High-dimensional factor copula models with estimation of latent variables

Xinyao Fan[1,*], Harry Joe[1]

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

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, 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.

Keywords: bi-factor, factor scores, identifiability, oblique factor, proxy variables, tail dependence

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 [12] and [13] 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, estimates of the latent variables referred to factor scores (see [11] and [8]) are of interest and useful for interpretation and further analysis. For similar reasons, the inference of the latent variables in factor copulas is also useful.

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. [12] and [13] 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 Gaussian-Legendre quadrature. Bi-factor and oblique factor models are useful when the observed variables can be placed into several non-overlapping homogeneous groups.

In this paper, one main focus for factor copulas is to show how 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 their previous numerical 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.

[14] initiate the use of proxies for latent variables to speed up numerical maximum likelihood estimation; their approach involved unweighted means in 1-factor and unweighted group means for oblique factor copula models. Their 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

---

*Corresponding author. Email address: `xinyao.fan@stat.ubc.ca`

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).

To justify the sequential method for latent variable and copula model estimation, several theoretical results in consistency, as the number D of observed variables increase to infinity, are needed. 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.

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 a super-population. We first obtain conditions for the proxies or conditional expectations to be asymptotically consistent estimates (of corresponding latent variables) when the the factor model is completely known; we also have results that suggest rates of convergence. In cases where consistency is not possible, then we know that we cannot expect consistency when 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 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 with $\sqrt{N}$ consistency and (ii) proxies are consistent under mild conditions. Because we need a method of proof that is valid for both Gaussian and factor copula models, our technique is different from that of [2]. Their approach does not provide insights for non-identifiability of latent variables such as in the bi-factor model.

The remainder of the paper is organized as follows. Section 2 provides the representations of Gaussian factor models and (structured) factor copula models. Section 3 has expressions for conditional expectations, and the proxies as estimates of the latent variables. Section 4 has 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. Simulation studies in Section 7 show the proxies are useful in selecting linking copula families and getting accurate parameter estimates with less computing time. Section 8 has sufficient conditions for using the proxies in the Section 3 when observed variables have weak dependence, rather than independence, conditional on the observed variables. Section 9 has a summary and discussion for further research.

2. Structured factor copula models

The 1-factor, bi-factor and oblique factor models that are the main focus of this paper are shown graphically in Figure 1, 2 and 3 respectively. The graphical representations are valid for the Gaussian factor models and their extensions to factor copula models.
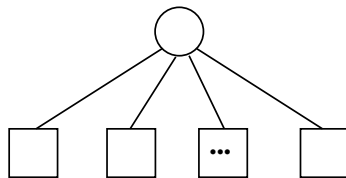

Fig. 1: 1-factor model, a sequence of the observed variables (rectangular shapes) are linked to the latent variable (circles).
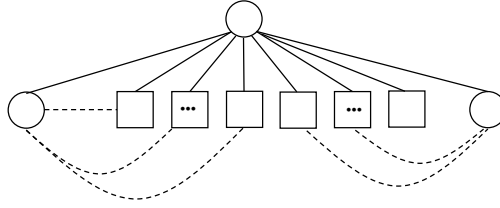
Fig. 2: Bi-factor model, G local latent variables (circles) and a sequence of observed variables (rectangles) are all linked to the global latent variable (circle; root variable) in solid lines; the dashed lines indicate the conditional dependence between observed variables and local latent variables conditioned on the global latent variable. Illustration here has $G = 2$ groups.
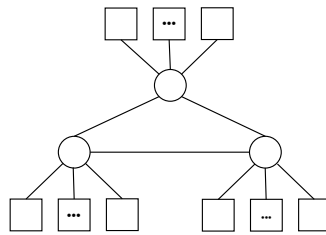

Fig. 3: Oblique factor model, observed variables (rectangles) are divided into the several non-overlapping groups, latent variables (circles) are dependent and the observed variables in each group are linked to the (local) group latent variable. Illustration here has $G = 3$ groups.

p-factor models with $p \geq 2$ are in general do not have an identifiable loading matrix in the Gaussian case because of 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 [9]. 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 leads 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 are their derivations are in Section 3.10 and 3.11 of [9], [12], [13] and [14].

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 densities for different vectors are indicated using C with subscripts for random vectors. The generic notation for a bivariate copula 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

3

used as arguments of densities or dummy variables of integrals.

1-factor copula model with D variables $\boldsymbol{U}_D = (U_1, \ldots, U_D)$, the copula density is:

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

where $c_{jV} = c_{U_j,V}$ for all j.

Bi-factor copula model with d_g variables in group g, fixed G groups, $\boldsymbol{D} = (d_1, \ldots, d_G)$ and $D = \sum_{g=1}^G d_g$ total number of variables. The copula density is:

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

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

Oblique factor copula model with d_g variables in group g as above. The copula density is:

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

The notation $c_{\boldsymbol{V}}(v)$ is the joint copula density of the 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

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

where $W, \epsilon_1, \epsilon_2, \ldots$ are mutually independent $N(0,1)$ random variables, and $-1 < \alpha_j < 1$ and $\psi_j^2 = 1 - \alpha_j^2$ for all j.

Gaussian bi-factor

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

where $W_0, \{W_g\}, \{\epsilon_{jg}\}$ are mutually independent $N(0,1)$, α's are in $(-1,1)$ 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

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

where $\{\epsilon_{jg}\}$ are mutually independent $N(0,1)$, independent of the multivariate normal vector $(W_1, \ldots, W_G)$, with zero mean vector and unit variances, α's are in $(-1,1)$ and $\psi_{jg}^2 = 1 - \alpha_{jg}^2$ for all (j,g). Let Σ_W be the correlation matrix of $\boldsymbol{W} = (W_1, \ldots, W_G)^T$.

Matrix representation

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

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 $\boldsymbol{W} = (W_0, W_1, \ldots, W_G)^T$, and for oblique factor, $p = G$ and $\boldsymbol{W} = (W_1, \ldots, W_G)^T$.

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

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

when $\boldsymbol{A}_D \boldsymbol{A}_D^T + \boldsymbol{\Psi}_D^2$ is non-singular and $\boldsymbol{\Psi}_D^2$ has no zeros on the diagonal. This identity is given in [11].

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, [12] and [13] 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 ([11]), 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 [12] and [13] 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 [11]. 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 [14].

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 [14] 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 Section 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 $E(W|Z_D = z_D)$ for 1-factor, 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.

p-**factor**: Let (w^0, z_D) be a realization of (W, Z_D). The proxy for W (or estimate of w^0) given z_D are:

$$\widetilde{w}_D = E(W|Z_D = z_D) = A_D{}^T(A_D A_D^T + \Psi_D^2)^{-1} z_D = (I_p + A_D{}^T \Psi_D^{-2} A_D)^{-1} A_D^T \Psi_D^{-2} z_D, \tag{9}$$

if $(A_D A_D^T + \Psi_D^2)$ is non-singular and Ψ_D has no zeros on diagonal. The above matrix equality follows from (8).

If Ψ_D has zeros on diagonal, then a linear combination of the latent variables is an observed variable. If $(A_D A_D^T + \Psi_D^2)$ is singular, then a linear combination of observed variables is a constant. These unrealistic cases will not be considered.

Bi-factor: Let $(w_0^0, w_1^0, \ldots, w_G^0, z_D)$ be a realization of $(W_0, W_1, \ldots, W_G, Z_D)$. The proxies (or estimates of $w^0, w_1^0, \ldots, w_G^0$) given z_D is:

$$\widetilde{w}_0 = \mathrm{E}(W_0 | Z_D = z_D) = (a_0)^T (A_D A_D^T + \Psi_D^2)^{-1} z_D, \tag{10}$$

$$\widetilde{w}_g(\widetilde{w}_0) = \mathrm{E}(W_g | Z_D = z_D, W_0 = \widetilde{w}_0) = (b_g^T, 0)(\Sigma_g)^{-1}(z_g^T, \widetilde{w}_0)^T, \tag{11}$$

where a_0 is the first column of the loading matrix, $z_g = (z_{1g}, \ldots, z_{d_g g})^T$, b_{0g} and b_g are the $d_g \times 1$ global and local loading vector for group g; let $B_g = [b_{0g}, b_g]$ (matrix of size $d_g \times 2$). Let Σ_g be the correlation matrix of (Z_g^T, W_0). Then $\Sigma_g = \begin{bmatrix} B_g B_g^T + \Psi_g^2 & b_{0g} \\ b_{0g}^T & 1 \end{bmatrix}$ for $g \in \{1, \ldots, G\}$. The proof that $\widetilde{w}_g(\widetilde{w}_0) = \mathrm{E}(W_g | Z_D = z_D)$ is given in the Appendix A.1.

Oblique factor: Let $(w_1^0, \ldots, w_D^0, z_D)$ be a realization of $(W_1, \ldots, W_G, Z_D)$. The proxies (or estimates of $w_1^0, \ldots, w_G^0$) given z_D are:

$$\widetilde{w}_g = \mathrm{E}(W_g | Z_g = z_g) = a_g^T (a_g a_g^T + \Psi_g^2)^{-1} z_g, \tag{12}$$

where $z_g = (z_{1g}, \ldots, z_{d_g g})^T$, Ψ_g is the g-th block diagonal of Ψ and a_g is the $d_g \times 1$ loading vector for group g, for $g \in \{1, \ldots, G\}$. This version, rather than $\mathrm{E}(W_g | Z_h = z_h, h \in \{1, \ldots, G\})$, has a version for the oblique factor copula that is numerically easier to handle.

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.

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

$$\widetilde{v}_D = \widetilde{v}_D(u_D) = \mathrm{E}(V | U_D = u_D) = \frac{\int_0^1 v c_{VU_D}(v, u_D) \, dv}{c_{U_D}(u_D)}. \tag{13}$$

Bi-factor copula model with density as defined 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 $V_1, \ldots, V_G$.

vector	joint density
(U_D, V_0, V)	$c_{U_D, V_0, V}(u_D, v_0, v) = \prod_{g=1}^{G} \prod_{j=1}^{d_g} \left\{ c_{U_{jg} V_0}(u_{jg}, v_0) \cdot c_{U_{jg} V_g; V_0}(C_{U_{jg}|V_0}(u_{jg}|v_0), v_g) \right\}$
(U_D, V_0)	$c_{U_D, V_0}(u, v_0) = \prod_{g=1}^{G} \left\{ \left[\prod_{j=1}^{d_g} c_{U_{jg} V_0}(u_{jg}, v_0) \right] \cdot f_g(u_g; v_0) \right\}$ $f_g(u_g; v_0) = \int_0^1 \prod_{j=1}^{d_g} c_{U_{jg} V_g; V_0}\left(C_{U_{jg}|V_0}(u_{jg}|v_0), v_g \right) dv_g, \quad g \in \{1, \ldots, G\}$
(U_g, V_0, V_g)	$c_{U_g, V_0, V_g}(u_g, v_0, v_g) = \prod_{j=1}^{d_g} \left\{ c_{U_{jg} V_0}(u_{jg}, v_0) \cdot c_{U_{jg} V_g; V_0}\left(C_{U_{jg}|V_0}(u_{jg}|v_0), v_g \right) \right\}$

Table 1: Densities in the bi-factor copula model (2).

The convenient form of conditional expectations is as follows:

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

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

$$
\begin{aligned}
\widetilde{v}_{gD}(\boldsymbol{u}_g, \widetilde{v}_{0D}) &= \mathrm{E}[V_g | V_0 = \widetilde{v}_{0D}, U_{jg} = u_{jg}, j \in \{1, \ldots, d_g\}] \\
&= \frac{\int_0^1 v_g c_{U_g, V_0, V_g}(\boldsymbol{u}_g, \widetilde{v}_{0D}, v_g) \, \mathrm{d}v_g}{\int_0^1 c_{U_g, V_0, V_g}(\boldsymbol{u}_g, \widetilde{v}_{0D}, v_g) \, \mathrm{d}v_g}.
\end{aligned}
\tag{15}
$$

The proxies are $\widetilde{v}_{0D}(\boldsymbol{u_D})$ and $\widetilde{v}_{gD}(\boldsymbol{u}_g, \widetilde{v}_{0D})$ for $g \in \{1, \ldots, G\}$.

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

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

The proxy for v_g^0 is $\widetilde{v}_{gD}(\boldsymbol{u}_g)$ 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.

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

model	conditional variance
p-factor	$\text{Cov}(\boldsymbol{W}\|\boldsymbol{Z}_D = z_D) = \mathbf{I}_p - A_D^T(A_DA_D^T + \boldsymbol{\Psi}_D^2)^{-1}A_D$ $= \mathbf{I}_p - (\mathbf{I}_p + A_D^T\boldsymbol{\Psi}_D^{-2}A_D)^{-1}A_D^T\boldsymbol{\Psi}_D^{-2}A_D$
bi-factor	$\text{Var}(W_0\|\boldsymbol{Z}_D = z_D) = 1 - \boldsymbol{a}_0^T(A_DA_D^T + \boldsymbol{\Psi}_D^2)^{-1}\boldsymbol{a}_0$ $\text{Var}(W_g\|\boldsymbol{Z}_D = z_D) = (1+q_g)^{-1}+\widetilde{q}_g^2(1+q_g)^{-2}\,\text{Var}(W_0\|\boldsymbol{Z}_D = z_D)$ (17)
oblique factor	$\text{Cov}(\boldsymbol{W}\|\boldsymbol{Z} = z_D) = \Sigma_W - \Sigma_W A_D^T(A_D\Sigma_W A_D^T + \boldsymbol{\Psi}_D^2)^{-1}A_D\Sigma_W$ (18)

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

Limit of covariance for p-factor In the (second) expression of conditional variance for the general p-factor models, define $\boldsymbol{Q}_D = A_D^T\boldsymbol{\Psi}_D^{-2}A_D$ as a $p \times p$ matrix. Suppose $\bar{\boldsymbol{Q}}_D =: D^{-1}\boldsymbol{Q}_D \to \boldsymbol{Q}$ 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 $A_D = \mathbf{0}$, then $\text{Cov}(\boldsymbol{W}\|\boldsymbol{Z}_D = z_D) = \mathbf{I}_p$, that is, $\boldsymbol{Z}_D$ provides no information about $\boldsymbol{W}$. If $\boldsymbol{Q}$ is a positive definite matrix, then

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

In the p-factor model, if the loading matrices in $\{A_D\}$ 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).
(a) If A_D does not have full column rank, then the latent variables in $\boldsymbol{W}$ are not identifiable.
(b) For the Gaussian bi-factor model as a special case of p-factor with $p = G + 1$, if 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. (a) Let $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_p$ be the columns of A_D. The columns of 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 $(t_2, \ldots, t_p)$ is a non-zero vector. Then, in (7),

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

Hence only some linear combinations of the latent variables can be identified.
(b) For bi-factor, let $\boldsymbol{a}_0, \boldsymbol{a}_1, \ldots, \boldsymbol{a}_G$ be the columns of A_D, and let the latent variables be $W_0, W_1, \ldots, W_G$. The above implies that

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

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

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 $\boldsymbol{Q}_D = A_D^T\boldsymbol{\Psi}_D^{-2}A_D$ because $\boldsymbol{Q}_D$ is not of full rank if 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.

8

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 (18) becomes:

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

and

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

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

Rate of convergence as a function of strength of dependence with latent variables.

- For 1-factor model, $\boldsymbol{Q}_D$ is 1×1 and can be denoted by the scalar q_D. Assume that $\bar{q}_D = D^{-1}q_D := D^{-1}\sum_{j=1}^{D}\alpha_j^2/(1-\alpha_j^2) \to q > 0$, so that the dominating term of the conditional variance is $D^{-1}q^{-1}$. If one or several entries of loadings increase, q^{-1} decreases; that is, the convergence will be faster.

- The bi-factor case with a full-rank $\boldsymbol{A}_D$ loading matrix is more complicated. We consider several cases to understand how the dependence affects the conditional variance, and all the claims are inferred from conditional variance decomposition formula (17) shown in Table 2 and numerical experiments.

  case a) If the global loadings are constant, then the group with stronger dependence has a smaller conditional variance, thus a faster convergence rate. From (17), the conditional variance of local latent factors will be most affected by the first term because in the second term, the conditional variance of W_0 is fixed for all groups, and the ratio $\widetilde{q_g^2}/(1 + q_g)^2$ will not differ much.

  case b) Suppose $\boldsymbol{A}_D$ is a well-conditioned matrix under small perturbations. If the global loadings are fixed, and the local loadings for one group, for example, g' increase a little. From the numerical experiments, the conditional variance for W_g, $g = 0$ or $g \neq g'$ could increase or decrease a little, but $\text{Var}(W_{g'}|\boldsymbol{Z}_D)$ will decrease. The argument can also be inferred from (17): as $\boldsymbol{b}_{g'}$ increases, the first term decreases, and it dominates the change of conditional variance.

  case c) Suppose $\boldsymbol{A}_D$ is a well-conditioned matrix under small perturbations. If the global loadings increase a little but the local loadings remain the same, $\text{Var}(W_0|\boldsymbol{Z}_D)$ will decrease according to the numerical results.

- In the oblique factor model, assuming that the groups are similar in size and increase to infinity, with no group dominating, the conditional variance of the latent variables is dominated by the term $D^{-1}\boldsymbol{Q}^{-1}$, where $\boldsymbol{Q}$ is diagonal and (g, g) entry being the limit of $D^{-1}\sum_{j=1}^{d_g}\alpha_{jg}^2/\psi_{jg}^2$ as $D \to \infty$. Therefore, the conditional covariance of latent variables is closely related to the strength of within-group dependence. The proxy variable in the group with relatively strong dependence will have smaller conditional variance.

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.

Theorem 2. *(Asymptotic properties of factor scores in 1-factor Gaussian model) For* (4), *suppose there is a realized infinite sequence of observed variables* $z_1, z_2, \ldots$ *and a realized infinite sequence of disturbance terms* $e_1, e_2, \ldots$ *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* $\mathbf{z}_D = (z_1, \ldots, z_D)^T$, $\mathbf{e}_D = (e_1, \ldots, e_D)^T$, *and let the loading matrix or vector be* $(\alpha_1, \ldots, \alpha_D)^T$. *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 equation (9), $\widetilde{w}_D - w^0 = O_p(D^{-1/2})$ *as* $D \to \infty$.

Remark 2. The proof is given in the 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). 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.

Corollary 3. *(Asymptotic properties of factor scores in oblique Gaussian model). Suppose there are realized infinite sequences of observed variables* $\mathbf{z}_1^T = (z_{1,1}, z_{2,1}, \ldots), \ldots, \mathbf{z}_G^T = (z_{1,G}, z_{2,G}, \ldots)$ *and realized sequences of disturbance terms* $\mathbf{e}_1^T = (e_{1,1}, e_{2,1}, \ldots), \ldots, \mathbf{e}_G^T = (e_{1,G}, e_{2,G}, \ldots)$, *with latent variables* $\mathbf{w}^0 = (w_1^0, \ldots, w_G^0)$ *from the oblique factor model with fixed G groups defined in* (6). *Truncate the sequences to the first* d_g *variables* z_{g,d_g} *for* $g \in \{1, \ldots, G\}$ *with no* d_g *dominating others. Let* $\mathbf{z}_D = (\mathbf{z}_{1,d_1}^T, \mathbf{z}_{2,d_2}^T, \ldots, \mathbf{z}_{G,d_G}^T)^T$, $\mathbf{e}_D = (\mathbf{e}_{1,d_1}^T, \mathbf{e}_{2,d_2}^T, \ldots, \mathbf{e}_{G,d_G}^T)$. *Assume* $d_g^{-1}\|\mathbf{a}_g\|_1 \not\to 0$ *as* $d_g \to \infty$ *for* $g \in \{1, 2, \ldots, G\}$. *Let* $\widetilde{\mathbf{w}}_D = (\widetilde{w}_1, \ldots, \widetilde{w}_G)$ *be the factor scores defined in* (12). *Then*

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

The consistency of proxies in the oblique factor model is a straightforward corollary of Theorem 2. Similar to the 1-factor model, the assumptions on the strength of dependence in the model suggests the with-in group dependence is not weak.

Theorem 4. *(Asymptotic properties of factor scores in bi-factor Gaussian model). Suppose there are realized infinite sequences of observed variables* $\mathbf{z}_1^T = (z_{1,1}, z_{2,1}, \ldots), \ldots, \mathbf{z}_G^T = (z_{1,G}, z_{2,G}, \ldots)$ *and realized sequences of disturbance terms* $\mathbf{e}_1^T = (e_{1,1}, e_{2,1}, \ldots), \ldots, \mathbf{e}_G^T = (e_{1,G}, e_{2,G}, \ldots)$, *with latent variables* $\mathbf{w}^0 = (w_0^0, w_1^0, \ldots, w_G^0)$ *from the bi-factor model with fixed G groups defined in* (5). *Truncate the sequences to the first* d_g *variables* z_{g,d_g} *for* $g \in \{1, \ldots, G\}$, *with no* d_g *dominating others. Let* $\mathbf{z}_D = (\mathbf{z}_{1,d_1}^T, \mathbf{z}_{2,d_2}^T, \ldots, \mathbf{z}_{G,d_G}^T)^T$, $\mathbf{e}_D = (\mathbf{e}_{1,d_1}^T, \mathbf{e}_{2,d_2}^T, \ldots, \mathbf{e}_{G,d_G}^T)$ *and assume that the loading matrices* $\mathbf{A}_D = [\mathbf{a}_0, \mathrm{diag}(\mathbf{a}_1, \ldots, \mathbf{a}_G)]$ *are of full rank, with bounded condition number over* $d_g \to \infty$ *for all g.* $d_g^{-1}\|\mathbf{a}_j\|_1 \not\to 0$ *for* $g \in \{0, 1, \ldots, G\}$. *Let* $\widetilde{\mathbf{w}}_D = (\widetilde{w}_0, \ldots, \widetilde{w}_G)$ *be the factor scores defined in* (10) *and* (11). *Then*

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

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, satisfy some standard regularity conditions, such as in [4]. For example, continuity of derivatives up to third-order of the log-densities of the bivariate linking copulas with respect to v's.

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.

Some results in [14] assume the observed variables are stochastically increasing in the latent variables, and this implies observed variables are monotonically related. The above consists of a mild condition, because one would not think of using factor models with variables are not monotonically related.

10

Theorem 5. *(Consistency of proxy in 1-factor copula model with known linking copulas) Suppose there is a realized infinite sequence $u_1, u_2, \ldots$, with latent variable v^0 (independent of dimension D) from the 1-factor model in (1). For the truncation to the first D variables, let $\boldsymbol{u_D} = (u_1, \ldots, u_D)^T$. Define the averaged negative log-likelihood in parameter v as*

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

Assume $\lim_{D \to \infty} \bar{L}_D$ asymptotically has a global minimum and is strictly locally convex at the minimum. Also, assume the likelihood function satisfies the usual regularity conditions. Consider the proxy defined in (13). As $D \to \infty$, then $\widetilde{v}_D - v^0 = O_p(D^{-1/2})$ as $D \to \infty$.

Remark 3. The theorem shows that under certain regularity conditions, the latent variables can be approximately recovered from the observed variables, 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, it's not hard to show 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) $(U_j, 2V)$ follows the Gaussian copula with parameter $\rho_j > 0$ if $1/2 \leq V < 1$ and (ii) $(U_j, 2V)$ follows the Gaussian copula with parameter $-\rho_j < 0$ if $0 < V < 1/2$, and the $\{\rho_j\}$ is uniformly distributed in a bounded interval such as $[0.2, 0.8]$.

Oblique factor copula model with known linking copulas: In the oblique factor copula model, 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). Suppose there are realized infinite sequences of observed variable values $\boldsymbol{u}_1^T, \boldsymbol{u}_2^T, \ldots, \boldsymbol{u}_G^T$ with latent variable values $\boldsymbol{v}^0 = (v_1^0, v_2^0, \ldots, v_G^0)^T$ from the oblique factor model with G groups defined in equation (3). 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 $\boldsymbol{u_D} = (\boldsymbol{u}_{1,d_1}^T, \boldsymbol{u}_{2,d_2}^T, \ldots, \boldsymbol{u}_{G,d_G}^T)^T$. With $D = \sum_{g=1}^{G} d_g$, define the averaged negative log-likelihood in group g as $\bar{L}_D^{(g)}(v_g) = -D^{-1} \sum_{j=1}^{d_g} \log c_{U_{jg}, V_g}(u_{jg}, v_g)$. Assume $\lim_{d_g \to \infty} \forall g \, \bar{L}_D^{(g)}$ asymptotically has a global minimum and is strictly locally convex at the minimum. Also assume these likelihood functions satisfy the usual regularity conditions. Then, consider proxy variable defined in (16), as $d_g \to \infty$, $\widetilde{v}_{g\boldsymbol{D}} - v_g^0 = O_p(D^{-1/2})$ for $g \in \{1, 2, \ldots, G\}$.*

Theorem 7. *(Consistency of proxies in bi-factor copula model with known linking copulas). Suppose there are realized infinite sequences of observed variable values $\boldsymbol{u}_1^T, \boldsymbol{u}_2^T, \ldots, \boldsymbol{u}_G^T$ with latent variables values $\boldsymbol{v}^0 = (v_0^0, v_1^0, v_2^0, \ldots, v_G^0)^T$ from the bi-factor model with G groups defined in (2). Truncate the sequences to the first d_g variables $\boldsymbol{u}_{g,d_g}$ for $g \in \{1, \ldots, G\}$. Let $\boldsymbol{u_D} = (\boldsymbol{u}_{1,d_1}^T, \ldots, \boldsymbol{u}_{G,d_G}^T)^T$. Let $L_0(v_0; \boldsymbol{u_D}) = \log c_{U_D, V_0}(\boldsymbol{u_D}, v_0)$ be a log-likelihood function in v_0 with observed variables $\boldsymbol{u_D}$. For $g \in \{1, \ldots, G\}$, let $L_g(v_g; v_0, \boldsymbol{u}_{g,d_g}) = \log c_{U_g, V_0, V_g}(\boldsymbol{u}_{g,d_g}, v_0, v_g)$ be a 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 $(\boldsymbol{U_D}, V_0)$ as $\bar{L}_{0\boldsymbol{D}}(v_0; \boldsymbol{u_D}) = -D^{-1} L_0(v_0; \boldsymbol{u_D})$ and the averaged negative log-likelihood for marginalized density of $(\boldsymbol{U}_{g,d_g}, V_0, V_g)$ as $\bar{L}_{g\boldsymbol{D}}(v_g; v_0, \boldsymbol{u}_{g,d_g}) = -D^{-1} L_g(v_g; v_0, \boldsymbol{u}_{g,d_g})$. Assume all the likelihood functions satisfy the usual regularity conditions. Assume $\lim_{d_g \to \infty} \bar{L}_{g\boldsymbol{D}}$ asymptotically has a global minimum and is strictly locally convex at the minimum. The same assumptions are applied to $\lim_{D \to \infty} \bar{L}_{0\boldsymbol{D}}$. Then, $\widetilde{v}_{0\boldsymbol{D}}$ defined in (14) and $\widetilde{v}_{g\boldsymbol{D}}(\widetilde{v}_{0\boldsymbol{D}})$ for $g = 1, \ldots, G$ 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.*

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 observed variables 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 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* $\mathbf{A}^T = (\mathbf{a}_1, \ldots, \mathbf{a}_D, \ldots)$, $\widehat{\mathbf{A}}^T = (\hat{\mathbf{a}}_1, \ldots, \hat{\mathbf{a}}_D, \ldots)$ *(infinite sequence) be the perturbation of* $\mathbf{A}^T$, *where* $\hat{\mathbf{a}}_j$ *and* $\mathbf{a}_j$ *denotes the jth column of matrix* $\widehat{\mathbf{A}}^T$ *and* $\mathbf{A}^T$ *respectively,* $j \in \{1, 2, \ldots\}$. *Suppose there is one realization* $\mathbf{z} = (z_1, \ldots, z_D, \ldots)$ *from the factor model with loading matrix A. Let* $\mathbf{A}_D$ *be A truncated to D rows, and similarly define* $\widehat{\mathbf{A}}_D$. *Suppose the entries of* $\mathbf{\Psi}_D^2, \widehat{\mathbf{\Psi}}_D^2$ *are bounded away from 0. Consider the factor scores vector in* (9) *as a function of the loading matrix:* $\widetilde{\mathbf{w}}_D = \widetilde{\mathbf{w}}_D(\mathbf{A}_D)$. *Let* $\mathbf{Q}_D = \mathbf{A}_D^T \mathbf{\Psi}_D^{-2} \mathbf{A}_D$, $\widehat{\mathbf{Q}}_D = \widehat{\mathbf{A}}_D^T \widehat{\mathbf{\Psi}}_D^{-2} \widehat{\mathbf{A}}_D$. *Suppose* $D^{-1}\mathbf{Q}_D \to \mathbf{Q}$, $D^{-1}\widehat{\mathbf{Q}}_D \to \widehat{\mathbf{Q}}$ *where* $\mathbf{Q}, \widehat{\mathbf{Q}}$ *are both positive definite and well-conditioned matrices. Then* $\|\widetilde{\mathbf{w}}_D(\widehat{\mathbf{A}}_D) - \widetilde{\mathbf{w}}_D(\mathbf{A}_D)\| \le K_D \cdot \sqrt{D^{-1} \sum_{j=1}^D \|\hat{\mathbf{a}}_j - \mathbf{a}_j\|^2}$, *where the constants* K_D *are bounded as* $D \to \infty$.

5.2. Parametric copula factor models

For parametric models we assume that there is parameter 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 $c_{U_s,V}(u, v; \boldsymbol{\theta}_s)$.

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* $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_D, \ldots)$. *Let* $\hat{\boldsymbol{\theta}}$ *be a perturbation of* $\boldsymbol{\theta}$. *Assume the parameters of both* $\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}$ *all lie in a bounded space such that all the linking copulas are bounded away from comonotonicity and countermonotonicity. Let* $\mathbf{u} = (u_1, \ldots, u_D, \ldots)$ *be one realization generated from model with* $\boldsymbol{\theta}$. *Let* $\boldsymbol{\theta}_D$ *(or* $\hat{\boldsymbol{\theta}}_D$), $\mathbf{u}_D$ *be truncated to the first D linking copulas and variables. Consider the proxy in* (13) *as a function of* $\boldsymbol{\theta}_D$: $\widetilde{v}_D = \widetilde{v}_D(\boldsymbol{\theta}_D)$. *Suppose* $\hat{\boldsymbol{\theta}}_D \in \bar{\mathbf{B}}(\boldsymbol{\theta}_D, \rho)$ *(ball of sufficiently small radius* $\rho > 0$). *Assume for each j, the following partial derivatives of the log copula densities exist:* $\partial^k \log c_{jV}(u_j, v; \boldsymbol{\theta}_j)/\partial v^k$, $k \in \{1, 2, 3\}$, *and* $\partial^{k+1} \log c_{jV}(u_j, v; \boldsymbol{\theta}_j)/\partial \boldsymbol{\theta}_j \partial v^k$ *for* $k \in \{0, 1, 2\}$. *Also, assume the derivatives* $\partial^{k+1} \log c_{jV}(u_j, v; \boldsymbol{\theta}_j)/\partial \boldsymbol{\theta}_j \partial v^k$ *are uniformly bounded for* $k \in \{0, 1, 2\}$ *in* $\bar{\mathbf{B}}(\boldsymbol{\theta}_D, \rho)$. *Then there exist a constant* B_D *which is bounded as* $D \to \infty$ *such that*

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

where $\|\hat{\boldsymbol{\theta}}_D - \boldsymbol{\theta}_D\|^* := \sqrt{D^{-1} \sum_{j=1}^D \|\hat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j\|_2^2}$.

Lemma 10. *Consider a G-group bi-factor copula model* (2) *with parametric linking copulas that have monotone dependence. 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* $\boldsymbol{\theta}_g^{(1)} = (\boldsymbol{\theta}_{1g,0}, \boldsymbol{\theta}_{2g,0}, \ldots)$. *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,* $j \in \{1, 2, \ldots\}$. *Let* $\boldsymbol{\theta}_g^{(2)} = (\boldsymbol{\theta}_{1g}, \boldsymbol{\theta}_{2g}, \ldots)$ *be the vector of parameters for* $g \in \{1, \ldots, G\}$. *Let* $\boldsymbol{\theta}^{(1)} = (\boldsymbol{\theta}_1^{(1)}, \ldots, \boldsymbol{\theta}_G^{(1)})$ *and* $\boldsymbol{\theta}^{(2)} = (\boldsymbol{\theta}_1^{(2)}, \ldots, \boldsymbol{\theta}_G^{(2)})$ *with perturbations* $\hat{\boldsymbol{\theta}}^{(1)}$ *and* $\hat{\boldsymbol{\theta}}^{(2)}$. *Assume the parameters in these four vectors are all in a bounded space such that all the linking copulas are bounded away from comonotonicity and countermonotonicty. Let* $\mathbf{D} = (d_1, \ldots, d_G)$ *and* $\mathbf{u}_D^T = (\mathbf{u}_1^T, \mathbf{u}_2^T, \ldots, \mathbf{u}_G^T)$ *be a truncation of infinite-dimensional realization from the model with parameters* $\boldsymbol{\theta}_D = (\boldsymbol{\theta}_D^{(1)}, \boldsymbol{\theta}_D^{(2)})$, *where* $\boldsymbol{\theta}_D^{(1)}, \boldsymbol{\theta}_D^{(2)}$ *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* $g \in \{1, \ldots, G\}$. *Similarly, define* $\hat{\boldsymbol{\theta}}_D, \hat{\boldsymbol{\theta}}_{g,d_g}$. *Let* $\widetilde{v}_{0D} = \widetilde{v}_{0D}(\boldsymbol{\theta}_D)$, $\widetilde{v}_{gD} = \widetilde{v}_{gD}(\boldsymbol{\theta}_D)$ *be the proxies defined in* (14) *and* (15). *Assume* $\hat{\boldsymbol{\theta}}_D \in \bar{\mathbf{B}}(\boldsymbol{\theta}_D, \rho)$ *(ball of sufficiently small radius* $\rho > 0$). *Assume for* $j \in \{1, \ldots, d_g\}$, $g \in \{1, \ldots, G\}$, *that the following partial derivatives of log copula densities exist:* $\partial^k \log c_{U_D,V_0}(\mathbf{u}_D, v_0, \boldsymbol{\theta}_D)/\partial v_0^k$, $\partial^k \log c_{U_g,V_g;V_0}(\mathbf{u}_g, v_g; v_0, \boldsymbol{\theta}_{g,d_g})/\partial v_g^k$, *for* $k \in \{1, 2, 3\}$ *and*

12

$\partial^{k+1} \log c_{U_D,V_0}(\boldsymbol{u}_D, v_0, \boldsymbol{\theta}_D)/\partial\theta_{jg,0}\partial v_0^k$, $\partial^{k+1} \log c_{U_D,V_0}(\boldsymbol{u}_D, v_0, \boldsymbol{\theta}_D)/\partial\theta_{jg}\partial v_0^k$, $\partial^{k+1} \log c_{U_g,V_g;V_0}(\boldsymbol{u}_g, v_g; v_0, \boldsymbol{\theta}_{g,d_g})/\partial\theta_{jg}\partial v_g^k$,
$\partial^{k+1} \log c_{U_g,V_g;V_0}(\boldsymbol{u}_g, v_g; v_0, \boldsymbol{\theta}_{g,d_g})/\partial\theta_{jg,0}\partial v_g^k$, for $k \in \{0, 1, 2\}$. Also, assume the partial derivatives $\partial^{k+1} \log c_{U_D,V_0}(\boldsymbol{u}_D, v_0, \boldsymbol{\theta}_D)/\partial\theta_{jg,0}\partial v_0^k$,
$\partial^{k+1} \log c_{U_D,V_0}(\boldsymbol{u}_D, v_0, \boldsymbol{\theta}_D)/\partial\theta_{jg}\partial v_0^k$, $\partial^{k+1} \log c_{U_g,V_g;V_0}(\boldsymbol{u}_g, v_g; v_0, \boldsymbol{\theta}_{g,d_g})/\partial\theta_{jg}\partial v_g^k$, $\partial^{k+1} \log c_{U_g,V_g;V_0}(\boldsymbol{u}_g, v_g; v_0, \boldsymbol{\theta}_{g,d_g})/\partial\theta_{jg,0}\partial v_g^k$
are uniformly bounded for $k \in \{0, 1, 2\}$ in $\bar{\boldsymbol{B}}(\boldsymbol{\theta}_D, \rho)$. Then there exists constants B_D and B_D^* that are bounded as $d_g \to \infty$
for all g, such that:

$$\|\widetilde{v}_{0D}(\hat{\boldsymbol{\theta}}_D) - \widetilde{v}_{0D}(\boldsymbol{\theta}_D)\| \le B_D\|\hat{\boldsymbol{\theta}}_D - \boldsymbol{\theta}_D\|^*$$
$$\|\widetilde{v}_{gD}(\hat{\boldsymbol{\theta}}_D) - \widetilde{v}_{gD}(\boldsymbol{\theta}_D)\| \le B_D^*\|\hat{\boldsymbol{\theta}}_{g,d_g} - \boldsymbol{\theta}_{g,d_g}\|^*, \quad g \in \{1, \dots, G\},$$

where $\|\hat{\boldsymbol{\theta}}_D - \boldsymbol{\theta}_D\|^* := \sqrt{(2D)^{-1} \sum_{g=1}^G \sum_{j=1}^{d_g} \|\hat{\boldsymbol{\theta}}_{jg} - \boldsymbol{\theta}_{jg}\|_2^2 + \|\hat{\boldsymbol{\theta}}_{jg,0} - \boldsymbol{\theta}_{jg,0}\|_2^2}$, and
$\|\hat{\boldsymbol{\theta}}_{g,d_g} - \boldsymbol{\theta}_{g,d_g}\|^* := \sqrt{(2d_g)^{-1} \sum_{j=1}^{d_g} \|\hat{\boldsymbol{\theta}}_{jg} - \boldsymbol{\theta}_{jg}\|_2^2 + \|\hat{\boldsymbol{\theta}}_{jg,0} - \boldsymbol{\theta}_{jg,0}\|_2^2}$.

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 order to have consistent estimates of parameters, an assumption is needed on the behavior as the number of variables increase 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. The next lemma summarizes the procedures for block estimation in the Gaussian and copula factor models. We unify the notation of observed variables to be X; in Gaussian models, $X = Z$, and in copula models, $X = U$. Assume that the data from a random sample with D variables are $\{X_i : i = 1, \dots, N\}$.

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. Standard maximum likelihood theory for a finite number of parameters can be applied, and there is no need to develop theory for simultaneous estimates of all parameters with the number of parameters increasing to ∞.

Lemma 11. *(Block estimation procedure for 1-factor, bi-factor and oblique factor model). 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.*

1. *(1-factor model): For $j \in 1, 2, \dots, D$, split D variables into K blocks of approximate size $B > 5$ in a sequential way, the partition $\mathcal{B} = \{\mathcal{B}_1, \mathcal{B}_2, \dots, \mathcal{B}_K\}$ where the cardinality of $\mathcal{B}_k$ is $B_k \approx B$ for $k \in \{1, \dots, K\}$. This leads to K 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 $\mathcal{B}_2, \dots, \mathcal{B}_K$. The estimates of the parameter associated with X_1 can be averaged over the blocks. The estimated parameters are $(\hat{\boldsymbol{\theta}}_{\mathcal{B}_1}, \hat{\boldsymbol{\theta}}_{\mathcal{B}_2}, \cdots, \hat{\boldsymbol{\theta}}_{\mathcal{B}_K})$ 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 blocks of approximate size $d_g^{(k)}$ dependent variables for $k \in \{1, \dots, K\}$. 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 $g \ne h$, $g, h \in \{1, 2, \dots, G\}$ in block k. The partition gives K blocks, for block k, $\{X_{ij \in \mathcal{B}_k}, : j \in \{\mathcal{B}_k^{(1)}, \dots, \mathcal{B}_k^{(g)}, \dots, \mathcal{B}_k^{(G)}\}; \quad i \in \{1, 2, \dots, N\}\}$ 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 $\mathcal{B}_2, \dots, \mathcal{B}_K$. Suppose the estimates of parameters involving the variables in block k of each group are $\hat{\boldsymbol{\theta}}_{\mathcal{B}_k}^{(1)}, \hat{\boldsymbol{\theta}}_{\mathcal{B}_k}^{(2)}, \dots, \hat{\boldsymbol{\theta}}_{\mathcal{B}_k}^{(G)}$, for $k \in \{1, \dots, K\}$. For parameters that are estimated over different blocks, such as $\boldsymbol{\Sigma}_W$, an average could be taken over the different blocks.*

13

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 $\mathcal{B}_2, \ldots, \mathcal{B}_K$.*

Lemma 12. *(Asymptotic properties of estimated parameters in 1-factor model, bi-factor model, oblique factor model). 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 θ_D be the parameter vector in the factor models and $\hat{\theta}_D$ 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*

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

Proof. For all factor models in Section 2, the block method in Lemma 11 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 the previous results in this section, 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 13. *(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 the Assumption 1 and the regularity conditions in Theorem 5, Theorem 7, and Corollary 6 respectively. Let the parameters θ be $(\theta_1, \theta_2, \ldots, \theta_D, \ldots)$, suppose $\hat{\theta}_j$ is a 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 θ_D be θ truncated to the first D random variables u_j's, then as $D \to \infty$, the following hold.*

(1) 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, $\|\widetilde{V}_{iD}(\hat{\theta}_D) - V_i\| = o_p(1)$.

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

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

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

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

For the first term in the right-hand side, Lemma 9 implies that it has the same order as $O_p(\sqrt{D^{-1} \sum_{j=1}^D \|\hat{\theta}_j - \theta_j\|^2}) = O_p(N^{-1/2})$. 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 we omit details here. □

Remark 4. 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 studied in this paper. But consistency and convergence rates are not affected because univariate distributions can be estimated well with a large sample size.

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 ([12]; Chapter 1 of [9]).

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 [14] 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 observe 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.

Suppose there is sample of size N from the model, in the 1-factor model, we denote the i-th sample as $\boldsymbol{u}_i = (u_{i1}, \ldots, u_{iD})$ and in the bi-factor or oblique factor, we denote the samples as $\boldsymbol{u}_i = (\boldsymbol{u}_{i,1}^T, \ldots, \boldsymbol{u}_{i,G}^T)$ (the dependence on d_g in $\boldsymbol{u}_{ig,d_g}$ is suppressed for simplicity.)

1-factor copula model. If the latent variable is assumed observed, then the complete log-likelihood is

$$\sum_{i=1}^{N} \log c_{U_{1:D},V}(u_{i1}, \ldots, u_{iD}, v_i; \boldsymbol{\theta}_D) = \sum_{i=1}^{N} \sum_{j=1}^{D} \log c_{jV}(u_{ij}, v_i; \boldsymbol{\theta}_j). \tag{20}$$

- 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. For each sample i, $\bar{u}_i = D^{-1} \sum_{j=1}^{D} u_{ij}$ 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 (20), and obtain the first-stage estimates of the parameters in $\boldsymbol{\theta}$ from the approximate log-likelihood. This is the method of [14].

- Stage 2: Construct the conditional expectations proxies based on (13) with the first-stage estimated parameters of $\boldsymbol{\theta}$. One-dimensional Gauss-Legendre quadrature can be used. Denote the proxies as $\widetilde{U}_0$. Substitute $v_i = \widetilde{u}_{i,0}$ in the log-likelihood (20), obtain the second-stage estimates of the parameters from the approximate log-likelihood.

Bi-factor copula model. If the latent variables are assumed observed, then the complete log-likelihood can be expressed as

$$\sum_{i=1}^{N} \log c_{U_{1:D},V_0,V_g}(\boldsymbol{u}_{i,1}^T, \ldots, \boldsymbol{u}_{i,G}^T, v_{i,0}, v_{i,g}; \boldsymbol{\theta}) = \sum_{i=1}^{N} \sum_{g=1}^{G} \sum_{i=1}^{d_g} \log c_{U_{jg},V_0}(u_{i,jg}, v_{i,0}; \boldsymbol{\theta}_{jg,0})$$

$$+ \log c_{U_{jg},V_g;V_0}(C_{U_{jg}|V_0}(u_{i,jg}|v_{i,0}), v_{i,g}; \boldsymbol{\theta}_{jg}). \tag{21}$$

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 $\widetilde{w}_{i0}, \widetilde{w}_{i1}, \ldots, \widetilde{w}_{iG}$. 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 $v_{i,0} = \widetilde{v}_{i,0}^{(1)}$, $v_{i,g} = \widetilde{v}_{i,g}^{(1)}$ in log-likelihood (21), obtain the first-stage estimates of the parameters from the approximate log-likelihood.

- Stage 2: Construct the conditional expectation proxies based on equation (14) and (15) with first-stage estimates plugged in. Nested 1-dimensional Gauss-Legendre quadrature can be used. Denote the conditional expectation proxies as $\widetilde{V}_0^{(2)}$, $\widetilde{V}_g^{(2)}$, $g \in \{1, \ldots, G\}$. Letting $v_{i,0} = \widetilde{v}_{i,0}^{(2)}$, $v_{i,g} = \widetilde{v}_{i,g}^{(2)}$ in log-likelihood (21), obtain the second-stage estimates of the parameters from the approximate log-likelihood.

Oblique factor copula model. Similar to the bi-factor copula model, assume the group structure of the model is known. If the latent variables are assumed observed, the complete log-likelihood is

$$\sum_{i=1}^{N} \log c_{\text{oblique},U_{1:D},V_{1:G}}(\boldsymbol{u}_{i,1}^T, \ldots, \boldsymbol{u}_{i,G}^T, v_{i,1}, \ldots, v_{i,G}; \boldsymbol{\theta}) = \sum_{i=1}^{N} \sum_{g=1}^{G} \sum_{j=1}^{d_g} \log c_{U_{jg},V_g}(u_{i,jg}, v_{i,g}; \boldsymbol{\theta}_{jg})$$
$$+ \log c_V(v_{i,1}, \ldots, v_{i,G}; \boldsymbol{\theta}_V), \tag{22}$$

where $\boldsymbol{\theta} = (\boldsymbol{\theta}_{d_1,1}^T, \ldots, \boldsymbol{\theta}_{d_G,G}^T, \boldsymbol{\theta}_V^T)$, and c_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 each sample, $\bar{u}_{i,g} = d_g^{-1} \sum_{j=1}^{d_g} u_{i,jg}$ and $u_{i,g} = [\text{rank}(\bar{u}_{i,g}) - 0.5)]/N$, where $\text{rank}(\bar{u}_{i,g})$ 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 (22) to get the first-stage estimates of the parameters from the approximate log-likelihood. This is the method of [14].

- Stage 2: Construct the conditional expectations proxies based on (16) with the first-stage estimated parameters. This requires 1-dimensional numerical integration. Denote the proxies as $\widetilde{U}_g$. Substitute $v_{i,g} = \widetilde{u}_{i,g}$ in the log-likelihood (22), 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 [14], but the above methods based on conditional expectations perform better from the simulation results shown in the next section. The estimation of proxies and copula parameters could be iterated further if desired stage 2 estimates differ a lot from stage 1 estimates.

For optimizing the above approximate log-likelihoods using proxies for the latent variables, we adopt a modified Newton-Raphson algorithm with analytic derivatives; see [13] for details of the numerical implementation.

7. Simulation experiments

This section has some simulation results to support and explain ideas in previous sections. In all the following settings, the parameters on the linking copulas are designed to be generated uniformly from a bounded subset of the parameter space; this is an example of sampling from some super-population. Many different scenarios were assessed and some representative summaries are given in three subsections for the 1-factor, bi-factor and oblique factor copula models.

The sequential approach of Section 6 is compared with the "exact" method from the implementation of [12][13] with R front-end and FORTRAN 90 back-end for minimizing the negative log-likelihood with a modified Newton-Raphson algorithm; the "exact" method is indicated with the superscript $m = 0$. The proxy approach in [14] for 1-factor and oblique factor, as summarized in Section 6 is indicated with the superscript $m = 1$. The stage 2 estimates for all three copula models is indicated with the superscript $m = 2$. When the linking copula families are assumed known, differences of copula parameter estimates are summarized in the Kendall's tau scale. If linking copula families are decided based on a few parametric choices that cover a range of tail asymmetry and strength of dependence in joint tails, additional summaries are based on the tail-weighted dependence measures defined in [15]. This is because different bivariate copula families can have members that are similar in tail properties.

7.1. One-factor copula model

Two settings are summarized in Table 3 to illustrate the performance of the sequential approach. The sample size is N and there are D variables with bivariate linking copulas $C_{jV}(\cdot, \cdot; \theta_j)$, $j \in \{1, \cdots, D\}$, to the latent variable. The parameters $\theta = (\theta_1, \ldots, \theta_D)$ are independent $U(\theta_L, \theta_U)$, where, θ_L and θ_U are chosen so that the Kendall's taus of the bivariate copulas range between 0.4 and 0.8 for moderate to strong dependence. The simulation size is 1000.

	N	D	linking families	$[\theta_L, \theta_U]$
setting1	500	(20,40,60,80)	Frank	[4.2.18.5]
setting2	500	(30,45,60,90)	Gumbel, t, Frank	$\{(1.67, 5), (4.2, 18.5), (0.59, 0.95)\}$

Table 3: Two simulation settings for the 1-factor copula model; in setting2, the number of linking copulas in different families are approximately $D/3$, and the ν parameter of Student-t copulas is fixed at 5. The parameters are chosen to let the Kendall's tau be in [0.4,0.8].

In setting1, the main summary is the mean absolute error (MAE) of estimated parameters for three different methods ($m = 0, 1, 2$ as indicated above): $\hat{\theta}^m_{\text{MAE}} = (ND)^{-1} \sum_{i=1}^N \sum_{j=1}^D |\hat{\theta}^m_{ij} - \theta_{ij}|$, where θ_{ij} is the parameter of $C_{j,0}$ generated at the ith simulation and $\hat{\theta}^m_{ij}$ is the corresponding estimate using the different approaches. An additional summary is the MAE of the differences between the estimates obtained from the proxy methods and the exact approach, as well as the differences of corresponding Kendall's taus (function of the estimated bivariate linking copula):

$$\hat{\theta}^m_{|\text{diff}|} = \frac{1}{ND} \sum_{i=1}^N \sum_{j=1}^D |\hat{\theta}^m_{ij} - \hat{\theta}^0_{ij}|, \quad \hat{\tau}^m_{|\text{diff}|} = \frac{1}{ND} \sum_{i=1}^N \sum_{j=1}^D |\hat{\tau}^m_{ij} - \hat{\tau}^0_{ij}|, \quad m = 1, 2. \tag{23}$$

In setting2, summaries include the averaged differences of the dependence measures between the true and fitted models over the D bivariate linking copulas:

$$[\hat{M}^{\text{diff}}]^m_{\text{mean}} = \frac{1}{D} \sum_{j=1}^D |[\widehat{M}^{\text{model}}]_j - [M^{\text{true}}]_j|, \quad m = 1, 2, \tag{24}$$

where the measure M can be Kendall's tau, and tail-weighted upper/lower tail dependence as defined in [15]. Denote them as $M = \tau, \zeta_{\alpha,U}(20), \zeta_{\alpha,L}(20)$ respectively.

To compare the proxies as estimated latent variables, a summary is the RMSE of the proxies of the two methods:

$$\hat{v}^m_{\text{RMSE}} = \left\{ \frac{1}{KN} \sum_{k=1}^K \sum_{i=1}^N (\hat{v}^m_{ki} - v_{ki})^2 \right\}^{1/2}, \quad m = 1, 2, \tag{25}$$

where v_{ki} is the latent variable for the ith observation vector in the kth simulation.

D	$\hat{\theta}^0_{\text{MAE}}$	$\hat{\theta}^1_{\text{MAE}}$	$\hat{\theta}^2_{\text{MAE}}$	$\hat{\theta}^1_{|\text{diff}|}$	$\hat{\theta}^2_{|\text{diff}|}$	$|\hat{\tau}^{\text{diff}}|^1_{\text{mean}}$	$|\hat{\tau}^{\text{diff}}|^2_{\text{mean}}$	$\hat{v}^{m=1}_{\text{RMSE}}$	$\hat{v}^{m=2}_{\text{RMSE}}$
20	0.493	0.680	0.750	0.545	0.578	0.010	0.008	0.041	0.032
40	0.474	0.582	0.552	0.374	0.283	0.006	0.004	0.032	0.025
60	0.472	0.565	0.505	0.320	0.184	0.005	0.002	0.028	0.023
80	0.471	0.544	0.485	0.270	0.138	0.004	0.002	0.026	0.016

Table 4: 1-factor copula models with all Frank linking copulas; simulation size 1000, sample size $N = 500$, θ uniform in (θ_L, θ_U) as specified in Table 3. Summaries from (23), (24) and (25) for 3 approaches — superscript $m = 0$: exact; superscript $m = 1$: unweighted average proxy; superscript $m = 2$: sequential.

D	$[\hat{\tau}^{\text{diff}}]^m_{\text{mean}}$	$[\hat{\zeta}^{\text{diff}}_{\alpha,U}]^m_{\text{mean}}$	$[\hat{\zeta}^{\text{diff}}_{\alpha,L}]^m_{\text{mean}}$	$\hat{\nu}^{m=1}_{\text{RMSE}}$	$\hat{\nu}^{m=2}_{\text{RMSE}}$
30	0.017/0.015	0.028/0.025	0.026/0.019	0.041	0.027
45	0.016/0.014	0.024/0.021	0.021/0.016	0.035	0.022
60	0.015/0.013	0.022/0.019	0.019/0.014	0.032	0.019
90	0.014/0.012	0.020/0.017	0.017/0.014	0.028	0.017

Table 5: 1-factor copula model with linking copulas from Gumbel, t and Frank families; Simulation size 1000, sample size $N = 500$. In each simulation θ is uniform in (θ_L, θ_U) as specified in Table 3. The summaries are for (24) and (25) and are shown in order $m = 1/m = 2$ respectively in columns 2 to 4.

From Table 4, the two proxy approaches can give accurate parameter estimates comparable to those obtained from the exact likelihood when $D \geq 40$. The sequential approach performs better than the "unweighted average" approach. The differences in dependence measures between the two proxy approaches and the exact approach decrease with the increasing dimension. The sequential approach gives the parameter estimates closer to the exact approach than that in [14]. In addition, the conditional expectation proxies are closer to the true realized latent variables. Similar observations can be seen in Table 5. In setting2, the proxy approach can identify the correct copula families in most cases, though sometimes the method selects BB1 copulas with similar tail behavior to the true linking Gumbel copulas.

7.2. Bi-factor copula model

Two settings are summarized in Table 6 to illustrate the sequential approach. The sample size is N and there are D variables and $2D$ linking copulas. The number of groups $G = 3$ and the size of each group is approximately $D/3$. The parameters of the D copulas linking the observed variables and the global latent variable are generated uniformly in (θ_L, θ_U) so that there is a wide range for the dependence between the observed variables and the global latent variable. For the D bivariate copulas for conditional dependence, the parameters are generated uniformly from $(\theta_{LL}, \theta_{UU})$ so that the within-group dependence is strong. Also, the parameter setting ensures the condition number (in Remark 1) of the Q_D, obtained from fitted bi-factor Gaussian factor structure on the data transformed to $N(0, 1)$ scales, is small enough for a reasonable convergence rate. In both setting, the simulation size is 1000.

	N	D	linking families	$[\theta_L, \theta_U]$ (τ)	$[\theta_{LL}, \theta_{UU}]$ (τ)
setting1	1200	(30,60,90,120)	Frank/Frank	[1.87,8] (0.2,0.6)	[4.2,11.5] (0.4,0.7)
setting2	2000	(30,60,90)	BB1/Frank	[0.3,1] $\times$ [1.1,2.5] (0.2,0.7)	[8.5,18.5] (0.6,0.8)

Table 6: Two simulation settings for the bi-factor copula model; In setting2, the D global linking copulas are in the BB1 family and the D local linking copulas are in the Frank family. For the comparison in setting2, $D = 120$ would take too much computational time for the exact approach. The range of Kendall's tau corresponds to the range of parameters are included after the parameter interval. The condition number of Q matrix is around 50 in setting1 and 70 in setting2.

As in the previous subsection, summaries are MAE of estimated parameters in the global and local linking copulas. Also summarized are differences in estimated parameters and corresponding Kendall's tau between the proxy method and the exact method. For setting2, the sequential approach is applied in two cases: (a) assuming the linking copula families are known; (b) assuming the linking copula families are to be decided. In setting1, the linking copula families are assumed known. In setting2 with case (b), summaries are as in (24). The simulation results are summarized in Table 7 and Table 8.

D	Global linking copulas			Local linking copulas			RMSE$_{\text{proxy}}$
Frank	$\hat{\theta}^m_{\text{glob:MAE}}$	$\hat{\theta}^{m=2}_{\text{glob:|diff|}}$	$\hat{\tau}^{m=2}_{\text{glob:|diff|}}$	$\hat{\theta}^m_{\text{loc:MAE}}$	$\hat{\theta}^{m=2}_{\text{loc:|diff|}}$	$\hat{\tau}^{m=2}_{\text{loc:|diff|}}$	$\hat{v}_0/\hat{v}_g$
30	0.160/0.352	0.306	0.020	0.260/1.000	0.959	0.029	0.073/0.100
60	0.145/0.256	0.201	0.012	0.239/0.461	0.377	0.012	0.054/0.073
90	0.139/0.234	0.180	0.010	0.230/0.320	0.201	0.007	0.045/0.061
120	0.135/0.233	0.183	0.010	0.227/0.283	0.159	0.006	0.040/0.053

Table 7: Bi-factor copula model with all linking copulas in the Frank family; Simulation size 1000, sample size $N = 1200$, number of variables in each group equal and set to be $d_g = 10, 20, 30, 40$, $g = 1, 2, 3$. The parameters are generated as specified in Table 6. For D global/local linking copulas, summaries of $\hat{\theta}_{\text{MAE}}, \hat{\theta}_{\text{diff}}, \hat{\tau}_{\text{diff}}$ for approaches superscript $m = 0$: exact; superscript $m = 2$: sequential are shown; For MAE, the results are shown for m=0/m=2 respectively.

From Table 7, the proxy method can give parameter estimates close to that of exact approach when $d_g \geq 20$. The differences in the estimates between the proxy and exact methods decrease as the dimension becomes large. Also, the sequential conditional expectation proxies are getting closer to the realizations of latent variables as d_g increases. The global latent variables can be estimated more accurately than the local ones since all the observed variables are used for estimation. A similar observation can be found in Table 8 in the case where copula families are not specified. From the results in the second part of Table 8, the dependence measures of the estimated linking copulas are close to those of the actual linking copulas. The method will also select t, Gumbel copula, or survival BB1 copula for the global linking copulas, and the selected copulas have similar tail behaviors to the actual ones. The RMSE of the proxies indicates that the latent variables can be estimated well even though some of the linking copula families are misspecified.

D	Global linking copulas			Local linking copulas			RMSE$_{\text{proxy}}$
case (a)	$\hat{\theta}^m_{\text{glob:MAE}}$	$\hat{\theta}^{m=2}_{\text{glob:|diff|}}$	$\hat{\tau}^{m=2}_{\text{glob:|diff|}}$	$\hat{\theta}^m_{\text{loc:MAE}}$	$\hat{\theta}^{m=2}_{\text{loc:|diff|}}$	$\hat{\tau}^{m=2}_{\text{loc:|diff|}}$	$\hat{v}_0/\hat{v}_g$
30	0.032/0.145	0.142	0.013	0.377/1.290	0.965	0.013	0.042/0.075
60	0.030/0.144	0.141	0.008	0.278/0.440	0.349	0.008	0.030/0.059
90	0.029/0.140	0.139	0.010	0.265/0.416	0.358	0.009	0.027/0.053
case (b)	$[\hat{\tau}^{\text{diff}}]^{m=2}_{\text{mean}}$	$[\hat{\zeta}^{\text{diff}}_{\alpha,U}]^{m=2}_{\text{mean}}$	$[\hat{\zeta}^{\text{diff}}_{\alpha,L}]^{m=2}_{\text{mean}}$	$[\hat{\tau}^{\text{diff}}]^m_{\text{mean}}$	$[\hat{\zeta}^{\text{diff}}_{\alpha,U}]^{m=2}_{\text{mean}}$	$[\hat{\zeta}^{\text{diff}}_{\alpha,L}]^{m=2}_{\text{mean}}$	$\hat{v}_0/\hat{v}_g$
30	0.013	0.024	0.037	0.015	0.018	0.018	0.042/0.075
60	0.012	0.023	0.028	0.007	0.008	0.008	0.031/0.058
90	0.013	0.025	0.022	0.008	0.009	0.009	0.027/0.053

Table 8: Bi-factor copula model with BB1 global linking copulas and Frank local linking copulas; Simulation size 1000, sample size $N = 2000$, number of variables in each group equal and set to be $d_g = 10, 20, 30$, $g = 1, 2, 3$. For D global/local linking copulas, $\hat{\theta}_{\text{MAE}}, \hat{\theta}_{\text{diff}}, \hat{\tau}_{\text{diff}}$ for approaches superscript $m = 0$: exact; superscript $m = 2$: sequential are shown. For MAE, the results are shown for m=0/m=2 respectively; In case (b), $[\hat{\tau}^{\text{diff}}]_{\text{mean}}, [\hat{\zeta}^{\text{diff}}_{\alpha,U}]_{\text{mean}}, [\hat{\zeta}^{\text{diff}}_{\alpha,L}]_{\text{mean}}$ are the averaged differences in the dependence measures between true and fitted models over D global/local linking copulas. The results are only shown for the sequential method.

7.3. Oblique factor model

A simulation setting consists of $K = 1000$ replications of sample size $N = 1000$ from a (nested) oblique copula model in (3) with $G = 3$ groups of equal group size d_g; $d_g = 10, 15, 20, 30$. The density c_V is assumed to have a one-factor structure with Frank linking copula, because this is the nested copula in [13] for which only 2-dimensional quadrature is needed instead of G-dimensional. The parameters in c_V are generated uniformly in $(\theta_L, \theta_U) = (3, 6)$ such that the Kendall's tau is between 0.3 to 0.5. The bivariate linking copulas in three groups are in the Gumbel, Frank and t families respectively. For each group, the parameters for the linking copulas are generated uniformly in $(\theta_{LL}, \theta_{UU})$, for Gumbel, BB1 and student-t copulas, with $(\theta_{LL}, \theta_{UU})$ in $(1.67, 5)$, $(0.25, 2) \times (1.5, 2.5)$, $(0.59, 0.95)$ respectively. The v parameter of Student-t copulas is fixed at 5. The Kendall's taus for linking copulas in each group are between 0.4 and 0.8. Proxy variables are used to decide on the families for the linking copulas in each group. The simulation results are summarized in the Table 9.

D	Global linking copulas		Local linking copulas			$\text{RMSE}_{\text{proxy}}$
	$\hat{\theta}_{\text{RMSE}}$	$[\hat{\tau}^{\text{diff}}]^m_{\text{mean}}$	$[\hat{\tau}^{\text{diff}}]^m_{\text{mean}}$	$[\hat{\zeta}^{\text{diff}}_{U,\alpha}]^m_{\text{mean}}$	$[\hat{\zeta}^{\text{diff}}_{L,\alpha}]^m_{\text{mean}}$	
30	0.514/0.489	0.028/0.025	0.022/0.029	0.029/0.023	0.024/0.019	0.067/0.056
45	0.488/0.475	0.026/0.024	0.016/0.021	0.025/0.018	0.022/0.016	0.056/0.045
60	0.471/0.460	0.025/0.023	0.014/0.017	0.022/0.016	0.020/0.015	0.049/0.039
90	0.479/0.478	0.025/0.024	0.012/0.014	0.019/0.015	0.018/0.013	0.040/0.031

Table 9: Oblique factor copula model with $G = 3$ groups of equal group size d_g; $d_g = 10, 15, 20, 30$; Sample size $N = 1000$; Simulation size is 1000. Summaries from (24) and (25) for 2 approaches — superscript $m = 1$: unweighted proxy; superscript $m = 2$: sequential. The results are shown for $m = 1/m = 2$ respectively. $\hat{\theta}^m_{\text{RMSE}} = \sqrt{\frac{1}{NG} \sum_{i=1}^N \sum_{g=1}^G |\hat{\theta}^m_{ig} - \theta_{ig}|^2}$, where θ_{ig} is the parameter of the copula linking the gth group-specific latent variable with the latent variable at the ith simulation.

From Table 9, the differences in Kendall's tau, and empirical dependency measures are decreasing as the dimension becomes large. The two proxy methods can perform well when $d_g \geq 20$ provided the within-group dependence is strong. From the results in local linking copulas, the unweighted average proxy approach has slightly smaller averaged differences in Kendall's tau while the sequential approach has smaller averaged differences in the empirical upper and lower dependence measures, as defined in (24). In addition, the conditional expectation proxies have smaller RMSEs and are closer to the true realized latent variables.

8. 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, as the number of variables increase, it is unlikely that factor models with conditional independence given latent variable continue to hold exactly. [14] 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 [10] 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 in this section 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 $\mathbf{\Omega}_D$ be the correlation matrix of $\epsilon_D = (\epsilon_1, \ldots, \epsilon_D)^T$, $\mathbf{\Gamma}_D = \mathbf{\Psi}_D \mathbf{\Omega}_D \mathbf{\Psi}_D$, and $\mathbf{\Psi}_D^2$ is a diagonal matrix with diagonal entries of $\mathbf{\Gamma}_D$. The factor scores are defined as $\widetilde{w}_D = (I + A_D^T \mathbf{\Psi}_D^{-2} A_D)^{-1} A_D^T \mathbf{\Psi}_D^{-2} z_D$.

Suppose the maximum eigenvalue of matrix $\mathbf{\Omega}_D$ is bounded as $D \to \infty$, the model is an approximate factor model from the definition in [6]. 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 [3].

Assumption 8.1. *Let $\mathbf{\Omega}_D = (\omega_{s,t})_{1 \leq s,t \leq D}$ be the correlation matrix of ϵ_D. Let $S_D = \sum_{j=1}^D \epsilon_j$ and $\bar{\epsilon}_D = S_D/D$, then* $\text{E}(S_D^2) = \sum_{s=1}^D \sum_{t=1}^D \omega_{s,t}$. *Assume*

$$0 < \lim_{D \to \infty} \inf \frac{\text{E}(S_D^2)}{D} < \lim_{D \to \infty} \sup \frac{\text{E}(S_D^2)}{D} < M$$

where M is a positive constant.

The above Assumption implies $\text{Var}(\bar{\epsilon}_D) = O(D^{-1})$, the same order as the case of iid. Under the Assumption 8.1 and the assumption on the loadings in Theorem 2, it is shown in B.2 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 5. The assumptions indicate the summation of entries in matrix residual correlation matrix $\mathbf{\Omega}_D$ is of $O(D)$. Let $\omega_{s+} = \sum_{t=1}^{D} \omega_{st}$. If ω_{s+} is $O(1)$ as $D \to \infty$ for all s, then Assumption 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 $k - j \geq 2$. If ω_{s+} is $O(D)$ as $D \to \infty$ for all s, then Assumption 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:

$$\sum_{j=1}^{D} \sum_{k=1}^{D} \mathrm{Cor}(C_{U_j|V}(U_j|V = v^0), C_{U_k|V}(U_k|V = v^0)) = O(D) \quad \forall\, 0 < v^0 < 1. \tag{26}$$

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

9. Discussion and further research

This paper proposes the conditional expectation proxies of the latent variables in some factor copula models and shows the consistency of proxy variables 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, [17] 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. [16] 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. Compared to their approaches, our inference method is more intuitive and does not need to fix a factor structure. The sequential procedures fit better with the use of Gaussian factor models as a started point to consider different factor structures that fit the data.

Our sequential proxy methods improve on the approach in [14] 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 at 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 [14] 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 in this paper 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 $p \geq 2$) 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 be needed for likelihood estimation and computations of proxies.

Appendices

A. Derivations: non-theorems

A.1. Bi-factor Gaussian model: Equivalence of two-stage factor scores defined in (10) *and* (11) *and regression factor scores*

Proof. Suppose there are G groups, and let $z_D = (z_1^T, z_2^T, \ldots, z_G^T)^T$, where $z_1, \ldots, z_G$ are the realization of observed variables $Z_1, \ldots, Z_G$. Let $\widehat{w}_0 = E(W_0 | Z_D = z_D)$. For proxies of local latent factors, let $\widehat{w}_1 = E(W_1 | Z_D = z_D)$. $\widetilde{w}_0, \widetilde{w}_1(\widetilde{w}_0)$ 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 $D = (d_1, d_2, \ldots, d_G)$, $d = \sum_{j=1}^{G} d_g$, $d_r = \sum_{j=2}^{G} d_2$, $z_D = (z_1^T, z_r^T)^T$. Then the loading matrix

$$A = [a_0, a_1, \ldots, a_p] = \begin{bmatrix} b_{01} & b_1 & 0 & \ldots & 0 \\ b_{02} & 0 & b_2 & \ldots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ b_{0G} & 0 & \ldots & 0 & b_G \end{bmatrix} = \begin{bmatrix} b_{01} & b_1 & 0 \\ b_{0r} & 0 & B_r \end{bmatrix}. \tag{27}$$

The partition of A leads to a 2×2 block matrix, where $b_{0r} = (b_{02}^T, \ldots, b_{0G}^T)^T$, $B_r = \text{diag}(b_2, \ldots, b_G)$. Also, partition $\Psi^2 = \text{diag}(\Psi_1^2, \ldots, \Psi_G^2) = \text{diag}(\Psi_1^2, \Psi_r^2)$ correspondingly.

Let $\Sigma_D = \text{Cor}(Z_D) = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} = \begin{bmatrix} b_{01}b_{01}^T + b_1b_1^T + \Psi_1^2 & b_{01}b_{0r}^T \\ b_{0r}b_{01}^T & b_{0r}b_{0r}^T + B_rB_r^T + \Psi_r^2 \end{bmatrix}$, $M = \Sigma_D^{-1} =: \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}$. Then the sizes of matrices $M_{11}, M_{12}, M_{21}, M_{22}$ 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 Σ_D have the same size. Let Σ_1 be the correlation matrix of (Z_1^T, W_0). Then $\Sigma_1 = \begin{bmatrix} b_{01}b_{01}^T + b_1b_1^T + \Psi_1^2 & b_{01} \\ b_{01}^T & 1 \end{bmatrix} =: \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$,

and $N = \Sigma_1^{-1} = \begin{bmatrix} N_{11} & N_{12} \\ N_{21} & N_{22} \end{bmatrix}$. 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 $z_D = (z_1^T, z_r^T)^T$, the regression factor scores defined in (9) are: $\widehat{w}_0 = a_0^T \Sigma_D^{-1} z_D = a_0^T M z_D$, $\widehat{w}_1 = a_1^T \Sigma_D^{-1} z_D = a_1^T M z_D$. Hence, by (27)

$$\widehat{w}_0 = (b_{01}^T, b_{0r}^T) \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} \begin{pmatrix} z_1 \\ z_r \end{pmatrix} = (b_{01}^T M_{11} + b_{0r}^T M_{21}) z_1 + (b_{01}^T M_{12} + b_{0r}^T M_{22}) z_r, \tag{28}$$

$$\widehat{w}_1 = (b_1^T, 0_r) \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} \begin{pmatrix} z_1 \\ z_r \end{pmatrix} = b_1^T M_{11} z_1 + b_1^T M_{12} z_r. \tag{29}$$

The expressions of $\widetilde{w}_0$ and $\widehat{w}_0$ equal $a_0^T M z_D$, so they are the same. After some algebraic calculations in (11), $\widetilde{w}_1 = b_1^T N_{11} z_1 + b_1^T N_{12} \widetilde{w}_0$. Substituting $\widehat{w}_0 = \widetilde{w}_0$ from (28) into $\widetilde{w}_1$ leads to

$$\widetilde{w}_1 = [b_1^T N_{11} + b_1^T N_{12}(b_{01}^T M_{11} + b_{0r}^T M_{21})] z_1 + b_1^T N_{12}(b_{01}^T M_{12} + b_{0r}^T M_{22}) z_r. \tag{30}$$

To conclude, it suffices to show that $\widetilde{w}_1$ in (30) and $\widehat{w}_1$ in (29) are equivalent, or that
(a) $N_{11} + N_{12}(b_{01}^T M_{11} + b_{0r}^T M_{21}) = M_{11}$, and (b) $N_{12}(b_{01}^T M_{12} + b_{0r}^T M_{22}) = M_{12}$.

Let $\Delta_1 = (b_1 b_1^T + \Psi_1^2)$, Δ_1 is positive definite and $\Delta_1 + b_{01}b_{01}^T = \Sigma_{11}$. Multiply Δ_1^{-1} on the left and Σ_{11}^{-1} on the right to get (c) $\Delta_1^{-1} b_{01} b_{01}^T \Sigma_{11}^{-1} - \Delta_1^{-1} = -\Sigma_{11}^{-1}$. From $\Sigma_D M = I$, we have (d) $M_{11} = \Sigma_{11}^{-1} - \Sigma_{11}^{-1} b_{01} b_{0r}^T M_{21}$ and (e) $M_{12} = -\Sigma_{11}^{-1} b_{01} b_{0r}^T M_{22}$. From $N\Sigma_1 = I$, we have (f) $N_{11}\Sigma_{11} + N_{12}b_{01}^T = N_{11}(\Delta_1 + b_{01}b_{01}^T) + N_{12}b_{01}^T = I$ and (g) $N_{11}b_{01} + N_{12} = 0$. In (g), multiply both sides by b_{01}^T to get (h) $N_{11}b_{01}b_{01}^T + N_{12}b_{01}^T = 0$. Then (f) and (h) together imply (i) $N_{11}\Delta_1 = I$. Hence, from (g) and (i), $N_{12} = -N_{11}b_{01} = -\Delta_1^{-1} b_{01}$, and from (f), $N_{11} = \Sigma_{11}^{-1} - N_{12}b_{01}^T \Sigma_{11}^{-1} = \Sigma_{11}^{-1} + \Delta_1^{-1} b_{01} b_{01}^T \Sigma_{11}^{-1}$. Substitute these expressions of N_{11} and N_{12} in the left-hand side of equation (a) to get:

$$N_{11} + N_{12}b_{01}^T M_{11} + N_{12}b_{0r}^T M_{21} = \Sigma_{11}^{-1} + \{\Delta_1^{-1} b_{01} b_{01}^T \Sigma_{11}^{-1} - \Delta_1^{-1} b_{01} b_{01}^T M_{11} - \Delta_1^{-1} b_{01} b_{0r}^T M_{21}\}.$$

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

$$\Delta_1^{-1}\boldsymbol{b}_{01}\boldsymbol{b}_{01}^T\Sigma_{11}^{-1} - \Delta_1^{-1}\boldsymbol{b}_{01}\boldsymbol{b}_{01}^T(\Sigma_{11}^{-1} - \Sigma_{11}^{-1}\boldsymbol{b}_{01}\boldsymbol{b}_{0r}^T\boldsymbol{M}_{21}) - \Delta_1^{-1}\boldsymbol{b}_{01}\boldsymbol{b}_{0r}^T\boldsymbol{M}_{21}$$
$$= \Delta_1^{-1}\boldsymbol{b}_{01}\boldsymbol{b}_{01}^T\Sigma_{11}^{-1}\boldsymbol{b}_{01}\boldsymbol{b}_{0r}^T\boldsymbol{M}_{21} - \Delta_1^{-1}\boldsymbol{b}_{01}\boldsymbol{b}_{0r}^T\boldsymbol{M}_{21}$$
$$= (\Delta_1^{-1}\boldsymbol{b}_{01}\boldsymbol{b}_{01}^T\Sigma_{11}^{-1} - \Delta_1^{-1})\boldsymbol{b}_{01}\boldsymbol{b}_{0r}^T\boldsymbol{M}_{21} = -\Sigma_{11}^{-1}\boldsymbol{b}_{01}\boldsymbol{b}_{0r}^T\boldsymbol{M}_{21}.$$

Thus, (a) is verified as $\boldsymbol{N}_{11} + \boldsymbol{N}_{12}\boldsymbol{b}_{01}^T\boldsymbol{M}_{11} + \boldsymbol{N}_{12}\boldsymbol{b}_{0r}^T\boldsymbol{M}_{21} = \Sigma_{11}^{-1} - \Sigma_{11}^{-1}\boldsymbol{b}_{01}\boldsymbol{b}_{0r}^T\boldsymbol{M}_{21} = \boldsymbol{M}_{11}$ 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

$$\boldsymbol{N}_{12}(\boldsymbol{b}_{01}^T\boldsymbol{M}_{12} + \boldsymbol{b}_{0r}^T\boldsymbol{M}_{22}) = \Delta_1^{-1}\boldsymbol{b}_{01}\boldsymbol{b}_{01}^T\Sigma_{11}^{-1}\boldsymbol{b}_{01}\boldsymbol{b}_{0r}^T\boldsymbol{M}_{22} - \Delta_1^{-1}\boldsymbol{b}_{01}\boldsymbol{b}_{0r}^T\boldsymbol{M}_{22}$$
$$= (\Delta_1^{-1}\boldsymbol{b}_{01}\boldsymbol{b}_{01}^T\Sigma_{11}^{-1} - \Delta_1^{-1})\boldsymbol{b}_{01}\boldsymbol{b}_{0r}^T\boldsymbol{M}_{22} = -\Sigma_{11}^{-1}\boldsymbol{b}_{01}\boldsymbol{b}_{0r}^T\boldsymbol{M}_{22} = \boldsymbol{M}_{12},$$

via (c) and then (e). $\qquad\qquad\square$

A.2. Proof for (17) in Table 2

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

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

$$\mathrm{Var}(W_g|\boldsymbol{Z}_D) = \mathrm{E}[\mathrm{Var}(W_g|\boldsymbol{Z}_D, W_0)|\boldsymbol{Z}_D] + \mathrm{Var}[\mathrm{E}(W_g|\boldsymbol{Z}_D, W_0)|\boldsymbol{Z}_D]$$
$$= \underbrace{[1 - \boldsymbol{b}_g^T(\boldsymbol{b}_g\boldsymbol{b}_g^T + \boldsymbol{\Psi}_g^2)^{-1}\boldsymbol{b}_g]}_{\text{term1}} + \underbrace{r^2\,\mathrm{Var}(W_0|\boldsymbol{Z}_D)}_{\text{term2}}, \qquad (31)$$

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

Since $\boldsymbol{Z}_g$ independent of Z's in other groups given W_0, $\mathrm{Var}(W_g|\boldsymbol{Z}_D, W_0) = \mathrm{Var}(W_g|\boldsymbol{Z}_g; W_0)$. The term1 in (31) follows because the joint distribution of $(\boldsymbol{Z}_g^T, W_g)^T$ given W_0 is multivariate normal with zero mean and covariance matrix $\begin{bmatrix} \boldsymbol{b}_g\boldsymbol{b}_g^T + \boldsymbol{\Psi}_g^2 & \boldsymbol{b}_g \\ \boldsymbol{b}_g^T & 1 \end{bmatrix}$. 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 $(1 + \boldsymbol{b}_g^T\boldsymbol{\Psi}_g^{-2}\boldsymbol{b}_g)^{-1} = (1 + q_g)^{-1}$. As for term2 in (31), in the expression of r, applying equation (8) as above, $r = -\tilde{q}_g(1 + q_g)^{-1}$. Combine the expression of two terms, the decomposition (17) is obtained.

B. Main Proofs in Sections 4 and 8

B.1. Proof of Theorem 2

Proof. In 1-factor model ($p = 1$), the loading matrix $\boldsymbol{A}_D$ is $D \times 1$, so we use notation $\boldsymbol{A}_D$ instead. 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 $q_D = \boldsymbol{A}_D^T\boldsymbol{\Psi}_D^{-2}\boldsymbol{A}_D$ (a positive real number). Since $D^{-1}\sum_{j=1}^D |\alpha_j| \le (D^{-1}\sum_{j=1}^D \alpha_j^2)^{1/2} \le (D^{-1}\sum_{j=1}^D |\alpha_j|)^{1/2}$ and $\lim_{D\to\infty} D^{-1}\sum_{j=1}^D |\alpha_j| \to \mathrm{const} \ne 0$, then $\bar{q}_D := D^{-1}q_D = D^{-1}\sum_{j=1}^D \alpha_j^2/\psi_j^2 \to q > 0$ (with the limit existing assuming sampling from a super-population). Since

$$\widetilde{w}_D = (1 + q_D)^{-1}\boldsymbol{A}_D^T\boldsymbol{\Psi}_D^{-2}(\boldsymbol{A}_Dw^0 + \boldsymbol{\Psi}_D\boldsymbol{e}_D)$$
$$= (1 + q_D)^{-1}\boldsymbol{A}_D^T\boldsymbol{\Psi}_D^{-1}\boldsymbol{e}_D + (1 + q_D^{-1})^{-1}w^0, \qquad (32)$$

then $\widetilde{w}_D - w^0 = (1 + q_D)^{-1} A_D^T \Psi_D^{-1} e_D + O(D^{-1})$. Next, $D^{-1/2} A_D^T \Psi_D^{-1} e_D$ is a realization of $D^{-1/2} A_D^T \Psi_D^{-1} \epsilon_D = D^{-1/2} \sum_{j=1}^D \alpha_j \epsilon_j / \psi_j$ which converges to $N(0, q)$ in distribution by the Continuity theorem, so it can be consider as $O_p(1)$. Hence,

$$(\widetilde{w}_D - w^0) = D^{-1/2}(D^{-1} + \bar{q}_D)^{-1}(D^{-1/2} A_D^T \Psi_D^{-1} e_D) + O(D^{-1})$$

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

B.2. Extension to weak residual dependence

An outline of the proof of consistency based on Assumption 8.1 is as follows.

Let $e_D = (e_1, \ldots, e_D)$ be one realization of ϵ_D. Let $q_D = A_D^T \Psi_D^{-2} A_D > 0$, and suppose $D^{-1} q_D \to q > 0$. Then, as in (32) $\widetilde{w}_D - w^0 = (1 + q_D)^{-1} A_D^T \Psi_D^{-1} e_D + (1 + q_D^{-1})^{-1} w^0 - w^0$. Note that $Y = D^{-1/2} A_D^T \Psi_D^{-1} e_D$ is a realization of $D^{-1/2} A_D^T \Psi_D^{-1} \epsilon_D = D^{-1/2} \sum_{j=1}^D \alpha_j \epsilon_j / \psi_j$ with variance $D^{-1} \sum_{j=1}^D \sum_{k=1}^D \alpha_j \alpha_k \omega_{jk} / [\psi_j \psi_k]$. By Assumption 8.1 and with loadings that are bounded away from ± 1, this variance is $O(1)$ so that Y can be considered as $O_p(1)$. Then $\widetilde{w}_D - w^0 = D^{-1/2}(D^{-1} + \bar{q}_D)^{-1}(D^{-1/2} A_D^T \Psi_D^{-1} \epsilon_D) + O(D^{-1})$ is asymptotically $O_p(D^{-1/2})$.

B.3. Proof of Theorem: 4

Proof. Using the technique in the proof of Theorem 2, let $Q_D = A_D^T \Psi_D^{-2} A_D$,

Then

$$\widetilde{w}_D - w^0 = (I_p + Q_D)^{-1} A_D^T \Psi_D^{-1} e_D + (I_p + Q_D^{-1})^{-1} w^0 - w^0$$

Since A_D is of full rank and the entries of Ψ_g for $g = 1, 2, \ldots, G$ are uniformly bounded away from 0. Then $\bar{Q}_D = D^{-1} A_D^T \Psi_D^{-2} A_D$ is positive definite for any fixed D, and $Q := \lim_{D \to \infty} \bar{Q}_D$ must be a semi positive definite matrix, with the limit existing assuming sampling from a super-population. Since $D^{-1} \|a_j\|_1 \nrightarrow 0$, then Q is a positive definite matrix. Since $\epsilon_D \sim N(0, I_D)$, then $D^{-1/2} A_D^T \Psi_D^{-1} \epsilon_D \to N(0, Q)$. Since $D^{-1/2} A_D^T \Psi_D^{-1} e_D$ is one realization, it can considered as $O_p(1)$. Hence,

$$\widetilde{w}_D - w^0 = D^{-1/2}(D^{-1} I_p + \bar{Q}_D)^{-1} D^{-1/2} A_D^T \Psi_D^{-1} \epsilon_D + O(D^{-1})$$

is asymptotically $O_p(D^{-1/2})$ by noticing that $(D^{-1} I_p + \bar{Q}_D)^{-1} \to Q^{-1}$. $\qquad\qquad\square$

B.4. Proof of consistency for proxies: Theorem 5, Theorem 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 Theorem 5 and Theorem 7 are similar. Both rely on the Laplace approximation for integrals (see [5]), and the asymptotic properties of maximum likelihood (ML) estimator for parameters. In our setting, the results in [4] 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 $\{f_{X_j}\}$.

The proof of Theorem 5 is given below.

Proof. In 1-factor copula model (1), there is a realized value v^0 for the latent variable. Then $(U_1, \ldots, U_D, \ldots)$ 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 $(u_1, \ldots, u_D)$, 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 [4]. Now apply the Laplace approximation. The numerator and denominator denoted as I_{1D} and I_{2D} in the expression of $\widetilde{v}_D$ in (9) can be approximated respectively by

$$I_{1D} = \int_0^1 v \exp\{-D \times g_D(v)\} dv = v_D^* \exp\{-D \times g(v_D^*)\} \sqrt{\frac{2\pi}{D|g''(v_D^*)|}} \{1 + O(D^{-1})\},$$

$$I_{2D} = \int_0^1 \exp\{-D \times g_D(v)\} dv = \exp\{-D \times g_D(v_D^*)\} \sqrt{\frac{2\pi}{D|g''(v_D^*)|}} \{1 + O(D^{-1})\}.$$

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

The proof of Theorem **7** for bi-factor copula model is given next.

Proof. There are realized value $v_0^0, v_1^0, \ldots, v_G^0$ for the latent variables. Then $(U_{1g}, \ldots, U_{d_g g}, \ldots)$ is an infinite sequence of dependent random variables for each $g = 1, \ldots, G$, 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 $C_{U_{jg}|V_g;V_0}(C_{U_{jg}|V_0}(\cdot|v_0^0), v_g^0)$ and its density is

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

If the values of $v^0, v_1^0, \ldots, v_G^0$ are to be estimated based on the realized d_g-vector $\boldsymbol{u}_{g,d_g} = (u_{1g}, \ldots, u_{d_g g})$ for $g \in \{1, \ldots, G\}$, then the integrated log-likelihood in v_0 is

$$L_0(v_0) = \log c_{\boldsymbol{U_D}, V_0}(\boldsymbol{u_D}, v_0),$$

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 $\Psi_{0,D}(v_0; \boldsymbol{u_D})$.

$$\Psi_{0,D}(v_0; \boldsymbol{u_D}) := \partial \log c_{\boldsymbol{U_D}, V_0}(\boldsymbol{u_D}, v_0)/\partial v_0 = \Psi_{01,D}(v_0; \boldsymbol{u_D}) + \Psi_{02,D}(v_0; \boldsymbol{u_D})$$

$$= \sum_{g=1}^{G} \sum_{j=1}^{d_g} \partial \log c_{U_{jg}, V_0}(u_{jg}, v_0)/\partial v_0 + \sum_{g=1}^{G} \partial \log f_g(\boldsymbol{u}_g; v_0)/\partial v_0.$$

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,

$$\frac{\partial \log f_g(\boldsymbol{u}_g; v_0)}{\partial v_0} = \frac{1}{f_g(\boldsymbol{u}_g; v_0)} \int_0^1 \Big(\partial \exp\Big\{\sum_{j=1}^{d_g} \log c_{U_{jg}, V_g; V_0}(C_{U_{jg}|V_0}(u_{jg}|v_0), v_g)\Big\} \Big/ \partial v_0\Big) dv_g$$

$$= \sum_{j=1}^{d_g} \int_0^1 \frac{c_{U_g, V_g; V_0}(\boldsymbol{u}_g, v_g; v_0)}{f_g(\boldsymbol{u}_g; v_0)} \frac{\partial \log c_{U_{jg} V_g; V_0}(C_{U_{jg}|V_0}(u_{jg}|v_0), v_g)}{\partial v_0} dv_g.$$

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 [7], under regularity conditions of the log-likelihood and assuming sampling from a super-population,

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

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

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):

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

The partial derivative of L_g 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 $v_{0,\boldsymbol{D}}^*$, 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 weak laws of large numbers, and the super-population assumption,

$$\lim_{d_g \to \infty} \bar{\Psi}_{g,D}(v_g; \boldsymbol{u}_{g,d_g}, v_{0,\boldsymbol{D}}^*) = \lim_{d_g \to \infty} d_g^{-1} \sum_{j=1}^{d_g} \frac{\partial \log c_{U_{jg}|V_g; V_0}\left(C_{U_{jg}|V_0}(u_{jg}|v_{0,\boldsymbol{D}}^*)\Big|v_g\right)}{\partial v_g}$$

exists. 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),

$$\widetilde{v}_{0\boldsymbol{D}}(\boldsymbol{u}_{\boldsymbol{D}}) = \frac{\int_0^1 v_0 \exp\{-D \cdot \bar{L}_{0\boldsymbol{D}}(v_0; \boldsymbol{u}_{\boldsymbol{D}})\}dv_0}{\int_0^1 \exp\{-D \cdot \bar{L}_{0\boldsymbol{D}}(v_0; \boldsymbol{u}_{\boldsymbol{D}})\}dv_0}. \tag{33}$$

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 (33) can be approximated by

$$v_{0,\boldsymbol{D}}^* \exp\{-D \times \bar{L}_{0\boldsymbol{D}}(v_{0,\boldsymbol{D}}^*)\} \sqrt{\frac{2\pi}{D\left|[\partial\bar{\Psi}_{0,D}/\partial v_0]|_{v_{0,\boldsymbol{D}}^*}\right|}} + O(D^{-1}),$$

$$\exp\{-D \times \bar{L}_{0\boldsymbol{D}}(v_{0,\boldsymbol{D}}^*)\} \sqrt{\frac{2\pi}{D\left|[\partial\bar{\Psi}_{0,D}/\partial v_0]|_{v_{0,\boldsymbol{D}}^*}\right|}} + O(D^{-1})$$

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 $g = 1, \ldots, G$ are consistent. $\square$

B.5. Proof of Theorem 8 (Lipschitz continuity of factor scores in Gaussian factor model)

Proof. 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

$$(\boldsymbol{I}_p + \widehat{\boldsymbol{Q}}_{\boldsymbol{D}})^{-1}\widehat{\boldsymbol{A}}_{\boldsymbol{D}}^T\widehat{\boldsymbol{\Psi}}_{\boldsymbol{D}}^{-2}\boldsymbol{z}_{\boldsymbol{D}} - (\boldsymbol{I}_p + \boldsymbol{Q}_{\boldsymbol{D}})^{-1}\boldsymbol{A}_{\boldsymbol{D}}^T\boldsymbol{\Psi}_{\boldsymbol{D}}^{-2}\boldsymbol{z}_{\boldsymbol{D}} \tag{34}$$

$$= \underbrace{(\boldsymbol{I}_p + \widehat{\boldsymbol{Q}}_{\boldsymbol{D}})^{-1}\widehat{\boldsymbol{A}}_{\boldsymbol{D}}^T\widehat{\boldsymbol{\Psi}}_{\boldsymbol{D}}^{-2}\boldsymbol{z}_{\boldsymbol{D}} - (\boldsymbol{I}_p + \boldsymbol{Q}_{\boldsymbol{D}})^{-1}\widehat{\boldsymbol{A}}_{\boldsymbol{D}}^T\widehat{\boldsymbol{\Psi}}_{\boldsymbol{D}}^{-2}\boldsymbol{z}_{\boldsymbol{D}}}_{\text{term1}}$$

$$+ \underbrace{(\boldsymbol{I}_p + \boldsymbol{Q}_{\boldsymbol{D}})^{-1}\widehat{\boldsymbol{A}}_{\boldsymbol{D}}^T\widehat{\boldsymbol{\Psi}}_{\boldsymbol{D}}^{-2}\boldsymbol{z}_{\boldsymbol{D}} - (\boldsymbol{I}_p + \boldsymbol{Q}_{\boldsymbol{D}})^{-1}\boldsymbol{A}_{\boldsymbol{D}}^T\boldsymbol{\Psi}_{\boldsymbol{D}}^{-2}\boldsymbol{z}_{\boldsymbol{D}}}_{\text{term2}}.$$

Let $\bar{\boldsymbol{Q}}_{\boldsymbol{D}} = D^{-1}\boldsymbol{Q}_{\boldsymbol{D}}, \bar{\widehat{\boldsymbol{Q}}}_{\boldsymbol{D}} = D^{-1}\widehat{\boldsymbol{Q}}_{\boldsymbol{D}}$ and recall we assume $\bar{\boldsymbol{Q}}_{\boldsymbol{D}} \to \boldsymbol{Q}, \bar{\widehat{\boldsymbol{Q}}}_{\boldsymbol{D}} \to \widehat{\boldsymbol{Q}}$ as $\boldsymbol{D} \to \infty$ where $\boldsymbol{Q}, \widehat{\boldsymbol{Q}}$ are positive definite matrix. Note $(\boldsymbol{I}_p + \boldsymbol{Q}_{\boldsymbol{D}})^{-1} = O(D^{-1}), (\boldsymbol{I}_p + \widehat{\boldsymbol{Q}}_{\boldsymbol{D}})^{-1} = O(D^{-1})$. Let $\boldsymbol{H}_{\boldsymbol{D}} = D \cdot (\boldsymbol{I}_p + \boldsymbol{Q}_{\boldsymbol{D}})^{-1}, \boldsymbol{H}_{\boldsymbol{D}} = O(1)$. 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, term 1 in (34) has the order of $(\bar{\widehat{\boldsymbol{Q}}}_{\boldsymbol{D}}^{-1} - \bar{\boldsymbol{Q}}_{\boldsymbol{D}}^{-1}) \cdot D^{-1} \cdot \widehat{\boldsymbol{A}}_{\boldsymbol{D}}^T\widehat{\boldsymbol{\Psi}}_{\boldsymbol{D}}^{-2}\boldsymbol{z}_{\boldsymbol{D}}$.

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

$$\bar{\widehat{\boldsymbol{Q}}} - \bar{\boldsymbol{Q}} = D^{-1}(\widehat{\boldsymbol{A}}^T\widehat{\boldsymbol{\Psi}}^{-2}\widehat{\boldsymbol{A}} - \boldsymbol{A}^T\boldsymbol{\Psi}^{-2}\boldsymbol{A})$$

$$= D^{-1}\left((\widehat{\boldsymbol{A}}^T - \boldsymbol{A}^T + \boldsymbol{A}^T)\widehat{\boldsymbol{\Psi}}^{-2}(\widehat{\boldsymbol{A}} - \boldsymbol{A} + \boldsymbol{A}) - \boldsymbol{A}^T\boldsymbol{\Psi}^{-2}\boldsymbol{A}\right)$$

$$= D^{-1}\left(\underbrace{\boldsymbol{A}^T\widehat{\boldsymbol{\Psi}}^{-2}(\widehat{\boldsymbol{A}} - \boldsymbol{A})}_{\text{term1}} + \underbrace{(\widehat{\boldsymbol{A}}^T - \boldsymbol{A}^T)\widehat{\boldsymbol{\Psi}}^{-2}\boldsymbol{A}}_{\text{term2}} + \underbrace{(\widehat{\boldsymbol{A}}^T - \boldsymbol{A}^T)\widehat{\boldsymbol{\Psi}}^{-2}(\widehat{\boldsymbol{A}} - \boldsymbol{A})}_{\text{term3}} + \underbrace{\boldsymbol{A}^T(\widehat{\boldsymbol{\Psi}}^{-2} - \boldsymbol{\Psi}^{-2})\boldsymbol{A}}_{\text{term4}}\right). \tag{35}$$

Since term3 is negligible compared to other terms, we only look at the order of the other three terms in the right-hand side of (35). For term1, term2, term4, multiplied by D^{-1}, the Cauchy-Schwartz inequality leads to:

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

$$\|D^{-1}\mathbf{A}^T(\widehat{\mathbf{\Psi}}^{-2} - \mathbf{\Psi}^{-2})\mathbf{A}\| = \left\|D^{-1}\sum_{j=1}^{D}\mathbf{a}_j\mathbf{a}_j^T(\widehat{\psi}_j^{-2} - \psi_j^{-2})\right\| \le \left(D^{-1}\sum_{j=1}^{D}\|\mathbf{a}_j\|^4\right)^{1/2} \cdot \left(D^{-1}\sum_{j=1}^{D}(\widehat{\psi}_j^{-2} - \psi_j^{-2})^2\right)^{1/2}.$$

For all j, $\|\mathbf{a}_j\| < 1$, ψ_j is bounded from zero, term1 and term2 in (35) multiplied by D^{-1} are all $O\left(\sqrt{D^{-1}\sum_{j=1}^{D}\|\hat{\mathbf{a}}_j - \mathbf{a}_j\|^2/\widehat{\psi}_j^2}\right)$. The term4 in (35) multiplied by D^{-1} is $O\left(\sqrt{D^{-1}\sum_{j=1}^{D}(\widehat{\psi}_j^{-2} - \psi_j^{-2})^2}\right)$. Also note

$$\|D^{-1}\widehat{\mathbf{A}}_D^T\widehat{\mathbf{\Psi}}_D^{-2}\mathbf{z}_D\| = \left\|D^{-1}\sum_{j=1}^{D}\frac{\hat{\mathbf{a}}_j z_j}{\widehat{\psi}_j^2}\right\| \le \sqrt{D^{-1}\sum_{j=1}^{D}\frac{\|\hat{\mathbf{a}}_j\|^2}{\widehat{\psi}_j^4}}\sqrt{D^{-1}\sum_{j=1}^{D}p \cdot z_j^2}. \tag{36}$$

Since the bound of (36) is $O_p(1)$, then term 1 in (34) is order $O_p\left(\sqrt{D^{-1}\sum_{j=1}^{D}\|\hat{\mathbf{a}}_j - \mathbf{a}_j\|^2/\widehat{\psi}_j^2}\right) + O_p\left(\sqrt{D^{-1}\sum_{j=1}^{D}(\widehat{\psi}_j^{-2} - \psi_j^{-2})^2}\right)$.

For term 2 in (34),

$$\mathbf{H}_D \cdot D^{-1} \cdot \left((\mathbf{A}_D^T + \widehat{\mathbf{A}}_D^T - \mathbf{A}_D^T)(\mathbf{\Psi}_D^{-2} + \widehat{\mathbf{\Psi}}_D^{-2} - \mathbf{\Psi}_D^{-2})\mathbf{z}_D - \mathbf{A}_D^T\mathbf{\Psi}_D^{-2}\mathbf{z}_D\right) \tag{37}$$

$$= \mathbf{H}_D \cdot D^{-1} \cdot \left(\underbrace{\mathbf{A}_D^T(\widehat{\mathbf{\Psi}}_D^{-2} - \mathbf{\Psi}_D^{-2})\mathbf{z}_D}_{\text{term1}} + \underbrace{(\widehat{\mathbf{A}}_D^T - \mathbf{A}_D^T)\mathbf{\Psi}_D^{-2}\mathbf{z}_D}_{\text{term2}} + \underbrace{(\widehat{\mathbf{A}}_D^T - \mathbf{A}_D^T)(\widehat{\mathbf{\Psi}}_D^{-2} - \mathbf{\Psi}_D^{-2})\mathbf{z}_D}_{\text{term3}}\right).$$

Since term3 in (37) is negligible in comparison, we only look at the first two terms. From the Cauchy-Schwartz inequality, for term1 and term2 in (37) multiplied by D^{-1},

$$D^{-1} \cdot \text{term1} = O_p\left(\sqrt{D^{-1}\sum_{j=1}^{D}(\hat{\psi}_j^{-2} - \psi_j^{-2})^2}\right), \quad D^{-1} \cdot \text{term2} = O_p\left(\sqrt{D^{-1}\sum_{j=1}^{D}\|\hat{\mathbf{a}}_j - \mathbf{a}_j\|^2}\right).$$

Recall $\mathbf{H}_D = O(1)$, then the term2 in (34) has the same order as term1 in (34). Also, due to ψ_j being bounded away from 0 and $\psi_j^2 = 1 - \|\mathbf{a}_j\|^2$, then $\|\widetilde{w}_D(\widehat{\mathbf{A}}_D) - \widetilde{w}_D(\mathbf{A}_D)\| = O_p\left(\sqrt{D^{-1}\sum_{j=1}^{D}\|\hat{\mathbf{a}}_j - \mathbf{a}_j\|^2}\right)$. $\qquad\square$

B.6. Proof of Lemma 9

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

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

$$= \left\|\int_0^1 \nabla\widetilde{v}(\gamma(t)) \cdot (\hat{\boldsymbol{\theta}}_D - \boldsymbol{\theta}_D)\,dt\right\| \le \|\hat{\boldsymbol{\theta}}_D - \boldsymbol{\theta}_D\|\int_0^1 \|\nabla\widetilde{v}(\gamma(t))\|\,dt < K_D\|\hat{\boldsymbol{\theta}}_D - \boldsymbol{\theta}_D\|,$$

where $K_D := \sup\{\|\nabla\widetilde{v}(\boldsymbol{\theta})\| : \boldsymbol{\theta} \in \bar{\mathbf{B}}(\boldsymbol{\theta}_D, \rho)\}$, and the norm are all l_2 norms. Next we derive order of the Lipschitz constant.

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

$$
\frac{\partial \widetilde{v}(\boldsymbol{\theta}_D)}{\partial \boldsymbol{\theta}_j} = \frac{\int_0^1 (v \partial f_D(v, \boldsymbol{\theta}_D)/\partial \boldsymbol{\theta}_j) \mathrm{d}v \cdot \int_0^1 f_D(v, \boldsymbol{\theta}_D) \mathrm{d}v - \int_0^1 (\partial f_D(v, \boldsymbol{\theta}_D)/\partial \boldsymbol{\theta}_j) \mathrm{d}v \cdot \int_0^1 v f_D(v, \boldsymbol{\theta}_D) \mathrm{d}v}{(\int_0^1 f_D(v, \boldsymbol{\theta}_D) \mathrm{d}v)^2}
$$

$$
= \left\{ \left(\int_0^1 f_D(v, \boldsymbol{\theta}_D) \mathrm{d}v \right)^{-1} \left[\int_0^1 (v \partial f_D(v, \boldsymbol{\theta}_D)/\partial \boldsymbol{\theta}_j) \mathrm{d}v - \widetilde{v}_D(\boldsymbol{\theta}_D) \times \int_0^1 (\partial f_D(v, \boldsymbol{\theta}_D)/\partial \boldsymbol{\theta}_j) \mathrm{d}v \right] \right\} \qquad (38)
$$

$$
= \left(\int_0^1 f_D(v, \boldsymbol{\theta}_D) \mathrm{d}v \right)^{-1} \left(\int_0^1 [v - \widetilde{v}_D(\boldsymbol{\theta}_D)] \times (\partial f_D(v, \boldsymbol{\theta}_D)/\partial \boldsymbol{\theta}_j) \mathrm{d}v \right).
$$

In (38), $\widetilde{v}_D = \widetilde{v}_D(\boldsymbol{\theta}_D)$ 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

$$
\frac{\int_0^1 (v - \widetilde{v}_D) \cdot |(\partial \log c_{jV}(u_j, v; \boldsymbol{\theta}_j)/\partial \boldsymbol{\theta}_j)| f_D(v, \boldsymbol{\theta}_D) \mathrm{d}v}{\int_0^1 f_D(v, \boldsymbol{\theta}_D) \mathrm{d}v}. \qquad (39)
$$

In (39), let $m_j(v) = \partial \log c_{jV}(u_j, v; \boldsymbol{\theta}_j)/\partial \boldsymbol{\theta}_j$, $h(v) = \bar{L}_D(v) = -D^{-1} \log f_D(v, \boldsymbol{\theta}_D)$. Let $v_D^* = \arg\min h(v)$. Let $t(v) = (v - \widetilde{v}_D) |m_j(v)|$, $t'(v) = |m_j(v)| + (v - \widetilde{v}_D)(\partial |m_j(v)|/\partial v)$ and $t''(v) = 2(\partial |m_j(v)|/\partial v) + (v - \widetilde{v}_D)(\partial^2 |m_j(v)|/\partial v^2)$. From equation (2.6) in [18], equation (39) becomes

$$
(v_D^* - \widetilde{v}_D) |m_j(v_D^*)| + (2D)^{-1} [h''(v_D^*)]^{-1} t''(v_D^*) - (2D)^{-1} [h''(v_D^*)]^{-2} t'(v_D^*) h'''(v_D^*) + O(D^{-2}).
$$

Under the assumptions on the bounded derivatives, together with the proof in Section B.4, $\partial \widetilde{v}(\boldsymbol{\theta}_D)/\partial \boldsymbol{\theta}_j = O(D^{-1})$. Then, the norm of derivatives $\|\nabla \widetilde{v}(\boldsymbol{\theta})\|$ equals to $\sqrt{\sum_{j=1}^D |\partial \widetilde{v}(\boldsymbol{\theta}_D)/\partial \boldsymbol{\theta}_j|^2} = O(D^{-1/2})$ and $|\widetilde{v}(\hat{\boldsymbol{\theta}}_D) - \widetilde{v}(\boldsymbol{\theta}_D)| = O\left(\sqrt{D^{-1} \sum_{j=1}^D \|\hat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j\|_2^2} \right)$.
□

B.7. Proof of Lemma 10

Proof. The proof technique is similar to that used in the 1-factor copula case. Let $f_D^{(1)}(v_0, \boldsymbol{\theta}_D) = c_{U_D, V_0}(\boldsymbol{u}_D, v_0; \boldsymbol{\theta}_D)$ be the marginal density function defined in Table 1. Then the components of the gradient of the global proxy $\widetilde{v}_{0D}$ with respect to the parameter vector $\boldsymbol{\theta}_D^{(1)}$ consists of

$$
\partial \widetilde{v}_{0,D}(\boldsymbol{\theta}_D)/\partial \theta_{jg,0} = \frac{\int_0^1 (v_0 \partial f_D^{(1)}(v_0, \boldsymbol{\theta}_D)/\partial \theta_{jg,0}) \mathrm{d}v_0 \cdot \int_0^1 f_D^{(1)}(v_0, \boldsymbol{\theta}_D) \mathrm{d}v_0}{(\int_0^1 f_D^{(1)}(v_0, \boldsymbol{\theta}_D) \mathrm{d}v_0)^2}
$$

$$
- \frac{\int_0^1 (\partial f_D^{(1)}(v_0, \boldsymbol{\theta}_D)/\partial \theta_{jg,0}) \mathrm{d}v_0 \int_0^1 v_0 f_D^{(1)}(v_0, \boldsymbol{\theta}_D) \mathrm{d}v_0}{(\int_0^1 f_D^{(1)}(v_0, \boldsymbol{\theta}_D) \mathrm{d}v_0)^2}
$$

$$
= \left(\int_0^1 f_D^{(1)}(v_0, \boldsymbol{\theta}_D) \mathrm{d}v_0 \right)^{-1} \left(\int_0^1 (v_0 - \widetilde{v}_{0,D}(\boldsymbol{\theta}_D)) \times (\partial f_D^{(1)}(v_0, \boldsymbol{\theta}_D)/\partial \theta_{jg,0}) \mathrm{d}v_0 \right). \qquad (40)
$$

As $\partial f_D^{(1)}(v_0, \boldsymbol{\theta}_D)/\partial \theta_{jg,0} = f_D^{(1)}(v_0, \boldsymbol{\theta}_D) \cdot \left(\partial \log c_{U, V_0}(\boldsymbol{u}_D, v_0, \boldsymbol{\theta}_D)/\partial \theta_{jg,0} \right)$, then (40) has the same order with

$$
\frac{\int_0^1 (v_0 - \widetilde{v}_{0,D}(\boldsymbol{\theta}_D)) \cdot |\partial \log c_{U, V_0}(\boldsymbol{u}_D, v_0, \boldsymbol{\theta}_D)/\partial \theta_{jg,0}| \cdot f_D^{(1)}(v_0, \boldsymbol{\theta}_D) \mathrm{d}v_0}{\int_0^1 f_D^{(1)}(v_0, \boldsymbol{\theta}_D) \mathrm{d}v_0}. \qquad (41)
$$

Let $m_{jg}(v_0) = \partial \log c_{U,V_0}(\boldsymbol{u_D}, v_0, \boldsymbol{\theta_D})/\partial\theta_{jg,0}$, $h(v_0) = \bar{L}_{0D}(v_0) = -D^{-1}\log f_D^{(1)}(v_0, \boldsymbol{\theta_D})$. Let $v_{0,D}^* = \arg\min h(v_0)$. With $\widehat{v}_{0,D} = \widetilde{v}_{0,D}(\boldsymbol{\theta_D})$, let $t(v_0) = (v_0 - \widetilde{v}_{0,D})|m_{jg}(v_0)|$, $t'(v_0) = (v_0 - \widetilde{v}_{0,D})(\partial|m_{jg}(v_0)|/\partial v_0) + |m_{jg}(v_0)|)$, $t''(v_0) = 2(\partial|m_{jg}(v_0)|/\partial v_0) + (v_0 - \widetilde{v}_{0,D})(\partial^2|m_{jg}(v_0)|/\partial v_0^2)$.

Then from equation (2.6) in [18], equation (41) becomes

$$(v_{0,D}^* - \widetilde{v}_{0,D})|m_{jg}(v_{0,D}^*)| + (2D)^{-1}[h''(v_{0,D}^*)]^{-1}t''(v_{0,D}^*) - (2D)^{-1}[h''(v_{0,D}^*)]^{-2}t'(v_{0,D}^*)h'''(v_{0,D}^*) + O(D^{-2}).$$

Under the assumptions on the bounded partial derivatives, together with the proof in Section B.4, $\partial\widetilde{v}_{0,D}(\boldsymbol{\theta_D})/\partial\theta_{jg,0} = O(D^{-1})$.

The same logic could be also applies to $\partial\widetilde{v}_{0D}(\boldsymbol{\theta_D})/\partial\theta_{jg}$, $\partial\widetilde{v}_{g,D}(\boldsymbol{\theta_D})/\partial\theta_{jg,0}$ and $\partial\widetilde{v}_{g,D}(\boldsymbol{\theta_D})/\partial\theta_{jg}$. From the above derivation, we conclude that

$$\|\widetilde{v}_{0D}(\hat{\boldsymbol{\theta}}_D) - \widetilde{v}_{0D}(\boldsymbol{\theta_D})\| = O\left(\sqrt{(2D)^{-1}\sum_{g=1}^{G}\sum_{j=1}^{d_g}\left\{\|\hat{\boldsymbol{\theta}}_{jg,0} - \boldsymbol{\theta}_{jg,0}\|_2^2 + \|\hat{\boldsymbol{\theta}}_{jg} - \boldsymbol{\theta}_{jg}\|_2^2\right\}}\right)$$

and

$$\|\widetilde{v}_{gD}(\hat{\boldsymbol{\theta}}_D) - \widetilde{v}_{gD}(\boldsymbol{\theta_D})\| = O\left(\sqrt{(2d_g)^{-1}\sum_{j=1}^{d_g}\left\{\|\hat{\boldsymbol{\theta}}_{jg,0} - \boldsymbol{\theta}_{jg,0}\|_2^2 + \|\hat{\boldsymbol{\theta}}_{jg} - \boldsymbol{\theta}_{jg}\|_2^2\right\}}\right), \quad g \in \{1, \ldots, G\}.$$

$\square$

References

[1] T. W. Anderson, Y. Amemiya, The asymptotic normal distribution of estimators in factor analysis under general conditions, Annals of Statistics 16 (1988) 759–771.

[2] J. Bai, K. Li, Statistical analysis of factor models of high dimension, Annals of Statistics 40 (2012) 436–465.

[3] J. Bai, K. Li, Maximum likelihood estimation and inference for approximate factor models of high dimension, Review of Economics and Statistics 98 (2016) 298–309.

[4] R. A. Bradley, J. J. Gart, The asymptotic properties of ML estimators when sampling from associated populations, Biometrika 49 (1962) 205–214.

[5] K. W. Breitung, Asymptotic Approximations for Probability Integrals, Springer, Berlin, 1994.

[6] G. Chamberlain, M. Rothschild, Arbitrage, factor structure, and mean-variance analysis on large asset markets, Econometrica (1983) 1281–1304.

[7] H. Cramér, Problems in probability theory, The Annals of Mathematical Statistics 18 (1947) 165–193.

[8] R. L. Gorsuch, Factor Analysis, Erlbaum Associates, Hillsdale, New Jersey, 1983.

[9] H. Joe, Dependence Modeling with Copulas, Chapman & Hall/CRC, Boca Raton, FL, 2014.

[10] H. Joe, Parsimonious graphical dependence models constructed from vines, Canadian Journal of Statistics 46 (2018) 532–555.

[11] R. A. Johnson, D. W. Wichern, Applied Multivariate Statistical Analysis, Prentice Hall, Englewood Cliffs, 5th edition, 2002.

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

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

[14] P. Krupskii, H. Joe, Approximate likelihood with proxy variables for parameter estimation in high-dimensional factor copula models, Statistical Papers 63 (2022) 543–569.

[15] D. Lee, H. Joe, P. Krupskii, Tail-weighted dependence measures with limit being the tail dependence coefficient, Journal of Nonparametric Statistics 30 (2018) 262–290.

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

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

[18] L. Tierney, R. E. Kass, J. B. Kadane, Fully exponential Laplace approximations to expectations and variances of nonpositive functions, Journal of the American Statistical Association 84 (1989) 710–716.