Research Article

Alex Verhoijsen* and Pavel Krupskiy

Fast inference methods for high-dimensional factor copulas

https://doi.org/10.1515/demo-2022-0117
received November 29, 2021; accepted July 12, 2022

Abstract: Gaussian factor models allow the statistician to capture multivariate dependence between variables. However, they are computationally cumbersome in high dimensions and are not able to capture multivariate skewness in the data. We propose a copula model that allows for arbitrary margins, and multivariate skewness in the data by including a non-Gaussian factor whose dependence structure is the result of a one-factor copula model. Estimation is carried out using a two-step procedure: margins are modelled separately and transformed to the normal scale, after which the dependence structure is estimated. We develop an estimation procedure that allows for fast estimation of the model parameters in a high-dimensional setting. We first prove the theoretical results of the model with up to three Gaussian factors. Then, simulation results confirm that the model works as the sample size and dimensionality grow larger. Finally, we apply the model to a selection of stocks of the S&P500, which demonstrates that our model is able to capture cross-sectional skewness in the stock market data.

Keywords: factor copulas, high-dimensional inference, fast inference, computational statistics

MSC 2020: Primary 62H05, 62H12, secondary 62-08

1 Introduction

In the literature, Gaussian factor models are used as a dimensionality reduction technique to describe a vector of $d$ variables $X = (X_1, \ldots, X_d)$ as a linear combination of a vector of $p$ Gaussian latent factors $Z = (Z_1, \ldots, Z_p)$, where for meaningful cases $p < d$. Thus, variables form groups based on their correlation, such that correlation within one particular group of variables can be high, while the correlation between variables assigned to different groups can be low [4,5]. Because of the minimal assumptions underlying factor models, this technique is amply used in a wide range of applications. This includes econometrics, where factor models are either used to identify the factor structure itself [1,2,18] or to control endogenous common factors in factor-augmented panel data regressions [15]. Other applications include the finance literature, see Chapter 5 in [10] or Chapter 9 in [21], and hydrology [19].

Nevertheless, the assumption of multivariate Gaussianity implies that these factor models cannot capture cross-sectional skewness or kurtosis in the underlying multivariate distribution. Copula models, on the other hand, provide a straightforward technique to capture the entire dependence structure between random variables, and allow for non-Gaussian skewness and kurtosis. A $d$-dimensional copula $C : [0, 1]^d \rightarrow [0, 1]$ is the cumulative distribution function (cdf) of a random vector $U = (U_1, \ldots, U_d)$, where $U_j \sim U(0, 1)$, for $j = 1, \ldots, d$. Let $F_1, \ldots, F_d$ be univariate cdfs, and let $x = (x_1, \ldots, x_d)$. Then by [17], the function $F(x) = \ldots$
\( C(F_1(x_1), \ldots, F_d(x_d)) \) is a \( d \)-dimensional cdf with margins \( F_1, \ldots, F_d \). In contrast, if \( F \) is a \( d \)-dimensional cdf with univariate cdfs \( F_1, \ldots, F_d \), then there exists a copula such that \( F(x) = C(F_1(x_1), \ldots, F_d(x_d)) \) holds \([17]\). If the margins are continuous, then one speaks of the copula, which is unique and equal to \( C(u) = F(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d)) \), where \( F^{-1} \) is the generalised inverse function of \( F \).

While copula models provide an elegant technique to model dependence between random variables, they come with certain restrictions. On one hand, they can be computationally demanding in a high-dimensional setting, while on the other hand they might not be flexible enough to model dependence between a large number of random variables. A solution was provided by Krupskii and Joe \([7]\), who proposed a class of factor copula models that are both parsimonious and flexible. The idea is to assume \( p \) latent factors \( V_1, \ldots, V_p \) that cause dependence between \( d \) observable variables \( U_1, \ldots, U_d \) via bivariate linking copulas. These types of factor copula models can be extended to allow for a more structured approach \([8]\). A different type of factor copula model was proposed by \([13]\). The latter model uses a simulated method of moments (SMM) estimation method described in \([12]\) and is extended to allow for time-varying parameters in \([14]\). However, estimation for this type of factor copula model is particularly slow in higher dimensions, as SMM is based on a resampling scheme. Bayesian estimation methods for factor copula models are explored by \([3, 6, 11, 16, 20]\), but are generally also computationally demanding. A novel method to decrease the computational time to estimate factor copula models is proposed by Krupskii and Joe \([9]\), who provide a procedure that allows for fast estimation of the factor copula parameters in non-overlapping factor copula models. The latter include the one-factor copula model in \([7]\) or the nested factor copula model in \([8]\).

The main contribution of this article is to provide a fast and computationally tractable procedure to model very high-dimensional data characterised by non-Gaussian skewness in the multivariate distribution. This is done in a two-step approach. First, marginals are modelled separately and transformed to normal scores. Second, dependence between the marginals is modelled by means of \( p + 1 \) common factors: one factor resulting from a one-factor copula model as described in \([7]\) and \( p \) Gaussian factors. Factor loadings and the residual correlation parameter are estimated using a generalised method of moments (GMM) approach, which provides consistent and asymptotically normal estimates as the sample size \( n \) tends to infinity. In a subsequent step, the unobserved Gaussian factors are estimated using a weighted combination of the observed variables. Weights are obtained by maximising the correlation between the estimated and the true Gaussian factors, while convergence is shown by letting the sample size \( n \) tend to infinity, and setting the dimension \( d \) sufficiently large. To allow for fast inference of the one-factor copula parameter, we use the proxy factor idea proposed by Krupskii and Joe \([9]\). We show that as \( n \) and \( d \) tend to infinity, the proxy converges towards the true factor.

In sum, the proposed model exploits both the versatility of Gaussian factor models and the fast inference scheme of non-overlapping factor copula models. The remainder of the article is organised as follows. The new model is proposed in Section 2, its main properties are discussed, and the statistical procedure is outlined. The theoretical results are outlined in Section 3, and the estimation strategy is illustrated with a simulation experiment in Section 4, where we propose a closed-form expression to quickly invert the variance-covariance matrix of the observed data. We illustrate the model by applying it to a real-world data example in Section 5. Section 6 wraps up the article with a discussion and ideas for future research.

## 2 Model outline and inference

In Section 2.1, a new factor copula model is presented that combines the characteristics of Gaussian factor models and of factor copulas, while still being computationally tractable in high dimensions. The estimation procedure is subsequently discussed in Section 2.2.

---

1 Note, however, that the copula model will only be unique in the case of continuous random variables. If the random variables are discrete, this is not necessarily the case, as illustrated by Example 5.5 in \([10]\).
2.1 Model outline

Assume one has at hand data from arbitrary margins which in a first step are transformed to \( N(0, 1) \) marginals. Then \( X = (X_1, \ldots, X_d)^T \) is a \( d \)-dimensional vector of observable variables so that \( X_j \sim N(0, 1) \), for \( j = 1, \ldots, d \). In a \( p \)-factor model, the observable variables \( X_j \) are a linear combination of \( p \) independent latent Gaussian factors and an error term:

\[
X_j = \sum_{k=1}^{p} \lambda_{jk} Z_k + \eta_j, \quad Z_1, \ldots, Z_p, \eta_j \sim \text{i.i.d. } (0, 1), \tag{2.1}
\]

where \( \lambda_{jk} = \text{Corr}(X_j, Z_k) \) and \( \eta_j = \sqrt{1 - \sum_{k=1}^{p} \lambda_{jk}^2} \), \( j = 1, \ldots, d \). If the errors \( \eta_1, \ldots, \eta_d \) are independent, then the model (2.1) is a classical Gaussian \( p \)-factor model. In this work, we assume that \( /DoubleCapE \) is \( \text{i.i.d. } (0, 1) \), and the errors are independent conditional on some unobserved factor \( V \sim U(0, 1) \) and the respective copula

\[
C(u_1, \ldots, u_d) = \prod_{j=1}^{d} C_{U|V}(u_j|V), \tag{2.2}
\]

where \( C_{U|V}(u_j|V) = \partial /\partial v C_{U|V}(u_j, v) \), and \( C_{U|V} \) is the copula linking \( (\epsilon_j, V) \). In other words, \( C \) is a one-factor copula; see [7] for more details. Casting the model in (2.1) into matrix notation gives

\[
X = \Lambda Z + \Gamma^{1/2} \epsilon, \tag{2.3}
\]

where \( X = (X_1, \ldots, X_d)^T \) is a column vector of \( d \) observable variables, and \( \Lambda = (\lambda_1, \ldots, \lambda_p) \) is a \( d \times p \) matrix of \( p \) column vectors \( \lambda_k = (\lambda_{1k}, \ldots, \lambda_{dk})^T \) consisting of \( d \) factor loadings, where \( \lambda_{jk} \) captures the impact of factor \( Z_k \) on the \( j \)-th variable \( X_j \). The \( p \) Gaussian factors are elements of the column vector \( Z = (Z_1, \ldots, Z_p)^T \), while the loadings on the \( d \) residuals \( \epsilon = (\epsilon_1, \ldots, \epsilon_d)^T \) are given in the \( d \times d \) diagonal matrix \( /DoubleCapE \) is the homogenous one-factor copula model in (2.2).

Denote

\[
H = \begin{pmatrix}
1 & \eta & \cdots & \eta \\
\eta & 1 & \cdots & \eta \\
\vdots & \ddots & \ddots & \vdots \\
\eta & \cdots & \eta & 1
\end{pmatrix},
\]

\( H \) is a \( d \times d \) matrix, with \( \eta = \text{E}(\epsilon_j\epsilon_{j_2}) \), where \( j_1 \neq j_2 \) and \( j_1, j_2 = 1, \ldots, d \). Then the correlation matrix of \( X \) is

\[
\Sigma_X = \Lambda \Lambda^T + H^{1/2} /DoubleCapE H^{1/2}. \tag{2.4}
\]

The difference with a standard Gaussian factor model lies in the assumption that \( /DoubleCapE = H \neq /DoubleCapE \), which implies the presence of correlation between the residual terms.

The following example shows that model 2.1 can capture asymmetric dependence.

**Example 2.1.** Let \( p = 2, d = 4 \) and assume that the copula linking \( (\epsilon_j, V) \) is the Clayton copula with parameter 2, which corresponds to moderate lower tail dependence. In this case, \( \eta \approx 0.52 \). Consider the model:

\[
X_j = 0.6Z_1 + 0.8\epsilon_j, \quad j = 1, 2,
X_j = 0.6Z_2 + 0.8\epsilon_j, \quad j = 3, 4.
\]

Figure 1 shows scatterplots of different pairs of the observed variables. It is seen that pairs \( (X_1, X_2) \) and \( (X_3, X_4) \) have stronger dependence in the lower tail, while \( (X_1, X_3) \) has much weaker dependence and it does not exhibit strong asymmetry.
2.2 Inference

In this section, we present the statistical procedure that ensures we can estimate the model parameters, the Gaussian factors, and the latent factor of the one-factor copula model. The corresponding theoretical results are outlined in Section 3.

We begin by estimating the parameters in the model, for which we implement a GMM approach. Let \( X_n = (X_1, \ldots, X_d) \) be a sample of size \( n \), where \( X_i = (X_{i1}, \ldots, X_{id}) \), for \( i = 1, \ldots, n \). Moreover, let \( \hat{\Sigma}_n = \hat{\Sigma}_n \) and define

\[
Q_n(\Lambda, \eta) = \sum_{j=1}^{d} \sum_{k=1}^{d} (X_{jk} - \hat{\Lambda}_{jk})^2,
\]

then the model parameters can be obtained by

\[
(\hat{\Lambda}, \hat{\eta}) = \arg\min_{(\Lambda, \eta)} Q_n(\Lambda, \eta).
\]

After obtaining the parameters, we want to estimate the Gaussian factors. This can be done by creating a proxy \( \hat{Z}_{knd}^* = \hat{w}_{knd} X_n \) of the true Gaussian factors \( Z_k \) (for \( k = 1, \ldots, p \)) by taking a weighted combination of the observable variables. Moreover, it is possible to obtain unique optimal weights \( \hat{w}_{knd} = (\hat{\Lambda}_k \hat{\Sigma}_n^{-1} \hat{\Lambda}^T_k)^{-1} \hat{\Sigma}_n^{-1} \hat{\Lambda}_k \) by minimising the \( L_2 \) norm between the proxy factor and the true factor. In Section 4.1, we propose a closed-form expression to quickly invert the \( d \times d \) variance-covariance matrix \( \Sigma \).

The next step is to recover the latent factor in the one-factor copula model. To this end, the estimated residuals can be recovered as:

\[
\hat{\epsilon}_{ind} = \frac{1}{\sqrt{n}} \left( X_j - \sum_{k=1}^{p} \hat{\Lambda}_{jk} \hat{Z}_{knd}^* \right),
\]

using the estimated values of the factor loadings. We can then use a proxy \( \hat{U}_{n} = \sum_{j=1}^{d} \hat{\epsilon}_{ind} / d \) to approximate the unknown factor \( V \) in the one-factor copula model. The density of the one-factor copula model is obtained by differentiating (2.2), that is

\[
c(u_1, \ldots, u_d) = \frac{C(u_1, \ldots, u_d)}{\partial u_1 \partial u_d} = \frac{1}{d} \prod_{j=1}^{d} c_{U,V}(u_j, v) dv.
\]

Thus, by plugging in \( \hat{U}_{n} \) as a proxy for \( V \), we can obtain the copula parameter \( \theta \) by maximising the log-likelihood

\[
\sum_{j=1}^{d} \sum_{i=1}^{n} \ln c_{U_{ij}, U_{nj}}(u_{ij}, v; \theta).
\]
3 Theoretical results

We prove these results for the model with $p \leq 3$ Gaussian factors. All proofs are deferred to Appendix B.

We assume the following conditions hold.

**Condition 3.1.** Let $\delta_{jk} = \lambda_k / y_j$, the vector $\Delta_k = [(\delta_{ik} - \delta_{mk}) / d_l]_{i,m=1}^d$, and $\phi_{kk}$ is the angle between $\Delta_k$ and $\Delta_{k_j}$. There exist $\xi_0$, $\xi_1$, $\xi_2 \in (0, \infty)$, and a constant integer $d_0 \in \mathbb{N}$ such that the following hold for any $d > d_0$ and all $j = 1, \ldots, d$, and for all $k, k_1, k_2, k_3 = 1, \ldots, p$:

(i) $\xi_0 < \sum_{k=1}^d \delta_{ik} < \xi_1 < \infty$;
(ii) $\Sigma_X$ is invertible and $\xi_0 < \|X\|_{kk} < 1 - \xi_0$ for $j \neq k$;
(iii) $\|\Delta_k\| > \xi_0$, $\phi_{kk} + \phi_{k_jk_j} - \phi_{kk_j} > \xi_0$ for $k_1 \neq k_2 \neq k_3$ and $\phi_{kk_j} + \phi_{k_jk_j} < 2\pi - \xi_0$, where $\|\Delta_k\| = \sum_{i=1}^d \sum_{m=1}^d (\delta_{ik} - \delta_{mk})^2$ is the Euclidean norm.²

The next theorem states that GMM estimation yields parameter estimates that converge in probability to their true values, given some regularity conditions. Moreover, the resulting estimators are asymptotically normal, as stated in the subsequent theorem.

**Theorem 3.2.** (Consistency of the GMM estimators) Let $f(X_n, \Lambda, \eta) = \Sigma_X - \tilde{\Sigma}_n$. The GMM estimators converge in probability to their true values as $n \to \infty$:

$$\hat{(\Lambda, \eta)} \overset{p}{\to} (\Lambda, \eta),$$

if the following conditions are satisfied:

(i) $\mathbb{E}[f(X, \Lambda, \eta)] = 0$ only for $(\Lambda, \eta) = (\Lambda_0, \eta_0)$;
(ii) The parameter space $\Theta \subset \mathbb{R}^{pd+1}$ is compact;
(iii) With probability one, $f(X, \Lambda, \eta)$ is continuous at each $(\Lambda, \eta)$;
(iv) $\mathbb{E}[\sup_{(\Lambda, \eta) \in \Theta} |f(X, \Lambda, \eta)|] < \infty$.

**Theorem 3.3.** (Asymptotic normality of GMM estimators) Let $f(X_n, \Lambda, \eta) = \Sigma_X - \tilde{\Sigma}_n$. The GMM estimators are asymptotically normal if the following conditions are satisfied:

(i) $(\hat{\Lambda}, \hat{\eta})$ is consistent;
(ii) The parameter space $\Theta \subset \mathbb{R}^{pd+1}$ is compact;
(iii) $f(X, \Lambda, \eta)$ is continuously differentiable in some neighbourhood of $(\Lambda_0, \eta_0)$ with probability one;
(iv) $\mathbb{E}[|f(X_n, \Lambda, \eta)|^2] < \infty$;
(v) $\sup_{(\Lambda, \eta) \in \Theta} \mathbb{E}[|f(X_n, \Lambda, \eta)|^2] < \infty$;
(vi) The matrix $(\nabla_{(\Lambda, \eta)} \Sigma_X) \nabla_{(\Lambda, \eta)} \Sigma_X$ is non-singular.

The following lemma states that we can obtain a proxy of the true Gaussian factors $Z_k$ (for $k = 1, \ldots, p$) by taking a weighted combination of the observable variables $X = (X_1, \ldots, X_d)^T$.

**Lemma 3.4.** (Optimal weights) Let $w_{kd} = (w_{k1}, \ldots, w_{kd})^T$ be a weight vector of length $d$ (for $k = 1, \ldots, p$), such that the weighted sum of the observed variables $\tilde{Z}_{kd} = w_{kd}^T X$ is a proxy of the Gaussian factor $Z_k$. Then one can obtain the unique optimal weight vector $w_{kd}^* = (\tilde{Z}_{kd} \Sigma_X^{-1} \Lambda_k)^{-1/2} \Sigma_X^{1/2} \Lambda_k$ by minimising $\mathbb{E}[(\tilde{Z}_{kd} - Z_k)^2]$ with respect to $w_{kd}$ such that $\text{Var}(Z_{kd}) = 1$.

Theorem 3.5 states that $\tilde{Z}_{kd}^* = w_{kd}^* X$ will converge to its true value if the dimension is sufficiently large.

² These conditions imply that vectors $\Delta_1, \Delta_2, \Delta_3$ are linearly independent. In Appendix A, we show that these are not restrictive conditions if factor loadings are not linearly dependent.
Theorem 3.5. (Convergence in $L_2$ of Gaussian factor proxies) Assume Condition 3.1 is satisfied. For $k = 1, \ldots, p$, let $\hat{Z}_{kd} = w_{kd}^T X$, with $w_{kd} = (\Lambda_k^{-1} \Sigma_k^{-1})^{1/2} \Lambda_k^{-1} \Lambda_k$. Then the following holds as $d \to \infty$,

$$\lim_{d \to \infty} \mathbb{E}[(\hat{Z}_{kd} - Z_k)^2] = 0.$$ 

Theorem 3.6 states that $\hat{Z}_{kd}$ will also converge to its true value.

Theorem 3.6. (Convergence in $L_2$ of estimated Gaussian factor proxies) Let $\hat{Z}_{kd} = \hat{w}_{kd}^T X$, with $\hat{w}_{kd} = (\hat{\Lambda}_n^{-1} \hat{\Sigma}_n^{-1})^{1/2} \hat{\Lambda}_n$. Let $\hat{\Sigma}_n^{-1}$ and $\hat{\Lambda}_n$ be the estimated counterparts of $\Sigma_k^{-1}$ and $\Lambda_k$, respectively. Then, for $k = 1, \ldots, p$,

$$\lim_{d \to \infty} \mathbb{E}[(\hat{Z}_{kd} - Z_k)^2] = 0.$$ 

Note that in Theorem 3.6, the order of the limits is crucial to ensure that convergence takes place.

Denote

$$\hat{e}_{jd} = \frac{1}{\gamma \gamma}(X_i - \sum_{k=1}^p \lambda_k \hat{Z}_{kd}).$$

A first step is to show in Lemma 3.7 that the estimated proxy for the latent factor $U_{nd} = \sum_{j=1}^d \hat{e}_{jd} / d$ converges towards the proxy for the latent factor $\bar{U}_d = \sum_{j=1}^d \hat{e}_{jd} / d$. Subsequently, in Theorem 3.8 we show that $\bar{U}_d$ converges in probability to a monotone function of the true latent factor, if the dimension is sufficiently large.

Lemma 3.7. Let $U_{nd} = \sum_{j=1}^d \hat{e}_{jd} / d$, and let $\bar{U}_d = \sum_{j=1}^d \hat{e}_{jd} / d$. Then

$$\bar{U}_d \to U_d, \quad \text{as } n \to \infty.$$ 

Theorem 3.8. (Convergence in probability to a monotonic function of the factor $V$ as $d \to \infty$) Let $\bar{U}_d = \sum_{j=1}^d \hat{e}_{jd} / d$. Under the assumptions of Theorem 5 in [9] and Theorem 3.5,

$$\bar{U}_d \to m(V),$$

as $n \to \infty$ and $d \to \infty$, where $m(\cdot)$ is a monotone function.

The next result shows that the observed variables in (2.1) are asymptotically independent so this model is not suitable for modelling data with strong tail dependence.

Theorem 3.9. (Asymptotic independence of $X$) If $0 < \gamma_l < 1$, $0 < \gamma_m < 1$, and $\gamma_l(1 - \gamma_m^{1/2}) \neq \gamma_m(1 - \gamma_l^{1/2})$ for $1 \leq l < m \leq d$, then

$$\lim_{d \to \infty} \frac{\Pr(X_i < z, X_m < z)}{\Phi(z)} = 0,$$

where $\Phi(\cdot)$ is the cdf of a standard normal distribution. This result implies that the copula linking $(X_i, X_m)$ has no lower tail dependence. Similar result holds for the upper tail.

4 Simulation study

In this section, we perform a simulation study with the goal to evaluate the performance of the estimation methods for model (2.1). The numerical implementation techniques that can be used to efficiently compute the presented estimators of the model are discussed in Section 4.1, and Section 4.2 sets up the simulation experiment for which the results are presented in Section 4.3.
4.1 Numerical implementation

The main hurdle in applying the calculations in the previous section is in obtaining the inverse of the correlation matrix, $\Sigma^{-1}$. Such an inversion becomes computationally infeasible whenever the dimension $d$ is too large. As mentioned in the proof of Theorem 3.5, the following expression can be obtained using Woodbury’s formula [Chapter 2 in 22]:

$$\Sigma^{-1} = (\Gamma^{1/2}H^{1/2})^{-1} - (\Gamma^{1/2}H^{1/2})^{-1} \Lambda (\Gamma^{1/2}H^{1/2})^{-1} A + I_p^{-1} \Lambda (\Gamma^{1/2}H^{1/2})^{-1}.$$

As one can see, the bottleneck lies in the inversion of the symmetric $d \times d$ matrix $\Gamma H\Gamma$. However, we can find that $(\Gamma^{1/2}H^{1/2})^{-1} = \{(y_{j,k})^T_{j,k=1}\}$, where

$$y_{jj} = \frac{1 + (d-2)\eta}{y_j^2(1 + (d-2)\eta - (d-1)\eta^2)},$$

$$y_{jk} = \frac{-\eta}{y_j y_k(1 + (d-2)\eta - (d-1)\eta^2)},$$

where $j \neq k$ for $j, k = 1, \ldots, d$. Using this closed-form expression for the inverse greatly improves computational efficiency, as we no longer have to deal with the inversion of a $d \times d$ matrix. Indeed, the remaining computations only require the inversion of the $p \times p$ matrix $\Lambda (\Gamma^{1/2}H^{1/2})^{-1} A + I_p$, for which in practical applications $p \ll d$.

4.2 Simulation setup

We assume that we have at hand a sample of size $n$ and dimension $d$ which admits the factor model in (2.1) with two Gaussian factors and residual one-factor copula dependence:

$$X_{ij} = \lambda_{i1} Z_{i1} + \lambda_{i2} Z_{i2} + \gamma e_{ij},$$

for $i = 1, \ldots, n$, and $j = 1, \ldots, d$. In addition, let $\lambda_{i1} \sim U(0.4, 0.8)$ and $\lambda_{i2} \sim U(0.2, 0.6)$. The following simulation procedure is used to generate the Gaussian factors and the residuals $e_{ij}$:

1. Draw a sample of size $n$ for $Z_{i1} \sim N(0, 1)$, $Z_{i2} \sim N(0, 1)$, and $W_{i1}, \ldots, W_{id}, V_i \sim U(0, 1)$ with $i = 1, \ldots, n$.

2. For each $j = 1, \ldots, d$, let $U_{ij} = C_{\theta}(W_{ij}, V_i)$.

3. Set $e_{ij} = \Phi^{-1}(U_{ij})$, where $\Phi^{-1}(\cdot)$ is the quantile function of the Gaussian cdf.

In a subsequent step, the factor model parameters are estimated using GMM, with parameter $\lambda_{i1}$ fixed at its true value to make the model parameters identifiable. Then, the factors are estimated using the weighted variables approach, where the weights are computed as in Theorem 3.6. Finally, the factor copula parameter is estimated using standard maximum likelihood.

This experiment is repeated for 1,000 independent simulations. For every experiment, one of the following settings is used for the sample size $n$, the dimension $d$, the copula parameter $\theta$, and the copula family $C$:

- The sample size is set to $n \in \{250, 500, 750, 1,000, 2,500, 5,000\}$.
- The dimension is set to $d \in \{10, 50, 100, 250, 500\}$.
- The copula parameter is calculated using Kendall’s tau, $\tau_C = 0.50$.
- The family of bivariate copulas are Clayton ($C_{\text{Cl}}$), and $t_6$ ($C_{\text{t}_6}$).

Next, over all simulation results, we calculate the mean and variance of the correlation between the true and the estimated Gaussian factors, and the Pearson correlation coefficient between the estimated copula factor $\hat{V}$, and the true copula factor $V$. Finally, we calculate the average bias of the copula parameter and its variance over all simulations. In the next section, we report these estimation results.

---

3 It is easy to see that minimising the $L_2$ norm between the true and the estimated factor is equivalent to maximising the correlation between the true and the estimated factor.
4.3 Simulation results

The results of the simulation experiments are presented in Tables 1 and 2. As one can see in both tables, for fixed $n$, the correlation between the estimated and the true factors increases as $d$ becomes larger. Similarly, for the bias of the estimated copula parameter reported in the last column of the tables, we can see that, for each $n$, there is a decrease in the bias as the dimension $d$ becomes larger. However, as $d$ increases from 250 to 500, the bias starts to increase again. On the other hand, as the sample size $n$ increases, this decrease becomes less important.

5 Data application

In this section, we apply the model outlined in the article to a real-world dataset. We consider three different sectors in the S&P500 between January 3, 2017, and December 30, 2020, which accounts for

| Table 1: Simulation results for 1,000 independent runs from a two-factor model with residual one-factor copula dependence with Clayton linking copulas |
|---|---|---|---|---|
| $n$ | $d$ | $\text{Cor}(\hat{Z}_1, \hat{Z}_{\text{med}})$ (s.d.) | $\text{Cor}(\hat{Z}_2, \hat{Z}_{\text{med}})$ (s.d.) | $\text{Cor}(V, \hat{V}_{\text{med}})$ (s.d.) | Bias($\hat{\theta}$) (s.d.) |
| 250 | 10 | 0.84 (0.05) | 0.68 (0.10) | 0.82 (0.09) | 0.13 (0.38) |
| 50 | 0.95 (0.01) | 0.92 (0.01) | 0.92 (0.01) | -0.07 (0.20) |
| 100 | 0.97 (0.01) | 0.95 (0.01) | 0.95 (0.01) | -0.10 (0.19) |
| 250 | 0.98 (0.01) | 0.98 (0.01) | 0.97 (0.01) | -0.07 (0.19) |
| 500 | 0.98 (0.03) | 0.97 (0.03) | 0.98 (0.01) | -0.07 (0.19) |
| 50 | 0.86 (0.04) | 0.74 (0.06) | 0.85 (0.04) | 0.25 (0.22) |
| 100 | 0.95 (0.01) | 0.93 (0.01) | 0.92 (0.01) | -0.07 (0.15) |
| 250 | 0.99 (0.01) | 0.95 (0.01) | 0.95 (0.01) | -0.05 (0.14) |
| 500 | 0.99 (0.02) | 0.98 (0.02) | 0.98 (0.01) | -0.04 (0.14) |
| 1,000 | 0.86 (0.03) | 0.75 (0.04) | 0.86 (0.01) | 0.29 (0.16) |
| 50 | 0.96 (0.01) | 0.93 (0.02) | 0.93 (0.01) | -0.06 (0.12) |
| 100 | 0.97 (0.01) | 0.95 (0.01) | 0.95 (0.01) | -0.04 (0.12) |
| 250 | 0.99 (0.01) | 0.98 (0.01) | 0.98 (0.01) | -0.03 (0.12) |
| 500 | 0.99 (0.02) | 0.98 (0.02) | 0.99 (0.01) | -0.05 (0.12) |
| 2,500 | 0.87 (0.03) | 0.76 (0.03) | 0.86 (0.01) | 0.29 (0.14) |
| 50 | 0.96 (0.01) | 0.93 (0.01) | 0.93 (0.01) | -0.06 (0.10) |
| 100 | 0.97 (0.01) | 0.95 (0.01) | 0.96 (0.01) | -0.04 (0.10) |
| 250 | 0.99 (0.01) | 0.98 (0.01) | 0.98 (0.01) | -0.02 (0.10) |
| 500 | 0.99 (0.01) | 0.98 (0.01) | 0.99 (0.01) | -0.05 (0.10) |
| 5,000 | 0.88 (0.02) | 0.78 (0.02) | 0.87 (0.01) | 0.31 (0.09) |
| 50 | 0.96 (0.01) | 0.93 (0.01) | 0.93 (0.01) | -0.04 (0.07) |
| 100 | 0.98 (0.01) | 0.96 (0.01) | 0.96 (0.01) | -0.03 (0.06) |
| 250 | 0.99 (0.01) | 0.98 (0.01) | 0.98 (0.01) | -0.02 (0.06) |
| 500 | 0.99 (0.01) | 0.98 (0.01) | 0.99 (0.01) | -0.04 (0.07) |

Note: Loadings on the first factors are uniformly chosen between 0.4 and 0.8, while loadings on the second factor are uniformly chosen between 0.2 and 0.6.
Table 2: Simulation results for 1,000 independent runs from a two-factor model with residual one-factor copula dependence with \( t_c \) linking copulas

| \( n \) | \( d \) | \( C_{or}(Z_t, \hat{Z}_{s_{2,od}}) \) (s.d.) | \( C_{or}(Z_t, \hat{Z}_{s_{2,od}}) \) (s.d.) | \( C_{or}(V, \hat{V}_{s_{2,od}}) \) (s.d.) | Bias(\( \hat{\theta} \)) (s.d.) |
|---|---|---|---|---|---|
| 250 | 10 | 0.83 (0.05) | 0.63 (0.11) | 0.79 (0.13) | -0.05 (0.02) |
| | 50 | 0.94 (0.02) | 0.90 (0.03) | 0.93 (0.02) | -0.02 (0.03) |
| | 100 | 0.96 (0.02) | 0.93 (0.03) | 0.96 (0.02) | -0.02 (0.03) |
| | 250 | 0.98 (0.02) | 0.97 (0.02) | 0.98 (0.01) | -0.02 (0.02) |
| | 500 | 0.96 (0.05) | 0.95 (0.06) | 0.98 (0.02) | -0.02 (0.03) |
| 500 | 10 | 0.88 (0.04) | 0.70 (0.08) | 0.86 (0.07) | 0.02 (0.12) |
| | 50 | 0.95 (0.01) | 0.92 (0.01) | 0.94 (0.01) | -0.01 (0.02) |
| | 100 | 0.97 (0.01) | 0.95 (0.01) | 0.97 (0.01) | -0.01 (0.02) |
| | 250 | 0.99 (0.01) | 0.98 (0.01) | 0.99 (0.01) | -0.01 (0.02) |
| | 500 | 0.98 (0.03) | 0.97 (0.03) | 0.99 (0.01) | -0.01 (0.02) |
| 750 | 10 | 0.85 (0.04) | 0.73 (0.06) | 0.88 (0.04) | 0.04 (0.06) |
| | 50 | 0.95 (0.01) | 0.92 (0.01) | 0.95 (0.01) | -0.01 (0.02) |
| | 100 | 0.97 (0.01) | 0.95 (0.01) | 0.97 (0.01) | -0.01 (0.01) |
| | 250 | 0.99 (0.01) | 0.98 (0.01) | 0.99 (0.01) | -0.01 (0.01) |
| | 500 | 0.99 (0.02) | 0.98 (0.02) | 0.99 (0.01) | -0.01 (0.01) |
| 1,000 | 10 | 0.86 (0.04) | 0.74 (0.04) | 0.88 (0.02) | 0.05 (0.02) |
| | 50 | 0.95 (0.01) | 0.92 (0.01) | 0.95 (0.01) | -0.01 (0.01) |
| | 100 | 0.97 (0.01) | 0.95 (0.01) | 0.97 (0.01) | -0.01 (0.01) |
| | 250 | 0.99 (0.01) | 0.98 (0.01) | 0.99 (0.01) | -0.01 (0.01) |
| | 500 | 0.99 (0.01) | 0.98 (0.01) | 0.99 (0.01) | -0.01 (0.01) |
| 2,500 | 10 | 0.87 (0.02) | 0.77 (0.02) | 0.89 (0.02) | 0.05 (0.01) |
| | 50 | 0.96 (0.01) | 0.93 (0.01) | 0.95 (0.01) | -0.01 (0.01) |
| | 100 | 0.97 (0.01) | 0.95 (0.01) | 0.97 (0.01) | 0.00 (0.01) |
| | 250 | 0.99 (0.01) | 0.98 (0.01) | 0.99 (0.01) | 0.00 (0.01) |
| | 500 | 0.99 (0.01) | 0.98 (0.01) | 0.99 (0.01) | -0.01 (0.01) |
| 5,000 | 10 | 0.88 (0.01) | 0.77 (0.01) | 0.89 (0.01) | 0.05 (0.01) |
| | 50 | 0.96 (0.01) | 0.93 (0.01) | 0.95 (0.01) | 0.00 (0.01) |
| | 100 | 0.97 (0.01) | 0.95 (0.01) | 0.97 (0.01) | 0.00 (0.01) |
| | 250 | 0.99 (0.01) | 0.98 (0.01) | 0.99 (0.01) | -0.01 (0.01) |
| | 500 | 0.99 (0.01) | 0.98 (0.01) | 0.99 (0.01) | -0.01 (0.01) |

Note: Loadings on the first factors are uniformly chosen between 0.4 and 0.8, while loadings on the second factor are uniformly chosen between 0.2 and 0.6.

1,006 observations. The sectors considered are Industrials (71 stocks), Information Technology (74 stocks), and Financials (64 stocks) as classified by the Global Industry Classification Standard.4

To model the univariate time series, we use an AR(1)-GARCH(1,1) process for the log-returns \( s_{jt} = \log(S_{jt}/S_{jt-1}) \) of stock price \( S_{jt} \), for each individual stock \( j = 1, \ldots, d \) and each time period \( t = 1, \ldots, T \). The univariate time series model can be written as follows:

\[
S_{jt} = \mu_j + \phi_j S_{jt-1} + \kappa_j \hat{\epsilon}_{jt},
\]

\[
\kappa_j^2 = \omega_j + \alpha_j \kappa_{jt-1}^2 + \beta_j \kappa_{jt-1}^2,
\]

where \( \mu_j \) is the unconditional mean, and \( \phi_j \) gives, for stock \( j \), the dynamic impact of a change in the log-returns in period \( t - 1 \). Furthermore, \( \kappa_j \) introduces, for stock \( j \), non-constant volatility to the process. The latter is a function of the average volatility \( \omega_j \), the dynamic impact \( \alpha_j \) of a change in the previous period of the squared series, and \( \beta_j \) captures the impact of a change in the volatility at time \( t - 1 \). Finally, error terms

---

4 Three stocks (CARR, IR, and OTIS) are excluded from the Industrials sector, as they were only included in the S&P500 in 2020.
\( \xi_{jt} \) are i.i.d. Skew-t distributed with \( \nu \) degrees of freedom and skewness parameter \( \gamma \). After estimating the univariate parameters, we transform the estimated residuals to a uniform scale.

Next, we want to identify the number of underlying factors. In Section 2.4 of [13] it is shown that, under regularity conditions, one can use a scree plot based on the eigenvalues of the rank correlation matrix to determine the appropriate number of factors. The scree plot for the AR(1)-GARCH(1,1) filtered data transformed to uniform ranks is given in Figure 2. One can see that for the Industrials sector, the majority of the variance in the data is explained by one factor, while for the Information Technology and Financials sectors, two common factors appear to be more appropriate to capture the variance in the data. Thus, when we look at the number of factors required to simultaneously model the Industrials, Information Technology, and Financials sectors, it is no surprise that the number of factors appears to be between 2 and 4.

Next, we model the multivariate dependence between the different log-returns of stock prices. First, we transform the data to normal scores and look at the empirical semi-correlation of the data. Semi-correlation can be defined as the correlation in the lower tail \((\rho_{<})\) or upper tail \((\rho_{>})\) of the data. We can see in Table 3 that the average absolute semi-correlation indicates that the data in each of the three sectors are skewed to the left.\(^5\) This is consistent with stylised facts in the financial literature that show stronger correlation during downturns. We therefore expect that a copula with stronger dependence in the lower tail is more appropriate to model these data. For comparison, we use different copula models to see which one fits these characteristics best.

---

\(^5\) The absolute semi-correlation is calculated to account for the seven negative values of upper semi-correlation in the Information Technology sector. However, these are small enough such that they do not affect the average up to two decimals.
First, we model dependence within each sector separately using only one Gaussian factor, and a Clayton, Frank, and normal one-factor copula model to capture the residual dependence. Because the lower tail dependence does not appear to be very strong, we also use a reflected Gumbel copula model to capture the residual dependence. Then, we subsequently use two and three Gaussian factors while using the same one-factor copula models. These estimation results for each sector are presented in Table 4. As one can see in the model with one Gaussian factor, all three sectors establish a moderate degree of dependence between its constituent stocks. This is not surprising, as linear dependence is already captured by the Gaussian factor. Similarly, when pairing industry sectors together in a model with two Gaussian factors, we can see that there is a moderate degree of dependence between the stocks in each of the pairs of industry sectors. Finally, when using three Gaussian factors, we can see that the estimated parameter for each of the three copula models is somewhat higher than the estimated dependence parameters in the model with two Gaussian factors.

To evaluate the performance of our model, we use the estimated parameters to simulate a sample of size 1,000,000. We then calculate the semi-correlations of this simulated sample and calculate the average (absolute) bias with respect to the empirical semi-correlations in Table 3. These results are reported in Tables 5 and 6. As we can see, the average absolute bias of the lower semi-correlation, \( b(\rho_{0}) \), is smallest for the model using reflected Gumbel linking copulas in the model with one Gaussian factor. This shows that our model is capable of capturing multivariate skewness in high-dimensional data. The average absolute bias of the upper semi-correlation is somewhat lower for the Frank and Normal copula, but the difference with that of the reflected Gumbel copula is relatively small, except for the model with one Gaussian factor with the Financials sector. Finally, we also modelled the data using a multivariate Student \( t_\nu \) copula with a three-factor covariance structure. However, the estimated degrees of freedom parameter is close to 20, which makes it indistinguishable from a normal copula model.

### Table 3: Sector-wise average absolute lower semi-correlation (\( \rho_{0} \)) and upper semi-correlation (\( \rho_{0} \)) using three different sectors in the S&P500

| Sector       | \( |\rho_{0}| \) | \( \rho_{0} \) |
|--------------|----------------|---------------|
| Industrials  | 0.38           | 0.25          |
| Information Technology | 0.39          | 0.21          |
| Financials   | 0.44           | 0.34          |

### Table 4: Estimated copula parameters for three different sectors in the S&P500

| Sectors         | r-Gumbel | Clayton | Frank | Normal | \( d \) |
|-----------------|----------|---------|-------|--------|--------|
| Industrials     | 1.46     | 0.73    | 3.51  | 0.50   | 71     |
| Information Technology | 1.50     | 0.79    | 3.75  | 0.52   | 74     |
| Financials      | 1.74     | 1.12    | 4.98  | 0.63   | 64     |
| Industrials-IIT | 1.35     | 0.56    | 2.86  | 0.43   | 145    |
| Industrials-Financials | 1.49     | 0.77    | 3.61  | 0.51   | 135    |
| IT-Financials   | 1.46     | 0.73    | 3.48  | 0.50   | 138    |
| Industrials-IIT-Financials | 1.61     | 0.91    | 4.32  | 0.57   | 209    |

Note: Top: one-factor model, middle: two-factor model, bottom: three-factor model.

---

6 Adding a second Gaussian factor to the model does not drastically change the parameter estimates.
Applying Gaussian factor models to high-dimensional data requires extensive computational power. Moreover, Gaussian factor models do not capture multivariate skewness in the data. In this article, we provided a method that exploits the high dimensionality of the data to accurately estimate the unobserved factors and simultaneously allows us to model multivariate skewness in the data. The proposed model is flexible, as it allows us to separately model the margins from the dependence structure. We proved that our model works for a factor model with up to three Gaussian factors. Simulation studies showed that the accuracy of our estimations improves as the sample size and dimension increase. To illustrate the model, we applied it to a dataset of three constituent sectors of the S&P500 stock data and found evidence that the stocks in the three sectors are skewed, consistent with findings in the finance literature.

There are different avenues for future research. One direction is to include fast inference models that allow for tail dependence in factor models or for a nested structure in the factor copula dependence. Another option is to include multiple variables that capture dependence between the data using different one-factor copula models.

Acknowledgments: The authors would like to thank the referee and Associate Editor for their constructive comments that have helped to improve this paper.

Funding information: This research was supported by the Andrew Sisson Grant and the Melbourne Research Scholarship from the University of Melbourne.

Conflict of interest: The authors state no conflict of interest.
Appendix

A Validity of Condition 3.1(iii)

We can illustrate Condition 3.1(iii) in the case where the factor loadings are uniformly spread out in the interval \([a_k, b_k]\), with \(-1 < a_k < b_k < 1\),

\[
\lambda_{lk} = a_k + \frac{\sigma_l(l)}{d}(b_k - a_k),
\]

where \(\sigma_k(l)\) is a permutation of indexes \(\{1, 2, \ldots, d\}\). Recall that \(\delta_{lk} = \lambda_{lk}/y_l\).

We first consider the case \(p = 1\) and we can assume without loss of generality \(\sigma_1(j) = j\) so that

\[
\delta_{l1} = \frac{a_1 + (1/d)(b_1 - a_1)}{[1 - \{a_1 + (1/d)(b_1 - a_1)\}^2]^{1/2}}.
\]

Let \(g_k(x) = a_k + x(b_k - a_k)\). We find that

\[
\lim_{d \to \infty} \|\Delta_k\|^2 = \lim_{d \to \infty} \frac{1}{d^2} \sum_{l=1}^d \sum_{m=1}^d (\delta_{l1} - \delta_{m1})^2
= 2 \lim_{d \to \infty} \frac{1}{d^2} \sum_{l=1}^d \delta_{l1}^2 - 2 \left( \lim_{d \to \infty} \frac{1}{d} \sum_{l=1}^d \delta_{l1} \right)^2
= 2 \int_0^1 \frac{g_k(x)^2}{1 - g_k(x)^2} \, dx - 2 \left( \int_0^1 \frac{g_k(x)}{\{1 - g_k(x)^2\}^{1/2}} \, dx \right)^2.
\]

(6.1)

We illustrate the validity of Condition 3.1(iii) in Tables A1 and A2 by plugging a range of values for \(a_1\) and \(b_1\) into (6.1).

The limiting norm of the vector \(\Delta_k\) is smaller if \(a_1\) and \(b_1\) are close. In practical applications, it indicates that more accurate parameter estimates can be obtained if the values of the loadings \(\Lambda_{lk}\) are more spread.

Now we consider the case \(p = 2\) with

\[
\delta_{lk} = \frac{a_k + (\sigma_k(l)/d)(b_k - a_k)}{[1 - \{a_1 + (\sigma_k(l)/d)(b_1 - a_1)\}^2 - \{a_2 + (\sigma_k(l)/d)(b_2 - a_2)\}^2]^{1/2}}, \quad k = 1, 2.
\]

Furthermore, we consider the case when the loadings \(\Lambda_1\) and \(\Lambda_2\) are monotonically dependent with \(\sigma_1(l) = \sigma_2(l)\) (case 1), and when they are independent (case 2).

In the first case, for \(k = 1, 2\),

\[
\lim_{d \to \infty} \|\Delta_k\|^2 = 2 \int_0^1 \frac{g_k(x)^2}{1 - g_k(x)^2 - g_k(x)^2} \, dx - 2 \left( \int_0^1 \frac{g_k(x)}{\{1 - g_k(x)^2 - g_k(x)^2\}^{1/2}} \, dx \right)^2,
\]

and in the second case, for \(k = 1, 2\),

\[
\lim_{d \to \infty} \|\Delta_k\|^2 = 2 \int_0^1 \frac{g_k(x_k)^2}{1 - g_k(x_k)^2 - g_k(x_k)^2} \, dx_k \, dx_k - 2 \left( \int_0^1 \frac{g_k(x_k)}{\{1 - g_k(x_k)^2 - g_k(x_k)^2\}^{1/2}} \, dx_k \, dx_k \right)^2.
\]

\[
\begin{array}{cccccccc}
\hline
a_1 & -0.99 & -0.80 & -0.50 & 0.00 & 0.50 & 0.80 \\
b_1 & 0.99 & 0.99 & 0.99 & 0.99 & 0.99 & 0.99 \\
\lim_{d \to \infty} \|\Delta_k\|^2 & 3.347 & 2.053 & 1.816 & 1.841 & 2.183 & 2.627 \\
\hline
\end{array}
\]

Table A1: Values for the integral in (6.1) for different values of \((a_1, b_1)\).
Furthermore, we need to check if \( \cos^2(\phi_{1,2}) = \langle \Delta_1, \Delta_2 \rangle^2 / (\|\Delta_1\|^2 \|\Delta_2\|^2) \) is bounded away from one as \( d \to \infty \) where \( \langle \Delta_i, \Delta_j \rangle \) is the inner product. We find

\[
\langle \Delta_1, \Delta_2 \rangle = \frac{1}{d^2} \left\{ \sum_{i=1}^{d} \sum_{m=1}^{d} (\delta_{i1} - \delta_{i2})(\delta_{m1} - \delta_{m2}) \right\} = 2 \frac{1}{d^2} \sum_{i=1}^{d} \delta_{i1} \delta_{i2} - 2 \left( \frac{1}{d^2} \sum_{i=1}^{d} \delta_{i1} \right) \left( \frac{1}{d^2} \sum_{i=1}^{d} \delta_{i2} \right).
\]

It implies that, in case 1,

\[
\lim_{d \to \infty} \langle \Delta_1, \Delta_2 \rangle = \frac{2}{\int \frac{1}{1-g(x)^2} \frac{g(x)}{g(x)^2} dx} - 2 \prod_{k=1}^{\infty} \left\{ \frac{1}{1-g(x)^2} \frac{g(x)}{g(x)^2} \right\} dx \frac{1}{\int \frac{1}{1-g(x)^2} \frac{g(x)}{g(x)^2} dx}.
\]

and in case 2,

\[
\lim_{d \to \infty} \langle \Delta_1, \Delta_2 \rangle = 2 \int \frac{1}{1-g(x)^2} \frac{g(x)}{g(x)^2} dx - 2 \prod_{k=1}^{\infty} \left\{ \frac{1}{1-g(x)^2} \frac{g(x)}{g(x)^2} \right\} dx \frac{1}{\int \frac{1}{1-g(x)^2} \frac{g(x)}{g(x)^2} dx}.
\]

Tables A3 and A4 show the limiting values for some pairs \((a_1, b_1)\) and \((a_2, b_2)\) for cases 1 and 2, respectively.

Note that in case 1, \( \cos^2(\phi_{1,2}) \) is very close to one, so the convergence of the parameter estimates to their true values may be very slow. In fact, \( \Delta_1 \) and \( \Delta_2 \) are linearly dependent with \( \Delta_1 = \Delta_2 \) when \((a_1, b_1) = (a_2, b_2) = (0.2, 0.6) \) and \( \Delta_2 = 0.5 \Delta_1 \) with \((a_1, b_1) = (0.2, 0.4) \) and \((a_2, b_2) = (0.4, 0.8) \). Condition 3.1(iii) fails in this case.

In case 2, \( \cos^2(\phi_{1,2}) \) is far from one in all cases, and accurate parameter estimates can be obtained in this case.

Similar results can be obtained when \( p = 3 \), so they are not reported here.

**Table A2:** Values for the integral in (6.1) for different values of \((a_1, b_1)\)

| \((a_1, b_1)\) | \(-0.20, -0.15\) | \(-0.10, 0.00\) | \(-0.10, 0.10\) | \(-0.10, 0.15\) |
|-----------------|-----------------|-----------------|-----------------|-----------------|
| \(\lim_{d \to \infty} \|\Delta_1\|^2\) | 0.0273 | 0.0208 | 0.0153 | 0.0069 | 0.0018 | 0.0005 |

**Table A3:** Limiting values of \(\|\Delta_k\|^2\), \(k = 1, 2\), and \(\cos^2(\phi_{1,2})\) for different values of \((a_1, b_1)\) and \((a_2, b_2)\); case 1

| \((a_1, b_1)\) | \((-0.5, 0.6)\) | \((0.2, 0.6)\) | \((0.2, 0.6)\) | \((0.2, 0.4)\) | \((0.2, 0.4)\) |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| \(\lim_{d \to \infty} \|\Delta_k\|^2\) | 0.427 | 0.119 | 0.067 | 0.219 | 0.061 |
| \(\lim_{d \to \infty} \|\Delta_1\|^2\) | 0.427 | 0.119 | 0.051 | 1.359 | 0.243 |
| \(\lim_{d \to \infty} \cos^2(\phi_{1,2})\) | 0.997 | 1.000 | 0.999 | 0.997 | 1.000 |

**Table A4:** Limiting values of \(\|\Delta_k\|^2\), \(k = 1, 2\), and \(\cos^2(\phi_{1,2})\) for different values of \((a_1, b_1)\) and \((a_2, b_2)\); case 2

| \((a_1, b_1)\) | \((-0.5, 0.6)\) | \((0.2, 0.6)\) | \((0.2, 0.6)\) | \((0.2, 0.4)\) | \((0.2, 0.4)\) |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| \(\lim_{d \to \infty} \|\Delta_k\|^2\) | 0.319 | 0.075 | 0.054 | 0.071 | 0.028 |
| \(\lim_{d \to \infty} \|\Delta_1\|^2\) | 0.319 | 0.075 | 0.040 | 0.796 | 0.180 |
| \(\lim_{d \to \infty} \cos^2(\phi_{1,2})\) | 0.004 | 0.189 | 0.041 | 0.703 | 0.416 |
B Proofs

In this section, proofs are provided for theoretical results presented in Section 3.

Proof of Theorem 3.2.

(i) In order for the parameter vector \((\Lambda_0, \eta_0)\) to be unique, one has to impose \(p(p - 1)/2\) restrictions on the factor loadings \(\Lambda_0\). Otherwise, one can always find a rotation \(\Lambda_{\text{rot}} = \Lambda A\), with \(A\) a \(p \times p\) orthogonal matrix, such that the variance-covariance matrix remains the same.

(ii) This is trivially satisfied, as the values of the correlation matrix by definition are restricted to the interval \([-1, 1]\).

(iii) This is trivially satisfied, as the correlation matrix is continuous at each \((\Lambda, \eta)\) in the parameter space.

(iv) Since the correlation parameter \(\sigma_{ij} = \text{Corr}(X_i, X_j)\) is bounded by the interval \([-1, 1]\), one can show that

\[
\mathbb{E} \left[ \sup_{(\Lambda, \eta) \in \Theta} \|f(X, \Lambda, \eta)\| \right] \leq d.
\]

□

Proof of Theorem 3.3.

(i) The GMM estimator is consistent if the conditions from Theorem 3.2 are satisfied.

(ii) This condition is satisfied, as the parameter space for the candidate loadings \(\lambda_{jk}\) and correlation \(\eta\), is restricted to the interval \([-1, 1]\).

(iii) We know that \(v_{(\Lambda, \eta)}f(X_i, \Lambda, \eta) = v_{(\Lambda, \eta)}\Sigma_{\xi, \eta}\). So for every element \(\sigma_{ij} = \sum_{k=1}^{p} \lambda_{jk} \lambda_{kj} + \eta \gamma_{ij}\), we have

\[
\frac{\partial}{\partial \lambda_{jk}} \sigma_{ij} = \gamma_{i} \delta_{jk} - \eta \delta_{jk}, \quad \frac{\partial}{\partial \eta} \sigma_{ij} = \eta \gamma_{ij},
\]

The derivatives exist and are finite under Condition 3.1(i).

(iv) It is easy to see that \(\mathbb{E}[\|f(X_i, \Lambda, \eta)\|^2] \leq d^2\).

(v) Note that

\[
\left( \frac{\partial}{\partial \lambda_{jk}} \sigma_{ij} \right)^2 \leq 2\delta^2_{ik} + 2\delta^2_{jk}, \quad \left( \frac{\partial}{\partial \lambda_{jk}} \sigma_{ij} \right)^2 \leq 2\delta^2_{ik} + 2\delta^2_{jk}, \quad \left( \frac{\partial}{\partial \eta} \sigma_{ij} \right)^2 \leq 1.
\]

Under Condition 3.1(i),

\[
\sup_{(\Lambda, \eta) \in \Theta} \| v_{(\Lambda, \eta)}f(X_i, \Lambda, \eta) \|^2 \leq 8 \sum_{k=1}^{d} \sum_{i=1}^{d} \delta_{ik}^2 + pd = pd(8\xi + 1).
\]

(vi) To check this condition, we have to show that the matrix

\[ G^T G = (v_{(\Lambda, \eta)}\Sigma_{\xi, \eta})^T v_{(\Lambda, \eta)}\Sigma_{\xi, \eta} \]

is invertible. The condition is illustrated for a model with one Gaussian factor, and the proof is similar in the general case. The model with one factor has the following derivatives:

\[
\frac{\partial \sigma_{ij}}{\partial \lambda_i} = \lambda_j - \eta \frac{Y_j}{Y_i} \lambda_i, \quad \frac{\partial \sigma_{ij}}{\partial \lambda_j} = \lambda_i - \eta \frac{Y_i}{Y_j} \lambda_j, \quad \frac{\partial \sigma_{ij}}{\partial \eta} = \gamma_{ij},
\]

for \(i, j = 1, \ldots, d\) and \(i \neq j\). Then it suffices to show that the following does not hold

\[
\alpha (\lambda_j - \eta \frac{Y_j}{Y_i} \lambda_i) + \alpha (\lambda_j - \eta \frac{Y_i}{Y_j} \lambda_j) = \gamma_{ij},
\]

\[
\delta_{ij} (\delta_{i} - \eta \delta_{i}) + \delta_{ij} (\delta_{j} - \eta \delta_{j}) = 1,
\]
where \( \tilde{a}_i = a_i / y_i \) and \( \delta_j = \lambda_j / y_j \). Then for \( i \neq j \), and \( i \neq k \) we have the following system of equations:

\[
\begin{cases}
\tilde{a}_i (\delta_j - \eta \delta_i) + \tilde{a}_j (\delta_i - \eta \delta_j) = 1 \\
\tilde{a}_i (\delta_k - \eta \delta_i) + \tilde{a}_k (\delta_i - \eta \delta_k) = 1,
\end{cases}
\]

subtracting the two equations gives another set of equations, where \( j \neq l \) and \( k \neq l \)

\[
\begin{cases}
\tilde{a}_i (\delta_j - \delta_k) + \delta_j (\tilde{a}_i - \tilde{a}_k) + \eta (\tilde{a}_k \delta_j - \tilde{a}_i \delta_k) = 0 \\
\tilde{a}_i (\delta_j - \delta_k) + \delta_j (\tilde{a}_i - \tilde{a}_k) + \eta (\tilde{a}_k \delta_j - \tilde{a}_i \delta_k) = 0,
\end{cases}
\]

and again subtracting the aforementioned equations gives

\[
(\tilde{a}_i - \tilde{a}_l)(\delta_j - \delta_l) = -(\tilde{a}_i - \tilde{a}_k)(\delta_j - \delta_k).
\] (A1)

Hence, there are two cases. In the first case, \( \tilde{a}_i - \tilde{a}_l = 0 \), which implies that \( \delta_j - \delta_l = 0 \). Otherwise, all differences \( \tilde{a}_i - \tilde{a}_l \) and \( \delta_j - \delta_k \) would be zero, for every possible combination of the pairs \(( i, l)\) and \(( j, k)\).

In the second case, \( \delta_i - \delta_l \) and \( \delta_j - \delta_k \) are different from zero, and we can rewrite the display in (A1) as

\[
\frac{\tilde{a}_i - \tilde{a}_l}{\delta_i - \delta_l} = \frac{\tilde{a}_i - \tilde{a}_k}{\delta_j - \delta_k}.
\]

The numerator and denominator on both sides of the above display are nonzero, even though the ratios are the same. However, looping through every possible combination of the pairs \(( i, l)\) and \(( j, k)\), the only valid result is that all differences \( \tilde{a}_i - \tilde{a}_l \) are zero, which implies that all \( \delta_l \)s are equal. However, this contradicts Condition 3.1(iii), which states that there can only be up to \( O(d^2) \) terms that are zero, which concludes the proof.

\[\square\]

Proof of Lemma 3.4. The optimal weight vector \( w^*_k \) can be derived by differentiating

\[
\E[(\tilde{Z}_k - Z_k)^2] = w_k^T \Sigma w_k - 2w_k^T \Lambda_k + 1
\] (A2)

with respect to \( w_k \) for each \( k = 1, \ldots, p \). Recall that \( \E(\tilde{Z}_k) = \E(Z_k) = 0 \). Then assuming that \( \Sigma \) is invertible, one can obtain the optimal weight vector \( w^*_k \) by optimising (A2) with respect to \( w_k \) subject to constraint

\[
\Var(\tilde{Z}_k) = w^*_k \Sigma w_k = 1.
\]

\[\square\]

Proof of Theorem 3.5. After substituting \( w^*_k = (\Lambda_k^T \Sigma_k^1 \Lambda_k)^{-1/2} \Sigma_k^1 \Lambda_k \) back into (A2), one obtains

\[
\lim_{d \to \infty} \E[(\tilde{Z}_k - Z_k)^2] = 2 - 2 \lim_{d \to \infty} (\Lambda_k^T \Sigma_k^1 \Lambda_k)^{1/2}.
\]

Therefore, we need to show that \( \lim_{d \to \infty} \Lambda_k^T \Sigma_k^1 \Lambda_k = 1 \). Using Woodbury’s formula, one can show that

\[
\Lambda_k^T \Sigma_k^1 \Lambda_k = \Lambda_k^T (\Gamma^{1/2} \Sigma_k^{1/2})^{-1} \Lambda_k = \Lambda_k^T \left( \Gamma^{1/2} \Sigma_k^{1/2} \right)^{-1} \Lambda_k = \Lambda_k^T \left( \Gamma^{1/2} \Sigma_k^{1/2} \right)^{-1} \Lambda_k.
\]

Without loss of generality, convergence of the factor estimates is illustrated for \( k = 1 \). However, the results hold for any \( k = 1, \ldots, p \). Furthermore, define \( b_{qr} = \Lambda_k^T (\Gamma^{1/2} \Sigma_k^{1/2})^{-1} \Lambda_k \), and \( a_k = b_{kk} \) for \( k, r = 1, \ldots, p \). We find

\[
a_k = \frac{1 + (d - 2)\eta}{1 + (d - 2)\eta - (d - 1)\eta^2} \sum_{l=1}^{d} \delta_{lk}^2 - \frac{\eta}{1 + (d - 2)\eta - (d - 1)\eta^2} \sum_{l=1}^{d} \delta_{lk} \delta_{mk},
\]

\[
b_{qr} = \frac{1 + (d - 2)\eta}{1 + (d - 2)\eta - (d - 1)\eta^2} \sum_{l=1}^{d} \delta_{ql} \delta_{lr} - \frac{\eta}{1 + (d - 2)\eta - (d - 1)\eta^2} \sum_{l=1}^{d} \delta_{ql} \delta_{mr},
\]

where \( \delta_{lk} = \lambda_{lk} / y_k \). Then \( a_i = \Lambda_i^T (\Gamma^{1/2} \Sigma_i^{1/2})^{-1} \Lambda_i \) is a scalar.
Proof for a one-factor model ($p = 1$, and $k = 1$). In the case where $p = 1$, it follows that $\Lambda = \Lambda_1$. Hence,

$$
\Lambda_1^T \Sigma X^T \Lambda_1 = \frac{a_1}{1 + a_1}.
$$

It remains to be shown that $a_1 \to \infty$ as $d \to \infty$. If Conditions 3.1(i) and 3.1(ii) are satisfied, then

$$
a_1 / d = \frac{1}{1 - \eta} \left( \frac{1}{d} \sum_{i=1}^{d} \delta_i - \frac{1}{d^2} \left( \sum_{i=1}^{d} \sum_{m=1}^{d} \delta_i \delta_m \right) \right) + O(1/d)
$$

$$
= \frac{1}{1 - \eta} \left( \frac{1}{d} \sum_{i=1}^{d} \delta_i^2 - \frac{1}{d^2} \left( \sum_{i=1}^{d} \delta_i \right)^2 \right) + O(1/d)
$$

$$
= \frac{1}{1 - \eta} \frac{1}{2d^2} \sum_{i=1}^{d} \left( \delta_i - \delta_m \right)^2 + O(1/d) = \frac{0.5 || \Delta ||^2}{1 - \eta} + O(1/d)
$$

$$
> \frac{0.5 \xi_0^2}{1 - \eta} + O(1/d),
$$

where Condition 3.1(iii) is used in the last step. It follows that $a_1 / d \to \infty$ as $d \to \infty$.

Proof for a two-factor model ($p = 2$, and $k = 1$).

In the case of a two-factor model,

$$
\Lambda (\Gamma^{1/2} H \Gamma^{1/2})^{-1} \Lambda + I_2 = \begin{pmatrix} 1 + a_1 & b_{12} \\ b_{12} & 1 + a_2 \end{pmatrix}.
$$

Computing the inverse, we obtain

$$
(\Lambda (\Gamma^{1/2} H \Gamma^{1/2})^{-1} \Lambda + I_2)^{-1} = \frac{1}{(1 + a_1)(1 + a_2) - b_{12}^2} \begin{pmatrix} 1 + a_2 & -b_{12} \\ -b_{12} & 1 + a_1 \end{pmatrix}.
$$

Finally,

$$
\Lambda_1^T (\Gamma^{1/2} H \Gamma^{1/2})^{-1} \Lambda = (a_1 \ b_{12}).
$$

Putting everything together, we obtain

$$
\Lambda_1^T \Sigma X^T \Lambda_1 = \frac{a_1(a_1 + a_2) - b_{12}^2}{(1 + a_1)(1 + a_2) - b_{12}^2} = 1 - \frac{1}{1 + a_1} \left( 1 - \frac{b_{12}^2}{(1 + a_1)(1 + a_2)} \right)^{-1}.
$$

It is required to show that $a_{12} = 1 - b_{12}^2 / [(1 + a_1)(1 + a_2)]$ is bounded away from zero. To start with,

$$
b_{12} / d = \frac{1}{1 - \eta} \left( \frac{1}{d} \sum_{i=1}^{d} \delta_i \delta_2 - \frac{1}{d^2} \right) + O(1/d)
$$

$$
= \frac{1}{1 - \eta} \left( \frac{1}{d} \sum_{i=1}^{d} \delta_i \delta_2 - \left( \frac{1}{d} \sum_{i=1}^{d} \delta_i \right) \left( \frac{1}{d} \sum_{m=1}^{d} \delta_m \right) \right) + O(1/d)
$$

$$
= \frac{0.5 || \Delta ||^2}{1 - \eta} + O(1/d) = \frac{0.5 || \Delta || || \Delta || \cos(\varphi_{12})}{1 - \eta} + O(1/d).
$$

Then Condition 3.1(iii) implies that $a_{12} = 1 - \cos^2(\varphi_{12}) + O(1/d) = \sin^2(\varphi_{12}) + O(1/d) > \sin^2(\xi_0/2) + O(1/d)$.

Proof for a three-factor model ($p = 3$, and $k = 1$). For a three-factor model, we obtain the following expression:

$$
\Lambda (\Gamma^{1/2} H \Gamma^{1/2})^{-1} \Lambda + I_p = \begin{pmatrix} 1 + a_1 & b_{12} & b_{13} \\ b_{12} & 1 + a_2 & b_{23} \\ b_{13} & b_{23} & 1 + a_3 \end{pmatrix}.
$$
After some calculations, we find

\[
\lambda^T \Sigma \lambda = 1 - \frac{b_{12}^2}{a_1 + 1} \left( 1 - \frac{b_{12}^2}{(a_1 + a_2)(a_1 + a_3)} \right),
\]

where

\[
b_{12} = 1 + 2 \cos(\phi_{1,2}) \cos(\phi_{1,3}) = 1 + 2 \cos(\phi_{1,2}) \cos(\phi_{1,3}) - \cos^2(\phi_{1,2}) - \cos^2(\phi_{1,3}) + O(1/d)
\]

\[
= 4 \sin(0.5(\phi_{1,2} + \phi_{1,3} - \phi_{2,3})) \sin(0.5(\phi_{1,2} + \phi_{2,3} - \phi_{1,3})) \sin(0.5(\phi_{1,2} + \phi_{2,3} - \phi_{1,3}))
\]

by Condition 3.1(iii). It implies that \(\lambda^T \Sigma \lambda \geq 1 - \sin^4(\xi_0/2)/(a_1 + 1) + O(1/d) \to 1\) as \(d \to \infty\).

Proof of Theorem 3.6. Using the properties of the GMM estimator, we know that \(\hat{\lambda}_k \to \lambda_k\) and \(\hat{\Sigma}_k \to \Sigma_k\). Under Condition 3.1(ii) \(\Sigma_k\) is invertible. Thus, by the continuous mapping theorem (CMT) it follows that

\[
\hat{\Sigma}_k^{-1} \to \Sigma_k^{-1}.
\]

As a result \(\hat{w}_{knd} = (\hat{\lambda}_k \Sigma_k^{-1/2} \hat{\lambda}_k^{-1})^{-1/2} \hat{\Sigma}_k^{-1} \hat{w}_{knd} = (\lambda_k^{-1} \Sigma^{-1} \lambda_k^{-1})^{-1/2} \Sigma_k^{-1} \hat{w}_{knd}\). By expanding, we obtain

\[
\lim_{d \to \infty} \lim_{n \to \infty} E((Z_{knd}^m - Z_k)^2) = \lim_{d \to \infty} \lim_{n \to \infty} E((X^T \hat{w}_{knd}^m)^2) = \lim_{d \to \infty} \lim_{n \to \infty} E((X^T \hat{w}_{knd})^2) + 1.
\]

The first and second terms in (A4) can be rewritten as

\[
E((X^T \hat{w}_{knd}^m)^2) = E((X^T \hat{w}_{knd}^T + \hat{w}_{knd}) X^T (\hat{w}_{knd}^T - \hat{w}_{knd})),
\]

\[
E((\hat{w}_{knd}^T X Z_k) = E((\hat{w}_{knd}^T - \hat{w}_{knd}) X Z_k) + E(\hat{w}_{knd}^T X Z_k),
\]

respectively. From Theorem 3.5, we know that

\[
\lim_{d \to \infty} \lim_{n \to \infty} E((X^T \hat{w}_{knd}^m)^2) = \lim_{d \to \infty} \lim_{n \to \infty} E((X^T \hat{w}_{knd})^2) = 0.
\]

Thus, it remains to be shown that the first term in (A5) and in (A6) converge to zero as \(n \to \infty\). Using Cauchy’s inequality, the first term in (A5) is bounded by

\[
E((X^T (\hat{w}_{knd} + \hat{w}_{knd}^*)^2) \leq E((X^T X)^2) \leq E((X^T (\hat{w}_{knd} + \hat{w}_{knd}^*)^T (\hat{w}_{knd} + \hat{w}_{knd}^*)^2)\leq E((X^T X)^2) \leq E((X^T (\hat{w}_{knd} + \hat{w}_{knd}^*)^T (\hat{w}_{knd} + \hat{w}_{knd}^*)^2)\leq E((X^T X)^2) \leq E((X^T (\hat{w}_{knd} + \hat{w}_{knd}^*)^T (\hat{w}_{knd} + \hat{w}_{knd}^*)^2)\leq E((X^T X)^2) \leq E((X^T (\hat{w}_{knd} + \hat{w}_{knd}^*)^T (\hat{w}_{knd} + \hat{w}_{knd}^*)^2)\leq E((X^T X)^2) \leq E((X^T (\hat{w}_{knd} + \hat{w}_{knd}^*)^T (\hat{w}_{knd} + \hat{w}_{knd}^*)^2)\leq E((X^T X)^2) \leq E((X^T (\hat{w}_{knd} + \hat{w}_{knd}^*)^T (\hat{w}_{knd} + \hat{w}_{knd}^*)^2)\leq E((X^T X)^2) \leq E((X^T (\hat{w}_{knd} + \hat{w}_{knd}^*)^T (\hat{w}_{knd} + \hat{w}_{knd}^*)^2)\leq E((X^T X)^2) \leq E((X^T (\hat{w}_{knd} + \hat{w}_{knd}^*)^T (\hat{w}_{knd} + \hat{w}_{knd}^*)^2).
\]

Note that \(E((X^T X)^2) = 3d\) and \(\hat{w}_{knd}^T \hat{\Sigma} \hat{w}_{knd} = 1\). From Condition 3.1(ii) we can conclude that \(\hat{w}_{knd}^* \to \hat{w}_{knd}\) and \(\hat{w}_{knd}^* \to \hat{w}_{knd}\) are bounded.

Since we know that \(\hat{w}_{knd}^* \to \hat{w}_{knd}^*\) and \((\hat{w}_{knd}^T - \hat{w}_{knd}^*)^T\) is uniformly integrable, we conclude that \(E((\hat{w}_{knd}^T - \hat{w}_{knd}^*)^T (\hat{w}_{knd}^T - \hat{w}_{knd}^*)^2) \to 0\) as \(n \to \infty\) by Vitali’s convergence theorem. It implies that \(E[X^T (\hat{w}_{knd} + \hat{w}_{knd}^*)^T (\hat{w}_{knd} - \hat{w}_{knd}^*)^T (\hat{w}_{knd} - \hat{w}_{knd}^*)^T] \to 0\) as \(n \to \infty\). Similarly, \(E((\hat{w}_{knd}^T - \hat{w}_{knd}^*)^T X Z_k) \to 0\) as \(n \to \infty\), which concludes the proof.

Proof of Lemma 3.7. This follows immediately from the fact that the GMM estimators converge in probability combined with the CMT, and from Theorem 3.6.

Proof of Theorem 3.8. From Lemma 3.7, we know that \(\hat{U}_{k,j} \to \hat{U}_{k,j}\) as \(n \to \infty\). Moreover, from Theorem 3.5, it follows that \(\hat{Z}_{k,j} \to Z_k\), for each \(k = 1, \ldots, p\). Rewriting, we obtain

\[
\hat{e}_{j} = \frac{1}{f_j} \left( X_j - \sum_{k=1}^{p} \lambda_k \hat{Z}_{k,j}^* \right) = e_j - \frac{1}{f_j} \sum_{k=1}^{p} \lambda_k (\hat{Z}_{k,j}^* - Z_k).
\]
and
\[
\tilde{U}_d = \tilde{e}_d - \frac{1}{d} \sum_{i=1}^{d} \sum_{k=1}^{p} \delta_{ik}(\tilde{Z}_{kd} - Z_k),
\]
where \(\tilde{e}_d = (1/d)\sum_{i=1}^{d} e_{id} \overset{p}{\to} m(V)\) by Theorem 5 in [9] and \(E[(\tilde{Z}_{kd} - Z_k)^2] < \xi^*/d\) for some constant \(\xi^* > 0\) as follows from the proof of Theorem 3.5 so that
\[
E\left[ \left( \frac{1}{d} \sum_{k=1}^{p} \sum_{j=1}^{d} \delta_{jk}(\tilde{Z}_{kd} - Z_k) \right)^2 \right] \leq \frac{\xi^* p}{d} \sum_{k=1}^{p} \left( \frac{1}{d} \sum_{j=1}^{d} \delta_{jk} \right)^2 \to 0 \quad \text{as } d \to \infty.
\]
It implies that \(\tilde{U}_d \overset{p}{\to} m(V)\), as \(n \to \infty\) and \(d \to \infty\).

**Proof of Theorem 3.9.** Let \(\tilde{y}_1 = (1 - y_1^2)^{1/2}\) and \(\tilde{y}_m = (1 - y_m^2)^{1/2}\). Denote \(W_i = \tilde{y}_1 \sum_{k=1}^{p} \lambda_{ik} Z_k\) and \(W_m = \tilde{y}_m \sum_{k=1}^{p} \lambda_{mk} Z_k\). Note that \((W_i, W_m)\) have the bivariate standard normal distribution. Let \(\phi_{l,m}\) be the joint density of \((W_i, W_m)\) and \(F_{l,m}\) be the joint cdf of \((\varepsilon_i, \varepsilon_m)\). We find
\[
\Pr(X_i < z, X_m < z) = \int_{\mathbb{R}^2} \phi_{l,m}(w_i, w_m) \frac{w_i}{y_i} \frac{w_m}{y_m} d\omega_i d\omega_m
\]
\[
\leq \int_{\mathbb{R}^2} \min \left( \phi\left( \frac{z - \tilde{y}_i W_i}{y_i} \right), \phi\left( \frac{z - \tilde{y}_m W_m}{y_m} \right) \right) \phi_{l,m}(w_i, w_m) d\omega_i d\omega_m
\]
\[
= \Pr(\tilde{y}_i W_i + y_i V < z, \tilde{y}_m W_m + y_m V < z) = \Phi_{\rho_{l,m}}(z, z),
\]
where \(\Phi_{\rho_{l,m}}\) is the cdf of a standard bivariate normal distribution with the correlation \(\rho_{l,m} = \tilde{y}_l \tilde{y}_m \text{Cor}(W_i, W_m) + y_l y_m < 1\), so that
\[
\lim_{z \to -\infty} \frac{\Pr(X_i < z, X_m < z)}{\Phi(z)} = \lim_{z \to -\infty} \frac{\Phi_{\rho_{l,m}}(z, z)}{\Phi(z)} = 0.
\]

**References**

[1] Bai, J., & Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1), 191–221.

[2] Bai, J., & Ng, S. (2008). Large dimensional factor analysis. *Foundations and Trends in Econometrics*, 3(2), 89–163.

[3] Creal, D. D., & Tsay, R. S. (2015). High dimensional dynamic stochastic copula models. *Journal of Econometrics*, 189(2), 335–345.

[4] Johnson, R. A., & Wichern, D. W. (2002). *Applied multivariate statistical analysis*. Volume 5. Upper Saddle River, NJ: Prentice Hall.

[5] Kent, J., Bibby, J., & Mardia, K. (1979). *Multivariate analysis*. Amsterdam: Academic Press.

[6] Kreuzer, A., & Czado, C. (2021). Bayesian inference for a single factor copula stochastic volatility model using Hamiltonian Monte Carlo. *Econometrics and Statistics*, 19, 130–150.

[7] Krupskii, P., & Joe, H. (2013). Factor copula models for multivariate data. *Journal of Multivariate Analysis*, 120, 85–101.

[8] Krupskii, P., & Joe, H. (2015). Structured factor copula models: Theory, inference and computation. *Journal of Multivariate Analysis*, 138, 53–73.

[9] Krupskii, P., & Joe, H. (2021). Approximate likelihood with proxy variables for parameter estimation in high-dimensional factor copula models. *Statistical Papers*, 63, 1–27.

[10] McNeil, A., Frey, R., & Embrechts, P. (2005). *Quantitative risk management: Concepts, techniques and tools*. Princeton, New Jersey: Princeton University Press.

[11] Nguyen, H., Ausín, M. C., & Galeano, P. (2020). Variational inference for high dimensional structured factor copulas. *Computational Statistics and Data Analysis*, 151, 107012.

[12] Oh, D. H., & Patton, A. J. (2013). Simulated method of moments estimation for copula-based multivariate models. *Journal of the American Statistical Association*, 108(502), 689–700.
[13] Oh, D. H., & Patton, A. J. (2017). Modeling dependence in high dimensions with factor copulas. *Journal of Business and Economic Statistics, 35*(1), 139–154.

[14] Oh, D. H., & Patton, A. J. (2018). Time-varying systemic risk: Evidence from a dynamic copula model of CDS spreads. *Journal of Business and Economic Statistics, 36*(2), 181–195.

[15] Pesaran, M. H. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica, 74*(4), 967–1012.

[16] Schamberger, B., Gruber, L., & Czado, C. (2017). Bayesian inference for latent factor copulas and application to financial risk forecasting. *Econometrics, 5*(2), 21.

[17] Sklar, M. (1959). Fonctions de répartition àn dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris, 8*, 229–231.

[18] Stock, J. H., & Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association, 97*(460), 1167–1179.

[19] Subbarao, C., Subbarao, N., & Chandu, S. (1996). Characterization of groundwater contamination using factor analysis. *Environmental Geology, 28*(4), 175–180.

[20] Tan, B. K., Panagiotelis, A., & Athanasopoulos, G. (2019). Bayesian inference for the one-factor copula model. *Journal of Computational and Graphical Statistics, 28*(1), 155–173.

[21] Tsay, R. S. (2005). *Analysis of financial time series*, Volume 543. New Jersey: John Wiley & Sons.

[22] Van Loan, C. F., & Golub, G. (1996). *Matrix computations*. Maryland: Johns Hopkins Studies in Mathematical Sciences.