The Canadian Journal of Statistics | La revue canadienne de statistique
Vol. , No. , , 1–27

Semiparametric and parametric distributional forecasting of univariate time series using non-Gaussian ARMA models based on D-vines

Martin BLADT
Department of Mathematical Sciences, University of Copenhagen, Copenhagen, Denmark

Alexandra DIAS
School for Business and Society, University of York, York, UK

Jialing HAN
School for Business and Society, University of York, York, UK

Alexander J. MCNEIL @
School for Business and Society, University of York, York, UK

Abstract A methodology for modelling and forecasting univariate time series using non-Gaussian ARMA and seasonal ARIMA models based on D-vine copulas is proposed. By combining a parametric D-vine process to describe serial dependence with a nonparametric or parametric model of the marginal distribution, the method offers improved modelling and distributional forecasting for time series that have a non-Gaussian distribution and a nonlinear dependence on past values. While D-vine copula-based models of univariate time series are known to generalize the classical Gaussian autoregressive (AR) model, an innovative method of parametrization based on the Kendall partial autocorrelation function is shown to permit models that generalize any ARMA model. Simulations and examples real data show the forecasting advantages of using non-Gaussian and non-linear serial dependence structures, as well as the advantages of improved marginal modelling that are offered by a copula approach.

Résumé [We will supply a French abstract for those authors who can't prepare it themselves.]

Keywords Time series; vine copulas; ARIMA models; non-Gaussian processes; semiparametric inference.
MSC2020 Primary 62G05, 62H05, 62M10, 91B84; Secondary 60G10, 60G15, 60G25, 60J05.
@ Corresponding author alexander.mcneil@york.ac.uk

1 Introduction

In this paper we consider a class of univariate non-linear and non-Gaussian time series models that generalizes the classical ARMA models and we investigate its potential in forecasting applications. The generalization was suggested in a recent paper by Bladt and McNeil (2022a) and is based on representing classical Gaussian ARMA models as stationary D-vine processes and then systematically replacing the pair copulas that describe their serial dependence structures with non-Gaussian copulas. Although it has previously been recognized that

stationary D-vines of finite order can be viewed as generalizations of classical Gaussian AR processes (see, for example, Joe, 2015, p.145), the new feature of the approach we apply is that it admits non-Gaussian analogues of ARMA models with both autoregressive and moving-average character, which permit the parsimonious modelling of a wider range of serial dependence structures.

While Bladt and McNeil (2022a) focussed on theory, in this paper we develop a practical methodology. The key enabling innovation is the use of a parametric function, known as the Kendall partial autocorrelation function (kpacf), to make the link between a classical ARMA(p,q) model and our generalized ARMA model of the same order; this allows us to define models that are equally as parsimonious as classical models. The kpacf of a stationary D-vine process $(X_t)_{t\in\mathbb{Z}}$ at lag k is, for any $t \in \mathbb{Z}$, the Kendall's tau value of the conditional distribution of the pair (X_t, X_{t+k}) given the variables $X_{t+1}, \dots, X_{t+k-1}$ lying in between (Bladt and McNeil, 2022a).

We extend Bladt and McNeil (2022a) in a number of ways. We propose an approach to non-Gaussian seasonal ARIMA modelling, which is important for capturing the well-known seasonal effects in many macroeconomic and physical time series. We compare the semiparametric and parametric approaches to copula inference and provide an original interpretation of the concept of semiparametric ARMA modelling. Finally, by considering both the semiparametric and parametric approaches, we shed light on the forecasting advantages of using non-Gaussian and non-linear serial dependence structures, the advantages of using well-specified parametric margins, particularly in samples of moderate size, and the disadvantages of using poorly specified parametric margins in some circumstances. We investigate these issues using the framework for evaluating distributional forecasts proposed by Gneiting and Ranjan (2011).

Our methodology is a contribution to the growing literature on vine copula models for time series. The use of vine copulas for modelling dependent data has been developed in many publications including Joe (1996, 1997), Bedford and Cooke (2001a,b, 2002), Kurowicka and Cooke (2006), Aas et al. (2009) and Smith et al. (2010). The specific application to time series is investigated in Darsow et al. (1992), Chen and Fan (2006), Beare (2010), Beare and Seo (2015) and Nagler et al. (2022), among others. The last of these papers introduces the terminology S-vine to describe vine copula structures that are particularly suitable for modelling stationary multivariate time series and this includes stationary univariate D-vines as a special case. In Bladt and McNeil (2022a) the terminology S-vine is adopted but in this paper we use the more familiar D-vine label for univariate models.

The semiparametric estimation approach to copula inference, in which marginal distributions are estimated by the empirical distribution function while copulas are modelled parametrically, has been widely applied to copula inference since the seminal paper of Genest et al. (1995); nonparametric approaches using a form of kernel density estimate are also possible, although care should be taken with heavy-tailed data since standard density estimators may give poor tail estimates (Buch-Kromann et al., 2005). On the other hand, the fully parametric approach to inference for models consisting of copulas and marginal distributions was first extensively investigated by Joe (1996) under the name "inference functions for margins". The semiparametric approach was extended to time series by Chen and Fan (2006) who investigated first-order Markov copula models, proposed a semiparametric forecasting procedure for the quantile function of the conditional distribution of X_{t+1} given X_t, and gave technical conditions for the consistency and asymptotic normality of the estimates of the copula parameter. More recently, theoretical results for semiparametric inference in D-vine analogues of AR(p) models have been provided by Zhao et al. (2022) and an extremely comprehensive set of results for semiparametric and parametric sequential inference in possibly-misspecified multivariate S-vines (subsuming univariate D-vines) has been established by Nagler et al. (2022). Note also, that the application of D-vine modelling to time series can be viewed as quantile autoregression, following the insights of Kraus and Czado (2017).

Yan and Genton (2019) group the most common approaches to non-Gaussian ARMA modelling into three categories: approaches in which Box-Cox and related transformations are used to make the modelled variable more normal; approaches based on ARMA models with non-Gaussian innovations (e.g. Li and McLeod (1988)); approaches which apply a GLM approach to time series (e.g. Benjamin et al. (2003)). All of these approaches impose limitations: Nelson and Granger (1979) report that the Box-Cox transformation is seldom successful in inducing normality in data; models with non-Gaussian innovations do not address the issue of non-linearity; the GLM approach of Benjamin et al. (2003) is only designed for exponential families. The approach that we take offers greater flexibility in principle, since any kind of marginal behaviour may be modelled and D-vine copulas offer a rich variety of non-linear serial dependence structures; limitations are of a more practical, computational nature and relate to the need for faster code for certain models.

The approach in this paper is primarily designed for applications where classical ARMA (or more generally

SARIMA) models are commonly used and where the serial dependencies between lagged variables are typically monotonic. By using cross-shaped copulas of the kind explored in Loaiza-Maya et al. (2018) and Bladt and McNeil (2022b), it could potentially also be applied in cases where serial dependencies are non-monotonic, for example, financial time series of asset returns displaying stochastic volatility and frequent sign changes.

The paper is structured as follows. Section 2 gives a concise overview of the theory of stationary D-vine processes while Section 3 proposes a new generalized seasonal ARMA model based on a D-vine and describes the method of parameterization via the kpacf. Estimation and forecasting are treated in Sections 4 and 5 and a simulation study shows the effectiveness of our proposed approach to inference. Applications of the forecasting methodology to quarterly U.S. inflation data and the absolute values of daily return data for the Nasdaq Composite index, relevant to forecasting market volatility, are provided in Section 6 while Section 7 concludes.

2 Stationary D-vine processes

2.1 Notation and assumptions

In this section we summarise the main elements of the presentation in Bladt and McNeil (2022a), making some small changes to notation to simplify exposition. Vectors $\mathbf{x} = (x_1, \dots, x_d)^\top$ are written in boldface and sequences are denoted $(x_t)_{t\in T}$ where the index set T is either the non-zero natural numbers $\mathbb{N}$ or the integers $\mathbb{Z}$. We write $\mathbf{x}_{[t:s]} = (x_t, \dots, x_s)^\top$ to refer to a sub-vector of $\mathbf{x}$ or a finite section of a sequence $(x_t)_{t\in T}$; note that we permit $t > s$, in which case the variables are taken in reverse temporal order. Random variables and sequences of random variables are denoted by capital letters.

Models are built from a a sequence of bivariate copulas $(C_k)_{k\in\mathbb{N}}$. We impose the technical assumption that each C_k belongs to the class $\mathcal{C}^\infty$ of smooth functions on $[0,1]^2$ with continuous partial derivatives of all orders and densities that are strictly positive on $(0,1)^2$. This ensures that finite-order stationary D-vine processes can be treated as Markov chains and is satisfied by all the standard pair copulas that are used in vine copula models, such as Gauss, t, Gumbel, Frank, Joe and Clayton; for the latter, we exclude the case with negative parameter since this does not have a strictly positive density throughout $(0,1)^2$.

2.2 Stationary D-vine copulas

Expressions for the densities of general simplified D-vines can be found in Aas et al. (2009) and Joe (2015) while the application to time series is studied in Smith et al. (2010). Under an assumption of strict stationarity the expressions for time series take a particularly simple form because the bivariate copulas in the so-called D-vine trees (Joe, 2015, Figure 3.5) must all be the same in each tree. We can write the density in dimension $d \geq 2$ as

$$c_{(d)}(u_1, \dots, u_d) = \prod_{k=1}^{d-1} \prod_{j=k+1}^{d} c_k\Big(B_{k-1}(u_{j-k}; \mathbf{u}_{[j-k+1:j-1]}), R_{k-1}(u_j; \mathbf{u}_{[j-1:j-k+1]})\Big) \tag{1}$$

where $(c_k)_{k\in\mathbb{N}}$ are the densities of the bivariate copulas in the sequence $(C_k)_{k\in\mathbb{N}}$ and where $R_k : (0,1) \times (0,1)^k \to (0,1)$ and $B_k : (0,1) \times (0,1)^k \to (0,1)$ are families of functions defined from $(C_k)_{k\in\mathbb{N}}$ in a recursive, interlacing fashion by $R_1(x;u) = h_1^{(1)}(u,x)$, $B_1(x;u) = h_1^{(2)}(x,u)$ and

$$\begin{aligned} R_k(x;\mathbf{u}) &= h_k^{(1)}\left(B_{k-1}(u_k; \mathbf{u}_{[k-1:1]}), R_{k-1}(x; \mathbf{u}_{[1:k-1]})\right) \\ B_k(x;\mathbf{u}) &= h_k^{(2)}\left(B_{k-1}(x; \mathbf{u}_{[1,k-1]}), R_{k-1}(u_k; \mathbf{u}_{[k-1:1]})\right) \end{aligned} \tag{2}$$

for $k \geq 2$, where $h_k^{(i)}(u_1,u_2) = \frac{\partial}{\partial u_i} C_k(u_1,u_2)$. Note that, by slight abuse of notation, R_0 and B_0 in formula (1) should be interpreted as the identity functions $R_0(x,\cdot) = B_0(x,\cdot) = x$ for all x, a convention we use throughout the paper.

As an example, when $d = 4$ the density in (1) would take the form

$$\begin{aligned} c_{(4)}(u_1,u_2,u_3,u_4) = c_1(u_1,u_2)\, c_1(u_2,u_3)\, c_1(u_3,u_4) \times c_2\left(B_1(u_1;u_2), R_1(u_3;u_2)\right)\, c_2\left(B_1(u_2;u_3), R_1(u_4;u_3)\right) \times \\ c_3\left(B_2(u_1,(u_2,u_3)^\top), R_2(u_4;(u_3,u_2)^\top)\right) \end{aligned}$$

where

$$R_1(x;u_1) = \frac{\partial}{\partial u} C_1(u,v)\Big|_{u=u_1,v=x}, \quad B_1(x;u_1) = \frac{\partial}{\partial v} C_1(u,v)\Big|_{u=x,v=u_1},$$

$$R_2\left(x;(u_1,u_2)^\top\right) = \frac{\partial}{\partial u}C_2(u,v)\Big|_{u=B_1(u_2;u_1),v=R_1(x;u_1)} \text{ and } B_2\left(x;(u_1,u_2)^\top\right) = \frac{\partial}{\partial v}C_2(u,v)\Big|_{u=B_1(x;u_1),v=R_1(u_2;u_1)}$$

so that it can be evaluated from the densities c_1, c_2 and c_3 and the partial derivatives (or h-functions) of the copulas C_1 and C_2, which are readily available for standard bivariate copula families.

2.3 Stationary D-vine processes

Following Bladt and McNeil (2022a), we say that a strictly stationary time series $(X_t)_{t\in\mathbb{Z}}$ is a stationary D-vine process if for every $t \in \mathbb{Z}$ and $d \geq 2$ the distribution of the vector $(X_t, \dots, X_{t+d-1})$ is absolutely continuous and admits a unique copula $C_{(d)}$ with a joint density $c_{(d)}$ of the form (1). We say that a stationary D-vine process $(U_t)_{t\in\mathbb{Z}}$ is a stationary D-vine copula process if its univariate marginal distribution is standard uniform.

The sequences of functions R_k and B_k for $k \geq 1$ are referred to as forward and backward Rosenblatt functions. If the copulas C_k are exchangeable for $k \in \{1, \dots, d\}$ and $d \geq 1$, then $B_k(x;\mathbf{u}) = R_k(x;\mathbf{u})$ for $k \in \{1, \dots, d\}$. In this case we can drop the forward and backward qualification and the recursions in (2) simplify to

$$R_k(x;\mathbf{u}) = h_k^{(1)}\left(R_{k-1}(u_k;\mathbf{u}_{[k-1:1]}), R_{k-1}(x;\mathbf{u}_{[1:k-1]})\right). \tag{3}$$

The Rosenblatt functions have important roles in prediction and forecasting. If we take a stationary D-vine process $(X_t)_{t\in\mathbb{Z}}$ with continuous marginal distribution F_X and transform it to a stationary D-vine copula process $(U_t)_{t\in\mathbb{Z}}$ by means of the componentwise transformation $U_t = F_X(X_t)$ then

$$\begin{aligned} R_k(x;\mathbf{u}) &= \mathrm{P}\left[U_t \leq x \mid U_{t-1} = u_1, \dots, U_{t-k} = u_k\right] \\ B_k(x;\mathbf{u}) &= \mathrm{P}\left[U_t \leq x \mid U_{t+1} = u_1, \dots, U_{t+k} = u_k\right]. \end{aligned} \tag{4}$$

In particular, the forward functions are the conditional distribution functions of terms in the process $(U_t)_{t\in\mathbb{Z}}$ given previous values. The set of functions $R_0, \dots, R_k$ are precisely the functions required to map the vector $(U_{t-k}, \dots, U_t)^\top$ into a vector of independent uniform random variables via the Rosenblatt transformation (Rosenblatt, 1952), hence their name.

The derivatives of the Rosenblatt forward functions $r_k(x;\mathbf{u}) = \frac{\partial}{\partial x}R_k(x;\mathbf{u})$ are conditional densities and satisfy the simpler equations

$$r_k(x;\mathbf{u}) = \frac{c_{(k+1)}(u_k, \dots, u_1, x)}{c_{(k)}(u_k, \dots, u_1)} = \prod_{j=1}^{k} c_j\big(B_{j-1}(u_j;\mathbf{u}_{[j-1:1]}), R_{j-1}(x;\mathbf{u}_{[1:j-1]})\big) \tag{5}$$

which facilitate density forecasts. The conditional distribution functions represented by the Rosenblatt forward functions have unique inverses $R_k^{-1}(z;\mathbf{u})$, referred to as Rosenblatt quantile functions, which can be used to generate realisations from a stationary D-vine process.

As explained in Bladt and McNeil (2022a), stationary D-vine processes can be thought of as extending the class of causal stationary Gaussian processes. Every such Gaussian process (for example, an ARMA or ARFIMA model) can be represented as a stationary D-vine process consisting of a sequence of bivariate Gaussian copulas $(C_k)_{k\in\mathbb{N}}$ and a Gaussian marginal distribution function F_X. For an AR(p) process the sequence can be truncated, meaning that the copulas $(C_k)_{k>p}$ are all independence copulas and can effectively be ignored since the copula densities appearing in expressions (1) and (5) satisfy $c_k(u,v) \equiv 1$ for $k > p$; for an ARMA(p,q) process with $q \neq 0$ the sequence cannot be truncated but the copulas C_k tend towards the independence copula in the limit as $k \to \infty$. In the more general family of stationary non-Gaussian D-vine processes we replace the Gaussian marginal distribution by an arbitrary continuous distribution and we replace the sequence of Gaussian copulas by a sequence of arbitrary bivariate copulas satisfying the technical assumption of Section 2.1. We refer to D-vine processes as being of finite and infinite order respectively according to whether the copula sequence is truncated or not.

The copula sequence $(C_k)_{k\in\mathbb{N}}$ determines the serial dependence properties of the process and is referred to as the partial copula sequence of the process. This is in analogy to the concept of partial correlation since the kth partial copula describes the dependence structure of the conditional distribution of two variables X_t and X_{t+k} conditional on the variables $X_{t+1}, \dots, X_{t+k-1}$ in between; in fact, when the stationary D-vine process is a Gaussian process, the parameter of the kth Gaussian copula in the partial copula sequence is precisely the kth partial correlation coefficient α_k. The Kendall partial autocorrelation coefficient τ_k is the Kendall's tau value of C_k and the Kendall partial autocorrelation function (kpacf) is the sequence $(\tau_k)_{k\in\mathbb{N}}$.

A stationary Gaussian process $(X_t)_{t\in\mathbb{Z}}$ is completely identified by its kpacf, up to scale and location. This follows because there is a bijective mapping from the kpacf to the partial correlation function (pacf) $(\alpha_k)_{k\in\mathbb{N}}$ of $(X_t)_{t\in\mathbb{Z}}$, via the transformation $\alpha_k = \sin(\pi\tau_k/2)$, and hence to the autocorrelation function (acf) of $(X_t)_{t\in\mathbb{Z}}$. This observation is key to our method of defining non-Gaussian analogues of classical Gaussian ARMA processes.

The advantage of considering a class of processes that allows non-Gaussian partial copulas is that it permits the modelling of serial dependencies that are asymmetric, so that joint large values behave differently to joint small values, and also tail dependent, so that the most extreme obswervations have a particular tendency to occur together in clusters. Joe et al. (2010) have shown that using a copula with upper tail dependence for C_1, such as Gumbel, Joe or survival Clayton, leads to multivariate tail dependence of the vectors $X_t, \dots, X_{t+k}$ for $k \geq 2$. Interestingly, in our empirical examples, we will find cases where tail-dependent copulas are selected at lags $k > 1$ in addition to lag $k = 1$, suggesting that tail dependence in the first tree of a D-vine is not sufficient to model joint multivariate tail behaviour of many real-world time series.

2.4 Theoretical issues

As soon as we allow non-Gaussian copulas C_k, there are some unresolved theoretical questions concerning the constraints that must be imposed on an infinite sequence $(C_k)_{k\in\mathbb{N}}$ in order to obtain a well-behaved stochastic process. Stationary D-vine processes are strictly stationary by design so the main theoretical issue is whether they are also ergodic processes with mixing behaviour that facilitates statistical inference. At a minimum, the copula sequence must converge to the independence copula but this alone is not sufficient and speed of convergence is important. More discussion of these issues is found in Bladt and McNeil (2022a).

In this paper we adopt the expedient of approximating infinite-order processes by finite-order processes of very high order. This makes negligible difference to the results of optimization procedures but, as well as sidestepping this theoretical issue, has the practical advantage of accelerating times required to fit models. Finite-order stationary D-vine processes are ergodic Markov processes and can be thought of as extensions of the Gaussian AR process. A number of authors have explored mixing and ergodic convergence rates for first-order Markov copula models (Chen and Fan, 2006; Beare, 2010; Chen et al., 2009; Longla and Peligrad, 2012) and it is known that models based on Gaussian, Student t, Frank, Clayton and Gumbel copulas are all geometrically β-mixing and geometrically ergodic. For higher-order Markovian D-vine models, Zhao et al. (2022) use a small set approach for Markov chains to give conditions for geometric ergodicity, although the conditions are extremely difficult to verify for an arbitrary set of copulas $C_1, \dots, C_K$.

3 D-vine processes with seasonal ARMA dependence

3.1 Seasonal ARIMA and ARMA processes

In this section we propose a new class of stationary D-vine processes with what we will refer to as seasonal ARMA dependence; this is the dependence structure of a differenced seasonal ARIMA (or SARIMA) model and includes the usual ARMA model as a special case.

We recall that a stochastic process $(Y_t)_{t\in\mathbb{Z}}$ is referred to as a SARIMA$(p,d,q)(P,D,Q)_s$ process if it satisfies the equations

$$\left(1-\sum_{i=1}^{p}\phi_i B^i\right)\left(1-\sum_{j=1}^{P}\Phi_j B^{js}\right)(1-B)^d(1-B^s)^D Y_t = \left(1+\sum_{k=1}^{q}\psi_k B^k\right)\left(1+\sum_{l=1}^{Q}\Psi_l B^{ls}\right)\epsilon_t$$

where B is the usual backshift operator, (ϕ_i), (Φ_j), (ψ_k) and (Ψ_l) are the AR, seasonal AR, MA and seasonal MA coefficients, $(\epsilon_t)_{t\in\mathbb{Z}}$ are white noise innovations and s is the periodicity or number of seasons per cycle. For example, for macroeconomic data showing an annual cycle we would typically set $s = 4$ for quarterly data and $s = 12$ for monthly data. The SARIMA process obviously constitutes a generalization of the usual ARIMA(p,d,q) process for which $P = D = Q = 0$.

If we define $X_t = (1-B)^d(1-B^s)^D Y_t$ then $(X_t)_{t\in\mathbb{Z}}$ is the stationary process that results from taking the dth ordinary difference (by iterated application of $\Delta Y_t = Y_t - Y_{t-1}$) together with the Dth seasonal difference (by iterated application of $\Delta^* Y_t = Y_t - Y_{t-s}$). The differenced process $(X_t)_{t\in\mathbb{Z}}$ could be said to follow a seasonal ARMA model of order $(p,q)(P,Q)_s$. In effect, this is simply an ARMA$(p+sP, q+sQ)$ process, but one in which the parameters are subject to a set of constraints that must be observed when estimating the model; there are only $p+P+q+Q$ free parameters.

Recall from the end of Section 2.3 that stationary Gaussian processes are uniquely identified, up to scale and location, by their kpacf. In view of the characterizing role of the kpacf in Gaussian processes we will refer more generally to any stationary D-vine process whose kpacf is identical to that of a Gaussian ARMA(p, q) process as an ARMA(p, q) D-vine process. We will also refer to D-vine processes as ARMA$(p, q)(P, Q)_s$ processes if we want to emphasise the seasonal character of the underlying models and the application of the corresponding parameter constraints in their estimation. We note that this form of label is more ambiguous since p, q, P and Q are not uniquely identified by the kpacf, only the values of $p+sP$ and $q+sQ$. For example, an ARMA$(1, 1)(1, 1)_4$ process shares its kpacf with an ARMA$(5, 1)(0, 1)_4$ model, an ARMA$(1, 5)(1, 0)_4$ model and an ARMA$(5, 5)$ model.

3.2 Model parameterization

In Bladt and McNeil (2022a) a parsimonious method of parameterizing stationary D-vine processes via their kpacf is proposed and this method may be applied to an ARMA$(p, q)(P, Q)_s$ D-vine process. Suppose the periodicity s and order (p, q, P, Q) of the process are fixed and let $\theta = (\phi_1, \dots, \phi_p, \psi_1, \dots, \psi_q, \Phi_1, \dots, \Phi_P, \Psi_1, \dots, \Psi_Q)^\top$ represent a vector of feasible parameters for a causal stationary and invertible Gaussian ARMA$(p, q)(P, Q)_s$ process. The kpacf of the Gaussian process is given by $(\tau_k(\theta))_{k\in\mathbb{N}}$ where $\tau_k(\theta) = (2/\pi) \arcsin \alpha_k(\theta)$ and $(\alpha_k(\theta))_{k\in\mathbb{N}}$ is the pacf, which can be readily calculated for any Gaussian ARMA process; for more details, see Appendix A. The idea is that we consider stationary D-vine models with sequences $(C_k)_{k\in\mathbb{N}}$ such that the Kendall's tau values satisfy $\tau(C_k) = (2/\pi) \arcsin \alpha_k(\theta)$ for all $k \in \mathbb{N}$ but where the copulas in the sequence may be taken from non-Gaussian copula families.

For non-Gaussian copulas it is convenient if there is an explicit one-to-one relationship between the parameter of the copula and Kendall's tau (for example, the Frank, Clayton, Gumbel and Joe copulas) but this is not necessary. If a copula C_k has more than one parameter (for example, the t copula or BB1 copula) the parameters are optimized under the constraint that $\tau(C_k) = (2/\pi) \arcsin \alpha_k(\theta)$ where θ is a feasible parameter vector for a Gaussian ARMA$(p, q)(P, Q)_s$ process. This means that for fixed θ a 2-parameter copula C_k contributes only one further free parameter. If we write the vector of additional free parameters as θ_2 the model is estimated by jointly optimizing θ over the set of feasible parameters for a causal stationary Gaussian ARMA $(p, q)(P, Q)_s$ process and θ_2 over the feasible domain for the additional free parameters.

3.3 Practical considerations

While we have a lot of flexibility in choosing copula sequences, our strategy in this paper is to begin with sequences of Gaussian copulas and to systematically replace a finite number of copulas in the sequence with non-Gaussian copulas. We make substitutions from the one-parameter Frank, Clayton, Gumbel and Joe copula families and the two-parameter t and BB1 families (Joe, 2015); the latter two families offer, respectively, symmetric and asymmetric tail dependence in both tails. The set of copula choices is partly dictated by programming considerations since the R package **tscopula** that is used draws on fast C++code for bivariate copulas accessed via the **rvinecopulib** package of Nagler and Vatter (2025).

For the copula families that are not radially symmetric (symmetric under a 180 degree rotation) we allow rotations through 180 degrees, since the rotated copulas can sometimes offer better fits than the unrotated copulas. For the copula families that are not comprehensive (able to attain any Kendall's tau value in $(-1, 1)$) we also allow rotations through 90 or 270 degrees; this applies to the Gumbel, Clayton, Joe and BB1 families. This may be necessary to model negative partial dependencies corresponding to negative values in the kpacf.

When we fit models to data, we use the initial Gaussian copula model to establish the possible order of the process (p, q, P and Q) and then fit processes with finitely many non-Gaussian substitutions which share their kpacf with a Gaussian ARMA$(p,q)(P,Q)_s$ process.

In practice we apply truncation to all infinite-order models (models with $q > 0$ or $Q > 0$) so that they are approximated by finite-order models, i.e. AR models of high order. For some small $\epsilon > 0$, we can always find a value K such that $|\tau_K(\theta)| \geq \epsilon$ and $|\tau_k(\theta)| < \epsilon$ for $k > K$. Since Gaussian copula sequences tend to the independence copula $C^\perp$ as $\tau_k \to 0$, then, for ϵ sufficiently small, we may assume that the copulas $(C_k)_{k>K}$ are almost indistinguishable from the independence copula and set $C_k = C^\perp$ for all $k > K$. We refer to K as the effective maximum lag at tolerance ϵ. If we describe a model as an ARMA$(p, q)(P, Q)_s$ truncated to lag K we mean that its kpacf coincides with that of a ARMA$(p, q)(P, Q)_s$ up to lag K; in reality it is an AR(K) model, but one where the kpacf is parameterized by only $p + q + P + Q \ll K$ parameters.

4 Estimation of stationary D-vines

4.1 Outline of procedure

Our proposed estimation approach has three main stages.

A. Determination of initial model order. We use the automatic SARIMA procedure of Hyndman and Khandakar (2008) to arrive at a plausible specification for the model order $(p,d,q)(P,D,Q)_s$ and then, if necessary, we difference the data according to the values of d and D to obtain data $\{x_t : t = 1,\dots,n\}$ that can be modelled as a stationary ARMA$(p,q)(P,Q)_s$ process. This procedure works best when the differenced data have a marginal distribution that is close to normal. When the marginal distribution is clearly non-Gaussian, a more accurate ARMA identification can generally be obtained by computing the scaled empirical distribution function (EDF) $\widehat{F}_X^{(n)} = (1/(n+1))\sum_{i=1}^{n} I_{\{x_i \le x\}}$, forming pseudo-copula data $\{\hat{u}_t = \widehat{F}_X^{(n)}(x_t) : t = 1,\dots,n\}$, and passing the probit-transformed data $\{\Phi^{-1}(\hat{u}_t) : t = 1,\dots,n\}$ through the automatic procedure again. We consider this issue in Section 4.4 and provide evidence to support our recommended methodology.

B. Estimation of SARMA model with Gaussian pair copulas. In the semiparametric version we take the pseudo-copula data $\{\hat{u}_t = \widehat{F}_X^{(n)}(x_t) : t = 1,\dots,n\}$ and maximize the log-likelihood

$$L(\theta;\mathbf{x}) = \log c_{(n)}(\hat{u}_1,\dots,\hat{u}_n) \tag{6}$$

with respect to the parameters θ of a Gaussian ARMA$(p,q)(P,Q)_s$ model; recall that $c_{(n)}$ is the n-dimensional density of a stationary D-vine copula defined in (1) and, at this stage, the copula sequence $(C_k)_{k\in\mathbb{N}}$ consists entirely of Gaussian copulas. Maximization can be achieved using a fast Kalman filter approach. In the parametric or density estimation version of stage B, the pseudo-copula data in (6) are replaced by data $\{\tilde{u}_t = \widehat{F}_X(x_t) : t = 1,\dots,n\}$ where $\widehat{F}_X$ denotes a parametric or kernel density estimate of F_X.

C. Replacement of low-order copulas with non-Gaussian alternatives. We fit a series of models in which C_1 is replaced by a non-Gaussian pair copula from our list of 1-parameter and 2-parameter alternatives; for the latter this introduces additional parameters θ_2 into the log-likelihood (6). Since the introduction of non-Gaussian copulas removes the linearity of the model, maximization of (6) can no longer use the Kalman filter approach and it is important to increase speed by truncating the copula sequence at an effective maximum lag K for some tolerance ϵ. The copula C_1 which reduces the AIC by the largest amount with respect to the initial Gaussian copula model is adopted. The process is then repeated for $C_2, C_3, \dots$ until the AIC can no longer be lowered significantly or until a predetermined number of replacements has been made.

Some remarks on the procedure. The ultimate test of a model derived through this procedure is whether it can give superior forecasts to a simpler alternative, such as the model with only Gaussian pair copulas obtained after stage B. Before looking in more detail at the subject of forecasting in Section 5, we devote the remainder of Section 4 to a discussion of the properties of parameter estimates.

It should first be noted that the procedure described above will, more often that not, deliver a model that is misspecified in some way. The estimated order of the model may be wrong since this is a difficult thing to identify exactly, even from perfectly Gaussian data. The use of model order in our methodology is a device to preserve the parsimony of classical ARMA models with moving average terms, but the exact identification of order is less important than the estimation of the values of the Kendall autocorrelation function (τ_k).

Moreover, some of the copulas comprising the sequence (C_k) may be wrongly identified. In our experience, Gaussian copulas can often be mistaken for Frank copulas, and vice versa, while other pairs that are often confused are Gumbel and survival Clayton or Clayton and survival Gumbel. The tendency to confuse members of these pairs increases as the absolute value of the Kendall correlation $|\tau_k|$ decreases, since the degree of differentiation weakens as the copulas approach independence. Again, some misspecification of copulas, particularly at higher lags k, may not be a serious problem if the kpacf can be estimated accurately.

4.2 Asymptotic theory

Although asymptotic theory exists for inference in finite-order D-vines, there remain gaps in its application to specific higher-order models. Two recent papers by Zhao et al. (2022) and Nagler et al. (2022) provide theoretical results that are relevant to the application of the semiparametric method based on the EDF to stationary D-vine models of finite order $p \ge 1$. The results in Nagler et al. (2022) are the most comprehensive to date and also encompass the fully parametric case and the case where the vine model may be misspecified. In particular, these authors consider the situation where the copula parameters are estimated sequentially, i.e. where the parameter(s) of $C_1,\dots,C_p$ are estimated in p successive steps, with previously estimated parameters held

fixed at each step. In contrast, we use a sequential method to identify copulas but then estimate all parameters of the kpacf in a single step.

In addition to typical regularity conditions for semiparametric maximum likelihood inference, the requirement for $\sqrt{n}$-consistency in Nagler et al. (2022) is that the underlying process $(X_t)_{t\in\mathbb{Z}}$ is absolutely regular, so that the β-mixing coefficients tend to zero; this is true under the assumptions of our model. However, for asymptotic normality, Nagler et al. (2022) show the rate of decay of the mixing coefficients is important and either a geometric or polynomial β-mixing rate is required, depending on moment conditions for the score function of the likelihood. The collective work of Chen and Fan (2006), Beare (2010), Chen et al. (2009) and Longla and Peligrad (2012) shows that first-order models ($p = 1$) with Gauss, Frank, Clayton, Gumbel or t copulas are in fact β-mixing with the faster geometric rate. However, to our knowledge, the mixing rates have not been established for combinations of $p > 1$ copulas, making it difficult to be certain whether the asymptotic normality results hold for specific higher-order models used in practice.

4.3 Design of simulation study

For the simulation study we consider models of order ARMA(1,1) with parameter vector $\theta = (\phi_1, \psi_1)^\top$ and SARMA(1,1)(1,1)$_4$ with parameter vector $\theta = (\phi_1, \psi_1, \Phi_1, \Psi_1)^\top$. In each replication of the experiment we generate random values of the model parameters independently subject to the following constraints; (1) the absolute value of each parameter must lie in $(0.15, 0.85)$; (2) the conditions $|\phi_1 + \psi_1| > 0.15$ and $|\Phi_1 + \Psi_1| > 0.15$ must hold; (3) for the corresponding kpacf $\tau_k(\theta)$ we must have $\min\{|\tau_1|, |\tau_2|, |\tau_3|\} > 0.1$; (4) the effective maximum lag K for a tolerance of $\epsilon = 0.0001$ must be no larger than 20. Condition (1) is to make sure that models are well differentiated from cases where autoregressive and moving-average parameters are zero, or cases where the characteristic polynomial of the model has unit roots; (2) distances models from the case of repeated roots; (3) ensures that the values of Kendall's tau at the first 3 lags are not too close to zero so that the first 3 partial copulas are well differentiated from the independence copula; (4) is imposed so that the kpacf does not decay too slowly, which would tend to lead to very long simulation times.

For each simulated set of parameters realisations of stationary D-vine copula processes with lengths $n \in \{200, 500, 1000\}$ are generated for (i) a copula sequence consisting only of Gaussian pair copulas and (ii) a copula sequence in which the first 3 copulas are randomly changed to non-Gaussian with the remaining copulas left as Gaussian; while a larger number of copulas could be changed, the choice of 3 ensures shorter simulation times and is sufficient to understand the effect of introducing non-Gaussian dependence. The non-Gaussian copulas are randomly chosen from the set of one-parameter copulas listed previously, including 180-degree rotations when partial Kendall correlations are positive, and rotations through 90 or 270 degrees when partial correlations are negative. We omit the 2-parameter t copula and BB1 copulas, which give similar results, albeit with longer running times.

For the ARMA(1,1) example, we endow the data with a heavy-tailed Student t marginal distribution F_X with 6 degrees of freedom and consider both the semiparametric and fully parametric approaches. The obvious non-normality of the data means that, for order identification in step A, we use the probit transformation of the pseudo-copula data rather than the raw data. The semiparametric EDF method is then robust to the choice of marginal distribution in the sense that, any choice of non-Gaussian margin would lead to identical estimates of the parameters of the copula process. For the fully parametric method the choice of margin has an influence on the estimation of the copula process; other choices of margin, for example including skewness, could lead to slightly different results. For the SARMA(1,1)(1,1)$_4$ example, the copula replacement procedure is much slower and we apply a normal marginal distribution to the simulated series and only consider the EDF method.

The data based on a Gaussian copula process are modelled by carrying out steps A and B of the procedure in Section 4.1 while the data generated from a non-Gaussian copula process are modelled by carrying out steps A, B and C; we are particularly interested in the improvement offered by step C. The automatic SARIMA procedure is told only that the data are stationary (i.e. we omit determination of the difference orders d and D) and that the data do or do not follow a seasonal process.

Interest centres on 3 aspects of the procedure. How often is the correct order, or a nearly correct order, determined in step A and does this vary for the different types of copula sequence? In the case of non-Gaussian copula sequences, how often are the non-Gaussian copulas at lags 1 to 3 correctly identified? How accurate are the estimates of the Kendall's tau values τ_1, τ_2 and τ_3 at the first 3 lags and are there differences for the Gaussian and non-Gaussian copula sequences?

4.4 Results for order identification

Table 1: Identification of model order using the automatic SARIMA procedure of Hyndman and Khandakar (2008) and the variant of this procedure described in Section 4.1. Numbers show the percentage of models for which the identified order is either exactly correct or close to correct and are based on 1000 model simulations.

			ARMA(1,1) Student t margin			SARMA(1,1)(1,1)$_4$ Normal margin		
Step A	Copula sequence	identification | n	200	500	1000	200	500	1000
raw	Gaussian	exact	55.5	58.5	61.2	6.3	11.5	15.7
		close	65.3	65.9	69.4	36.1	38.6	38.3
	non-Gaussian	exact	38.9	33.8	28.5	3.7	8.4	11.5
		close	58.6	50.3	44.6	28.8	29.4	28.5
probit	Gaussian	exact	55.3	57.7	59.0			
		close	63.4	63.8	67.3			
	non-Gaussian	exact	43.6	39.3	30.8			
		close	59.6	54.1	46.4			

Table 1 addresses order identification. A close identification refers to the case where the estimated order is out by 1; in the case of ARMA(1,1) this means that the chosen model is AR(1), MA(1), ARMA(2,1) or ARMA(1,1). The rows in which Step A is labelled 'raw' refer to the order suggested by the automatic SARIMA procedure applied to the raw data. For the ARMA(1,1) example the data have a Student t6 marginal distribution and are clearly non-Gaussian. In this case, the rows labelled 'probit' refer to the order obtained when the procedure is applied to the probit-transformed pseudo-copula data, as advocated in Section 4.1. It may be noted that, while this leads to slightly worse identification rates in the case of the Gaussian copula sequence, it leads to better identification rates for the non-Gaussian copula sequence. Hence subsequent analyses of the ARMA(1,1) data are based on the model orders identified by the 'probit' method.

We note that the identification success rates are lower when non-Gaussian copulas are present in the copula sequence but identification is by no means perfect in the case of models with only Gaussian copulas. While the identification success rate improves with sample size for the case of Gaussian copulas, it actually deteriorates for the case of non-Gaussian copulas. Order identification is particularly difficult for the seasonal ARMA models. However, as discussed in Section 4.1, identification is a means to the end of parsimoniously parameterizing a copula sequence, so the results for copula choice and parameter accuracy should be accorded more importance.

4.5 Results for semiparametric approach

We now consider copula identification and parameter estimation using the EDF method for both the ARMA(1,1) and SARMA(1,1)(1,1)$_4$ example. Table 2 addresses copula identification and shows the percentage of non-Gaussian models for which the first few copulas are correctly identified. The percentages are high if one considers only the first copula C_1 but, as would be expected, they decrease as one considers more copulas. Kendall's tau typically decreases in absolute value with lag so the problem of copula identification tends to become more difficult for higher-order partial copulas. Moreover, if, for example, an AR(2) model is identified in step A, then C_3 will be incorrectly identified as 'missing'. Copula identification clearly improves with increasing sample size.

Table 3 addresses the accuracy of the estimates of the true Kendall correlations τ_1, τ_2 and τ_3, which can be thought of as the parameters of the first 3 copulas. In every simulation run the values of τ_1, τ_2 and τ_3 are different, but common to both the Gaussian and the non-Gaussian D-vines. For this reason, we do not tabulate parameter estimates and standard errors, but instead give the square roots of the average squared errors, or RMSEs. To analyse whether the errors are materially different, we carry out two-sided Kolmogorov-Smirnov tests. First, we test the hypothesis that, for the non-Gaussian case, the errors obtained from the full procedure comprising the steps A, B and C have a similar distribution to those obtained from the procedure comprising steps A and B only. Next, we test the hypothesis that the errors obtained from carrying out steps A and B in the Gaussian case have a similar distribution to those obtained from carrying out the full procedure (A+B+C)

Table 2: Percentage of models with non-Gaussian dependence for which the first 3 copulas are correctly identified. Results are based on 1000 model simulations. The true copulas C_1, C_2 and C_3 are drawn randomly from the Frank, Clayton, Gumbel, Joe and Gaussian familes; random rotations of Clayton, Gumbel and Joe through 90, 180 and 270 degrees are allowed. The Kendall's tau values for these copulas are in the range $[-0.74, -0.1] \cup [0.1, 0.75]$ for the ARMA experiment and $[-0.82, -0.1] \cup [0.1, 0.84]$ for the SARMA experiment.

	ARMA(1,1)			SARMA(1,1)(1,1)$_4$		
correct identification | n	200	500	1000	200	500	1000
C_1	67.2	80.9	85.6	54.6	67.9	73.1
C_1, C_2	33.1	58.4	70.3	24.7	41.8	54.9
C_1, C_2, C_3	11.1	28.5	44.3	9.5	23.4	34.8

in the non-Gaussian case. Rejection of the null hypothesis at the 5% level is indicated by a bold value for the RMSE of the first named method.

We find that when the underlying data come from a model with non-Gaussian copulas, the full procedure comprising the steps A, B and C leads to significantly smaller errors than are obtained by ending the procedure after step B. There is obviously an advantage to adding step C in terms of the accuracy with which the Kendall partial autocorrelations can be estimated.

Table 3: Root mean-squared errors (RMSEs) of estimates of first 3 terms of kpacf (τ_k). Results are based on 1000 model simulations. Bold figures indicate that Kolmogorov-Smirnov tests of the equality of the distributions of the errors give significant results at the 5% level; see the text in Section 4.5 for exact details of the comparisons that are carried out.

			ARMA(1,1)			SARMA(1,1)(1,1)$_4$		
D-vine type	procedure	parameter	200	500	1000	200	500	1000
Gaussian	A,B	τ_1	**0.0355**	0.0222	0.0157	0.0570	0.0347	0.0248
		τ_2	0.0382	**0.0232**	**0.0161**	0.0551	**0.0314**	**0.0224**
		τ_3	0.0366	0.0190	0.0128	0.0556	0.0342	0.0240
		all	0.0368	**0.0215**	**0.0149**	0.0559	**0.0335**	**0.0238**
Non-Gaussian	A,B	τ_1	0.0493	0.0367	0.0341	0.0657	0.0440	0.0338
		τ_2	0.0659	0.0625	0.0629	0.0657	0.0485	0.0447
		τ_3	0.0427	0.0302	0.0271	0.0634	0.0421	0.0337
		all	0.0535	0.0454	0.0442	0.0649	0.0450	0.0378
	A,B,C	τ_1	**0.0450**	**0.0283**	**0.0212**	0.0614	**0.0397**	**0.0289**
		τ_2	**0.0400**	**0.0242**	**0.0175**	**0.0535**	**0.0326**	**0.0247**
		τ_3	**0.0369**	**0.0206**	**0.0149**	**0.0566**	**0.0365**	**0.0266**
		all	**0.0408**	**0.0246**	**0.0180**	**0.0573**	**0.0364**	**0.0268**

When the underlying data come from a model with only Gaussian pair copulas, and steps A and B, are applied the resulting errors are smaller than those resulting from the application of the full procedure to data from a model with non-Gaussian dependencies. However, the differences are not always significant, suggesting that the presence of non-Gaussian dependencies can be addressed with our proposed procedure with a modest deterioration in the accuracy that would be expected in the Gaussian case.

4.6 Comparison with the fully parametric approach

To compare with the fully parametric approach, we consider only the ARMA(1,1) model, since results for the seasonal ARMA model are very similar, but run times are much longer. We recall that all the simulated ARMA(1,1) datasets have a Student t6 marginal distribution.

We consider the case where a parametric modeller uses a Student t distribution in step B and the case where the modeller uses a normal distribution; in both cases the modeller estimates the parameters of their choice

of marginal distribution from the data. The aim of this exercise is to understand whether there is an improvement in inference about the structure and parameters of the D-vine relative to the semiparametric approach when a modeller uses a correctly specified parametric model and whether there is any deterioration when an incorrectly specified parametric model is used. The idea that one would misspecify a Student t marginal distribution with a normal distribution is perhaps unrealistic in practice, particularly with samples of 200 or more data; other forms of marginal misspecification, such as overlooked skewness, might be more common. The experiment is included for its pedagogical value and complements the work of Kim et al. (2007) in a non-time-series context. Results for copula identification and parameter estimation are found in Tables 4 and 5.

In these tables the modeller using a semiparametric approach with non-Gaussian copulas is labelled SP, the parametric modeller using a correctly specified marginal distribution (Student t) is labelled P and the parametric modeller using a misspecified marginal distribution (normal) is labelled PM. It is evident from Table 4 that modeller P does a slightly better job than modeller SP at identifying copulas, while modeller PM does worse in all cases but one (identifying C_1 when $n = 200$). The performance in identifying C_2 and C_3 is much worse for the misspecified parametric modeller PM, suggesting that errors are compounded for higher order partial copulas C_k.

Table 4: Percentage of models with non-Gaussian dependence for which the first 3 copulas are correctly identified. Results are based on 1000 model simulations and the true marginal distribution is Student t with 6 degrees of freedom. SP, P, and PM refer to the semiparametric, parametric and misspecified parametric modellers respectively. The true copulas C_1, C_2 and C_3 are drawn randomly from the Frank, Clayton, Gumbel, Joe and Gaussian familes; random rotations of Clayton, Gumbel and Joe through 90, 180 and 270 degrees are allowed. The Kendall's tau values for these copulas are in the range $[-0.74, -0.1] \cup [0.1, 0.75]$.

	SP			P			PM		
correct identification | n	200	500	1000	200	500	1000	200	500	1000
C_1	67.2	80.9	85.6	74.8	83.8	87.1	69.1	77.5	79.8
C_1, C_2	33.1	58.4	70.3	39.5	59.7	73.7	32.3	45.8	50.3
C_1, C_2, C_3	11.1	28.5	44.3	13.2	29.5	46.7	10.8	18.7	24.0

The findings of Table 4 are also apparent in Table 5. The estimation errors for modeller P are slightly smaller than those for modeller SP, but quite a lot larger for modeller PM. These results suggest that the parametric approach brings advantages over the semiparametric approach when a good marginal model can be found. Similar to the findings of Kim et al. (2007) in the i.i.d. case, they underscore that marginal analysis needs to be done with care if the fully parametric method is used. In Section 5.5 we consider how these conclusions carry over to forecasting.

Table 5: Root mean-squared errors (RMSEs) of estimates of first 3 terms of kpacf (τ_k). Results are based on 1000 model simulations and the true marginal distribution is Student t with 6 degrees of freedom. SP, P, and PM refer to the semiparametric, parametric and misspecified parametric modellers respectively.

	SP			P			PM		
par.	200	500	1000	200	500	1000	200	500	1000
τ_1	0.0450	0.0283	0.0212	0.0428	0.0254	0.0180	0.0473	0.0328	0.0263
τ_2	0.0400	0.0242	0.0174	0.0361	0.0223	0.0167	0.0402	0.0281	0.0250
τ_3	0.0369	0.0206	0.0149	0.0346	0.0196	0.0144	0.0396	0.0245	0.0189
all	0.0408	0.0246	0.0180	0.0380	0.0225	0.0164	0.0415	0.0281	0.0234

5 Forecasting using stationary D-vines

5.1 Estimates of the predictive distribution

We now turn our attention to forecasting with stationary D-vine models, beginning with the formulas that are used to describe the predictive distribution of a stationary D-vine process $(X_t)_{t\in\mathbb{Z}}$. The predictive distribution

function of X_{t+1} conditional on the previous k values $\mathbf{X}_{[t-k+1:t]}$ will be denoted $F_{t+1|k}(x)$, suppressing the explicit dependence on the realized values $\mathbf{x}_{[t-k+1:t]}$ for notational convenience. The predictive distribution function and its density and quantile function are given by

$$F_{t+1|k}(x) = R_k\left(F_X(x); F_X(\mathbf{x}_{[t:t-k+1]})\right) \tag{7}$$

$$f_{t+1|k}(x) = r_k\left(F_X(x); F_X(\mathbf{x}_{[t:t-k+1]})\right) f_X(x) \tag{8}$$

$$F^{-1}_{t+1|k}(u) = F_X^{-1}\left(R_k^{-1}(u; F_X(\mathbf{x}_{[t:t-k+1]}))\right) \tag{9}$$

where F_X and f_X are the common marginal distribution function and density of the s-vine process and R_k and r_k are as in (2) and (5).

In the EDF method it is natural to simply use the (rescaled) empirical distribution function $\widehat{F}_X^{(n)}(x) = (1/(n+1))\sum_{i=1}^{n} I_{\{x_{t-i+1}\le x\}}$ as a plug-in estimator of $F_X(x)$ in (7) and (8) since this is what is used in stage B of the estimation procedure in Section 4.1. Combined with an estimate $\widehat{R}_k$ of the Rosenblatt forward function from the fittted copula process, this yields the estimator

$$\widehat{F}^{(n)}_{t+1|k}(x) = \widehat{R}_k\left(\widehat{F}_X^{(n)}(x); \widehat{F}_X^{(n)}(\mathbf{x}_{[t:t-k+1]})\right) \tag{10}$$

for the predictive distribution function (7). For an estimator of the predictive density $f_{t+1|k}$ in (8) the natural plug-in estimator of f_X is a kernel density estimator $\widehat{f}_X^{(n)}$. To estimate $F^{-1}_{t+1|k}$ in (9), a nonparametric estimate of the marginal quantile function $F_X^{-1}(\alpha)$ is required; an empirical quantile of the data $x_{t-n+1}, \dots, x_t$ can be computed using one of the definitions in Hyndman and Fan (1996). In the fully parametric method, we simply insert the parametric estimates $\widehat{F}_X$ and $\widehat{f}_X$ of the marginal distribution function and its density in formulas (7)–(9). This would also apply if we used nonparametric kernel density estimate instead of a parametric one.

5.2 Evaluating distributional forecasts

For $i \in \{1, \dots, m\}$ suppose we have a sequence of estimates $\widehat{F}_{t+i|k}$ of the one-step predictive distributions $F_{t+i|k}$ together with subsequent realised values x_{t+i} of the random variables X_{t+i}. One method of evaluating these distributional forecasts, following Diebold et al. (1998), is to calculate the probability-integral transform (or PIT values), $\{u_{t+1}, \dots, u_{t+m}\}$ given by $u_{t+i} = \widehat{F}_{t+i|k}(x_{t+i})$ for $i \in \{1, \dots, m\}$, and to test these for uniformity and independence.

In addition to testing PIT values, we also compare competing forecast models using the scoring approach of Gneiting and Ranjan (2011) based on the continuous ranked probability score (CRPS). This takes the equivalent forms

$$\mathrm{CRPS}\left(\widehat{F}_{t+i|k}, x_{t+i}\right) = \int_{-\infty}^{\infty} \mathrm{S}^B\left(\widehat{F}_{t+i|k}(y), I_{\{x_{t+i}\le y\}}\right) \mathrm{d}y = 2\int_0^1 \mathrm{S}^Q_\alpha\left(\widehat{F}^{-1}_{t+i|k}(\alpha), x_{t+i}\right) \mathrm{d}\alpha \tag{11}$$

where $\mathrm{S}^B(p,q) = (p-q)^2$ is the Brier score for a probabilistic forecast p of a binary outcome q and where $\mathrm{S}^Q_\alpha(y,x) = \left(I_{\{x\le y\}} - \alpha\right)(y-x)$ is a consistent scoring function for evaluating a forecast y of the α-quantile when the realized value is x. Gneiting and Ranjan (2011) suggest weighted versions of both integrals in (11) to emphasise different features of the forecast distribution (such as the centre or the tails)

We adopt the approach based on $\mathrm{S}^Q_\alpha(y,x)$, which seems appropriate for a copula-based methodology. That is, we consider scores which take the form $\int_0^1 \mathrm{S}^Q_\alpha\left(\widehat{F}^{-1}_{t+i|k}(\alpha), x_{t+i}\right) \mathrm{d}\nu(\alpha)$ where ν is a Lebesgue-Stieltjes measure which is designed to apply different weights to different quantiles. In particular we will consider a discrete measure ν and a score that we call the average weighted quantile score (AWQS) given by

$$\mathrm{AWQS}\left(\widehat{F}_{t+i|k}, x_{t+i}\right) = \frac{1}{J-1}\sum_{j=1}^{J-1} \mathrm{S}^Q_{\alpha_j}\left(\widehat{F}^{-1}_{t+i|k}(\alpha_j), x_{t+i}\right) \upsilon(\alpha_j), \qquad \alpha_j = \frac{j}{J}. \tag{12}$$

where $\upsilon(\alpha)$ is the weight function and the integer J determines the granularity of the discrete measure.

For tail weighting, Gneiting and Ranjan (2011) suggest the quadratic function $\upsilon(\alpha) = (2\alpha-1)^2$ but, in our view, this may not give enough weight to the more extreme quantiles that might be the focus of interest in some applications. We choose $\upsilon(\alpha)$ to be the reciprocal of what we will call a template function, that is a

function of the form $\alpha \mapsto \mathrm{E}\left[S_\alpha^Q(H^{-1}(\alpha), Z)\right]$ where Z is a random variable with distribution function H; the template function at α is the expected quantile score for an ideal forecaster who forecasts the α-quantile of a random variable Z using the correct quantile function H^{-1}. When interest focusses on forecasting the copula process itself, we could take the reciprocal of the uniform template function given by $\upsilon(\alpha)^{-1} = \mathrm{E}\left[S_\alpha^Q(\alpha, U)\right] = (1/2)\alpha(1-\alpha)$ where $U \sim \mathcal{U}(0,1)$. Otherwise our recommendation would be to take H to be an estimate of the unknown marginal distribution F_X. The estimate need not be particularly precise, since our simulations will show that results are not particularly sensitive to the exact form of H. In the fully parametric approach we could take the parametric estimate $H = \widehat{F}_X$ used in the inference functions for margins approach; in the semiparametric case we would estimate a parametric model that roughly fits the empirical data.

In the context of a one-step ahead backtesting experiment of length m we compare competing forecasting models $\widehat{F}_{t+i|k}^{(1)}$ and $\widehat{F}_{t+i|k}^{(2)}$ by computing average values $\overline{\mathrm{AWQS}}_m^{(1)}$ and $\overline{\mathrm{AWQS}}_m^{(2)}$ over all m forecasts; we favour the model giving the lower average value. The values can be compared formally using the test of equal forecast performance proposed by Diebold and Mariano (1995). To understand the differences in forecast performance at different quantile levels, we apply a graphical method proposed by Laio and Tamea (2007) in which the points

$$\left(\alpha_j, \quad \frac{\upsilon(\alpha_j)}{m}\sum_{i=1}^{m} \mathrm{S}_{\alpha_j}^Q\left(\widehat{F}_{t+i|k}^{(l)-1}(\alpha_j), x_{t+i}\right)\right), \quad \alpha_j = \frac{j}{J}, \quad j = 1, \ldots, J-1, \tag{13}$$

are plotted for each model $l = 1, 2$. We refer to this as a quantile score plot when we apply the uniform weight function $\upsilon(\alpha) = 1$ and a weighted quantile score plot otherwise.

5.3 Simulation study of semiparametric forecasting

In the simulation study we use the forecast evaluation methods of Section 5.2 to reveal differences between models based on copula processes with different degrees of misspecification. We base the study on the examples of Section 4.3, in which the true data generating mechanisms are processes with random non-Gaussian copula substitutions at the first 3 lags, and we consider two forecasters. The first semiparametric forecaster implements steps A and B of the method described in Section 4 while the second semiparametric forecaster implements steps A, B and C. The second forecaster, referred to as the non-Gaussian forecaster, will tend to use a better specified model than the first forecaster, referred to as the Gaussian forecaster. A comparison with fully parametric forecasting is given in Section 5.5.

In each repetition of our main simulation experiment we assume that the first n observations are used to estimate the Gaussian and non-Gaussian forecast models as in Section 4.3. The selected copulas of the non-Gaussian model and the parameters of both copula processes are subsequently held fixed while the marginal distribution and its quantiles are estimated used a rolling window of n observations and a series of m one-step distributional forecasts are made. In a second experiment (Section 5.4) we also consider re-estimating the parameters of the copula process at regular intervals. The number of conditioning variables k used in the forecasts is always the effective maximum lag K for the forecast model being used. In estimating the quantile function of the predictive distribution (9) to evaluate AWQS scores (12), we use the default definition of the empirical quantile in R, which is calculated according to method 7 of Hyndman and Fan (1996).

Figure 1 shows the quantile score plot and weighted quantile score plot (13) for a single simulation experiment where $n = 500$ and $m = 100$. In this case we have set $J = 20$ and applied the standard normal template function. The black and red lines shows the score curves for the non-Gaussian and Gaussian forecasters respectively; the green line shows the score curve for an ideal forecaster who has oracular knowledge of the true underlying process. Note how the weighted score plot blows up the differences in both tails for the Gaussian forecaster. The curves for the non-Gaussian forecaster are close to the curves for the ideal forecaster in this experiment, but the curves for the Gaussian forecaster are clearly worse. The Diebold-Mariano test of forecast equality yields a p-value of 0.0002 without weighting and a p-value of 0.0003 when weighting is based on the normal template function; in this case, the weighting makes no difference to the final inference (that the non-Gaussian forecaster does significantly better).

In the simulation study, using different values of n and m, we obtain the results in Table 6 based on 1000 replications. Note that, for example, the first line where $n = 200$ and $m = 40$ represents a situation where a model is fitted to 50 years of quarterly data and validated on a further 10 years of data, so this case is relevant to macroeconomic models based on quarterly data (see Section 6.1); other cases are more relevant to data collected at a higher frequency, or for a longer period. For consistency across simulations, we always choose the standard

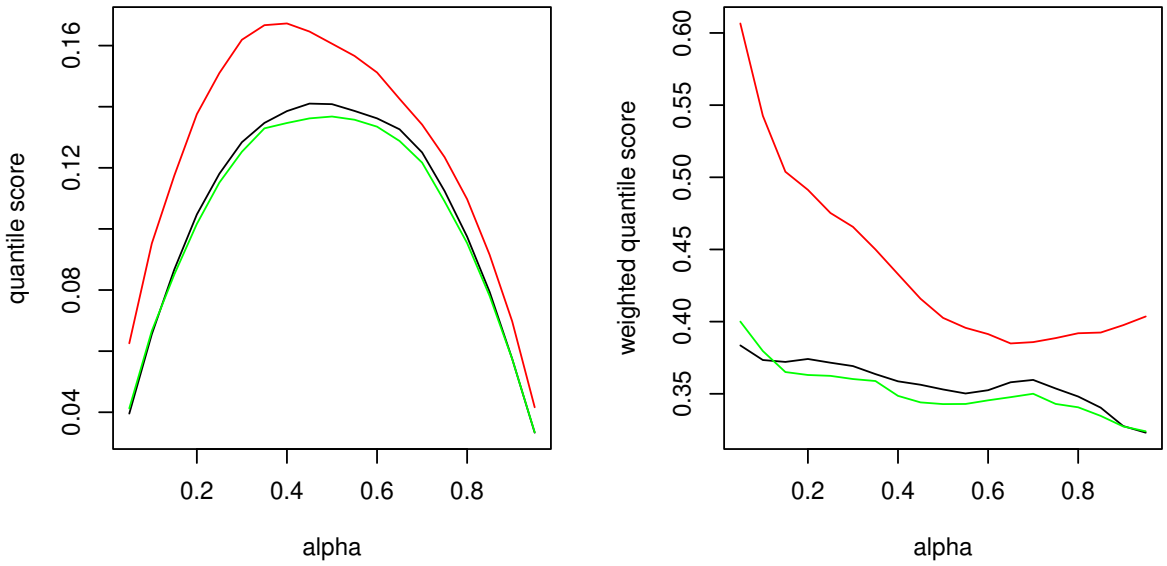

Figure 1: Quantile score plot and weighted quantile scoreplot for $m = 100$ forecasts based on an estimation window of length $n = 500$. Weighting is by the reciprocal of the normal template function and $J = 20$. Black line - non-Gaussian forecaster; red line - Gaussian forecaster; green line - ideal forecaster. Results are based on 500 simulations.

Table 6: Columns in which the ordering is labelled 'correct' show the percentage of experiments in which $\overline{\text{AWQS}}_m$ is smaller for the non-Gaussian forecaster than for the Gaussian forecaster. Columns labelled 'signif.' show the percentage of experiments in which this ordering is backed up by a significant Diebold-Mariano test result at the 5% level. The experiment is repeated using uniform weighting and using a weighting based on the reciprocal of the normal template function. All results are based on 1000 replications.

model		ARMA(1,1)				SARMA(1,1)(1,1)$_4$			
weight		none		normal template		none		normal template	
m	n | ordering	correct	signif.	correct	signif.	correct	signif.	correct	signif.
40	200	75.5	17.1	75.7	17.9	74.7	15.4	75.4	15.9
	500	79.2	21.1	80.8	22.7	77.7	17.6	79.0	18.5
	1000	81.3	21.9	82.2	24.0	78.9	17.9	80.0	18.7
100	200	80.5	30.6	81.5	31.6	77.5	25.6	77.7	26.7
	500	89.2	36.5	89.4	37.3	87.3	31.9	87.7	33.7
	1000	91.7	41.2	92.7	43.1	87.3	31.9	87.7	33.7

normal template function for weighting quantiles, since this is the true marginal distribution of the data and a reasonable forecaster would tend to estimate a marginal distribution that is close to standard normal.

Across all the experiments the non-Gaussian forecaster is found to give superior distributional forecasts, according to the $\overline{\text{AWQS}}_m$ score, between 74% and 93% of the time. The superiority of the non-Gaussian forecaster becomes more apparent as the number of forecasts m increases and the amount of data n used to calibrate the D-vine model increases. The percentage of occasions on which a formal Diebold-Mariano test returns a significant result in favour of the non-Gaussian forecaster also increases with n and m; for a model calibrated to

$n = 1000$ data evaluated using $m = 100$ forecasts it is close to 1/2. There is relatively little difference between the results for no weighting and for weighting based on a normal template function; the latter leads to the non-Gaussian forecaster being favoured slightly more often and to more significant test results, showing that the superiority of the non-Gaussian forecaster may often be manifested in better estimates of quantiles in the tails. Results are relatively similar for the ARMA(1,1) and SARMA(1,1)$(1,1)_4$ models, although the superiority of the non-Gaussian forecaster is slightly more apparent in the former case.

Table 7: Results of Kolmogorov-Smirnov (KS) tests of uniformity and Ljung-Box (LB) tests for serial correlation carried out on the PIT values produced by the Gaussian (G) and non-Gaussian (NG) forecasters when the underlying models contain non-Gaussian dependencies. Tabulated numbers show percentage of tests yielding significant results at the 5% level based on 1000 replications. The value of the lag used in the Ljung-Box test is 5.

model		ARMA(1,1)				SARMA(1,1)$(1,1)_4$			
test		K-S		L-B		K-S		L-B	
m	n | forecaster	G	NG	G	NG	G	NG	G	NG
40	200	10.4	7.6	14.8	12.1	8.4	6.1	12.4	9.8
	500	8.9	5.8	11.8	8.4	7.2	6.0	8.6	7.6
	1000	7.7	5.1	9.8	7.5	6.6	5.3	8.7	6.8
100	200	12.5	7.9	21.1	17.7	9.3	8.5	23.1	19.5
	500	12.8	6.1	12.0	9.1	9.2	5.3	13.5	10.2
	1000	10.8	4.5	10.9	7.1	9.9	6.1	10.6	7.3

In contrast, Table 7 shows that testing based on PIT values is not as effective at revealing the differences between the two forecasters that are clearly evident in the scoring analysis. The rejection rate for the null hypothesis of uniformity in the Kolmogorov-Smirnov test is close to the nominal test size of 5% for the non-Gaussian forecaster, particularly for larger values of n, but is only very slightly larger for the Gaussian forecaster. The rejection rate for the null hypothesis of no serial correlation in the Ljung-Box test tends to be larger than 5% for the non-Gaussian forecaster (particularly when the model is calibrated to $n = 200$ values). The rejection rate is systematically higher for the Gaussian forecaster but, again, the differences between the forecasters are not particularly pronounced. Testing PIT values for uniformity and independence does not appear to be an effective way of differentiating between forecasters who differ only in their specification of the serial dependence model.

We draw the following conclusions from this section. First, we infer that the non-Gaussian forecaster who uses the methodology we propose does tend to obtain superior forecasts, according to the scoring techniques for comparing forecasts developed by Gneiting and Ranjan (2011). If non-Gaussian serial dependencies are present, then attempting to model them, even allowing for some inevitable misspecification of model order and copula choice, brings forecasting advantages. As might be expected, these advantages become more apparent as more data are used to develop models and as more forecasts are issued. Second, we conclude that scoring methods are a better approach than PIT-value tests for evaluating these advantages. This may be particularly true when forecasters differ with respect to their approach to modelling serial dependence structure, but not their approach to marginal modelling.

5.4 Re-estimating the copula process

In this section we investigate whether any advantage can be gained by re-estimating the copula parameters at regular intervals. In the context of our simulation study it would be prohibitively slow to re-determine the structure of the D-vine and re-estimate its parameters on a rolling basis, since the algorithm involves a greedy search over all possible copulas at each lag. Instead, we fix the structure of the D-vine using the original training dataset, but re-estimate the copula parameters after every 5 forecasts. We restrict attention to the ARMA(1,1) experiment since conclusions are identical for the seasonal ARMA experiment.

In Table 8 we compare the scores for the non-Gaussian forecaster who re-estimates the parameters periodically with the scores for the non-Gaussian forecaster who holds the parameters of the copula process fixed throughout the forecasting exercise.

Table 8: Comparison of non-Gaussian forecaster who re-estimates copula process with non-Gaussian forecaster who folds parameters fixed throughout forecasting exercise. The columns labelled 'refit vs. no refit' show the percentage of experiments in which $\overline{\text{AWQS}}_m$ is smaller for the forecaster who re-estimates copula parameters. Columns labelled 'signif.' show the percentage of experiments in which this ordering is backed up by a significant Diebold-Mariano test result at the 5% level. The experiment is repeated using uniform weighting and using a weighting based on the reciprocal of the normal template function. All results are based on 1000 replications.

weight		none		normal template	
m	n	refit vs. no refit	signif.	refit vs. no refit	signif.
40	200	47.9	1.2	48.7	2.1
	500	48.3	1.7	48.7	2.1
	1000	50.7	2.2	49.4	2.4
100	200	49.0	2.8	49.3	2.3
	500	46.8	2.0	46.9	3.0
	1000	47.3	2.8	46.5	2.9

Re-estimating the copula parameters actually appears to slightly worsen the forecasting scores in most cases, although differences are seldom significant. This may be due to the introduction of some additional estimation noise. It should be noted, however, that our simulation study takes place in an ideal, perfectly stationary environment. Re-estimation of the copula parameters might be reasonable in practice if we believed that the stationarity assumption was tenuous and changes of regime were taking place periodically.

5.5 Comparison with fully parametric forecasters

This section is a continuation of the comparison between a semiparametric and fully parametric approach in Section 4.6. It might be assumed that it is particularly in forecasting distributions that the fully parametric approach may show advantages since empirical distribution estimates are subject to high error in the tails and we are comparing forecasts at quantiles ranging from 5% to 95% in increments of 5%.

Again we consider only the ARMA(1,1) case and we compare the semiparametric forecaster SP, the fully parametric forecaster P who uses a correctly specified marginal distribution and the fully parametric forecaster PM who uses a misspecified marginal model. In view of the poor discriminatory power of forecast comparisons based on PIT values established in Section 5.3, we restrict attention to comparing forecasts using the scoring approach.

The left panel of Table 9 shows the comparison between the fully parametric forecaster P and the semiparametric forecaster SP. The former obtains lower scores than the latter and the advantages of the parametric approach increase as the training sample size n decreases and the number of forecasts m increases. For $n = 200$ and $m = 100$ the parametric forecaster obtains lower scores around 80% of the time and the difference is significant around 20% of the time. In this experiment we carry out a weighted comparison using the inverse of the template function of the Student t distribution with 6 degrees of freedom. However, this does not make much difference to the results.

The right panel of Table 9 shows the comparison between the semiparametric forecaster SP and the parametric forecaster PM using a misspecified normal distribution instead of a Student t distribution. The first point to notice is that there is seldom a significant difference between the performance of the forecasters according to the Diebold-Mariano test. When the training sample size n is 1000 the semiparametric method is preferred more often; when $n = 200$ the parametric method is preferred more often (despite misspecification); when $n = 500$ the two approaches are preferred approximately equally often.

These results suggest that the fully parametric approach should be preferred for forecasting when (i) there is high confidence in the quality of the marginal model or (ii) there is less confidence in the quality of the marginal model but very few data n to use in the training set. The semiparametric approach would mainly come into consideration for forecasting when it is difficult to find a good parametric marginal distribution and when the size of the training dataset n is relatively large (over 500).

Table 9: Columns labelled 'P < SP' show the percentage of experiments in which the parametric forecaster obtains a lower score $\overline{\text{AWQS}}_m$ than the semiparametric forecaster; columns labelled 'SP < PM' show a similar comparison for the semiparametric versus the parametric forecaster who uses a misspecified marginal distribution. Columns labelled 'signif.' show the percentage of experiments in which these orderings are backed up by a significant Diebold-Mariano test result at the 5% level. The experiment is repeated using uniform weighting and using a weighting based on the reciprocal of the Student t template function with 6 degrees of freedom. All results are based on 1000 replications.

Comparison		P vs. SP				SP vs. PM			
weight		none		Student template		none		Student template	
m	n | ordering	P < SP	signif.	P < SP	signif.	SP < PM	signif.	SP < PM	signif.
40	200	71.0	8.6	71.2	8.7	42.0	2.4	40.0	2.5
	500	66.2	5.4	66.0	5.8	51.4	3.4	51.3	3.4
	1000	61.3	5.0	60.7	5.2	59.5	4.9	60.6	5.5
100	200	78.5	21.5	81.0	21.9	35.6	3.6	34.1	2.9
	500	75.7	10.5	75.9	11.1	49.3	4.8	50.2	4.8
	1000	69.2	7.1	68.9	7.8	63.5	5.2	64.1	6.2

6 Empirical examples

We present two examples with real data. In Section 6.1 we illustrate the use of a seasonal ARMA dependence structure to model changes in the U.S. Consumer Prices Index (CPI). In Section 6.2 we apply non-seasonal ARMA models to a much larger dataset of absolute log-returns of the Nasdaq Composite index.

6.1 Forecasting the rate of inflation

We model changes in the U.S. Consumer Prices Index (CPI), which is a weighted average of consumer prices in a representative basket of goods and services. There is a vast empirical literature on modelling and forecasting CPI and inflation. Our intention is not to benchmark the stationary D-vine methodology against a wide range of competing forecast approaches but, more modestly, to show how non-Gaussian forecasting using stationary D-vines improves on forecasting using conventional Gaussian seasonal ARIMA models.

ARIMA models are highly stylized models which it would be difficult to view as the "true" data-generating processes of macroeconomic data. Integrated models may drift to arbitrarily large positive or negative values and this is unrealistic behaviour for variables like CPI which attract strong government intervention. Despite such drawbacks, ARIMA models are quite commonly applied to short-term inflation forecasting. We note that there are many alternative methods ranging from judgmental survey-based forecasts to Phillips-curve-based forecasts, and forecasts based on more complex models such as Markov switching models. An extensive review of inflation forecasting can be found in Faust and Wright (2013) including a comparison of the performances of a variety of different methods for the U.S., the U.K., Canada, Germany and Japan. More recently, machine learning approaches, such as lasso regression, random forests, and neural networks, have been applied to inflation forecasting; see, for example, Barkan et al. (2023) and references therein. Almosova and Andresen (2019) use recurrent neural networks to model non-linearities in inflation data. Non-linear effects are also reported by Faust and Wright (2013) and our use of stationary D-vines can be viewed as an alternative approach to addressing this phenomenon.

The raw data for the analysis in this section are taken from the OECD website and are the quarterly total consumer price index (CPI) values from the final quarter of 1959 to the end of 2022. Following the approach of Faust and Wright (2013) in the Handbook of Economic Forecasting, we compute quarterly values for the annualized rate of inflation and express these as a percentage. In other words, the data we model are observations of $Y_t = 100 \times 4 \times \ln(\text{CPI}_t/\text{CPI}_{t-1})$, where CPI_t denotes the CPI in quarter t, as shown in Figure 2.

We form a training dataset consisting of observations of Y_t up to and including the final quarter of 2011. In stage A of the procedure in Section 4.1 the automatic ARIMA modelling procedure selects a seasonal ARIMA model of order $(1,1,2)(1,0,0)_4$ for the training data implying that a single ordinary difference is required to obtain data that admit a stationary model. Thus we model the first difference of the rate of inflation X_t =

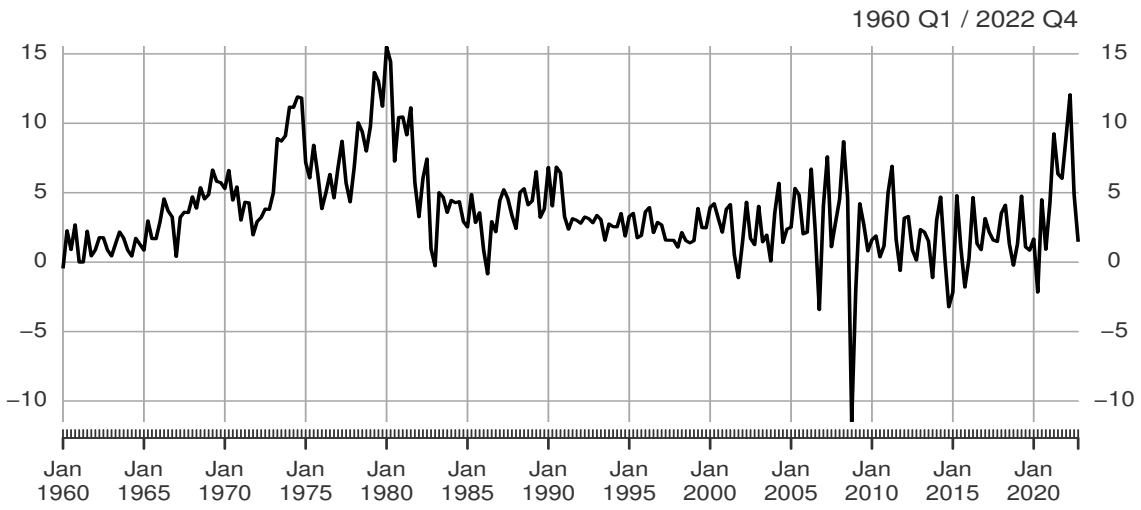

Figure 2: Quarterly data on the annual rate of inflation in the US from 1960 - 2022. The rate is calculated as the log-return of CPI and expressed as an annualized percentage.

$Y_t - Y_{t-1}$ in our stationary D-vine models; this yields $n = 207$ observations in the training set.

In stage B we first use the semiparametric model with Gaussian pair copulas to fit a SARMA model of order $(1,2)(1,0)_4$ to the training data. This model yields the parameter estimates and AIC and BIC values in the first line of Table 10; standard errors for the parameter estimates are given in the second line. We then apply the fully parametric approach to the same data. After trying the normal, Student t, skewed Student t, hyperbolic, normal-inverse Gaussian, generalized hyperbolic and Johnson's S_U distributions as marginal models, we conclude that the best of these is the Student t distribution with $\nu = 3.39$ degrees of freedom and with scale and location parameters $\mu = 0.069$ and $\sigma = 1.67$. In combination with Gaussian pair copulas this yields the parameter estimates for the kpacf and the AIC and BIC values in the fifth line of Table 10 as well as the standard errors in the sixth line. Note that, in both cases the AIC and BIC values relate to the model that is fitted to the pseudo-copula data, i.e. the model that is fitted to the data that are obtained after transformation of the training data by either the EDF or the parametric marginal distribution. Thus the values for the semiparametric and parametric approaches are not strictly comparable, since the models are fitted to slightly different data.

In stage C we replace Gaussian copulas with non-Gaussian copulas in both the semiparametric and parametric approaches. We replace copulas using the AIC criterion and stop at the point where the AIC is reduced by less than 0.5 on 3 consecutive occasions; it would be possible to continue making replacements but there are diminishing returns in doing so. In the non-parametric approach 14 of the first 15 copulas are replaced while in the parametric approach the first 10 copulas are replaced. Parameter estimates and AIC and BIC values in the Non-Gaussian copula case are found in the third and seventh lines of Table 10; the corresponding standard errors are in the fourth and eighth lines. A summary of the selected copulas and their Kendall's tau values is given in Table 11. In both approaches it may be noted that the substitutions at lags 1 and 2 are t copulas, which introduce 2 additional parameters into the model, relative to the Gaussian model. The in-sample fits, as measured by the AIC and BIC values, are greatly improved by the incorporation of non-Gaussian copulas. In both approaches, the effective maximum lag of the model with non-Gaussian copulas is $K = 39$ at tolerance $\epsilon = 0.001$.

In Table 11 there is a remarkable degree of correspondence between the copulas selected at the first 10 lags and the values of Kendall's tau in both approaches, the only difference in the selected copulas being at lag $k = 8$. Figure 3 shows the Kendall partial autocorrelation function (kpacf) for the semiparametric approach in red and for the parametric approach in blue; the latter almost entirely covers the former. Both exhibit negative partial dependence at the first four lags and then a decaying pattern of alternating positive and negative partial dependence thereafter. The black bars are empirical estimates of the Kendall's tau values derived from the data using a recursive method referred to as generalized lagging and explained in Bladt and McNeil (2022b).

Table 10: Parameter estimates and AIC and BIC values for ARMA $(1,2)(1,0)_4$ models based on Gaussian (G) and non-Gaussian (NG) copula sequences fitted to training data in both the semiparametric (SP) and fully parametric (P) approaches.

Approach	Copulas	df	ϕ_1	θ_1	θ_2	Φ_1	ν_1	ν_2	AIC	BIC
SP	G	4	-0.446	0.139	-0.666	0.301			-77.2	-63.8
	(s.e.)		0.094	0.076	0.052	0.072				
	NG	6	-0.497	0.154	-0.671	0.237	2.51	4.53	-106.9	-86.9
	(s.e.)		0.013	0.013	0.013	0.013	0.615	1.69		
P	G	4	-0.436	0.135	-0.672	0.279			-79.3	-66.0
	(s.e.)		0.093	0.076	0.053	0.071				
	NG	6	-0.483	0.146	-0.678	0.244	3.10	7.41	-104.6	-84.6
	(s.e.)		0.013	0.013	0.012	0.013	0.846	4.24		

Table 11: Summary of selected copulas and Kendall's tau values in both the semiparametric (SP-NG) and fully parametric (P-NG) approaches. The numbers 90, 180 and 270 appended to the names refer to rotations of the basic copulas.

k	SP-NG copulas C_k	τ_k	P-NG copulas C_k	τ_k
1	t ($\nu = 2.51$)	-0.12	t ($\nu = 3.10$)	-0.12
2	t ($\nu = 4.53$)	-0.34	t ($\nu = 7.41$)	-0.35
3	Joe90	-0.09	Joe90	-0.09
4	Clayton90	-0.07	Clayton90	-0.07
5	Gumbel180	0.03	Gumbel180	0.03
6	Gumbel90	-0.05	Gumbel90	-0.05
7	Joe	0.03	Joe	0.03
8	Gauss	-0.04	Clayton	-0.04
9	Joe180	0.02	Joe180	0.02
10	Clayton90	-0.03	Clayton90	-0.03
11	Frank	0.02	Gauss	0.02
12	Clayton90	-0.02	Gauss	-0.02
13	Frank	0.02	Gauss	0.02
14	Clayton270	-0.02	Gauss	-0.02
15	Clayton180	0.01	Gauss	0.01

The presence of t copulas with negative Kendall correlations at the first two lags in the non-Gaussian model is interesting and suggests that this model captures tail dependence in large changes in inflation following large changes with opposite sign; this could be a consequence of corrective measures taken by the Federal Reserve to target an inflation level of 2%.

We use $m = 44$ out-of-sample forecasts to evaluate the relative forecast performance of five models. In addition to the four models in Table 10, we also consider a fully parametric model with Gaussian margins and Gaussian copulas as obtained from standard statistical software. We use the same general forecasting methodology applied in Section 5.3, where copulas are not re-estimated, and we compare models using the scoring methodology.

All of the models in Table 10 give vastly superior forecasts to the fully Gaussian model from standard software. Whether we use non-parametric or Student t margins, the Diebold-Mariano test gives strongly significant results, showing the advantages of a D-vine copula-based methodology where margins and dependence structure can be modelled separately. Ranking the four forecasters in Table 10 by score, the ordering from best to worst is P-NG, SP-NG, P-G, SP-G, both for AQS scores and for AWQS scores based on the Student t template function with 3.4 degrees of freedom (as estimated for the margin). The quantile score and weighted quantile score plots for the four forecasters are shown in Figure 4. In a Diebold-Mariano test of the best method (P-NG) against the second best method (SP-NG) the result based on the AWQS is marginally significant ($p = 0.06$).

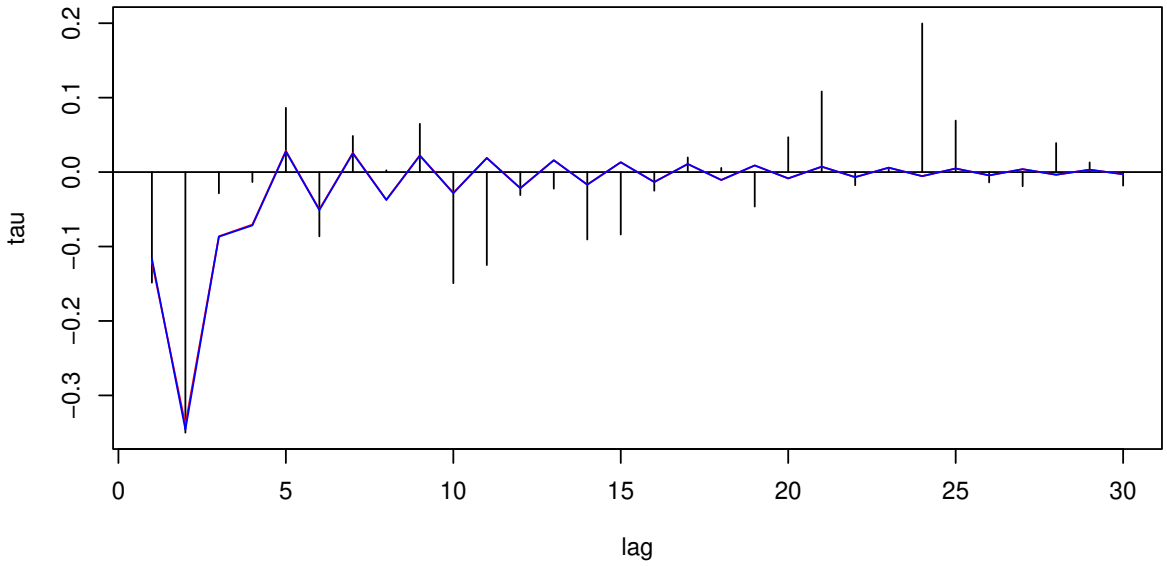

Figure 3: Kendall partial autocorrelation function (kpacf) for the non-Gaussian parametric approach in blue and for the non-Gaussian semiparametric approach in red; the latter is almost entirely covered by the former.

This example thus illustrates the advantages of the fully parametric approach when the size of the data sample is relatively modest.

To complete the analysis and indicate a possible use for the methodology, we show in Figure 5 our forecasts of the distribution of Y_t for each quarter from the start of 2012 based on the best model. The distributional forecasts are summarised by blue lines at the estimated α_j-quantiles, where $\alpha_j = j/20$ for $j = 1, \dots, 19$. The actual realized value of Y_t is superimposed as a black line. This picture underscores the idea that the methodology we present can be viewed as a form of quantile autoregression which provides properly ordered estimates for different quantiles of the forecast distribution.

6.2 Absolute returns of Nasdaq Composite index

We give a second example with a larger dataset where the difference in forecast performance between D-vines using Gaussian and non-Gaussian copulas is strongly significant. The data comprise the absolute values of the log-returns of the Nasdaq Composite equity index for the 5-year period 2016–2020. The models in this paper are appropriate for modelling time series in which the dependencies between lagged variables are monotonic, which is the case for absolute or squared log-returns but not the returns themselves; to model the latter we require more bespoke models constructed from cross-shaped copulas (Loaiza-Maya et al., 2018). The absolute values are of independent interest as measures and predictors of market volatility (Forsberg and Ghysels, 2007; Giles, 2008). Moreover, they show strongly persistent serial dependence which is better modelled by ARMA rather than AR processes. It may be noted that the popular GARCH(p,q) model (Bollerslev, 1986) for asset returns can be thought of as an ARMA(p,q) model for the squared returns; see, for example, McNeil et al. (2015).

The full time series of absolute log-returns is shown in Figure 6. We use the first four years (1004 values) as training data and the final year of 2020 as testing data for one-step forecasts (253 values). The final year was the year in which the Covid pandemic led to lockdowns and market turbulence throughout the world and it is of interest to see whether time series models with non-Gaussian dependence structures can capture this volatility better than models with Gaussian dependence structures.

The automatic ARIMA selection procedure applied to the probit-transformed pseudo-copula observations derived from the training data selects an ARMA(1,1) model. For a parametric marginal distribution we tried

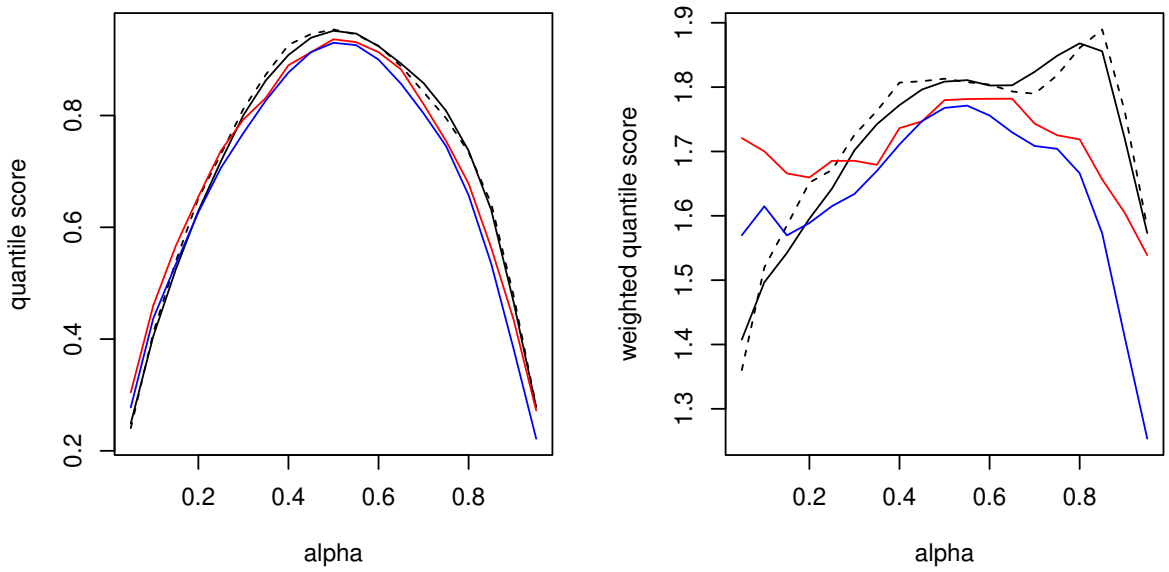

Figure 4: Quantile score plot and weighted quantile scoreplot for $m = 44$ forecasts of the first difference of the rate of inflation based on an estimation window of length $n = 207$. Weighting is by the reciprocal of the Student template function with 3.4 degrees of freedom and $J = 20$. The coding of the lines is: P-G (black), SP-G (dashed black), SP-NG (red) and P-NG (blue).

gamma, Weibull, inverse gamma and inverse Gaussian. The best of these according to AIC and BIC was the gamma distribution with shape parameter 1.02 and scale parameter 0.69, which provides an acceptable QQ-plot as shown in Figure 7, although there is a slight underestimation of the tail of the data. This illustrates the difficulty of finding a perfect parametric fit to real data.

As in the previous example, we consider semiparametric (SP) and fully parametric (P) forecasters using models with only Gaussian (G) or mixed non-Gaussian and Gaussian copulas (NG). Parameter estimates for the copula processes and AIC/BIC values are found in Table 12 while the substituted copulas are listed in Table 13. The copula replacement algorithm was stopped when the AIC could not be lowered by more than 0.2 in 3 successive steps; this resulted in non-Gaussian substitutions at 5 of the first 6 lags in both the semiparametric and fully parametric approaches. The effective maximum lag of both of the final models is $K = 29$ at a tolerance of $\epsilon = 0.001$. There are some differences in the substitutions for the two approaches but copulas with upper tail dependence (Clayton180, Gumbel, Joe, t) are selected at lags 1, 2, 3, 5 and 6 in both approaches.

The forecasting performance for the year 2020 of the models with non-Gaussian copula substitutions is greatly superior to the models with only Gaussian copulas, as illustrated in Figure 8. Ranking the forecasters from best to worst gives the ordering P-NG, SP-NG, P-G, SP-G for both AQS and AWQS scores. Considering the two middle methods, the Diebold-Mariano test shows that SP-NG gives significantly better forecasts than P-G (p-values 0.018 and 0.022 for AQS and AWQS respectively); the comparison between P-NG and SP-NG gives a non-significant result in both cases. The picture clearly shows that the advantage of the non-Gaussian dependence models stems from improved forecasts in the right tail of the forecast distribution and can be attributed to the asymmetry and upper tail dependence of the selected copulas.

7 Conclusion

The methodology described in this paper can offer improved forecasting for time series that are routinely modelled by ARMA or ARIMA processes and their seasonal extensions, particularly where non-Gaussian and non-linear behaviour may be present.

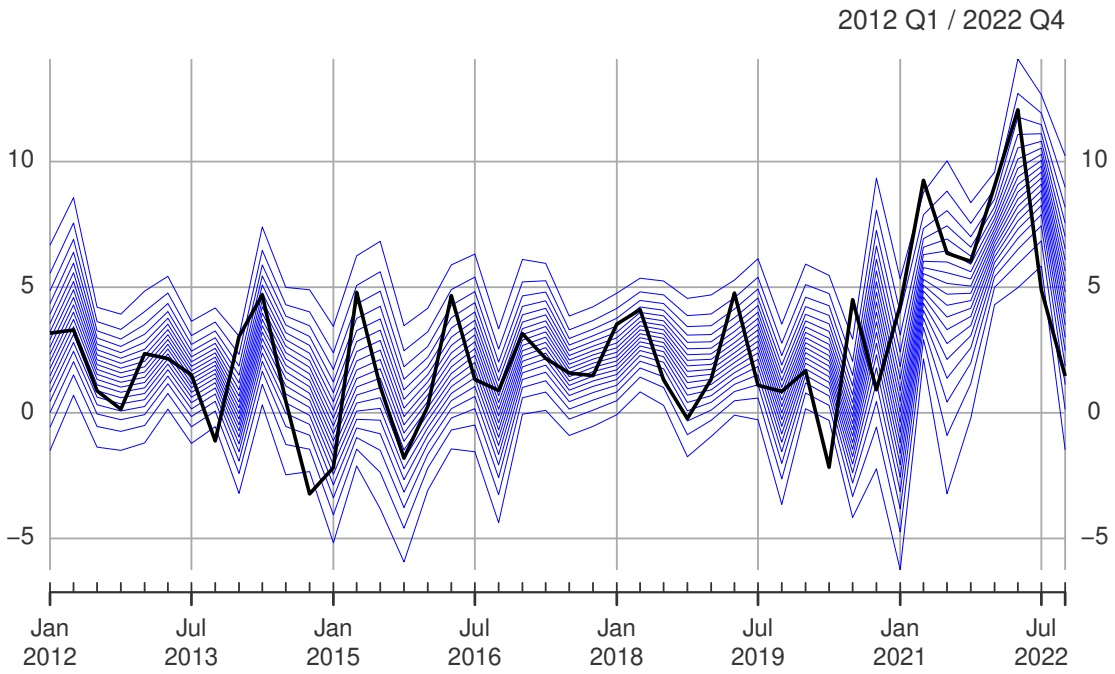

Figure 5: Distributional forecasts of the annual rate of inflation in the US from 2012–2022 based on the fully parametric forecasting model with non-Gaussian copulas (P-NG). The blue lines show estimated quantiles of the forecast distribution at the levels $\alpha_j = j/20$ for $j = 1, \dots, 19$. The black line shows the realized values.

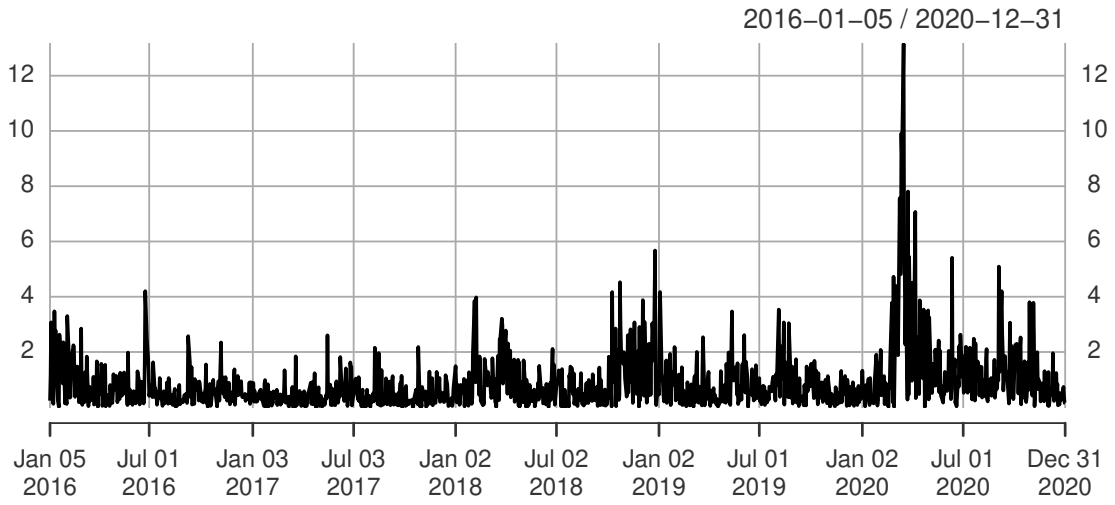

Figure 6: Absolute daily log-returns ($\times$ 100) of the Nasdaq Composite index in the period 2016–2020.

In contrast to other approaches that are designed to accomodate non-Gaussian data in the classical linear paradigm, such as Box-Cox transformations of the data or the use of non-Gaussian innovations, our approach allows a fully flexible nonparametric or parametric estimate of the marginal distribution as well as a detailed modelling of serial dependence using a parametric stationary D-vine that may contain non-Gaussian pair copulas. The methodology is relatively straightforward to implement and tools for estimation and model validation are available in the R package **tscopula** (McNeil and Bladt, 2023) which takes advantage of C++code for vine copulas accessed via the **rvinecopulib** package of Nagler and Vatter (2025)

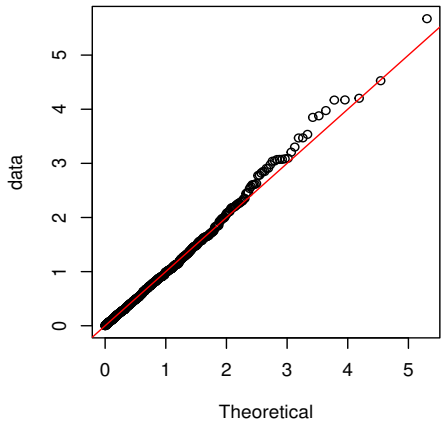

Figure 7: QQ-plot of training data against gamma distribution with shape and scale parameters 1.02 and 0.69.

Table 12: Parameter estimates and AIC and BIC values for ARMA $(1, 1)$ models based on Gaussian (G) and non-Gaussian (NG) copula sequences fitted to training data for Nasdaq absolute log-return series in both the semiparametric (SP) and fully parametric (P) approaches.

Approach	Copulas	df	ϕ_1	θ_1	ν	AIC	BIC
SP	G	2	0.962	-0.871		-97.9	-88.1
	(s.e.)		0.015	0.029			
	NG	2	0.955	-0.854		-140.1	-130.3
	(s.e.)		0.005	0.006			
P	G	2	0.961	-0.868		-102.8	-93.0
	(s.e.)		0.015	0.029			
	NG	3	0.951	-0.858	13.0	-139.8	-125.1
	(s.e.)		0.005	0.007	6.5		

In general, for forecasting distributions at a range of quantiles that include tail values such as 5% or 95%, the fully parametric approach to constructing D-vines should be preferred and this is particularly important when training samples have less than $n = 500$ observations. In situations where samples are large and where it is difficult to find a good parametric model, the semiparametric approach is a viable alternative. A compromise between the two approaches might be to use extreme value theory (EVT) to construct semi-parametric estimates of the marginal distribution with parametric tails, for example by using the generalized Pareto (GPD) tail approximation (McNeil and Frey, 2000).

Some open theoretical questions remain for infinite-order stationary D-vine models, particularly conditions on the copula sequence that guarantee ergodicity. However, these issues can be circumvented in practical applications by truncating copula sequences to obtain Markov processes whose behaviour is better understood.

The methodology we propose could be extended to financial price log-return series in which sign changes are considered; the analysis of absolute returns in Section 6.2 does not consider important features of the raw return data, such as the fact that negative returns may contribute more to volatility than positive returns, a phenomenon known as the leverage effect. Modelling the raw returns would require different copulas to the ones we have used in this paper; in particular they would need to be broadly cross-shaped to model non-monotonic serial dependence (Loaiza-Maya et al., 2018). Models of this kind, using parametric copulas constructed from so-called v-transforms, are proposed in McNeil (2021) and Bladt and McNeil (2022b). Cross-shaped copula behaviour can also be modelled non-parametrically using the method implemented in the **rvinecopulib** R

Table 13: Summary of selected copulas and Kendall's tau values in both the semiparametric (SP-NG) and fully parametric (P-NG) approaches applied to the Nasdaq absolute log-return data; clayton180 refers to the Clayton survival copula.

k	SP-NG copulas C_k	τ_k	P-NG copulas C_k	τ_k
1	Clayton180	0.12	Clayton180	0.11
2	Gumbel	0.10	Clayton180	0.09
3	Gauss	0.08	Gauss	0.07
4	Gumbel	0.06	Gumbel	0.06
5	Joe	0.05	Clayton180	0.05
6	Clayton	0.05	t ($\nu = 13.0$)	0.04

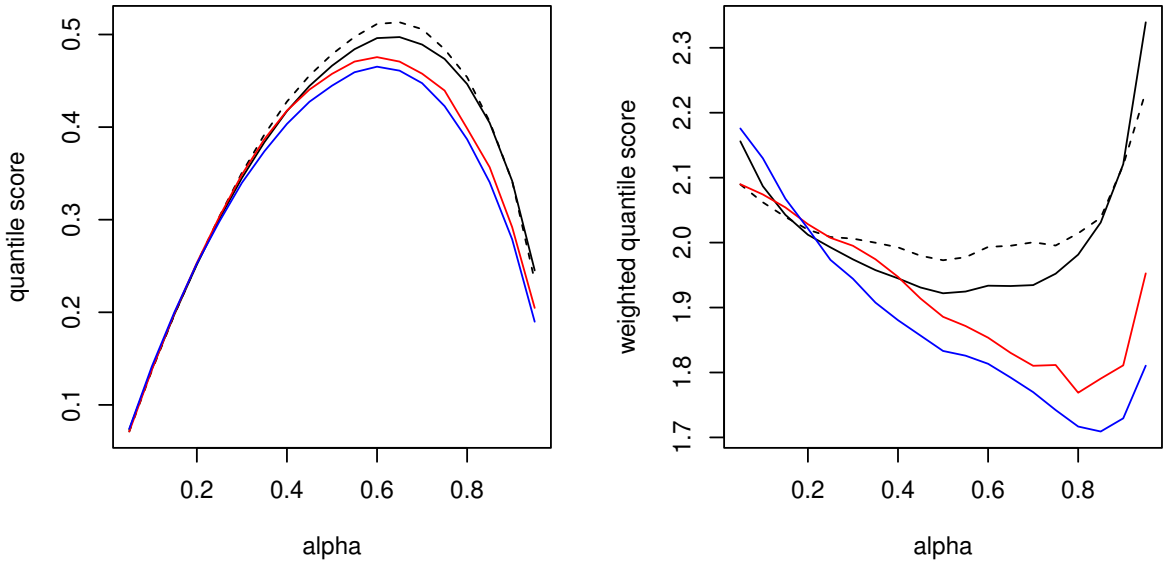

Figure 8: Quantile score plot and weighted quantile scoreplot for forecasts of Nasdaq absolute log-returns. Weighting is by the reciprocal of the template function of the fitted gamma marginal distribution and $J = 20$. The coding of the lines is: P-G (black), SP-G (dashed black), SP-NG (red) and P-NG (blue).

package (Nagler and Vatter, 2025).

The methodology could also be extended to multivariate time series. One possibility would be to extend Gaussian vector autoregressive (VAR) models by allowing non-Gaussian pair copulas for both serial and cross-sectional dependencies in the s-vine framework proposed by Nagler et al. (2022). Another possibility would be to embed our univariate stationary D-vine approach in the copula-linked univariate D-vine approach of Zhao et al. (2022). In both cases it may make sense to switch from nonparametric models to parametric models for marginal distributions to facilitate inference about multivariate phenomena and avoid the curse of dimensionality.

Data sharing

The analyses in this paper were carried out using version 0.4.7 of the R package **tscopula** which is available at `github.com/ajmcneil/tscopula`. The simulation studies, datasets and empirical examples are documented by code in the repository `github.com/ajmcneil/papers`.

Acknowledgements

The authors are grateful for insightful comments and suggestions from two anonymous referees which have been of great help in improving the paper.

ORCID

Martin Bladt: https://orcid.org/0000-0002-3121-3936
Alexandra Dias: https://orcid.org/0000-0003-0210-552X
Alexander J. McNeil: https://orcid.org/0000-0002-6137-2890

A Pacf of Gaussian processes

Suppose that $(X_t)_{t\in\mathbb{Z}}$ follows a Gaussian ARMA process with autocorrelation function (acf) $(\rho_k)_{k\in\mathbb{N}}$. The acf can be calculated from the AR and MA coefficients of the process using standard methods available in most statistical software (Brockwell and Davis, 1991, Section 3.3).

For $k \in \mathbb{N}$ let $\rho_k = (\rho_1, \dots, \rho_k)^\top$ and let P_k denote the correlation matrix of $(X_1, \dots, X_k)$. Clearly $P_1 = 1$ and, for $k > 1$, P_k is a symmetric Toeplitz matrix whose diagonals are filled by the first $k-1$ elements of ρ_k; moreover, P_k is non-singular for all k (Brockwell and Davis, 1991, Proposition 4). To calculate the pacf $(\alpha_k)_{k\in\mathbb{N}}$ from the acf $(\rho_k)_{k\in\mathbb{N}}$ we set $\alpha_1 = \rho_1$ and, for $k > 1$,

$$\alpha_k = \frac{\rho_k - \rho_{k-1}^\top P_{k-1}^{-1} \overline{\rho}_{k-1}}{1 - \rho_{k-1}^\top P_{k-1}^{-1} \rho_{k-1}};$$

see Joe (Joe, 2006) or the Durbin-Levinson Algorithm (Brockwell and Davis, 1991, Proposition 5.2.1).

References

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

Almosova, A. and Andresen, N. (2019). *Nonlinear inflation forecasting with recurrent neural networks*. Tech. rep., European Central Bank (ECB).

Barkan, O., Benchimol, J., Caspi, I., Cohen, E., Hammer, A., and Koenigstein, N. (2023). Forecasting CPI inflation components with hierarchical recurrent neural networks. *International Journal of Forecasting*, 39(3):1145–1162.

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

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

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

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

Benjamin, M. A., Rigby, R., and Stasinopoulos, D. (2003). Generalized autoregressive moving average models. *Journal of the American Statistical Association*, 98(461):214–223.

Bladt, M. and McNeil, A. (2022a). Time series with infinite-order partial copula dependence. *Dependence Modeling*, 10:87–107.

Bladt, M. and McNeil, A. J. (2022b). Time series copula models using d-vines and v-transforms. *Econometrics and Statistics*, 24:27–48.

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

Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods*. 2nd ed. Springer, New York.

Buch-Kromann, T., Nielson, J. P., Guillen, M., and Bolancé, C. (2005). Kernel density estimation for heavy-tailed distributions using the Champernowne transformation. *Statistics*, 39:503–516.

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

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

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

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

Diebold, F. and Mariano, R. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13:253–265.

Faust, J. and Wright, J. H. (2013). Forecasting inflation. In G. Elliot, C. Granger, and A. Timmermann, editors, *Handbook of economic forecasting*, vol. 2A. Elsevier, Amsterdam, North Holland.

Forsberg, L. and Ghysels, E. (2007). Why do absolute returns predict volatility so well? *Journal of Financial Econbometrics*, 5(1):31–67.

Genest, C., Ghoudi, K., and Rivest, L. (1995). A semi-parametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, 82:543–552.

Giles, D. (2008). Some properties of absolute returns as a proxy for volatility. *Applied Financial Economics Letters*, 4(5):347–350.

Gneiting, T. and Ranjan, R. (2011). Comparing density forecasts using threshold- and quantile-weighted scoring rules. *Journal of Business & Economic Statistics*, 29(3):411–422.

Hyndman, R. and Khandakar, Y. (2008). Automatic time series forecasting: the forecast package for R. *Journal of Statistical Software*, 26(3):1–22.

Hyndman, R. J. and Fan, Y. (1996). Sample quantiles in statistical packages. *American Statistician*, 50:361–365.

Joe, H. (1996). Families of m-variate distributions with given margins and $m(m-1)/2$ bivariate dependence parameters. In L. Rüschendorf, B. Schweizer, and M. D. Taylor, editors, *Distributions with fixed marginals and related topics*, vol. 28 of *Lecture Notes–Monograph Series*, pages 120–141. Institute of Mathematical Statistics, Hayward, CA.

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

Joe, H. (2006). Generating random correlation matrices based on partial correlations. *Journal of Multivariate Analysis*, 97:2177–2189.

Joe, H. (2015). *Dependence Modeling with Copulas*. CRC Press, Boca Raton.

Joe, H., Li, H., and Nikoloulopoulos, A. K. (2010). Tail dependence functions and vine copulas. *Journal of Multivariate Analysis*, 101(1):252–270.

Kim, G., Silvapulle, M. J., and Silvapulle, P. (2007). Comparison of semiparametric and parametric methods for estimating copulas. *Computational Statistics & Data Analysis*, 51:2836–2850.

Kraus, D. and Czado, C. (2017). D-vine copula based quantile regression. *Computational Statistics and Data Analysis*, 110:1–18.

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

Laio, F. and Tamea, S. (2007). Verification tools for probabilistic forecasts of continuous hydrological variables. *Hydrology and Earth Sciences*, 11:1267–1277.

Li, W. K. and McLeod, A. I. (1988). ARMA modelling with non-Gaussian innovations. *Journal of Time Series Analysis*, 9(2):155–168.

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

Longla, M. and Peligrad, M. (2012). Some aspects of modeling dependence in copula-based Markov chains. *Journal of Multivariate Analysis*, 111:234–240.

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

McNeil, A. and Bladt, M. (2023). *tscopula: Time Series Copula Models*. R package version 0.3.9.

McNeil, A. J. and Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: An extreme value approach. *Journal of Empirical Finance*, 7:271–300.

McNeil, A. J., Frey, R., and Embrechts, P. (2015). *Quantitative Risk Management: Concepts, Techniques and Tools*. 2nd ed. Princeton University Press, Princeton.

Nagler, T., Krüger, D., and Min, A. (2022). Stationary vine copula models for multivariate time series. *Journal of Econometrics*, 227(2):305–324.

Nagler, T. and Vatter, T. (2025). *rvinecopulib: High Performance Algorithms for Vine Copula Modeling*. R package version 0.7.2.1.0.

Nelson, H. L. and Granger, C. (1979). Experience with using the Box-Cox transformation when forecasting economic time series. *Journal of Econometrics*, 10(1):57–69.

Rosenblatt, M. (1952). Remarks on a multivariate transformation. *Annals of Mathematical Statistics*, 23:470–472.

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

Yan, Y. and Genton, M. G. (2019). Non-Gaussian autoregressive processes with Tukey g-and-h transformations. *Environmetrics*, 30(2).

Zhao, Z., Shi, P., and Zhang, Z. (2022). Modeling multivariate time series with copula linked univariate D-vines. *Journal of Business and Economic Statistics*, 40(2):690–704.