Sparse Copula-based Black-Litterman Portfolio Optimization with Normalizing Flows

Stanislas Muhinyuza* Maziar Sahamkhadam† Andreas Stephan‡

Abstract

This paper extends the Black-Litterman (BL) portfolio optimization to a vine copula-based normalizing flow that minimizes the forward Kullback-Leibler (KL) divergence between the transformed prior and the view distribution. We refer to the suggested model as the copula-based BL with normalizing flows (CBL-NF). The vine-structured flow is a composition of rational-quadratic spline transforms aligned with the vine tree sequence, which imposes sparsity on the conditional dependence structure. Applying the CBL-NF approach to a sample of S&P 100 historical constituents, our out-of-sample analysis indicates that the CBL-NF portfolios achieve a lower CVaR than the classical BL portfolios for most of the portfolio strategies, at a similar level of average return and turnover. Furthermore, the suggested approach reaches a higher terminal wealth than the classical BL and the equally weighted portfolios for the reward-based strategies. We also show that our results are robust to different scaling parameters, data frequencies, and the truncation level in R-vine copula modeling.

Keywords: Black-Litterman model, vine copulas, normalizing flows, Bayesian portfolio optimization, GRU neural network, conditional value at risk

JEL Classification: C11, C32, C45, C58, G11

*Department of Economics and Statistics, School of Business and Economics, Linnaeus University, Sweden.

†**Corresponding author**: Department of Economics and Statistics, School of Business and Economics, Linnaeus University, Sweden. Corresponding author. E-mail: maziar.sahamkhadam@lnu.se

‡Faculty of Technology, Linnaeus University, Sweden.

1 Introduction

The Black-Litterman (BL) model of Black and Litterman (1991, 1992), further developed by He and Litterman (1999), is a widely used approach for incorporating investor views into portfolio construction. The model combines the equilibrium returns implied by the Capital Asset Pricing Model (CAPM) with the investor's views through a Bayesian update, and addresses several shortcomings of the mean-variance optimization of Markowitz (1952), namely the sensitivity to the input parameters and the resulting extreme and unstable portfolio weights (Chopra and Ziemba, 1993; DeMiguel et al., 2009; Kolm et al., 2014). In its original form, the BL model assumes a Gaussian prior on the expected returns and a Gaussian view likelihood, and the conjugacy between the two yields a closed-form posterior that adjusts the first two moments of the return distribution. Several studies relax the Gaussian assumption by introducing non-Gaussian priors or views, for instance through stable distributions, GARCH dynamics, or copula-based forecasts, but the update in these extensions remains linear and conjugate. A notable exception is the entropy-pooling approach of Meucci (2006, 2008), and its copula-based extension, which reweight a fixed set of prior scenarios to match the views. Since the reweighting keeps the same set of scenarios, it cannot strengthen the tail dependence beyond what the prior already contains.

In this article, we generalize the BL update from an adjustment of the first two moments to a transformation of the entire joint return distribution. We construct a non-Gaussian joint prior on returns and update it given the investor views using a normalizing flow, which is an invertible map that transforms one density into another through the change-of-variables formula (Papamakarios et al., 2021). Unlike the linear conjugate update, the flow can modify the higher-order features of the prior, namely the tail dependence and the asymmetry, in addition to the first two moments. Hickling and Prangle (2025) develop a flow with adjustable tail behavior to model the heavy tails of financial returns, and Yang and Hospedales (2023) use a flow to estimate the risk-neutral density for option pricing. We refer to the suggested approach as the copula-based Black-Litterman normalizing flow (CBL-NF).

Several other studies extend the BL model to non-Gaussian distributions. Beyond the entropy-pooling approaches discussed above, Giacometti et al. (2007) replace the Gaussian prior with stable distributions, and Harris et al. (2017) and Deng (2018) incorporate multivariate GARCH dynamics. Kolm and Ritter (2017) and Palczewski and Palczewski (2019) provide Bayesian and continuous-distribution interpretations of the BL formula. More recently, Teplova et al. (2023) use ARMA-GARCH-copula expected returns as investor views, and Barua and Sharma (2022) apply neural networks to generate dynamic BL views.

We construct the joint prior on returns following the copula-based forecasting litera-

ture (Patton, 2006; Chen and Fan, 2006; Joe, 2014). In particular, we fit ARMA-GARCH marginals with skewed Student-t innovations (Bollerslev, 1986) to each asset, transform the standardized residuals to uniform marginals, and fit a truncated R-vine copula (Aas et al., 2009; Bedford and Cooke, 2002; Dißmann et al., 2013) on the uniform marginals. The truncation sets the conditional pair-copula densities at tree depths beyond K to the independence copula (Brechmann et al., 2012).

To incorporate the investor views, we replace the linear conjugate update with a normalizing flow aligned with the vine tree sequence, namely a composition of invertible rational-quadratic spline transforms (Durkan et al., 2019). The flow has the same sparsity as the truncated vine, since we introduce a multilayer perceptron for each pair copula in the trees $\mathcal{T}_1, \ldots, \mathcal{T}_K$. We assume a Gaussian prior with an identity initialization on the flow parameters, and obtain the maximum a posteriori estimate by minimizing the forward Kullback-Leibler divergence between the transformed prior and the view distribution. This formulation follows the generalized Bayesian inference framework of Bissiri et al. (2016) and Knoblauch et al. (2022). To construct the views from data, we use a Gated Recurrent Unit (GRU) network for each asset, as in the recent literature on machine learning in asset pricing (Cho et al., 2014; Gu et al., 2020). We notice, however, that the suggested CBL-NF framework does not depend on the source of the views and applies to any views given by a forecast mean and an uncertainty estimate.

The contribution of this paper is twofold. First, we extend the BL update from the linear conjugate Gaussian framework to a generalized Bayesian framework (Bissiri et al., 2016; Knoblauch et al., 2022), in which the posterior is obtained by minimizing the forward Kullback-Leibler divergence between the transformed prior and the view distribution. The resulting posterior modifies the dependence structure of the prior, including the lower and upper tail dependence, which is not possible under the moment-based BL extensions of Harris et al. (2017), Deng (2018), and Teplova et al. (2023). The CBL-NF approach reduces to the prior when the views are uninformative, and to the classical BL formula under Gaussian restrictions. Second, we suggest a vine-structured normalizing flow for the BL update. The closest construction is the vine copula flow of Mitskopoulos et al. (2022), who use neural spline flows to estimate vine copula densities for neural spike data. Our vine-structured normalizing flow differs from theirs in several respects. To start with, they fit each pair copula separately and tree by tree, whereas we compose the pair transforms into a single map and estimate all parameters jointly. Furthermore, they impose the simplifying assumption of Haff et al. (2010), under which the conditional copulas do not depend on the conditioning variables, whereas we condition each transform on its conditioning variables through a multilayer perceptron, which allows the dependence to vary across the conditioning set. Finally, we truncate the vine at tree K, so that the parameter count grows linearly in the number of assets. To the best of our knowledge, this is the first study to use a vine-structured normalizing flow as the

Bayesian view update in the BL framework.

We complement the empirical analysis with a simulation study based on a known C-vine data-generating process with $d = 5$ assets. The simulation results show that the CBL-NF approach preserves the asymmetric tail dependence of the prior at small view weights and approaches the Gaussian view distribution at large view weights. Using a sample of historical constituents of the S&P 100 index from 4 January 2016 to 26 January 2026 at the daily frequency, we apply the CBL-NF approach to examine both its in-sample behavior and out-of-sample portfolio performance. The in-sample analysis on the S&P 100 returns confirms the findings from the simulation study. We further show the advantages of the suggested CBL-NF approach when applying to out-of-sample portfolio optimization and backtesting. We show that the CBL-NF portfolios generally achieve a lower tail risk than the classical BL portfolios at a similar level of average return and turnover, as well as robustness of these results to different values of λ, data frequencies, and R-vine truncation levels.

The remainder of this article is structured as follows. Section 2 presents the methodology, including the prior construction via econometric marginals and vine copulas, the BL approach and its generalization, the vine-structured normalizing flow, the investor views, and the portfolio optimization methods. Section 3 provides the steps for constructing the CBL-NF posterior and portfolios. Section 4 presents the simulation study. Section 5 describes the data and provides the in-sample empirical analysis. Section 7 gives the concluding remarks.

2 Methodology

The CBL-NF posterior is constructed by combining (i) a non-Gaussian prior from econometric marginals and a vine copula and (ii) a normalizing flow aligned with the vine tree sequence, which updates the prior using the investor views. We first describe the prior construction, then present the BL framework and its generalization. The view generation is presented next, followed by the Bayesian inference for the flow and the portfolio optimization methods.

2.1 Truncated R-vine copula

Let $r_{i,t}$ denote the log-return of asset i at time t, for $i = 1, \ldots, d$ and $t = 1, \ldots, T$. As equity returns show volatility clustering, fat tails, and distributional asymmetry (Cont, 2001), we follow Bollerslev (1986) and Engle (1982) and assume that, for each asset, the

returns follow a univariate ARMA(p_i, q_i)-GARCH$(1,1)$ process:

$$\begin{aligned} r_{i,t} &= \mu_{i,t} + \epsilon_{i,t}, \\ \epsilon_{i,t} &= \sigma_{i,t}\, z_{i,t}, \qquad z_{i,t} \sim \text{iid}, \\ \sigma^2_{i,t} &= \omega_i + \alpha_i\, \epsilon^2_{i,t-1} + \beta_i\, \sigma^2_{i,t-1}, \\ \omega_i &> 0, \quad \alpha_i \geq 0, \quad \beta_i \geq 0, \quad \alpha_i + \beta_i < 1, \end{aligned} \tag{1}$$

where $\mu_{i,t}$ is the conditional mean from the ARMA component, $\epsilon_{i,t}$ is the residual, $\sigma^2_{i,t}$ is the conditional variance, and $z_{i,t}$ is the standardized residual. To accommodate the heavy tails and asymmetry of equity returns, $z_{i,t}$ is assumed to follow a skewed Student-t distribution.[1]

Given the d-dimensional uniform marginals $\boldsymbol{u}_t = (u_{1,t}, \ldots, u_{d,t})$ for $t = 1, \ldots, T$, we model the contemporaneous dependence using a regular vine (R-vine) copula. Sklar (1959) shows how to obtain the d-dimensional copula C which, for the distribution function F with marginal distributions $F_1, \ldots, F_d$, satisfies

$$F(z_1, \ldots, z_d) = C\big(F_1(z_1), \ldots, F_d(z_d)\big), \tag{2}$$

where $z_j = F_j^{-1}(u_j)$ and $u_j \sim U[0,1]$. Drawing upon the idea of Joe (1996) to decompose the joint density into pair-copula densities, Bedford and Cooke (2001, 2002) derive a graphical representation of the pair-copula construction (PCC) in the form of nested trees. Aas et al. (2009) show how to derive different vine structures from R-vine representations and develop the inference algorithms for these models. Further properties and statistical theory are developed in Joe (2014) and Czado (2019).

Following Bedford and Cooke (2002), a sequence of trees $\mathcal{V} = (\mathcal{T}_1, \ldots, \mathcal{T}_{d-1})$ on d nodes is called a regular vine if (i) $\mathcal{T}_1$ is a tree with node set $\mathcal{N}_1 = \{1, \ldots, d\}$ and edge set $\mathcal{E}_1$, (ii) for $k = 2, \ldots, d-1$, $\mathcal{T}_k$ is a tree with node set $\mathcal{N}_k = \mathcal{E}_{k-1}$ and edge set $\mathcal{E}_k$, and (iii) proximity condition: for $k \geq 2$, two edges in $\mathcal{T}_{k-1}$ are connected by an edge in $\mathcal{T}_k$ only if they share a common node in $\mathcal{T}_{k-1}$. Each edge $e \in \mathcal{E}_k$ of the regular vine corresponds to a bivariate copula density $c_{j(e),k(e);D(e)}$ for the conditional pair $(j(e), k(e))$ given the conditioning set $D(e)$. The vine copula density is the product of these pair-copula densities across all edges.

For high-dimensional applications such as $d = 100$ assets, the full vine includes 4,950 pair copulas, which is computationally extensive. Following Brechmann et al. (2012) and Brechmann and Joe (2015), we truncate the vine at level K by setting all pair copulas in trees $\mathcal{T}_{K+1}, \ldots, \mathcal{T}_{d-1}$ to the independence copula. The truncated R-vine density is given

[1] The ARMA orders (p_i, q_i) are selected by BIC, and the parameters are estimated by quasi-maximum likelihood (QML) for each asset separately. The standardized residuals $z_{i,t}$ are then mapped to uniform marginals through their marginal distribution function F_i, and the inverse F_i^{-1} is used for simulation from the fitted vine copula.

by

$$c_{\text{trunc}}(\boldsymbol{u};\boldsymbol{\eta}) = \prod_{k=1}^{K} \prod_{e\in\mathcal{E}_k} c_{j(e),k(e);D(e)}\big(u_{j(e)|D(e)},\, u_{k(e)|D(e)};\, \boldsymbol{\eta}_e\big), \tag{3}$$

where $\boldsymbol{\eta}_e$ is the parameter of the pair-copula family chosen for edge e, and $\boldsymbol{\eta} = (\boldsymbol{\eta}_e)_{e\in\mathcal{E}_1\cup\cdots\cup\mathcal{E}_K}$ collects the pair-copula parameters across all edges in the truncated vine. The truncation imposes a sparsity assumption on the conditional dependence structure, namely the conditional pair-copula densities at tree depths beyond K are set to the independence copula, in line with the broader literature on sparse covariance and dependence estimation in high dimensions (Bickel and Levina, 2008; Friedman et al., 2008; Nagler et al., 2019).

We follow Dißmann et al. (2013) for the structure selection. In tree $\mathcal{T}_1$, we construct a maximum spanning tree on the complete graph with d nodes using the absolute empirical Kendall's τ as edge weights. We construct trees $\mathcal{T}_2, \ldots, \mathcal{T}_K$ sequentially while respecting the proximity condition. For each edge, the pair-copula family is selected from a candidate set including Gaussian, Student-t, Clayton, Gumbel, Frank, and Joe families, with all rotations. Parameters are estimated by inversion of Kendall's τ followed by maximum likelihood, and the pseudo-observations for subsequent trees are computed from the conditional CDFs of the fitted pair copulas (see Aas et al., 2009). [2]

2.2 Copula-based CAPM Equilibrium

In the original BL model, a CAPM-based market equilibrium is used to compute the implied expected returns. The equilibrium return can be interpreted as the expected return that makes a mean-variance investor hold the market portfolio. In particular, considering the investor's utility function $\max_{\boldsymbol{w}}\{\boldsymbol{w}'\boldsymbol{\pi} - (\delta/2)\,\boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w}\}$ and applying reverse optimization given the optimal portfolio weights $\boldsymbol{w} = \boldsymbol{w}_{\text{mkt}}$, the equilibrium returns are obtained as

$$\boldsymbol{\pi} = \delta\,\hat{\boldsymbol{\Sigma}}\,\boldsymbol{w}_{\text{mkt}}, \tag{4}$$

where δ is the investor's risk aversion coefficient and $\boldsymbol{w}_{\text{mkt}}$ is the vector of market capitalization weights. Following He and Litterman (1999) and similar to Harris et al. (2017), we set $\delta = 2.5$ at the standard value used in the BL literature.

Unlike the original BL model, where $\hat{\boldsymbol{\Sigma}}$ is the sample covariance of historical returns, we estimate $\hat{\boldsymbol{\Sigma}}$ from the copula model fitted in Section 2.1. In particular, drawing J samples $\{\boldsymbol{z}^{(j)}\}_{j=1}^{J}$ of standardized residuals from the inverse Rosenblatt transform of the truncated R-vine, we compute

$$\hat{\boldsymbol{\Sigma}} = \operatorname{diag}(\hat{\boldsymbol{\sigma}})\,\hat{\boldsymbol{R}}\,\operatorname{diag}(\hat{\boldsymbol{\sigma}}), \tag{5}$$

[2] One can select the truncation level K using the modified BIC for vine copulas suggested by Nagler et al. (2019). Here we use a predefined fix $K = 5$ and perform robustness checks on the truncation level.

where $\hat{\boldsymbol{\sigma}} = (\hat{\sigma}_{1,T+1|T}, \ldots, \hat{\sigma}_{d,T+1|T})$ is the vector of one-step-ahead GARCH volatility forecasts from Eq. (1), and $\hat{\boldsymbol{R}}$ is the sample correlation matrix of the simulated standardized residuals. This step retains the dependence structure estimated by the vine copula and contrasts with the standard BL model, where $\hat{\boldsymbol{\Sigma}}$ is constructed from the sample second moment under an implicit Gaussian assumption.

In the next step, the conditional mean $\mu_{i,t}$ of each asset in Eq. (1) is replaced with the CAPM-implied equilibrium return π_i obtained by reverse optimization, while the copula structure on the uniform marginals is kept unchanged. Since copulas are invariant to monotone transformations of the marginals (Nelsen, 2006), the location shift preserves the entire dependence structure, namely the pairwise tail dependence, the rank correlations, and the asymmetric co-movements estimated from the data remain exactly as fitted.

The prior return distribution is constructed by combining the CAPM equilibrium mean $\boldsymbol{\pi}$ with the volatility forecasts and the dependence structure from the copula-GARCH model. The prior samples are generated via the inverse Rosenblatt transform: we draw $\boldsymbol{u}^{(j)} \sim c_{\text{trunc}}(\cdot; \hat{\boldsymbol{\eta}})$ from the fitted vine copula and compute

$$r_{0,i}^{(j)} = \pi_i + \hat{\sigma}_i \, F_i^{-1}\big(u_i^{(j)}\big), \qquad i = 1, \ldots, d, \quad j = 1, \ldots, J, \tag{6}$$

where $\hat{\sigma}_i$ is the one-step-ahead GARCH volatility forecast of asset i from Eq. (1) and F_i^{-1} is the quantile function of the skewed Student-t innovation distribution of asset i. Since copulas are invariant to monotone transformations of the marginals (Nelsen, 2006), the sampling procedure in Eq. (6) preserves the entire dependence structure estimated by the vine copula. The resulting prior samples $\{\boldsymbol{r}_0^{(j)}\}_{j=1}^{J}$ combine the CAPM-implied equilibrium returns $\boldsymbol{\pi}$ with the dependence structure from the fitted vine copula.

The joint prior density of $\boldsymbol{r}_0 = (r_{0,1}, \ldots, r_{0,d})$ implied by the sampling procedure in Eq. (6) is, by Sklar's theorem,

$$f_0(\boldsymbol{r}) = c_{\text{trunc}}\big(F_1(z_1), \ldots, F_d(z_d); \hat{\boldsymbol{\eta}}\big) \prod_{i=1}^{d} \frac{1}{\hat{\sigma}_i} f_i(z_i)\,, \qquad z_i = \frac{r_i - \pi_i}{\hat{\sigma}_i}, \tag{7}$$

where f_i is the skewed Student-t density of the standardized residual of asset i. The marginal of f_0 for asset i has mean π_i from the CAPM equilibrium and scale $\hat{\sigma}_i$ from the GARCH volatility forecast.

2.3 Black-Litterman Approach: Generalization

Black and Litterman (1991, 1992) show how to incorporate the investor views when constructing portfolio strategies. Let $\boldsymbol{r}_t = (r_{1,t}, \ldots, r_{d,t})$ be the vector of excess returns with the $d \times d$ covariance matrix $\boldsymbol{\Sigma}$ and the d-vector of expected returns $\boldsymbol{\mu}$. With $\boldsymbol{\pi}$ as the equilibrium returns and τ as a scaling parameter, the BL prior is $\boldsymbol{\mu} \sim \mathcal{N}(\boldsymbol{\pi}, \tau\boldsymbol{\Sigma})$. Given a

$k \times d$ pick matrix $\boldsymbol{P}$ representing k investor views, a k-vector $\boldsymbol{q}$ of expected values for these views, and a $k \times k$ confidence matrix $\boldsymbol{\Omega}$, the views are modelled as $\boldsymbol{q} \mid \boldsymbol{\mu} \sim \mathcal{N}(\boldsymbol{P}\boldsymbol{\mu}, \boldsymbol{\Omega})$. The posterior distribution of $\boldsymbol{\mu}$ has mean and covariance (He and Litterman, 1999)

$$\boldsymbol{\mu}_{\mathrm{BL}} = \left[(\tau\boldsymbol{\Sigma})^{-1} + \boldsymbol{P}'\boldsymbol{\Omega}^{-1}\boldsymbol{P}\right]^{-1}\left[(\tau\boldsymbol{\Sigma})^{-1}\boldsymbol{\pi} + \boldsymbol{P}'\boldsymbol{\Omega}^{-1}\boldsymbol{q}\right], \tag{8}$$

$$\boldsymbol{\Sigma}_{\mathrm{BL}} = \left[(\tau\boldsymbol{\Sigma})^{-1} + \boldsymbol{P}'\boldsymbol{\Omega}^{-1}\boldsymbol{P}\right]^{-1}. \tag{9}$$

The closed-form expression in Eqs. (8)–(9) relies on the conjugacy between the Gaussian prior on $\boldsymbol{\mu}$ and the Gaussian view likelihood. Although the BL formula is widely applied in portfolio optimization, the Gaussian assumption restricts the posterior to be Gaussian as well, which captures only the first two moments of the return distribution (Palczewski and Palczewski, 2019).

In the suggested CBL-NF approach, the Gaussian prior on $\boldsymbol{\mu}$ is replaced with the non-Gaussian copula-GARCH prior f_0 in Eq. (7), and the conjugacy is lost. To obtain a posterior under f_0, we replace the linear conjugate update with a normalizing flow $T_{\boldsymbol{\theta}} : \mathbb{R}^d \to \mathbb{R}^d$ that maps f_0 to the posterior density $f_{\boldsymbol{\theta}} = T_{\boldsymbol{\theta}} \# f_0$, where $\boldsymbol{\theta}$ is the flow parameter. The flow $T_{\boldsymbol{\theta}}$ is described in Section 2.4.

In the classical BL framework, the views $\boldsymbol{q}$ are treated as noisy observations of $\boldsymbol{P}\boldsymbol{\mu}$, and the standard Bayesian update is applied to obtain the posterior of $\boldsymbol{\mu}$. In the CBL-NF approach, the views are treated as a Gaussian target distribution on returns, $p_{\mathrm{view}}(\boldsymbol{r}) \propto \exp\{-\frac{1}{2}(\boldsymbol{q} - \boldsymbol{P}\boldsymbol{r})'\boldsymbol{\Omega}^{-1}(\boldsymbol{q} - \boldsymbol{P}\boldsymbol{r})\}$, and the prior on $\boldsymbol{\theta}$ is updated using the discrepancy between p_{view} and the flow-transported distribution $q_{\boldsymbol{\theta}} = T_{\boldsymbol{\theta}} \# f_0$. Following Bissiri et al. (2016) and Knoblauch et al. (2022), the generalized Bayesian framework updates the prior on $\boldsymbol{\theta}$ from such a discrepancy without requiring a likelihood on the returns. The generalized posterior is given by

$$\pi_\lambda(\boldsymbol{\theta}) \propto \exp\{-\lambda\, \ell(\boldsymbol{\theta})\}\, p(\boldsymbol{\theta}), \tag{10}$$

where $\ell(\boldsymbol{\theta})$ is a loss function that measures the discrepancy between p_{view} and $q_{\boldsymbol{\theta}}$, $p(\boldsymbol{\theta})$ is the prior on the flow parameters, and $\lambda > 0$ is a learning rate. Bissiri et al. (2016) show that Eq. (10) is the unique coherent posterior under the loss ℓ, and it reduces to the standard Bayesian posterior when ℓ is the negative log-likelihood and $\lambda = 1$.

We use the forward Kullback-Leibler divergence from $q_{\boldsymbol{\theta}}$ to p_{view} as the loss function, namely $\ell(\boldsymbol{\theta}) = \mathrm{KL}(q_{\boldsymbol{\theta}} \,\|\, p_{\mathrm{view}})$. The Monte Carlo computation of this loss is derived in Section 2.5. We notice that the suggested CBL-NF approach includes the classical BL formula as a special case. In particular, when f_0 is Gaussian, $T_{\boldsymbol{\theta}}$ is a location shift, $\lambda = 1$, and $\tau\boldsymbol{\Sigma} \propto \boldsymbol{I}$, the maximum a posteriori estimate from Eq. (10) reduces to the He-Litterman posterior mean in Eq. (8) (see the footnote in Section 2.5 for the derivation).

The learning rate λ controls the relative weight of the loss and the parameter prior.

Under the classical Bayesian interpretation with a Gaussian view likelihood, λ would be fixed at one. The generalized framework allows λ to be set by the investor, which lets the investor express confidence in the views beyond the view variance $\boldsymbol{\Omega}$. A large value of λ corresponds to high confidence in the views, in which case the posterior is dominated by the views. A small value of λ corresponds to low confidence in the views, in which case the posterior is close to the prior and the flow stays close to the identity.

2.4 Vine-structured Normalizing Flow

A normalizing flow is an invertible and differentiable map $T : \mathbb{R}^d \to \mathbb{R}^d$ that transforms a base density f_0 into a target density f_1 via the change-of-variables formula (Papamakarios et al., 2021):

$$f_1(\boldsymbol{r}) = f_0\big(T^{-1}(\boldsymbol{r})\big) \cdot \big|\det J_{T^{-1}}(\boldsymbol{r})\big|, \tag{11}$$

where $J_{T^{-1}}$ is the Jacobian of the inverse map. The main computational step in Eq. (11) is the Jacobian determinant. Following Papamakarios et al. (2017), the autoregressive flow constructs T with a triangular Jacobian, whose determinant is the product of the diagonal entries.

We construct the normalizing flow based on the vine tree sequence in Eq. (3). For each tree $\mathcal{T}_k$ with $k = 1, \ldots, K$, we define a flow block F_k. The block F_k transforms the coordinates appearing in the edges of $\mathcal{T}_k$ and leaves the other coordinates unchanged. For each edge $e = (j(e), k(e); D(e))$ in $\mathcal{T}_k$, the block applies an element-wise spline transform to one of the two coordinates of the edge. The spline parameters are obtained from a multilayer perceptron (MLP) with the conditioning variables $\boldsymbol{r}_{k(e)}$ and $\boldsymbol{r}_{D(e)}$ as input. The MLP has one hidden layer of n_{hid} units with softplus activation $\log(1 + \exp(\cdot))$ and a linear output layer of size $3M + 1$. The edge transformations within F_k are ordered autoregressively, namely the spline parameters of each edge depend only on coordinates that are not transformed by subsequent edges within F_k. The Jacobian of F_k is then triangular.

For the element-wise transforms, we use the monotone rational-quadratic spline of Durkan et al. (2019). The spline is defined on the bounded interval $[-B, B]$ with $B > 0$ and is the identity outside this interval. Each spline is parameterized by M bin widths, M bin heights, and $M + 1$ boundary derivatives. The $3M + 1$ unconstrained outputs of the MLP are mapped to the spline parameters as follows. The first M outputs are mapped to the bin widths and the next M to the bin heights by a softmax, so that the widths and the heights are positive and sum to the interval length. The final $M + 1$ outputs are mapped to positive boundary derivatives by a softplus. Let s'_e denote the derivative of the spline transform at edge e.

We define the full flow as the composition $T_{\boldsymbol{\theta}} = F_K \circ \cdots \circ F_2 \circ F_1$, where $\boldsymbol{\theta}$ collects the MLP parameters across all edges in the truncated vine. By the chain rule and the

triangular Jacobian within each block F_k, the log-determinant of the full Jacobian is

$$\log\left|\det J_{T_{\boldsymbol{\theta}}}\right| = \sum_{k=1}^{K} \sum_{e \in \mathcal{E}_k} \log\left|s'_e\right|. \tag{12}$$

The truncation at K trees yields $\sum_{k=1}^{K}(d-k) = Kd - K(K+1)/2$ edges. The total number of flow parameters is therefore $O(Kd)$, with $K \ll d$. At the identity initialization $\boldsymbol{\theta}_0$, all MLP weights and biases are zero except the output biases. The width and height logits are set to zero, which gives equal bin widths and heights, and the derivative logits are set to $\log(e-1) \approx 0.541$, which gives unit boundary derivatives. At $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, the flow $T_{\boldsymbol{\theta}_0}$ is the identity map and does not modify the prior.

The idea of aligning a normalizing flow with a vine tree sequence is introduced by Mitskopoulos et al. (2022), who use neural spline flows to estimate vine copula densities for neural spike data. In their approach, each pair-wise copula is fitted separately under the simplifying assumption of Haff et al. (2010), so that the conditional copulas do not depend on the conditioning variables. Our construction differs in two respects. First, we compose the pair transforms into a single map and estimate all flow parameters jointly. Second, we condition each transform on its conditioning variables through a multilayer perceptron, and we do not impose the simplifying assumption. We also use the flow to update the prior towards the investor views, rather than to estimate a density.

2.5 MAP Estimation of the Flow Parameters

We assume an isotropic Gaussian prior on the flow parameters, $p(\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\theta}_0, \sigma_\theta^2 \boldsymbol{I})$, centered at the identity initialization $\boldsymbol{\theta}_0$ from Section 2.4. By the change-of-variables formula, the forward Kullback-Leibler divergence in the loss function of Section 2.3 is given by

$$\mathrm{KL}(q_{\boldsymbol{\theta}} \,\|\, p_{\text{view}}) = \mathbb{E}_{\boldsymbol{r}_0 \sim f_0}\Big[-\log p_{\text{view}}\big(T_{\boldsymbol{\theta}}(\boldsymbol{r}_0)\big) - \log\left|\det J_{T_{\boldsymbol{\theta}}}(\boldsymbol{r}_0)\right|\Big] + C, \tag{13}$$

where $C = \mathbb{E}_{f_0}[\log f_0(\boldsymbol{r}_0)]$ does not depend on $\boldsymbol{\theta}$ and is dropped from the optimization. We substitute the Gaussian view density p_{view} and approximate the expectation by a sample average over J prior samples $\boldsymbol{r}_0^{(1)}, \ldots, \boldsymbol{r}_0^{(J)}$ from Eq. (6). The loss function is then

$$\ell(\boldsymbol{\theta}) = \frac{1}{J} \sum_{j=1}^{J} \Big[\tfrac{1}{2}\big(\boldsymbol{q} - \boldsymbol{P}\, T_{\boldsymbol{\theta}}(\boldsymbol{r}_0^{(j)})\big)' \boldsymbol{\Omega}^{-1} \big(\boldsymbol{q} - \boldsymbol{P}\, T_{\boldsymbol{\theta}}(\boldsymbol{r}_0^{(j)})\big) - \log\left|\det J_{T_{\boldsymbol{\theta}}}(\boldsymbol{r}_0^{(j)})\right|\Big]. \tag{14}$$

The first term in Eq. (14) is the average squared discrepancy between the transformed samples and the views. The log-Jacobian term $-\log|\det J_{T_{\boldsymbol{\theta}}}|$ penalizes the volume contraction of the flow and prevents the trivial solution in which the flow collapses all prior

samples to the point $\boldsymbol{q}$.

Combining the parameter prior $p(\boldsymbol{\theta})$ with the loss in Eq. (14) gives the MAP objective

$$f(\boldsymbol{\theta}) = \frac{\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|^2}{2\sigma_\theta^2} + \lambda\, \ell(\boldsymbol{\theta}). \tag{15}$$

We compute the MAP estimate $\hat{\boldsymbol{\theta}} = \arg\min f(\boldsymbol{\theta})$ using L-BFGS optimization.[3] The first-order condition of the MAP estimate is $(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)/\sigma_\theta^2 + \lambda\, \nabla \ell(\hat{\boldsymbol{\theta}}) = \mathbf{0}$, namely the deviation of the MAP estimate from the identity scales with the product $\sigma_\theta^2 \lambda$.

The posterior predictive distribution under the MAP estimate is obtained by pushing the J prior samples through the optimized flow:

$$p(\boldsymbol{r} \mid \text{views}) \approx \frac{1}{J} \sum_{j=1}^{J} \delta_{\{T_{\hat{\boldsymbol{\theta}}}(\boldsymbol{r}_0^{(j)})\}}(\boldsymbol{r}), \tag{16}$$

where $\boldsymbol{r}_0^{(j)}$ are the prior samples from Eq. (6). The J transformed samples are used as input to the portfolio optimization in Section 2.7.

2.6 Investor Views

The CBL-NF framework requires three inputs from the investor: the pick matrix $\boldsymbol{P}$, the view vector $\boldsymbol{q}$, and the confidence matrix $\boldsymbol{\Omega}$. The framework is agnostic to the source of these inputs, which can be obtained from factor-model regressions, analyst forecasts, or other forms of subjective input. In the empirical analysis, we generate $\boldsymbol{q}$ from a Gated Recurrent Unit (GRU) network of Cho et al. (2014), which has been applied in recent studies on machine learning in asset pricing (Gu et al., 2020).

For each asset $i \in \{1, \ldots, d\}$, we train a separate GRU on a lookback window of L past returns to produce a one-step-ahead forecast $\hat{\mu}_i$ of the conditional mean return. The GRU is a recurrent neural network with gating mechanisms that capture nonlinear temporal dependence in the asset's return series. We use one network for each asset to avoid pooling across assets with heterogeneous return dynamics, which would smooth out asset-specific signals (Gu et al., 2020). The GRU is trained on the first 80% of the rolling estimation window by minimizing the mean squared error between the predicted and observed returns, and the out-of-sample MSE is computed on the remaining 20%.

We construct the BL inputs from the GRU forecasts as follows. The pick matrix is set to the identity, $\boldsymbol{P} = \boldsymbol{I}_d$, which corresponds to one absolute view per asset. The view vector is set to the GRU forecasts, $\boldsymbol{q} = \hat{\boldsymbol{\mu}}$. The confidence matrix is set to the diagonal of

[3] The suggested CBL-NF approach includes the classical BL formula as a special case. When f_0 is Gaussian, $T_{\boldsymbol{\theta}}$ is a location shift $T_{\boldsymbol{\theta}}(\boldsymbol{r}) = \boldsymbol{r} + \boldsymbol{b}$, $\lambda = 1$, and $\tau\boldsymbol{\Sigma} \propto \boldsymbol{I}$, the log-Jacobian in Eq. (14) vanishes and the first-order condition gives $\hat{\boldsymbol{b}} = (\sigma_\theta^{-2}\boldsymbol{I} + \boldsymbol{P}'\boldsymbol{\Omega}^{-1}\boldsymbol{P})^{-1}\boldsymbol{P}'\boldsymbol{\Omega}^{-1}(\boldsymbol{q} - \boldsymbol{P}\boldsymbol{\pi})$. The posterior mean $\boldsymbol{\pi} + \hat{\boldsymbol{b}}$ then coincides with the He-Litterman posterior mean in Eq. (8).

the out-of-sample MSEs, $\boldsymbol{\Omega} = \text{diag}(\text{MSE}_1, \ldots, \text{MSE}_d)$, namely the view variance of each asset reflects the GRU's forecast error on data not used in training.

2.7 Portfolio Optimization

Let $\hat{\boldsymbol{r}}_t^{(j)}$ for $j = 1, \ldots, J$ be the posterior predictive samples from Eq. (16). We denote by $\hat{\boldsymbol{\mu}}_t$ and $\hat{\boldsymbol{\Sigma}}_t$ the sample mean and sample covariance of the posterior samples. We denote by $\text{CVaR}_\alpha(X)$ the conditional Value-at-Risk, also known as expected shortfall, of the random variable X at the confidence level α. We consider eight portfolio strategies, where the first four are optimization-based strategies and the last four are risk-parity strategies.

For the optimization-based strategies, the minimum variance (MinV) portfolio of Markowitz (1952) is given by

$$\min_{\hat{\boldsymbol{w}}_t} \ \hat{\boldsymbol{w}}_t' \hat{\boldsymbol{\Sigma}}_t \hat{\boldsymbol{w}}_t, \quad \text{s.t.} \quad \hat{\boldsymbol{w}}_t' \mathbf{1} = 1, \quad \hat{\boldsymbol{w}}_t \geq \mathbf{0}. \tag{17}$$

The minimum CVaR (MinCVaR) portfolio of Rockafellar and Uryasev (2000, 2002) is given by

$$\begin{aligned} \min_{\hat{\boldsymbol{w}}_t, \hat{\text{VaR}}_\alpha, \boldsymbol{\upsilon}_t} \ & \hat{\text{VaR}}_\alpha + \frac{1}{J(1-\alpha)} \sum_{j=1}^{J} \upsilon_{j,t} \\ \text{s.t.} \ & \hat{\boldsymbol{w}}_t' \hat{\boldsymbol{r}}_t^{(j)} + \hat{\text{VaR}}_\alpha + \upsilon_{j,t} \geq 0, \quad j = 1, \ldots, J, \\ & \hat{\boldsymbol{w}}_t' \mathbf{1} = 1, \quad \hat{\boldsymbol{w}}_t \geq \mathbf{0}, \quad \boldsymbol{\upsilon}_t \geq \mathbf{0}. \end{aligned} \tag{18}$$

The maximum Sharpe ratio (MaxSR) and the maximum stable tail-adjusted return ratio (MaxSTARR) portfolios are obtained by maximizing the ratio of the expected portfolio return to the portfolio standard deviation and to the portfolio CVaR, respectively. Following Charnes and Cooper (1962) and Stoyanov et al. (2007), the MaxSTARR portfolio is computed as

$$\begin{aligned} \min_{\tilde{\boldsymbol{w}}_t, \hat{\text{VaR}}_\alpha, \boldsymbol{\upsilon}_t, \nu} \ & \hat{\text{VaR}}_\alpha + \frac{1}{J(1-\alpha)} \sum_{j=1}^{J} \upsilon_{j,t} \\ \text{s.t.} \ & \tilde{\boldsymbol{w}}_t' \hat{\boldsymbol{\mu}}_t = 1, \\ & \tilde{\boldsymbol{w}}_t' \hat{\boldsymbol{r}}_t^{(j)} + \hat{\text{VaR}}_\alpha + \upsilon_{j,t} \geq 0, \quad j = 1, \ldots, J, \\ & \tilde{\boldsymbol{w}}_t' \mathbf{1} = \nu, \quad \tilde{\boldsymbol{w}}_t \geq \mathbf{0}, \quad \boldsymbol{\upsilon}_t \geq \mathbf{0}, \quad \nu > 0, \end{aligned} \tag{19}$$

with the optimal weights recovered as $\hat{w}_{j,t} = \tilde{w}_{j,t}/\nu$. The MaxSR portfolio is computed by replacing the CVaR linearization in Eq. (19) with the portfolio variance $\tilde{\boldsymbol{w}}_t' \hat{\boldsymbol{\Sigma}}_t \tilde{\boldsymbol{w}}_t$.

For the risk-parity strategies, the variance risk parity (VRP) and CVaR risk parity

(CVaRRP) portfolios are given by

$$\hat{w}_{i,t}^{\mathrm{VRP}} = \frac{1/\hat{\sigma}_{i,t}^2}{\sum_{j=1}^{d} 1/\hat{\sigma}_{j,t}^2}, \qquad \hat{w}_{i,t}^{\mathrm{CVaRRP}} = \frac{1/\mathrm{CVaR}_\alpha(r_i)}{\sum_{j=1}^{d} 1/\mathrm{CVaR}_\alpha(r_j)}, \tag{20}$$

where $\hat{\sigma}_{i,t}^2$ is the posterior predictive variance of asset i and $\mathrm{CVaR}_\alpha(r_i)$ is the CVaR estimated from the posterior distribution of asset i at level α. The VRRT and CVaRRRT incorporate the posterior mean forecast $\hat{\mu}_{i,t}$:

$$\hat{w}_{i,t}^{\mathrm{VRRT}} = \frac{\max(\hat{\mu}_{i,t}, 0)/\hat{\sigma}_{i,t}^2}{\sum_{j=1}^{d} \max(\hat{\mu}_{j,t}, 0)/\hat{\sigma}_{j,t}^2}, \qquad \hat{w}_{i,t}^{\mathrm{CVaRRRT}} = \frac{\max(\hat{\mu}_{i,t}, 0)/\mathrm{CVaR}_\alpha(r_i)}{\sum_{j=1}^{d} \max(\hat{\mu}_{j,t}, 0)/\mathrm{CVaR}_\alpha(r_j)}. \tag{21}$$

3 Steps for the CBL-NF Approach

In this section, we present the steps for constructing the CBL-NF posterior and the corresponding portfolios. We set the following parameters: T_{tot} is the total sample length, L_{est} is the rolling estimation window length, J is the number of prior samples, K is the vine truncation level, M is the number of spline bins, B is the spline domain, n_{hid} is the MLP hidden dimension, σ_θ is the prior standard deviation on the flow parameters, λ is the view weight, L is the GRU lookback, H is the GRU hidden dimension, and δ is the risk aversion coefficient. We repeat the following steps for all out-of-sample iterations $t_i \in \{t_{L_{\mathrm{est}}+1}, t_{L_{\mathrm{est}}+2}, \ldots, t_{T_{\mathrm{tot}}}\}$:

Step 1. For all $t \in [t_{i-L_{\mathrm{est}}}, t_{i-1}]$, compute the log-returns $\boldsymbol{r}_t = (r_{1,t}, \ldots, r_{d,t})$ with $r_{i,t} = \ln(p_{i,t}/p_{i,t-1})$ from the observed adjusted prices $p_{i,t}$.

Step 2. For each asset $i \in \{1, \ldots, d\}$, select the ARMA orders $(p_i, q_i) \in \{0, 1\}^2$ by BIC and fit the ARMA(p_i, q_i)-GARCH$(1, 1)$ model with skewed Student-t innovations in Eq. (1) by QML. Obtain the volatility forecast $\hat{\sigma}_{i,t_i}$ and the uniform marginals $\hat{u}_{i,t}$ from the fitted skewed Student-t distribution function.

Step 3. Fit the truncated R-vine copula in Eq. (3) at truncation level K to the uniform marginals $(\hat{U}_1, \ldots, \hat{U}_d)$ via Dissmann's algorithm (Dißmann et al., 2013). Draw J standardized residual vectors from the fitted vine through the inverse Rosenblatt transform and the inverse skewed Student-t quantile function of each asset.

Step 4. Compute the copula-implied covariance matrix $\hat{\boldsymbol{\Sigma}}$ in Eq. (5) and the equilibrium returns $\boldsymbol{\pi} = \delta\, \hat{\boldsymbol{\Sigma}}\, \boldsymbol{w}_{\mathrm{mkt}}$ in Eq. (4). Construct the prior samples $\{\boldsymbol{r}_0^{(j)}\}_{j=1}^{J}$ as in Eq. (6).

Step 5. For each asset $i \in \{1, \ldots, d\}$, train a GRU with hidden dimension H on the first 80% of the estimation window and compute the one-step-ahead forecast $\hat{\mu}_i$ and the out-of-sample MSE on the remaining 20%. Set $\boldsymbol{q} = \hat{\boldsymbol{\mu}}$, $\boldsymbol{P} = \boldsymbol{I}_d$, and $\boldsymbol{\Omega} = \mathrm{diag}(\mathrm{MSE}_1, \ldots, \mathrm{MSE}_d)$.

Step 6. Initialize the vine-structured normalizing flow at the identity $\boldsymbol{\theta}_0$ and compute the MAP estimate $\hat{\boldsymbol{\theta}} = \arg\min f(\boldsymbol{\theta})$ in Eq. (15) by L-BFGS. At each iteration, we evaluate

the flow by passing the J prior samples through the K vine trees in sequence. For each edge e in tree $\mathcal{T}_k$, the MLP ϕ_e takes the conditioning variables $\boldsymbol{r}_{k(e)}$ and $\boldsymbol{r}_{D(e)}$ as input and returns the parameters of the rational-quadratic spline that transforms the variable $\boldsymbol{r}_{j(e)}$, and the log-derivative of the spline is added to the running log-determinant in Eq. (12). After the K trees, we evaluate the loss in Eq. (14) and the objective in Eq. (15) from the transformed samples and the log-determinant, and update $\boldsymbol{\theta}$ until convergence.

Step 7. Compute the posterior predictive samples $\hat{\boldsymbol{r}}^{(j)} = T_{\hat{\boldsymbol{\theta}}}(\boldsymbol{r}_0^{(j)})$ for $j = 1, \ldots, J$ in Eq. (16), and their sample mean $\hat{\boldsymbol{\mu}}_{\text{post}}$ and covariance $\hat{\boldsymbol{\Sigma}}_{\text{post}}$.

Step 8. Insert the posterior predictive samples and moments into the portfolio optimization in Section 2.7 and obtain the optimal weights $\hat{\boldsymbol{w}}_{t_i}$. Given the proportional transaction cost C and the realized asset returns $\boldsymbol{r}_{t_i}$, compute the realized portfolio return as

$$R_{t_i} = \left[1 - C \sum_{i=1}^{d} \left| \hat{w}_{i,t_i} - \hat{w}_{i,t_i^-} \right| \right] \left(1 + \hat{\boldsymbol{w}}_{t_i}' \boldsymbol{r}_{t_i}\right) - 1,$$

where $\hat{w}_{i,t_i^-}$ is the portfolio weight of asset i at the end of the previous rebalancing period.

By the log-determinant in Eq. (12), the objective and its gradient at each L-BFGS iteration are evaluated in $O(KdJ)$ operations.

4 Simulation Study

In this section, we examine the behavior of the suggested CBL-NF approach using simulated data from a known data-generating process (DGP).

We simulate $T = 1{,}000$ observations of $d = 5$ asset returns from a C-vine copula centered at Asset 1. The pair copulas in the first tree are reported in Table 1, and the pair copulas in the higher trees are set to the independence copula. The marginals follow a constant mean with a GJR-GARCH$(1,1)$ specification (Glosten et al., 1993) and skewed Student-t innovations, where the mean, the degrees of freedom, and the skewness differ across the five assets. We treat the simulated returns as if the DGP is unknown and apply the steps in Section 3. We fit ARMA-GARCH marginals with skewed-t innovations and a truncated R-vine copula at $K = 4$, so that the marginal model is misspecified relative to the GJR-GARCH DGP. We compute the equilibrium returns under $\delta = 2.5$ and an equally weighted market portfolio, based on $J = 2{,}000$ samples from the fitted vine. Since five assets are not sufficient to estimate the GRU, we set the investor views directly as $\boldsymbol{q} = \boldsymbol{\pi} + \boldsymbol{\delta}_q$, with $\boldsymbol{\delta}_q = (0.20, -0.15, 0, -0.20, 0.10)'$, where the shifts are chosen to oppose the tail dependence of the prior. We set $\boldsymbol{P} = \boldsymbol{I}_5$ and $\boldsymbol{\Omega} = 2\,\text{diag}(\hat{\boldsymbol{\Sigma}})$. The flow uses $M = 4$ spline bins, the domain $B = 0.5$, and the hidden dimension $n_{\text{hid}} = 4$, and is initialized at the identity. We consider four pairs of

hyperparameters, $(\lambda, \sigma_\theta) \in \{(1, 0.05), (5, 0.1), (100, 0.2), (200, 0.3)\}$, which cover a wide range of the product $\lambda\sigma_\theta^2$ that controls the strength of the update.

Table 1: Pair copulas in the first tree of the C-vine DGP

Edge	Family	θ_1	θ_2	Kendall's τ	(λ_L, λ_U)
$(1, 2)$	Clayton	2.00	–	0.500	$(0.71, 0.00)$
$(1, 3)$	Rotated Clayton 270°	-1.00	–	-0.333	$(0.00, 0.50)$
$(1, 4)$	Gumbel	2.00	–	0.500	$(0.00, 0.59)$
$(1, 5)$	Student-t	0.60	3.00	0.408	$(0.34, 0.34)$

Notes: The table reports the pair copulas in the first tree of the C-vine, centered at Asset 1, that we use to simulate the returns. The families are selected to give different tail and rank dependence. θ_1 and θ_2 are the copula parameters, and λ_L and λ_U are the lower- and upper-tail dependence coefficients. The pair copulas in the higher trees $\mathcal{T}_2$, $\mathcal{T}_3$, and $\mathcal{T}_4$ are set to the independence copula.

Figure 1 reports the marginal posterior densities under the CBL-NF approach for the four pairs of hyperparameters, together with the vine copula prior and the Gaussian views. When λ and σ_θ are small, the posterior densities are close to the prior densities for all assets, and the flow remains near the identity. When λ and σ_θ increase, the posterior densities move toward the views. At $(\lambda, \sigma_\theta) = (100, 0.2)$, the posterior mean of Asset 1 increases toward its investor view and the posterior mean of Asset 4 decreases toward its investor view, while the skewness and the heavy tails of the prior remain. At $(\lambda, \sigma_\theta) = (200, 0.3)$, the posterior densities are close to the Gaussian views, although the heavy-tailed assets, such as Asset 4 with $\nu = 3$, remain non-Gaussian.

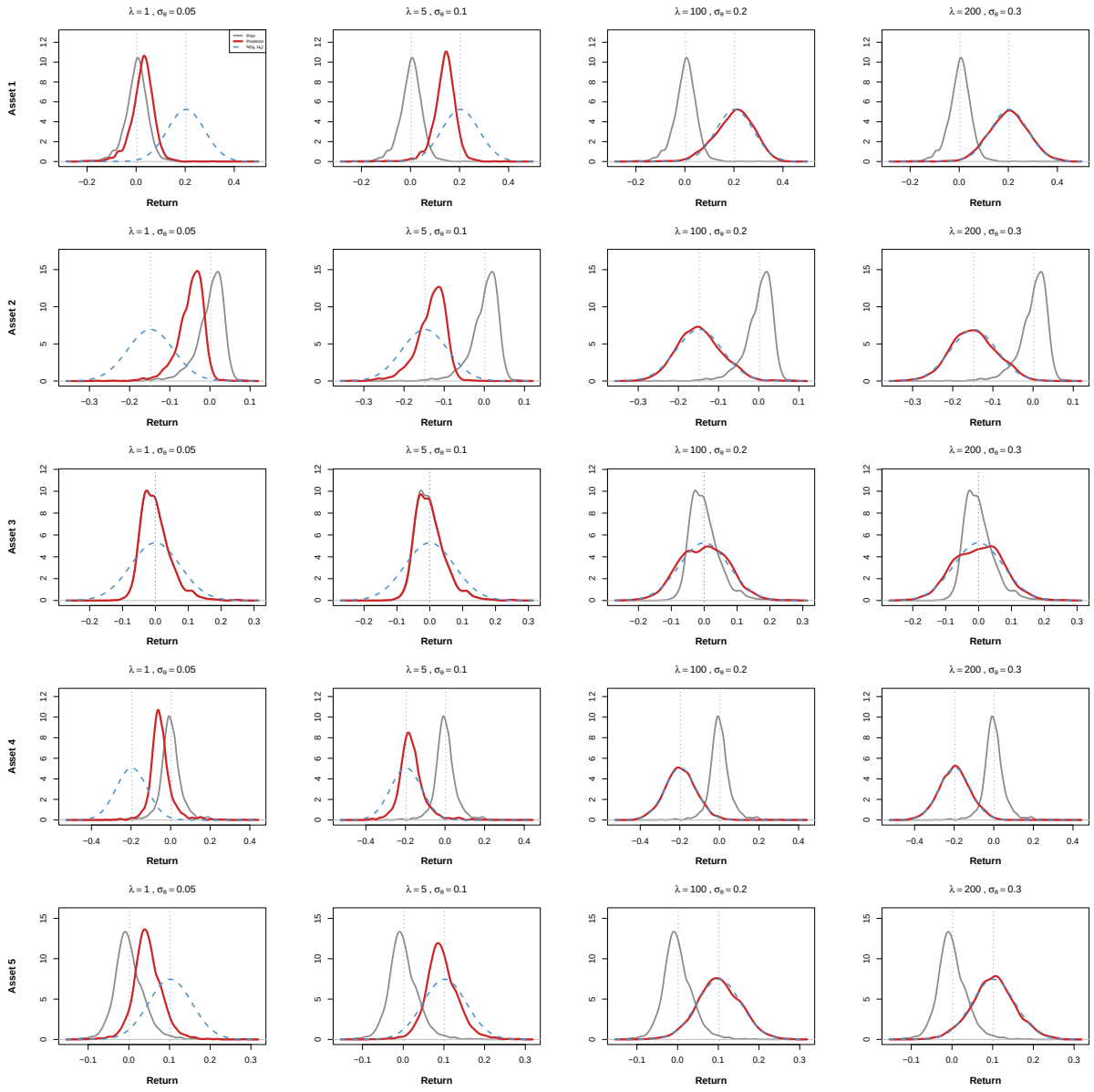


Figure 1: Marginal posterior densities under the CBL-NF approach for the simulated data. Each row corresponds to one of the five assets, and each column corresponds to one of the four pairs of hyperparameters (λ, σ_θ). The grey line is the vine copula prior, the red line is the CBL-NF posterior, and the blue dashed line is the Gaussian view $\mathcal{N}(q_i, \Omega_{ii})$. The grey and blue dotted lines show the equilibrium return π_i and the investor's view q_i.

Figure 2 reports the bivariate posterior samples for the four vine pairs in the first tree, together with the posterior Kendall's τ. When λ and σ_θ are small, the posterior values of Kendall's τ (0.479, -0.365, 0.492, and 0.424) are close to the prior values. The difference between these values and the population values of the DGP is due to the finite-sample fit of the C-vine at $T = 1{,}000$, and is not caused by the flow. The CBL-NF approach therefore retains the dependence of the prior when the views carry little weight. When λ and σ_θ increase, the posterior dependence changes at different rates across the four

pairs. At $(\lambda, \sigma_\theta) = (100, 0.2)$, the Kendall's τ of the Clayton pair $(1, 2)$ falls to 0.089, while the values for the rotated Clayton, Gumbel, and Student-t pairs (-0.364, 0.490, and 0.421) remain close to the prior. At $(\lambda, \sigma_\theta) = (200, 0.3)$, the dependence in all four pairs is close to zero, so the posterior is close to the Gaussian views. The Clayton pair is the first to lose its dependence because the views on Assets 1 and 2 move in opposite directions and oppose the lower-tail dependence of the Clayton copula. A moment-based BL update represents the dependence through a covariance matrix, which captures only linear dependence and cannot separate the lower-tail from the upper-tail behavior. The dependence change shown here is therefore not possible under the extensions of Harris et al. (2017), Deng (2018), and Sahamkhadam et al. (2022).

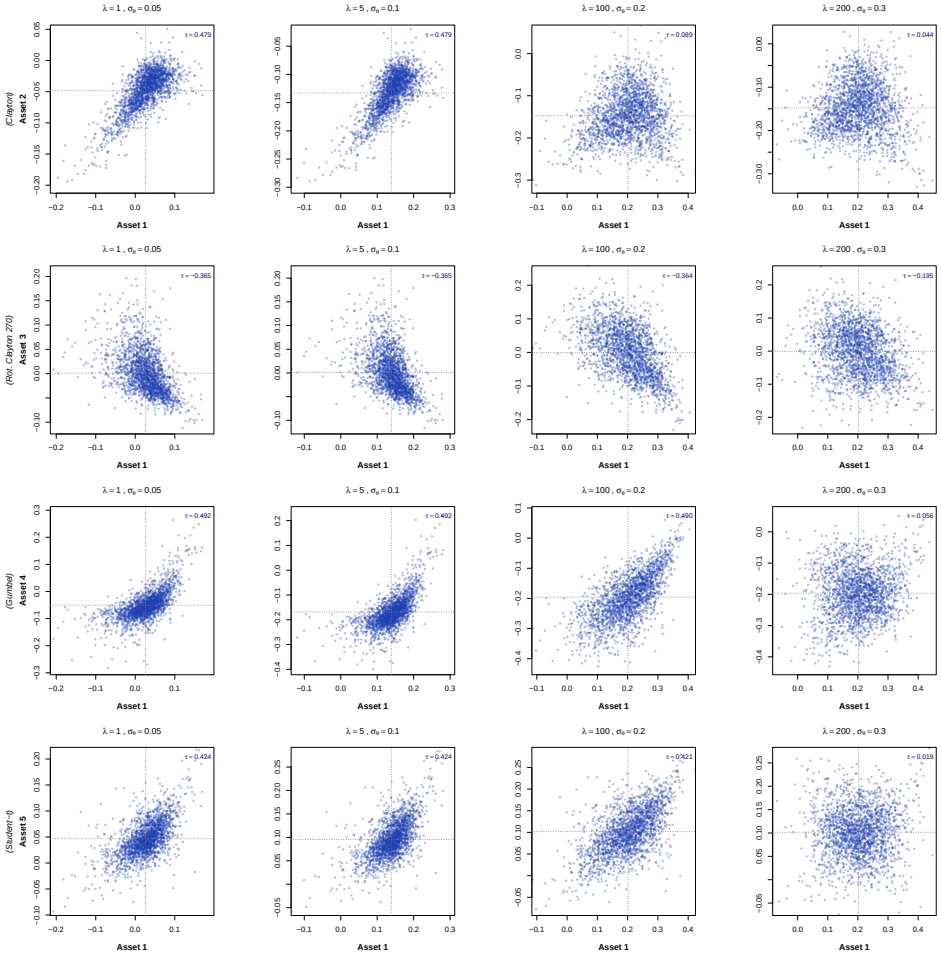


Figure 2: Bivariate posterior samples under the CBL-NF approach for the four vine pairs in the first tree. Each row corresponds to one pair, namely $(1, 2)$ Clayton, $(1, 3)$ rotated Clayton 270°, $(1, 4)$ Gumbel, and $(1, 5)$ Student-t, and the columns correspond to the four pairs of hyperparameters (λ, σ_θ) as in Figure 1. Each panel reports the posterior Kendall's τ.

5 Empirical Results

5.1 Data

To construct the CBL-NF portfolios, we use a sample of the historical constituents of the S&P 100 index. For each constituent, we obtain the total return index, which accounts for dividends and stock splits, and the market capitalization from Refinitiv LSEG Datastream. The sample runs from January 4, 2016 to January 26, 2026 at a daily frequency, and we compute the log-returns as the first differences of the natural logarithm of the total return index. At each rebalancing date, we use the index constituents at that date and keep an asset if it has no missing prices over the estimation window, a non-zero log-return on more than 20% of the days, and an available market capitalization. The number of assets varies between 88 and 95 across the out-of-sample iterations. For robustness, we also use weekly and monthly data, where the weekly sample runs from February 26, 1998 to February 19, 2026 and the monthly sample from February 1998 to January 2026.

5.2 In-sample analysis

In this subsection, we examine the behavior of the suggested CBL-NF approach on the first estimation window of the empirical sample. The window covers the first $T = 501$ daily observations, starting on January 4, 2016, and contains $d = 95$ assets. We apply the R-vine copula with truncation at $K = 5$ trees, which results in 460 pair-wise copulas. The flow uses $M = 4$ spline bins, domain $B = 0.1$, and MLP hidden dimension $n_{\text{hid}} = 2$. We fix $\sigma_\theta = 0.3$ and vary $\lambda \in \{5, 25, 50, 200\}$. For the GRU model, we set $L = 5$ and $H = 32$, and we use $J = 1000$ prior samples for the MAP estimation and $J = 2000$ for the posterior predictive.

Figure 3 reports the convergence at $(\sigma_\theta, \lambda) = (0.3, 50)$. The total objective decreases from approximately 3000 at the identity initialization to approximately 2580 within 30 iterations. The regularization term rises to 44.05, with a parameter deviation at the MAP of $\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_2 \approx 2.82$, and the log-Jacobian across the $J = 1000$ prior samples has mean 2.96 and standard deviation 4.68. The convergence in 30 iterations on a vine with 460 edges is in line with the linear cost in Eq. (12).

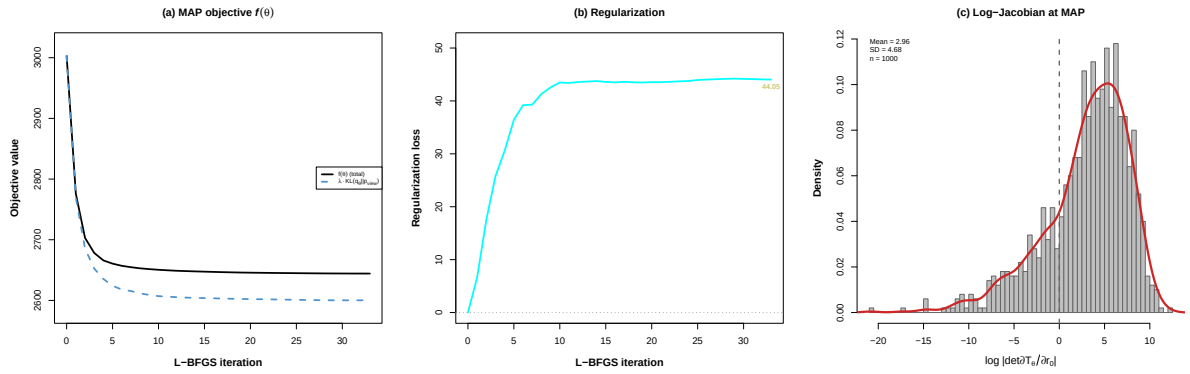


Figure 3: MAP convergence for the daily in-sample window at $\sigma_\theta = 0.3$ and $\lambda = 50$. Panel (a) shows the total objective $f(\boldsymbol{\theta})$ and the scaled forward KL component $\lambda\,\mathrm{KL}(q_\theta \,\|\, p_{\text{view}})$ across the iterations. Panel (b) shows the regularization term $\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|^2/(2\sigma_\theta^2)$. Panel (c) shows the distribution of the log-Jacobian $\log|\det \partial T_{\hat{\boldsymbol{\theta}}}/\partial \boldsymbol{r}_0|$ at the MAP across the $J = 1000$ prior samples, with mean 2.96 and standard deviation 4.68.

Figure 4 reports the marginal posterior densities for three S&P 100 constituents, with the ISINs given in the captions, across the four values of λ. For the first constituent, which has a positive investor view, the posterior mean changes from +9 basis points at $\lambda = 5$ to +13, +14, and +13 basis points at $\lambda \in \{25, 50, 200\}$, with a posterior standard deviation around 0.015 to 0.017. For the second constituent, which has a small positive view, the posterior mean moves from +2 to +4 basis points. For the third constituent, which has a negative view, the posterior mean shifts from −4 to −5 basis points. At all four values of λ, the posterior densities keep the asymmetry, the kurtosis, and the modal location of the prior, and remain different from the Gaussian view density $\mathcal{N}(q_i, \Omega_{ii})$ even at $\lambda = 200$. This is in accordance with the simulation results in Section 4 for the assets with fat tails.

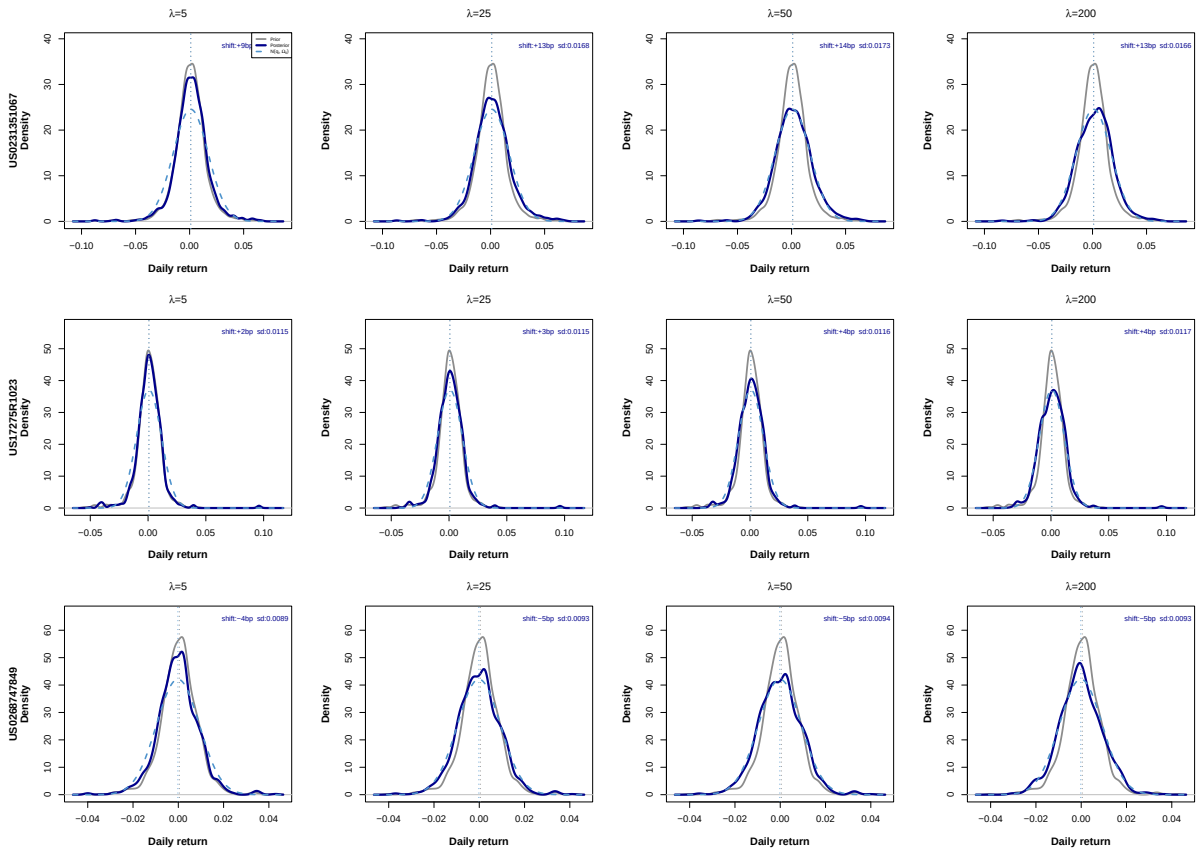


Figure 4: Marginal posterior densities for three S&P 100 constituents at $\sigma_\theta = 0.3$. Each row corresponds to one constituent (US0231351067, US17275R1023, US0268747849), and each column to one value of $\lambda \in \{5, 25, 50, 200\}$. The grey line is the vine copula prior, the blue line is the CBL-NF posterior, and the blue dashed line is the Gaussian view $\mathcal{N}(q_i, \Omega_{ii})$. The annotations report the posterior mean shift relative to the prior, in basis points, and the posterior standard deviation.

Figure 5 reports the posterior bivariate samples for the three pairs of S&P 100 constituents with the strongest prior dependence, across the four values of λ.[4] The posterior τ is constant across $\lambda \in \{5, 25, 50\}$ for all three pairs (0.693, 0.676, and 0.561), and falls to near zero at $\lambda = 200$ (-0.013, 0.045, and -0.011). There is no value of λ in this grid at which τ takes an intermediate value. In the simulation of Section 4, the Clayton pair starts to lose its dependence at $(\lambda, \sigma_\theta) = (100, 0.2)$, whereas these pairs keep their dependence up to $\lambda = 50$ and fall only at $\lambda = 200$. The prior τ values of these pairs (0.56 to 0.69) are higher than the 0.50 of the Clayton pair in the simulation, so a larger λ is needed to remove the dependence. At a large value of λ, the view quadratic in Eq. (14) moves each posterior sample toward $\mathcal{N}(\boldsymbol{q}, \boldsymbol{\Omega})$, which is diagonal, and the rank dependence falls.

[4]We use Kendall's τ rather than Pearson's correlation, since the latter is sensitive to outliers in heavy-tailed return data (Embrechts et al., 2002).

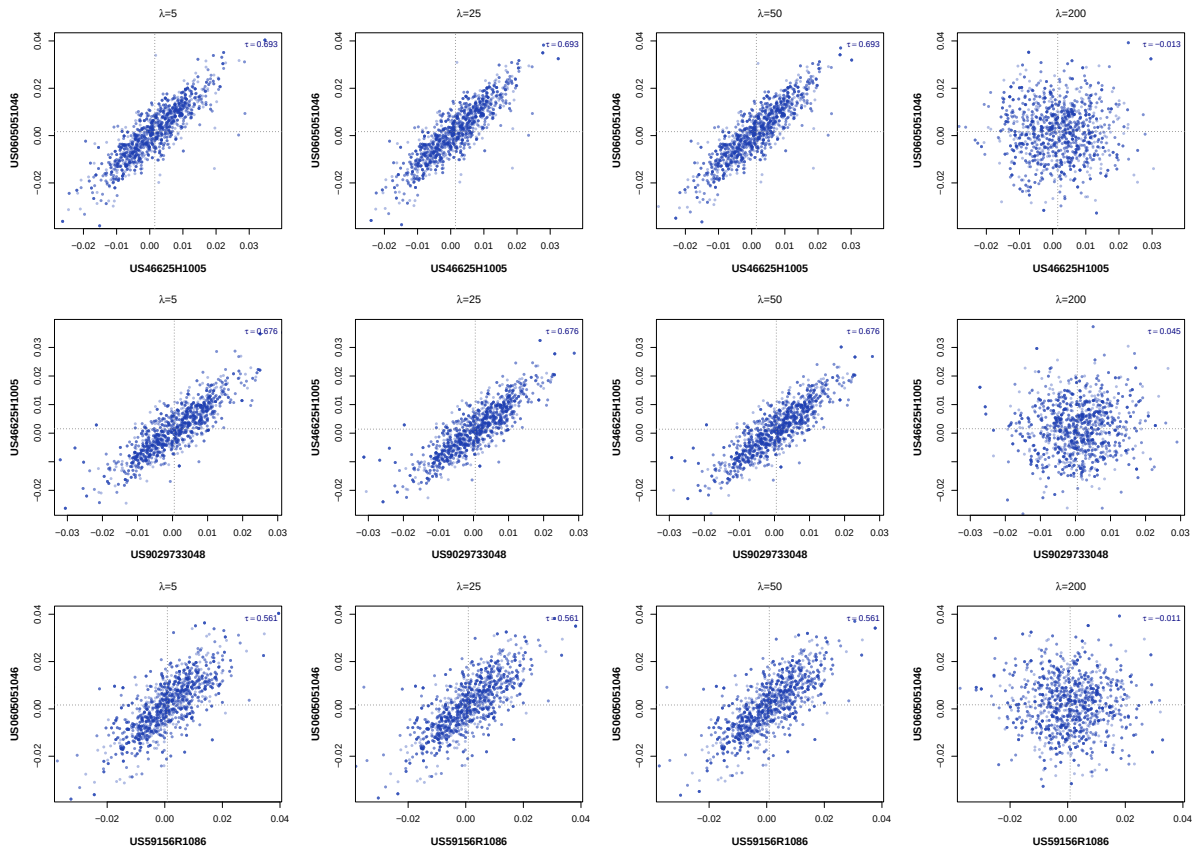


Figure 5: Posterior bivariate samples for three pairs of S&P 100 constituents at $\sigma_\theta = 0.3$. Each row corresponds to one pair, namely (US46625H1005, US0605051046), (US9029733048, US46625H1005), and (US59156R1086, US0605051046), and each column to one value of $\lambda \in \{5, 25, 50, 200\}$. Each panel reports the posterior Kendall's τ.

Figure 6 reports the rational-quadratic spline transform $T_\theta(x)$ against the identity for the most active edge in each of trees 1 to 4 at $(\sigma_\theta, \lambda) = (0.3, 50)$, evaluated at the median of the conditioning variables. We measure the activity by the deviation $\|\Delta\boldsymbol{\theta}_e\|_2$ of the edge MLP parameters from the identity initialization. The displayed edges have $\|\Delta\boldsymbol{\theta}_e\|_2$ of 0.45 (tree 1), 0.43 (tree 2), 0.27 (tree 3), and 0.51 (tree 4). Tree 3 has the smallest deviation and tree 4 the largest, so the deviation does not decrease from the first to the higher trees, and all the trees contribute to the flow.

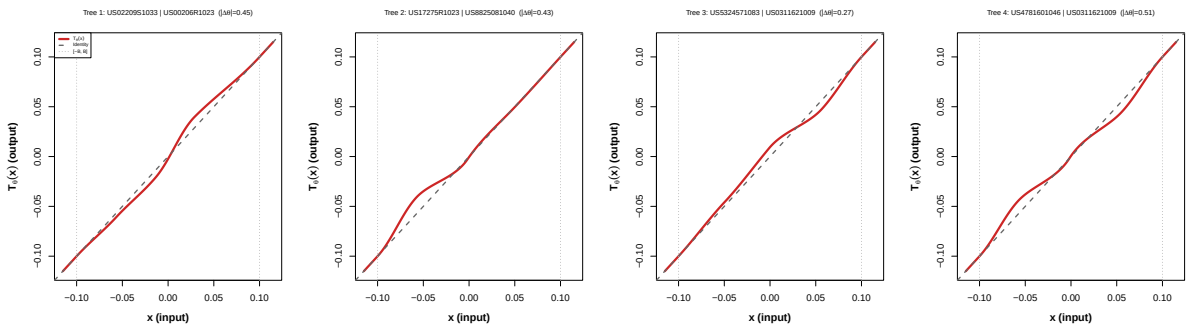


Figure 6: Rational-quadratic spline transforms at $\sigma_\theta = 0.3$ and $\lambda = 50$. For each tree $k \in \{1, 2, 3, 4\}$, the displayed edge is the one with the largest parameter deviation $\|\Delta\boldsymbol{\theta}_e\|_2$. The red line is $T_\theta(x)$ at the median of the conditioning variables, and the grey dashed line is the identity. The vertical dotted lines mark the spline domain boundaries $\pm B = \pm 0.1$.

Figures 7 and 8 report the nonparametric lower and upper tail dependence for the rank-transformed posterior samples of each pair. At $\lambda \in \{5, 25, 50\}$, the lower and upper tail dependence are close to the prior over the entire range, with a lower tail dependence around 0.55 to 0.65 at $u = 0.10$ for the first pair and an upper tail dependence of similar size. At $\lambda = 200$, both tails fall, with a lower tail dependence around 0.10 and an upper tail dependence around 0.13 at $u = 0.10$ for the first pair. The reduction is similar in the two tails, which follows from the Gaussian investor's views with zero tail dependence.

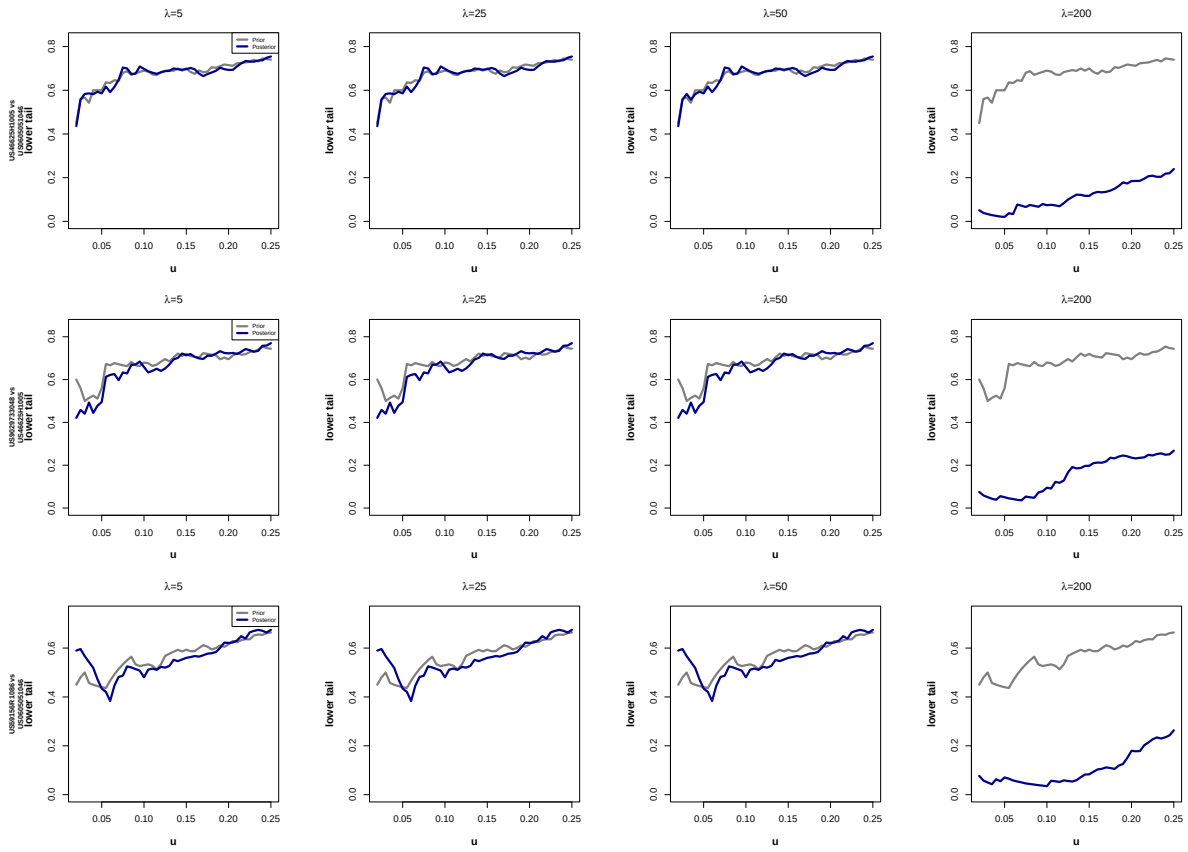


Figure 7: Nonparametric lower tail dependence for the three pairs of Figure 5 at $\sigma_\theta = 0.3$. The grey line is the vine copula prior and the blue line is the CBL-NF posterior. Each column corresponds to one value of $\lambda \in \{5, 25, 50, 200\}$.

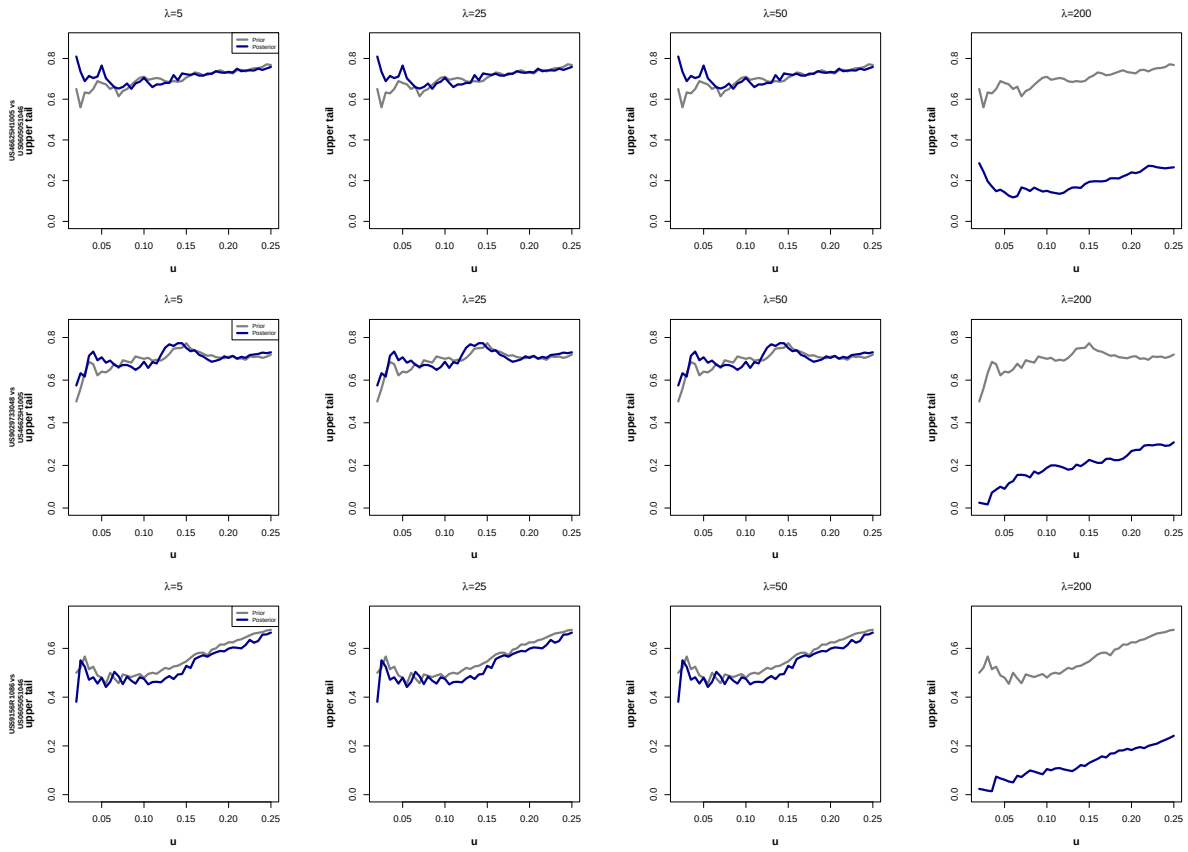


Figure 8: Nonparametric upper tail dependence for the three pairs of Figure 5 at $\sigma_\theta = 0.3$. The grey line is the vine copula prior and the blue line is the CBL-NF posterior. Each column corresponds to one value of $\lambda \in \{5, 25, 50, 200\}$.

At $\lambda \in \{5, 25, 50\}$, the posterior keeps the dependence structure of the prior while the marginal means move toward the investor views. In summary, the in-sample results indicate that (i) the MAP estimation converges in less than 100 iterations on a flow with $d = 95$ and 460 pair-wise copulas, (ii) the posterior moves the marginal means toward the views while keeping the asymmetry and the heavy tails of the prior, (iii) the spline deviations do not decrease from the first to the higher trees, and (iv) the posterior preserves the dependence of the prior up to a larger λ than in the simulation, since these pairs have stronger prior dependence.

6 Out-of-sample portfolio performance

In this section, we evaluate the out-of-sample performance of the portfolios constructed from the CBL-NF posterior. At each rebalancing date, we estimate the model on the rolling window, draw the posterior predictive samples, and solve the eight portfolio problems in Section 2.7. We hold the portfolios until the next rebalancing date, at which point we update the weights and compute the realized return net of a one basis point

proportional transaction cost. We report the average return, the standard deviation, the CVaR at the 1% level, the Sharpe ratio (SR), the STARR ratio, the average turnover, and the accumulated wealth of an initial investment of 100. We compare the portfolios against the equally weighted portfolio (EQW). Unless stated otherwise, we use the daily frequency and set $\lambda = 25$ and $\sigma_\theta = 0.1$. We first compare the portfolios constructed from the prior, namely the copula-based CAPM equilibrium (CEQ), with the portfolios constructed from the historical returns. The historical portfolios use the empirical returns over the rolling window, without any forecasting or simulation, while the CEQ portfolios use the prior equilibrium returns. We then compare the portfolios constructed from the CBL-NF posterior with those obtained from the classical BL formula, where the BL mean and covariance are estimated from the same prior and the same views.

Table 2 reports the out-of-sample performance of the historical and the CEQ portfolios. The two approaches achieve similar average returns and risk-adjusted ratios, and neither performs better across all portfolio problems. For the risk-based portfolios in Panel C, such as the minimum-variance and the minimum-CVaR, the CEQ achieves a lower CVaR (4.084 and 4.179) than the corresponding historical portfolios. For the reward-based portfolios, such as the maximum Sharpe ratio and the maximum STARR, the historical portfolios achieve higher average returns, since they load on the assets with the highest realized returns on the rolling window, while the CEQ shrinks the expected returns toward the equilibrium. However, the CEQ portfolios lead to a higher turnover.

Table 3 reports the out-of-sample performance of the classical BL and the CBL-NF portfolios. The CBL-NF portfolios in Panel C achieve a lower CVaR than the corresponding BL portfolios for most of the portfolio problems, while the average returns and the Sharpe ratios remain close. For instance, the minimum-variance and the maximum STARR portfolios achieve a lower CVaR (3.975 and 4.546) under the CBL-NF approach. This lower tail risk is in line with the in-sample result that the posterior keeps the joint left-tail dependence of the prior, while the classical BL update, which adjusts only the mean and the covariance, does not. For the CVaRRRT portfolio, the CBL-NF further achieves a higher average return (0.038) and a higher accumulated wealth (198), while the turnover remains comparable across the two approaches. In general, the CBL-NF approach reduces the tail risk compared to the classical BL update, at a similar level of average return and turnover.

Table 2: Portfolio out-of-sample performance: copula-based CAPM equilibrium versus historical.

Portfolio Strategy	Ave. Return	St. Deviation	CVaR	SR	STARR	Turnover	Wealth
Panel A: Benchmark portfolio							
EQW	0.037	1.168	5.329	0.032	0.007	0.011	190
Panel B: Historical portfolios							
MinV	0.014	0.884	4.196	0.016	0.003	0.038	124
VRP	0.033	1.058	4.989	0.031	0.007	0.010	179
VRRT	0.033	1.053	4.717	0.031	0.007	0.043	180
MaxSR	0.039	1.233	5.110	0.032	0.008	0.108	197
MinCVaR	0.023	0.927	3.947	0.025	0.006	0.039	150
CVaRRP	0.034	1.098	5.090	0.031	0.007	0.011	182
CVaRRRT	0.036	1.113	4.829	0.033	0.008	0.043	190
MaxSTARR	0.049	1.319	5.348	0.037	0.009	0.181	235
Panel C: Copula-based CAPM equilibrium portfolios							
MinV	0.019	0.921	4.084	0.021	0.005	0.617	138
VRP	0.034	1.023	4.730	0.033	0.007	0.157	183
VRRT	0.033	1.148	5.021	0.029	0.007	0.802	177
MaxSR	0.011	1.126	4.562	0.010	0.002	1.754	110
MinCVaR	0.013	0.936	4.179	0.014	0.003	1.220	121
CVaRRP	0.034	1.086	5.021	0.031	0.007	0.161	182
CVaRRRT	0.036	1.243	5.298	0.029	0.007	0.811	183
MaxSTARR	0.017	1.115	4.618	0.016	0.004	1.761	127

Notes: The table reports the portfolio out-of-sample performance at the daily frequency, with $\lambda = 25$ and $\sigma_\theta = 0.1$. Panel A reports the equally weighted portfolio (EQW). Panel B reports the portfolios constructed from the historical returns on the rolling window. Panel C reports the portfolios constructed from the copula-based CAPM equilibrium. Portfolio returns are obtained using rolling window estimation, with 501 days as the training sample, and portfolio back-testing from January 4, 2016 until January 26, 2026, with a one basis point proportional transaction cost. Except for the wealth, all measures are expressed in percent. CVaR is reported at the 1% level. The Sharpe ratio (SR) and the STARR ratio are the average return over the standard deviation and over the CVaR. Accumulated portfolio wealth is calculated at the end of the out-of-sample period assuming an initial investment of 100.

Table 3: Portfolio out-of-sample performance: CBL-NF versus classical BL.

Portfolio Strategy	Ave. Return	St. Deviation	CVaR	SR	STARR	Turnover	Wealth
Panel A: Benchmark portfolio							
EQW	0.037	1.168	5.329	0.032	0.007	0.011	190
Panel B: Classical BL portfolios							
MinV	0.019	0.918	4.078	0.020	0.005	0.591	136
VRP	0.034	1.022	4.729	0.033	0.007	0.098	185
VRRT	0.030	1.142	5.034	0.026	0.006	0.684	164
MaxSR	0.007	1.122	4.894	0.006	0.001	1.746	102
MinCVaR	0.014	0.929	4.215	0.016	0.003	0.995	124
CVaRRP	0.036	1.084	4.991	0.033	0.007	0.068	188
CVaRRRT	0.033	1.229	5.281	0.027	0.006	0.670	171
MaxSTARR	0.012	1.128	4.926	0.011	0.002	1.760	112
Panel C: CBL-NF portfolios							
MinV	0.020	0.905	3.975	0.022	0.005	0.642	140
VRP	0.033	1.015	4.685	0.033	0.007	0.147	182
VRRT	0.032	1.054	4.777	0.031	0.007	0.704	176
MaxSR	0.013	1.055	4.753	0.012	0.003	1.193	117
MinCVaR	0.013	0.928	4.060	0.014	0.003	1.219	120
CVaRRP	0.034	1.083	5.000	0.032	0.007	0.171	183
CVaRRRT	0.038	1.133	5.059	0.034	0.008	0.748	198
MaxSTARR	0.012	1.041	4.546	0.012	0.003	1.398	116

Notes: The table reports the portfolio out-of-sample performance at the daily frequency, with $\lambda = 25$ and $\sigma_\theta = 0.1$. Panel A reports the equally weighted portfolio (EQW). Panel B reports the portfolios constructed from the classical BL formula. Panel C reports the portfolios constructed from the CBL-NF posterior. The BL mean and covariance are estimated from the same prior and the same views as the CBL-NF posterior. Portfolio returns are obtained using rolling window estimation, with 500 days as the training sample, and portfolio back-testing from January 4, 2016 until January 26, 2026, with a one basis point proportional transaction cost. Except for the wealth, all measures are expressed in percent. CVaR is reported at the 1% level. The Sharpe ratio (SR) and the STARR ratio are the average return over the standard deviation and over the CVaR. Accumulated portfolio wealth is calculated at the end of the out-of-sample period assuming an initial investment of \$100.

To assess whether the differences between the CBL-NF and the BL portfolios are statistically significant, we apply a block bootstrap. For each portfolio problem and each performance measure, we draw $B = 1000$ blocks of the out-of-sample returns, compute the measure on each block for the CBL-NF and the BL portfolio, and test the difference with a one-sided Wilcoxon test. Figure 9 reports the bootstrap differences for the CVaR,

variance, Sharpe ratio, as well as the STARR ratio.

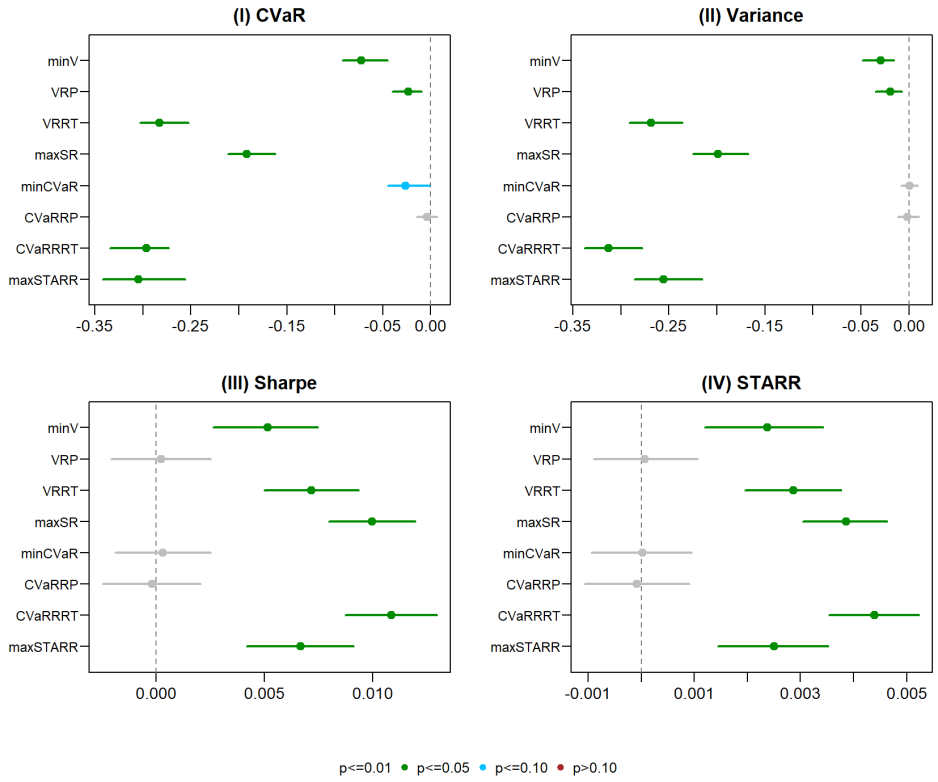


Figure 9: Bootstrap differences between the CBL-NF and the classical BL portfolios at the daily frequency, with $\lambda = 25$. Each panel corresponds to one performance measure, namely (I) the CVaR, (II) the variance, (III) the Sharpe ratio, and (IV) the STARR ratio. A point to the left of the dashed line in panels (I) and (II) indicates a lower CVaR or variance for the CBL-NF portfolio, and a point to the right of the dashed line in panels (III) and (IV) indicates a higher Sharpe ratio or STARR ratio.

The bootstrap results in Figure 9 confirm the findings from Table 3. For the CVaR in panel (I), the CBL-NF portfolios achieve a significantly lower CVaR than the BL portfolios for most of the problems, with the difference significant at the 1% level for the minimum-variance, VRP, VRRT, maximum Sharpe ratio, CVaRRRT, and maximum STARR portfolios. A similar pattern holds for the variance in panel (II). For the Sharpe ratio in panel (III) and the STARR ratio in panel (IV), the CBL-NF achieves a significantly higher ratio for the minimum-variance, VRRT, maximum Sharpe ratio, CVaRRRT, and maximum STARR portfolios, while the difference is not significant for the VRP, the minimum-CVaR, and the CVaRRP portfolios. The two risk parity portfolios, namely the VRP and the CVaRRP, show the smallest and least significant differences.

6.1 Robustness checks

In this subsection, we evaluate the robustness of the out-of-sample results to the value of λ, the data frequency, and the truncation level K. For each check, we report the

bootstrap differences between the CBL-NF and the classical BL portfolios, as in Figure 9.

We first vary λ, which is the weight on the views in the posterior, and set $\lambda = 5$ and $\lambda = 100$ at the daily frequency. Figures 10 and 11 report the bootstrap differences for the two values. At $\lambda = 5$, the posterior is close to the prior, and the differences between the CBL-NF and the BL portfolios appear mostly in the CVaR. At $\lambda = 100$, the differences are larger and significant for most of the portfolios, both in the CVaR and in the reward-based ratios. For the two values, the VRP and the CVaRRP portfolios show the smallest differences, since their weights depend on the marginal risks rather than on the joint dependence.

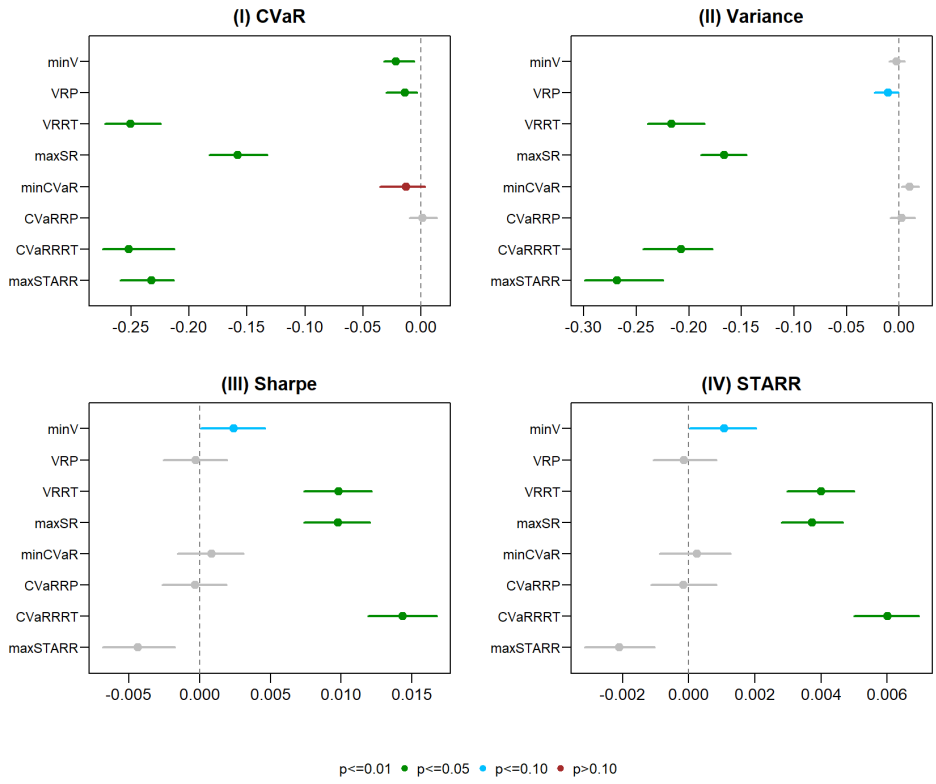


Figure 10: Bootstrap differences between the CBL-NF and the classical BL portfolios at the daily frequency, with $\lambda = 5$. See the notes to Figure 9.

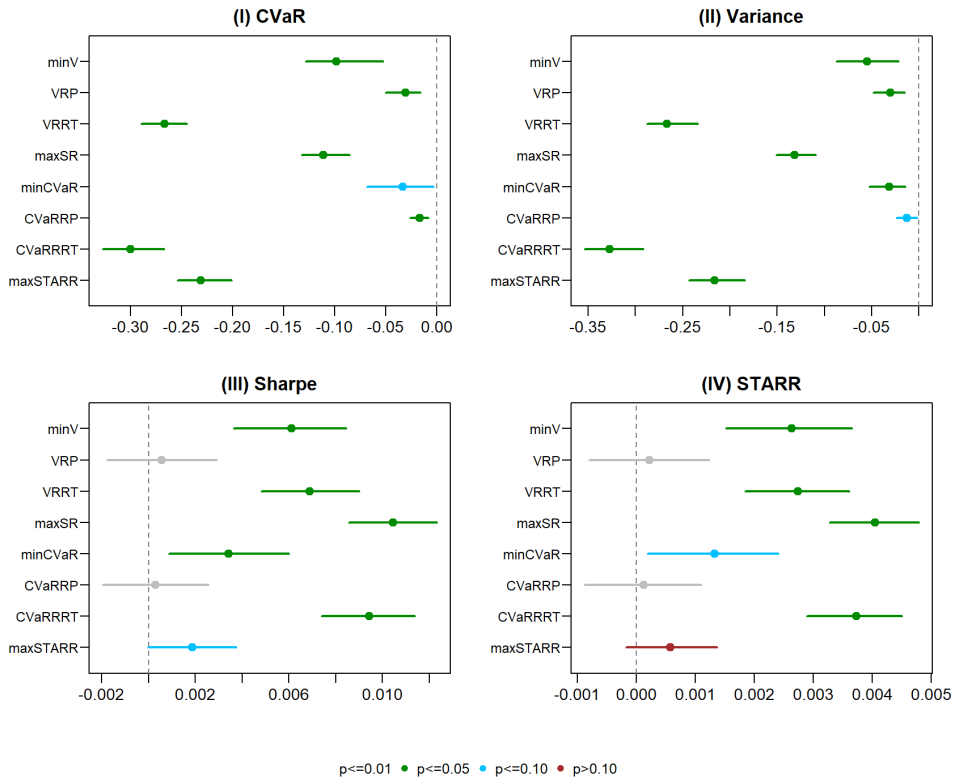


Figure 11: Bootstrap differences between the CBL-NF and the classical BL portfolios at the daily frequency, with $\lambda = 100$. See the notes to Figure 9.

We next apply the CBL-NF approach to the weekly and the monthly data, since the dependence and the tail behavior of the returns differ across frequencies. Figures 12 and 13 report the bootstrap differences at the two frequencies. The CBL-NF approach achieves a significantly lower CVaR and a significantly higher STARR ratio for the VRRT, the maximum Sharpe ratio, the CVaRRRT, and the maximum STARR portfolios at both frequencies, while the differences for the minimum-variance, the VRP, and the CVaRRP portfolios are not significant. Figure 14 reports the wealth at the weekly frequency, where the CBL-NF portfolios result in a higher terminal wealth, compared to that of the BL and the EQW portfolios for the reward-based portfolios.

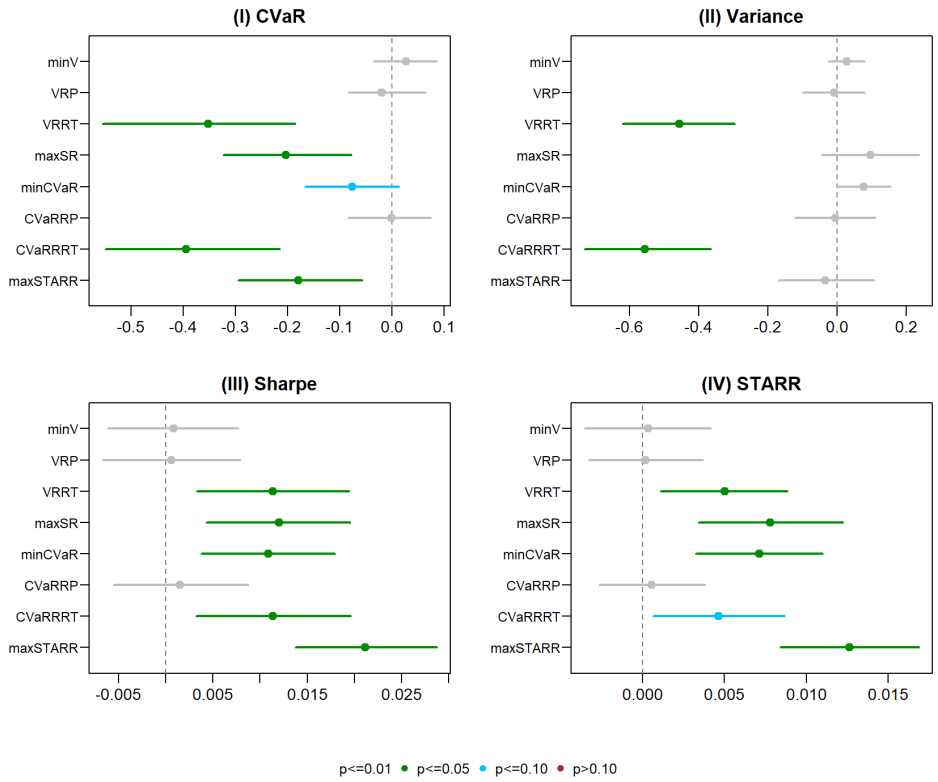


Figure 12: Bootstrap differences between the CBL-NF and the classical BL portfolios at the weekly frequency, with $\lambda = 25$. See the notes to Figure 9.

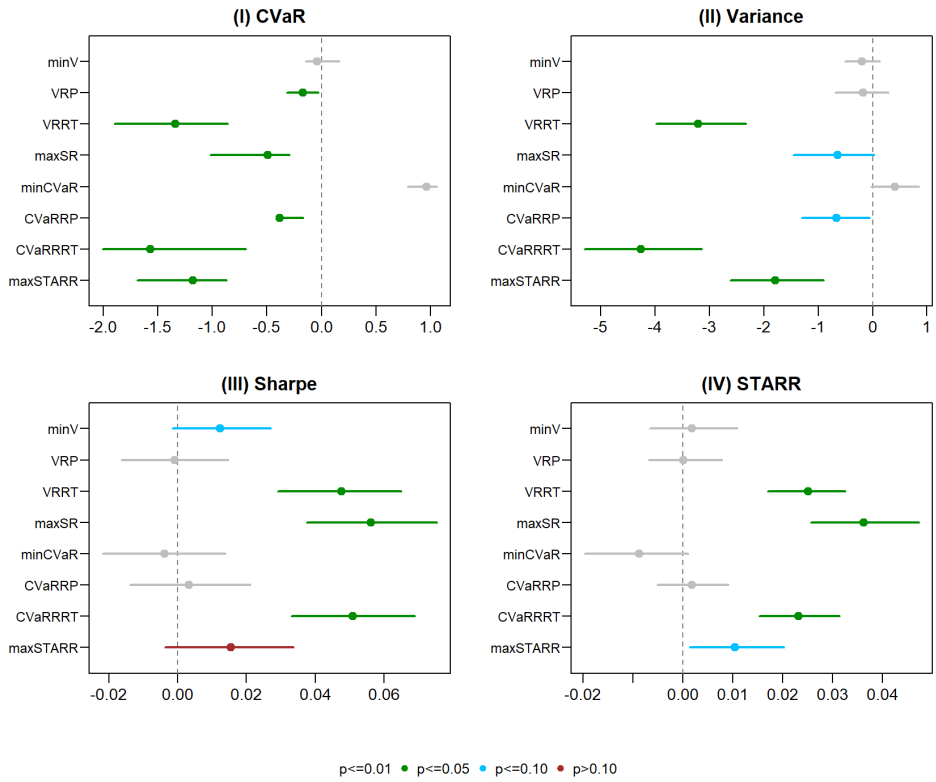


Figure 13: Bootstrap differences between the CBL-NF and the classical BL portfolios at the monthly frequency, with $\lambda = 25$. See the notes to Figure 9.

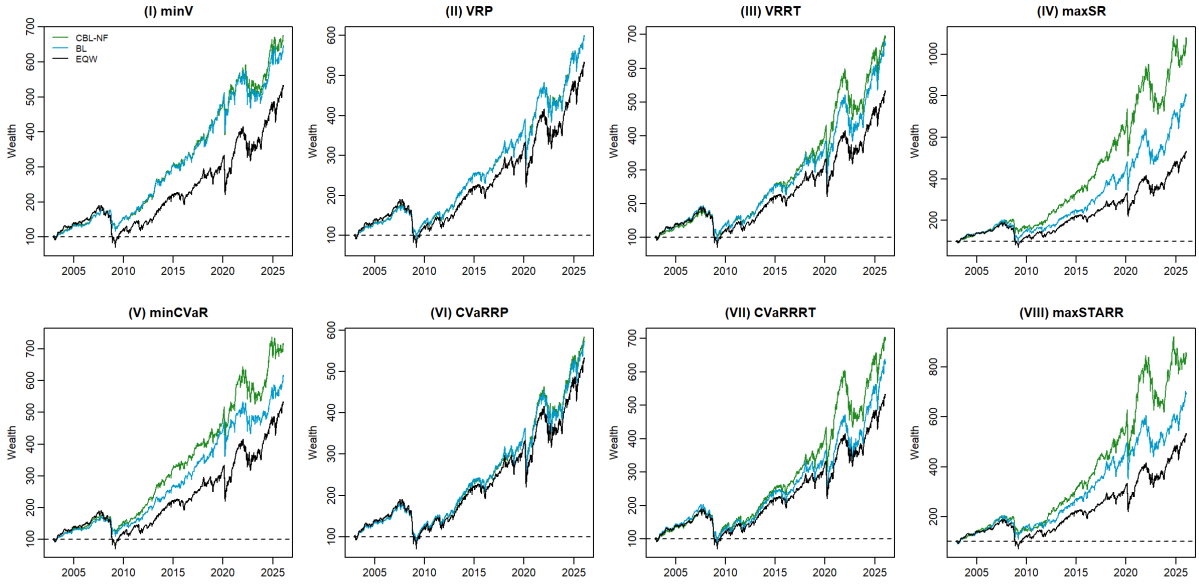


Figure 14: Wealth trajectory for the CBL-NF (green), the classical BL (blue), and the equally weighted (black) portfolios at the weekly frequency, with $\lambda = 25$. The wealth is calculated using portfolio net returns with an initial investment of \$100 at the start of the out-of-sample period.

Finally, we vary the truncation level K, which is the number of vine trees included in

the flow, and set $K = 3$ and $K = 7$ at the weekly frequency. Figures 15 and 16 report the bootstrap differences for the two values. At $K = 3$, the CBL-NF approach achieves a significantly lower CVaR for the VRRT, the CVaRRRT, and the maximum STARR portfolios, as at $K = 5$ in the main analysis. At $K = 7$, the CVaR differences are smaller and significant for fewer portfolios. The additional trees therefore do not improve the comparison, which indicates that a truncation at $K = 5$ already includes the dependence that the flow updates, while the higher trees add little.

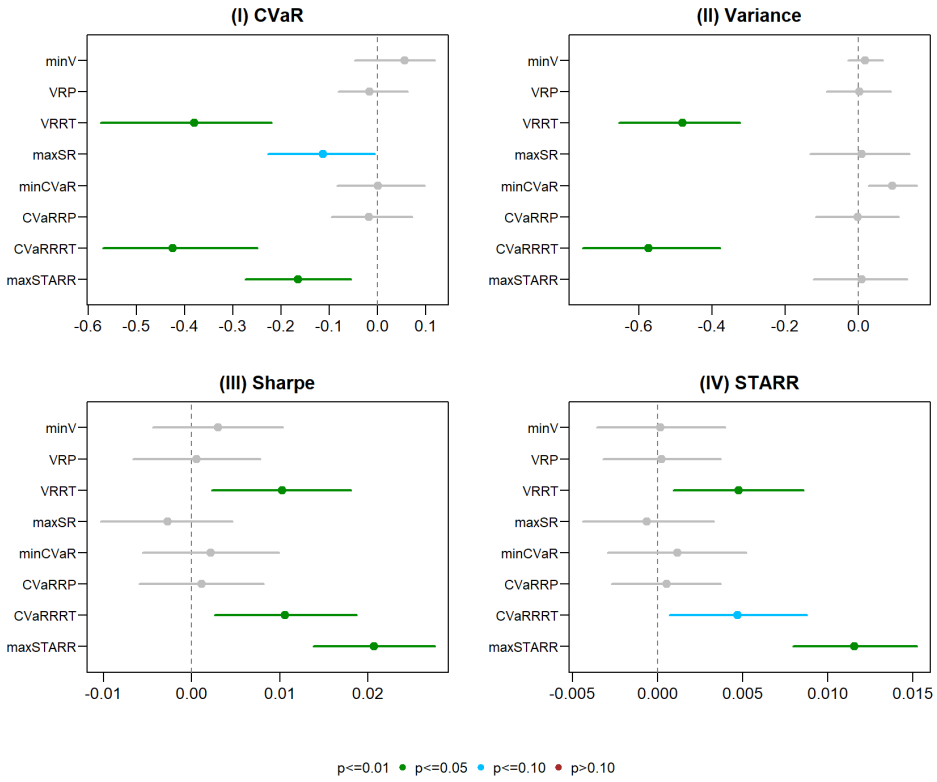


Figure 15: Bootstrap differences between the CBL-NF and the classical BL portfolios at the weekly frequency, with $\lambda = 25$ and $K = 3$. See the notes to Figure 9.

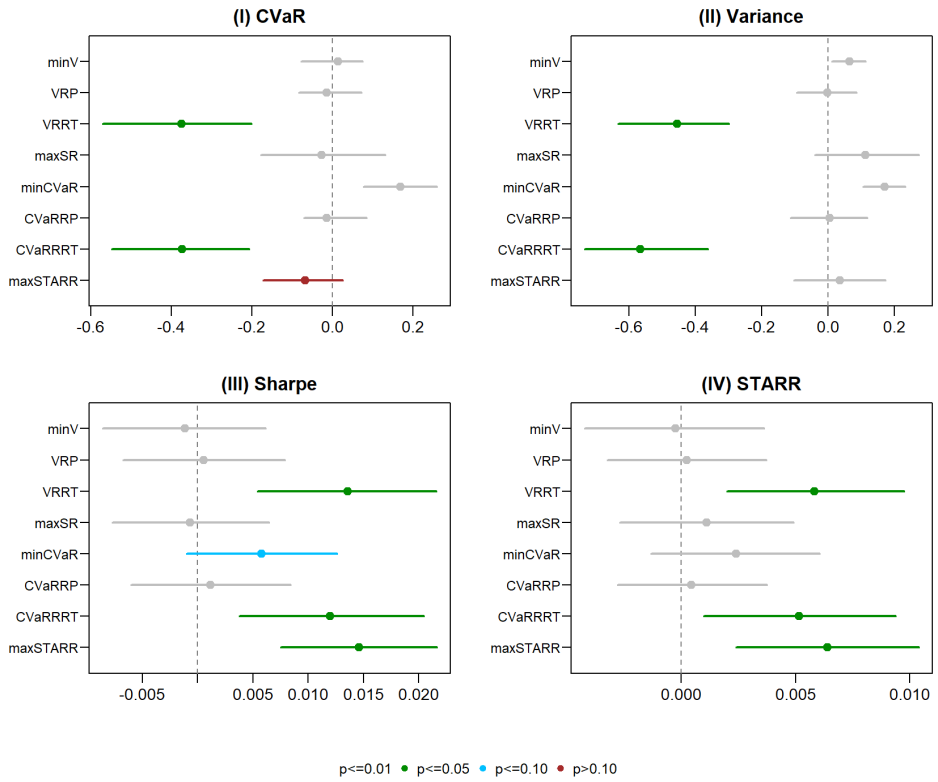


Figure 16: Bootstrap differences between the CBL-NF and the classical BL portfolios at the weekly frequency, with $\lambda = 25$ and $K = 7$. The panels and the color coding are as in Figure 9.

7 Conclusions

This article suggests a generalized Bayesian extension of the Black-Litterman model, in which a normalizing flow transforms the entire joint return distribution toward the investor views, rather than adjusting only the first two moments. We construct the prior from a truncated R-vine copula, generate investor views from a GRU model, and estimate the flow by minimizing the forward Kullback-Leibler divergence between the transformed prior and the view distribution, with a Gaussian prior on the flow parameters. We align the flow with the vine tree sequence, which follows the vine copula flows of Mitskopoulos et al. (2022), while our flow conditions each transform on its conditioning variables and is utilized as the Bayesian update of the prior rather than as a density estimator. We evaluate the approach on a simulation study and on the historical constituents of the S&P 100 index, and compare the out-of-sample portfolio performance with the historical approach and the classical BL formula.

The simulation and the in-sample analysis indicate that the CBL-NF posterior keeps the dependence structure of the prior at small and moderate values of λ, and approaches the Gaussian view distribution at large values of λ. The simulation results show that this holds regardless of the type of the prior dependence, since the flow moves the lower-

tail, the upper-tail, and the rotated pair copulas toward the tail-independent Gaussian distribution as λ increases. These findings show that the flow updates the dependence structure of the prior, which is not possible under the moment-based BL extensions, such as Harris et al. (2017), Deng (2018), and Teplova et al. (2023), that adjust only the mean and the covariance. The out-of-sample results indicate that the CBL-NF portfolios achieve a lower CVaR than the classical BL portfolios for most of the portfolio problems, at a similar level of average return and turnover, since the posterior keeps the joint lower tail dependence that affects the portfolio CVaR. These results hold across different values of λ, across the weekly and the monthly frequencies, and across the truncation level.

For future research, we suggest that the CBL-NF approach can be extended by allowing for non-Gaussian investor views. In this article, we use a Gaussian view distribution, so that the posterior approaches a tail-independent distribution as λ increases. The forward Kullback-Leibler divergence, however, only requires a view distribution with a likelihood function that can be evaluated at the transformed prior samples. The investor views can therefore be expressed through any such distribution, for instance a multivariate skewed or a heavy-tailed distribution, in which case the posterior would move toward views that themselves capture, possibly different, joint taill dependence.

Maziar Sahamkhadam and Andreas Stephan gratefully acknowledge financial support from the Nasdaq Nordic Foundation for the project entitled *Optimal Bond and Stock Portfolios Under Consideration of ESG*. We also thank participants at the XY, for their valuable comments and suggestions.

References

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

Barua, R. and Sharma, A. K. (2022). Dynamic black-litterman portfolios with views derived via cnn-bilstm predictions. *Finance Research Letters*, 49:103111.

Bedford, T. and Cooke, R. M. (2001). Probability density decomposition for conditionally dependent random variables modeled by vines. *Annals of Mathematics and Artificial Intelligence*, 32(1):245–268.

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

Bickel, P. J. and Levina, E. (2008). Regularized estimation of large covariance matrices. *The Annals of Statistics*, 36(1):199–227.

Bissiri, P. G., Holmes, C. C., and Walker, S. G. (2016). A general framework for updating belief distributions. *Journal of the Royal Statistical Society: Series B*, 78(5):1103–1130.

Black, F. and Litterman, R. (1991). Asset allocation: Combining investor views with market equilibrium. *Journal of Fixed Income*, 1(2):7–18.

Black, F. and Litterman, R. (1992). Global portfolio optimization. *Financial Analysts Journal*, 48(5):28–43.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327.

Brechmann, E. C., Czado, C., and Aas, K. (2012). Truncated regular vines in high dimensions with application to financial data. *Canadian Journal of Statistics*, 40(1):68–85.

Brechmann, E. C. and Joe, H. (2015). Truncation of vine copulas using fit indices. *Journal of Multivariate Analysis*, 138:19–33.

Charnes, A. and Cooper, W. W. (1962). Programming with linear fractional functionals. *Naval Research Logistics Quarterly*, 9(3-4):181–186.

Chen, X. and Fan, Y. (2006). Estimation and model selection of semiparametric copula-based multivariate dynamic models under copula misspecification. *Journal of Econometrics*, 135(1-2):125–154.

Cho, K., van Merriënboer, B., Bahdanau, D., and Bengio, Y. (2014). On the properties of neural machine translation: Encoder-decoder approaches. In *Proceedings of SSST-8, Workshop on Syntax, Semantics and Structure in Statistical Translation*, pages 103–111.

Chopra, V. K. and Ziemba, W. T. (1993). The effect of errors in means, variances, and covariances on optimal portfolio choice. *Journal of Portfolio Management*, 19(2):6–11.

Cont, R. (2001). Empirical properties of asset returns: Stylized facts and statistical issues. *Quantitative Finance*, 1(2):223–236.

Czado, C. (2019). *Analyzing Dependent Data with Vine Copulas: A Practical Guide with R*. Springer, Cham.

DeMiguel, V., Garlappi, L., and Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *The Review of Financial Studies*, 22(5):1915–1953.

Deng, Q. (2018). A generalized VECM/VAR-DCC/ADCC framework and its application in the Black-Litterman model: Illustrated with a China portfolio. *China Finance Review International*, 8(4):453–467.

Dißmann, J., Brechmann, E. C., Czado, C., and Kurowicka, D. (2013). Selecting and estimating regular vine copulae and application to financial returns. *Computational Statistics & Data Analysis*, 59:52–69.

Durkan, C., Bekasov, A., Murray, I., and Papamakarios, G. (2019). Neural spline flows. In *Advances in Neural Information Processing Systems*, volume 32, pages 7511–7522.

Embrechts, P., McNeil, A., and Straumann, D. (2002). Correlation and dependence in risk management: Properties and pitfalls. *Risk Management: Value at Risk and Beyond*, pages 176–223.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4):987–1007.

Friedman, J., Hastie, T., and Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441.

Giacometti, R., Bertocchi, M., Rachev, S. T., and Fabozzi, F. J. (2007). Stable distributions in the Black-Litterman approach to asset allocation. *Quantitative Finance*, 7(4):423–433.

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

Gu, S., Kelly, B., and Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, 33(5):2223–2273.

Haff, I. H. k., Aas, K., and Frigessi, A. (2010). On the simplified pair-copula construction — simply useful or too simplistic? *Journal of Multivariate Analysis*, 101(5):1296–1310.

Harris, R. D. F., Stoja, E., and Tan, L. (2017). The dynamic Black-Litterman approach to asset allocation. *European Journal of Operational Research*, 259(3):1085–1096.

He, G. and Litterman, R. (1999). The intuition behind Black-Litterman model portfolios. Technical report, Goldman Sachs Investment Management Research.

Hickling, T. and Prangle, D. (2025). Flexible tails for normalizing flows, with application to the modelling of financial return data. In *Proceedings of the 42nd International Conference on Machine Learning (ICML)*.

Joe, H. (1996). Families of m-variate distributions with given margins and $m(m-1)/2$ bivariate dependence parameters. *Lecture Notes-Monograph Series*, 28:120–141.

Joe, H. (2014). *Dependence Modeling with Copulas*. Chapman and Hall/CRC, Boca Raton.

Knoblauch, J., Jewson, J., and Damoulas, T. (2022). An optimization-centric view on Bayes' rule: Reviewing and generalizing variational inference. *Journal of Machine Learning Research*, 23:1–109.

Kolm, P. N. and Ritter, G. (2017). On the Bayesian interpretation of Black-Litterman. *European Journal of Operational Research*, 258(2):564–572.

Kolm, P. N., Tütüncü, R., and Fabozzi, F. J. (2014). 60 years of portfolio optimization: Practical challenges and current trends. In *European Journal of Operational Research*, volume 234, pages 356–371.

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, 7(1):77–91.

Meucci, A. (2006). Beyond Black-Litterman: Views on non-normal markets. *Risk*, 19:87–92.

Meucci, A. (2008). Fully flexible views: Theory and practice. *Risk*, 21(10):97–102.

Mitskopoulos, L., Amvrosiadis, T., and Onken, A. (2022). Mixed vine copula flows for flexible modeling of neural dependencies. *Frontiers in Neuroscience*, 16:910122.

Nagler, T., Bumann, C., and Czado, C. (2019). Model selection in sparse high-dimensional vine copula models with an application to portfolio risk. *Journal of Multivariate Analysis*, 172:180–192.

Nelsen, R. B. (2006). *An Introduction to Copulas*. Springer, New York, 2 edition.

Palczewski, A. and Palczewski, J. (2019). Black-Litterman model for continuous distributions. *European Journal of Operational Research*, 273(2):708–720.

Papamakarios, G., Nalisnick, E., Rezende, D. J., Mohamed, S., and Lakshminarayanan, B. (2021). Normalizing flows for probabilistic modeling and inference. *Journal of Machine Learning Research*, 22:1–64.

Papamakarios, G., Pavlakou, T., and Murray, I. (2017). Masked autoregressive flow for density estimation. In *Advances in Neural Information Processing Systems*, volume 30, pages 2338–2347.

Patton, A. J. (2006). Modelling asymmetric exchange rate dependence. *International Economic Review*, 47(2):527–556.

Rockafellar, R. T. and Uryasev, S. (2000). Optimization of conditional value-at-risk. *Journal of Risk*, 2:21–42.

Rockafellar, R. T. and Uryasev, S. (2002). Conditional value-at-risk for general loss distributions. *Journal of Banking & Finance*, 26(7):1443–1471.

Sahamkhadam, M., Stephan, A., and Östermärk, R. (2022). Copula-based Black-Litterman portfolio optimization. *European Journal of Operational Research*, 297(3):1055–1070.

Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut Statistique de l'Université de Paris*, 8:229–231.

Stoyanov, S. V., Rachev, S. T., and Fabozzi, F. J. (2007). Optimal financial portfolios. *Applied Mathematical Finance*, 14(5):401–436.

Teplova, T., Mikova, E., Munir, Q., and Pivnitskaya, N. (2023). Black-Litterman model with copula-based views in mean-CVaR portfolio optimization framework with weight constraints. *Economic Change and Restructuring*, 56:515–535.

Yang, Y. and Hospedales, T. M. (2023). Mixture of normalizing flows for european option pricing. In *Proceedings of the 39th Conference on Uncertainty in Artificial Intelligence (UAI)*, volume 216 of *Proceedings of Machine Learning Research*, pages 2390–2399. PMLR.