(\mathcal{T}_1)$ become the nodes of $\mathcal{T}_2$, a process which is completed in a recursive manner, such that $N(\mathcal{T}_{l+1})=\xi(\mathcal{T}_l)$, with the edges of each tree joining the adjacent nodes, and with the mutual elements between the nodes becoming the conditioning set.

Figure 1: D-vine PCC for the T-variate case



Note: D-vine for a sample size T consists of $T-1$ trees. The first tree consists of the conditional marginals ordered consecutively as nodes, with the edges connecting the adjacent nodes, and with the elements shared by the two nodes going in the conditioning set. The edges of each tree $\mathcal{T}_l$ become the nodes of the tree $\mathcal{T}_{l+1}$. In this figure, $\underline{\mathrm{S}}_{\backslash i-}$ and $\underline{\mathrm{S}}_{\backslash j+}$ correspond to the elements $\underline{\mathrm{S}}_{\backslash j-} := \{2, 3, \cdots, T-2\}$ and $\underline{\mathrm{S}}_{\backslash j+} := \{3, \cdots, T-1\}$ respectively, with $\underline{\mathrm{S}}_{\backslash j} := \{2, 3, \cdots, T-1\}$.

To express the conditional likelihood function (10) as a D-vine, we begin by calculating the conditional marginals $F_1, \cdots, F_T$, where $F_t \sim Bernoulli(p_t)$, for $t = 1, \cdots, T$, with conditional CDFs that are expressed as (9). Therefore, under assumptions (1) and (2), we have

$$F_t(s_t \mid X) = \begin{cases} 0, & \text{for} \quad s_t < 0 \\ 1 - P[\varepsilon_t \geq -\beta' x_{t-1} \mid X], & \text{for} \quad 0 \leq s_t < 1 \\ 1, & \text{for} \quad s_t \geq 1 \end{cases}, \quad t = 1, \cdots, T. \tag{14}$$

Once the marginals are obtained, the next step consists of evaluating the copulas in the first tree - i.e. $C_{12}(F_1, F_2), \cdots, C_{T-1,T}(F_{T-1}, F_T)$, corresponding to the edges $\xi(\mathcal{T}_1)$). In the second tree, the copulas $C_{13|2}(F_{1|2}, F_{3|2}), \cdots, C_{T-2,T|T-1}(F_{T-2|T-1}, F_{T|T-1})$ are evaluated, then $C_{14|23}(F_{1|23}, F_{4|23}), \cdots, C_{T-3,T|T-2,T-1}(F_{T-3|T-2,T-1}, F_{T|T-2,T-1})$ in the third tree, and so on.

In the case of continuous variables, say $\{s^*(y_t) \in \mathbb{R}, t = 1, \cdots, T\}$, the construction of the D-vine involves an iterative copula evaluation process for the trees $\mathcal{T}_1, \cdots, \mathcal{T}_{T-1}$. This leads to $T(T-1)/2$ copula evaluations, which correspond to one copula evaluation for each edge [see Appendix]. On the other hand, for discrete variables, the conditional p.m.fs are expressed as in (12), which requires the evaluation of the following four copulas

$$\begin{aligned} &C^{++}_{t,j|\backslash\mathbf{j}}(F^{+}_{t|\backslash\mathbf{j}}, F^{+}_{j|\backslash\mathbf{j}}), \quad C^{+-}_{t,j|\backslash\mathbf{j}}(F^{+}_{t|\backslash\mathbf{j}}, F^{-}_{j|\backslash\mathbf{j}}), \\ &C^{-+}_{t,j|\backslash\mathbf{j}}(F^{-}_{t|\backslash\mathbf{j}}, F^{+}_{j|\backslash\mathbf{j}}), \quad C^{--}_{t,j|\backslash\mathbf{j}}(F^{-}_{t|\backslash\mathbf{j}}, F^{-}_{j|\backslash\mathbf{j}}), \end{aligned}$$

where $F^{+}_{t|\backslash\mathbf{j}} = P[s(y_t) \leq s_t \mid \underline{S}^{\backslash j}_{t-1} = \underline{s}^{\backslash j}_{t-1}, X]$ and $F^{-}_{t|\backslash\mathbf{j}} = P[s(y_t) \leq s_t - 1 \mid \underline{S}^{\backslash j}_{t-1} = \underline{s}^{\backslash j}_{t-1}, X]$. Henceforth, $4 \times T(T-1)/2$ bivariate copulas need to be evaluated in the case of discrete data.

Let us express the conditional joint p.m.f of the signs as follows

$$P_1[s(y_1) = s_1 \mid X] \times \prod_{t=2}^{T} P_{t|1:t-1}[s(y_t) = s_t \mid s(y_1) = s_1, \cdots, s(y_{t-1}) = s_{t-1}, X], \tag{15}$$

where following the results by Stoeber et al. (2013), if the D-vine is expressed as a vine-array $A = (\sigma_{lt})_{1 \leq l \leq t \leq T}$, where $l = 2, \cdots, T-1$ is the row with tree $\mathcal{T}_l$, and column t has the permutation $\underline{\sigma}_{t-1} =$

$(\sigma(1,t),\cdots,\sigma(t-1,t))$ of the previously added variables, such that

$$\begin{bmatrix} - & 12 & 23 & 34 & \cdots & T-1,T \\ & - & 13\mid 2 & 24\mid 3 & \cdots & T-2,T\mid T-1 \\ & & \ddots & \cdots & \cdots & \vdots \\ & & & - & 1,T-1\mid \underline{\mathbb{S}}_{\backslash j-} & 2,T\mid \underline{\mathbb{S}}_{\backslash j+} \\ & & & & - & 1,T\mid \underline{\mathbb{S}}_{\backslash j} \\ & & & & & - \end{bmatrix}, \quad A = \begin{bmatrix} 1 & 1 & 2 & 3 & \cdots & T-1 \\ & 2 & 1 & 2 & \cdots & T-2 \\ & & \ddots & \cdots & \cdots & \vdots \\ & & & T-2 & 1 & 2 \\ & & & & T-1 & 1 \\ & & & & & T \end{bmatrix}$$

then

$$P_{t|1:t-1}[s(y_t)=s_t \mid s(y_1)=s_1:s(y_{t-1})=s_{t-1},X] = \left\{\prod_{l=t-1}^{2} c_{\sigma_{lt}t,|\sigma_{1t},\cdots,\sigma_{t-1,t}}\right\} \times c_{\sigma_{lt}t} \times P_t[s(y_t)=s_t \mid X], \tag{16}$$

where following Joe (2014), the copula densities in expression (16) are calculated by

$$c_{t,j|\backslash\mathbf{j}} = \frac{C^{++}_{t,j|\backslash\mathbf{j}}(F^{+}_{t|\backslash\mathbf{j}},F^{+}_{j|\backslash\mathbf{j}}) - C^{-+}_{t,j|\backslash\mathbf{j}}(F^{-}_{t|\backslash\mathbf{j}},F^{+}_{j|\backslash\mathbf{j}}) - C^{+-}_{t,j|\backslash\mathbf{j}}(F^{+}_{t|\backslash\mathbf{j}},F^{-}_{j|\backslash\mathbf{j}}) + C^{--}_{t,j|\backslash\mathbf{j}}(F^{-}_{t|\backslash\mathbf{j}},F^{-}_{j|\backslash\mathbf{j}})}{P_{t|\backslash\mathbf{j}}[s(y_t)=s_t \mid \underline{\mathbb{S}}^{\backslash j}_{t-1}=\underline{\mathbb{s}}^{\backslash j}_{t-1},X]P_{j|\backslash\mathbf{j}}[s(y_j)=s_j \mid \underline{\mathbb{S}}^{\backslash j}_{t-1}=\underline{\mathbb{s}}^{\backslash j}_{t-1},X]}, \tag{17}$$

which leads to the following proposition.

Proposition 1 *Let $A=(\sigma_{lt})_{1\le l\le t\le T}$ be a D-vine array for the signs $s(y_1),\cdots,s(y_T)$. Under assumptions (1) and (2), let H_0 and H_1 be defined by (3) - (4),*

$$SL_T(\boldsymbol{\beta}_1) = \sum_{t=2}^{T}\sum_{l=t-1}^{2} \ln c_{\sigma_{lt}t,|\sigma_{1t},\cdots,\sigma_{t-1,t}} + \sum_{t=2}^{T} \ln c_{\sigma_{1t}t} + \sum_{t=1}^{T} s(y_t)a_t(\boldsymbol{\beta}_1) > c_1(\boldsymbol{\beta}_1),$$

where

$$a_t(\boldsymbol{\beta}_1) = \ln\left\{\frac{1-P_t[\varepsilon_t \le -\boldsymbol{\beta}'\boldsymbol{x}_{t-1} \mid X]}{P_t[\varepsilon_t \le -\boldsymbol{\beta}'\boldsymbol{x}_{t-1} \mid X]}\right\},$$

and suppose the constant $c_1(\boldsymbol{\beta}_1)$ satisfies $P\left[SL_T(\boldsymbol{\beta}_1) > c_1(\boldsymbol{\beta}_1)\right] = \alpha$ under H_0, with $0<\alpha<1$. Then the test that rejects H_0 when

$$SL_T(\boldsymbol{\beta}_1) > c_1(\boldsymbol{\beta}_1) \tag{18}$$

is most powerful for testing H_0 against H_1 among level-α tests based on the signs $\big(s(y_1),\cdots,s(y_T)\big)'$.

Under the null hypothesis, the signs $s(y_1),\cdots,s(y_T)$ are i.i.d. according to Bernoulli $Bi(1,0.5)$, with the distribution of $SL_T(\boldsymbol{\beta}_1)$ only depending on the weights $a_t(\boldsymbol{\beta}_1)$, without the presence of any nuisance parameters. Assumption (2) implies that tests based on $SL_T(\boldsymbol{\beta}_1)$, such as the test given by (27), are distribution-free and robust against heteroskedasticity of unknown form. On the other hand, under the

alternative hypothesis, the power function of the test depends on the form of the distribution of ε_t. A special case is where $\varepsilon_1, \cdots, \varepsilon_T$ are independently distributed according to $N(0,1)$, which leads to the optimal test statistic assuming the following form

$$SL_T(\boldsymbol{\beta}_1) = \sum_{t=2}^{T} \sum_{l=t-1}^{2} \ln c_{\sigma_{lt}t,|\sigma_{1t},\cdots,\sigma_{t-1,t}} + \sum_{t=2}^{T} \ln c_{\sigma_{1t}t} + \sum_{t=1}^{T} s(y_t) a_t(\boldsymbol{\beta}_1) > c_1(\boldsymbol{\beta}_1),$$

where

$$a_t(\boldsymbol{\beta}_1) = \ln \left\{ \frac{\Phi(\boldsymbol{\beta}' \boldsymbol{x}_{t-1})}{1 - \Phi(\boldsymbol{\beta}' \boldsymbol{x}_{t-1})} \right\},$$

where $\Phi(.)$ is the standard normal distribution function. The distribution of $SL_T(\boldsymbol{\beta}_1)$ can be simulated under the null hypothesis with sufficient number of replications, and the critical values can be obtained to any degree of precision.

3.2 Testing general full coefficient hypothesis in nonlinear predictive regressions

We now consider the nonlinear predictive regression model

$$y_t = f(\boldsymbol{x}_{t-1}, \boldsymbol{\beta}) + \varepsilon_t, \ t = 1, \cdots, T, \tag{19}$$

where $\boldsymbol{x}_{t-1}$ is a $(k+1)\times 1$ vector of stochastic explanatory variables, such that $\boldsymbol{x}_{t-1} = [1, x_{1,t-1}, \cdots, x_{k,t-1}]'$, $f(\,\cdot\,)$ is a scalar function, $\boldsymbol{\beta} \in \mathbb{R}^{(k+1)}$ is an unknown vector of parameters and

$$\varepsilon_t \mid X \sim F_t(. \mid X)$$

where as before $F_t(. \mid X)$ is a distribution function and $X = [\boldsymbol{x}_0', \cdots, \boldsymbol{x}_{T-1}']$ is an $T \times (k+1)$ matrix. Suppose that the error process $\{\varepsilon_t, t = 1, 2, \cdots\}$ is a strict conditional mediangale, such that

$$P[\varepsilon_t > 0 \mid \boldsymbol{\varepsilon}_{t-1}, X] = P[\varepsilon_t < 0 \mid \boldsymbol{\varepsilon}_{t-1}, X] = \frac{1}{2}, \tag{20}$$

with

$$\boldsymbol{\varepsilon}_0 = \{\emptyset\}, \quad \boldsymbol{\varepsilon}_{t-1} = \{\varepsilon_1, \cdots, \varepsilon_{t-1}\}, \quad \text{for} \quad t \geq 2$$

and where (20) entails that $\varepsilon_t \mid X$ has no mass at zero, *i.e.* $P[\varepsilon_t = 0 \mid X]$=0 for all t. We do not require that the parameter vector $\boldsymbol{\beta}$ be identified.

We consider the problem of testing the null hypothesis

$$H(\boldsymbol{\beta}_0) : \boldsymbol{\beta} = \boldsymbol{\beta}_0, \tag{21}$$

against the alternative hypothesis

$$H(\boldsymbol{\beta}_1) : \boldsymbol{\beta} = \boldsymbol{\beta}_1, \tag{22}$$

We construct a test statistic for testing $H(\beta_0)$ against $H(\beta_1)$ in a similar manner to Section 3.1, by transforming model (19) to

$$\tilde{y}_t = g(\boldsymbol{x}_{t-1}, \boldsymbol{\beta}, \boldsymbol{\beta}_0) + \varepsilon_t, \quad t = 1, \cdots, T$$

where $\tilde{y}_t = y_t - f(\boldsymbol{x}_{t-1}, \boldsymbol{\beta}_0)$ and $g(\boldsymbol{x}_{t-1}, \boldsymbol{\beta}, \boldsymbol{\beta}_0) = f(\boldsymbol{x}_{t-1}, \boldsymbol{\beta}) - f(\boldsymbol{x}_{t-1}, \boldsymbol{\beta}_0)$. Notice that testing $H(\boldsymbol{\beta}_0)$ against $H(\boldsymbol{\beta}_1)$ is equivalent to testing

$$\bar{H}_0 : g(\boldsymbol{x}_{t-1}, \boldsymbol{\beta}, \boldsymbol{\beta}_0) = \mathbf{0}, \quad \text{for} \quad t = 1, \cdots, T$$

where $\mathbf{0}$ is a $(k+1) \times 1$ zero vector, against the alternative

$$\bar{H}_A : g(\boldsymbol{x}_{t-1}, \boldsymbol{\beta}, \boldsymbol{\beta}_0) = f(\boldsymbol{x}_{t-1}, \boldsymbol{\beta}_1) - f(x_{t-1}, \boldsymbol{\beta}_0), \quad \text{for} \quad t = 1, \cdots, T.$$

For $\tilde{U}(T) = (s(\tilde{y}_1), \cdots, s(\tilde{y}_T))'$, where for $1 \leq t \leq T$

$$s(\tilde{y}_t) = \begin{cases} 1, \, if \;\; \tilde{y}_t \geq 0 \\ 0, \, if \;\; \tilde{y}_t < 0 \end{cases}.$$

As before, the conditional joint p.m.f of the process of signs is expressed as

$$P_1[s(\tilde{y}_1) = \tilde{s}_1 \mid X] \times \prod_{t=2}^{T} P_{t|1:t-1}[s(\tilde{y}_t) = \tilde{s}_t \mid s(\tilde{y}_1) = \tilde{s}_1, \cdots, s(\tilde{y}_{t-1}) = \tilde{s}_{t-1}, X]. \tag{23}$$

Furthermore, the D-vine-array $\tilde{A} = (\tilde{\sigma}_{lt})_{1 \leq l \leq t \leq T}$, is such that $l = 2, \cdots, T-1$ is the row with tree $\mathcal{T}_l$, and column t has the permutation $\underline{\tilde{\sigma}}_{t-1} = (\tilde{\sigma}_{1t}, \cdots, \tilde{\sigma}_{t-1,t})$ of the previously added variables. Then

$$P_{t|1:t-1}[s(\tilde{y}_t) = \tilde{s}_t \mid s(\tilde{y}_1) = \tilde{s}_1, \cdots, s(\tilde{y}_{t-1}) = \tilde{s}_{t-1}, X] = \left\{ \prod_{l=t-1}^{2} c_{\tilde{\sigma}_{lt}t, |\tilde{\sigma}_{1t}, \cdots, \tilde{\sigma}_{t-1,t}} \right\} \times c_{\tilde{\sigma}_{lt}t} \times P_t[s(\tilde{y}_t) = \tilde{s}_t \mid X], \tag{24}$$

which leads to the following corollary.

Corollary 1 *Let $\tilde{A} = (\tilde{\sigma}_{lt})_{1 \leq l \leq t \leq T}$ be a D-vine array for the signs $s(\tilde{y}_1), \cdots, s(\tilde{y}_T)$. Under assumptions (19) and (2), let $H(\boldsymbol{\beta}_0)$ and $H(\boldsymbol{\beta}_1)$ be defined by (21) - (22),*

$$SN_T(\boldsymbol{\beta}_0 \mid \boldsymbol{\beta}_1) = \sum_{t=2}^{T} \sum_{l=t-1}^{2} \ln c_{\tilde{\sigma}_{lt}t, |\tilde{\sigma}_{1t}, \cdots, \tilde{\sigma}_{t-1,t}} + \sum_{t=2}^{T} \ln c_{\tilde{\sigma}_{1t}t} + \sum_{t=1}^{T} s(y_t - f(\boldsymbol{x}_{t-1}, \boldsymbol{\beta}_0)) \tilde{a}_t(\boldsymbol{\beta}_0 \mid \boldsymbol{\beta}_1) > c_1(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1),$$

where

$$\tilde{a}_t(\boldsymbol{\beta}_0 \mid \boldsymbol{\beta}_1) = \ln\left\{\frac{1-p_t[\boldsymbol{x}_{t-1},\boldsymbol{\beta}_0,\boldsymbol{\beta}_1 \mid X]}{p_t[\boldsymbol{x}_{t-1},\boldsymbol{\beta}_0,\boldsymbol{\beta}_1 \mid X]}\right\}, \quad p_t[\boldsymbol{x}_{t-1},\boldsymbol{\beta}_0,\boldsymbol{\beta}_1 \mid X] = P_t[\varepsilon_t \leq f(\boldsymbol{x}_{t-1},\boldsymbol{\beta}_0) - f(\boldsymbol{x}_{t-1},\boldsymbol{\beta}_1) \mid X]$$

and suppose the constant $c_1(\boldsymbol{\beta}_0,\boldsymbol{\beta}_1)$ *satisfies the constraint* $P\left[SN_T(\boldsymbol{\beta}_0 \mid \boldsymbol{\beta}_1) > c_1(\boldsymbol{\beta}_0,\boldsymbol{\beta}_1)\right] = \alpha$ *under* $H(\boldsymbol{\beta}_0)$, *with* $0 < \alpha < 1$. *Then the test that rejects* $H(\boldsymbol{\beta}_0)$ *when*

$$SN_T(\boldsymbol{\beta}_0 \mid \boldsymbol{\beta}_1) > c_1(\boldsymbol{\beta}_0,\boldsymbol{\beta}_1) \tag{25}$$

is most powerful for testing $H(\boldsymbol{\beta}_0)$ *against* $H(\boldsymbol{\beta}_1)$ *among level-*α *tests based on the signs* $\left(s(\tilde{y}_1),\cdots,s(\tilde{y}_T)\right)'$.

Consider a linear function $f(\boldsymbol{x}_{t-1},\boldsymbol{\beta}) = \boldsymbol{\beta}'\boldsymbol{x}_{t-1}$, and assume that under the alternative hypothesis ε_t for $t = 1,\cdots,T$ follows a standard normal distribution (i.e. $\varepsilon_t \sim N(0,1)$). Then the statistic for testing $H(\boldsymbol{\beta}_0)$ against the alternative $H(\boldsymbol{\beta}_1)$ is given by

$$SN_T(\boldsymbol{\beta}_0 \mid \boldsymbol{\beta}_1) = \sum_{t=2}^{T}\sum_{l=t-1}^{2} \ln c_{\tilde{\delta}_{lt}t,|\tilde{\delta}_{1t},\cdots,\tilde{\delta}_{t-1,t}} + \sum_{t=2}^{T} \ln c_{\tilde{\delta}_{1t}t} + \sum_{t=1}^{T} s(y_t - \boldsymbol{\beta}_0'\boldsymbol{x}_{t-1})\tilde{\delta}_t(\boldsymbol{\beta}_0 \mid \boldsymbol{\beta}_1) > c_1(\boldsymbol{\beta}_0,\boldsymbol{\beta}_1),$$

where

$$\tilde{a}_t(\boldsymbol{\beta}_0 \mid \boldsymbol{\beta}_1) = \ln\left\{\frac{\Phi((\boldsymbol{\beta}_1-\boldsymbol{\beta}_0)'\boldsymbol{x}_{t-1})}{1-\Phi((\boldsymbol{\beta}_1-\boldsymbol{\beta}_0)'\boldsymbol{x}_{t-1})}\right\},$$

such that $\Phi(.)$ is the standard normal distribution function. As in Section 3, the distribution of $SN_T(\boldsymbol{\beta}_0 \mid \boldsymbol{\beta}_1)$ can be simulated under the null hypothesis with sufficient number of replications and the relevant critical values can be obtained to any degree of precision.

4 Estimation

In this Section, we first consider the issue of estimating the bivariate copulas in the D-vine decomposition and suggest a sequential estimation strategy for the parameters of the copulas. We then turn our attention to the problem of selecting a class of parametric bivariate copulas. The choice of the latter has an important implication on introducing dependence to the vector of signs.

4.1 Sequential estimation of the D-vine

The calculation of the test statistics in Section 3 requires four bivariate copula evaluations at $T(T-1)/2$ distinct points, leading to a total of $2T(T-1)$ copula evaluations. The estimation of the D-vine is often facilitated with the maximum likelihood (MLE hereafter). However, since the latter requires optimization with respect to at least $2T(T-1)$ copula parameters, sequential estimation procedures are favored

for faster computation times, with the caveat that the increased speed comes at the cost of efficiency. Furthermore, the sequential estimates may be provided as starting points for the simultaneous numerical optimization using MLE [see Czado et al. (2012), Haff (2012) and Dissmann et al. (2013) among others]. We assume that the copulas are specified parametrically, given by an appropriate parameter (vector). More specifically, let $\boldsymbol{\theta}_l = (\theta'_{1,k}, \cdots, \theta'_{T-l,k})'$ be the set of all the parameters to be estimated for tree $\mathcal{T}_l$, $l = 1, \cdots, T-1$ of the D-vine, with $k = l-1$ conditioning variables. Therefore, $\boldsymbol{\theta} = (\boldsymbol{\theta}'_1, \cdots, \boldsymbol{\theta}'_{T-1})'$ is the entire set of the parameters that need to be estimated for the D-vine decomposition. To estimate the parameter vector $\boldsymbol{\theta}$, we follow a sequential estimation strategy proposed by Czado et al. (2012), whereby first, the parameters of the unconditional bivariate copulas are estimated. These parameters are then utilized as means of estimating the parameters of bivariate copulas with a single conditioning variable. The latter are then used to estimate the pair-copulas with two conditioning variables, and so on. This bivariate copula estimation approach is continued sequentially until all parameters are estimated.

In the first step, the marginals are obtained by computing the conditional Bernoulli CDFs (14) using an arbitrary distribution, such as the standard normal distribution considered in Section 3. The second step of the process involves estimating the parameters of the unconditional copula, by fixing the marginals with their aforementioned estimates and maximizing the bivariate likelihood corresponding to each copula in each tree $\mathcal{T}_l$ to obtain $\hat{\boldsymbol{\theta}}_l = (\hat{\theta}'_{1,k}, \cdots, \hat{\theta}'_{T-l,k})$ for $l = 1, \cdots, T-1$ and $k = l-1$. As all the variables are discrete, the log-likelihood function, say, for the unconditional copula $C_{t,t+1}$ for $t = 1, \cdots, T-1$ for the signs $(s(y_{i,t}), s(y_{i,t+1}))$, $i = 1, \cdots, n-1$ is expressed as

$$L(\theta_{t,0}) = \sum_{i=1}^{T-1} \log \left\{ \sum_{\{a_1,a_2\} \in \{-,+\}^2} (-1)^{a_j} C_{t,t+1} \left(F_t(s_{i,t}^{a_1} \mid X; \hat{\boldsymbol{\beta}}_{\mathbf{1}}), F_{t+1}(s_{i,t+1}^{a_2} \mid X; \hat{\boldsymbol{\beta}}_{\mathbf{1}}); \theta_{t,0} \right) \right\}.$$

The estimate of the copula parameter, $\hat{\theta}_{t,0}$ for $t = 1, \cdots, T-1$, is then obtained as follows

$$\hat{\theta}_{t,0} = \arg\max_{\theta_{t,0}} L(\theta_{t,0}),$$

which under regularity conditions solves

$$\frac{\partial L(\theta_{t,0})}{\partial \theta_{t,0}} = 0.$$

Let us illustrate this process with an example: once the marginals are obtained, the next step involves estimating the parameters $\theta_{t,0}$ for $t = 1, \cdots, T-1$ of the unconditional copulas. Next, we are interested in estimating $\theta_{t,1}$ for $t = 1, \cdots, T-2$. Define

$$\hat{u}_{t|t+1} = F_{t|t+1} \left(s_t \mid s_{t+1}, X; \hat{\theta}_{t,0} \right),$$

and

$$\hat{v}_{t+2|t+1} = F_{t+2|t+1}\left(s_{t+2} \mid s_{t+1}, X; \hat{\theta}_{t+1,0}\right),$$

for $t = 1, \cdots, T-2$. The data $\hat{u}_{t|t+1}$ and $\hat{v}_{t+2|t+1}$ is then used to estimate the parameters $\theta_{t,1}$ for $t = 1, \cdots, T-2$, denoted by $\hat{\theta}_{t,1}$. This procedure is repeated sequentially until all parameters are estimated. Haff et al. (2010) show that under regularity conditions, the sequential estimates are asymptotically normal; however, as noted earlier their asymptotic covariance is "intractable" and the faster computation time comes at the cost of efficiency. Therefore, the sequential estimates can be utilized as the starting values of the high-dimensional MLE.

Another approach for estimating one-parameter pair-copulas in the sequential estimation procedure for copula families with a known relationship to Kendall's τ consists of inverting the empirical Kendall's τ based on, say, $\hat{u}_t$ and $\hat{u}_{t+1}$ for $t = 1, \cdots, T-1$ for the edges of the first tree. However, we provide a caveat that the Kendall's τ of discrete data does not correspond to the Kendall's τ of the bivariate copulas [see Denuit and Lambert (2005)]. Denuit and Lambert (2005) show that by continuous extension of the discrete variables with a perturbation with values in $[0, 1]$, the continuous features of Kendall's τ are adaptable to discrete data. In other words

$$\tau(s^*(y_t), s^*(y_{t+1})) = 4 \int \int_{[0,1]^2} C^*_{t,t+1}\left(\hat{u}_t, \hat{u}_{t+1}\right) dC^*_{t,t+1}\left(\hat{u}_t, \hat{u}_{t+1}\right) - 1$$

for $t = 1, \cdots, T-1$, such that $\hat{u}_t = F\left(s^*_t \mid X; \hat{\boldsymbol{\beta}}_1\right)$,

$$s^*(y_t) = s(y_t) + U - 1$$

where U is a continuous random variable in $[0, 1]$. A natural choice for U is the uniform distribution.

4.2 Selection of the copula family

Many different classes of parametric bivariate copulas have extensively been studied and reviewed by Joe (2014). These include the Archimedean, elliptical, extreme value or max-id copula families that can specify the dependence structure of the vector of signs. As the dependence is introduced by the copula family, the type and the degree of the dependence between the signs depends on the choice of the copula. The literature surrounding the goodness-of-fit of copulas is extensive and has been analyzed by Genest et al. (2006), Genest et al. (2009), and Berg (2009), among many others. Genest et al. (2009) categorize goodness-of-fit tests into three broad categories: procedures for testing particular dependence structures such as Gaussian or Clayton family; procedures that may be used for any classes of copulas, but require a

strategic choice for their implementation; and finally, the so-called "blanket tests"that apply to all classes of copulas and require no strategic choice for their use. A simple procedure proposed by Joe (1997) involves specifying the Akaike information critetion (AIC) to different copulas and using it as a copula selection criterion, which is particularly attractive as it allows for the automation of the copula selection process [see. Czado et al. (2012)]. The AIC specified to the copulas of, say, the first tree of the D-vine, can be expressed as follows

$$AIC = -2\sum_{i=1}^{t} \log c_{t,t+1}(\hat{u}_{i,t}, \hat{u}_{i,t+1}; \hat{\theta}_{1,k}) + 2l$$

for $t = 1, \cdots, T-1$ and $k = 1, \cdots, T-1$, and where l is the number of parameters $\theta_{1,k}$. Panagiotelis et al. (2012) suggest that while dependence structures such as tail dependence are weak in discrete data, the choice of the copulas could still have a significant effect on the joint pmf of the signs. They considered Gaussian, Clayton and Gumbel copulas in constructing the D-vines for Bernoulli margins by keeping the marginal probabilities and dependence constant, and have found that in the case where the probabilities of zero marginals and joint probabilities in the data is high, preference goes to the use of the Gumbel copula over the other two alternatives.

Within the context of our work, the mediangale assumption (2) implies that the signs $s(y_1), \cdots, s(y_T)$ conditional on X, and in turn $\hat{u}_1, \cdots, \hat{u}_T$ exhibit serial nonlinear dependence. Thus, the issue with specifying a copula-based model for the signs is that the distribution of the signs must imply an identity correlation matrix, where independence is only sufficient for uncorrelatedness. The literature generally deals with this issue by capturing the serial nonlinear dependence by imposing the identity correlation matrix on a multivariate Student's t distribution. Henceforth, we consider the "jointly symmetric" copulas proposed by Oh and Patton (2016), where the latter can be constructed with any given (possibly asymmetric) copula family. In addition, when they are combined with symmetric marginals, they ensure an identity correlation matrix. A "jointly symmetric" copula is defined as follows

Definition 2 *The n dimensional copula C^{JS}, is jointly symmetric:*

$$C^{JS}(u_1, \cdots, u_n) = \frac{1}{2^n} \sum_{k_1=0}^{2} \cdots \sum_{k_n=0}^{2} (-1)^R C(\tilde{u}_1, \cdots, \tilde{u}_i, \cdots, \tilde{u}_n),$$

$$\textit{where} \quad R = \sum_{i=1}^{n} \mathbf{1}\{k_i = 2\}, \quad \textit{and} \quad \tilde{u}_i = \begin{cases} 1, & k_i = 0 \\ u_i, & k_i = 1 \\ 1 - u_i, & k_i = 2 \end{cases}.$$

The general idea is that the average of mirror image rotations of a possibly asymmetric copula along each axis generates a jointly symmetric copula [see Oh and Patton (2016)]. For instance, the marginals can be assumed to possess standard normal distributions, while the nonlinear dependency is modeled using jointly symmetric copulas.

Therefore, using the jointly symmetric copula family, the bivariate copulas can be evaluated by considering

$$C^{JS}\left(\phi_t(s_t \mid X;\hat{\boldsymbol{\beta}}_{\mathbf{1}}),\phi_{t+1}(s_{t+1} \mid X;\hat{\boldsymbol{\beta}}_{\mathbf{1}})\right)=\frac{1}{4}\sum_{k_1=0}^{2}\sum_{k_2=0}^{2}(-1)^R C(\phi_t(s_t \mid X;\hat{\boldsymbol{\beta}}_{\mathbf{1}}),\phi_{t+1}(s_{t+1} \mid X;\hat{\boldsymbol{\beta}}_{\mathbf{1}}))$$

$$\text{where}\quad R=\sum_{i=1}^{n}\mathbf{1}\{k_i=2\},\quad \text{and}\quad \tilde{u}_i=\begin{cases}1, & k_i=0\\ \phi_i(.), & k_i=1\\ 1-\phi_i(.), & k_i=2\end{cases},\quad \forall i\in\{t,t+1\}.$$

for $t=1,\cdots,T-1$.

4.3 Truncated D-vines

Following Joe (2014), we refer to a D-vine as a p-truncated D-vine, if the copulas in the trees $\mathcal{T}_{p+1},\cdots,\mathcal{T}_T$ are $C^{\perp}$, where by definition

$$C^{\perp}(u_1,\cdots,u_n)=\prod_{t=1}^{T}u_t,\quad \text{with}\quad (U_1,\cdots,U_T)\sim U(0,1),$$

implying $U_1\perp U_2\perp\cdots\perp U_T$. The POS-based tests constructed using the Markov assumption of order one can be regarded as a special case of the PCC-POS based tests, whereby the former can be constructed by a 1-truncated D-vine, which only depends on $C_{12},C_{23},\cdots,C_{T-1,T}$, or rather $C_{12},C_{\sigma_{13}3},\cdots,,C_{\sigma_{1T}T}$ using a vine array representation, given the following array

$$\begin{bmatrix}
- & 12 & 23 & 34 & \cdots & T-1,T\\
 & - & 13\mid 2^{\perp} & 24\mid 3^{\perp} & \cdots & T-2,T\mid T-1^{\perp}\\
 & & \ddots & \cdots & \cdots & \vdots\\
 & & & - & 1,T-1\mid \underline{\mathrm{S}}^{\perp}_{\setminus j-} & 2,T\mid \underline{\mathrm{S}}^{\perp}_{\setminus j+}\\
 & & & & - & 1,T\mid \underline{\mathrm{S}}^{\perp}_{\setminus j}\\
 & & & & & -
\end{bmatrix},\quad A=\begin{bmatrix}
1 & 1 & 2 & 3 & \cdots & T-1\\
 & 2 & 1^{\perp} & 2^{\perp} & \cdots & T-2^{\perp}\\
 & & \ddots & \cdots & \cdots & \vdots\\
 & & & T-2 & 1^{\perp} & 2^{\perp}\\
 & & & & T-1 & 1^{\perp}\\
 & & & & & T
\end{bmatrix}$$

Similarly, a 2-truncated D-vine, depends on the copulas $C_{12}, C_{23}, \cdots, C_{T-1,T}$ and $C_{13|2}, C_{24|3}, \cdots, C_{T-2,T|T-1}$ or $C_{\sigma_{1t}t}$ for $t = 2, \cdots, T$ and $C_{\sigma_{2t}t|\sigma_{1t}}$ for $t = 3, \cdots, T$ using the vine array representation. Therefore, for a p-truncated D-vine, (24) is modified to

$$P_{t|1:t-1}[s(\tilde{y}_t) = \tilde{s}_t \mid s(\tilde{y}_1) = \tilde{s}_1, \cdots, s(\tilde{y}_{t-1}) = \tilde{s}_{t-1}, X] = P_{t|1:p}[s(\tilde{y}_t) = \tilde{s}_t \mid s(\tilde{y}_1) = \tilde{s}_1, \cdots, s(\tilde{y}_p) = \tilde{s}_p, X]$$
$$= \left\{ \prod_{l=p\wedge(t-1)}^{2} c_{\tilde{\sigma}_{lt}t,|\tilde{\sigma}_{1t},\cdots,\tilde{\sigma}_{t-1,t}} \right\} \times c_{\tilde{\sigma}_{lt}t} \times P_t[s(\tilde{y}_t) = \tilde{s}_t \mid X], \tag{26}$$

for $t - 1 \geq p$.

Corollary 2 *Let $\tilde{A} = (\tilde{\sigma}_{lt})_{1\leq l\leq t\leq T}$ be a D-vine array for the signs $s(\tilde{y}_1), \cdots, s(\tilde{y}_T)$, where the signs $\{s(\tilde{y}_t)\}_{t=0}^{\infty}$ follow a Markov process of order p. Under assumptions (19) and (2), let $H(\boldsymbol{\beta}_0)$ and $H(\boldsymbol{\beta}_1)$ be defined by (21) - (22),*

$$SN_T(\boldsymbol{\beta}_0 \mid \boldsymbol{\beta}_1) = \sum_{t=2}^{T} \sum_{l=p\wedge(t-1)}^{2} \ln c_{\tilde{\sigma}_{lt}t,|\tilde{\sigma}_{1t},\cdots,\tilde{\sigma}_{t-1,t}} + \sum_{t=2}^{T} \ln c_{\tilde{\sigma}_{1t}t} + \sum_{t=1}^{T} s(y_t - f(\boldsymbol{x}_{t-1}, \boldsymbol{\beta}_0))\tilde{a}_t(\boldsymbol{\beta}_0 \mid \boldsymbol{\beta}_1) > c_1(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1),$$

where

$$\tilde{a}_t(\boldsymbol{\beta}_0 \mid \boldsymbol{\beta}_1) = \ln\left\{\frac{1 - p_t[\boldsymbol{x}_{t-1}, \boldsymbol{\beta}_0, \boldsymbol{\beta}_1 \mid X]}{p_t[\boldsymbol{x}_{t-1}, \boldsymbol{\beta}_0, \boldsymbol{\beta}_1 \mid X]}\right\}, \quad p_t[\boldsymbol{x}_{t-1}, \boldsymbol{\beta}_0, \boldsymbol{\beta}_1 \mid X] = P_t[\varepsilon_t \leq f(\boldsymbol{x}_{t-1}, \boldsymbol{\beta}_0) - f(\boldsymbol{x}_{t-1}, \boldsymbol{\beta}_1) \mid X]$$

and suppose the constant $c_1(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1)$ satisfies the constraint $P\left[SN_T(\boldsymbol{\beta}_0 \mid \boldsymbol{\beta}_1) > c_1(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1)\right] = \alpha$ under $H(\boldsymbol{\beta}_0)$, with $0 < \alpha < 1$. Then the test that rejects $H(\boldsymbol{\beta}_0)$ when

$$SN_T(\boldsymbol{\beta}_0 \mid \boldsymbol{\beta}_1) > c_1(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1) \tag{27}$$

is most powerful for testing $H(\boldsymbol{\beta}_0)$ against $H(\boldsymbol{\beta}_1)$ among level-α tests based on the signs $\left(s(\tilde{y}_1), \cdots, s(\tilde{y}_T)\right)'$.

5 Choice of the optimal alternative hypothesis

In this Section, we follow Dufour and Taamouti (2010) by first showing the analytical derivation of the power envelope function of the PCC-POS-based tests. We then suggest using simulations as means of approximating the said function, by showing the difficulty of inverting the latter to find the optimal alternative. Thereafter, we propose an adaptive approach based on the split-sample technique to choose an alternative which has a power function close to that of the power envelope.

5.1 Power envelope of PCC-POS tests

Point-optimal tests trace out the power envelope (i.e. the maximum attainable power) for any given testing problem [see King (1987)]. However, in practice the alternative hypothesis $\boldsymbol{\beta}_1$ is unknown and a problem consists of finding an approximation for it, such that the power function is maximized and is close to that of the power envelope. Following Dufour and Taamouti (2010) and Dufour and Jasiak (2001), we propose an adaptive approach based on the split-sample technique to choose an alternative $\boldsymbol{\beta}_1$ that yields the greatest power function and makes size control easier [see Dufour and Jasiak (2001) and Dufour and Taamouti (2010) for an overview]. We follow Dufour and Taamouti (2010) by presenting the analytical derivation of the power envelope of the PCC-POS tests for predictive regressions, which can be purposed as a benchmark for comparing the power functions of the PCC-POS tests for different sample splits.

We have shown in Section 3.2 that the PCC-POS tests are a function of $\boldsymbol{\beta}_1$. In other words,

$$SN_T(\boldsymbol{\beta}_0 \mid \boldsymbol{\beta}_1) = \sum_{t=2}^{T}\sum_{l=t-1}^{2} \ln c_{\tilde{\sigma}_{lt}t|\tilde{\sigma}_{1t},\cdots,\tilde{\sigma}_{t-1,t}} + \sum_{t=2}^{T} \ln c_{\tilde{\sigma}_{1t}t} + \sum_{t=1}^{T} \ln\left\{\frac{1-p_t[\boldsymbol{x}_{t-1},\boldsymbol{\beta}_0,\boldsymbol{\beta}_1 \mid X]}{p_t[\boldsymbol{x}_{t-1},\boldsymbol{\beta}_0,\boldsymbol{\beta}_1 \mid X]}\right\} s(y_t - f(\boldsymbol{x}_{t-1},\boldsymbol{\beta}_0)).$$

which in turn implies that its power function, say $\Pi(\boldsymbol{\beta}_0,\boldsymbol{\beta}_1)$, is also a function of $\boldsymbol{\beta}_1$

$$\Pi(\boldsymbol{\beta}_0,\boldsymbol{\beta}_1) = P[SN_T(\boldsymbol{\beta}_0 \mid \boldsymbol{\beta}_1) > c_1(\boldsymbol{\beta}_0,\boldsymbol{\beta}_1) \mid H(\boldsymbol{\beta}_1)]$$

where $c_1(\boldsymbol{\beta}_0,\boldsymbol{\beta}_1)$ is the smallest constant that satisfies $P[SN_T(\boldsymbol{\beta}_0 \mid \boldsymbol{\beta}_1) > c_1(\boldsymbol{\beta}_0,\boldsymbol{\beta}_1) \mid H(\boldsymbol{\beta}_0)] \leq \alpha$, and where α is an arbitrary significance level. Theorem 1 provides the theoretical results for the power function of the PCC-POS tests.

Theorem 1 *Under assumption (2) and and given model (19), and further under the condition that $s(\tilde{y}_1),\cdots,s(\tilde{y}_T)$ conditional on X follow a Regularity Markov Type process (RMT hereafter), the power function of $SN_T(\boldsymbol{\beta}_0 \mid \boldsymbol{\beta}_1)$ is given by*

$$\Pi(\boldsymbol{\beta}_0,\boldsymbol{\beta}_1) = P[SN_T(\boldsymbol{\beta}_0 \mid \boldsymbol{\beta}_1) > c_1(\boldsymbol{\beta}_0,\boldsymbol{\beta}_1) \mid X] = \frac{1}{2} + \frac{1}{\pi}\int_0^{\infty} \frac{\mathrm{Im}\{\exp(iuc_1(\boldsymbol{\beta}_0,\boldsymbol{\beta}_1))\phi_{SN_T}(u)\}}{u} du$$

$\forall u \in \mathbb{R}$, $i = \sqrt{-1}$, and with $\mathrm{Im}\{z\}$ denoting the imaginary part of the complex number z. $\phi_{SN_T}(u)$ is given by

$$\phi_{SN_T}(u) = \prod_{t=1}^{T}\left(\mathbb{E}_X\left[\exp\left(iu\left\{R_{t,t-1} + \ln\left\{\frac{1-p_t[\boldsymbol{x}_{t-1},\boldsymbol{\beta}_0,\boldsymbol{\beta}_1 \mid X]}{p_t[\boldsymbol{x}_{t-1},\boldsymbol{\beta}_0,\boldsymbol{\beta}_1 \mid X]}\right\} s(\tilde{y}_t)\right\}\right)\right] + \rho_t(u)\right),$$

where $R_{1,0} = 0$, $R_{t,t-1} = \sum_{l=t-1}^{2} \ln c_{\tilde{\sigma}_{lt}t|\tilde{\sigma}_{1t},\cdots,\tilde{\sigma}_{t-1,t}} + \ln c_{\tilde{\sigma}_{1t}t}$ *for* $t = 2, \cdots, T$, *such that for D-vine-array* $\tilde{A} = (\tilde{\sigma}_{lt})_{1\leq l\leq t\leq T}$, $l = 2, \cdots, T-1$ *is the row with tree* $\mathcal{T}_l$, *and column* t *has the permutation* $\underset{\sim}{\tilde{\sigma}}_{t-1} = (\tilde{\sigma}_{1t}, \cdots, \tilde{\sigma}_{t-1,t})$ *of the previously added variables,* $p_t[\boldsymbol{x}_{t-1}, \boldsymbol{\beta}_0, \boldsymbol{\beta}_1 \mid X] = P_t[\varepsilon_t \leq f(\boldsymbol{x}_{t-1}, \boldsymbol{\beta}_0) - f(\boldsymbol{x}_{t-1}, \boldsymbol{\beta}_1) \mid X]$, $\underset{\sim}{\tilde{S}}_{t-1} = s(y_{t-1} - f(\boldsymbol{x}_{t-2}, \boldsymbol{\beta}_0)), \cdots, s(y_1 - f(\boldsymbol{x}_0, \boldsymbol{\beta}_0))$. *Finally,* $c_1(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1)$ *is the smallest constant that satisfies* $P[SN_T(\boldsymbol{\beta}_0 \mid \boldsymbol{\beta}_1) > c_1(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1) \mid H(\boldsymbol{\beta}_0)] \leq \alpha$, *where* α *is an arbitrary significance level.*

Under the assumption that the signs follow an RMT-process, $\rho_t(u)$ can be estimated using the results from Theorem 2 of Heinrich (1982). Given that point-optimal tests are optimal at a specific point in the alternative parameter space, the power envelope of the PCC-POS tests, say $\bar{\Pi}(\boldsymbol{\beta}_1)$, is obtained for values of $\boldsymbol{\beta}$, such that $\{\boldsymbol{\beta} : \boldsymbol{\beta} = \boldsymbol{\beta}_1, \forall \boldsymbol{\beta}_1 \in \mathbb{R}^{(k+1)}\}$. Finding values of $\boldsymbol{\beta}_1$ for a PCC-POS test at level α, with a power function that is close to the power envelope can be achieved by inverting the power envelope function. However, in a much simpler case of POS tests for i.n.i.d data, Dufour and Taamouti (2010) show that the inversion of the power function is not a straightforward task and obtaining an exact solution is not feasible. Therefore, simulations are used as means of approximating the power envelope function and finding the optimal alternative for the PCC-POS test.

5.2 Split-sample technique for choosing the optimal alternative

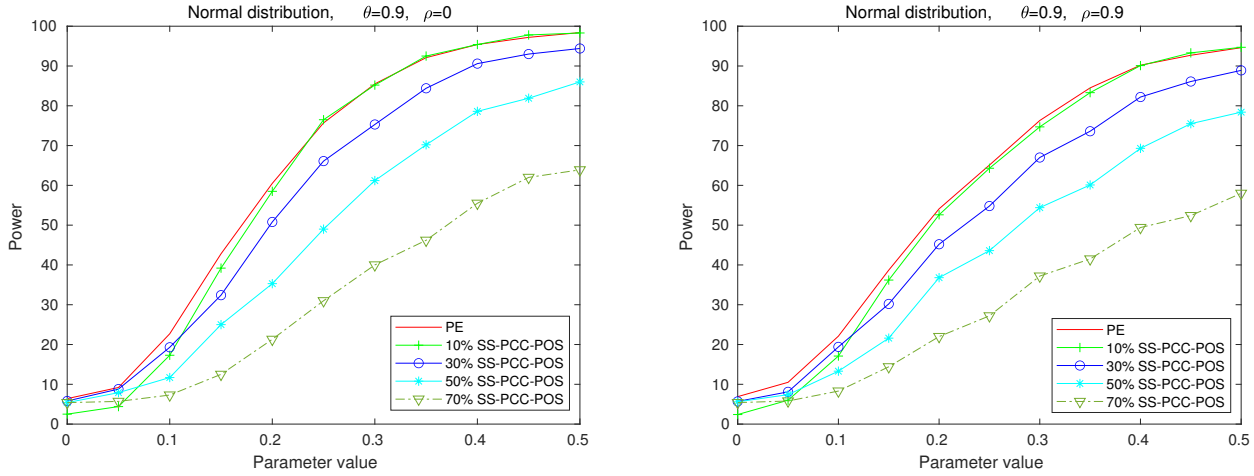
As noted earlier, the power function of the PCC-POS test statistic depends on the alternative $\boldsymbol{\beta}_1$, which in practice is unknown and needs to be approximated. To make size control easier and to choose an approximation to $\boldsymbol{\beta}_1$ such that the power function of the test statistic is close to that of the power envelope, we follow Dufour and Taamouti (2010) by proposing an adaptive approach based on the split-sample technique for choosing the alternative. For an extensive review of adaptive statistical methods, we refer the reader to O'Gorman (2004). Furthermore, the application of the split-sample technique in parametric settings can be studied by consulting Dufour and Taamouti (2003) and Dufour et al. (2008).

The split-sample technique involves splitting a sample of size T into two independent subsamples, say T_1 and T_2, such that $T = T_1 + T_2$. The first subsample is then used to estimate the alternative $\boldsymbol{\beta}_1$, while the other is purposed for computing the PCC-POS test statistic. Assuming that $f(\boldsymbol{x}_{t-1}, \boldsymbol{\beta}) = \boldsymbol{x}'_{t-1}\boldsymbol{\beta}$, the alternative $\boldsymbol{\beta}_1$ can be estimated using OLS

$$\hat{\boldsymbol{\beta}}_{(1)} = (X'_{(1)}X_{(1)})^{-1}X'_{(1)}y_{(1)}.$$

We provide a caveat that the OLS estimator is sensitive to extreme outliers, which motivates the use of robust estimators [see. Maronna et al. (2019) for a review of robust estimators]. Using $\hat{\boldsymbol{\beta}}_{(1)}$ and the

Figure 2: Power comparisons: different split-samples. Normal error distributions with different values of ρ in (33) and $\theta = 0.9$ in (32)



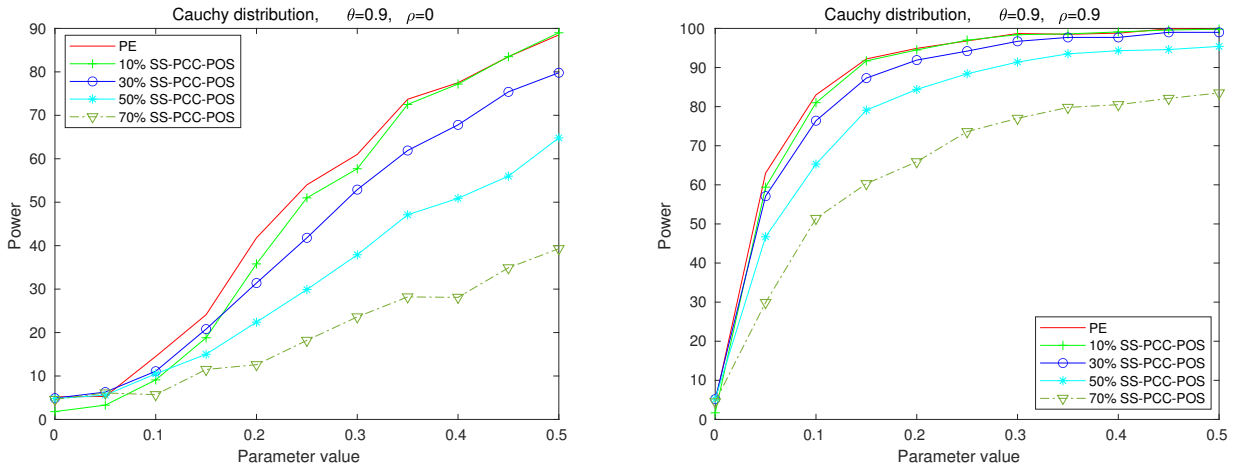
Note: These figures compare the power envelope the PCC-POS test statistic using different split-samples: 10%, 30%, 50%, 70%. "PE"refers to the power envelope of the PCC-POS test.

observations in the second independent subsample, we compute the test-statistic as follows

$$SN_T(\boldsymbol{\beta}_0 \mid \boldsymbol{\beta}_{(1)}) = \sum_{t=T_1+2}^{T} \sum_{l=t-1}^{2} \ln c_{\tilde{\sigma}_{lt} t \mid \tilde{\sigma}_{(T_1+1)t}, \cdots, \tilde{\sigma}_{t-1,t}} + \sum_{t=T_1+2}^{T} \ln c_{\tilde{\sigma}_{(T_1+1)t} t} + \sum_{t=T_1+1}^{T} \ln \left\{ \frac{1 - p_t[\boldsymbol{x}_{t-1}, \boldsymbol{\beta}_0, \boldsymbol{\beta}_{(1)} \mid X]}{p_t[\boldsymbol{x}_{t-1}, \boldsymbol{\beta}_0, \boldsymbol{\beta}_{(1)} \mid X]} \right\} s(y_t - \boldsymbol{x}'_{t-1}\boldsymbol{\beta}_0).$$

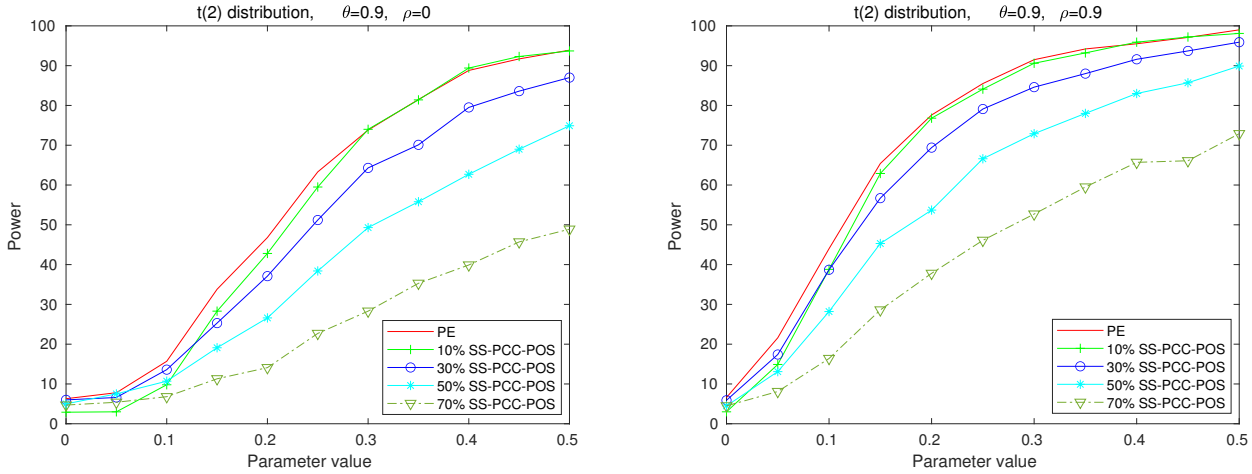
where for $t = T_1 + 2, \cdots, T$ and D-vine-array $\tilde{A} = (\tilde{\sigma}_{lt})_{1 \le l \le t \le n}$, $l = T_1 + 1, \cdots, T - 1$ is the row with tree $\mathcal{T}_l$, and column t has the permutation $\tilde{\underline{\sigma}}_{t-1} = (\tilde{\sigma}_{(T_1+1)t}, \cdots, \tilde{\sigma}_{t-1,t})$ of the previously added variables, $p_t[\boldsymbol{x}_{t-1}, \boldsymbol{\beta}_0, \boldsymbol{\beta}_{(1)} \mid X] = P_t[\varepsilon_t \le \boldsymbol{x}'_{t-1}(\boldsymbol{\beta}_0 - \boldsymbol{\beta}_1) \mid X]$, and $\tilde{\mathbb{S}}_{t-1} = s(y_{t-1} - \boldsymbol{x}'_{t-2}\boldsymbol{\beta}_0), \cdots, s(y_{T_1+2} - \boldsymbol{x}'_{T_1+1}\boldsymbol{\beta}_0)$. The choices for the subsamples T_1 and T_2 can be arbitrary. However, our simulations show that the proportion of the observations retained for estimating the alternative and in turn for computing the PCC-POS test statistic has an impact on the power of the test. We find that the power function of the split-sample PCC-POS test (SS-PCC-POS test hereafter) is closest to that of the power envelope, when a relatively small number of observations is retained for estimating the alternative, with the rest used for computing the test statistic - findings that are in line with Dufour and Taamouti (2010). Specifically, by considering all the DGPs in our simulations study, we find that the subsamples T_1 and T_2 must in turn contain roughly 10% and 90% of the observations in the entire sample respectively.

Figure 3: Power comparisons: different split-samples. Cauchy error distributions with different values of ρ in (33) and $\theta = 0.9$ in (32)



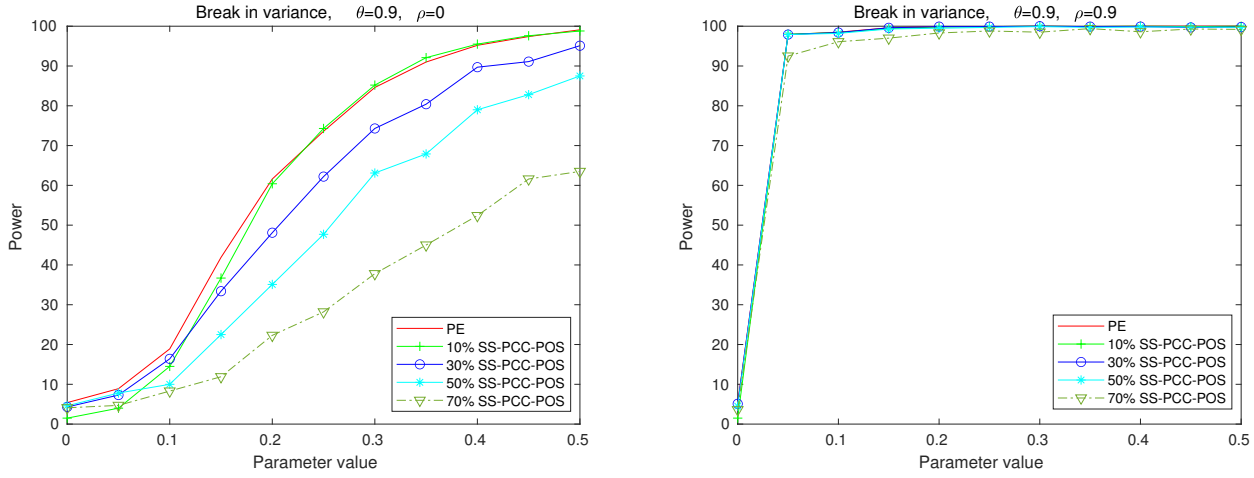
Note: These figures compare the power envelope the PCC-POS test statistic using different split-samples: 10%, 30%, 50%, 70%. "PE"refers to the power envelope of the PCC-POS test.

Figure 4: Power comparisons: different split-samples. Student's t error distributions with 2 degrees of freedom [i.e $t(2)$] with different values of ρ in (33) and $\theta = 0.9$ in (32)



Note: These figures compare the power envelope the PCC-POS test statistic using different split-samples: 10%, 30%, 50%, 70%. "PE"refers to the power envelope of the PCC-POS test.

Figure 5: Power comparisons: different split-samples. Normal error distributions with break in variance, with different values of ρ in (33) and $\theta = 0.9$ in (32)



Note: These figures compare the power envelope the PCC-POS test statistic using different split-samples: 10%, 30%, 50%, 70%. "PE"refers to the power envelope of the PCC-POS test.

6 PCC-POS confidence regions

In this Section, we lay out the theoretical framework for building confidence regions for a vector (sub-vector) of the unknown parameters $\boldsymbol{\beta}$, say $C_{\boldsymbol{\beta}}(\boldsymbol{\alpha})$, at a given significance level α, using the proposed PCC-POS tests. Consider model (19) such that $f(\boldsymbol{x}_{t-1}, \boldsymbol{\beta}) = \boldsymbol{x}'_{t-1}\boldsymbol{\beta}$. Suppose we wish to test the null hypothesis (21) against the alternative hypothesis (22). Formally, this implies finding all the values of $\boldsymbol{\beta}_0 \in \mathbb{R}^k$ such that

$$SN_T(\boldsymbol{\beta}_0 \mid \boldsymbol{\beta}_1) = \sum_{t=2}^{T} \sum_{l=t-1}^{2} \ln c_{\tilde{\sigma}_{lt}t,|\tilde{\sigma}_{1t},\cdots,\tilde{\sigma}_{t-1,t}} + \sum_{t=2}^{T} \ln c_{\tilde{\sigma}_{1t}t} + \sum_{t=1}^{T} s(y_t - \boldsymbol{\beta}'_0 - \boldsymbol{x}_{t-1})\tilde{a}_t(\boldsymbol{\beta}_0 \mid \boldsymbol{\beta}_1) < c_1(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1). \tag{28}$$

where the critical value is given by the smallest constant $c(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1)$ such that

$$P\left[SN_T(\boldsymbol{\beta}_0 | \boldsymbol{\beta}_1) > c(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1) \mid \boldsymbol{\beta} = \boldsymbol{\beta}_0\right] \leq \alpha$$

Thus, the confidence region of the vector of parameters $\boldsymbol{\beta}$ can be defined as follows:

$$C_{\boldsymbol{\beta}}(\alpha) = \{\boldsymbol{\beta}_0 : SN_T(\boldsymbol{\beta}_0 | \boldsymbol{\beta}_1) < c(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1) | P[SN_T(\boldsymbol{\beta}_0 | \boldsymbol{\beta}_1) > c(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1) \mid \boldsymbol{\beta} = \boldsymbol{\beta}_0] \leq \alpha\}.$$

Given the confidence region $C_{\boldsymbol{\beta}}(\alpha)$, confidence intervals for the components and sub-vectors of vector $\boldsymbol{\beta}$ can be derived using the projection techniques [see Dufour and Taamouti (2010) and Coudin and Dufour (2009)]. Confidence sets in the form of transformations T of $\boldsymbol{\beta} \in \mathbb{R}^m$, for $m \leq (k+1)$, say $T(C_{\boldsymbol{\beta}}(\alpha))$, can easily be found using these techniques. Since, for any set $C_{\boldsymbol{\beta}}(\alpha)$

$$\boldsymbol{\beta} \in C_{\boldsymbol{\beta}}(\alpha) \implies T(\boldsymbol{\beta}) \in T(C_{\boldsymbol{\beta}}(\alpha)), \tag{29}$$

we have

$$P[\boldsymbol{\beta} \in C_{\boldsymbol{\beta}}(\alpha)] \geq 1-\alpha \implies P[T(\boldsymbol{\beta}) \in T(C_{\boldsymbol{\beta}}(\alpha))] \geq 1-\alpha \tag{30}$$

where

$$T(C_{\boldsymbol{\beta}}(\alpha)) = \{\delta \in \mathbb{R}^m : \exists \boldsymbol{\beta} \in C_{\boldsymbol{\beta}}(\alpha), T(\boldsymbol{\beta}) = \delta\}.$$

From (29) and (30) , the set $T(C_{\boldsymbol{\beta}}(\alpha))$ is a conservative confidence set for $T(\boldsymbol{\beta})$ with level $1-\alpha$. If $T(\boldsymbol{\beta})$ is a scalar, then we have

$$P[\inf\{T(\boldsymbol{\beta}_0), \quad \text{for} \quad \boldsymbol{\beta}_0 \in C_{\boldsymbol{\beta}}(\alpha)\} \leq T(\boldsymbol{\beta}) \leq \sup\{T(\boldsymbol{\beta}_0), \quad \text{for} \quad \boldsymbol{\beta}_0 \in C_{\boldsymbol{\beta}}(\alpha)\}] > 1-\alpha.$$

7 Monte Carlo study

In this Section, we assess the performance of the proposed 10% SS-PCC-POS tests (in terms of size control and power) by comparing it to other tests that are intended to be robust against non-standard distributions and heteroskedasticity of unknown form. We consider DGPs under different distributional assumptions and heteroskedasticities. For each DGP, we further consider different correlation coefficients between the errors of the predictive regression and the disturbances of the regressors. In the first subsection, the DGPs are formally introduced, and in the following subsection, the performance of the proposed 10% SS-PCC-POS tests are compared to that of the other tests considered in our study.

7.1 Simulation setup

We assess the performance of the proposed 10% SS-PCC-POS tests in terms of size and power, by considering various DGPs with different symmetric and asymmetric distributions and forms of heteroskedasticity. The DGPs under consideration are supposed to mimic different scenarios that are often encountered in practical settings. The performance of the 10% SS-PCC-POS tests is compared to that of a few other

tests, by considering the following linear predictive regression model

$$y_t = \beta x_{t-1} + \varepsilon_t \tag{31}$$

where β is an unknown parameter and

$$x_t = \theta x_{t-1} + u_t \tag{32}$$

where $\theta = 0.9$ and

$$u_t = \rho\varepsilon_t + w_t\sqrt{1-\rho^2} \tag{33}$$

for $\rho = 0, 0.1, 0.5, 0.9$ and ε_t and w_t are assumed to be independent. The initial value of x is given by: $x_0 = \frac{w_0}{\sqrt{1-\theta^2}}$ and w_t are generated from $N(0,1)$. The errors ε_t are i.n.i.d and are categorized by two groups in our simulation study. In the first group, we consider DGPs where the errors ε_t possess symmetric and asymmetric distributions:

1. normal distribution: $\varepsilon_t \sim N(0,1)$;

2. Cauchy distribution: $\varepsilon_t \sim Cauchy$;

3. Student's t distribution with two degrees of freedom: $\varepsilon_t \sim t(2)$;

4. Mixture of Cauchy and normal distributions: $\varepsilon_t \sim \mid \varepsilon_t^C \mid -(1-s_t) \mid \varepsilon_t^N \mid$, where ε_t^C follows Cauchy distribution, ε_t^N follows $N(0,1)$ distribution, and

$$P(s_t = 1) = P(s_t = 0) = \frac{1}{2}$$

The second group of DGPs represents different forms of heteroskedasticity:

5. break in variance:

$$\varepsilon_t \sim \begin{cases} N(0,1) & \text{for } t \neq 25 \\ \sqrt{1000}N(0,1) & \text{for } t = 25 \end{cases};$$

6. GARCH(1, 1) plus jump variance:

$$\sigma_\varepsilon^2(t) = 0.00037 + 0.0888\varepsilon_{t-1}^2 + 0.9024\sigma_\varepsilon^2(t-1)\,,$$

$$\varepsilon_t \sim \begin{cases} N(0, \sigma_\varepsilon^2(t)) & \text{for } t \neq 25 \\ 50N(0, \sigma_\varepsilon^2(t)) & \text{for } t = 25 \end{cases};$$

We consider the problem of testing the null hypothesis $H_0 : \beta = 0$. Our Monte Carlo simulations compare the size and power of the 10%-PCC-POS test to those of t-test, t-test based on White (1980) variance-correction (WT-test hereafter), and the sign-based test proposed by Campbell and Dufour (1995). Due to computational constraints, we perform only $M_1 = 999$ simulations to evaluate the probability distribution of the 10% SS-PCC-POS test statistic and $M_2 = 1,000$ iterations for approximating the power functions of the proposed PCC-POS test and other tests. In all simulations, we consider a sample size of $T = 50$. As the sign-based statistic of Campbell and Dufour (1995) has a discrete distribution, it is not possible to obtain test with a precise size of 5%; therefore, the size of the test is 5.95% for $T = 50$.

7.2 Simulation results

The results of the Monte Carlo study corresponding to DGPs described in Section 7.1 are presented in figures 6-11. These figures compare the performance of the 10% SS-PCC-POS test in terms of size and power, to those of the t-test, t-test based on White's (1980) variance-correction, as well as the sign-based procedure proposed by Dufour and Taamouti (2010). The results are described in detail below.

First, figure 6 considers the case where the error terms ε_t are normally distributed. At first glance, we note that all tests control size. Evidently, our test is outperformed by the t-test, as well as the t-test based on White's (1980) variance-correction. The former is expected, since for normally distributed error terms, the t-test is the most powerful test. However, the 10% SS-PCC-POS test outperforms the sign-based procedure proposed by Campbell and Dufour (1995) [CD (1995) hereafter]. Furthermore, changing the correlation coefficient ρ does not seem to lead to visually significant differences in the performance of the tests.

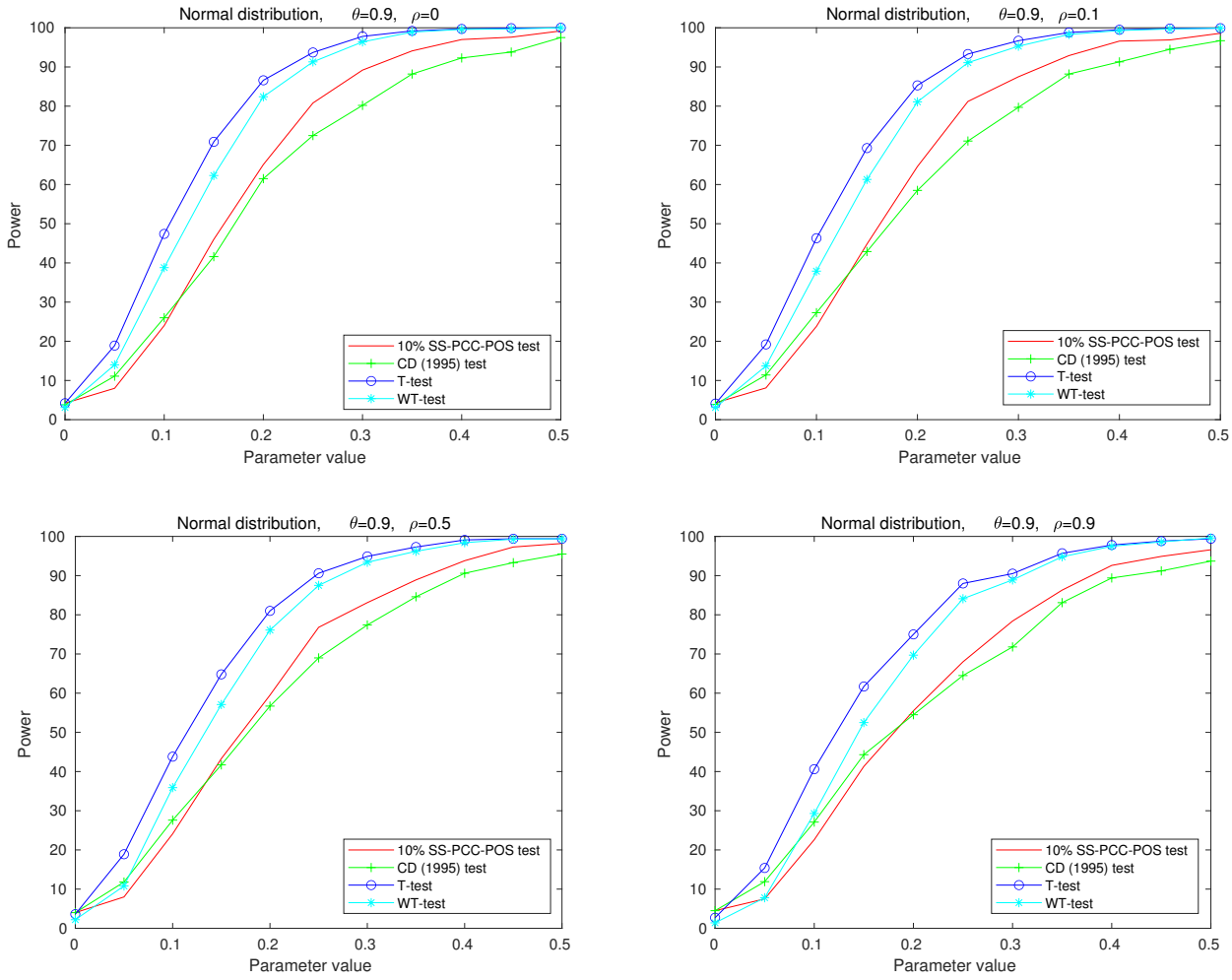
Second, figure 7 presents the results of the performance of the aforementioned tests, when the errors ε_t follows Cauchy distribution. It is evident that the 10% SS-PCC-POS test outperforms all other tests. Moreover, the t-test and WT-test do not possess much power for low correlation coefficient (0 and 0.1) values, ρ. However, as the correlation between u_t and w_t increases, the gap between the power functions narrows significantly.

Third, in figures 8 and 9, we have considered the cases where the errors in turn follow $t(2)$ and mixture distributions. In the former case of $t(2)$ distributed errors, the 10% SS-PCC-POS test outperforms the rest; however, for almost all correlation coefficients ρ, the gap between the power functions is rather small, albeit it is narrowest for $\rho = 0.9$. In the case of errors with mixture distribution, our 10% SS-PCC-POS test is still the most powerful test. On other hand, it is evident that the t-test and WT-test do not possess much power for small values (0 and 0.1) of correlation coefficient ρ. However, the power functions increase and converge to those of the other tests, as the correlation increases.

An interesting observation is the stark contrast between the power of the 10% SS-PCC-POS test and the t-test, when the errors follow the Cauchy, $t(2)$ and normal distributions respectively. The Cauchy and $t(2)$ distributions possess heavy tails, in the presence of which the standard error of the regression coefficients is inflated, which in turn leads to low power when the mean is used as a measure of central tendency. For instance, the Cauchy distribution has the heaviest tails among the considered DGPs, as a result of which the t-test and WT-test have very low power. By noting that a Student's t distribution with ν degrees of freedom converges to the Cauchy distribution for $\nu = 1$ and to the normal distribution as $\nu \to \infty$, it would be interesting to see at which degree of freedom the 10% SS-PCC-POS test is outperformed by the t-test and WT-tests. Figures 12-15 suggest that for different values of ρ in (33) the t-test and WT test outperform the 10% SS-PCC-POS test for $\nu = 4$. Interestingly, figure 16 shows that the tails of the $t(2)$ distribution are substantially heavier than that of the $t(4)$, which may explain the transition.

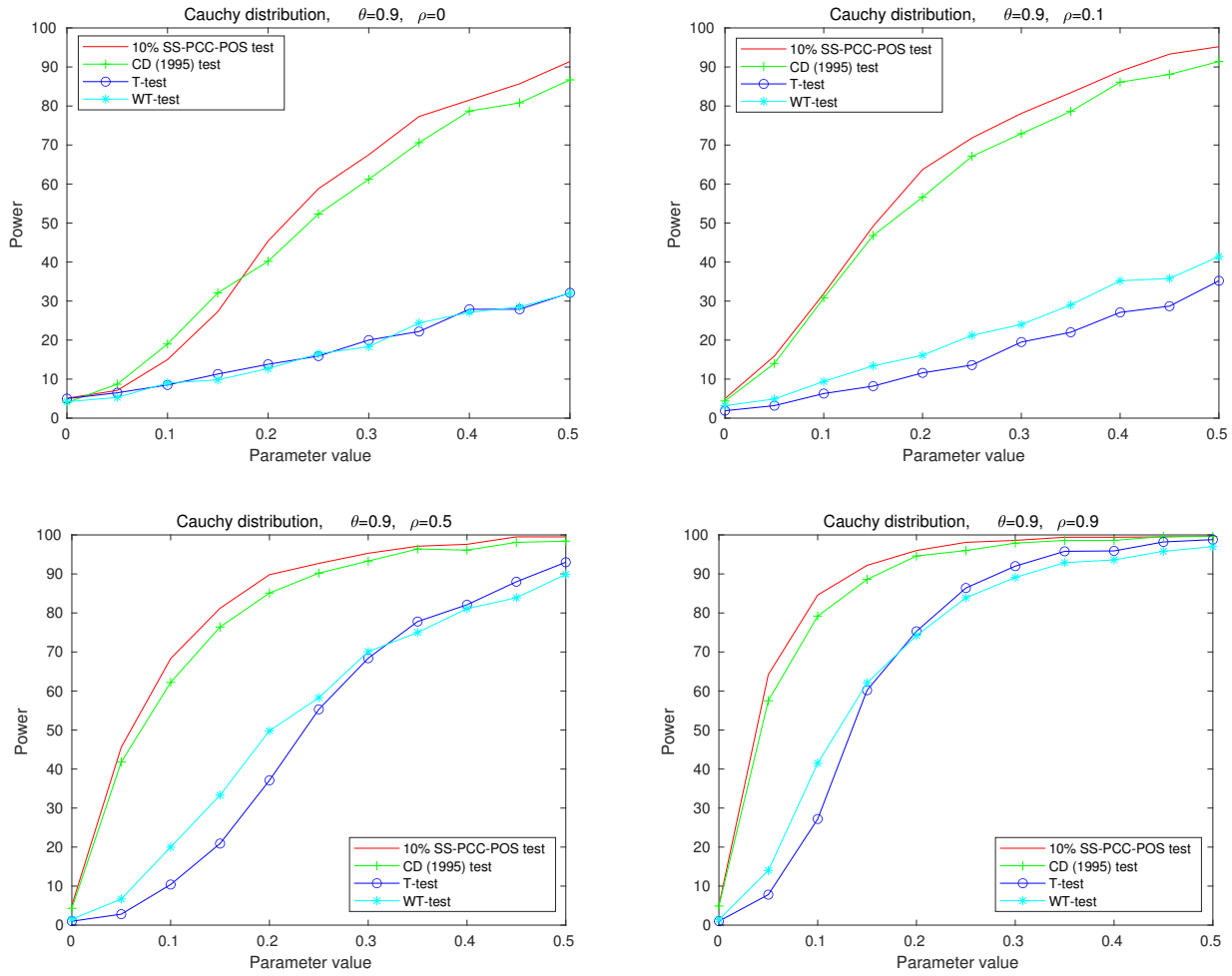
Finally, in figures 8 and 9, the errors are normally distributed with different forms of heteroskedasticity. In the first case [see figure 8], there is a break in variance, in the presence of which our test outperforms the other tests. Furthermore, the t-test and WT-test do not possess any power for low correlation (0 and 0.1) values of ρ. However, with increasing values of the correlation coefficient the power curves of all test appear to converge. In the other case [see figure 9], the variance follows a GARCH(1,1) model with a jump in variance. In this case, our test is only outperformed by the CD (1995) test, which has the greatest power. Nevertheless, the 10% SS-PCC-POS test still outperforms the t-test and WT-test.

Figure 6: Power comparisons: different tests. Normal error distributions with different values of ρ in (33) and $\theta = 0.9$ in (32)



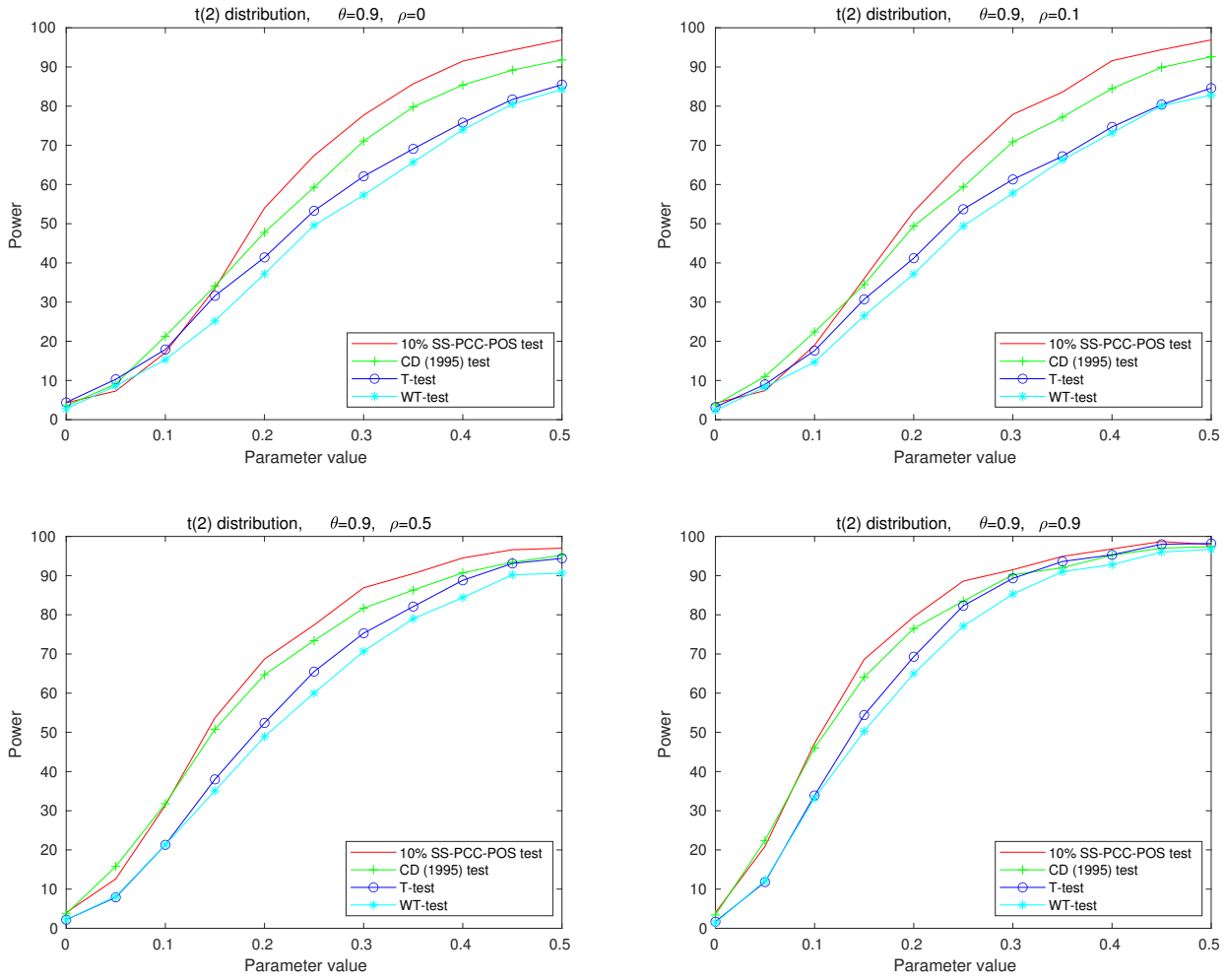
Note: These figures compare the power curves of the 10% split-sample PCC-POS test [10% SS-PCC-POS test] with: (1) the t-test; (2) the sign-based test proposed by Campbell and Dufour (1995) [CD (1995) test]; and (3) the t-test based on White's (1980) variance correction [WT-test].

Figure 7: Power comparisons: different tests. Cauchy error distributions with different values of ρ in (33) and $\theta = 0.9$ in (32)



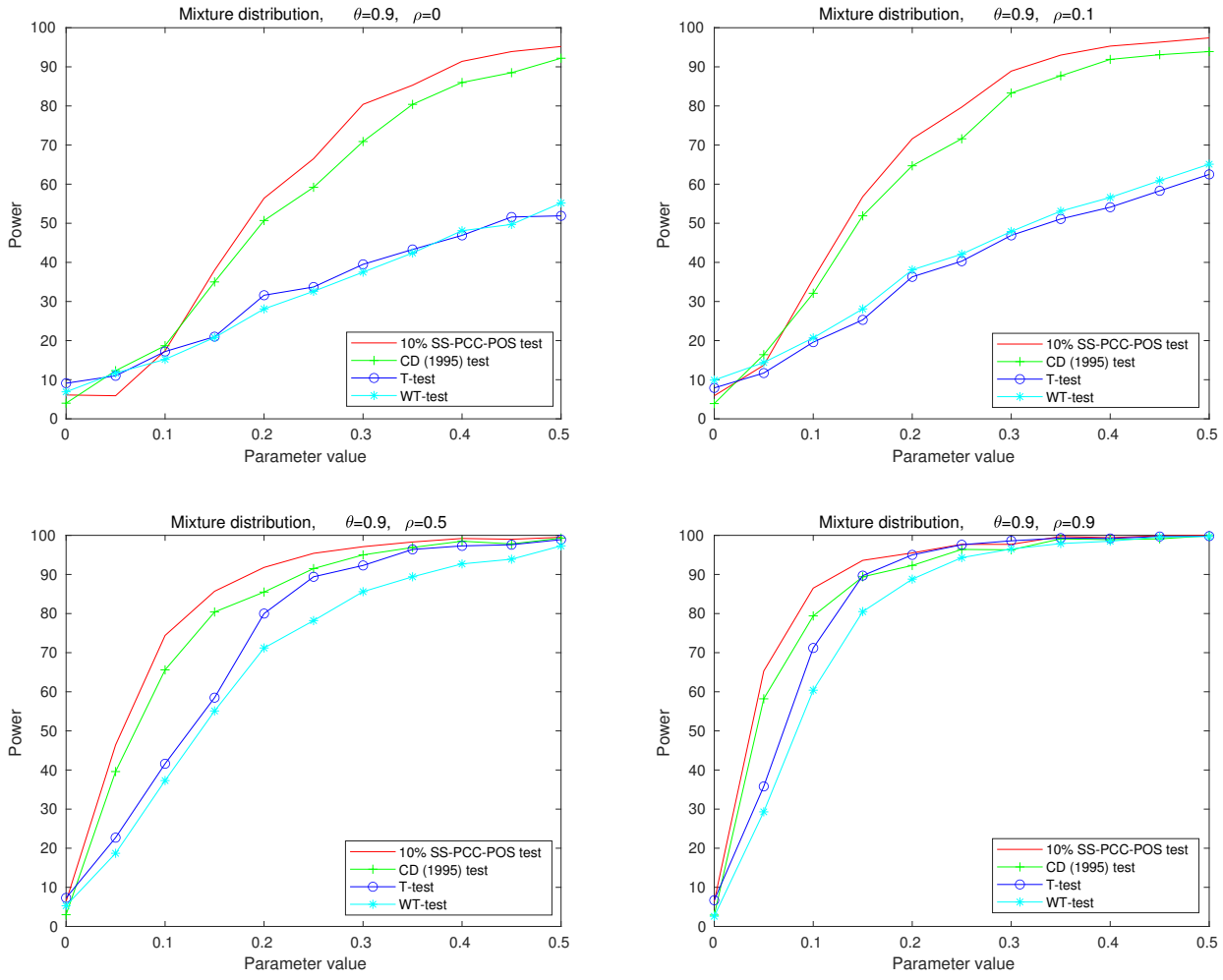
Note: These figures compare the power curves of the 10% split-sample PCC-POS test [10% SS-PCC-POS test] with: (1) the t-test; (2) the sign-based test proposed by Campbell and Dufour (1995) [CD (1995) test]; and (3) the t-test based on White's (1980) variance correction [WT-test].

Figure 8: Power comparisons: different tests. Student's t error distributions with 2 degrees of freedom [i.e $t(2)$], with different values of ρ in (33) and $\theta = 0.9$ in (32)



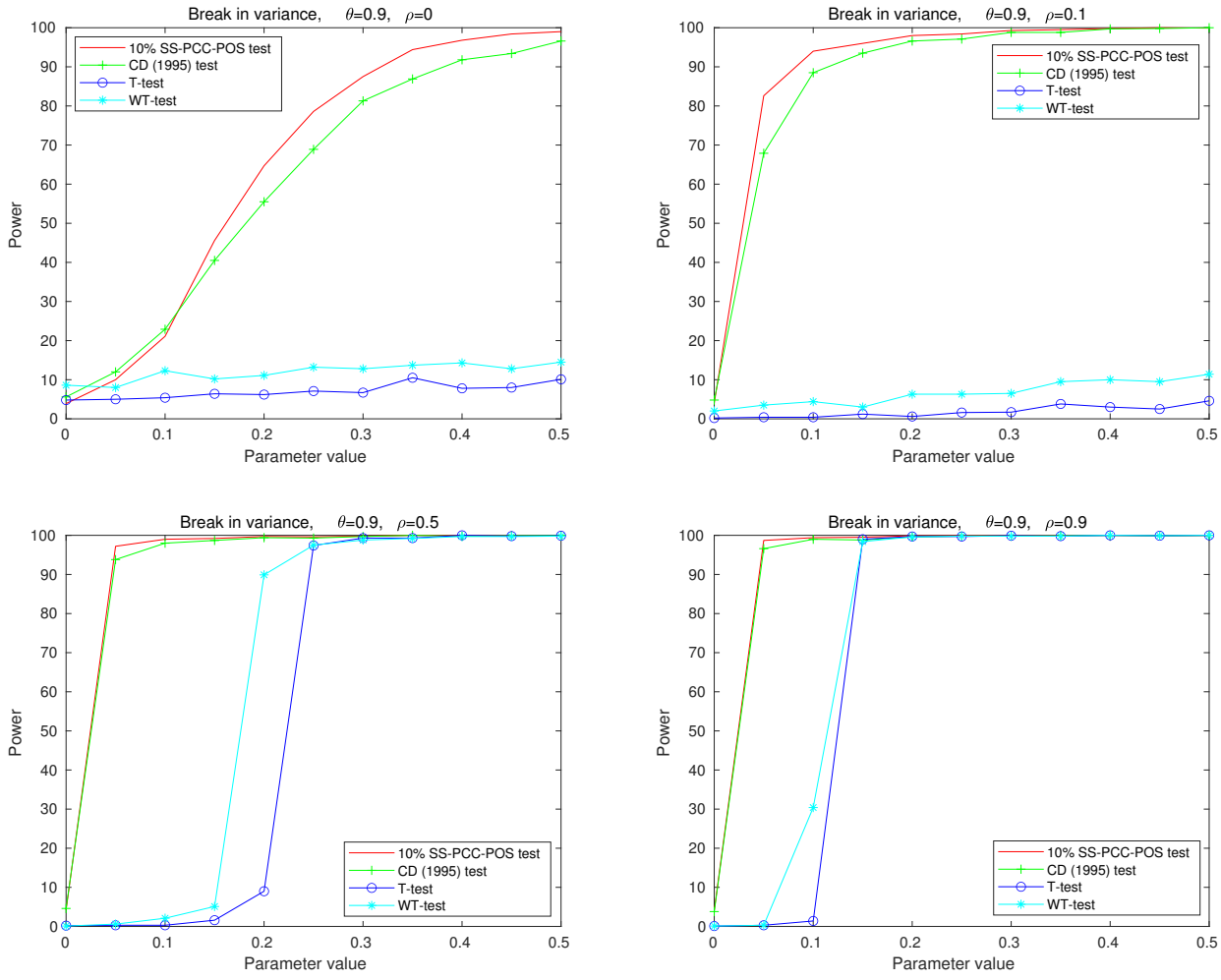
Note: These figures compare the power curves of the 10% split-sample PCC-POS test [10% SS-PCC-POS test] with: (1) the t-test; (2) the sign-based test proposed by Campbell and Dufour (1995) [CD (1995) test]; and (3) the t-test based on White's (1980) variance correction [WT-test].

Figure 9: Power comparisons: different tests. Mixture error distributions with different values of ρ in (33) and $\theta = 0.9$ in (32)



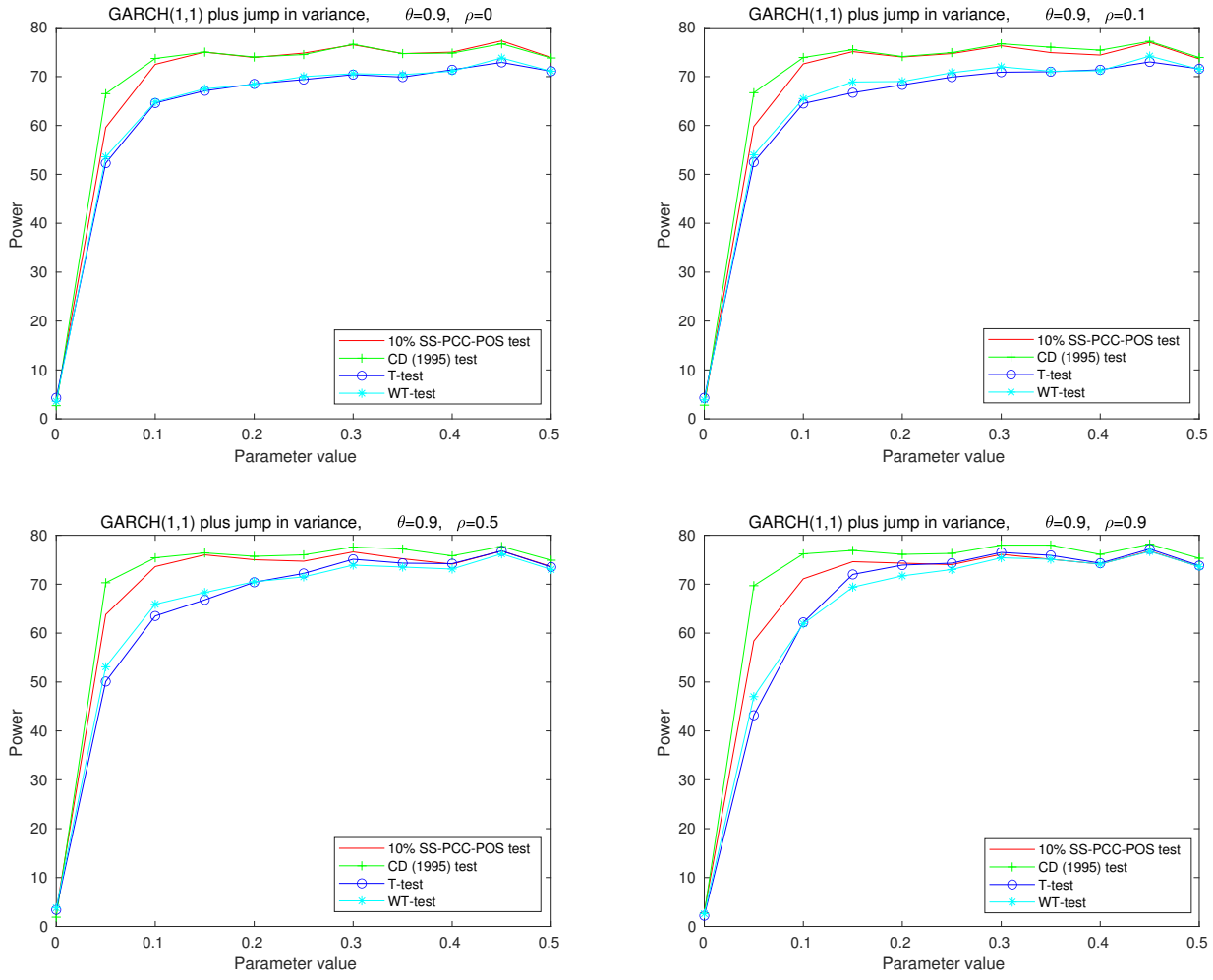
Note: These figures compare the power curves of the 10% split-sample PCC-POS test [10% SS-PCC-POS test] with: (1) the t-test; (2) the sign-based test proposed by Campbell and Dufour (1995) [CD (1995) test]; and (3) the t-test based on White's (1980) variance correction [WT-test].

Figure 10: Power comparisons: different tests. Normal error distributions with break in variance, with different values of ρ in (33) and $\theta = 0.9$ in (32)



Note: These figures compare the power curves of the 10% split-sample PCC-POS test [10% SS-PCC-POS test] with: (1) the t-test; (2) the sign-based test proposed by Campbell and Dufour (1995) [CD (1995) test]; and (3) the t-test based on White's (1980) variance correction [WT-test].

Figure 11: Power comparisons: different tests. Normal error distributions GARCH(1,1) plus jump invariance, with different values of ρ in (33) and $\theta = 0.9$ in (32)



Note: These figures compare the power curves of the 10% split-sample PCC-POS test [10% SS-PCC-POS test] with: (1) the t-test; (2) the sign-based test proposed by Campbell and Dufour (1995) [CD (1995) test]; and (3) the t-test based on White's (1980) variance correction [WT-test].

8 Conclusion

In this paper, we extend the exact point-optimal sign-based procedures proposed by Dufour and Taamouti (2010) to a predictive regression framework. We showed that by implementing the procedures for pair copula constructions of discrete data, we can derive exact and distribution-free sign-based statistics for dependent data in the context of linear and nonlinear predictive regressions, without imposing any potentially restrictive assumptions. The proposed tests are valid, distribution-free and robust against heteroskedasticity of unknown form. Furthermore, they may be inverted to produce a confidence region for the vector (sub-vector) of parameters of the regression model.

We further suggest a sequential estimation strategy for the D-vine PCC and discuss the choice of the copula family. As the proposed sign statistics depend on the alternative hypothesis, another problem consists of finding an alternative that controls size and maximizes the power. In line with Dufour and Taamouti (2010), we find that when 10% of sample is used to estimate the alternative and the rest to compute the test-statistic, our procedures have the optimal power and are closest to the power envelope.

Finally, we present a Monte Carlo study to assess the performance of the proposed tests in terms of size control and power, by comparing them to some other tests that are intended to be robust against heteroskedasticity. We consider a variety of different DGPs and we show that the 10% split-sample point-optimal sign-test based on pair copula constructions is superior to the t-test, Campbell and Dufour (1995) sign-based test, and the t-test based on White (1980) variance correction in most cases.

9 Appendix

Derivation of the Neyman-Pearson type sign-based statistic for testing the unpredictability hypothesis for $T \leq 3$.

The likelihood function of sample in terms of signs $s(y_1), \cdots, s(y_T)$ conditional on X is

$$L\left(U(T), \boldsymbol{\beta}, X\right) = P\left[s(y_1) = s_1, \cdots, s(y_T) = s_T \mid X\right] = \prod_{t=1}^{T} P\left[s(y_t) = s_t \mid \underline{\mathrm{S}}_{t-1} = \underline{\mathrm{s}}_{t-1}, X\right],$$

for

$$\underline{\mathrm{S}}_0 = \{\emptyset\}, \quad \underline{\mathrm{S}}_{t-1} = \{s(y_1), \cdots, s(y_{t-1})\}, \text{ for } t \geq 2,$$

and

$$P\left[s(y_1) = s_1 \mid \underline{\mathrm{S}}_0 = \underline{\mathrm{s}}_0, X\right] = P\left[s(y_1) = s_1 \mid X\right],$$

where each s_t, for $1 \leq t \leq T$, takes two possible values 0 and 1. Given model (1) and assumption (2), under the null hypothesis of unpredictability, the signs $s(\varepsilon_t)$, for $1 \leq t \leq T$, are i.i.d conditional on X according to $Bi(1, 0.5)$. Then, the signs $s(y_t)$, for $1 \leq t \leq T$, will also be i.i.d conditional on X with

$$P\left[s(y_t) = 1 \mid X\right] = P\left[s(y_t) = 0 \mid X\right] = \frac{1}{2}, \quad \text{for} \quad t = 1, \cdots, T.$$

Consequently, under H_0

$$L_0\left(U(T), \mathbf{0}, X\right) = \prod_{t=1}^{T} P\left[s(y_t) = s_t \mid X\right] = \left(\frac{1}{2}\right)^T$$

and under H_1 we have

$$L_1\left(U(T), \boldsymbol{\beta}_1, X\right) = \prod_{t=1}^{T} P\left[s(y_t) = s_t \mid \underline{\mathrm{S}}_{t-1} = \underline{\mathrm{s}}_{t-1}, X\right]$$

where now, for $t = 1, \cdots, T$,

$$y_t = \boldsymbol{\beta}_1' \boldsymbol{x}_{t-1} + \varepsilon_t$$

The log-likelihood ratio is given by

$$\ln\left\{\frac{L_1\left(U(T), \boldsymbol{\beta}_1, X\right)}{L_0\left(U(T), \mathbf{0}, X\right)}\right\} = \sum_{t=1}^{T} \ln\left\{P\left[s(y_t) = s_t \mid \underline{\mathrm{S}}_{t-1} = \underline{\mathrm{s}}_{t-1}, X\right]\right\} - \mathrm{T} \ln\left\{\frac{1}{2}\right\}.$$

According to Neyman-Pearson lemma [see e.g. Lehmann (1959), page 65], the best test to test H_0 against

H_1, based on $s(y_1), \cdots, s(y_T)$, rejects H_0 when

$$SL_T(\boldsymbol{\beta}_1) = \ln\left\{\frac{L_1\left(U(T), \boldsymbol{\beta}_1, X\right)}{L_0\left(U(T), \mathbf{0}, X\right)}\right\} \geq c$$

or when

$$\sum_{t=1}^{T} \ln\left\{P\left[s(y_t) = s_t \mid \underline{\mathrm{S}}_{t-1} = \underline{\mathrm{s}}_{t-1}, X\right]\right\} \geq c_1 \equiv c + T\ln\left(\frac{1}{2}\right),$$

The critical value, say c_1 is given by the smallest constant c_1 such that

$$P\left(\sum_{t=1}^{T} \ln\left\{P\left[s(y_t) = s_t \mid \underline{\mathrm{S}}_{t-1} = \underline{\mathrm{s}}_{t-1}, X\right]\right\} > c_1 \mid H_0\right) \leq \alpha.$$

We have

$$\begin{aligned}
\ln\left\{P[s(y_1) = s_1 \mid \underline{\mathrm{S}}_0 = \underline{\mathrm{s}}_0, X]\right\} &= \ln\left\{P[s(y_1) = s_1 \mid X]\right\} \\
&= s(y_1)\ln P[y_1 \geq 0 \mid X] + (1 - s(y_1))\ln P[y_1 < 0 \mid X] \\
&= s(y_1)\ln\left\{\frac{P[y_1 \geq 0 \mid X]}{P[y_1 < 0 \mid X]}\right\} + \ln P[y_1 < 0 \mid X] \\
&= s(y_1)\ln\left\{\frac{P[\varepsilon_1 \geq -\boldsymbol{\beta}_1'\boldsymbol{x}_{t-1} \mid X]}{P[\varepsilon_1 < -\boldsymbol{\beta}_1'\boldsymbol{x}_{t-1} \mid X]}\right\} + \ln P[\varepsilon_1 < -\boldsymbol{\beta}_1'\boldsymbol{x}_{t-1} \mid X],
\end{aligned}$$

and for $t = 2, \cdots, T$, with $T \leq 3$ we have

$$\begin{aligned}
\sum_{t=2}^{T} \ln P\left[s(y_t) = s_t \mid \underline{\mathrm{S}}_{t-1} = \underline{\mathrm{s}}_{t-1}, X\right] &= \sum_{t=2}^{T} \ln\left(\frac{P[s(y_t) = s_t, s(y_{t-1}) = s_{t-1} \mid \underline{\mathrm{S}}_{t-2} = \underline{\mathrm{s}}_{t-2}, X]}{P[s(y_{t-1}) = s_{t-1} \mid \underline{\mathrm{S}}_{t-2} = \underline{\mathrm{s}}_{t-2}, X]}\right) \\
&= \sum_{t=2}^{T} \ln\Bigg(\sum_{k_t=0,1}\sum_{k_{t-1}=0,1} (-1)^{k_t+k_{t-1}} \\
&\quad \times \left\{P[s(y_t) \leq s_t - k_t, s(y_{t-1}) \leq s_{t-1} - k_{t-1} \mid \underline{\mathrm{S}}_{t-2} = \underline{\mathrm{s}}_{t-2}, X]\right\} \\
&\quad /P[s(y_{t-1}) = s_{t-1} \mid \underline{\mathrm{S}}_{t-2} = \underline{\mathrm{s}}_{t-2}, X]\Bigg) \\
&= \sum_{t=2}^{T} \ln\Bigg(\sum_{k_t=0,1}\sum_{k_{t-1}=0,1} (-1)^{k_t+k_{t-1}} \\
&\quad \times \Big\{C_{s(y_t),s(y_{t-1})|\underline{\mathrm{S}}_{t-2}}\Big(F_{s(y_t)|\underline{\mathrm{S}}_{t-2}}(s_t - k_t \mid \underline{\mathrm{s}}_{t-2}, X), \\
&\qquad\qquad F_{s(y_{t-1})|\underline{\mathrm{S}}_{t-2}}(s_{t-1} - k_{t-1} \mid \underline{\mathrm{s}}_{t-2}, X)\Big)\Big\} \\
&\quad /P[s(y_{t-1}) = s_{t-1} \mid \underline{\mathrm{S}}_{t-2} = \underline{\mathrm{s}}_{t-2}, X]\Bigg)
\end{aligned}$$

$$= \sum_{t=2}^{T} \ln \Bigg\{ \sum_{k_t=0,1} \sum_{k_{t-1}=0,1} (-1)^{k_t+k_{t-1}} \times \Big\{ C_{s(y_t),s(y_{t-1})|\mathbb{S}_{t-2}} \Big(F_{s(y_t)|\mathbb{S}_{t-2}}(s_t - k_t \mid \mathbb{s}_{t-2}, X), F_{s(y_{t-1})|\mathbb{S}_{t-2}}(s_{t-1} - k_{t-1} \mid \mathbb{s}_{t-2}, X) \Big) \Big\} \Bigg\}$$
$$- \sum_{t=2}^{T} \ln \left\{ P[s(y_{t-1}) = s_{t-1} \mid \mathbb{S}_{t-2} = \mathbb{s}_{t-2}, X] \right\}$$

Each argument $F_{s(y_t)|\mathbb{S}_{t-2}}(s_t - k_t \mid \mathbb{s}_{t-2}, X)$ and $F_{s(y_{t-1})|\mathbb{S}_{t-2}}(s_{t-1} - k_{t-1} \mid \mathbb{s}_{t-2}, X)$ in the copula expression above can be evaluated as follows

$$F_{s(y_t)|\mathbb{S}_{t-2}}(s_t - k_t \mid \mathbb{s}_{t-2}, X) =$$
$$\Big\{ C_{s(y_t),s(y_{t-2})|\mathbb{S}_{t-3}}(F(s_t - k_t \mid \mathbb{s}_{t-3}, X), F(s_{t-2} \mid \mathbb{s}_{t-3}, X))$$
$$- C_{s(y_t),s(y_{t-2})|\mathbb{S}_{t-3}}(F(s_t - k_t \mid \mathbb{s}_{t-3}, X), F(s_{t-2} - 1 \mid \mathbb{s}_{t-3}, X)) \Big\} / P[s(y_{t-2}) = s_{t-2} \mid \mathbb{S}_{t-3} = \mathbb{s}_{t-3}, X]$$

and similarly

$$F_{s(y_{t-1})|\mathbb{S}_{t-2}}(s_{t-1} - k_{t-1} \mid \mathbb{s}_{t-2}, X) =$$
$$\Big\{ C_{s(y_{t-2}),s(y_{t-1})|\mathbb{S}_{t-3}}(F(s_{t-2} \mid \mathbb{s}_{t-3}, X), F(s_{t-1} - k_{t-1} \mid \mathbb{s}_{t-3}, X))$$
$$- C_{s(y_{t-2}),s(y_{t-1})|\mathbb{S}_{t-3}}(F(s_{t-2} - 1 \mid \mathbb{s}_{t-3}, X), F(s_{t-1} - k_{t-1} \mid \mathbb{s}_{t-3}, X)) \Big\} / P[s(y_{t-2}) = s_{t-2} \mid \mathbb{S}_{t-3} = \mathbb{s}_{t-3}, X]$$

Thus, for $T \leq 3$ the Neyman-Pearson type test statistic based on $s(y_1), \cdots, s(y_T)$, can be expressed as

$$SL_T(\boldsymbol{\beta}_1) = s(y_1) \ln \left\{ \frac{P[\varepsilon_1 \geq -\boldsymbol{\beta}_1' \boldsymbol{x}_{t-1} \mid X]}{P[\varepsilon_1 < -\boldsymbol{\beta}_1' \boldsymbol{x}_{t-1} \mid X]} \right\} + \ln P[\varepsilon_1 < -\boldsymbol{\beta}_1' \boldsymbol{x}_{t-1} \mid X] + \sum_{t=2}^{T} \ln \Bigg\{ \sum_{k_t=0,1} \sum_{k_{t-1}=0,1} (-1)^{k_t+k_{t-1}}$$
$$\times \Big(C_{s(y_t),s(y_{t-1})|\mathbb{S}_{t-2}} \Big(F_{s(y_t)|\mathbb{S}_{t-2}}(s_t - k_t \mid \mathbb{s}_{t-2}, X), F_{s(y_{t-1})|\mathbb{S}_{t-2}}(s_{t-1} - k_{t-1} \mid \mathbb{s}_{t-2}, X) \Big) \Big) \Bigg\}$$
$$- \sum_{t=2}^{T} \ln \left\{ P[s(y_{t-1}) = s_{t-1} \mid \mathbb{S}_{t-2} = \mathbb{s}_{t-2}, X] \right\} - n \ln \left\{ \frac{1}{2} \right\}$$

■

Vine decomposition in the continuous case.

In Section 4, it is shown that the signs $s(y_1), \cdots, s(y_T)$ may have a continuous extension with a perturbation in $[0, 1]$ [see Denuit and Lambert (2005)]. This can be achieved by employing a transformation of the form $s^*(y_t) = s(y_t) + U - 1$ for $t = 1, \cdots, T$, where a natural choice for U is the uniform distribution.

Thus, for $\{s^*(y_t) \in \mathbb{R}, t = 1, \cdots, T\}$ consider the continuous equivalent of the conditional probability mass function (11) - i.e. the conditional density function. Further, by letting $\underline{\mathrm{S}}^*_{t-1}$ be the continuous extension of $\underline{\mathrm{S}}_{t-1}$, the conditional density function may be expressed as

$$f_{s^*(y_t)|\underline{\mathrm{S}}^{*\backslash j}_{t-1} \cup s^*(y_j)} = \frac{f_{s^*(y_t),s^*(y_j)|\underline{\mathrm{S}}^{*\backslash j}_{t-1}}}{f_{s^*(y_j)|\underline{\mathrm{S}}^{*\backslash j}_{t-1}}}. \tag{34}$$

From the Theorem of Sklar (1959), we know that

$$\begin{aligned} f_{s^*(y_t),s^*(y_j)|\underline{\mathrm{S}}^{*\backslash j}_{t-1}}(s^*_t, s^*_j \mid \underline{\mathrm{s}}^{*\backslash j}_{t-1}, X) = c_{s^*(y_t),s^*(y_j)|\underline{\mathrm{S}}^{*\backslash j}_{t-1}} \left(F_{s^*(y_t)|\underline{\mathrm{S}}^{*\backslash j}_{t-1}}(s^*_t \mid \underline{\mathrm{s}}^{*\backslash j}_{t-1}, X), F_{s^*(y_j)|\underline{\mathrm{S}}^{*\backslash j}_{t-1}}(s^*_j \mid \underline{\mathrm{s}}^{*\backslash j}_{t-1}, X) \right) \\ \times \quad f_{s^*(y_t)|\underline{\mathrm{S}}^{*\backslash j}_{t-1}} f_{s^*(y_j)|\underline{\mathrm{S}}^{*\backslash j}_{t-1}}, \end{aligned} \tag{35}$$

where $c()$ is the copula density function. Thus,

$$f_{s^*(y_t)|\underline{\mathrm{S}}^{*\backslash j}_{t-1} \cup s^*(y_j)} = c_{s^*(y_t),s^*(y_j)|\underline{\mathrm{S}}^{*\backslash j}_{t-1}} \left(F_{s^*(y_t)|\underline{\mathrm{S}}^{*\backslash j}_{t-1}}(s^*_t \mid \underline{\mathrm{s}}^{*\backslash j}_{t-1}, X), F_{s^*(y_j)|\underline{\mathrm{S}}^{*\backslash j}_{t-1}}(s^*_j \mid \underline{\mathrm{s}}^{*\backslash j}_{t-1}, X) \right) f_{s^*(y_t)|\underline{\mathrm{S}}^{*\backslash j}_{t-1}}, \tag{36}$$

with

$$c_{s^*(y_t),s^*(y_j)|\underline{\mathrm{S}}^{*\backslash j}_{t-1}} \left(F_{s^*(y_t)|\underline{\mathrm{S}}^{*\backslash j}_{t-1}}(s^*_t \mid \underline{\mathrm{s}}^{*\backslash j}_{t-1}, X), F_{s^*(y_j)|\underline{\mathrm{S}}^{*\backslash j}_{t-1}}(s^*_j \mid \underline{\mathrm{s}}^{*\backslash j}_{t-1}, X) \right) = \frac{\partial^2 C_{s^*(y_t),s^*(y_j)|\underline{\mathrm{S}}^{*\backslash j}_{t-1}} \left(F_{s^*(y_t)|\underline{\mathrm{S}}^{*\backslash j}_{t-1}}(s^*_t|\underline{\mathrm{s}}^{*\backslash j}_{t-1}, X), F_{s^*(y_j)|\underline{\mathrm{S}}^{*\backslash j}_{t-1}}(s^*_j|\underline{\mathrm{s}}^{*\backslash j}_{t-1}, X) \right)}{\partial F_{s^*(y_t)|\underline{\mathrm{S}}^{*\backslash j}_{t-1}} \left(s^*_t|\underline{\mathrm{s}}^{*\backslash j}_{t-1}, X \right) \partial F_{s^*(y_j)|\underline{\mathrm{S}}^{*\backslash j}_{t-1}} \left(s^*_j|\underline{\mathrm{s}}^{*\backslash j}_{t-1}, X \right)} \tag{37}$$

can express (34), and the arguments of the copulas, say, $F_{s^*(y_t)|\underline{\mathrm{S}}^{*\backslash j}_{t-1}}(s^*_t \mid \underline{\mathrm{s}}^{*\backslash j}_{t-1}, X)$ are obtained using the expression by Joe (1996), such that

$$F_{s^*(y_t)|\underline{\mathrm{S}}^{*\backslash j}_{t-1}}(s^*_t \mid \underline{\mathrm{s}}^{*\backslash j}_{t-1}, X) = \frac{\partial C_{s^*(y_t),s^*(y_i)|\underline{\mathrm{S}}^{*\backslash j,i}_{t-1}} \left(F_{s^*(y_t)|\underline{\mathrm{S}}^{*\backslash j,i}_{t-1}}(s^*_t \mid \underline{\mathrm{s}}^{*\backslash j,i}_{t-1}, X), F_{s^*(y_i)|\underline{\mathrm{S}}^{*\backslash j,i}_{t-1}}(s^*_i \mid \underline{\mathrm{s}}^{*\backslash j,i}_{t-1}, X) \right)}{\partial F_{s^*(y_i)|\underline{\mathrm{S}}^{*\backslash j,i}_{t-1}}(s^*_i \mid \underline{\mathrm{s}}^{*\backslash j,i}_{t-1}, X)}. \tag{38}$$

Therefore, when the data is continuous, the marginals in the copula expressions of, say, the third tree, $F_{t|t+1,t+2}$ for $t = 1, \cdots, T-2$ and $F_{t+3|t+1,t+2}$ for $t = 1, \cdots, T-3$ are obtained by

$$F_{t|t+1,t+2} = \frac{\partial C_{t,t+1|t+2}(F_{t|t+2}(s^*_t \mid s^*_{t+2}, X), F_{t+1|t+2}(s^*_{t+1} \mid s^*_{t+2}, X))}{\partial F_{t+1|t+2}(s^*_{t+1} \mid s^*_{t+2}, X)}, \tag{39}$$

where $F_{t+3|t+1,t+2}$ is obtained in a similar way. ■

Proof of Proposition 1.

The likelihood function of the sample in terms of signs $s(y_1), \cdots, s(y_T)$ conditional on X is given by

$$L\left(U(T), \boldsymbol{\beta}, X\right) = P\left[s(y_1) = s_1, \cdots, s(y_T) = s_T \mid X\right]$$

where each s_t, for $1 \leq t \leq T$, takes two possible values 0 and 1. Given model (1) and assumption (2), under the null hypothesis the signs $s(\varepsilon_t)$, for $1 \leq t \leq T$, are i.i.d conditional on X according to $Bi(1, 0.5)$. Then, the signs $s(y_t)$, for $1 \leq t \leq T$, will also be i.i.d conditional on X

$$P\left[s(y_t) = 1 \mid X\right] = P\left[s(y_t) = 0 \mid X\right] = \frac{1}{2}, \text{ for } t = 1, \cdots, T.$$

Consequently, under H_0 we have

$$L_0\left(U(T), \mathbf{0}, X\right) = \prod_{t=1}^{T} P\left[s(y_t) = s_t \mid X\right] = \left(\frac{1}{2}\right)^T$$

and under H_1 the likelihood function conditional on X can be expressed as

$$L_1\left(U(T), \boldsymbol{\beta}_1, X\right) = P_1[s(y_1) = s_1 \mid X] \times \prod_{t=2}^{T} P_{t|1:t-1}[s(y_t) = s_t \mid s(y_1) = s_1 : s(y_{t-1}) = s_{t-1}, X].$$

which can further be decomposed using the D-vine array $A = (\sigma_{lt})_{1 \leq l \leq t \leq T}$ to obtain

$$L_1\left(U(T), \boldsymbol{\beta}_1, X\right) = P_1[s(y_1) = s_1 \mid X] \times \prod_{t=2}^{T} \prod_{l=t-1}^{2} c_{\sigma_{lt}t, |\sigma_{1t}, \cdots, \sigma_{t-1,t}} \times c_{\sigma_{1t}t} \times P_t[s(y_t) = s_t \mid X]$$

where now for $t = 1, \cdots, T$,

$$y_t = \boldsymbol{\beta}_1' \boldsymbol{x}_{t-1} + \varepsilon_t$$

Under assumption (1) and (2), the likelihood function conditional on X, under the alternative hypothesis can be expressed as

$$\begin{aligned} L_1\left(U(T), \boldsymbol{\beta}_1, X\right) = & \left(1 - P_1[\varepsilon_1 < -\boldsymbol{\beta}_1' \boldsymbol{x}_0 \mid X]\right)^{s(y_1)} \times P_1[\varepsilon_1 < -\boldsymbol{\beta}_1' \boldsymbol{x_0} \mid X]^{1-s(y_1)} \\ & \times \prod_{t=2}^{T} \prod_{l=t-1}^{2} c_{\sigma_{lt}j, |\sigma_{1t}, \cdots, \sigma_{t-1,t}} \times c_{\sigma_{1t}t} \times \left(1 - P_t[\varepsilon_t < -\boldsymbol{\beta}_1' \boldsymbol{x}_{t-1} \mid X]\right)^{s(y_t)} \\ & \times P_t[\varepsilon_t < -\boldsymbol{\beta}_1' \boldsymbol{x}_{t-1} \mid X]^{1-s(y_t)} \end{aligned}$$

The log-likelihood ratio is given by

$$\ln\left\{\frac{L_1\left(U(T),\boldsymbol{\beta}_1,X\right)}{L_0\left(U(T),\mathbf{0},X\right)}\right\} = s(y_1)\ln\left\{\frac{1-P_1[\varepsilon_1<-\boldsymbol{\beta}_1'\boldsymbol{x_0}\mid X]}{P_1[\varepsilon_1<-\boldsymbol{\beta}_1'\boldsymbol{x_0}\mid X]}\right\}+\ln\left(1-P_1[\varepsilon_1<-\boldsymbol{\beta}_1'\boldsymbol{x_0}\mid X]\right)$$
$$+\sum_{t=2}^{T}\sum_{l=t-1}^{2}\ln c_{\sigma_{lt}t,|\sigma_{1t},\cdots,\sigma_{t-1,t}}+\sum_{t=2}^{T}\ln c_{a_{1t}t}+\sum_{t=2}^{T}s(y_t)\ln\left\{\frac{1-P_t[\varepsilon_t<-\boldsymbol{\beta}_1'\boldsymbol{x}_{t-1}\mid X]}{P_t[\varepsilon_t<-\boldsymbol{\beta}_1'\boldsymbol{x}_{t-1}\mid X]}\right\}$$
$$+\sum_{t=2}^{T}\ln\left(1-P_t[\varepsilon_t<-\boldsymbol{\beta}_1'\boldsymbol{x}_{t-1}\mid X]\right)-T\ln\left(\frac{1}{2}\right)$$

According to Neyman-Pearson Lemma [see e.g. Lehmann (1959), page 65], the best test for testing H_0 against H_1, based on $s(y_1),\cdots,s(y_T)$, rejects H_0 when

$$\ln\left\{\frac{L_1\left(U(T),\boldsymbol{\beta}_1,X\right)}{L_0\left(U(T),\mathbf{0},X\right)}\right\}\geq c$$

or when

$$\ln\left\{\frac{L_1\left(U(T),\boldsymbol{\beta}_1,X\right)}{L_0\left(U(T),\mathbf{0}\right)}\right\} = \sum_{t=2}^{T}\sum_{l=t-1}^{2}\ln c_{\sigma_{lt}t,|\sigma_{1t},\cdots,\sigma_{t-1,t}}+\sum_{t=2}^{T}\ln c_{\sigma_{1t}t}$$
$$+\sum_{t=1}^{T}s_t\ln\left\{\frac{1-P_t[\varepsilon_t<-\boldsymbol{\beta}_1'\boldsymbol{x}_{t-1}\mid X]}{P_t[\varepsilon_t<-\boldsymbol{\beta}_1'\boldsymbol{x}_{t-1}\mid X]}\right\}>c_1(\boldsymbol{\beta}_1)$$

The critical value, say $c_1(\beta_1)$ is given by the smallest constant $c_1(\beta_1)$ such that

$$P\left(\ln\left\{\frac{L_1\left(U(T),\boldsymbol{\beta}_1,X\right)}{L_0\left(U(T),\mathbf{0},X\right)}\right\}>c_1(\boldsymbol{\beta}_1)\mid H_0\right)\leq\alpha.$$

■ **Algorithm for the likelihood function of the signs under the alternative hypothesis.**

In this Section, we adapt the algorithm for the joint pmf for D-vine for discrete variables of Panagiotelis et al. (2012) and Joe (2014) to the context of our study. Let $U(n)=(s(y_1),s(y_2),\cdots,s(y_T))'$ be a binary valued T-vector. Furthermore, for a vector of integers $\mathbf{i}$, let $\mathbf{S_i}=\{s(y_i),i\in\mathbf{i}\}$, where $\mathbf{s_i}$ is a mass point of $\mathbf{S_i}$ and s_g is a mass point of $s(y_g)$. Let

$$F^{+}_{g|\mathbf{i}}:=P\left[s(y_g)\leq s_g\mid\mathbf{S_i}=\mathbf{s_i},X\right],\quad F^{-}_{g|\mathbf{i}}:=P\left[s(y_g)<s_g\mid\mathbf{S_i}=\mathbf{s_i},X\right],$$
$$f_{g|\mathbf{i}}:=P[s(y_g)=s_g\mid\mathbf{S_i}=\mathbf{s_i},X].$$

noting that when $\mathbf{i}=\{\emptyset\}$, these conditional probabilities, correspond to marginal probabilities. Furthermore,

let $C_{gh|\mathbf{i}}$ be a bivariate copula for the conditional CDFs $F_{g|\mathbf{i}}$ and $F_{h|\mathbf{i}}$, and denote

$$C^{++}_{gh|\mathbf{i}} := C_{gh|\mathbf{i}}\left(F^{+}_{g|\mathbf{i}}, F^{+}_{h|\mathbf{i}}\right), \quad C^{+-}_{gh|\mathbf{i}} := C_{gh|\mathbf{i}}\left(F^{+}_{g|\mathbf{i}}, F^{-}_{h|\mathbf{i}}\right),$$
$$C^{-+}_{gh|\mathbf{i}} := C_{gh|\mathbf{i}}\left(F^{-}_{g|\mathbf{i}}, F^{+}_{h|\mathbf{i}}\right), \quad C^{--}_{gh|\mathbf{i}} := C_{gh|\mathbf{i}}\left(F^{-}_{g|\mathbf{i}}, F^{-}_{h|\mathbf{i}}\right).$$

The main elements of the algorithm is the following recursions:

(I) $F^{+}_{j-t|(j-t+1):(j-1)} = \left[C^{++}_{j-t,j-1|(j-t+1):(j-2)} - C^{+-}_{j-t,j-1|(j-t+1):(j-2)}\right] / f_{j-1|(j-t+1):(j-2)}$;

(II) $F^{-}_{j-t|(j-t+1):(j-1)} = \left[C^{-+}_{j-t,j-1|(j-t+1):(j-2)} - C^{--}_{j-t,j-1|(j-t+1):(j-2)}\right] / f_{j-1|(j-t+1):(j-2)}$;

(III) $f_{j-t|(j-t+1):(j-1)} = F^{+}_{j-t|(j-t+1):(j-1)} - F^{-}_{j-t|(j-t+1):(j-1)}$;

(IV) $F^{+}_{j|(j-t+1):(j-1)} = \left[C^{++}_{j-t+1,j|(j-t+2):(j-1)} - C^{-+}_{j-t+1,j|(j-t+2):(j-1)}\right] / f_{j-t+1|(j-t+2):(j-1)}$;

(V) $F^{-}_{j|(j-t+1):(j-1)} = \left[C^{+-}_{j-t+1,j|(j-t+2):(j-1)} - C^{--}_{j-t+1,j|(j-t+2):(j-1)}\right] / f_{j-t+1|(j-t+2):(j-1)}$;

(VI) $f_{j|(j-t+1):(j-1)} = F^{+}_{j|(j-t+1):(j-1)} - F^{-}_{j|(j-t+1):(j-1)}$;

(VII) The values based on $C_{j-t,j|(j-t+1):(j-1)}$ is computed;

(VIII) t is incremented by 1 and back to (I).

The identity employed in the recursions is

$$P\left[s(y_g) \le s_g \mid s(y_h) = s_h, \mathbf{S_i} = \mathbf{s_i}, X\right] = \frac{P[s(y_g)\le s_g, s(y_h)\le s_h|\mathbf{S_i}=\mathbf{s_i}, X] - P[s(y_g)\le s_g, s(y_h)< s_h|\mathbf{S_i}=\mathbf{s_i}, X]}{P[s(y_h)=s_h|\mathbf{S_i}=\mathbf{s_i}, X]}.$$

The algorithm is as follows

1. Input $\mathbf{s}_T = (s_1, \cdots, s_T)$.
2. Allocate an $T \times T$ matrix π, where $\pi_{tj} = f_{(j-t+1):j}$ for $t = 1, \cdots, T$ and $j = t+1, \cdots, T$ and the likelihood function $P[s(y_1) = s_1, \cdots, s(y_T) = s_T]$ under the alternative will appear as π_{TT}.
3. Allocate C^{++}, C^{+-}, C^{-+}, C^{--}, U'^{+}, U'^{-}, U^{+}, U^{-}, u', u, w', w, as vectors of length T.
4. Evaluate F^{+}_{j}, F^{-}_{j}, and $f_j = F^{+}_{j} - F^{-}_{j}$ using (14) and let $\pi_{1j} \leftarrow f_j$ for $j = 1, \cdots, T$;
5. Let $C^{++}_{j} \leftarrow C_{j-1,j}\left(F^{+}_{j-1}, F^{+}_{j}\right)$, $C^{+-}_{j} \leftarrow C_{j-1,j}\left(F^{+}_{j-1}, F^{-}_{j}\right)$, $C^{-+}_{j} \leftarrow C_{j-1,j}\left(F^{-}_{j-1}, F^{+}_{j}\right)$, and $C^{--}_{j} \leftarrow C_{j-1,j}\left(F^{-}_{j-1}, F^{-}_{j}\right)$ for $j = 2, \cdots, T$;
6. Set $P_{2j} \leftarrow C^{++}_{j} - C^{+-}_{j} - C^{-+}_{j} + C^{--}_{j}$ for $j = 2, \cdots, T$;

7. **for** $j = 2, \cdots, T : (\mathcal{T}_1)$ **do**

8. $U_j^{'+} \leftarrow F_{j-1|j}^{+} = \left(C_j^{++} - C_j^{+-}\right)/f_j$, $U_j^{'-} \leftarrow F_{j-1|j}^{-} = \left(C_j^{-+} - C_j^{--}\right)/f_j$, and $u'_j \leftarrow f_{j-1|j} = F_{j-1|j}^{+} - F_{j-1|j}^{-}$;

9. $U_j^{+} \leftarrow F_{j|j-1}^{+} = \left(C_j^{++} - C_j^{-+}\right)/f_{j-1}$, $U_j^{-} \leftarrow F_{j|j-1}^{-} = \left(C_j^{-+} - C_j^{--}\right)/f_{j-1}$, and $u_j \leftarrow f_{j|j-1} = F_{j|j-1}^{+} - F_{j|j-1}^{-}$;

10. **end for**

11. **for** $t = 2, \cdots, T-1 : (\mathcal{T}_2, \cdots, \mathcal{T}_{T-1})$ **do**

12. let $C_j^{\alpha\beta} \leftarrow C_{j-t,j|(j-t+1):(j-1)}\left(U_{j-1}^{'\alpha}, U_j^{\beta}\right)$, for $j = t+1, \cdots, T$ and $\alpha, \beta \in \{+, -\}$;

13. let $w'_j \leftarrow u'_j$, $w_j \leftarrow u_j$ for $j = t, \cdots, T$;

14. **for** $j = t+1, \cdots, T$ **: do**

15. $U_j^{'+} \leftarrow \left(C_j^{++} - C_j^{+-}\right)/w_j$, $U_j^{'-} \leftarrow \left(C_j^{-+} - C_j^{--}\right)/w_j$ and $u'_j \leftarrow U_j^{'+} - U_j^{'-}$;

16. $U_j^{+} \leftarrow \left(C_j^{++} - C_j^{-+}\right)/w'_{j-1}$, $U_j^{-} \leftarrow \left(C_j^{+-} - C_j^{--}\right)/w'_{j-1}$ and $u_j \leftarrow U_j^{+} - U_j^{-}$;

17. **end for**

18. let $\pi_{t+1,j} \leftarrow \pi_{t,j-1} \times u_j$ for $j = t+1, \cdots, T$.

19. **end for**

20. Return the likelihood function π_{TT}.

■

Proof of Theorem 1. The characteristic function of the test statistic $SN_T(\boldsymbol{\beta}_0 \mid \boldsymbol{\beta}_1)$ conditional on X is given by

$$\begin{aligned}\phi_{SN_T}(u) &= \mathbb{E}_X\left[\exp(iuSN_T(\boldsymbol{\beta}_0 \mid \boldsymbol{\beta}_1))\right] \\ &= \mathbb{E}_X\left[\exp\left(iu\left(\sum_{t=1}^{T} R_{t,t-1} + \sum_{t=1}^{T} \ln\left\{\frac{1 - p_t[\boldsymbol{x}_{t-1}, \boldsymbol{\beta}_0, \boldsymbol{\beta}_1 \mid X]}{p_t[\boldsymbol{x}_{t-1}, \boldsymbol{\beta}_0, \boldsymbol{\beta}_1 \mid X]}\right\} s(\tilde{y}_t)\right)\right)\right],\end{aligned}$$

which may be expressed as

$$\phi_{SN_T}(u) = \mathbb{E}_X\left[\prod_{t=1}^{T} \exp\left(iu\left(R_{t,t-1} + \ln\left\{\frac{1 - p_t[\boldsymbol{x}_{t-1}, \boldsymbol{\beta}_0, \boldsymbol{\beta}_1 \mid X]}{p_t[\boldsymbol{x}_{t-1}, \boldsymbol{\beta}_0, \boldsymbol{\beta}_1 \mid X]}\right\} s(\tilde{y}_t)\right)\right)\right],$$

with $R_{1,0} = 0$, and $R_{t,t-1} = \sum_{l=t-1}^{2} \ln c_{\tilde{\sigma}_{lt}t|\tilde{\sigma}_{1t},\cdots,\tilde{\sigma}_{t-1,t}} + \ln c_{\tilde{\sigma}_{1t}t}$ for $t = 2,\cdots,T$, for the D-vine-array $\tilde{A} = (\tilde{\sigma}_{lt})_{1\leq l\leq t\leq T}$, such that $l = 2,\cdots,T-1$ is the row with tree $\mathcal{T}_l$, and column t has the permutation $\underline{\tilde{\sigma}}_{t-1} = (\tilde{\sigma}_{1t},\cdots,\tilde{\sigma}_{t-1,t})$ of the previously added variables, $p_t[\boldsymbol{x}_{t-1},\boldsymbol{\beta}_0,\boldsymbol{\beta}_1 \mid X] = P_t[\varepsilon_t \leq f(\boldsymbol{x}_{t-1},\boldsymbol{\beta}_0) - f(\boldsymbol{x}_{t-1},\boldsymbol{\beta}_1) \mid X]$, and $s(\tilde{y}_t) = s(y_t - f(\boldsymbol{x}_{t-1},\boldsymbol{\beta}_0))$. Furthermore, $u \in \mathbb{R}$ and the complex number $i = \sqrt{-1}$. Unlike Dufour and Taamouti (2010), $\tilde{y}_t$ for $t = 1,\cdots,T$ are no longer necessarily independent conditional on X. Therefore, we follow Heinrich (1982) by expressing the characteristic function $\phi_{SN_T}(u)$ as follows

$$\phi_{SN_T}(u) = \prod_{t=1}^{T} \varphi_t(u)$$

where $\varphi_1(u) = \mathbb{E}_X\left[\exp\left(iu\left(\ln\left\{\frac{1-p_1[\boldsymbol{x}_0,\boldsymbol{\beta}_0,\boldsymbol{\beta}_1|X]}{p_1[\boldsymbol{x}_0,\boldsymbol{\beta}_0,\boldsymbol{\beta}_1|X]}\right\} s(\tilde{y}_1)\right)\right)\right]$ and for $t = 2,\cdots,T$

$$\varphi_t(u) = \frac{f_t(u)}{f_{t-1}(u)}, \quad \text{where,} \quad f_t(u) = \mathbb{E}_X\left[\exp(iuSN_t(\boldsymbol{\beta}_0 \mid \boldsymbol{\beta}_1))\right]$$

Heinrich (1982) shows that $\varphi_t(u)$ can alternatively be expressed as

$$\varphi_t(u) = \mathbb{E}_X\left[\exp\left(iu\left\{R_{t,t-1} + \ln\left\{\frac{1-p_t[\boldsymbol{x}_{t-1},\boldsymbol{\beta}_0,\boldsymbol{\beta}_1 \mid X]}{p_t[\boldsymbol{x}_{t-1},\boldsymbol{\beta}_0,\boldsymbol{\beta}_1 \mid X]}\right\} s(\tilde{y}_t)\right\}\right)\right] + \rho_t(u)$$

where

$$\begin{aligned}\rho_t(u) = \Big\{ & \mathbb{E}_X\left[\exp\left(iu\left\{SN_t\left(\boldsymbol{\beta}_0 \mid \boldsymbol{\beta}_1\right)\right\}\right)\right] - \\ & \mathbb{E}_X\left[\exp\left(iu\left\{R_{t,t-1} + \ln\left\{\frac{1-p_t[\boldsymbol{x}_{t-1},\boldsymbol{\beta}_0,\boldsymbol{\beta}_1 \mid X]}{p_t[\boldsymbol{x}_{t-1},\boldsymbol{\beta}_0,\boldsymbol{\beta}_1 \mid X]}\right\} s(\tilde{y}_t)\right\}\right)\right] \times \\ & \mathbb{E}_X\left[\exp\left(iu\left\{SN_{t-1}\left(\boldsymbol{\beta}_0 \mid \boldsymbol{\beta}_1\right)\right\}\right)\right]\Big\} \Big/ \mathbb{E}_X\left[\exp\left(iu\left\{SN_{t-1}\left(\boldsymbol{\beta}_0 \mid \boldsymbol{\beta}_1\right)\right\}\right)\right].\end{aligned}$$

Therefore, the characteristic function of the PCC-POS test statistic can be expressed as

$$\begin{aligned}\phi_{SN_T}(u) &= \prod_{t=1}^{T} \varphi_t(u) \\ &= \prod_{t=1}^{T}\left(\mathbb{E}_X\left[\exp\left(iu\left\{R_{t,t-1} + \ln\left\{\frac{1-p_t[\boldsymbol{x}_{t-1},\boldsymbol{\beta}_0,\boldsymbol{\beta}_1 \mid X]}{p_t[\boldsymbol{x}_{t-1},\boldsymbol{\beta}_0,\boldsymbol{\beta}_1 \mid X]}\right\} s(\tilde{y}_t)\right\}\right)\right] + \rho_t(u)\right),\end{aligned} \tag{40}$$

where $\rho_1(u) = 0$, $R_{1,0} = \rho_1(u) = 0$.

Let $Z_t = R_{t,t-1} + \ln\left\{\frac{1-p_t[\boldsymbol{x}_{t-1},\boldsymbol{\beta}_0,\boldsymbol{\beta}_1|X]}{p_t[\boldsymbol{x}_{t-1},\boldsymbol{\beta}_0,\boldsymbol{\beta}_1|X]}\right\} s(\tilde{y}_t)$ for $t = 1,\cdots,T$. Then following Heinrich (1982), and by assuming that $Z_1,\cdots,Z_T$ are weakly dependent, the term $\rho_t(u)$ can further be factorized. For instance, a case of such weakly dependent random variables for which a Theorem exists is the regularity

Markov type process (i.e. RMT-process). Let $\mathcal{B}_s^{s+m} = \sigma(Z_s, \cdots, Z_{s+m})$ be the Borel σ-field generated by $\{Z_t, t = s, \cdots, s+u\}$. The process $\{Z_t\}_{t=1,2,\cdots}$ is an RMT-process, if for $1 \leq s \leq t$, the uniform mixing coefficient $\phi(m) \leq \gamma(s,t)$ with probability one, where

$$\phi(m) \equiv \sup_{s \geq 1} \phi(\mathcal{B}_1^s, \mathcal{B}_{s+m}^{\infty})$$

and where $\phi(\mathcal{B}_1^s, \mathcal{B}_{s+m}^{\infty})$

$$\phi(\mathcal{B}_1^s, \mathcal{B}_{s+m}^{\infty}) \equiv \sup_{G \in \mathcal{B}_{s+m}^{\infty}, H \in \mathcal{B}_1^s} |P[H \mid G] - P[H]|,$$

with $\sup_{s \geq 1} \gamma(s, s+m) \to 0$ as $m \to \infty$. Given such dependence, $\rho_t(u)$ can be factorized using the results of Theorem 2 of Heinrich (1982).

The conditional CDF of $SN_T(\boldsymbol{\beta}_0 \mid \boldsymbol{\beta}_1)$ evaluated at a constant $c_1(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1)$, where $c_1(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1) \in \mathbb{R}$, given by the conditional characteristic functions $\phi_{SN_T}(u)$ can then be obtained using the Fourier-inversion formula [see Gil-Pelaez (1951)] as follows

$$P[SN_T(\boldsymbol{\beta}_0 \mid \boldsymbol{\beta}_1) \leq c_1(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1)] = \frac{1}{2} - \frac{1}{\pi} \int_0^{\infty} \frac{\mathrm{Im}\{\exp(-iuc_1(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1))\phi_{SN_T}(u)\}}{u} du$$

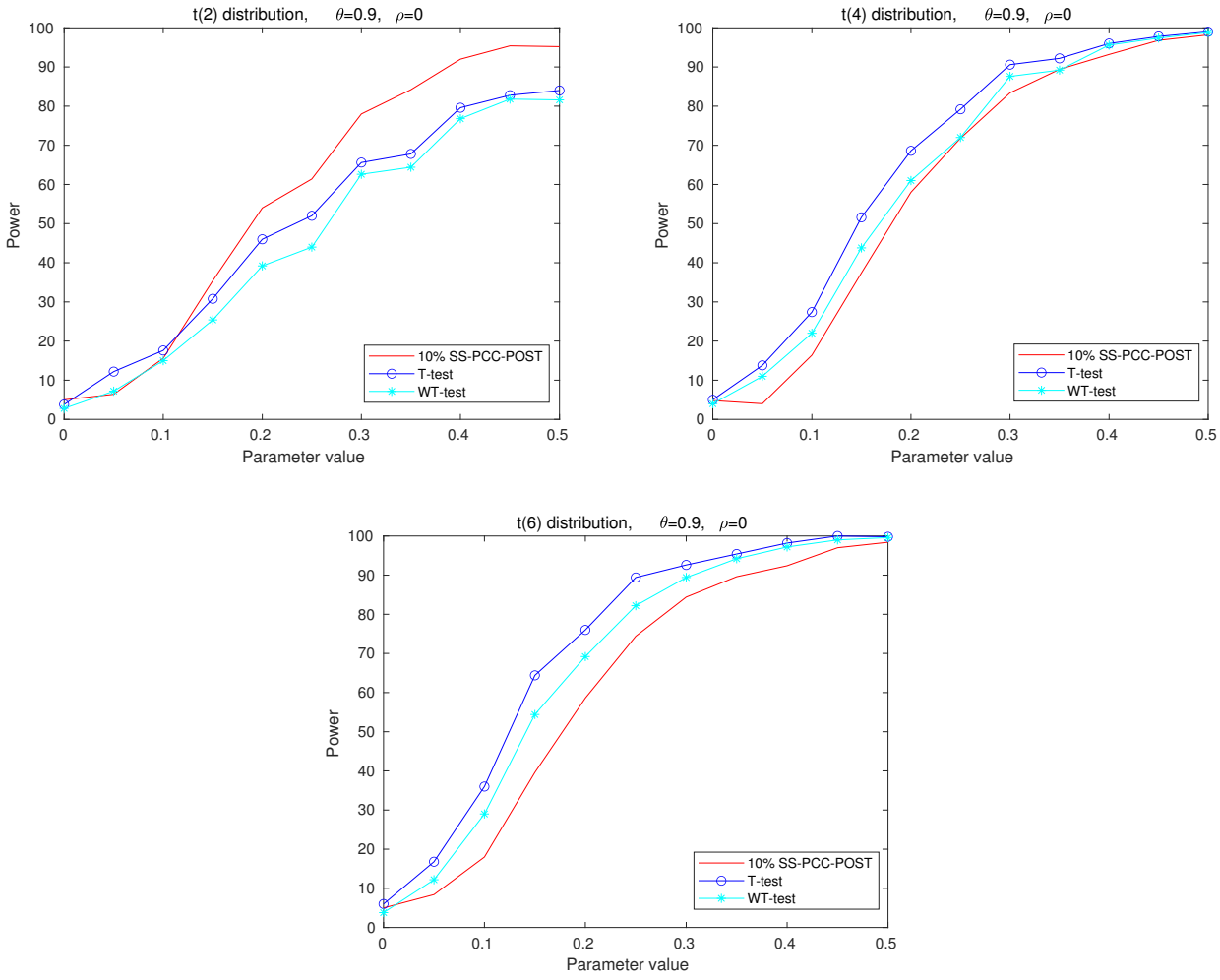
where $\forall u \in \mathbb{R}$, the conditional characteristic function $\phi_{SN_T}(u)$ is expressed by (40) and $\mathrm{Im}\{z\}$ denotes the imaginary part of the complex number z. Therefore, the power function can be obtained as follows

$$\Pi(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1) = P[SN_T(\boldsymbol{\beta}_0 \mid \boldsymbol{\beta}_1) > c_1(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1)] = \frac{1}{2} + \frac{1}{\pi} \int_0^{\infty} \frac{\mathrm{Im}\{\exp(-iuc_1(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1))\phi_{SN_T}(u)\}}{u} du$$

■

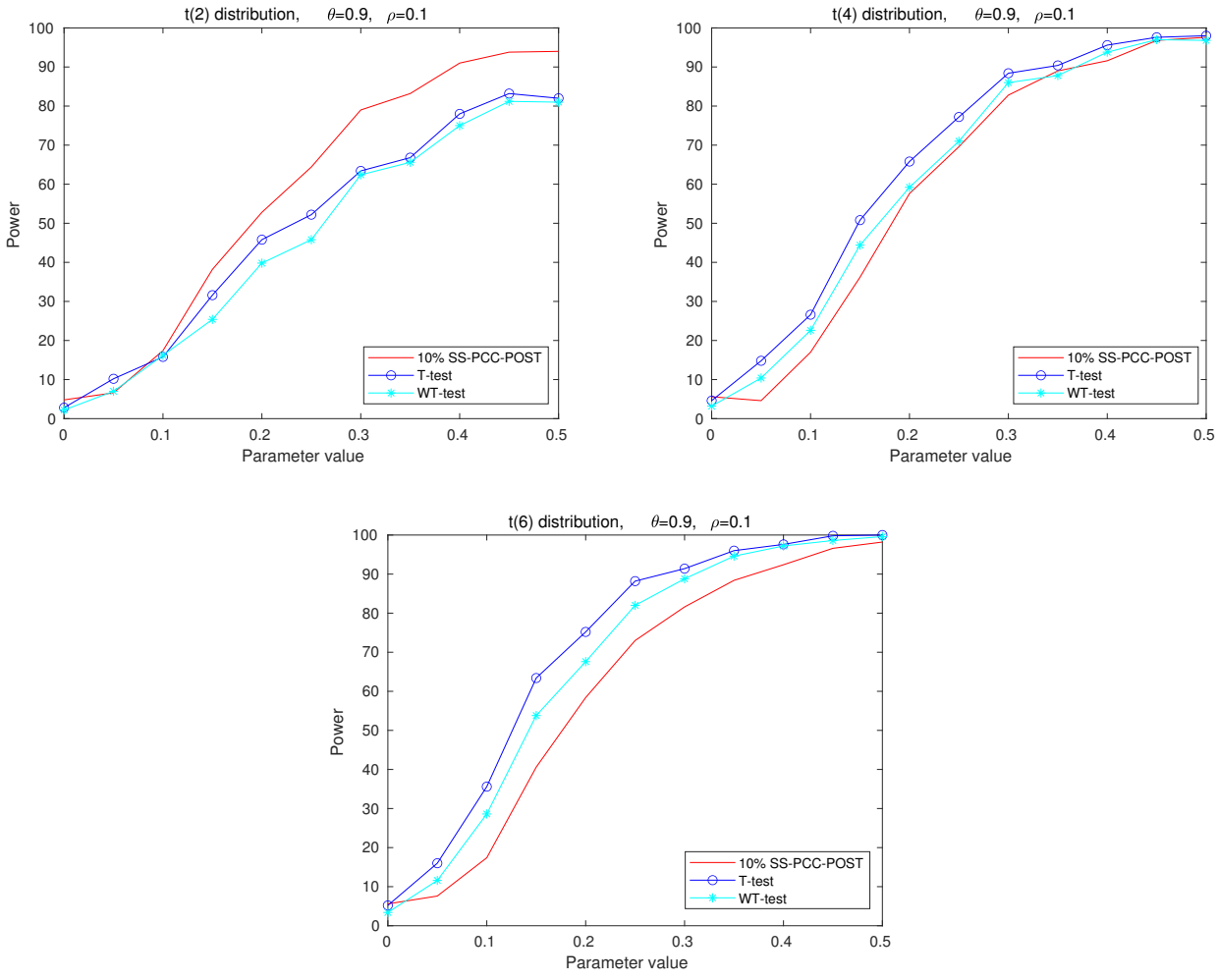
Additional simulations.

Figure 12: Power comparisons: different tests. Student's $t(\nu)$ error distributions, with different degrees of freedom ν, $\rho = 0$ in (33) and $\theta = 0.9$ in (32)



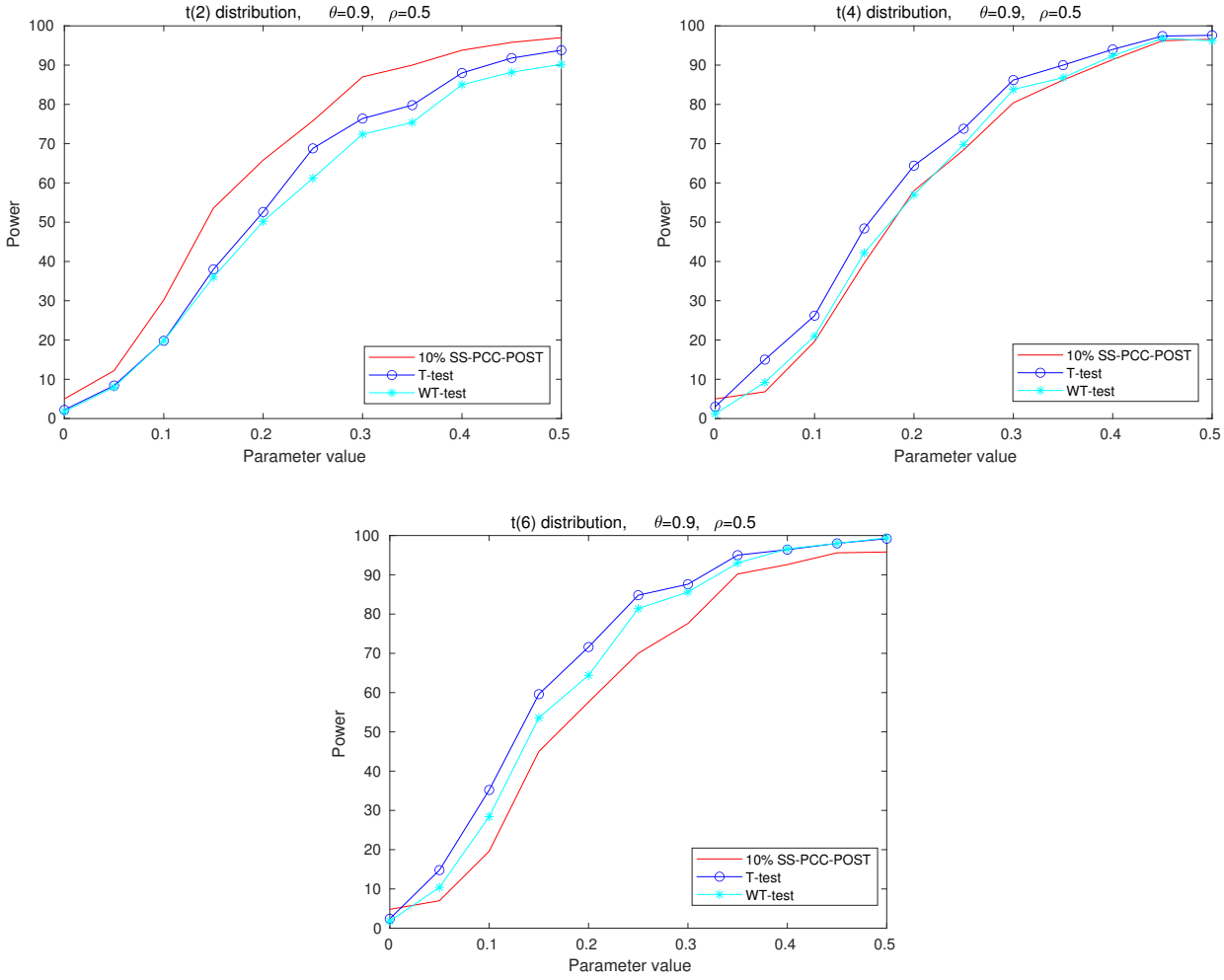
Note: These figures compare the power curves of the 10% split-sample PCC-POS test [10% SS-PCC-POS test] with: (1) the t-test and (2) the t-test based on White's (1980) variance correction [WT-test].

Figure 13: Power comparisons: different tests. Student's $t(\nu)$ error distributions, with different degrees of freedom ν, $\rho = 0.1$ in (33) and $\theta = 0.9$ in (32)



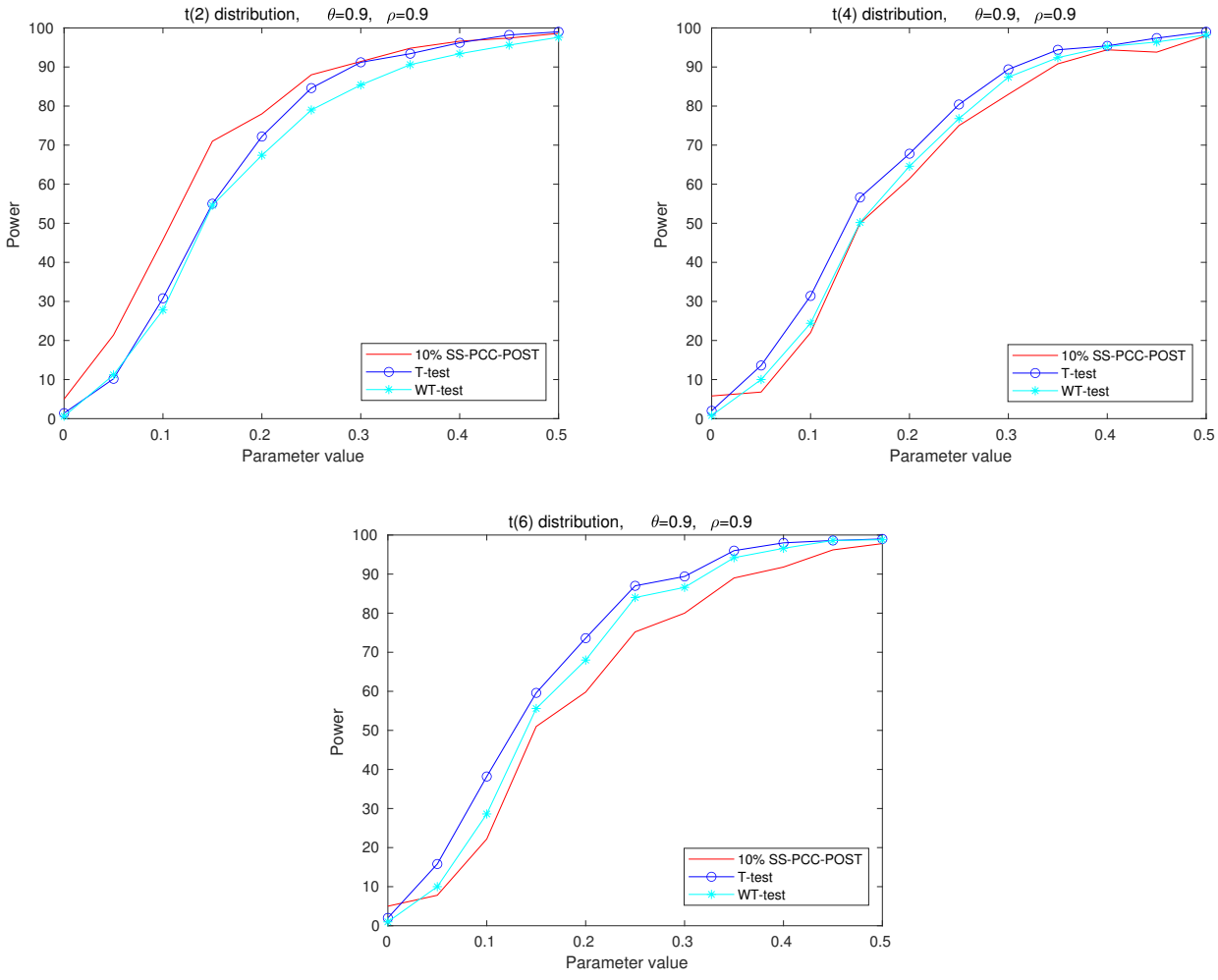
Note: These figures compare the power curves of the 10% split-sample PCC-POS test [10% SS-PCC-POS test] with: (1) the t-test and (2) the t-test based on White's (1980) variance correction [WT-test].

Figure 14: Power comparisons: different tests. Student's $t(\nu)$ error distributions, with different degrees of freedom ν, $\rho = 0.5$ in (33) and $\theta = 0.9$ in (32)



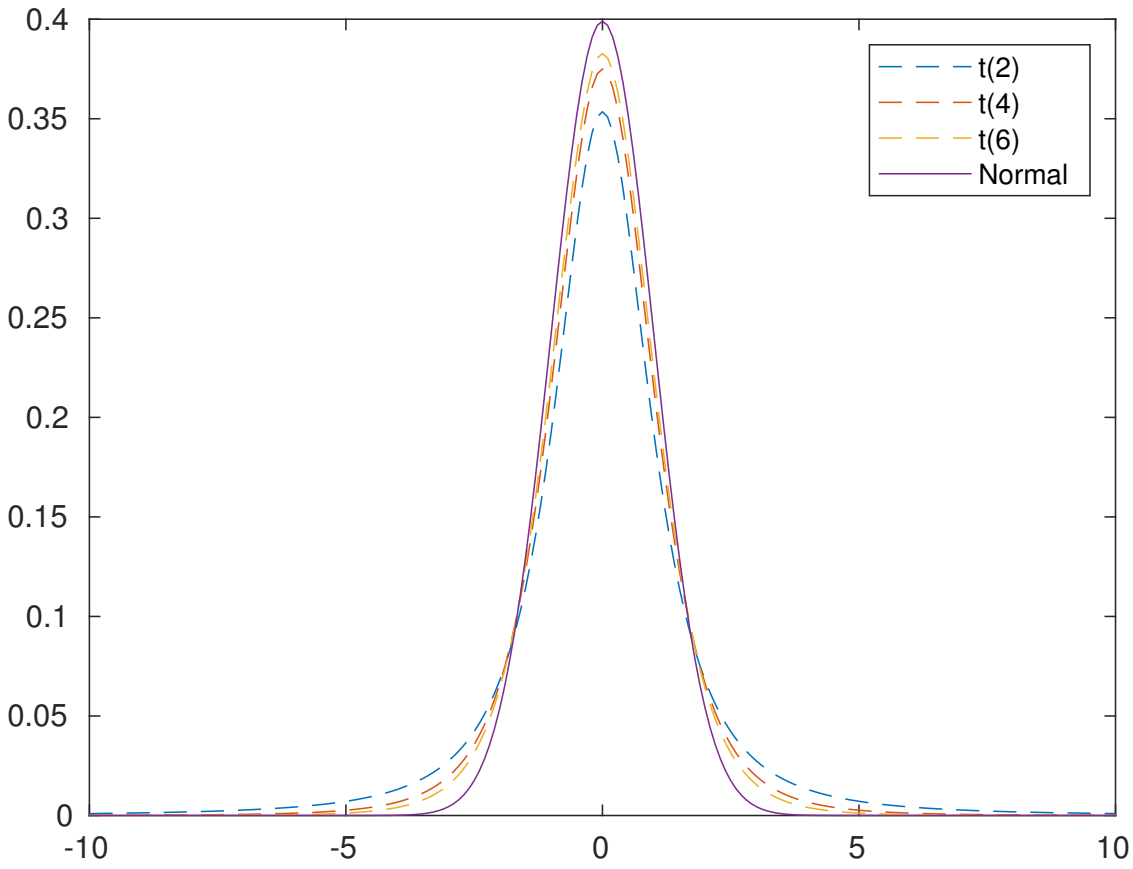
Note: These figures compare the power curves of the 10% split-sample PCC-POS test [10% SS-PCC-POS test] with: (1) the t-test and (2) the t-test based on White's (1980) variance correction [WT-test].

Figure 15: Power comparisons: different tests. Student's $t(\nu)$ error distributions, with different degrees of freedom ν, $\rho = 0.9$ in (33) and $\theta = 0.9$ in (32)



Note: These figures compare the power curves of the 10% split-sample PCC-POS test [10% SS-PCC-POS test] with: (1) the t-test and (2) the t-test based on White's (1980) variance correction [WT-test].

Figure 16: Comparison of the student's t distribution with various degrees of freedom to the normal distribution



Note: In this figure, we compare the Normal and Student's distribution with two, four and six degrees of freedom - i.e. $\nu = 2, \quad \nu = 4, \quad \nu = 6$.

■

References

Berg, D. (2009). Copula goodness-of-fit testing: an overview and power comparison. *The European Journal of Finance*, 15(7-8):675–701.

Boldin, M. V., Simonova, G. I., and Tyurin, I. N. (1997). *Sign-based methods in linear statistical models*, volume 162. American Mathematical Soc.

Campbell, B. and Dufour, J.-M. (1995). Exact nonparametric orthogonality and random walk tests. *The Review of Economics and Statistics*, 77(1):1–16.

Campbell, B. and Dufour, J.-M. (1997). Exact nonparametric tests of orthogonality and random walk in the presence of a drift parameter. *International Economic Review*, pages 151–173.

Coudin, E. and Dufour, J.-M. (2009). Finite-sample distribution-free inference in linear median regressions under heteroscedasticity and non-linear dependence of unknown form. *The Econometrics Journal*, 12:S19–S49.

Czado, C., Schepsmeier, U., and Min, A. (2012). Maximum likelihood estimation of mixed c-vines with application to exchange rates. *Statistical Modelling*, 12(3):229–255.

Denuit, M. and Lambert, P. (2005). Constraints on concordance measures in bivariate discrete data. *Journal of Multivariate Analysis*, 93(1):40–57.

Dissmann, J., Brechmann, E. C., Czado, C., and Kurowicka, D. (2013). Selecting and estimating regular vine copulae and application to financial returns. *Computational Statistics & Data Analysis*, 59:52–69.

Dufour, J., Jouneau, F., and Torrès, O. (2008). On (non)-testability: Applications to linear and nonlinear semiparametric and nonparametric regression models. Technical report, Technical report, Department of Economics and CIREQ, McGill University, and ....

Dufour, J.-M. and Jasiak, J. (2001). Finite sample limited information inference methods for structural equations and models with generated regressors. *International Economic Review*, 42(3):815–844.

Dufour, J.-M. and Taamouti, A. (2010). Exact optimal inference in regression models under heteroskedasticity and non-normality of unknown form. *Computational Statistics & Data Analysis*, 54(11):2532–2553.

Dufour, J.-M. and Taamouti, M. (2003). Point-optimal instruments and generalized anderson-rubin procedures for nonlinear models. Technical report, Technical report, CRDE, Université de Montréal.

Genest, C., Quessy, J.-F., and Rémillard, B. (2006). Goodness-of-fit procedures for copula models based on the probability integral transformation. *Scandinavian Journal of Statistics*, 33(2):337–366.

Genest, C., Rémillard, B., and Beaudoin, D. (2009). Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and economics*, 44(2):199–213.

Gil-Pelaez, J. (1951). Note on the inversion theorem. *Biometrika*, 38(3-4):481–482.

Haff, I. H. (2012). Comparison of estimators for pair-copula constructions. *Journal of Multivariate Analysis*, 110:91–105.

Haff, I. H., Aas, K., and Frigessi, A. (2010). On the simplified pair-copula construction—simply useful or too simplistic? *Journal of Multivariate Analysis*, 101(5):1296–1310.

Heinrich, L. (1982). Factorization of the characteristic function of a sum of dependent random variables. *Lithuanian Mathematical Journal*, 22(1):92–100.

Joe, H. (1996). Families of m-variate distributions with given margins and m (m-1)/2 bivariate dependence parameters. *Lecture Notes-Monograph Series*, pages 120–141.

Joe, H. (1997). *Multivariate models and multivariate dependence concepts*. CRC Press.

Joe, H. (2014). *Dependence modeling with copulas*. Chapman and Hall/CRC.

King, M. L. (1987). Towards a theory of point optimal testing. *Econometric Reviews*, 6(2):169–218.

Lehmann, E. L. and Romano, J. P. (2006). *Testing statistical hypotheses*. Springer Science & Business Media.

Li, J. and Wong, W. K. (2010). Two-dimensional toxic dose and multivariate logistic regression, with application to decompression sickness. *Biostatistics*, 12(1):143–155.

Luger, R. (2003). Exact non-parametric tests for a random walk with unknown drift under conditional heteroscedasticity. *Journal of Econometrics*, 115(2):259–276.

Magdalinos, T. and Phillips, P. C. (2009). Limit theory for cointegrated systems with moderately integrated and moderately explosive regressors. *Econometric Theory*, 25(2):482–526.

Mankiw, N. G. and Shapiro, M. D. (1986). Do we reject too often?: Small sample properties of tests of rational expectations models. *Economics Letters*, 20(2):139–145.

Maronna, R. A., Martin, R. D., Yohai, V. J., and Salibián-Barrera, M. (2019). *Robust statistics: theory and methods (with R)*. John Wiley & Sons.

Nelson, C. R. and Kim, M. J. (1993). Predictable stock returns: The role of small sample bias. *The Journal of Finance*, 48(2):641–661.

Nikoloulopoulos, A. K. and Karlis, D. (2008). Multivariate logit copula model with an application to dental data. *Statistics in Medicine*, 27(30):6393–6406.

Nikoloulopoulos, A. K. and Karlis, D. (2009). Finite normal mixture copulas for multivariate discrete data modeling. *Journal of Statistical Planning and Inference*, 139(11):3878–3890.

O'Gorman, T. W. (2004). *Applied adaptive statistical methods: tests of significance and confidence intervals*, volume 12. Siam.

Oh, D. H. and Patton, A. J. (2016). High-dimensional copula-based distributions with mixed frequency data. *Journal of Econometrics*, 193(2):349–366.

Panagiotelis, A., Czado, C., and Joe, H. (2012). Pair copula constructions for multivariate discrete data. *Journal of the American Statistical Association*, 107(499):1063–1072.

Phillips, P. C. (2014). On confidence intervals for autoregressive roots and predictive regression. *Econometrica*, 82(3):1177–1195.

Phillips, P. C. and Lee, J. H. (2013). Predictive regression under various degrees of persistence and robust long-horizon regression. *Journal of Econometrics*, 177(2):250–264.

Pratt, J. W. and Gibbons, J. D. (2012). *Concepts of nonparametric theory*. Springer Science & Business Media.

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

Song, P. X.-K., Li, M., and Yuan, Y. (2009). Joint regression analysis of correlated data using gaussian copulas. *Biometrics*, 65(1):60–68.

Stambaugh, R. F. (1999). Predictive regressions. *Journal of Financial Economics*, 54(3):375–421.

Stoeber, J., Joe, H., and Czado, C. (2013). Simplified pair copula constructions—limitations and extensions. *Journal of Multivariate Analysis*, 119:101–118.

Taamouti, A. (2015). Finite-sample sign-based inference in linear and nonlinear regression models with applications in finance. *L'Actualité économique*, 91(1-2):89–113.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *econometrica*, 48(4):817–838.