Author/s:
Deng, L;Smith, MS;Maneesoonthorn, W

Title:
Large Skew-t Copula Models and Asymmetric Dependence in Intraday Equity Returns

Date:
2025-01-01

Citation:
Deng, L., Smith, M. S. & Maneesoonthorn, W. (2025). Large Skew-t Copula Models and Asymmetric Dependence in Intraday Equity Returns. Journal of Business and Economic Statistics, 43 (2), pp.269-285. https://doi.org/10.1080/07350015.2024.2360592.

Persistent Link:
https://hdl.handle.net/11343/352250

Taylor & Francis
Taylor & Francis Group

Journal of Business & Economic Statistics

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/ubes20

Large Skew-t Copula Models and Asymmetric Dependence in Intraday Equity Returns

Lin Deng, Michael Stanley Smith & Worapree Maneesoonthorn

To cite this article: Lin Deng, Michael Stanley Smith & Worapree Maneesoonthorn (15 Jul 2024): Large Skew-t Copula Models and Asymmetric Dependence in Intraday Equity Returns, Journal of Business & Economic Statistics, DOI: 10.1080/07350015.2024.2360592

To link to this article: https://doi.org/10.1080/07350015.2024.2360592

JOURNAL OF BUSINESS & ECONOMIC STATISTICS
2024, VOL. 00, NO. 0, 1–17
https://doi.org/10.1080/07350015.2024.2360592

Large Skew-t Copula Models and Asymmetric Dependence in Intraday Equity Returns

Lin Deng[a], Michael Stanley Smith[a], and Worapree Maneesoonthorn[b]

[a]Melbourne Business School, University of Melbourne, Carlton, Australia; [b]Department of Econometrics and Business Statistics, Monash University, Clayton, Australia

ABSTRACT

Skew-t copula models are attractive for the modeling of financial data because they allow for asymmetric and extreme tail dependence. We show that the copula implicit in the skew-t distribution of Azzalini and Capitanio allows for a higher level of pairwise asymmetric dependence than two popular alternative skew-t copulas. Estimation of this copula in high dimensions is challenging, and we propose a fast and accurate Bayesian variational inference (VI) approach to do so. The method uses a generative representation of the skew-t distribution to define an augmented posterior that can be approximated accurately. A stochastic gradient ascent algorithm is used to solve the variational optimization. The methodology is used to estimate skew-t factor copula models with up to 15 factors for intraday returns from 2017 to 2021 on 93 U.S. equities. The copula captures substantial heterogeneity in asymmetric dependence over equity pairs, in addition to the variability in pairwise correlations. In a moving window study we show that the asymmetric dependencies also vary over time, and that intraday predictive densities from the skew-t copula are more accurate than those from benchmark copula models. Portfolio selection strategies based on the estimated pairwise asymmetric dependencies improve performance relative to the index.

ARTICLE HISTORY
Received March 2024
Accepted May 2024

KEYWORDS
Asymmetric dependence; Bayesian data augmentation; Factor copula; Generative representation; Intraday equity returns; Variational inference

1. Introduction

Copula models are widely employed for multivariate data because they separate the selection of marginals from the copula function. When modeling financial data, the copula should possess two features. First, it should allow for asymmetric dependence, where the level of dependence varies for different quantiles of the variables. Second, it should allow for high tail dependence, where extreme values of both variables are positively or negatively dependent. Skew-t copulas can capture both features in high dimensions. However, there exist multiple variants of skew-t copulas, with no consensus on the most effective choice, and their estimation in high dimensions is challenging. This article addresses both issues by showing that the skew-t copula implicit in the Azzalini and Capitanio (2003) distribution, combined with Bayesian variational inference, offers an attractive solution. We employ the methodology in a study using 15 min intraday returns on 93 U.S. equities over five years, and show that pairwise asymmetric dependence varies in a complex fashion over both equity pairs and time.

A skew-t copula is that implicit in a parametric skew-t distribution, and three types have been used previously. The first was proposed by Demarta and Mcneil (2005) and is based on the generalized hyperbolic (GH) skew-t distribution. It is the most popular of the three in financial studies, with applications by Christoffersen et al. (2012), Creal and Tsay (2015), Oh and Patton (2017), Lucas, Schwaab, and Zhang (2017), and Oh and Patton (2023). The second was suggested by Smith, Gan, and Kohn (2010) and is based on the distribution of Sahu, Dey, and Branco (2003), while the third proposed by Kollo and Pettere (2010) and Yoshiba (2018) is based on the distribution of Azzalini and Capitanio (2003), which we label the "AC skew-t copula." All three skew-t distributions have latent generative representations for which factor models can be used, and their implicit copulas are called "factor copulas" by Oh and Patton (2017).[1] In this article we show the AC skew-t factor copula allows for higher levels of asymmetry in pairwise dependence than the other two copulas, making it an attractive choice for financial data.

Yoshiba (2018) studies maximum likelihood estimation of the AC skew-t copula in low dimensions. However, the likelihood (and therefore also the posterior) has a complex geometry and its direct optimization is hard in high dimensions, which has precluded its use with large financial panels. Here we propose a scalable Bayesian solution that uses a conditionally Gaussian generative representation of the AC skew-t distribution. Latent variables are introduced and augmented with the copula parameters to provide a tractable joint (or "augmented") posterior. However, evaluation of this augmented posterior using Markov chain Monte Carlo (MCMC) methods is prohibitively slow in

[1]This type of factor copula is not to be confused with the similarly named vine-based factor copulas of Krupskii and Joe (2013) and others, which are not the implicit copulas of high-dimensional skew-t distributions studied here.

CONTACT Michael Stanley Smith mike.smith@mbs.edu Melbourne Business School, 200 Leicester Street, Carlton 3053, Victoria, Australia.
Supplementary materials for this article are available online. Please go to www.tandfonline.com/UBES.

high dimensions. To overcome this we develop a new variational inference estimator for evaluating the proposed augmented posterior of the AC skew-t copula in high dimensions.

Variational inference (VI) approximates a target density with a tractable density called a variational approximation (VA). The VA is calibrated by solving an optimization problem where the Kullback-Leibler divergence between it and the target density is minimized. Well designed VI methods are effective for estimating models with large numbers of parameters and/or big datasets; see Blei, Kucukelbir, and McAuliffe (2017) for an introduction. But here, the complex geometry of the posterior of the AC skew-t copula parameters is difficult to approximate directly using standard VI methods. Therefore, we instead propose using the augmented posterior as the target density, and then adopt a flexible VA of the form suggested by Loaiza-Maya et al. (2022). To solve the optimization problem stochastic gradient methods (Ranganath, Gerrish, and Blei 2014) are used with re-parameterization gradients (Kingma and Welling 2014). To increase the speed of the method we derive these gradients analytically. The speed and accuracy of the resulting VI method is demonstrated using simulated data.

We use our new copula methodology in two studies using 15 min intraday financial data. The first illustrates the flexibility of the AC skew-t copula by modeling returns on two equities and the VIX index during two 40 day periods: a low market volatility period, and the high market volatility period of the COVID-19 pandemic crash. We find strong asymmetries in the pairwise dependencies, which change in direction and scale from the low to high market volatility periods; asymmetries that are not identified using the two other variants of skew-t copula.

The second example is our main study, and uses an AC skew-t factor copula to capture dependence between returns on 93 equities. Using a rolling window of width 40 trading days from January 2017 to December 2021, we make three findings. First, we show that one-step-ahead forecast densities are more accurate than those from benchmark copula models. Forecast accuracy is maximized with 10–15 factors, suggesting a rich factor structure is worthwhile. This is consistent with Oh and Patton (2017), Opschoor et al. (2021), and Oh and Patton (2023), who find that industry or other group-based factors increase accuracy. Second, a major finding is that pairwise dependence asymmetries can differ substantially over equity pairs in addition to overall correlations. Third, we show that (absent of trading costs) portfolio selection strategies based on pairwise quantile dependencies from the copula model can improve performance relative to index returns. Our findings are consistent with evidence of cross-sectional heterogeneity in asymmetric dependence of equity returns (Bollerslev, Patton, and Quaedvlieg 2022; Ando, Greenwood-Nimmo, and Shin 2022) and its importance in risk management (Patton 2004; Harvey et al. 2010; Giacometti, Torri, and Paterlini 2021).

Skew-t factor copulas have been used previously to capture dependence in high dimensional financial data. Oh and Patton (2013) use the implicit copula of the GH distribution, and of a mixture of GH and t distributions. Creal and Tsay (2015) and Oh and Patton (2023) consider the implicit copula of the GH distribution with a single global factor, which the latter authors augment with latent group specific factors. Lucas, Schwaab, and Zhang (2017) considers the same copula with a block equicorrelation structure. In comparison, ours is the first application of which we are aware of the AC skew-t copula to a large financial panel. We adopt an unrestricted specification for the skew parameters of the skew-t distribution, allowing for greater flexibility in its implicit copula. Previous studies consider dynamic latent factor specifications for fitting daily or weekly financial time series. In contrast, we employ intraday data at the 15 min frequency in a rolling window study using a static factor copula model with up to 15 global factors. Our objective is to allow for a high degree of heterogeneity in the level of asymmetric dependence over equity pairs, rather than capturing time variation at a lower data frequency.

Our paper also contributes to the literature on Bayesian estimation of skew-t copulas. Creal and Tsay (2015) use MCMC to estimate their skew-t factor copula, where a particle filter is used to evaluate the intractable likelihood. Smith, Gan, and Kohn (2010) use MCMC with data augmentation, but for the skew-t copula implicit in the distribution of Sahu, Dey, and Branco (2003). Both approaches can also be adopted to estimate the AC skew-t copula, but they incur a much higher computational burden than the VI approach suggested here. A few recent studies have used VI methods to estimate other high-dimensional copula models. Loaiza-Maya and Smith (2019) use VI to estimate vine copulas with discrete margins, while Nguyen, Ausín, and Galeano (2020) use VI to estimate factor copulas based on pair-copula constructions. However, both studies employ simple mean field or other VAs that are ineffective approximations for the complex geometry of the posterior of the AC skew-t copula.

The rest of the article is organized as follows. Section 2 outlines the three skew-t copulas. Section 3 develops the variational inference method for estimating the AC skew-t factor copula, and its efficacy is demonstrated using simulated data in Section 4. Sections 5 and 6 contain the two empirical studies of intraday equity returns, while Section 7 concludes.

2. Skew-t Copulas

2.1. Copula Models

A copula model (Nelsen 1999; Joe 2014) expresses the joint distribution function of a continuous-valued random vector $\boldsymbol{Y} = (Y_1, \ldots, Y_d)^\top$ as

$$F_Y(\boldsymbol{y}) = C(F_{Y_1}(y_1), \ldots, F_{Y_d}(y_d)) \,. \tag{1}$$

Here, $\boldsymbol{y} = (y_1, \ldots, y_d)^\top$, F_{Y_j} is the marginal distribution function of Y_j, and $C : [0,1]^d \to \mathbb{R}$ is a copula function that captures the dependence structure. Differentiating (1) gives the joint density

$$f_Y(\boldsymbol{y}) = \frac{\partial}{\partial \boldsymbol{y}} F_Y(\boldsymbol{y}) = c(F_{Y_1}(y_1), \ldots, F_{Y_d}(y_d)) \prod_{j=1}^{d} f_{Y_j}(y_j) \,, \tag{2}$$

where $f_{Y_j} = \frac{\partial}{\partial y_j} F_{Y_j}(y_j)$, $c(\boldsymbol{u}) = \frac{\partial}{\partial \boldsymbol{u}} C(\boldsymbol{u})$ is widely called the "copula density", and $\boldsymbol{u} = (u_1, \ldots, u_d)^\top$.[2]

The main advantage of using a copula model is that the marginals and the copula function can be modeled separately.

[2]The notations $C(u_1, \ldots, u_d)$ and $C(\boldsymbol{u})$ are used interchangeably, as are the notations $c(u_1, \ldots, u_d)$ and $c(\boldsymbol{u})$.

When the elements of $\boldsymbol{Y}$ are asset returns, their distribution features asymmetric and high tail dependence; for example see Patton (2006) and Christoffersen et al. (2012). A skew-t copula is one of only a few copulas that can account for these features in high dimensions.

2.2. Three Skew-t Copulas

Let the continuous random vector $\boldsymbol{Z} = (Z_1, \ldots, Z_d)^\top \sim F_Z$, with marginals $F_{Z_1}, \ldots, F_{Z_d}$. Then if $U_j = F_{Z_j}(Z_j)$, the distribution function of $\boldsymbol{U} = (U_1, \ldots, U_d)^\top$ is $C(\boldsymbol{u}) = F_Z(F_{Z_1}^{-1}(u_1), ..., F_{Z_d}^{-1}(u_d))$. This is called an implicit copula, and the elements of $\boldsymbol{Z}$ are called pseudo or auxiliary variables because their values are unobserved; see Smith (2023) for an introduction. The copula density is

$$c(\boldsymbol{u}) = \frac{\partial}{\partial \boldsymbol{u}} C(\boldsymbol{u}) = f_Z(\boldsymbol{z}) / \prod_{j=1}^{d} f_{Z_j}(z_j) , \tag{3}$$

where $\boldsymbol{z} = (z_1, \ldots, z_d)^\top$, $f_Z(\boldsymbol{z}) = \frac{\partial}{\partial \boldsymbol{z}} F_Z(\boldsymbol{z})$ and $f_{Z_j} = \frac{\partial}{\partial z_j} F_{Z_j}(z_j)$. A skew-t copula is where F_Z is a multivariate skew-t distribution. Because there are multiple skew-t distributions, there are also multiple copulas. Three types have been used previously, which we briefly outline below. The marginal moments of $\boldsymbol{Z}$ are unidentified in C, so that location parameters are unnecessary and the scale matrices are restricted to be a correlation matrix $\bar{\Omega}$ for all three skew-t copulas.

2.2.1. GH Skew-t Copula

Demarta and Mcneil (2005) construct the implicit copula of the generalized hyperbolic (GH) skew-t distribution (Barndorff-Nielsen 1977). This distribution has the generative representation

$$\boldsymbol{Z}_{\text{GH}} = \boldsymbol{\delta} W + W^{-1/2} \boldsymbol{X} ,$$

where $\boldsymbol{X} \sim N_d(\boldsymbol{0}, \bar{\Omega})$, $W \sim \text{Gamma}(\nu/2, \nu/2)$, and $\boldsymbol{\delta} \in \mathbb{R}^d$ is the skewness parameter. The joint density of $\boldsymbol{Z}_{\text{GH}}$ is available in closed form (see Part A of the Web Appendix). However, the marginal distributions and quantile functions are not and are difficult to compute using numerical methods, so that they are usually evaluated by Monte Carlo simulation from the generative representation. For high d this is slow. Oh and Patton (2023) note that if $\boldsymbol{\delta} = \delta(1, 1, \ldots, 1)^\top$ with δ a scalar, then the d marginals of the GH distribution are the same. This speeds computation greatly, but at the cost of restricting the flexibility of the pairwise asymmetric dependencies in its implicit copula.

2.2.2. SDB Skew-t Copula

Smith, Gan, and Kohn (2010) construct the implicit copula of the skew-t distribution proposed by Sahu, Dey, and Branco (2003). This distribution can be constructed from a t-distribution by hidden conditioning as follows. Let $\boldsymbol{X}$ and $\boldsymbol{L}$ be d-dimensional random variables jointly distributed

$$\begin{pmatrix} \boldsymbol{X} \\ \boldsymbol{L} \end{pmatrix} \sim t_{2d}\left(\boldsymbol{0}, \Omega, \nu\right), \quad \Omega = \begin{pmatrix} \bar{\Omega} + D^2 & D \\ D & I \end{pmatrix},$$

with $D = \text{diag}(\boldsymbol{\delta})$, $\boldsymbol{\delta} \in \mathbb{R}^d$, $\bar{\Omega}$ a positive definite matrix, and $t_{2d}(\boldsymbol{0}, \Omega, \nu)$ denoting a t-distribution with mean zero, scale matrix Ω and ν degrees of freedom. Then, $\boldsymbol{Z}_{\text{SDB}} \equiv (\boldsymbol{X}|\boldsymbol{L} > \boldsymbol{0})$ is a skew-t distribution[3] with skewness parameters $\boldsymbol{\delta}$ and joint density $f_{Z_{\text{SDB}}}(\boldsymbol{z}; \bar{\Omega}, \boldsymbol{\delta}, \nu) = 2^d f_t(\boldsymbol{z}; \boldsymbol{0}, \bar{\Omega} + D^2, \nu) \Pr(\boldsymbol{V} > \boldsymbol{0}; \boldsymbol{z})$ where $f_t(\boldsymbol{z}; \boldsymbol{0}, \bar{\Omega}, \nu)$ is the density of a $t_d(\boldsymbol{0}, \bar{\Omega}, \nu)$ distribution evaluated at $\boldsymbol{z}$, and $\boldsymbol{V}$ follows a d-dimensional t-distribution with parameters that are functions of $\boldsymbol{z}$. Sahu, Dey, and Branco (2003) show that the j^{th} marginal has density $f_{Z_{\text{SDB}}}(z_j; 1, \delta_j, \nu)$ with distribution and quantile functions that are computed numerically. Evaluation of the copula density at (3) for large d is difficult because computation of the term $\Pr(\boldsymbol{V} > \boldsymbol{0}; \boldsymbol{z})$ is also. However, Smith, Gan, and Kohn (2010) show how to estimate the distribution and its implicit copula using Bayesian data augmentation.

2.2.3. AC Skew-t Copula

Kollo and Pettere (2010) and Yoshiba (2018) construct the implicit copula of the skew-t distribution proposed by Azzalini and Capitanio (2003). This distribution is formed via hidden conditioning as follows. Let $\boldsymbol{X}$ be a d-dimensional random vector and L a random variable with joint distribution

$$\begin{pmatrix} \boldsymbol{X} \\ L \end{pmatrix} \sim t_{d+1}(\boldsymbol{0}, \Omega, \nu), \quad \Omega = \begin{pmatrix} \bar{\Omega} & \boldsymbol{\delta} \\ \boldsymbol{\delta}^\top & 1 \end{pmatrix}, \tag{4}$$

where $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_d)^\top$. Then, $\boldsymbol{Z}_{\text{AC}} \equiv (\boldsymbol{X}|L > 0)$ is a skew-t distribution with joint density

$$f_{Z_{\text{AC}}}(\boldsymbol{z}; \bar{\Omega}, \boldsymbol{\delta}, \nu) = 2 f_t(\boldsymbol{z}; \boldsymbol{0}, \bar{\Omega}, \nu) T\left(\boldsymbol{\alpha}^\top \boldsymbol{z} \sqrt{\frac{\nu + d}{\nu + \mathcal{M}(z)}}; \nu + d\right), \tag{5}$$

where $\mathcal{M}(\boldsymbol{z}) = \boldsymbol{z}^\top \bar{\Omega}^{-1} \boldsymbol{z}$, and $T(x; \nu)$ is the distribution function of a univariate student-t with degrees of freedom ν, evaluated at x. There is a one-to-one relationship between $\boldsymbol{\alpha}$ and $\boldsymbol{\delta}$, given by

$$\boldsymbol{\alpha} = (1 - \boldsymbol{\delta}^\top \bar{\Omega}^{-1} \boldsymbol{\delta})^{-1/2} \bar{\Omega}^{-1} \boldsymbol{\delta}, \text{ and}$$
$$\boldsymbol{\delta} = (1 + \boldsymbol{\alpha}^\top \bar{\Omega} \boldsymbol{\alpha})^{-1/2} \bar{\Omega} \boldsymbol{\alpha} .$$

It is more convenient to employ $\boldsymbol{\alpha} \in \mathbb{R}^d$ as the skewness parameter, rather than $\boldsymbol{\delta}$, because it is unconstrained. The j^{th} marginal of (5) has density $f_{Z_{\text{AC}}}(z_j; 1, \delta_j, \nu)$, and the distribution function $F_j(z_j; \delta_j, \nu) = \int_{-\infty}^{z_j} f_{Z_{\text{AC}}}(\xi; 1, \delta_j, \nu) d\xi$ is computed using numerical integration.

A draw from the AC skew-t distribution can be obtained by first generating $W \sim \text{Gamma}(\nu/2, \nu/2)$, followed by $\tilde{L} \sim N(0, 1)$ constrained so that $\tilde{L} > 0$, and then from

$$\boldsymbol{Z}_{\text{AC}} | \tilde{L}, W \sim N_d\left(\frac{\tilde{L}}{\sqrt{W}} \boldsymbol{\delta}, \frac{1}{W}\left(\bar{\Omega} - \boldsymbol{\delta}\boldsymbol{\delta}^\top\right)\right), \tag{6}$$

where $\tilde{L} = W^{1/2} L$, and L is defined at (4). An alternative generative representation for the AC skew-t distribution is given in Appendix B. To convert a draw $\boldsymbol{Z}_{\text{AC}} = (Z_1, \ldots, Z_d)^\top$ from the skew-t distribution to a draw $\boldsymbol{U} \sim C$ from its implicit copula, set $U_j = F_j(Z_j; \delta_j, \nu)$ for all j.

[3]The notation $\boldsymbol{X}|\boldsymbol{L} > \boldsymbol{0}$ corresponds to the conditional distribution of $\boldsymbol{X}$ given that all elements of $\boldsymbol{L}$ are positive. It does not denote the conditional distribution of $\boldsymbol{X}$ given a specific value for $\boldsymbol{L}$. This notational convention is used throughout the article.

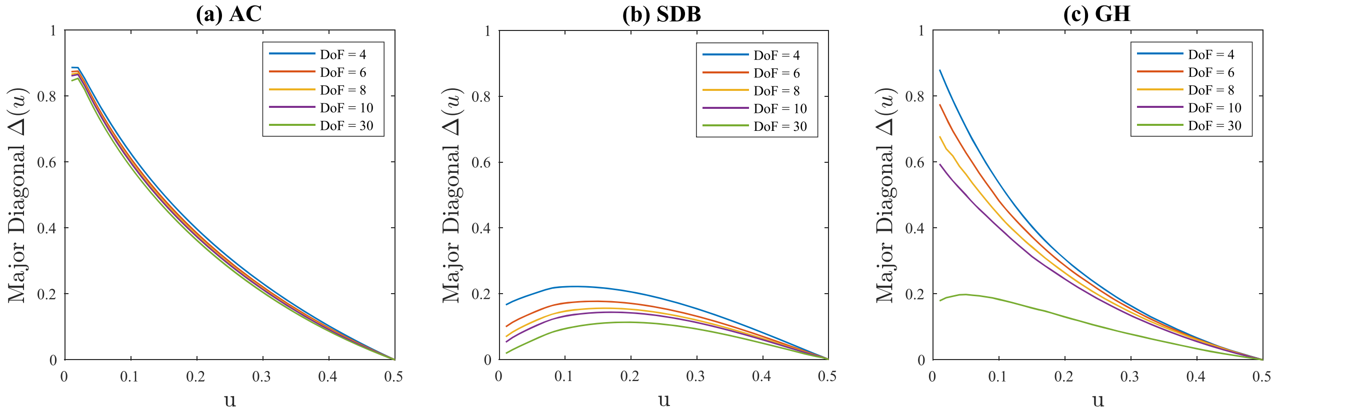

Figure 1. Maximum asymmetric dependence $\max_{\delta,\rho}\{\Delta(u)\}$ of the three skew-t copulas (a) AC, (b) SDB, and (c) GH. Results are given for degrees of freedom $\nu = 4, 6, 8, 10, 30$, and plotted as a function of u.

2.3. Asymmetric Dependence

The motivation for using any skew-t copula over a t-copula is to capture asymmetric dependence. To measure this here we compute the four pairwise quantile dependence metrics, and define measures of asymmetry in the major and minor diagonals as follows. Let (Y_1, Y_2) follow a bivariate copula model[4] with copula function $C(u_1, u_2) = \Pr(U_1 \leq u_1, U_2 \leq u_2)$. Then, for $0 < u < 0.5$, we define the four quantile dependencies as: lower left $\lambda_{\text{LL}}(u) = P(U_2 \leq u | U_1 \leq u)$, upper right $\lambda_{\text{UR}}(u) = P(U_2 > 1-u | U_1 > 1-u)$, lower right $\lambda_{\text{LR}}(u) = P(U_2 \leq u | U_1 > 1-u)$, and upper left $\lambda_{\text{UL}}(u) = P(U_2 > 1-u | U_1 \leq u)$. They are computed for the AC and SDB skew-t copulas using numerical integration, and by Monte Carlo simulation for the GH skew-t copula; see Appendix A for more details. For a given quantile value u, the metrics

$$\Delta_{\text{Major}}(u) \equiv \lambda_{\text{UR}}(u) - \lambda_{\text{LL}}(u), \quad \text{and} \quad \Delta_{\text{Minor}}(u) \equiv \lambda_{\text{UL}}(u) - \lambda_{\text{LR}}(u),$$

measure asymmetry in the major and minor diagonals, respectively. Ando, Greenwood-Nimmo, and Shin (2022) suggest similar quantile metrics, although based on a topological model, rather than a copula model.

The maximum asymmetry for each of the three copulas is obtained by solving the optimization

$$\max_{\delta,\rho} \{\Delta(u)\}$$

for given values of $0 < u < 0.5$ and ν. The solution is the same for $\Delta_{\text{Major}}(u)$ and $\Delta_{\text{Minor}}(u)$, and optimization is with respect to $\boldsymbol{\delta} = (\delta_1, \delta_2)$ and the off-diagonal element ρ of the (2×2) correlation matrix $\bar{\Omega}$. Figure 1 plots the maximums against u. At every quantile the GH skew-t copula has a higher maximum asymmetric dependence than the SDB skew-t copula. However, the AC skew-t copula has a higher maximum level of asymmetric dependence than both alternatives for all but the lowest values of ν, at which $\Delta(u)$ is almost equal for the AC and GH skew-t copulas. Figure A1 in the Web Appendix plots contours of the bivariate densities for the three copulas with maximal $\Delta_{\text{Major}}(0.01)$, further illustrating the strong differences between the skew-t copulas

In summary, the AC skew-t copula allows for a higher level of asymmetric dependence, making it an attractive choice. In addition, the computational demands of evaluating the marginals of the GH skew-t distribution complicates the use of its implicit copula in higher dimensions. A comparison of the differing skew-t copulas is undertaken later in our study of intraday data.

2.4. Factor Copula

For large d, an unrestricted correlation matrix $\bar{\Omega}$ is difficult to estimate and one solution is adopt a factor model for the auxiliary variables $\boldsymbol{Z}$. Following Murray et al. (2013), a static factor copula is used here, which corresponds to adopting the factorization

$$\bar{\Omega} = V_1 V V_1 = V_1 (GG^\top + D) V_1.$$

Here, $G = \{g_{ij}\}$ is an $(d \times k)$ loadings matrix with $k < d$, zero upper triangle and positive leading diagonal elements (i.e., $g_{ii} > 0$), and to identify the parameters $D = I_d$. The diagonal matrix $V_1 = \text{diag}(V)^{-1/2}$ normalizes V to a correlation matrix. If $\widetilde{G}$ equals G with the leading diagonal elements replaced by their logarithmic values, then $\bar{\Omega}$ is parameterized by $\text{vech}(\widetilde{G})$. We adopt this transformation because our VI method is applied to unconstrained real-valued model parameters.

3. Variational Inference for the AC Skew-t Copula

3.1. Likelihood and Extended Likelihood

Consider observations $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{id})^\top$, for $i = 1, \ldots, n$, drawn independently from the copula model at (1) with copula parameters $\boldsymbol{\theta}$. The pseudo variables $\boldsymbol{z}_i = (z_{i1}, \ldots, z_{id})^\top$ are given by

$$z_{ij} = F_{Z_j}^{-1}\left(F_{Y_j}\left(y_{ij}\right); \boldsymbol{\theta}_j\right), \tag{7}$$

where F_{Z_j} is a function of $\boldsymbol{\theta}_j \subseteq \boldsymbol{\theta}$. If $\boldsymbol{y} = \{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n\}$, then from (2) and (3), the likelihood is

$$p(\boldsymbol{y}|\boldsymbol{\theta}) = \prod_{i=1}^{n} \left\{ p(\boldsymbol{z}_i|\boldsymbol{\theta}) \prod_{j=1}^{d} \frac{f_{Y_j}(y_{ij})}{f_{Z_j}(z_{ij}; \boldsymbol{\theta}_j)} \right\}.$$

[4]Because all three skew-t distributions are closed under marginalization, so are the skew-t copulas. Therefore, maximum asymmetric tail dependence in the bivariate case is equal to that for variable pairs in higher dimensions.

For the AC skew-t copula, this likelihood has a complex geometry, making direct maximization challenging for large d; see Part D.2 of the Web Appendix for an illustration. However, a more tractable extended likelihood of the AC skew-t copula model can be obtained from the generative representation for the AC skew-t distribution given in Section 2.2.3 as follows.

Let $\boldsymbol{w} = (w_1, \ldots, w_n)^\top$ be a vector n draws of W, and $\tilde{\boldsymbol{l}} = (\tilde{l}_1, \ldots, \tilde{l}_n)^\top$ be a vector of n draws of $\tilde{L}$, then

$$p(\boldsymbol{y}, \tilde{\boldsymbol{l}}, \boldsymbol{w}|\boldsymbol{\theta}) = \prod_{i=1}^{n} p(\boldsymbol{y}_i, \tilde{l}_i, w_i|\boldsymbol{\theta}) = \prod_{i=1}^{n} \left\{ p(\boldsymbol{z}_i, \tilde{l}_i, w_i|\boldsymbol{\theta}) \prod_{j=1}^{d} \frac{f_{Y_j}(y_{ij})}{f_{Z_j}(z_{ij}; \boldsymbol{\theta}_j)} \right\}, \quad (8)$$

is the extended likelihood. The product over j in (8) is the Jacobian of the transformation from $\boldsymbol{z}$ to $\boldsymbol{y}$ at (7). From the generative representation of the AC skew-t distribution, the joint density

$$p(\boldsymbol{z}_i, \tilde{l}_i, w_i|\boldsymbol{\theta}) = p(\boldsymbol{z}_i|\tilde{l}_i, w_i, \boldsymbol{\theta}) p(\tilde{l}_i|w_i) p(w_i|\nu), \quad (9)$$

where $p(w_i|\nu)$ is the density of a Gamma$(\nu/2, \nu/2)$ distribution, $p(\tilde{l}_i|w_i) = 2\phi_1(\tilde{l}_i; 0, 1)\mathbb{1}(\tilde{l}_i > 0)$ is the density of a constrained standard normal, and $p(\boldsymbol{z}_i|\tilde{l}_i, w_i, \boldsymbol{\theta}) = \phi_d(\boldsymbol{z}_i; \boldsymbol{\delta}\tilde{l}_i w_i^{-1/2}, w_i^{-1}(\bar{\Omega} - \boldsymbol{\delta}\boldsymbol{\delta}^\top))$. Here, $\phi_m(\cdot; \boldsymbol{\mu}, \Sigma)$ denotes the density of a $N_m(\boldsymbol{\mu}, \Sigma)$ distribution, and the indicator function $\mathbb{1}(X) = 1$ if X is true, and zero otherwise. The copula model likelihood can be recovered by integrating over the two latent vectors; that is $p(\boldsymbol{y}|\boldsymbol{\theta}) = \int p(\boldsymbol{y}, \tilde{\boldsymbol{l}}, \boldsymbol{w}|\boldsymbol{\theta}) \mathrm{d}(\tilde{\boldsymbol{l}}, \boldsymbol{w})$.

This extended likelihood is tractable and fast to compute. Evaluation of $F_{Z_1}^{-1}, \ldots, F_{Z_d}^{-1}$ at (7) is undertaken using the spline interpolation method outlined by Smith and Maneesoonthorn (2018) and Yoshiba (2018) that uses the closed form densities $f_{Z_1}, \ldots, f_{Z_d}$. These authors show this approach is highly accurate and is scalable with respect to n. An alternative extended likelihood can be specified using the other generative representation given in Appendix B. However, we found it to be less effective than that adopted here; see Parts C.3 and D.2 of the Web Appendix.

3.2. Priors and Posterior

The copula parameters with a factor decomposition for $\bar{\Omega}$ are $\boldsymbol{\theta} = \{\text{vech}(\widetilde{G}), \boldsymbol{\alpha}, \nu\}$, where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_d)^\top$ is a one-to-one function of $\boldsymbol{\delta}$. Bayesian estimation uses the posterior density $p(\boldsymbol{\theta}|\boldsymbol{y})$, which is the marginal in $\boldsymbol{\theta}$ of the augmented posterior

$$p(\boldsymbol{\theta}, \tilde{\boldsymbol{l}}, \boldsymbol{w}|\boldsymbol{y}) \propto p(\boldsymbol{y}, \tilde{\boldsymbol{l}}, \boldsymbol{w}|\boldsymbol{\theta}) p(\boldsymbol{\theta}) \equiv h(\boldsymbol{\psi}), \quad (10)$$

where $\boldsymbol{\psi} \equiv \{\boldsymbol{\theta}, \tilde{\boldsymbol{l}}, \boldsymbol{w}\}$. The augmented posterior is the product of the extended likelihood at (8) and the prior $p(\boldsymbol{\theta}) = p(\text{vech}(\widetilde{G}))p(\boldsymbol{\alpha})p(\nu)$. We adopt the generalized double Pareto distribution prior for each element of vech$(\widetilde{G})$ suggested by Murray et al. (2013). The prior $\alpha_j \sim N(0, 5^2)$, which places 99% mass on $\delta_j \in (-0.997, 0.997)$. The prior for ν is constrained so that $\nu > 2$, with $(\nu - 2) \sim$ Gamma$(3, 0.2)$. This places 99% mass on the range $\nu \in (3.69, 48.47)$.

An MCMC scheme that produces draws from the augmented posterior of the AC skew-t copula is outlined in Appendix D. However, it is slow and for high d (including when $d = 93$ as in our equity application in Section 6) it does not evaluate the augmented posterior in reasonable computing time. Variational inference is a faster and scalable alternative, as we now discuss.

3.3. Hybrid Variational Inference

VI approximates the augmented posterior using a density $q(\boldsymbol{\psi}) \in \mathcal{Q}$ called the "variational approximation" (VA) from a family of tractable approximations $\mathcal{Q}$. Typically, the VA is obtained by minimizing the Kullback-Leibler divergence (KLD) between the target density $p(\boldsymbol{\psi}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{\psi})p(\boldsymbol{\psi}) = h(\boldsymbol{\psi})$ and its approximation $q(\boldsymbol{\psi})$. It is easily shown that this is equivalent to maximizing the Evidence Lower Bound (ELBO)

$$\mathcal{L} = E_q \left[\log h(\boldsymbol{\psi}) - \log q(\boldsymbol{\psi}) \right]$$

over $q \in \mathcal{Q}$, where the expectation is with respect to $\boldsymbol{\psi} \sim q$.

The selection of families $\mathcal{Q}$ that balance accuracy of the VA and the time required to maximize the ELBO function, is the topic of much current research. For models with a large number of latent variables, such as the augmented posterior at (10), Loaiza-Maya et al. (2022) point out that assuming simple fixed form approximations (such as the widely used mean field approximation) can be very inaccurate. Instead, they suggest using a VA family of the form

$$q_\lambda(\boldsymbol{\psi}) = p(\tilde{\boldsymbol{l}}, \boldsymbol{w}|\boldsymbol{\theta}, \boldsymbol{y}) q_\lambda^0(\boldsymbol{\theta}), \quad (11)$$

where $p(\tilde{\boldsymbol{l}}, \boldsymbol{w}|\boldsymbol{\theta}, \boldsymbol{y})$ is the conditional posterior of the latent variables, and $q_\lambda^0(\boldsymbol{\theta})$ is the density of a fixed form VA with parameters $\boldsymbol{\lambda}$ which we discuss further below.[5] These authors show that when using the VA at (11), the ELBO of the target density $p(\boldsymbol{\psi}|\boldsymbol{y})$ is exactly equal to the ELBO for the target density $p(\boldsymbol{\theta}|\boldsymbol{y})$; that is,

$$\begin{aligned} \mathcal{L} &= E_{q_\lambda} \left[\log h(\boldsymbol{\psi}) - \log q_\lambda(\boldsymbol{\psi}) \right] \\ &= E_{q^0} \left[\log \left(p(\boldsymbol{y}|\boldsymbol{\theta}) p(\boldsymbol{\theta}) \right) - \log q_\lambda^0(\boldsymbol{\theta}) \right]. \end{aligned}$$

Thus, maximizing $\mathcal{L}$ is equivalent to solving the variational optimization for the parameter posterior with the latent variables $\{\tilde{\boldsymbol{l}}, \boldsymbol{w}\}$ integrated out exactly, which makes (11) more accurate than other choices of VA for this target density.

To maximize $\mathcal{L}$ with respect to $\boldsymbol{\lambda}$, stochastic gradient ascent (SGA) is the most frequently used algorithm (Ranganath, Gerrish, and Blei 2014). From an initial value $\boldsymbol{\lambda}^{(0)}$, SGA recursively updates

$$\boldsymbol{\lambda}^{(t+1)} = \boldsymbol{\lambda}^{(t)} + \boldsymbol{\rho}^{(t)} \odot \widehat{\nabla_\lambda \mathcal{L}}\big|_{\boldsymbol{\lambda}=\boldsymbol{\lambda}^{(t)}}, \; t = 0, 1, \ldots \quad (12)$$

until reaching convergence. Here, $\boldsymbol{\rho}^{(t)}$ is a vector of adaptive learning rates set using the momentum method of Zeiler (2012), "$\odot$" is the Hadamard product, and $\widehat{\nabla_\lambda \mathcal{L}}$ is an unbiased estimator of the gradient, which is evaluated at $\boldsymbol{\lambda} = \boldsymbol{\lambda}^{(t)}$. The key to fast convergence and computational efficiency of this SGA algorithm is that $\widehat{\nabla_\lambda \mathcal{L}}$ has low variability and is fast to evaluate. One of the most effective ways to achieve this is to use the re-parameterization trick of Kingma and Welling (2014). For the VA at (11), the re-parameterization is a transformation from $\boldsymbol{\theta}$

[5] In the machine learning literature a parametric density q_λ with parameters $\boldsymbol{\lambda}$ is often called a fixed form density.

to $\boldsymbol{\varepsilon} \sim f_\varepsilon$, where $\boldsymbol{\theta} = \tau(\boldsymbol{\varepsilon}, \boldsymbol{\lambda})$ and τ is a deterministic function. In this case, Loaiza-Maya et al. (2022) show that

$$\nabla_\lambda \mathcal{L} = E_{f_{\varepsilon,\tilde{l},w}} \left[\frac{\partial \boldsymbol{\theta}^\top}{\partial \boldsymbol{\lambda}} \left(\nabla_\theta \log h(\boldsymbol{\psi}) - \nabla_\theta \log q_\lambda^0(\boldsymbol{\theta}) \right) \right], \quad (13)$$

where the expectation is with respect to the density $f_{\varepsilon,\tilde{l},w}(\boldsymbol{\varepsilon}, \tilde{\boldsymbol{l}}, \boldsymbol{w}) = f_\varepsilon(\boldsymbol{\varepsilon}) p(\tilde{\boldsymbol{l}}, \boldsymbol{w}|\boldsymbol{\theta}, \boldsymbol{y})$. Typically only a single draw from $f_{\varepsilon,\tilde{l},w}$ is required to obtain a low variance estimate of the expectation in (13), resulting in a major computational advantage.

For q_λ^0 we use a Gaussian density with a factor model covariance matrix as suggested by Miller, Foti, and Adams (2017) and Ong, Nott, and Smith (2018) with r factors. This is not to be confused with the factor model used to define the copula. Ong, Nott, and Smith (2018) give the transformation τ required for the re-parameterization trick, fast to compute closed form expressions for the derivatives $\frac{\partial \boldsymbol{\theta}^\top}{\partial \boldsymbol{\lambda}}$ and $\nabla_\theta \log q_\lambda^0(\boldsymbol{\theta})$ in (13), along with MATLAB routines for their evaluation. The derivative

$$\nabla_\theta \log h(\boldsymbol{\psi}) = (\nabla^\top_{\text{vech}(\widetilde{G})} \log h(\boldsymbol{\psi}), \nabla^\top_\alpha \log h(\boldsymbol{\psi}), \nabla^\top_\nu \log h(\boldsymbol{\psi}))^\top,$$

is specific to the target density $p(\boldsymbol{\psi}|\boldsymbol{y})$, which is the augmented posterior at (10) here. Table C1 in Appendix C provide this gradient, which is evaluated recursively from the bottom of the columns upwards. Its derivation is given in Part B of the Web Appendix, and uses the trace operator to express the gradient in a computationally efficient directional derivative form. The gradient can also be computed using automatic differentiation, although this is much slower.

Loaiza-Maya et al. (2022) call an approach that combines the VA at (11) with SGA optimization and the re-parameterized gradient at (13), a "hybrid VI" method. This is because it nests an MCMC step within a well-defined stochastic optimization algorithm for solving the VI problem. Algorithm 1 details our proposed hybrid VI algorithm for estimating the AC skew-t copula. At Step (b) a small number of Gibbs steps that first draw from $\tilde{\boldsymbol{l}}|\boldsymbol{w}, \boldsymbol{\theta}, \boldsymbol{y}$, and then from $\boldsymbol{w}|\tilde{\boldsymbol{l}}, \boldsymbol{\theta}, \boldsymbol{y}$, are used as outlined in Appendix D. We demonstrate that this works well here, although other methods, such as the Hamiltonian Monte Carlo sampler of Hoffman and Gelman (2014), can also be used at this step. The output of the algorithm is $q_{\hat{\lambda}}^0(\boldsymbol{\theta})$, which is usually called the "variational posterior", because it is the optimal VA to the parameter posterior $p(\boldsymbol{\theta}|\boldsymbol{y})$.

Algorithm 1 Hybrid VI for AC Skew-t Copula

Initiate $\boldsymbol{\lambda}^{(0)}$ and set $t = 0$
repeat
(a) Generate $\boldsymbol{\theta}^{(t)}$ using its re-parameterized representation
(b) Generate $\{\tilde{\boldsymbol{l}}^{(t)}, \boldsymbol{w}^{(t)}\} \sim p(\tilde{\boldsymbol{l}}, \boldsymbol{w}|\boldsymbol{\theta}^{(t)}, \boldsymbol{y})$
(c) Compute the gradient $\widehat{\nabla_\lambda \mathcal{L}}$ using (13) evaluated at the values $\{\boldsymbol{\lambda}^{(t)}, \boldsymbol{\theta}^{(t)}, \tilde{\boldsymbol{l}}^{(t)}, \boldsymbol{w}^{(t)}\}$
(d) Compute step size $\boldsymbol{\rho}^{(t)}$ using an automatic adaptive method (e.g., an ADA method)
(e) Set $\boldsymbol{\lambda}^{(t+1)} = \boldsymbol{\lambda}^{(t)} + \boldsymbol{\rho}^{(t)} \circ \widehat{\nabla_\lambda \mathcal{L}}(\boldsymbol{\lambda}^{(t)})$
(f) Set $t = t + 1$
until either a stopping rule is satisfied or a fixed number of steps is taken
Set $\widehat{\boldsymbol{\lambda}} = \frac{1}{500} \sum_{s=1}^{500} \boldsymbol{\lambda}^{(t+1-s)}$, and the variational parameter posterior to $q_{\hat{\lambda}}^0(\boldsymbol{\theta})$

4. Simulation

We first demonstrate the efficacy of the VI method using simulated data from lower dimensional examples where the exact posterior can be calculated using MCMC to measure accuracy.

4.1. Design

A sample of size $n = 1024$ is generated from each of two AC skew-t factor copulas. The first is a low-dimensional single factor model ($d = 5, k = 1$; Case 1), and the second is higher dimensional five factor model ($d = 30, k = 5$; Case 2). The copula parameters were obtained from fitting these models to returns data; see Part C of the Web Appendix for details of the data generating processes.

4.2. Approximation Accuracy

For both examples, the variational posterior is compared to the exact posterior $p(\boldsymbol{\theta}|\boldsymbol{y})$ computed using the (slower) MCMC scheme in Appendix D. MCMC evaluates the posterior up to an arbitrary error, which we made small using a large Monte Carlo sample size. A skew-t copula with very different parameter values can have similar densities $c(\boldsymbol{u})$ and thus dependence structures. Therefore, estimation accuracy is best measured using pairwise dependence metrics of the copula for all possible variable pairs, rather than parameter values. The metrics computed here include the quantile dependencies computed at the 1% and 5% quantiles, and the Spearman correlation. For the pair (Y_i, Y_j), the latter is $\rho^S_{ij} = 12 \int \int C_{ij}(u'_i, u'_j) du'_i du'_j - 3$, with C_{ij} the bivariate marginal copula for (Y_i, Y_j) from the skew-t copula evaluated as in Appendix A.

The exact and approximate posterior means of the metrics are evaluated by averaging their values computed at Monte Carlo draws from $p(\boldsymbol{\theta}|\boldsymbol{y})$ and $q_{\hat{\lambda}}^0(\boldsymbol{\theta})$, respectively. Figure 2 plots the exact and variational posterior means of the pairwise Spearman correlations. Figures 3 and 4 do the same for the quantile dependence metrics at the 5% level. They show that VI produces a copula estimate with a dependence structure that is close to that of the exact posterior; further results are given in Part C of the Web Appendix. Variational inference is an approximate method and it typically gives estimates of the second moments of the Bayesian posterior that are less accurate than the posterior mean, which we also observe here, although this has limited impact on prediction; see Frazier, Loaiza-Maya, and Martin (2023) for a discussion.

4.3. Approximation Speed

In general, we implement Algorithm 1 using 25 Gibbs draws at step (b). To illustrate that the algorithm is robust to the number

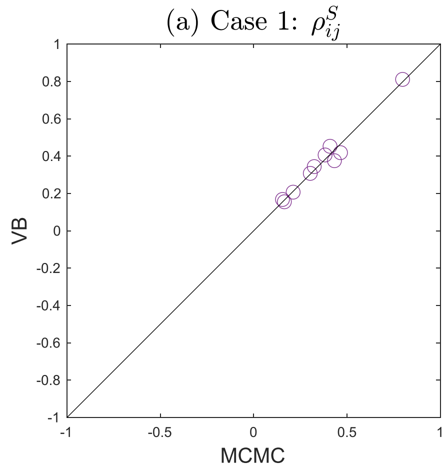

Figure 2. Comparison of the posterior mean estimates of the pairwise Spearman correlations ρ^S_{ij} computed exactly using MCMC (horizontal axis) and approximately using VI (vertical axis). Panel (a) plots these for the five-dimensional skew-t copula in Case 1, and panel (b) for the 30-dimensional skew-t copula in Case 2.

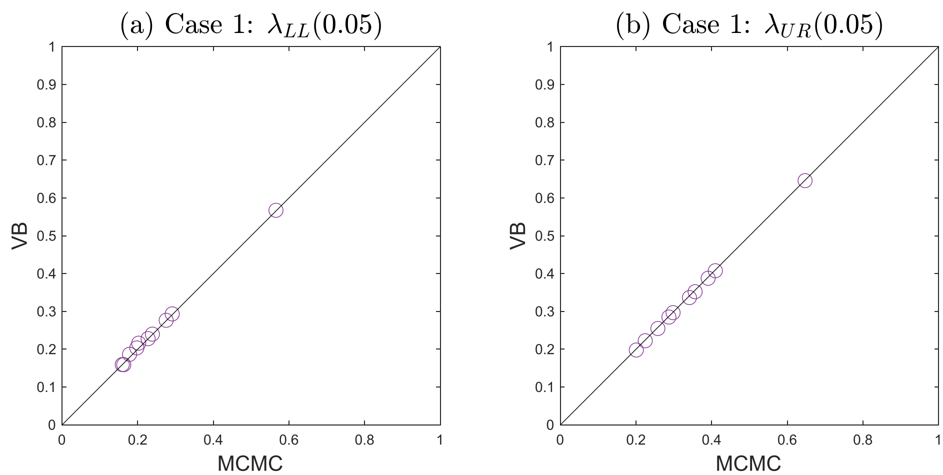

Figure 3. Comparison of the posterior mean estimates of the pairwise 5% quantile dependence metrics $\lambda_{LL}(0.05)$ and $\lambda_{UR}(0.05)$ for Case 1 computed exactly using MCMC (horizontal axis) and approximately using VI (vertical axis).

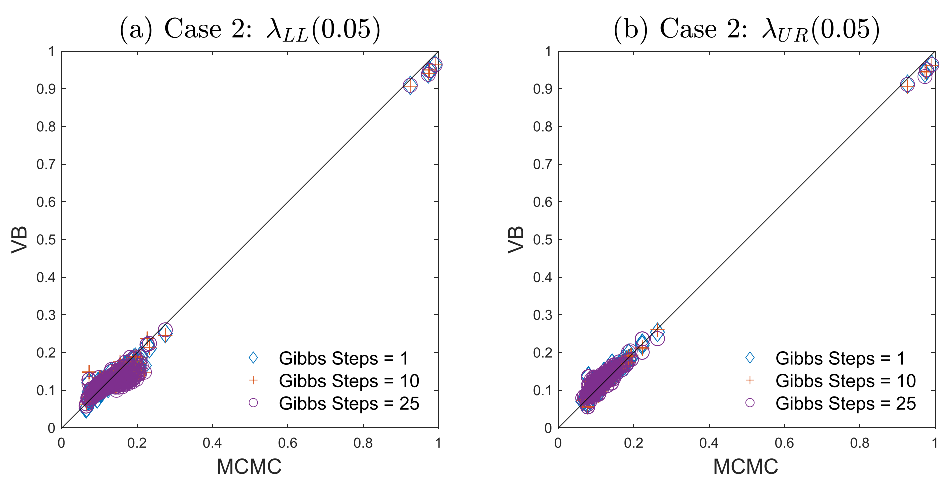

Figure 4. Comparison of the posterior mean estimates of pairwise metrics $\lambda_{LL}(0.05)$ and $\lambda_{UR}(0.05)$ for Case 2 computed exactly using MCMC (horizontal axis) and approximately using VI (vertical axis). Results are given for three VI implementations where 1, 10, and 25 random walk Metropolis-Hastings draws were used to obtain a draw at step (b) of Algorithm 1, with almost identical values.

of draws, Figure 4 plots estimates for three implementations with 1, 10, and 25 Gibbs draws. The VI estimates are very similar, which is consistent with the finding of Loaiza-Maya et al. (2022) for some other models.

Table 1 compares the computation time for MCMC and VI (with 25 Gibbs draws) using a standard laptop. The diagnostic of Geweke (1991) was used to determine the number of MCMC sweeps required. The number of steps in the SGA algorithm was

Table 1. Computation times to estimate skew-t copulas for simulated datasets.

	Time/step (seconds)	Number of steps	Total time (hr)
	Case 1 (Dimension $d = 5$ with $k = 1$ factor)		
MCMC	0.207	27,000	1.553
VI	0.268	5000	0.372
	Case 2 (Dimension $d = 30$ with $k = 5$ factors)		
MCMC	1.580	33,000	14.480
VI	1.446	5000	2.008

NOTE: Computation times in MATLAB for a 2018 MacBook Pro (with 2.3GHz Intel i5 CPU) laptop.

5000, which is a conservative number as judged by the stability of the optimal VA; for example, in Case 1 the VA based on only 2500 steps is very similar. The VI estimator is faster, and its computational advantage grows with d and k. Finally, in the empirical work in Section 6 where $d = 93$, we found the VI estimator stable, whereas the MCMC algorithm in Appendix D exhibited poor mixing and was unable to estimate the model in reasonable time.

4.4. Choice of Data Augmentation

Our approach uses the generative representation at (6) (labelled "GR2") to construct a variational data augmentation method. One drawback is that a nested (albeit fast) MCMC sampler to draw from $p(\tilde{\boldsymbol{l}}, \boldsymbol{w}|\boldsymbol{\theta}, \boldsymbol{y})$ is required at step (b) of Algorithm 1. An alternative generative representation "GR1" in Appendix B with latent vector $\boldsymbol{l} = (l_1, \ldots, l_n)^\top$ can also be used to construct another hybrid VI algorithm that targets $p(\boldsymbol{l}, \boldsymbol{\theta}|\boldsymbol{y})$, where at step (b) direct generation from $p(\boldsymbol{l}|\boldsymbol{\theta}, \boldsymbol{y})$ is possible without the use of a nested MCMC sampler. We implemented this alternative hybrid VI algorithm based on GR1, but found it to be less efficient than that based on GR2; see Part C.3 of the Web Appendix for more details. Additional discussion of why GR2 is an effective target distribution is also given in Part D.2 of the Web Appendix.

5. Asymmetric Dependence Between Equity Returns and the VIX

The first example applies the AC skew-t copula model, with $k = 1$ factor and fitted using the proposed VI method, to two equity returns series and the VIX. The latter is an index of overall market volatility published by the Chicago Board of Exchange (Britten-Jones and Neuberger 2000). The objective is to illustrate the flexibility of the dependence structure of the AC skew-t copula.

5.1. Data Description and Marginal Models

Trade prices for Bank of America (BAC) and JP Morgan (JPM) were obtained from Revinitiv DataScope Select database. For each equity, the $N = 26$ fifteen min returns between 09:30 and 16:00 EDT of each trading day are constructed as follows. If $P_{\tau,t}$ denotes the last trade price of a given equity in the τ^{th} 15 min interval on trading day t, then the return for that interval is $r_{\tau,t} = \log(P_{\tau,t}) - \log(P_{\tau-1,t})$, with $P_{0,t} \equiv P_{N,t-1}$ the last trade price on the previous day. Days where trading is suspended are removed.

We follow Patton (2006) and many others by using the copula to capture cross-sectional dependence in the conditional distribution of the financial variables. The intraday GARCH model of Engle and Sokalska (2012) is used as a marginal model for the returns on each equity, which decomposes the return into components

$$r_{\tau,t} = (h_t s_\tau)^{1/2} \sigma_{\tau,t} \varepsilon_{\tau,t} \,.$$

Here, $h_t = \sum_{\tau=1}^{N} (r_{\tau,t})^2$ is the realized variance of day t, $s_\tau = \frac{1}{T} \sum_{t=1}^{T} (r_{\tau,t})^2 / h_t$ is an estimate of the diurnal pattern in volatility for the τth interval, and the conditional volatility movement

$$\sigma^2_{\tau,t} = \beta_0 + \beta_1 \epsilon^2_{\tau-1,t} + \beta_2 \sigma^2_{\tau-1,t} \,,$$

with $\epsilon_{\tau,t} = r_{\tau,t}/(h_t s_\tau)^{1/2}$, $\epsilon_{0,t} = \epsilon_{N,t-1}$ and $\sigma_{0,t} = \sigma_{N,t-1}$. In our analysis, we assume that the innovation $\varepsilon_{\tau,t} \sim t_1(0, 1, \tilde{\nu})$, and estimate the parameters $(\beta_0, \beta_1, \beta_2, \tilde{\nu})$ for each equity using maximum likelihood. The copula data is computed as $T(\varepsilon_{\tau,t}; \tilde{\nu})$ at the model estimates.[6]

Fifteen minute observations of the VIX were also obtained from the Revinitiv DataScope Select database. The VIX marginal model is a first order autoregression with a nonparametric disturbance estimated using a kernel density estimator.

5.2. Empirical Results

The copula model was fit to data from two periods of 40 trading days: a low volatility period between October 24 and December 22, 2017, and a high volatility period between February 11 and April 15, 2020.[7] Table 2 reports the estimates of the copula parameters, correlations and asymmetry measures. The two equity returns are positively correlated, whereas they are both negatively correlated with the VIX. There are strong asymmetries in the quantile dependencies, which change in direction and scale from the low to high volatility periods. For example, between the two equity returns, $\Delta_{\text{Major}}(0.05)$ changes from 0.169 to -0.279. This suggests that during the period of high market volatility, negative returns are more dependent than positive returns. Whereas, between the equity returns and the VIX, $\Delta_{\text{Minor}}(0.05)$ changes from $-0.059, -0.061$ to $0.240, 0.211$. This suggests that during the period of high market volatility there is an increase in the dependence between high VIX values and negative equity returns.

Figures 5 and 6 further summarize the dependence structure of the estimated copula for the low and high market volatility periods, respectively. Following Oh and Patton (2013), panels (d, g, h) give "quantile dependence plots" for the pairs BAC-VIX, JPM-VIX and JPM-BAC, respectively. These visualize asymmetric dependence along the major diagonal by plotting the quantile dependencies $\lambda_{\text{LL}}(u)$ & $\lambda_{\text{UR}}(1-u)$ against u (blue & red lines), and along the minor diagonal by plotting $\lambda_{\text{LR}}(u)$ & $\lambda_{\text{UL}}(1-u)$ versus u (yellow & purple lines). The dependence between the two equity

[6] To match this with the copula model notation in Section 3.1, note that if $i = (N-1)t+\tau$ (so that $i = 1, 2, \ldots, NT$) and the equity return is the jth marginal, then $y_{ij} = r_{\tau,t}$ and $F_{Y_j}(y_{ij}) = T(\varepsilon_{\tau,t}; \tilde{\nu})$ in (7).

[7] We select these because the first period contains the lowest VIX value from 2017 to 2021, whereas the second contains the highest VIX value that corresponds to the COVID-19 crash.

Table 2. Estimated AC skew-t copula with $d = 3$ for the BAC-JPM-VIX example.

	Low volatility period (October 24, 2017–December 22, 2017)						
	Correlation		5% Quantile asymmetry		Parameters ($\boldsymbol{\theta}$)		
Pair	Spearman	Kendall	$\Delta_{Major}(0.05)$	$\Delta_{Minor}(0.05)$	ρ	(δ_1, δ_2)	ν
BAC – VIX	−0.322	−0.219	0.003	−0.059	−0.459	0.987, −0.422	
	(0.029)	(0.021)	(0.003)	(0.023)	(0.036)	(0.041), (0.006)	
JPM – VIX	−0.355	−0.242	0.000	−0.061	−0.484	0.829, −0.422	25.77
	(0.041)	(0.029)	(0.001)	(0.013)	(0.037)	(0.034), (0.006)	(4.76)
JPM – BAC	0.759	0.566	0.169	0.000	0.881	0.829, 0.987	
	(0.028)	(0.025)	(0.041)	(0.000)	(0.024)	(0.034), (0.041)	
	High volatility period (February 11, 2020–April 15, 2020)						
	Correlation		5% Quantile asymmetry		Parameters ($\boldsymbol{\theta}$)		
Pair	Spearman	Kendall	$\Delta_{Major}(0.05)$	$\Delta_{Minor}(0.05)$	ρ	(δ_1, δ_2)	ν
BAC – VIX	−0.536	−0.381	−0.019	0.240	−0.697	−0.987, 0.649	
	(0.024)	(0.019)	(0.004)	(0.030)	(0.023)	(0.004), (0.030)	
JPM – VIX	−0.547	−0.393	−0.008	0.211	−0.708	−0.865, 0.649	4.11
	(0.028)	(0.023)	(0.002)	(0.019)	(0.023)	(0.025), (0.030)	(0.34)
JPM – BAC	0.805	0.620	−0.279	−0.002	0.911	−0.865, −0.987	
	(0.023)	(0.024)	(0.038)	(0.001)	(0.017)	(0.025), (0.004)	

NOTE: The variational posterior means, along with standard deviations in parentheses below, of the correlation coefficients (Spearman, and Kendall), 5% quantile asymmetry metrics defined in the text, and copula parameters are reported. Results are given for two windows of 40 days of 15 min data that correspond to low and high market volatility periods.

returns in the major diagonal is positively asymmetric during the low volatility period in Figure 5(h), and negatively asymmetric during the high volatility period in Figure 6(h). In contrast, dependence in the minor diagonal between the returns on both equities and the VIX is close to symmetric in the low volatility period in Figure 5(d) and (g), but positively asymmetric during the high volatility period in Figure 6(d) and (g), as also indicated by the metrics in Table 2.

To visualize the flexibility of the predictive distributions from the copula model, we compute these for the first 15-min trading interval of the following day of each sample, which are December 23, 2017 (low volatility period) and April 16, 2020 (high volatility period). Figures 5 and 6 plot the marginal predictive densities in panels (a, e, i), and the bivariate slices of the joint predictive density in panels (b, c, f). In the high volatility period both the tails and asymmetric dependence of the distributions are greatly accentuated, compared to the low volatility period.

For comparison, we also fit the SDB and GH skew-t copula models with the same marginals. To fit the former we used the MCMC algorithm outlined by Smith, Gan, and Kohn (2010), and for the latter we used the code provided by Oh and Patton (2023). For SDB and AC we use $k = 1$ factors for $\bar{\Omega}$, and for the GH we use 1 global factor plus 3 group-specific factors. Part D.1 of the Web Appendix reports the dependence metrics and parameter estimates for both copulas. We observe that all copulas have similar correlation estimates, but both SDB and GH exhibit much lower asymmetry in dependence compared to AC. Moreover, during the second period $\hat{\nu} = 4.11$ for AC, whereas $\hat{\nu} = 20.5$ and 21.2 for SDB and GH, so that the AC skew-t copula also captures higher extremal tail dependence.

6. S&P100 Portfolio

We now apply the AC skew-t copula model to the 15 min returns on the constituents of the S&P100 index for the period January 1, 2017 to December 31, 2021, which is our main application. Only the $d = 93$ equities that are listed over the entire period are used in our empirical analysis. The returns were constructed as in the previous example, and the same marginal models used. Rolling windows of width $T = 40$ trading days are employed, and the copula model is estimated in each window using $n = 26 \times 40 = 1040$ intraday returns for each equity. The model is refit every 20 trading days, resulting in 60 overlapping windows. Our empirical study focuses on three research questions. The first is whether adopting the AC skew-t copula improves the accuracy of 15 min ahead density forecasts of portfolio returns, relative to benchmarks. The second is what is the degree of asymmetric dependence over the five year period of our data. The third question is whether, or not, exploiting asymmetry in dependence when forming investment portfolios can improve returns.

6.1. Density Forecasting

We consider 15 min (i.e., one-step-ahead) density forecasts of the return on a market value weighted portfolio of the equities. The weights are recalculated at the same frequency as the model estimation; that is, every 20 trading days. The forecasts are Bayesian posterior predictive distributions for every 15 min period from March 3, 2017 to January 25, 2022, giving a total of $1200 \times 26 = 31{,}200$ forecasts. Each of these is evaluated by drawing 10,000 iterates from the posterior predictive joint distribution of returns (see Part E.1 of the Web Appendix), from which draws of the portfolio return are obtained. Kernel density estimates of these are the density forecasts of the portfolio return.

Two density forecasting metrics are calculated: the log-score (LS) and the continuous ranked probability score (CRPS) of Gneiting, Balabdaoui, and Raftery (2007). Higher values of the LS and lower values of the CRPS correspond to increased accuracy. Table 3 reports the mean of these over all portfolio return density forecasts. Results are given for different numbers of factors k in the copula model, and estimated using VI where the

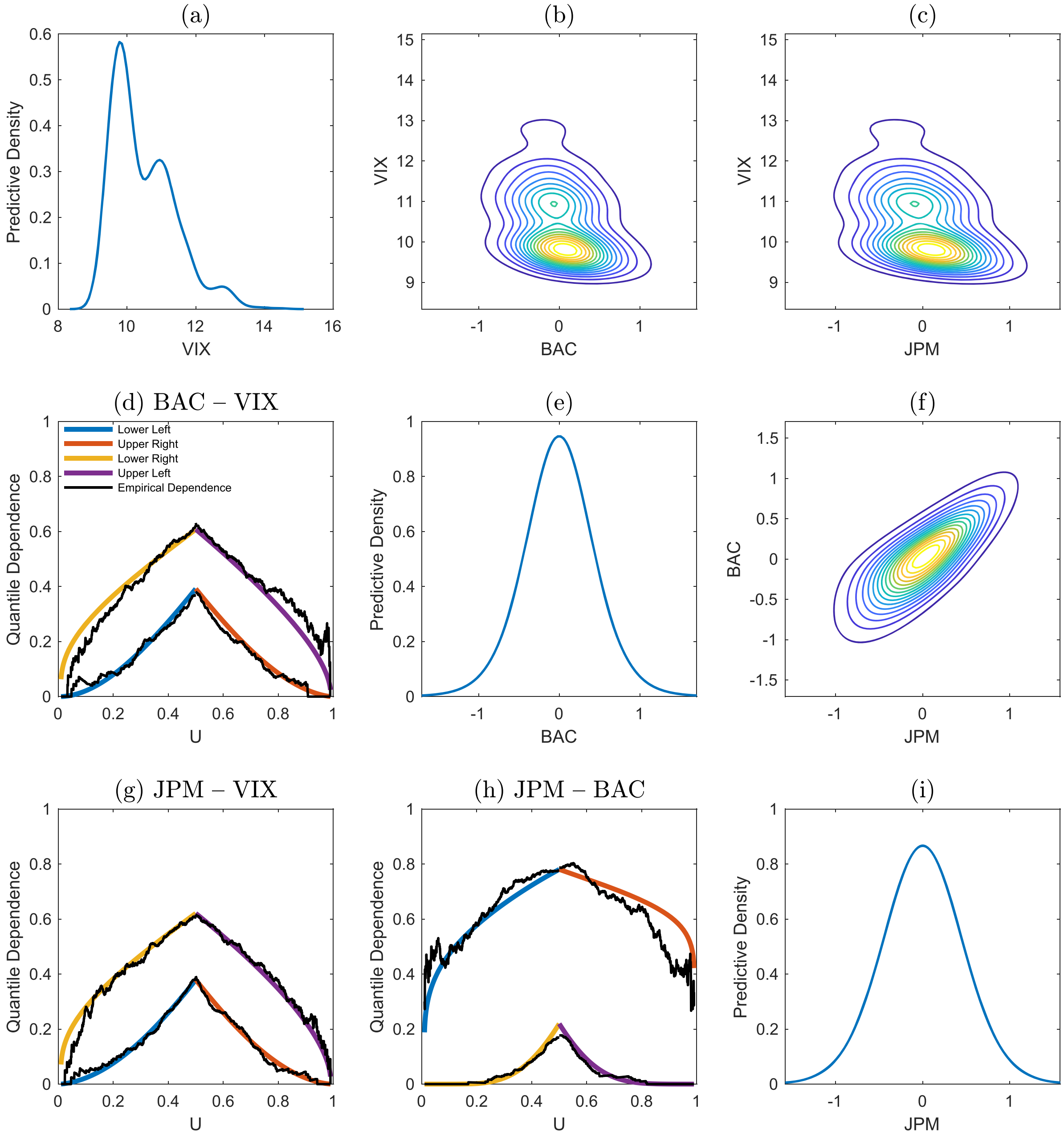

Figure 5. *Low Volatility Period* results from the AC skew-t copula model for variables (VIX, BAC, JPM). Panels (a, e, g) give marginal predictive distributions for the series, and panels (b, c, f) give bivariate predictive distributions for the variable pairs, all on December 23, 2017 at 09:45. Estimated quantile dependence plots for the pairs are given in panels (d) BAC–VIX, (g) JPM–VIX, and (h) JPM–BAC. Each quantile dependence plot visualizes asymmetry along the major diagonal by plotting $\lambda_{LL}(u)$ & $\lambda_{UR}(1-u)$ against u (blue and red lines respectively), and along the minor diagonal by plotting $\lambda_{LR}(u)$ & $\lambda_{UL}(1-u)$ versus u (yellow and purple lines respectively). Empirical quantile dependence plots are given for comparison (black lines).

covariance matrix of the Gaussian VA q_λ^0 has different numbers of factors r. Note that while setting $r=0$ corresponds to a fully factorized VA for q_λ^0, we stress that the VA q_λ at (11) for the target density $p(\boldsymbol{\psi}|\boldsymbol{y})$ is not of a mean field type. We make two observations. First, using higher values of k (i.e., $k=10$ and $k=15$) increases accuracy. This result is consistent with Oh and Patton (2023), who found latent group factors, over-and-above a single global factor, improve the dependence structure in their skew-t copula model. Second, the results are similar for $r=3,5,10$, which is consistent with Ong, Nott, and Smith (2018) and subsequent authors who find lower values of r typically work well. We focus on results for $k=10$ factors (which corresponds to a total of 979 copula parameters in $\boldsymbol{\theta}$) with $r=3$ for the VA. Estimation for a single window takes 29.7 hr using 20,000 steps of the SGA algorithm on a 2018 MacBook Pro.

The predictive accuracy of the AC skew-t copula model is compared with five benchmarks models. The first four are copula models with the same marginals. Three copulas are subtypes of the AC skew-t copula; namely, the AC skew-normal copula ($\nu \to \infty$), the t-copula ($\boldsymbol{\delta} = \mathbf{0}$), and the Gaussian copula ($\nu \to \infty$ and $\boldsymbol{\delta} = \mathbf{0}$). The fourth copula is the GH skew-t copula, implemented using the code and best fitting factor structure in Oh and Patton (2023); see Part E.3 of the Web Appendix for implementation details. The fifth benchmark model is an intraday Constant Conditional Correlation GARCH (CCC-GARCH) (Bollerslev 1990) specified as in Section 5.1, but

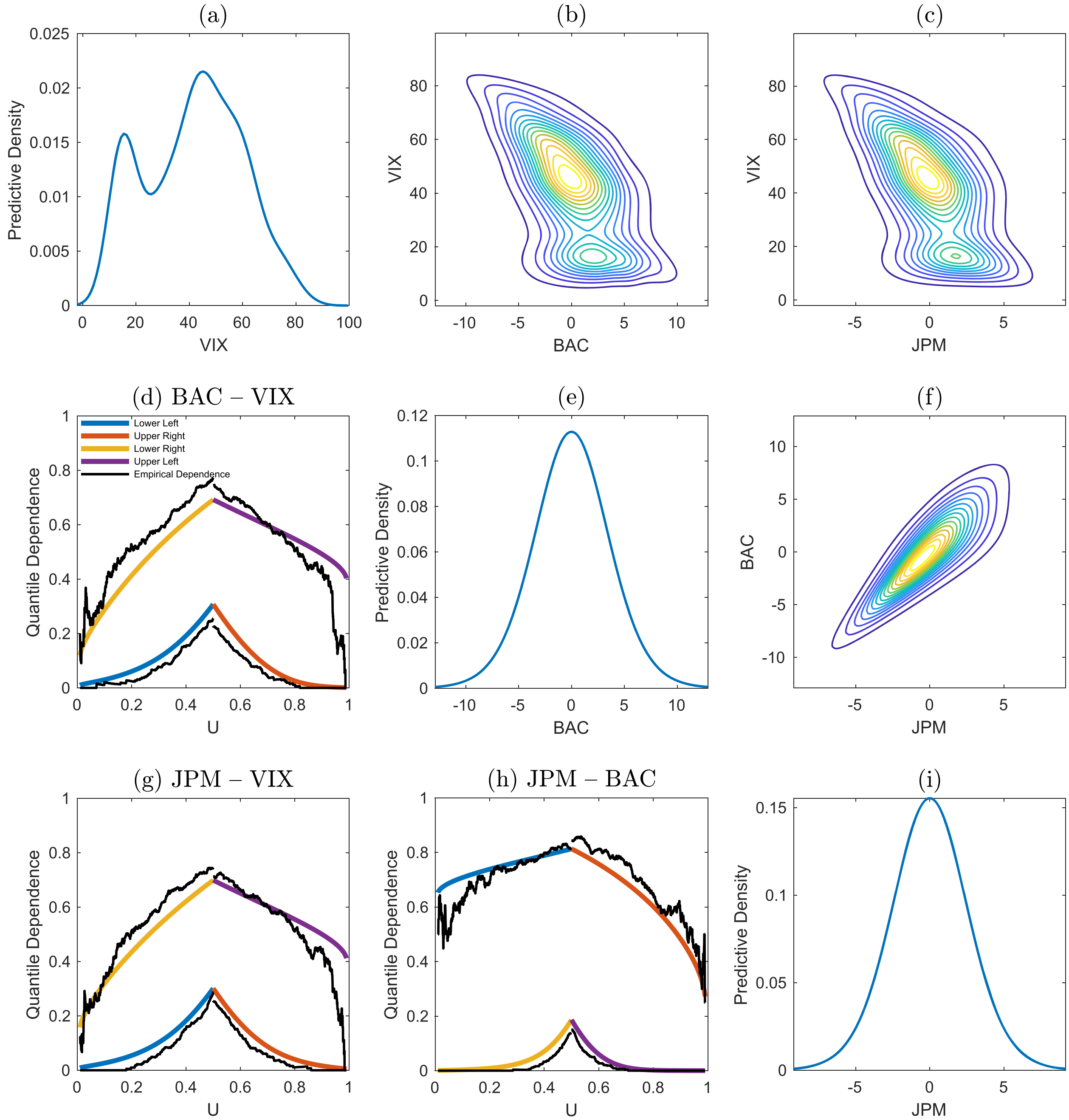

Figure 6. *High Volatility Period* results from the AC skew-t copula model for variables (JPM, BAC, VIX). Panels (a, e, g) give marginal predictive distributions for the series, and panels (b, c, f) give bivariate predictive distributions for the variable pairs, all on April 16, 2020 at 09:45. Estimated quantile dependence plots for the pairs are given in panels (d) BAC–VIX, (g) JPM–VIX, and (h) JPM–BAC. Each quantile dependence plot visualizes asymmetry along the major diagonal by plotting $\lambda_{LL}(u)$ & $\lambda_{UR}(1-u)$ versus u (blue and red lines respectively), and along the minor diagonal by plotting $\lambda_{LR}(u)$ and $\lambda_{UL}(1-u)$ versus u (yellow and purple lines respectively). Empirical quantile dependence plots are given for comparison (black lines).

Table 3. Mean log-score and CRPS of the equally-weighted S&P100 portfolio return forecasts.

	Panel A: mean log-score				Panel B: mean CRPS			
# Copula factors k	# Variational factors r				# Variational factor r			
	0	3	5	10	0	3	5	10
1	6.097	6.179	6.111	6.325	8.468	8.462	8.456	8.404
3	6.856	6.744	6.853	6.890	8.334	8.344	8.345	8.307
5	7.106	7.131	7.117	7.106	8.277	8.276	8.254	8.285
10	7.156	**7.185**	7.132	7.167	8.251	**8.246**	8.252	8.288
15	7.132	7.174	7.143	7.193	8.246	8.246	8.246	8.248

NOTE: LS values are multiplied by 10, and CRPS values are multiplied by 100, for presentation. The means are computed over all 15 min density forecasts between March 03, 2017 and January 25, 2020. Higher values of LS and lower values of CRPS correspond to greater accuracy. Figures in bold correspond to the optimal values of k and r.

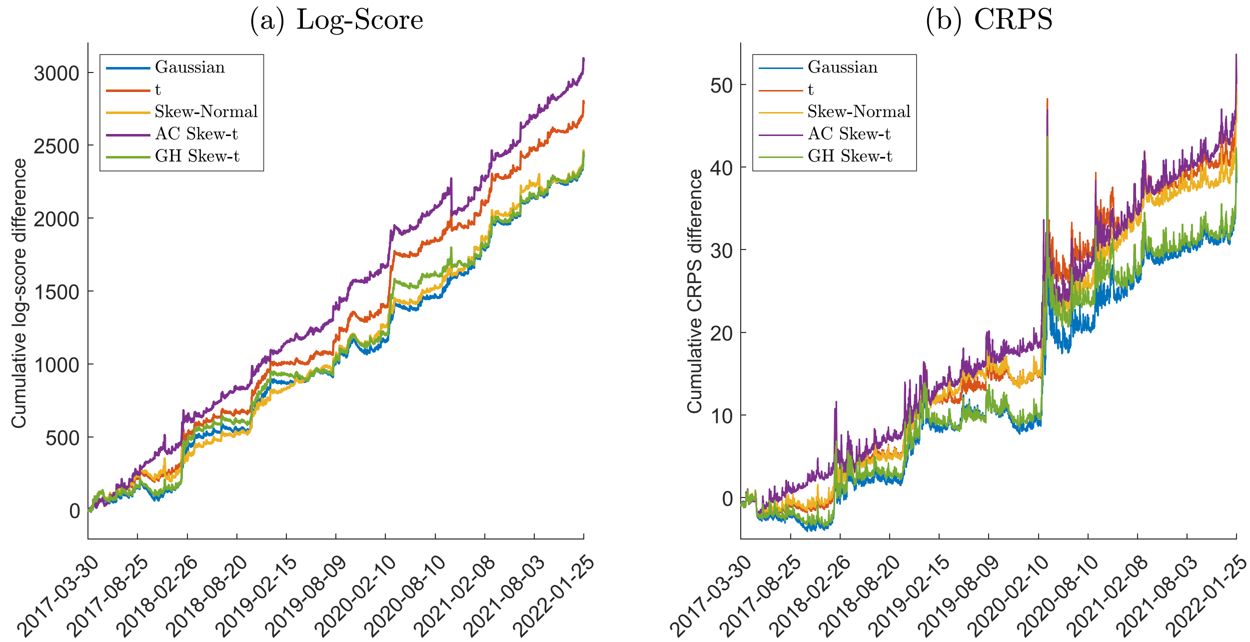

Figure 7. Cumulative difference in CRPS and LS scores between each of (i) Gaussian copula, (ii) t-copula, (iii) AC skew-normal copula, (iv) AC skew-t copula, and (v) static GH skew-t coupla, and the baseline CCC-GARCH model. All copulas are factor copulas as described in the text.

with $\varepsilon_{\tau,t} \sim N(0,1)$ and cross-asset correlation estimated using the sample correlation. We treat the CCC-GARCH model as the baseline, and compute the cumulative difference of each metric for each copula model and this baseline model. Figure 7 plots the differences over the validation period for (a) LS and (b) CRPS. The positive slope indicates that the five copula models all outperform the baseline model throughout the period, and the AC skew-t factor copula dominates the other factor copula models on both metrics.

6.2. *Asymmetric Dependence*

If $\Delta_{\text{Major},i,j}(u)$ denotes asymmetric quantile dependence for dimensions i, j, then we propose

$$\mathcal{T}_{i,j} = \int_{\epsilon}^{0.5} \left|\Delta_{\text{Major},i,j}(u)\right| \mathrm{d}u\,, \tag{14}$$

as a measure of Total Asymmetric Dependence (TAD) in the major diagonal. We set $\epsilon = 0.001$ and compute the integral over all quantiles numerically. Asymmetry along the major diagonal is measured, rather than the minor diagonal, because most equity pairs exhibit positive overall dependence. When dependence is symmetric $\mathcal{T}_{i,j} = 0$, while the maximum value of the TAD for any pair of variables in the AC skew-t copula can be computed as $\max_{\rho,\delta}\{\mathcal{T}_{i,j}\} = 0.193$ with $\nu = 2$.

This metric is computed for all $93 \times 92/2 = 4278$ pairs of equities and for the AC skew-t copulas fitted at each of the 60 windows. Figure 8 summarizes the evolution of the $\mathcal{T}_{i,j}$ values over the estimation windows as follows. The gray shaded area is the interval between $\min_{i,j}\{\mathcal{T}_{i,j}\}$ and $\max_{i,j}\{\mathcal{T}_{i,j}\}$, while the dashed black line depicts the mean value across all equity pairs. The windows with the highest level of TAD are December 18, 2018 to February 15, 2019, and February 11, 2020 to April 15, 2020. The former window corresponds to an escalating trade war between the U.S. and China, while the second window corresponds to the height of the COVID-19 equity market crash. The $\mathcal{T}_{i,j}$ values for the two pairs Apple–Microsoft and Google–Tesla are also plotted, illustrating how the level of TAD can vary substantially over different equity pairs.

Finally, to further highlight the heterogeneity in asymmetric quantile dependence, Figure 9 depicts $\Delta_{\text{Major},i,j}(0.01)$ over the final ten non-overlapping windows for the ten equities with the largest market capitalization. Positive values of $\Delta_{\text{Major},i,j}(0.01)$ (green) indicate upside tail dependence, while negative values of $\Delta_{\text{Major},i,j}(0.01)$ (red) indicate downside tail dependence. In particular, during the early stages of the COVID-19 crash in panel (h), the copula captures extreme downside tail dependence across all equity pairs, with the exception of Apple.

6.3. *Portfolio Equity Selection*

Dependence between equity returns is one of the major attributes of asset selection strategies for investment portfolios. Traditionally, equities that are negatively correlated are preferred because they achieve a superior efficient investment frontier. More recently, consideration has been given to "tail risk" of extreme downside losses. For example, Guidolin and Timmermann (2008) and Harvey et al. (2010) proposed portfolio selection based on skewness and/or kurtosis, Giacometti, Torri, and Paterlini (2021) propose a portfolio allocation strategy with penalty terms based on tail risk measures, Zhao (2021) proposes a new portfolio optimization method using a bivariate peak-over-threshold approach; and Bollerslev, Patton, and Quaedvlieg (2022) made use of semi-beta risk measures, separating market upside and downside movements, for portfolio construction. We consider using pairwise asymmetric quantile dependencies from the skew-t copula to construct portfolios.

The following two strategies are considered to select five equity pairs from the 4278 possible:

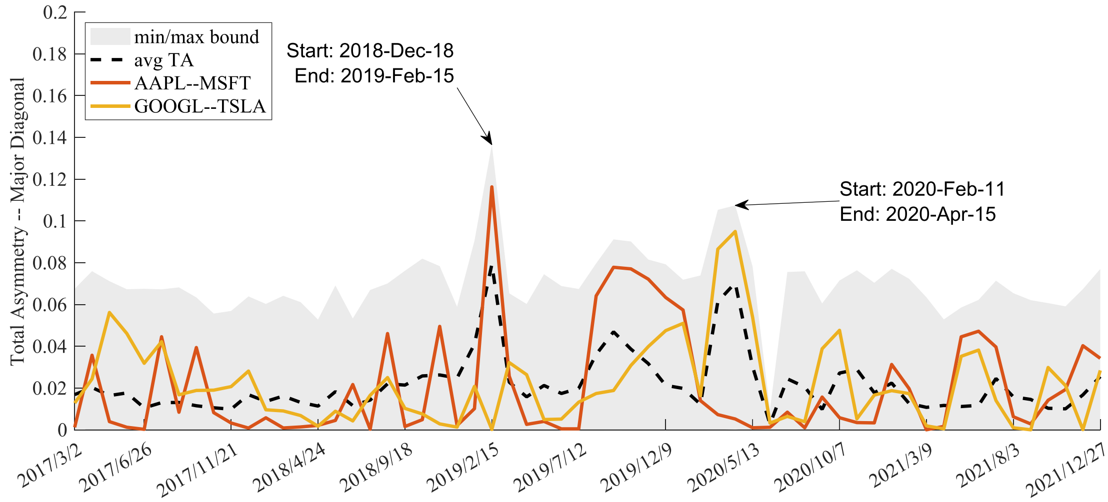

Figure 8. Summary of the total asymmetric dependence metric values $\mathcal{T}_{i,j}$ over the rolling windows in the five year period. The gray shaded area is the interval between $\min_{i,j}\{\mathcal{T}_{i,j}\}$ and $\max_{i,j}\{\mathcal{T}_{i,j}\}$, while the dashed black line depicts the mean value across all equity pairs. The values of $\mathcal{T}_{i,j}$ for the pairs Apple–Microsoft and Google–Tesla are plotted in red and orange, respectively.

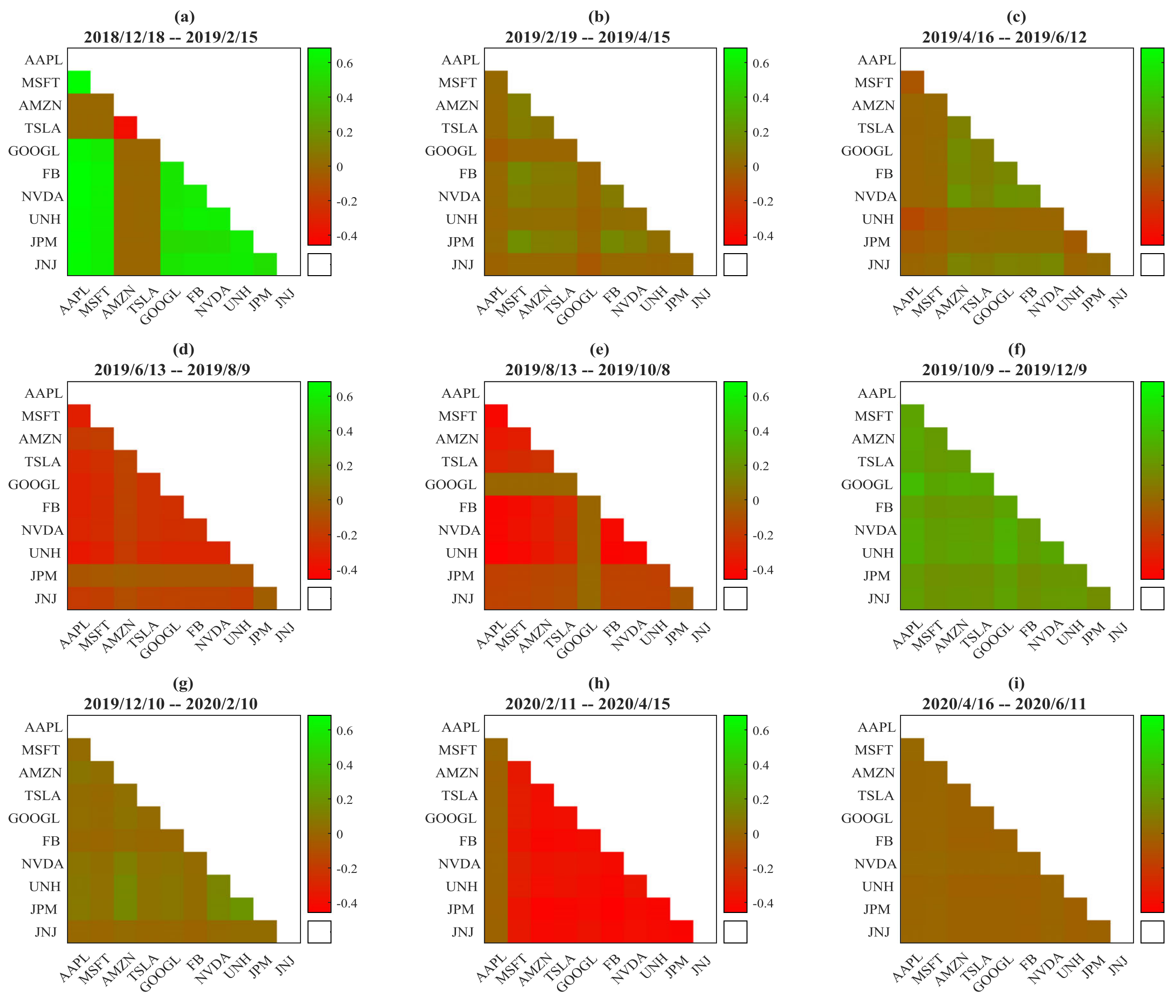

Figure 9. Heatmap of $\Delta_{\text{Major},i,j}(0.01)$ across the 10 equity pairs (i,j) with largest market capitalization between December 18, 2018 and August 19, 2020.

- *Upside Gains*: to maximize upside gains, equities that move together strongly in the upper tail are preferred, so that equity pairs with the largest values of $\Delta_{\text{Major}}(0.05)$ are selected.
- *High Minor Tail Dependence*: to maximize negative tail dependence, equity pairs that maximize the sum of the minor diagonal tail dependencies, $\lambda_{UL}(0.05)+\lambda_{LR}(0.05)$, are selected.

The equities that make up the top five pairs according to both criteria are used to construct portfolios using market-value weights.

Table 4. Backtesting results of the four equity selection strategies.

	93-Equities (benchmark S&P100)	Pearson's correlation (largest negative)	Upside gain (max $\Delta_{\text{Major}}(0.05)$) AC copula	GH copula	High minor tail dependence (max $\lambda_{UL}(0.05)+\lambda_{LR}(0.05)$) AC copula	GH copula
Sharpe ratio	0.293	0.159	0.248	0.155	0.381	0.276
Return (p.a.)	15.736%	12.695%	17.443%	10.606%	26.204%	18.123%
Stdev (p.a.)	15.087%	22.449%	19.759%	19.205%	19.327%	18.425%
VaR(5%)	−6.054%	−8.064%	−7.633%	−8.247%	−5.804%	−7.303%

NOTE: The Sharpe ratio, annualized realized portfolio return and standard deviation, and the 5% Value-at-Risk are reported. "93 Equities" is the S&P 100 portfolio corrected for stock selection. The two asymmetric dependence-based stock selection strategies "Upside Gain" and "High Minor Tail Dependence" are implemented using estimates from both the AC and GH skew-t copulas.

Table 4 reports backtesting results from our equity selection strategies implemented for the period from March 3, 2017 to January 25, 2022. For comparison, we also include the market-value weighted portfolio with all 93 equities (i.e., the S&P100 index corrected for sample selection), along with a strategy selecting the five equity pairs with the most negative Pearson's correlation, as our benchmarks. All strategies are rebalanced with the most current estimates and market values every 20 trading days. The high minor tail dependence strategy performs strongly, even against the benchmark S&P100 portfolio. It has the highest Sharpe ratio and lower downside risk as measured by the Value-at-Risk (VaR) at the 5% level.[8] While this study does not incorporate trading costs, it suggests that trading strategies based on quantile dependencies from our AC skew-t copula model have the potential to improve the risk profile—particularly downside risk—of portfolios. Finally, we repeat the study using the GH skew-t copula, which performed poorly.

7. Discussion

This article makes three main contributions. First, we show that the AC skew-t copula can capture greater asymmetry in dependence than two other types of skew-t copulas. This is important because capturing asymmetry is the sole reason to adopt a skew-t copula over the t-copula. Second, we propose a VI method that can estimate the AC skew-t copula parameters for large panels of financial data. Third, our empirical work shows that the direction and extent of asymmetric dependence can vary over both equity pair and time, and that exploiting this in portfolio formation can improve investment performance. Below, we make some additional comments on the methodology.

While VI methods are increasingly employed in econometrics (see e.g., Loaiza-Maya and Nibbering 2023; Chan and Yu 2022; Gefang, Koop, and Poon 2023) for more complex models the choice of both target and approximating densities is crucial. Standard VI methods cannot be used to approximate the posterior $p(\boldsymbol{\theta}|\boldsymbol{y})$ because it has a complex geometry. However, the augmented posterior $p(\boldsymbol{\psi}|\boldsymbol{y})$ is well approximated using (11) and hybrid VI. While the SGA algorithm is widely used to solve variational optimization problems (e.g., see Ranganath, Gerrish, and Blei 2014) it is the efficient re-parameterization gradient at (13) that makes our VI algorithm fast.

When estimating any implicit copula, it is necessary to compute the quantile functions $F_{Z_j}^{-1}$, for $j = 1, \ldots, d$, at every observation. For the GH skew-t distribution this is difficult and large Monte Carlo samples (e.g., one million draws) are typically used to compute them to a necessary level of accuracy. One way to speed this computation is to set $\boldsymbol{\delta} = \delta(1, 1, \ldots, 1)^\top$, however, this restricts variation in pairwise asymmetric dependence, which is the key focus of our study. This restriction, plus the difficulty in computing the MLE, is likely to contribute to the GH skew-t copula's poor relative performance in Section 6. In contrast, for the AC skew-t distribution $F_{Z_j}^{-1}$ can be computed with high accuracy numerically, and there is no need to restrict $\boldsymbol{\delta}$. Most previous factor copula studies use only one or two global factors, often enriched with an industry or latent group-specific factor. The computational efficiency of our VI method allows for a larger number of global factors to be used. This proves important in our empirical work, where we find 10 global factors improves performance, compared to a smaller number.

Finally, we conclude by mentioning two possible extensions of our work. In our study of intraday returns we use a static copula with a rolling window of width 40 trading days. But for studies of lower frequency returns, allowing for time variation in the AC skew-t copula can be achieved by following previous authors and adopting a dynamic specification. A second promising extension would be to consider generalizations of the AC skew-t distribution underlying the implicit copula to provide greater flexibility in capturing tail dependence. However, in both extensions identification of the copula parameters and an effective estimation methodology would need to be considered.

Appendix A: Quantile Dependence

For each skew-t copula, the pairwise quantile dependence metrics in Section 2.3 are evaluated from their bivariate copula function $C(u_1, u_2) = F_{Z_1,Z_2}\left(F_{Z_1}^{-1}(u_1), F_{Z_2}^{-1}(u_2)\right)$. Computation of $F_{Z_j}^{-1}(u_j)$ is discussed in the manuscript for each distribution. The function F_{Z_1,Z_2} is obtained for the SDB skew-t distribution by numerical integration. For the AC skew-t distribution it is computed as the trivariate student t distribution function $F_{Z_{\text{AC}}}\left((z_1, z_2)^\top; \boldsymbol{\delta}, \bar{\Omega}, \nu\right) = 2F_t\left((z_1, z_2, 0)^\top; \Omega^*, \nu\right)$, where

$$\Omega^* = \begin{pmatrix} \bar{\Omega} & -\boldsymbol{\delta} \\ -\boldsymbol{\delta}^\top & 1 \end{pmatrix},$$

which can be derived from (4). The GH skew-t distribution function cannot be computed in closed form, and numerical integration of its

[8] The risk-free rate used to calculate the Sharpe Ratio is the Federal Reserve Effective Rate, while higher values of VaR(5%) indicate a decrease in downside risk at this quantile.

joint density is difficult. Therefore, a kernel-based approximation based on one million Monte Carlo draws is employed.

Appendix B: Generative Representations for AC Skew-t

We list below two different generative representations for $\boldsymbol{Z}_{\mathrm{AC}}$ that can be derived from (4).

GR1: The first is to generate from the marginal $L \sim t_1(0,1,\nu)$ constrained so that $L > 0$, and then from the conditional $\boldsymbol{Z}_{\mathrm{AC}} = (\boldsymbol{X}|L) \sim t_d\left(\mathbf{0}, \frac{\nu+L^2}{\nu+1}(\bar{\Omega} - \boldsymbol{\delta}\boldsymbol{\delta}^\top), \nu+1\right)$.

GR2: The second uses a scale mixture of normals representation for a t-distribution, where if $W \sim \text{Gamma}(\nu/2, \nu/2)$, then $(\boldsymbol{X}^\top, L)^\top = W^{-1/2}(\tilde{\boldsymbol{X}}^\top, \tilde{L})^\top$ with $(\tilde{\boldsymbol{X}}^\top, \tilde{L})^\top \sim N_{d+1}(\mathbf{0}, \Omega)$. Generating sequentially from W, $\tilde{L}$ and then from $\boldsymbol{Z}_{\mathrm{AC}} = (\boldsymbol{X}|\tilde{L}, W) \sim N_d\left(W^{-1/2}\boldsymbol{\delta}\tilde{L}, W^{-1}(\bar{\Omega} - \boldsymbol{\delta}\boldsymbol{\delta}^\top)\right)$ as in Section 2.2.3, gives a generative representation.

The extended likelihood in Section 3.1 is based on generative representation GR2.

Appendix C: Gradient

Table C1. Computationally efficient closed form expression for the gradient $\nabla_\theta \log p(\boldsymbol{\theta}, \tilde{\boldsymbol{l}}, \boldsymbol{w}|\boldsymbol{y})$.

Computing $\nabla_G \log p(\boldsymbol{\theta}, \tilde{\boldsymbol{l}}, \boldsymbol{w}|\boldsymbol{y})$

$$\nabla_G \log p(\boldsymbol{\theta}, \tilde{\boldsymbol{l}}, \boldsymbol{w}|\boldsymbol{y}) = \nabla_G\Big\{-\frac{n}{2}T_{G1} - \frac{1}{2}\sum_{i=1}^n (T_{G2i} + T_{G3i} + T_{G4i})\Big\} + \nabla_G \log p(G)$$

$$\nabla_G T_{G1} = -2(V_2 \odot VV_1 C_4 + V_1 C_4 V_1)G$$

$$\nabla_G T_{G2i} = -2(V_2 \odot VV_1 C_{7i} + V_1 C_{7i} V_1)G$$

$$\nabla_G T_{G3i} = -2(V_2 \odot VV_1 C_{11i} + V_1 C_{11i} V_1)G$$

$$\nabla_G T_{G4i} = -2(V_2 \odot VV_1 C_{15i} + V_1 C_{15i} V_1)G$$

$$\nabla_G \log p(G) = -(a_1+1)\,\text{sign}(G)\frac{1}{b_1 + |G|}$$

$$C_0 = \bar{\Omega}^{-1} + \frac{\bar{\Omega}^{-1}\boldsymbol{\delta}\boldsymbol{\delta}^\top\bar{\Omega}^{-1}}{1 - \boldsymbol{\delta}^\top\bar{\Omega}^{-1}\boldsymbol{\delta}}$$

$$C_1 = 1 + \boldsymbol{\alpha}^\top\bar{\Omega}\boldsymbol{\alpha}$$

$$C_2 = \boldsymbol{\alpha}\boldsymbol{\alpha}^\top\bar{\Omega}$$

$$C_3 = C_1^{-1}C_2C_0 + C_1^{-1}C_0C_2^\top - C_1^{-2}C_2C_0C_2^\top$$

$$C_4 = C_0 - C_3$$

$$C_{5i} = -w_i C_0 \boldsymbol{y}_i \boldsymbol{y}_i^\top C_0$$

$$C_{6i} = C_1^{-1}C_2C_{5i} + C_1^{-1}C_{5i}C_2^\top - C_1^{-2}C_2C_{5i}C_2^\top$$

$$C_{7i} = C_{5i} - C_{6i}$$

$$C_{8i} = -2\tilde{l}_i w_i^{1/2} C_1^{-1/2} C_0 \boldsymbol{y}_i \boldsymbol{\alpha}^\top + \tilde{l}_i w_i^{1/2} C_1^{-3/2} C_2 C_0 \boldsymbol{y}_i \boldsymbol{\alpha}^\top$$

$$C_{9i} = 2\tilde{l}_i w_i^{1/2} C_0 \boldsymbol{y}_i \boldsymbol{\delta}^\top C_0$$

$$C_{10i} = C_1^{-1}C_2C_{9i} + C_1^{-1}C_{9i}C_2^\top - C_1^{-2}C_2C_{9i}C_2^\top$$

$$C_{11i} = C_{8i} + C_{9i} - C_{10i}$$

$$C_{12i} = 2\tilde{l}_i^2 C_1^{(-1/2)} \boldsymbol{\alpha}\boldsymbol{\delta}^\top C_0 - \tilde{l}_i^2 C_1^{-3/2} \boldsymbol{\alpha}\boldsymbol{\delta}^\top C_0 C_2^\top$$

$$C_{13i} = -\tilde{l}_i^2 C_0 \boldsymbol{\delta}\boldsymbol{\delta}^\top C_0$$

$$C_{14i} = C_1^{-1}C_2C_{13i} + C_1^{-1}C_{13i}^\top C_2^\top - C_1^{-2}C_2C_{13i}C_2^\top$$

$$C_{15i} = C_{12i} + C_{13i} - C_{14i}$$

$$V_1 = \text{diag}(V)^{-1/2} \quad \& \quad V_2 = \text{diag}(V)^{-3/2}$$

Computing $\nabla_{\widetilde{G}} \log p(\boldsymbol{\theta}, \tilde{\boldsymbol{l}}, \boldsymbol{w}|\boldsymbol{y})$

$$\nabla_{\widetilde{G}} \log p(\boldsymbol{\theta}, \tilde{\boldsymbol{l}}, \boldsymbol{w}|\boldsymbol{y}) = \nabla_G \log p(\boldsymbol{\theta}, \tilde{\boldsymbol{l}}, \boldsymbol{w}|\boldsymbol{y}) \times \exp(\widetilde{G}) + 1$$

Computing $\nabla_{\boldsymbol{\alpha}} \log p(\boldsymbol{\theta}, \tilde{\boldsymbol{l}}, \boldsymbol{w}|\boldsymbol{y})$

$$\nabla_{\boldsymbol{\alpha}} \log p(\boldsymbol{\theta}, \tilde{\boldsymbol{l}}, \boldsymbol{w}|\boldsymbol{y}) = \nabla_{\boldsymbol{\alpha}}\Big\{-\frac{n}{2}T_{\boldsymbol{\alpha}1} - \frac{1}{2}\sum_{i=1}^n (T_{\boldsymbol{\alpha}2i} + T_{\boldsymbol{\alpha}3i} + T_{\boldsymbol{\alpha}4i})\Big\} + \nabla_{\boldsymbol{\alpha}} \log p(\boldsymbol{\alpha})$$

$$\nabla_{\boldsymbol{\alpha}} T_{\boldsymbol{\alpha}1} = 2C_1^{-3/2}\bar{\Omega}\boldsymbol{\alpha}\boldsymbol{\alpha}^\top\bar{\Omega}C_0\boldsymbol{\delta} - 2C_1^{-1/2}\bar{\Omega}C_0\boldsymbol{\delta}$$

$$\nabla_{\boldsymbol{\alpha}} T_{\boldsymbol{\alpha}2i} = C_1^{-2}\bar{\Omega}\boldsymbol{\alpha}\boldsymbol{\alpha}^\top\bar{\Omega}(C_{5i} + C_{5i}^\top)\bar{\Omega}\boldsymbol{\alpha} - C_1^{-1}\bar{\Omega}(C_{5i} + C_{5i}^\top)\bar{\Omega}\boldsymbol{\alpha}$$

$$\nabla_{\boldsymbol{\alpha}} T_{\boldsymbol{\alpha}3i} = C_{16i} + C_1^{-1}\bar{\Omega}(C_{17i} + C_{17i}^\top)\bar{\Omega}\boldsymbol{\alpha} - C_1^{-2}\bar{\Omega}\boldsymbol{\alpha}\boldsymbol{\alpha}^\top\bar{\Omega}(C_{17i} + C_{17i}^\top)\bar{\Omega}\boldsymbol{\alpha}$$

$$\nabla_{\boldsymbol{\alpha}} T_{\boldsymbol{\alpha}4i} = 2C_{18i} + C_1^{-1}\bar{\Omega}(C_{19i} + C_{19i}^\top)\bar{\Omega}\boldsymbol{\alpha} - C_1^{-2}\bar{\Omega}\boldsymbol{\alpha}\boldsymbol{\alpha}^\top\bar{\Omega}(C_{19i} + C_{19i}^\top)\bar{\Omega}\boldsymbol{\alpha}$$

$$\nabla_{\boldsymbol{\alpha}} \log p(\boldsymbol{\alpha}) = -\frac{1}{\sigma^2}\sum_i \alpha_i$$

$$C_{16i} = -2\tilde{l}_i w_i^{1/2} C_1^{-1/2} \bar{\Omega} C_0 \boldsymbol{y}_i + 2\tilde{l}_i w_i^{1/2} C_1^{-3/2} \boldsymbol{\alpha}^\top \bar{\Omega} C_0 \boldsymbol{y}_i \bar{\Omega}\boldsymbol{\alpha}$$

$$C_{17i} = -2\tilde{l}_i w_i^{1/2} C_0 \boldsymbol{y}_i \boldsymbol{\delta}^\top C_0$$

$$C_{18i} = \tilde{l}_i^2 C_1^{-1/2} \bar{\Omega} C_0 \boldsymbol{\delta} - \tilde{l}_i^2 C_1^{-3/2} \bar{\Omega}\boldsymbol{\alpha}\boldsymbol{\alpha}^\top \bar{\Omega} C_0 \boldsymbol{\delta}$$

$$C_{19i} = \tilde{l}_i^2 C_0 \boldsymbol{\delta}\boldsymbol{\delta}^\top C_0$$

Computing $\nabla_{\tilde{\nu}} \log p(\boldsymbol{\theta}, \tilde{\boldsymbol{l}}, \boldsymbol{w}|\boldsymbol{y})$

$$\nabla_{\tilde{\nu}} \log p(\boldsymbol{\theta}, \tilde{\boldsymbol{l}}, \boldsymbol{w}|\boldsymbol{y}) = \nabla_\nu\Big\{\mathcal{L}(\boldsymbol{y}, \tilde{\boldsymbol{l}}, \boldsymbol{w}|\boldsymbol{\theta}) + \log(p(\nu))\Big\} \times \exp(\tilde{\nu}) + 1$$

$$\nabla_\nu \mathcal{L}(\boldsymbol{y}, \tilde{\boldsymbol{l}}, \boldsymbol{w}|\boldsymbol{\theta}) = \frac{1}{2}\sum_{i=1}^n \log(w_i) + n\Big(\frac{1}{2}\log(\frac{\nu}{2}) + \frac{1}{2} - \frac{1}{2}\psi_0(\frac{\nu}{2})\Big) - \frac{1}{2}\sum_{i=1}^n w_i$$

$$\nabla_\nu \log(p(\nu)) = (a_2 - 1)\frac{1}{\nu} - b_2$$

NOTE: "$\odot$" denotes the Hadamard (i.e., element-wise) product, and the four gradients are computed by evaluating the terms sequentially from the bottom upwards. Their derivation is found in Part B of the Web Appendix. MATLAB routines to evaluate the gradients are provided.

Appendix D: MCMC Scheme

Algorithm D1 is an MCMC sampling scheme to evaluate the augmented posterior of the AC skew-t factor copula parameters $\boldsymbol{\theta}$. Steps 1 and 2 are also used in the hybrid VI Algorithm 1.

Algorithm D1 MCMC Scheme for AC Skew-t Copula

Step 0. Initialize feasible values for $\boldsymbol{\theta}, \tilde{\boldsymbol{l}}$ and $\boldsymbol{w}$

Step 1. Generate from $p(\tilde{\boldsymbol{l}}|\boldsymbol{\theta},\boldsymbol{w},\boldsymbol{y}) = \prod_{i=1}^n p(\tilde{l}_i|\boldsymbol{\theta}, w_i, \boldsymbol{y}_i)$

Step 2. Generate from $p(\boldsymbol{w}|\boldsymbol{\theta},\tilde{\boldsymbol{l}},\boldsymbol{y}) = \prod_{i=1}^n p(w_i|\boldsymbol{\theta}, \tilde{l}_i, \boldsymbol{y}_i)$

Step 3. Generate from $p(\alpha_j|\{\boldsymbol{\theta}\backslash\alpha_j\},\tilde{\boldsymbol{l}},\boldsymbol{w},\boldsymbol{y})$ for $j = 1,\ldots,d$.

Step 4. Generate from $p(\tilde{g}_{ij}|\{\boldsymbol{\theta}\backslash\tilde{g}_{ij}\},\tilde{\boldsymbol{l}},\boldsymbol{w},\boldsymbol{y})$ for nonzero elements $\tilde{g}_{ij}$ of $\widetilde{G}$.

Step 5. Generate from $p(\nu|\{\boldsymbol{\theta}\backslash\nu\},\tilde{\boldsymbol{l}},\boldsymbol{w},\boldsymbol{y})$

The conditional posteriors at Steps 3–5 can be derived from (10), but are unrecognizable, so we employ adaptive random walk Metropolis-Hastings schemes to generate each element.

To derive the posteriors at Steps 1 and 2, note that from (8) and (9),

$$p(\tilde{\boldsymbol{l}},\boldsymbol{w}|\boldsymbol{\theta},\boldsymbol{y}) \propto p(\tilde{\boldsymbol{l}},\boldsymbol{w},\boldsymbol{y}|\boldsymbol{\theta}) \propto \prod_{i=1}^n p(\boldsymbol{z}_i|\tilde{l}_i, w_i,\boldsymbol{\theta})p(\tilde{l}_i|w_i)p(w_i|\nu)$$
$$= \prod_{i=1}^n \phi_d(\boldsymbol{z}_i;\boldsymbol{\mu}_{z,i},\Sigma_{z,i})\phi_1(\tilde{l}_i;0,1)\mathbb{1}(\tilde{l}_i > 0)p(w_i|\nu),$$

with $\boldsymbol{\mu}_{z,i} = \boldsymbol{\delta}\tilde{l}_i w_i^{-1/2}$, $\Sigma_{z,i} = w_i^{-1}(\bar{\Omega} - \boldsymbol{\delta}\boldsymbol{\delta}^\top)$ and $p(w_i|\nu)$ is a Gamma$(\nu/2,\nu/2)$ density.

From the above, at Step 1 the conditional posterior $p(\tilde{\boldsymbol{l}}|\boldsymbol{w},\boldsymbol{\theta},\boldsymbol{y}) = \prod_{i=1}^n p(\tilde{l}_i|w_i,\boldsymbol{\theta},\boldsymbol{y})$, where $\tilde{l}_i|\boldsymbol{\theta},w_i,\boldsymbol{y}_i \sim N_+(A^{-1}B_i, A^{-1})$, with $A = 1+\boldsymbol{\delta}^\top(\bar{\Omega}-\boldsymbol{\delta}\boldsymbol{\delta}^\top)^{-1}\boldsymbol{\delta}$, $B_i = w_i^{1/2}\boldsymbol{\delta}^\top(\bar{\Omega}-\boldsymbol{\delta}\boldsymbol{\delta}^\top)^{-1}\boldsymbol{z}_i$ and N_+ denotes a univariate normal distribution constrained to positive values. Similarly, at Step 2 the conditional posterior $p(\boldsymbol{w}|\tilde{\boldsymbol{l}},\boldsymbol{\theta},\boldsymbol{y}) = \prod_{i=1}^n p(w_i|\tilde{l}_i,\boldsymbol{\theta},\boldsymbol{y})$, with

$$p(w_i|\boldsymbol{\theta},\tilde{l}_i,\boldsymbol{y}) \propto w_i^{\frac{d+\nu}{2}-1}\exp\left\{-\frac{1}{2}\left(w_i\boldsymbol{z}_i^\top(\bar{\Omega}-\boldsymbol{\delta}\boldsymbol{\delta}^\top)^{-1}\boldsymbol{z}_i - 2\tilde{l}_i w_i^{1/2}\boldsymbol{\delta}^\top(\bar{\Omega}-\boldsymbol{\delta}\boldsymbol{\delta}^\top)^{-1}\boldsymbol{z}_i + w_i\nu\right)\right\}.$$

An adaptive random walk Metropolis-Hastings step is used to draw from this posterior.

In Step (b) of Algorithm 1, a draw of $\tilde{\boldsymbol{l}},\boldsymbol{w}$ is obtained by repeatedly drawing from the conditionals 25 times. We stress this is fast because all demanding computations do not involve $\boldsymbol{l}$ and $\boldsymbol{w}$, so that they only need to be computed once per SGA step. We found 25 draws to be adequate, which is consistent with the findings in Loaiza-Maya et al. (2022) for other models.

Supplementary Materials

The Supplementary Materials consist of a Web Appendix and MATLAB code. The Web Appendix provides further technical details on the copula and the VI algorithm, as well as additional empirical results for Sections 5 and 6 of the paper. The MATLAB code estimates the skew-t copula and can be used to replicate the results of the study, or be employed using other datasets. The latest version of the code can also be found at *https://github.com/lindenglab/ACCopVB-MATLAB*.

Acknowledgments

We thank an Associate Editor and two referees for comments that have helped improve the article, and the Coeditor Ivan A. Canay for his guidance.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

Funding

This research was supported by The University of Melbourne's Research Computing Services and the Petascale Campus Initiative. A/Prof. Maneesoonthorn has been supported by the Australian Research Council (ARC) Discovery Project Grant (DP200101414).

References

Ando, T., Greenwood-Nimmo, M., and Shin, Y. (2022), "Quantile Connectedness: Modeling Tail Behavior in the Topology of Financial Networks," *Management Science*, 68, 2401–2431. [2,4]

Azzalini, A., and Capitanio, A. (2003), "Distributions Generated by Perturbation of Symmetry with Emphasis on a Multivariate Skew t - Distribution," *Journal of the Royal Statistical Society*, Series B, 65, 367–389. [1,3]

Barndorff-Nielsen, O. (1977), "Exponentially Decreasing Distributions for the Logarithm of Particle Size," *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, 353, 401–419. [3]

Blei, D. M., Kucukelbir, A., and McAuliffe, J. D. (2017), "Variational Inference: A Review for Statisticians," *Journal of the American Statistical Association*, 112, 859–877. [2]

Bollerslev, T. (1990), "Modelling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized ARCH Model," *The Review of Economics and Statistics*, 72, 498. [10]

Bollerslev, T., Patton, A. J., and Quaedvlieg, R. (2022), "Realized Semibetas: Disentangling "good" and "bad" Downside Risks," *Journal of Financial Economics*, 144, 227–246. [2,12]

Britten-Jones, M., and Neuberger, A. (2000), "Option Prices, Implied Price Processes, and Stochastic Volatility," *The Journal of Finance*, 55, 839–866. [8]

Chan, J. C., and Yu, X. (2022), "Fast and Accurate Variational Inference for Large Bayesian VARs with Stochastic Volatility," *Journal of Economic Dynamics and Control*, 143, 104505. [14]

Christoffersen, P., Errunza, V., Jacobs, K., and Langlois, H. (2012), "Is the Potential for International Diversification Disappearing? A Dynamic Copula Approach," *Review of Financial Studies*, 25, 3711–3751. [1,3]

Creal, D. D., and Tsay, R. S. (2015), "High Dimensional Dynamic Stochastic Copula Models," *Journal of Econometrics*, 189, 335–345. [1,2]

Demarta, S., and Mcneil, A. J. (2005), "The t Copula and Related Copulas," *International Statistical Review*, 73, 111–129. [1,3]

Engle, R. F., and Sokalska, M. E. (2012), "Forecasting Intraday Volatility in the US Equity Market. Multiplicative Component GARCH," *Journal of Financial Econometrics*, 10, 54–83. [8]

Frazier, D. T., Loaiza-Maya, R., and Martin, G. M. (2023), "Variational Bayes in State Space Models: Inferential and Predictive Accuracy," *Journal of Computational and Graphical Statistics*, 32, 793–804. [6]

Gefang, D., Koop, G., and Poon, A. (2023), "Forecasting Using Variational Bayesian Inference in Large Vector Autoregressions with Hierarchical Shrinkage," *International Journal of Forecasting*, 39, 346–363. [14]

Geweke, J. (1991), "Evaluating the Accuracy of Sampling-based Approaches to the Calculation of Posterior Moments," Staff Report 148, Federal Reserve Bank of Minneapolis. [7]

Giacometti, R., Torri, G., and Paterlini, S. (2021), "Tail Risks in Large Portfolio Selection: Penalized Quantile and Expectile Minimum Deviation Models," *Quantitative Finance*, 21, 243–261. [2,12]

Gneiting, T., Balabdaoui, F., and Raftery, A. E. (2007), "Probabilistic Forecasts, Calibration and Sharpness," *Journal of the Royal Statistical Society*, Series B, 69, 243–268. [9]

Guidolin, M., and Timmermann, A. (2008), "International Asset Allocation Under Regime Switching, Skew, and Kurtosis Preferences," *The Review of Financial Studies*, 21, 889–935. [12]

Harvey, C. R., Liechty, J. C., Liechty, M. W., and Müller, P. (2010), "Portfolio Selection with Higher Moments," *Quantitative Finance*, 10, 469–485. [2,12]

Hoffman, M. D., and Gelman, A. (2014), "The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo," *Journal of Machine Learning Research*, 15, 1593–1623. [6]

Joe, H. (2014), *Dependence Modeling with Copulas*, New York: Chapman and Hall/CRC. [2]

Kingma, D. P., and Welling, M. (2014), "Auto-Encoding Variational Bayes," arXiv:1312.6114. [2,5]

Kollo, T., and Pettere, G. (2010), "Parameter Estimation and Application of the Multivariate Skew t-copula," in *Copula Theory and its Applications*, pp. 289–298, Berlin: Springer. [1,3]

Krupskii, P., and Joe, H. (2013), "Factor Copula Models for Multivariate Data," *Journal of Multivariate Analysis*, 120, 85–101. [1]

Loaiza-Maya, R., and Nibbering, D. (2023), "Fast Variational Bayes Methods for Multinomial Probit Models," *Journal of Business & Economic Statistics*, 41, 1352–1363. [14]

Loaiza-Maya, R., and Smith, M. S. (2019), "Variational Bayes Estimation of Discrete-Margined Copula Models with Application to Time Series," *Journal of Computational and Graphical Statistics*, 28, 523–539. [2]

Loaiza-Maya, R., Smith, M. S., Nott, D. J., and Danaher, P. J. (2022), "Fast and Accurate Variational Inference for Models with Many Latent Variables," *Journal of Econometrics*, 230, 339–362. [2,5,6,7,16]

Lucas, A., Schwaab, B., and Zhang, X. (2017), "Modeling Financial Sector Joint Tail Risk in the Euro Area," *Journal of Applied Econometrics*, 32, 171–191. [1,2]

Miller, A. C., Foti, N. J., and Adams, R. P. (2017), "Variational Boosting: Iteratively Refining Posterior Approximations," in *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of Proceedings of Machine Learning Research, eds. D. Precup, Y. W. Teh, pp. 2420–2429, PMLR. [6]

Murray, J. S., Dunson, D. B., Carin, L., and Lucas, J. E. (2013), "Bayesian Gaussian Copula Factor Models for Mixed Data," *Journal of the American Statistical Association*, 108, 656–665. [4,5]

Nelsen, R. B. (1999), *An Introduction to Copulas*, New York: Springer. [2]

Nguyen, H., Ausín, M. C., and Galeano, P. (2020), "Variational Inference for High Dimensional Structured Factor Copulas," *Computational Statistics & Data Analysis*, 151, 107012. [2]

Oh, D. H., and Patton, A. J. (2013), "Simulated Method of Moments Estimation for Copula-based Multivariate Models," *Journal of the American Statistical Association*, 108, 689–700. [2,8]

Oh, D. H., and Patton, A. J. (2017), "Modeling Dependence in High Dimensions with Factor Copulas," *Journal of Business & Economic Statistics*, 35, 139–154. [1,2]

Oh, D. H., and Patton, A. J. (2023), "Dynamic Factor Copula Models with Estimated Cluster Assignments," *Journal of Econometrics*, 237, 105374. [1,2,3,9,10]

Ong, V. M.-H., Nott, D. J., and Smith, M. S. (2018), "Gaussian Variational Approximation with a Factor Covariance Structure," *Journal of Computational and Graphical Statistics*, 27, 465–478. [6,10]

Opschoor, A., Lucas, A., Barra, I., and Van Dijk, D. (2021), "Closed-Form Multi-Factor Copula Models with Observation-Driven Dynamic Factor Loadings," *Journal of Business & Economic Statistics*, 39, 1066–1079. [2]

Patton, A. J. (2004), "On the Out-of-Sample Importance of Skewness and Asymmetric Dependence for Asset Allocation," *Journal of Financial Econometrics*, 2, 130–168. [2]

Patton, A. J. (2006), "Modelling Asymmetric Exchange Rate Dependence," *International Economic Review*, 47, 527–556. [3,8]

Ranganath, R., Gerrish, S., and Blei, D. (2014), "Black Box Variational Inference," in *Artificial Intelligence and Statistics*, PMLR, pp. 814–822. [2,5,14]

Sahu, S. K., Dey, D. K., and Branco, M. D. (2003), "A New Class of Multivariate Skew Distributions with Applications to Bayesian Regression Models," *Canadian Journal of Statistics*, 31, 129–150. [1,2,3]

Smith, M. S. (2023), "Implicit Copulas: An Overview," *Econometrics and Statistics*, 28, 81–104. [3]

Smith, M. S., Gan, Q., and Kohn, R. J. (2010), "Modelling Dependence Using Skew t Copulas: Bayesian Inference and Applications," *Journal of Applied Econometrics*, 27, 500–522. [1,2,3,9]

Smith, M. S., and Maneesoonthorn, W. (2018), "Inversion Copulas from Nonlinear State Space Models with an Application to Inflation Forecasting," *International Journal of Forecasting*, 34, 389–407. [5]

Yoshiba, T. (2018), "Maximum Likelihood Estimation of Skew-t Copulas with its Applications to Stock Returns," *Journal of Statistical Computation and Simulation*, 88, 2489–2506. [1,3,5]

Zeiler, M. D. (2012), "ADADELTA: An Adaptive Learning Rate Method," arXiv:1212.5701 [cs]. [5]

Zhao, Z. (2021), "Dynamic Bivariate Peak over Threshold Model for Joint Tail Risk Dynamics of Financial Markets," *Journal of Business & Economic Statistics*, 39, 892–906. [12]