White Rose Research Online

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of York.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/170011/

Version: Published Version

Article:
McNeil, Alexander John (2021) Modelling volatile time series with v-transforms and copulas. Risks. 14. ISSN: 2227-9091

https://doi.org/10.3390/risks9010014

Reuse
This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:
https://creativecommons.org/licenses/

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

UNIVERSITY OF LEEDS
University of Sheffield
UNIVERSITY of York

risks

MDPI

Article

Modelling Volatile Time Series with V-Transforms and Copulas

Alexander J. McNeil

The York Management School, University of York, Freboys Lane, York YO10 5GD, UK;
alexander.mcneil@york.ac.uk

Abstract: An approach to the modelling of volatile time series using a class of uniformity-preserving transforms for uniform random variables is proposed. V-transforms describe the relationship between quantiles of the stationary distribution of the time series and quantiles of the distribution of a predictable volatility proxy variable. They can be represented as copulas and permit the formulation and estimation of models that combine arbitrary marginal distributions with copula processes for the dynamics of the volatility proxy. The idea is illustrated using a Gaussian ARMA copula process and the resulting model is shown to replicate many of the stylized facts of financial return series and to facilitate the calculation of marginal and conditional characteristics of the model including quantile measures of risk. Estimation is carried out by adapting the exact maximum likelihood approach to the estimation of ARMA processes, and the model is shown to be competitive with standard GARCH in an empirical application to Bitcoin return data.

Keywords: time series; volatility; probability-integral transform; ARMA model; copula

check for updates

Citation: McNeil, Alexander J. 2021. Modelling Volatile Time Series with V-Transforms and Copulas. *Risks* 9: 14. http://doi.org/10.3390/risks9010014

Received: 29 November 2020
Accepted: 29 December 2020
Published: 5 January 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

In this paper, we show that a class of uniformity-preserving transformations for uniform random variables can facilitate the application of copula modelling to time series exhibiting the serial dependence characteristics that are typical of volatile financial return data. Our main aims are twofold: to establish the fundamental properties of v-transforms and show that they are a natural fit to the volatility modelling problem; to develop a class of processes using the implied copula process of a Gaussian ARMA model that can serve as an archetype for copula models using v-transforms. Although the existing literature on volatility modelling in econometrics is vast, the models we propose have some attractive features. In particular, as copula-based models, they allow the separation of marginal and serial dependence behaviour in the construction and estimation of models.

A distinction is commonly made between genuine stochastic volatility models, as investigated by Taylor (1994) and Andersen (1994), and GARCH-type models as developed in a long series of papers by Engle (1982), Bollerslev (1986), Ding et al. (1993), Glosten et al. (1993) and Bollerslev et al. (1994), among others. In the former an unobservable process describes the volatility at any time point while in the latter volatility is modelled as a function of observable information describing the past behaviour of the process; see also the review articles by Shephard (1996) and Andersen and Benzoni (2009). The generalized autoregressive score (GAS) models of Creal et al. (2013) generalize the observation-driven approach of GARCH models by using the score function of the conditional density to model time variation in key parameters of the time series model. The models of this paper have more in common with the observation-driven approach of GARCH and GAS but have some important differences.

In GARCH-type models, the marginal distribution of a stationary process is inextricably linked to the dynamics of the process as well as the conditional or innovation distribution; in most cases, it has no simple closed form. For example, the standard GARCH mechanism serves to create power-law behaviour in the marginal distribution, even when the innovations come from a lighter-tailed distribution such as Gaussian (Mikosch and

Stărică 2000). While such models work well for many return series, they may not be sufficiently flexible to describe all possible combinations of marginal and serial dependence behaviour encountered in applications. In the empirical example of this paper, which relates to log-returns on the Bitcoin price series, the data appear to favour a marginal distribution with sub-exponential tails that are lighter than power tails and this cannot be well captured by standard GARCH models. Moreover, in contrast to much of the GARCH literature, the models we propose make no assumptions about the existence of second-order moments and could also be applied to very heavy-tailed situations where variance-based methods fail.

Let $X_1, \ldots, X_n$ be a time series of financial returns sampled at (say) daily frequency and assume that these are modelled by a strictly stationary stochastic process (X_t) with marginal distribution function (cdf) F_X. To match the stylized facts of financial return data described, for example, by Campbell et al. (1997) and Cont (2001), it is generally agreed that (X_t) should have limited serial correlation, but the squared or absolute processes (X_t^2) and $(|X_t|)$ should have significant and persistent positive serial correlation to describe the effects of volatility clustering.

In this paper, we refer to transformed series like $(|X_t|)$, in which volatility is revealed through serial correlation, as *volatility proxy series*. More generally, a volatility proxy series $(T(X_t))$ is obtained by applying a transformation $T : \mathbb{R} \mapsto \mathbb{R}$ which (i) depends on a change point μ_T that may be zero, (ii) is increasing in $X_t - \mu_T$ for $X_t \geq \mu_T$ and (iii) is increasing in $\mu_T - X_t$ for $X_t \leq \mu_T$.

Our approach in this paper is to model the probability-integral transform (PIT) series (V_t) of a volatility proxy series. This is defined by $V_t = F_{T(X)}(T(X_t))$ for all t, where $F_{T(X)}$ denotes the cdf of $T(X_t)$. If (U_t) is the PIT series of the original process (X_t), defined by $U_t = F_X(X_t)$ for all t, then a *v-transform* is a function describing the relationship between the terms of (V_t) and the terms of (U_t). Equivalently, a v-transform describes the relationship between quantiles of the distribution of X_t and the distribution of the volatility proxy $T(X_t)$. Alternatively, it characterizes the dependence structure or copula of the pair of variables $(X_t, T(X_t))$. In this paper, we show how to derive flexible, parametric families of v-transforms for practical modelling purposes.

To gain insight into the typical form of a v-transform, let $x_1, \ldots, x_n$ represent the realized data values and let $u_1, \ldots, u_n$ and $v_1, \ldots, v_n$ be the samples obtained by applying the transformations $v_t = F_n^{(|X|)}(|x_t|)$ and $u_t = F_n^{(X)}(x_t)$, where $F_n^{(X)}(x) = \frac{1}{n+1}\sum_{t=1}^{n} I_{\{x_t \leq x\}}$ and $F_n^{(|X|)}(x) = \frac{1}{n+1}\sum_{t=1}^{n} I_{\{|x_t| \leq x\}}$ denote scaled versions of the empirical distribution functions of the x_t and $|x_t|$ samples, respectively. The graph of (u_t, v_t) gives an empirical estimate of the v-transform for the random variables $(X_t, |X_t|)$. In the left-hand plot of Figure 1 we show the relationship for a sample of $n = 1043$ daily log-returns of the Bitcoin price series for the years 2016–2019. Note how the empirical v-transform takes the form of a slightly asymmetric 'V'.

The right-hand plot of Figure 1 shows the sample autocorrelation function (acf) of the data given by $z_t = \Phi^{-1}(v_t)$ where Φ is the standard normal cdf. This reveals a persistent pattern of positive serial correlation which can be modelled by the implied ARMA copula. This pattern is not evident in the acf of the raw x_t data in the centre plot.

To construct a volatility model for (X_t) using v-transforms, we need to specify a process for (V_t). In principle, any model for a series of serially dependent uniform variables can be applied to (V_t). In this paper, we illustrate concepts using the Gaussian copula model implied by the standard ARMA dependence structure. This model is particularly tractable and allows us to derive model properties and fit models to data relatively easily.

There is a large literature on copula models for time series; see, for example, the review papers by Patton (2012) and Fan and Patton (2014). While the main focus of this literature has been on cross-sectional dependencies between series, there is a growing literature on models of serial dependence. First-order Markov copula models have been investigated by Chen and Fan (2006), Chen et al. (2009) and Domma et al. (2009) while higher-order Markov copula models using D-vines are applied by Smith et al. (2010). These models are based

on the pair-copula apporoach developed in Joe (1996), Bedford and Cooke (2001, 2002) and Aas et al. (2009). However, the standard bivariate copulas that enter these models are not generally effective at describing the typical serial dependencies created by stochastic volatility, as observed by Loaiza-Maya et al. (2018).

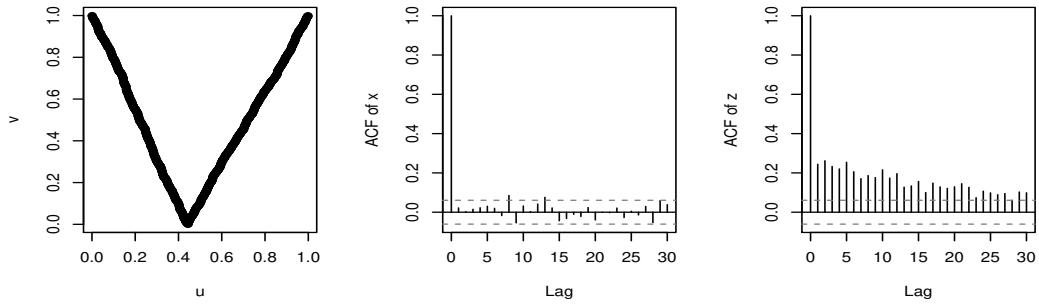

Figure 1. Scatterplot of v_t against u_t (left), sample acf of raw data x_t (centre) and sample acf of $z_t = \Phi^{-1}(v_t)$ (right). The transformed data are defined by $v_t = F_n^{(|X|)}(|x_t|)$ and $u_t = F_n^{(X)}(x_t)$ where $F_n^{(X)}$ and $F_n^{(|X|)}$ denote versions of the empirical distribution function of the x_t and $|x_t|$ values, respectively. The sample size is $n = 1043$ and the data are daily log-returns of the Bitcoin price for the years 2016–2019.

The paper is structured as follows. In Section 2, we provide motivation for the paper by constructing a symmetric model using the simplest example of a v-transform. The general theory of v-transforms is developed in Section 3 and is used to construct the class of VT-ARMA processes and analyse their properties in Section 4. Section 5 treats estimation and statistical inference for VT-ARMA processes and provides an example of their application to the Bitcoin return data; Section 6 presents the conclusions. Proofs may be found in the Appendix A.

2. A Motivating Model

Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we construct a symmetric, strictly stationary process $(X_t)_{t\in\mathbb{N}\setminus\{0\}}$ such that, under the even transformation $T(x) = |x|$, the serial dependence in the volatility proxy series $(T(X_t))$ is of ARMA type. We assume that the marginal cdf F_X of (X_t) is absolutely continuous and the density f_X satisfies $f_X(x) = f_X(-x)$ for all $x > 0$. Since F_X and $F_{|X|}$ are both continuous, the properties of the probability-integral (PIT) transform imply that the series (U_t) and (V_t) given by $U_t = F_X(X_t)$ and $V_t = F_{|X|}(|X_t|)$ both have standard uniform marginal distributions. Henceforth, we refer to (V_t) as the *volatility PIT process* and (U_t) as the *series PIT process*.

Any other volatility proxy series that can be obtained by a continuous and strictly increasing transformation of the terms of $(|X_t|)$, such as (X_t^2), yields exactly the same volatility PIT process. For example, if $\tilde{V}_t = F_{X^2}(X_t^2)$, then it follows from the fact that $F_{X^2}(x) = F_{|X|}(\sqrt[+]{x})$ for $x \geq 0$ that $\tilde{V}_t = F_{X^2}(X_t^2) = F_{|X|}(|X_t|) = V_t$. In this sense, we can think of classes of equivalent volatility proxies, such as $(|X_t|)$, (X_t^2), $(\exp|X_t|)$ and $(\ln(1+|X_t|))$. In fact, (V_t) is itself an equivalent volatility proxy to $(|X_t|)$ since $F_{|X|}$ is a continuous and strictly increasing transformation.

The symmetry of f_X implies that $F_{|X|}(x) = 2F_X(x) - 1 = 1 - 2F_X(-x)$ for $x \geq 0$. Hence, we find that

$$V_t = F_{|X|}(|X_t|) = \begin{cases} F_{|X|}(-X_t) = 1 - 2F_X(X_t) = 1 - 2U_t, & \text{if } X_t < 0 \\ F_{|X|}(X_t) = 2F_X(X_t) - 1 = 2U_t - 1, & \text{if } X_t \geq 0 \end{cases}$$

which implies that the relationship between the volatility PIT process (V_t) and the series PIT process (U_t) is given by

$$V_t = \mathcal{V}(U_t) = |2U_t - 1| \tag{1}$$

where $\mathcal{V}(u) = |2u - 1|$ is a perfectly symmetric v-shaped function that maps values of U_t close to 0 or 1 to values of V_t close to 1, and values close to 0.5 to values close to 0. $\mathcal{V}$ is the canonical example of a v-transform. It is related to the so-called tent-map transformation $\mathcal{T}(u) = 2\min(u, 1-u)$ by $\mathcal{V}(u) = 1 - \mathcal{T}(u)$.

Given (V_t), let the process (Z_t) be defined by setting $Z_t = \Phi^{-1}(V_t)$ so that we have the following chain of transformations:

$$X_t \xrightarrow{F_X} U_t \xrightarrow{\mathcal{V}} V_t \xrightarrow{\Phi^{-1}} Z_t \quad . \tag{2}$$

We refer to (Z_t) as a *normalized volatility proxy series*. Our aim is to construct a process (X_t) such that, under the chain of transformations in (2), we obtain a Gaussian ARMA process (Z_t) with mean zero and variance one. We do this by working back through the chain.

The transformation $\mathcal{V}$ is not an injection and, for any $V_t > 0$, there are two possible inverse values, $\frac{1}{2}(1 - V_t)$ and $\frac{1}{2}(1 + V_t)$. However, by randomly choosing between these values, we can 'stochastically invert' $\mathcal{V}$ to construct a random variable U_t such that $\mathcal{V}(U_t) = V_t$, This is summarized in Lemma 1, which is a special case of a more general result in Proposition 4.

Lemma 1. *Let V be a standard uniform variable. If $V = 0$, set $U = \frac{1}{2}$. Otherwise, let $U = \frac{1}{2}(1 - V)$ with probability 0.5 and $U = \frac{1}{2}(1 + V)$ with probability 0.5. Then, U is uniformly distributed and $\mathcal{V}(U) = V$.*

This simple result suggests Algorithm 1 for constructing a process (X_t) with symmetric marginal density f_X such that the corresponding normalized volatility proxy process (Z_t) under the absolute value transformation (or continuous and strictly increasing functions thereof) is an ARMA process. We describe the resulting model as a VT-ARMA process.

Algorithm 1:

1. Generate (Z_t) as a causal and invertible Gaussian ARMA process of order (p, q) with mean zero and variance one.
2. Form the volatility PIT process (V_t) where $V_t = \Phi(Z_t)$ for all t.
3. Generate a process of iid Bernoulli variables (Y_t) such that $\mathbb{P}(Y_t = 1) = 0.5$.
4. Form the PIT process (U_t) using the transformation $U_t = 0.5(1 - V_t)^{I_{\{Y_t=0\}}}(1 + V_t)^{I_{\{Y_t=1\}}}$.
5. Form the process (X_t) by setting $X_t = F_X^{-1}(U_t)$.

It is important to state that the use of the Gaussian process (Z_t) as the fundamental building block of the VT-ARMA process in Algorithm 1 has no effect on the marginal distribution of (X_t), which is F_X as specified in the final step of the algorithm. The process (Z_t) is exploited *only for its serial dependence structure*, which is described by a family of finite-dimensional Gaussian copulas; this dependence structure is applied to the volatility proxy process.

Figure 2 shows a symmetric VT-ARMA(1,1) process with ARMA parameters $\alpha_1 = 0.95$ and $\beta_1 = -0.85$; such a model often works well for financial return data. Some intuition for this observation can be gained from the fact that the popular GARCH(1,1) model is known to have the structure of an ARMA(1,1) model for the squared data process; see, for example, McNeil et al. (2015) (Section 4.2) for more details.

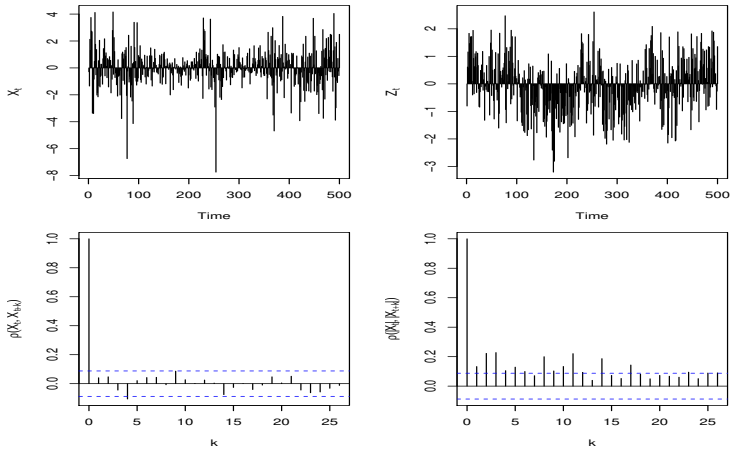

Figure 2. Realizations of length $n = 500$ of (X_t) and (Z_t) for a VT-ARMA(1,1) process with a marginal Student t distribution with $\nu = 3$ degrees of freedom and ARMA parameters $\alpha = 0.95$ and $\beta = -0.85$. ACF plots for (X_t) and $(|X_t|)$ are also shown.

3. V-Transforms

To generalize the class of v-transforms, we admit two forms of asymmetry in the construction described in Section 2: we allow the density f_X to be skewed; we introduce an asymmetric volatility proxy.

Definition 1 (Volatility proxy transformation and profile)**.** *Let T_1 and T_2 be strictly increasing, continuous and differentiable functions on $\mathbb{R}^+ = [0, \infty)$ such that $T_1(0) = T_2(0)$. Let $\mu_T \in \mathbb{R}$. Any transformation $T : \mathbb{R} \to \mathbb{R}$ of the form*

$$T(x) = \begin{cases} T_1(\mu_T - x) & x \leq \mu_T \\ T_2(x - \mu_T) & x > \mu_T \end{cases} \tag{3}$$

is a volatility proxy transformation. The parameter μ_T is the change point of T and the associated function $g_T : \mathbb{R}^+ \to \mathbb{R}^+$, $g_T(x) = T_2^{-1} \circ T_1(x)$ is the profile function of T.

By introducing μ_T, we allow for the possibility that the natural change point may not be identical to zero. By introducing different functions T_1 and T_2 for returns on either side of the change point, we allow the possibility that one or other may contribute more to the volatility proxy. This has a similar economic motivation to the *leverage* effects in GARCH models (Ding et al. 1993); falls in equity prices increase a firm's leverage and increase the volatility of the share price.

Clearly, the profile function of a volatility proxy transformation is a strictly increasing, continuous and differentiable function on $\mathbb{R}^+$ such that $g_T(x) = 0$. In conjunction with μ_T, the profile contains all the information about T that is relevant for constructing v-transforms. In the case of a volatility proxy transformation that is symmetric about μ_T, the profile satisfies $g_T(x) = x$.

The following result shows how v-transforms $V = \mathcal{V}(U)$ can be obtained by considering different continuous distributions F_X and different volatility proxy transformations T of type (3).

Proposition 1. *Let X be a random variable with absolutely continuous and strictly increasing cdf F_X on $\mathbb{R}$ and let T be a volatility proxy transformation. Let $U = F_X(X)$ and $V = F_{T(X)}(T(X))$. Then, $V = \mathcal{V}(U)$ where*

$$\mathcal{V}(u) = \begin{cases} F_X\Big(\mu_T + g_T\Big(\mu_T - F_X^{-1}(u)\Big)\Big) - u, & u \leq F_X(\mu_T) \\ u - F_X\Big(\mu_T - g_T^{-1}\Big(F_X^{-1}(u) - \mu_T\Big)\Big), & u > F_X(\mu_T)\,. \end{cases} \tag{4}$$

The result implies that any two volatility proxy transformations T and $\tilde{T}$ which have the same change point μ_T and profile function g_T belong to an equivalence class with respect to the resulting v-transform. This generalizes the idea that $T(x) = |x|$ and $T(x) = x^2$ give the same v-transform in the symmetric case of Section 2. Note also that the volatility proxy transformations $T^{(V)}$ and $T^{(Z)}$ defined by

$$\begin{aligned} T^{(V)}(x) &= F_{T(X)}(T(x)) = \mathcal{V}(F_X(x)) \\ T^{(Z)}(x) &= \Phi^{-1}(T^{(V)}(x)) = \Phi^{-1}\Big(\mathcal{V}(F_X(x))\Big) \end{aligned} \tag{5}$$

are in the same equivalence class as T since they share the same change point and profile function.

Definition 2 (v-transform and fulcrum)**.** *Any transformation $\mathcal{V}$ that can be obtained from Equation (4) by choosing an absolutely continuous and strictly increasing cdf F_X on $\mathbb{R}$ and a volatility proxy transformation T is a v-transform. The value $\delta = F_X(\mu_T)$ is the fulcrum of the v-transform.*

3.1. A Flexible Parametric Family

In this section, we derive a family of v-transforms using construction (4) by taking a tractable asymmetric model for F_X using the construction proposed by Fernández and Steel (1998) and by setting $\mu_T = 0$ and $g_T(x) = kx^\xi$ for $k > 0$ and $\xi > 0$. This profile function contains the identity profile $g_T(x) = x$ (corresponding to the symmetric volatility proxy transformation) as a special case, but allows cases where negative or positive returns contribute more to the volatility proxy. The choices we make may at first sight seem rather arbitrary, but the resulting family can in fact assume many of the shapes that are permissible for v-transforms, as we will argue.

Let f_0 be a density that is symmetric about the origin and let $\gamma > 0$ be a scalar parameter. Fernandez and Steel suggested the model

$$f_X(x;\gamma) = \begin{cases} \frac{2\gamma}{1+\gamma^2}\, f_0(\gamma x) & x \le 0 \\ \frac{2\gamma}{1+\gamma^2}\, f_0\left(\frac{x}{\gamma}\right) & x > 0\,. \end{cases} \tag{6}$$

This model is often used to obtain skewed normal and skewed Student distributions for use as innovation distributions in econometric models. A model with $\gamma > 1$ is skewed to the right while a model with $\gamma < 1$ is skewed to the left, as might be expected for asset returns. We consider the particular case of a Laplace or double exponential distribution $f_0(x) = 0.5\exp(-|x|)$ which leads to particularly tractable expressions.

Proposition 2. *Let $F_X(x;\gamma)$ be the cdf of the density (6) when $f_0(x) = 0.5\exp(-|x|)$. Set $\mu_T = 0$ and let $g_T(x) = kx^\xi$ for $k, \xi > 0$. The v-transform (4) is given by*

$$\mathcal{V}_{\delta,\kappa,\xi}(u) = \begin{cases} 1 - u - (1-\delta)\exp\Big(-\kappa\big(-\ln\big(\frac{u}{\delta}\big)\big)^\xi\Big) & u \le \delta, \\ u - \delta\exp\Big(-\kappa^{-1/\xi}\Big(-\ln\Big(\frac{1-u}{1-\delta}\Big)\Big)^{1/\xi}\Big) & u > \delta, \end{cases} \tag{7}$$

where $\delta = F_X(0) = (1+\gamma^2)^{-1} \in (0,1)$ and $\kappa = k/\gamma^{\xi+1} > 0$.

It is remarkable that (7) is a uniformity-preserving transformation. If we set $\xi = 1$ and $\kappa = 1$, we get

$$\mathcal{V}_\delta(u) = \begin{cases} (\delta - u)/\delta & u \le \delta, \\ (u-\delta)/(1-\delta) & u > \delta \end{cases} \tag{8}$$

which obviously includes the symmetric model $\mathcal{V}_{0.5}(u) = |2u - 1|$. The v-transform $\mathcal{V}_\delta(u)$ in (8) is a very convenient special case, and we refer to it as the *linear* v-transform.

In Figure 3, we show the v-transform $\mathcal{V}_{\delta,\kappa,\xi}$ when $\delta = 0.55$, $\kappa = 1.4$ and $\xi = 0.65$. We will use this particular v-transform to illustrate further properties of v-transforms and find a characterization.

3.2. Characterizing v-Transforms

It is easily verified that any v-transform obtained from (4) consists of two arms or branches, described by continuous and strictly monotonic functions; the left arm is decreasing and the right arm increasing. See Figure 3 for an illustration. At the fulcrum δ, we have $\mathcal{V}(\delta) = 0$. Every point $u \in [0,1] \setminus \{\delta\}$ has a *dual point* u^* on the opposite side of the fulcrum such that $\mathcal{V}(u^*) = \mathcal{V}(u)$. Dual points can be interpreted as the quantile probability levels of the distribution of X that give rise to the same level of volatility.

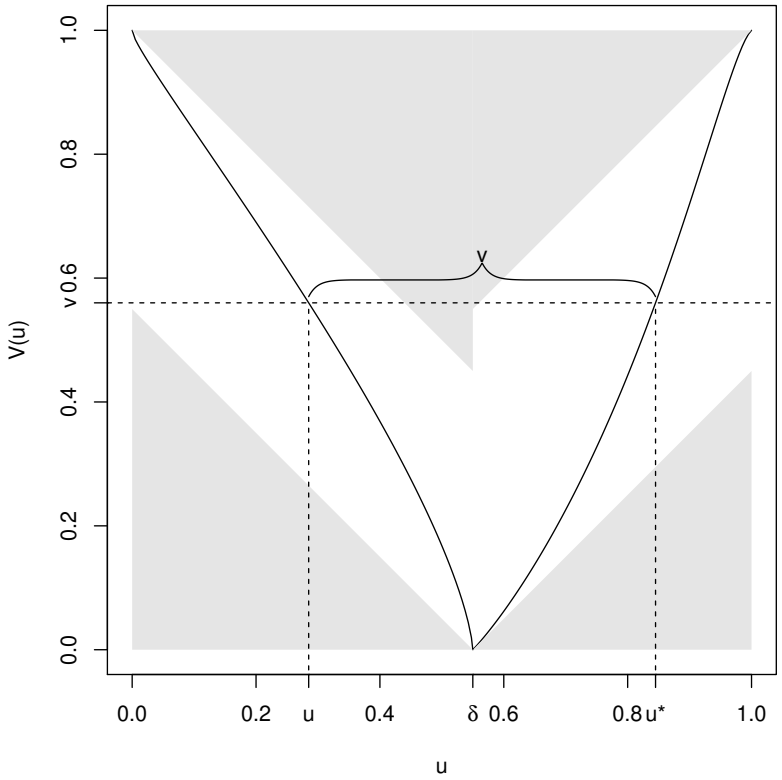

Figure 3. An asymmetric v-transform from the family defined in (7). For any v-transform, if $v = \mathcal{V}(u)$ and u^* is the dual of u, then the points $(u, 0)$, (u, v), $(u^*, 0)$ and (u^*, v) form the vertices of a square. For the given fulcrum δ, a v-transform can never enter the gray shaded area of the plot.

We collect these properties together in the following lemma and add one further important property that we refer to as the *square property* of a v-transform; this property places constraints on the shape that v-transforms can take and is illustrated in Figure 3.

Lemma 2. *A v-transform is a mapping $\mathcal{V} : [0,1] \to [0,1]$ with the following properties:*

1. $\mathcal{V}(0) = \mathcal{V}(1) = 1$;
2. *There exists a point δ known as the fulcrum such that $0 < \delta < 1$ and $\mathcal{V}(\delta) = 0$;*
3. *$\mathcal{V}$ is continuous;*
4. *$\mathcal{V}$ is strictly decreasing on $[0, \delta]$ and strictly increasing on $[\delta, 1]$;*

5. *Every point $u \in [0,1] \setminus \{\delta\}$ has a dual point u^* on the opposite side of the fulcrum satisfying $\mathcal{V}(u) = \mathcal{V}(u^*)$ and $|u^* - u| = \mathcal{V}(u)$ (square property).*

It is instructive to see why the square property must hold. Consider Figure 3 and fix a point $u \in [0,1] \setminus \{\delta\}$ with $\mathcal{V}(u) = v$. Let $U \sim U(0,1)$ and let $V = \mathcal{V}(U)$. The events $\{V \le v\}$ and $\{\min(u,u^*) \le U \le \max(u,u^*)\}$ are the same and hence the uniformity of V under a v-transform implies that

$$v = \mathbb{P}(V \le v) = \mathbb{P}(\min(u,u^*) \le U \le \max(u,u^*)) = |u^* - u|\,. \tag{9}$$

The properties in Lemma 2 could be taken as the basis of an alternative definition of a v-transform. In view of (9), it is clear that any mapping $\mathcal{V}$ that has these properties is a uniformity-preserving transformation. We can characterize the mappings $\mathcal{V}$ that have these properties as follows.

Theorem 1. *A mapping $\mathcal{V} : [0,1] \to [0,1]$ has the properties listed in Lemma 2 if and only if it takes the form*

$$\mathcal{V}(u) = \begin{cases} (1-u) - (1-\delta)\Psi\left(\frac{u}{\delta}\right) & u \le \delta, \\ u - \delta\Psi^{-1}\left(\frac{1-u}{1-\delta}\right) & u > \delta, \end{cases} \tag{10}$$

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

Our arguments so far show that every v-transform must have the form (10). It remains to verify that every uniformity-preserving transformation of the form (10) can be obtained from construction (4), and this is the purpose of the final result of this section. This allows us to view Definition 2, Lemma 2, and the characterization (10) as three equivalent approaches to the definition of v-transforms.

Proposition 3. *Let $\mathcal{V}$ be a uniformity-preserving transformation of the form (10) and F_X a continuous distribution function. Then, $\mathcal{V}$ can be obtained from construction (4) using any volatility proxy transformation with change point $\mu_T = F_X^{-1}(\delta)$ and profile*

$$g_T(x) = F_X^{-1}(F_X(\mu_T - x) + \mathcal{V}(F_X(\mu_T - x))) - \mu_T, \quad x \ge 0. \tag{11}$$

Henceforth, we can view (10) as the general equation of a v-transform. Distribution functions Ψ on $[0,1]$ can be thought of as *generators* of v-transforms. Comparing (10) with (7), we see that our parametric family $\mathcal{V}_{\delta,\kappa,\xi}$ is generated by $\Psi(x) = \exp(-\kappa(-(\ln x)^{\xi}))$. This is a 2-parameter distribution whose density can assume many different shapes on the unit interval including increasing, decreasing, unimodal, and bathtub-shaped forms. In this respect, it is quite similar to the beta distribution which would yield an alternative family of v-transforms. The uniform distribution function $\Psi(x) = x$ gives the family of linear v-transforms $\mathcal{V}_\delta$.

In applications, we construct models starting from the building blocks of a tractable v-transform $\mathcal{V}$ such as (7) and a distribution F_X; from these, we can always infer an implied profile function g_T using (11). The alternative approach of starting from g_T and F_X and constructing $\mathcal{V}$ via (4) is also possible but can lead to v-transforms that are cumbersome and computationally expensive to evaluate if F_X and its inverse do not have simple closed forms.

3.3. V-Transforms and Copulas

If two uniform random variables are linked by the v-transform $V = \mathcal{V}(U)$, then the joint distribution function of (U,V) is a special kind of copula. In this section, we derive the form of the copula, which facilitates the construction of stochastic processes using v-transforms.

To state the main result, we use the notation $\mathcal{V}^{-1}$ and $\mathcal{V}'$ for the the inverse function and the gradient function of a v-transform $\mathcal{V}$. Although there is no unique inverse $\mathcal{V}^{-1}(v)$

(except when $v = 0$), the fact that the two branches of a v-transform mutually determine each other allows us to define $\mathcal{V}^{-1}(v)$ to be the inverse of the left branch of the v-transform given by $\mathcal{V}^{-1} : [0,1] \to [0,\delta]$, $\mathcal{V}^{-1}(v) = \inf\{u : \mathcal{V}(u) = v\}$. The gradient $\mathcal{V}'(u)$ is defined for all points $u \in [0,1] \setminus \{\delta\}$, and we adopt the convention that $\mathcal{V}'(\delta)$ is the left derivative as $u \to \delta$.

Theorem 2. *Let V and U be random variables related by the v-transform $V = \mathcal{V}(U)$.*

1. *The joint distribution function of (U, V) is given by the copula*

$$C(u,v) = \mathbb{P}(U \le u, V \le v) = \begin{cases} 0 & u < \mathcal{V}^{-1}(v) \\ u - \mathcal{V}^{-1}(v) & \mathcal{V}^{-1}(v) \le u < \mathcal{V}^{-1}(v) + v \\ v & u \ge \mathcal{V}^{-1}(v) + v\,. \end{cases} \tag{12}$$

2. *Conditional on $V = v$, the distribution of U is given by*

$$U = \begin{cases} \mathcal{V}^{-1}(v) & \text{with probability } \Delta(v) \text{ if } v \neq 0 \\ \mathcal{V}^{-1}(v) + v & \text{with probability } 1 - \Delta(v) \text{ if } v \neq 0 \\ \delta & \text{if } v = 0 \end{cases} \tag{13}$$

where

$$\Delta(v) = -\frac{1}{\mathcal{V}'(\mathcal{V}^{-1}(v))}\,. \tag{14}$$

3. $\mathbb{E}(\Delta(V)) = \delta$.

Remark 1. *In the case of the symmetric v-transform $\mathcal{V}(u) = |1 - 2u|$, the copula in (12) takes the form $C(u,v) = \max(\min(u + \frac{v}{2} - \frac{1}{2}, v), 0)$. We note that this copula is related to a special case of the tent map copula family $C_\theta^{\mathcal{T}}$ in Rémillard (2013) by $C(u,v) = u - C_1^{\mathcal{T}}(u, 1-v)$.*

For the linear v-transform family, the conditional probability $\Delta(v)$ in (14) satisfies $\Delta(v) = \delta$. This implies that the value of V contains no information about whether U is likely to be below or above the fulcrum; the probability is always the same regardless of V. In general, this is not the case and the value of V does contain information about whether U is large or small.

Part (2) of Theorem 2 is the key to stochastically inverting a v-transform in the general case. Based on this result, we define the concept of stochastic inversion of a v-transform. We refer to the function Δ as the *conditional down probability* of $\mathcal{V}$.

Definition 3 (Stochastic inversion function of a v-transform)**.** *Let $\mathcal{V}$ be a v-transform with conditional down probability Δ. The two-place function $\boldsymbol{\mathcal{V}}^{-1} : [0,1] \times [0,1] \to [0,1]$ defined by*

$$\boldsymbol{\mathcal{V}}^{-1}(v,w) = \begin{cases} \mathcal{V}^{-1}(v) & \text{if } w \le \Delta(v) \\ v + \mathcal{V}^{-1}(v) & \text{if } w > \Delta(v). \end{cases} \tag{15}$$

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

The following proposition, which generalizes Lemma 1, allows us to construct general asymmetric processes that generalize the process of Algorithm 1.

Proposition 4. *Let V and W be iid $U(0,1)$ variables and let $\mathcal{V}$ be a v-transform with stochastic inversion function $\boldsymbol{\mathcal{V}}$. If $U = \boldsymbol{\mathcal{V}}^{-1}(V,W)$, then $\mathcal{V}(U) = V$ and $U \sim U(0,1)$.*

In Section 4, we apply v-transforms and their stochastic inverses to the terms of time series models. To understand the effect this has on the serial dependencies between

random variables, we need to consider multivariate componentwise v-transforms of random vectors with uniform marginal distributions and these can also be represented in terms of copulas. We now give a result which forms the basis for the analysis of serial dependence properties. The first part of the result shows the relationship between copula densities under componentwise v-transforms. The second part shows the relationship under the componentwise stochastic inversion of a v-transform; in this case, we assume that the stochastic inversion of each term takes place independently given $\boldsymbol{V}$ so that all serial dependence comes from $\boldsymbol{V}$.

Theorem 3. *Let $\mathcal{V}$ be a v-transform and let $\boldsymbol{U} = (U_1, \dots, U_d)'$ and $\boldsymbol{V} = (V_1, \dots, V_d)'$ be vectors of uniform random variables with copula densities $c_{\boldsymbol{U}}$ and $c_{\boldsymbol{V}}$, respectively.*

1. *If $\boldsymbol{V} = (\mathcal{V}(U_1), \dots, \mathcal{V}(U_d))'$, then*

$$c_{\boldsymbol{V}}(v_1, \dots, v_d) = \sum_{j_1=1}^{2} \cdots \sum_{j_d=1}^{2} c_{\boldsymbol{U}}(u_{1j_1}, \dots, u_{dj_d}) \prod_{i=1}^{d} \Delta(v_i)^{I_{\{j_i=1\}}} (1 - \Delta(v_i))^{I_{\{j_i=2\}}} \tag{16}$$

   where $u_{i1} = \mathcal{V}^{-1}(v_i)$ and $u_{i2} = \mathcal{V}^{-1}(v_i) + v_i$ for all $i \in \{1, \dots, d\}$.
2. *If $\boldsymbol{U} = (\boldsymbol{\mathcal{V}}^{-1}(V_1, W_1), \dots, \boldsymbol{\mathcal{V}}^{-1}(V_d, W_d))'$ where $W_1, \dots, W_d$ are iid uniform random variables that are also independent of $V_1, \dots, V_d$, then*

$$c_{\boldsymbol{U}}(u_1, \dots, u_d) = c_{\boldsymbol{V}}(\mathcal{V}(u_1), \dots, \mathcal{V}(u_d)). \tag{17}$$

4. VT-ARMA Copula Models

In this section, we study some properties of the class of time series models obtained by the following algorithm, which generalizes Algorithm 1. The models obtained are described as VT-ARMA processes since they are stationary time series constructed using the fundamental building blocks of a v-transform $\mathcal{V}$ and an ARMA process.

We can add any marginal behaviour in the final step, and this allows for an infinitely rich choice. We can, for instance, even impose an infinite-variance or an infinite-mean distribution, such as the Cauchy distribution, and still obtain a strictly stationary process for (X_t). We make the following definitions.

Definition 4 (VT-ARMA and VT-ARMA copula process)**.** *Any stochastic process (X_t) that can be generated using Algorithm 2 by choosing an underlying ARMA process with mean zero and variance one, a v-transform $\mathcal{V}$, and a continuous distribution function F_X is a VT-ARMA process. The process (U_t) obtained at the penultimate step of the algorithm is a VT-ARMA copula process.*

Algorithm 2:

1. Generate (Z_t) as a causal and invertible Gaussian ARMA process of order (p, q) with mean zero and variance one.
2. Form the volatility PIT process (V_t) where $V_t = \Phi(Z_t)$ for all t.
3. Generate iid $U(0, 1)$ random variables (W_t).
4. Form the series PIT process (U_t) by taking the stochastic inverses $U_t = \boldsymbol{\mathcal{V}}^{-1}(V_t, W_t)$.
5. Form the process (X_t) by setting $X_t = F_X^{-1}(U_t)$ for some continuous cdf F_X.

Figure 4 gives an example of a simulated process using Algorithm 2 and the v-transform $\mathcal{V}_{\delta,\kappa,\xi}$ in (7) with $\kappa = 0.9$ and MA parameter $\xi = 1.1$. The marginal distribution is a heavy-tailed skewed Student distribution of type (6) with degrees-of-freedom $\nu = 3$ and skewness $\gamma = 0.8$, which gives rise to more large negative returns than large positive

returns. The underlying time series model is an ARMA(1,1) model with AR parameter $\alpha = 0.95$ and MA parameter $\beta = -0.85$. See the caption of the figure for full details of parameters.

In the remainder of this section, we concentrate on the properties of VT-ARMA copula processes (U_t) from which related properties of VT-ARMA processes (X_t) may be easily inferred.

4.1. Stationary Distribution

The VT-ARMA copula process (U_t) of Definition 4 is a strictly stationary process since the joint distribution of $(U_{t_1}, \dots, U_{t_k})$ for any set of indices $t_1 < \cdots < t_k$ is invariant under time shifts. This property follows easily from the strict stationarity of the underlying ARMA process (Z_t) according to the following result, which uses Theorem 3.

Proposition 5. *Let (U_t) follow a VT-ARMA copula process with v-transform $\mathcal{V}$ and an underlying ARMA(p,q) structure with autocorrelation function $\rho(k)$. The random vector $(U_{t_1}, \dots, U_{t_k})$ for $k \in \mathbb{N}$ has joint density $c^{Ga}_{P(t_1,\dots,t_k)}(\mathcal{V}(u_1), \dots, \mathcal{V}(u_k))$, where $c^{Ga}_{P(t_1,\dots,t_k)}$ denotes the density of the Gaussian copula $C^{Ga}_{P(t_1,\dots,t_k)}$ and $P(t_1, \dots, t_k)$ is a correlation matrix with (i, j) element given by $\rho(|t_j - t_i|)$.*

An expression for the joint density facilitates the calculation of a number of dependence measures for the bivariate marginal distribution of (U_t, U_{t+k}). In the bivariate case, the correlation matrix of the underlying Gaussian copula $C^{\text{Ga}}_{P(t,t+k)}$ contains a single off-diagonal value $\rho(k)$ and we simply write $C^{\text{Ga}}_{\rho(k)}$. The Pearson correlation of (U_t, U_{t+k}) is given by

$$\rho(U_t, U_{t+k}) = 12 \int_0^1 \int_0^1 u_1 u_2 c^{\text{Ga}}_{\rho(k)}(\mathcal{V}(u_1), \mathcal{V}(u_2)) \mathrm{d}u_1 \mathrm{d}u_2 - 3 \,. \tag{18}$$

This value is also the value of the Spearman rank correlation $\rho_S(X_t, X_{t+k})$ for a VT-ARMA process (X_t) with copula process (U_t) (since the Spearman's rank correlation of a pair of continuous random variables is the Pearson correlation of their copula). The calculation of (18) typically requires numerical integration. However, in the special case of the linear v-transform $\mathcal{V}_\delta$ in (8), we can get a simpler expression as shown in the following result.

Proposition 6. *Let (U_t) be a VT-ARMA copula process satisfying the assumptions of Proposition 5 with linear v-transform $\mathcal{V}_\delta$. Let (Z_t) denote the underlying Gaussian ARMA process. Then,*

$$\rho(U_t, U_{t+k}) \quad = \quad (2\delta - 1)^2 \rho_S(Z_t, Z_{t+k}) = \frac{6(2\delta - 1)^2 \arcsin\left(\frac{\rho(k)}{2}\right)}{\pi} \,. \tag{19}$$

For the symmetric v-transform $\mathcal{V}_{0.5}$, Equation (19) obviously yields a correlation of zero so that, in this case, the VT-ARMA copula process (U_t) is a white noise with an autocorrelation function that is zero, except at lag zero. However, even a very asymmetric model with $\delta = 0.4$ or $\delta = 0.6$ gives $\rho(U_t, U_{t+k}) = 0.04 \rho_S(Z_t, Z_{t+k})$ so that serial correlations tend to be very weak.

When we add a marginal distribution, the resulting process (X_t) has a different autocorrelation function to (U_t), but the same rank autocorrelation function. The symmetric model of Section 2 is a white noise process. General asymmetric processes (X_t) are not perfect white noise processes but have only very weak serial correlation.

4.2. Conditional Distribution

To derive the conditional distribution of a VT-ARMA copula process, we use the vector notation $\boldsymbol{U}_t = (U_1, \dots, U_t)'$ and $\boldsymbol{Z}_t = (Z_1, \dots, Z_t)'$ to denote the history of processes up to

time point t and $\boldsymbol{u}_t$ and $\boldsymbol{z}_t$ for realizations. These vectors are related by the componentwise transformation $\boldsymbol{Z}_t = \Phi^{-1}(\mathcal{V}(\boldsymbol{U}_t))$. We assume that all processes have a time index set given by $t \in \{1, 2, \ldots\}$.

Proposition 7. *For $t > 1$, the conditional density $f_{U_t|\boldsymbol{U}_{t-1}}(u \mid \boldsymbol{u}_{t-1})$ is given by*

$$f_{U_t|\boldsymbol{U}_{t-1}}(u \mid \boldsymbol{u}_{t-1}) = \frac{\phi\left(\frac{\Phi^{-1}(\mathcal{V}(u))-\mu_t}{\sigma_\epsilon}\right)}{\sigma_\epsilon \phi(\Phi^{-1}(\mathcal{V}(u)))} \tag{20}$$

where $\mu_t = \mathbb{E}(Z_t \mid \boldsymbol{Z}_{t-1} = \Phi^{-1}(\mathcal{V}(\boldsymbol{u}_{t-1})))$ and σ_ϵ is the standard deviation of the innovation process for the ARMA model followed by (Z_t).

When (Z_t) is iid white noise $\mu_t = 0$, $\sigma_\epsilon = 1$ and (20) reduce to the uniform density $f_{U_t|\boldsymbol{U}_{t-1}}(u \mid \boldsymbol{u}_{t-1}) = 1$ as expected. In the case of the first-order Markov AR(1) model $Z_t = \alpha_1 Z_{t-1} + \epsilon_t$, the conditional mean of Z_t is $\mu_t = \alpha_1 \Phi^{-1}(\mathcal{V}(u_{t-1}))$ and $\sigma_\epsilon^2 = 1 - \alpha_1^2$. The conditional density (20) can be easily shown to simplify to $f_{U_t|U_{t-1}}(u \mid u_{t-1}) = c^{\mathrm{Ga}}_{\alpha_1}(\mathcal{V}(u), \mathcal{V}(u_{t-1}))$ where $c^{\mathrm{Ga}}_{\alpha_1}(\mathcal{V}(u_1), \mathcal{V}(u_2))$ denotes the copula density derived in Proposition 5. In this special case, the VT-ARMA model falls within the class of first-order Markov copula models considered by Chen and Fan (2006), although the copula is new.

If we add a marginal distribution F_X to the VT-ARMA copula model to obtain a model for (X_t) and use similar notational conventions as above, the resulting VT-ARMA model has conditional density

$$f_{X_t|\boldsymbol{X}_{t-1}}(x \mid \boldsymbol{x}_{t-1}) = f_X(x) f_{U_t|\boldsymbol{U}_{t-1}}(F_X(x) \mid F_X(\boldsymbol{x}_{t-1})) \tag{21}$$

with $f_{U_t|\boldsymbol{U}_{t-1}}$ as in (20). An interesting property of the VT-ARMA process is that the conditional density (21) can have a pronounced bimodality for values of μ_t in excess of zero that is in high volatility situations where the conditional mean of Z_t is higher than the marginal mean value of zero; in low volatility situations, the conditional density appears more concentrated around zero. This phenomenon is illustrated in Figure 4. The bimodality in high volatility situations makes sense: in such cases, it is likely that the next return will be large in absolute value and relatively less likely that it will be close to zero.

The conditional distribution function of (X_t) is $F_{X_t|\boldsymbol{X}_{t-1}}(x \mid \boldsymbol{x}_{t-1}) = F_{U_t|\boldsymbol{U}_{t-1}}(F_X(x) \mid F_X(\boldsymbol{x}_{t-1}))$ and hence the ψ-quantile $x_{\psi,t}$ of $F_{X_t|\boldsymbol{X}_{t-1}}$ can be obtained by solving

$$\psi = F_{U_t|\boldsymbol{U}_{t-1}}(F_X(x_{\psi,t}) \mid F_X(\boldsymbol{x}_{t-1}))\,. \tag{22}$$

For $\psi < 0.5$, the negative of this value is often referred to as the conditional $(1-\psi)$-VaR (value-at-risk) at time t in financial applications.

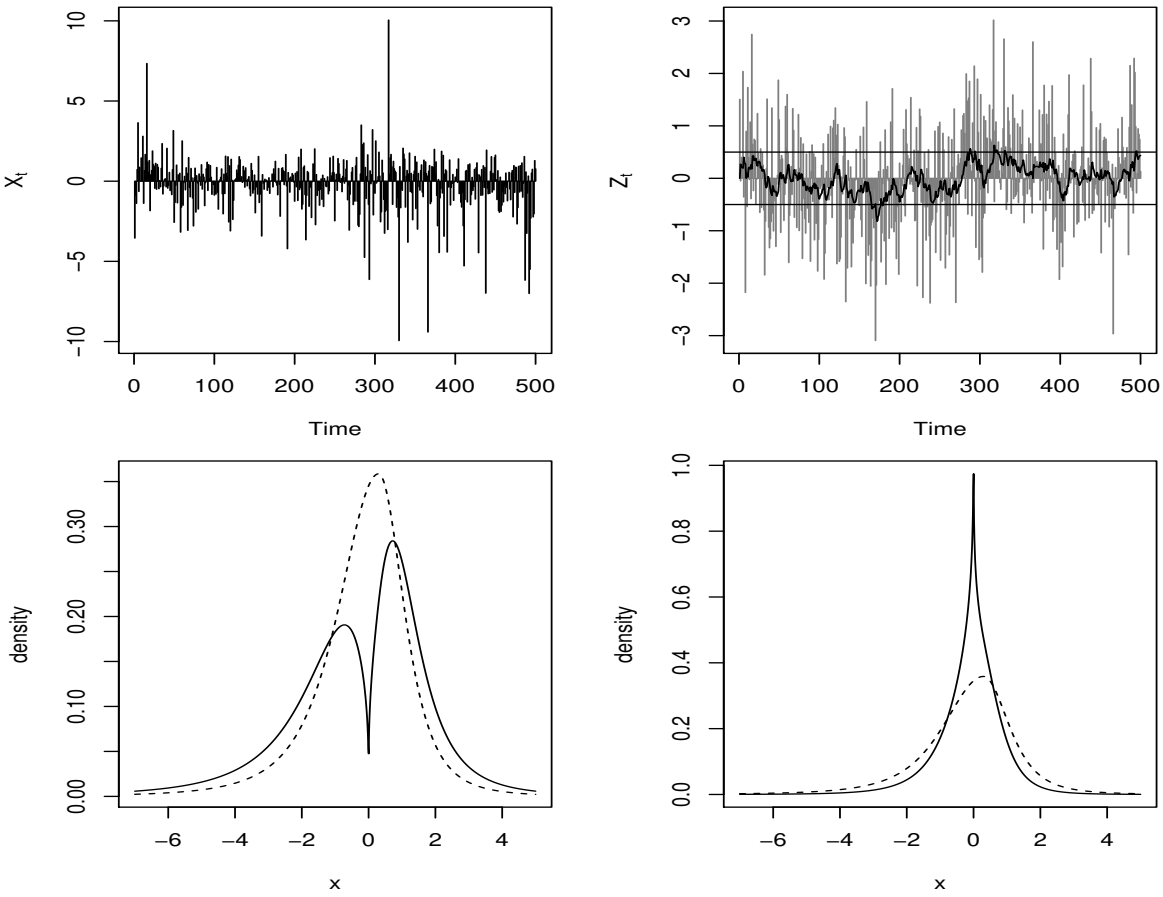

Figure 4. Top left: realization of length $n = 500$ of (X_t) for a process with a marginal skewed Student distribution (parameters: $\nu = 3$, $\gamma = 0.8$, $\mu = 0.3$, $\sigma = 1$) a v-transform of the form (7) (parameters: $\delta = 0.50$, $\kappa = 0.9$, $\xi = 1.1$) and an underlying ARMA process ($\alpha = 0.95$, $\beta = -0.85$, $\sigma_\epsilon = 0.95$). Top right: the underlying ARMA process (Z_t) in gray with the conditional mean (μ_t) superimposed in black; horizontal lines at $\mu_t = 0.5$ (a high value) and $\mu_t = -0.5$ (a low value). The corresponding conditional densities are shown in the bottom figures with the marginal density as a dashed line.

5. Statistical Inference

In the copula approach to dependence modelling, the copula is the object of central interest and marginal distributions are often of secondary importance. A number of different approaches to estimation are found in the literature. As before, let $x_1, \ldots, x_n$ represent realizations of variables $X_1, \ldots, X_n$ from the time series process (X_t).

The semi-parametric approach developed by Genest et al. (1995) is very widely used in copula inference and has been applied by Chen and Fan (2006) to first-order Markov copula models in the time series context. In this approach, the marginal distribution F_X is first estimated non-parametrically using the scaled empirical distribution function $F_n^{(X)}$ (see definition in Section 1) and the data are transformed onto the $(0,1)$ scale. This has the effect of creating pseudo-copula data $u_t = \text{rank}(x_t)/(n+1)$ where $\text{rank}(x_t)$ denotes the rank of x_t within the sample. The copula is fitted to the pseudo-copula data by maximum likelihood (ML).

As an alternative, the inference-functions-for-margins (IFM) approach of Joe (2015) could be applied. This is also a two-step method although in this case a parametric model $\widehat{F}_X$ is estimated under an iid assumption in the first step and the copula is fitted to the data $u_t = \widehat{F}_X(x_t)$ in the second step.

The approach we adopt for our empirical example is to first use the semi-parametric approach to determine a reasonable copula process, then to estimate marginal parameters

under an iid assumption, and finally to estimate all parameters jointly using the parameter estimates from the previous steps as starting values.

We concentrate on the mechanics of deriving maximum likelihood estimates (MLEs). The problem of establishing the asymptotic properties of the MLEs in our setting is a difficult one. It is similar to, but appears to be more technically challenging than, the problem of showing consistency and efficiency of MLEs for a Box-Cox-transformed Gaussian ARMA process, as discussed in Terasaka and Hosoya (2007). We are also working with a componentwise transformed ARMA process, although, in our case, the transformation $(X_t) \to (Z_t)$ is via the nonlinear, non-increasing volatility proxy transformation $T^{(Z)}(x)$ in (5), which is not differentiable at the change point μ_T. We have, however, run extensive simulations which suggests good behaviour of the MLEs in large samples.

5.1. Maximum Likelihood Estimation of the VT-ARMA Copula Process

We first consider the estimation of the VT-ARMA copula process for a sample of data $u_1, \ldots, u_n$. Let $\boldsymbol{\theta}^{(V)}$ and $\boldsymbol{\theta}^{(A)}$ denote the parameters of the v-transform and ARMA model, respectively. It follows from Theorem 3 (part 2) and Proposition 5 that the log-likelihood for the sample $u_1, \ldots, u_n$ is simply the log density of the Gaussian copula under componentwise inverse v-transformation. This is given by

$$L(\boldsymbol{\theta}^{(V)}, \boldsymbol{\theta}^{(A)} \mid u_1, \ldots, u_n) = L^*(\boldsymbol{\theta}^{(A)} \mid \Phi^{-1}(\mathcal{V}_{\boldsymbol{\theta}^{(V)}}(u_1)), \ldots, \Phi^{-1}(\mathcal{V}_{\boldsymbol{\theta}^{(V)}}(u_n))) - \sum_{t=1}^{n} \ln \phi\Big(\Phi^{-1}(\mathcal{V}_{\boldsymbol{\theta}^{(V)}}(u_t))\Big) \tag{23}$$

where the first term L^* is the log-likelihood for an ARMA model with a standard N(0,1) marginal distribution. Both terms in the log-likelihood (23) are relatively straightforward to evaluate.

The evaluation of the ARMA likelihood $L^*(\boldsymbol{\theta}^{(A)} \mid z_1, \ldots, z_n)$ for parameters $\boldsymbol{\theta}^{(A)}$ and data $z_1, \ldots, z_n$ can be accomplished using the Kalman filter. However, it is important to note that the assumption that the data $z_1, \ldots, z_n$ are standard normal requires a bespoke implementation of the Kalman filter, since standard software always treats the error variance σ_ϵ^2 as a free parameter in the ARMA model. In our case, we need to constrain σ_ϵ^2 to be a function of the ARMA parameters so that $\mathrm{var}(Z_t) = 1$. For example, in the case of an ARMA(1,1) model with AR parameter α_1 and MA parameter β_1, this means that $\sigma_\epsilon^2 = \sigma_\epsilon^2(\alpha_1, \beta_1) = (1 - \alpha_1^2)/(1 + 2\alpha_1\beta_1 + \beta_1^2)$. The constraint on σ_ϵ^2 must be incorporated into the state-space representation of the ARMA model.

Model validation tests for the VT-ARMA copula can be based on residuals

$$r_t = z_t - \widehat{\mu}_t, \quad z_t = \Phi^{-1}(\mathcal{V}_{\widehat{\boldsymbol{\theta}}^{(V)}}(u_t)) \tag{24}$$

where z_t denotes the implied realization of the normalized volatility proxy variable and where an estimate $\widehat{\mu}_t$ of the conditional mean $\mu_t = \mathbb{E}(Z_t \mid \boldsymbol{Z}_{t-1} = \boldsymbol{z}_t)$ may be obtained as an output of the Kalman filter. The residuals should behave like an iid sample from a normal distribution.

Using the estimated model, it is also possible to implement a likelihood-ratio (LR) test for the presence of stochastic volatility in the data. Under the null hypothesis that $\boldsymbol{\theta}^{(A)} = \mathbf{0}$, the log-likelihood (23) is identically equal to zero. Thus, the size of the maximized log-likelihood $L(\widehat{\boldsymbol{\theta}}^{(V)}, \widehat{\boldsymbol{\theta}}^{(A)}; u_1, \ldots, u_n)$ provides a measure of the evidence for the presence of stochastic volatility.

5.2. Adding a Marginal Model

If F_X and f_X denote the cdf and density of the marginal model and the parameters are denoted $\boldsymbol{\theta}^{(M)}$, then the full log-likelihood for the data $x_1, \ldots, x_n$ is simply

$$L^{\text{full}}(\boldsymbol{\theta} \mid x_1, \ldots, x_n) = \sum_{t=1}^{n} \ln f_X(x_t\,;\, \boldsymbol{\theta}^{(M)}) + L\Big(\boldsymbol{\theta}^{(V)}, \boldsymbol{\theta}^{(A)} \mid F_X(x_1\,;\, \boldsymbol{\theta}^{(M)}), \ldots, F_X(x_n\,;\, \boldsymbol{\theta}^{(M)})\Big) \quad (25)$$

where the first term is the log-likelihood for a sample of iid data from the marginal distribution F_X and the second term is (23).

When a marginal model is added, we can recover the implied form of the volatility proxy transformation using Proposition 3. If $\widehat{\delta}$ is the estimated fulcrum parameter of the v-transform, then the estimated change point is $\widehat{\mu}_T = F_X^{-1}(\widehat{\delta}; \widehat{\boldsymbol{\theta}}^{(M)})$ and the implied profile function is

$$\widehat{g}_T(x) \quad = \quad \widehat{F}_X^{-1}\Big(\widehat{F}_X(\widehat{\mu}_T - x) - \mathcal{V}_{\widehat{\boldsymbol{\theta}}^{(V)}}\Big(\widehat{F}_X(\widehat{\mu}_T - x)\Big)\Big) - \widehat{\mu}_T. \quad (26)$$

Note that is is possible to force the change point to be zero in a joint estimation of marginal model and copula by imposing the constraint $F_X(0; \boldsymbol{\theta}^{(M)}) = \delta$ on the fulcrum and marginal parameters during the optimization. However, in our experience, superior fits are obtained when these parameters are unconstrained.

5.3. Example

We analyse $n = 1043$ daily log-returns for the Bitcoin price series for the period 2016–2019; values are multiplied by 100. We first apply the semi-parametric approach of Genest et al. (1995) using the log-likelihood (23) which yields the results in Table 1. Different models are referred to by VT(n)-ARMA(p, q), where (p, q) refers to the ARMA model and n indexes the v-transform: 1 is the linear v-transform $\mathcal{V}_\delta$ in (8); 3 is the three-parameter transform $\mathcal{V}_{\delta,\kappa,\xi}$ in (7); 2 is the two-parameter v-transform given by $\mathcal{V}_{\delta,\kappa} := \mathcal{V}_{\delta,\kappa,1}$. In unreported analyses, we also tried the three-parameter family based on the beta distribution, but this had negligible effect on the results.

The column marked L gives the value of the maximized log-likelihood. All values are large and positive showing strong evidence of stochastic volatility in all cases. The model VT(1)-ARMA(1,0) is a first-order Markov model with linear v-transform. The fit of this model is noticeably poorer than the others suggesting that Markov models are insufficient to capture the persistence of stochastic volatility in the data. The column marked SW contains the p-value for a Shapiro–Wilks test of normality applied to the residuals from the VT-ARMA copula model; the result is non-significant in all cases.

Table 1. Analysis of daily Bitcoin return data 2016–2019. Parameter estimates, standard errors (below estimates) and information about the fit: SW denotes Shapiro–Wilks p-value; L is the maximized value of the log-likelihood and AIC is the Akaike information criterion.

Model	α_1	β_1	δ	κ	ξ	**SW**	L	**AIC**
VT(1)-ARMA(1,0)	0.283		0.460			0.515	37.59	−71.17
	0.026		0.001					
VT(1)-ARMA(1,1)	0.962	−0.840	0.416			0.197	92.91	−179.81
	0.012	0.028	0.004					
VT(2)-ARMA(1,1)	0.965	−0.847	0.463	0.920		0.385	94.73	−181.45
	0.011	0.026	0.001	0.131				
VT(3)-ARMA(1,1)	0.962	−0.839	0.463	0.881	0.995	0.407	94.82	−179.64
	0.012	0.028	0.001	0.123	0.154			

According to the AIC values, the VT(2)-ARMA(1,1) is the best model. We experimented with higher order ARMA processes, but this did not lead to further significant improvements. Figure 5 provides a visual of the fit of this model. The pictures in the panels show the QQplot of the residuals against normal, acf plots of the residuals and squared residuals and the estimated conditional mean process $(\widehat{\mu}_t)$, which can be taken as an indicator of high and low volatility periods. The residuals and absolute residuals show very little evidence of serial correlation and the QQplot is relatively linear, suggesting that the ARMA filter has been successful in explaining much of the serial dependence structure of the normalized volatility proxy process.

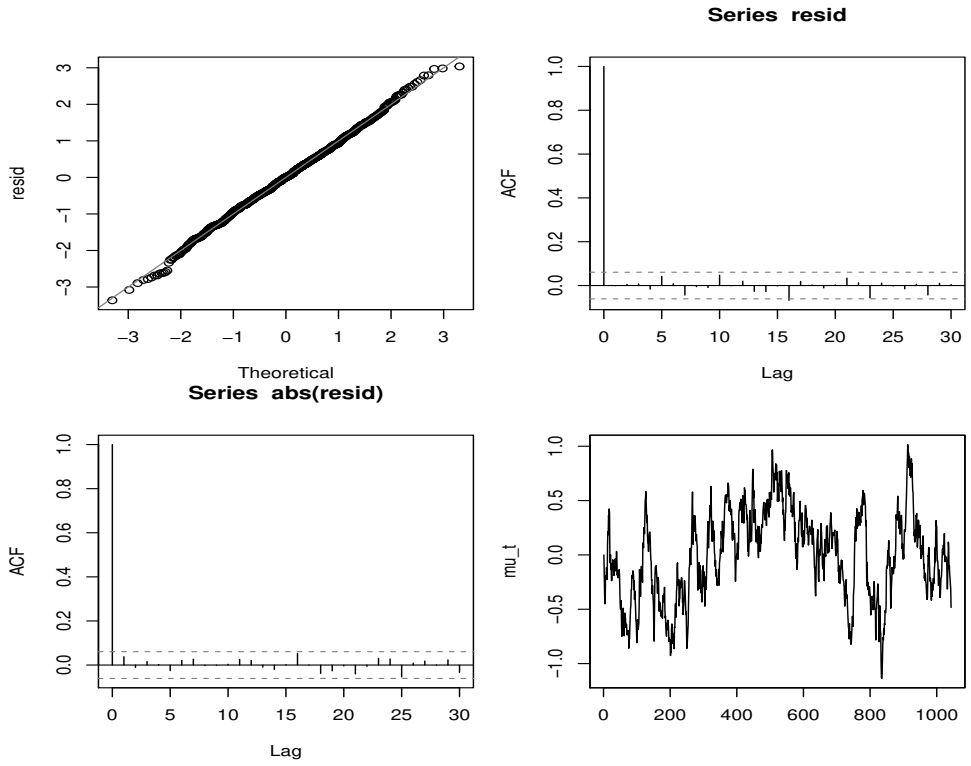

Figure 5. Plots for a VT(2)-ARMA(1,1) model fitted to the Bitcoin return data: QQplot of the residuals against normal (**upper left**); acf of the residuals (**upper right**); acf of the absolute residuals (**lower left**); estimated conditional mean process (μ_t) (**lower right**).

We now add various marginal distributions to the VT(2)-ARMA(1,1) copula model and estimate all parameters of the model jointly. We have experimented with a number of location-scale families including Student-t, Laplace (double exponential), and a double-Weibull family which generalizes the Laplace distribution and is constructed by taking back-to-back Weibull distributions. Estimation results are presented for these three distributions in Table 2. All three marginal distributions are symmetric around their location parameters μ, and no improvement is obtained by adding skewness using the construction of Fernández and Steel (1998) described in Section 3.1; in fact, the Bitcoin returns in this time period show a remarkable degree of symmetry. In the table, the shape and scale parameters of the distributions are denoted η and σ, respectively; in the case of Student, an infinite-variance distribution with degree-of-freedom parameter $\eta = 1.94$ is fitted, but this model is inferior to the models with Laplace and double-Weibull margins; the latter is the favoured model on the basis of AIC values.

Figure 6 shows some aspects of the joint fit for the fully parametric VT(2)-ARMA(1,1) model with double-Weibull margin. A QQplot of the data against the fitted marginal distribution confirms that the double-Weibull is a good marginal model for these data. Although this distribution is sub-exponential (heavier-tailed than exponential), its tails

do not follow a power law and it is in the maximum domain of attraction of the Gumbel distribution (see, for example, McNeil et al. 2015, Chapter 5).

Table 2. VT(2)-ARMA(1,1) model with three different margins: Student-t, Laplace, double Weibull. Parameter estimates, standard errors (alongside estimates) and information about the fit: SW denotes Shapiro–Wilks p-value; L is the maximized value of the log-likelihood and AIC is the Akaike information criterion.

	Student		**Laplace**		**dWeibull**	
α_1	0.954	0.012	0.953	0.012	0.965	0.021
β_1	−0.842	0.026	−0.847	0.025	−0.847	0.035
δ	0.478	0.001	0.480	0.002	0.463	0.000
κ	0.790	0.118	0.811	0.129	0.939	0.138
η	1.941	0.005			0.844	0.022
μ	0.319	0.002	0.315	0.002	0.192	0.001
σ	2.427	0.003	3.194	0.004	2.803	0.214
SW	0.585		0.551		0.376	
L	−2801.696		−2791.999		−2779.950	
AIC	5617.392		5595.999		5573.899	

Using (26), the implied volatility proxy profile function $\widehat{g}_T$ can be constructed and is found to lie just below the line $y = x$ as shown in the upper-right panel. The change point is estimated to be $\widehat{\mu}_T = 0.06$. We can also estimate an implied volatility proxy transformation in the equivalence class defined by $\widehat{g}_T$ and $\widehat{\mu}_T$. We estimate the transformation $T = T^{(Z)}$ in (5) by taking $\widehat{T}(x) = \Phi^{-1}(\mathcal{V}_{\widehat{\boldsymbol{\theta}}^{(V)}}(F_X(x; \widehat{\boldsymbol{\theta}}^{(M)})))$. In the lower-left panel of Figure 6, we show the empirical v-transform formed from the data $(x_t, \widehat{T}(x_t))$ together with the fitted parametric v-transform $\mathcal{V}_{\widehat{\boldsymbol{\theta}}^{(V)}}$. We recall from Section 1 that the empirical v-transform is the plot (u_t, v_t) where $u_t = F_n^{(X)}(x_t)$ and $v_t = F_n^{(\widehat{T}(X))}(\widehat{T}(x_t))$. The empirical v-transform and the fitted parametric v-transform show a good degree of correspondence. The lower-right panel of Figure 6 shows the volatility proxy transformation $\widehat{T}(x)$ as a function of x superimposed on the points $(x_t, \Phi^{-1}(v_t))$. Using the curve, we can compare the effects of, for example, a log-return ($\times$ 100) of -10 and a log-return of 10. For the fitted model, these are 1.55 and 1.66 showing that the up movement is associated with slightly higher volatility.

As a comparison to the VT-ARMA model, we fit standard GARCH(1,1) models using Student-t and generalized error distributions for the innovations; these are standard choices available in the popular `rugarch` package in R. The generalized error distribution (GED) contains normal and Laplace as special cases as well as a model that has a similar tail behaviour to Weibull; note, however, that, by the theory of Mikosch and Stărică (2000), the tails of the marginal distribution of the GARCH decay according to a power law in both cases. The results in Table 3 show that the VT(2)-ARMA(1,1) models with Laplace and double-Weibull marginal distributions outperform both GARCH models in terms of AIC values.

Table 3. Comparison of three VT(2)-ARMA(1,1) models with different marginal distributions with two GARCH(1,1) models with different innovation distributions.

	Parameters	**AIC**
VT-ARMA (Student)	7	5617.39
VT-ARMA (Laplace)	6	5596.00
VT-ARMA (dWeibull)	7	5573.90
GARCH (Student)	5	5629.02
GARCH (GED)	5	5611.53

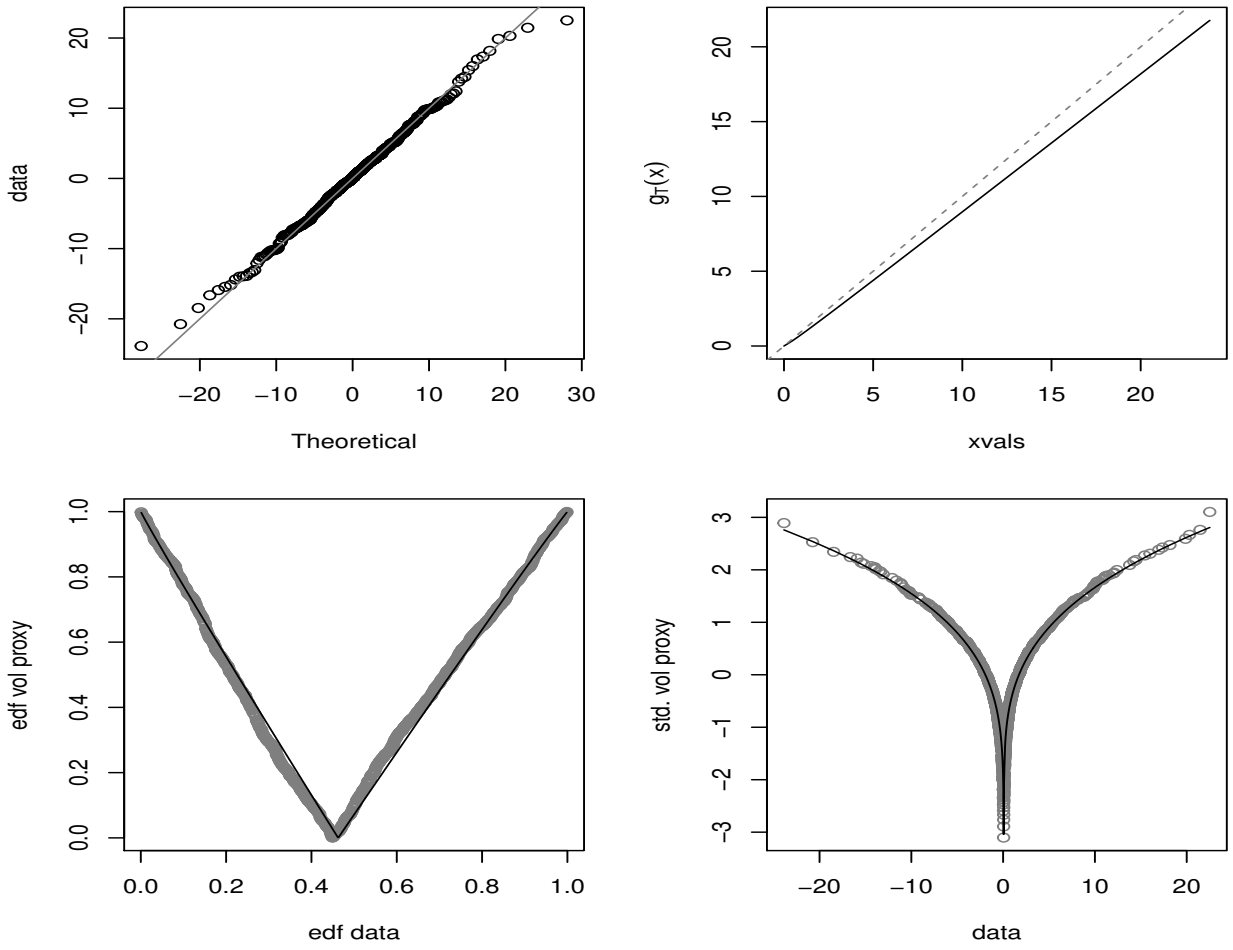

Figure 6. Plots for a VT(2)-ARMA(1,1) model combined with a double Weibull marginal distribution fitted to the Bitcoin return data: QQplot of the data against fitted double Weibull model (**upper left**); estimated volatility proxy profile function g_T (**upper right**); estimated v-transform (**lower left**); implied relationship between data and volatility proxy variable (**lower right**).

Figure 7 shows the in-sample 95% conditional value-at-risk (VaR) estimate based on the VT(2)-ARMA(1,1) model which has been calculated using (22). For comparison, a dashed line shows the corresponding estimate for the GARCH(1,1) model with GED innovations.

Finally, we carry out an out-of-sample comparison of conditional VaR estimates using the same two models. In this analysis, the models are estimated daily throughout the 2016–2019 period using a 1000-day moving data window and one-step-ahead VaR forecasts are calculated. The VT-ARMA model gives 47 exceptions of the 95% VaR and 11 exceptions of the 99% VaR, compared with expected numbers of 52 and 10 for a 1043 day sample, while the GARCH model leads to 57 and 12 exceptions; both models pass binomial tests for these exception counts. In a follow-up paper (Bladt and McNeil 2020), we conduct more extensive out-of-sample backtests for models using v-transforms and copula processes and show that they rival and often outperform forecast models from the extended GARCH family.

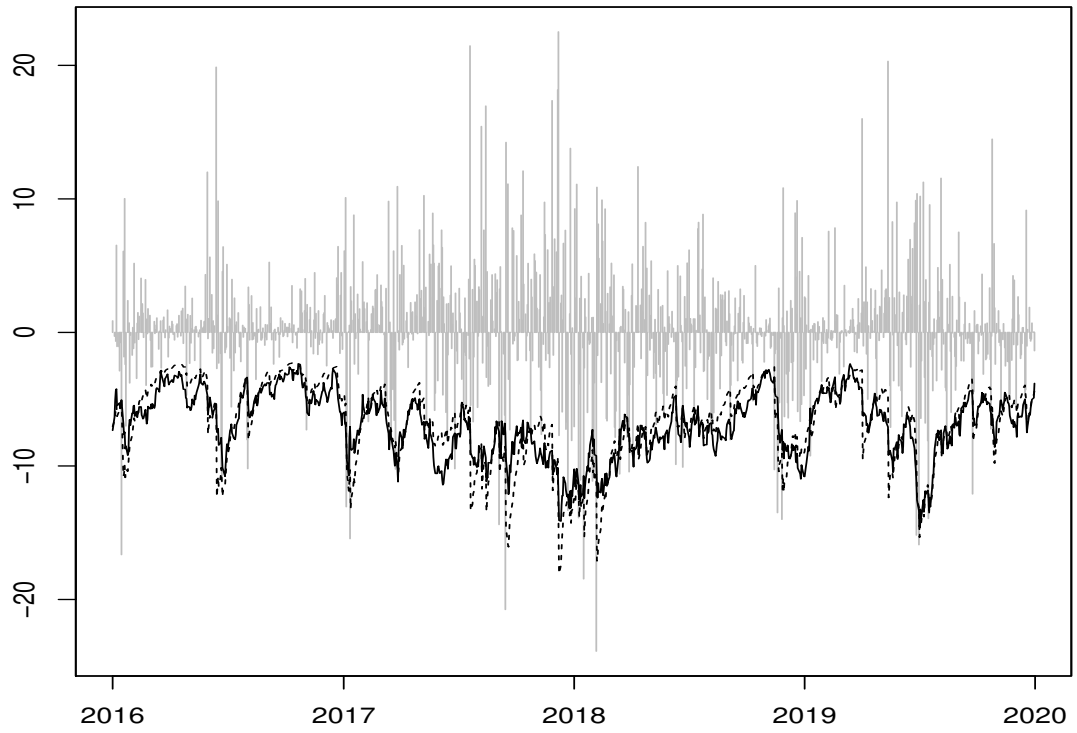

Figure 7. Plot of estimated 95% value-at-risk (VaR) for Bitcoin return data superimposed on log returns. Solid line shows VaR estimated using the VT(2)-ARMA(1,1) model combined with a double-Weibull marginal distribution; the dashed line shows VaR estimated using a GARCH(1,1) model with GED innovation distribution.

6. Conclusions

This paper has proposed a new approach to volatile financial time series in which v-transforms are used to describe the relationship between quantiles of the return distribution and quantiles of the distribution of a predictable volatility proxy variable. We have characterized v-transforms mathematically and shown that the stochastic inverse of a v-transform may be used to construct stationary models for return series where arbitrary marginal distributions may be coupled with dynamic copula models for the serial dependence in the volatility proxy.

The construction was illustrated using the serial dependence model implied by a Gaussian ARMA process. The resulting class of VT-ARMA processes is able to capture the important features of financial return series including near-zero serial correlation (white noise behaviour) and volatility clustering. Moreover, the models are relatively straightforward to estimate building on the classical maximum-likelihood estimation of an ARMA model using the Kalman filter. This can be accomplished in the stepwise manner that is typical in copula modelling or through joint modelling of the marginal and copula process. The resulting models yield insights into the way that volatility responds to returns of different magnitude and sign and can give estimates of unconditional and conditional quantiles (VaR) for practical risk measurement purposes.

There are many possible uses for VT-ARMA copula processes. Because we have complete control over the marginal distribution, they are very natural candidates for the innovation distribution in other time series models. For example, they could be applied to the innovations of an ARMA model to obtain ARMA models with VT-ARMA errors; this

might be particularly appropriate for longer interval returns, such as weekly or monthly returns, where some serial dependence is likely to be present in the raw return data.

Clearly, we could use other copula processes for the volatility PIT process (V_t). The VT-ARMA copula process has some limitations: the radial symmetry of the underlying Gaussian copula means that the serial dependence between large values of the volatility proxy must mirror the serial dependence between small values; moreover, this copula does not admit tail dependence in either tail and it seems plausible that very large values of the volatility proxy might have a tendency to occur in succession.

To extend the class of models based on v-transforms, we can look for models for the volatility PIT process (V_t) with higher dimensional marginal distributions given by asymmetric copulas with upper tail dependence. First-order Markov copula models as developed in Chen and Fan (2006) can give asymmetry and tail dependence, but they cannot model the dependencies at longer lags that we find in empirical data. D-vine copula models can model higher-order Markov dependencies and Bladt and McNeil (2020) show that this is a promising alternative specification for the volatility PIT process.

Funding: This research received no external funding.

Data Availability Statement: The analyses were carried out using R 4.0.2 (R Core Team, 2020) and the `tscopula` package (Alexander J. McNeil and Martin Bladt, 2020) available at https://github.com/ajmcneil/tscopula. The full reproducible code and the data are available at https://github.com/ajmcneil/vtarma.

Acknowledgments: The author is grateful for valuable input from a number of researchers including Hansjoerg Albrecher, Martin Bladt, Valérie Chavez-Demoulin, Alexandra Dias, Christian Genest, Michael Gordy, Yen Hsiao Lok, Johanna Nešlehová, Andrew Patton, and Ruodu Wang. Particular thanks are due to Martin Bladt for providing the Bitcoin data and advice on the data analysis. The paper was completed while the author was a guest at the Forschungsinstitut für Mathematik (FIM) at ETH Zurich.

Conflicts of Interest: The author declares no conflict of interest.

Appendix A. Proofs

Appendix A.1. Proof of Proposition 1

We observe that, for $x \geq 0$,

$$F_{T(X)}(x) = \mathbb{P}(\mu_T - T_1^{-1}(x) \leq X_t \leq \mu_T + T_2^{-1}(x)) = F_X(\mu_T + T_2^{-1}(x)) - F_X(\mu_T - T_1^{-1}(x)).$$

$\{X_t \leq \mu_T\} \iff \{U \leq F_X(\mu_T)\}$ and in this case

$$\begin{aligned} V = F_{T(X)}(T(X_t)) = F_{T(X)}(T_1(\mu_T - X_t)) &= F_X(\mu_T + T_2^{-1}(T_1(\mu_T - X_t))) - F_X(X_t) \\ &= F_X\Big(\mu_T + g_T\Big(\mu_T - F_X^{-1}(U)\Big)\Big) - U. \end{aligned}$$

$\{X_t > \mu_T\} \iff \{U > F_X(\mu_T)\}$ and in this case

$$\begin{aligned} V = F_{T(X)}(T(X_t)) = F_{T(X)}(T_2(X_t - \mu_T)) &= F_X(X_t) - F_X(\mu_T - T_1^{-1}(T_2(X_t - \mu_T))) \\ &= U - F_X\Big(\mu_T - g_T^{-1}\Big(F_X^{-1}(U) - \mu_T\Big)\Big). \end{aligned}$$

Appendix A.2. Proof of Proposition 2

The cumulative distribution function $F_0(x)$ of the double exponential distribution is equal to $0.5e^x$ for $x \leq 0$ and $1 - 0.5e^{-x}$ if $x > 0$. It is straightforward to verify that

$$F_X(x;\gamma) = \begin{cases} \delta e^{\gamma x} & x \leq 0 \\ 1 - (1-\delta)e^{-\frac{x}{\gamma}} & x > 0 \end{cases} \quad \text{and} \quad F_X^{-1}(u;\gamma) = \begin{cases} \frac{1}{\gamma}\ln\left(\frac{u}{\delta}\right) & u \leq \delta \\ -\gamma\ln\left(\frac{1-u}{1-\delta}\right) & u > \delta. \end{cases}$$

When $g_T(x) = kx^\xi$, we obtain for $u \le \delta$ that

$$\mathcal{V}_{\delta,\kappa,\xi}(u) = F_X\left(\frac{k}{\gamma^\xi}\left(\ln\left(\frac{\delta}{u}\right)^\xi\right);\gamma\right) - u = 1 - u - (1-\delta)\exp\left(-\frac{k}{\gamma^{\xi+1}}\left(-\ln\left(\frac{u}{\delta}\right)\right)^\xi\right).$$

For $u > \delta$, we make a similar calculation.

Appendix A.3. Proof of Theorem 1

It is easy to check that Equation (10) fulfills the list of properties in Lemma 2. We concentrate on showing that a function that has these properties must be of the form (10). It helps to consider the picture of a v-transform in Figure 3. Consider the lines $v = 1 - u$ and $v = \delta - u$ for $u \in [0, \delta]$. The areas above the former and below the latter are shaded gray.

The left branch of the v-transform must start at $(0, 1)$, end at $(\delta, 0)$, and lie strictly between these lines in $(0, \delta)$. Suppose, on the contrary, that $v = \mathcal{V}(u) \le \delta - u$ for $u \in (0, \delta)$. This would imply that the dual point u^* given by $u^* = u + v$ satisfies $u^* \le \delta$ which contradicts the requirement that u^* must be on the opposite side of the fulcrum. Similarly, if $v = \mathcal{V}(u) \ge 1 - u$ for $u \in (0, \delta)$, then $u^* \ge 1$ and this is also not possible; if $u^* = 1$, then $u = 0$, which is a contradiction.

Thus, the curve that links $(0, 1)$ and $(\delta, 0)$ must take the form

$$\mathcal{V}(u) = (\delta - u)\Psi\left(\frac{u}{\delta}\right) + (1-u)\left(1 - \Psi\left(\frac{u}{\delta}\right)\right) = (1-u) - (1-\delta)\Psi\left(\frac{u}{\delta}\right)$$

where $\Psi(0) = 0$, $\Psi(1) = 1$ and $0 < \Psi(x) < 1$ for $x \in (0, 1)$. Clearly, Ψ must be continuous to satisfy the conditions of the v-transform. It must also be strictly increasing. If it were not, then the derivative would satisfy $\mathcal{V}'(u) \ge -1$, which is not possible: if at any point $u \in (0, \delta)$, we have $\mathcal{V}'(u) = -1$, then the opposite branch of the v-transform would have to jump vertically at the dual point u^*, contradicting continuity; if $\mathcal{V}'(u) > -1$, then $\mathcal{V}$ would have to be a decreasing function at u^*, which is also a contradiction.

Thus, Ψ fulfills the conditions of a continuous, strictly increasing distribution function on $[0, 1]$, and we have established the necessary form for the left branch equation. To find the value of the right branch equation at $u > \delta$, we invoke the square property. Since $\mathcal{V}(u) = \mathcal{V}(u^*) = \mathcal{V}(u - \mathcal{V}(u))$, we need to solve the equation $x = \mathcal{V}(u - x)$ for $x \in [0, 1]$ using the formula for the left branch equation of $\mathcal{V}$. Thus, we solve $x = 1 - u + x - (1-\delta)\Psi(\frac{u-x}{\delta})$ for x, and this yields the right branch equation as asserted.

Appendix A.4. Proof of Proposition 3

Let $g_T(x)$ be as given in (11) and let $u(x) = F_X(\mu_T - x)$. For $x \in \mathbb{R}^+$, $u(x)$ is a continuous, strictly decreasing function of x starting at $u(0) = \delta$ and decreasing to 0. Since Ψ is a cumulative distribution function, it follows that

$$u^*(x) = u(x) + \mathcal{V}(u(x)) = 1 - (1-\delta)\Psi\left(\frac{u(x)}{\delta}\right)$$

is a continuous, strictly increasing function starting at $u^*(0) = \delta$ and increasing to 1. Hence, $g_T(x) = F_X^{-1}(u^*(x)) - \mu_T$ is continuous and strictly increasing on $\mathbb{R}^+$ with $g_T(0) = 0$ as required of the profile function of a volatility proxy transformation. It remains to check that, if we insert (11) in (4), we recover $\mathcal{V}(u)$, which is straightforward.

Appendix A.5. Proof of Theorem 2

1. For any $0 \le v \le 1$, the event $\{U \le u, V \le v\}$ has zero probability for $u < \mathcal{V}^{-1}(v)$. For $u \ge \mathcal{V}^{-1}(v)$, we have

$$\{U \le u, V \le v\} = \{\mathcal{V}^{-1}(v) \le U \le \min(u, \mathcal{V}^{-1}(v) + v)\}$$

and hence $\mathbb{P}(U \le u, V \le v) = \min(u, \mathcal{V}^{-1}(v) + v) - \mathcal{V}^{-1}(v)$ and (12) follows.

2. We can write $\mathbb{P}(U \le u, V \le v) = C(u,v)$, where C is the copula given by (12). It follows from the basic properties of a copula that

$$\mathbb{P}(U \le u, V = v) = \frac{\mathrm{d}}{\mathrm{d}v} C(u,v) = \begin{cases} 0 & u < \mathcal{V}^{-1}(v) \\ -\frac{\mathrm{d}}{\mathrm{d}v}\mathcal{V}^{-1}(v) & \mathcal{V}^{-1}(v) \le u < \mathcal{V}^{-1}(v) + v \\ 1 & u \ge \mathcal{V}^{-1}(v) + v \end{cases}$$

This is the distribution function of a binomial distribution, and it must be the case that $\Delta(v) = -\frac{\mathrm{d}}{\mathrm{d}v}\mathcal{V}^{-1}(v)$. Equation (14) follows by differentiating the inverse.

3. Finally, $\mathbb{E}(\Delta(V)) = \delta$ is easily verified by making the substitution $x = \mathcal{V}^{-1}(v)$ in the integral $\mathbb{E}(\Delta(V)) = -\int_0^1 \frac{1}{\mathcal{V}'(\mathcal{V}^{-1}(v))}\mathrm{d}v$.

Appendix A.6. Proof of Proposition 4

It is obviously true that $\mathcal{V}(\boldsymbol{\mathcal{V}}^{-1}(v, W)) = v$ for any W. Hence, $\mathcal{V}(U) = \mathcal{V}(\boldsymbol{\mathcal{V}}^{-1}(V, W)) = V$. The uniformity of U follows from the fact that

$$\mathbb{P}\Big(\boldsymbol{\mathcal{V}}^{-1}(V, W) = \mathcal{V}^{-1}(v) \mid V = v\Big) = \mathbb{P}(W \le \Delta(v) \mid V = v) = \mathbb{P}(W \le \Delta(v)) = \Delta(v)\,.$$

Hence, the pair of random variables (U, V) has the conditional distribution (13) and is distributed according to the copula C in (12).

Appendix A.7. Proof of Theorem 3

1. Since the event $\{V_i \le v_i\}$ is equal to the event $\{\mathcal{V}^{-1}(v_i) \le U_i \le \mathcal{V}^{-1}(v_i) + v_i\}$, we first compute the probability of a box $[a_1, b_1] \times \cdots \times [a_d, b_d]$ where $a_i = \mathcal{V}^{-1}(v_i) \le \mathcal{V}^{-1}(v_i) + v_i = b_i$. The standard formula for such probabilities implies that the copulas $C_{\boldsymbol{V}}$ and $C_{\boldsymbol{U}}$ are related by

$$C_{\boldsymbol{V}}(v_1, \ldots, v_d) = \sum_{j_1=1}^{2} \cdots \sum_{j_d=1}^{2} (-1)^{j_1+\cdots+j_d} C_{\boldsymbol{U}}(u_{1j_1}, \ldots, u_{dj_d})\,;$$

see, for example, McNeil et al. (2015), p. 221. Thus, the copula densities are related by

$$c_{\boldsymbol{V}}(v_1, \ldots, v_d) = \sum_{j_1=1}^{2} \cdots \sum_{j_d=1}^{2} c_{\boldsymbol{U}}(u_{1j_1}, \ldots, u_{dj_d}) \prod_{i=1}^{d} \frac{\mathrm{d}}{\mathrm{d}v_i}(-1)^{j_i} u_{ij_i},$$

and the result follows if we use (14) to calculate that

$$\frac{\mathrm{d}}{\mathrm{d}v_i}(-1)^j u_{ij} = \begin{cases} \frac{\mathrm{d}}{\mathrm{d}v_i}(-\mathcal{V}^{-1}(v_i)) = \Delta(v_i) & \text{if } j = 1, \\ \frac{\mathrm{d}}{\mathrm{d}v_i}(v_i + \mathcal{V}^{-1}(v_i)) = 1 - \Delta(v_i) & \text{if } j = 2. \end{cases}$$

2. For the point $(u_1, \ldots, u_d) \in [0,1]^d$, we consider the set of events $A_i(u_i)$ defined by

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

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

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

The event $A_i(u_i)$ can be written

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

and hence we can use Theorem 2 to write

$$\mathbb{P}\left(\bigcap_{i=1}^{d} A_i(u_i)\right) = \int_{\mathcal{V}(u_1)}^{1} \cdots \int_{\mathcal{V}(u_d)}^{1} c_{\boldsymbol{V}}(v_1, \ldots, v_d) \prod_{i=1}^{d} \Delta(v_i)^{I_{\{u_i \leq \delta\}}} (1 - \Delta(v_i))^{I_{\{u_i > \delta\}}} \mathrm{d}v_1 \cdots \mathrm{d}v_d \,.$$

The derivative is given by

$$\frac{\mathrm{d}^d}{\mathrm{d}u_1 \cdots \mathrm{d}u_d} P\left(\bigcap_{i=1}^{d} A_i(u_i)\right) = (-1)^d c_{\boldsymbol{V}}(\mathcal{V}(u_1), \ldots, \mathcal{V}(u_d)) \prod_{i=1}^{d} p(u_i)^{I_{\{u_i \leq \delta\}}} (1 - p(u_i))^{I_{\{u_i > \delta\}}} \mathcal{V}'(u_i)$$

where $p(u_i) = \Delta(\mathcal{V}(u_i))$ and hence we obtain

$$c_{\boldsymbol{U}}(u_1, \ldots, u_d) = c_{\boldsymbol{V}}(\mathcal{V}(u_1), \ldots, \mathcal{V}(u_d)) \prod_{i=1}^{d} (-p(u_i))^{I_{\{u_i \leq \delta\}}} (1 - p(u_i))^{I_{\{u_i > \delta\}}} \mathcal{V}'(u_i).$$

It remains to verify that each of the terms in the product is identically equal to 1. For $u_i \leq \delta$, this follows easily from (14) since $-p(u_i) = -\Delta(\mathcal{V}(u_i)) = 1/\mathcal{V}'(u_i)$. For $u_i > \delta$, we need an expression for the derivative of the right branch equation. Since $\mathcal{V}(u_i) = \mathcal{V}(u_i - \mathcal{V}(u_i))$, we obtain

$$\mathcal{V}'(u_i) = \mathcal{V}'(u_i - \mathcal{V}(u_i))(1 - \mathcal{V}'(u_i)) = \mathcal{V}'(u_i^*)(1 - \mathcal{V}'(u_i)) \Longrightarrow \mathcal{V}'(u_i) = \frac{\mathcal{V}'(u_i^*)}{1 + \mathcal{V}'(u_i^*)}$$

implying that

$$1 - p(u_i) = 1 - \Delta(\mathcal{V}(u_i)) = 1 - \Delta(\mathcal{V}(u_i^*)) = 1 + \frac{1}{\mathcal{V}'(u_i^*)} = \frac{1 + \mathcal{V}'(u_i^*)}{\mathcal{V}'(u_i^*)} = \frac{1}{\mathcal{V}'(u_i)} \,.$$

Appendix A.8. Proof of Proposition 5

Let $V_t = \mathcal{V}(U_t)$ and $Z_t = \Phi^{-1}(V_t)$ as usual. The process (Z_t) is an ARMA process with acf $\rho(k)$ and hence $(Z_{t_1}, \ldots, Z_{t_k})$ are jointly standard normally distributed with correlation matrix $P(t_1, \ldots, t_k)$. This implies that the joint distribution function of $(V_{t_1}, \ldots, V_{t_k})$ is the Gaussian copula with density $c^{\mathrm{Ga}}_{P(t_1,\ldots,t_k)}$ and hence by Part 2 of Theorem 3 the joint distribution function of $(U_{t_1}, \ldots, U_{t_k})$ is the copula with density $c^{\mathrm{Ga}}_{P(t_1,\ldots,t_k)}(\mathcal{V}(u_1), \ldots, \mathcal{V}(u_k))$.

Appendix A.9. Proof of Proposition 6

We split the integral in (18) into four parts. First, observe that, by making the substitutions $v_1 = \mathcal{V}(u_1) = 1 - u_1/\delta$ and $v_2 = \mathcal{V}(u_2) = 1 - u_2/\delta$ on $[0, \delta] \times [0, \delta]$, we get

$$\begin{aligned} \int_0^\delta \int_0^\delta u_1 u_2 c^{\mathrm{Ga}}_{\rho(k)}(\mathcal{V}(u_1), \mathcal{V}(u_2)) \mathrm{d}u_1 \mathrm{d}u_2 &= \delta^4 \int_0^1 \int_0^1 (1 - v_1)(1 - v_2) c^{\mathrm{Ga}}_{\rho(k)}(v_1, v_2) \mathrm{d}v_1 \mathrm{d}v_2 \\ &= \delta^4 \mathbb{E}((1 - V_t)(1 - V_{t+k})) \\ &= \delta^4 (1 - \mathbb{E}(V_t) - \mathbb{E}(V_{t+k}) + \mathbb{E}(V_t V_{t+k})) = \delta^4 \mathbb{E}(V_t V_{t+k}) \end{aligned}$$

where (V_t, V_{t+k}) has joint distribution given by the Gaussian copula $C^{\mathrm{Ga}}_{\rho(k)}$. Similarly, by making the substitutions $v_1 = \mathcal{V}(u_1) = 1 - u_1/\delta$ and $v_2 = \mathcal{V}(u_2) = (u_2 - \delta)/(1 - \delta)$ on $[0, \delta] \times [\delta, 1]$, we get

$$\begin{aligned}
&\int_0^\delta \int_\delta^1 u_1 u_2 c^{\mathrm{Ga}}_{\rho(k)}(\mathcal{V}(u_1), \mathcal{V}(u_2)) \mathrm{d}u_1 \mathrm{d}u_2 \\
&= \int_0^1 \int_0^1 \delta^2 (1-\delta)(1-v_1)\Big(\delta + (1-\delta)v_2\Big) c^{\mathrm{Ga}}_{\rho(k)}(v_1, v_2) \mathrm{d}v_1 \mathrm{d}v_2 \\
&= \delta^3(1-\delta)\mathbb{E}(1-V_t) + \delta^2(1-\delta)^2\mathbb{E}((1-V_t)V_{t+k}) = \frac{\delta^2(1-\delta)}{2} - \delta^2(1-\delta)^2\mathbb{E}(V_t V_{t+k})
\end{aligned}$$

and the same value is obtained on the quadrant $[\delta, 1] \times [0, \delta]$. Finally, making the substitutions $v_1 = \mathcal{V}(u_1) = (u_1 - \delta)/(1 - \delta)$ and $v_2 = \mathcal{V}(u_2) = (u_2 - \delta)/(1 - \delta)$ on $[\delta, 1] \times [\delta, 1]$, we get

$$\begin{aligned}
&\int_\delta^1 \int_\delta^1 u_1 u_2 c^{\mathrm{Ga}}_{\rho(k)}(\mathcal{V}(u_1), \mathcal{V}(u_2)) \mathrm{d}u_1 \mathrm{d}u_2 \\
&= \int_0^1 \int_0^1 (1-\delta)^2 \Big(\delta + (1-\delta)v_1\Big)\Big(\delta + (1-\delta)v_2\Big) c^{\mathrm{Ga}}_{\rho(k)}(v_1, v_2) \mathrm{d}v_1 \mathrm{d}v_2 \\
&= \int_0^1 \int_0^1 (1-\delta)^2 \Big(\delta^2 + \delta(1-\delta)v_1 + \delta(1-\delta)v_2 + (1-\delta)^2 v_1 v_2\Big) c^{\mathrm{Ga}}_{\rho(k)}(v_1, v_2) \mathrm{d}v_1 \mathrm{d}v_2
\end{aligned}$$

$$\begin{aligned}
&= \delta^2(1-\delta)^2 + \delta(1-\delta)^3\mathbb{E}(V_t) + \delta(1-\delta)^3\mathbb{E}(V_{t+k}) + (1-\delta)^4\mathbb{E}(V_t V_{t+k}) \\
&= \delta(1-\delta)^2 + (1-\delta)^4\mathbb{E}(V_t V_{t+k})
\end{aligned}$$

Collecting all of these terms together yields

$$\int_0^1 \int_0^1 u_1 u_2 c^{\mathrm{Ga}}_{\rho(k)}(\mathcal{V}(u_1), \mathcal{V}(u_2)) \mathrm{d}u_1 \mathrm{d}u_2 = \delta(1-\delta) + (2\delta - 1)^2 \mathbb{E}(V_t V_{t+k})$$

and, since $\rho_S(Z_t, Z_{t+k}) = 12\mathbb{E}(V_t V_{t+k}) - 3$, it follows that

$$\begin{aligned}
\rho(U_t, U_{t+k}) = 12\mathbb{E}(U_t U_{t+k}) - 3 &= 12 \int_0^1 \int_0^1 u_1 u_2 c^{\mathrm{Ga}}_{\rho(k)}(\mathcal{V}(u_1), \mathcal{V}(u_2)) \mathrm{d}u_1 \mathrm{d}u_2 - 3 \\
&= 12\delta(1-\delta) + 12(2\delta - 1)^2 \mathbb{E}(V_t V_{t+k}) - 3 \\
&= 12\delta(1-\delta) + (2\delta - 1)^2(\rho_S(Z_t, Z_{t+k}) + 3) - 3 \\
&= (2\delta - 1)^2 \rho_S(Z_t, Z_{t+k}) \,.
\end{aligned}$$

The value of Spearman's rho $\rho_S(Z_t, Z_{t+k})$ for the bivariate Gaussian distribution is well known; see, for example, McNeil et al. (2015).

Appendix A.10. Proof of Proposition 7

The conditional density satisfies

$$f_{U_t|\boldsymbol{U}_{t-1}}(u \mid \boldsymbol{u}_{t-1}) \;=\; \frac{c_{\boldsymbol{U}_t}(u_1, \ldots, u_{t-1}, u)}{c_{\boldsymbol{U}_{t-1}}(u_1, \ldots, u_{t-1})} = \frac{c^{\mathrm{Ga}}_{P(1,\ldots,t)}(\mathcal{V}(u_1), \ldots, \mathcal{V}(u_{t-1}), \mathcal{V}(u))}{c^{\mathrm{Ga}}_{P(1,\ldots,t-1)}(\mathcal{V}(u_1), \ldots, \mathcal{V}(u_{t-1}))} \,.$$

The Gaussian copula density is given in general by

$$c^{\mathrm{Ga}}_P(v_1, \ldots, v_d) = \frac{f_{\boldsymbol{Z}}\big(\Phi^{-1}(v_1), \ldots, \Phi^{-1}(v_d)\big)}{\prod_{i=1}^d \phi\big(\Phi^{-1}(v_i)\big)}$$

where $\boldsymbol{Z}$ is a multivariate Gaussian vector with standard normal margins and correlation matrix P. Hence, it follows that we can write

$$\begin{aligned} f_{U_t|\boldsymbol{U}_{t-1}}(u \mid \boldsymbol{u}_{t-1}) &= \frac{f_{\boldsymbol{Z}_t}\left(\Phi^{-1}(\mathcal{V}(u_1)), \ldots, \Phi^{-1}(\mathcal{V}(u_{t-1})), \Phi^{-1}(\mathcal{V}(u))\right)}{f_{\boldsymbol{Z}_{t-1}}\left(\Phi^{-1}(\mathcal{V}(u_1)), \ldots, \Phi^{-1}(\mathcal{V}(u_{t-1}))\right)\phi(\Phi^{-1}(\mathcal{V}(u)))} \\ &= \frac{f_{Z_t|\boldsymbol{Z}_{t-1}}\left(\Phi^{-1}(\mathcal{V}(u)) \mid \Phi^{-1}(\mathcal{V}(\boldsymbol{u}_{t-1}))\right)}{\phi(\Phi^{-1}(\mathcal{V}(u)))} \end{aligned}$$

where $f_{Z_t|\boldsymbol{Z}_{t-1}}$ is the conditional density of the ARMA process, from which (20) follows easily.

References

Aas, Kjersti, Claudia Czado, Arnoldo Frigessi, and Henrik Bakken. 2009. Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics* 44: 182–98. [CrossRef]

Andersen, Torben G. 1994. Stochastic autoregressive volatility: A framework for volatility modeling. *Mathematical Finance* 4: 75–102.

Andersen, Torben G., and Luca Benzoni. 2009. Stochastic Volatility. In *Complex Systems in Finance and Econometrics*. Edited by Robert A. Meyers. New York: Springer.

Bedford, Tim, and Roger M. Cooke. 2001. Probability density decomposition for conditionally independent random variables modeled by vines. *Annals of Mathematics and Artificial Intelligence* 32: 245–68. [CrossRef]

Bedford, Tim and Roger M. Cooke. 2002. Vines—A new graphical model for dependent random variables. *Annals of Statistics* 30: 1031–68. [CrossRef]

Bladt, Martin, and Alexander J. McNeil. 2020. Time series copula models using d-vines and v-transforms: An alternative to GARCH modelling. *arXiv* arXiv:2006.11088.

Bollerslev, Tim. 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics 31*: 307–27. [CrossRef]

Bollerslev, Tim, Robert F. Engle, and Daniel B. Nelson. 1994. ARCH models. In *Handbook of Econometrics*. Edited by Robert F. Engle and Daniel L. McFadden. Amsterdam: North-Holland, vol. 4, pp. 2959–3038.

Campbell, John Y., Anrew W. Lo, and A. Craig MacKinlay. 1997. *The Econometrics of Financial Markets*. Princeton: Princeton University Press.

Chen, Xiaohong, and Yanqin Fan. 2006. Estimation of copula-based semiparametric time series models. *Journal of Econometrics* 130: 307–35. [CrossRef]

Chen, Xiaohong, Wei Biao Wu, and Yanping Yi. 2009. Efficient estimation of copula-based semiparametric Markov models. *Annals of Statistics 37*: 4214–53. [CrossRef]

Cont, Rama. 2001. Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance* 1: 223–36. [CrossRef]

Creal, Drew, Siem Jan Koopman, and André Lucas. 2013. Generalized autoregressive score models with applications. *Journal of Applied Econometrics* 28: 777–95. [CrossRef]

Ding, Zhuanxin, Clive W. Granger, and Robert F. Engle. 1993. A long memory property of stock market returns and a new model. *Journal of Empirical Finance* 1: 83–106. [CrossRef]

Domma, Filippo, Sabrina Giordano, and Pier Francesco Perri. 2009. Statistical modeling of temporal dependence in financial data via a copula function. *Communications if Statistics: Simulation and Computation* 38: 703–28. [CrossRef]

Engle, Robert F. 1982. Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. *Journal of the Econometric Society* 50: 987–1008. [CrossRef]

Fan, Yanqin, and Andrew J. Patton. 2014. Copulas in econometrics. *Annual Review of Economics* 6: 179–200. [CrossRef]

Fernández, Carmen, and Mark F. J. Steel. 1998. On Bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association* 93: 359–71.

Genest, Christian, Kilani Ghoudi, and Louis-Paul Rivest. 1995. A semi-parametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika* 82: 543–52. [CrossRef]

Glosten, Lawrence R., Ravi Jagannathan, and David E. Runkle. 1993. On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance* 48: 1779–801. [CrossRef]

Joe, Harry. 1996. Families of m-variate distributions with given margins and $m(m-1)/2$ bivariate dependence parameters. In *Distributions with Fixed Marginals and Related Topics*. Edited by Ludger Rüschendorf, Berthold Schweizer and Michael D. Taylor. Lecture Notes–Monograph Series. Hayward: Institute of Mathematical Statistics, vol. 28, pp. 120–41.

Joe, Harry. 2015. *Dependence Modeling with Copulas*. Boca Raton: CRC Press.

Loaiza-Maya, Rubén, Michael S. Smith, and Worapree Maneesoonthorn. 2018. Time series copulas for heteroskedastic data. *Journal of Applied Econometrics* 33: 332–54. [CrossRef]

McNeil, Alexander J., Rüdiger Frey, and Paul Embrechts. 2015. *Quantitative Risk Management: Concepts, Techniques and Tools*, 2nd ed. Princeton: Princeton University Press.

Mikosch, Thomas, and Catalin Stărică. 2000. Limit theory for the sample autocorrelations and extremes of a GARCH(1,1) process. *The Annals of Statistics* 28: 1427–51.
Patton, Andew J. 2012. A review of copula models for economic time series. *Journal of Multivariate Analysis* 110: 4–18. [CrossRef]
Rémillard, Bruno. 2013. *Statistical Methods for Financial Engineering*. London: Chapman & Hall.
Shephard, Neil. 1996. Statistical aspects of ARCH and stochastic volatility. In *Time Series Models in Econometrics, Finance and Other Fields*. Edited by David R. Cox, David V. Hinkley and Ole E. Barndorff-Nielsen. London: Chapman & Hall, pp. 1–55.
Smith, Michael S., Aleksey Min, Carlos Almeida, and Claudia Czado. 2010. Modeling Longitudinal Data Using a Pair-Copula Decomposition of Serial Dependence. *Journal of the American Statistical Association* 105: 1467–79. [CrossRef]
Taylor, Stephen J. 1994. Modeling stochastic volatility: A review and comparative study. *Mathematical Finance* 4: 183–204. [CrossRef]
Terasaka, Takahiro, and Yuzo Hosoya. 2007. A modified Box-Cox transformation in the multivariate ARMA model. *Journal of the Japan Statistical Society* 37: 1–28. [CrossRef]