UNIVERSITY *of York*

This is a repository copy of *Time series copula models using d-vines and v-transforms*.

White Rose
university consortium
Universities of Leeds, Sheffield & York

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Time series copula models using d-vines and v-transforms

Martin Bladt [a,*], Alexander J. McNeil [b]

[a] *Faculty of Business and Economics, University of Lausanne, Dorigny, Quartier, 1015 Lausanne, Switzerland*
[b] *The York Management School, University of York*

ABSTRACT

An approach to modelling volatile financial return series using stationary d-vine copula processes combined with Lebesgue-measure-preserving transformations known as v-transforms is proposed. By developing a method of stochastically inverting v-transforms, models are constructed that can describe both stochastic volatility in the magnitude of price movements and serial correlation in their directions. In combination with parametric marginal distributions it is shown that these models can rival and sometimes outperform well-known models in the extended GARCH family.[1]

1. Introduction

The concept of a v-transform (McNeil, 2021) facilitates the application of copula models to time series where the dominant feature is stochastic volatility, such as financial asset return series. In the copula modelling approach to a single time series $\{x_1, \ldots, x_n\}$ the idea is to find an appropriate strictly stationary stochastic process (X_t) consisting of a continuous marginal distribution F_X and a copula process (U_t), given by $U_t = F_X(X_t)$ for all t, which models the serial dependencies in the data; the latter is a process of standard uniform variables with higher-dimensional marginal distributions that may be described by a family of copulas $C_U(u_1, \ldots, u_d)$ for $d \geqslant 2$.

For volatile return data, it is well known that serial dependence becomes more apparent under transformations like the absolute-value transformation $T(x) = |x|$ or the squared-value transformation $T(x) = x^2$, which remove directional information and summarise the magnitude of price movements in what we term a volatility proxy time series.

In McNeil (2021) general asymmetric volatility proxy transformations $T(x)$ with change points μ_T are considered; these are continuous functions which are increasing in $(x - \mu_T)$ for $x > \mu_T$ and increasing in $(\mu_t - x)$ for $x \leqslant \mu_T$. Under such transformations, the relationship between the terms of the copula process (U_t) of (X_t) and the terms of the copula process (V_t) of the volatility proxy process $(T(X_t))$ can be described by a v-shaped function, known as a v-transform, which

---

is a mapping $\mathcal{V}: [0, 1] \to [0, 1]$ that preserves the uniformity of uniform random variables. The relationships between the transformations are shown in diagram (1).

$$
\begin{array}{ccc}
X_t & \xrightarrow{\;\; F_X \;\;} & U_t \\
\downarrow{\scriptstyle T} & & \downarrow{\scriptstyle \mathcal{V}} \\
T(X_t) & \xrightarrow{\;\; F_{T(X)} \;\;} & V_t
\end{array}
\tag{1}
$$

The key idea is that, rather than modelling serial dependence of the time series at the level of (U_t), we can model it at the level of (V_t) and create a composite copula model consisting of a family of copulas $C_V(v_1, \ldots, v_d)$, $d = 2, \ldots, n$ and a v-transform $\mathcal{V}$. In McNeil (2021) C_V is modelled using the implied copula process of an ARMA model while in this paper we apply d-vine copula models (Aas et al., 2009; Smith et al., 2010).

The resulting copula models, when combined with suitable marginal distributions, yield an extremely flexible class of non-linear time series models for volatile data. In common with the popular GARCH family (Engle, 1982; Bollerslev, 1986; Ding et al., 1993; Glosten et al., 1993) and models from the more general GAS (generalized autoregressive score) family (Creal et al., 2013), the resulting models are observation-driven volatility models, but they employ a rather different mechanism in which the key feature is the strict separation of marginal and serial dependence modelling. In the GARCH paradigm the mechanism employed to capture serial dependence behaviour can have a constraining effect on the resulting marginal behaviour, which is only partly mitigated by varying the innovation distribution. For example, the standard GARCH model yields marginal distributions with tails following power laws for a wide class of innovation distributions (Mikosch and Stărică, 2000) and we will show in our examples that these are not always appropriate for observed return series.

There is a large literature on copula models for time series and good starting points are the comprehensive review papers by Patton (2012) and Fan and Patton (2014). While the main focus of this literature has been on cross-sectional dependence between multiple time series, there is also a growing literature on modelling serial dependence within single series and lagged dependence across series. Markov copula models were first investigated by Darsow et al. (1992) and are further studied in Chen and Fan (2006). In the latter paper and in Chen et al. (2009) the theory of semi-parametric estimation for these models is developed, while Beare (2010) studies mixing properties of the resulting processes; an application to data is given in Domma et al. (2009). Although theoretically interesting, first-order Markov models are not realistic candidates for modelling the persistent dependence and stochastic volatility that is found in typical financial return series.

A distinct approach to copula modelling in statistics uses pair-copula constructions. A key reference for applications in risk modelling is Aas et al. (2009), which builds on underpinning work on joint density decompositions by Joe (1996, 1997) and on graphical dependence models by Bedford and Cooke (2001a, 2001b, 2002) and Kurowicka and Cooke (2006). The application of this methodology to modelling longitudinal dependence in time series with d-vines was developed in Smith et al. (2010) and the extension of this approach to bivariate processes with both serial and cross-sectional dependence using alternative vine structures is treated in Beare and Seo (2015), Brechmann and Czado (2015), Smith (2015) and Nagler et al. (2020).

The first-order Markov models in Chen and Fan (2006) are special cases of the d-vine copula approach of Smith et al. (2010) but the general d-vine pair-copula model allows higher-order Markov dependence of the kind analysed in Ibragimov (2009). In applications of these models, the pair-copula building blocks used by researchers have tended to be limited to a number of well known bivariate copulas, such as the Gumbel, Clayton, Gaussian, Frank, Joe and t copulas, as well as rotations of certain copulas through 90, 180 and 270 degrees. With the possible exception of the t copula with very small degrees of freedom ($\nu < 1$), these basic copulas are not particularly effective at capturing the particular forms of serial dependence created by stochastic volatility, in which large price movements are followed be other large price movements, but of frequently changing sign.

Loaiza-Maya et al. (2018) observe that these sign changes tend to lead to lag-plots of log returns on the copula scale that are cross-shaped. They address the shortcomings of standard pair copulas in d-vine models by creating mixtures of pair copulas and rotated pair copulas which can emulate these cross-shaped patterns. In this paper we will show that the standard copulas can be combined with v-transforms and d-vines to offer a parsimonious method of obtaining a similar effect. Moreover the approach has more econometric interpretability in that the driver of serial dependence is identified with a volatility proxy series.

The contributions of the paper are threefold: we extend the theory of v-transformed copula processes as presented in McNeil (2021) to allow models that can describe both the phenomenon of stochastic volatility, as well as serial correlation in the direction of price movements; we show how to apply the modelling framework to copula processes based on d-vines and develop an approach to estimation; we demonstrate that the resulting models, when combined with suitable marginal distributions, can rival and sometimes outperform popular models in the GARCH class.

The paper is structured as follows. In Section 2 we extend the theory of copula processes constructed using v-transforms; in particular, we explore a generalization of the concept of stochastic inversion of v-transforms. Section 3 shows how the theory applies to d-vine copula processes and Section 4 explains our approach to the estimation of models and empirical examples are presented in Section 5. We apply the fitted models to value-at-risk (VaR) estimation and analyse their out-of-sample forecasting performance in Section 6; Section 7 concludes.

2. V-transforms and time series copula processes

2.1. V-transforms of uniform random variables

In McNeil (2021) three equivalent definitions of v-transforms are provided. Suppose we consider absolutely continuous and strictly increasing cdfs F_X on $\mathbb{R}$ and volatility proxy transformations T that are (i) continuous, (ii) strictly increasing for $x \geqslant \mu_T$, (iii) strictly decreasing for $x < \mu_T$ and (iv) differentiable everywhere except at a change point parameter μ_T which may or may not be zero; examples are $T(x) = |x|$ as well as alternatives that are asymmetric around μ_T. Then a v-transform is as a function $\mathcal{V} : [0, 1] \to [0, 1]$ constructed from F_X and T by

$$\mathcal{V}(u) = F_{T(X)}\big(T\big(F_X^{-1}(u)\big)\big), \tag{2}$$

where $F_{T(X)}$ denotes the cdf of $T(X)$ for any random variable X with cdf F_X. Observe that the composite transformation in (2) represents an excursion round three sides of the rectangle in diagram (1), from top right to bottom right.

$\mathcal{V}$ is thus a mapping of the probability-integral transform (PIT transform) of X to the PIT transform of $T(X)$ since $\mathcal{V}(F_X(X)) = F_{T(X)}(T(X))$. Clearly, by the properties of the PIT transform, such a transformation will preserve the uniformity of uniform random variables: if $U \sim U(0, 1)$ and $V = \mathcal{V}(U)$ then $V \sim U(0, 1)$. More formally, $\mathcal{V}$ is a Lebesgue measure-preserving transformation on the Borel subsets of [0,1].

A more visually interpretable definition is the following:

Definition 1. A v-transform is a mapping $\mathcal{V} : [0, 1] \to [0, 1]$ with the following properties:

1. $\mathcal{V}(0) = \mathcal{V}(1) = 1$;
2. There exists a point δ known as the fulcrum such that $0 < \delta < 1$ and $\mathcal{V}(\delta) = 0$;
3. $\mathcal{V}$ is continuous;
4. $\mathcal{V}$ is strictly decreasing on $[0, \delta]$ and strictly increasing on $[\delta, 1]$;
5. Every point $u \in [0, 1] \setminus \{\delta\}$ has a dual point u^* on the opposite side of the fulcrum satisfying $\mathcal{V}(u) = \mathcal{V}(u^*)$ and $|u^* - u| = \mathcal{V}(u)$ (square property).

Finally it is useful to have a definition which shows how v-transforms can be easily constructed and this is afforded by the following characterization.

Theorem 1. *A mapping $\mathcal{V} : [0, 1] \to [0, 1]$ is a v-transform if and only if it takes the form*

$$\mathcal{V}(u) = \begin{cases} (1 - u) - (1 - \delta)\Psi\big(\frac{u}{\delta}\big) & u \leqslant \delta, \\ u - \delta\Psi^{-1}\big(\frac{1-u}{1-\delta}\big) & u > \delta, \end{cases} \tag{3}$$

where Ψ is a continuous and strictly increasing distribution function on [0,1].

Parametric families of v-transforms may be obtained by assuming, for example, that Ψ is the cdf of a beta distribution or $\Psi(x) = \exp(-\kappa(-(\ln x)^\xi))$ for $\kappa > 0$ and $\xi > 0$, which is the main family considered in McNeil (2021). Both families include the important special case of the linear v-transform

$$\mathcal{V}_\delta(u) = \begin{cases} (\delta - u)/\delta & u \leqslant \delta, \\ (u - \delta)/(1 - \delta) & u > \delta, \end{cases} \tag{4}$$

which corresponds to a uniform cdf for Ψ and which subsumes the symmetric case $\mathcal{V}_{0.5}(u) = |2u - 1|$.

If we write a volatility proxy transformation in the form

$$T(x) = \begin{cases} T_1(\mu_T - x) & x \leqslant \mu_T, \\ T_2(x - \mu_T) & x > \mu_T, \end{cases}$$

for strictly increasing and continuous T_1 and T_2 satisfying $T_1(0) = T_2(0)$, then it can be shown that the v-transform $\mathcal{V}$ in (2) is determined by F_X, the value μ_T and the *profile function* $g_T(x) = T_2^{-1} \circ T_1(x)$. Different volatility proxy transformations may share the same change point μ_T and profile function g_T. For a fixed distribution F_X the pair (μ_T, g_T) partitions the set of volatility proxy transformations into equivalence classes, each corresponding to a unique v-transform. For example, the volatility proxy transformations $T(x) = |x|$, $T(x) = x^2$ and $T(x) = \ln|x|$ are all in the same equivalence class described by $\mu_T = 0$ and $g_T(x) = x$. From this point of view, having selected F_X, the selection of a v-transform amounts to an implicit choice of a class of volatility proxy transformations.

Example 1. Let F_X be the distribution function of a standard Student t distribution with 5 degrees of freedom. Consider the v-transform $\mathcal{V}$ obtained by using the generator $\Psi(x) = \exp(-\kappa(-(\ln x)^\xi))$ in (3) with parameter choices $\delta = 0.3$, $\kappa = 2.5$ and $\xi = 0.5$.

The left panel of Figure 1 shows $\mathcal{V}$ together with the admissible area (in white) corresponding to the fulcrum value $\delta = 0.3$; the restriction arises from the aforementioned square property. The centre and right panels show the implied volatility proxy transformation $T(x) = \Phi^{-1}(\mathcal{V}(F_X(x)))$, where Φ is the standard normal df, and the profile function g_T; note that T is just one possible member of the equivalence class defined by F_X and $\mathcal{V}$. The changepoint $\mu_T = F_X(\delta)$ is negative and is
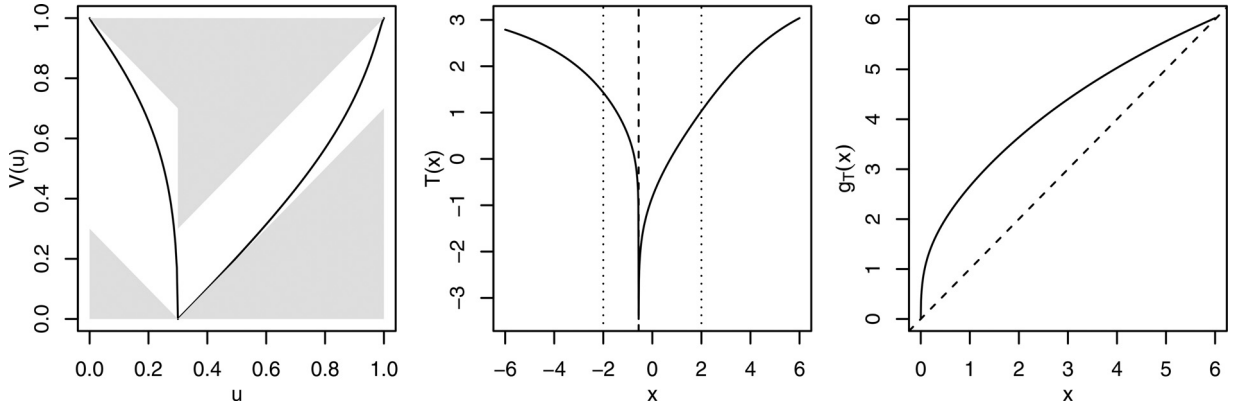
Fig. 1. Illustrations of the v-transform $\mathcal{V}$ the implied volatility proxy transformation T and the profile function g_T in Example 1.

marked by a dashed vertical line; two further dotted vertical lines show that an X value at -2 is associated with higher volatility than a value at 2. The profile function is determined by (see McNeil, 2021, Proposition 3)

$$g_T(x) = F_X^{-1}(F_X(\mu_T - x) + \mathcal{V}(F_X(\mu_T - x))) - \mu_T, \quad x \geqslant 0.$$

The dashed line shows the profile function for a symmetric volatility proxy transformation. In this example, the shape of the curve $g_T(x)$ indicates that, for $x \geqslant 0$, a realization at $\mu_T - x$ has a greater effect on volatility than a realization at $\mu_T + x$, but that this effect wears off for larger x.

Note that, as the fulcrum parameter $\delta \to 0$ in (3) the v-transform tends towards the function $\mathcal{V}_0(u) = u$ on (0,1] and as $\delta \to 1$ it tends towards the function $\mathcal{V}_1(u) = 1 - u$ on [0,1). We refer to $\mathcal{V}_0 : [0, 1] \to [0, 1]$, $\mathcal{V}_0(u) = u$ and $\mathcal{V}_1 : [0, 1] \to [0, 1]$, $\mathcal{V}_1(u) = 1 - u$ as degenerate v-transforms; the former corresponds to any strictly increasing transformations of a continuously distributed random variable X and the latter to any strictly decreasing function.

2.2. Stochastic inversion of a v-transform

Our aim is to develop stochastic processes for the process of uniform random variables (V_t) depicted in diagram (1), and hence to build processes (X_t) to model financial returns. To this end we need to be able to invert a v-transform $\mathcal{V}$, but this is complicated by the fact that $\mathcal{V}$ is not one-to-one. Stochastic inversion refers to the process of randomly reversing a v-transform to arrive back at one of the two dual points that yield the same value. To develop stochastic processes for (X_t) in which we have control over the marginal distribution, we need to be able to do this in such a way that uniformity is preserved under the stochastic inversion.

We introduce some further notation. Let $\mathcal{V}^{-1}$ denote the partial inverse given by $\mathcal{V}^{-1} : [0, 1] \to [0, \delta]$, $\mathcal{V}^{-1}(v) = \inf\{u : \mathcal{V}(u) = v\}$ and let $\mathcal{V}'$ denote the gradient of $\mathcal{V}$. The gradient of a v-function is defined for all points $u \in [0, 1] \setminus \{\delta\}$ and we adopt the convention that $\mathcal{V}'(\delta)$ is the left derivative as $u \to \delta$.

If two uniform random variables are linked by the v-transform $V = \mathcal{V}(U)$ then the joint distribution function of (U, V) is a special kind of copula. McNeil (2021) showed that, conditional on $V = v$,

$$U = \begin{cases} \mathcal{V}^{-1}(v) & \text{with probability } \Delta(v), \\ v + \mathcal{V}^{-1}(v) & \text{with probability } 1 - \Delta(v), \end{cases} \tag{5}$$

where the function

$$\Delta(v) = -\frac{1}{\mathcal{V}'(\mathcal{V}^{-1}(v))}. \tag{6}$$

is referred to as the conditional down probability of the v-transform and satisfies $\mathbb{E}(\Delta(V)) = \delta$. This allows the concept of *the stochastic inversion function* of a v-transform to be defined. This is simply a function that facilitates the construction of a Bernoulli event by which a value of V is randomly assigned to one of the dual points U and U^* such that $\mathcal{V}(U) = \mathcal{V}(U^*) = V$.

Definition 2 (Stochastic inversion function of a v-transform). Let $\mathcal{V}$ be a v-transform with conditional down probability $\Delta(\cdot)$. The two-place function $\boldsymbol{\mathcal{V}}^{-1} : [0, 1] \times [0, 1] \to [0, 1]$ defined by

$$\boldsymbol{\mathcal{V}}^{-1}(v, w) = \begin{cases} \mathcal{V}^{-1}(v) & \text{if } w \leqslant \Delta(v) \\ v + \mathcal{V}^{-1}(v) & \text{if } w > \Delta(v). \end{cases} \tag{7}$$

is the stochastic inversion function of $\mathcal{V}$.

It is obviously true that $\mathcal{V}(\mathcal{V}^{-1}(v, w)) = v$ for any w. It is also simple to show that if V and W are independent $U(0, 1)$ random variables and $U = \mathcal{V}^{-1}(V, W)$, then $U \sim U(0, 1)$. This is because

$$\mathbb{P}\big(\mathcal{V}^{-1}(V, W) = \mathcal{V}^{-1}(v) \mid V = v\big) = \mathbb{P}(W \leqslant \Delta(v) \mid V = v) = \mathbb{P}(W \leqslant \Delta(v)) = \Delta(v), \tag{8}$$

so U has the conditional distribution given in (5) and must be uniformly distributed.

When we apply a v-transform $\mathcal{V}(u)$ followed by a stochastic inversion of the v transform, then we either arrive back at the point u or at its dual point u^*. In the next result we consider the sequence of uniformity-preserving transformations $U \to \mathcal{V}(U) \to \mathcal{V}^{-1}(\mathcal{V}(U), W)$ for U and W independent and quantify the probability of arriving back at our starting point.

Proposition 1. *Let $U \sim U(0, 1)$ and $W \sim U(0, 1)$ be independent random variables and let $\mathcal{V}$ be a v-transform with fulcrum δ and conditional down probability $\Delta(v)$. If $V = \mathcal{V}(U)$ and $\tilde{U} = \mathcal{V}^{-1}(V, W)$ then*

$$\mathbb{P}\big(\tilde{U} = U\big) = \delta^2 + (1 - \delta)^2 + 2\operatorname{var}(\Delta(V)) \geqslant \delta^2 + (1 - \delta)^2 \geqslant 0.5 \,.$$

We see that the probability $\mathbb{P}(\tilde{U} = U)$ that we recover the original value of U is bounded below by $\delta^2 + (1 - \delta)^2$. This value is attained for the linear v-transform in (7) since $\Delta(v) = \delta$ for all v for that family. The global minimum value is 0.5, which is attained only for the symmetric v-transform $\mathcal{V}_{0.5}$. Interestingly, when asymmetry is present, there is a greater than 50% chance of recovering the original value.

2.3. V-transforms and inverse v-transforms of copulas

V-transforms and their stochastic inversion functions can be applied componentwise to random vectors. For vectors $\boldsymbol{u}$, $\boldsymbol{v}$ and $\boldsymbol{w}$ in $[0, 1]^d$ we write

$$\mathcal{V}(\boldsymbol{u}) = (\mathcal{V}(u_1), \ldots, \mathcal{V}(u_d))' \quad \text{and} \quad \mathcal{V}^{-1}(\boldsymbol{v}, \boldsymbol{w}) = (\mathcal{V}^{-1}(v_1, w_1), \ldots, \mathcal{V}^{-1}(v_d, w_d))',$$

for the componentwise operations. Let $\{(V_1, W_1), \ldots, (V_d, W_d)\}$ be a set of pairs of uniform random variables with the property that V_i is independent of W_i for all i. Note that these pairs need *not* be independent of each other. Let $\boldsymbol{U} = \mathcal{V}^{-1}(\boldsymbol{V}, \boldsymbol{W})$ where $\boldsymbol{V} = (V_1, \ldots, V_d)'$ and $\boldsymbol{W} = (W_1, \ldots, W_d)'$. Then we know that:

1. $\boldsymbol{U}$ is a uniform random vector or, in other words, its joint distribution is a copula. This is guaranteed by the independence of V_i and W_i for all i according to (8).
2. $\mathcal{V}(\boldsymbol{U}) = \boldsymbol{V}$ regardless of the exact nature of the joint distribution of $(\boldsymbol{V}, \boldsymbol{W})$ since $\mathcal{V}(\mathcal{V}^{-1}(v, w)) = v$ for all v, w. It is possible to create different joint models $(\boldsymbol{V}_1, \boldsymbol{W}_1)$ and $(\boldsymbol{V}_2, \boldsymbol{W}_2)$ such that $\boldsymbol{V}_1 \stackrel{\mathrm{d}}{=} \boldsymbol{V}_2$. In this case the implied copulas $\boldsymbol{U}_1 = \mathcal{V}^{-1}(\boldsymbol{V}_1, \boldsymbol{W}_1)$ and $\boldsymbol{U}_2 = \mathcal{V}^{-1}(\boldsymbol{V}_2, \boldsymbol{W}_2)$ are different but $\mathcal{V}(\boldsymbol{U}_1) \stackrel{\mathrm{d}}{=} \mathcal{V}(\boldsymbol{U}_2)$.

It is of interest to be able to determine the joint distribution of $\boldsymbol{U}$ under various assumptions about $(\boldsymbol{V}, \boldsymbol{W})$. We give a result for the general case as well as the case where these vectors are independent of each other. This generalizes a result given in McNeil (2021) for the case where $W_1, \ldots, W_d$ are also iid. To state this result compactly we introduce the notation

$$\delta(u) = \Delta(\mathcal{V}(u)), \quad I_{\delta,u}(x) = \begin{cases} [0, x] & u \leqslant \delta, \\ [x, 1] & u > \delta, \end{cases} \quad p_{\delta,u}(x) = \begin{cases} x & u \leqslant \delta, \\ 1 - x & u > \delta, \end{cases} \quad \delta \in (0, 1), \quad u, x \in [0, 1], \tag{9}$$

and the vector form of the latter $p_{\delta,\boldsymbol{u}}(\boldsymbol{x}) = \big(p_{\delta,u_1}(x_1), \ldots, p_{\delta,u_d}(x_d)\big)'$. Note that $\delta(u)$ is the probability that the v-transform of an observation at u is assigned to the left side of the fulcrum under stochastic inversion. We now state the main result of this section.

Theorem 2. *Let $\mathcal{V}$ be a v-transform and let $\{(V_1, W_1), \ldots, (V_d, W_d)\}$ be a set of pairs of uniform random variables with the property that V_i is independent of W_i for all i. Assume the copula of $(\boldsymbol{V}, \boldsymbol{W})$ has a joint density. Then the copula density $c_{\boldsymbol{U}}(u_1, \ldots, u_d)$ of $\boldsymbol{U} = \mathcal{V}^{-1}(\boldsymbol{V}, \boldsymbol{W})$ is*

$$\frac{\int_{I_{\delta,u_1}(\delta(u_1))} \cdots \int_{I_{\delta,u_d}(\delta(u_d))} c_{\boldsymbol{V},\boldsymbol{W}}(\mathcal{V}(u_1), \ldots, \mathcal{V}(u_d), z_1, \ldots, z_d) \mathrm{d}z_1 \cdots \mathrm{d}z_d}{\prod_{i=1}^{d} p_{\delta,u_i}\big(\delta(u_i)\big)}, \tag{10}$$

where $c_{\boldsymbol{V},\boldsymbol{W}}$ denotes the joint copula density of $(\boldsymbol{V}, \boldsymbol{W})$. When $\boldsymbol{V}$ and $\boldsymbol{W}$ are independent this reduces to the simpler form

$$c_{\boldsymbol{U}}(u_1, \ldots, u_d) = c_{\boldsymbol{V}}\big(\mathcal{V}(u_1), \ldots, \mathcal{V}(u_d)\big) \frac{C_{p_{\delta,\boldsymbol{u}}(\boldsymbol{W})}\Big(p_{\delta,u_1}\big(\delta(u_1)\big), \ldots, p_{\delta,u_d}\big(\delta(u_d)\big)\Big)}{\prod_{i=1}^{d} p_{\delta,u_i}\big(\delta(u_i)\big)}, \tag{11}$$

where $c_{\boldsymbol{V}}$ denotes the copula density of $\boldsymbol{V}$ and $C_{p_{\delta,\boldsymbol{u}}(\boldsymbol{W})}$ denotes the copula of $p_{\delta,\boldsymbol{u}}(\boldsymbol{W})$. When, in addition, $W_1, \ldots, W_d$ are independent, (11) reduces further to $c_{\boldsymbol{V}}(\mathcal{V}(u_1), \ldots, \mathcal{V}(u_d))$.

In the applied sections of this paper the focus will be on models of type (11) and it is instructive to consider the structure of the copula in more detail. For $d = 2$, Figure 2 illustrates $c_{\boldsymbol{V}}(\mathcal{V}(u_1), \mathcal{V}(u_2))$ and $c_{\boldsymbol{U}}(u_1, u_2)$ for particular choices of
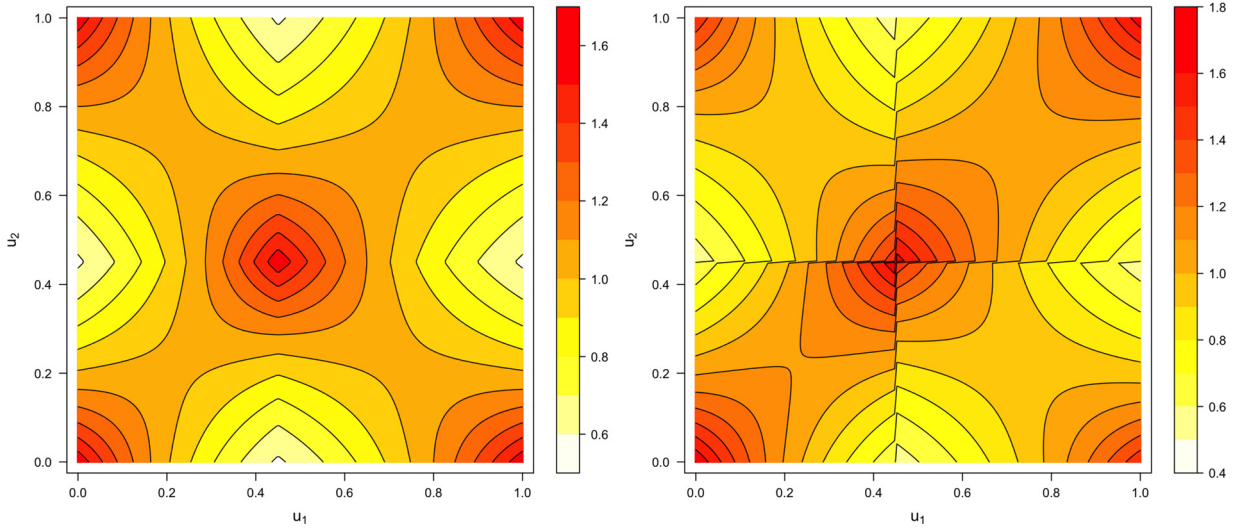
Fig. 2. Contour plot of $c_V(\mathcal{V}(u_1), \mathcal{V}(u_2))$ (left) and $c_U(u_1, u_2)$ (right) for model in which V follows Frank(1) copula and W follows Frank(0.5) copula; v-transform is linear with fulcrum $\delta = 0.45$.
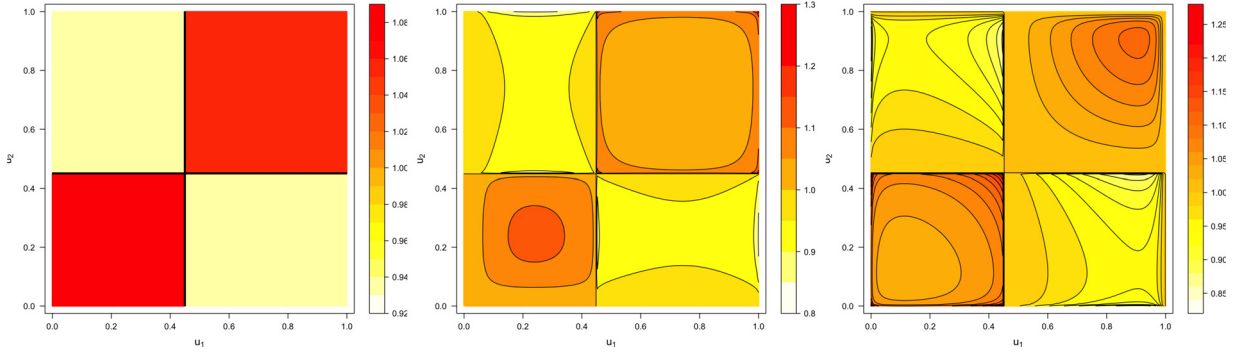

Fig. 3. Copula density $c_{W^*}(u_1, u_2)$ when W follows Frank(0.5) copula for three different v-transforms, all with fulcrum $\delta = 0.45$: linear (left); $\kappa = \xi = 1.5$ (middle); $\kappa = 1.5$, $\xi = 0.5$ (right).

parametric copulas for V and W and for the linear v-transform. We observe the characteristic cross-shape often observed in lag-plots for processes with stochastic volatility.

The density in (11) is itself the product of two copula densities, $c_V(\mathcal{V}(u_1), \ldots, \mathcal{V}(u_d))$ and the density

$$c_{W^*}(u_1, \ldots, u_d) := \frac{C_{p_{\delta,u}(W)}\Big(p_{\delta,u_1}\big(\delta(u_1)\big), \ldots, p_{\delta,u_d}\big(\delta(u_d)\big)\Big)}{\prod_{i=1}^d p_{\delta,u_i}\big(\delta(u_i)\big)} \, . \tag{12}$$

To see that this is a density observe that $c_U(u_1, \ldots, u_d) = c_{W^*}(u_1, \ldots, u_d)$ when V is a vector of independent uniform variables so that $c_V(v_1, \ldots, v_d) = 1$. In Figure 3 we illustrate, for three different choices of v-transform, the copula density $c_{W^*}(u_1, u_2)$, which controls the dependencies in directions of movements. It is interesting to note that the presence of serial dependence in W leads to a discontinuous copula density c_{W^*} in Figure 3 and a discontinuous density c_U in Figure 2 (although the copula itself is continuous). Empirical data sometimes show markedly different behaviour in the four quadrants of the unit square corresponding to the four possible combinations of positive and negative returns (with respect to the changepoint μ_T) at times t and $t + 1$. A discontinous copula density is one possible way of accommodating these differences.

2.4. V-transforms of time series copula processes

The theory presented in the previous section can obviously be applied to the construction of time series copula processes $(U_t)_{t \in \mathbb{Z}}$ that are suitable for modelling financial return data. By constructing a strictly stationary bivariate process $(V_t, W_t)_{t \in \mathbb{Z}}$ we obtain a strictly stationary process for (U_t) through the stochastic inversion construction of Theorem 2. In these models

(V_t) can be thought of as accounting mainly for stochastic volatility (serial dependence in the magnitude of movements) while (W_t), if it is not an iid process, can account for extra serial dependence in the direction of price movements.

We restrict attention to the special case where (V_t) and (W_t) are independent processes as this results in the tractable joint density for (U_t) in (11) and permits likelihood-based inference. Within this framework we consider models where (W_t) is either strict white noise (an iid process) or a first-order Markov process. In comparison with stochastic volatility, dependence in the signs of asset returns is a relatively weak and transient phenomenon and first-order Markov models for (W_t) appear to be sufficient in the majority of datasets we have considered.

The process (V_t) will be modelled by using a stationary d-vine copula process of Markov order k yielding the class of vt-d-vine models, which complement the vt-ARMA class of models in McNeil (2021). First-order Markov dependence in (W_t) will be modelled using a d-vine process of order $k = 1$, i.e. the kind of model considered in Chen and Fan (2006) and Domma et al. (2009).

3. D-vine and vt-d-vine copula processes

The aim of this section is to develop copula processes for (V_t) and (U_t) that fulfill the requirement of strict stationarity.

3.1. D-vine copula processes

Using the theory described in Smith et al. (2010) the multivariate copula density $c_{\boldsymbol{V}}$ of a random vector $\boldsymbol{V} = (V_1, \ldots, V_d)'$ can be decomposed as a d-vine taking the form

$$c_{\boldsymbol{V}}(v_1, \ldots, v_d) = \prod_{i=1}^{d-1} \prod_{t=1}^{d-i} c_{t,t+i|S_{t,t+i}}(v_{t|S_{t,t+i}}, v_{t+i|S_{t,t+i}}), \tag{13}$$

where $S_{t,t+i} = \{t+1, \ldots, t+i-1\}$ denotes the set of indices of the variables lying between V_t and V_{t+i}, $c_{t,t+i|S_{t,t+i}}$ is a pair copula density (i.e. a bivariate copula density) describing the dependence between variables V_t and V_{t+i} conditional on these variables and

$$v_{j|S_{t,t+i}} = \mathbb{P}(V_j \leqslant v_j | V_{t+1} = v_{t+1}, \ldots, V_{t+i-1} = v_{t+i-1}), \quad j \in \{t, t+i\},$$

denotes the conditional cdf of variable j conditional on the intermediate variables; note that $S_{t,t+1} = \emptyset$ and so the conditioning set is dropped in this case.

The decomposition (13) is not the unique d-vine expression for $c_{\boldsymbol{V}}(v_1, \ldots, v_d)$ when $d > 2$, since the variables $v_1, \ldots, v_d$ could be arranged in other orders. However, when the variables have a natural ordering, as they do for a time series, then (13) is the canonical expression for a d-vine. It should also be noted that there are other ways of decomposing a joint density using pair copulas which are not d-vines but which belong to the more general class of regular vines investigated in Bedford and Cooke (2001b); we do not consider these further as d-vines seem well adapted to the univariate time series context.

A final important point is that the decomposition (13) of an arbitrary joint density may result in pair copulas whose functional forms depend on the values of the conditioning variables in the sets $S_{j,i}$. However, in applied statistics, (13) is used as a framework for constructing rather than deconstructing models and interest is usually confined to so-called *simplified pair copula constructions* in which the copula forms are invariant to the values of the conditioning variables and are chosen from a number of well-known parametric families. In this case we can simplify the copula notation to $c_{t,t+i} = c_{t,t+i|S_{t,t+i}}$. There has been quite a lot of interest in the question of whether simplified constructions are sufficiently flexible and robust to model all dependence structures; see Haff et al. (2010), Stöber et al. (2013), Spanhel and Kurz (2019) and Mroz et al. (2021). While these authors draw attention to limitations, the simplifying assumption still admits a rich class of copulas that generalize the dependence inherent in classical autoregressive time series models, as explained below.

It is possible to construct strictly stationary time series $(V_t)_{t\in\mathbb{Z}}$ in which the d-dimensional joint densities of random vectors $(V_{t+1}, \ldots, V_{t+d})$ for $d \geqslant 2$ are given by the simplified form of the decomposition (13). In this case the stationarity requirement imposes the restriction that the pair copula densities $c_{t,t+i}$ may only depend on i and we can further simplify notation by writing $c_i = c_{t,t+i}$. Moreover, by setting $c_i = 1$ (corresponding to the independence copula) for $i > k$ we obtain Markov processes of order k and reduce the number of copulas that need to be determined. Models of this kind are investigated by Brechmann and Czado (2015) (under the name COPAR), Smith (2015) and Nagler et al. (2020); in the latter paper it is shown that d-vines are the only regular vines that can be used to construct stationary univariate time series.

Definition 3. A time series $(V_t)_{t\in\mathbb{Z}}$ is a d-vine(k) copula process if, for any $d \geqslant 2$ and $t \in \mathbb{Z}$, the d-dimensional marginal density of the random vector $(V_t, \ldots, V_{t+d-1})$ has the form

$$c_{\boldsymbol{V}}(v_1, \ldots, v_d) = \prod_{i=1}^{\min(k,d-1)} \prod_{t=1}^{d-i} c_i\big(v_{t|S_{t,t+i}}, \ v_{t+i|S_{t,t+i}}\big), \tag{14}$$

for a sequence of bivariate copula densities $c_1, \ldots, c_k$.

A d-vine(k) copula process is fully determined by the k copula densities corresponding to each of its conditional dependencies, or *generalized lags*. When $k = 1$, the copula process reduces to that of Chen and Fan (2006). When the copula densities are Gaussian, the dependence structure is that of a Gaussian AR(k) model, and the parameters of the pair copulas in the d-vine model are the partial autocorrelations of the underlying AR(k) process. In general, for $k > 1$, the main practical difficulty lies in the calculation of the expressions $v_{t|S_{t,t+i}}$ and $v_{t+i|S_{t,t+i}}$ for $i > 1$. This can be done using the recursive identities

$$v_{t|S_{t,t+i}} = h_{i-1}^{(2)}\big(v_{t|S_{t,t+i-1}}, \ v_{t+i-1|S_{t,t+i-1}}\big)$$
$$v_{t+i|S_{t,t+i}} = h_{i-1}^{(1)}\big(v_{t+1|S_{t+1,t+i}}, \ v_{t+i|S_{t+1,t+i}}\big),$$

$$(15)$$

where $h_i^{(j)}(v_1, v_2) = \frac{\partial}{\partial v_j} C_i(v_1, v_2)$, $j \in \{1, 2\}$, denotes the partial derivative or h-function of the copula C_i. Thus the problem is recursively reduced to the problem of evaluating h-functions of bivariate copulas (Joe, 1996). Note that when C_i is an exchangeable copula (satisfying $C_i(v_1, v_2) = C_i(v_2, v_1)$ for all v_1, v_2) the calculation is further simplified by the fact that $h_i^{(2)}(v_1, v_2) = h_i^{(1)}(v_2, v_1)$ but when C_i is non-exchangeable then both partial derivatives must be calculated explicitly.

The d-vine process of Definition 3 is strictly stationary by design but the questions of ergodicity and mixing are trickier. A number of authors including Chen and Fan (2006), Beare (2010) and Longla and Peligrad (2012) have studied the first-order process ($k = 1$). Longla and Peligrad (2012) show that, if the density of the absolutely continuous part of a copula is strictly positive almost everywhere, the resulting process is β-mixing (absolutely regular) and therefore ergodic. A variety of results have been obtained on the rate of mixing and these typically depend on the tail dependence characteristics of the copula. The Gauss and Frank copulas can be shown to be ϕ-mixing (Longla and Peligrad, 2012) and thus geometrically β-mixing, while the Gumbel, Clayton and t copulas satisfy the weaker property of geometric ρ-mixing (Beare, 2010); see Bradley (2005) for more details of mixing conditions. Zhao et al. (2018) extend the Markov chain approach to models with order $k > 1$ and give general conditions for geometric ergodicity, but do not address specific copula choices.

3.2. Vt-d-vine copula processes

Definition 4. Let $\mathcal{V}$ be a v-transform, let $(V_t)_{t \in \mathbb{Z}}$ be a d-vine(k) copula process and let $(W_t)_{t \in \mathbb{Z}}$ be any strictly stationary copula process that is independent of (V_t). Let $(U_t)_{t \in \mathbb{Z}}$ be defined componentwise by setting $U_t = \mathcal{V}^{-1}(V_t, W_t)$.

1. If (W_t) is an iid process (strict white noise) we say that (U_t) is a vt-d-vine(k) copula process.
2. Otherwise (U_t) is a generalized vt-d-vine(k) copula process, or a gvt-d-vine(k) process.

It follows immediately from Theorem 2, by inserting the d-vine(k) density (14) in the general expression (11) and using the notation (12), that the joint density of a gvt-d-vine(k) process is

$$c_{\boldsymbol{U}}(u_1, \ldots, u_d) = c_{\boldsymbol{W}^*}(u_1, \ldots, u_d) \prod_{i=1}^{\min(k, d-1)} \prod_{t=1}^{d-i} c_i\big(v_{t|S_{t,t+i}}, \ v_{t+i|S_{t,t+i}}\big)\big|_{v_1 = \mathcal{V}(u_1), \ldots, v_d = \mathcal{V}(u_d)}.$$

$$(16)$$

Remark 1. Under the degenerate v-transform $\mathcal{V}_0$ with fulcrum set at zero we have that $\mathcal{V}_0(u) = u$ for $u \in (0, 1]$. In this case the joint density of a vt-d-vine(k) process reduces to $c_{\boldsymbol{U}}(u_1, \ldots, u_d) = c_{\boldsymbol{V}}(u_1, \ldots, u_d)$ so that a d-vine(k) copula process model can be considered as a boundary case of a vt-d-vine(k) model.

As noted earlier, we will use d-vine(1) models for the process (W_t). In this case the term $c_{\boldsymbol{W}^*}(u_1, \ldots, u_d)$ in the joint density (16) can be fully expressed in terms of bivariate copula densities. When $d = 2$ let the joint distribution function of (W_1, W_2) be denoted $C_W = C_{\boldsymbol{W}}$ and let $c_{W^*} = c_{\boldsymbol{W}^*}$. Furthermore let $C_{W'}$, $C_{W''}$ $C_{W'''}$ denote the joint cdfs of $(W_1, 1 - W_2)$, $(1 - W_1, 1 - W_2)$ and $(1 - W_1, W_2)$ respectively; these are the copulas obtained by rotating the distribution described by C_W through 90, 180 and 270 degrees clockwise. We then have the following expression for $c_{\boldsymbol{W}^*}$.

Proposition 2. *When (W_t) is first-order Markov*

$$c_{\boldsymbol{W}^*}(u_1, \ldots, u_d) = \prod_{i=2}^{d} c_{W^*}(u_{i-1}, u_i), \quad c_{W^*}(u_1, u_2) = \begin{cases} \frac{C_W\big(\delta(u_1), \delta(u_2)\big)}{\delta(u_1)\delta(u_2)} & u_1 \leqslant \delta, u_2 \leqslant \delta, \\ \frac{C_{W'}\big(\delta(u_1), 1-\delta(u_2)\big)}{\delta(u_1)(1-\delta(u_2))} & u_1 \leqslant \delta, u_2 > \delta, \\ \frac{C_{W''}\big(1-\delta(u_1), 1-\delta(u_2)\big)}{(1-\delta(u_1))(1-\delta(u_2))} & u_1 > \delta, u_2 > \delta, \\ \frac{C_{W'''}\big(1-\delta(u_1), \delta(u_2)\big)}{(1-\delta(u_1))\delta(u_2)} & u_1 > \delta, u_2 \leqslant \delta. \end{cases}$$

$$(17)$$

The conditional density of the resulting gvt-d-vine(k) copula process can be calculated from (16) and (17) and takes the form
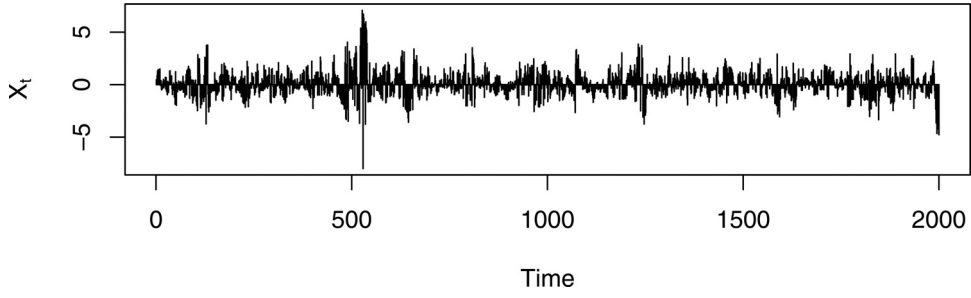
Fig. 4. Time series plot for a sample of size $n = 2000$ from the process described in Example 2.
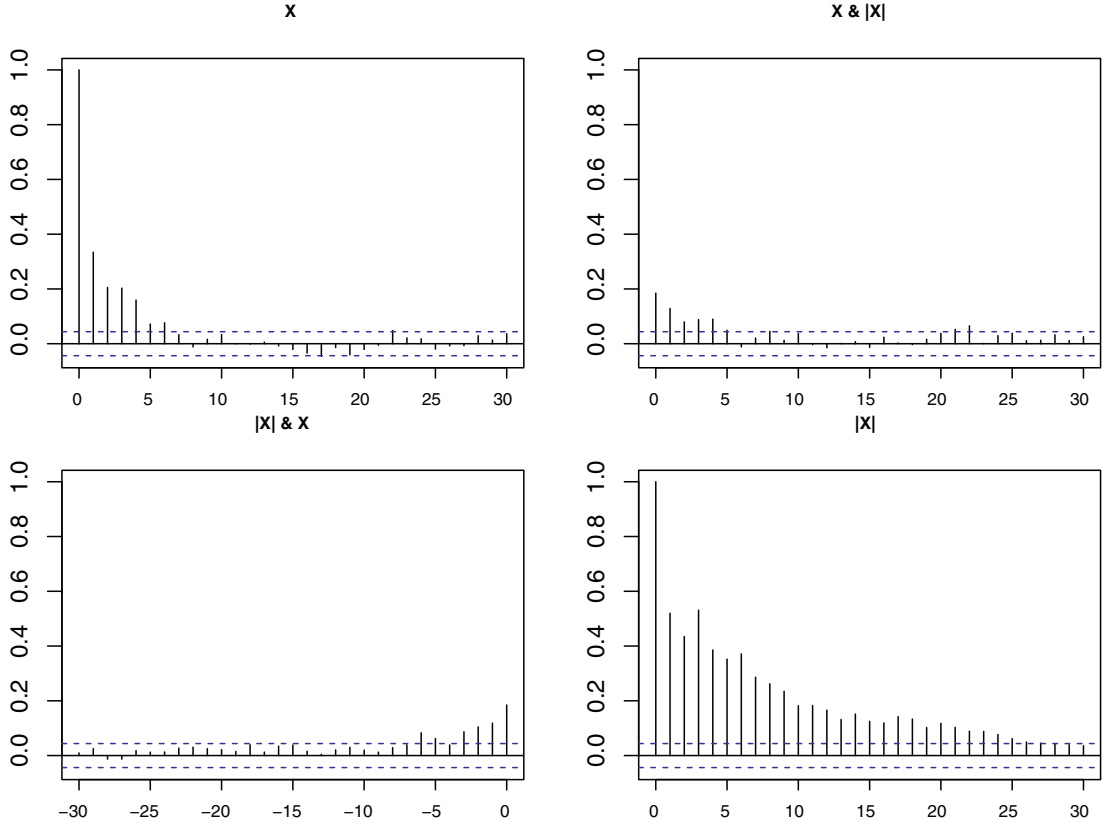
Fig. 5. Sample acf plots for a sample of size $n = 2000$ from the process described in Example 2.

$$f_{U_t|U_{t-1},\ldots,U_{t-k}}(u_t|u_{t-1},\ldots,u_{t-k}) = \frac{c_{\boldsymbol{U}}(u_{t-k},\ldots,u_t)}{c_{\boldsymbol{U}}(u_{t-k},\ldots,u_{t-1})}$$

$$= c_{W^*}(u_{t-1},u_t)\prod_{i=1}^{k} c_i(v_{t-i|S_{t-i,t}},v_{t|S_{t-i,t}})\Big|_{v_{t-k}=\mathcal{V}(u_{t-k}),\ldots,v_t=\mathcal{V}(u_t)}. \tag{18}$$

For $k = 1$ this is simply $f_{U_t|U_{t-1}}(u_t|u_{t-1}) = c_{\boldsymbol{U}}(u_{t-1},u_t) = c_{W^*}(u_{t-1},u_t)c_1(\mathcal{V}(u_{t-1}),\mathcal{V}(u_t))$ which is the conditional density of a first-order Markov copula model with copula $C_{\boldsymbol{U}}$. If the copula densities c_1 and c_{W^*} are strictly positive almost everywhere, the result of Longla and Peligrad (2012) implies the process is β-mixing and ergodic. The requirements for mixing and ergodicity in the case $k > 1$ are an open question.

Example 2. We construct a strictly stationary time series based on a gvt-d-vine(3) process with the follow specification. The underlying d-vine(3) copula process has a 180-degree-rotated Clayton copula ($\theta = 0.7$) at lag 1, a t copula ($\nu = 5$ and $\rho = 0.2$) at lag 2 and a Joe copula ($\theta = 1.5$) at lag 3. The v-transform $\mathcal{V}$ and marginal distribution F_X are as in Example 1; in particular, the marginal distribution is Student t. The d-vine(1) copula process for (W_t) uses the Gaussian copula ($\rho = 0.7$).

Figure 4 shows a realization of length $n = 2000$ from the resulting process (X_t) and Figure 5 shows that there is strong serial dependence in $(|X_t|)$ as well as serial dependence in (X_t) and between (X_t) and $(|X_t|)$.

4. Estimation

We now turn to the statistical estimation of the processes defined in the previous section. Let $\boldsymbol{x} = \{x_1, \ldots, x_n\}$ denote a realization from a strictly stationary process with parametric marginal distribution $F_X(x; \boldsymbol{\theta}_m)$ and joint copula density $c_{\boldsymbol{U}}(u_1, \ldots, u_n; \boldsymbol{\theta}_c)$. The full log-likelihood is

$$L(\boldsymbol{\theta}_c, \boldsymbol{\theta}_m; \boldsymbol{x}) = \log\left(c_{\boldsymbol{U}}(F_X(x_1; \boldsymbol{\theta}_m), \ldots, F_X(x_n; \boldsymbol{\theta}_m); \boldsymbol{\theta}_c)\right) + \sum_{i=1}^{n} \log\left(f_X(x_i; \boldsymbol{\theta}_m)\right). \tag{19}$$

In copula-based inference it is common to build up the model componentwise by estimating marginal distribution and copula model in successive steps. We follow the inference-functions-for-margins (IFM) approach (Joe, 1997) in which we first maximize $L_1(\boldsymbol{\theta}_m; \boldsymbol{x}) = \sum_{i=1}^{n} \log\left(f_X(x_i; \boldsymbol{\theta}_m)\right)$ to obtain estimates of the marginal parameters $\widehat{\boldsymbol{\theta}}_m$ and then maximize $L_2(\boldsymbol{\theta}_c; \boldsymbol{u}) = \log\left(c_{\boldsymbol{U}}(u_1 \ldots, u_n; \boldsymbol{\theta}_c)\right)$ using pseudo-copula data $\{u_i = F_X(x_i; \widehat{\boldsymbol{\theta}}_m), i = 1, \ldots, n\}$ to obtain $\widehat{\boldsymbol{\theta}}_c$. Moreover, in the second step we use a variant on the sequential procedure proposed by Aas et al. (2009) to select pair copulas one by one; we refer to this as *incremental* copula inference. The final stage of our method is a joint maximization of (19) over all model parameters (with the exception of the fulcrum δ, for reasons we later explain) using the estimates from IFM as starting values.

Statistical theory for copula inference for time series is a developing area. For first-order Markov processes, Chen and Fan (2006) showed consistency and asymptotic normality of a two-step procedure in which the margin is estimated non-parametrically and the copula is estimated parametrically. Nagler et al. (2020) have recently proposed multivariate time series models combining parametric margins and s-vine copula processes, which subsume our models in the case where the v-transform is degenerate; they formulate regularity conditions under which the multi-step sequential method of Aas et al. (2009) yields consistent and asymptotically normal parameter estimates.

In the following sections we elaborate first on the critical marginal modelling phase before discussing inference for the different elements of the copula density $c_{\boldsymbol{U}}$.

4.1. Marginal modelling

We are free to choose the best-fitting marginal distributions we can find for the data. In contrast, well-known econometric models are more constrained in the marginal behaviour they can model. In particular, many time series models in the GARCH family have the property that the resulting tails of the marginal distribution are regularly varying regardless of the choice of innovation distribution, i.e. they follow a power law (Mikosch and Stărică, 2000). However, in real asset return data we often encounter situations where tail behaviour differs in the two tails and one or both may be lighter than a regularly-varying law would dictate.

See, for instance, Figure 6. For three financial datasets exhibiting stochastic volatility, which will be analysed in Section 5, we show the well-known Hill estimator of the tail index of a regularly-varying law (see Hill (1975) and De Haan and Resnick (1998) for the basic properties of the Hill estimator). These plots should stabilize towards the left-hand end at a value greater than zero if a power tail is justified, but all these plots appear to continue to decay towards zero.

To model this behaviour, as well as the asymmetry of tails, we introduce a simple mixture of positive-valued distributions as a model for marginal distributions. The specific families that we consider are the generalized gamma distribution and the Burr distribution. In terms of extreme value theory (EVT) the first of these belongs to the Gumbel domain of attraction, and the second belongs to the Fréchet domain of attraction (Embrechts et al., 1997). In both cases, the associated Hill plots have similar shapes to the ones shown in our examples. A convenient feature of the generalized gamma distribution is the fact that for different choices of parameters, it can have tails which are both heavier or lighter than the exponential distribution, and it contains the Weibull distribution as a special case.

In general, we define a two-sided mixture corresponding by taking a density $f_0(\cdot; \boldsymbol{\eta})$ on the positive half-line and a parameter $p \in [0, 1]$ and setting

$$f(x | \boldsymbol{\eta}^+, \boldsymbol{\eta}^-) = \begin{cases} (1-p) f_0(-x | \boldsymbol{\eta}^-); & x < 0 \\ p f_0(x | \boldsymbol{\eta}^+); & x \geqslant 0. \end{cases} \tag{20}$$

A mixture of this form can also be interpreted as the result of splicing densities at the origin in a non-continuous manner. It is also possible to splice densities continuously but our results suggest that this results in slightly inferior fits. In the case of the generalized gamma distribution

$$f_0(x | \boldsymbol{\eta}) = \frac{\nu y^{\nu-1}}{(\mu/\sigma)^{\nu\sigma}\Gamma(\sigma)} y^{\nu(\sigma-1)} \exp\left(-(y\sigma/\mu)^{\nu}\right), \quad \boldsymbol{\eta} = (\sigma, \mu, \nu)' > \boldsymbol{0},$$

while for the Burr distribution

$$f_0(x | \boldsymbol{\eta}) = \frac{\alpha\beta(x/\sigma)^{\beta}}{x[1 + (x/\sigma)^{\beta}]^{\alpha+1}}, \quad \boldsymbol{\eta} = (\alpha, \beta, \sigma)' > \boldsymbol{0}.$$
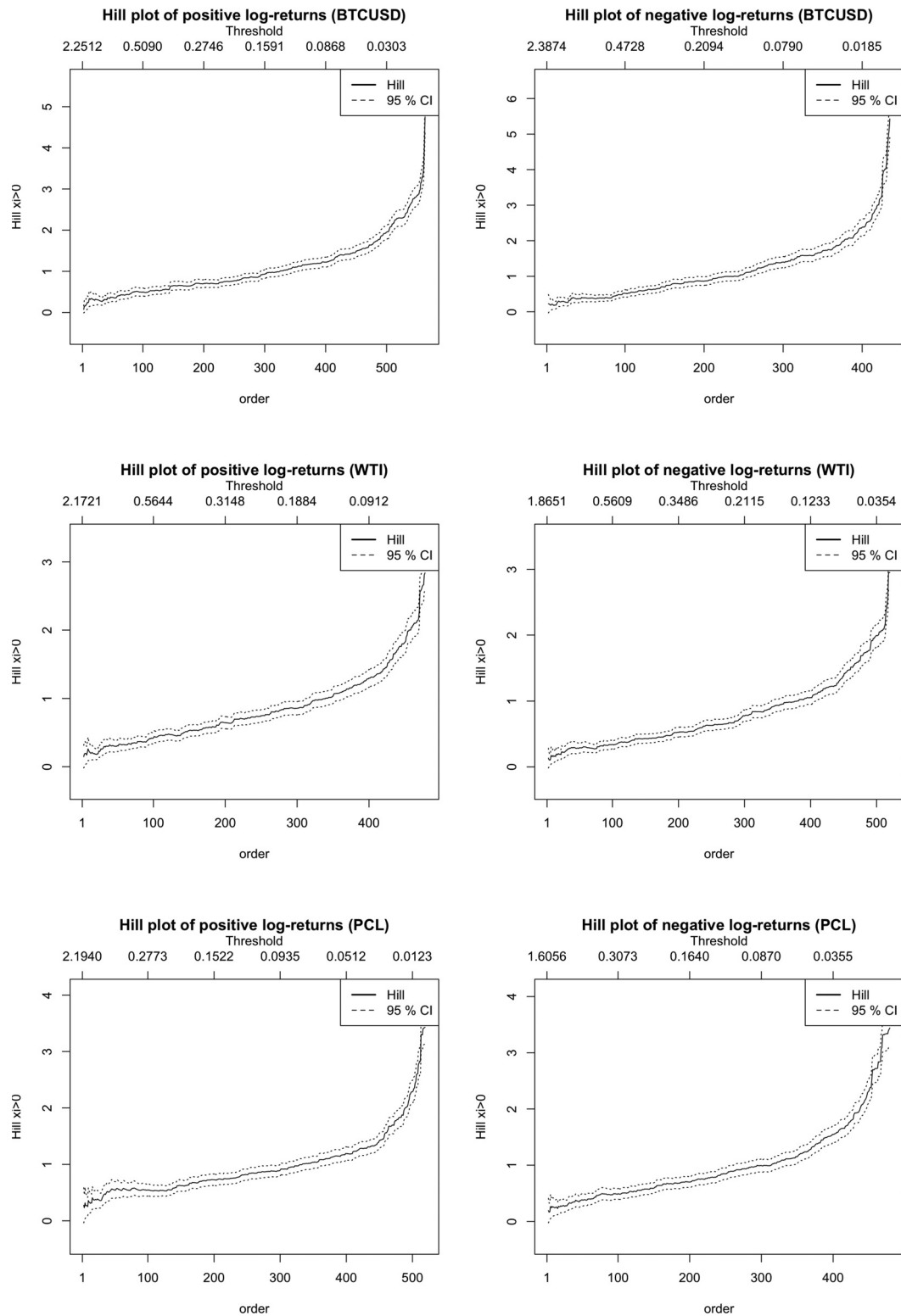
Fig. 6. Hill plots of the positive and negative log-returns. Top: Bitcoin-USD log-returns (2016-01-01/2019-11-01); middle: WTI crude oil log-returns (2015-01-01/2019-02-12); bottom: PCL log-returns (2006-01-01/2010-01-10).

Other choices for f_0 such as the log-gamma distribution can also give good results but the generalized gamma and Burr are good models for the data we analyse in this paper.

We have also compared these models with various two-sided distributions that are common in modelling log-returns such as the Student t, skewed Student t and normal inverse-Gaussian (NIG) distributions. The mixtures generally give statistically superior fits.

4.2. Incremental copula inference

In our model, the first term $L_2(\boldsymbol{\theta}_c; \boldsymbol{u}) = \log\left(c_{\boldsymbol{U}}(u_1 \ldots, u_n; \boldsymbol{\theta}_c)\right)$ in the log-likelihood (19) is maximized with respect to the copula parameters $\boldsymbol{\theta}_c$ using pseudo-copula data $\{u_i = F_X(x_i; \widehat{\boldsymbol{\theta}}_m), i = 1, \ldots, n\}$. The term $L_2(\boldsymbol{\theta}_c; \boldsymbol{u})$ further splits into a term coming from the copula c_{W^*} in (16) and a term coming from the d-vine copula model for (V_t) which may be written as $\sum_{i=1}^{k} L_i(\boldsymbol{\theta}_i, \boldsymbol{\theta}_v; \boldsymbol{u})$ where

$$L_i(\boldsymbol{\theta}_i, \boldsymbol{\theta}_v; \boldsymbol{u}) = \sum_{t=1}^{n-i} \log c_i\left(v_{t|S_{t,t+i}}, \ v_{t+i|S_{t,t+i}}; \boldsymbol{\theta}_i\right)\Big|_{v_1 = \mathcal{V}(u_1; \boldsymbol{\theta}_v), \ldots, v_n = \mathcal{V}(u_n; \boldsymbol{\theta}_v)}, \tag{21}$$

where $\boldsymbol{\theta}_i = (\theta_1, \ldots, \theta_i)$ and θ_i denotes the parameter of the ith pair copula. Note that the terms $c_i\left(v_{t|S_{t,t+i}}, \ v_{t+i|S_{t,t+i}}; \boldsymbol{\theta}_i\right)$ depend on θ_i through the copula c_i and on $\theta_1, \ldots, \theta_{i-1}$ through the conditional distributions $v_{t|S_{t,t+i}}$ and $v_{t+i|S_{t,t+i}}$. We assume to begin with that the parameter(s) $\boldsymbol{\theta}_v$ of the v-transform are fixed and postpone a discussion of their estimation to the next section.

We select pair copulas from a number of widely-used one-parameter families, these being Gauss, Gumbel, Clayton, Frank and Joe, as well as rotations of the Gumbel, Clayton and Joe copulas through 180 degrees. We have experimented with non-exchangeable generalizations of these copulas using the construction of Liebscher (2008). These break the time reversibility of the resulting process but only rarely deliver gains in fit to justify the extra complexity and additional parameters. We also omit the two-parameter t copula from consideration since it adds little in the examples we present and since the restriction to one-parameter copulas permits the use of fast sequential method-of-moments estimation in our forecasting study.

In the incremental method we choose the copula families for $c_1, \ldots, c_k$ sequentially but we continually re-estimate all parameters of previously chosen copulas. We also use the incremental method to estimate the order k of the process. The algorithm consists of an initialization step and a continuation step:

Step 1: Select the copula family c_1 and parameter value θ_1 that maximize $L_1(\boldsymbol{\theta}_1, \boldsymbol{\theta}_v; \boldsymbol{u})$.

Step j: With copula families $c_1, \ldots, c_{j-1}$ already determined, select the copula family c_j and parameter values $\boldsymbol{\theta}_j$ that maximize $\sum_{i=1}^{j} L_i(\boldsymbol{\theta}_i, \boldsymbol{\theta}_v; \boldsymbol{u})$. If the AIC of the resulting model of order j is lower than the previous model of order $j-1$, continue; otherwise stop and set $k = j - 1$.

This differs from the sequential method used by Aas et al. (2009) and analysed in Nagler et al. (2020) since we do not hold the parameters of previously selected copulas fixed at their previously estimated values. This should reduce the risk of parameter error percolating through the procedure and lead to estimates that are even closer to full ML estimates than those obtained using the sequential method.

4.3. Estimating the v-transform

A natural approach to estimating the parameters $\boldsymbol{\theta}_v$ of the v-transform is to optimize over these in an outer loop, while applying the incremental procedure of the previous section in an inner loop. This is feasible, albeit very computationally intensive, since the incremental procedure is itself a greedy algorithm.

Care must be taken in the optimization with respect to the fulcrum parameter δ. If $\delta = u_i = F_X(x_i; \widehat{\boldsymbol{\theta}}_m)$ for some original data point x_i, then $\mathcal{V}(u_i) = 0$ and the log-likelihood for the copula takes the value $-\infty$. This implies that the profile likelihood of δ is not differentiable at such points and has multiple local maxima. We thus use a grid search for an optimal estimate of δ, avoiding the u_i values, rather than continuous optimization in the interval [0,1]. The key observation is that δ is a *threshold* parameter and pseudo copula observations u_i on either side lead to different responses. The situation is akin to that which arises for TAR and other threshold models in econometrics; see, for example, Tong (1983) and Hansen (2000). In these models it is typical to use conditional (or profile likelihood) inference for other parameters while the threshold parameter δ takes fixed values on a grid.

In our empirical examples we use a procedure that is simpler on two counts, and therefore faster. First, we restrict attention to the linear v-transform. This is sufficient to capture the main features of stochastic volatility and outperform competitor models in the examples we present; moreover, it facilitates value-at-risk (VaR) forecasting and backtesting with the fitted models, as we show in Section 6. Second, we optimize over the fulcrum parameter δ in the d-vine(1) model of Step 1 above, rather than the full d-vine(k) model.

4.4. Estimating the copula C_W

In a gvt-d-vine copula process the log-likelihood contains an extra additive term which can be fully expressed in terms of $C_W(\delta(u_{t-1}), \delta(u_t); \boldsymbol{\theta}_W)$ for the bivariate copula C_W in (17). A final step is added to the stepwise method in which this additional term is maximized with respect to $\boldsymbol{\theta}_W$. In principle the term could be maximized jointly with other terms, since it also depends on the parameters of the v-transform, but we chose the simpler stepwise approach to preserve speed.

In practice, the extent to which the parameter(s) $\boldsymbol{\theta}_W$ of a copula C_W can be identified from data depends on the extent to which the set of points $S = \{(\delta(u_{t-1}), \delta(u_t)), \, t = 2, \ldots, d\}$ fills the unit square $[0, 1]^2$, which depends in turn upon the choice of v-transform. In the special case of the linear v-transform, S is identical to the singleton $\{(\delta, \delta)\}$ meaning that we can only identify the value of C_W at this point (see the left panel of Figure 3).

For the case of a non-linear v-transform we recommend picking a copula family C_W that can model positive dependence, negative dependence and independence; an ideal candidate is the radially symmetric Frank copula which involves only simple functions and is quick to evaluate. For the linear transform we can actually circumvent the evaluation of C_W entirely. The likelihood can be reparameterized in terms of $\lambda = C_W(\delta, \delta) = \mathbb{P}(W_{t-1} \leqslant \delta, W_t \leqslant \delta) = \mathbb{P}(U_{t-1} \leqslant \delta, U_t \leqslant \delta)$, or in terms of a correlation parameter

$$\rho_U = \rho(I_{\{U_{t-1} \leqslant \delta\}}, I_{\{U_t \leqslant \delta\}}) = \rho(I_{\{W_{t-1} \leqslant \delta\}}, I_{\{W_t \leqslant \delta\}}) = \frac{\lambda - \delta^2}{\delta - \delta^2}. \tag{22}$$

4.5. A graphical method using generalized lagging

We use a graphical method to gain insight into the fitted pair copula model. Having fitted the model, we first construct pseudo-copula data $\{v_i = \mathcal{V}(F_X(x_i; \widehat{\boldsymbol{\theta}}_m); \widehat{\boldsymbol{\theta}}_V), i = 1, \ldots, n\}$ from c_V. We then use a sequential method to reconstruct, for $i = 1, \ldots, k$, the datasets

$$\mathcal{D}_i = \left\{ (v_{1,t}^{(i)}, v_{2,t}^{(i)}), \; t = 1, \ldots, n - i \right\}, \quad v_{1,t}^{(i)} = v_{t|S_{t,t+i}}, \quad v_{2,t}^{(i)} = v_{t+i|S_{t,t+i}}, \tag{23}$$

which are modelled by copula c_i in the likelihood contribution (21). The points of the lag 1 dataset $\mathcal{D}_1$ are given by $\left(v_{1,t}^{(1)}, v_{2,t}^{(1)} \right) = (v_t, v_{t+1})$, $t = 1, \ldots, n - 1$, and these are the points that would be used in a first-order lagplot. For $i < k$ the recursive formulas (15) then allow the recursive construction of dataset $\mathcal{D}_{i+1}$ from $\mathcal{D}_i$ by setting

$$v_{1,t}^{(i+1)} = h_1^{(i)}(v_{1,t}^{(i)}, v_{2,t}^{(i)}), \; v_{2,t}^{(i+1)} = h_2^{(i)}(v_{1,t+1}^{(i)}, v_{2,t+1}^{(i)}), \quad t = 1, \ldots, n - i - 1.$$

The h-functions of the previous copula C_i are always required to construct the next dataset $\mathcal{D}_{i+1}$. We refer to the iterative procedure as *generalized lagging* since a scatterplot of the points in the dataset $\mathcal{D}_i$ may be thought of as a kind of lagplot at generalized lag i. In the graphical method we plot sample Kendall's tau values for each dataset $\mathcal{D}_i$ against the theoretical values for the fitted copulas c_i to see the degree of correspondence between sample and fitted measures.

The generalized lagging procedure also suggests a fast alternative to the sequential ML estimation method of Aas et al. (2009). We could use the dataset $\mathcal{D}_i$ at lag i to estimate θ_i by a method-of-moments procedure that exploits the one-to-one relationsip between θ_i and Kendall's tau for the copula families of interest. This is a much faster method than our main incremental method and we use it in a rolling estimation and backtesting study at the end of the paper.

5. Empirical Results

5.1. Data and models

In this section we consider three different financial datasets and their estimation using the models and methods introduced in the previous sections. The first dataset consists of the log-returns of the Bitcoin-USD exchange rate from January 1, 2016 to November 1, 2019. The second dataset comprises the log-returns of the WTI crude-oil price from January 1, 2015 to February 12, 2019. Finally, the third series consists of the log returns of the price of the PCL stock from January 1, 2006 to January 10, 2010. Each series consists of 1000 observations and in the presented results the values of log-returns have been multiplied by 10 for improved stability in fitting some of the alternative GARCH-type specifications.

The first two datasets do not exhibit strong serial correlation (as is usual for log-returns), but they do show stochastic volatility. In contrast, both serial correlation and stochastic volatility are present in the third dataset. The third dataset was identified by taking all S&P500 stocks for the 2006–10 time period and selecting those with the largest absolute values of first-order serial correlation in log-returns. Thus it can be considered as a more extreme example of the level of serial correlation that is present in raw log-returns and a good candidate series for exploring the added value of a gvt-d-vine model over a standard vt-d-vine-model.

In addition to the vt-dvine model introduced in this paper we also consider a vt-ARMA model of the type presented in McNeil (2021). To examine the extent to which higher-order Markov models are necessary, we also compare with a vt-d-vine model of order 2. From the wider GARCH family inspired by the ideas of Engle (1982) we fit the standard GARCH model of Bollerslev (1986), the exponential GARCH model of Nelson (1991) and the GJR-GARCH model of Glosten et al. (1993). We also vary the choice of innovation distribution for these models. The optimal vt-ARMA and GARCH model orders are selected by minimization of the AIC criterion.

Table 1
Summary of parameter estimates for vt-d-vine copula models.

Parameter	BTCUSD	WTI	PCL
θ_1	1.4212 (frank)	1.1714 (joe)	1.3485 (gumbel)
θ_2	1.2649 (frank)	0.1272 (gaussian)	0.3424 (clayton)
θ_3	0.8665 (frank)	1.0341 (gumbel)	0.2713 (clayton)
θ_4	0.1591 (clayton)	1.0381 (joe)	0.2010 (clayton)
θ_5	0.9710 (frank)	0.0717 (gaussian)	1.1779 (joe)
θ_6	0.1050 (clayton)	1.0526 (joe)	0.1567 (clayton)
θ_7	0.0208 (clayton)	0.0722 (gaussian)	0.1233 (clayton)
θ_8	0.0253 (clayton)		0.1058 (gaussian)
θ_9	0.3023 (frank)		1.0497 (gumbel)
θ_{10}	0.5885 (frank)		0.5165 (frank)
θ_{11}			0.0783 (clayton)
θ_{12}			0.1179 (clayton)
θ_{13}			0.0626 (clayton)
δ	0.45	0.3	0.55
η_1^+	3.6121	8.9880	1.1573
η_2^+	0.0426	0.0049	1.5252
η_3^+	0.4479	0.3518	0.1078
η_1^-	1.6237	1.7246	2.4315
η_2^-	0.1859	0.3123	1.1650
η_3^-	0.6446	0.9053	0.2851
p_0	0.5534	0.4795	0.5309

Table 2
Summary of all model fits for all datasets.

Dataset	Model	LogLik	NumPars	AIC	rank	extra AIC
BTCUSD	vt-ARMA(1,1)	-373.54	9	765.08	2	8.91
	vt-d-vine(2)	-407.77	9	833.54	6	77.37
	vt-d-vine(10)	-361.09	17	756.18	1	0.00
	sGARCH(1,1)	-401.08	6	814.17	5	57.99
	eGARCH(1,1)	-384.61	7	783.21	3	27.04
	gjrGARCH(1,1)	-399.06	7	812.13	4	55.95
WTI	vt-ARMA(1,1)	-591.34	9	1200.69	3	12.41
	vt-d-vine(2)	-588.92	9	1195.83	2	7.56
	vt-d-vine(7)	-580.14	14	1188.28	1	0.00
	sGARCH(1,1)	-601.43	6	1214.86	5	26.58
	eGARCH(6,6)	-579.28	22	1202.56	4	14.28
	gjrGARCH(1,1)	-601.40	7	1216.79	6	28.52
PCL	vt-ARMA(1,1)	159.64	10	-299.29	5	44.67
	vt-d-vine(2)	82.51	9	-147.02	6	196.93
	vt-d-vine(13)	191.98	20	-343.95	1	0.00
	sGARCH(1,1)	172.80	6	-333.61	2	10.34
	eGARCH(1,1)	173.77	7	-333.53	3	10.42
	gjrGARCH(1,1)	173.36	7	-332.73	4	11.23

5.2. Parameter estimates and model comparison

Table 1 shows the estimates for the vt-d-vine models fitted to the datasets. The parameters of the pair copulas are simply denoted θ_i and the names of the selected copulas are given. The marginal model used for both the vt-processes is a mixture of positive and negative distributions as in (20) amounting to 7 marginal parameters when using the generalized gamma (for the first two datasets) and Burr (for the third dataset) distributions. Marginal parameters are denoted η_j^- and η_j^+ for the negative and positive tails respectively. The fulcrum estimates are $\delta = 0.45, 0.3, 0.55$ for the three datasets.

Table 2 summarizes measures of fit for the best fitting model within each of the five considered classes with each table relating to a different dataset. We see that vt-d-vine models (and sometimes also vt-ARMA models) fare favourably against the alternatives from the GARCH family. It is evident that the vt-d-vine model of order 2 is not at all competitive for the BTCUSD and PCL data, but it does much better for the WTI dataset, where it gives the second lowest AIC model after the vt-d-vine model of order 7.

For the PCL data, a gvt-d-vine model incorporating a Frank copula for the (W_t) process gave a significant improvement over the best vt-d-vine model according to a likelihood ratio test. In terms of log-likelihood, it was also superior to any combined ARMA-GARCH model using the GARCH specifications of Table 2 although the AIC value of the best ARMA-GARCH model does 'catch up' with the AIC value of the best gvt-d-vine model in this case.

Note that the Markov nature of the d-vine models means that more lags need to be considered in this approach to obtain similar behaviour to models in the GARCH family which effectively have a 'moving-average term' for modelling volatility.

5.3. Graphical analysis of fit

We assess the fit of the models graphically in Figures 7, 8 and 9. In addition to a time series plot of each dataset and a QQ-plot against the fitted marginal distribution, we show four other plots. The first of these is a plot of a *partial rank autocorrelation function* in which the empirical values of Kendall's tau calculated from the generalized lagged datasets $\mathcal{D}_i$ in (23) are plotted as bars and the implied Kendall tau values for the corresponding fitted pair-copulas are plotted as points. These plots show good correspondence between data and model values.

The fourth plot for each series shows an estimate of the profile function g_T of the volatility proxy transformation T as defined in Section 2.1. The fifth plot shows the volatility proxy function $T(x) = \Phi^{-1}(\mathcal{V}(F_X(x)))$. The points show an empirical estimate of this relationship while the curve is the parametric estimate implied by the fitted model. The final plot is an illustration of an interval-valued risk measure that we call ViVaR which is described in Section 6.1.

Overall, these sets of plots suggest satisfactory fits. For the PCL dataset, the curvature of the marginal QQ-plot does suggest a slight lack of fit in the tails. This seems to be introduced in the final joint optimization over all parameters and suggests a lack of fit in the copula component, possibly caused by some remaining misspecification of the tail dependencies in the copula model. The estimated marginal distribution obtained in the first step of the IFM model yields a better looking QQ-plot (omitted).

The estimated profile function g_T for the Bitcoin data is that of a symmetric volatility proxy function. Moreover the smallest value of the volatility proxy corresponds with a zero return. For the WTI and PCL data there is more asymmetry. The estimated values of μ_T are not equal to zero and both profile functions satisfy $g_T(x) > x$ for large x. This can be interpreted as large negative log-returns contributing more to the volatility proxy variable.

6. Using the Model for Forecasting

In this section we consider the out-of-sample performance of the best models from the previous section, looking in particular at predictions of value-at-risk (VaR).

6.1. Prediction of value-at-risk measures

In our most general model, conditional quantiles of the forecast distribution $F_{U_t|U_{t-1},\dots,U_{t-k}}$ can be calculated via numerical integration of the conditional density in equation (18). In the model with linear v-transform considerable simplification is possible. In this case, using the shorthand notation $\boldsymbol{U_{t-1}} = (U_{t-1},\dots,U_{t-k})$, the conditional distribution function satisfies, for $u_t \leqslant \delta$,

$$
\begin{aligned}
F_{U_t|\boldsymbol{U_{t-1}}}(u_t|\boldsymbol{u_{t-1}}) &= \mathbb{P}(U_t \leqslant u_t \mid \boldsymbol{U_{t-1}} = \boldsymbol{u_{t-1}}) \\
&= \mathbb{P}(V_t \geqslant \mathcal{V}(u_t) \mid \boldsymbol{U_{t-1}} = \boldsymbol{u_{t-1}})\mathbb{P}(W_t \leqslant \delta \mid U_{t-1} = u_{t-1}),
\end{aligned}
\tag{24}
$$

where the second term satisfies

$$
\mathbb{P}(W_t \leqslant \delta \mid U_{t-1} = u_{t-1}) = d(u_{t-1};\delta,\rho_U) = \begin{cases} \delta + \rho_U(1-\delta) & \text{if } u_{t-1} \leqslant \delta \\ \delta(1-\rho_U) & \text{if } u_{t-1} > \delta \end{cases}
$$

where ρ_U is the correlation parameter describing serial correlation in the (U_t) process in (22). In the case where $\rho_U = 0$ we simply have $d(u_{t-1};\delta,\rho_U) = \delta$.

To find the α-VaR at time t for $\alpha \gg 0.5$ we have to solve the equation

$$
\mathbb{P}(V_t \geqslant v_t \mid \boldsymbol{U_{t-1}} = \boldsymbol{u_{t-1}}) = \frac{1-\alpha}{d(u_{t-1};\delta,\rho_U)},
$$

for v_t and then compute $\text{VaR}_{t,\alpha} = F_X^{-1}(\mathcal{V}^{-1}(v_t))$. The calculation is facilitated by the fact that V_t depends on the past values of $\boldsymbol{U_{t-1}}$ only through the past values of $\boldsymbol{V_{t-1}} = \mathcal{V}(\boldsymbol{U_{t-1}})$ and hence

$$
\mathbb{P}(V_t \geqslant v_t \mid \boldsymbol{U_{t-1}} = \boldsymbol{u_{t-1}}) = \mathbb{P}(V_t \geqslant v_t \mid \boldsymbol{V_{t-1}} = \boldsymbol{v_{t-1}}) = 1 - F_{V_t|\boldsymbol{V_{t-1}}}(v_t \mid \boldsymbol{v_{t-1}}) = 1 - v_{t|S_{t-k-1,t}},
$$

where the final expression uses the notation of Section 3 and we recall that the conditional probabilities $v_{t|S_{t-k-1,t}}$ can be calculated recursively.

While full calculation of VaR is relatively straightforward in models with a linear v-transform, the above calculations suggest that in general models an alternative measure of risk could be considered by concentrating on quantiles of the
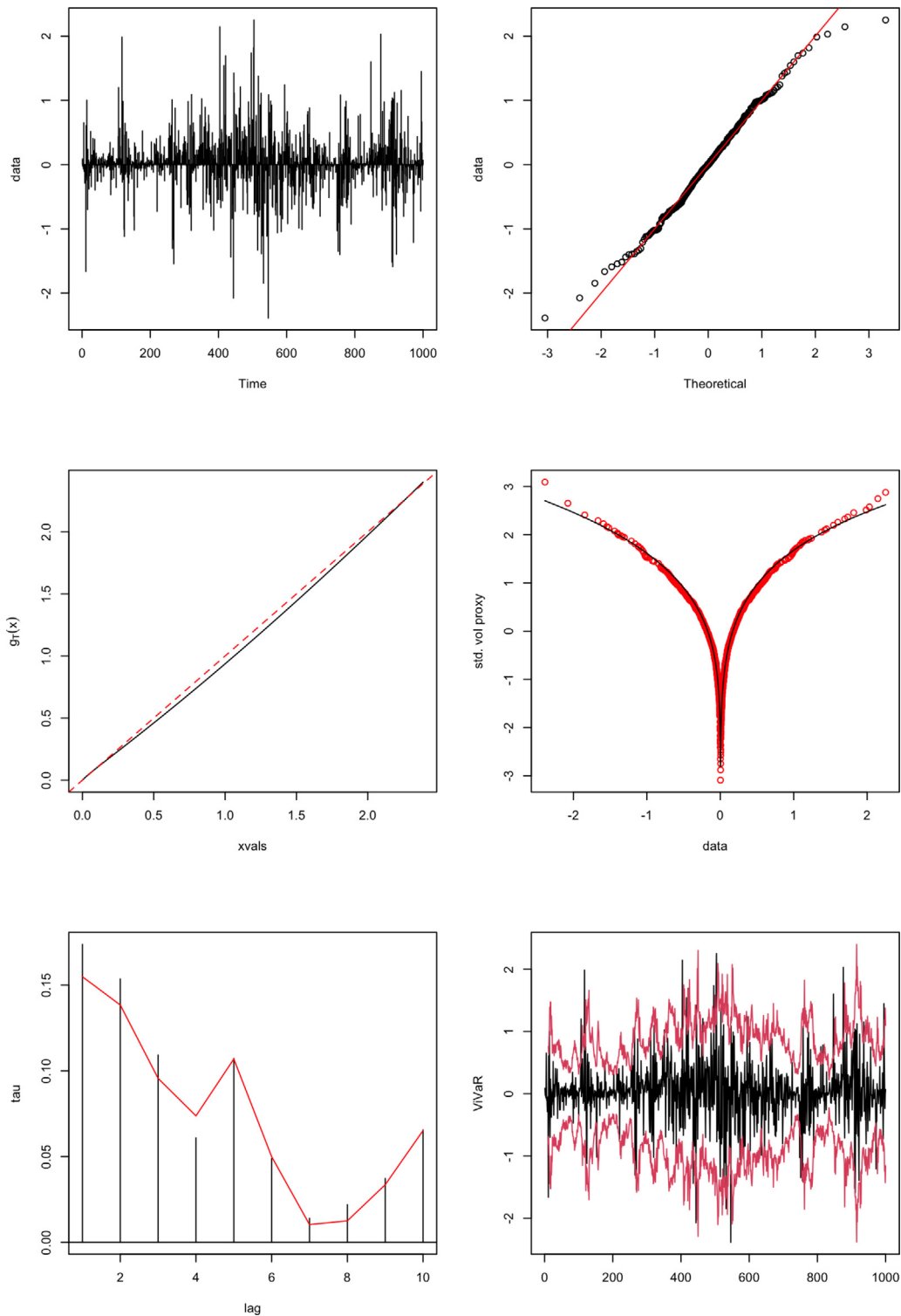
Fig. 7. Bitcoin data. Top left: log-returns; top right: QQ-plot of the fitted versus empirical marginal quantiles; middle: implied profile function g_T and standardized volatility proxy; bottom left: implied Kendall's tau values from the fitted model (line) and from sequential method of moments estimation (bars); bottom right: plot of 95% ViVaR interval.

Fig. 8. WTI data. Top left: log-returns; top right: QQ-plot of the fitted versus empirical marginal quantiles; middle: implied profile function g_T and standardized volatility proxy; bottom left: implied Kendall's tau values from the fitted model (line) and from sequential method of moments estimation (bars); bottom right: plot of 95% ViVaR interval.
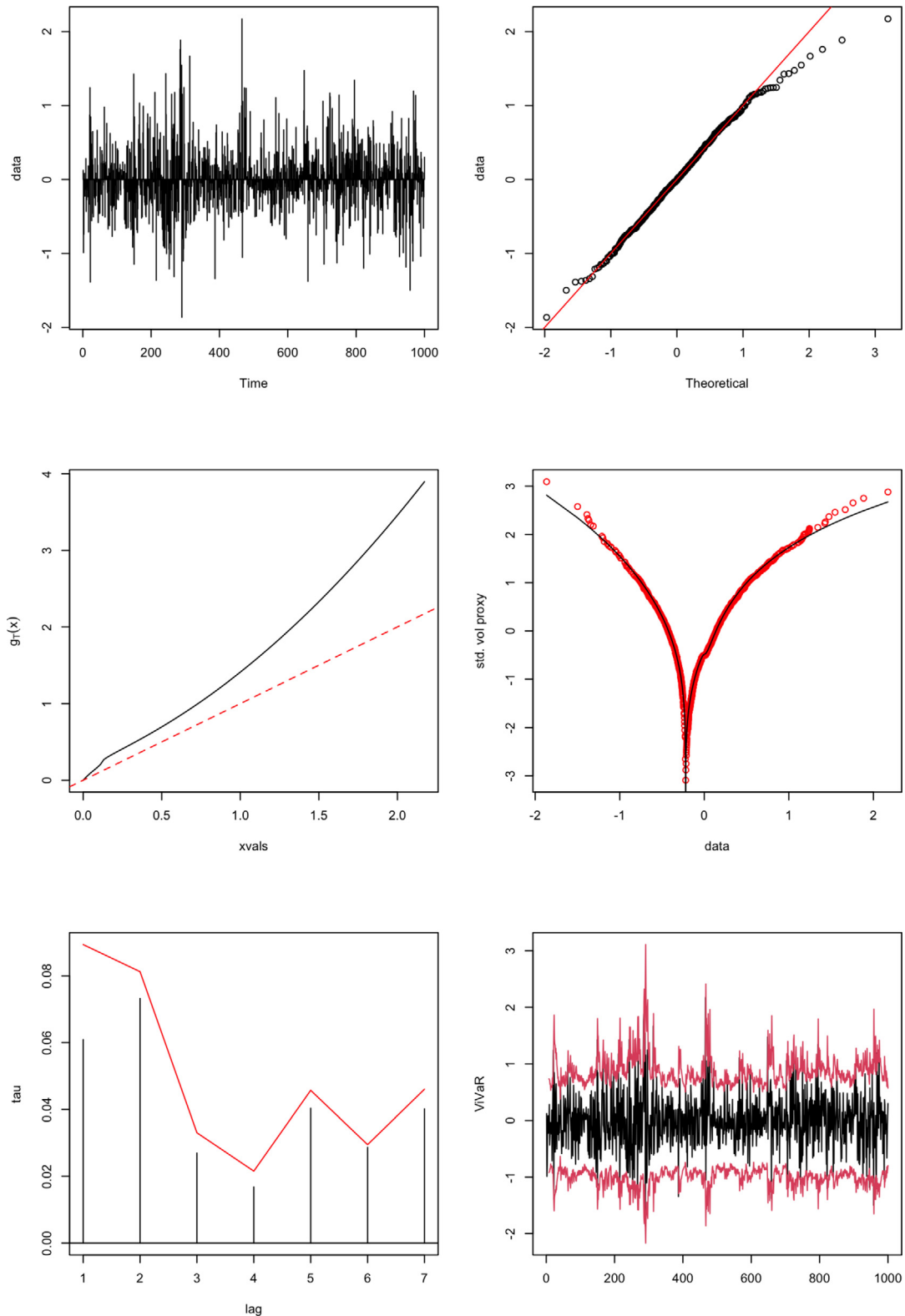
Fig. 9. PCL data. Top left: log-returns; top right: QQ-plot of the fitted versus empirical marginal quantiles; middle: implied profile function g_T and standardized volatility proxy; bottom left: implied Kendall's tau values from the fitted model (line) and from sequential method of moments estimation (bars); bottom right: plot of 95% ViVaR interval.
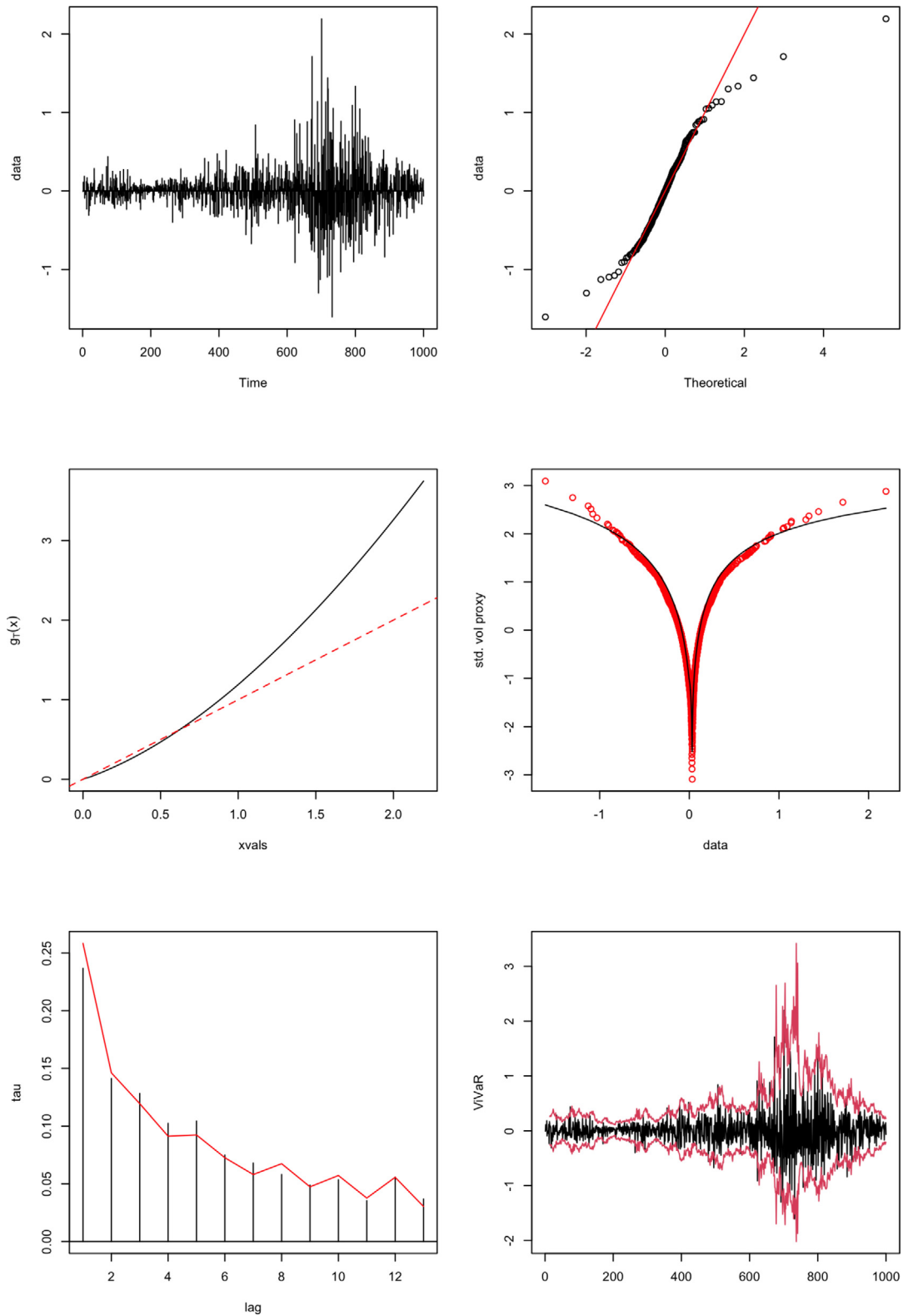
Table 3

p-values for tests for uniformity and independence of PIT values applied to u-PIT and v-PIT values; p_{KS}, p_R and p_{LB10} refer to Kolmogorov-Smirnov test, runs test and Ljung-Box test using 10 lags.

Dataset	n	Tests on u-PITs			Tests on v-PITs		
		p_{KS}	p_R	p_{LB10}	p_{KS}	p_R	p_{LB10}
BTCUSD	803	0.53	0.31	0.03	0.62	0.92	0.14
WTI	720	0.23	0.77	0.50	0.13	0.99	0.61
PCL	744	0.51	0.00	0.01	0.20	0.88	0.82

conditional distribution of V_t in the first term of (24). For $\alpha \gg 0.5$ let $v_{\alpha,t}$ be given by $1 - \alpha = \mathbb{P}(V_t \geqslant v_{\alpha,t} \mid \boldsymbol{V}_{t-1} = \boldsymbol{v}_{t-1})$. The event $\{V_t \geqslant v_{\alpha,t}\}$ can be written as

$$\{V_t \geqslant v_{\alpha,t}\} = \left\{U_t \leqslant \mathcal{V}^{-1}(v_{\alpha,t})\right\} \cup \left\{U_t \geqslant \mathcal{V}^{-1}(v_{\alpha,t}) + v_{\alpha,t}\right\}$$
$$= \left\{X_t \in \underbrace{\left(F_X^{-1}\left(\mathcal{V}^{-1}(v_{\alpha,t})\right), F_X^{-1}\left(\mathcal{V}^{-1}(v_{\alpha,t}) + v_{\alpha,t}\right)\right)}_{I_{\alpha,t}}\right\}^{\complement},$$

and thus the interval labelled $I_{\alpha,t}$ plays the role of a set-valued risk measure with the interpretation that its probability is exactly equal to α and all points outside it lead to values of the volatility proxy that exceed the α-quantile of volatility. We refer to $I_{\alpha,t}$ as a volatility-implied VaR interval (ViVaR) and have calculated it for the examples of Section 5.

6.2. Out-of-sample forecasting experiments

We evaluate the predictive performance of the vt-d-vine models using a rolling backtesting procedure typical for financial risk applications; we construct models using a moving window of $n = 500$ data, calculate the one-step-ahead predictive distribution and then validate the model using probability-integral-transform (PIT) values and value-at-risk (VaR) exceptions. Data are taken for a 5-year period; these are the years 2015–19 for BTCUSD and WTI and the years 2006–10 for PCL.

To facilitate backtesting of vt-d-vine models in reasonable time, we use an accelerated fitting procedure based on the sequential estimation idea described at the end of Section 4.5. We follow the IFM approach and first estimate the marginal distribution F_X. We then fix k so that we do not optimize over model order; $k = 5$ is sufficient to give good results for BTCUSD and WTI, while $k = 10$ is required for the PCL dataset. We estimate parameters using the sequential method-of-moments procedure based on Kendall's tau. For PCL we restrict attention to vt-d-vine models without serial dependence in (W_t).

The PIT method follows the general model validation approach of Diebold et al. (1998). Suppose at time t we have data $x_{t-n}, \ldots, x_{t-1}$. We estimate the marginal model F_X and use this to obtain transformed data $u_{t-n}, \ldots, u_{t-1}$ on the u-scale, which we use to estimate the copula model. The conditional cdf is given by (24) and this can be used to calculate a PIT value $\widehat{u}_t = F_{U_t \mid \boldsymbol{U}_{t-1}}(u_t \mid \boldsymbol{u}_{t-1})$ where $u_t = F_X(x_t)$ is calculated from a new out-of sample observation x_t. We refer to the sequence of values $(\widehat{u}_t)$ as *u-PIT values*; if the forecast models are reasonable they should be iid uniform. We test uniformity with a Kolmogorov-Smirnov test and independence using the Wald-Wolfowitz runs test and the Ljung-Box test based on the first 10 autocorrelation estimates; p-values are denoted p_{KS}, p_R and p_{L10B} respectively and given in Table 3. There is a question mark over the independence of the u-PIT values for PCL but not their uniformity; this is explained by the dropping of the

Table 4

Results of two-sided binomial score test: n is number of trials; n_e number of exceptions and p_B is the test p-value. VaR is estimated at level $\alpha = 0.95$.

Dataset	Model	n	n_e	n_e/n	p_B
BTCUSD	vt-d-vine(5)	803	48	0.060	0.20
	sGARCH(1,1)	803	59	0.073	0.00
	eGARCH(1,1)	803	46	0.057	0.34
	gjr-GARCH(1,1)	803	62	0.077	0.00
WTI	vt-d-vine(5)	720	47	0.065	0.06
	sGARCH(1,1)	720	41	0.057	0.39
	eGARCH(1,1)	720	47	0.065	0.06
	gjr-GARCH(1,1)	720	44	0.061	0.17
PCL	vt-d-vine(10)	744	44	0.059	0.25
	sGARCH(1,1)	744	35	0.047	0.71
	eGARCH(1,1)	744	32	0.043	0.38
	gjr-GARCH(1,1)	744	33	0.044	0.48

model for (W_t) for computational speed. The Ljung-Box test also gives a significant result for the Bitcoin data although this was not confirmed by the runs test. Otherwise the p-values are non-significant.

The u-PIT values reveal very little about the quality of stochastic volatility modelling which is key to making reasonable predictions of VaR; to validate the modelling of volatility we consider *v-PIT values* given by $\widehat{v}_t = F_{V_t|\boldsymbol{V}_{t-1}}(v_t \mid \boldsymbol{v}_{t-1})$ where the values $v_t = \mathcal{V}_\delta(u_t)$ are obtained using the estimated linear v-transform and where $F_{V_t|\boldsymbol{V}_{t-1}}$ is the predictive distribution of the underlying d-vine model as before. The v-PIT values should also be iid uniform and we run the same tests as for u-PIT values and show results in Table 3. The tests of uniformity and independence of v-PIT values are not rejected for any of the time series.

The VaR exception tests are based on comparing a sequence of α-VaR estimates $(\text{VaR}_{t,\alpha})$ at times t with the corresponding realized values (x_t) and counting exceptions $(x_t < -\text{VaR}_{t,\alpha})$. In Table 4 we give the results of a two-sided binomial score test applied to the VaR exception data for $\alpha = 0.95$ and compare with the other models. The only significant test results are for the standard GARCH and GJR-GARCH models applied to the Bitcoin data; in all other cases the exception counts are in line with the expected number under the hypothesis that the forecast distributions are correct.

7. Conclusion

The results in this paper and further unreported analyses suggest that many volatile time series of log-returns on financial assets can be successfully modelled using d-vine copula processes in conjunction with v-transforms and asymmetric mixed marginal distributions. In many cases the in-sample fits and out-of-sample predictions are superior to those obtained from all the most widely used members of the extended GARCH family. It is noteworthy that the best fitting models are higher-order Markov models with relatively short memory. In the examples of this paper the orders are 7, 10 and 13, corresponding to no more than 3 weeks of trading days.

The class of models in this paper could be extended in a number of directions. An obvious topic of interest is multivariate time series models for multiple volatile return series. The m-vine model of Beare and Seo (2015) and the recent work on s-vine models by Nagler et al. (2020) suggest directions for generalizing the vt-d-vine model to the bivariate and multivariate case.

Remaining in the univariate case, it would be interesting to find further tractable models for the bivariate process (V_t, W_t) that is used to construct the process (U_t) under the stochastic inversion operation of Theorem 2. We have restricted attention to processes (V_t) and (W_t) that are fully independent of each other, although the requirement of the result is simply that the variables V_t and W_t are contemporaneously independent for all t. Cross dependencies between the processes at different lags could be useful for modelling feedback effects between the signs of log-returns at time t and volatility at future times $t + k$, or vice versa, although statistical estimation of the resulting models is likely to be challenging.

Despite the current popularity of vine copula models, there are two areas where more theoretical work is needed. The first is the study of the serial dependence or mixing properties of vine-copula-based time series. The papers of Beare (2010) and Longla and Peligrad (2012) make important contributions in the first-order Markov case but mixing and ergodicity results for more general processes would be valuable, particularly if the conditions can be checked for concrete combinations of pair copulas. The second area is statistical inference for vine-copula-based time series, where the recent paper of Nagler et al. (2020) is one of the first to address the challenge of underpinning widely-used and intuitive stepwise procedures with regularity conditions to guarantee the usual desirable properties of estimators.

Declaration of Competing Interest

Martin Bladt and Alexander J. McNeil both declare that they have no conflict of interest regarding the submission of the manuscript "Time series copula models using d-vines and v-transforms" to Econometrics and Statistics.

Acknowledgment

Appendix A. Proofs

A1. Proof of Proposition 1

We first observe that, conditional on $V = v$, we have

$$\left\{\tilde{U} = U\right\} = \left\{\mathcal{V}^{-1}(v, W) = \mathcal{V}^{-1}(v), U = \mathcal{V}^{-1}(v)\right\} \cup \left\{\mathcal{V}^{-1}(v, W) = v + \mathcal{V}^{-1}(v), U = v + \mathcal{V}^{-1}(v)\right\}$$

$$= \left\{W \leqslant \Delta(v), U = \mathcal{V}^{-1}(v)\right\} \cup \left\{W > \Delta(v), U = v + \mathcal{V}^{-1}(v)\right\}$$

Hence, by the independence of W and U, it follows that

$$\mathbb{P}\left(\tilde{U} = U \mid V = v\right) = \Delta(v)^2 + (1 - \Delta(v))^2$$

and integrating over v we obtain

$$
\begin{aligned}
\mathbb{P}\big(\tilde{U} = U\big) &= 2\mathbb{E}\big(\Delta(V)^2\big) + 1 - 2\mathbb{E}(\Delta(V)) \\
&= 2\big(\mathrm{var}\,(\Delta(V)) + \delta^2\big) + 1 - 2\delta \\
&= 2\,\mathrm{var}\,(\Delta(V)) + \delta^2 + (1 - \delta)^2\,.
\end{aligned}
$$

A2. Proof of Theorem 2

We will need the fact that

$$
p_{\delta, u}(\delta(u)) = \frac{(-1)^{I_{\{u \leqslant \delta\}}}}{\mathcal{V}'(u)} \quad \text{for all } u \in [0, 1] \tag{A.1}
$$

which we first prove. When $u \leqslant \delta$ we clearly have that $p_{\delta,u}(\delta(u)) = \delta(u) = \Delta(\mathcal{V}(u)) = -1/\mathcal{V}'(u)$ where we have used equation (6). When $u > \delta$ we have that $p_{\delta,u}(\delta(u)) = 1 - \Delta(\mathcal{V}(u)) = 1 + 1/\mathcal{V}'(u^*)$ where u^* is the dual point of u_i as defined in Definition 1. We have to show that $1 + 1/\mathcal{V}'(u^*) = 1/\mathcal{V}'(u)$. This follows from the fact that for any $u > \delta$ with dual point $u^* = u - \mathcal{V}(u)$ we have $\mathcal{V}(u) = \mathcal{V}(u^*) = \mathcal{V}(u - \mathcal{V}(u))$ so that $\mathcal{V}'(u) = \mathcal{V}'(u^*)(1 - \mathcal{V}'(u))$ and hence $\mathcal{V}'(u) = \mathcal{V}'(u^*)/(1 + \mathcal{V}'(u^*))$.

Now fix the point $(u_1, \ldots, u_d) \in [0, 1]^d$ and consider the set of events $A_i(u_i)$ defined by

$$
A_i(u_i) = \begin{cases} \{U_i \leqslant u_i\} & \text{if } u_i \leqslant \delta \\ \{U_i > u_i\} & \text{if } u_i > \delta \end{cases}
$$

The probability $\mathbb{P}(A_1(u_1), \ldots, A_d(u_d))$ is the probability of the orthant defined by the point $(u_1, \ldots, u_d)$ and the copula density at this point is given by

$$
c_{\boldsymbol{U}}(u_1, \ldots, u_d) = (-1)^{\sum_{i=1}^{d} I_{\{u_i > \delta\}}} \frac{\mathrm{d}^d}{\mathrm{d}u_1 \cdots \mathrm{d}u_d} \mathbb{P}\left(\bigcap_{i=1}^{d} A_i(u_i)\right)\,. \tag{A.2}
$$

The event $A_i(u_i)$ can be written

$$
A_i(u_i) = \begin{cases} \{V_i \geqslant \mathcal{V}(u_i),\ W_i \leqslant \Delta(V_i)\} & \text{if } u_i \leqslant \delta \\ \{V_i > \mathcal{V}(u_i),\ W_i > \Delta(V_i)\} & \text{if } u_i > \delta \end{cases}
$$

and hence the probability of the event $\bigcap_{i=1}^{d} A_i(u_i)$ can be written

$$
\int_{\mathcal{V}(u_1)}^{1} \cdots \int_{\mathcal{V}(u_d)}^{1} \int_{I_{\delta, u_1}(\Delta(v_1))} \cdots \int_{I_{\delta, u_d}(\Delta(v_d))} c_{\boldsymbol{V}, \boldsymbol{W}}(v_1, \ldots, v_d, z_1, \ldots, z_d) \mathrm{d}v_1 \cdots \mathrm{d}v_d \mathrm{d}z_1 \cdots \mathrm{d}z_d\,.
$$

Taking the derivative of this expression with respect to $u_1, \ldots, u_d$ and using (Appendix A.1) and (Appendix A.2) yields (10).

When $\boldsymbol{V}$ and $\boldsymbol{W}$ are independent the joint density factorizes and (11) follows by noting that

$$
\{W_i \in I_{\delta, u}(\delta(u))\} = \begin{cases} \{W_i \leqslant \delta(u)\} & \text{if } u \leqslant \delta \\ \{1 - W_i \leqslant 1 - \delta(u)\} & \text{if } u > \delta \end{cases} = \{p_{\delta, u}(W_i) \leqslant p_{\delta, u}(\delta(u))\}\,.
$$

Clearly (11) reduces to the simple form $c_{\boldsymbol{V}}(\mathcal{V}(u_1), \ldots, \mathcal{V}(u_d))$ when $C_{\boldsymbol{W}}$ is the independence copula.

A3. Proof of Proposition 2

The decomposition of $c_{\boldsymbol{W}^*}$ into a product of bivariate densities c_{W^*} is a straightforward consequence of the first-order Markov assumption. We then have that

$$
c_{W^*}(u_1, u_2) = \frac{C_{p_{\delta, \boldsymbol{u}}(\boldsymbol{W})}\Big(p_{\delta, u_1}\big(\delta(u_1)\big), p_{\delta, u_2}\big(\delta(u_2)\big)\Big)}{p_{\delta, u_1}\big(\delta(u_1)\big) p_{\delta, u_2}\big(\delta(u_2)\big)} = \frac{\mathbb{P}\Big(p_{\delta, u_1}(W_1) \leqslant p_{\delta, u_1}\big(\delta(u_1)\big), p_{\delta, u_2}(W_2) \leqslant p_{\delta, u_2}\big(\delta(u_2)\big)\Big)}{p_{\delta, u_1}\big(\delta(u_1)\big) p_{\delta, u_2}\big(\delta(u_2)\big)}
$$

and (17) follows easily by considering the four possible combinations of values for $p_{\delta, u_1}(\cdot)$ and $p_{\delta, u_2}(\cdot)$.

References

Aas, K., Czado, C., Frigessi, A., Bakken, H., 2009. Pair-copula constructions of multiple dependence. Insurance: Mathematics and Economics 44 (2), 182–198.

Beare, B.K., 2010. Copulas and temporal dependence. Econometrica 78 (395–410).

Beare, B.K., Seo, J., 2015. Vine copula specifications for stationary multivariate Markov chains. Journal of Time Series Analysis 36, 228–246.

Bedford, T., Cooke, R., 2001a. Probabilistic Risk Analysis: Foundations and Methods. Cambridge University Press, Cambridge.

Bedford, T., Cooke, R.M., 2001b. probability density decomposition for conditionally independent random variables modeled by vines. Annals of Mathematics and Artificial Intelligence 32, 245–268.

Bedford, T., Cooke, R.M., 2002. Vines–a new graphical model for dependent random variables. Annals of Statistics 30 (4), 1031–1068.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31, 307–327.

Bradley, R.C., 2005. Basic properties of strong mixing conditions: a survey and some open questions. Probability Surveys 2, 107–144.

Brechmann, E.C., Czado, C., 2015. Copar-multivariate time series modeling using the copula autoregressive model. Applied Stochastic Models in Business and Industry 31 (4), 495–514.

Chen, X., Fan, Y., 2006. Estimation of copula-based semiparametric time series models. Journal of Econometrics 130 (2), 307–335.

Chen, X., Wu, W.B., Yi, Y., 2009. Efficient estimation of copula-based semiparametric Markov models. Annals of Statistics 37 (6B), 4214–4253.

Creal, D., Koopman, S.J., Lucas, A., 2013. Generalized autoregressive score models with applications. Journal of Applied Econometrics 28, 777–795.

Darsow, W.F., Nguyen, B., Olsen, E.T., 1992. Copulas and Markov processes. Illinois Journal of Mathematics 36 (4), 600–642.

De Haan, L., Resnick, S., 1998. On asymptotic normality of the hill estimator. Stochastic Models 14 (4), 849–866.

Diebold, F.X., Gunther, T.A., Tay, A.S., 1998. Evaluating density forecasts with applications to financial risk management. International Economic Review 39 (4), 863–883.

Ding, Z., Granger, C.W., Engle, R.F., 1993. A long memory property of stock market returns and a new model. Journal of Empirical Finance 1, 83–106.

Domma, F., Giordano, S., Perri, P.F., 2009. Statistical modeling of temporal dependence in financial data via a copula function. Communications if Statistics: Simulation and Computation 38 (4), 703–728.

Embrechts, P., Klüppelberg, C., Mikosch, T., 1997. Modelling Extremal Events for Insurance and Finance. Springer, Berlin.

Engle, R.F., 1982. Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. Econometrica. Journal of the Econometric Society 50, 987–1008.

Fan, Y., Patton, A.J., 2014. Copulas in econometrics. Annual Review of Economics 6, 179–200.

Glosten, L.R., Jagannathan, R., Runkle, D.E., 1993. On the relation between the expected value and the volatility of the nominal excess return on stocks. The Journal of Finance 48 (5), 1779–1801.

Haff, H., Aas, K., Frigessi, A., 2010. On the simplified pair copula construction - simply useful or too simplistic? Journal of Multivariate Analysis 101, 1296–1310.

Hansen, B.E., 2000. Sample splitting and threshold estimation. Econometrica 68 (3), 575–603.

Hill, B.M., 1975. A simple general approach to inference about the tail of a distribution. The annals of statistics 1163–1174.

Ibragimov, R., 2009. Copula-based characterizations for higher-order Markov processes. Econometric Theory 25 (819–846).

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

Joe, H., 1997. Multivariate Models and Dependence Concepts. Chapman & Hall, London.

Kurowicka, D., Cooke, R., 2006. Uncertainty Analysis with High Dimensional Dependence Modelling. Wiley, Chichester.

Liebscher, E., 2008. Construction of asymmetric multivariate copulas. Journal of Multivariate Analysis 99, 2234–2250.

Loaiza-Maya, R., Smith, M.S., Maneesoonthorn, W., 2018. Time series copulas for heteroskedastic data. Journal of Applied Econometrics 33, 332–354.

Longla, M., Peligrad, M., 2012. Some aspects of modeling dependence in copula-based Markov chains. Journalof Multivariate Analysis 111, 234–240.

McNeil, A., & Bladt, M. (2021). tscopula: Time series copula models. R package version 0.2.1.

McNeil, A.J., 2021. Modelling volatility with v-transforms and copulas. Risks 9 (1), 14.

Mikosch, T., Stărică, C., 2000. Limit theory for the sample autocorrelations and extremes of a GARCH(1,1) process. The Annals of Statistics 28, 1427–1451.

Mroz, T., Fuchs, S., Trutschnig, W., 2021. How simplifying and flexible is the simplifying assumption in pair copula constructions - analytic answers in dimension three and a glimpse beyond. Electronic Journal of Statistics 15 (1), 1951–1992.

Nagler, T., Krüger, D., & Min, A. (2020). Stationary vine copula models for multivariate time series. Working paper.

Nagler, T., & Vatter, T. (2020). rvinecopulib: high performance algorithms for vine copula modeling. R package version 0.5.5.1.1.

Nelson, D.B., 1991. Conditional heteroskedasticity in asset returns: A new approach. Econometrica 59, 347–370.

Patton, A.J., 2012. A review of copula models for economic time series. Journal of Multivariate Analysis 110, 4–18.

Smith, M., 2015. Copula modelling of dependence in multivariate time series. International Journal of Forecasting 31, 815–833.

Smith, M., Min, A., Almeida, C., Czado, C., 2010. Modeling Longitudinal Data Using a Pair-Copula Decomposition of Serial Dependence. Journal of the American Statistical Association 105 (492), 1467–1479.

Spanhel, F., Kurz, M.S., 2019. Simplified vine copula models: approximations based on the simplifying assumption. Electronic Journal of Statistics 13 (1), 1254–1291.

Stöber, J., Joe, H., Czado, C., 2013. Simplified pair copula extensions—limitations and extensions. Journal of Multivariate Analysis 119, 101–118.

Tong, H., 1983. Threshold models in nonlinear time series analysis. Lecture Notes in Statistics. Springer, Berlin. Number 21

Zhao, Z., Shi, P., Zhang, Z., 2018. Modeling multivariate time series with copula-linked univariate d-vines. Preprint on arXiv.