Semiparametric forecasting using non-Gaussian ARMA models based on s-vines

Martin Bladt[1], Alexandra Dias[2], Jialing Han[2], and Alexander J. McNeil[2]

[1]Department of Mathematical Sciences, University of Copenhagen

[2]School for Business and Society, University of York

December 15, 2022

Abstract

A semiparametric method for forecasting time series based on the s-vine copula approach for stationary time series developed in Bladt and McNeil (2022) is proposed. By combining a parametric s-vine process to describe serial dependence with a nonparametric model of the marginal distribution, the method offers improved modelling and forecasting for time series that have a non-Gaussian distribution and a nonlinear dependence on past values. The methodologogy gives a clear meaning to the concept of a non-Gaussian ARMA model in which a parametric object known as the Kendall partial autocorrelation function plays the central role. To demonstrate the potential forecasting gains that can be obtained by using non-Gaussian models, an approach to comparing distributional forecasts proposed by Gneiting and Ranjan (2011) is applied. The methodology is illustrated with an application to forecasting the force of inflation in the US.

Keywords: Time series; vine copulas; Gaussian processes; ARMA processes; ARFIMA processes.

1 Introduction

In this paper we propose a semiparametric method for modelling and forecasting time series based on a stationary vine-copula (or s-vine) approach in which we combine ARMA partial rank correlation functions with non-Gaussian copulas. Our approach builds on a recent paper by Bladt and McNeil (2022) which shows how any classical stationary Gaussian process (such as ARMA, ARFIMA and fractional Gaussian noise) may be made non-Gaussian, both in terms of its marginal and its serial

dependence behaviour. In this paper we concentrate on the ARMA case (including the ARMA models implied by differenced seasonal ARIMA models); the further extension to long-memory ARFIMA models with fractional integration is straightforward. Because the s-vine models extend the class of Gaussian ARMA models, they have the potential to provide superior inference and prediction in any of the settings where these classical models are used but where non-Gaussian and non-linear behaviour may be present.

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; maximum likelihood inference in models with non-Gaussian innovations is challenging to implement and details of algorithms and estimator properties have generally only been established for special cases; the GLM approach of Benjamin et al. (2003) is designed for exponential families.

The approach that we take is extremely flexible and allows non-Gaussianity in two senses. On the one hand, by adopting the copula approach, we can decouple the marginal distribution from the serial dependence model and apply any form of marginal distribution to the Gaussian copula structure implied by a seasonal ARMA model; this preserves the linear structure of the ARMA process under a transformation of the data. On the other hand by considering s-vine decompositions of the multivariate copulas of a stationary Gaussian process, we can systematically replace Gaussian pair copulas with non-Gaussian pair copulas to obtain non-linear ARMA models with non-Gaussian serial dependence structure. In the latter case, the link between a classical ARMA model and our generalized ARMA model of the same order is no longer through a shared linear equation system but rather through a shared parametric object that we call the Kendall partial autocorrelation function (kpacf). The kpacf of a strictly stationary time series $(X_t)_{t\in\mathbb{Z}}$ at lag k is the Kendall rank correlation of the conditional distribution of the pair (X_t, X_{t+k}) given any variables $X_{t+1}, \dots, X_{t+k-1}$ lying in between (Bladt and McNeil, 2022).

The main contribution of this paper is to show how the s-vine approach of Bladt and McNeil (2022) may be adapted to yield a semiparametric estimation and forecasting procedure. Let $\{X_{t-n+1}, \dots, X_t\}$ denote a finite segment from $(X_t)_{t\in\mathbb{Z}}$ and let F_X be the marginal distribution function. We estimate the latter non parametrically to obtain an estimator $\widehat{F}_X$ with values in

$(0,1)$ and then fit fully parametric s-vine copula models to the data $\{\widehat{U}_{t-n+1},\dots,\widehat{U}_t\}$ given by $\widehat{U}_t = \widehat{F}_X(X_s)$ for $s \in \{t-n+1,\dots,t\}$. The objective is then to derive estimates of the conditional distribution of X_{t+1} given k previous values $X_{t-k+1},\dots,X_t$ where $1 \leqslant k \leqslant n$.

Semiparametric methods have been widely applied to copula inference since the seminal paper of Genest et al. (1995), which proposed the two-stage pseudo-maximum-likelihood method. It has been observed that the use of a non-parametric estimator for marginal distributions can lead to better inferences about the dependence structure than are obtained by imposing parametric models, which are often an imperfect fit to data (Shih and Louis, 1995). The semiparametric estimation approach was extended to time series by Chen and Fan (2006) who investigated first-order Markov copula models and proposed a semiparametric forecasting procedure for the quantile function of the conditional distribution of X_{t+1} given X_t. Our methodology can be viewed as generalizing their methodology to allow forecasts that depend explicitly on an arbitrarily large number of past values.

Further contributions of this paper relate to the practical implementation of the s-vine methodology for real data. These include: the use of the kpacf as primary parametric tool for specifying models; the construction of models using sequences of pair copulas that have an arbitrary but finite number of non-Gaussian terms; the extension of the notion of seasonal ARMA dependence to the non-Gaussian case; the development of an evaluation strategy for distributional forecasts in s-vine models using ideas based on Gneiting and Ranjan (2011).

The paper is structured as follows. Section 2 sets out the theory of s-vine processes and explains their relationship to s-vine copulas. Section 3 presents the forecasting methodology and this is then critically evaluated in a simulation study in Section 4. An application of the forecasting methodology to US inflation data is provided in Section 5 while Section 6 concludes.

2 S-vine processes

Notation. Vectors are written $\boldsymbol{x} = (x_1,\dots,x_d)^\top$ and sequences are denoted $(x_t)_{t\in T}$ where T is an index set T, typically the non-zero natural numbers $\mathbb{N}$ or the integers $\mathbb{Z}$. We write $\boldsymbol{x}_{[t:s]} = (x_t,\dots,x_s)^\top$ to refer to a sub-vector of $\boldsymbol{x}$ or a finite section of the sequence $(x_t)_{t\in\mathbb{Z}}$; note that we permit $t > s$, in which case the variables are taken in reverse order to their natural ordering. Random variables and sequences of random variables are denoted by capital letters.

2.1 From s-vine copulas to s-vine processes

Let $(C_k)_{k\in\mathbb{N}}$ be a sequence of bivariate copulas satisfying the following assumption.

Assumption 1. 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 assumption applies to all the standard pair copulas that are used in vine copula models (e.g. Gauss, Clayton[1], Gumbel, Frank, Joe and t), as well as non-exchangeable copulas following the extension of Liebscher (2008) or mixtures of the kind considered by Loaiza-Maya et al. (2018).

An s-vine copula density in dimension $d \geqslant 2$ takes the form

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

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 $R^*_k : (0,1)\times(0,1)^k \to (0,1)$ are families of functions defined from $(C_k)_{k\in\mathbb{N}}$ in a recursive, interlacing fashion by $R_1(x;u) = h^{(1)}_1(u,x)$, $R^*_1(x;u) = h^{(2)}_1(x,u)$ and

$$\begin{aligned} R_k(x;\boldsymbol{u}) &= h^{(1)}_k\left(R^*_{k-1}(u_k;\boldsymbol{u}_{[k-1:1]}), R_{k-1}(x;\boldsymbol{u}_{[1:k-1]})\right) \\ R^*_k(x;\boldsymbol{u}) &= h^{(2)}_k\left(R^*_{k-1}(x;\boldsymbol{u}_{[1,k-1]}), R_{k-1}(u_k;\boldsymbol{u}_{[k-1:1]})\right) \end{aligned} \tag{2}$$

for $k \geqslant 2$, where $h^{(i)}_k(u_1,u_2) = \frac{\partial}{\partial u_i}C_k(u_1,u_2)$. Note that, by slight abuse of notation, R_0 and R^*_0 in formula (1) should be interpreted as the identity functions $R_0(x,\cdot) = R^*_0(x,\cdot) = x$ for all x.

S-vine copulas are d-vine copulas that are subject to additional translation-invariance restrictions which render them suitable to serve as higher-dimensional marginal distributions of stationary processes (Nagler et al., 2022).

Definition 1 (S-vine process)**.** A strictly stationary time series $(X_t)_{t\in\mathbb{Z}}$ is an *s-vine process* if for every $t\in\mathbb{Z}$ and $d\geqslant 2$ 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). An s-vine process $(U_t)_{t\in\mathbb{Z}}$ is an *s-vine copula process* if its univariate marginal distribution is standard uniform.

We refer to the sequences of functions R_k and R^*_k for $k\geqslant 1$ as *forward and backward Rosenblatt functions*. If the copulas C_k are exchangeable for $k = 1,\dots,d$ and $d\geqslant 1$, then a simple inductive

[1]We exclude the case where the Clayton copula has parameter less than zero.

argument shows that $R_k^*(x;\boldsymbol{u}) = R_k(x;\boldsymbol{u})$ for $k = 1,\dots,d$. In this case we can drop the forward and backward qualification and the recursions in (2) simplify to

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

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

$$\begin{aligned} R_k(x;\boldsymbol{u}) &= \mathbb{P}(U_t \leqslant x \mid U_{t-1} = u_1, \dots, U_{t-k} = u_k) \\ R_k^*(x;\boldsymbol{u}) &= \mathbb{P}(U_t \leqslant x \mid U_{t+1} = u_1, \dots, U_{t+k} = u_k). \end{aligned} \tag{4}$$

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 Rosenbaltt forward functions $r_k(x;\boldsymbol{u}) = \frac{\partial}{\partial x} R_k(x;\boldsymbol{u})$ are conditional densities and satisfy

$$r_k(x;\boldsymbol{u}) = \frac{c_{(k+1)}(u_k,\dots,u_1,x)}{c_{(k)}(u_k,\dots,u_1)} = c_1(u_1,x)\prod_{j=2}^{k} c_j\big(R_{j-1}^*(u_j;\boldsymbol{u}_{[j-1:1]}), R_{j-1}(x;\boldsymbol{u}_{[1:j-1]})\big). \tag{5}$$

Under Assumption 1 the conditional distribution functions represented by the Rosenblatt forward functions have unique inverses $Q_k(z;\boldsymbol{u})$ satisfying $R_k(Q_k(z;\boldsymbol{u});\boldsymbol{u}) = z$ for all $(z,\boldsymbol{u}) \in (0,1)\times(0,1)^k$; these inverses, which we refer to as Rosenblatt quantile functions, are used in the sequential generation of realisations from an s-vine process.

As explained in Bladt and McNeil (2022), s-vine processes can be thought of as extending the class of causal stationary Gaussian processes. Every such Gaussian process (such as an ARMA or ARFIMA model) can be represented an an s-vine process consisting of a sequence of bivariate Gaussian copulas $(C_k)_{k\in\mathbb{N}}$ and a Gaussian marginal distrbution function F_X; the sequence of copulas can be finite or infinite depending on the particular Gaussian process. In the more general family of s-vine processes we can replace the Gaussian marginal distribution by an arbitrary continuous distribution and we can replace the sequence of Gaussian copulas by a sequence of arbitrary bivariate copulas satisfying Assumption 1.

The copula sequence $(C_k)_{k\in\mathbb{N}}$ determines the serial dependence properties of the process and we refer to it 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 s-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.

A useful function for describing the serial dependence of an s-vine process is the Kendall partial autocorrelation function or kpacf $(\tau_k)_{k\in\mathbb{N}}$. This is simply the sequence of Kendall partial correlation coefficients $\tau_k = \tau(C_k)$ of the partial copula sequence. If $(X_t)_{t\in\mathbb{Z}}$ is a Gaussian process and $(U_t)_{t\in\mathbb{Z}}$ the associated Gaussian copula process, given by $U_t = F_X(X_t)$ for all t, then the kpacf completely characterizes $(U_t)_{t\in\mathbb{Z}}$. This follows from the fact that there is a bijective mapping from the kpacf to the partial correlation function (pacf) $(\alpha_k)_{k\in\mathbb{N}}$ (via the transformation $\alpha_k = \sin(\pi\tau_k/2)$) and hence to the autocorrelation function (acf) of $(X_t)_{t\in\mathbb{Z}}$. Since Gaussian processes are characterized by their first two moments, this implies that the kpacf $(\tau_k)_{k\in\mathbb{N}}$ of a Gaussian copula process identifies a unique process.

2.2 Stability of s-vine processes

S-vine processes are strictly stationary by design since they exploit the translation-invariance of s-vine copulas. A finite sequence $U_1, \dots, U_n$ of any length n from an s-vine copula process $(U_t)_{t\in\mathbb{Z}}$ process with copula sequence $(C_k)_{k\in\mathbb{N}}$ can be constructed from a sequence of iid uniform variates $(Z_t)_{t\in\mathbb{N}}$ by setting $U_1 = Z_1$ and $U_k = Q_{k-1}(Z_k; \boldsymbol{U}_{[k-1:1]})$ for $k = 2, \dots, n$. The random vector $\boldsymbol{U} = (U_1, \dots, U_n)^\top$ will have the joint density $c_{(n)}$ of an s-vine copula in (1).

The question of ergodicity is more complicated than that of stationarity. We distinguish between s-vine processes of finite and infinite order. The former case describes models where the copula sequence $(C_k)_{k\in\mathbb{N}}$ satisfies $C_p \neq C^\perp$ and $C_k = C^\perp$ for $k > p$ for some $p \geqslant 1$, where $C^\perp$ denotes the bivariate independence copula. Finite-order s-vine processes are ergodic Markov processes on a p-dimensional state space and can be thought of as extensions of the Gaussian AR(p) process. A number of authors have explored mixing and ergodic convergence rates in the case where $p = 1$ (Chen and Fan, 2006; Beare, 2010; Chen et al., 2009; Longla and Peligrad, 2012); it is known that first-order Markov models based on Gaussian, Student t, Frank, Clayton and Gumbel copulas are all geometrically β-mixing or, equivalently for Markov chains, geometrically ergodic. For $p \geqslant 1$, Zhao et al. (2022) use a small set approach for Markov chains to give conditions under which an s-vine is

geometrically ergodic. However, the conditions are difficult to verify for an arbitrary set of copulas $C_1, \dots, C_p$ with $p > 1$.

An s-vine process of infinite order is one in which, for every $p \geqslant 1$ there exists a $k \geqslant p$ such that $C_k \neq C^\perp$. Any Gaussian ARMA(p, q) process with $q > 0$ provides an example of an infinite-order s-vine process. To formulate a stability condition, let the functions $S_k : (0,1) \times (0,1)^k \to (0,1)$, $k \in \mathbb{N}$, be defined from the Rosenblatt quantile functions by $S_1(x; z) = Q_1(x; z)$ and

$$S_k(x; \boldsymbol{z}) = Q_k\left(x; \left(S_{k-1}(z_1; \boldsymbol{z}_{[2:k]}), \dots, S_1(z_{k-1}; z_k), z_k\right)^\top\right), \quad k \geqslant 2.$$

If we set $S_0(x; \cdot) = x$ then the functions $S_0, \dots, S_{n-1}$ describe the mapping by which independent uniform innovations $Z_1, \dots, Z_n$ are transformed into process values $U_1, \dots, U_n$ via the transformations $U_k = S_{k-1}(Z_k; \boldsymbol{Z}_{[k-1:1]})$ for $k = 1, \dots, n$, a transformation also known as the multivariate quantile transform (Rüschendorf, 2009).

Assumption 2. For a sequence $(Z_t)_{t\in\mathbb{N}}$ of iid uniform innovations the limit

$$S_\infty(x; (Z_t)_{t\in\mathbb{N}}) = \lim_{k\to\infty} S_k(x; \boldsymbol{Z}_{[1:k]}) \tag{6}$$

exists almost surely for all $x \in (0,1)$ and, for any sample path $(Z_t(\omega))_{t\in\mathbb{N}}$, is a continuous and increasing function such that $\lim_{x\to 0} S_\infty(x; (Z_t(\omega))_{t\in\mathbb{N}}) = 0$ and $\lim_{x\to 1} S_\infty(x; (Z_t(\omega))_{t\in\mathbb{N}}) = 1$.

Under Assumption 2 the mapping S_∞ describes a convergent non-linear filter of noise of the kind discussed by Wu (2005). This makes it possible to define a causal, stationary and ergodic process $(U_t)_{t\in\mathbb{Z}}$ of the form

$$U_t = S_\infty(Z_t; (Z_{t-1}, Z_{t-2}, \dots)^\top)$$

for which the copula sequence $(C_k)_{k\in\mathbb{N}}$ describes the partial dependencies. However, the existence of the limit (6) is hard to verify analytically for non-Gaussian copulas sequences. It is certainly necessary that $C_k \to C^\perp$ as $k \to \infty$, but a counterexample in (Bladt and McNeil, 2022) shows that this is not sufficient; assumptions on the speed of covergence to independence are required. In the practical application of these models in this paper, we truncate all copula sequences so that we remain within the realm of finite-order processes.

2.3 S-vine processes with ARMA dependence structure

In this section we explain how to construct s-vine processes which have the serial dependence structure of ARMA processes. We use a generalized definition of an ARMA process that allows the kind of seasonal features implied by seasonal ARIMA models. These seem particularly useful for many modelling problems in economics and finance.

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 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 and where 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 $(X_t)_{t\in\mathbb{Z}}$ will be said to follow an ARMA model of order $(p,q)(P,Q)_s$. This is effectively an ARMA process of order $(p+sP, q+sQ)$ in the usual notation, 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.

Since every Gaussian copula process can be characterized by its kpacf, it follows that there is a one-to-one mapping from the set of causal stationary and invertible Gaussian ARMA processes with fixed mean and variance to the set of Kendall partial autocorrelation functions of Gaussian ARMA processes; eack kpacf indentifies a unique ARMA process (up to location and scaling) and it makes sense to talk, for example, of the kpacf of an ARMA(1,1) process. We note, however, that there is some ambiguity in referring to the kpacf of an ARMA$(p,q)(P,Q)_s$ process; for example, an ARMA$(1,1)(1,1)_4$ shares its kpacf with an ARMA$(5,1)(0,1)_4$ model and an ARMA$(1,5)(1,0)_4$ model. With this minor caveat we will use the following terminology.

Definition 2. Any s-vine process whose kpacf is identical to the kpacf of a causal stationary and

invertible Gaussian ARMA$(p,q)(P,Q)_s$ process is said to be an ARMA$(p,q)(P,Q)_s$ s-vine process.

In Bladt and McNeil (2022) a parsimonious method of parameterizing s-vine processes via their kpacf is proposed and this method may be applied to an s-vine process with ARMA $(p,q)(P,Q)_s$ Kendall partial autocorrelation structure. Suppose the periodicity s and order (p,q,P,Q) of the copula process are fixed and let $\boldsymbol{\theta}$ represent a vector of feasible parameters for a causal stationary Gaussian ARMA$(p,q)(P,Q)_s$ process. The kpacf of the Gaussian process is given by $(\tau_k(\boldsymbol{\theta}))_{k\in\mathbb{N}}$ where $\tau_k(\boldsymbol{\theta}) = \frac{2}{\pi}\arcsin\alpha_k(\boldsymbol{\theta})$ and $(\alpha_k(\boldsymbol{\theta}))_{k\in\mathbb{N}}$ is the pacf, which can be readily calculated for any Gaussian ARMA process. The idea is that we consider s-vine models with sequences $(C_k)_{k\in\mathbb{N}}$ such that the Kendall rank correlations satisfy $\tau(C_k) = \frac{2}{\pi}\arcsin\alpha_k(\boldsymbol{\theta})$ 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 relationship between the parameter(s) of the copula and Kendall's tau, as is the case for the Frank, Clayton, Gumbel, Joe and t copulas. If the copulas have more than one parameter a set of additional parameters $\boldsymbol{\phi}$ is introduced into the model. To fit such a model to data, the parameters $\boldsymbol{\theta}$ are optimized over the set of feasible parameters for a causal stationary Gaussian ARMA $(p,q)(P,Q)_s$ process and any additional parameters $\boldsymbol{\phi}$ are either fixed or optimized over their feasible domain.

While we have a lot of flexibility in our choice of copula sequence, in this paper we will consider Gaussian copula sequences with finitely many non-Gaussian copula substitutions. Our intuition is that the largest benefits of including non-Gaussian copulas are likely to be seem at lower lags. As the partial copula sequence tends to the independence copula and partial dependencies become weaker, the differences between Gaussian and non-Gaussian copulas will become less pronounced.

The substituted non-Gaussian copulas may be required to attain any Kendall rank correlation in the interval $(-1,1)$; copulas with this property are called comprehensive. The t and Frank copulas are comprehensive but the Gumbel and Joe copulas (and their corresponding survival copulas) can only model positive dependence and attain rank correlations in $[0,1)$. The Clayton family is comprehensive but when the Kendall rank correlation is negative the copula does not have a strictly positive density on $(0,1)^2$ and violates Assumption 1; we thus restrict attention to Clayton copulas (and survival copulas) with positive dependence. When positive-dependence copulas are substituted we allow the option of rotating these through 90 degrees (clockwise or anticlockwise) to obtain negative partial dependence.

A proof of ergodicity for ARMA s-vine processes based on infinite Gaussian copula sequences with finitely-many non-Gaussian substitutions has not yet been given. To avoid this theoretical

issue, and to increase the speed of estimation algorithms, we work in practice with finite-order models by truncating copula sequences at some high order (e.g. 50 or 100) where the addition of further terms leads to a negligible improvement in fit.

3 Forecasting methodology

3.1 Forecasting formulas

Given an s-vine process $(X_t)_{t\in\mathbb{Z}}$, suppose we want to forecast X_{t+1} conditional on the previous k values, i.e. conditional on the event $\{\boldsymbol{X}_{[t-k+1:t]} = \boldsymbol{x}_{[t-k+1:t]}\}$. We denote the predictive distribution function by $F_{t+1|k}(x)$, suppressing the explicit dependence on $\boldsymbol{x}_{[t-k+1:t]}$ for notational convenience. The predictive distribution function and its density and quantile function are given by

$$\begin{aligned} F_{t+1|k}(x) &= R_k\left(F_X(x); F_X(\boldsymbol{x}_{[t:t-k+1]})\right) \\ f_{t+1|k}(x) &= r_k\left(F_X(x); F_X(\boldsymbol{x}_{[t:t-k+1]})\right) f_X(x) \\ F^{-1}_{t+1|k}(u) &= F_X^{-1}\left(Q_k(u; F_X(\boldsymbol{x}_{[t:t-k+1]}))\right) \end{aligned} \tag{7}$$

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).

3.2 The semiparametric method

Let us suppose we want to estimate the functions in (7) using the last n observations of the process $x_{t-n+1}, \ldots, x_t$. Semiparametric estimates may be obtained when F_X is estimated by a version of the empirical distribution function (edf) of the data, denoted $\widehat{F}_X^{(n)}$. Similarly, f_X can be replaced by a kernel estimate $\widehat{f}_X^{(n)}$ derived from these data and $F_X^{-1}(\alpha)$ can be estimated by calculating an empirical α-quantile, either by inverting $\widehat{F}_X^{(n)}$ or by using an alternative definition of an empirical quantile as in Hyndman and Fan (1996).

Suppose that we know the structure of the s-vine copula process $(U_t)_{t\in\mathbb{Z}}$ of $(X_t)_{t\in\mathbb{Z}}$, and therefore the form of the Rosenblatt function R_n, but the marginal distribution F_X is unknown and needs to be estimated. Consider the estimator of the distribution function of the predictive distribution given by

$$\widehat{F}^{(n)}_{t+1|k}(x) = R_k\left(\widehat{F}_X^{(n)}(x); \widehat{F}_X^{(n)}(\boldsymbol{x}_{[t:t-k+1]})\right). \tag{8}$$

Provided $(X_t)_{t\in\mathbb{Z}}$ is an ergodic process, then for any x in the interior of the support of F_X, we have

$\widehat{F}_X^{(n)}(x) \to F_X(x) \in (0,1)$ as $n \to \infty$, almost surely. Assumption 1 implies that R_k is a continuous function on $(0,1) \times (0,1)^k$. Provided we use a version of the empirical distribution function such that $\widehat{F}_X^{(n)}(\boldsymbol{x}_{[t:t-k+1]}) \in (0,1)^k$ for any $k \leqslant n$ we can conclude that $\widehat{F}_{t+1|k}^{(n)}(x) \to F_{t+1|k}(x)$ as $n \to \infty$. One possibility is to use $\widehat{F}_X^{(n)}(x) = \frac{1}{n+1} \sum_{i=1}^{n} I_{\{x_{t-n+i} \leqslant x\}}$ but kernel estimators are also possible. In practice we don't know R_k and this will be estimated parametrically by a function $\widehat{R}_k$ using the fitted s-vine.

3.3 Evaluating distributional forecasts

In this section we consider a sequence of estimates $\widehat{F}_{t+i|k}, i = 1, \ldots, m$ of the one-step predictive distributions $F_{t+i|k}$ of the values X_{t+i}, in each case conditioning on the last k values. One method of evaluating these distributional forecasts is based on the probability-integral transform using the so-called PIT values $\{u_{t+1}, \ldots, u_{t+m}\}$ given by $u_{t+i} = \widehat{F}_{t+i|k}(x_{t+i})$ for $i = 1, \ldots, m$, where x_{t+i} is the realized value of X_{t+i}. Under perfect estimation these should form an iid uniform sample and this can be easily tested, for example by using Kolmogorov-Smirnov goodness-of-fit tests and Ljung-Box tests of serial dependence. However, when m is small, these tests turn out to be quite weak tests of the quality of the forecast distributions. The null hypothesis of independence and uniformity may fail to be rejected for a number of competing forecast models including some in which the serial dependence structure may be misspecified.

To choose between models we can also apply a scoring approach following the methodology proposed by Gneiting and Ranjan (2011). A commonly-used proper scoring function for comparing distributional forecasts $\widehat{F}_{t+i|k}$ with realized values x_{t+i} is the continuous ranked probability score (CRPS) which takes the equivalent forms

$$\mathrm{CRPS}\left(\widehat{F}_{t+i|k}, x_{t+i}\right) = \int_{-\infty}^{\infty} \mathrm{S}^B\left(\widehat{F}_{t+i|k}(y), I_{\{x_{t+i} \leqslant y\}}\right) \mathrm{d}y \tag{9}$$

$$= 2 \int_0^1 \mathrm{S}_\alpha^Q\left(\widehat{F}_{t+i|k}^{-1}(\alpha), x_{t+i}\right) \mathrm{d}\alpha \tag{10}$$

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

$$\mathrm{S}_\alpha^Q(y,x) = \left(I_{\{x \leqslant y\}} - \alpha\right)(y - x) \tag{11}$$

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 (9) and (10) to emphasise different

features of the forecast distribution (such as centre or tails) and we will adopt the approach based on (10). That is, we consider scores which take the form $\int_0^1 \mathrm{S}_\alpha^Q\left(\widehat{F}_{t+i|k}^{-1}(\alpha), x_{t+i}\right) \mathrm{d}\nu(\alpha)$ 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

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

where $v(\alpha)$ is the weight function and J is some integer (e.g. $J = 100$) that determines the granularity of the discrete measure. For uniform weighting we can set $v(\alpha) = 1$ and in this case we will simply refer to the AQS, or average quantile score.

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

$$\alpha \to \mathbb{E}\left(S_\alpha^Q(F^{-1}(\alpha), X\right) \tag{13}$$

where F is a distribution function that roughly corresponds to the marginal distribution of the data and X is a random variable with distribution function F. Thus the template function at α is the expected quantile score for an ideal forecaster who forecasts the α-quantile of a random variable X using the correct quantile function F^{-1}. In some applications it may be appropriate to take the reciprocal of the normal template function, i.e. $v(\alpha)^{-1} = \mathbb{E}\left(S_\alpha^Q(\Phi^{-1}(\alpha), X\right)$ where $X \sim N(0,1)$. For a copula process with uniform marginal distribution one might take the reciprocal of the uniform template function which has a simple analytical form: $v(\alpha)^{-1} = \mathbb{E}\left(S_\alpha^Q(\alpha, U\right) = \frac{1}{2}\alpha(1-\alpha)$ where $U \sim U(0,1)$.

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 values for

$$\overline{\mathrm{AWQS}}_m^{(l)} = \frac{1}{m}\sum_{i=1}^{m} \mathrm{AWQS}\left(\widehat{F}_{t+i|k}^{(l)}, x_{t+i}\right), \quad l = 1, 2.$$

To judge whether there is a significant different in forecast performance we can use the test of Diebold

and Mariano (1995) based on the statistic

$$T_m = \sqrt{m}\frac{\overline{\text{AWQS}}_m^{(1)} - \overline{\text{AWQS}}_m^{(2)}}{\widehat{\sigma}_m}$$

where

$$\widehat{\sigma}_m^2 = \frac{1}{m}\sum_{i=1}^{m}\left(\text{AWQS}\left(\widehat{F}_{t+i|k}^{(1)}, x_{t+i}\right) - \text{AWQS}\left(\widehat{F}_{t+i|k}^{(2)}, x_{t+i}\right)\right)^2.$$

This statistic T_m is compared with a normal reference distribution and, in the case that the null hypothesis of equal forecast performance is rejected, the model giving the smaller value of $\overline{\text{AWQS}}_m^{(l)}$ should be preferred.

To understand the differences between models we also use a graphical exploration of the differences between the forecast models based on a suggestion by Laio and Tamea (2007). This is a plot of the points

$$\left(\alpha_j, \quad \frac{v(\alpha_j)}{m}\sum_{i=1}^{m}\text{S}_\alpha^Q\left(\widehat{F}_{t+i|k}^{(l)-1}(\alpha_j), x_{t+i}\right)\right), \quad \alpha_j = \frac{j}{J}, \quad j = 1, \ldots, J-1. \tag{14}$$

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

4 A simulation study

In the simulation study we focus on examining whether the forecast evaluation methods of Section 3.3 can distinguish between the performance of models based on correctly and incorrectly specified copula processes. The marginal distribution of the data is always estimated nonparametrically using the (scaled) empirical distribution function.

The true data-generating process is a Clayton AR(3) model, in other words a model in which the partial copulas at the first three lags are all Clayton or rotated Clayton (and partial copulas of higher order are all equal to the independence copula). Use of an AR model (as opposed to MA or ARMA models) leads to faster simulation times and thus accelerates the Monte Carlo study. The parameters of the three Clayton copulas are respectively 1.0, 0.8 and 0.6, but the second copula is rotated through 90 degrees to the left; the corresponding Kendall rank correlations are 0.333, -0.286, 0.231.

We consider two forecasters: the first forecaster correctly deduces the order of the model and the

fact that the copula sequence is Clayton; the second forecaster correctly deduces the order of the model but uses a sequence of 3 Gaussian partial copulas. Both forecasters estimate the parameters of their models from the data.

In each repetition of the simulation experiment we assume that the first n observations from the data-generating process are used to estimate the parameters of the copula process and these parameters are subsequently held fixed. The marginal distribution is estimated used a rolling window of n observations and a series of m one-step forecasts are made. To calculate PIT values, we use formula (8) with $k = 3$ (since both forecasters use models of finite order 3) combined with the scaled version of the empirical distribution function $\widehat{F}_X^{(n)}(x) = \frac{1}{n+1}\sum_{i=1}^{n} I_{\{x_{t-n+i}\leqslant x\}}$ to ensure that the first and second arguments of the function R_k are in $(0,1)$ and $(0,1)^k$ respectively; this guarantees the PIT values are also in $(0,1)$.

To calculate quantile scores, AQS and AWQS values we estimate $F^{-1}_{t+1|k}(u)$, the u-quantile of the predictive distribution, by taking the empirical $Q_k(u; \widehat{F}_X^{(n)}(\boldsymbol{x}_{[t:t-k+1]}))$-quantile of the data $\{x_{t-n+1}, \ldots, x_t\}$; 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 (14) for a single simulation experiment where $n = 500$ and $m = 100$. In this case we have applied the uniform template function and we have set $J = 100$. The black line show the score curve for the first forecaster, who deduces the correct copula sequence (but still has to estimate their parameters); the red line shows the score curve for the second forecaster, who incorrectly uses a sequence of Gaussian copulas; the green line shows the score curve for an ideal forecaster who knows the correct copula sequence and its parameters. Note how the weighted score plot blows up the differences in both tails and potentially forms a better basis for comparing forecast performance when the tails of the predictive distribution are important. The Diebold-Mariano test of forecast equality yields a p-value of 4.9×10^{-5} for the unweighted comparison based on the AQS and a (smaller) p-value of 2.9×10^{-5} for the weighted comparison based on the AWQS. In this case the weighting makes no difference to the inference (that the first forecaster is significantly better) but this is not always the case.

In the simulation study, using different values of n and m, we obtain the results in Table 1 based on 1000 replications. Note that, for example, the case 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 5); other cases are more relevant to data collected at a higher frequency, or for a longer period.

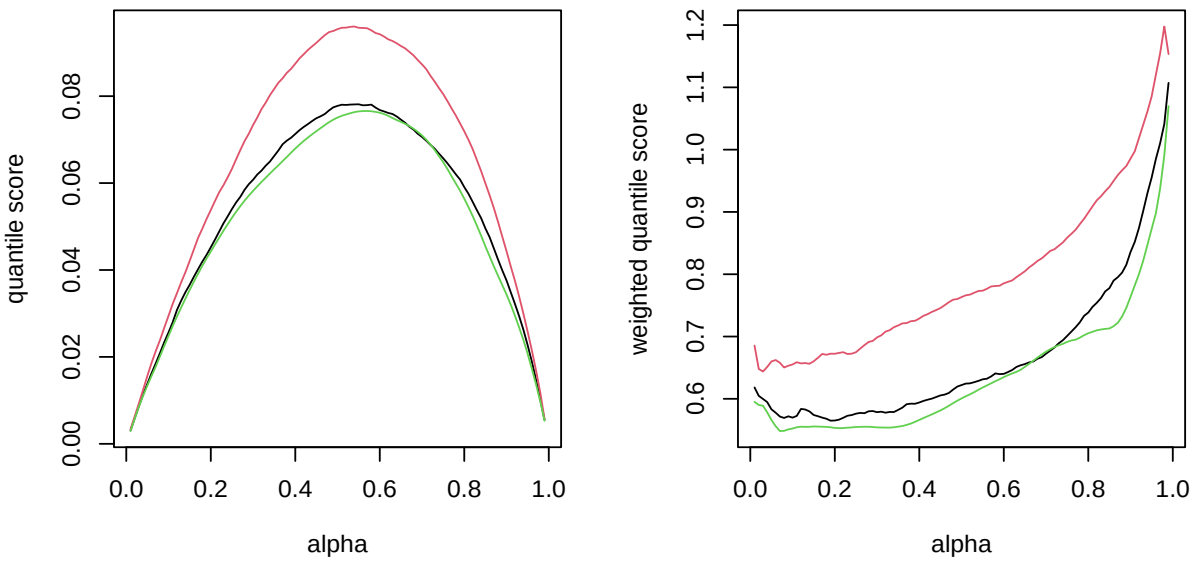

Figure 1: Quantile score plot and weighted quantile scoreplot for $m = 100$ forecasts based on an estimation window of length $n = 500$. Weighting is by the reciprocal of the uniform template function and $J = 100$. Black line - forecaster using correct non-Gaussian s-vine copula model; red line - forecaster using incorrect Gaussian copula model; green line - ideal forecaster.

The results show that the AIC values from the in-sample estimation of the parametric copula process always favour the correctly specified model. The tests based on out-of-sample PIT values are relatively ineffective at distinguishing between the correctly and incorrectly specified models. The Kolmogorov-Smirnov test for the uniformity of the PITs rejects the correctly specified model a little more often than desired for tests of size 5%. Conversely, for the misspecified model the rejection rates, while larger than for the correctly specified model, are very modest and never greater than 17.5%. The Ljung-Box test for abence of serial correlation in the PITs gives similarly weak results.

The AQS and AWQS statistics calculated from the out-of-sample forecasts both tend to be smaller for the correctly specified model than the incorrectly specified model. This tendency increases with m, the number of forecasts, and with n, the size of the estimation window. The weighted AWQS and the unweighted AQS both lead to a similar percentage of correct model orderings.

There is, however, some difference between the Diebold-Mariano test results for equivalent forecast performance for the AWQS and AQS statistics. The AWQS leads to a higher percentage of significant test results than the AQS for all values of n and m. However, we need to make at least $m = 100$ forecasts to see significant differences in AWQS values at least 50% of the time.

n	100		200		500	
m	40	100	40	100	40	100
AIC correctly ordered	95.2		99.0		100	
PIT KS-test significant (correct model)	10.0	8.6	7.8	9.5	5.6	7.2
PIT KS-test significant (misspecified model)	12.0	12.3	13.0	17.5	14.5	15.7
PIT LB-test significant (correct model)	9.2	22.2	7.5	14.1	4.9	7.9
PIT LB-test significant (misspecified model)	9.8	27.5	8.4	16.6	5.6	10.1
AQS correctly ordered	86.1	93.1	91.3	97.3	93.3	98.1
AQS DM-test significant	24.8	48.5	28.9	58.5	31.5	63.3
AWQS correctly ordered	86.6	93.1	92.1	97.3	93.9	98.5
AWQS DM-test significant	26.4	49.8	32.4	61.1	38.5	68.4

Table 1: Simulation study on evaluation of forecasts. Numbers in body of table are percentages based on 1000 replications. KS = Kolmogorov Smirnov test for uniformity. LB = Ljung-Box test for absence of serial dependence (based on 3 lags). DM = Diebold-Mariano test for equivalent forecast performance. The values of J used were 25, 50 and 100 for n values of 100, 200 and 500 respectively.

These results suggest that the forecast comparison approach based on scoring developed by Gneiting and Ranjan (2011) is a useful method for distinguishing between competing nonparametric ARMA models based on different s-vine specifications. It seems reasonable to use the ordering of AQS or AWQS values (particularly when this accords with the ordering of in-sample AIC values) to choose a favoured forecasting model, although a large number of forecasts are required before clear superiority of one model over another can be demonstrated.

5 Forecasting the force of inflation in the US

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 first quarter of 2022. From these we compute quarterly values for the annualized force of inflation and express these as a percentage. In other words, our data 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, and we have 249 quarterly observations of the force of inflation from the first quarter of 1960, 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. An automatic ARIMA modelling procedure in R suggests 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 force of inflation $X_t = Y_t - Y_{t-1}$ in our s-vine models; this yields $n = 207$ observations in the training set.

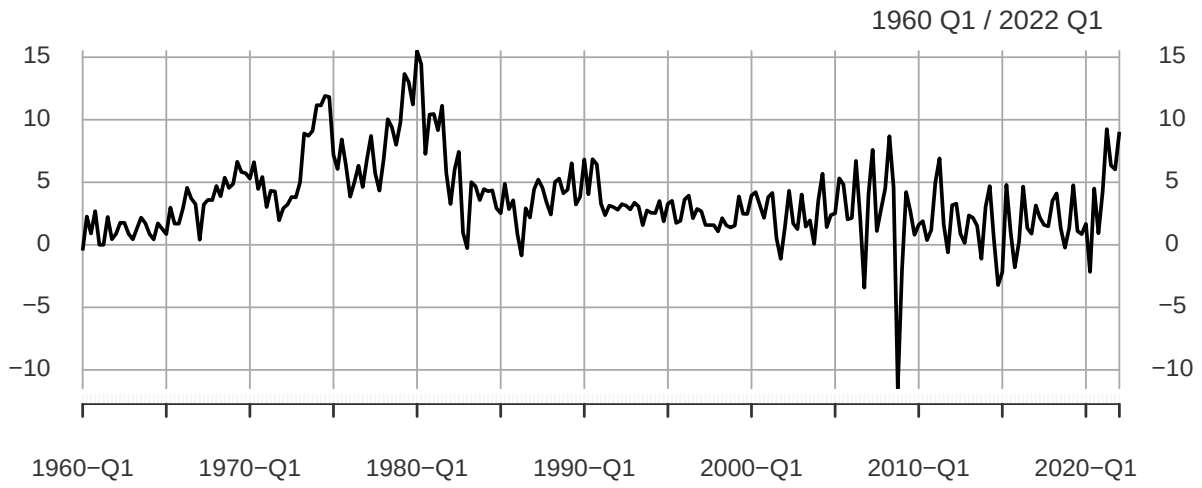

Figure 2: Quarterly data on the annual force of inflation in the US from 1960 - 2022.

We use the semi-parametric method to fit various s-vine models of order $(1,2)(1,0)_4$ to the data $x_1, \ldots, x_n$ starting with a model where all pair copulas are Gaussian, which we consider to be the baseline model. We systematically replace pair copulas at the first few lags with non-Gaussian copulas from the Clayton, Joe, Frank and Gumbel families; in these cases the copula sequences are all truncated at lag $k = 40$ since longer copula sequences yield no improvement in fit. The most promising models result from Gumbel substitutions and our final chosen model of order $(1,2)(1,0)_4$ is based on substituting Gumbel copulas at the first 15 lags. The parameter estimates, AIC and BIC values for the baseline Gaussian copula model and the model with Gumbel substitutions (which we simply refer to as the Gumbel model) are shown in Table 2. Figure 3 shows the Kendall partial autocorrelation function (kpacf) of the Gumbel model, showing negative partial dependence at the first three gags and then a decaying pattern of alternating positive and negative partial dependence thereafter; this picture provides some justification for truncation of the copula sequence at lag $k = 40$.

	ar1	ma1	ma2	sar1	AIC	BIC
Gaussian model	-0.45	0.14	-0.67	0.30	-77.20	-63.80
Gumbel model	-0.47	0.14	-0.61	0.34	-83.80	-70.40

Table 2: Estimates for AR, MA and seasonal AR parameters as well as AIC and BIC values for two ARMA $(1,2)(1,0)_4$ models fitted to training data.

We conducted a residual analysis for both models based on an idea proposed in Bladt and McNeil

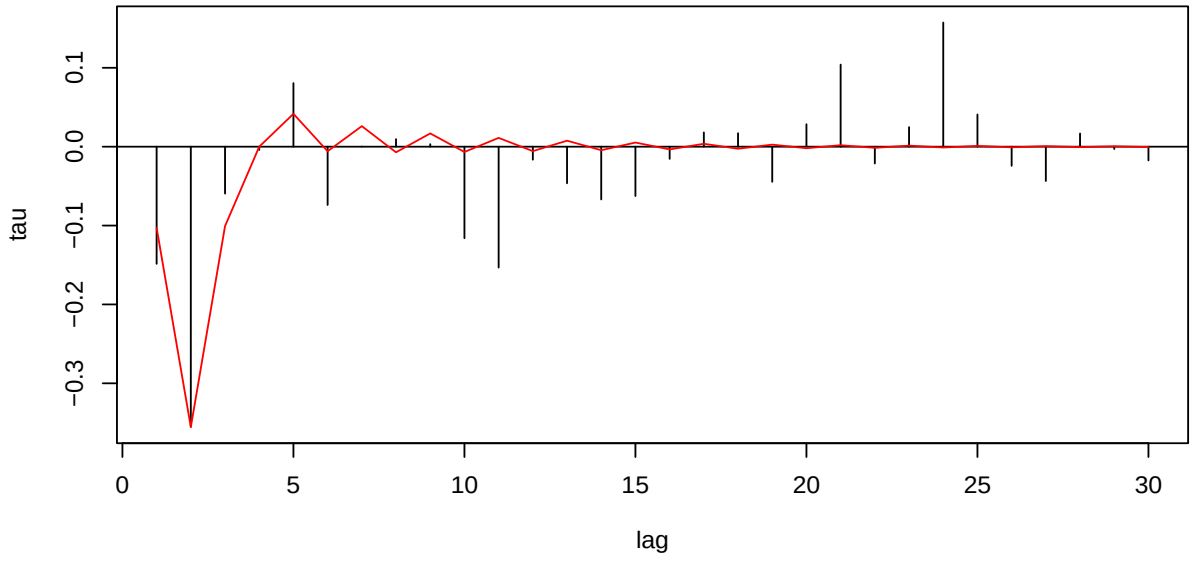

Figure 3: Kendall partial autocorrelation function (kpacf) of the Gumbel copula model.

(2022). We construct the series $\{z_1, \ldots, z_n\}$ by setting $z_1 = \widehat{F}_X^{(n)}(x_1)$ and

$$z_{t+1} = \widehat{R}_t \left(\widehat{F}_X^{(n)}(x_{t+1}); \widehat{F}_X^{(n)}(\boldsymbol{x}_{[t:1]}) \right), \quad t = 1, \ldots, n-1,$$

and then test whether $\{\Phi^{-1}(z_1), \ldots, \Phi^{-1}(z_n)\}$ forms a random sample from a normal distribution. For both models the normal hypothesis is not rejected in a Shapiro-Wilks test.

We now turn to one-step, out-of-sample forecasting using $m = 41$ values of the first difference $X_t = Y_t - Y_{t-1}$ starting in the first quarter of 2012 and ending in the first quarter of 2022. To compare the forecasting performance of the models we form the score plot and weighted score plot (this time based on the normal template function) as shown in Figure 4. The Gumbel model tends to give lower scores and lower weighted scores than the Gaussian model. The AQS values are respectively 0.643 and 0.659 while the AWQS values (using the normal template function) are 2.26 and 2.32. While these both indicate a preference for the Gumbel model, the Diebold-Mariano tests do not indicate a significant different in forecasting performance in this case (p-values are 0.49 for AQS and 0.50 for AWQS).

To complete the analysis we show in Figure 5 our forecasts of the distribution of Y_t for each quarter from the start of 2012. The distributional forecasts are summarised by red lines at the

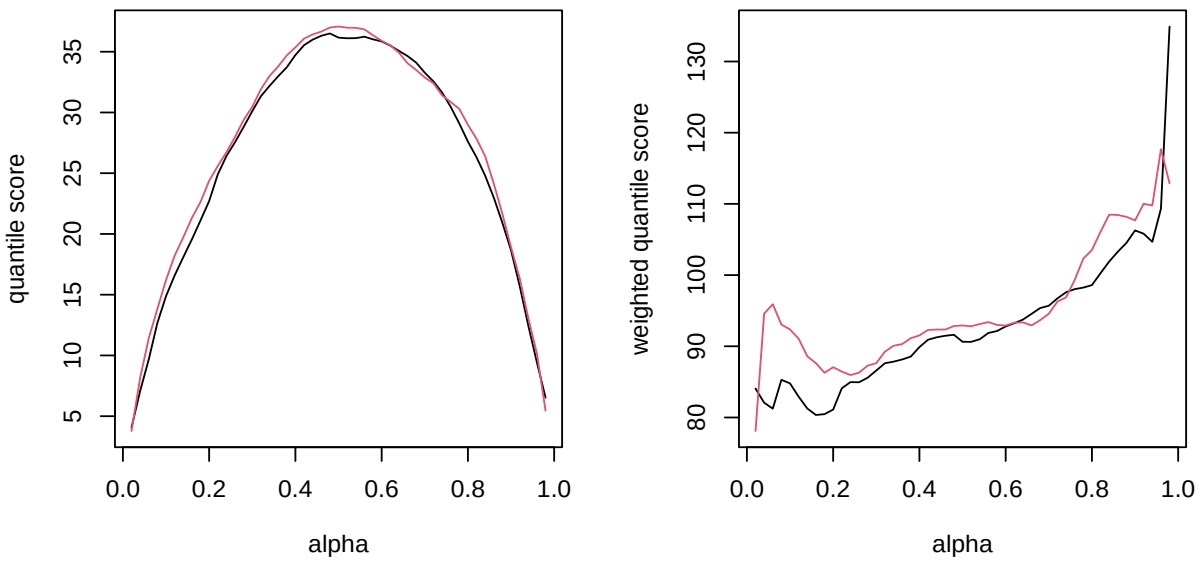

Figure 4: Quantile score plot and weighted quantile scoreplot for $m = 41$ forecasts of the first difference of the force of inflation based on an estimation window of length $n = 207$. Weighting is by the reciprocal of the normal template function and $J = 50$. Black line - forecaster using Gumbel copula model; red line - forecaster using Gaussian copula model.

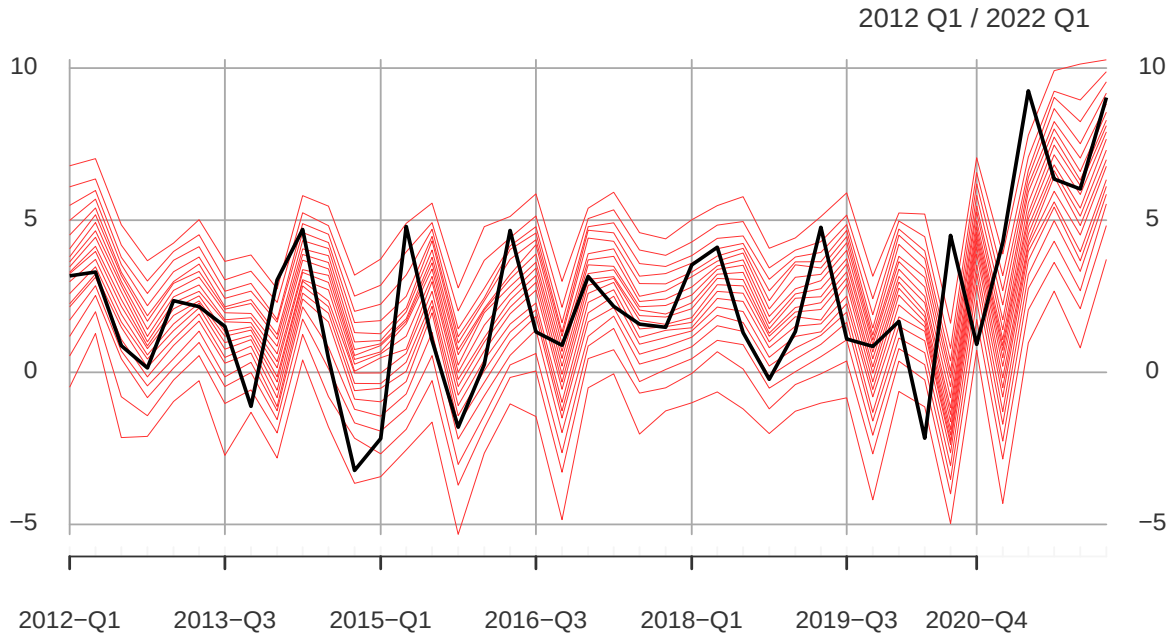

Figure 5: Forecasting the annual force of inflation in the US from 2012–2022. The red lines show estimated quantiles of the forecast distribution at the levels $\alpha_j = j/20$ for $j = 1, \ldots, 19$.

estimated α_j-quantiles, where $\alpha_j = j/20$ for $j = 1, \ldots, 19$. The actual realized value of Y_t is superimposed as a black line.

6 Conclusion

The methodology described in this paper offers improved forecasting for time series that are poorly modelled by classical ARMA processes. 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 non-parametric estimate of the marginal distribution as well as a detailed modelling of serial dependence using a parametric s-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` on CRAN.

References

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

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. (2022). Time series with infinite-order partial copula dependence. *Dependence Modeling*, 10:87–107.

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. *Annals of Statistics*, 37(6B):4214–4253.

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

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.

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

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.

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

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.

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

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.

Rüschendorf, R. (2009). On the distributional transform, Sklar's Theorem, and the empirical copula process. *Journal of Statistical Planning and Inference*, 139:3921–3927.

Shih, J. H. and Louis, T. A. (1995). Inferences on the association parameter in copula models for bivariate survival data. *Biometrics*, 51:1384–1399.

Wu, W. (2005). Nonlinear system theory: another look at dependence. *Proceedings of the National Academy of Sciences USA*, 102(40):14150–14154.

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.