Journal of Multivariate Analysis 201 (2024) 105263

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

High-dimensional factor copula models with estimation of latent variables

Xinyao Fan [*], Harry Joe

Department of Statistics, University of British Columbia, Vancouver, Canada V6T 1Z4

ARTICLE INFO

AMS 2020 subject classifications:
primary 62H12
secondary 62F12

Keywords:
Bi-factor
Factor scores
Identifiability
Oblique factor
Proxy variables
Tail dependence

ABSTRACT

Factor models are a parsimonious way to explain the dependence of variables using several latent variables. In Gaussian 1-factor and structural factor models (such as bi-factor and oblique factor) and their factor copula counterparts, factor scores or proxies are defined as conditional expectations of latent variables given the observed variables. With mild assumptions, the proxies are consistent for corresponding latent variables as the sample size and the number of observed variables linked to each latent variable go to infinity. When the bivariate copulas linking observed variables to latent variables are not assumed in advance, sequential procedures are used for latent variables estimation, copula family selection and parameter estimation. The use of proxy variables for factor copulas means that approximate log-likelihoods can be used to estimate copula parameters with less computational effort for numerical integration.

1. Introduction

Factor models are flexible and parsimonious ways to explain the dependence of variables with one or more latent variables. The general factor copula models in [15,16] are extensions of classical Gaussian factor models and are useful for joint tail inference if the variables have stronger tail dependence that can be expected with Gaussian models, such as with asset return data.

In classical factor analysis, the estimates of the latent variables referred to factor scores (see [10,13]) are of interest and useful for interpretation and further regression and prediction analysis. The inference of the latent variables in factor copulas is helpful for selection of parametric copula families to link observed and latent variables, and lead to simpler parameter estimation procedures.

For maximum likelihood estimation in parametric factor copula models, the copula density and likelihood involve integrals with dimension equal to the number of latent variables. In [15,16], the authors provide procedures for computationally efficient evaluations of the log-likelihood and its gradient and Hessian for 1-factor, 2-factor, bi-factor and a special case of the oblique factor copula models. These are the cases for which integrals can be evaluated via 1-dimensional or 2-dimensional Gauss-Legendre quadrature. Bi-factor and oblique structured factor models are useful when the observed variables belong to several non-overlapping homogeneous groups. Factor copula models would mainly be considered if the observed variables are monotonically related and have at least moderate dependence. There may be more dependence in the joint tails than expected with Gaussian dependence, but the Gaussian factor models can be considered as first-order models.

In this paper, one main focus for some structured factor copula models is to show how the use of "proxies" to estimate latent variables (a) can help in diagnostic steps for deciding on the bivariate copula families that link observed variables to the latent variables and (b) lead to approximate log-likelihoods for which numerical maximum likelihood estimation is much faster. The 1-factor, bi-factor and oblique factor copula models are used to illustrate the theory because with the previous numerical

* Corresponding author.
E-mail address: xinyao.fan@stat.ubc.ca (X. Fan).

https://doi.org/10.1016/j.jmva.2023.105263
Received 30 May 2022; Received in revised form 13 July 2023; Accepted 23 November 2023
Available online 25 November 2023

implementations for maximum likelihood, we can make comparisons with the faster proxy-based methods introduced within. The theory developed here can be applied in other factor copula models, and this is discussed in the final section on further research.

The use of proxies for latent variables is initiated in [17] to speed up numerical maximum likelihood estimation; the approach involves unweighted means in 1-factor and unweighted group means for oblique factor copula models. The approach does not extend to bi-factor and other structural p-factor copula models. In order to accommodate these other factor copula models, we use two-stage proxies, with stage 1 being factor scores based on the estimated loading matrix after each variable has been empirically transformed to standard normal, and stage 2 based on conditional expectations of latent variables given the observed variables (using a copula model fitted from the stage 1 proxies). In [17], it is shown that the joint distribution of observed variables and these simpler proxy variables converges to the underlying joint distribution as the dimension goes to infinity. We present a stronger weak consistency result for the conditional expectation proxy variable as it also provides the convergence rate of the proxy variables in the 1-factor and oblique factor copula models. Moreover, our method can handle bi-factor models where proxies based on unweighted averages are not valid.

To justify the sequential method for latent variable and copula model estimation, several theoretical results in consistency, as the number D of observed variables increases to infinity, are needed. There is no general statistical theory for the number of parameters increasing to infinity; using the property that 1-factor, bi-factor and oblique factor models are closed under margins, we obtain asymptotic results with the assumption of sampling from super-populations.

The proxies as estimates of latent variables are extensions of Gaussian regression factor scores because these are based on conditional expectations of latent variables given the observed. For D increasing, we consider the observed variables (or their correlations, partial correlations, or linking copulas) as being sampled from appropriate super-populations. We first obtain conditions for the proxies or conditional expectations to be asymptotically consistent estimates (of corresponding latent variables) when the factor model is completely known; we also have results that suggest rates of convergence. In cases where factor model is known but consistency is not possible, then one cannot expect consistency when the parameters in the factor model must be estimated. One such case involves the Gaussian bi-factor model where (a) a loading matrix of less than full column rank implies that the latent variables are not identifiable, and the model can be reduced to an oblique factor model; and (b) the rate of convergence of the proxies is slow if a product matrix involving the loading matrix has a large condition number. With a sample of size N, the assumption of a super-population, combined with factor models being closed under margins, suggest that in the case of estimated parameters, (i) all parameters can be estimated in blocks with $\sqrt{N}$ consistency and (ii) proxies are consistent under mild conditions. The "closed-under-margin" property indicates that the same latent variables apply for different subsets of variables. Because we need a method of proof that is valid for both Gaussian and factor copula models, our technique is different from that of [3]. Their approach does not provide insights for non-identifiability of latent variables such as in the bi-factor model.

The remainder is organized as follows. Section 2 provides the representations of Gaussian factor models and (structured) factor copula models. Section 3 has the expressions for conditional expectations, and the proxies as estimates of the latent variables. Section 4 contains some sufficient or necessary conditions for asymptotic consistency of proxy variables with known loading matrix or known linking copulas. Section 5 has results and conditions for the consistency of proxy variables with estimated parameters in linking copulas (with copula families known). Section 6 proposes a sequential method for the practical use of proxy variables in cases where the linking copula families are not specified. Section 7 has sufficient conditions for using the proxies in Section 3 when observed variables have weak dependence, rather than independence, conditional on the observed variables. Section 8 has a summary and discussion for further research. The Supplementary material has simulation studies that show that the proxies are useful in selecting linking copula families and getting accurate parameter estimates with less computing time.

2. Structured factor copula models

The 1-factor, bi-factor and oblique factor models that are the main focus are shown graphically in Figures 1, 2, 3 in the Supplementary material. The graphical representations are valid for the Gaussian factor models and their extensions to factor copula models.

Gaussian p-factor models with $p \geq 2$ in general do not have an identifiable loading matrix because of the orthogonal transform of the loading matrix. The bi-factor structure is a special of the p-factor model with many structural zeros. The bi-factor and oblique factor models are two parsimonious factor models that can be considered when variables can be divided into G non-overlapping groups.

In the graphs for 1-factor and bi-factor, each observed variable links to the (global) latent variable; the edges of the graphs have a correlation (of observed with latent) for multivariate Gaussian and bivariate linking copula for the factor copula.

For the bi-factor graph, there are additional edges linking each observed variable to its (local) group latent variable. For multivariate Gaussian, these edges have partial correlation of observed variables with corresponding group latent variable, conditioned on the global latent variable; this can be converted to a linear representation with a loading matrix — see Section 6.16 of [11]. For the bi-factor copula, these additional edges are summarized with bivariate copulas linking observed variables with the corresponding group latent variable, conditioned on the global latent variable. The group latent variables are independent of each other and are independent of the global latent variable. There is dependence of all variables from the common link to the global latent variable. There is additional within-group dependence from links to the group latent variable.

For the oblique factor graph, each observed variable is linked to a (local) group latent variable, so that there is within-group dependence. The group latent variables are dependent, and these lead to between-group dependence.

The linear representations (Gaussian) and copula densities are given below, with notation in a form that allows for their study as the number of observed variables D increases to ∞. References and their derivations are in Sections 3.10 and 3.11 of [11], and [15–17].

For notation, observed variables are denoted as U_j or U_{jg} after transform to $U(0,1)$, or Z_j or Z_{jg} after transform to $N(0,1)$, and latent variables are denotes as V, V_0 or V_g on the $U(0,1)$ scale and W, W_0 or W_g on the $N(0,1)$ scale. Copula distributions for different vectors are indicated using C with subscripts for random vectors. The generic notation for a bivariate copula cumulative distribution function (cdf) has the form $C_{U,V}(u,v)$ and its partial derivatives are denoted as $C_{U|V}(u|v) = \partial C_{U,V}(u,v)/\partial v$ and $C_{V|U}(v|u) = \partial C_{U,V}(u,v)/\partial u$ because these are conditional distributions. Lower case variables are used as arguments of densities or dummy variables of integrals.

For the 1-factor copula model for $\boldsymbol{U}_D = (U_1,\dots,U_D)$, with $c_{jV} = c_{U_j,V}$ for all j, the copula density is:

$$c_{\boldsymbol{U}_D}(\boldsymbol{u}_D) = \int_0^1 \prod_{j=1}^{D} c_{jV}(u_j, v)\,\mathrm{d}v. \tag{1}$$

For the bi-factor copula model with G groups, d_g variables in group g, $\boldsymbol{D} = (d_1,\dots,d_G)$ and $D = \sum_{g=1}^{G} d_g$ total number of variables, the copula density is:

$$c_{\boldsymbol{U}_{\boldsymbol{D}}}(\boldsymbol{u}_{\boldsymbol{D}}) = \int_0^1 \prod_{g=1}^{G} \Big\{ \int_0^1 \prod_{j=1}^{d_g} c_{U_{jg}V_0}(u_{jg}, v_0) \cdot c_{U_{jg}V_g;V_0}\big(C_{U_{jg}|V_0}(u_{jg}|v_0), v_g\big)\,\mathrm{d}v_g \Big\} \mathrm{d}v_0, \tag{2}$$

where $c_{U_{jg}V_g;V_0}$ is the copula density assigned to the edge with connecting U_{jg}, V_g given V_0.

For the oblique factor copula model with d_g variables in group g as above, the copula density is:

$$c_{\boldsymbol{U}_{\boldsymbol{D}}}(\boldsymbol{u}_{\boldsymbol{D}}) = \int_0^1 \cdots \int_0^1 \prod_{g=1}^{G} \prod_{j=1}^{d_g} c_{U_{jg},V_g}(u_{jg}, v_g)\, c_{\boldsymbol{V}}(\boldsymbol{v})\,\mathrm{d}v_1 \dots \mathrm{d}v_G. \tag{3}$$

where $c_{\boldsymbol{V}}(v)$ is the joint copula density of the G latent variables.

When all linking copulas are bivariate Gaussian copulas, the usual representation of Gaussian factor models result after transforms of $U(0,1)$ variables to standard normal $N(0,1)$ variables.

- Gaussian 1-factor with $-1 < \alpha_j < 1$ for all j:

$$Z_j = \alpha_j W + \psi_j \epsilon_j, \quad j \in \{1,\dots,D\}, \tag{4}$$

  where $W, \epsilon_1, \epsilon_2, \dots$ are mutually independent $N(0,1)$ random variables, and $\psi_j^2 = 1 - \alpha_j^2$ for all j.
- Gaussian bi-factor with $\alpha_{jg,0}, \alpha_{jg} \in (-1,1)$:

$$Z_{jg} = \alpha_{jg,0} W_0 + \alpha_{jg} W_g + \psi_{jg}\epsilon_{jg}, \quad j \in \{1,\dots,d_g\}, \quad g \in \{1,\dots,G\}, \tag{5}$$

  where $W_0, \{W_g\}, \{\epsilon_{jg}\}$ are mutually independent $N(0,1)$, and $\psi_{jg}^2 = 1 - \alpha_{jg,0}^2 - \alpha_{jg}^2 < 1$ for all (j,g). Note that $\rho_{Z_{jg},W_g;W_0} = \alpha_{jg}/(1-\alpha_{jg,0}^2)^{1/2}$ is the partial correlation of Z_{jg} with W_g given W_0.
- Gaussian oblique factor with $\alpha_{jg} \in (-1,1)$:

$$Z_{jg} = \alpha_{jg} W_g + \psi_{jg}\epsilon_{jg}, \quad j \in \{1,\dots,d_g\}, \quad g \in \{1,\dots,G\}, \tag{6}$$

  where $\{\epsilon_{jg}\}$ are mutually independent $N(0,1)$, independent of the multivariate normal vector $(W_1,\dots,W_G)$, with zero mean vector and unit variances, and $\psi_{jg}^2 = 1 - \alpha_{jg}^2$ for all (j,g). Let $\boldsymbol{\Sigma}_W$ be the correlation matrix of $\boldsymbol{W} = (W_1,\dots,W_G)^\top$.

In all cases, the matrix representation has form:

$$\boldsymbol{Z}_D = \boldsymbol{A}_D \boldsymbol{W} + \boldsymbol{\Psi}_D \boldsymbol{\epsilon}_D, \tag{7}$$

where the loading matrix $\boldsymbol{A}_D$ is of size $D \times p$, $\boldsymbol{\Psi}_D^2$ is a $D \times D$ diagonal matrix of individual variances (ψ_j or ψ_{jg}), $\boldsymbol{\epsilon}_D$ is a $D \times 1$ column vector of ϵ_j or ϵ_{jg}, and $\boldsymbol{Z}_D$ is a $D \times 1$ column vector of Z_j or Z_{jg}. For 1-factor, $p = 1$; for bi-factor, $p = G + 1$ and $\boldsymbol{W} = (W_0, W_1, \dots, W_G)^\top$, and for oblique factor, $p = G$ and $\boldsymbol{W} = (W_1,\dots,W_G)^\top$.

A matrix identity that is useful in calculations of conditional expectation and covariance of $\boldsymbol{W}$ given $\boldsymbol{Z}_D$ is the following

$$\boldsymbol{A}_D^\top (\boldsymbol{A}_D \boldsymbol{A}_D^\top + \boldsymbol{\Psi}_D^2)^{-1} = (\mathbf{I}_p + \boldsymbol{A}_D^\top \boldsymbol{\Psi}_D^{-2} \boldsymbol{A}_D)^{-1} \boldsymbol{A}_D^\top \boldsymbol{\Psi}_D^{-2}, \tag{8}$$

when $\boldsymbol{A}_D \boldsymbol{A}_D^\top + \boldsymbol{\Psi}_D^2$ is non-singular and $\boldsymbol{\Psi}_D^2$ has no zeros on the diagonal. This identity is given in [13].

In general, except for the case of bivariate Gaussian linking copulas, the integrals in the above copula densities do not simplify, and numerical maximum likelihood involves numerical integration when there is a random sample of size N. With a parametric family for each bivariate linking copula, [15,16] outline numerically efficient approaches for a modified Newton–Raphson method for optimizing the negative log-likelihood for 1-factor, bi-factor and a special nested factor subcase of the oblique factor copulas. 1-dimensional or 2-dimensional Gauss–Legendre quadrature is used to evaluate the integrals and their partial derivatives order 1 and 2 for the gradient and Hessian of the negative log-likelihood. The factor copulas are extensions of their Gaussian counterparts and are useful when plots of normal scores data (after empirical transforms to $N(0,1)$) show tail dependence or tail asymmetry in bivariate scatterplots.

In classical factor analysis (e.g., [13]), factor scores or estimates of latent variables are considered after estimating a loading matrix. The information on the latent variables may be used in subsequent analysis following the factor analysis; for example, regression analysis incorporated the factor information. For factor copulas, these could be obtained after fitting a parametric model. Factor copula models for practical use are considered when Gaussian factor models are considered as first-order approximations, so that (transformed) factor scores could be considered as a starting point. An approach to estimate the latent variables without the need to fit a parametric model by numerical procedures in [15,16] is proposed, and more details will be illustrated in later sections.

The next section has proxies as estimates of latent variables based on conditional expectations given observed variables.

3. Proxies for the latent variables

In the Gaussian factor models, factor scores are defined as the estimates of unobserved latent variables, see [13]. The form of factor scores that extend to factor copula models are the regression factor scores, which are conditional expectations of latent variables given the observed variables.

For factor copula models, having reasonable estimates of latent variables is also of interest since these can lead to simpler and more efficient numerical procedures for determining parametric bivariate linking copula families and estimating their parameters. For factor copula models, we use the term "proxies" for the estimates of latent variables, as in [17].

The study of the conditional expectation of latent variables (W's or V's) given observed variables (Z's or U's) is done in three stages for the models in Section 2.

A. The loading matrix is known or all of the bivariate linking copulas are known. In this case, the proxy variables are defined as the conditional expectation of latent variables given the observed variables, and we refer to "conditional expectation" proxies.

B1. Gaussian factor models with estimated loading matrix. Since the general Gaussian factor model is non-identifiable in terms of rotation of the loading matrix, for a model with two or more factors, consistency of estimation requires a structured loading matrix such as that of the bi-factor model or oblique factor model. In the models, the proxies are defined in the same way as in case A but with an estimated loading matrix (in blocks), where parameter estimates have a variance of order $O(1/N)$ for sample size N.

B2. Factor copula models with known parametric families for each linking copula. In the models, the parameters are estimated via sequential maximum likelihood with a variance of order $O(1/N)$ for sample size N. The proxies are defined in the same way as in case A with the estimated linking copulas.

C. Linking copula families are not known or specified in advance (the situation in practice). A sequential method is used starting with unweighted averages as estimates in [17] or regression factor scores computed from an estimated loading matrix after observed variables are transformed to have $N(0,1)$ margins. Then, the "conditional expectation" proxies are constructed with the copula families and estimated parameters determined in the first stage.

In Sections 3.1 and 3.2, the conditional expectations (for case A) are given. The asymptotic properties of proxies for case A and for cases B1, B2 are given in Sections 4 and 5 respectively, and the sequential method of case C is in Section 6.

3.1. Proxies in Gaussian factor models

In this section, we summarize $\mathrm{E}(\boldsymbol{W}|\boldsymbol{Z}_D=\boldsymbol{z}_D)$ for p-factor (1-factor is a special case when $p=1$), bi-factor and oblique factor Gaussian models with observed variables that are in $N(0,1)$. These are called (regression) factor scores in the factor analysis literature; see [13].

For p-factor, let $(\boldsymbol{w}^0,\boldsymbol{z}_D)$ be a realization of $(\boldsymbol{W},\boldsymbol{Z}_D)$. The proxy for $\boldsymbol{W}$ (or estimate of $\boldsymbol{w}^0$) given $\boldsymbol{z}_D$ is:

$$\widetilde{\boldsymbol{w}}_{\boldsymbol{D}}=\mathrm{E}(\boldsymbol{W}|\boldsymbol{Z}_D=\boldsymbol{z}_D)=\boldsymbol{A}_{\boldsymbol{D}}^\top(\boldsymbol{A}_{\boldsymbol{D}}\boldsymbol{A}_{\boldsymbol{D}}^\top+\boldsymbol{\Psi}_{\boldsymbol{D}}^2)^{-1}\boldsymbol{z}_{\boldsymbol{D}}=(\mathbf{I}_p+\boldsymbol{A}_{\boldsymbol{D}}^\top\boldsymbol{\Psi}_{\boldsymbol{D}}^{-2}\boldsymbol{A}_{\boldsymbol{D}})^{-1}\boldsymbol{A}_{\boldsymbol{D}}^\top\boldsymbol{\Psi}_{\boldsymbol{D}}^{-2}\boldsymbol{z}_{\boldsymbol{D}},\tag{9}$$

if $(\boldsymbol{A}_{\boldsymbol{D}}\boldsymbol{A}_{\boldsymbol{D}}^\top+\boldsymbol{\Psi}_{\boldsymbol{D}}^2)$ is non-singular and $\boldsymbol{\Psi}_{\boldsymbol{D}}$ has no zeros on diagonal. The above matrix equality follows from (8).

If $\boldsymbol{\Psi}_{\boldsymbol{D}}$ has zeros on diagonal, then a linear combination of the latent variables is an observed variable. If $(\boldsymbol{A}_{\boldsymbol{D}}\boldsymbol{A}_{\boldsymbol{D}}^\top+\boldsymbol{\Psi}_{\boldsymbol{D}}^2)$ is singular, then a linear combination of observed variables is a constant. These unrealistic cases will not be considered.

For bi-factor, let $(w_0^0,w_1^0,\ldots,w_G^0,\boldsymbol{z}_D)$ be a realization of $(W_0,W_1,\ldots,W_G,\boldsymbol{Z}_D)$. The proxies (or estimates of $w^0,w_1^0,\ldots,w_G^0$) given $\boldsymbol{z}_D$ are:

$$\widetilde{w}_0=\mathrm{E}(W_0|\boldsymbol{Z}_{\boldsymbol{D}}=\boldsymbol{z}_{\boldsymbol{D}})=(\boldsymbol{a}_0)^\top(\boldsymbol{A}_{\boldsymbol{D}}\boldsymbol{A}_{\boldsymbol{D}}^\top+\boldsymbol{\Psi}_{\boldsymbol{D}}^2)^{-1}\boldsymbol{z}_{\boldsymbol{D}},\tag{10}$$

$$\widetilde{w}_g(\widetilde{w}_0)=\mathrm{E}(W_g|\boldsymbol{Z}_{\boldsymbol{D}}=\boldsymbol{z}_{\boldsymbol{D}},W_0=\widetilde{w}_0)=(\boldsymbol{b}_g^\top,0)(\boldsymbol{\Sigma}_g)^{-1}(\boldsymbol{z}_g^\top,\widetilde{w}_0)^\top,\tag{11}$$

where $\boldsymbol{a}_0$ is the first column of the loading matrix, $\boldsymbol{z}_g=(z_{1g},\ldots,z_{d_g g})^\top$, $\boldsymbol{b}_{0g}$ and $\boldsymbol{b}_g$ are the $d_g\times 1$ global and local loading vector for group g; let $\boldsymbol{B}_g=[\boldsymbol{b}_{0g},\boldsymbol{b}_g]$ (matrix of size $d_g\times 2$). Let $\boldsymbol{\Sigma}_g$ be the correlation matrix of $(\boldsymbol{Z}_g^\top,W_0)$. Then $\boldsymbol{\Sigma}_g=\begin{bmatrix}\boldsymbol{B}_g\boldsymbol{B}_g^\top+\boldsymbol{\Psi}_g^2 & \boldsymbol{b}_{0g}\\ \boldsymbol{b}_{0g}^\top & 1\end{bmatrix}$ for $g\in\{1,\ldots,G\}$. The proof that $\widetilde{w}_g(\widetilde{w}_0)=\mathrm{E}(W_g|\boldsymbol{Z}_{\boldsymbol{D}}=\boldsymbol{z}_{\boldsymbol{D}})$ is given in Appendix A.1. The two-stage formulation of (10) and (11) is needed for the copula extension.

For oblique factor, let $(w_1^0,\ldots,w_D^0,\boldsymbol{z}_D)$ be a realization of $(W_1,\ldots,W_G,\boldsymbol{Z}_D)$. The proxies (or estimates of $w_1^0,\ldots,w_G^0$) given $\boldsymbol{z}_D$ are:

$$\widetilde{w}_g=\mathrm{E}(W_g|\boldsymbol{Z}_g=\boldsymbol{z}_g)=\boldsymbol{a}_g^\top(\boldsymbol{a}_g\boldsymbol{a}_g^\top+\boldsymbol{\Psi}_g^2)^{-1}\boldsymbol{z}_g,\tag{12}$$

where $\boldsymbol{z}_g=(z_{1g},\ldots,z_{d_g g})^\top$, $\boldsymbol{\Psi}_g$ is the gth block diagonal of $\boldsymbol{\Psi}$ and $\boldsymbol{a}_g$ is the $d_g\times 1$ loading vector for group g, for $g\in\{1,\ldots,G\}$. This version, rather than $\mathrm{E}(W_g|\boldsymbol{Z}_h=\boldsymbol{z}_h,h\in\{1,\ldots,G\})$, has a version for the oblique factor copula that is numerically easier to handle.

Table 1
Densities for tuple random vectors $(\boldsymbol{U_D}, V_0, \boldsymbol{V})$, $(\boldsymbol{U_D}, \boldsymbol{V}_0)$, and $(\boldsymbol{U}_g, \boldsymbol{V}_0, \boldsymbol{V}_g)$ in the bi-factor copula model (2), where $\boldsymbol{V} = (V_1, \ldots, V_G)$.

Vector	Joint density
$(\boldsymbol{U_D}, V_0, \boldsymbol{V})$	$c_{\boldsymbol{U_D},V_0,\boldsymbol{V}}(\boldsymbol{u_D}, v_0, \boldsymbol{v}) = \prod_{g=1}^{G}\prod_{j=1}^{d_g}\left\{c_{U_{jg}V_0}(u_{jg}, v_0)\cdot c_{U_{jg}V_g;V_0}\left(C_{U_{jg}\|V_0}(u_{jg}\|v_0), v_g\right)\right\}$
$(\boldsymbol{U_D}, \boldsymbol{V}_0)$	$c_{\boldsymbol{U_D},V_0}(\boldsymbol{u_D}, v_0) = \prod_{g=1}^{G}\left\{\prod_{j=1}^{d_g} c_{U_{jg}V_0}(u_{jg}, v_0)\cdot f_g(\boldsymbol{u}_g; v_0)\right\}$, where
	$f_g(\boldsymbol{u}_g; v_0) = \int_0^1 \prod_{j=1}^{d_g} c_{U_{jg}V_g;V_0}\left(C_{U_{jg}\|V_0}(u_{jg}\|v_0), v_g\right)\mathrm{d}v_g, \quad g \in \{1, \ldots, G\}$
$(\boldsymbol{U}_g, \boldsymbol{V}_0, \boldsymbol{V}_g)$	$c_{\boldsymbol{U}_g,V_0,V_g}(\boldsymbol{u}_g, v_0, v_g) = \prod_{j=1}^{d_g}\left\{c_{U_{jg}V_0}(u_{jg}, v_0)\cdot c_{U_{jg}V_g;V_0}\left(C_{U_{jg}\|V_0}(u_{jg}\|v_0), v_g\right)\right\}$

3.2. Proxies in factor copula models

This subsection has the corresponding conditional expectations of latent variables given $U(0,1)$ distributed observed variables for the 1-factor, bi-factor, and oblique factor copula models.

For the 1-factor copula model (1), with $c_{V\boldsymbol{U_D}}(v, \boldsymbol{u_D}) = \prod_{j=1}^{D} c_{jV}(u_j, v)$, let $(v^0, u_1, \ldots, u_D)$ be one realization of $(V, U_1, \ldots, U_D)$. Then the proxy, as an estimate of v^0, is:

$$\tilde{v}_D = \tilde{v}_D(\boldsymbol{u_D}) = \mathrm{E}(V|\boldsymbol{U}_D = \boldsymbol{u_D}) = \int_0^1 v\, c_{V\boldsymbol{U}_D}(v, \boldsymbol{u_D})\,\mathrm{d}v \Big/ \int_0^1 c_{V\boldsymbol{U}_D}(v, \boldsymbol{u_D})\,\mathrm{d}v \tag{13}$$

For the bi-factor copula model in (2), let $(v_0^0, \{v_g^0\}, \{u_{jg}\})$ be one realization of $(V_0, \{V_g\}, \{U_{jg}\})$. Table 1 has densities involving the global latent variable V_0 and group latent variables $V_1, \ldots, V_G$.

The convenient form of conditional expectations is as follows: with $\boldsymbol{D} = (d_1, \ldots, d_g)$,

$$\tilde{v}_{0\boldsymbol{D}}(\boldsymbol{u_D}) = \mathrm{E}[V_0|U_{jg} = u_{jg}, j \in \{1, \ldots, d_g\}, g \in \{1, \ldots, G\}] = \frac{\int_0^1 v_0 c_{\boldsymbol{U_D},V_0}(\boldsymbol{u_D}, v_0)\,\mathrm{d}v_0}{\int_0^1 c_{\boldsymbol{U_D},V_0}(\boldsymbol{u_D}, v_0)\,\mathrm{d}v_0}. \tag{14}$$

For $g \in \{1, \ldots, G\}$,

$$\tilde{v}_{g\boldsymbol{D}}(\boldsymbol{u}_g, \tilde{v}_{0\boldsymbol{D}}) = \mathrm{E}[V_g|V_0 = \tilde{v}_{0\boldsymbol{D}}, U_{jg} = u_{jg}, j \in \{1, \ldots, d_g\}] = \frac{\int_0^1 v_g c_{\boldsymbol{U}_g,V_0,V_g}(\boldsymbol{u}_g, \tilde{v}_{0\boldsymbol{D}}, v_g)\,\mathrm{d}v_g}{\int_0^1 c_{\boldsymbol{U}_g,V_0,V_g}(\boldsymbol{u}_g, \tilde{v}_{0\boldsymbol{D}}, v_g)\,\mathrm{d}v_g}. \tag{15}$$

The proxies are $\tilde{v}_{0\boldsymbol{D}}(\boldsymbol{u_D})$ (as an estimate of v_0^0) and $\tilde{v}_{g\boldsymbol{D}}(\boldsymbol{u}_g, \tilde{v}_{0\boldsymbol{D}})$ (as an estimate of v_g^0) for $g \in \{1, \ldots, G\}$.

For the oblique factor copula model in (3), for $g \in \{1, \ldots, G\}$ let $c_{\boldsymbol{U}_g,V_g}(v_g, \boldsymbol{u}_g) = \prod_{j=1}^{d_g} c_{U_{jg},V_g}(u_{jg}, v_g)$. Let $(\{v_g^0\}, \{u_{jg}\})$ be one realization of $(\{V_g\}, \{U_{jg}\})$. Then

$$\tilde{v}_{g\boldsymbol{D}}(\boldsymbol{u}_g) = \mathrm{E}[V_g|U_{jg} = u_{jg}, j \in \{1, \ldots, d_g\}] = \frac{\int_0^1 v_g c_{\boldsymbol{U}_g,V_g}(\boldsymbol{u}_g, v_g)\,\mathrm{d}v_g}{c_{\boldsymbol{U}_g}(\boldsymbol{u}_g)} = \frac{\int_0^1 v_g c_{\boldsymbol{U}_g,V_g}(\boldsymbol{u}_g, v_g)\,\mathrm{d}v_g}{\int_0^1 c_{\boldsymbol{U}_g,V_g}(\boldsymbol{u}_g, v_g)\,\mathrm{d}v_g}. \tag{16}$$

The proxy for v_g^0 is $\tilde{v}_{g\boldsymbol{D}}(\boldsymbol{u}_g)$ for $g \in \{1, \ldots, G\}$.

4. Consistency of proxies and rate of convergence: model known

In this section, we obtain conditions so that the proxies defined in Section 3 are consistent as $D \to \infty$ for 1-factor and $d_g \to \infty$ for all g for bi-factor or oblique factor models. More direct calculations are possible for Gaussian models and these provide insights into behavior for factor copulas.

4.1. Conditional variance for Gaussian factor models

We start with the conditional variance of the latent variables given the observed variables. If the conditional variance does not go to zero as $D \to \infty$, then the latent variable cannot be consistently estimated; this can happen if the overall dependence with the latent variable is weak, even as more variables are added. If the conditional variance is 0 for a finite D or $\boldsymbol{D}$, then the latent variable can be determined exactly (this can happen for the 1-factor model if $\psi_j = 0$ for some j). The practical case is when the dependence is moderate to strong, so that intuitively we have a better idea of the value of the latent variable as D or $\boldsymbol{D}$ increases.

We summarize the expressions of the conditional variances in Gaussian factor models in Table 2. The 1-factor model and bi-factor model are special cases of the p-factor model. For the bi-factor model, we decompose the conditional variance of group latent variables into two parts; one part only depends on the within-group dependence, and another part comes from the conditional variance of the global latent variable.

Table 2
Conditional variance of latent factors in general and structured Gaussian factor models; the matrix equality for p-factor follows from (8); $q_g = \boldsymbol{b}_g^\top \boldsymbol{\Psi}_g^{-2} \boldsymbol{b}_g$, $\tilde{q}_g = \boldsymbol{b}_g^\top \boldsymbol{\Psi}_g^{-2} \boldsymbol{b}_{0g}$, where $\boldsymbol{b}_g$, $\boldsymbol{b}_{0g}$ are the $d_g \times 1$ local and global loading vectors respectively for group g, $g \in \{1, \dots, G\}$.

Model	Conditional variance	
p-factor	$\mathrm{Cov}(\boldsymbol{W}\mid \boldsymbol{Z}_D = \boldsymbol{z}_D) = \mathbf{I}_p - \boldsymbol{A}_D^\top(\boldsymbol{A}_D\boldsymbol{A}_D^\top + \boldsymbol{\Psi}_D^2)^{-1}\boldsymbol{A}_D$	
	$= \mathbf{I}_p - (\mathbf{I}_p + \boldsymbol{A}_D^\top \boldsymbol{\Psi}_D^{-2}\boldsymbol{A}_D)^{-1}\boldsymbol{A}_D^\top\boldsymbol{\Psi}_D^{-2}\boldsymbol{A}_D$	
bi-factor	$\mathrm{Var}(W_0\mid \boldsymbol{Z}_D = \boldsymbol{z}_D) = 1 - \boldsymbol{a}_0^\top(\boldsymbol{A}_D\boldsymbol{A}_D^\top + \boldsymbol{\Psi}_D^2)^{-1}\boldsymbol{a}_0$ and	
	$\mathrm{Var}(W_g\mid \boldsymbol{Z}_D = \boldsymbol{z}_D) = (1+q_g)^{-1} + \tilde{q}_g^2(1+q_g)^{-2}\,\mathrm{Var}(W_0\mid \boldsymbol{Z}_D = \boldsymbol{z}_D)$	(17)
oblique factor	$\mathrm{Cov}(\boldsymbol{W}\mid \boldsymbol{Z} = \boldsymbol{z}_D) = \boldsymbol{\Sigma}_W - \boldsymbol{\Sigma}_W\boldsymbol{A}_D^\top(\boldsymbol{A}_D\boldsymbol{\Sigma}_W\boldsymbol{A}_D^\top + \boldsymbol{\Psi}_D^2)^{-1}\boldsymbol{A}_D\boldsymbol{\Sigma}_W$	(18)

Details of the derivations for the decomposition formula in (17) of the conditional variance can be found in Appendix A.2. The expression of the conditional variance in the oblique factor model is slightly different from the p-factor models, but the derivations are similar so the details are omitted.

In the (second) expression of conditional variance for the general p-factor models, define $\boldsymbol{Q}_D = \boldsymbol{A}_D^\top\boldsymbol{\Psi}_D^{-2}\boldsymbol{A}_D$ as a $p\times p$ matrix. Suppose $\overline{\boldsymbol{Q}}_D =: D^{-1}\boldsymbol{Q}_D \to \boldsymbol{Q}$ as $D \to \infty$. Note that if dependence in the loading matrices is weak, then $\boldsymbol{Q}$ can be the zero matrix. In the boundary case with $\boldsymbol{A}_D = \boldsymbol{0}$, then $\mathrm{Cov}(\boldsymbol{W}\mid \boldsymbol{Z}_D = \boldsymbol{z}_D) = \mathbf{I}_p$, that is, $\boldsymbol{Z}_D$ provides no information about $\boldsymbol{W}$. If $\boldsymbol{Q}$ is a positive definite matrix, then

$$\mathbf{I}_p - (\mathbf{I}_p + \boldsymbol{A}_D^\top\boldsymbol{\Psi}_D^{-2}\boldsymbol{A}_D)^{-1}\boldsymbol{A}_D^\top\boldsymbol{\Psi}_D^{-2}\boldsymbol{A}_D \approx \mathbf{I}_p - (\mathbf{I}_p + D\boldsymbol{Q})^{-1}D\boldsymbol{Q} = \mathbf{I}_p - (D^{-1}\boldsymbol{Q}^{-1} + \mathbf{I}_p)^{-1} = D^{-1}\boldsymbol{Q}^{-1} + o(D^{-1}).$$

In the p-factor model, if the loading matrices in $\{\boldsymbol{A}_D\}$ have full column rank, and the strength of dependence between observed variables and the latent factors is strong enough such that the limiting matrix $\boldsymbol{Q}$ is invertible, then the limit of conditional covariance for p-factor is $O(D^{-1})$. The next theorem indicates what happens if the condition of full column rank does not hold.

Theorem 1. *Consider the p-factor model with $p \geq 2$ with matrix representation (7).*

(i) If $\boldsymbol{A}_D$ does not have full column rank, then the latent variables in $\boldsymbol{W}$ are not identifiable.

(ii) For the Gaussian bi-factor model as a special case of p-factor with $p = G+1$, if $\boldsymbol{A}_D$ does not have full column rank, then the Gaussian bi-factor model can be rewritten as an oblique factor model with fewer parameters.

Proof. (i) Let $\boldsymbol{a}_1, \dots, \boldsymbol{a}_p$ be the columns of $\boldsymbol{A}_D$. The columns of $\boldsymbol{A}_D$ are linearly dependent. Without loss of generality, assume $\boldsymbol{a}_1 = t_2\boldsymbol{a}_2 + \cdots + t_p\boldsymbol{a}_p$ where $(t_2, \dots, t_p)$ is a non-zero vector. Then, in (7),

$$\boldsymbol{Z}_D - \boldsymbol{\Psi}_D\boldsymbol{\epsilon}_D = \sum_{j=2}^{p} t_j\boldsymbol{a}_j \cdot W_1 + \sum_{j=2}^{p}\boldsymbol{a}_j W_j = \sum_{j=2}^{p}\boldsymbol{a}_j(t_jW_1 + W_j).$$

Hence only some linear combinations of the latent variables can be identified.

(ii) For bi-factor, let $\boldsymbol{a}_0, \boldsymbol{a}_1, \dots, \boldsymbol{a}_G$ be the columns of $\boldsymbol{A}_D$, with $\boldsymbol{a}_0$ being a linear combination of the remaining columns (because this is the only column with non-zero elements for each group). Let the latent variables be $W_0, W_1, \dots, W_G$. Part (i) implies that

$$\boldsymbol{Z}_D = \sum_{g=1}^{G}\boldsymbol{a}_g W_g^* + \boldsymbol{\Psi}_D\boldsymbol{\epsilon}_D = \boldsymbol{A}_D^*\boldsymbol{W}^* + \boldsymbol{\Psi}_D\boldsymbol{\epsilon}_D,$$

where $W_g^* = t_gW_0 + W_g$ for a non-zero vector $(t_1, \dots, t_G)$ and $\boldsymbol{A}_D^* = (\boldsymbol{a}_1, \dots, \boldsymbol{a}_G)$ is $D \times G$. The identifiable latent variables W_g^* are dependent. □

Remark 1. For the bi-factor model, if the global loading vector is roughly equal to a linear combination of the group loading vectors, then the latent factors are close to non-identifiable, and the oblique factor model may be a good approximation. A useful diagnostic tool is the condition number of $\boldsymbol{Q}_D = \boldsymbol{A}_D^\top\boldsymbol{\Psi}_D^{-2}\boldsymbol{A}_D$ because $\boldsymbol{Q}_D$ is not of full rank if $\boldsymbol{A}_D$ is not of full column rank. If the condition number is small enough, then the bi-factor model is appropriate to use; otherwise, oblique factor model can be a good fit.

For the oblique factor model, with $\boldsymbol{B}_D = \boldsymbol{A}_D\boldsymbol{\Sigma}_W^{1/2}$ and using (8), the right-hand-side of in Table 2 becomes:

$$\boldsymbol{\Sigma}_W - \boldsymbol{\Sigma}_W^{1/2}\boldsymbol{B}_D^\top(\boldsymbol{B}_D\boldsymbol{B}_D^\top + \boldsymbol{\Psi}_D^2)^{-1}\boldsymbol{B}_D\boldsymbol{\Sigma}_W^{1/2} = \boldsymbol{\Sigma}_W - \boldsymbol{\Sigma}_W^{1/2}(\mathbf{I}_p + \boldsymbol{B}_D^\top\boldsymbol{\Psi}_D^{-2}\boldsymbol{B}_D)^{-1}\boldsymbol{B}_D^\top\boldsymbol{\Psi}_D^{-2}\boldsymbol{B}_D\boldsymbol{\Sigma}_W^{1/2}, \tag{19}$$

and

$$\boldsymbol{B}_D^\top\boldsymbol{\Psi}_D^{-2}\boldsymbol{B}_D = \boldsymbol{\Sigma}_W^{1/2}\boldsymbol{A}_D^\top\boldsymbol{\Psi}_D^{-2}\boldsymbol{A}_D\boldsymbol{\Sigma}_W^{1/2} = \boldsymbol{\Sigma}_W^{1/2}\boldsymbol{Q}_D\boldsymbol{\Sigma}_W^{1/2}.$$

After some algebraic calculations, (19) simplifies to $\boldsymbol{\Sigma}_W - (\mathbf{I}_p + \boldsymbol{Q}_D^{-1}\boldsymbol{\Sigma}_W^{-1})^{-1}\boldsymbol{\Sigma}_W$. The limit of conditional covariance is $O(D^{-1})$ if $D^{-1}\boldsymbol{Q}_D \to \boldsymbol{Q}$ with $\boldsymbol{Q}$ being non-singular.

The rate of convergence as a function of strength of dependence with latent variables is discussed in the Supplementary material.

Suppose the conditional variance of a latent variable given the observed variables does not go to 0 as $D \to \infty$, then one cannot expect the corresponding proxy estimates in Section 2 to be consistent. For factor copula models, the conditional variances of the latent variables do not have closed forms. The results in this section for Gaussian factor models provide insights into conditions for consistency of proxy estimates in factor copula models as well as the connection between the rate of convergence and strength of dependence with the latent variables. Under some regularity conditions on the bivariate linking copulas, it can be shown that the limit of the conditional variance in factor copula models is also $O(D^{-1})$.

4.2. Consistency in Gaussian factor models

In this subsection, mild conditions are obtained for consistency of proxy estimates via conditional expectations in Section 3. The cases that are covered in the theorems have moderate to strong dependence, without loading parameters going to ± 1 as $D \to \infty$. In the latter case, with even stronger dependence with latent variables, there is consistency, but the proofs would be different because identity (8) would not hold in the limit. The conditions in the theorems match practical uses of factor models — one might have idea of latent factors that affect dependence within groups of variables; there is at least moderate dependence among observed variables and dependence is not so strong that one variable could be considered as a proxy for the latent variable.

The consistency of proxy variables in 1-factor, bi-factor and oblique Gaussian factor models are shown in Theorem 2, Theorem 4 and Corollary 3 respectively. The framework of theory is built on the assumption that correlations or partial correlations are being sampled from some appropriate super-population. In [3], the authors consider the asymptotic properties of regression factor scores (proxies) in high-dimensional unstructured p-factor models under several identifying restrictions, while we deal with structured p-factor models. The overlap is the 1-factor model, as the bi-factor and oblique factor models are not addressed in [3]. The assumptions for consistency of proxies in the structured Gaussian-factor models shed light into the corresponding factor copula model.

Theorem 2 (*Asymptotic Properties of Factor Scores in 1-Factor Gaussian Model*). *For the 1-factor model in (4), with given sequence $\{\alpha_1, \alpha_2, \dots\}$, suppose there is a realized infinite sequence of observed variables $z_1, z_2, \dots$ and a realized infinite sequence of disturbance terms $e_1, e_2, \dots$ with realized latent variable w^0 (independent of dimension D) from the 1-factor model. For the truncated sequence to the first D variables, let $\boldsymbol{z_D} = (z_1, \dots, z_D)^\top$, $\boldsymbol{e_D} = (e_1, \dots, e_D)^\top$, and let the loading matrix or vector be $(\alpha_1, \dots, \alpha_D)^\top$. Assume*

$$-1 < \liminf_{j\to\infty} \alpha_j < \limsup_{j\to\infty} \alpha_j < 1, \quad \lim_{D\to\infty} D^{-1} \sum_{j=1}^{D} |\alpha_j| \to const, \quad const \neq 0.$$

Then for the factor scores defined in Eq. (9), $\widetilde{w}_D - w^0 = O_p(D^{-1/2})$ as $D \to \infty$.

Remark 2. The proof is given in Appendix B.1. The assumption on α_j uniformly bounded away from ± 1 ensures that the proxies are well-defined in two equivalent forms in (9). If some α_j's approach ± 1, then one or more of the observed variables can be considered as the proxy and then (9) is not needed. The second assumption about the averaged absolute loadings ensures that the dependence is strong enough, because from Section 4.1, consistency does not hold in the case of sufficiently weak dependence. If the α_j's are randomly sampled from a super-population or distribution with values bounded away from -1 and 1 with absolute mean that is non-zero, then the condition in Theorem 2 for $\{\alpha_j\}$ is satisfied with probability 1. Similar comments on super-populations apply below for the oblique and bi-factor models.

Corollary 3 (*Asymptotic Properties of Factor Scores in Oblique Factor Gaussian Model*). *For the oblique factor model with fixed G groups defined in (6), with given sequences $\{\alpha_{1g}, \alpha_{2g}, \dots\}$ for $g \in \{1, \dots, G\}$, suppose there are realized infinite sequences of observed variables $\boldsymbol{z}_1^\top = (z_{1,1}, z_{2,1}, \dots)$, $\dots$, $\boldsymbol{z}_G^\top = (z_{1,G}, z_{2,G}, \dots)$ and realized sequences of disturbance terms $\boldsymbol{e}_1^\top = (e_{1,1}, e_{2,1}, \dots)$, $\dots$, $\boldsymbol{e}_G^\top = (e_{1,G}, e_{2,G}, \dots)$, with latent variables $\boldsymbol{w}^0 = (w_1^0, \dots, w_G^0)$. Truncate the sequences to the first d_g variables $\boldsymbol{z}_{g,d_g}$ for $g \in \{1, \dots, G\}$ with no d_g dominating others. Let $\boldsymbol{z_D} = (\boldsymbol{z}_{1,d_1}^\top, \boldsymbol{z}_{2,d_2}^\top, \dots, \boldsymbol{z}_{G,d_G}^\top)^\top$, $\boldsymbol{e_D} = (\boldsymbol{e}_{1,d_1}^\top, \boldsymbol{e}_{2,d_2}^\top, \dots, \boldsymbol{e}_{G,d_G}^\top)$. With $\boldsymbol{a}_{d_g,g} = (\alpha_{1g}, \dots, \alpha_{d_g g})$, assume $d_g^{-1} \|\boldsymbol{a}_{d_g,g}\|_1 \not\to 0$ as $d_g \to \infty$ for $g \in \{1, 2, \dots, G\}$. Also assume that $\limsup_{j\to\infty} |\alpha_{jg}| < 1$ for all g. Let $\widetilde{\boldsymbol{w}}_{\boldsymbol{D}} = (\widetilde{w}_1, \dots, \widetilde{w}_G)$ be the factor scores defined in (12). Then*

$$\widetilde{w}_g - w_g^0 = O_p(D^{-1/2}), \quad g \in \{1, \dots, G\}.$$

Proof. The oblique model implies a 1-factor model for each group g, so Theorem 2 applies. □

The bi-factor model which also has variables in non-overlapping groups must be handled differently. For example, the theory of [17] can handle 1-factor and oblique factor models but not bi-factor models.

Theorem 4 (*Asymptotic Properties of Factor Scores in Bi-factor Gaussian Model*). *For the bi-factor model with fixed G groups defined in (5), with given sequences $\{\alpha_{1g}, \alpha_{2g}, \dots\}$ for $g \in \{1, \dots, G\}$, suppose there are realized infinite sequences of observed variables $\boldsymbol{z}_1^\top = (z_{1,1}, z_{2,1}, \dots)$, $\dots$, $\boldsymbol{z}_G^\top = (z_{1,G}, z_{2,G}, \dots)$ and realized sequences of disturbance terms $\boldsymbol{e}_1^\top = (e_{1,1}, e_{2,1}, \dots)$, $\dots$, $\boldsymbol{e}_G^\top = (e_{1,G}, e_{2,G}, \dots)$, with latent variables $\boldsymbol{w}^0 = (w_0^0, w_1^0, \dots, w_G^0)$. Truncate the sequences to the first d_g variables $\boldsymbol{z}_{g,d_g}$ for $g \in \{1, \dots, G\}$, with no d_g dominating others. Let $\boldsymbol{z_D} = (\boldsymbol{z}_{1,d_1}^\top, \boldsymbol{z}_{2,d_2}^\top, \dots, \boldsymbol{z}_{G,d_G}^\top)^\top$, $\boldsymbol{e_D} = (\boldsymbol{e}_{1,d_1}^\top, \boldsymbol{e}_{2,d_2}^\top, \dots, \boldsymbol{e}_{G,d_G}^\top)$ and assume that the loading matrices $\boldsymbol{A_D} = [\boldsymbol{a}_{D,0}, \mathrm{diag}(\boldsymbol{a}_{d_1,1}, \dots, \boldsymbol{a}_{d_G,G})]$ are of full rank, with bounded condition number over $d_g \to \infty$ for all g. With $\boldsymbol{a}_{d_g,g} = (\alpha_{1g}, \dots, \alpha_{d_g g})$, assume $D^{-1}\|\boldsymbol{a}_{D,0}\|_1 \not\to 0$, $d_g^{-1}\|\boldsymbol{a}_{d_g,g}\|_1 \not\to 0$ for $g \in \{1, \dots, G\}$, and $\liminf_{j,g} \psi_{jg}^2 > 0$. Let $\widetilde{\boldsymbol{w}}_{\boldsymbol{D}} = (\widetilde{w}_0, \dots, \widetilde{w}_G)$ be the factor scores defined in (10) and (11). Then*

$$\widetilde{w}_g - w_g^0 = O_p(D^{-1/2}), \quad g \in \{0, 1, 2, \dots, G\}.$$

The proof is in Appendix B.1.

4.3. Consistency in factor copula models

In this section, we state results with mild conditions for the consistency of the proxy variables in the factor copula models with known parameters. The conditions and interpretation parallel those in the preceding Section 4.2.

We next state some assumptions that are assumed throughout this section.

Assumption 1.

(a) The bivariate linking copulas have monotonic dependence, that is, the observed variables are monotonically related to the latent variables.

(b) For any fixed dimension $\boldsymbol{D}$, the log-likelihood function with latent variables considered as parameters to be estimated satisfies some standard regularity conditions, such as in [5], including log-densities of the bivariate linking copulas that are continuously differentiable with respect to v up to third order.

(c.1) In the 1-factor copula model (1), assume all linking copulas $C_{jV}(\cdot, v)$ with densities $c_{jV}(\cdot, v)$ are such that $|\partial \log c_{jV}/\partial v|$ are uniformly bounded over all j for $v_A \le v \le v_B$, where $v_A > 0$ is close to 0 and $v_B < 1$ is close to 1. In the oblique factor model (3), this is assumed with the linking copulas C_{U_{jg},V_g} replacing the C_{jV}.

(c.2) In the bi-factor copula model (2), assume the global and local linking copulas $C_{jg,V_0}(\cdot, v_0)$ with density $c_{jg,V_0}(\cdot, v_0)$ and $C_{jg,V_g;V_0}(\cdot, v_g; v_0)$ with density $c_{jg,V_g;V_0}(\cdot, v_g; v_0)$ are such that the derivatives $|\partial \log c_{jg,V_0}/\partial v_0|$ and $|\partial \log c_{jg,V_g;V_0}/\partial v_g|$ are uniformly bounded over all j, g for $v_A \le v_0, v_g \le v_B$, where v_A is close to 0 and v_B is close 1.

For the conditional expectations for factor copula models, the v's are treated as parameters and the u's are realization of independent random variables when the latent variable are fixed. The proofs make use of the Laplace approximation method.

Theorem 5 (*Consistency of Proxy in 1-Factor Copula Model with Known Linking Copulas*)**.** *For the 1-factor model in (1), suppose there is a realized infinite sequence $u_1, u_2, \ldots$, with latent variable v^0 (independent of dimension D) For the truncation to the first D variables, let $\boldsymbol{u}_D = (u_1, \ldots, u_D)^\top$. Define the averaged negative log-likelihood in parameter v as*

$$\bar{L}_D(v) = -D^{-1} \sum_{j=1}^{D} \log c_{jV}(u_j, v). \tag{20}$$

Assume $\bar{L}_D$ has a global minimum for all $D \ge d$ (some large $d > 0$) and the second derivative $\bar{L}''_D$ at the global minimum has a limit or liminf that is strictly positive. With Assumption 1, for the proxy $\tilde{v}_D$ defined in (13). then $\tilde{v}_D - v^0 = O_p(D^{-1/2})$ as $D \to \infty$.

Proof. See Appendix B.2. □

Remark 3. The theorem shows that under certain regularity conditions, the latent variables can be approximately recovered from the observed variables when they are not too close to the boundary of 0 or 1, assuming that the number of variables monotonically linked to the latent variable is large enough and that the dependence is strong enough. If the overall dependence of the c_{jV} is weak, with many copulas approaching independence ($c_{jV}(u, v) \approx u$ for many j), then it is possible that $\bar{L}_D(v)$ is a constant function in the limit. Our assumption on the limiting function of the averaged log-likelihood is mild; if all the bivariate copulas are strictly stochastically increasing, the limiting function is locally convex around the true realized value v^0, and v^0 is a global minimum of the function. If the U_j are not monotonically related to the latent variable, then it is possible for $\bar{L}_D(v)$ to have more than one local minimum. An example consists of: C_{jV} is the copula of (U_j, V) such that (i) $(U_j, 2V - 1)$ follows the Gaussian copula with parameter $\rho_j > 0$ if $1/2 \le V < 1$ and (ii) $(U_j, 2V)$ follows the Gaussian copula with parameter $-\rho_j < 0$ if $0 < V < 1/2$, and the ρ_j's are uniformly distributed in a bounded interval such as $[0.2, 0.8]$. Further remarks about this theorem are in the Supplementary material.

In the oblique factor copula model with known linking copulas, the variables in one group are linked to the same latent variable and these variables satisfy a 1-factor copula model. Hence, the assumptions and conclusion for the 1-factor copula model extend to the oblique factor model.

Corollary 6 (*ConsistEncy of Proxies in Oblique Factor Copula Model with Known Linking Copulas*)**.** *For the oblique factor model with fixed G groups in (3), suppose there are realized infinite sequences of observed variable values $\boldsymbol{u}_1^\top, \boldsymbol{u}_2^\top, \ldots, \boldsymbol{u}_G^\top$ with latent variable values $\boldsymbol{v}^0 = (v_1^0, v_2^0, \ldots, v_G^0)^\top$. Truncate the sequences to the first d_g variables u_{g,d_g} for $g \in \{1, \ldots, G\}$ with no d_g dominating. Let $\boldsymbol{u}_{\boldsymbol{D}} = (\boldsymbol{u}_{1,d_1}^\top, \boldsymbol{u}_{2,d_2}^\top, \ldots, \boldsymbol{u}_{G,d_G}^\top)^\top$. With $D = \sum_{g=1}^{G} d_g$, define the averaged negative log-likelihood in group g as $\bar{L}_D^{(g)}(v_g) = -D^{-1} \sum_{j=1}^{d_g} \log c_{U_{jg},V_g}(u_{jg}, v_g)$. Suppose Assumption 1 holds. For each g, assume $\bar{L}_D^{(g)}$ has a global minimum for $d_g \ge d$ (some large $d > 0$) and the second derivative of $\bar{L}_D^{(g)}$ at the global minimum has a limit or liminf that is strictly positive. For the proxy variable $\tilde{v}_{g\boldsymbol{D}}$ defined in (16), $\tilde{v}_{g\boldsymbol{D}} - v_g^0 = O_p(D^{-1/2})$ as $d_g \to \infty$, for $g \in \{1, 2, \ldots, G\}$.*

Proof. $\tilde{v}_{g\boldsymbol{D}}$ is the proxy by considering a 1-factor model for group g. □

Similar to the Gaussian case, consistency for the bi-factor copula model is handled differently.

Theorem 7 (*Consistency of Proxies in Bi-Factor Copula Model with Known Linking Copulas*). *For the bi-factor model with fixed G groups in (2), suppose there are realized infinite sequences of observed variable values $\boldsymbol{u}_1^\top, \boldsymbol{u}_2^\top, \ldots, \boldsymbol{u}_G^\top$ with latent variables values $\boldsymbol{v}^0 = (v_0^0, v_1^0, v_2^0, \ldots, v_G^0)^\top$. Truncate the sequences to the first d_g variables $\boldsymbol{u}_{g,d_g}$ for $g \in \{1, \ldots, G\}$. Let $\boldsymbol{u}_{\boldsymbol{D}} = (\boldsymbol{u}_{1,d_1}^\top, \ldots, \boldsymbol{u}_{G,d_G}^\top)^\top$. Let $L_{0D}(v_0; \boldsymbol{u}_{\boldsymbol{D}}) = -\log c_{\boldsymbol{U}_{\boldsymbol{D}}, V_0}(\boldsymbol{u}_{\boldsymbol{D}}, v_0)$ be the negative log-likelihood function in v_0 with observed variables $\boldsymbol{u}_{\boldsymbol{D}}$. For $g \in \{1, \ldots, G\}$, let $L_{g\boldsymbol{D}}(v_g; v_0, \boldsymbol{u}_{g,d_g}) = -\log c_{\boldsymbol{U}_g, V_0, V_g}(\boldsymbol{u}_{g,d_g}, v_0, v_g)$ be the negative log-likelihood function in v_g with observed $\boldsymbol{u}_{g,d_g}$ and given v_0. Define the averaged negative log-likelihood for marginalized density of $(\boldsymbol{U}_{\boldsymbol{D}}, V_0)$ as $\bar{L}_{0D}(v_0; \boldsymbol{u}_{\boldsymbol{D}}) = D^{-1} L_{0D}(v_0; \boldsymbol{u}_{\boldsymbol{D}})$ and the averaged negative log-likelihood for marginalized density of $(\boldsymbol{U}_{g,d_g}, V_0, V_g)$ as $\bar{L}_{g\boldsymbol{D}}(v_g; v_0, \boldsymbol{u}_{g,d_g}) = D^{-1} L_{gD}(v_g; v_0, \boldsymbol{u}_{g,d_g})$. Suppose the regularity conditions in Assumption 1 holds. Assume $\lim_{d_g \to \infty} \bar{L}_{0\boldsymbol{D}}$ has a global minimum for $D \geq d$ (some large $d > 0$) and the second derivative of $\bar{L}_{0\boldsymbol{D}}$ at the global minimum has a limit or liminf that is strictly positive. The same assumptions apply to $\bar{L}_{g\boldsymbol{D}}$. The proxy variables, $\tilde{v}_{0\boldsymbol{D}}$ defined in (14) and $\tilde{v}_{g\boldsymbol{D}}(\tilde{v}_{0\boldsymbol{D}})$ for $g \in \{1, \ldots, G\}$ defined in (15) are consistent for $v_0^0, v_1^0, \ldots, v_G^0$ respectively as $d_g \to \infty$ for all g.*

Proof. See Appendix B.2. □

Remark 4. The asymptotic properties of the unweighted average version of proxy variables are illustrated in [17]. Under some mild assumptions, the joint distribution of the observed variables and these simpler proxy variables converges to the underlying joint distribution as the dimension goes to infinity. Here, the conditional expectation version of proxies are weighted averages of observed variables in Gaussian factor models and non-linear transforms of observed variables in factor copula models. The weak consistency result of the conditional expectation proxy variable is stronger than the results in [17] as it also provides the convergence rate of the proxy variables in the 1-factor and oblique factor copula models. Additionally, our method can handle bi-factor models for which the unweighted average proxies are not valid.

5. Consistency of proxies with estimated parameters

With a parametric model for the loading matrix in the Gaussian factor models, and parametric bivariate linking copulas in the factor copula models, the parameters can be estimated and then the proxies in Section 3 can be applied with the usual plug-in method. For factor copulas, numerical integration would be needed to evaluate the proxies for a random sample of size N.

In this section, we prove the consistency of proxies with estimated parameters in two steps. For step 1, we prove the equations for the proxy variables are locally Lipschitz in the parameters (even as number of parameters increase as D increases). For step 2, under the assumption that the parameters can be considered as a sample from a super-population, all parameters can be estimated with $O(N^{-1/2})$ accuracy.

The theoretical results of this section and the preceding section support the use of proxies as estimates of latent variables for D large enough, and N large enough in factor models. The practicality of $D \to \infty$ is that the case of $D > N$ can be handled, with N large enough for reasonable estimates of parameters. The next Section 6 outlines a sequential method for determining proxies for factor copulas.

5.1. Gaussian factor models

Lemma 8. *In the p-factor model (7), let the loading matrix be $\boldsymbol{A}^\top = (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_D, \ldots)$, and let $\widehat{\boldsymbol{A}}^\top = (\hat{\boldsymbol{a}}_1, \ldots, \hat{\boldsymbol{a}}_D, \ldots)$ (infinite sequence) be a perturbation of $\boldsymbol{A}^\top$, where $\hat{\boldsymbol{a}}_j$ and $\boldsymbol{a}_j$ denotes the jth column of matrix $\widehat{\boldsymbol{A}}^\top$ and $\boldsymbol{A}^\top$ respectively, $j \in \{1, 2, \ldots\}$. Specifically, $\widehat{\boldsymbol{A}}^\top$ satisfies $\sqrt{D^{-1} \sum_{j=1}^D \|\hat{\boldsymbol{a}}_j - \boldsymbol{a}_j\|^2} \leq \rho$ for a small positive ρ. Let $\boldsymbol{z} = (z_1, \ldots, z_D, \ldots)$ be one realization from the factor model with loading matrix $\boldsymbol{A}$ and suppose $\limsup D^{-1} \sum_{j=1}^D z_i^2 < \infty$. Let $\boldsymbol{A}_D$ be $\boldsymbol{A}$ truncated to D rows, and similarly define $\widehat{\boldsymbol{A}}_D$. Suppose there is a positive constant $K > 0$ such that $\psi_j > K$ and $\hat{\psi}_j > K$ for all j. Consider the factor scores vector in (9) as a function of the loading matrix: $\widetilde{\boldsymbol{w}}_D = \widetilde{\boldsymbol{w}}_D(\boldsymbol{A}_D)$. Let $\boldsymbol{Q}_D = \boldsymbol{A}_D^\top \boldsymbol{\Psi}_D^{-2} \boldsymbol{A}_D$, $\widehat{\boldsymbol{Q}}_D = \widehat{\boldsymbol{A}}_D^\top \widehat{\boldsymbol{\Psi}}_D^{-2} \widehat{\boldsymbol{A}}_D$. Suppose $D^{-1}\boldsymbol{Q}_D \to \boldsymbol{Q}$, $D^{-1}\widehat{\boldsymbol{Q}}_D \to \widehat{\boldsymbol{Q}}$ where $\boldsymbol{Q}, \widehat{\boldsymbol{Q}}$ are both positive definite and well-conditioned matrices. Then $\|\widetilde{\boldsymbol{w}}_D(\widehat{\boldsymbol{A}}_D) - \widetilde{\boldsymbol{w}}_D(\boldsymbol{A}_D)\| \leq K_D \cdot \sqrt{D^{-1} \sum_{j=1}^D \|\hat{\boldsymbol{a}}_j - \boldsymbol{a}_j\|^2}$ for a constant K_D. The sequence $\{K_D\}$ is bounded as $D \to \infty$ if the $\boldsymbol{a}_j$'s are assumed to be sampled from a super-population with the corresponding ψ_j^2 bounded away from 0.*

Proof. See Appendix B.3. □

5.2. Parametric copula factor models

For parametric models we assume that there is a parameter representing monotone dependence associated with each bivariate linking copula, so that generically, the copulas of form $c_{U_s,V}(u, v)$ in Section 2 are now written as $c_{U_s,V}(u, v; \boldsymbol{\theta}_s)$. We next clarify the notation of the increasing parameters in the considered factor copula models.

1. For the 1-factor copula model (1), the parameter vector of linking copulas is the infinite sequence $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots)$.
2. For the bi-factor copula model (2) with G groups, let the global linking copula densities be $c_{U_{jg},V_0}(u_{jg}, v_0; \boldsymbol{\theta}_{jg,0})$ in group g, $j \in \{1, 2, \ldots\}$. Let the local linking copula densities be $c_{U_{jg}V_g;V_0}(C_{U_{jg}|V_0}(u_{jg}|v_0), v_g; \boldsymbol{\theta}_{jg})$ in group g, $j \in \{1, 2, \ldots\}$. The vectors of parameters are $\boldsymbol{\theta}_g^{(1)} = (\boldsymbol{\theta}_{1g,0}, \boldsymbol{\theta}_{2g,0}, \ldots)$ for global linking copulas and $\boldsymbol{\theta}_g^{(2)} = (\boldsymbol{\theta}_{1g}, \boldsymbol{\theta}_{2g}, \ldots)$, $g \in \{1, \ldots, G\}$ for local linking copulas. The concatenated parameter vectors over groups are $\boldsymbol{\theta}^{(1)} = (\boldsymbol{\theta}_1^{(1)}, \ldots, \boldsymbol{\theta}_G^{(1)})$ and $\boldsymbol{\theta}^{(2)} = (\boldsymbol{\theta}_1^{(2)}, \ldots, \boldsymbol{\theta}_G^{(2)})$. $\boldsymbol{\theta} = (\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)})$.

We state some regularity assumptions on the linking copulas that guarantee the Lipschitz continuity.

Assumption 2.

(a) In the 1-factor copula model, for any fixed dimension D, the partial derivatives $\partial^k \log c_{jV}(u_j, v; \theta_j)/\partial v^k$ exist for each j, where $k \in \{1, 2, 3\}$. Similarly, in the bi-factor copula model, for any fixed dimension $\boldsymbol{D}$, the partial derivatives $\partial^k \log c_{\boldsymbol{U_D},V_0}(\boldsymbol{u_D}, v_0, \boldsymbol{\theta_D})/\partial v_0^k$, and $\partial^k \log c_{\boldsymbol{U}_g,V_g;V_0}(\boldsymbol{u}_g, v_g; v_0, \boldsymbol{\theta}_{g,d_g})/\partial v_g^k$ exist, for $k \in \{1, 2, 3\}$, $g \in \{1, \dots, G\}$.

(b) The mixed partial derivatives of log linking copula densities with respect to $\boldsymbol{\theta}$ and v (up to the second-order) exist and are uniformly bounded over the range of parameters $\boldsymbol{\theta}$. In the 1-factor copula model, the mixed partial derivatives $\partial^{k+1} \log c_{jV}(u_j, v; \boldsymbol{\theta}_j)/\partial \boldsymbol{\theta}_j \partial v^k$ exist for each j, where $k \in \{0, 1, 2\}$. The sequences of partial derivatives are uniformly bounded for $k \in \{0, 1, 2\}$, and $v_A \le v \le v_B$ where $v_A > 0$ is close to 0 and $v_B < 1$ is close to 1.
In the bi-factor copula model, there are more mixed partial derivatives as there are global and local linking copulas taking derivatives with respect to combinations of two parameter vectors. The mixed partial derivatives $\partial^{k+1} \log c_{\boldsymbol{U_D},V_0}(\boldsymbol{u_D}, v_0, \boldsymbol{\theta_D})/\partial \boldsymbol{\theta}_s \partial v_t^k$ and $\partial^{k+1} \log c_{\boldsymbol{U}_g,V_g;V_0}(\boldsymbol{u}_g, v_g; v_0, \boldsymbol{\theta}_{g,d_g})/\partial \boldsymbol{\theta}_s \partial v_{t^*}^k$ for $s = jg, 0$ or $s = jg$, $t = 0$ and $t^* = g$ exist, where $k \in \{0, 1, 2\}$. The four sequences of partial derivatives are uniformly bounded for $k \in \{0, 1, 2\}$, and $v_A \le v_0, v_g \le v_B$ where $v_A > 0$ is close to 0 and $v_B < 1$ is close to 1, $g \in \{1, \dots, G\}$.

(c) The parameter vectors are all in bounded space such that all linking copulas are bounded away from comonotonicity and countermonotonicity.

Remark 5. The conditions (a,b) are satisfied by all commonly used parametric bivariate copulas with monotone dependence. Assumption 2(c) indicates there are no observed variables that approach perfect dependence with the latent variable. If the contrary holds, the conditional expectation proxy is not necessary as one or more of the observed variables can be considered as the proxy.

Lemma 9. *Consider a 1-factor copula model (1) with parametric linking copulas that have monotone dependence and the bivariate linking copulas satisfy Assumption 1. Let the parameter vector of linking copulas be the infinite sequence $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_D, \dots)$. Let $\hat{\boldsymbol{\theta}}$ be a perturbation of $\boldsymbol{\theta}$. Suppose the linking copulas with parameter vectors $\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}$ satisfy Assumption 2. Let $\boldsymbol{u} = (u_1, \dots, u_D, \dots)$ be one realization generated from model with $\boldsymbol{\theta}$. Let $\boldsymbol{\theta}_D$ (or $\hat{\boldsymbol{\theta}}_D$), $\boldsymbol{u}_D$ be truncated to the first D linking copulas and variables. Consider the proxy in (13) as a function of $\boldsymbol{\theta}_D$: $\tilde{v}_D = \tilde{v}_D(\boldsymbol{\theta}_D)$. Suppose $\hat{\boldsymbol{\theta}}_D \in \bar{\boldsymbol{B}}(\boldsymbol{\theta}_D, \rho)$ (ball of sufficiently small radius $\rho > 0$); that is, $\sqrt{D^{-1} \sum_{j=1}^D \|\hat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j\|_2^2} \le \rho$. Assume $\bar{L}_D$ in (20) has a global minimum for all $D \ge d$ (some large $d > 0$) and the second derivative $\bar{L}_D''$ at the global minimum has a limit or liminf that is strictly positive. Then there exist a constant B_D such that*

$$\|\tilde{v}_D(\hat{\boldsymbol{\theta}}_D) - \tilde{v}_D(\boldsymbol{\theta}_D)\| \le B_D \|\hat{\boldsymbol{\theta}}_D - \boldsymbol{\theta}_D\|^*$$

where $\|\hat{\boldsymbol{\theta}}_D - \boldsymbol{\theta}_D\|^* := \sqrt{D^{-1} \sum_{j=1}^D \|\hat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j\|_2^2}$. The sequence $\{B_D\}$ is bounded as $D \to \infty$ if the $\boldsymbol{\theta}_j$'s are assumed to be sampled from a super-population satisfying Assumption 2(c).

Proof. See Appendix B.3. □

Lemma 10. *Consider a G-group bi-factor copula model (2) with parametric linking copulas satisfy Assumption 1. Let the parameter vectors be $\boldsymbol{\theta}^{(1)} = (\boldsymbol{\theta}_1^{(1)}, \dots, \boldsymbol{\theta}_G^{(1)})$ and $\boldsymbol{\theta}^{(2)} = (\boldsymbol{\theta}_1^{(2)}, \dots, \boldsymbol{\theta}_G^{(2)})$ with perturbations $\hat{\boldsymbol{\theta}}^{(1)}$ and $\hat{\boldsymbol{\theta}}^{(2)}$. Suppose the global and linking copulas have parameter vectors satisfying Assumption 2. Let $\boldsymbol{D} = (d_1, \dots, d_G)$ and $\boldsymbol{u}_{\boldsymbol{D}}^\top = (\boldsymbol{u}_1^\top, \boldsymbol{u}_2^\top, \dots, \boldsymbol{u}_G^\top)$ be a truncation of infinite-dimensional realization from the model with parameters $\boldsymbol{\theta_D} = (\boldsymbol{\theta}_{\boldsymbol{D}}^{(1)}, \boldsymbol{\theta}_{\boldsymbol{D}}^{(2)})$, where $\boldsymbol{\theta}_{\boldsymbol{D}}^{(1)}$, $\boldsymbol{\theta}_{\boldsymbol{D}}^{(2)}$ consist of $\boldsymbol{\theta}^{(1)}$, $\boldsymbol{\theta}^{(2)}$ truncated to the first d_g in group g. Let $\boldsymbol{\theta}_{g,d_g} = (\boldsymbol{\theta}_{g,d_g}^{(1)}, \boldsymbol{\theta}_{g,d_g}^{(2)})$, where $\boldsymbol{\theta}_{g,d_g}^{(1)}$, $\boldsymbol{\theta}_{g,d_g}^{(2)}$ consist of $\boldsymbol{\theta}_g^{(1)}$, $\boldsymbol{\theta}_g^{(2)}$ truncated to the first d_g in group g, for $g \in \{1, \dots, G\}$. Similarly, define $\hat{\boldsymbol{\theta}}_{\boldsymbol{D}}, \hat{\boldsymbol{\theta}}_{g,d_g}$. Let $\tilde{v}_{0\boldsymbol{D}} = \tilde{v}_{0\boldsymbol{D}}(\boldsymbol{\theta_D})$, $\tilde{v}_{g\boldsymbol{D}} = \tilde{v}_{g\boldsymbol{D}}(\boldsymbol{\theta_D})$ be the proxies defined in (14) and (15). Suppose $\hat{\boldsymbol{\theta}}_{\boldsymbol{D}} \in \left\{ \bar{\boldsymbol{B}}(\boldsymbol{\theta_D}, \rho) : \sqrt{(2D)^{-1} \sum_{j=1}^D \|\hat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j\|_2^2} \le \rho \right\}$. Let L_{0D} and L_{gD} ($g \in \{1, \dots, G\}$) be defined as in Theorem 7 and suppose they satisfy the conditions in the statement of this theorem. Then there exists constants $B_{\boldsymbol{D}}$ and $B_{\boldsymbol{D}}^*$ such that:*

$$\|\tilde{v}_{0\boldsymbol{D}}(\hat{\boldsymbol{\theta}}_{\boldsymbol{D}}) - \tilde{v}_{0\boldsymbol{D}}(\boldsymbol{\theta_D})\| \le B_{\boldsymbol{D}} \|\hat{\boldsymbol{\theta}}_{\boldsymbol{D}} - \boldsymbol{\theta_D}\|^*, \quad \|\tilde{v}_{g\boldsymbol{D}}(\hat{\boldsymbol{\theta}}_{\boldsymbol{D}}) - \tilde{v}_{g\boldsymbol{D}}(\boldsymbol{\theta_D})\| \le B_{\boldsymbol{D}}^* \|\hat{\boldsymbol{\theta}}_{g,d_g} - \boldsymbol{\theta}_{g,d_g}\|^*, \quad g \in \{1, \dots, G\},$$

where
$\|\hat{\boldsymbol{\theta}}_{\boldsymbol{D}} - \boldsymbol{\theta_D}\|^* := \sqrt{(2D)^{-1} \sum_{g=1}^G \sum_{j=1}^{d_g} (\|\hat{\boldsymbol{\theta}}_{jg} - \boldsymbol{\theta}_{jg}\|_2^2 + \|\hat{\boldsymbol{\theta}}_{jg,0} - \boldsymbol{\theta}_{jg,0}\|_2^2)}$*, and* $\|\hat{\boldsymbol{\theta}}_{g,d_g} - \boldsymbol{\theta}_{g,d_g}\|^* := \sqrt{(2d_g)^{-1} \sum_{j=1}^{d_g} (\|\hat{\boldsymbol{\theta}}_{jg} - \boldsymbol{\theta}_{jg}\|_2^2 + \|\hat{\boldsymbol{\theta}}_{jg,0} - \boldsymbol{\theta}_{jg,0}\|_2^2)}$. *The sequences $\{B_D\}$ and $\{B_D^*\}$ are bounded as $d_g \to \infty$ for all g if the $\boldsymbol{\theta}_j$'s are assumed to be sampled from a super-population satisfying Assumption 2(c).*

Proof. The proof is similar to that of Lemma 9, see the Supplementary material for the detailed proof. □

5.3. Estimation of parameters in blocks

Since the general Gaussian factor model is non-identifiable in terms of orthogonal rotation of the loading matrix, for a model with two or more factors, consistency of proxy variables requires a structured loading matrix such as that of the bi-factor model or oblique factor model. In these structured factor-models, the identification restrictions to eliminate the indeterminacy of the rotation like IC2 in [3] can be avoided.

In order to have consistent estimates of parameters, an assumption is needed on the behavior as the number of variables increases to ∞. A realistic assumption is that the observed variables (or their correlations, partial correlations, linking copulas) are sampled from a super-population. Then block estimation of parameters is possible, with a finite number of parameters in each block. In the structured factor copula models with known parametric family for each linking copula, the parameters are estimated via sequential maximum likelihood with variance of order $O(1/N)$ for sample size N. We unify the notation of observed variables to be X; in Gaussian models, $\boldsymbol{X} = \boldsymbol{Z}$, and in copula models, $\boldsymbol{X} = \boldsymbol{U}$. Assume that the data from a random sample with D variables are $\{\boldsymbol{X}_i : 1 \le i \le N\}$.

The idea behind block estimation is that, with the super-population assumption, the factor models are closed under margins (the same latent variables apply to different margins), and parameters can be estimated from appropriate subsets or blocks. Therefore, it is sufficient to guarantee that factor models can be identified in each block. Standard maximum likelihood theory for a finite number of parameters can be applied in each block, and there is no need to develop theory for simultaneous estimates of all parameters with the number of parameters increasing to ∞.

The identifiability conditions in [2] for 1-factor model, [9] for bi-factor model, and [14] for oblique-factor model can be applied to each block. Suppose the factor models have moderate to strong dependence and have a large number of variables in each group (for 1-factor model, there is only one group). With the assumption that the number of groups are fixed and the groups are ordered in a fixed index, the loading matrix with structured zero loadings in each block can be identified up to signs.

The block estimation procedures are explained for the 1-factor, bi-factor and oblique factor models. The observed variables are split into several blocks and each block is a marginal factor model linking to the same latent variables. Estimates of parameters in different blocks are concatenated. In the Gaussian case, the maximum likelihood (ML) estimates of factor loadings are unique up to signs, the signs of the estimates in each block can be adjusted appropriately. The number of blocks K_D increases with D. The Supplementary material has a few details to show the form of the estimating equations when block estimation is used.

1. (1-factor model): For $j \in \{1, 2, \ldots, D\}$, split D variables into K_D blocks of approximate size $B > 5$ in a sequential way, the partition $\mathcal{B} = \{\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_{K_D}\}$ where the cardinality of $\mathcal{B}_k$ is $B_k \approx B$ for $k \in \{1, \ldots, K_D\}$. This leads to K_D marginal 1-factor models with the same latent variable. For the convenience of determining the signs of estimated parameters (or positive or negative dependence in the linking copulas), add the first variable X_1 in block $\mathcal{B}_2, \ldots, \mathcal{B}_{K_D}$. The estimates of the parameter associated with X_1 can be averaged over the blocks. The estimated parameters are $(\hat{\boldsymbol{\theta}}_{\mathcal{B}_1}, \hat{\boldsymbol{\theta}}_{\mathcal{B}_2}, \cdots, \hat{\boldsymbol{\theta}}_{\mathcal{B}_{K_D}})$ after adjusting the signs of estimated parameters in each block.
2. (Oblique factor model): Under the assumption of oblique factor model, there are G groups with d_g dependent variables in the gth group. For each group, split d_g variables into K_D blocks of approximate size $d_g^{(k)}$ dependent variables for $k \in \{1, \ldots, K_D\}$. Keep the ratio of size of G groups invariant in each block when splitting, that is $d_g^{(k)}/d_h^{(k)} \approx d_g/d_h$ for $g \neq h$, $g, h \in \{1, 2, \ldots, G\}$ in block k. The partition gives K_D blocks, for block k, $\{\boldsymbol{X}_{ij\in\mathcal{B}_k} : j \in \{\mathcal{B}_k^{(1)}, \ldots, \mathcal{B}_k^{(g)}, \ldots, \mathcal{B}_k^{(G)}\};\quad i \in \{1, 2, \ldots, N\}\}$ where $\mathcal{B}_k^{(g)}$ denotes the kth block in the gth group. For the convenience of determining the signs (or positive or negative dependence in the linking copulas), add G auxiliary variables which are the first variable in G groups for block $\mathcal{B}_1$ to groups in blocks $\mathcal{B}_2, \ldots, \mathcal{B}_{K_D}$. Suppose the estimates of parameters involving the variables in block k of each group are $\hat{\boldsymbol{\theta}}_{\mathcal{B}_k}^{(1)}, \hat{\boldsymbol{\theta}}_{\mathcal{B}_k}^{(2)}, \ldots, \hat{\boldsymbol{\theta}}_{\mathcal{B}_k}^{(G)}$, for $k \in \{1, \ldots, K_D\}$. For parameters that are estimated over different blocks, such as $\boldsymbol{\Sigma}_{\boldsymbol{W}}$, an average could be taken over the different blocks.
3. (Bi-factor model): The block method to estimate parameters in bi-factor model is similar as that used in oblique factor model. The auxiliary variables that help to determine the signs of parameters are now the first variable X_1 and $G-1$ variables which are the first variable in group g, $g \in \{2, \ldots, G\}$ for block $\mathcal{B}_1$. The variables are added to G groups in blocks $\mathcal{B}_2, \ldots, \mathcal{B}_{K_D}$.

The remaining results in this section assume one of the block estimation procedures.

Lemma 11 (*Asymptotic Properties of Estimated Parameters in 1-Factor, Bi-Factor and Oblique Factor Models*)**.** *Suppose there is a sample of size N. For the 1-factor model in (1) and (4), bi-factor model in (2) and (5) and oblique factor model in (3) and (6), for any fixed dimension D, let $\boldsymbol{\theta}_D$ be the parameter vector in the factor models and $\hat{\boldsymbol{\theta}}_D$ be the corresponding estimates of the parameters, using the block estimation method. Suppose parameters from the linking copulas behave like a sample from a super-population (bounded parameter space, bounded away from comonotonicity/countermonotonicity), then*

$$\hat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j = O_p(N^{-1/2}), \quad \textit{uniformly for } j \in \{1, \ldots, D\}.$$

Proof. For all factor models in Section 2, the block method gives a partition of variables into several blocks. In each block, they are marginal factor models. In the Gaussian models, the results of maximum likelihood estimation [1] could be applied in each block. In the copula factor models, asymptotic maximum likelihood theory could be applied under standard regularity conditions. The super-population and other regularity assumptions imply that the expected Fisher information matrices (and standard errors) can be uniformly bounded over different block sizes. □

Combined with the previous results, we show the consistency of proxy variables with $N, D \to \infty$. Due to the consistency of proxy variable with the known parameters, with the Lipschitz inequalities, the consistency still hold when the parameters are estimated, as both $N, D \to \infty$. The results in Gaussian and factor copula models are similar, so the results are only stated in the copula case.

Theorem 12 (*Consistency of Proxies in 1-Factor, Bi-Factor and Oblique Factor Copula Models*)**.** *Suppose the 1-factor, bi-factor and oblique factor copula models satisfy Assumption 1 and the regularity conditions in Theorems 5, 7, and Corollary 6 respectively. Let the parameters θ be considered as a sample from a super-population satisfying Assumption 2. Suppose $\hat{\theta}_j$ is an estimate of θ_j. Assume the factor models are identifiable with respect to parameters, and $\|\hat{\theta}_j - \theta_j\| = O_p(N^{-1/2})$ for all j. Let $\theta_{\boldsymbol{D}}$ be θ truncated to the first $\boldsymbol{D}$ random variables u_j's, then as $\boldsymbol{D} \to \infty$, the following hold.*

(i) For 1-factor copula model, let $(U_{i1}, U_{i2}, \ldots, U_{iD}, \ldots, V_i)$ be a random infinite sequence for $i \in \{1, \ldots, N\}$. With $\widetilde{V}_D(\hat{\theta}_D)$ being the proxy random variable, $\|\widetilde{V}_{iD}(\hat{\theta}_D) - V_i\| = o_p(1)$.

(ii) For bi-factor copula model with G groups, let $(\boldsymbol{U}_{i1}^\top, \ldots, \boldsymbol{U}_{ig}^\top, \ldots, \boldsymbol{U}_{iG}^\top, V_{i0}, V_{i1}, \ldots, V_{iG})$ be a random infinite sequence for $i \in \{1, \ldots, N\}$, where $\boldsymbol{U}_{ig}^\top = (U_{i1g}, U_{i2g}, \ldots, U_{id_g g})$. With $\widetilde{V}_{0\boldsymbol{D}}(\hat{\theta}_{\boldsymbol{D}}), \widetilde{V}_{1\boldsymbol{D}}(\hat{\theta}_{\boldsymbol{D}}), \ldots, \widetilde{V}_{G\boldsymbol{D}}(\hat{\theta}_{\boldsymbol{D}})$ for the proxy variables, then $\|\widetilde{V}_{ig,\boldsymbol{D}}(\hat{\theta}_{\boldsymbol{D}}) - V_{ig}\| = o_p(1)$ for $g \in \{0, 1, \ldots, G\}$.

(iii) For oblique factor model with G groups, let $(\boldsymbol{U}_{i1}^\top, \ldots, \boldsymbol{U}_{ig}^\top, \ldots, \boldsymbol{U}_{iG}^\top, V_{i1}, \ldots, V_{iG})$ be a random infinite sequence for $i \in \{1, \ldots, N\}$, where $\boldsymbol{U}_{ig}^\top = (U_{i1g}, U_{i2g}, \ldots, U_{id_g g})$. With $\widetilde{V}_{1\boldsymbol{D}}(\hat{\theta}_{\boldsymbol{D}})$, ..., $\widetilde{V}_{G\boldsymbol{D}}(\hat{\theta}_{\boldsymbol{D}})$ for the proxy variables, then $\|\widetilde{V}_{ig,\boldsymbol{D}}(\hat{\theta}_{\boldsymbol{D}}) - V_{ig}\| = o_p(1)$, $g \in \{1, \ldots, G\}$.

Proof. (1) For 1-factor copula model, based on triangle inequality,

$$\|\widetilde{V}_{iD}(\hat{\theta}_D) - V_i\| = \|\widetilde{V}_{iD}(\hat{\theta}_D) - \widetilde{V}_{iD}(\theta_D) + \widetilde{V}_{iD}(\theta_D) - V_i\| \le \|\widetilde{V}_{iD}(\hat{\theta}_D) - \widetilde{V}_{iD}(\theta_D)\| + \|\widetilde{V}_{iD}(\theta_D) - V_i\|$$

For the first term in the right-hand side, Lemma 9 implies that it has the same order as $O_p\left(\sqrt{D^{-1}\sum_{j=1}^D \|\hat{\theta}_j - \theta_j\|^2}\right) = O_p(N^{-1/2})$. By Theorem 5, the second term is of order $O_p(D^{-1/2})$, then $\|\widetilde{V}_{iD}(\hat{\theta}_D) - V_i\| = o_p(1)$ as $D, N \to \infty$. For cases (2) and (3), the same proof technique can be applied in bi-factor and oblique factor model and the details are omitted. □

Remark 6. In the proof for consistency of the proxy variables, we assume univariate margins are known. For Gaussian factor models with margins, we assume μ's and σ's are known or estimated before transforming to standard normal. For Gaussian factor dependence models and non-Gaussian margins, we assume univariate cdfs are known or have been estimated before transforming to standard normal. For factor copulas models, we assume univariate cdfs are known or have been estimated before transforming to $U(0,1)$. In practice, proxies are estimated after the estimation of univariate margins, and there is one more source of variability beyond what we have considered here.

6. Sequential estimation for parametric factor copula models

In this section, sequential methods are suggested for estimating the latent variables and the parameters of the linking copulas, allowing for choice among several candidate families for each observed variable. Preliminary diagnostic plots can help to check for deviations for the Gaussian copula in terms of tail dependence or tail asymmetry ([15]; Chapter 1 of [11]).

For high-dimensional multivariate data for which initial data analysis and the correlation matrix of normal scores suggest a copula dependence structure of 1-factor, bi-factor or oblique factor, a sequential procedure is presented to estimate the latent variables with proxies, decide on suitable families of linking copulas, and estimate parameters of the linking copulas without numerical integration. Suppose the parametric linking copula families are not known or specified in advance (the situation in practice), the sequential method starts with unweighted averages estimates in [17] or factor scores computed from an estimated loading matrix after observed variables are transformed to have $N(0,1)$ margins. Then, the "conditional expectation" proxies are constructed and are used to estimate the parameters by optimizing the approximate (complete) log-likelihood with the latent variables assumed observed at the values of the proxy variables. The copula density which includes latent variables does not require the integrals in Section 2. More details are illustrated below.

Let the random vector from the 1-factor copula model in (1) be $\boldsymbol{U}_D = (U_1, \ldots, U_D)$. Suppose there is sample of size N from the model, we denote the ith observation which is the realization of $\boldsymbol{U}_D$ of length D as $\boldsymbol{u}_i = (u_{i1}, \ldots, u_{iD})$. Let the random vector from the bi-factor copula model in (2) or oblique factor copula model in (3) be $\boldsymbol{U}_{\boldsymbol{D}}$, and $\boldsymbol{U}_{\boldsymbol{D}} = (\boldsymbol{U}_{1,d_1}^\top, \ldots, \boldsymbol{U}_{G,d_G}^\top)$ where $\boldsymbol{U}_{g,d_g} = (U_{1g}, \ldots, U_{d_g g})$. We denote the ith observation which is the realization of $\boldsymbol{U}_{\boldsymbol{D}}$ as $\boldsymbol{u}_i = (\boldsymbol{u}_{i,1}^\top, \ldots, \boldsymbol{u}_{i,G}^\top)$ (the dependence on d_g in $\boldsymbol{u}_{ig,d_g}$ is suppressed for simplicity.)

For the 1-factor copula model, if the latent variables are assumed observed, then the complete log-likelihood is

$$\sum_{i=1}^N \log c_{U_{1:D},V}(u_{i1}, \ldots, u_{iD}, v_i; \theta_D) = \sum_{i=1}^N \sum_{j=1}^D \log c_{jV}(u_{ij}, v_i; \theta_j). \tag{21}$$

- Stage 1: Define the "unweighted average" proxy variable as $U_0 = P_D^U(D^{-1}\sum_{j=1}^D U_j)$, where P_D^U is the cdf of $\bar{U}_D := D^{-1}\sum_{j=1}^D U_j$. With enough dependence, $\bar{U}_D$ does not converge in probability to a constant. An extreme case consists of bivariate linking copulas all near the independence copula, with $\bar{U}_D \to 1/2$ as $D \to \infty$. For case i, let $\bar{u}_i = D^{-1}\sum_{j=1}^D u_{ij}$ and $u_{i,0} = [\text{rank}(\bar{u}_i) - 0.5]/N$; $\text{rank}(\bar{u}_i)$ is defined as the rank of $\bar{u}_i$ based on $\bar{u}_1, \ldots, \bar{u}_N$. Substitute $v_i = u_{i,0}$ in the log-likelihood (21), and obtain the first-stage estimates of the parameters in θ_D from the approximate log-likelihood. This is the method of [17].
- Stage 2: Construct the conditional expectations proxies based on (13) with the first-stage estimated parameters of θ_D. One-dimensional Gauss–Legendre quadrature can be used. Denote the conditional expectation proxy variable as $\widetilde{U}_0$. Substitute the ith proxy $v_i = \widetilde{u}_{i,0}$ which are realizations of $\widetilde{U}_0$ in the log-likelihood (21), obtain the second-stage estimates of the parameters from the approximate log-likelihood.

For the bi-factor copula model, if the latent variables are assumed observed, then the complete log-likelihood is

$$\begin{aligned}&\sum_{i=1}^{N}\log c_{\boldsymbol{U}_{1:D},V_0,\boldsymbol{V}}(\boldsymbol{u}_{i,1}^\top,\ldots,\boldsymbol{u}_{i,G}^\top,v_{i,0},v_{i,1},\ldots,v_{i,G};\boldsymbol{\theta})\\&=\sum_{i=1}^{N}\sum_{g=1}^{G}\sum_{j=1}^{d_g}\Big[\log c_{U_{jg},V_0}(u_{i,jg},v_{i,0};\boldsymbol{\theta}_{jg,0})+\log c_{U_{jg},V_g;V_0}(C_{U_{jg}|V_0}(u_{i,jg}|v_{i,0}),v_{i,g};\boldsymbol{\theta}_{jg})\Big].\end{aligned}\tag{22}$$

Suppose the bi-factor structure is known, i.e., the number of groups and the number of variables in each group, estimation can be performed in two stages.

- Stage 1: Assume the variables are monotonically related and that the Gaussian copula is reasonable as a first-order model. Convert data into normal scores and fit a Gaussian model with bi-factor structure. Compute the factor scores in (10) and (11), and denote as $\widetilde{w}_{i0},\widetilde{w}_{i1},\ldots,\widetilde{w}_{iG}$. Let the random version of factor scores be $\widetilde{W}_{i0},\widetilde{W}_{i1},\ldots,\widetilde{W}_{iG}$. The first-stage proxy variable are defined as $\widetilde{V}_0^{(1)}=P_{D,0}^U(\widetilde{W}_0)$, $\widetilde{V}_g^{(1)}=P_{d_g,g}^U(\widetilde{W}_g)$, $g\in\{1,\ldots,G\}$; where $P_{D,0}^U$ is the cdf of $\widetilde{W}_0$, and $P_{d_g,g}^U$ is the cdf of $\widetilde{W}_g$, $g\in\{1,\ldots,G\}$. Letting $v_{i,0}=\widetilde{v}_{i,0}^{(1)}$, $v_{i,g}=\widetilde{v}_{i,g}^{(1)}$ which are the realizations of $\widetilde{V}_0^{(1)}$ and $\widetilde{V}_g^{(1)}$ respectively in log-likelihood (22), obtain the first-stage estimates of the parameters from the approximate log-likelihood.
- Stage 2: Construct the conditional expectation proxies based on Eqs. (14) and (15) with first-stage estimates plugged in. Nested 1-dimensional Gauss–Legendre quadrature can be used. Denote the conditional expectation proxy variables as $\widetilde{V}_0^{(2)}$, $\widetilde{V}_g^{(2)}$, $g\in\{1,\ldots,G\}$. Letting $v_{i,0}=\widetilde{v}_{i,0}^{(2)}$, $v_{i,g}=\widetilde{v}_{i,g}^{(2)}$ which are the realizations of $\widetilde{V}_0^{(2)}$ and $\widetilde{V}_g^{(2)}$ respectively in log-likelihood (22), obtain the second-stage estimates of the parameters from the approximate log-likelihood.

For the oblique factor copula model, if the latent variables are assumed observed, the complete log-likelihood is

$$\begin{aligned}&\sum_{i=1}^{N}\log c_{\mathrm{oblique},\boldsymbol{U}_{1:D},V_{1:G}}(\boldsymbol{u}_{i,1}^\top,\ldots,\boldsymbol{u}_{i,G}^\top,v_{i,1},\ldots,v_{i,G};\boldsymbol{\theta})\\&=\sum_{i=1}^{N}\left\{\sum_{g=1}^{G}\sum_{j=1}^{d_g}\log c_{U_{jg},V_g}(u_{i,jg},v_{i,g};\boldsymbol{\theta}_{jg})+\log c_{\boldsymbol{V}}(v_{i,1},\ldots,v_{i,G};\boldsymbol{\theta}_{\boldsymbol{V}})\right\},\end{aligned}\tag{23}$$

where $\boldsymbol{\theta}=(\boldsymbol{\theta}_{d_1,1}^\top,\ldots,\boldsymbol{\theta}_{d_G,G}^\top,\boldsymbol{\theta}_V^\top)$, and $c_{\boldsymbol{V}}$ is the copula density of the latent variables.

- Stage 1: For $g\in\{1,\ldots,G\}$, let $\bar{U}_g=P_{d_g,g}^U(d_g^{-1}\sum_{j=1}^{d_g}U_{jg})$, where $P_{d_g,g}^U$ is the cdf of $\bar{U}_g:=d_g^{-1}\sum_{j=1}^{d_g}U_{jg}$. For the ith case, $\bar{u}_{i,g}=d_g^{-1}\sum_{j=1}^{d_g}u_{i,jg}$ and $u_{i,g}=[\mathrm{rank}(\bar{u}_{i,g})-0.5]/N$, where $\mathrm{rank}(\bar{u}_{i,g})$ is defined as the rank of $\bar{u}_{i,g}$ based on $\bar{u}_{1,g},\ldots,\bar{u}_{N,g}$. Let $v_{i,g}=u_{i,g}$ in log-likelihood (23) to get the first-stage estimates of the parameters from the approximate log-likelihood. This is the method of [17].
- Stage 2: Construct the conditional expectations proxies based on (16) with the first-stage estimated parameters. his requires 1-dimensional numerical integration. Denote the proxies as $\widetilde{U}_g$. Substitute $v_{i,g}=\widetilde{u}_{i,g}$ which are the realizations of $\widetilde{U}_g$ in the log-likelihood (23), and obtain the second-stage estimates of the parameters.

For 1-factor and oblique factor models, unweighted averages can be consistent under some mild conditions [17], but the above methods based on conditional expectations perform better from the simulation results shown in the Supplementary material. The estimation of proxies and copula parameters could be iterated further if the stage 2 estimates differ a lot from stage 1 estimates. In our simulation examples, the two-step sequential procedure is adequate to provide good estimates of the latent variables.

For optimizing the above approximate log-likelihoods using proxies for the latent variables, we adopt a modified Newton–Raphson algorithm with analytic derivatives; see [16] for details of the numerical implementation.

7. Factor models with residual dependence

It is important that we can show that proxy estimates for latent variables can be adequate for some factor models when the sample size is large enough and there are enough observed variables linked to each latent variable. However, for real data, as the number of variables increases, it is unlikely that factor models with conditional independence given latent variables continue to hold exactly. [17] have a partial study of their simple proxies in the case of weak conditional dependence of observed variables given the latent variables. This is called weak residual dependence; see also [12] and references therein.

For the proxies in Section 3, we have obtained conditions for weak residual dependence for which these proxies (derived based on assumption of conditional independence) are still consistent. We indicate a result for the Gaussian 1-factor model. There are analogous conditions for the 1-factor, bi-factor and oblique factor copula models.

With the linear representation as the Gaussian 1-factor model defined in (4), the residual dependence indicates that ϵ_j are not independent. Let $\boldsymbol{\Omega}_D$ be the correlation matrix of $\boldsymbol{\epsilon}_D=(\epsilon_1,\ldots,\epsilon_D)^\top$, $\boldsymbol{\Gamma}_D=\boldsymbol{\Psi}_D\boldsymbol{\Omega}_D\boldsymbol{\Psi}_D$, and $\boldsymbol{\Psi}_D^2$ is a diagonal matrix with diagonal entries of $\boldsymbol{\Gamma}_D$. The factor scores are defined as $\widetilde{w}_D=(\boldsymbol{I}+\boldsymbol{A}_D^\top\boldsymbol{\Psi}_D^{-2}\boldsymbol{A}_D)^{-1}\boldsymbol{A}_D^\top\boldsymbol{\Psi}_D^{-2}\boldsymbol{z}_D$.

Suppose the maximum eigenvalue of matrix $\boldsymbol{\Omega}_D$ is bounded as $D\to\infty$, the model is an approximate factor model from the definition in [7]. This assumption is sufficient for the defined proxy to be asymptotically consistent. An equivalent assumption, which is easier to check, is given below. Similar assumptions are presented in [4].

Assumption 3. Let $\boldsymbol{\Omega}_D = (\omega_{s,t})_{1\le s,t\le D}$ be the correlation matrix of ϵ_D. Let $S_D = \sum_{j=1}^{D} \epsilon_j$ and $\bar{\epsilon}_D = S_D/D$, then $\mathrm{E}(S_D^2) = \sum_{s=1}^{D}\sum_{t=1}^{D} \omega_{s,t}$. For a positive constant M, assume

$$0 < \lim\inf_{D\to\infty} D^{-1}\mathrm{E}(S_D^2) < \lim\sup_{D\to\infty} D^{-1}\mathrm{E}(S_D^2) < M,$$

Assumption 3 implies $\mathrm{Var}(\bar{\epsilon}_D) = O(D^{-1})$, the same order as the case of independent and identically distributed. Under Assumption 3 and the assumption on the loadings in Theorem 2, it is shown in Appendix B.1 that $\widetilde{w}_D - w^0 = O_p(D^{-1/2})$ as $D \to \infty$. That is, if the residual dependence is weak, the consistency of the proxy variable defined from a slightly misspecified model still holds with the same convergence rate.

Remark 7. Let $\omega_{s+} = \sum_{t=1}^{D} \omega_{st}$. If ω_{s+} is $O(1)$ as $D \to \infty$ for all s, then Assumption 3 is satisfied; e.g., ϵ's are indexed to have ante-dependence of order 1: $0 < r_1 < \omega_{j,j+1} < r_2 < 1$ for all j, and $\omega_{jk} = \prod_{i=j}^{k-1} \omega_{i,i+1}$ for $k - j \ge 2$. If ω_{s+} is $O(D)$ as $D \to \infty$ for all s, then Assumption 3 is not satisfied; e.g., ϵ_1 is dominating: $0 < r_1 < \omega_{1,j} < r_2 < 1$ for all j, and $\omega_{jk} = \omega_{1j} \cdot \omega_{1,k}$.

In the 1-factor copula model with weak residual dependence, with the same notations and assumptions in Theorem 5, there are similar sufficient conditions. If the copula for residual dependence is multivariate Gaussian, a sufficient condition is:

$$\sum_{j=1}^{D}\sum_{k=1}^{D} \mathrm{Cor}\big(C_{U_j|V}(U_j|V = v^0), C_{U_k|V}(U_k|V = v^0)\big) = O(D), \quad \forall\, 0 < v^0 < 1. \tag{24}$$

Similar ideas extend to residual dependence for bi-factor and oblique factor copulas.

8. Discussion and further research

Conditional expectation proxies of the latent variables are proposed in some factor copula models and they are consistent under some mild conditions. For high-dimensional factor copula models with a large sample size (large N, large D), simulation studies show that the sequential estimation approach can efficiently estimate the latent variables and select the families of linking copulas as well as estimate the copula parameters.

There are other recent methods for factor copula models that use Bayesian computing methods. For 1-factor copula model, [19] use reversible jump MCMC to select the bivariate copula links during the sampling process and to make inferences of the model parameters and latent variables. [18] utilize a Bayesian variational inference algorithm to make inferences for structured factor models but they make a strong assumption on the form of posterior distributions.

Our inference method is more intuitive and does not need a fixed factor structure. The sequential procedures fit better with the use of Gaussian factor models as a starting point to consider different factor structures that fit the data.

Our sequential proxy methods improve on the approach in [17] for 1-factor and oblique factor models, and can handle bi-factor copula models under some conditions. The sequential proxy procedures require numerical integration to compute second-stage proxies but not for maximum likelihood iterations for copula parameters, and hence the computation effort is reduced a lot. The simulation studies show the conditional expectation proxies are usually closer to the realized latent variables, leading to more accurate estimates of the parameters than that obtained from the "unweighted average" proxy approach in [17] in the 1-factor or oblique factor models.

Applications of factor copula models making use of the theory in this paper will be developed separately. Topics of further research and applications include the following.

(a) If the 1-factor structure is not adequate and group structure of observed variables cannot be determined from context, then a p-factor structure with varimax rotation can be fit to observed variables in the normal scores scale to check if an interpretable loading matrix with many zeros, corresponding to variables in overlapping groups, can be found. If so, for the factor copula counterpart, the sequential approach for the bi-factor copula can be extended. If the number p of latent variables is three of more, the exact copula likelihood would require p-dimensional Gaussian quadrature and we would not be able to compare estimation of copula parameters via proxies and via the exact likelihood. However the theory and examples suggest that the proxy approach will work if the number of variables linked to each latent variable is large enough.

(b) If one latent variable can explain much of the dependence but any p-factor loading matrix (with $p \ge 2$) is not interpretable, one could consider a 1-factor model with weak or moderate residual dependence. Starting with a preliminary 1-factor copula with residual dependence, one can iterate as in Section 6 and get proxies from the conditional expectation of the latent variable given the observed variables, from which to get better choices for the bivariate linking copulas to the latent variable. At most 1-dimensional Gaussian quadrature would be needed for likelihood estimation and computations of proxies.

CRediT authorship contribution statement

Xinyao Fan: Conceptualization, Methodology, Software, Simulation, Writing – original draft, Writing – review & editing. **Harry Joe:** Conceptualization, Methodology, Software, Simulation, Writing – review & editing.

Acknowledgments

This research has been support with NSERC Discovery Grant GR010293. Thanks to the referees for constructive comments and careful checks of proofs that have led to improved presentation.

Appendix A. Proofs of some derivations

A.1. Two-stage proxies in the bi-factor Gaussian model

Proof of equivalence of two-stage factor scores defined in (10) and (11) and regression factor scores. Suppose there are G groups, and let $\boldsymbol{z_D} = (\boldsymbol{z}_1^\top, \boldsymbol{z}_2^\top, \dots, \boldsymbol{z}_G^\top)^\top$, where $\boldsymbol{z}_1, \dots, \boldsymbol{z}_G$ are the realization of observed variables $\boldsymbol{Z}_1, \dots, \boldsymbol{Z}_G$. Let $\widehat{w}_0 = \mathrm{E}(W_0|\boldsymbol{Z_D} = \boldsymbol{z_D})$. For proxies of local latent factors, let $\widehat{w}_1 = \mathrm{E}(W_1|\boldsymbol{Z_D} = \boldsymbol{z_D})$. $\widetilde{w}_0, \widetilde{w}_1(\widetilde{w}_0)$ are defined in (10) and (11). Let $\widetilde{w}_1 = \widetilde{w}_1(\widetilde{w}_0)$ for notation simplicity. With loss of generality, it suffices to prove that $\widetilde{w}_0 = \widehat{w}_0$, $\widetilde{w}_1 = \widehat{w}_1$, as all indices of local latent factors could be permuted to be in the first group. Let $\boldsymbol{D} = (d_1, d_2, \dots, d_G)$, $d = \sum_{j=1}^{G} d_g$, $d_r = \sum_{j=2}^{G} d_j$, $\boldsymbol{z_D} = (\boldsymbol{z}_1^\top, \boldsymbol{z}_r^\top)^\top$. Then the loading matrix is

$$\boldsymbol{A} = [\boldsymbol{a}_0, \boldsymbol{a}_1, \dots, \boldsymbol{a}_p] = \left[\begin{array}{cc|ccc} \boldsymbol{b}_{01} & \boldsymbol{b}_1 & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \hline \boldsymbol{b}_{02} & \boldsymbol{0} & \boldsymbol{b}_2 & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{b}_{0G} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{b}_G \end{array}\right] = \begin{bmatrix} \boldsymbol{b}_{01} & \boldsymbol{b}_1 & \boldsymbol{0} \\ \boldsymbol{b}_{0r} & \boldsymbol{0} & \boldsymbol{B}_r \end{bmatrix}. \tag{25}$$

The partition of $\boldsymbol{A}$ leads to a 2×2 block matrix, where $\boldsymbol{b}_{0r} = (\boldsymbol{b}_{02}^\top, \dots, \boldsymbol{b}_{0G}^\top)^\top$, $\boldsymbol{B}_r = \mathrm{diag}(\boldsymbol{b}_2, \dots, \boldsymbol{b}_G)$. Also, partition $\boldsymbol{\Psi}^2 = \mathrm{diag}(\boldsymbol{\Psi}_1^2, \dots, \boldsymbol{\Psi}_G^2) = \mathrm{diag}(\boldsymbol{\Psi}_1^2, \boldsymbol{\Psi}_r^2)$ correspondingly.

Let $\boldsymbol{\Sigma_D} = \mathrm{Cor}(\boldsymbol{Z_D}) = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix} = \begin{bmatrix} \boldsymbol{b}_{01}\boldsymbol{b}_{01}^\top + \boldsymbol{b}_1\boldsymbol{b}_1^\top + \boldsymbol{\Psi}_1^2 & \boldsymbol{b}_{01}\boldsymbol{b}_{0r}^\top \\ \boldsymbol{b}_{0r}\boldsymbol{b}_{01}^\top & \boldsymbol{b}_{0r}\boldsymbol{b}_{0r}^\top + \boldsymbol{B}_r\boldsymbol{B}_r^\top + \boldsymbol{\Psi}_r^2 \end{bmatrix}$, $\boldsymbol{M} = \boldsymbol{\Sigma_D}^{-1} =: \begin{bmatrix} \boldsymbol{M}_{11} & \boldsymbol{M}_{12} \\ \boldsymbol{M}_{21} & \boldsymbol{M}_{22} \end{bmatrix}$. Then the sizes of matrices $\boldsymbol{M}_{11}, \boldsymbol{M}_{12}, \boldsymbol{M}_{21}, \boldsymbol{M}_{22}$ are $d_1 \times d_1$, $d_1 \times d_r$, $d_r \times d_1$, $d_r \times d_r$ respectively, and the corresponding blocks in $\boldsymbol{\Sigma_D}$ have the same size. Let $\boldsymbol{\Sigma}_1$ be the correlation matrix of $(\boldsymbol{Z}_1^\top, W_0)$. Then $\boldsymbol{\Sigma}_1 = \begin{bmatrix} \boldsymbol{b}_{01}\boldsymbol{b}_{01}^\top + \boldsymbol{b}_1\boldsymbol{b}_1^\top + \boldsymbol{\Psi}_1^2 & \boldsymbol{b}_{01} \\ \boldsymbol{b}_{01}^\top & 1 \end{bmatrix} =: \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$, and $\boldsymbol{N} = \boldsymbol{\Sigma}_1^{-1} = \begin{bmatrix} \boldsymbol{N}_{11} & \boldsymbol{N}_{12} \\ \boldsymbol{N}_{21} & \boldsymbol{N}_{22} \end{bmatrix}$. The sizes of the matrices N_{11}, N_{12}, N_{21} are $d_1 \times d_1$, $d_1 \times 1$, $1 \times d_1$ respectively, and N_{22} is a scalar. Let $\boldsymbol{z_D} = (\boldsymbol{z}_1^\top, \boldsymbol{z}_r^\top)^\top$, the regression factor scores defined in (9) are: $\widehat{w}_0 = \boldsymbol{a}_0^\top \boldsymbol{\Sigma_D}^{-1}\boldsymbol{z_D} = \boldsymbol{a}_0^\top \boldsymbol{M}\boldsymbol{z_D}$, $\quad \widehat{w}_1 = \boldsymbol{a}_1^\top \boldsymbol{\Sigma_D}^{-1}\boldsymbol{z_D} = \boldsymbol{a}_1^\top \boldsymbol{M}\boldsymbol{z_D}$. Hence, by (25),

$$\widehat{w}_0 = (\boldsymbol{b}_{01}^\top, \boldsymbol{b}_{0r}^\top) \begin{bmatrix} \boldsymbol{M}_{11} & \boldsymbol{M}_{12} \\ \boldsymbol{M}_{21} & \boldsymbol{M}_{22} \end{bmatrix} \begin{pmatrix} \boldsymbol{z}_1 \\ \boldsymbol{z}_r \end{pmatrix} = (\boldsymbol{b}_{01}^\top \boldsymbol{M}_{11} + \boldsymbol{b}_{0r}^\top \boldsymbol{M}_{21})\boldsymbol{z}_1 + (\boldsymbol{b}_{01}^\top \boldsymbol{M}_{12} + \boldsymbol{b}_{0r}^\top \boldsymbol{M}_{22})\boldsymbol{z}_r, \tag{26}$$

$$\widehat{w}_1 = (\boldsymbol{b}_1^\top, \boldsymbol{0}_r) \begin{bmatrix} \boldsymbol{M}_{11} & \boldsymbol{M}_{12} \\ \boldsymbol{M}_{21} & \boldsymbol{M}_{22} \end{bmatrix} \begin{pmatrix} \boldsymbol{z}_1 \\ \boldsymbol{z}_r \end{pmatrix} = \boldsymbol{b}_1^\top \boldsymbol{M}_{11}\boldsymbol{z}_1 + \boldsymbol{b}_1^\top \boldsymbol{M}_{12}\boldsymbol{z}_r. \tag{27}$$

The expressions of $\widetilde{w}_0$ and $\widehat{w}_0$ equal $\boldsymbol{a}_0^\top \boldsymbol{M}\boldsymbol{z_D}$, so they are the same. After some algebraic calculations in (11), $\widetilde{w}_1 = \boldsymbol{b}_1^\top \boldsymbol{N}_{11}\boldsymbol{z}_1 + \boldsymbol{b}_1^\top \boldsymbol{N}_{12}\widetilde{w}_0$. Substituting $\widehat{w}_0 = \widetilde{w}_0$ from (26) into $\widetilde{w}_1$ leads to

$$\widetilde{w}_1 = [\boldsymbol{b}_1^\top \boldsymbol{N}_{11} + \boldsymbol{b}_1^\top \boldsymbol{N}_{12}(\boldsymbol{b}_{01}^\top \boldsymbol{M}_{11} + \boldsymbol{b}_{0r}^\top \boldsymbol{M}_{21})]\boldsymbol{z}_1 + \boldsymbol{b}_1^\top \boldsymbol{N}_{12}(\boldsymbol{b}_{01}^\top \boldsymbol{M}_{12} + \boldsymbol{b}_{0r}^\top \boldsymbol{M}_{22})\boldsymbol{z}_r. \tag{28}$$

To conclude, it suffices to show that $\widetilde{w}_1$ in (28) and $\widehat{w}_1$ in (27) are equivalent, or that (a) $\boldsymbol{N}_{11} + \boldsymbol{N}_{12}(\boldsymbol{b}_{01}^\top \boldsymbol{M}_{11} + \boldsymbol{b}_{0r}^\top \boldsymbol{M}_{21}) = \boldsymbol{M}_{11}$, and (b) $\boldsymbol{N}_{12}(\boldsymbol{b}_{01}^\top \boldsymbol{M}_{12} + \boldsymbol{b}_{0r}^\top \boldsymbol{M}_{22}) = \boldsymbol{M}_{12}$.

Let $\Delta_1 = (\boldsymbol{b}_1\boldsymbol{b}_1^\top + \boldsymbol{\Psi}_1^2)$; Δ_1 is positive definite and $\Delta_1 + \boldsymbol{b}_{01}\boldsymbol{b}_{01}^\top = \boldsymbol{\Sigma}_{11}$. Multiply Δ_1^{-1} on the left and $\boldsymbol{\Sigma}_{11}^{-1}$ on the right to get (c) $\Delta_1^{-1}\boldsymbol{b}_{01}\boldsymbol{b}_{01}^\top \boldsymbol{\Sigma}_{11}^{-1} - \Delta_1^{-1} = -\boldsymbol{\Sigma}_{11}^{-1}$. From $\boldsymbol{\Sigma_D}\boldsymbol{M} = \mathbf{I}$, we have (d) $\boldsymbol{M}_{11} = \boldsymbol{\Sigma}_{11}^{-1} - \boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{b}_{01}\boldsymbol{b}_{0r}^\top \boldsymbol{M}_{21}$ and (e) $\boldsymbol{M}_{12} = -\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{b}_{01}\boldsymbol{b}_{0r}^\top \boldsymbol{M}_{22}$. From $\boldsymbol{N}\boldsymbol{\Sigma}_1 = \mathbf{I}$, we have (f) $\boldsymbol{N}_{11}\boldsymbol{\Sigma}_{11} + \boldsymbol{N}_{12}\boldsymbol{b}_{01}^\top = \boldsymbol{N}_{11}(\Delta_1 + \boldsymbol{b}_{01}\boldsymbol{b}_{01}^\top) + \boldsymbol{N}_{12}\boldsymbol{b}_{01}^\top = \mathbf{I}$ and (g) $\boldsymbol{N}_{11}\boldsymbol{b}_{01} + \boldsymbol{N}_{12} = 0$. In (g), multiply both sides by $\boldsymbol{b}_{01}^\top$ to get (h) $\boldsymbol{N}_{11}\boldsymbol{b}_{01}\boldsymbol{b}_{01}^\top + \boldsymbol{N}_{12}\boldsymbol{b}_{01}^\top = 0$. Then (f) and (h) together imply (i) $\boldsymbol{N}_{11}\Delta_1 = \mathbf{I}$. Hence, from (g) and (i), $\boldsymbol{N}_{12} = -\boldsymbol{N}_{11}\boldsymbol{b}_{01} = -\Delta_1^{-1}\boldsymbol{b}_{01}$, and from (f), $\boldsymbol{N}_{11} = \boldsymbol{\Sigma}_{11}^{-1} - \boldsymbol{N}_{12}\boldsymbol{b}_{01}^\top \boldsymbol{\Sigma}_{11}^{-1} = \boldsymbol{\Sigma}_{11}^{-1} + \Delta_1^{-1}\boldsymbol{b}_{01}\boldsymbol{b}_{01}^\top \boldsymbol{\Sigma}_{11}^{-1}$. Substitute these expressions of $\boldsymbol{N}_{11}$ and $\boldsymbol{N}_{12}$ in the left-hand side of equation (a) to get:

$$\boldsymbol{N}_{11} + \boldsymbol{N}_{12}\boldsymbol{b}_{01}^\top \boldsymbol{M}_{11} + \boldsymbol{N}_{12}\boldsymbol{b}_{0r}^\top \boldsymbol{M}_{21} = \boldsymbol{\Sigma}_{11}^{-1} + \left\{\Delta_1^{-1}\boldsymbol{b}_{01}\boldsymbol{b}_{01}^\top \boldsymbol{\Sigma}_{11}^{-1} - \Delta_1^{-1}\boldsymbol{b}_{01}\boldsymbol{b}_{01}^\top \boldsymbol{M}_{11} - \Delta_1^{-1}\boldsymbol{b}_{01}\boldsymbol{b}_{0r}^\top \boldsymbol{M}_{21}\right\}.$$

For the right-hand side of the above, substitute $\boldsymbol{M}_{11}$ from (d) and then $-\boldsymbol{\Sigma}_{11}^{-1}$ in (c), so that the sum of the last three terms in braces becomes

$$\begin{aligned} &\Delta_1^{-1}\boldsymbol{b}_{01}\boldsymbol{b}_{01}^\top \boldsymbol{\Sigma}_{11}^{-1} - \Delta_1^{-1}\boldsymbol{b}_{01}\boldsymbol{b}_{01}^\top(\boldsymbol{\Sigma}_{11}^{-1} - \boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{b}_{01}\boldsymbol{b}_{0r}^\top \boldsymbol{M}_{21}) - \Delta_1^{-1}\boldsymbol{b}_{01}\boldsymbol{b}_{0r}^\top \boldsymbol{M}_{21} \\ &= \Delta_1^{-1}\boldsymbol{b}_{01}\boldsymbol{b}_{01}^\top \boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{b}_{01}\boldsymbol{b}_{0r}^\top \boldsymbol{M}_{21} - \Delta_1^{-1}\boldsymbol{b}_{01}\boldsymbol{b}_{0r}^\top \boldsymbol{M}_{21} = (\Delta_1^{-1}\boldsymbol{b}_{01}\boldsymbol{b}_{01}^\top \boldsymbol{\Sigma}_{11}^{-1} - \Delta_1^{-1})\boldsymbol{b}_{01}\boldsymbol{b}_{0r}^\top \boldsymbol{M}_{21} = -\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{b}_{01}\boldsymbol{b}_{0r}^\top \boldsymbol{M}_{21}. \end{aligned}$$

Thus, (a) is verified as $\boldsymbol{N}_{11} + \boldsymbol{N}_{12}\boldsymbol{b}_{01}^\top \boldsymbol{M}_{11} + \boldsymbol{N}_{12}\boldsymbol{b}_{0r}^\top \boldsymbol{M}_{21} = \boldsymbol{\Sigma}_{11}^{-1} - \boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{b}_{01}\boldsymbol{b}_{0r}^\top \boldsymbol{M}_{21} = \boldsymbol{M}_{11}$ via (d). Next, substitute $\boldsymbol{N}_{12}$ in (g) and $\boldsymbol{M}_{12}$ in (e) in the left-hand side of (b), so that (b) is verified as

$$\begin{aligned} \boldsymbol{N}_{12}(\boldsymbol{b}_{01}^\top \boldsymbol{M}_{12} + \boldsymbol{b}_{0r}^\top \boldsymbol{M}_{22}) &= \Delta_1^{-1}\boldsymbol{b}_{01}\boldsymbol{b}_{01}^\top \boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{b}_{01}\boldsymbol{b}_{0r}^\top \boldsymbol{M}_{22} - \Delta_1^{-1}\boldsymbol{b}_{01}\boldsymbol{b}_{0r}^\top \boldsymbol{M}_{22} \\ &= (\Delta_1^{-1}\boldsymbol{b}_{01}\boldsymbol{b}_{01}^\top \boldsymbol{\Sigma}_{11}^{-1} - \Delta_1^{-1})\boldsymbol{b}_{01}\boldsymbol{b}_{0r}^\top \boldsymbol{M}_{22} = -\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{b}_{01}\boldsymbol{b}_{0r}^\top \boldsymbol{M}_{22} = \boldsymbol{M}_{12}, \end{aligned}$$

via (c) and then (e). □

A.2. Eq. (17) in Table 2

Proof. In the bi-factor model (5), recall the notation $\boldsymbol{b}_g$, $\boldsymbol{b}_{0g}$, $\boldsymbol{\Sigma}_g$ defined in Section 3.1, and q_g, $\tilde{q}_g$ in the caption of Table 2. Let $\boldsymbol{r}_g = (\boldsymbol{b}_g^\top, 0)\boldsymbol{\Sigma}_g^{-1}$ be a vector of length $d_g + 1$. Let r be the last entry of vector $\boldsymbol{r}_g$. Define $\boldsymbol{\Sigma}_{gg} = \boldsymbol{B}_g\boldsymbol{B}_g^\top + \boldsymbol{\Psi}_g^2$. Let the last column of $\boldsymbol{\Sigma}_g^{-1}$ be $[\boldsymbol{s}_{12}^\top, s_{22}]^\top$. From $\boldsymbol{\Sigma}_g\boldsymbol{\Sigma}_g^{-1} = \mathbf{I}$, two equations are obtained. (a) $\boldsymbol{\Sigma}_{gg}\boldsymbol{s}_{12} + \boldsymbol{b}_{0g}s_{22} = 0$ and (b) $\boldsymbol{b}_{0g}^\top \boldsymbol{s}_{12} + s_{22} = 1$. Multiply both sides of (b) by $\boldsymbol{b}_{0g}$ to get (c) $\boldsymbol{b}_{0g}\boldsymbol{b}_{0g}^\top \boldsymbol{s}_{12} + \boldsymbol{b}_{0g} s_{22} = \boldsymbol{b}_{0g}$. Then (a) and (c) together implies $(\boldsymbol{b}_{0g}\boldsymbol{b}_{0g}^\top - \boldsymbol{\Sigma}_{gg})\boldsymbol{s}_{12} = \boldsymbol{b}_{0g}$. Hence, $\boldsymbol{s}_{12} = (\boldsymbol{b}_{0g}\boldsymbol{b}_{0g}^\top - \boldsymbol{\Sigma}_{gg})^{-1}\boldsymbol{b}_{0g}$. Since $\boldsymbol{r}_g = (\boldsymbol{b}_g^\top, 0)\boldsymbol{\Sigma}_g^{-1} = (\boldsymbol{b}_g^\top, 0)\begin{bmatrix} *, \boldsymbol{s}_{12} \\ *, s_{22} \end{bmatrix} = [*, \boldsymbol{b}_g^\top \boldsymbol{s}_{12}]$, the last entry of $\boldsymbol{r}_g$ is $r = \boldsymbol{b}_g^\top \boldsymbol{s}_{12}$.

In the definition of factor scores in (11), $\mathrm{E}(W_g|\boldsymbol{Z}_{\boldsymbol{D}}, W_0) = \boldsymbol{r}_g(\boldsymbol{Z}_g^\top, W_0)^\top = h(\boldsymbol{Z}_g) + rW_0$, where $h(\boldsymbol{Z}_g)$ is a linear function of $\boldsymbol{Z}_g$. Then $\mathrm{Var}[\mathrm{E}(W_g|\boldsymbol{Z}_{\boldsymbol{D}}, W_0)|\boldsymbol{Z}_D] = r^2\,\mathrm{Var}(W_0|\boldsymbol{Z}_D)$. From the conditional variance decomposition formula,

$$\begin{aligned}\mathrm{Var}(W_g|\boldsymbol{Z}_{\boldsymbol{D}}) &= \mathrm{E}[\mathrm{Var}(W_g|\boldsymbol{Z}_{\boldsymbol{D}}, W_0)|\boldsymbol{Z}_{\boldsymbol{D}}] + \mathrm{Var}[\mathrm{E}(W_g|\boldsymbol{Z}_{\boldsymbol{D}}, W_0)|\boldsymbol{Z}_{\boldsymbol{D}}] \\ &= \underbrace{[1 - \boldsymbol{b}_g^\top(\boldsymbol{b}_g\boldsymbol{b}_g^\top + \boldsymbol{\Psi}_g^2)^{-1}\boldsymbol{b}_g]}_{\text{term1}} + \underbrace{r^2\,\mathrm{Var}(W_0|\boldsymbol{Z}_{\boldsymbol{D}})}_{\text{term2}},\end{aligned} \tag{29}$$

where $r = \boldsymbol{b}_g^\top \boldsymbol{s}_{12} = -\boldsymbol{b}_g^\top(\boldsymbol{\Sigma}_{gg} - \boldsymbol{b}_{0g}\boldsymbol{b}_{0g}^\top)^{-1}\boldsymbol{b}_{0g} = -\boldsymbol{b}_g^\top(\boldsymbol{b}_g\boldsymbol{b}_g^\top + \boldsymbol{\Psi}_g^2)^{-1}\boldsymbol{b}_{0g}$.

Since $\boldsymbol{Z}_g$ is independent of Z's in other groups given W_0, $\mathrm{Var}(W_g|\boldsymbol{Z}_{\boldsymbol{D}}, W_0) = \mathrm{Var}(W_g|\boldsymbol{Z}_g; W_0)$. The term1 in (29) follows because the joint distribution of $(\boldsymbol{Z}_g^\top, W_g)^\top$ given W_0 is multivariate normal with zero mean and covariance matrix $\begin{bmatrix} \boldsymbol{b}_g\boldsymbol{b}_g^\top + \boldsymbol{\Psi}_g^2 & \boldsymbol{b}_g \\ \boldsymbol{b}_g^\top & 1 \end{bmatrix}$. From assuming $\psi_{jg} > 0$ for all j, g, applying (8) with $\boldsymbol{A}_D = \boldsymbol{b}_g$ and $\boldsymbol{\Psi}_D = \boldsymbol{\Psi}_g$, term1 simplifies into $(1 + \boldsymbol{b}_g^\top \boldsymbol{\Psi}_g^{-2}\boldsymbol{b}_g)^{-1} = (1 + q_g)^{-1}$. As for term2 in (29), in the expression of r, applying (8) as above, $r = -\tilde{q}_g(1 + q_g)^{-1}$. Combine the expression of two terms, the decomposition (17) is obtained. $\square$

Appendix B. Main proofs in Sections 4 and 7

B.1. Proof of consistency for regression factor scores: Theorems 2 and 4

Proof of Theorem 2. In 1-factor model ($p = 1$), the loading matrix $\boldsymbol{A}_D$ is $D \times 1$. Due to assumption on α_j uniformly bounded away from ± 1, $\boldsymbol{\Psi}_D^{-1}$ is well-defined for all D. Thus the regression factor scores can be expressed in two equivalent forms. In the expression (9), let $q_D = \boldsymbol{A}_D^\top \boldsymbol{\Psi}_D^{-2}\boldsymbol{A}_D$ (a positive real number). Then

$$\widetilde{w}_D - w^0 = (1 + q_D)^{-1}\boldsymbol{A}_D^\top \boldsymbol{\Psi}_D^{-1}\boldsymbol{e}_D + O(D^{-1}).$$

Next, $D^{-1/2}\boldsymbol{A}_D^\top \boldsymbol{\Psi}_D^{-1}\boldsymbol{e}_D$ is a realization of $D^{-1/2}\boldsymbol{A}_D^\top \boldsymbol{\Psi}_D^{-1}\boldsymbol{\epsilon}_D = D^{-1/2}\sum_{j=1}^D \alpha_j\epsilon_j/\psi_j$, with $\mathrm{Var}(D^{-1/2}\sum_{j=1}^D \alpha_j\epsilon_j/\psi_j) = q_D/D =: \bar{q}_D$.

With the assumptions, (1) $\liminf \psi_j > k$, where k is a small positive number, (2) $D^{-1}\sum_{j=1}^D |\alpha_j| \le (D^{-1}\sum_{j=1}^D \alpha_j^2)^{1/2} \le (D^{-1}\sum_{j=1}^D |\alpha_j|)^{1/2}$ and $D^{-1}\sum_{j=1}^D |\alpha_j| \to \text{const} \ne 0$. It can be concluded that $\bar{q}_D = D^{-1}\sum_{j=1}^D \alpha_j^2\psi_j^{-2} < k^{-2}D^{-1}\sum_{j=1}^D \alpha_j^2$ is bounded for all large D. So $\bar{q}_D = O(1)$. Hence, $D^{-1/2}\sum_{j=1}^D \alpha_j\epsilon_j/\psi_j$ and the realization $D^{-1/2}\boldsymbol{A}_D^\top \boldsymbol{\Psi}_D^{-1}\boldsymbol{e}_D$ can be consider as $O_p(1)$. Therefore,

$$(\widetilde{w}_D - w^0) = D^{-1/2}(D^{-1} + \bar{q}_D)^{-1}(D^{-1/2}\boldsymbol{A}_D^\top \boldsymbol{\Psi}_D^{-1}\boldsymbol{e}_D) + O(D^{-1})$$

is asymptotically $O_p(D^{-1/2})$. $\square$

Proof of extension to weak residual dependence. An outline of the proof of consistency based on Assumption 3 is as follows. Let $\boldsymbol{e}_D = (e_1, \ldots, e_D)$ be one realization of $\boldsymbol{\epsilon}_D$. Let $q_D = \boldsymbol{A}_D^\top \boldsymbol{\Psi}_D^{-2}\boldsymbol{A}_D > 0$. Then, as in (9),

$$\widetilde{w}_D - w^0 = (1 + q_D)^{-1}\boldsymbol{A}_D^\top \boldsymbol{\Psi}_D^{-1}\boldsymbol{e}_D + (1 + q_D^{-1})^{-1}w^0 - w^0.$$

Note that $Y = D^{-1/2}\boldsymbol{A}_D^\top \boldsymbol{\Psi}_D^{-1}\boldsymbol{e}_D$ is a realization of $D^{-1/2}\boldsymbol{A}_D^\top \boldsymbol{\Psi}_D^{-1}\boldsymbol{\epsilon}_D = D^{-1/2}\sum_{j=1}^D \alpha_j\epsilon_j/\psi_j$. The latter random quantity has variance $D^{-1}\sum_{j=1}^D\sum_{k=1}^D \alpha_j\alpha_k\omega_{jk}/[\psi_j\psi_k]$. By Assumption 3 and with loadings that are bounded away from ± 1, this variance is $O(1)$ so that Y can be considered as $O_p(1)$. Following the same logic as in the proof of Theorem 2, $\bar{q}_D = O(1)$. Then $\widetilde{w}_D - w^0 = D^{-1/2}(D^{-1} + \bar{q}_D)^{-1}(D^{-1/2}\boldsymbol{A}_D^\top \boldsymbol{\Psi}_D^{-1}\boldsymbol{\epsilon}_D) + O(D^{-1})$ is asymptotically $O_p(D^{-1/2})$. $\square$

Proof of Theorem 4. Using the technique in the proof of Theorem 2, let $\boldsymbol{Q}_{\boldsymbol{D}} = \boldsymbol{A}_{\boldsymbol{D}}^\top \boldsymbol{\Psi}_{\boldsymbol{D}}^{-2}\boldsymbol{A}_{\boldsymbol{D}}$, and $\overline{\boldsymbol{Q}}_{\boldsymbol{D}} = D^{-1}\boldsymbol{A}_{\boldsymbol{D}}^\top \boldsymbol{\Psi}_{\boldsymbol{D}}^{-2}\boldsymbol{A}_{\boldsymbol{D}}$. Since $\boldsymbol{A}_{\boldsymbol{D}}$ is of full rank and the entries of $\boldsymbol{\Psi}_g$ for $g \in \{1, \ldots, G\}$ are uniformly bounded away from 0. Also by assumption $d_g^{-1}\|\boldsymbol{a}_{d_g,g}\|_1 \not\to 0$, $\overline{\boldsymbol{Q}}_{\boldsymbol{D}}$ is positive definite for any fixed D and $\overline{\boldsymbol{Q}}_{\boldsymbol{D}} = O(1)$. Hence,

$$\widetilde{\boldsymbol{w}}_{\boldsymbol{D}} - \boldsymbol{w}^0 = (\mathbf{I}_p + \boldsymbol{Q}_{\boldsymbol{D}})^{-1}\boldsymbol{A}_{\boldsymbol{D}}^\top \boldsymbol{\Psi}_{\boldsymbol{D}}^{-1}\boldsymbol{e}_{\boldsymbol{D}} + (\mathbf{I}_p + \boldsymbol{Q}_{\boldsymbol{D}}^{-1})^{-1}\boldsymbol{w}^0 - \boldsymbol{w}^0 = D^{-1/2}(D^{-1}\mathbf{I}_p + \overline{\boldsymbol{Q}}_{\boldsymbol{D}})^{-1}D^{-1/2}\boldsymbol{A}_{\boldsymbol{D}}^\top \boldsymbol{\Psi}_{\boldsymbol{D}}^{-1}\boldsymbol{e}_{\boldsymbol{D}} + O(D^{-1}). \tag{30}$$

The second equation in (30) is based on Taylor expansion

$$(\mathbf{I}_p + \boldsymbol{Q}_D^{-1})^{-1}\boldsymbol{w}^0 - \boldsymbol{w}^0 = (\mathbf{I}_p + D^{-1}\overline{\boldsymbol{Q}}_D^{-1})^{-1}\boldsymbol{w}^0 - \boldsymbol{w}^0 = \left(\mathbf{I}_p - D^{-1}\overline{\boldsymbol{Q}}_D^{-1} + o(D^{-1})\right)\boldsymbol{w}^0 - \boldsymbol{w}^0 = O(D^{-1}).$$

Let $\boldsymbol{Y} = D^{-1/2}\boldsymbol{A}_{\boldsymbol{D}}^\top \boldsymbol{\Psi}_{\boldsymbol{D}}^{-1}\boldsymbol{\epsilon}_{\boldsymbol{D}}$, so that $\mathrm{Var}(\boldsymbol{Y}) = D^{-1}\boldsymbol{Q}_{\boldsymbol{D}} = \overline{\boldsymbol{Q}}_{\boldsymbol{D}} = O(1)$ and $\mathrm{E}(\boldsymbol{Y}) = 0$. Hence $\boldsymbol{Y} = O_p(1)$. Since $D^{-1/2}\boldsymbol{A}_{\boldsymbol{D}}^\top \boldsymbol{\Psi}_{\boldsymbol{D}}^{-1}\boldsymbol{e}_{\boldsymbol{D}}$ is one realization of random variable $\boldsymbol{Y}$, it can also be considered as $O_p(1)$. Hence, $\widetilde{\boldsymbol{w}}_{\boldsymbol{D}} - \boldsymbol{w}^0$ is asymptotically $O_p(D^{-1/2})$ by noticing that $(D^{-1}\mathbf{I}_p + \overline{\boldsymbol{Q}}_{\boldsymbol{D}})^{-1} = O(1)$. $\square$

B.2. Proof of consistency for proxies: Theorems 5 and 7

For the conditional expectations for 1-factor and bi-factor copulas, the v's should be treated as parameters, and the u's are the realization of independent random variables when the latent variables are fixed. The proof techniques of Theorems 5 and 7 are similar. Both rely on the Laplace approximation for integrals (see [6] and [20]), and the asymptotic properties of maximum likelihood (ML) estimator for parameters. In our setting, the results in [5] are used for the asymptotics of a log-likelihood for a sample $X_j \sim f_{X_j}$ from independent but not identically distributed observations with common parameters over the $\{f_{X_j}\}$.

Proof of Theorem 5. With the assumptions that (a) all $\log c_{jV}$ are continuously differentiable up to third order and (b) $|\partial \log c_{jV}/\partial v|$ are uniform bounded for $v_A \le v \le v_B$, the sufficient conditions in [5] are satisfied.

In 1-factor copula model (1), there is a realized value v^0 for the latent variable. Then $(U_1,\dots,U_D,\dots)$ is an infinite sequence of independent random variables with $U_j \sim c_{jV}(\cdot, v^0)$. If the value of v^0 is to be estimated based on the realized D-vector $(u_1,\dots,u_D)$, then the averaged negative log-likelihood in v is $g_D(v) = -D^{-1}\sum_{j=1}^D \log c_{jV}(u_j, v)$. The maximum likelihood estimate v_D^* satisfies $v_D^* = v^0 + O_p(D^{-1/2})$ from results in [5]. With the Laplace approximation via equation (2.3) in [20], the numerator and denominator denoted as I_{1D} and I_{2D} in the expression of $\widetilde{v}_D$ in (13) can be approximated respectively by

$$I_{1D} = \int_0^1 v\exp\{-D\times g_D(v)\}\,\mathrm{d}v = v_D^*\exp\{-D\times g_D(v_D^*)\}\cdot(2\pi)^{1/2}[D|g_D''(v_D^*)|]^{-1/2}\cdot\{1+O(D^{-1})\},$$

$$I_{2D} = \int_0^1 \exp\{-D\times g_D(v)\}\,\mathrm{d}v = \exp\{-D\times g_D(v_D^*)\}\cdot(2\pi)^{1/2}[D|g_D''(v_D^*)|]^{-1/2}\cdot\{1+O(D^{-1})\}.$$

Hence $\widetilde{v}_D - v_D^* = I_{1D}/I_{2D} - v_D^* = O(D^{-1})$ and $\widetilde{v}_D - v^0 = O_p(D^{-1/2})$. □

Proof of Theorem 7. With Assumption 1, the sufficient conditions in [5] are satisfied. There are realized value $v_0^0, v_1^0, \dots, v_G^0$ for the latent variables. Then $(U_{1g},\dots,U_{d_g g},\dots)$ is an infinite sequence of dependent random variables for each $g \in \{1,\dots,G\}$, and the G sequences are mutually independent given latent variables. For the bi-factor copula model, from Algorithm 24 in Joe (2014), the cdf of U_{jg} is $C_{U_{jg}|V_g;V_0}\big(C_{U_{jg}|V_0}(\cdot|v_0^0), v_g^0\big)$ and its density is

$$c_{U_{jg}V_0}(\cdot, v_0^0)\cdot c_{U_{jg},V_g;V_0}\big(C_{U_{jg}|V_0}(\cdot|v_0^0), v_g^0\big).$$

If the values of $v_0^0, v_1^0, \dots, v_G^0$ are to be estimated based on the realized d_g-vector $\boldsymbol{u}_{g,d_g} = (u_{1g},\dots,u_{d_g g})$ for $g \in \{1,\dots,G\}$, then the integrated log-likelihood in v_0 is $-L_{0D}(v_0) = \log c_{\boldsymbol{U_D},V_0}(\boldsymbol{u_D}, v_0)$, where $\log c_{\boldsymbol{U_D},V_0}(\boldsymbol{u_D}, v_0)$ is defined in Table 1. Take the partial derivative with respect to v_0 leads the first inference function $\Psi_{0,D}(v_0;\boldsymbol{u_D})$.

$$\begin{aligned}\Psi_{0,D}(v_0;\boldsymbol{u_D}) &:= \partial\log c_{\boldsymbol{U_D},V_0}(\boldsymbol{u_D}, v_0)/\partial v_0 = \Psi_{01,D}(v_0;\boldsymbol{u_D}) + \Psi_{02,D}(v_0;\boldsymbol{u_D})\\ &:= \sum_{g=1}^G\sum_{j=1}^{d_g}\partial\log c_{U_{jg},V_0}(u_{jg}, v_0)/\partial v_0 + \sum_{g=1}^G \partial\log f_g(\boldsymbol{u}_g; v_0)/\partial v_0.\end{aligned}$$

Let $v_{0,\boldsymbol{D}}^*$ be the maximum likelihood estimate and assume it is the unique solution of $\bar{\Psi}_{0,D} = D^{-1}\Psi_{0,D}$. Note that assuming regularity assumptions include the exchange of integration and the partial differentiation,

$$\begin{aligned}\frac{\partial\log f_g(\boldsymbol{u}_g;v_0)}{\partial v_0} &= \frac{1}{f_g(\boldsymbol{u}_g;v_0)}\int_0^1\left(\partial\exp\left\{\sum_{j=1}^{d_g}\log c_{U_{jg},V_g;V_0}(C_{U_{jg}|V_0}(u_{jg}|v_0), v_g)\right\}\Big/\,\partial v_0\right)\mathrm{d}v_g\\ &= \sum_{j=1}^{d_g}\int_0^1\frac{h_g(\boldsymbol{u}_g;v_0)}{f_g(\boldsymbol{u}_g;v_0)}\frac{\partial\log c_{U_{jg}V_g;V_0}(C_{U_{jg}|V_0}(u_{jg}|v_0), v_g)}{\partial v_0}\,\mathrm{d}v_g,\end{aligned}$$

where $h_g(\boldsymbol{u}_g;v_0) = \prod_{j=1}^{d_g} c_{U_{jg},V_g;V_0}(C_{U_{jg}|V_0}(u_{jg}|v_0), v_g)$. The derivatives of $\bar{\Psi}_{0,D}$ can be written as $\partial\bar{\Psi}_{0,D}(v_0)/\partial v_0 := \partial\bar{\Psi}_{01,D}(v_0)/\partial v_0 + \partial\bar{\Psi}_{02,D}(v_0)/\partial v_0$. From the laws of large numbers in page 174 of [8], and with the given assumptions,

$$\lim_{D\to\infty}\bar{\Psi}_{01,D}(v_0,\boldsymbol{u_D}) = \lim_{D\to\infty} D^{-1}\sum_{g=1}^G\sum_{j=1}^{d_g}\partial\log c_{U_{jg},V_0}(u_{jg}, v_0)/\partial v_0,$$

$$\lim_{D\to\infty}\bar{\Psi}_{02,D}(v_0,\boldsymbol{u_D}) = \lim_{D\to\infty} D^{-1}\sum_{g=1}^G\sum_{j=1}^{d_g}\int_0^1\frac{c_{\boldsymbol{U}_g,V_g;V_0}(\boldsymbol{u}_g, v_g; v_0)}{f_g(\boldsymbol{u}_g;v_0)}\frac{\partial\log c_{U_{jg}V_g;V_0}(C_{U_{jg}|V_0}(u_{jg}|v_0), v_g)}{\partial v_0}\mathrm{d}v_g$$

exist. With enough dependence on the latent variable, the derivative of $\bar{\Psi}_{0,D}(v_0,\boldsymbol{u_D})$ is bounded away from 0. Then as $d_g\to\infty$ for all g, $v_{0,\boldsymbol{D}}^* = v_0^0 + o_p(1)$.

Furthermore, the profile log-likelihood in v_g given v_0 is (from Table 1):

$$-L_{gD}(v_g;\boldsymbol{u}_{g,d_g}, v_0) = \log c_{\boldsymbol{U}_g,V_0,V_g}(\boldsymbol{u}_{g,d_g}, v_0, v_g) = \sum_{j=1}^{d_g}\log\Big\{c_{U_{jg}V_0}(u_{jg}, v_0) + \log c_{U_{jg}|V_g;V_0}\big(C_{U_{jg}|V_0}(u_{jg}|v_0), v_g\big)\Big\}.$$

The partial derivative of $-L_{gD}$ with respect to v_g leads to the inference function $\Psi_{g,D}(v_g;\boldsymbol{u}_{g,d_g},v_0)$. For v_0 in a neighborhood of v_0^0, let $v^*_{g,d_g}(v_0)$ be maximum profile likelihood estimate and assume it is the unique solution of $\bar{\Psi}_{g,D}(v_g;\boldsymbol{u}_{g,d_g},v_0) := d_g^{-1}\Psi_{g,D}(v_g;\boldsymbol{u}_{g,d_g},v_0)$. From the uniform laws of large numbers, and with the given assumptions,

$$\lim_{d_g\to\infty}\bar{\Psi}_{g,D}(v_g;\boldsymbol{u}_{g,d_g},v_0)=\lim_{d_g\to\infty} d_g^{-1}\sum_{j=1}^{d_g}\frac{\partial\log c_{U_{jg}|V_g;V_0}\left(C_{U_{jg}|V_0}(u_{jg}|v_0)\;\middle|\;v_g\right)}{\partial v_g}$$

uniformly converge to the limiting function in a neighborhood around v_0^0. As $d_g\to\infty$ for all g, $v^*_{g,d_g}(v^*_{0,\boldsymbol{D}})=v_g^0+o_p(1)$.

For the proxy defined in (14),

$$\widetilde{v}_{0\boldsymbol{D}}(\boldsymbol{u}_{\boldsymbol{D}})=\int_0^1 v_0\exp\{-D\cdot\bar{L}_{0\boldsymbol{D}}(v_0;\boldsymbol{u}_{\boldsymbol{D}})\}\mathrm{d}v_0\Big/\int_0^1\exp\{-D\cdot\bar{L}_{0\boldsymbol{D}}(v_0;\boldsymbol{u}_{\boldsymbol{D}})\}\mathrm{d}v_0. \tag{31}$$

Since $\bar{L}_{0\boldsymbol{D}}$ attains the global minimum at $v^*_{0,\boldsymbol{D}}$, then from the Laplace approximation, when $\boldsymbol{D}\to\infty$, the numerator and denominator in (31) can be approximated by

$$\begin{aligned}
&v^*_{0,\boldsymbol{D}}\exp\{-D\times\bar{L}_{0\boldsymbol{D}}(v^*_{0,\boldsymbol{D}})\}\cdot(2\pi)^{1/2}\left[D\left|\{\partial\bar{\Psi}_{0,D}/\partial v_0\}|_{v^*_{0,\boldsymbol{D}}}\right|\right]^{-1/2}+O(D^{-1}),\\
&\exp\{-D\times\bar{L}_{0\boldsymbol{D}}(v^*_{0,\boldsymbol{D}})\}\cdot(2\pi)^{1/2}\left[D\left|\{\partial\bar{\Psi}_{0,D}/\partial v_0\}|_{v^*_{0,\boldsymbol{D}}}\right|\right]^{-1/2}+O(D^{-1})
\end{aligned}$$

respectively. Then, $\widetilde{v}_{0\boldsymbol{D}}(\boldsymbol{u}_{\boldsymbol{D}})=v^*_{0,\boldsymbol{D}}+O(D^{-1})$. Similarly for (15), from the Laplace approximation, $\widetilde{v}_{g\boldsymbol{D}}(\boldsymbol{u}_{g,d_g};\widetilde{v}_{0\boldsymbol{D}})=v^*_{g,d_g}(v^*_{0,\boldsymbol{D}})+O(D^{-1})$ for all g. Thus, the proxies $\widetilde{v}_{0\boldsymbol{D}}$ and $\widetilde{v}_{g\boldsymbol{D}}$ for $g\in\{1,\dots,G\}$ are consistent. □

B.3. Inequality for proof of Lipschitz continuity of proxies in factor (copula) models: Lemmas 8–10

Matrix Cauchy–Schwarz inequality: Let $\boldsymbol{S}$ and $\boldsymbol{T}$ be $D\times p_1$ and $D\times p_2$ matrices respectively. Let the ith row vector of the two matrices be $\boldsymbol{s}_i^\top$, $\boldsymbol{t}_i^\top$ respectively. Assume F-norm for the matrix norm and L^2-norm for vectors. For the $p_1\times p_2$ matrix $\boldsymbol{S}^\top\boldsymbol{T}$, then $\|D^{-1}\boldsymbol{S}^\top\boldsymbol{T}\|^2\le\left(D^{-1}\sum_{j=1}^D\|\boldsymbol{s}_j\|^2\right)\left(D^{-1}\sum_{j=1}^D\|\boldsymbol{t}_j\|^2\right)$.

Proof. Write

$$D^{-1}\boldsymbol{S}^\top\boldsymbol{T}=D^{-1}\sum_{j=1}^D\boldsymbol{s}_j\boldsymbol{t}_j^\top=\begin{bmatrix}\frac{1}{D}\sum_{j=1}^D s_{j1}t_{j1} & \cdots & \frac{1}{D}\sum_{j=1}^D s_{j1}t_{jp_2}\\ \vdots & \ddots & \vdots\\ \frac{1}{D}\sum_{j=1}^D s_{jp_1}t_{j1} & \cdots & \frac{1}{D}\sum_{j=1}^D s_{jp_1}t_{jp_2}\end{bmatrix}$$

For each entry in the matrix, the Cauchy–Schwarz inequality can apply. Take the (m,k) entry,

$$\left(\frac{1}{D}\sum_{j=1}^D s_{jm}t_{jk}\right)^2\le\left(\frac{1}{D}\sum_{j=1}^D s_{jm}^2\right)\left(\frac{1}{D}\sum_{j=1}^D t_{jk}^2\right)$$

Computing a bound on the sum of matrix components leads to:

$$\begin{aligned}
\left\|\frac{1}{D}\sum_{j=1}^D\boldsymbol{s}_j\boldsymbol{t}_j^\top\right\|^2 &\le \left(\frac{1}{D}\sum_{j=1}^D s_{j1}^2+\cdots+\frac{1}{D}\sum_{j=1}^D s_{jp_1}^2\right)\left(\frac{1}{D}\sum_{j=1}^D t_{j1}^2\right)\\
&+\ldots+\left(\frac{1}{D}\sum_{j=1}^D s_{j1}^2+\cdots+\frac{1}{D}\sum_{j=1}^D s_{jp_1}^2\right)\left(\frac{1}{D}\sum_{j=1}^D t_{jp_2}^2\right)=\left(\frac{1}{D}\sum_{j=1}^D\|\boldsymbol{s}_j\|^2\right)\left(\frac{1}{D}\sum_{j=1}^D\|\boldsymbol{t}_j\|^2\right).\quad\square
\end{aligned}$$

Proof of Lemma 8. The difference between $\widetilde{\boldsymbol{w}}_{\boldsymbol{D}}(\widehat{\boldsymbol{A}}_{\boldsymbol{D}})$ and $\widetilde{\boldsymbol{w}}_{\boldsymbol{D}}(\boldsymbol{A}_{\boldsymbol{D}})$ can be written as

$$\begin{aligned}
&(\mathbf{I}_p+\widehat{\boldsymbol{Q}}_{\boldsymbol{D}})^{-1}\widehat{\boldsymbol{A}}_{\boldsymbol{D}}^\top\widehat{\boldsymbol{\Psi}}_{\boldsymbol{D}}^{-2}\boldsymbol{z}_{\boldsymbol{D}}-(\mathbf{I}_p+\boldsymbol{Q}_{\boldsymbol{D}})^{-1}\boldsymbol{A}_{\boldsymbol{D}}^\top\boldsymbol{\Psi}_{\boldsymbol{D}}^{-2}\boldsymbol{z}_{\boldsymbol{D}}\\
&=\underbrace{(\mathbf{I}_p+\widehat{\boldsymbol{Q}}_{\boldsymbol{D}})^{-1}\widehat{\boldsymbol{A}}_{\boldsymbol{D}}^\top\widehat{\boldsymbol{\Psi}}_{\boldsymbol{D}}^{-2}\boldsymbol{z}_{\boldsymbol{D}}-(\mathbf{I}_p+\boldsymbol{Q}_{\boldsymbol{D}})^{-1}\widehat{\boldsymbol{A}}_{\boldsymbol{D}}^\top\widehat{\boldsymbol{\Psi}}_{\boldsymbol{D}}^{-2}\boldsymbol{z}_{\boldsymbol{D}}}_{\text{term1}}+\underbrace{(\mathbf{I}_p+\boldsymbol{Q}_{\boldsymbol{D}})^{-1}\widehat{\boldsymbol{A}}_{\boldsymbol{D}}^\top\widehat{\boldsymbol{\Psi}}_{\boldsymbol{D}}^{-2}\boldsymbol{z}_{\boldsymbol{D}}-(\mathbf{I}_p+\boldsymbol{Q}_{\boldsymbol{D}})^{-1}\boldsymbol{A}_{\boldsymbol{D}}^\top\boldsymbol{\Psi}_{\boldsymbol{D}}^{-2}\boldsymbol{z}_{\boldsymbol{D}}}_{\text{term2}}.
\end{aligned}\tag{32}$$

Let $\bar{\boldsymbol{Q}}_{\boldsymbol{D}}=D^{-1}\boldsymbol{Q}_{\boldsymbol{D}}$, $\bar{\widehat{\boldsymbol{Q}}}_{\boldsymbol{D}}=D^{-1}\widehat{\boldsymbol{Q}}_{\boldsymbol{D}}$ with (by assumption) $\bar{\boldsymbol{Q}}_{\boldsymbol{D}}\to\boldsymbol{Q}$, $\bar{\widehat{\boldsymbol{Q}}}_{\boldsymbol{D}}\to\widehat{\boldsymbol{Q}}$ as $\boldsymbol{D}\to\infty$ where $\boldsymbol{Q},\widehat{\boldsymbol{Q}}$ are positive definite matrix. Note that $(\mathbf{I}_p+\boldsymbol{Q}_{\boldsymbol{D}})^{-1}=O(D^{-1})$, $(\mathbf{I}_p+\widehat{\boldsymbol{Q}}_{\boldsymbol{D}})^{-1}=O(D^{-1})$. Let $\boldsymbol{H}_{\boldsymbol{D}}=D\cdot(\mathbf{I}_p+\boldsymbol{Q}_{\boldsymbol{D}})^{-1}$, so that $\boldsymbol{H}_{\boldsymbol{D}}=O(1)$. Since $\bar{\boldsymbol{Q}}_{\boldsymbol{D}}$ and $\bar{\widehat{\boldsymbol{Q}}}_{\boldsymbol{D}}$ are both positive definite and well-conditioned, then there is bound on the condition numbers of $\bar{\boldsymbol{Q}}_{\boldsymbol{D}}$ and $\bar{\widehat{\boldsymbol{Q}}}_{\boldsymbol{D}}$ for all large D, and $\|\bar{\widehat{\boldsymbol{Q}}}_{\boldsymbol{D}}^{-1}-\bar{\boldsymbol{Q}}_{\boldsymbol{D}}^{-1}\|=O(\|\bar{\widehat{\boldsymbol{Q}}}_{\boldsymbol{D}}-\bar{\boldsymbol{Q}}_{\boldsymbol{D}}\|)$. Then, term1 in (32) has the order of $(\bar{\widehat{\boldsymbol{Q}}}_{\boldsymbol{D}}^{-1}-\bar{\boldsymbol{Q}}_{\boldsymbol{D}}^{-1})\cdot D^{-1}\cdot\widehat{\boldsymbol{A}}_{\boldsymbol{D}}^\top\widehat{\boldsymbol{\Psi}}_{\boldsymbol{D}}^{-2}\boldsymbol{z}_{\boldsymbol{D}}$.

For simplicity, we suppress the subscript of $\bar{\boldsymbol{Q}}_{\boldsymbol{D}}$, $\bar{\widehat{\boldsymbol{Q}}}_{\boldsymbol{D}}$, $\boldsymbol{A}_{\boldsymbol{D}}$ and $\boldsymbol{\Psi}_{\boldsymbol{D}}$ below. Then,

$$\begin{aligned}
\bar{\widehat{\boldsymbol{Q}}}-\bar{\boldsymbol{Q}}&=D^{-1}(\widehat{\boldsymbol{A}}^\top\widehat{\boldsymbol{\Psi}}^{-2}\widehat{\boldsymbol{A}}-\boldsymbol{A}^\top\boldsymbol{\Psi}^{-2}\boldsymbol{A})\\
&=D^{-1}\left((\widehat{\boldsymbol{A}}^\top-\boldsymbol{A}^\top+\boldsymbol{A}^\top)\widehat{\boldsymbol{\Psi}}^{-2}(\widehat{\boldsymbol{A}}-\boldsymbol{A}+\boldsymbol{A})-\boldsymbol{A}^\top\boldsymbol{\Psi}^{-2}\boldsymbol{A}\right)
\end{aligned}$$

$$= D^{-1}\left(\underbrace{A^\top\hat{\Psi}^{-2}(\hat{A}-A)}_{\text{term3}}+\underbrace{(\hat{A}^\top-A^\top)\hat{\Psi}^{-2}A}_{\text{term4}}+\underbrace{(\hat{A}^\top-A^\top)\hat{\Psi}^{-2}(\hat{A}-A)}_{\text{term5}}+\underbrace{A^\top(\hat{\Psi}^{-2}-\Psi^{-2})A}_{\text{term6}}\right). \tag{33}$$

Since term5 is negligible compared to other terms, we only look at the order of the other three terms in the right-hand side of (33). For term3, term4, term6, multiplied by D^{-1}, the matrix Cauchy–Schwarz inequality, with $S=\hat{\Psi}^{-1}A$ and $T=\hat{\Psi}^{-1}(\hat{A}-A)$ for term3 (term4), and $S=(\hat{\Psi}^{-2}-\Psi^{-2})A$ and $T=A$ for term6, leads to:

$$\|D^{-1}A^\top\hat{\Psi}^{-2}(\hat{A}-A)\| = \left\|D^{-1}\sum_{j=1}^{D}\frac{a_j(\hat{a}_j-a_j)^\top}{\hat{\psi}_j^2}\right\| \le \left(D^{-1}\sum_{j=1}^{D}\frac{\|a_j\|^2}{\hat{\psi}_j^2}\right)^{1/2}\cdot\left(D^{-1}\sum_{j=1}^{D}\frac{\|\hat{a}_j-a_j\|^2}{\hat{\psi}_j^2}\right)^{1/2},$$

$$\|D^{-1}A^\top(\hat{\Psi}^{-2}-\Psi^{-2})A\| = \left\|D^{-1}\sum_{j=1}^{D}a_ja_j^\top(\hat{\psi}_j^{-2}-\psi_j^{-2})\right\| \le \left(D^{-1}\sum_{j=1}^{D}\|a_j(\hat{\psi}_j^{-2}-\psi_j^{-2})\|^2\right)\cdot\left(D^{-1}\sum_{j=1}^{D}\|a_j\|^2\right)^{1/2}.$$

For all j, $\|a_j\|<1$, ψ_j is bounded from zero, term3 and term4 in (33) multiplied by D^{-1} are all $O\left(\sqrt{D^{-1}\sum_{j=1}^{D}\|\hat{a}_j-a_j\|^2/\hat{\psi}_j^2}\right)$. The term6 in (33) multiplied by D^{-1} is $O\left(\sqrt{D^{-1}\sum_{j=1}^{D}(\hat{\psi}_j^{-2}-\psi_j^{-2})^2}\right)$.

Also, from the matrix Cauchy–Schwarz inequality with $S=\hat{\Psi}_D^{-2}A_D$ and $T=z_D$,

$$\|D^{-1}\hat{A}_D^\top\hat{\Psi}_D^{-2}z_D\| = \left\|D^{-1}\sum_{j=1}^{D}\frac{\hat{a}_jz_j}{\hat{\psi}_j^2}\right\| \le \left(D^{-1}\sum_{j=1}^{D}\|\hat{a}_j\|^2/\hat{\psi}_j^4\right)^{1/2}\left(D^{-1}\sum_{j=1}^{D}z_j^2\right)^{1/2}. \tag{34}$$

Since the bound of (34) is $O(1)$, then term1 in (32) is order $O\left(\sqrt{D^{-1}\sum_{j=1}^{D}\|\hat{a}_j-a_j\|^2/\hat{\psi}_j^2}\right)+O\left(\sqrt{D^{-1}\sum_{j=1}^{D}(\hat{\psi}_j^{-2}-\psi_j^{-2})^2}\right)$.

For term2 in (32),

$$H_D\cdot D^{-1}\cdot\left((A_D^\top+\hat{A}_D^\top-A_D^\top)(\Psi_D^{-2}+\hat{\Psi}_D^{-2}-\Psi_D^{-2})z_D-A_D^\top\Psi_D^{-2}z_D\right) \tag{35}$$

$$= H_D\cdot D^{-1}\cdot\left(\underbrace{A_D^\top(\hat{\Psi}_D^{-2}-\Psi_D^{-2})z_D}_{\text{term7}}+\underbrace{(\hat{A}_D^\top-A_D^\top)\Psi_D^{-2}z_D}_{\text{term8}}+\underbrace{(\hat{A}_D^\top-A_D^\top)(\hat{\Psi}_D^{-2}-\Psi_D^{-2})z_D}_{\text{term9}}\right).$$

Since term9 in (35) is negligible in comparison, we only look at the first two terms. From applying the Cauchy–Schwartz inequality as above, for term7 and term8 in (35) multiplied by D^{-1},

$$D^{-1}\cdot\text{term7}=O\left(\left\{D^{-1}\sum_{j=1}^{D}(\hat{\psi}_j^{-2}-\psi_j^{-2})^2\right\}^{1/2}\right),\quad D^{-1}\cdot\text{term8}=O\left(\left\{D^{-1}\sum_{j=1}^{D}\|\hat{a}_j-a_j\|^2\right\}^{1/2}\right).$$

Since $H_D=O(1)$, then the term2 in (32) has the same order as term1 in (32). Also, due to the ψ_j being bounded away from 0 and $\psi_j^2=1-\|a_j\|^2$, then $\|\tilde{\omega}_D(\hat{A}_D)-\tilde{\omega}_D(A_D)\|=O\left(\sqrt{D^{-1}\sum_{j=1}^{D}\|\hat{a}_j-a_j\|^2}\right)$. This indicates there exists a Lipschitz constant K_D such that $\|\tilde{\omega}_D(\hat{A}_D)-\tilde{\omega}_D(A_D)\|\le K_D\cdot\sqrt{D^{-1}\sum_{j=1}^{D}\|\hat{a}_j-a_j\|^2}$. $\{K_D\}$ is bounded as $D\to\infty$ if the a_j are assumed to be sampled from a super-population with the corresponding ψ_j^2 bounded away from 0. □

Proof of Lemma 9. Let $\gamma:[0,1]\to\bar{B}(\theta_D,\rho)$ be the path $\gamma(t)=t\hat{\theta}_D+(1-t)\theta_D$ from θ_D to $\hat{\theta}_D$ in $\bar{B}(\theta_D,\rho)$. For simplicity, suppress the subscript for $\tilde{v}_D$ below. Then

$$\|\tilde{v}(\hat{\theta}_D)-\tilde{v}(\theta_D)\|=\|\tilde{v}(\gamma(1))-\tilde{v}(\gamma(0))\|=\left\|\int_0^1\frac{\mathrm{d}\tilde{v}(\gamma(t))}{\mathrm{d}t}\,\mathrm{d}t\right\|$$

$$=\left\|\int_0^1\nabla\tilde{v}(\gamma(t))\cdot(\hat{\theta}_D-\theta_D)\,\mathrm{d}t\right\|\le\|\hat{\theta}_D-\theta_D\|\int_0^1\|\nabla\tilde{v}(\gamma(t))\|\,\mathrm{d}t<K_D\|\hat{\theta}_D-\theta_D\|,$$

where $K_D:=\sup\{\|\nabla\tilde{v}(\theta)\|:\theta\in\bar{B}(\theta_D,\rho)\}$, and $\|\cdot\|$ is the l_2 norm. Next we derive order of the Lipschitz constant.

Let $f_D(v,\theta_D)=c_{VU_D}(u_{1:D},v;\theta_D)=\exp\{\sum_{j=1}^{D}\log c_{jV}(u_j,v;\theta_j)\}$ be the density function of 1-factor copula model defined in (1), the jth element of the gradient vector is

$$\frac{\partial\tilde{v}(\theta_D)}{\partial\theta_j}=\frac{\int_0^1(v\,\partial f_D(v,\theta_D)/\partial\theta_j)\,\mathrm{d}v\cdot\int_0^1 f_D(v,\theta_D)\,\mathrm{d}v-\int_0^1(\partial f_D(v,\theta_D)/\partial\theta_j)\,\mathrm{d}v\cdot\int_0^1 vf_D(v,\theta_D)\,\mathrm{d}v}{\left(\int_0^1 f_D(v,\theta_D)\,\mathrm{d}v\right)^2}$$

$$=\left(\int_0^1 f_D(v,\theta_D)\,\mathrm{d}v\right)^{-1}\left[\int_0^1(v\,\partial f_D(v,\theta_D)/\partial\theta_j)\,\mathrm{d}v-\tilde{v}_D(\theta_D)\cdot\int_0^1(\partial f_D(v,\theta_D)/\partial\theta_j)\,\mathrm{d}v\right] \tag{36}$$

$$=\left(\int_0^1 f_D(v,\theta_D)\,\mathrm{d}v\right)^{-1}\left(\int_0^1[v-\tilde{v}_D(\theta_D)]\cdot(\partial f_D(v,\theta_D)/\partial\theta_j)\,\mathrm{d}v\right).$$

In (36), $\widetilde{v}_D = \widetilde{v}_D(\boldsymbol{\theta}_D)$ is the proxy variable (13) defined in 1-factor copula model (1). Also since $\partial f_D(v, \boldsymbol{\theta}_D)/\partial \boldsymbol{\theta}_j = f_D(v, \boldsymbol{\theta}_D)\left(\partial \log c_{jV}(u_j, v; \boldsymbol{\theta}_j)/\partial \boldsymbol{\theta}_j\right)$, then

$$\frac{\partial \widetilde{v}(\boldsymbol{\theta}_D)}{\partial \boldsymbol{\theta}_j} = \frac{\int_0^1 (v - \widetilde{v}_D)\, f_D(v, \boldsymbol{\theta}_D) \cdot (\partial \log c_{jV}(u_j, v; \boldsymbol{\theta}_j)/\partial \boldsymbol{\theta}_j)\, \mathrm{d}v}{\int_0^1 f_D(v, \boldsymbol{\theta}_D)\, \mathrm{d}v}.$$

It has the same order as

$$\int_0^1 (v - \widetilde{v}_D) \cdot |(\partial \log c_{jV}(u_j, v; \boldsymbol{\theta}_j)/\partial \boldsymbol{\theta}_j)|\, f_D(v, \boldsymbol{\theta}_D)\, \mathrm{d}v \Big/ \int_0^1 f_D(v, \boldsymbol{\theta}_D)\mathrm{d}v. \tag{37}$$

In (37), let $m_j(v) = \partial \log c_{jV}(u_j, v; \boldsymbol{\theta}_j)/\partial \boldsymbol{\theta}_j$ and $h_D(v) = \bar{L}_D(v) = -D^{-1} \log f_D(v, \boldsymbol{\theta}_D)$. Let $v_D^* = \arg\min h_D(v)$. Let $t_j(v) = (v - \widetilde{v}_D)\,|m_j(v)|$, $t_j'(v) = |m_j(v)| + (v - \widetilde{v}_D)\,(\partial |m_j(v)|/\partial v)$ and $t_j''(v) = 2(\partial |m_j(v)|/\partial v) + (v - \widetilde{v}_D)\,(\partial^2 |m_j(v)|/\partial v^2)$. The given assumptions on $\bar{L}_D(v)$ imply that the Laplace approximation can be used. From equation (2.6) in [20], Eq. (37) becomes

$$(v_D^* - \widetilde{v}_D)\,|m_j(v_D^*)| + (2D)^{-1}\,[h_D''(v_D^*)]^{-1}\, t_j''(v_D^*) - (2D)^{-1}[h_D''(v_D^*)]^{-2}\, t_j'(v_D^*)\, h_D'''(v_D^*) + O(D^{-2}). \tag{38}$$

Under the assumptions on the bounded derivatives, analogous to the proof in Appendix B.2, $\partial \widetilde{v}(\boldsymbol{\theta}_D)/\partial \theta_j = O(D^{-1})$. Then, the norm of derivatives $\|\nabla \widetilde{v}(\boldsymbol{\theta})\|$ equals to $\{\sum_{j=1}^D |\partial \widetilde{v}(\boldsymbol{\theta}_D)/\partial \boldsymbol{\theta}_j|^2\}^{1/2} = O(D^{-1/2})$ (or $K_D = O(D^{-1/2})$), and $|\widetilde{v}(\hat{\boldsymbol{\theta}}_D) - \widetilde{v}(\boldsymbol{\theta}_D)| = O\left(\{D^{-1} \sum_{j=1}^D \|\hat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j\|_2^2\}^{1/2}\right)$. This indicates there exists a Lipschitz constant $\boldsymbol{B}_D$ such that $\|\widetilde{v}(\hat{\boldsymbol{\theta}}_D) - \widetilde{v}(\boldsymbol{\theta}_D)\| \le \boldsymbol{B}_D \|\hat{\boldsymbol{\theta}}_D - \boldsymbol{\theta}_D\|^*$. $\boldsymbol{B}_D$ is bounded as $D \to \infty$ if the $\boldsymbol{\theta}_j$ are assumed to be sampled from a super-population satisfying the assumptions. □

The proof of Lemma 10 is similar to that of Lemma 9 and appears in the Supplementary material.

Appendix C. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jmva.2023.105263.

References

[1] T.W. Anderson, Y. Amemiya, The asymptotic normal distribution of estimators in factor analysis under general conditions, Ann. Statist. 16 (2) (1988) 759–771.
[2] T.W. Anderson, H. Rubin, Statistical inference in factor analysis, in: Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability. vol. 5, University of California Press, Berkeley, CA, 1956, pp. 111–150.
[3] J. Bai, K. Li, Statistical analysis of factor models of high dimension, Ann. Statist. 40 (1) (2012) 436–465.
[4] J. Bai, K. Li, Maximum likelihood estimation and inference for approximate factor models of high dimension, Rev. Econ. Stat. 98 (2) (2016) 298–309.
[5] R.A. Bradley, J.J. Gart, The asymptotic properties of ML estimators when sampling from associated populations, Biometrika 49 (1/2) (1962) 205–214.
[6] K.W. Breitung, Asymptotic Approximations for Probability Integrals, Springer, Berlin, 1994.
[7] G. Chamberlain, M. Rothschild, Arbitrage, factor structure, and mean-variance analysis on large asset markets, Econometrica (1983) 1281–1304.
[8] H. Cramér, Problems in probability theory, Ann. Math. Stat. 18 (2) (1947) 165–193.
[9] G. Fang, J. Guo, X. Xu, Z. Ying, S. Zhang, Identifiability of bifactor models, Statist. Sinica 31 (2021) 2309–2330.
[10] R.L. Gorsuch, Factor Analysis, Erlbaum Associates, Hillsdale, New Jersey, 1983.
[11] H. Joe, Dependence Modeling with Copulas, Chapman & Hall/CRC, Boca Raton, FL, 2014.
[12] H. Joe, Parsimonious graphical dependence models constructed from vines, Canad. J. Statist. 46 (4) (2018) 532–555.
[13] R.A. Johnson, D.W. Wichern, Applied Multivariate Statistical Analysis, fifth ed., Prentice Hall, Englewood Cliffs, NJ, 2002.
[14] K.G. Jöreskog, D. Sörbom, J. Magidson, Advances in Factor Analysis and Structural Equation Models, Abt Books, Cambridge, MA, 1979.
[15] P. Krupskii, H. Joe, Factor copula models for multivariate data, J. Multivariate Anal. 120 (2013) 85–101.
[16] P. Krupskii, H. Joe, Structured factor copula models: Theory, inference and computation, J. Multivariate Anal. 138 (2015) 53–73.
[17] P. Krupskii, H. Joe, Approximate likelihood with proxy variables for parameter estimation in high-dimensional factor copula models, Statist. Papers 63 (2022) 543–569.
[18] H. Nguyen, M.C. Ausin, P. Galeano, Variational inference for high dimensional structured factor copulas, Comput. Statist. Data Anal. 151 (2020) 107012.
[19] B.K. Tan, A. Panagiotelis, G. Athanasopoulos, Bayesian inference for the one-factor copula model, J. Comput. Graph. Statist. 28 (1) (2019) 155–173.
[20] L. Tierney, R.E. Kass, J.B. Kadane, Fully exponential Laplace approximations to expectations and variances of nonpositive functions, J. Amer. Statist. Assoc. 84 (407) (1989) 710–716.